Structural and biochemical insight into the interactions of Cdc42 with TOCA1 and N-WASP

Joanna Rose Watson

Corpus Christi College
University of Cambridge

June 2017

This dissertation is submitted for the degree of Doctor of Philosophy.
Summary

Cdc42 is a member of the Rho family of small GTPases, which, together with its homologues RhoA and Rac1, controls a multitude of cellular functions via the actin cytoskeleton. Cdc42 exerts its effects on the cytoskeleton via effector proteins of the Wiskott-Aldrich Syndrome (WASP) family and the Transducer of Cdc42-dependent Actin assembly (TOCA) family. The WASP family and their activation by Cdc42 have been thoroughly studied in vitro and are well understood. Conversely, understanding of the TOCA family remains limited by a lack of biochemical, biophysical and structural insight.

An investigation of the TOCA1-Cdc42 interaction is described here, revealing a relatively low affinity interaction with a dissociation constant in the micromolar range. This is 10-100x weaker than other Rho-effector interactions and suggests that TOCA1 must first be co-localised with Cdc42 to achieve stable binding in vivo. The solution NMR structure of the Cdc42 binding HR1 domain of TOCA1 provides the first structural data on this protein and reveals some interesting structural features that may relate to binding affinity and specificity. A structural model of the Cdc42-HR1 complex provides further insight into differential specificities and affinities of GTPase-effector interactions. NMR and actin polymerisation assays provide insight into the pathway of Cdc42/TOCA1/WASP-dependent actin assembly, suggesting unidirectional displacement of TOCA1 by N-WASP. A comparison of the Cdc42-TOCA1 model with an NMR structure of Cdc42 in complex with the GTPase binding domain of WASP reveals a possible mechanism by which an ‘effector handover’ from TOCA1 to N-WASP could take place.

Small GTPases such as Cdc42 are lipid modified and membrane anchored via their C-termini in vivo, so in vitro studies using truncated, unmodified GTPases are limited in their biological interpretation. This project also aimed to develop methods to study full length and membrane-anchored GTPases in vitro. Lipid modified protein was produced, which showed a weak affinity for liposomes, and so structural studies of membrane anchored protein are within reach. Further method development is now required to achieve stable membrane anchoring of lipid modified GTPases for detailed NMR studies.
Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

It does not exceed the prescribed word limit for the Degree Committee of the Faculty of Biology.

Findings from this work have been published as follows:


Acknowledgements

To begin, I would like to thank my supervisors, Helen Mott and Darerca Owen, who have provided invaluable training, support, advice and mentorship throughout the past five years. Together they have helped me to develop as a researcher and ultimately made this thesis possible. I extend particular thanks to Helen, who has guided and inspired me to pursue a career in Structural Biology.

Secondly, I would like to thank those who have contributed financially to the completion of this project. Herchel Smith have funded me for four years and three months and the Cambridge Philosophical Society provided funding for a further three months. I must also thank Corpus Christi College, who enabled me to survive through undergraduate and postgraduate studies by providing financial and pastoral support.

I also owe gratitude to everyone else who has provided support, training and guidance over the past four and a half years. Jenny Gallop offered specialist insight into the pathways of actin dynamics and also provided training, materials and access to the equipment needed for the liposome and pyrene actin experiments. Jenny’s input into this project also helped me to achieve my first publication. Katherine Stott provided training in biophysical techniques and Daniel Nietlispach was instrumental in teaching me NMR theory. All of the NMR pulse sequences were also provided by Daniel. Astrid Walrant of the Gurdon Institute, UK, helped with pyrene actin assays and Helen Fox, also of the Gurdon Institute, provided some of the TOCA1 proteins. The Protein and Nucleic Acid Centre (PNAC) at the Department of Biochemistry carried out all of the mass spectrometry and amino acid analysis.

My thanks further extend to all who made up the substantial support network I have benefitted from throughout my PhD. These include all of the members of the Mott/Owen, Thomas and Nietlispach groups at the Department of Biochemistry. In particular I would like to thank Natasha Clayton for emotional support and encouragement and Mark Bostock for his gentle guidance through the minefield of NMR theory. In addition, I thank all of my family who have played a part in my earlier education that led me to undertake this PhD. In particular I thank my Dad who gave up his career and dedicated his life to ensuring I had the best possible chance at a stable childhood and the best opportunities for education and development. I thank also my church family at St Andrew the Great church in Cambridge, including my close friend and pillar of support, Vic Richie. She has constantly grounded me and served to put my momentary struggles into perspective over the past two years. Finally, my heartfelt thanks go to my husband and friend, Ben Watson, who has picked me up again and again, without complaint.
# Table of Contents

**Abbreviations** ................................................................................................. i

1 **Introduction** .................................................................................................. 1
  1.1 The Ras Superfamily ......................................................................................... 1
  1.1.1 The G Domain ................................................................................................. 2
  1.1.2 The GTPase Switch ......................................................................................... 6
  1.1.3 Family Specific Variations on the G Domain .................................................. 8
  1.1.4 Guanine Nucleotide Exchange Factors .......................................................... 10
  1.1.5 GTP Hydrolysis and GTPase Activating Proteins ............................................ 12
  1.1.6 Post Translational Modifications and Membrane Localisation ....................... 14
  1.1.7 Guanine Nucleotide Dissociation Inhibitors ................................................... 16
  1.1.8 Effector Binding ............................................................................................. 16
  1.2 The Rho family ................................................................................................. 19
    1.2.1 The Diverse Roles of a Small Family ............................................................ 19
    1.2.2 Cdc42 and the Actin Cytoskeleton ................................................................ 20
  1.3 WASP and N-WASP ......................................................................................... 21
    1.3.1 The G protein Binding Domain ...................................................................... 21
    1.3.2 Activation of WASP and N-WASP ............................................................... 24
  1.4 TOCA1 ............................................................................................................. 28
    1.4.1 Functions of TOCA1 ..................................................................................... 29
    1.4.2 Structural Features and Functional Implications ............................................ 30
      1.4.2.1 The F-BAR Domain ................................................................................. 30
      1.4.2.2 The SH3 Domain ..................................................................................... 31
      1.4.2.3 The HR1 Domain ................................................................................... 32
    1.4.3 Activation of N-WASP by TOCA1 ............................................................... 33
  1.5 Open questions ............................................................................................... 35
    1.5.1 Specificity and Selectivity of GTPase-Effector Interactions ......................... 35
    1.5.2 The Cdc42-TOCA1 Interaction ..................................................................... 36
    1.5.3 Roles and Regulation of TOCA1 ................................................................. 37
      1.5.3.1 Autoregulation of TOCA1 ..................................................................... 37
      1.5.3.2 Cdc42, TOCA1 and N-WASP ............................................................... 39
    1.5.4 The Role of the Membrane in GTPase Signalling ......................................... 40
  1.6 Aims .................................................................................................................. 42
  1.7 Experimental Techniques ............................................................................... 43
    1.7.1 An Introduction to NMR Spectroscopy ......................................................... 43
      1.7.1.1 The NMR Signal .................................................................................... 44
      1.7.1.2 The Basic NMR Experiment .................................................................. 46
      1.7.1.3 Multi-Dimensional NMR ....................................................................... 52
      1.7.1.4 The NOE ............................................................................................... 55

2 **Materials and Methods** .................................................................................. 56
  2.1 Materials .......................................................................................................... 56
    2.1.1 Chemicals ...................................................................................................... 56
    2.1.2 Commercial Enzymes ................................................................................... 56
    2.1.3 Commercial Kits ........................................................................................... 56
NMR Spectroscopy

Biochemical and Biophysical Assays

Nucleotide Exchange and HPLC Analysis

Protein Expression and Purification

Cloning

Media and Buffers

Bacterial Strains

Protein Expression Constructs

2.2 Methods

2.2.1 General

2.2.1.1 SDS-PAGE

2.2.1.2 Agarose Gel Electrophoresis

2.2.1.3 Transformation of E. coli (Calcium Treatment)

2.2.2 Cloning

2.2.2.1 The Polymerase Chain Reaction (PCR)

2.2.2.2 TOPO cloning

2.2.2.3 Digestion and Purification of Expression Vectors

2.2.2.4 Ligation of DNA Inserts into Expression Vectors

2.2.2.5 Site-Directed Mutagenesis

2.2.3 Protein Expression and Purification

2.2.3.1 Small Scale Protein Expression

2.2.3.2 Large Scale Protein Expression

2.2.3.3 Cdc42/FTase Co-expression

2.2.3.4 Expression of Isotope Labelled Proteins

2.2.3.5 Affinity Purification - GST

2.2.3.6 Affinity Purification - Ni\(^{2+}\)-affinity Beads

2.2.3.7 Affinity Purification - Farnesyl Transferase

2.2.3.8 Concentration of Protein Solutions

2.2.3.9 Purification by Size Exclusion Chromatography

2.2.3.10 Purification by Ion Exchange

2.2.3.11 Protein Analysis

2.2.3.12 In Vitro Farnesylation

2.2.4 Nucleotide Exchange and HPLC Analysis

2.2.4.1 Nucleotide Exchange

2.2.4.2 HPLC Analysis

2.2.5 Biochemical and Biophysical Assays

2.2.5.1 Scintillation Proximity Assays (SPA)

2.2.5.2 Isothermal Titration Calorimetry (ITC)

2.2.5.3 Analytical Ultracentrifugation (AUC)

2.2.5.4 Liposome Sedimentation Assays

2.2.5.5 Pyrene actin assays

2.2.5.6 Circular Dichroism

2.2.5.7 Nucleotide Hydrolysis Assays

2.2.6 NMR Spectroscopy

2.2.6.1 NMR Experiments

2.2.6.2 NMR Data Processing

2.2.6.3 Chemical Shift Mapping
3 Biochemical Studies of TOCA1 .................................................................... 78
3.1 Introduction................................................................................................. 78
3.2 TOCA1-Cdc42 Binding ............................................................................. 79
  3.2.1 Expression and Purification of the TOCA1 HR1 Domain ..................... 79
    3.2.1.1 Definition of Domain Boundaries ................................................. 79
    3.2.1.2 Protein Expression and Purification ............................................. 80
  3.2.2 Expression and Purification of Cdc42AΔ7Q61L ..................................... 85
  3.2.3 Binding Assays with HR1TOCA1 and Cdc42Δ7Q61L ......................... 86
  3.2.4 Purification of Full Length Cdc42ΔQ61L ............................................. 89
  3.2.5 Binding Assays with HR1TOCA1 and Full Length Cdc42 ................. 90
  3.2.6 Binding Assays with other TOCA1 Domains .................................... 91
  3.2.7 Other TOCA Family Members .......................................................... 94
    3.2.7.1 Defining the Domain Boundaries for FBP17 and CIP4 .......... 94
    3.2.7.2 Protein Expression and Purification ............................................. 95
    3.2.7.3 Competition Binding Assays ...................................................... 97
  3.2.8 The Oligomeric State of the HR1 Domain .......................................... 98
  3.2.9 Measuring Cdc42-HR1CIP4 Binding with AUC ............................... 100
  3.2.10 Other Binding Assays ....................................................................... 103
3.3 Autoregulation of TOCA1 ........................................................................ 104
  3.3.1 Purification and Proteolysis of Full Length TOCA1 ............................ 104
  3.3.2 The HR1 and SH3 Domains ............................................................... 107
    3.3.2.1 Protein Purification ..................................................................... 107
    3.3.2.2 Analysis of the HR1SH3 Di-domain by NMR ............................ 110
    3.3.2.3 Purification of the HR1SH3 Di-domain from pMAT10-P ........... 112
    3.3.2.4 Binding of the HR1SH3 Di-domain to Cdc42 .................................... 113
    3.3.2.5 Purification of the SH3 Domain .................................................. 114
    3.3.2.6 An NMR titration with the HR1 and SH3 Domains .................. 115
  3.3.3 The F-BAR and SH3 Domains ............................................................ 116
    3.3.3.1 Protein Purification ..................................................................... 116
    3.3.3.2 ITC with the F-BAR and SH3 Domains ....................................... 119
    3.3.3.3 Liposome Sedimentation Assays ............................................... 119
  3.3.4 The F-BAR and HR1 Domains .......................................................... 120
3.4 Conclusions and Perspective.................................................................... 122
  3.4.1 TOCA1-Cdc42 Binding: A Low Affinity Interaction ....................... 122
  3.4.2 Regulation of TOCA1: No Evidence for Intramolecular Interactions ... 123
  3.4.3 Where next? ...................................................................................... 124

4 Structural studies of HR1TOCA1 ............................................................. 126
4.1 Introduction............................................................................................... 126
4.2 The NMR Structure of the TOCA1 HR1 Domain ................................... 126
  4.2.1 Preliminary NMR Experiments ........................................................... 126
    4.2.1.1 Two-dimensional NOESY and TOCSY of HR1TOCA1 ............ 126
    4.2.1.2 Purification of 15N-labelled HR1TOCA1 ................................ 128
    4.2.1.3 HSQC Experiments ................................................................. 130
5 Structural Investigation of the Cdc42-HR1\textsuperscript{TOCA1} Interaction........................172
5.1 Introduction..................................................172
5.2 Investigating the Cdc42-HR1\textsuperscript{TOCA1} Interaction by NMR........173
  5.2.1 Mapping the Cdc42 Binding Site onto HR1\textsuperscript{TOCA1}........................................175
    5.2.1.1 HSQC Titrations.........................................175
    5.2.1.2 Analysis of the Binding Surface..............................183
  5.2.2 Mapping the HR1\textsuperscript{TOCA1} Binding Site onto Cdc42................187
    5.2.2.1 Protein Purification.......................................187
    5.2.2.2 HSQC Titrations.........................................187
    5.2.2.3 Analysis of the Titration Data.............................190
    5.2.2.4 Analysis of the Binding Surface..............................193
  5.2.3 Optimising the NMR data..............................................................195
    5.2.3.1 NMR Experiments at Different Temperatures and Field Strengths...195
  5.2.4 Modelling the Complex..................................................201
    5.2.4.1 Ambiguous Interaction Restraints..........................201
    5.2.4.2 Deriving Distance Information from AIRs.....................204
    5.2.4.3 Rigid Body Docking, Semi-flexible Annealing and Water
             Refinement..................................................206
    5.2.4.4 Analysis of the HADDOCK Results..................................207
  5.2.5 Analysis of the Model: Sequence Specific Contacts..................214
5.3 A Possible Cdc42/N-WASP/TOCA1 Ternary Complex..........................219
  5.3.1 Purification of the N-WASP GBD........................................220
  5.3.2 Competitive NMR Titrations........................................223
  5.3.3 Pyrene Actin Assays...........................................225
5.4 Conclusions and Perspective..............................................227
  5.4.1 Structural Insight into Differential Affinities and Specificities....227
  5.4.2 Unidirectional Competition between N-WASP and TOCA1............231
  5.4.3 The Membrane...................................................232

6 Cdc42: Towards Structural Studies on the Membrane..........................233
  6.1 Introduction..................................................233
  6.2 The Oligomeric State of Cdc42........................................234
    6.2.1 Protein Purification........................................235
6.2.2 Analysis of the Oligomeric State ................................................................. 237
6.3 Lipid Modification of Cdc42 ........................................................................... 241
  6.3.1 Co-expression of Cdc42 with Farnesyl Transferase ............................... 242
    6.3.1.1 Preliminary Expressions ..................................................................... 242
    6.3.1.2 Stabilising Cdc42 with Detergent ...................................................... 245
    6.3.1.3 Comparing L191A with L191M .......................................................... 247
  6.3.2 Separation of Farnesylated and Unmodified Protein ............................... 248
    6.3.3 In vitro Farnesylation with Purified Proteins ........................................... 250
      6.3.3.1 Purification of Farnesyl Transferase from E. coli ......................... 250
      6.3.3.2 In vitro Farnesylation ...................................................................... 251
      6.3.3.3 Farnesylation and the Nucleotide State .......................................... 253
6.4 The C-terminus in Nucleotide Hydrolysis .................................................. 255
6.5 An Interaction between the C-terminus and the G Domain ....................... 260
  6.5.1 Assessing Changes in Secondary Structure using CD ............................ 260
  6.5.2 NMR ........................................................................................................ 264
6.6 Membrane Anchoring of Cdc42 ................................................................. 272
6.7 Conclusions ................................................................................................... 276
  6.7.1 Cdc42 is Monomeric in Solution .............................................................. 276
  6.7.2 The C-terminus Affects the G Domain .................................................... 278
  6.7.3 Membrane Anchoring of Cdc42 for Structural Studies ......................... 278

7 Wider Conclusions and Future Directions .................................................... 280
  7.1 Biochemical and Structural Insight into the Pathways of TOCA1-dependent
    Actin Assembly .............................................................................................. 281
    7.1.1 Expansion of the Model of Cdc42-dependent Actin Assembly .............. 284
  7.2 Future Directions .......................................................................................... 287
    7.2.1 Determinants of Binding Specificity and Affinity ................................. 287
    7.2.2 Confirmation of the Effector Handover .................................................. 288
    7.2.3 Structure and Function of Cdc42 .......................................................... 288
    7.2.4 The C-terminus and Effector Binding .................................................. 289
    7.2.5 Membrane-anchoring of Cdc42 ............................................................. 289
    7.2.6 Final Remarks ......................................................................................... 290

Appendices .......................................................................................................... 291
  Appendix 1 Mass Spectrometry Analysis of FBP17 ......................................... 292
  Appendix 2 Dihedral Angle Predictions from TALOS-N ................................. 293
  Appendix 3 Mass Spectrometry Analysis of Cdc42 ......................................... 295

Bibliography ....................................................................................................... 299
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACK</td>
<td>Activated Cdc42 kinase</td>
</tr>
<tr>
<td>AID</td>
<td>Ambiguous interaction distance</td>
</tr>
<tr>
<td>AIR</td>
<td>Ambiguous interaction restraint</td>
</tr>
<tr>
<td>AUC</td>
<td>Analytical ultracentrifugation</td>
</tr>
<tr>
<td>BLI</td>
<td>Biolayer interferometry</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>CIP4</td>
<td>Cdc42 interacting protein 4</td>
</tr>
<tr>
<td>CRIB</td>
<td>Cdc42- and Rac-interactive binding region</td>
</tr>
<tr>
<td>CSP</td>
<td>Chemical shift perturbation</td>
</tr>
<tr>
<td>CT-HSQC</td>
<td>Constant time-heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>EtBr</td>
<td>Ethidium bromide</td>
</tr>
<tr>
<td>F-BAR</td>
<td>FER/CIP4 homology-bar</td>
</tr>
<tr>
<td>FBP17</td>
<td>Formin binding protein 17</td>
</tr>
<tr>
<td>FCC</td>
<td>Fraction of common contacts</td>
</tr>
<tr>
<td>FPP</td>
<td>Farnesyl pyrophosphate</td>
</tr>
<tr>
<td>FTase</td>
<td>Farnesyl transferase</td>
</tr>
<tr>
<td>GAP</td>
<td>GTPase activating protein</td>
</tr>
<tr>
<td>GBD</td>
<td>G protein binding domain</td>
</tr>
<tr>
<td>GDI</td>
<td>Guanine nucleotide dissociation inhibitor</td>
</tr>
<tr>
<td>GDP</td>
<td>Guanosine diphosphate</td>
</tr>
<tr>
<td>GEF</td>
<td>Guanine nucleotide exchange factor</td>
</tr>
<tr>
<td>GGTase</td>
<td>Geranylgeranyl transferase</td>
</tr>
<tr>
<td>GMPPNP</td>
<td>Guanosine 5’-β, γ-imido]triphosphate</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferase</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine triphosphate</td>
</tr>
<tr>
<td>HMQC</td>
<td>Heteronuclear multiple quantum coherence</td>
</tr>
<tr>
<td>HPLC</td>
<td>High pressure liquid chromatography</td>
</tr>
<tr>
<td>HR1</td>
<td>Homology region 1</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>ICMT</td>
<td>Isoprenylcysteine carboxylmethyltransferase</td>
</tr>
<tr>
<td>INEPT</td>
<td>Insensitive nuclei enhanced by polarization transfer</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal titration calorimetry</td>
</tr>
<tr>
<td>N-WASP</td>
<td>Neural-wiskott-aldrich syndrome protein</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser effect</td>
</tr>
<tr>
<td>NOESY</td>
<td>Nuclear overhauser effect spectroscopy</td>
</tr>
<tr>
<td>OG</td>
<td>Octyl glucoside</td>
</tr>
<tr>
<td>PAK</td>
<td>p21-activated kinase</td>
</tr>
<tr>
<td>PC</td>
<td>Phosphatidylcholine</td>
</tr>
<tr>
<td>PCH</td>
<td><em>Pombe</em> Cdc15 homology</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PI(3)P</td>
<td>Phosphatidylinositol-3-phosphate</td>
</tr>
<tr>
<td>PI(4,5)P₂</td>
<td>Phosphatidylinositol-4,5-bisphosphate</td>
</tr>
<tr>
<td>PRK</td>
<td>Protein kinase C-related kinase</td>
</tr>
<tr>
<td>PS</td>
<td>Phosphatidylserine</td>
</tr>
<tr>
<td>Ras</td>
<td>Rat sarcoma</td>
</tr>
<tr>
<td>RBD</td>
<td>Ras binding domain</td>
</tr>
<tr>
<td>RCE1</td>
<td>Ras-converting CAAX endopeptidase 1</td>
</tr>
<tr>
<td>Rho</td>
<td>Ras homology</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root mean squared deviation</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SH3</td>
<td>Src-homology 3</td>
</tr>
<tr>
<td>SNX9</td>
<td>Sortin nexin 9</td>
</tr>
<tr>
<td>SPA</td>
<td>Scintillation proximity assays</td>
</tr>
<tr>
<td>TOCA1</td>
<td>Transducer of Cdc42-dependent actin assembly protein 1</td>
</tr>
<tr>
<td>TOCSY</td>
<td>Total correlation spectroscopy</td>
</tr>
<tr>
<td>VCA</td>
<td>Verprolin-homology, coflin-homology, acidic region</td>
</tr>
<tr>
<td>WASP</td>
<td>Wiskott-aldrich syndrome protein</td>
</tr>
<tr>
<td>WH1</td>
<td>WASP homology region 1</td>
</tr>
<tr>
<td>WIP</td>
<td>WASP interacting protein</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 The Ras Superfamily

The Ras (Rat sarcoma) superfamily of small GTPases are ‘molecular switches’ that preside over an astounding array of cell signalling pathways. They influence a broad range of cellular processes, from gene expression to cytoskeletal organisation to vesicle trafficking (reviewed in Takai et al. 2001). The superfamily comprises more than 160 members in humans (Rojas et al. 2012), united by their ability to bind to and hydrolyse guanosine triphosphate (GTP) but divided by their diverse range of cellular functions.

The superfamily can be divided into five families (Ras, Rho, Rab, Arf and Ran) based on structural and functional similarities but all share a well-defined core protein structure of ~20 kDa, known as the G domain (Bourne et al. 1991). The G domain is responsible for high affinity binding to GDP and GTP ($K_d = 1 \text{ pM}-10 \text{ nM}$) and for GTP hydrolysis. It is this nucleotide binding and hydrolysis that underlies their functions as molecular switches as they bind to downstream effector proteins and exert their effects within the cell only when they are GTP bound. GTP hydrolysis therefore switches the proteins off whilst nucleotide exchange from GDP to GTP switches them on.

The molecular switch, or GTPase cycle, centres on the G domain but also depends upon auxiliary proteins that regulate the nucleotide binding and hydrolysis. Guanine nucleotide exchange factors (GEFs) stimulate exchange of GDP for GTP and so switch the GTPase on whilst the GTPase activating proteins (GAPs) stimulate intrinsic GTP hydrolysis and so switch the GTPase off (reviewed in Cherfils and Zeghouf 2013). Guanine nucleotide dissociation inhibitors (GDIs) have also been identified for Rho and Rab family members, which sequester the GDP bound GTPase in the cytosol in the off state (reviewed in DerMardirossian and Bokoch 2005; Cherfils and Zeghouf 2013). The general cycle is illustrated in Figure 1.1.
The G Domain

The G domain is found in more than 37,000 proteins across all kingdoms of life, including all known Ras superfamily members (reviewed in Wittinghofer and Vetter 2011). It is 18-20 kDa in size, has a universally conserved structure and can be recognised by five conserved sequence motifs known as G1-G5 (Bourne et al. 1990; Bourne et al. 1991). It constitutes the basic unit for guanine nucleotide binding, binds to GDP and GTP and mediates a universal GDP/GTP switch mechanism (reviewed in

Figure 1.1. The GTPase Cycle. GTPases are activated by exchange of GDP for GTP. Specific Guanine nucleotide Exchange Factors (GEFs) stimulate this process in response to upstream signals. GTPases bind to a range of effector proteins in their active, GTP-bound state. Intrinsic GTP hydrolysis is stimulated by GTPase-Activating Proteins (GAPs), thus returning the GTPase to the inactive state. For the Rab and Rho families, Guanine nucleotide Dissociation Inhibitors (GDIs) sequester the GTPase in the inactive form, opposing its activation by GEFs.
The definitive structure consists of a hydrophobic core of 6 β-strands surrounded by 5 α-helices connected by hydrophobic loops as reviewed in (Bourne et al. 1991; Wittinghofer and Vetter 2011) and shown in Figure 1.2. The five conserved sequence motifs, G1-G5, are involved in nucleotide binding.

**Figure 1.2. A Schematic of the G Domain Structure.** The β-strands are represented by dark blue arrows and the α-helices as cyan cylinders. The connecting loops are coloured grey and shown by lines. G1/P-loop is coloured in orange, G2/Switch 1 red, G3/Switch 2 green, G4 magenta and G5 purple.

Ras was the first G protein to have high resolution structures available (de Vos et al. 1988; Pai et al. 1989; Milburn et al. 1990; Brünger et al. 1990; Pai et al. 1990) (Figure 1.3) and is considered the archetypal example of the G domain and the GDP/GTP switch. G1-G5 were initially defined based on the structure of Ras but are found across the Ras superfamily and in other GTP binding proteins. G1-G4 are the most conserved across GTP binding proteins whilst G5 is only weakly conserved and is not always present. Variations of G1-G3 are also found in ATP binding proteins whereas G4 and G5 confer Guanine nucleotide specificity.

G1 comprises residues 10-17 of Ras, connecting the first β-strand (β1) to the first amphipathic α-helix (α1). It contains the consensus sequence GxGK(S/T). The main chain amides of Gly15 and Lys16 and the sidechain amino group of Lys16 contact the β-phosphate of GDP or GTP (or ADP/ATP in ATPases). The Serine or Threonine residue (Ser17 in Ras) coordinates a magnesium ion that is essential for nucleotide binding. This region is also known as the P-loop as it constitutes a highly conserved phosphate binding region found across GTP binding and ATP binding proteins.
**Figure 1.3. The G Domain of Ras.** A) A crystal structure of Ras-GMPPNP (PDB code: 5B2Z) coloured as in Figure 1.2. The β-strands are dark blue, α-helices cyan and loops grey. G1-G5 are marked in orange, red, green, magenta and cyan, respectively. The nucleotide is shown with black lines and the magnesium ion as a grey sphere. B) The amino acid sequence of Ras is shown with regions G1-G5 boxed according the same colour scheme as in A. The consensus sequence for each region is shown below the sequence. Φ indicates a hydrophobic residue and ζ indicates a hydrophilic residue. Switch 2 is boxed in grey and labelled.
G2, also known as Switch 1, comprises residues 32-40 in Ras and encompasses part of β2 and the loop preceding it. This region is defined by an invariant Thrionine residue (Thr35 in Ras). The hydroxyl sidechain is also conserved in the ATP binders, myosin and kinesin, which have a Serine at this position. This invariant residue is central to the GTPase switch mechanism as it forms main chain hydrogen bonds with the γ-phosphate of GTP, which are broken when GTP is hydrolysed (Pai et al. 1989; Milburn et al. 1990; Pai et al. 1990). Its sidechain OH contacts the magnesium.

G3 comprises residues 53-62 of Ras. The consensus sequence is D-x₂-G followed by a hydrophilic residue and the Glycine residue (Gly60 in Ras) is as invariant as the Thrionine of G2. It also forms main chain hydrogen bonds with the γ-phosphate of GTP (Pai et al. 1989; Milburn et al. 1990; Pai et al. 1990) and is conserved in ATP binders (Vetter and Wittinghofer 2001). The Aspartate of G3 indirectly contacts the magnesium ion via water. G3 overlaps with the region known as Switch 2 (residues 58-67 of Ras), which includes the invariant Glycine. G2/Switch 1 and G3/Switch 2 are critical players in the universal molecular switch mechanism of GTP binders that will be described in section 1.1.2.

G4 comprises 112-119 of Ras and consists of four hydrophobic residues followed by (N/T)(K/Q)xD. This region is the main determinant for guanine base specificity and also has a role in stabilising the nucleotide binding site via hydrogen bonds with residues of G1. It makes direct contacts with the guanine ring via the conserved Aspartate (Asp119 of Ras), or occasionally Glutamate, which forms hydrogen bonds with the guanine base. G5 (including the consensus, E(A/C/S/T)SA(K/L)) stabilises G4 and also contributes to guanine nucleotide specificity via a direct contact between the main chain amide of Ala146Ras and O6 of the guanine ring. The close contact of this Alanine with the base sterically and electrostatically excludes Adenine.

These important G domain-nucleotide contacts derived from the Ras structures are summarised in Figure 1.4. The greatest contributions to tight nucleotide binding come from contacts between G1 and the β-phosphate of GDP/GTP and between G4/G5 and the Guanine base (Rensland et al. 1995).
1.1.2 The GTPase Switch

Comparisons of the crystal structures of GTP and GDP bound Ras (de Vos et al. 1988; Pai et al. 1989; Milburn et al. 1990; Schlichting et al. 1990; Brünger et al. 1990; Pai et al. 1990) showed that the overall conformation of the G domain in the GTP and GDP bound states is similar. Significant differences were, however, observed in Switch 1 (residue 32-40) and Switch 2 (residues 58-67). There are now more than 590 structures of GTPases bound to a range of nucleotides and nucleotide analogues and these have revealed that changes to Switches 1 and 2 between GDP and GTP forms are common across the Ras superfamily (Wittinghofer and Vetter 2011).

The changes in G2/Switch 1 arise from a change in the orientation of the invariant Thrreonine (Thr35 in Ras), which contacts GTP directly via the γ-phosphate and indirectly via the magnesium ion. Following GTP hydrolysis, this residue points away from the nucleotidte allowing the switch region to undergo conformational changes. The changes to G3/Switch 2 are mediated by the invariant Glycine (Gly60 in Ras), which also contacts the γ-phosphate prior to hydrolysis. The switch mechanism has...
been simplistically described as the ‘loaded spring’, in which the switches become free to change conformation upon loss of their contacts with the γ-phosphate of the nucleotide (Rudolph et al. 2001), as depicted in Figure 1.5. The invariance of these contacts across GTPases suggest that the loaded spring mechanism is universal.

Figure 1.5. A ‘Loaded Spring’. A diagram illustrating the loaded spring mechanism. Contacts between the γ-phosphate of GTP and the invariant Threonine and Glycine of the switch regions of the GTPase are lost upon GTP hydrolysis. The switch regions are free to reorient following the loss of these contacts. This figure was taken from (Wittinghofer and Vetter 2011).

The structural details of the conformational changes are poorly defined for most of the Ras superfamily because the conformations of the two switch regions are rarely well defined in NMR or crystal structures of either GDP or GTP bound forms. They are often only visible when the protein is bound to an effector or regulatory protein, or as a result of crystal packing. This is due to their inherent flexibility, which has been demonstrated using NMR relaxation experiments (Ito et al. 1997; Loh et al. 1999).

In addition to conformational alterations, differences have been observed in the dynamics of the two switches in different nucleotide states. The switches are flexible relative to the rest of the G domain in both the GDP and GTP states but show greater rigidity in the GTP state (Vetter and Wittinghofer 2001). Furthermore, conformational sub-states have been identified within the GTP bound state of Ras using solution $^{31}$P-NMR. One of the two states is the same as that bound to the Ras binding domain (RBD) of effector proteins so is termed the active state, whilst the other is termed the
inactive state (Spoerner et al. 2001; Spoerner et al. 2005a; Ye et al. 2005; Spoerner et al. 2005b; Liao et al. 2008; Spoerner et al. 2010). Such sub-states have also been demonstrated for other members of the Ras superfamily (Geyer et al. 1999) and using alternative techniques such as FTIR (Kötting and Gerwert 2005) and solid state $^{31}$P-NMR (Stumber et al. 2002).

The structural and dynamic alterations in the switch regions confer the ability for only GTP-bound GTPases to bind to their cognate downstream effector proteins, through which they exert their extensive roles within the cell. In this way, the two switch regions ‘sense’ the nucleotide state and transmit the signal downstream.

### 1.1.3 Family Specific Variations on the G Domain

The Ras superfamily is divided into 5 families (Ras, Rho, Rab, Arf and Ran). The Rho, Arf and Ran families have distinctive structural features in addition to the core G domain. The Rho family is defined by a 13-residue $\alpha$-helical insertion between $\alpha_4$ and $\beta_4$, the function of which is unknown (Hirshberg et al. 1997; Wei et al. 1997). The Arf family is recognised by an N-terminal amphipathic helix that rests on the G domain in the GDP bound form and inserts into membranes following GDP/GTP exchange (Carlos Amor et al. 1994; Pasqualato et al. 2002). Comparably, the Ran family is marked by a C-terminal extension, which binds to the body of GDP-bound Ran but is freed for downstream interactions upon GTP binding (Scheffzek et al. 1995). The defining structural features of each family are illustrated in Figure 1.6.
The switch regions are well conserved within each family of the Ras superfamily but, other than the Threonine and Glycine, are less conserved between families and across other G proteins. Therefore, the extent of conformational change is different in different proteins despite the universal loaded spring mechanism that hinges on the invariant Threonine and Glycine. In addition, regions outside of the two switches also undergo conformational changes that are family-specific. For example, the conformational changes in the Arf and Ran families are far more pronounced than the minor changes seen for the Ras, Rho and Rab families. For Arf proteins, there is a profound change in Switch 1 and detachment of the N-terminal helix extension when going from the GDP to GTP bound form (Antonny et al. 1997; Goldberg 1998) and for Ran proteins, there is a large change in Switch 1, unfolding of a β-strand and relocation of the C-terminal extension (Vetter et al. 1999b; Chook and Blobel 1999; Vetter et al. 1999a). For the Ras, Rho and Rab families the changes are less well

Figure 1.6. The Ras, Rho, Rab, Arf and Ran Families: Extensions to the G Domain. Cartoon representations of A) Ras·GDP (1AA9), B) RhoA·GDP (1FTN), C) Rab7·GDP (1VG1), D) Arf6·GDP (1E0S) and E) Ran·GDP (3GJ0). The G domain is coloured similarly to Figure 1.1, with β-strands in dark blue, α-helices in cyan and loops in grey. Additional structural features of the Rho, Arf and Ran families are coloured red. The Rho family contain an additional α-helix between β5 and α4 (the Rho insert), the Arf family have an additional N-terminal α-helix and the Ran family have a C-terminal extension including an α-helix.
characterised due to the flexibility of the switch regions and may be more subtle, involving changes in dynamics as well as conformation.

In addition to their structural differences, the different families have distinct functional roles (reviewed in Etienne-Manneville and Hall 2002; Wennerberg et al. 2005; Mizuno-Yamasaki et al. 2012): the Ras family are involved in controlling cell proliferation, differentiation and protection from apoptosis; the Rho family is famous for its role in cell shape and cytoskeletal regulation; the Rab and Arf families regulate vesicle-dependent processes and vesicle trafficking; and the Ran family, consisting of just one member, is responsible for nuclear transport. Generally, the structural features of each family do not explain the functional differences but rather their primary structure and therefore their specificities for particular upstream regulators and downstream effectors are responsible. Specific post translational modifications that lead to localisation of GTPases to specific membranes may also play a role in mediating their diverse and distinct roles.

1.1.4 Guanine Nucleotide Exchange Factors
The small GTPases bind GDP and GTP with high affinities in the low nanomolar to picomolar range ($K_d = 10$ pM for Ras). Given the cellular concentrations of GDP (0.01 mM) and GTP (0.1 mM) (Bourne et al. 1991), they are saturated in vivo. Exchange of GDP for GTP (switching on) is therefore very slow, on the order of hours (Bos et al. 2007). To accelerate this process there are a number of Guanine nucleotide Exchange Factors (GEFs), which reduce the affinity for the nucleotide and increase the rate of nucleotide dissociation by several orders of magnitude (reviewed in Bos et al. 2007; Cherfils and Zeghouf 2013). Following dissociation of GDP, it is thought that either nucleotide can bind but that GTP binds preferentially because it is present at 10-fold higher concentrations in the cell.

GEFs have been identified for every family within the Ras superfamily. For example, there are ~30 RasGEFs (Diez et al. 2011) and more than 80 RhoGEFs (Rossman et al. 2005; Côté and Vuori 2007) currently known. They are generally multi-domain proteins with protein-protein and protein-lipid binding domains alongside their GEF domains (reviewed in Bos et al. 2007; Cherfils and Zeghouf 2013).
The GEF domains for different GTPase families are largely structurally unrelated and they bind their GTPase in vastly different orientations but, despite the lack of conservation in structure and GTPase binding mode, the mechanisms by which GEFs cause nucleotide dissociation are generally shared (reviewed in Bos et al. 2007; Cherfils and Zeghouf 2013). The mechanism of GEF action typically involves changes to the magnesium and phosphate binding regions (G1/P-loop, G2/Switch 1 and G3/Switch 2), leaving the guanine base binding regions (G4/G5) largely unaffected. Changes to G1-G3 typically achieve a number of things, including steric occlusion of the magnesium and phosphate binding regions, electrostatic repulsion of the phosphates, hydrophobic repulsion of the magnesium and stabilisation of the empty GTPase structure by replacement of electrostatic contacts lost upon removal of the nucleotide. Insertion of GEF residues near to the magnesium and/or phosphates is also commonly seen.

The Dbh-homology domains of the RhoGEFs and the structurally unrelated CDC25-homology domains of the RasGEFs exemplify a number of these mechanistic features. Firstly, they remodel Switch 2 upon GTPase binding, reorienting the conserved Alanine of Switch 2 (Ala69 of Rho/Ras) near to the magnesium binding site and causing hydrophobic repulsion of the magnesium ion. Secondly, they reorient Glu62 of Switch 2 to form a salt bridge with Lys16 of the P-loop and so stabilise the nucleotide-free GTPase. Thirdly, they sterically remove Switch 1 from the nucleotide binding site, thus removing favourable Switch 1-nucleotide/magnesium contacts. The RhoGEFs also insert a Leucine residue near to the magnesium, causing further hydrophobic repulsion (Bouquier et al. 2009).

The regulation of GTPases by GEFs provides clear links between upstream signals and GTPase activation, as GEFs are regulated by protein-protein and protein-lipid interactions, second messengers and post-translational modifications (reviewed in Bos et al. 2007). These regulatory events can affect GEFs via translocation: for example, the RacGEF, Tiam1, is localised to the plasma membrane by interactions with Ras·GTP (Lambert et al. 2002). Furthermore, calcium/diacylglycerol signalling causes localisation of RasGEFs to the plasma membrane (Ebinu et al. 1998) and phosphorylation of the Ras/RacGEF, SOS, causes its dissociation from adaptor
proteins that localise it to specific regions of the plasma membrane (Waters et al. 1995; Dong Chen et al. 1996).

In addition to their effects on localisation, the protein-protein interactions, second messengers and post-translational modifications affect autoinhibitory interactions within the GEF and allosterically activate the GEF catalytic domains. For example, cAMP activates RapGEFs by relieving an autoinhibitory interaction between its regulatory and catalytic domains (Rehmann et al. 2006), as does phosphorylation of the RhoGEF, Vav1 (Aghazadeh et al. 2000). Furthermore, a number of RhoGEFs are allosterically activated by interactions with the Gα subunit of heterotrimeric G proteins (reviewed in Siehler 2009). The GEFs, therefore, constitute a means by which diverse upstream signals are communicated to specific small GTPases.

1.1.5 GTP Hydrolysis and GTPase Activating Proteins

Much like nucleotide exchange, intrinsic GTP hydrolysis is very slow in most small G proteins, in the region of $10^{-4}$-$10^{-5}$ s$^{-1}$ (Hall et al. 2002). This slow intrinsic rate of hydrolysis is stimulated by GAPs by several orders of magnitude (reaching 1-10 s$^{-1}$). GAPs, therefore, mediate the return to the inactive, GDP bound state (reviewed in Cherfils and Zeghouf 2013). The sequence and structure conservation of GAPs across families is low, as was seen for the GEFs, except for the Ras and RhoGAPs which share some structural similarities. The mechanism of GTPase activation is, however, generally shared and involves three main features: orientation and polarisation of a water molecule for in-line nucleophilic attack on the γ-phosphate, occlusion of water from the catalytic site and stabilisation of the transition state.

Two of the first GAP/small GTPase complexes to be structurally characterised were p120GAP/Ras·GDP·AlF$_4$ (Scheffzek et al. 1997) and p50-RhoGAP/RhoA·GDP·AlF$_4$ (Rittinger et al. 1997). The AlF$_4$ locks the GTPase-GAP complex in a transition state. The two GAPs are unrelated in primary structure and only distantly related in tertiary structure but their mechanism of action is shared. In both cases the conserved Glutamine of Switch 2 (Q61$_{\text{Ras}}$/Q63$_{\text{Rho}}$) is positioned by the GAP to coordinate and activate a water molecule for nucleophilic attack (Figure 1.7A). Also for both GAPs, an Arginine residue is inserted near to the phosphates, which stabilises the partial
negative charges in the transition state. This Arginine, termed the Arginine ‘finger’, is common to many known GAPs including Ras, Rho and Arf GAPs (Figure 1.7B) but is sometimes replaced with an Asparagine ‘thumb’, for example in RapGAP (Daumke et al. 2004). Both are analogous to a catalytic Arginine found within Switch 1 of the heterotrimeric G proteins, which is responsible for their higher rates of intrinsic hydrolysis compared with the Ras superfamily (Coleman et al. 1994).

**Figure 1.7. GAPs in GTP Hydrolysis.** A) A diagram illustrating the mechanism of GAP-induced GTP hydrolysis, based on Figure 3 of (Bos et al. 2007). The GAP is represented in purple and the GTPase in red. A Glutamine residue in Switch 2 of the GTPase is positioned by the GAP such that it aligns and activates the nucleophilic water for attack on the γ-phosphate of GTP. The Arginine finger of the GAP is inserted near the phosphates, stabilising the developing negative charges of the transition state. B) Examples of Arginine fingers seen in Ras, Arf, Rho and Arl3 GAPs, taken from Figure 6 of (Cherfils and Zeghouf 2013).
Like the GEFs, the GAPs are subject to upstream regulation from protein-protein and protein-lipid interactions, second messengers and post translational modifications. For example, interactions between the SH2 domains of RasGAP and receptor tyrosine kinases modulate the GAP localisation and so activity (Bernards and Settleman 2004). Moreover, membrane localisation and activity of the RacGAP, β2-chimaerin, is affected by second messengers such as diacylglycerol (Caloca et al. 2003) and the RhebGAP, TCS2, is inhibited by phosphorylation (Avruch et al. 2006). As such, GAPs provide another clear link between upstream signals and downstream effects.

1.1.6 Post Translational Modifications and Membrane Localisation
The GTPase cycle relies upon the G domain and auxiliary proteins as described thus far. There is, however, another dimension to the picture. In the early 1980s it was discovered that Ras is post translationally modified and that only the mature protein associates with the plasma membrane (Willingham et al. 1980; Sefton et al. 1982; Shih et al. 1982). Furthermore, the C-terminal region of Ras was shown to be required for lipid binding, membrane association and cell transformation (Willumsen et al. 1984). In the following decade a number of GTPases across the Ras, Rho, Rab and Arf families were found to be modified by the addition of either myristic acid (myristoylation) or isoprenoid groups (prenylation) (reviewed in Glomset and Farnsworth 1994; Lane 2006; Wang and Casey 2016). The Arf family are myristoylated with a 14-carbon saturated fatty acid on their N-terminal Glycine whilst the Ras, Rho and Rab families are prenylated with 15- or 20-carbon isoprenoids at their C-termini.

Prenylation of the Ras, Rho and Rab families depends upon a 4-residue motif at their C-terminus known as the CAAX motif (A = aliphatic, X = any). The final residue (X) determines the specificity for one of two prenyl transferases, Farnesyl or Geranylgeranyl Transferase (FTase or GGTase) (Roskoski and Ritchie 1998). For Ras, X is either Serine or Methionine, which is recognised by FTase. FTase transfers a 15-carbon farnesyl group to the Cysteine residue of the CAAX motif, a process which is essential for the correct membrane localisation of Ras in cells (Brunsveld et al. 2009). For other members of the Ras family such as Rap1 and RalA/B, X is a Leucine, which confers GGTase specificity. GGTase transfers a 20-carbon geranylgeranyl
group to the Cysteine. The CAAX of many Rho family members, including RhoA, Rac1 and Cdc42, also ends in a GGTase-specific amino acid (usually Leucine or Phenylalanine) but other less well studied Rho proteins such as Rnd1, TC10 and TCL have FTase-specific CAAX motifs (Roberts et al. 2008).

Following prenylation of the Cysteine, the AAX is cleaved by membrane-associated Ras-converting CAAX endopeptidase 1 (RCE1) at the endoplasmic reticulum prior to methylesterification of the new C-terminus by isoprenylcysteine carboxymethyltransferase (ICMT) (reviewed in Wang and Casey 2016) (Figure 1.8). These two enzymes are required for proper development in mice, as reviewed in Wang and Casey 2016, but the importance of the cleavage and methylesterification steps in GTPase localisation and downstream function is unclear as the two enzymes have a huge number of protein targets.

**Figure 1.8. Post Translational Modification of the Ras, Rho and Rab Families.** The proteins are first modified at the Cysteine of the C-terminal CAAX motif by addition of a farnesyl (15 carbon) or geranylgeranyl (20 carbon) group. This is carried out by Farnesyl or Geranylgeranyl Transferase (FTase/GGTase). Ras-converting CAAX endopeptidase 1 (RCE1) recognises the lipid modified protein and cleaves the three C-terminal residues of the CAAX. Isoprenylcysteine carboxymethyltransferase (ICMT) completes the modifications by methylesterification of the new C-terminus.

Many GTPases contain polybasic regions, phosphorylatable Serine/Threonine residues and/or further lipid modification sites upstream of the modification. For example, palmitoylation of the brain isoform of Cdc42 near to the prenylated Cysteine affects its interactions with regulatory proteins (Nishimura and Linder 2013) and
phosphorylation within the prenylated C-terminal regions of other Rho family members leads to inhibition of their activity (Riou et al. 2013).

Lipid modification and membrane anchoring of Ras does not appear to affect the overall structure of the G domain (Güldenhaupt et al. 2008) but it is essential for proper localisation, biological function and regulation of most small GTPases (Maurer-Stroh et al. 2007; Roberts et al. 2008). The importance of prenylation in GTPase regulation is exemplified by RhoB, a small GTPase that can be farnesylated or geranylgeranylated to carry out distinct functions (Du et al. 1999). Membrane anchoring also affects GEF-mediated activation of GTPases (Peurois et al. 2017) and, for the Rho family, lipid modification affects interactions with guanine nucleotide dissociation inhibitors (GDIs) (reviewed in DerMardirossian and Bokoch 2005; Cherfils and Zeghouf 2013). Lipid modification and membrane association therefore provides another layer of regulation upon the GTPase cycle.

### 1.1.7 Guanine Nucleotide Dissociation Inhibitors

Guanine nucleotide Dissociation Inhibitors (GDIs) have been identified for Rho and Rab GTPases (reviewed in DerMardirossian and Bokoch 2005; Cherfils and Zeghouf 2013). They bind to the prenylated GTPases, preferentially in the GDP form, and sequester the GTPase in the cytosol in its inactive state, protecting the lipid group from the aqueous environment and preventing downstream interactions with effectors. The binding of GEFs, GAPs and GDIs are mutually exclusive such that nucleotide exchange and hydrolysis is prevented when the GTPase is GDI bound.

### 1.1.8 Effector Binding

Despite the highly conserved G domain and molecular switch mechanism, the downstream target proteins of the Ras superfamily do not contain particular recognizable sequence motifs, nor can they be predicted based on their structural features. There are currently more than 60 small GTPase-effector complexes in the Protein Data Bank (PDB) and the effectors represent a range of protein folds (reviewed in Mott and Owen 2015). Given that there are more than 160 members of the Ras superfamily and individual members have been shown to recognise multiple
effectors, it is likely that there are hundreds of effectors that have not yet been identified or studied in detail.

The GTPase effector complexes that have been studied so far can be classified into 4 main groups based on their structural features (reviewed in Mott and Owen 2015). These structural classes are not specific to single GTPases, families or functions and a single GTPase can recognise multiple effectors of different classes. The most populous group is the helical pair effectors, which bind GTPases via a pair of α-helices. The second most populous and the first to be identified was the intermolecular β-sheet effectors, which form intermolecular β-sheets with the central β-sheet of the G domain, incorporating Switch 1. The third, sparsely populated group is the PH domain effectors that use different regions of their 7-stranded anti-parallel β-sheet and C-terminal α-helix to bind GTPases in a diverse number of orientations. The fourth group contains all of the other effectors that do not fit into these categories. These final two classes will not be discussed further.

Helical pair effectors have been identified for the Ras, Rho, Rab and Arf families. Both parallel and anti-parallel coiled-coils bind GTPases in a variety of orientations, using a range of salt bridges, hydrogen bonds and hydrophobic contacts. The interactions most commonly involve anti-parallel helix-helix interactions with Switch 2 but there are also examples of parallel helix-helix interactions. Switch 1 is also involved in most cases but to a lesser extent (reviewed in Mott and Owen 2015). Famous examples include the Protein Kinase C-Related Kinases (PRKs) that interact with members of the Rho family (Watanabe et al. 1996).

Intermolecular β-sheet effectors have been found for the Ras, Rho and Arf families. The first example to be structurally defined was the Ras binding domain (RBD) of c-Raf with the Ras family member, Rap1a (Nassar et al. 1995). Since then, several more complexes between Ras family members and β-sheet effectors have been studied and some general features of the interactions have been identified. The effector domain consists of a fully folded β-sheet, which contacts Switch 1/β2 of the GTPase to form an antiparallel intermolecular β-sheet and is largely unchanged upon binding. Parallel intermolecular β-sheets have also been observed, for example when Exo84 binds the Ras family member, RalA (Jin et al. 2005). The Ras surface is generally negatively
charged and the effector surface is positively charged with very few hydrophobic contributions (reviewed in Mott and Owen 2015). The Arf family bind to folded β-sheet proteins in a similar way.

In contrast to the pre-formed β-sheet effectors of the Ras and Arf families, the β-sheet of the Rho binders is induced only upon GTPase binding, after which the intermolecular β-sheet is formed. For example, the Cdc42- and Rac-Interacting Binding regions (CRIB) of Activated Cdc42 Kinase (ACK), p21-Activated Kinase (PAK) and the Wiskott-Aldrich Syndrome Protein (WASP) are largely disordered until they bind Cdc42 (Abdul-Manan et al. 1999; Mott et al. 1999; Morreale et al. 2000). The intermolecular β-sheet that is formed is antiparallel, as in most of the Ras-effector complexes but, with the exception of PAK, is more irregular. These interactions are also more hydrophobic, involving contacts between α5 of Cdc42 and hydrophobic residues of the region N-terminal to the CRIB β-sheet.

Even within the different structural classes of effectors it is difficult to predict the details of the interaction as the β-sheet class can form parallel and anti-parallel intermolecular sheets and the helical class can be divided into six different types. The interactions are also not easily predicted based on sequence and residue types because hydrophobic, charged and mixed interfaces have been observed. However, in general, Switch 1, Switch 2, parts of β1 and α1 in the inter-switch region and β3 (see Figure 1.3) are involved in many GTPase-effector interactions. For the Rho, Rab and Arf families the end of α3 and the following loop also commonly contacts effectors. α5 is also implicated in many Rho interactions and the region C-terminal to the G domain has been implicated in Rab interactions. Ran is rather different, with known contacts across most of its sequence.

Specificity determinants for some interactions have been described, but for others there are no clear sequence specific contacts. It is thought that, in some cases, structural and dynamic features may also contribute to specificity (Hutchinson et al. 2013). How different GTPases bind to a number of effectors selectively and specifically therefore remains an important unanswered question in GTPase biology.
1.2 The Rho family

1.2.1 The Diverse Roles of a Small Family

The Rho family comprises 20 members in mammals (reviewed in Etienne-Manneville and Hall 2002). The founding members of the family, RhoA, Rac1 and Cdc42, have been relatively well studied and more than 60 effector proteins have been identified for these three family members alone. The remaining 17 GTPases of the Rho family have been less well studied and, for many, their roles and effectors have not been defined at all.

RhoA, Rac1 and Cdc42 are best known for their roles in the regulation of the cytoskeleton (Nobes and Hall 1995; Machesky and Hall 1996; Hall 1998; Machesky and Insall 1999). RhoA rearranges existing assemblies of actin and myosin filaments leading to formation of stress fibres (Ridley and Hall 1992), whilst Rac1 and Cdc42 induce de novo actin polymerisation to form actin-rich surface protrusions known as lamellipodia (Ridley et al. 1992) and filopodia (Nobes and Hall 1995; Kozma et al. 1995), respectively. A number of the RhoA, Rac1 and Cdc42 effector proteins have, unsurprisingly, also been directly linked to the pathways that govern cytoskeletal dynamics. These include Formins (Kühn and Geyer 2014), members of the Protein Kinase C-related Kinase (PRK) family (Watanabe et al. 1996), the Wiskott-Aldrich Syndrome (WASP) family (Symons et al. 1996) and the Transducer of Cdc42-dependent Actin assembly (TOCA) family (Aspenström 1997; Ho et al. 2004; Wu et al. 2013) amongst many others.

In addition to their roles in cytoskeletal organisation, which have now been confirmed in mammalian cells, yeast, flies and worms, many more roles have been identified for RhoA, Rac1 and Cdc42 (reviewed in Etienne-Manneville and Hall 2002). For example, Cdc42 is required for correctly localised cell division in yeast as well as in the formation of surface protrusions, known as shmoos, during yeast mating (Pruyne and Bretscher 2000). Furthermore, Rac1 and Cdc42 cooperate to establish epithelial cell polarity and maintain cell shape (reviewed in Etienne-Manneville and Hall 2002) and RhoA, Rac1 and Cdc42 all contribute to G1 cell cycle progression in yeast (Olson et al. 1995). Altogether, RhoA, Rac1 and Cdc42 have been found to regulate cell
polarity, gene transcription, G1 cell cycle progression, microtubule dynamics and vesicular transport pathways.

1.2.2 Cdc42 and the Actin Cytoskeleton
Following the discovery that Cdc42 regulates actin dynamics (Ridley and Hall 1992; Nobes and Hall 1995; Kozma et al. 1995), the search began for effector proteins that would link Cdc42 to actin and filopodia formation. The first link came from studies of Wiskott-Aldrich Syndrome patients, who have mutations in the gene encoding WASP and show cytoskeletal disarray in T lymphocytes. It was therefore postulated and subsequently demonstrated that WASP interacts with Cdc42 (Symons et al. 1996; Kolluri et al. 1996).

These studies provided evidence that WASP binds specifically to Cdc42·GTP and identified a CRIB region similar to those previously identified in the Cdc42 binders, ACK and PAK (Manser et al. 1993; Manser et al. 1994). Mutations in the CRIB region and inactivating mutations in Cdc42 also showed that disruption of the Cdc42-WASP interaction abolishes Cdc42-induced clustering of WASP with actin and has downstream effects on the cytoskeleton. This showed that Cdc42 affects actin via an interaction with WASP. The ubiquitously expressed WASP homologue, N-WASP, also binds Cdc42 via its CRIB region and this interaction is required for Cdc42/N-WASP-induced filopodia formation (Miki et al. 1998a).

Around the same time, studies using cytoplasmic extracts from Xenopus eggs showed that actin polymerisation is globally inhibited but can be activated by Cdc42 and the signalling lipid phosphatidylinositol-4,5-bisphosphate (PI(4,5)P₂) (Ma et al. 1998a; Ma et al. 1998b). These studies also showed that N-WASP is essential for Cdc42-dependent actin assembly and identified the ubiquitous actin regulating complex, the Arp2/3 complex, as the direct link between N-WASP and actin polymerisation (Ma et al. 1998b; Rohatgi et al. 1999).
1.3 WASP and N-WASP
WASP and N-WASP share 50% sequence homology and comprise multiple protein-protein and protein-lipid binding domains as illustrated in Figure 1.9. The N-terminal WASP-homology 1 (WH1) domain is the least conserved region and mediates interactions with regulatory proteins, upstream signals and actin. For example, it has been shown to interact with the WASP interacting protein (WIP) family, calmodulin, PI(4,5)P₂, and actin filaments (Miki et al. 1996; Ramesh et al. 1997; Egile et al. 1999). The poly proline region mediates interactions with SH3-domain proteins such as TOCA1, Nck and Grb2 (Carlier et al. 2000; Rohatgi et al. 2001; Cory et al. 2003; Ho et al. 2004). The basic region has been implicated in lipid binding (Rohatgi et al. 2000) and the Verprolin-homology, Cofilin-homology, acidic (VCA) domain binds the Arp2/3 complex and actin (Ma et al. 1998b; Rohatgi et al. 1999).

![Figure 1.9. The Domain Structure of WASP/N-WASP.](image)

**Figure 1.9. The Domain Structure of WASP/N-WASP.** A simplified representation of the domain structure of WASP/N-WASP (not to scale). The key domains are illustrated as boxes and labelled: WH1 = WASP-homology 1; B = basic region; GBD = G protein binding domain; PP = Polyproline region; VCA = Verprolin-homology, Cofilin-homology, acidic domain. Arrows indicate interactions. Ellipses indicate that a list of binding partners is not exhaustive. Domain boundaries (taken from Rohatgi et al. 2000) are based on human WASP.

1.3.1 The G protein Binding Domain
The G protein binding domain (GBD), including a 14-residue CRIB region, is 70% identical in WASP and N-WASP compared to 50% across the full length protein. CRIB motifs have been identified in a number of signalling proteins and the consensus sequence is ISxPxxFxHxxHVG, where ‘x’ is any amino acid. The importance of particular residues of various CRIB proteins in GTPase binding has been studied (for examples see Thompson et al. 1998; Rudolph et al. 1998; Rudolph et al. 2001; Hemsath et al. 2005) and NMR structures of GTPase-CRIB effector complexes have been solved (for example Abdul-Manan et al. 1999; Mott et al. 1999; Morreale et al. 2000). All of these studies have shown the importance of extra-CRIB residues in GTPase binding and so CRIB motifs are generally described as falling
within larger GBDs. Regions C-terminal to the CRIB motif commonly contribute to GTPase binding (reviewed in Mott and Owen 2015).

The minimal high affinity Cdc42-binding region of WASP has been defined as residues 230-288 (Abdul-Manan et al. 1999; Rudolph et al. 2001). This will be referred to as the GBD hereafter. The smallest WASP fragment found to bind Cdc42, however, comprised only residues 221-257, which includes the CRIB motif (residues 238-251) but is much smaller than the high affinity GBD (Rudolph et al. 1998). Another fragment omitting residues from the N-terminus (residues 235-268) can also bind to Cdc42 (Symons et al. 1996).

The GBD can be divided into regions, with residues 235-257 representing the core GBD (i.e. the minimal region for Cdc42 binding, red in Figure 1.10). Residues N- and C-terminal to the core GBD increase the affinity of the GBD for Cdc42. For example, a construct comprising residues 201-321 bound Cdc42 with a $K_d$ of 77 nM, compared to 470 nM for residues 201-268 (Symons et al. 1996) and residues 230-288 maintain high affinities in the range of 20-30 nM (Abdul-Manan et al. 1999). Therefore, residues 230-235 are considered to be an N-terminal extension (dark blue in Figure 1.10) and 258-288 a C-terminal extension (cyan in Figure 1.10) to the core GBD, both of which are required for high affinity binding. The N-terminal extension includes positive residues that have been implicated in an electrostatic steering mechanism of binding (residues 230-232). Mutation of these residues has a large effect on the rate of association but does not affect the rate of dissociation (Hemsath et al. 2005).

![NMR structure of the high affinity GBD in complex with Cdc42](image)

**Figure 1.10. The WASP/N-WASP GBD.** The sequences of the WASP and N-WASP GBDs are shown. The core GBD is highlighted red, representing the minimal region required for Cdc42 binding based on the data from Symons et al. 1996 and Rudolph et al. 1998. The N-terminal extension is highlighted dark blue and the C-terminal extension cyan. The residues implicated in an electrostatic steering mechanism and the residues of the CRIB motif are labelled.
C-terminal residues of Switch 1 and the entire adjacent β strand (Figure 1.11). It forms the classic intermolecular β-sheet but with irregular bulges (residues 240-241, 243-244 and 247-248) similar to that seen for Cdc42-ACK (Mott et al. 1999). Residues N-terminal to the CRIB (residues 231-237) form a loop that packs against the β2-β3 turn and is adjacent to α5 of Cdc42. Residues C-terminal to the CRIB region (residues 252-274) form a β-hairpin (residues 252-253 and 257-258) and a short α-helix (265-274) (Figure 1.11A). These two structural features are both amphipathic and contact a hydrophobic patch on Cdc42 within the C-terminal portion of Switch 1, Switch 2 and β3.

Figure 1.11. The NMR Structure of the WASP GBD in Complex with Cdc42. A) The NMR ensemble of structures of the Cdc42-GBD complex is shown. This figure was adapted from Figure 1 of (Abdul-Manan et al. 1999). Cdc42 is shown in blue and the WASP GBD in yellow. The structural features of the GBD that are of particular interest are circled and labelled. B) The lowest energy structure is shown as a cartoon representation. Cdc42 is coloured in grey and the GBD coloured as in Figure 1.10: the N-terminal extension is dark blue, the core GBD is red and the C-terminal extension is cyan.

The lowest energy structure in the NMR ensemble is shown in Figure 1.11B with the three regions of the GBD coloured as in Figure 1.10. The β-hairpin (residues 252-253 and 257-258) involves the core GBD and one residue in the C-terminal extension. Therefore, the higher affinity of the C-terminally extended constructs can be explained in part by the completion of this structural feature that would be incomplete.
in the core GBD alone. The C-terminal extension also includes the short \( \alpha \)-helix (265-274) but the remaining residues are poorly defined in the NMR structure, being unstructured and relatively flexible. Residues within the \( \beta \)-hairpin and \( \alpha \)-helix interact with residues of Cdc42 and so the higher affinity of the extended constructs can also be explained in terms of direct contacts (Abdul-Manan et al. 1999).

Mutational studies with Cdc42 have shown some distinctions between binding to WASP and N-WASP. For example, the Y40C mutant of Cdc42 binds to N-WASP but not to WASP (Rohatgi et al. 1999). The importance and biological relevance of this difference has not been explored. Despite this finding, the homology between the WASP and N-WASP GBDs suggests that the N-WASP GBD is likely to bind Cdc42 similarly to WASP.

1.3.2 Activation of WASP and N-WASP

The VCA domain of WASP/N-WASP has been shown to interact with the Arp2/3 complex (Rohatgi et al. 1999) and is essential for N-WASP-induced cytoskeletal reorganisation (Miki and Takenawa 1998). Full length N-WASP, however, only affects actin polymerisation in the same way when Cdc42\( \cdot \)GTP is present (Miki et al. 1998b). This led to the discovery of an elegant mechanism of autoinhibition of WASP/N-WASP involving an intramolecular, autoinhibitory interaction between the GBD and VCA domains that is relieved by Cdc42 (Miki and Takenawa 1998; Kim et al. 2000; Rohatgi et al. 2000).

Residues 151-277 of N-WASP (equivalent to 188-312 of WASP), encompassing the GBD and the upstream basic region, are sufficient to bind to the VCA domain, and the V region is dispensable (Rohatgi et al. 2000). For WASP, the basic region upstream of the GBD is dispensable, as is the A region of the VCA domain (Kim et al. 2000). Instead, residues C-terminal to the CRIB are required. The precise nature of the interaction may, therefore, differ slightly between N-WASP and WASP but the overall mechanism of regulation is comparable.

The VCA-GBD interaction inhibits VCA-Arp2/3 binding and the addition of either Cdc42\( \cdot \)GTP\( \gamma \)S or liposomes containing the signalling lipid, PI(4,5)P\( _2 \), opposes the
VCA-GBD interaction and promotes the VCA-Arp2/3 interaction (Rohatgi et al. 2000). The activation of N-WASP by PI(4,5)P$_2$ and Cdc42 is synergistic but the effect of PI(4,5)P$_2$ is not well understood and may involve allosteric effects upon the downstream GBD. It may also act via membrane-localisation (reviewed in Padrick and Rosen 2010). Conversely, the mechanism of activation by Cdc42 has been characterised in detail following extensive structural analyses (Kim et al. 2000; Buck et al. 2001; Leung and Rosen 2005; Padrick and Rosen 2010).

The VCA-binding region of the WASP GBD (242-310) is largely unfolded in solution but adopts a predominantly α-helical fold in the presence of the VCA domain or organic solvents, as shown using NMR and circular dichroism (CD) (Kim et al. 2000). The existence of the helical conformation in the presence of organic solvents implies that the GBD has a tendency to adopt the helical state but that this state is not well populated in aqueous solution. This autoinhibited fold contrasts with the structure of the GBD when bound to Cdc42 and so the two folds are mutually exclusive (Figure 1.12).

The β-hairpin and α-helix that form upon Cdc42 binding can also be seen in the autoinhibited structure (orange in Figure 1.12B) but there are important differences. For example, in the autoinhibited structure the orientation of α1 relative to the β-hairpin is altered by the insertion of the β2/α1 loop and the α2 helix between them, and α1 forms part of the VCA binding site. Following Cdc42 binding, α1 packs against Cdc42 disrupting the VCA binding site. Furthermore, almost all of the WASP residues that are buried in the hydrophobic core of the autoinhibited conformation are also buried by Cdc42. The hydrophobic core of the VCA-bound GBD is disrupted upon Cdc42 binding, destabilising α2 and 3 and further disrupting the VCA binding site (Kim et al. 2000; Buck et al. 2001; Padrick and Rosen 2010). Therefore, the steric incompatibility between the autoinhibited (VCA bound) and the active (Cdc42 bound) conformations explains the activation of WASP/N-WASP by Cdc42.
The extensive disruption of the autoinhibited conformation by Cdc42 incurs a significant energetic penalty compared to Cdc42 binding to the free, unstructured GBD. This manifests in the remarkably higher affinity of Cdc42 for the free, unfolded GBD (~20 nM) compared to its affinity for a GBD-VCA construct or the GBD in 5 % organic solvent (~11 µM). An equilibrium between the Cdc42 and VCA binding states that would be perturbed in the presence of Cdc42 or the VCA has therefore been suggested (Kim et al. 2000) and expanded into a two state allosteric model of activation (Figure 1.13) (Buck et al. 2004; Leung and Rosen 2005), which is comparable to the classical example of haemoglobin.

Hydrogen/Deuterium exchange NMR experiments showed that the protons of a GBD construct tethered to the C region of the VCA domain were well protected from solvent exchange, especially in the helical regions, indicating a folded structure. This state is termed the ‘tense’ (‘T’) state in analogy to the ‘T’ state of Haemoglobin. The protons of the GBD alone, which does not form the helical fold, were less protected, indicating an extended conformation. This is defined as the ‘relaxed’ (‘R’) state in analogy to the ‘R’ state of Haemoglobin. The ‘R’ state (free GBD in aqueous buffer)

Figure 1.12. The Mutually Exclusive Folds of the GBD when Bound to Cdc42 or the VCA Domain. A) The lowest energy NMR structure of the WASP GBD (residues 230-288) bound to Cdc42 (PDB code: 1CEE). Cdc42 is shown in cyan and the WASP GBD in orange. B) The lowest energy NMR structure of most of the GBD plus a C-terminal extension (residues 341-310) bound to the C region of the VCA domain, which was tethered to the GBD by a covalent linker (PDB code: 1EJ5). The GBD (up to residue 288) is shown in orange, the C-terminal extension to the GBD (GBD+) is shown in green and the ‘C’ region of the VCA domain in magenta. Pertinent structural features are labelled in both A and B.
has a high affinity for Cdc42 (~20 nM) whilst the ‘T’ state (GBD in the presence of
organic solvent or VCA domain) has a lower, micromolar affinity. The addition of
Cdc42 to the ‘T’ state leads to global unfolding, indicated by an increase in H/D
exchange towards the rates recorded for the ‘R’ state. The ‘T’ state is favoured by
VCA binding whilst the ‘R’ state is favoured by Cdc42 binding (Figure 1.13).

![Figure 1.13. The Two-State Model of WASP/N-WASP Activation by Cdc42.](image)
The WASP GBD exists in equilibrium between tense (‘T’) and relaxed (‘R’)
states, shown here as cyan and dark blue boxes respectively. The ‘T’ state is
favoured by the VCA whilst the ‘R’ state is favoured by Cdc42. When the GBD is
in the ‘R’ state, the VCA domain is freed to interact with the Arp2/3 complex.

Whilst the structural and thermodynamic mechanisms underlying Cdc42-dependent
WASP/N-WASP activation were understood in great detail, it soon became clear that
the understanding of the overall pathway of Cdc42/WASP-dependent actin assembly
was far from complete. Firstly, N-WASP is bound to the WASP-interacting protein
(WIP) in vivo (Ramesh et al. 1997; Martinez-Quiles et al. 2001) and, whilst Cdc42 is
capable of binding the N-WASP/WIP complex, the presence of WIP inhibits Cdc42-
dependent activation of N-WASP and downstream filopodia formation. PI(4,5)P₂
overcomes this effect (Martinez-Quiles et al. 2001). Secondly, other activators of N-
WASP have been found, such as the Src homology 3 (SH3) domain proteins TOCA1,
Sortin Nexin 9 (SNX9), Grb2 and Nck (Carlier et al. 2000; Rohatgi et al. 2001; Ho et
al. 2004; Badour et al. 2007) and kinases including Casein Kinase 2 (Suetsugu et al.
The mechanisms of activation by SH3 domain proteins and kinases are not well understood but the SH3 domain proteins may stabilise the ‘R’ state of WASP/N-WASP by interactions with the polyproline region. Thirdly, the activation of WASP/N-WASP is also dependent on effective oligomerisation resulting from clustering of WASP/N-WASP at the membrane, which may result from binding to SH3 domain proteins (Padrick and Rosen 2010).

The equilibrium model of Cdc42-dependent WASP/N-WASP activation is therefore over simplified and ignores the effective oligomerisation of WASP by membrane clustering and activation of WASP by other activators such as SH3 domain proteins and PI(4,5)P₂. Additional equilibria as well as contributions to the Cdc42-dependent equilibrium by other activators are therefore assumed to be involved. Moreover, Cdc42 can itself be represented as an equilibrium between inactive and active states, regulated by GEFs, GAPs and GDIs. The signalling pathways from upstream signals to downstream actin polymerisation should therefore be considered as a complex network of interlinked equilibria between inactive and active and free and bound pathway components. More details of the complexity of WASP/N-WASP regulation can be found in Padrick and Rosen 2010.

1.4 TOCA1

The Arp2/3 complex and N-WASP were previously found to be essential in the pathway of Cdc42-dependent actin polymerisation by fractionation of Xenopus egg extracts (Ma et al. 1998b; Rohatgi et al. 1999) and so purified systems were used to delineate the activation of N-WASP by Cdc42 as described above. These purified systems differed, however, from the situation seen in Xenopus lysates where Cdc42 and PI(4,5)P₂ were not sufficient to activate N-WASP and induce Arp2/3-mediated actin polymerisation. Rather, a third proteinaceous fraction was required. A single protein was purified from this fraction and confirmed to be an additional factor required for Cdc42/N-WASP-dependent actin assembly. This protein was identified as the Transducer of Cdc42-actin assembly protein (TOCA1) (Ho et al. 2004).
1.4.1 Functions of TOCA1

The functional importance of TOCA1 in actin dynamics, and specifically in filopodia formation, has since been demonstrated in a range of in vitro and in vivo studies (Bu et al. 2009; Bu et al. 2010; Lee et al. 2010; Hu et al. 2011; Gallop et al. 2013).

Additional roles for TOCA1 in endocytosis (Tsujita et al. 2006; Bu et al. 2009; Giuliani et al. 2009; Bu et al. 2010), cell motility (Hu et al. 2011), neurite elongation (Kakimoto et al. 2006), nuclear actin dynamics that regulate gene expression (Miyamoto et al. 2011), actin assembly at the tight junction (Van Itallie et al. 2015), and formation of invadopodia in tumour invasion and metastasis (Chander et al. 2013) have also been published. In these cases, TOCA1 functions via the classic Cdc42/N-WASP/Arp2/3 pathway of actin polymerisation but TOCA1 has also been shown to function via the N-WASP homologue, WAVE1, to regulate retrograde endocytic recycling in C. elegans (Bai and Grant 2015).

Thus, TOCA1 has been implicated in various pathways that centre around actin dynamics and membrane dynamics, providing a link between actin polymerisation and membrane trafficking. Moreover, these studies have shown the importance of TOCA1 in Xenopus (Ho et al. 2004; Lee et al. 2010), C. elegans (Giuliani et al. 2009; Bai and Grant 2015), Drosophila (Fricke et al. 2009) and a range of mammalian systems (Kakimoto et al. 2006; Hu et al. 2011; Chander et al. 2013; Van Itallie et al. 2015).

Overall, the physiological roles of TOCA1 are extensive and varied and some of the reported roles remain contentious. Functional studies are confused by redundancy between TOCA homologues (Tsujita et al. 2006; Shimada et al. 2007), by splice variant-specific functions (Van Itallie et al. 2015), by N-WASP-independent activation of the Arp2/3 complex (Snapper et al. 2001; Lee et al. 2010) and by alternative TOCA1-independent pathways of N-WASP activation (Gallop et al. 2013). Nonetheless, it is clear that TOCA1 functions in actin pathways upstream of N-WASP and that it is required for N-WASP activation in many systems (Ho et al. 2004; Bu et al. 2009; Bu et al. 2010; Lee et al. 2010).
1.4.2 Structural Features and Functional Implications

TOCA1 belongs to the *Pombe* Cdc15 Homology (PCH) family, comprising a number of adaptor proteins with limited sequence conservation but shared domain architecture (reviewed in Chitu and Stanley, 2007). All members of the PCH family feature an N-terminal FER/CIP4 homology (FCH) domain adjacent to a BAR domain (known together as the F-BAR domain) and one or two C-terminal SH3 domains. TOCA1 and its homologues, Cdc42-Interacting Protein 4 (CIP4) and Formin Binding Protein 17 (FBP17), also contain a Rho GTPase binding module known as the HR1 domain, which lies between the F-BAR and SH3 domains. The domain structure of the TOCA family is shown in Figure 1.14, along with the structures of the CIP4 F-BAR (Shimada et al. 2007), HR1 (Kobashigawa et al. 2009) and SH3 (Miyamoto *et al.* 2005, unpublished) domains.

![Figure 1.14. The Domain Structure of the TOCA Family.](image)

**Figure 1.14. The Domain Structure of the TOCA Family.** A simplified representation of the domain structures of TOCA family proteins (not to scale). The key domains are illustrated as boxes and labelled: F-BAR = FER/CIP4 Homology (FCH)-BAR; HR1 = Homology Region 1; SH3 = Src Homology 3. Arrows indicate interactions. The crystal structure of the CIP4 F-BAR domain (PDB code: 2EFK) and the NMR structures of the HR1 (PDB code: 2KE4) and SH3 domains (PDB code: 2CT4) of CIP4 are shown below the diagram.

1.4.2.1 The F-BAR Domain

The F-BAR domain is a known dimerisation, membrane binding and membrane deforming domain belonging to the BAR superfamily (Itoh *et al.* 2005; Tsujita *et al.* 2006; Henne *et al.* 2007). BAR domains can bind to negatively charged membranes as monomers or dimers and the $K_d$ of dimerisation is in the micromolar range (Gallop *et al.* 2006; Henne *et al.* 2007). The core BAR domain fold consists of three $\alpha$-helices
that come together to form a homodimeric 6-helix bundle (reviewed in Qualmann et al. 2011) and F-BARs contain one or two additional helices at the N-terminus (~60 amino acids) (Shimada et al. 2007). As dimers, F-BAR domains form crescent shapes with a positively charged concave face responsible for binding to negatively charged membranes (Shimada et al. 2007; Shimada et al. 2010).

The F-BAR domains of the TOCA family members CIP4 and FBP17 have been studied using X-ray crystallography, revealing large hydrophobic interfaces of dimerisation and the classic positively charged concave face. Cryo electron microscopy revealed that they undergo further oligomerisation and filament formation mediated by hydrogen bonding and electrostatic contacts between neighbours (Frost et al. 2008). Mutation of the residues involved in this oligomerisation and filament formation (Arg47, Lys51, Lys66, Asp161, Asn163, Lys273, Phe276 Glu285 and Asp286 in FBP17) or of the positive residues in the concave face that contact the membrane (Arg27, Lys30, Lys33, Lys110, Arg113, Lys114, Arg139, Arg140, Arg146 and Arg150 in FBP17) decreases or abolishes membrane binding and downstream function of TOCA family proteins (Shimada et al. 2007; Frost et al. 2008).

The TOCA1, CIP4 and FBP17 F-BAR domains have been implicated in recruitment of TOCA proteins to PI(4,5)P2-containing membranes (Kamioka et al. 2004; Itoh et al. 2005; Tsujita et al. 2006; Shimada et al. 2007; Bu et al. 2009; Bu et al. 2010). Furthermore, the F-BAR domain is essential for membrane invagination and suppression of neurite elongation (Kakimoto et al. 2006) and mutations in the positively charged, membrane binding face of the TOCA1 F-BAR domain (K33Q/R35Q, R51Q/K52Q, R112Q/K113Q) abolishes the ability of TOCA1 to induce membrane tubulation, vesicle formation and filopodia growth in mammalian cells (Bu et al. 2009; Bu et al. 2010).

1.4.2.2 The SH3 Domain

The SH3 domain is a well known protein-protein interaction domain that recognizes proline rich sequences (Pawson and Gish 1992). The SH3 domain of TOCA1 has many known binding partners, including the classic actin regulatory proteins such as N-WASP (Ho et al. 2004), Dynamin (Itoh et al. 2005) and WAVE1 (Giuliani et al. 2006).
2009; Bai and Grant 2015). A mutation in the SH3 domain that prevents N-WASP binding (W518K) abolishes normal membrane tubulation, vesicle formation and filopodia formation and reduces colocalisation of TOCA1 and N-WASP to endocytic vesicles (Bu et al. 2009; Bu et al. 2010). Moreover, the mutation opposes recruitment of N-WASP/WIP to tight junctions (Van Itallie et al. 2015), inhibits TOCA1 and TOCA2 dependent retrograde endocytic recycling in C. elegans (Bai and Grant 2015) and abolishes the effect of TOCA1 on nuclear actin and gene expression (Miyamoto et al. 2011).

1.4.2.3 The HR1 Domain

The HR1 domain is a known GTPase-binding domain belonging to the helical pair class of GTPase effectors described in section 1.1.8. It is found in a number of Rho family effectors, for example members of the PRK family have three HR1 domains in tandem (HR1a, HR1b and HR1c) (Mukai 2003). The structures of the PRK HR1a and HR1b domains and the CIP4 HR1 domain reveal an antiparallel coiled-coil fold that is shared between the PRK and TOCA families (Maesaki et al. 1999; Modha et al. 2008; Kobashigawa et al. 2009). The TOCA1, CIP4 and FBP17 HR1 domains share a high degree of sequence identity (60-70 %) and so all three TOCA family members are expected to form similar anti-parallel coiled-coils.

The HR1 domains of the PRK family bind to RhoA, RhoB, RhoC and Rac1 (Watanabe et al. 1996; Vincent and Settleman 1997; Hutchinson et al. 2013). TOCA1 does not bind to RhoA or Rac1 but instead binds specifically to Cdc42·GTPγS (Ho et al. 2004). Unsurprisingly, the interaction depends upon the HR1 domain, as shown by a triple mutation in the HR1 domain (M383I/G384S/D385T) abolishing binding, but additional regions of TOCA1 may also contribute to binding (Ho et al. 2004). The triple HR1 domain mutation abolishes N-WASP activation and actin polymerisation in vitro (Ho et al. 2004), membrane tubulation in mammalian cells (Bu et al. 2009; Bu et al. 2010) and retrograde endocytic recycling in C. elegans (Bai and Grant 2015). The HR1-Cdc42 interaction is therefore functionally important, along with the SH3 and F-BAR domain interactions described above.
1.4.3 Activation of N-WASP by TOCA1

The precise roles and functions of TOCA1 in actin dynamics are evidently complex and remain unclear but it is apparent that the interactions mediated by each domain of TOCA1 are important for N-WASP-dependent actin polymerisation (Ho et al. 2004; Bu et al. 2009; Bu et al. 2010). The mechanisms by which TOCA1 activates N-WASP are not yet known but several possible actions have been suggested.

TOCA1 is not required for activation of actin polymerisation in a purified system of Cdc42, N-WASP and the Arp2/3 complex but it is needed in Xenopus extracts. It was therefore proposed that TOCA1 opposes the activity of an N-WASP inhibitor present in the extracts (Ho et al. 2004). A prime candidate for the inhibitor is WIP, which can suppress the activation of N-WASP by Cdc42 in vitro (Martinez-Quiles et al. 2001). Consistent with this idea, TOCA1 recruits WIP to membranes along with N-WASP (Van Itallie et al. 2015) and activation of the N-WASP/WIP complex by Cdc42 depends more strongly on TOCA1 than does N-WASP alone. Having said this, TOCA1 has an activatory effect on N-WASP in the purified system independent of WIP and its direct interactions with N-WASP and membranes are functionally important (Section 1.4.2). TOCA1 may, therefore, also affect N-WASP directly.

The F-BAR-SH3 domain architecture allows for coupling of membrane-based signals with downstream protein-mediated effects and so the TOCA proteins may be responsible for sensing particular membrane curvatures, fluidity or shape and communicating them downstream via N-WASP. FBP17, for example, has been described as a translator of membrane tension into downstream cytoskeletal effects (Tsujita et al. 2006; Zegers and Friedl 2015). Furthermore, F-BAR proteins couple membrane curvature to actin signalling via N-WASP and the deformability of the membrane is important for N-WASP activation (Takano et al. 2008). The TOCA proteins may also sense specific membrane compositions, as specific signalling lipids within particular contexts of membrane curvature result in a switch between different BAR protein pathways (Gallop et al. 2013). Specifically, PI(4,5)P_2 signalling depends on TOCA1 whilst PI(3)P within a curved membrane context signals instead via SNX9. The role of TOCA1 may, therefore, be to promote N-WASP clustering at specific membranes of particular composition and curvature via its F-BAR and SH3...
mediated interactions. The interaction between TOCA1 and Cdc42 may further localise clustering of N-WASP near to activated Cdc42.

Given that other SH3 domain proteins have been suggested to allosterically activate N-WASP (Carlier et al. 2000; Rohatgi et al. 2001; Badour et al. 2007), TOCA1 may also activate N-WASP allosterically. TOCA1 may influence the equilibrium between the ‘T’ and ‘R’ states of N-WASP by binding to the polyproline region that lies between the GBD and VCA. The pathway of WASP/N-WASP-induced actin polymerisation, including the postulated mechanisms of TOCA1 action, are summarised in Figure 1.15. It is not clear where the Cdc42-HR1 interaction fits into the pathway but it may relate to activation of TOCA1 prior to N-WASP activation.

**Figure 1.15. The Pathway of Cdc42/N-WASP-dependent Actin Assembly.** A diagram to illustrate the pathway of N-WASP-induced actin polymerisation in response to Cdc42. The pathway as it was understood before the discovery of TOCA1 is shown in black. The possible involvement and interactions of TOCA1 within the pathway are indicated by red lines.
1.5 Open questions

1.5.1 Specificity and Selectivity of GTPase-Effector Interactions

The Ras superfamily share a high degree of sequence and structural homology, particularly in the nucleotide binding and effector binding regions, and individual GTPases can bind to multiple different effectors belonging to vastly different structural classes. Moreover, seemingly very similar effectors can bind specifically to a single GTPase within a family whilst not possessing cross-specificity for other, very similar GTPases. Therefore, how small GTPases bind selectively and specifically to a range of effectors, each with different sequence and structural features, remains a prominent question in GTPase research.

The Rho-interactive HR1 domains provide an interesting example of homologous effectors selectively binding to specific members of the Rho family that are also similar. For example, the HR1a, HR1b and HR1c domains of PRK proteins share a moderate degree of sequence similarity (for example HR1a and HR1b of PRK1 share 25% sequence identity) and high structural similarity and their GTPase binding partners are even more conserved (for example RhoA and RhoC share 92% sequence identity and RhoA and Rac1 share 57%). Nonetheless, each domain shows specificity for a different Rho family member, or subset of members.

For example, the HR1b domain of PRK1 binds only to Rac1 with high affinity (Owen et al. 2003). The HR1a domain of PRK1 binds to RhoA, RhoB, RhoC and Rac1, but its affinity for each Rho family member is significantly different (Hutchinson et al. 2013). Moreover, the affinities of the PRK2 and PRK3 HR1 domains for the Rho members are much lower despite high sequence similarity with the PRK1 HR1 domains (~40-70%) (Hutchinson et al. 2013). There are no obvious sequence specific determinants of selectivity or affinity as all but two of the energetically important HR1-Rho contacts are fully conserved between the PRKs and the other two are conservative changes from Leucine to Isoleucine. There are also no obvious structural differences but subtle differences in their stabilities and the rigidity of the coiled-coils may contribute to their different specificities and affinities (Hutchinson et al. 2013).
The HR1 domains of the TOCA family are 20% identical to the PRK HR1 domains and share the classic coiled-coil fold (Kobashigawa et al. 2009). However, the binding specificities of the PRK and TOCA protein families do not appear to overlap, as TOCA1 has been shown to bind to Cdc42 but not RhoA or Rac1 (Ho et al. 2004) and CIP4 interacts with Cdc42 and TC10 (Aspenström 1997; Chang et al. 2002). The TOCA1-Cdc42 and CIP4-Cdc42 interactions represent the first Cdc42-specific HR1 domain interactions to be found and so biochemical characterization of these interactions is of interest. Structural characterisation of a novel Cdc42-specific HR1 domain interaction will also contribute to the growing body of information relating to Rho-HR1 complexes.

1.5.2 The Cdc42-TOCA1 Interaction

Despite the deleterious effect of the M383I/G384S/D385T mutation on TOCA1-Cdc42 binding and downstream TOCA1 function, it has not been formally shown that the HR1 domain is sufficient for Cdc42-TOCA1 binding. Moreover, it has been suggested that other regions of TOCA1 may contribute to the interaction (Ho et al. 2004). Preliminary work from a previous project has suggested that the TOCA1 HR1 domain binds Cdc42 with a modest, micromolar affinity (Joanna R Watson, unpublished), far lower than the nanomolar affinities of other Rho-binding HR1 domains (Owen et al. 2003; Hutchinson et al. 2013). This further indicates that the TOCA1 HR1 domain may be insufficient for maximal TOCA1-Cdc42 binding. This previous binding study was, however, incomplete and there were stability issues with the HR1 domain construct used.

The concentration of TOCA1 in *Xenopus* extracts has been determined to be 5-10 nM based on western blotting and kinetic measurements and, in agreement with this, the amount of TOCA1 required to rescue actin polymerisation in TOCA1-depleted *Xenopus* extracts is 5 nM (Ho et al. 2004). This is consistent with a nanomolar affinity interaction, comparable to the PRK-Rho interactions. It is therefore necessary to determine whether the TOCA1 HR1 domain is sufficient for Cdc42 binding or whether other regions of TOCA1 contribute to binding. Should other regions of TOCA1 be required for tight binding, it would then be interesting to discover whether
their effects on binding are direct or whether they affect binding indirectly by modulating the structure or stability of the HR1 domain.

1.5.3 Roles and Regulation of TOCA1

The molecular basis of the activation for WASP/N-WASP by Cdc42 is well understood thanks to a wealth of biophysical and structural studies that have elucidated an intricate mechanism of autoregulation and disinhibition by Cdc42. This over-simplified pathway of Cdc42/WASP-dependent actin polymerisation has evidently suffered from a lack of wider biological insight prior to the discovery of additional regulatory factors such as TOCA1. In contrast, the understanding of TOCA1 has suffered from the reverse problem: there is a wealth of functional, in vivo information about the TOCA family, but biophysical and structural insight is scarce.

Functional studies into TOCA1 have primarily been carried out in complex systems such as mammalian cell lines, Xenopus extracts, C. elegans and mouse models. These have been confounded by redundancy in protein families, compensatory biological pathways, tissue and organism specific variations and more. The understanding of the roles and regulation of TOCA1 is limited by a lack of in vitro studies aimed at providing biophysical and structural insight into specific steps of the pathway.

1.5.3.1 Autoregulation of TOCA1

Given that TOCA1 activity is influenced by Cdc42 and PI(4,5)P₂, TOCA1 may be regulated in a similar way to WASP/N-WASP. For example, binding of Cdc42 to TOCA1 may disrupt an autoinhibitory interaction between any of the TOCA1 domains. Regulatory F-BAR-SH3 interactions have been reported for other PCH family members such as Syndapin 1 (Wang et al. 2009; Rao et al. 2010) and so this is a tempting candidate for TOCA regulation, although interactions between the HR1 and SH3 and the HR1 and F-BAR domains have also been suggested.

F-BAR domains dimerise via a predominantly hydrophobic interface and bind to membranes via a positively charged concave face. Dimers also form lateral contacts between dimers via salt bridges and hydrophobic contacts leading to higher order oligomerisation, clustering on membranes and subsequent membrane tubulation (Frost
et al. 2008). The F-BAR-SH3 interaction identified in Syndapin 1 is also predominantly electrostatic, involving positively charged residues of the F-BAR domain (including Lys141, Lys145 and Lys148) and negatively charged residues of the SH3 domain (Gln396, Glu397 and Glu400). The interaction does not abolish membrane binding but it significantly inhibits F-BAR-mediated membrane tubulation, presumably by opposing higher order oligomerisation and clustering (Rao et al. 2010). The F-BAR and proline-rich protein binding activities of the SH3 domain are mutually exclusive and so the F-BAR-SH3 interaction is abrogated upon addition of proline rich proteins such as Dynamin. Dynamin binding therefore releases the F-BAR domain and promotes membrane tubulation (Rao et al. 2010).

A similar mechanism of regulation may exist in TOCA1, in which N-WASP, or indeed Dynamin, binds to the SH3 domain and relieves the autoinhibitory interaction, freeing all three domains of TOCA1 to interact with their binding partners. The Cdc42-HR1 interaction could also affect the suggested autoinhibited conformation allosterically, leading to disruption of the F-BAR-SH3 interaction.

Regulatory interactions between BAR domains and Rho binding regions have also been reported (Krugmann et al. 2001) and it is easier to envisage how Cdc42 could relieve an autoinhibitory interaction involving the HR1 domain. Furthermore, mutation of the HR1 domain has been shown to affect F-BAR-dependent membrane tubulation in mammalian cells, suggesting a role for the Cdc42-HR1 interaction in regulation of the F-BAR activity of TOCA1 (Bu et al. 2009). An HR1-F-BAR interaction is, therefore, also a prime candidate for an autoregulatory interaction.

To complete the possible combinations of domains implicated in autoregulation, protection of the linker between the HR1 and SH3 domains from proteolysis was observed, suggesting that an HR1-SH3 interaction may exist (Dr J Peterson, Fox Chase, personal communication). The potential mechanisms of TOCA1 autoregulation are summarised in Figure 1.16A. The interactions may not be mutually exclusive but rather the overall structure of TOCA1 may involve multiple interdomain interactions as illustrated in Figure 1.16B. This has been observed for the I-BAR protein IRSp53, in which the I-BAR domain interacts with the central region of the
protein masking a partial CRIB responsible for Rac1 and Cdc42 binding and also masking the SH3 domain (Krugmann et al. 2001).

1.5.3.2 Cdc42, TOCA1 and N-WASP

It is well documented that individual GTPases can recognise multiple effectors of different structural classes but it is not known how small GTPases select between their different effectors. It is easy to comprehend how this may work by GEF, GAP, GDI and lipid-mediated localisation of GTPases near to particular effectors, but only when considering effectors that are active in distinct pathways occurring at different subcellular locations. The identification of two vastly different effectors of Cdc42 from entirely dissimilar structural classes that are so closely linked within the same functional pathway is, therefore, particularly interesting.

Figure 1.16. Possible Autoinhibitory Interactions within TOCA1. Diagrams to illustrate the postulated intramolecular interactions that may regulate TOCA1. A) The F-BAR-SH3, HR1-SH3 and F-BAR-HR1 interactions are illustrated. B) The interactions shown in A may not be mutually exclusive and so some possible combinations of interactions are illustrated.
The question remains as to whether Cdc42 binds to one effector before the other or whether a single molecule of Cdc42 can bind to both the WASP/N-WASP GBD and the TOCA1 HR1 simultaneously. Should the former be true, it is important to understand how Cdc42 switches between the TOCA1 and N-WASP binding modes and in which order the binding occurs. Should the latter be true, it would also be of general interest in small GTPase signalling as no examples of two effectors bound to a single small GTPase have been found to date.

There is some evidence for a ternary complex between Cdc42, N-WASP and TOCA1 (Bu et al. 2009) but it remains unclear whether the complex involves simultaneous interactions of the two effectors with a single molecule of Cdc42. Moreover, the vastly different structures of the N-WASP GBD compared with the TOCA HR1 domain indicate that a ternary complex may be possible. Demonstration of a ternary complex could have huge implications for other small GTPases known to bind both β-sheet and α-helical effectors and also for Cdc42/TOCA1/N-WASP signalling, as the roles and relative contributions of the Cdc42, TOCA1 and N-WASP interactions to N-WASP activation are not yet well understood.

1.5.4 The Role of the Membrane in GTPase Signalling

Despite the centrality of the C-terminal sequences, lipid modification and membrane binding to the biological functions of the Ras superfamily (section 1.1.6), the majority of in vitro studies have focused on unmodified, C-terminally truncated forms in the absence of membranes. Whilst in vitro studies on truncated GTPases have been necessary for understanding the GTPase cycle and effector interactions, it is becoming increasingly apparent that these studies do not provide a complete picture.

Although membrane anchoring of Ras does not appear to affect the overall structure of the G domain (Güldenhaupt et al. 2008), a number of observations suggest that the C-terminal regions, lipid modification and membrane anchoring of some GTPases affects their interaction with nucleotides, with effectors and with each other. Firstly, the polybasic C-terminal region of the Rho family member, Rac1, is required for binding to several regulators and effectors (reviewed in Lam and Hordijk 2013).
Secondly, the C-terminal region of the Ras family member, RalA binds to Calmodulin in a nucleotide-dependent manner (A. Shafiq and H.R. Mott, unpublished). Thirdly, an interaction between the C-terminal region of Cdc42 and the membrane lipid PI(4,5)P₂ affects nucleotide binding (Zheng et al. 1996). Finally, the C-terminal regions of Cdc42, Rac1, RhoA and Rac2 have been connected to functionally important dimerisation (Zhang and Zheng 1998; Zhang et al. 2001).

Hence, the biological limitations of in vitro studies of truncated GTPases in solution are becoming increasingly recognized. Research into Cdc42, TOCA1 and N-WASP, and indeed into all small GTPases, is hindered by technical limitations that prevent structural studies in the presence of membranes and methods to study these proteins in the presence of membranes are therefore needed. Given the inherent importance of dynamics in GTPase signalling (section 1.1.2), solution NMR with membrane mimetics would provide the most useful information.
1.6 Aims

The project described herein aimed to provide much needed biophysical and structural insight into the roles and regulation of TOCA1 in the pathways of Cdc42/N-WASP-dependent actin polymerisation and into Rho-HR1 interactions. Specifically, the initial aims were as follows:

1. To measure the affinity of the Cdc42-TOCA1 interaction, determine the minimal region of TOCA1 required for maximal affinity Cdc42 binding and to investigate the possible mechanisms of TOCA1 autoregulation

2. To solve the solution NMR structure of the minimal Cdc42 binding region of TOCA1

3. To characterise the Cdc42-TOCA1 interaction structurally in order to address questions surrounding GTPase-effector selectivity and specificity and to determine whether TOCA1 and N-WASP can bind to a single molecule of Cdc42 simultaneously

The pathway of actin polymerisation described in the preceding sections occurs at membranes and many of the central proteins involved bind directly to membranes. The results pertaining to these three initial aims, presented in Chapters 3-5, were therefore limited by the absence of membranes. For this reason the project evolved to pursue methods to lipid modify and membrane anchor Cdc42 in order to study the effects of membrane anchoring on the switch regions, on nucleotide binding and hydrolysis and on effector binding. The key aims and results pertaining to membrane binding are presented in Chapter 6.
1.7 Experimental Techniques

A number of biochemical and biophysical techniques were used to address these aims, the results from which are presented herein. These techniques included Scintillation Proximity Assays (SPA), Analytical UltraCentrifugation (AUC), liposome sedimentation assays and Nuclear Magnetic Resonance (NMR) Spectroscopy amongst others. With the exception of NMR, sufficient information is given on each technique and experiment within the Results and Discussion chapters (Chapters 3-6) such that the results presented can be fully understood. It is necessary, however, to include a short introduction to the principles of NMR spectroscopy before moving on to the results.

1.7.1 An Introduction to NMR Spectroscopy

NMR spectroscopy is a powerful tool for studying the structures of macromolecules and more than 11,500 solution NMR structures of proteins have been deposited in the PDB to date. It can be used to produce de novo structures of proteins up to ~30 kDa using the conventional methods that will be described in Chapter 4 but more recent advances in protein labelling schemes (reviewed in Zhang and van Ingen 2016), NMR pulse sequences and data processing have allowed for much larger proteins to be studied (50-1000 kDa).

NMR can be used to study proteins in solution and so it offers a significant advantage over alternative techniques in maintaining natural protein dynamics and flexibility. Moreover, it is uniquely appropriate for studying a wide range of protein motions including the conformational and dynamic changes that are central to GTPase function. NMR can also be used to map protein-protein interactions (as described in Chapter 5) and report on the timescales of internal dynamics, conformational exchange and ligand binding.

The key concepts of NMR spectroscopy will be briefly introduced so that the reader may understand the NMR data and analysis that is presented in the Results and Discussion chapters. A more in depth description of NMR theory is beyond the scope of this thesis but can be found elsewhere (Cavanagh et al. 1998; Keeler 2005).
NMR relies upon a quantum mechanical property of atomic nuclei known as the nuclear spin angular momentum, which is described using the nuclear spin quantum number, I. This is commonly referred to more simply as the ‘nuclear spin’ but it does not arise from spinning of a nucleus. Rather, it is an intrinsic property of the nucleus like mass or charge. Nuclei with \( I \neq 0 \) are said to be ‘spin active’ and, for protein NMR, it is the nuclei with \( I = \frac{1}{2} \) that are the most useful. Nuclei present in proteins with \( I = \frac{1}{2} \) are \(^1\text{H}, ^{15}\text{N}\) and \(^{13}\text{C}\). \(^1\text{H}\) is naturally abundant (99.99%), whilst \(^{13}\text{C}\) (~1% natural abundance) and \(^{15}\text{N}\) (~0.4%) can be incorporated during protein expression in order to increase the abundance to >98%.

The nuclear spin angular momentum of a nucleus with \( I \neq 0 \) gives rise to a magnetic moment. The z-component of the magnetic moment (\( \mu \)) is simply the z-component of the angular momentum (\( \hat{I}_z \)) multiplied by the gyromagnetic ratio (\( \gamma \)), which is a fundamental property of the nucleus:

\[
\mu = \gamma \hat{I}_z
\]  
Equation 1.1

The interaction between this magnetic moment and an applied magnetic field (\( B_0 \)) underlies NMR spectroscopy.

The energy of the interaction depends upon the angle (\( \theta \)) between the magnetic moment and the applied field, such that the energy is lowest when the two are parallel (\( \theta = 0 \)) and highest when antiparallel (\( \theta = \pi \) radians). Therefore, there is an energetic advantage for a nuclear magnetic moment to align with the applied field and over time (seconds) the lower energy orientations become preferentially populated. The energetic preference is, however, opposed by thermal motion and so the net alignment that arises is very slight (1 in \( 10^5 \) spins). Nonetheless, the slight alignment is such that the sample acquires its own magnetic moment, known as the ‘bulk magnetisation’ (or equilibrium magnetisation). This is in the same direction as the applied magnetic field, i.e. along the z-axis by convention. The extent of the spin alignment (i.e. the bulk magnetisation) increases with increasing field strength and, for protein NMR, magnetic fields of more than 500 MHz are needed to achieve sufficient magnetisation.

By definition, the equilibrium magnetisation does not change in size or direction over time (Figure 1.17A). However, if this magnetic moment is perturbed away from the z-
axis towards the xy-plane, the vector precesses about the z-axis (Figure 1.17B). The frequency of precession depends upon the gyromagnetic ratio ($\gamma$) and the applied field ($B_0$) and is known as the Larmor frequency ($\omega_0$):

$$\omega_0 = -\gamma B_0 \quad \text{Equation 1.2}$$

The direction of this precession is negative (y to x using the right-handed axes system shown in Figure 1.17) if the gyromagnetic ratio is positive and vice versa. The NMR signal comes from the precession of this magnetisation vector, which can be detected using a coil of wire placed around the sample along the x-axis. When the magnetisation vector ‘cuts’ the coil, a current is induced and recorded as the free induction decay (FID) (Figure 1.17C). The signal decays due to relaxation of the magnetisation vector back to equilibrium, a concept which will not be discussed further here.

**Figure 1.17. Magnetisation and the NMR Signal.** A) The equilibrium magnetisation that builds up after a few seconds in a magnetic field is, by convention, along the z-axis. The direction and magnitude of the equilibrium magnetisation vector does not change over time. B) Following perturbation of the magnetisation vector away from the z-axis into the xy-plane, the vector precesses about the z-axis at the Larmor frequency ($\omega_0 = -\gamma B_0$). The precession shown here is for a nucleus with a positive gyromagnetic ratio such as a proton. C) As the vector ‘cuts’ the receiver coil, which is placed around the x-axis, an electrical current is produced. This current, known as the free induction decay (FID), constitutes the NMR signal.

The precise magnetic field experienced by a nucleus depends on its chemical environment and so each nucleus within a protein effectively experiences a different ‘$B_0$’. Therefore, the Larmor frequency of each nucleus is also different. This can be
described mathematically by incorporating an additional term into the equation for the
Larmor precession known as the ‘chemical shift’:

$$\omega_0 = -\gamma B_0 \times \text{‘chemical shift’} \quad \text{Equation 1.3}$$

The FID, which is converted to a spectrum of frequencies using the Fourier Transform
(described in Keeler 2005), contains the precession frequency of each spin active
nucleus in the protein. Of course, the NMR experiment does not detect a single spin
but rather the NMR sample contains in the region of $10^{17}$ molecules of the protein and
so $10^{17}$ of each spin. The frequency of each spin is therefore detected as a population
average.

These frequencies are given relative to the receiver reference frequency, which is
typically placed by the user in the centre of the frequencies of interest. For example, if
one was interested in peaks between 0 and 1000 Hz, the receiver would be placed at
500 Hz resulting in a spectrum going between -500 and +500 Hz. Further to this, the
frequencies ($v$) are normalised relative to the frequency of a reference compound ($v_{ref}$)
and quoted as chemical shifts ($\delta$) in parts per millions (ppm), in order to remove their
dependence on the applied field:

$$\delta \text{(ppm)} = 10^6 \times (v - v_{ref})/v_{ref} \quad \text{Equation 1.4}$$

The final NMR spectrum contains a peak for each spin active nucleus in the protein at
its Larmor frequency, given as a chemical shift in ppm.

1.7.1.2 The Basic NMR Experiment

The rotation of the equilibrium magnetization away from the z-axis and into the xy-
plane is fundamental to the NMR signal. Switching off the applied field ($B_0$) along z
and replacing it with a field along x or y is the most obvious means of achieving this,
but it is simply not possible to switch the powerful superconducting magnets used to
establish the bulk magnetisation off and on in this way. However, the applied field can
be overcome by an additional small, oscillating magnetic field ($B_1$) applied along the
x-axis (or y-axis), known as the radio frequency (RF) field (or pulse).

The $B_1$ field is generated using the detector coil and is set to linearly oscillate at or
close to the Larmor frequency. It can be considered as two counter-rotating fields ($B_1^+$
and $B_1^-$), rotating at a frequency of $\omega_1$ (Figure 1.18A). The two components of $B_1$ are
equal in magnitude and so the sum of their y-components is always zero throughout their rotation, whilst the sum of their x-components goes between $\pm 2B_1$. The overall effect is a field that oscillates linearly along x (Figure 1.18B). The rotating field significantly interacts with the Larmor precession of a nucleus if it is rotating in the same direction but has no significant interaction if rotating in the opposite direction. This means that one of the two components can be ignored. For example, for a nucleus with a negative gyromagnetic ratio such as $^{15}$N, only the $B_1^+$ component of the RF field is rotating with the same sense as the Larmor frequency (Equation 1.2) and so the $B_1^-$ component can be ignored. Now, by observing $B_1^+$ from a ‘rotating frame’ that rotates with the same sense and frequency as $B_1^+$, $B_1^+$ appears static (Figure 1.18C).

**Figure 1.18. The RF Field and the Rotating Frame.** A) A linearly oscillating magnetic field can be considered as two components ($B_1^-$ and $B_1^+$) rotating in opposite directions as shown here. B) The y-component of each always sums to zero whilst the x-components sum to give an overall field that oscillates along x. C) The rotating frame is shown, in which the xy-plane is rotated in the same direction and at the same frequency as $B_1^+$, the $B_1^+ \text{ field therefore appears static.}$ For a nucleus with a negative gyromagnetic ratio, the Larmor precession is in the opposite direction to $B_1^-$ and so $B_1^-$ can be ignored.
Observing the Larmor precession in this same rotating frame causes an apparent change to the Larmor frequency. The apparent Larmor frequency in the rotating frame is known as the offset ($\Omega$) and is given by:

$$\Omega = \omega_0 - \omega_{\text{rotating frame}}$$  \hspace{1cm} \text{Equation 1.5}

The Larmor frequency is related to the applied magnetic field (Equation 1.2) and so, in the same way, the offset is related to the apparent field experienced in the rotating frame ($\Delta B$):

$$\Omega = -\gamma \Delta B$$  \hspace{1cm} \text{Equation 1.6}

This proportional relationship between the apparent field (also known as the reduced field) and the offset allows the apparent field to be reduced to zero simply by setting the frequency of the rotating frame ($\omega_{\text{rotating frame}}$) equal to the Larmor frequency ($\omega_0$).

The rotating frame therefore gives us the power to apply an oscillating magnetic field along $x$ or $y$ that appears stationary as well as effectively converting $B_0$ into a reduced field ($\Delta B$) that is very small. The field now experienced by each spin is a vector sum of the apparent field ($\Delta B$) and the RF field ($B_1$) and is known as the effective field ($B_{\text{eff}}$). When the frequency of the RF field is selected to match the Larmor frequency exactly, $\Omega = 0$ (Equation 1.5), $\Delta B = 0$ (Equation 1.6) and so $B_{\text{eff}} = B_1$. This is called an ‘on-resonance’ pulse and it causes the magnetisation to rotate about the axis to which it was applied. The angle through which the magnetisation vector rotates (flip angle, $\beta$) depends on the frequency of the RF pulse and pulse length ($t_{\text{app}}$):

$$\beta = \omega_1 t_{\text{app}}$$  \hspace{1cm} \text{Equation 1.7}

When the frequency of the RF field is very different to the Larmor frequency, $\Omega \approx \omega_0$ and so $B_{\text{eff}}$ is dominated by $\Delta B$. This is called an ‘off-resonance’ pulse and it essentially has no effect on the magnetisation. The Larmor precession of $^1\text{H}$, $^{15}\text{N}$ and $^{13}\text{C}$ differ significantly due to their different gyromagnetic ratios (Equation 1.2) and so a pulse that is on-resonance with protons will not affect the other nuclei.

The basic ‘pulse-acquire’ NMR experiment consists of a $90^\circ$ pulse on the $x$-axis, on-resonance with the nucleus of interest, followed by signal acquisition (Figure 1.19A). In terms of the ‘vector model’, the $90^\circ$ pulse on $x$ simply rotates the magnetisation vector from $z$ onto $-y$ and the magnetisation then precesses in the $xy$-plane during acquisition of the FID (Figure 1.19B). The addition of a $180^\circ$ pulse before acquisition turns the pulse-acquire into the ‘spin-echo’ (Figure 1.19D). In the spin-echo, the
equilibrium magnetization is rotated into the transverse plane as in the pulse-acquire and the spins precess within the xy-plane for time, \( \tau \). The spins do not precess at identical frequencies and so after \( \tau \), the magnetisation of each spin will have reached a different point along a circular path around \( z \) (Figure 1.19E). When a 180° pulse is applied, all of the spins are rotated to be exactly opposite where they were before the pulse. The spins continue to precess at the same frequency and in the same direction as before and so after another period of exactly \( \tau \), they will be aligned on the y axis. The chemical shift is said to be ‘refocused’. The utility of this experiment will become apparent in section 1.7.1.3.
Figure 1.19. The Pulse-Acquire and the Spin Echo. A) The basic ‘pulse-acquire’ experiment. The RF line contains the information on the RF pulses and the acquisition (Acq) line illustrates when the signal is recorded. The acquired FID is processed using the Fourier Transform to produce a spectrum with a peak for each spin active nucleus at its Larmor frequency, expressed as the chemical shift in ppm. B) A vector model diagram illustrating the changes to the bulk magnetisation that occur at points (1), (2) and (3) during the experiment shown in A. C) A description of the same experiment using product operators. D) A diagram of the spin-echo pulse sequence. E) A vector model diagram showing how the magnetisation changes after the 180° pulse and final period of free precession (τ). F) A description of the same experiment using product operators.
These simple experiments can also be described using the product operator approach (Figure 1.19C and F), which has been covered elsewhere (Cavanagh et al. 1998; Keeler 2005) and will be only briefly described here. This approach uses quantum mechanics to describe in more detail how the magnetisation changes during pulses and free precession and is particularly important for understanding systems with two or more coupled spins (discussed in section 1.7.1.3). In the interest of brevity, the quantum mechanical derivation of the mathematical rules needed to master this approach will not be described but rather the necessary rules will simply be stated.

The magnetisation vector of a single spin is described by three operators, $\hat{I}_x$, $\hat{I}_y$ and $\hat{I}_z$. The ways in which these operators change over time under certain conditions can be described using a quantum mechanical assessment of the Hamiltonian operators ($\hat{H}$) that act upon the magnetisation operators. For example, the Hamiltonian operator for a single spin when an RF pulse is applied about the x-axis is

$$\hat{H}_{\text{pulse}} = \omega_1 \hat{I}_x$$  \hspace{1cm} \text{Equation 1.8}

and quantum mechanics provides a simple rule to describe how the magnetisation operator changes under this Hamiltonian:

Original operator $\rightarrow$ \cos($\omega_1 \tau$)(Original operator) + \sin($\omega_1 \tau$)(New operator)

The original operator at the start of an experiment is, by convention, $\hat{I}_z$ and the new operator can be deduced using the axis system shown in Figure 1.19B. The rotation caused by the pulse is anti-clockwise about the x-axis, from z towards $-$y, and so the new operator is $-$\hat{I}_y. During an RF pulse, therefore, $\hat{I}_z$ becomes $\cos(\omega_1 \tau)\hat{I}_z$ - $\sin(\omega_1 \tau)\hat{I}_y$. If the flip angle of the pulse is set to 90°, $\cos(90^\circ)$ is zero and $\sin(90^\circ)$ is 1, and so the $\hat{I}_z$ magnetisation simply becomes $-$\hat{I}_y.

The Hamiltonian operator for a single spin during free precession in the rotating frame is:

$$\hat{H}_{\text{free precession}} = \Omega \hat{I}_z$$  \hspace{1cm} \text{Equation 1.9}

and so the simple rule for describing the change in magnetisation during free precession is essentially the same as during a pulse but with $\Omega \tau$ in place of $\omega_1 \tau$:

Original operator $\rightarrow$ \cos($\Omega \tau$)(Original operator) + \sin($\Omega \tau$)(New operator)

The precession is anti-clockwise about the z-axis (i.e. from $-$y towards x) and the new operator can again be deduced from the same axes system. For example, after time, $\tau$, $-$\hat{I}_y becomes -\cos($\Omega \tau$)$\hat{I}_y$ + $\sin($\Omega \tau$)$\hat{I}_x$. This is consistent with what we see from the
vector model shown in Figure 1.19B as it describes the appearance of an x-component as the magnetisation rotates away from the -y-axis and towards the x-axis in a circular motion. The product operators describing how the magnetisation changes during the pulse-acquire and spin-echo experiments are given in Figure 1.19C and F.

### 1.7.1.3 Multi-Dimensional NMR

The real treasure of the product operator approach is seen when dealing with two or more spin-active nuclei that are interacting via scalar coupling. Put simply, scalar coupling is an interaction between the magnetic moments of two spin-active nuclei that are bonded together. The size of the interaction depends on the gyromagnetic ratios of each nuclei, the number of bonds between them, the bond lengths and the bond angles and is described by the scalar coupling constant, $J$. Scalar coupling can, therefore, report upon connectivity of nuclei within a protein and bond angles, which depend on protein secondary structure. Most importantly, the coupling allows for the deliberate and selective transfer of magnetisation between nuclei.

For the single spin system, $\hat{I}_z$, $\hat{I}_x$ and $\hat{I}_y$ are the only operators needed to describe the magnetisation vector and the Hamiltonian for free precession has just one term (Equation 1.9). For the two spin system there are sixteen operators, fifteen of which are useful for describing pulse sequences:

<table>
<thead>
<tr>
<th>Magnetisation</th>
<th>Operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>z-magnetisation on one spin</td>
<td>$\hat{I}_z$, $\hat{I}_2z$</td>
</tr>
<tr>
<td>z-magnetisation on two spins</td>
<td>$2\hat{I}_1\hat{I}_2$</td>
</tr>
<tr>
<td>In-phase transverse</td>
<td>$\hat{I}<em>{1x}, \hat{I}</em>{2x}, \hat{I}<em>{1y}, \hat{I}</em>{2y}$</td>
</tr>
<tr>
<td>Anti-phase transverse</td>
<td>$2\hat{I}<em>{1x}\hat{I}</em>{2x}, 2\hat{I}<em>{1y}\hat{I}</em>{2y}, 2\hat{I}<em>{1z}\hat{I}</em>{2z}$</td>
</tr>
<tr>
<td>Multiple-quantum coherance</td>
<td>$2\hat{I}<em>{1x}\hat{I}</em>{2x}, 2\hat{I}<em>{1x}\hat{I}</em>{2y}, 2\hat{I}<em>{1y}\hat{I}</em>{2x}, 2\hat{I}<em>{1y}\hat{I}</em>{2y}$</td>
</tr>
</tbody>
</table>

The Hamiltonian now has three terms:

$$\hat{H}_{\text{two spins}} = \Omega_1 \hat{I}_{1z} + \Omega_2 \hat{I}_{2z} + 2\pi J_{12} \hat{I}_1 \hat{I}_2$$  \hspace{1cm} \text{Equation 1.10}

where $\Omega_1$ and $\Omega_2$ are the offsets of each spin and $J_{12}$ is the scalar coupling constant between the two spins. The three terms of this Hamiltonian can be treated separately and in any order. The first two terms describe the changes to the chemical shift of each nucleus during free precession and can be treated as in the one spin case described above. The third term describes the scalar coupling, which is active during periods of free precession.
The effect of coupling on the magnetisation can be described using:

Original operator $\rightarrow \cos(\pi J_{12}\tau)(\text{Original operator}) + \sin(\pi J_{12}\tau)(\text{New operator})$

but this time the axes in Figure 1.19B cannot predict the new operator. Rather, it is described by the axis system shown in Figure 1.20A. For example, $-\hat{I}_{1y}$ becomes

$-\cos(\pi J_{12}\tau)\hat{I}_{1y} + \sin(\pi J_{12}\tau)2\hat{I}_{1z}\hat{I}_{2z}$ after a period of free precession, $\tau$. By setting $\tau$ such that $\cos(\pi J_{12}\tau)$ becomes zero and $\sin(\pi J_{12}\tau)$ becomes 1, the coupling results in complete loss of $-\hat{I}_{1y}$ magnetisation and gain of $2\hat{I}_{1z}\hat{I}_{2z}$. The overall result of the free precession with scalar coupling active is therefore to produce a product operator term featuring both spins. The operator, $2\hat{I}_{1z}\hat{I}_{2z}$, is known as an anti-phase operator as the magnetisation of spin 1 is on x (i.e. transverse) and the magnetisation of spin 2 is on z (i.e. not transverse). 90° pulses can then be used to convert this term into $2\hat{I}_{1z}\hat{I}_{2y}$ and hence the transverse magnetisation is transferred from spin 1 to spin 2.

The use of this is best illustrated using an extension of the spin-echo experiment that was described in section 1.7.1.2, known as the Insensitive Nuclei Enhanced by Polarisation Transfer (INEPT) sequence. The gyromagnetic ratio of protons is 10x larger than that of $^{15}$N and, as the magnetic moment of a nucleus and the energetic preference for it to align with the applied magnetic field is proportional to the gyromagnetic ratio (Equation 1.1), the bulk magnetisation of protons is 10x larger at a given field strength. Hence, protons give the largest NMR signal. The INEPT sequence is therefore fundamental in the study of proteins by NMR as it allows the signal from an insensitive nucleus such as $^{15}$N or $^{13}$C to be enhanced using the more sensitive proton signal. Moreover, it forms the basis of a number of 2D and 3D NMR experiments that rely on directional magnetisation transfer such as those used in the protein backbone assignment described in section 4.2.2.1.

The pulse sequence of the INEPT is shown in Figure 1.20B and the path of the magnetisation throughout the experiment is shown below it using simplified product operators. For clarity, $\hat{I}_{1}$ is referred to as $H$ (proton) and $\hat{I}_{2}$ as $N$ ($^{15}$N). To begin, a 90° (x) pulse on the protons rotates the equilibrium magnetisation from $H_z$ to $-H_y$. The proton chemical shift is refocused during $2\tau_1$ due to the 180° pulse as shown in the description of the spin-echo in section 1.7.1.2 but the simultaneous 180° pulse on the $^{15}$N causes the coupling to remain active over $2\tau_1$. When $\tau_1$ is set to $\frac{1}{4}J_{NH}$, $\cos(2\pi J_{NHT})$ becomes zero and $\sin(2\pi J_{NHT})$ becomes 1, and the resulting magnetisation

53
is described by $-2H_xN_z$. A 90° y pulse on the protons changes this to $2H_xN_z$ and a simultaneous 90° x pulse on $^{15}$N results in $-2H_xN_y$. The magnetisation of spin 1 is now on z and that of spin 2 is now on y. The transverse magnetisation has, therefore, been transferred from the proton onto $^{15}$N by the end of the sequence. The result is an observable signal on Nitrogen, the size of which depends upon the equilibrium magnetisation of the proton.

**Figure 1.20. Magnetisation Transfer in the INEPT and HSQC.** A) The diagram shows how the magnetisation evolves during periods of free precession when scalar coupling is active. B) A diagram of the INEPT pulse sequence. The evolution of the magnetisation during pulses and periods of free precession is shown using the product operator formalism below the pulse sequence. The red arrow indicates the transfer of magnetisation from proton to $^{15}$N. C) A pulse sequence for the $^{15}$N-HSQC is shown with the important product operators below. The red arrow indicates the magnetisation transfer from proton to $^{15}$N and back. The signal is detected on the proton channel as indicated.

The INEPT is an important building block for many NMR experiments, including the Heteronuclear Single Quantum Coherance (HSQC) experiment. In its simplest form, this contains two INEPT sequences separated by an evolution time, $t_1$ (Figure 1.20C). During $t_1$, the proton chemical shift is refocused by a 180° pulse but the $^{15}$N chemical
shift is allowed to evolve. This is necessary in order to encode the $^{15}$N shifts in the final signal. Stepping through the product operators using the rules outlined above would show that the coupling between the two nuclei is also refocused by the 180° on the proton and by stage 5, after $t_1$ and the 90° pulses, the magnetisation is described by $-\cos(\Omega N t_1)2H_y N_z$. Thus, the $^{15}$N chemical shifts are encoded and the magnetisation has been transferred back to the proton. The second INEPT sequence is completed to leave $\cos(\Omega N t_1)H_x$. The final signal encodes the $^{15}$N chemical shifts modulated by $t_1$ and is detected on the proton channel. The resulting $^{15}$N-HSQC spectrum contains a signal for every NH group in the protein. $^{15}$N- and $^{13}$C-HSQCs were used extensively within this project and several examples of each are shown in Chapters 3-6.

Further extensions to the HSQC include the triple resonance experiments used for protein backbone assignments such as the HNCA and HN(CO)CA (see Chapter 4, section 4.2.2.1), in which the magnetisation is selectively transferred using scalar couplings between protons, $^{15}$N and $^{13}$C nuclei. This brief introduction to NMR theory should provide sufficient information for the reader to understand how magnetisation is created, transferred using scalar couplings and RF pulses and detected in multi-dimensional NMR experiments. The individual pulse sequences of each experiment used within this project will not be described. Rather, the overall magnetisation transfer pathways for each experiment will be shown in simple diagrams alongside the results where necessary.

1.7.1.4 The NOE

An additional set of NMR experiments that will be described in the results chapters rely upon an interaction between nuclei that are not directly bonded but that are close in space, typically up to a maximum of 5 Å apart. A dipolar interaction between two nearby spins gives rise to a relaxation phenomenon known as the Nuclear Overhauser Effect (NOE), which causes the magnetisation of one spin to be affected by the other. The size of this NOE is inversely proportional to $r^6$, where $r$ is the distance between the nuclei. The NOE is crucial for structure determination, as will be described in more detail in Chapter 4. The forms of various NMR spectra that involve the NOE are described in more detail alongside the results presented in Chapters 3 and 4.
2 Materials and Methods

2.1 Materials

2.1.1 Chemicals
General chemicals were obtained from Sigma-Aldrich, Thermo Fisher, New England BioLabs, Formedium, Invitrogen, GE Healthcare, Abcam, Anatrace and Agilent Technologies. Farnesyl Pyrophosphate and $^{15}$N-Ammonium Chloride (98 %) was obtained from Sigma-Aldrich and $^{13}$C-glucose (99 %) from Cambridge Isotope Laboratories.

2.1.2 Commercial Enzymes
Restriction enzymes were obtained from New England BioLabs, T4 DNA Ligase and Calf-intestinal Alkaline Phosphatase from Sigma-Aldrich and restriction grade Thrombin from Merck Millipore.

2.1.3 Commercial Kits

2.1.3.1 DNA Purification
The GenElute Plasmid Miniprep kit (Sigma-Aldrich) and Illustra plasmidPrep Mini Spin kit (GE Life Sciences) were used for plasmid purification from bacteria. The QIAquick Gel Extraction kit (Qiagen) was used for gel extraction of DNA.

2.1.3.2 Site-directed Mutagenesis
The QuikChange Lightning Multi Site-Directed Mutagenesis kit from Agilent Technologies was used for all mutagenesis.

2.1.3.3 TOPO Cloning
The TOPO® TA Cloning® Kit with PCR®2.1 TOPO® from Invitrogen was used in conjunction with One Shot® TOP10F’ Chemically Competent E. coli.

2.1.4 Oligonucleotides
The customised PCR primers (section 2.2.2.1) were obtained from Sigma-Aldrich.
### 2.1.5 Expression Vectors

<table>
<thead>
<tr>
<th>Vector</th>
<th>Fusion Tag</th>
<th>Protease Cleavage Site</th>
<th>Source</th>
<th>Antibiotic Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEX-6P-1</td>
<td>GST (N)</td>
<td>3C Protease</td>
<td>GE Healthcare</td>
<td>Amp</td>
</tr>
<tr>
<td>pGEX-HisP</td>
<td>GST (N)</td>
<td>3C Protease</td>
<td>Based on pGEX-6P-1, Ref: (Hutchinson et al. 2011)</td>
<td>Amp</td>
</tr>
<tr>
<td>pGEX-2T</td>
<td>GST (N)</td>
<td>Thrombin</td>
<td>GE Healthcare</td>
<td>Amp</td>
</tr>
<tr>
<td>pMAT10</td>
<td>His&lt;sub&gt;6&lt;/sub&gt;-MBP (N)</td>
<td>Thrombin</td>
<td>K. Littlefield and D. Owen unpublished</td>
<td>Amp</td>
</tr>
<tr>
<td>pMAT10-P</td>
<td>His&lt;sub&gt;6&lt;/sub&gt;-MBP (N)</td>
<td>3C Protease</td>
<td>D. Owen, unpublished.</td>
<td>Amp</td>
</tr>
<tr>
<td>pET-6xHis-SNAP</td>
<td>His&lt;sub&gt;6&lt;/sub&gt;-SNAP (N)</td>
<td>-</td>
<td>J. Gallop, unpublished.</td>
<td>Amp</td>
</tr>
<tr>
<td>pRSF-DUET-1</td>
<td>Cloning site 1: His&lt;sub&gt;6&lt;/sub&gt; (N) Cloning site 2: none</td>
<td>-</td>
<td>Addgene</td>
<td>Kan</td>
</tr>
</tbody>
</table>

**Table 2-1**
## 2.1.6 Protein Expression Constructs

<table>
<thead>
<tr>
<th>Protein</th>
<th>Gene source</th>
<th>Amino Acids</th>
<th>Mutations</th>
<th>Vectors</th>
<th>Restriction Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOCA1</td>
<td><em>Xenopus tropicalis</em></td>
<td>HR1: 330-426</td>
<td>-</td>
<td>pGEX-6P-1, pGEX-HisP, pMAT10</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FBAR: 1-287</td>
<td>-</td>
<td>pGEX-6P-1, pMAT10</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HR1SH3: 330-545</td>
<td>-</td>
<td>pGEX-6P-1, pGEX-HisP, pMAT10, pMAT10-P</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SH3: 479-545</td>
<td>-</td>
<td>pGEX-6P-1, pMAT10</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FL: 1-545</td>
<td>-</td>
<td>pGEX-6P-1, pMAT10-P</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td>FPB17</td>
<td><em>Xenopus laevis</em></td>
<td>HR1: 385-486</td>
<td>-</td>
<td>pGEX-6P-1</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td>CIP4</td>
<td><em>Homo sapiens</em></td>
<td>HR1: 388-481</td>
<td>-</td>
<td>pGEX-6P-1</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td>Cdc42</td>
<td><em>Homo sapiens</em></td>
<td>FL: 1-191</td>
<td>Q61L</td>
<td>pGEX-2T, pGEX-6P-1</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L191A</td>
<td>pGEX-6P-1</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Q61L, L191A</td>
<td>pGEX-6P-1</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Q61L, L191M</td>
<td>pGEX-6P-1</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Δ7</td>
<td>Q61L</td>
<td>pGEX-2T</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Δ4-His10</td>
<td>Q61L</td>
<td>pGEX-6P-1</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Δ3</td>
<td>Q61L, C188S</td>
<td>pGEX-6P-1</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
<tr>
<td>NWASP</td>
<td><em>Xenopus tropicalis</em></td>
<td>GBD: 197-255</td>
<td>-</td>
<td>pGEX-6P-1</td>
<td><em>BamHI, EcoRI</em></td>
</tr>
</tbody>
</table>

**Constructs from Helen Fox (cloned from IMAGE: 5157175)**

| TOCA1         | *Xenopus tropicalis* | FL: 1-550 | -         | pET-6xHis-SNP, FseI, AscI |
| F-BAR: 1-287  | -         | pET-6xHis-SNP, FseI, AscI |
| ΔSH3: 1-480   | -         | pET-6xHis-SNP, FseI, AscI |

**Construct from Gerrit Praefke**

<table>
<thead>
<tr>
<th>Farnesyl Transferase</th>
<th><em>Homo sapiens</em></th>
<th>FL</th>
<th>-</th>
<th>pRSF-DUET-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloning site 1: α subunit, <em>BamHI, EcoRI</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloning site 2: β subunit, <em>BglII, XhoI</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2-2
## 2.1.7 Bacterial Strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Phenotype</th>
<th>Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL21 (Invitrogen)</td>
<td>F′ompT hsds(rBmB)</td>
<td>Deficient in lon8 and ompT proteases</td>
<td>None</td>
</tr>
<tr>
<td>BL21 DE3 (Invitrogen)</td>
<td>F′ompT hsds(rBmB) gal dcm (DE3)</td>
<td>Based on BL21. Contains DE3 lysogen encoding T7 polymerase</td>
<td>None</td>
</tr>
<tr>
<td>BL21-CodonPlus-RP (DE3) (Agilent)</td>
<td>F′ompT hsds(rBmB) gal dcm argU (AGA, AGG), proL (CCC), gal dcm (DE3)</td>
<td>Based on BL21 (DE3) with tRNAs for AGA, AGG and CCC</td>
<td>Cam</td>
</tr>
<tr>
<td>BL21-CodonPlus-RIL (DE3) (Agilent)</td>
<td>F′ompT hsds(rBmB) gal dcm argU (AGA, AGG), ileY (AUA), leuW (CUA), gal dcm (DE3)</td>
<td>Based on BL21 (DE3) with tRNAs for AGA, AGG, AUA and CUA</td>
<td>Cam</td>
</tr>
<tr>
<td>BL21-Rosetta2™ (DE3) pLysS (Novagen)</td>
<td>F′ompT hsds(rBmB) gal dcm (DE3) pLysSRARE2 (CamR)</td>
<td>Based on BL21 (DE3) with tRNAs for AGA, AGG, AUA, CUA, CCC, and GGA</td>
<td>Cam</td>
</tr>
<tr>
<td>BL21 (DE3) OverExpress™ C41</td>
<td>F′ompT hsdsSB (rB-mB-) gal dcm (DE3)</td>
<td>Effective in expressing toxic and membrane proteins</td>
<td>None</td>
</tr>
<tr>
<td>BL21 OverExpress™ C41 + pLysSRARE2 (Novagen)</td>
<td>F′ompT hsdsSB (rB-mB-) gal dcm (DE3) pLysSRARE2 (CamR)</td>
<td>Effective in expressing toxic and membrane proteins. tRNAs: AUA, AGG, AUA, CUA, CCC, and GGA</td>
<td>Cam</td>
</tr>
<tr>
<td>BL21 TRX (Novagen)</td>
<td>F′ompT hsds(rBmB) trxB-11 KanR</td>
<td>Based on BL21 with thioredoxin B mutation, facilitates protein folding</td>
<td>Kan</td>
</tr>
<tr>
<td>XL1 Blue MRF+ (shortened to XL1 hereafter)</td>
<td>recA1 endA1 gyrA96 thi-1 hsdr17 supE44 relA1 lac [F′ proAB lacqZΔM15 Tn10 (Tetr)]</td>
<td>Cloning site within LacZ mini-gene for blue-white selection</td>
<td>Tet</td>
</tr>
<tr>
<td>One Shot® TOP10 F′ Chemically Competent E. coli (Thermo Fisher)</td>
<td>F′[lacI Tn10 (Tetr)] mcrA Δ(mrr-hsdRMSP-mrBC) Δ80lacZΔM15 ΔlacX74 recA1 araD139</td>
<td>The F′ episome carries the lacI9 repressor for inducible expression with IPTG and blue-white screening</td>
<td>Tet</td>
</tr>
<tr>
<td>XL10 Gold® Ultracompetent (Agilent Technologies)</td>
<td>TetrD(mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′ proAB lacqZDM15 Tn10 (Tetr) Amy Camr]</td>
<td>High transformation efficiency for large DNA molecules, deficient in endonuclease A, recombination and all known restriction systems.</td>
<td>Tet, Cam</td>
</tr>
</tbody>
</table>

Table 2-3
2.1.8 Media and Buffers

2.1.8.1 Cloning and Microbiology

### Agarose Gels

- 50x TAE Buffer: 2 M Tris-base, 5% Acetic acid, 50 mM EDTA, pH 8.0
- 5x STOP Buffer: 50% sucrose, 50 mM EDTA pH 7.5, 0.1% Bromophenol blue

### Bacterial Growth Media

**2TY**
- 16 g/L Tryptone, 10 g/L Yeast Extract, 5 g/L NaCl

**Agar plates**
- 16 g/L Tryptone, 10 g/L Yeast Extract, 5 g/L NaCl, 16 g/L Agar

**Celtone-enriched MOPS**
- 10x MOPS: 400 mM K-MOPS pH 7.4, 40 mM Tricine pH 7.4, 5.28 mM MgCl₂, 2.76 mM K₂SO₄, 500 mM NaCl, 100 µM FeSO₄, 5 µM CaCl₂, 100x Micronutrients: 0.4 mM H₃BO₃, 80 µM MnCl₂, 30 µM CoCl₂, 10 µM CuSO₄, 10 µM ZnSO₄, 3 µM (NH₄)₆Mo₇O₂₄₃
- Final MOPS medium (1L): 100 mL 10x MOPS, 1 mL 100x Micronutrients, 1.32 mM KH₂PO₄, pH 7.4, 0.4% (w/v) glucose, 20 mM NH₄Cl, 5% (v/v) Celtone (Martek)

**M9**
- 47 mM Na₂HPO₄, 22 mM KH₂PO₄, 8.6 mM NaCl, 4 µM ZnSO₄, 1 pM MnSO₄, 0.7 µM H₃BO₃, 0.7 µM CuSO₄, 0.4% (w/v) glucose, 2 nM FeCl₃, 2 µM MgSO₄, 20 mM NH₄Cl, 0.1 µM CaCl₂

### Bacterial Transformations

- **TF1**: 10 mM MOPS pH 6.8, 10 mM CaCl₂
- **TF2**: 10 mM MOPS pH 6.8, 75 mM CaCl₂
- **TF3**: 50 mM MOPS pH 7.5, 75 mM CaCl₂, 5% glycerol, 1 mM EDTA, 100 mM KCl

#### Table 2-4

2.1.8.2 Protein Purification

### Lysis and Wash Buffers

- **Cdc42**: 50 mM Tris·HCl pH 7.5, 150 mM NaCl, 5 mM DTT, 5 mM MgCl₂
- **TOCA1 HR1**: MT-PBS (20 mM sodium phosphate pH 7.5, 150 mM NaCl)
- **TOCA1 FL**: 20 mM Sodium Phosphate pH 7.4, 300 mM NaCl, 5 mM DTT
- **TOCA1 SH3**: MT-PBS
- **FBP17 HR1**: MT-TBS (50 mM Tris·HCl pH 7.5, 150 mM NaCl), 5 mM EDTA
- **CIP4 HR1**: MT-TBS, 5 mM EDTA
- **N-WASP GBD**: MT-PBS
- **Ni²⁺-affinity Lysis Buffer**: 20 mM Tris·HCl pH 7.9, 500 mM NaCl
- **Ni²⁺-affinity Wash Buffer**: 20 mM Tris·HCl pH 7.9, 250 mM NaCl

### Protease Cleavage Buffers

- **3C Protease Cleavage Buffer**: 50 mM Tris·HCl pH 7.5, 150 mM NaCl, 1 mM DTT (with 0.5 mM EDTA added for all except Cdc42)
- **Thrombin Cleavage Buffer**: 50 mM Tris·HCl pH 8.4, 150 mM NaCl, 2.5 mM CaCl₂

### Ion Exchange Buffers

- **Buffer A**: 20 mM Tris·HCl pH 7.5
- **Buffer B**: 20 mM Tris·HCl pH 7.5, 1 M NaCl

#### Table 2-5
2.1.8.3  Biochemistry and Biophysics

<table>
<thead>
<tr>
<th>Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Laemmli Resolving Gel</strong></td>
</tr>
<tr>
<td>Component</td>
</tr>
<tr>
<td>Acrylamide: Bis-Acrylamide (30 %, 29:1)</td>
</tr>
<tr>
<td>1.5 mM Tris:HCl pH 8.8</td>
</tr>
<tr>
<td>MiliQ Water</td>
</tr>
<tr>
<td>10 % SDS</td>
</tr>
<tr>
<td>10 % APS</td>
</tr>
<tr>
<td>TEMED</td>
</tr>
<tr>
<td><strong>Laemmli Stacking Gel</strong></td>
</tr>
<tr>
<td>Component</td>
</tr>
<tr>
<td>Acrylamide: Bis-Acrylamide (30 %, 29:1)</td>
</tr>
<tr>
<td>0.5 mM Tris:HCl pH 6.8</td>
</tr>
<tr>
<td>MiliQ Water</td>
</tr>
<tr>
<td>10 % SDS</td>
</tr>
<tr>
<td>10 % APS</td>
</tr>
<tr>
<td>TEMED</td>
</tr>
<tr>
<td><strong>Final Volume</strong></td>
</tr>
<tr>
<td><strong>2x SDS Loading Buffer (SDS-LB)</strong></td>
</tr>
<tr>
<td>3 % SDS, 6 M UREA, 33 mM Tris:HCl pH 7.4, 66 mM β-mercaptoethanol, 0.1 % Bromophenol blue</td>
</tr>
<tr>
<td><strong>Running Buffer</strong></td>
</tr>
<tr>
<td>25 mM Tris-base, 192 mM Glycine, 0.1 % SDS</td>
</tr>
<tr>
<td><strong>SPA Dilution Buffer</strong></td>
</tr>
<tr>
<td>50 mM Tris:HCl pH 7.5, 0.2 mg/mL BSA, 2 mM DTT, 1 mM MgCl₂</td>
</tr>
<tr>
<td><strong>SPA Reaction Mix</strong></td>
</tr>
<tr>
<td>50 mM Tris:HCl pH 7.5, 0.2 mg/mL BSA, 2 mM DTT, 1 mM MgCl₂, 29 % SPA beads (PerkinElmer) in 50 mM Tris:HCl pH 7.5/0.1 % azide, 0.7 % anti-GST (A5800 from Life Technologies) or 0.6 % anti-his antibody (H1029 from Sigma).</td>
</tr>
<tr>
<td><strong>Other</strong></td>
</tr>
<tr>
<td><strong>HPLC Buffer</strong></td>
</tr>
<tr>
<td>0.6 M monobasic ammonium phosphate pH 4.0</td>
</tr>
<tr>
<td><strong>NMR Titration Buffer</strong></td>
</tr>
<tr>
<td>50 mM sodium phosphate, 25 mM NaCl, 5 mM MgCl₂, 5 mM DTT</td>
</tr>
</tbody>
</table>

Table 2-6

2.2  Methods

2.2.1  General

2.2.1.1  SDS-PAGE

Laemmli gels were prepared (Table 2-6) (percentage indicated in the results) in 1 mm gel cassettes (Thermo Fisher). Gel samples were prepared in SDS-LB (Table 2-6), boiled for 5 min and 10 µL loaded into the gel with unstained broad range protein markers (New England Biolabs) as indicated. Gels were run in Running Buffer at 200 V for 1 h, stained with InstantBlue (Expedeon) and washed in water.
2.2.1.2 *Agarose Gel Electrophoresis*

Agarose gels were prepared in TAE buffer (Table 2-4) with 0.5 µg/mL Ethidium Bromide (EtBr) at the desired percentage (1-3 %). DNA samples were made up in 5x STOP buffer and 20 µL loaded into the gel. The gels were run in TAE buffer with 0.5 µg/ml ethidium bromide at 50-100 V until the desired separation was reached and visualised in UV light.

2.2.1.3 *Transformation of E. coli (Calcium Treatment)*

0.75 mL of an overnight culture of *E. coli* (strains indicated) was added to 20 mL 2TY and grown at 37 ºC until A₆₀₀ = 0.4-0.6. Cells were pelleted at 6000 g for 10 minutes, resuspended in 10 mL TF1, pelleted again and resuspended in 1 mL TF2. Cells were kept at 4 ºC for between 2 and 48 h. 100 µL of cell suspension was added to 80-99 µL of TF3 with 1-20 µL of DNA (final volume 200 µL) and incubated on ice for 45 min. The cells were heat-shocked at 42 ºC for 10 min before being returned to ice for 2 min. 1 mL of 2TY was added and the cells incubated at 37 ºC for 30 min. 200 µL was spread on an agar plate with the appropriate antibiotics. The remaining 1 mL of culture was spread on another plate.

2.2.2 Cloning

2.2.2.1 *The Polymerase Chain Reaction (PCR)*

PCR was used to clone cDNA regions according to the following hot start PCR cycle:

<table>
<thead>
<tr>
<th>Process</th>
<th>Time (min)</th>
<th>Temperature (ºC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denature</td>
<td>6</td>
<td>95</td>
</tr>
<tr>
<td>Anneal</td>
<td>2</td>
<td>45-65</td>
</tr>
<tr>
<td>Pause</td>
<td>1 µL (5 units) of Taq polymerase added</td>
<td></td>
</tr>
<tr>
<td>Extend</td>
<td>2</td>
<td>72</td>
</tr>
<tr>
<td>Denature</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>Anneal</td>
<td>2</td>
<td>45-65</td>
</tr>
<tr>
<td>Extend</td>
<td>10</td>
<td>72</td>
</tr>
</tbody>
</table>

**Reaction Mix:** 55 µL sterile analytical grade water (SAW), 8 µL of 7.5 mM dNTPs, 6 µL of 25 mM MgCl₂, 10 µL of 10x Taq buffer (New England Biolabs), 10 µL of each primer at 100 pmoles/µL and 1 µL of template DNA

*Table 2-7*
The constructs cloned are detailed in Table 2.8, along with the primers and annealing temperatures used. The restriction sites that were incorporated into the primer are coloured red.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Template</th>
<th>Construct</th>
<th>Annealing Temp (°C)</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOCA1</td>
<td>Xenopus tropicalis cDNA Accession number: NM_001005148 (isoform 4) from Dr J Gallop, Gurdon Institute UK</td>
<td>HR1: 330-426</td>
<td>45</td>
<td>F:CGCGGGATCCCATATGAAGGGGTCC GGCCTTGGGAAGATTTTG TACTGAGTTTGGGAC R:GCGGAAATTCAGATCTTCACCTCACT TCTCTGGGATACCTTTGCTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FBAR: 1-287</td>
<td>45</td>
<td>F:CGCGGGATCCCATATGAGGTCGG TAAGTTTGTGGGAC R:GCGGAAATTCAGATCTTCACCTCACT TCTCTGGGATACCTTTGCTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HR1SH3: 330-545</td>
<td>45</td>
<td>F:CGCGGGATCCCATATGAGGAGTTCC GGCCTTGGGAAGATTTTG R:GCGGAAATTCAGATCTTCACCTCACT TCTCTGGGATACCTTTGCTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SH3: 479-545</td>
<td>45</td>
<td>F:CGCGGGATCCCATATGCCAGCTATC GGACACTGCAAATCA R:GCGGAAATTCAGATCTTCACCTCACT TCTCTGGGATACCTTTGCTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FL: 1-545</td>
<td>45</td>
<td>F:CGCGGGATCCCATATGCCAGCTATC GGACACTGCAAATCA R:GCGGAAATTCAGATCTTCACCTCACT TCTCTGGGATACCTTTGCTT</td>
</tr>
<tr>
<td>FPB17</td>
<td>Xenopus laevis cDNA IMAGE: 5514481</td>
<td>HR1: 385-486</td>
<td>60-54°C in 2°C increments, then 45°C</td>
<td>F:CGCGGGATCCCATATGCTAAAGCG AGGGGGC R:GCGGAAATTCAGATCTTCACCTCACT TCTCTGGGATACCTTTGCTT</td>
</tr>
<tr>
<td>CIP4</td>
<td>Homo sapiens cDNA IMAGE: 3532036</td>
<td>HRI: 388-481</td>
<td>Same as FPB17 HR1</td>
<td>F:CGCGGGATCCATGACCGAGGATTTT AGCCACTTG R:GCGGAAATTCAGATCTTCACCTCACT TCTCTGGGATACCTTTGCTT</td>
</tr>
<tr>
<td></td>
<td>Homo sapiens cDNA from lab stocks: pGEX2T-Cdc42ΔQ61L</td>
<td>Δ4 Q61L His10</td>
<td>65, 60, 58, 56, 54, 50, 45°C then 44°C for remaining cycles</td>
<td>F:CGCGGGATCCCATATGCAAGACAAA TAAGTTTGTGGGAC R:GCGGAAATTCAGATCTTCACCTCACT TCTCTGGGATACCTTTGCTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Δ3 Q61L C188S</td>
<td>60-45°C in 5°C increments, then 45°C</td>
<td>F:CGCGGGATCCCATATGCAAGACAAA TAAGTTTGTGGGAC R:GCGGAAATTCAGATCTTCACCTCACT TCTCTGGGATACCTTTGCTT</td>
</tr>
<tr>
<td>NWASP</td>
<td>Xenopus tropicalis cDNA IMAGE: 5379332</td>
<td>GBD: 197-255</td>
<td>47-45°C in 1°C increments, then 45°C</td>
<td>F:CGCGGGATCCCATATGCAAGACAAA TAAGTTTGTGGGAC R:GCGGAAATTCAGATCTTCACCTCACT TCTCTGGGATACCTTTGCTT</td>
</tr>
</tbody>
</table>

Table 2-8

The PCR products were verified based on size using agarose gel electrophoresis (section 2.2.1.2) before cloning into PCR®2.1 TOPO®.
2.2.2.2 **TOPO cloning**

The PCR products were ligated into PCR®2.1 TOPO®, transformed into chemically competent TOP10F’ cells (Invitrogen) and selected on kanamycin/tetracycline plates. Plasmid DNA was purified using the GenElute Plasmid Miniprep kit or the Illustra plasmidPrep Mini Spin kit and verified by DNA sequencing (DNA sequencing facility, Department of Biochemistry).

2.2.2.3 **Digestion and Purification of Expression Vectors**

Expression vectors were purified from XL1 cells. The DNA concentration was measured by UV absorbance at 260 nm and calculated according to the following equation:

\[ A_{260} \times 50 \mu g/mL = X \mu g/mL \text{ of DNA} \]

5 µg of vector was incubated with 50 units of *BamHI* and 50 units of *EcoRI* in restriction digest Buffer B (Roche). Sterile analytical grade water (SAW) was added to a final reaction volume of 100 µL and the reaction incubated at 37 °C for 2.5 h. 15 µL of 1M Tris:HCl pH 8.0, 15 µL of 100 mM MgCl₂, 15 µL of SAW and 15 units of Calf Intestinal Phosphatase (Sigma-Aldrich) were added and the reaction incubated at 37 °C for 20 min. The reaction was incubated at 65 °C for 20 min before 150 µL of SAW was added. 300 µL of phenol was added and the reaction vortexed for 30 s. The mixture was centrifuged at 13000 g for 1 min and the top layer (containing the vector) put into a new tube. This was repeated 3 times. 300 µL of ether was added, the mixture vortexed and centrifuged and the top layer was discarded. This was repeated until the mixture was clear. The ether was evaporated for 30 min. The vector was stored at -20 °C.

2.2.2.4 **Ligations of DNA Inserts into Expression Vectors**

The constructs in Table 2-8 were made as follows: 100 ng of each PCR2.1-construct was digested with *BamHI* and *EcoRI* (50 units of each in 100 µL) in Buffer B (Roche). The insert was purified by agarose gel electrophoresis and gel extraction of the correctly sized DNA fragment. 16 µL of the purified insert was mixed with 1 µL of pre-cut expression vector (*BamHI/EcoRI*, see above), 2 µL of 10x T4 ligase buffer and 1 µL T4 ligase (1 unit/µL) and incubated at 16 °C overnight. For TOCA1 HR1SH3, FBAR and FL, the inserts were first diluted to achieve a final molar ends.
ratio of 2:1, insert:vector. The ligation reactions (20 µL) were transformed into Calcium-competent *E. coli* XL1 and selected on ampicillin/tetracycline plates.

### 2.2.2.5 Site-Directed Mutagenesis

The mutants outlined in Table 2.9 were produced from pGEX-6P-1 Cdc42Q61L using the QuikChange Lightning Multi Site-Directed Mutagenesis kit (Agilent Technologies) with an extension time of 165 s using one primer (Table 2-9). The mutated DNA was transformed into XL10-Gold Ultracompetent cells. Plasmids were purified and sequenced (DNA sequencing facility, Department of Biochemistry).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Mutation</th>
<th>Mutagenesis Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cdc42</td>
<td>L191A</td>
<td>CGC AGG TGT GTG CTG GCA TGA ATT CCC GGG TCG</td>
</tr>
<tr>
<td></td>
<td>L191M</td>
<td>GC CGC AGG TGT GTG CTG ATG TGA ATT CCC GGG TCG AC</td>
</tr>
<tr>
<td></td>
<td>L61Q</td>
<td>GA CTT TTT GAT ACT GCA GGG CAA GAG GAT TAT GAC AGA TTA CG</td>
</tr>
</tbody>
</table>

**Table 2-9**

### 2.2.3 Protein Expression and Purification

#### 2.2.3.1 Small Scale Protein Expression

Expression constructs (verified by sequencing) were transformed into expression strains of *E. coli* (Table 2-3). 1 mL of overnight culture was added to 9 mL of 2TY with the appropriate antibiotics and grown at 37 °C until $A_{600} = 0.8$-1.0. Cultures were cooled and 0, 0.1 or 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) added. Cultures were incubated at 37 °C for 5 h or at 20 °C for 20 h. 1 mL of each culture was pelleted at 13000 g for 5 min and the pellets resuspended in the appropriate volume of Milli-Q water and SDS-LB (1:1) in order to normalise samples, as determined by:

$$\text{Total volume (µL)} = 1000 \times \frac{A_{600}}{6}$$

A further 1 mL sample from each of the induced cultures was pelleted and resuspended in the calculated volume of water. 2 µL of 20 mg/mL lysozyme and 1 µL of 10 mg/mL DNase were added and the samples subjected to multiple rounds of freeze-thawing (dry ice to 42 °C) until clear. The cell debris was pelleted at 13000 g for 5 min and the supernatant was added to SDS-LB (1:1). The pellet was resuspended in the calculated volume of SDS-LB. Samples were analysed by SDS-PAGE.
2.2.3.2 Large Scale Protein Expression

6x50 mL overnight cultures of *E. coli* (strains indicated below) were added to 6x500 mL 2TY with the appropriate antibiotic(s) in baffled flasks and grown at 37 °C with shaking until A_{600} ≈ 0.8. The total culture volume was 3 L unless otherwise indicated. The cultures were cooled and induced with 0.1 or 1 mM IPTG for 5 h at 37 °C or 20 hours at 20 °C (Table 2.10).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Strain</th>
<th>[IPTG] (mM)</th>
<th>Temperature (°C)</th>
<th>Time (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOCA1 HR1</td>
<td>BL21</td>
<td>0.1</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>TOCA1 FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOCA1 SH3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cdc42 (all)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBP17 HR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIP4 HR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NWASP GBD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
<th>Strain</th>
<th>[IPTG] (mM)</th>
<th>Temperature (°C)</th>
<th>Time (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOCA1 HR1</td>
<td>BL21</td>
<td>0.1</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>TOCA1 SH3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cdc42 (all)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBP17 HR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIP4 HR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NWASP GBD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
<th>Strain</th>
<th>[IPTG] (mM)</th>
<th>Temperature (°C)</th>
<th>Time (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOCA1 HR1</td>
<td>BL21</td>
<td>0.1</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>TOCA1 SH3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cdc42 (all)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBP17 HR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIP4 HR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NWASP GBD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2-10

2.2.3.3 Cdc42/FTase Co-expression

BL21 cultures were made competent (section 2.2.1.3) and transformed with pGEX-6P-1-Cdc42Q61L/L191A or L191M and pRSF-DUET-1-FTase. Doubly transformed cells were selected on the appropriate antibiotics (see Table 2-1 and Table 2-3). Expressions were carried out as described in section 2.2.3.2 and in Chapter 6.
2.2.3.4 Expression of Isotope Labelled Proteins

For small scale expression trials, 10 mL overnight cultures grown in 2TY were pelleted and resuspended in 10 mL minimal media (M9 or MOPS). 1 mL was then added to 10 mL minimal media. For large scale expressions, 20 mL overnight cultures grown in 2TY were pelleted and resuspended in 10 mL minimal media, which was added to 500 mL minimal media. For \(^{15}\)N-labelled proteins, \(^{15}\)N-ammonium chloride was added to the growth medium at 1 mg/mL with 0.4 % unlabelled glucose. For \(^{15}\)N/\(^{13}\)C-labelled proteins, \(^{15}\)N-ammonium chloride was added to 1 mg/mL with 0.3 % \(^{13}\)C-glucose. The cultures were induced and grown as for the unlabelled proteins.

2.2.3.5 Affinity Purification - GST

Bacterial cultures were pelleted at 7000 g for 15 minutes and resuspended in 20 mL of buffer (Table 2-5) with SIGMAFAST Protease Inhibitor Cocktail tablets, EDTA-free (Sigma-Aldrich). 1 mM Phenylmethylsulfonyl fluoride (PMSF) was added to all but the Cdc42 preparations. Cells were lysed at 10000 psi using an Emulsiflex (Avestin). 1 mM PMSF was added again with 0.1 % Triton (except for Cdc42 preparations, for which no PMSF was added and 0.1 % octyl glucoside was used instead of Triton). The lysates were spun at 48000 x g for 30 min and the supernatant added to 2x 7.5 mL glutathione-agarose beads (Sigma-Aldrich). The beads were rotated at 4 °C for 1 h and washed three times with 40 mL of wash buffer (Table 2-5) with 0.1 % Triton (or 0.1 % octyl glucoside) then 3x in 3C protease buffer. 15 mL of 3C protease buffer was added with 3C protease at 0.025 mg/ml and the beads incubated overnight at 4 °C with rotation. For proteins expressed from pGEX-2T or pMAT10, Thrombin cleavage buffer was used with 25 units of Thrombin (Merck-Millipore). The beads were pelleted at 1000 g for 10 minutes and the supernatant collected. Three washes were collected (buffers in Table 2-5). Samples of washes and bead samples before and after cleavage and washing were analysed by SDS-PAGE.

2.2.3.6 Affinity Purification - Ni\(^{2+}\)-affinity Beads

Cultures were processed as above in 10 mL of Ni\(^{2+}\)-affinity Lysis Buffer. The supernatant collected after lysis was added to 2x 5 mL Ni\(^{2+}\)-NTA-agarose resin (Qiagen). The resin was rotated at 4 °C for 1 h and washed three times in 40 mL Ni\(^{2+}\)-affinity Wash Buffer with 0.1 % Triton and a further 2 times with 40 mL Ni\(^{2+}\)-affinity Wash Buffer supplemented with 20 mM Imidazole. The resin was then washed 3 times into Thrombin cleavage buffer (proteins expressed from pMAT10) or 3C
Protease cleavage buffer without DTT (proteins expressed from pMAT10-P). 15 mL of protease cleavage buffer and 25 units of Thrombin or 0.025 mg/ml 3C protease were added. The resin was incubated overnight at 4 °C with rotation then spun at 1000 g for 10 minutes and the supernatant collected. Three washes with Ni^{2+}-affinity Wash Buffer were collected. Samples from every stage were analysed by SDS-PAGE.

2.2.3.7 Affinity Purification - Farnesyl Transferase
10 mL of Ni^{2+}-IDA sepharose resin (GE Healthcare) was packed into a 50 mL column (GE Healthcare). The resin was washed with 5 volumes of 50 mM EDTA pH 7.9, 5 volumes of water and 5 volumes of 0.5 M NiSO_{4}. The column was left in NiSO_{4} for 1 hour before being washed into Ni^{2+}-affinity Wash Buffer supplemented with 20 µM ZnCl_{2}. 50 mL of bacterial cell lysate was loaded onto the column using a 50 mL superloop (GE Healthcare) at 1 mL/min. The column was washed with Ni^{2+}-affinity Wash Buffer at 1 mL/min until the absorbance returned to baseline and then with 1.5 volumes of 20 mM Imidazole in Ni^{2+}-affinity Wash Buffer supplemented with 20 µM ZnCl_{2}. A gradient of 20 to 500 mM Imidazole over 5 column volumes was used to elute the protein in 2 mL fractions, which were analysed by SDS-PAGE.

2.2.3.8 Concentration of Protein Solutions
Protein samples were concentrated using an Amicon® stirred cell with YM3 or YM10 membrane prior to size exclusion chromatography or ion exchange and on 4 or 15 mL Amicon® Ultra Centrifugal Filters with 3000 or 10000 NMWL (Merck-Millipore) following size exclusion chromatography or ion exchange.

2.2.3.9 Purification by Size Exclusion Chromatography
A Superdex 16/60 S30, S75 or S200 column (GE Healthcare) was used depending on the size of the protein as indicated in the figure legends within the results chapters. The columns were equilibrated in the buffers indicated in the results prior to loading of 1-2 mL protein sample. The samples were eluted with 1.5 column volumes at 1 mL/min and collected in 2 mL fractions, which were analysed by SDS-PAGE.

2.2.3.10 Purification by Ion Exchange
A 1 mL ResourceQ column (GE Healthcare) was used. The column was first washed in 10 column volumes of Ion Exchange Buffer A followed by 5 column volumes of Ion Exchange Buffer B and another 10 volumes of Ion Exchange Buffer A (Table
2 mL of protein sample was loaded onto the column. The column was washed with Buffer A until the A$_{280}$ returned to baseline. The protein was eluted with a gradient of 0-100 % Buffer B, which was reduced to 0-25 % in subsequent runs.

### 2.2.3.11 Protein Analysis

TOCA1 HR1 and HR1SH3, FBP17 HR1 and the Cdc42 proteins were analysed using mass spectrometry to ensure no degradation had taken place (PNAC facility, Department of Biochemistry, University of Cambridge). All protein concentrations were determined using amino acid analysis (PNAC facility), apart from the FBP17 HR1 domain, which was determined by UV absorbance at 280 nm, using the Beer-Lambert Law with its theoretical extinction coefficient ($\varepsilon = 6990$, estimated using EMBOSS Pepstats).

### 2.2.3.12 In Vitro Farnesylation

Initial *in vitro* farnesylation reactions were carried out with 200 µM Cdc42, 10 µM FTase and 300 µM FPP for 1 h at 30 °C. Further reactions were carried out with altered conditions as described in Chapter 6. Farnesylated protein was purified by size exclusion chromatography on a Superdex 16/60 S75 column (GE Healthcare) in 50 mM Tris:HCl, 150 mM NaCl, 5 mM DTT and 5 mM MgCl$_2$.

### 2.2.4 Nucleotide Exchange and HPLC Analysis

#### 2.2.4.1 Nucleotide Exchange

**[³H]-GTP for SPAs:** Cdc42 was loaded with [³H]GTP using [8,5'-³H]GTP (0.15 mCi, GE Healthcare) as described previously (Bailey et al. 2009). Excess nucleotide was removed using 5 mL Sephadex G25 spin columns (GE Healthcare) in Cdc42 Buffer (Table 2-5). The concentration of protein was determined by Bradford Assay as follows: 5 µL of protein was added to 800 µL water and 200 µL Bradford Reagent (BioRad) and the A$_{595}$ recorded. The concentration was estimated by the following equation, as determined previously based on BSA standards:

$$\text{Concentration (µg)} = (A-0.0136)/(0.0383)$$

**GMPPNP for NMR:** 15 mg of Cdc42 in 500 µL of Cdc42 Buffer was incubated with 333 µL of 20 mM GMPPNP (Sigma-Aldrich), 333 µL alkaline phosphatase beads (Sigma-Aldrich) in 10 mM Tris:HCl pH 7.5 (80 units), 91 µL of 3 M ammonium
sulphate and a final concentration of 0.1 mM ZnCl₂ at 37 °C for 4.5 h. The proteins were then further purified by size exclusion chromatography. HPLC was used to confirm the exchange (section 2.2.4.2). For full length and lipid modified Cdc42, 0.1 % octyl glucoside was included in the reaction and the amount of ammonium sulphate reduced by half.

2.2.4.2 HPLC Analysis
50 µg of Cdc42 was precipitated with 2 µL of 0.9 M perchloric acid to release the bound nucleotide. 100 µL of HPLC Buffer (Table 2-6) was added and the sample spun at 13000 g for 10 min to remove precipitated protein. The supernatant was loaded onto a 1.5 mL Partisphere SAX column (Whatman) that had been pre-equilibrated in HPLC Buffer and eluted in the same buffer at 1.5 mL/min. The retention time was compared with nucleotide standards prepared at 0.1 mM.

2.2.5 Biochemical and Biophysical Assays

2.2.5.1 Scintillation Proximity Assays (SPA)
**Direct binding assay:** 30 nM GST-PAK, ACK or WASP or 50 nM TOCA1-His₆ was added to SPA Reaction Mix (Table 2-6) and, as such, the GST- or His-tagged proteins were attached to a fluromicrosphere via an anti-GST or anti-His antibody. A 12 x serial dilution of Cdc42Δ7Q61L·[^³H]GTP or full length Cdc42Q61L·[^³H]GTP was set up in SPA Dilution Buffer (Table 2-6) with a maximum concentration of 16 µM and 25 µL at each concentration was added to 175 µL of the SPA Reaction Mix with the protein of interest, in a 96-well plate. The plate was incubated with shaking at 18 °C for 30 min and spun at 1000 g for 2 min. Each well was counted in a MicroBeta Scintillation counter (PerkinElmer) for 5 min. Control experiments were carried out in SPA Reaction Mix with no effector protein and subtracted from the experimental data. Binding curves were fitted using a direct binding isotherm to obtain K⁺ values for the interactions as described previously (Thompson et al. 1998), using GraFit5 (Erithacus Software).

**Competition assays:** 30 nM GST-PAK, ACK or WASP or 50 nM TOCA1-His₆ was attached to a fluromicrosphere as above in the presence of equimolar Cdc42Δ7Q61L·[^³H]GTP or full length Cdc42Q61L·[^³H]GTP. 12 x serial dilutions of each competitor (ACK GBD, TOCA1 HR1, HR1SH3 or N-WASP GBD) were
prepared in SPA Dilution Buffer in order to achieve the final concentrations that are indicated in the figures when 25 µL of each concentration was added to 175 µL of the SPA Reaction Mix. The reaction mixes were prepared with the anti-GST or anti-His antibody and supplemented with the GST- or His- fusion protein and equimolar Cdc42. The plate was incubated, spun and counted as above. The data were fitted to competition binding isotherms to obtain $K_d$ values as described previously (Owen et al. 2008). Each experiment was repeated (see Chapter 3) and the mean $K_d$ calculated. The data was analysed in Microsoft Excel to determine confidence intervals based on a 95 % confidence level.

### 2.2.5.2 Isothermal Titration Calorimetry (ITC)

ITC experiments were carried out in ITC buffer on an iTC$_{200}$ Microcalorimeter (MicroCal). 300 µL of buffer or protein was loaded into the cell at the lower concentration (indicated in the results) and 100 µL of buffer or protein at the high concentration in the syringe. Buffer into buffer, protein into buffer and buffer into protein controls were carried out. Each experiment was performed two ways round, e.g. with F-BAR in the syringe and SH3 in the cell and vice versa. A range of concentrations and temperatures (indicated in the results) was used for each pair. Each experiment was performed across 50 min, with 18 injections of 2 µL after one initial injection of 0.4 µL.

### 2.2.5.3 Analytical Ultracentrifugation (AUC)

Sedimentation velocity experiments were performed with an An60 Ti four-hole rotor in an Optima XL-I centrifuge (Beckman Coulter). Standard double-sector Epon centrepieces with sapphire windows were filled with 400 µL of protein at the indicated concentrations. The rotor speed was set to 50000 rpm and temperature to 20 °C. Interference data were acquired in the continuous mode with systematic noise subtracted, without averaging and with radial increments of 0.003 cm. Absorbance data were also collected, at a wavelength of 280 nm. The time between scans was set to zero. The density and viscosity of the buffer and the partial specific volume of the protein were calculated using Sednterp (Laue et al. 1992). Multi-component sedimentation coefficient distributions were obtained from 100 scans by direct boundary modelling of the Lamm equation using Sedfit v14.1 (Schuck 2000). For protein-protein interaction studies, the proteins were mixed and incubated for more than 1 h before the AUC experiment. The data were acquired as above.
2.2.5.4 Liposome Sedimentation Assays

Liposomes were made according to previously described methods (Walrant et al. 2015) using brain or liver lipids (Avanti Lipids, chloroform stocks). For 0.5 mL of phosphatidyl choline/phosphatidyl serine (PC/PS) liposomes, 35 µL of PC at 10 mg/mL was first added to a round bottom glass tube. 15 µL of PS at 10 mg/mL was added and mixed slowly. For PC/PS/PI(4,5)P₂ liposomes, the volume of PC was reduced to 30 µL and 50 µL of PI(4,5)P₂ was added last. 5 µL of methanol was added. The lipids were dried using N₂ whilst rotating the glass tube so that the bottom was coated with an even layer of lipid. The lipids were further dried in a vacuum dessicator for 1 h. 0.5 mL of buffer (see below) was added to the dried lipids to achieve a final lipid concentration of 1 mg/mL with 70 % PC and 30 % PS or 60 % PC, 30 % PS and 10 % PI(4,5)P₂. After 10 minutes, the mixture was agitated by pipetting and shaking, then sonicated in a water bath for 10 min at room temperature.

15 µL of the liposome suspension was added to 35 µL protein solution in the appropriate buffer and mixed gently. The HisMBP-F-BAR spin assays were carried out in 20 mM HEPES pH 7.5/50 mM NaCl with a final concentration of 2 µM HisMBP-F-BAR. The SH3 or HR1 domains were added to the protein solution prior to addition of liposomes to achieve final concentrations of 5, 10, 20 or 50 µM SH3 or 5, 15 or 30 µM HR1. The protein solutions were incubated with liposomes for 10 minutes at room temperature before centrifugation at 100,000 g for 15 minutes. The supernatant was collected and 12.5 µL of 4x NuPAGE LDS sample buffer (Thermo Scientific) with 40 mM β-mercaptoethanol was added. The pellets were resuspended in 50 µL of buffer and 12.5 µL sample buffer added. 15 µL of each sample were assessed by SDS-PAGE.

Spin assays with Cdc42 were carried out in 50 mM Tris:HCl pH 7.5 with 150 mM NaCl and 5 mM MgCl₂ and with a final concentration of 9 µM Cdc42.

2.2.5.5 Pyrene actin assays

Pyrene actin assays were carried out as previously described in (Walrant et al. 2015). *Xenopus* high speed supernatant (provided by A. Walrant) was used at 5 mg/mL and supplemented with 0.12 mg/mL pyrene actin. TOCA1 HR1 domain or N-WASP GBD was added to 200 µL of the high speed supernatant/pyrene actin mixture at the
concentrations indicated before the mixture was transferred into a 200 µL quartz microcuvette. The assays were initiated by the addition of 5 µL of liposomes (section 2.2.5.4) per 200 µL of lysate. The polymerisation of pyrene actin was measured using a fluorescence spectrophotometer set to $\lambda_{ex} = 365$ and $\lambda_{em} = 407$.

2.2.5.6 Circular Dichroism

Sample preparation: Cdc42 samples were buffer exchanged into 20 mM NaH$_2$PO$_4$/Na$_2$HPO$_4$ pH 7.4, 150 mM NaF using 3.5 mL PD10 desalting columns (GE Healthcare). 0.5 mL samples were prepared at 0.1 mg/mL protein concentration and octyl glucoside was added to the concentrations indicated in Chapter 6.

Secondary structure analyses: The ellipticity of each sample was measured between 185 and 260 nm. Buffer-only control data was recorded and subtracted from the experimental data. The helical content was estimated using the SELCON3 programme in DichroWeb (Sreerama and Woody 1993; Whitmore and Wallace 2008), by comparison with data sets 3 and 7. The normalised root-mean-square deviation (NRMSD) for each data fit is given in Chapter 6.

Thermal melts: The absorbance at 222 nm was recorded at 2 °C increments between 18 and 86 °C.

2.2.5.7 Nucleotide Hydrolysis Assays

Protein samples were incubated in 50 mM Tris:HCl, 150 mM NaCl, 5 mM DTT, 5 mM MgCl$_2$ and 0.1 % (w/v) octyl glucoside at 25 °C for 100 h. 50 µg samples were taken at various time points (details in Chapter 6) and analysed by HPLC (section 2.2.4.2).
### 2.2.6 NMR Spectroscopy

#### 2.2.6.1 NMR Experiments

<table>
<thead>
<tr>
<th>Protein</th>
<th>Conc (mM)</th>
<th>Buffer</th>
<th>Spectra Acquired</th>
<th>Spectrometer</th>
<th>Temp (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unlabelled TOCA1 HR1</td>
<td>0.9</td>
<td>50 mM sodium phosphate pH 5.5, 25 mM NaCl, 10 % D$_2$O</td>
<td>2D NOESY 2D TOCSY</td>
<td>Bruker DRX500</td>
<td>298</td>
</tr>
<tr>
<td>$^{15}$N TOCA1 HR1</td>
<td>0.7</td>
<td>50 mM sodium phosphate pH 5.5, 25 mM NaCl, 10 % D$_2$O</td>
<td>$^{15}$N-HSQC</td>
<td>Bruker DRX500</td>
<td>298</td>
</tr>
<tr>
<td>$^{15}$N/$^{13}$C TOCA1 HR1</td>
<td>0.9</td>
<td>20 mM sodium phosphate pH 7.4, 150 mM NaCl, 10 % D$_2$O</td>
<td>$^{15}$N-HSQC $^{13}$C-HSQC HNCA HNCA CB HNCO HN(CO)CA HN(CO)CAB</td>
<td>Bruker DRX500</td>
<td>298</td>
</tr>
<tr>
<td>$^{15}$N TOCA1 SH3</td>
<td>0.36</td>
<td>20 mM Tris:HCl pH 7.0, 250 mM NaCl, 1 mM EDTA, 10 % D$_2$O</td>
<td>2D NOESY 2D TOCSY</td>
<td>Bruker DRX500</td>
<td>298</td>
</tr>
<tr>
<td>$^{15}$N/$^{13}$C Cdc42Δ7</td>
<td>0.65</td>
<td>50 mM sodium phosphate pH 5.5, 25 mM NaCl, 5 mM MgCl$_2$, 5 mM DTT, 10 % D$_2$O</td>
<td>$^{15}$N-HSQC $^{13}$C-HSQC constant time $^{13}$C-SOFAST-HMQC $^{13}$C-NOESY-HSQC</td>
<td>Avance AV600</td>
<td>283, 288, 293, 298, 303, 308</td>
</tr>
<tr>
<td>$^{15}$N Cdc42Δ7</td>
<td>0.65</td>
<td>50 mM Tris:HCl pH 7.5 with 150 mM NaCl, 5 mM DTT, 5 mM MgCl$_2$ and 10 % D$_2$O</td>
<td>$^{15}$N-HSQC $^{13}$C-HSQC constant time $^{13}$C-SOFAST-HMQC $^{13}$C-NOESY-HSQC</td>
<td>Avance AV600</td>
<td>283, 288, 293, 298, 303, 308</td>
</tr>
<tr>
<td>$^{15}$N Cdc42Δ3</td>
<td>0.6</td>
<td>50 mM Tris:HCl pH 7.5 with 150 mM NaCl, 5 mM DTT, 5 mM MgCl$_2$ and 10 % D$_2$O</td>
<td>$^{15}$N-HSQC</td>
<td>Avance AV600</td>
<td>283, 288, 293, 298, 303, 308</td>
</tr>
</tbody>
</table>

Table 2-11
For effector binding experiments, all of the experiments outlined below were recorded at 25 °C in 50 mM sodium phosphate pH 5.5, 25 mM NaCl, 5 mM MgCl₂, 5 mM DTT, 10 % D₂O.

<table>
<thead>
<tr>
<th>Labelled Protein</th>
<th>Conc (mM)</th>
<th>Unlabelled Protein</th>
<th>Titration Points</th>
<th>Spectra Acquired</th>
<th>Spectrometer</th>
<th>Temp (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOCA1 HR1</td>
<td>0.2</td>
<td>Cdc42Δ7</td>
<td>1:0 1:0.25 1:0.5</td>
<td>15N-HSQC</td>
<td>Bruker DRX500 Avance AV600</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1:1 1:1.5 1:2</td>
<td>13C-HSQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1:3 1:4</td>
<td>13C-SOFAST-HMQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>Cdc42Δ7</td>
<td>1:0 1:3</td>
<td>13C-HSQC constant time</td>
<td>Avance AV600</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>Cdc42Δ7, N-WASP GBD</td>
<td>1:1:0 1:1:0.25 1:1:0.5 1:1:0.75 1:1:1</td>
<td>15N-HSQC</td>
<td>Bruker DRX500</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>Cdc42Δ7</td>
<td>1:0 1:1.2</td>
<td>15N-HSQC</td>
<td>Bruker DRX500 Avance AV600 Avance AV800</td>
<td>288, 298, 303</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1:0 1:1.2</td>
<td>13C-NOESY-HSQC</td>
<td>Bruker DRX500</td>
<td>303</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>TOCA1 HR1</td>
<td>1:0 1:0.25 1:0.5 1:2.2</td>
<td>15N-HSQC</td>
<td>Avance AV600</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13C-HSQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>TOCA1 HR1</td>
<td>1:2:2</td>
<td>15N-HSQC</td>
<td>Avance AV600</td>
<td>288, 298, 303</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13C-HSQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.65</td>
<td>TOCA1 HR1</td>
<td>1:0 1:0.5 1:2</td>
<td>13C-HSQC constant time</td>
<td>Avance AV600</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13C-SOFAST-HMQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>N-WASP GBD</td>
<td>1:0 1:0.25 1:0.5 1:0.75 1:1</td>
<td>15N-HSQC</td>
<td>Bruker DRX500</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.145</td>
<td>N-WASP GBD, TOCA1 HR1</td>
<td>1:1:0 1:1:0.25 1:1:0.5 1:1:0.75 1:1:1 1:1:5</td>
<td>15N-HSQC</td>
<td>Bruker DRX500</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td>0.36</td>
<td>TOCA1 HR1</td>
<td>1:0 1:0.25 1:0.5 1:0.75 1:1</td>
<td>15N-HSQC</td>
<td>Bruker DRX500</td>
<td>298</td>
</tr>
</tbody>
</table>

Table 2-12
2.2.6.2 NMR Data Processing

NMR data were processed using Azara (W. Boucher, University of Cambridge) using processing scripts provided by Helen Mott.

2.2.6.3 Chemical Shift Mapping

The chemical shift changes, δ, were calculated using:

\[ \delta = \sqrt{\delta_{1H}^2 + (0.15\delta_{15N})^2} \]

\(\delta_{1H}\) and \(\delta_{15N}\) are the chemical shift changes for the \(^1\)H and \(^{15}\)N dimensions. Chemical shift differences between Cdc42Δ7 and Cdc42Δ3 were calculated using the same equation. Residues that were not trackable due to line broadening were assigned a \(\delta\) value larger than the maximum calculated \(\delta\) for the data set and residues that were too overlapped to be reliably assigned in the complex spectra were assigned \(\delta = 0\) and are indicated in the results. Changes that were more than the mean \(\delta\) across the spectrum (HR1) or mean plus one standard deviation (Cdc42) were classed as significant.

2.2.7 Structure Calculations and Structure Modelling

2.2.7.1 Structure Calculations

Structures were calculated iteratively using Aria 2.3.1 (Rieping et al. 2007) interfaced to CNS 1.3 (Brünger et al. 1998). The PROSLQ force field was used for non-bonded parameters. The dihedral angle restraints and distance restraints were derived as described in Chapter 4. The flat-bottom calibration curve was used for calibration of the \(^{15}\)N- and \(^{13}\)C-NOESY-HSQC experiments, from which distance restraints were derived. 100 structures were calculated in the final iteration and the 50 lowest energy structures were water refined. The 35 lowest energy structures constituted the final ensemble that was submitted to the PDB (PDB code: 5FRG).

2.2.7.2 Structure Validation

An unminimized average structure was calculated from the ensemble and the backbone RMSD of each structure from the average was determined using CNS scripts provided by Helen Mott. The RMSDs of the backbone and sidechain atoms were assessed across the ensemble for the whole protein and for the ordered regions only (Helix 1: 342-379 and Helix 2: 386-420). The structure with the lowest RMSD
from the average was taken as the most representative structure. The experimental data and the final structure ensemble was assessed using iCING (Doreleijers et al. 2012), which integrates a number of processes including dihedral angle validation using PROCHECK_NMR (Laskowski et al. 1996). MolProbity version 3.2.1 (Chen et al. 2010) was also used to assess the dihedral angles.

2.2.7.3 **HADDOCK Modelling**

10000 models of Cdc42-HR1<sup>TOCA1</sup> were calculated in the first stages of rigid body docking. The 400 lowest energy structures were carried forwards to semi-flexible simulated annealing and water refinement. Details of the ambiguous interaction restraints (AIRs) are given in Chapter 5. The models were clustered based on the fraction of common contacts (FCC).

2.2.8 **Software and Online Servers**

<table>
<thead>
<tr>
<th>Task</th>
<th>Software/Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence alignments</td>
<td>ClustalX version 2.0 (Larkin et al. 2007) or Expresso (Armougom et al. 2006)</td>
</tr>
<tr>
<td></td>
<td>as indicated in the figure legends</td>
</tr>
<tr>
<td>Sequence-based secondary</td>
<td>PSIPRED (Jones 1999; Buchan et al. 2013)</td>
</tr>
<tr>
<td>structure predictions</td>
<td></td>
</tr>
<tr>
<td>Analysis of SPA data</td>
<td>GraFit5 (Erithacus Software)</td>
</tr>
<tr>
<td>Analysis of AUC data</td>
<td>Sednterp (Laeu et al. 1992) and Sedfit v14.1 (Schuck 2000)</td>
</tr>
<tr>
<td>Images of protein structures</td>
<td>MacPyMol (<a href="http://pymol.sourceforge.net">http://pymol.sourceforge.net</a>)</td>
</tr>
<tr>
<td>NMR data processing</td>
<td>Azara (Wayne Boucher, University of Cambridge)</td>
</tr>
<tr>
<td>NMR assignments</td>
<td>CCPN Analysis v2.3 (Vranken et al. 2005)</td>
</tr>
<tr>
<td>Structure Validation</td>
<td>iCING (Doreleijers et al. 2012) and MolProbity version 3.2.1 (Chen et al. 2010)</td>
</tr>
<tr>
<td>Definition of secondary</td>
<td>Stride (Heimig and Frishman 2004)</td>
</tr>
<tr>
<td>structure from PDB files</td>
<td></td>
</tr>
<tr>
<td>Definition of contacts from</td>
<td>CCP4 Contact v6.4.0 (Winn et al. 2011)</td>
</tr>
<tr>
<td>NMR structures</td>
<td></td>
</tr>
<tr>
<td>Assessing solvent accessibility</td>
<td>NACCESS (Hubbard and Thornton 1993)</td>
</tr>
<tr>
<td>Analysis of helices</td>
<td>HELANAL-Plus (Kumar and Bansal 2012)</td>
</tr>
<tr>
<td>Rigid body docking</td>
<td>HADDOCK online server (Dominguez et al. 2003; de Vries et al. 2010; van Zundert et al. 2016)</td>
</tr>
<tr>
<td>HPLC data collection and</td>
<td>DIONEX Chromeleon version 6,50SP1 (Thermo Fisher)</td>
</tr>
<tr>
<td>analysis</td>
<td></td>
</tr>
<tr>
<td>Size exclusion chromatography</td>
<td>Äkta Pure (GE Healthcare) with Unicorn 6.3 (GE Healthcare)</td>
</tr>
<tr>
<td>and ion exchange</td>
<td></td>
</tr>
<tr>
<td>Production of figures</td>
<td>Inkscape 0.91 (inkscape.org).</td>
</tr>
</tbody>
</table>

**Table 2-13**
3 Biochemical Studies of TOCA1

3.1 Introduction
As discussed in Chapter 1, TOCA1 was identified as a Cdc42 effector protein acting upstream of N-WASP in the pathway of Cdc42-dependent actin polymerisation (Ho et al. 2004). It was shown to be characteristic of an effector in that it bound to Cdc42-GTPγS but not to Cdc42-GDP. It contains a known Rho binding module known as the HR1 domain, mutation of which (M383I/G384S/D385T) abolishes Cdc42 binding \textit{in vitro} and has significant functional effects on actin-dependent processes (Ho et al. 2004; Bu et al. 2009; Bu et al. 2010; Bai and Grant 2015). Despite the effect of the MGD mutation, it has not yet been determined whether the HR1 domain is sufficient for Cdc42-TOCA1 binding or whether other regions of the protein contribute to the interaction (Ho et al. 2004).

The role of the Cdc42-TOCA1 interaction in the pathway of Cdc42-dependent actin polymerisation is not yet known, but it may relate to the regulation of TOCA1. The activity of TOCA1, like N-WASP, is influenced by Cdc42 and PI(4,5)P$_2$ and so TOCA1 may be autoregulated in a similar manner to N-WASP. Specifically, interactions between the F-BAR and SH3, F-BAR and HR1 and HR1 and SH3 domains have all been suggested based on homologous proteins and previous studies of TOCA1 (Krugmann et al. 2001; Wang et al. 2009; Bu et al. 2010; Rao et al. 2010; Dr J Peterson, Fox Chase, personal communication).

The aim of this chapter was to study TOCA1 and its interaction with Cdc42 biochemically, in order to determine the affinity of the interaction, to assess whether the HR1 domain is sufficient for Cdc42 binding and to investigate possible autoregulatory mechanisms of TOCA1.
3.2 TOCA1-Cdc42 Binding

3.2.1 Expression and Purification of the TOCA1 HR1 Domain

3.2.1.1 Definition of Domain Boundaries

The HR1 domain is a well known Rho-binding module comprising an antiparallel coiled-coil (Maesaki et al. 1999; Modha et al. 2008; Hutchinson et al. 2011; Hutchinson et al. 2013). Another TOCA family member, CIP4, has previously been studied structurally (Kobashigawa et al. 2009) and so a sequence alignment was carried out between the CIP4 construct used in the structural studies and TOCA1 using ClustalX (Larkin et al. 2007) (Figure 3.1). The two HR1 domains were found to be homologous (68 % sequence identity) and a sequence-based secondary structure prediction for TOCA1, carried out using PSIPRED (Jones 1999; Buchan et al. 2013), was broadly in agreement with the CIP4 helices. The predicted helices are shown by underlining in Figure 3.1.

| TOCA1 324 | LEFKIKGPALEDFSLPPEQRRKRLQRIDSELQKRMDOKDLNKMKDYYEK |
| CIP4  332 | ----------------TEDFSHLPPEQIRRKLQOQLEERSRELQKEVDQREALKKMKDYYEK |

**Figure 3.1. Defining the TOCA1 HR1 Domain by Sequence Alignment.** A sequence alignment was carried out with full length TOCA1 and the CIP4 HR1 domain (as used in structural studies) using ClustalX version 2.0. A section of the alignment focusing on the HR1 domain is shown. The degree of conservation for each residue is indicated below the sequence and the secondary structure elements coloured as indicated.

The N-terminal region of the CIP4 HR1 domain exhibits a series of turns (residues 334-338) in the NMR structure (Kobashigawa et al. 2009) and so it is possible that further secondary structure extends N-terminal to this construct. For this reason, an N-terminally extended construct of TOCA1 (324-426) was initially produced but later found to undergo N-terminal degradation up to residue 330 (Joanna R Watson, Part III.
Project). Therefore, a construct beginning at residue 330 and ending at 426 to correspond with the C-terminus of the CIP4 HR1 construct was engineered and is referred to as HR1\textsuperscript{TOCA1} hereafter.

### 3.2.1.2 Protein Expression and Purification

HR1\textsuperscript{TOCA1} was cloned into the expression vectors pGEX-6P-1 and pGEX-HisP as described in Chapter 2. Small scale expression trials from the pGEX-6P-1 construct were then carried out in \textit{E. coli} BL21 and CodonPlus RP cells, which contain a plasmid encoding tRNAs for rare Arginine and Proline codons. These were tested because the TOCA1 HR1 DNA contains rare Arginine and Proline codons. The gel images in Figure 3.2 show samples that were normalized to the optical density of the bacterial cultures and so show the protein expression for a given cell density. From this, it is clear that the \textit{E. coli} BL21 cells induced with 0.1 mM IPTG produced the most soluble protein (the band of interest is marked with a white asterisk in Figure 3.2). Large scale expressions were therefore carried out in \textit{E. coli} BL21 induced with 0.1 mM IPTG at 20 °C for 20 h.

![Figure 3.2. Small Scale Expression Trials of HR1\textsuperscript{TOCA1}](image)

**Figure 3.2. Small Scale Expression Trials of HR1\textsuperscript{TOCA1}**. A 15 % Laemmli gel showing samples taken from the indicated bacterial strains transformed with pGEX-6P-1-HR1\textsuperscript{TOCA1}, grown at 20 °C for 20 h. U – uninduced, I – induced (with indicated concentration of IPTG), S – soluble, P – pellet. The band showing the most soluble protein is marked with a white asterisk and the condition chosen for large scale expression is marked with a red asterisk.
The protein was purified using glutathione-agarose as described in Chapter 2, an example of which is shown in Figure 3.3A. HR1\textsuperscript{TOCA1} was removed from the GST-tag by proteolysis with 3C Protease (Figure 3.3B) and further purified by size exclusion chromatography (Figure 3.3C-D). The protein eluted in a single peak (Figure 3.3C) and was free from significant proteinous contaminants (Figure 3.3D). 20 mg of purified protein was obtained from 3 L of E. coli, which was frozen in 500 µL aliquots at 0.9 mM at -80 °C. HR1\textsuperscript{TOCA1} appeared to be stable following sequential freeze-thawing and at room temperature for 15 days (seen by a lack of precipitation) and was not subject to continued proteolysis, as shown in Figure 3.3E.
Figure 3.3. Purification of HR1^TOCA1. A) A 15 % Laemmli gel showing samples taken from each wash of the glutathione agarose beads (washes) and from a 50 % bead slurry (beads). B) An 18 % Laemmli gel showing gel samples taken from the beads post-3C protease cleavage (B1), bead washes and a final bead sample (B2). C) The A280 trace showing the purification by size exclusion chromatography on a Superdex 16/60 S30 column in 50 mM sodium phosphate pH 5.5, 25 mM NaCl. D) An 18 % Laemmli gel showing a 1 in 10 dilution of the sample that was loaded onto the gel filtration column (L) and samples taken from each 2 mL fraction corresponding to the void and peak 1. E) An 18 % Laemmli gel showing samples taken after freeze-thawing as indicated and after incubation at room temperature (RT) or 4 °C for 15 days.
The HR1 domain was also expressed from pGEX-6P as an N-terminal GST- and C-terminal His6- fusion protein for use in direct binding assays, where the HR1 domain would be immobilised via the His-tag. Following affinity purification, proteolysis of the GST-tag and size exclusion chromatography (Figure 3.4A), the protein sample contained breakdown products (Figure 3.4B). The non-His-tagged HR1 protein was run on the same gel (lane 6) for a size comparison. The breakdown products migrated through the gel similarly to the untagged HR1 domain and so could have resulted from proteolysis of the C-terminal His-tag. The sample was therefore further purified using Nickel-NTA-agarose beads, to which only the His-tagged species would bind.

The protein was incubated with the beads and eluted with imidazole. Figure 3.4C shows samples from the washes and the elutions, confirming that the eluted protein was free from any significant contamination. The imidazole was diluted out by repeated concentration and dilution in 20 mM Tris:HCl pH 7.3, 150 mM NaCl, 5 mM EDTA. After concentration, 8 mg of protein was obtained, giving a concentration of 650 µM in 600 µL. The protein was frozen at -80 °C until required and the concentration of the protein was confirmed by amino acid analysis before use in binding assays.
Figure 3.4. Purification of C-terminally His-tagged HR1^{TOCA1}. A) An $A_{280}$ trace showing the purification of HR1-His$_6$ by size exclusion chromatography on a Superdex 16/60 S30 in 20 mM sodium phosphate pH 7.4, 150 mM NaCl. B) An 18 % Laemmli gel showing samples taken from post-cleavage washes (lanes 1-4), the previously purified untagged HR1 domain (lane 6) and each 2 mL fraction corresponding to the void (lane 7) and peak 1 (lanes 8-12) from the size exclusion chromatography. C) An 18 % Laemmli gel showing samples taken from the supernatant (SN), washes and elutions following incubation on Nickel-NTA-agarose beads.
3.2.2 Expression and Purification of Cdc42Δ7Q61L

The 7 C-terminal residues of most small GTPases are not required for effector interactions, including the interaction between RhoA and the PRK1 HR1a domain (Owen et al. 2003). Furthermore, this region of Cdc42 is unstructured, flexible and prone to proteolysis and is not required for any of the Cdc42-effector interactions studied thus far and so this region is generally removed for structural work. A pGEX-2T-Cdc42Δ7Q61L construct was available from laboratory stocks and the protein had been expressed from this vector previously. The Q61L mutation renders the protein deficient in GTP hydrolysis activity and so locks it in the active state.

The protein was expressed in 3 L of *E. coli* BL21, purified using glutathione-agarose beads (Figure 3.5A) and removed from the GST tag using thrombin protease (Figure 3.5B). The cleaved protein was further purified by size exclusion chromatography (Figure 3.5C) and 64 mg of pure Cdc42 was obtained (Figure 3.5D). Small GTPases are frequently seen as multiple bands in SDS-PAGE, comparable to the doublet seen in Figure 3.5D, presumably due to incomplete denaturation. Nonetheless, mass spectrometry was used to confirm that the sample contained only the expected species with no breakdown products (PNAC, Department of Biochemistry).

The protein was then loaded with [³H]GTP for use in binding assays as described in Chapter 2. 0.7 mg of protein was exchanged in each reaction and approximately 0.25-0.6 mg remained after the exchange and purification. The radiolabelled protein was frozen at 100-250 µM in 120 µL aliquots at -80 °C.
3.2.3 Binding Assays with HR1TOCA1 and Cdc42Δ7Q61L

The interaction between Cdc42 and HR1TOCA1 was investigated using Scintillation Proximity Assays (SPA), the concept of which is illustrated in Figure 3.6. SPAs have been used extensively to study GTPase-effector interactions within the lab and the affinities derived are highly reproducible. They require very small amounts of protein (<1 mg) and a large amount of data can be obtained in a short amount of time. Furthermore, the radioactivity that is used to measure binding is readily incorporated simply by loading the GTPase with [3H]GTP.

In direct SPAs, the effector protein is immobilised on scintillation beads via an affinity tag as shown in Figure 3.6A and the radiolabelled GTPase is added at increasing concentrations. When the GTPase binds to the immobilised effector, the
radioactive decay from the tritiated nucleotide excites the scintillant contained within the beads leading to a measurable emission of blue light. In competition SPAs, another effector with known affinity for the GTPase is immobilised on the beads and equimolar GTPase is added. The effector of interest is titrated in and the competition measured as a decrease in the scintillation signal (Figure 3.6B).

**Figure 3.6. Scintillation Proximity Assays.** A) An illustration of a direct binding SPA. The scintillation beads are coated with protein A, which is used to conjugate an antibody such as an anti-His antibody. The effector protein is immobilised to this antibody via its fusion-tag. The GTPase is added at increasing concentrations and it is brought into close proximity with the scintillation beads when it binds to the effector. The radiolabelled nucleotide causes excitation of the scintillation beads and emission of light in the blue region of the spectrum. This can be measured using a scintillation counter. B) An illustration of a competition SPA. An effector with known affinity for the GTPase is immobilised in the same way as in A and equimolar GTPase added. A competitor is added at increasing concentrations and the binding measured as a decrease in scintillation at higher concentrations as the GTPase is sequestered away from the beads.
A representative binding isotherm for the interaction between Cdc42Δ7Q61L-[³H]GTP and HRTDNA TOCA1 measured by direct SPA is shown in Figure 3.7A. The Cdc42-PAK interaction was included as a positive control (Thompson et al. 1998). TOCA1-Cdc42 binding is evident from the increase in radioactive counts at higher concentrations of Cdc42 and fitting of the data to a direct binding isotherm (Chapter 2) revealed a $K_d > 1 \mu M$. Data from a previous project showed that the binding affinity was not increased in a range of conditions, including several NaCl concentrations (Joanna R Watson, Part III Project).

This modest affinity is outside of the accuracy range for direct SPAs (0.1-1 µM) as the higher concentrations of radiolabelled Cdc42 required to reach saturation would result in non-specific signals. Furthermore, the immobilization of HRTDNA TOCA1 could have a negative impact on the interaction and so the affinity was instead determined using competition SPA (Figure 3.7B-D). GST-ACK GBD, which binds with a high affinity to Cdc42 (Mott et al. 1999), was immobilised on the SPA beads with equimolar Cdc42Δ7Q61L-[³H]GTP. Increasing concentration of untagged HRTDNA TOCA1 were added and the ability of HRTDNA TOCA1 to compete with GBDACK was detected as a decrease in radioactive counts with increasing HRTDNA TOCA1. Self-competition of GBDACK with GST-GBDACK was used as a control and to establish the background counts when Cdc42 is fully displaced.

Free ACK competed with GST-ACK with an affinity of 33 nM (Figure 3.7B), similar to the value obtained by direct binding, 23 nM (Mott et al. 1999). When HRTDNA TOCA1 was tested using the same concentration range, the competition data could be fit to an affinity of 1.8 nM but with an error of 38 nM (Figure 3.7C). Also, the curve did not saturate at the level observed for the GST-ACK vs ACK control (~1300 compared to ~400 counts per minute). The assay was therefore repeated with higher concentration serial dilutions of HRTDNA TOCA1. This assay, a representative example of which is shown in Figure 3.7D, was repeated four times with separate preparations of the HR1 domain and Cdc42. These yielded $K_d$ values of 5.0, 5.0, 5.1 and 6.1 µM. The mean $K_d$ can therefore be given as $5.3 \pm 0.51$ based on a confidence level of 95%.
3.2.4 Purification of Full Length Cdc42Q61L

The C-terminus of Rac1 contains a polybasic sequence that is important for binding to the HR1b domain of PRK1 (Owen et al. 2003; Modha et al. 2008). Deletion of any of the basic residues in the C-terminal region of Rac1 (KKRKRKCLLL) has a negative effect on HR1-domain binding, whilst PAK binding is unaffected (Owen et al. 2003). The C-terminus of Cdc42 (PKKSRRCVLLL) is similar to that of Rac1 and so the C-terminal residues could be required for maximal Cdc42-HR1TOCA1 binding. Therefore,
full-length Cdc42Q61L was cloned into the pGEX-6P-1 expression vector and the protein was expressed and purified similarly to Cdc42Δ7 (Figure 3.8). 11 mg of purified protein was obtained. Mass spectrometry was used to confirm the full length protein (PNAC, Department of Biochemistry) before it was radiolabelled and frozen at -80 °C in 120 µL aliquots at 100-250 µM.

Figure 3.8. Purification of Full Length Cdc42Q61L. A) An A280 trace from the purification of full length Cdc42 on a Superdex 16/60 S75 size exclusion column in 50 mM Tris:HCl pH 7.5, 150 mM NaCl, 5 mM DTT and 5 mM MgCl2. B) A 15 % Laemmli gel showing samples taken from the loaded sample (L) and each 2 mL fraction corresponding to the void and peak 1.

3.2.5 Binding Assays with HRI\textsuperscript{TOCA1} and Full Length Cdc42

Competition SPAs were carried out as with truncated Cdc42. The affinity of full-length Cdc42 for HRI\textsuperscript{TOCA1} ($K_d = 5.39$ µM, Figure 3.9) was comparable with that of truncated Cdc42 ($K_d = 5.3 \pm 0.51$). Self-competition of ACK was included as a control.
(Figure 3.9A). The competition SPA with full length Cdc42 was repeated giving a mean $K_d = 4.2 \pm 2.4$ within a 95 % confidence level.

![Figure 3.9A](image)

**Figure 3.9. Binding Assays with Full Length Cdc42 and HR1$^{TOCA}$.** Representative binding curves from competition SPA experiments carried out with the indicated concentrations of A) GBD$^{ACK}$ or B) HR1$^{TOCA}$ titrated into 30 nM GST-ACK and 30 nM Cdc42Δ7Q61L·[^3H]GTP or full length Cdc42Q61L·[^3H]GTP. The $K_d$ values derived from fitting to a competition binding isotherm were 0.032 ± 0.01 and 0.011 ± 0.01 µM for GBD$^{ACK}$ and 6.05 ± 1.96 and 5.39 ± 1.69 µM for HR1$^{TOCA}$ binding to Cdc42Δ7 or full length, respectively. The errors are curve fitting errors.

### 3.2.6 Binding Assays with Other TOCA1 Domains

The data presented thus far indicate that the affinity of the HR1$^{TOCA}$-Cdc42 interaction is relatively low compared to other Cdc42-effector interactions. It has not yet, however, been determined whether the HR1 domain is solely responsible for the TOCA1-Cdc42 interaction or whether other TOCA1 domains contribute to Cdc42 binding. Both SH3 and BAR domains have been implicated in G protein interactions (ten Klooster et al. 2006; Nakamura et al. 2012) and so the SH3 and F-BAR domains
of TOCA1 are prime candidates for Cdc42 binding. Furthermore, TOCA1 dimerizes via its F-BAR domain and this dimerization could affect Cdc42 binding, for example by presenting two HR1 domains for Cdc42 interactions.

Full-length TOCA1, the F-BAR domain and a truncated construct missing the SH3 domain (Figure 3.10A-B) were kindly provided as His-SNAP fusion proteins by Helen M Fox (Gurdon Institute UK). These constructs were tested for binding to full length Cdc42 in direct SPAs (Figure 3.10C). The isolated F-BAR domain showed no binding to Cdc42. Full length TOCA1 and ΔSH3 TOCA1 bound with micromolar affinity, similarly to the isolated HR1 domain. Repeats of this assay with fresh preparations of each protein gave comparable results.

The $K_d$s measured using these direct binding assays were semi-quantitative as the affinities were outside of the accuracy range. Competition binding assays would provide quantitative estimates of the affinities but were not possible as the proteins could not be produced at high enough concentrations.
Figure 3.10. Binding of TOCA1 Proteins to Cdc42. A) Domain maps showing the TOCA1 constructs tested for Cdc42 binding. B) A 10% Laemmli gel showing samples of each construct. FL* indicates a W518K mutant of full length TOCA1, which was not included in the SPAs. C) The binding curves derived from direct SPAs, in which the indicated concentrations of full length Cdc42Q61L·[^3]H]GTP were incubated with 30 nM GST-ACK, His-SNAP-full length TOCA1, His-SNAP-F-BAR$^{TOCA1}$ or His-SNAP-ΔSH3$^{TOCA1}$. The SPA signal was corrected by subtraction of control data with no GST-ACK or HisSNAP-TOCA1. The data, expressed here as a percentage of the maximum signal, were fitted to a direct binding isotherm to give an apparent $K_d$. 
3.2.7 Other TOCA Family Members

To determine whether the micromolar affinity interaction is common to the TOCA family members, FBP17 (Wu et al. 2013) and CIP4 (Aspenström 1997), the HR1 domains were cloned and expressed for use in competition SPAs.

3.2.7.1 Defining the Domain Boundaries for FBP17 and CIP4

The human FBP17 HR1 domain boundaries were determined based on a sequence alignment with *X. tropicalis* TOCA1 (Figure 3.11). There is a 48 residue insertion in FBP17 relative to TOCA1, N-terminal to the HR1 domain and so the N-terminal domain boundary was less certain. Therefore, an FBP17 HR1 construct that was N-terminally extended relative to the TOCA1 construct, as shown in Figure 3.11, was designed to ensure the complete HR1 domain was incorporated. This HR1 construct was cloned into pGEX-6P-1 and a CIP4 construct that is identical to the construct used for structural studies (Kobashigawa et al. 2009) was cloned from *H. sapiens* cDNA as described in Chapter 2.

| TOCA1  | 300 STPKQESLKPDRVTGKAKWLFGKKPK--------------------------- 330 |
| FBP17  | 301 T---KACCKSDPKVPS-KSNKLFPIKNNKPPPPASSSAPVNGPHSKQQREPLS 357 |
| CIP4   | 266 --LSIESKVSKPFLASFRSRLR-----GSRG------------------------ 328 |

| TOCA1  | 331 -------------------------------GPALEDFSHPPEQRRKLQRIDELESRLQKE 363 |
| FBP17  | 358 HRFNFMTSKKHCFRSLKRGKSLKGLSTPEDSFLPQRRKLQKVDMLNKDQKE 417 |
| CIP4   | 329 -------------------------------TVVTEDFSLPPEQRRKLQQGLEERSRLQKE 362 |

| TOCA1  | 364 MDQKDALNKMDVVEKNPQMDSSLHPKIAETTSNIEKREIHKNEAWLSEVEGK-VS 422 |
| FBP17  | 418 MDQKDALNKMDVVEKNPQMDSSSSLHPKIAETTSNIEKREIHKNEASLSEVEGK-VS 476 |
| CIP4   | 463 MDQKDALNKMDVVEKNPQMDSPLEQIAETTSNIEKREIHKNEASLSEVEGK-VS 422 |

| TOCA1  | 423 OR---SERHRS----------AEANHLVAQGREGSPEGYTEDANQGRVQPAHPEFDDEFD 473 |
| FBP17  | 477 ARNDSRQRQGLYQDQMNQTVNCAQDRESGPQTYEQEPEVKTP---TEFDDEDFD 533 |
| CIP4   | 423 HRBLSRHARPPDPP-ASAPPSNSSNSASQDTSESSEPPSESQDTDP--YTEFDEDFE 480 |

Figure 3.11. A Sequence Alignment of the TOCA Family HR1 Domains. A sequence alignment carried out using ClustalX. The degree of conservation for each residue is shown below the sequence and the HR1 domain constructs cloned are highlighted yellow, magenta or green as indicated.
3.2.7.2 Protein Expression and Purification

Figure 3.12 shows purifications of the CIP4 and FBP17 HR1 domains. 5 mM EDTA was included in all of the buffers prior to size exclusion chromatography to reduce the activity of any contaminating proteases. The fractions from size exclusion chromatography of CIP4 showed a protein at the correct size for the CIP4 HR1 domain, with no visible breakdown or proteinaceous impurities. 21.6 mg of protein was obtained, concentrated to 1.5 mM and frozen at -80 °C in 400 µL aliquots.

For the FBP17 HR1 domain purification, approximately 22 mg of protein was obtained but the sample appeared to contain 3 species that were only partially resolvable by size exclusion chromatography (Figure 3.12F). The highest molecular weight species (running at ~15 kDa) eluted from the size exclusion chromatography column slightly earlier than the lowest molecular weight species (running at ~10 kDa) indicating that there were multiple species present rather than a single species running as a triplet as might be caused by incomplete denaturation.

Fractions 33-35 contained most of the higher molecular weight species, whilst fractions 36-37 comprised mainly the two lower molecular weight species. Fractions 33-35 were analysed by mass spectrometry (PNAC, Department of Biochemistry, Appendix 1). The dominant species fit to ~14.2 kDa, bigger than the 12.3 kDa expected for the FBP17 HR1 domain. The other two major species present in the sample fit to 12.3 kDa and 11.0 kDa (Appendix 1A). MALDI-ISD, used to detect a partial sequence ladder at the N-terminus (Appendix 1B), confirmed that the dominant species in fractions 33-35 comprised the correct N-terminal sequence and that the 14.2 kDa species is a C-terminal extension of the expected HR1 domain, with an identical N-terminus, perhaps arising from stop codon readthrough.
Figure 3.12. Purification of HR1\textsuperscript{CIP4} and HR1\textsuperscript{FBP17}. A-D) The HR1\textsuperscript{CIP4} purification: A) a 12 % Laemmli gel showing samples taken from each wash of the glutathione agarose beads (Washes) and from a 50 % bead slurry (Beads); B) an 18 % Laemmli gel showing gel samples taken from the post-cleavage bead washes and final bead samples; C) an \( A_{280} \) chromatogram showing the purification by size exclusion chromatography, performed on a Superdex 16/60 S75 in MT-TBS; D) an 18 % Laemmli gel showing samples taken from each 2 mL fraction corresponding to the void, peak 1 and peak 2. E-F) The HR1\textsuperscript{FBP17} purification: E) an \( A_{280} \) chromatogram showing the purification by size exclusion chromatography performed on a Superdex 16/60 S75 in MT-TBS; F) an 18 % Laemmli gel showing samples taken from each 2 mL fraction corresponding to the void, peak 1 and peak 2. Fractions 33-37 are boxed and labelled.
3.2.7.3 Competition Binding Assays

The affinity of each HR1 domain for Cdc42 was determined using competition SPAs, analogous to those carried out with HR1TOCA1. Binding assays were carried out with both of the FBP17 samples (fractions 33-35 and 36-37) for comparison. The two samples of the FBP17 HR1 domain bound with micromolar affinities (29 and 11 µM), with fractions 36-37 binding slightly tighter (Figure 3.13A-B). Given the curve fitting errors, the difference is unlikely to be significant. The affinity of the CIP4 HR1 domain was also in the low micromolar range (5 µM) (Figure 3.13C). The assays were repeated and the mean $K_d$s calculated (CIP4 $K_d = 5.1 - 5.4$ µM, $n = 2$; FBP17 $K_d = 16.8 \pm 6.3$ µM based on a 95 % confidence level, $n = 3$).

![Graphs showing competition binding assays](image)

**Figure 3.13. Competition SPAs with Cdc42 and the CIP4 and FBP17 HR1 Domains.** Competition experiments were carried out with the indicated concentrations of A) HR1$^{FBP17}$ least pure fractions, B) HR1$^{FBP17}$ purest fractions and C) HR1$^{CIP4}$, titrated into 30 nM GST-ACK and 30 nM full length Cdc42Q61L-[³H]GTP. The $K_d$ values derived from fitting to a competition binding isotherm are shown with their curve fitting errors.
3.2.8 The Oligomeric State of the HR1 Domain

It has been assumed thus far that the TOCA family HR1 domains are monomeric. The size exclusion chromatograms of each HR1 domain comprise a single peak, consistent with a single species, but the HR1 domain has not been formally been shown to be monomeric. An obligate dimer, for example, would not be inconsistent with the size exclusion chromatograms. Oligomerisation of the HR1 domain, whether physiological or as an artifact of *in vitro* conditions, could affect HR1-Cdc42 binding and so AUC was employed to investigate the oligomeric state of HR1<sub>CIP4</sub>. The oligomeric state at the high concentrations needed for NMR experiments was also of interest prior to embarking on the structural studies discussed in Chapter 4.

AUC is a powerful method for assessing oligomeric state, aggregation and protein-protein binding in solution (reviewed in Cole et al. 2008). Protein solutions are simply spun at high speeds and their sedimentation is measured using absorbance or interference optics. The sedimentation coefficient (s) is directly measured in the experiment. This is proportional to the buoyant molar mass and inversely proportional to the frictional ratio. The sedimenting boundary changes shape over time due to diffusion, described by the translational diffusion coefficient (D), which is also measured. The time evolution of the boundary can be defined in terms of s and D by an equation known as the Lamm equation, which in turn can be used to describe the nature of the sedimenting species within a solution. Modern methods of model dependent analysis of AUC data allows for detailed descriptions of the species present based simply on the values for s and D. Details of the mathematical treatment of the Lamm equation and model dependent analysis of AUC data can be found in Schuck 2000, Balbo and Schuck 2005 and Cole et al. 2008.

Two concentrations (0.4 and 0.8 mM) of the HR1 domain were analysed in a sedimentation velocity experiment, as described in Chapter 2 and based on methods described by (Cole et al. 2008). The data is shown in Figure 3.14A. The data was fit using the continuous c(s) distribution model in Sedfit (Schuck 2000; Balbo and Schuck 2005) and c(s) distributions are shown in Figure 3.14A and B. This model is based on numerical solutions to the Lamm equations. Data is simulated based on these solutions and fit to the experimental data using least squares fitting. The fitting
produces a distribution of s-values present within the data and also provides an estimation of the frictional ratio (a measure of how spherical the protein is). The quality of the fit is described by the RMSD between the simulated and experimental data. The RMSDs are given in Figure 3.14B and C and the residuals are shown in Figure 3.14A below the raw data.

![Figure 3.14](image)

**Figure 3.14. AUC Analysis of HR1^{CIP4}.** Sedimentation velocity data of HR1^{CIP4}. The data was fit with the continuous c(s) distribution model. A) The fringes and the residuals from the fit for the 0.4 mM sample are shown. B) The component sedimentation coefficient distribution (c(s) distribution) for 0.4 mM HR1^{CIP4}. C) The c(s) distribution for 0.8 mM HR1^{CIP4}. The c(s) distributions for each concentration show a single species with s^{0}_{20,w} values indicated, assuming a uniform frictional ratio of \( F_{k,w} = 1.36 \) that was derived during the data fitting. The expected molecular weight of HR1^{CIP4} is 11.8 kDa.

The fits were of good quality as indicated by the low RMSDs (<10 % of the maximum signal). Each distribution contained two peaks with s=0 and s=1.3 or 1.2. The peak at s=0 results from buffer mismatch between the protein sample and buffer reference.
cell. The peak at $s=1.3$ for the 0.4 mM sample or 1.2 for the 0.8 mM sample represents a single oligomeric species of HR1$^{CIP4}$. The fitted data was used to derive an estimation of the molecular mass, assuming a uniform frictional ratio of 1.36, in order to identify the species with $s=1.2$ or 1.3. The calculated masses, indicated on the $c(s)$ distributions shown in Figure 3.14, were indicative of monomeric HR1 domain.

The concentrations chosen were far in excess of those used in binding assays and more in line with those needed for future NMR experiments. The high protein concentrations used, however, led to deviations in sedimentation behaviour, which cause hydrodynamic and thermodynamic non-ideality such that the diffusion coefficient (D) and the sedimentation coefficient ($s$) become concentration dependent (Balbo and Schuck 2005). The concentration-dependence of $s$ can be seen in the small difference between the $s$-value for HR1$^{CIP4}$ at the two different concentrations. The estimated molecular weight (12.1 and 10.3 kDa) was evidently less accurate for the higher concentration sample exhibiting the most ‘non-ideal’ sedimentation behaviour but both were reasonably close to the actual molecular weight (11.8 kDa).

### 3.2.9 Measuring Cdc42-HR1$^{CIP4}$ Binding with AUC

AUC was employed as an alternative, direct binding, in solution method to confirm the affinity of the Cdc42-HR1 interaction measured by SPA. Cdc42Δ7Q61L was first loaded with the non-hydrolysable GTP analogue, GMPPNP, as described in Chapter 2 to ensure that the sample was locked in the active, effector-binding state. HPLC was used to confirm that the protein was 100 % GMPPNP-bound as shown in Figure 3.15.

![Figure 3.15. HPLC Analysis to Confirm GMPPNP Loading of Cdc42. HPLC traces showing nucleotide standards (GDP, GTP and GMPPNP) and nucleotide samples from the protein as indicated. Samples were run on a 1.5 mL Partisphere SAX HPLC column (Whatman) in 0.6 M Ammonium Phosphate pH 4.0.](image)
AUC experiments were carried out with 5 µM Cdc42Δ7Q61L·GMPPNP, 5 µM HR1CIP4 or both 5 µM Cdc42Δ7Q61L·GMPPNP and 5 µM HR1CIP4 pre-incubated for >1 hour (Figure 3.16A-C). Further experiments were carried out with 45 µM Cdc42Δ7Q61L·GMPPNP, 45 µM HR1CIP4 and with both 45 µM Cdc42Δ7Q61L·GMPPNP and 45 µM HR1CIP4 pre-incubated for >1 hour (Figure 3.16D-F). Interference optics were used for detection due to the vastly different extinction coefficients of the two proteins and the low A280 of HR1CIP4 at 5 µM. Figure 3.16 shows the c(s) distributions for each experiment. Figure 3.16A, B, D and E show single peaks for Cdc42Δ7Q61L and HR1CIP4 when run alone, with resolvable s-values (2.1 and 1.3 respectively). The estimated molecular weights for HR1CIP4 and Cdc42Δ7Q61L (12.5 kDa and 20.1 kDa) based on the fitted data were close to the expected values (11.8 kDa and 20.7 kDa).

When 5 µM of Cdc42 and HR1CIP4 were run together, there were two overlapping peaks (Figure 3.16C). The s-value of the first peak (1.6) was slightly larger than that of HR1CIP4 when run alone (1.3). This indicates exchange between free and bound HR1CIP4 on a millisecond timescale. Such exchange results in higher s-values and, in turn, overestimation of the molecular weight following fitting. The peak was therefore estimated to represent a species of 15.8 kDa, which was slightly larger than the expected molecular weight (11.8 kDa). The s-value of the second peak (2.2) was larger than that of Cdc42 alone (2.1) and, as the peaks for Cdc42 and Cdc42-HR1CIP4 complex could not be resolved, this peak most likely encompasses free Cdc42 and the Cdc42-HR1CIP4 complex. Its s-value likely represents a weighted average between the individual s-values of free Cdc42 and the Cdc42-HR1CIP4 complex. The s-value will also have been influenced by exchange between free and bound states. The estimated molecular weight based on the s-value of the second peak (23.7 kDa) lies somewhere in between that of Cdc42 alone (20.7 kDa) and a Cdc42-HR1 complex (32.5 kDa).

Integration of the two peaks revealed that 20-30 % of the protein signal is accounted for by HR1CIP4 and 70-80 % by Cdc42 and complex (i.e. a ratio of ~1:3). At a concentration near to the $K_d$ (~5 µM), 50 % of the protein would be free and 50 % complexed. Thus, the AUC data would be expected to show peaks for HR1CIP4, Cdc42 and complex at 1:1:2 (i.e. 50 % free HR1 plus free Cdc42 and 50 % complex). As the
Cdc42 and complex peaks are indistinguishable, there are instead two peaks at a ratio of 1:3. The data is, therefore, consistent with 5 μM being close to the $K_d$.

With 45 μM of each protein, the two peaks were better resolved. The first peak had an s-value of 1.3, matching that of HR1$^{\text{CIP4}}$ on its own (Figure 3.16D and F), and fitted to a molecular weight of 14.6 kDa. The second peak had an s-value of 2.3 and fitted to a molecular weight of 31.2 kDa. At this higher concentration, this peak is expected to be dominated by the complex, which is reflected in the higher s-value and corresponding higher molecular weight estimate (31.2 kDa). This is close to the expected molecular weight of the complex (32.5 kDa). The first peak (HR1$^{\text{CIP4}}$) accounted for ~5-10 % of the overall signal and the second peak (Cdc42 and complex) accounted for 90-95 %. At 9x the $K_d$, 10 % of the protein would be free and the remainder complexed. Overall, three peaks with 5 % CIP4, 5 % Cdc42 and 90 % complex would be expected were Cdc42 and complex resolved. Therefore, this data is also consistent with a $K_d$ in the range of 5 μM.

The TOCA1 and CIP4 HR1 domains share 68 % sequence identity and 97 % amino acid similarity. Moreover, the affinity of the two HR1 domains for Cdc42 as determined by competition SPA experiments were comparable. The results from these AUC experiments were, therefore, assumed to be transferable to the TOCA1 HR1 domain and its interaction with Cdc42. Importantly, given the similarities between the two domains, it can be reasonably assumed that the TOCA1 HR1 domain is also monomeric and that the HR1-Cdc42 interaction does not involve oligomerisation of either protein or the complex.
As the AUC data was only semi-quantitative, another method to confirm the affinity more quantitatively was desirable. ITC experiments were therefore carried out with a range of concentrations of each protein (up to 0.5 mM) and at temperatures between 10 and 35 °C. No heat changes were seen in any of the experiments, indicating that the enthalpy change associated with the interaction is not detectable (data not shown).

Other higher affinity G protein-HR1 interactions have also failed to show heat changes in ITC experiments (D. Owen, unpublished).

Figure 3.16. Confirmation of the Cdc42-HR1 Affinity by AUC. Analytical ultracentrifugation sedimentation velocity data of A and D) HR1\textsuperscript{CIP4}, B and E) Cdc42 and C and F) equimolar Cdc42 and HR1\textsuperscript{CIP4} at the indicated concentrations. The data was fit with the continuous c(s) distribution model. The component sedimentation coefficient distributions are shown. The $s^\text{20,w}$ values of each peak and corresponding estimated molecular weights are indicated, assuming a uniform frictional ratio of $F_{k,w} = 1.36$. 

3.2.10 Other Binding Assays

As the AUC data was only semi-quantitative, another method to confirm the affinity more quantitatively was desirable. ITC experiments were therefore carried out with a range of concentrations of each protein (up to 0.5 mM) and at temperatures between 10 and 35 °C. No heat changes were seen in any of the experiments, indicating that the enthalpy change associated with the interaction is not detectable (data not shown). Other higher affinity G protein-HR1 interactions have also failed to show heat changes in ITC experiments (D. Owen, unpublished).
Biolayer interferometry (BLI) was also attempted, using GST-Cdc42, C-terminally His6-tagged Cdc42 or biotinylated Cdc42 attached to a sensor and untagged HR1\(^{\text{TOCA1}}\) in solution. His6-HR1\(^{\text{TOCA1}}\) was also tested with untagged Cdc42. In all cases the non-specific binding of the untagged protein to the sensors used (streptavidin, anti-GST and anti-His) could not be overcome in a range of buffers and detergents (data not shown). No data could therefore be obtained for determining the affinity of the interaction or any kinetic parameters.

### 3.3 Autoregulation of TOCA1

A number of potential autoregulatory interactions have been suggested for TOCA1, as discussed in Chapter 1 (section 1.5.3.1). Contrary to these suggestions, the TOCA1-Cdc42 binding assays described above showed that all of the HR1 domain-containing constructs bound to activated Cdc42 with equivalent affinities. The data therefore suggested that the activation of TOCA1 by Cdc42 is not comparable to the situation seen for (N-)WASP. Importantly, full length TOCA1 and the HR1 domain alone were equally competent for Cdc42 binding, unlike with N-WASP where full length protein binds with lower affinity than the GBD alone (Buck et al. 2004). These data suggest that an autoregulatory interaction that affects Cdc42-binding does not exist but an autoregulatory interaction that does not affect HR1-Cdc42 binding remains possible.

The linkers between each domain of TOCA1 are predicted to be unstructured by sequence-based structure prediction servers but the HR1-SH3 linker has been suggested to be protected from proteolysis (J. Peterson, personal communication). An investigation into the proteolytic vulnerability of full-length TOCA1 was therefore of interest, to look for indications of higher order structure.

#### 3.3.1 Purification and Proteolysis of Full Length TOCA1

Full length TOCA1 was cloned into pGEX-6P-1 as described in Chapter 2. The full length protein was then expressed in *E. coli* BL21 and purified as a GST-fusion protein by affinity purification. The GST-tag was cleaved using 3C protease and the protein further purified by size exclusion chromatography (Figure 3.17A-B). Following gel filtration, the protein was contaminated with breakdown products
(Figure 3.17B) and so attempts were made to further purify the full length protein. Anion exchange was used, an example of which is shown in Figure 3.17C. The breakdown was so pronounced that no full length protein was visible following anion exchange (Figure 3.17D).

**Figure 3.17. Purification of Full Length TOCA1.** A) An $A_{280}$ trace showing the purification of full length TOCA1 on a Superdex 16/60 S200 run in 20 mM sodium phosphate pH 7.4, 250 mM NaCl. B) A 10 % Laemmli gel showing a sample of the protein loaded onto the column (L) and samples taken from 2 mL fractions across peaks 1 and 2. C) An $A_{280}$ trace from ion exchange using a 1 mL ResourceQ column (GE Healthcare). The sample was eluted in an NaCl gradient as indicated. D) A 10 % Laemmli gel showing a sample of the loaded protein (L) and samples taken from each 0.5 mL fraction across peaks 1-4. The suggested identity of each band is shown, based on the apparent molecular weight (MW).
Further attempts were made to suppress the proteolysis and so purify the full length protein such that controlled, limited proteolysis could be carried out. These attempts included the addition of EDTA and various specific protease inhibitors. Despite this, the protein continued to break down. The most successful attempt at ‘limited proteolysis’ with Subtilisin is shown in Figure 3.18. The protein was significantly broken down in the absence of any Subtilisin. Attempts to purify the different breakdown products by size exclusion chromatography and to identify them by mass spectrometry also failed due to continued breakdown. There was no evidence for protection of either linker, but the data were far from ideal. It was decided therefore to focus on the individual pairs of domains and look for interactions using biochemical and biophysical techniques.

**Figure 3.18. ‘Limited’ proteolysis of TOCA1.** A 15 % Laemmli gel showing samples taken at two time points, as indicated, having been incubated with the indicated concentration of Subtilisin. The suggested identity of each band based on the apparent molecular weight (MW) is shown.
3.3.2 The HR1 and SH3 Domains

3.3.2.1 Protein Purification

A construct encompassing the HR1 and SH3 domains (residues 330-545) was cloned into pGEX-6P-1 and pMAT10 expression vectors and expression trials were carried out in a range of E. coli strains and conditions (Figure 3.19A-C). The protein did not express well as a GST fusion in any strain or under any of the conditions tested. An example GST-fusion expression is shown in Figure 3.19A. The protein did, however, express well as a His₆MBP fusion in BL21-CodonPlus-RP (Figure 3.19C). The His₆MBP protein was purified by affinity purification on Ni-NTA-agarose (Figure 3.19D) and cleaved from the affinity tag using thrombin protease (Figure 3.19E).

The protein was further purified by size exclusion chromatography (Figure 3.20A and B), but multiple peaks were observed, with HR1SH3 eluting in peaks 2 and 3. None of the fractions were free from significant contamination and so anion exchange was used to purify the HR1SH3 construct further (Figure 3.20C). The protein, separated from its breakdown products (Figure 3.20D), was assessed for further breakdown at room temperature and 4 °C overnight, in the absence or presence of 1 mM EDTA. EDTA was found to reduce breakdown (Figure 3.20E), indicating that a contaminating protease had not been fully removed but could be suppressed with EDTA. 9 mg of protein was obtained, which was kept in 1 mM EDTA, concentrated to 0.7 mM and frozen at -80 °C.
Figure 3.19. Expression and Purification of the TOCA1 HR1SH3 Di-domain. A-C) 10 % Laemmli gels showing samples taken from the indicated bacterial strains transformed with A) pGEX-6P-1-HR1SH3TOCA1 or B and C) pMAT10-HR1SH3TOCA1, grown under the indicated conditions. U − uninduced, I − induced (with indicated concentration of IPTG), S − soluble, P − pellet. The chosen condition is marked with a red asterisk. D) A 12 % Laemmli gel showing samples taken from each wash of the Ni-NTA-agarose beads (Washes) and from a 50 % bead slurry (Beads). E) A 12 % Laemmli gel showing samples taken from the bead washes post-thrombin cleavage and a final bead sample.
Figure 3.20. Further Purification of the HR1SH3 Di-domain. A) An $A_{280}$ trace showing purification of the di-domain by size exclusion chromatography on a Superdex 16/60 S75 in 20 mM Tris-HCl pH 7.9, 250 mM NaCl, 1 mM EDTA. B) A 15 % Laemmli gel showing a 1 in 10 dilution of the protein loaded onto the column (L) and samples taken from each 2 mL fraction across the indicated volume. The fractions boxed in red were pooled and further purified by ion exchange. C) An $A_{280}$ trace from ion exchange using a 1 mL ResourceQ column (GE Healthcare). The sample was eluted in a NaCl gradient as indicated. D) A 12 % Laemmli gel showing samples of the loaded protein (L), the no salt wash (A) and each 1 mL fraction across peaks 2 and 3. E) A 12 % gel showing samples taken after overnight incubation at the indicated temperature, with or without EDTA.
3.3.2.2 Analysis of the HR1SH3 Di-domain by NMR

Simple two-dimensional NMR experiments can be used to study unlabelled proteins to report on their overall fold. For example, 2D Nuclear Overhauser Effect Spectroscopy (NOESY) records interactions (NOEs) between protons that are near in space, typically up to a maximum of 5 Å apart. These NOEs report on the secondary structure of the protein as the distances between particular protons are distinctive within each secondary structure class. In a random coil or β-sheet, the protein backbone is extended and so the amide protons are not close in space and NOEs are not expected in the amide region of the spectrum. Conversely, sequential amides are within short distances in an α-helix and so an abundance of cross peaks in the amide region would indicate an helical fold. Uniquely in β-sheet proteins, the amides of residues in one strand are close to Hαs in adjacent strands and so an abundance of cross peaks in the HN-Hα region would indicate a β-sheet.

2D TOtal Correlation Spectroscopy (TOCSY) experiments are also useful. These rely on through bond (scalar) coupling and contain cross peaks between all of the protons across a residue, for example between the amide proton and the Hα. The resolution of the peaks in the N-Hα region of the TOCSY, also known as the ‘fingerprint’ region, can give an indication of protein size. Therefore, the resolution in this region could indicate whether the HR1SH3 di-domain behaves as a ~30 kDa globular protein or whether the HR1 and SH3 domains behave as two separate ~15 kDa domains.

2D NOESY and TOCSY experiments were recorded on the HR1SH3 di-domain. There was an abundance of HN-HN cross-peaks in NOESY spectrum (outlined in blue in Figure 3.21), which indicated the α-helical structure expected for the HR1 domain. There was also an abundance of sequential HN+1-Hα NOEs observed in the NOESY (outlined in red), which is indicative of the β-sheet structure expected for the SH3 domain. Both spectra showed sharp peaks that were consistent with two small (~15 kDa) domains behaving independently of one another and the resolution in the fingerprint region of the TOCSY (outlined in red) was indicative of two independent domains rather than a 30 kDa globular protein.
Analysis of the NMR sample following these experiments revealed extensive breakdown (Figure 3.22). The expected positions for the individual HR1 and SH3 domains in SDS-PAGE are indicated. It appears that the HR1SH3 di-domain was cleaved into 3 major bands that most likely represent the two individual domains with some of the linker residues still attached, as each band is running a few kDa larger than the individual domains (see Figures 3.3 and 3.25 for gels of HR1 and SH3 alone). Despite the breakdown (HR1SH3, HR1- and -SH3 all present), there was a single set of peaks in the NOESY and TOCSY spectra in Figure 3.21, which is again consistent with two separate domains behaving independently from one another.

Figure 3.21. Two-dimensional NOESY and TOCSY of the HR1SH3 Di-domain. The spectra were recorded on 0.7 mM HR1SH3 in 20 mM Tris:HCl pH 7.0, 250 mM NaCl, 1 mM EDTA and 10 % D2O on a Bruker DRX500 at 25 °C. A and B show the left-hand-side of a 2D NOESY (A), in which through-space transfer of magnetization between nearby protons leads to cross-peaks that are visible off the diagonal line, and a 2D TOCSY (B), in which through-bond transfer of magnetisation results in the cross-peaks.
3.3.2.3 Purification of the HR1SH3 Di-domain from pMAT10-P

The contaminating protease could be thrombin, left over from the cleavage of the His<sub>6</sub>MBP fusion tag. Thrombin has an extended cleavage specificity and can cleave between Glycine and Arginine residues at many sites that do not strictly match its consensus recognition sequence. It can also cleave non-specifically between other amino acids, although cleavage after a positive residue is preferred (Gallwitz et al. 2012). The HR1-SH3 linker contains four Arginines, one of which is preceded by a Glycine, so is vulnerable to cleavage. The HR1SH3 di-domain was therefore cloned into pMAT10-P, which has a 3C protease recognition site in place of the thrombin site. 3C protease shows a higher degree of specificity and so if the contaminating protease was thrombin, the problem of continued proteolysis would be abolished.

The protein was purified by Ni<sup>2+</sup>-affinity purification as previously described and the His<sub>6</sub>MBP tag removed with 3C protease. Size exclusion chromatography, shown in Figure 3.22, yielded sufficiently pure protein, free from significant breakdown. The higher molecular weight band, which was present in a concentration-dependent manner, was consistent with HR1SH3 dimer. The SH3 domain contains a Cysteine residue, which may mediate dimerisation if the sample was not sufficiently reduced. The size exclusion chromatography, performed in buffer containing fresh reducing agent, is consistent with a single monomeric species.
3.3.2.4 Binding of the HR1SH3 Di-domain to Cdc42

The HR1SH3 di-domain competed with GST-ACK GBD in competition SPAs, with a similar affinity to the HR1 domain alone ($K_d = 4.6 \pm 4 \text{ µM}$, Figure 3.24). This implies no inhibition of the HR1-Cdc42 interaction by the SH3 domain. It does not, however, rule out an HR1SH3 interaction that is independent of Cdc42 binding.

Figure 3.23. Purification of the HR1SH3 Di-domain from pMAT10-P. A) An $A_{280}$ trace showing the purification of the HR1SH3 di-domain by size exclusion chromatography on a Superdex 16/60 S75 in MT-PBS with 5 mM EDTA. B) A 12 % Laemmli gel showing a sample of the protein loaded onto the column (L) and samples taken from each 2 mL fraction across the void and peak 1.
Purification of the SH3 Domain

NMR spectroscopy can be used to detect and map protein-protein interactions by observing one labelled protein in the presence of an unlabelled protein. For example, $^{15}$N-SH3$_{TOCA1}$ could be monitored in the presence of unlabelled HR1$_{TOCA1}$ to determine whether the two domains interact. If they interact, some of the peaks visible in the NMR spectrum of the SH3 domain would be shifted due to the new chemical environment imposed by the HR1 domain. The SH3 domain was therefore cloned and the $^{15}$N-labelled protein purified for use in NMR experiments.

The domain boundaries of the TOCA1 SH3 domain were defined based on a sequence alignment with CIP4 SH3 domain (Figure 3.25) as the structure of the CIP4 SH3 domain has previously been solved (Myamoto et al. unpublished. PDB code: 2CT4). The $^{15}$N-labelled protein was expressed in *E. coli* BL21 grown in minimal media (MOPS) and purified in a similar way to HR1$_{TOCA1}$ (Figure 3.26). 9 mg of protein was obtained, concentrated to 0.4 mM and frozen at -80 °C in 500 µL aliquots.

Figure 3.24. A Competition SPA with the HR1SH3 Di-domain and Cdc42.

A competition SPA experiment was carried out with the indicated concentrations of HR1SH3$_{TOCA1}$ titrated into 30 nM GST-ACK and 30 nM Cdc42Δ7Q61L·[$^3$H]GTP. The $K_d$ value, derived from fitting to a competition binding isotherm, is shown along with the curve-fitting error.
An NMR titration with the HR1 and SH3 Domains

The $^{15}$N-HSQC is one of the most commonly used NMR experiments when studying proteins. It has been described briefly in the Introduction (section 1.7.1.3) and in detail elsewhere (Cavanagh et al. 1998). In theory, the resulting spectrum comprises a set of peaks representing every NH group in the protein. The $^{15}$N-HSQC of the TOCA1 SH3 domain (Figure 3.27) was very well resolved and contained the expected number of
peaks based on the amino acid sequence. Unlabelled HR1 domain was next titrated into $^{15}$N-SH3 and a series of $^{15}$N-HSQCs recorded. The final spectrum of the 1:1 sample is shown overlaid with the spectrum of free SH3 domain in Figure 3.27. No changes were seen in the HSQC, indicating that there is no interaction between the SH3 and HR1 domains.

![Figure 3.27](image)

**Figure 3.27.** $^{15}$N-HSQCs of the SH3 Domain in the Presence of the HR1 Domain. The spectra were recorded on 0.36 mM $^{15}$N-SH3 in 20 mM sodium phosphate, 150 mM NaCl, 5 mM DTT, 5 mM MgCl$_2$ and 10 % D$_2$O, free (black) or in the presence of equimolar unlabelled HR1 domain (orange).

### 3.3.3 The F-BAR and SH3 Domains

#### 3.3.3.1 Protein Purification

Small scale expression trials of the TOCA1 F-BAR domain were carried out in a variety of conditions. The protein expressed well as a GST-fusion but was insoluble in all conditions (Figure 3.28A and B). The protein also expressed well as a His$_6$MBP-fusion (Figure 3.28C-E) and the protein was soluble when expressed at 20 °C in BL21-CodonPlus-RIL (Figure 3.28E).
The protein was expressed in 3 L of *E. coli* RIL and affinity purified on Ni-NTA-agarose (Figure 3.29A). Cleavage of the His<sub>6</sub>MBP fusion tag was unsuccessful on the beads (Figure 3.29B) and so the protein was eluted using 500 mM Imidazole (Figure 3.29C). Half of the protein was further purified by size exclusion chromatography (Figure 3.29D) and 25 mg of protein was obtained. The other half was cleaved in solution with thrombin protease but no protein was obtained as the F-BAR domain precipitated. It was decided to proceed with His<sub>6</sub>MBP-F-BAR.
Figure 3.29. Purification of the TOCA1 F-BAR domain. A) A 12 % Laemmli gel showing samples taken from the supernatant (SN) and each wash of the glutathione agarose beads (washes) and from a 50 % bead slurry (beads). B) A 12 % gel showing samples taken from the post-3C protease cleavage washes and a final bead sample. C) A 12 % gel showing samples taken from further washes and elutions with the indicated concentration of imidazole. D) An A$_{280}$ chromatogram from the purification on a Superdex 16/60 S200 run in 20 mM Tris:HCl pH 7.9, 250 mM NaCl, 1 mM EDTA. E) A 12 % gel showing a 1 in 10 dilution of the loaded sample (L) and samples taken from 2 mL fractions across peaks 1-4.
3.3.3.2 **ITC with the F-BAR and SH3 Domains**

ITC experiments were carried out with the F-BAR and SH3 domains at a range of concentrations but showed no heat changes above those seen for the F-BAR into buffer control (data not shown). The ITC experiments were repeated at 15 °C, 25 °C and 35 °C, with a range of concentrations up to 260 µM F-BAR domain and 20 µM SH3 domain. The F-BAR domain could not be used at higher concentrations due to aggregation.

3.3.3.3 **Liposome Sedimentation Assays**

Data from ITC experiments were inconclusive as a lack of heat changes does not prove the absence of an interaction. As the F-BAR-SH3 interaction would not be expected to disrupt F-BAR-liposome binding, liposome sedimentation assays were employed with the F-BAR and SH3 domains to investigate whether the F-BAR domain pulls down the SH3 domain (Figure 3.30). The F-BAR domain was present in the supernatant and pellet samples. A small amount pelleted in the absence of liposomes but ~50 % pelleted when liposomes were present. This suggested that the F-BAR domain bound to the liposomes with a $K_d$ close to the concentration of F-BAR domain used (2 µM). The amount of F-BAR domain in the pellet was unchanged in the presence of the SH3 domain, suggesting that, as expected, the SH3 domain did not inhibit localisation of the F-BAR domain to the liposomes. Furthermore, there was no more SH3 domain visible in the pellet than in the ‘no liposome’ or ‘no F-BAR’ negative controls, indicating that it was not pulled down by the F-BAR domain.

Alternative pull downs were also performed with high concentrations of GST-SH3 immobilised on glutathione agarose beads incubated with His$_6$MBP-F-BAR and with His$_6$MBP-F-BAR immobilised on Ni-NTA-agarose beads incubated with the SH3 domain. No binding was seen in a range of buffer conditions (data not shown).
For completion, the liposome sedimentation assay was repeated with the HR1 domain (Figure 3.31). The results were comparable to those seen with the SH3 domain. Importantly, the HR1 domain did not affect F-BAR-membrane binding and was not pulled down with the F-BAR domain. Alternative pull downs with either the F-BAR or the HR1 domain immobilised on Ni-NTA-agarose beads were also carried out, along with extensive ITC experiments, but again no binding was observed.

Figure 3.30. Liposome Sedimentation Assays with the F-BAR and SH3 Domains. A 4-12 % gradient gel (NuPAGE, Thermo Fisher) showing samples taken from the supernatant (SN) and pellet (P) fractions following liposome sedimentation. The F-BAR protein was incubated for 10 minutes with liposomes, and with the SH3 domain at the indicated concentrations, prior to centrifugation.

3.3.4 The F-BAR and HR1 Domains

For completion, the liposome sedimentation assay was repeated with the HR1 domain (Figure 3.31). The results were comparable to those seen with the SH3 domain. Importantly, the HR1 domain did not affect F-BAR-membrane binding and was not pulled down with the F-BAR domain. Alternative pull downs with either the F-BAR or the HR1 domain immobilised on Ni-NTA-agarose beads were also carried out, along with extensive ITC experiments, but again no binding was observed.
Figure 3.31. Liposome Sedimentation Assays with the F-BAR and HR1 Domains. A 4-12 % gradient gel (NuPAGE, Thermo Fisher) showing samples taken from the supernatant (SN) and pellet (P) fractions following liposome sedimentation. The F-BAR protein was incubated for 10 minutes with liposomes, and with the HR1 domain at the indicated concentrations, prior to centrifugation.
3.4 Conclusions and Perspective

3.4.1 TOCA1-Cdc42 Binding: A Low Affinity Interaction

The data presented here indicate that the TOCA1 HR1 domain is sufficient for Cdc42 binding but that it binds Cdc42 with a modest affinity (~5 µM). The F-BAR domain alone does not bind Cdc42, whilst the HR1, ΔSH3, HR1SH3 and full length proteins bind with comparable affinities, suggesting that no region outside of the HR1 domain contributes to binding. The C-terminus of Cdc42 was not required for maximal binding as full length Cdc42 bound HR1\textsuperscript{TOCA1} with comparable affinity to Cdc42Δ7. This is in contrast to the Rac1-PRK1 interaction (Owen et al. 2003) but similar to most other G protein-effector interactions (Mott and Owen 2015).

The HR1 domains of CIP4 and FBP17 were also found to bind Cdc42 with moderate, micromolar affinities (CIP4: $K_d = 5.1-5.4$ µM, FBP17: $K_d = 16.8 \pm 6.3$ µM), corroborated by AUC data. Thus, the low affinity Cdc42 interaction appears to be common amongst the TOCA family but is unlike other HR1 domains. For example, the HR1\textsubscript{a} domain of PRK1 binds to RhoA with a $K_d$ of 60 nM and the HR1\textsubscript{b} domain of PRK1 binds Rac1 with a $K_d$ of 68 nM (Owen et al. 2003).

The low affinity of the HR1\textsuperscript{TOCA1}-Cdc42 interaction in the context of the physiological concentration of TOCA1 in Xenopus extracts (~5-10 nM) (Ho et al. 2004) suggests that binding between TOCA1 and Cdc42 is likely to occur in vivo only when TOCA1 is at high local concentrations. This implies that TOCA1 must be membrane localized, and therefore in close proximity to activated Cdc42, prior to Cdc42 binding. Once at the membrane, high local concentrations of TOCA1 could exceed the $K_d$ of the Cdc42-HR1\textsuperscript{TOCA1} interaction. Cdc42-HR1\textsuperscript{TOCA1} binding would then be favourable, as long as coincident activation of Cdc42 had occurred.

The low affinity of the Cdc42-HR1\textsuperscript{TOCA1} interaction is consistent with a pathway with tight spatial and temporal regulation. Such tightly regulated pathways involve combinatorial signals leading to a series of coincident, weak interactions that together elicit full activation of downstream effects (reviewed in Schmid and McMahon 2007; Perkins et al. 2010; Acuner Ozbabacan et al. 2011). As such, weak, transient protein-protein interactions are functionally significant in several systems. For example, the
binding of adaptor proteins to protein cargo during the formation of clathrin-coated vesicles in endocytosis involves multiple interactions of micromolar affinity (Praefcke et al. 2004; Höning et al. 2005).

Consistent with a pathway of multiple low affinity interactions, the liposome spin assays showed that the F-BAR domain was ~50 % membrane bound when present at 2 µM (section 3.3.3.3) and so the $K_d$ is likely to be ~2 µM. This modest affinity of the TOCA1-membrane interaction along with the modest affinity of TOCA1 dimerisation (which is likely to be comparable to that of the FCHo2 F-BAR domain (2.5 µM) (Henne et al. 2007)) and of the TOCA1-Cdc42 interaction (~5 µM) implies the need for coincident interactions in order to achieve robust TOCA1 recruitment and clustering at the membrane. A critical threshold level of clustered TOCA1 may be necessary prior to robust N-WASP recruitment and activation.

Similar regulatory thresholds have been demonstrated previously in relation to N-WASP activation. For example, in vitro activation of N-WASP by PI(4,5)P$_2$ occurs above a sharp threshold, which arises from the cooperativity of the multivalent interaction of the basic region of N-WASP with PI(4,5)P$_2$ and the autoinhibitory VCA domain (Papayannopoulos et al. 2005). It is possible that the TOCA1-N-WASP interaction will also contribute to a sharp threshold of N-WASP activation in vivo.

### 3.4.2 Regulation of TOCA1: No Evidence for Intramolecular Interactions

The affinity of full length TOCA1 for Cdc42 is similar to that of the HR1 domain alone. This is in contrast to (N-)WASP, for which the affinity of full length is 100-fold lower (3.2 µM) than of the GBD alone (~30 nM) (Buck et al. 2004). The mechanism of (N-)WASP autoinhibition and activation described by (Leung and Rosen 2005) relies upon an equilibrium that is driven forward by the significantly higher affinity of the unfolded GBD compared to the folded full length protein. Such a mechanism evidently does not exist for TOCA1 and it does not seem that the HR1-Cdc42 interaction could be responsible for driving an equilibrium between inactive and active TOCA1. This is consistent with the hypothesis described above in which the role of the HR1-Cdc42 interaction is to contribute another low affinity interaction that leads to clustering of TOCA1 at sites rich in activated Cdc42.
The comparable affinities of full length TOCA1 and the HR1 domain alone do not preclude other autoregulatory interactions that are independent of Cdc42 binding. However, no binding was observed between any of the TOCA1 domains in a variety of biochemical and biophysical experiments. It therefore appears unlikely that TOCA1 is autoregulated simply by inter-domain interactions. Instead its regulation may be influenced by other proteins or post-translational modifications that have not been addressed here or indeed via the combinatorial effects of multiple, low affinity interactions.

It is important to note the limitations of the methods used to investigate the proposed inter-domain interactions. Pull downs are prone to false negatives caused by immobilization of one of the proteins, buffer composition and steric effects. Moreover, a lack of heat changes in ITC does not prove that there is no interaction, as an interaction may be predominantly entropically driven and have heat changes too small to detect within the normal concentration ranges and conditions used for ITC. The data is also limited in that the domains were tested in isolation and so do not preclude a very low affinity interaction that occurs only between the naturally tethered domains. The 2D NMR experiments carried out with the HR1SH3 di-domain, however, strongly suggest that these two domains behave independently of one another. Altogether, the multitude of negative binding results in a range of buffers and conditions strongly suggest that no intramolecular interactions exist for TOCA1. The SPA data with full length TOCA1 and various truncations is also consistent with this.

### 3.4.3 Where next?

The results presented in this chapter have addressed the first main aim of the project: to measure the affinity of Cdc42-TOCA1 binding, determine the minimal region of TOCA1 required for maximal affinity binding and to investigate possible mechanisms of TOCA1 autoregulation. It is now of particular interest to investigate the underlying molecular explanation for the low affinity of this G protein-HR1 interaction.

The interaction may be markedly different from previously studied G protein-HR1 interactions or the structure of the HR1 domain itself could be atypical. After all, the TOCA family HR1 domains share 60-70 % sequence identity with one another but are
more divergent from the previously studied PRK1 HR1 domains (~20 %). The HR1 domains of PRK1 are also relatively divergent from each other and their relatively low sequence identities (~25 %) are reflected in subtle structural differences. For example, they differ in the relative lengths of the two helices and in the secondary structure of the region N-terminal to the coiled-coil (Maesaki et al. 1999; Owen et al. 2003).

Structural studies of the TOCA1 HR1 domain and its interaction with Cdc42 using NMR spectroscopy are, therefore, of interest. The results in this chapter have shown that it is possible to produce and purify HR1$^{\text{TOCA1}}$ and that the protein is stable at concentrations as high as 0.9 mM. Furthermore, the purified protein is free from proteinous contaminants and is not susceptible to proteolysis over the course of 15 days (Figure 3.3). It is therefore amenable to the long experiments needed for structural studies by NMR. The AUC data pertaining to the CIP4 HR1 domain (68 % identical to HR1$^{\text{TOCA1}}$) is indicative of a monodisperse sample comprising only monomer, which is also encouraging for NMR studies.

The following chapters describe the NMR structure of the TOCA1 HR1 domain along with a structural analysis of the HR1-Cdc42 binding interface.
4 Structural studies of HR1\textsuperscript{TOCA1}

4.1 Introduction

The results in the preceding chapter revealed that the affinity of the Cdc42-HR1\textsuperscript{TOCA1} interaction was 100x weaker than previously studied small GTPase-HR1 interactions. This raised the question as to whether the Cdc42-HR1\textsuperscript{TOCA1} interaction is remarkably different or indeed whether the structure of the TOCA1 HR1 domain itself is atypical. Structural studies of the TOCA1 HR1 domain and its interaction with Cdc42 were therefore of interest. Based on the previous chapter, HR1\textsuperscript{TOCA1} (330-426) is sufficient for maximal Cdc42-binding and so this construct was used in the structural studies.

4.2 The NMR Structure of the TOCA1 HR1 Domain

4.2.1 Preliminary NMR Experiments

4.2.1.1 Two-dimensional NOESY and TOCSY of HR1\textsuperscript{TOCA1}

The 2D NOESY and TOCSY can be used to report on the secondary structure of a protein as described in section 3.3.2.2. The resolution of the peaks in the H\textsuperscript{N}-H\textsuperscript{\alpha} region of the TOCSY, also known as the ‘fingerprint’ region, can also provide an indication of whether the protein can be studied by NMR. If the peaks in this region are very broad it could indicate oligomerisation or aggregation of the protein, that the protein is too large for classical NMR studies and that further NMR experiments will not provide the necessary resolution for structural studies. Good resolution in this region indicates that the much higher resolution double and triple resonance (\textsuperscript{1}H/\textsuperscript{15}N or \textsuperscript{1}H/\textsuperscript{15}N/\textsuperscript{13}C) experiments will be of ample resolution for backbone and sidechain assignments and for structural studies.

Therefore, 2D NOESY (Figure 4.1A) and TOCSY (Figure 4.1B) experiments were recorded on HR1\textsuperscript{TOCA1}. In the NOESY, there were many cross-peaks in the H\textsuperscript{N}-H\textsuperscript{N} region, which is outlined in red in Figure 4.1A, arising from through-space NOE transfer of magnetisation between adjacent NH groups. The abundance of cross-peaks in this region indicated the expected \alpha-helical structure. The cross-peaks in the ‘fingerprint’ region of the TOCSY, outlined in blue in Figure 4.1B, were well
resolved. Together, these spectra indicate that the HR1 domain is α-helical and that it is amenable to study by NMR.

**Figure 4.1. 2D NOESY and TOCSY of HR1\textsuperscript{TOCA1}.** NMR spectra of 0.9 mM HR1\textsuperscript{TOCA1} in 50 mM sodium phosphate pH 5.5 with 25 mM NaCl and 10 % D\textsubscript{2}O were recorded on a Bruker DRX500 at 25 °C. The figure shows the left-hand-side of a 2D NOESY (A) and a 2D TOCSY (B). The amide region of the NOESY is outlined in red and the fingerprint region of the TOCSY is outlined in blue.
4.2.1.2 Purification of $^{15}$N-labelled HR1$^{TOCAl}$

Structural studies of proteins involve a number of 2D and 3D NMR experiments that provide information on the backbone (NH, CH and CO) and sidechain (CH and NH) groups. The spin active forms of Nitrogen and Carbon, $^{15}$N and $^{13}$C, have extremely low natural abundance and so these nuclei must first be incorporated into the protein in place of $^{14}$N and $^{12}$C. This can be readily achieved in bacterial expressions by simply growing the bacteria in media containing $^{15}$N-ammonium chloride and $^{13}$C-glucose.

To begin, small scale expression trials were carried out with E. coli BL21 grown in minimal media. Figure 4.2 shows that the expression was higher when cells were grown in Celtone-enriched MOPS than in M9 (compare the lanes marked with white asterisks). $^{15}$N-HR1 was therefore expressed in E. coli grown in MOPS enriched with $^{15}$N-Celtone and containing $^{15}$N-NH$_4$Cl and purified as previously described for unlabelled protein. 22 mg of $^{15}$N-labelled protein was obtained from 3 L of E. coli following size exclusion chromatography and frozen at -80 °C in 0.5 mL aliquots at 0.9 mM.
Figure 4.2. Purification of $^{15}$N-HR1$^{\text{TOCA1}}$ for NMR studies. A) A 12 % Laemmli gel showing samples taken from *E. coli* BL21 transformed with pGEX-6P-1-HR1$^{\text{TOCA1}}$, grown in 2TY, M9 or MOPS as indicated. U – uninduced, I – induced with 0.1 mM IPTG, S – soluble, P – pellet. Comparing the bands marked with white asterisks showed that expression in MOPS was greater than in M9. B) A 15 % Laemmli gel showing the affinity purification of GST-HR1$^{\text{TOCA1}}$. Gel samples were taken from the supernatant (SN), each wash of the glutathione agarose beads and a 50 % bead slurry. C) An 18 % Laemmli gel showing gel samples taken from the supernatant (SN) and bead washes following overnight cleavage with 3C protease and a final bead sample. D) An $A_{280}$ trace from size exclusion chromatography carried out on a Superdex 16/60 S30 in 20 mM sodium phosphate pH 7.4, 150 mM NaCl. E) An 18 % Laemmli gel showing a 1 in 10 dilution of the sample that was loaded onto the size exclusion column (L) and samples taken from each 2 mL fraction corresponding to the void and peak 1.
4.2.1.3 HSQC Experiments

The $^{15}$N-HSQC is one of the most commonly used NMR experiments when studying proteins. It has been briefly described in the Introduction (section 1.7.1.3) and a more detailed textbook description of the experiment can be found in (Cavanagh et al. 1998; Keeler 2005). The resulting spectrum comprises a set of peaks representing every NH group in the protein. This includes the backbone NH group of every residue (absent in Prolines), the sidechain NH of Tryptophan, Histidine and Arginine, the sidechain NH$_2$ of Arginine, Glutamine and Asparagine and the sidechain NH$_3^+$ of Lysine. For Arginine, however, the NH and NH$_2$ groups are often very weak due to exchange broadening and so may not be observed. The NH of Histidine and the NH$_3^+$ of Lysine are also not observed. The NH$_2$ groups of Glutamine and Asparagine, however, give stronger peaks in the top right hand region of the HSQC ($\delta^N = 110-115$ ppm and $\delta^H = 6.5-8.0$, labelled in Figure 4.3). Two peaks are observed with equivalent Nitrogen shifts, representing each proton attached to a single Nitrogen.

The HSQC of HR$^1$TOCA$^1$ was recorded, processed in Azara (W. Boucher, Department of Biochemistry) and analysed in CCPN Analysis (Vranken et al. 2005). The spectrum was well dispersed (Figure 4.3) and contained the expected number of peaks. The construct comprised residues 330-426 of TOCA1 along with 7 additional residues, 3 encoded by the vector and resulting from proteolysis using 3C protease (GPL), and 4 that were encoded in the primers during cloning (GSHM). The $^{15}$N-HSQC of the TOCA1 HR1 domain was therefore expected to contain a total of 119 peaks: 96 backbone NH peaks (104 residues minus the N-terminus and 7 prolines), 11 NH$_2$ pairs (7 Glutamine and 4 Asparagine) and 1 Tryptophan NH peak. The N-terminus was not visible due to exchange broadening.
As the $^{15}$N-HSQC was well dispersed and so most likely fully assignable, $^{15}$N/$^{13}$C-labelled HR1 domain was expressed and purified for use in the $^{13}$C- and triple resonance experiments required for backbone and sidechain assignment. The protein was expressed as described in Chapter 2 and purified using the same methods shown for the singly labelled protein shown above. 6 mg of protein was obtained from 3 L of E. coli and a $^{13}$C-HSQC spectrum was recorded.

The $^{13}$C-HSQC is expected to contain a peak for every CH group in the protein, including the backbone and sidechains. The spectrum is therefore far more crowded than the $^{15}$N-HSQC and overlap is a much greater problem even for small proteins. The $^{13}$C-HSQC of HR1$^{TOCA1}$ was, however, relatively well resolved and much of it was expected to be assignable (Figure 4.4). The key regions are annotated.
Figure 4.4. The $^{13}$C-HSQC of HR1$^{\text{TOCA1}}$. A $^{13}$C-HSQC of 0.9 mM HR1$^{\text{TOCA1}}$ in 20 mM sodium phosphate pH 7.4 with 150 mM NaCl and 10 % D$_2$O recorded on a Bruker DRX500 at 25 °C.
4.2.2 Resonance Assignment

Structure determination of proteins by NMR relies upon several sets of restraints, including dihedral angle restraints derived from chemical shifts and distance restraints derived from through space NOE experiments. Data must therefore be recorded on the backbone and sidechain atoms and, before the necessary restraints can be derived from the NMR data sets, the backbone and sidechain resonances must be assigned as completely as possible. Such an assignment relies upon a series of through bond experiments to be analysed with reference to the HSQC s shown above. The following sections describe the experiments and the procedure used for the backbone and sidechain assignment of HR1^TOCA1.

4.2.2.1 The Backbone Assignment

The backbone assignment was performed using five triple resonance experiments: HNCO, HNCA, HN(CO)CA, HNCACB and HN(CO)CACB. These experiments depend upon the transfer of magnetisation via scalar couplings between the amide proton, amide nitrogen and carbon atoms. For example, the transfer in the HNCA utilises \( ^1J_{NH} \) and \( ^1J_{NC\alpha} \). There may also be transfer across the carbonyl to the Ca of residue i-1 via \( ^2J_{NC\alpha} \). The magnetisation transfer pathways for each experiment are illustrated in Figure 4.5.
Figure 4.5. Magnetisation Transfer Pathways in Triple Resonance NMR Experiments. The diagrams illustrate the transfer of magnetisation via one or two bond J couplings, indicated by double headed arrows. In each case the magnetisation begins and ends on the amide proton. The following spectra are illustrated: A) the HNCA, B) the HNCACB, C) the HN(CO)CA, D) the HN(CO)CACB and E) the HNCO. The chemical shift is allowed to evolve on all of the nuclei highlighted in red and, as such, the chemical shift of each is encoded. The nuclei highlighted in green are utilised for magnetisation transfer in order to restrict the transfer pathway to the i-1 residue but their shifts are refocused and so not encoded.
The HNCA thus contains a cross-peak resulting from the one-bond scalar coupling ($^{1}J_{NC\alpha}$) between the NH and Cα and as such encodes the chemical shift of the NH and Cα for each residue. The HNCACB contains two cross-peaks for each NH group, one to the Cα and one to the Cβ. These experiments are therefore used to assign the Cα and Cβ corresponding to each NH group. In some cases, the HNCA and HNCACB contain additional weaker peaks due to correlation of the NH to the Cα or Cβ of the preceding residue via the $^{2}J_{NC\alpha}$ coupling. The HN(CO)CA and HN(CO)CACB experiments contain only cross-peaks arising from the coupling between the NH of each residue (i) to the Cα and Cβ of the preceding residue (i-1), as the magnetisation is transferred selectively via the carbonyl group. These experiments are therefore used to link sequential residues in the first step of resonance assignment, the sequential backbone assignment, as described below and illustrated in Figure 4.6.

First, the NH strip for a residue (i) can be located in the HNCA and HN(CO)CA (which are overlaid) by navigation from the HSQC (illustrated by arrow 1 in Figure 4.6A, shown inside the blue box). The Cα of the preceding residue (i-1) is visible in the HN(CO)CA so can be used to locate the Cα in the HNCA of i-1 by matching the chemical shift (illustrated with arrow 2 in Figure 4.6A). The corresponding peak is then located in the HSQC and linked sequentially to the starting HSQC peak (illustrated by arrow 3). The i+1 strip can also be located using the Cα in the HNCA of residue i to find the corresponding Cα peak in the HN(CO)CA of residue i+1 (illustrated with arrow 4). In this way it is possible to step along the protein backbone.

The overall chemical shift range for the 104 discernable Cα peaks for HR1$^{\text{TOCA1}}$ was 43-68 ppm, with 14 % of them falling in the lower half of the range and 86 % in the upper half (between 55.5 and 68 ppm). This equates to 89 peaks in a 12.5 ppm window. The Cα chemical shift degeneracy was such that the HNCA/HN(CO)CA pair were not sufficient to complete the sequential assignment. The HNCACB and HN(CO)CACB spectra were used to resolve the ambiguity as illustrated in Figure 4.6B (inside the orange box). The Cβ resonances covered a broader range of chemical shifts (17-69 ppm), with 86 % of the peaks spread over 26 ppm. Once sequential links had been made between peaks in the HNCA and HNCACB, sequential links were
easily made in the HNCO as it contains a single cross-peak for each NH group to the carbonyl of residue i-1, as illustrated in Figure 4.6C.

**Figure 4.6. Backbone Assignments.** The indicated experiments were recorded on 0.9 mM HR1\textsuperscript{TOCA1} in 20 mM sodium phosphate pH 7.4 with 150 mM NaCl and 10 % D\textsubscript{2}O on a Bruker DRX500 at 25 °C. Section A, outlined in blue, shows the $^{15}$N-HSQC alongside the Cα (57-70 ppm) region of the HNCA and HN(CO)CA for each of four NH strips corresponding to four sequential residues. This section illustrates the first steps of sequential backbone assignment: 1) navigation from the HSQC to the triple resonance experiments, 2) location of the Cα of i-1 in the HNCA from the HN(CO)CA of residue i, 3) navigation from the triple resonance strip of residue i-1 to the HSQC and 4) location of the i+1 NH strip, and so on. Section B, outlined in orange, shows the Cβ (27-34 ppm) region of the HNCA and HN(CO)CA for the same four residues and illustrates how these experiments were used to reduce ambiguity in the assignment. Section C shows the CO (176-180 ppm) region of the HNCO for the same four residues.
Sequence specific assignments were made concurrently with the sequential linkages by comparison of the Cα and Cβ chemical shifts with reference values. More than 7 million chemical shift values from NMR assignments are contained within an online database, the BioMagResBank (Markley et al. 2008). These values have been used to calculate the minimum, maximum and average value for the chemical shift of each ^1^H, ^15^N and ^13^C atom in specific amino acid types and this information has been tabulated. For example, the minimum chemical shift of the Cα of an Aspartate is recorded as 41.11, the maximum is 67.17 and the average (mean) is 54.69. This information has also been further stratified by secondary structure (random coil, α-helices and β-strands). For example, when in a β-strand the Cα of an Aspartate falls within 42.5-59.3 ppm and the mean (52.9), median (53.0) and mode (52.4) are relatively low. When in an α-helix, however, the range is shifted (46-66.8) and the mean, median and mode are higher (57.1).

The Cα and Cβ chemical shifts for each residue were compared with the reference values for different residue types and so the possible identities of each residue were deduced. The different amino acid types have overlapping chemical shift profiles so a single Cα/Cβ pair could not be assigned in isolation but, by matching the possible identities for stretches of sequentially linked residues with the amino acid sequence of HR1^TOCA1, some sequence specific assignments could be made.

To overcome the limitations caused by Cα and Cβ chemical shift degeneracy between amino acid types, two further experiments were used in addition to the triple resonance experiments: the ^15^N-NOESY-HSQC and ^15^N-TOCSY-HSQC. The NOESY relies on through space NOE transfer, whilst the TOCSY transfer relies on through bond scalar couplings. In both cases, the protons are first excited and their chemical shifts recorded. For the NOESY, the magnetisation is then put back on the z-axis prior to mixing. For the TOCSY, the magnetisation is transverse during the mixing step. It is during this mixing step that the magnetisation is transferred across all of the protons via either through space NOE transfer or through bond J couplings. Both experiments end with a ^15^N-HSQC and so the magnetisation is transferred back to amide protons via nitrogen before detection. As a result, only correlations between
the NH groups and other protons are recorded. The magnetisation transfer pathways for these two experiments are shown in Figure 4.7.

![Figure 4.7](image)

**Figure 4.7. Magnetisation Transfer Pathways in the $^{15}$N-NOESY-HSQC and $^{15}$N-TOCSY-HSQC.** A) A diagram of the magnetisation transfer in the $^{15}$N-NOESY-HSQC. Through space NOE transfer is illustrated with arrows. Transfer of magnetisation via the Nitrogen before detection results in selection of only the NOEs to the NH. The nuclei for which the chemical shift is recorded are shown in red circles. The spectrum also includes NOEs from other residues that are close in space that are not shown here. B) A diagram of through bond magnetisation transfer in the $^{15}$N-TOCSY-HSQC. The $^{3}J_{HN-H\alpha}$ coupling is shown with red lines and the nuclei for which the chemical shift is recorded are shown in red circles. Cross-peaks to the H$\beta$ and H$\gamma$ protons, shown in pink circles, can also be seen but may be too weak for detection in $\alpha$-helical proteins as $^{3}J_{HN-H\alpha}$ is small (4-6 Hz).

The NOESY mixing time is selected so that only direct sidechain proton to NH group NOEs are recorded. This is typically 100 msec for small proteins. If the mixing time is too short, only the NOEs between the closest protons will be detected and longer distance information will be lost. If it is too long, the NOEs will contain additional magnetisation transferred from other nearby protons and so will not provide an accurate measure of distance between sidechain protons and the amide groups. This is called spin diffusion and is addressed later (section 4.2.3.2). The TOCSY mixing time must be long enough for the magnetisation to reach the more distant protons but if it is too long the signals will be lost due to relaxation.

In the TOCSY, all of the cross peaks are intra-residue as the magnetisation is transferred through scalar couplings and is not transferred via the N-CO or C$\alpha$-CO bonds and the protons that are connected via the fewest bonds to the NH group will
give the strongest cross peaks. In theory, the TOCSY provides the proton shifts across the complete sidechain, providing a useful fingerprint for amino acid type. However, for most residues of HR1\textsuperscript{TOCA1}, the TOCSY transfer was only sufficient to identify the H\textalpha. This is because the $^{3}J_{HN-H\textalpha}$ coupling constant is small (4-6 Hz) in helical proteins and so magnetisation transfer from all but the H\textalpha requires impractically long mixing times that would result in pronounced signal loss to relaxation.

Conversely, the NOESY experiment relies upon through space correlations that are efficient up to 5 Å and so for HR1\textsuperscript{TOCA1} it did provide sidechain fingerprints. The NOESY contains intra- and inter-residue NOEs and the intra-residue NOEs are often strong as the intra-residue protons are close in space to the NH group. Navigation from the sequentially assigned $^{15}$N-HSQC to the $^{15}$N-NOESY and $^{15}$N-TOCSY provided a more complete sidechain fingerprint including the chemical shifts of most of the sidechain protons. The sequence specific backbone assignments were therefore completed by comparison of these shifts with reference values.

The inter-residue NOEs observed in the NOESY, which may be short or long range in terms of primary structure, can later be used to derive distance restraints for structure calculations (section 4.2.3.2). These increase spectral crowding but can also aid the assignment by the provision of helpful short and medium range NOEs. For HR1\textsuperscript{TOCA1}, short range NH-NH NOEs (i→i+1) along the sequence and medium range NOEs in the $\alpha$-helical regions (i→i+3) were useful in confirming the assignments.

The amide chemical shifts of each residue are encoded in the $^{15}$N-HSQC and in the other experiments described above. As such, each spectrum is inherently linked and so can be assigned concomitantly. The navigation between the spectra is simple, using the $^{15}$N-HSQC coordinates as the reference point. The fully assigned $^{15}$N-HSQC of HR1\textsuperscript{TOCA1} is shown in Figure 4.8. This assigned spectrum can be used to navigate to regions of the triple resonance experiments and to regions of the $^{15}$N-NOESY and TOCSY experiments corresponding to each specific residue.
For the assignment of sidechain CH groups, a 3D through-bond experiment was needed, the HCCH-TOCSY. In this experiment, the magnetisation is transferred from the sidechain protons to their directly attached $^{13}$C nuclei and then transferred across the entire sidechain by isotropic $^{13}$C mixing (Figure 4.9). It is then transferred back to the proton for detection. The experiment therefore shows correlations between all of the aliphatic CH groups in the residue, including the sidechain CHs and the $\text{CaH}$.

**Figure 4.8. The assigned $^{15}$N-HSQC of HR1$^{\text{TOCA1}}$.** The HSQC was recorded on 0.9 mM HR1$^{\text{TOCA1}}$ in 20 mM sodium phosphate pH 7.4 with 150 mM NaCl and 10 % D$_2$O on a Bruker DRX500 at 25 °C. Sidechain NH$_2$S are indicated by a dotted line connecting the pair of peaks.
A well-resolved, short sidechain (Asn380) is used as an example in Figure 4.10 to illustrate the assignment of sidechain CH groups. Beginning from the $^{15}$N-HSQC, the NH strip for Asn380 was located in the HNCA, HNCACB and $^{15}$N-NOESY-HSQC as indicated by the orange arrows in Figure 4.10. The C\(\alpha\) and C\(\beta\) planes were then located in the HCCH-TOCSY by navigation from the assigned HNCA and HNCACB as indicated by the purple and maroon arrows. These planes in the HCCH-TOCSY were then compared with the appropriate NH strip in the $^{15}$N-NOESY-HSQC to locate the proton shifts for H\(\alpha\) and H\(\beta\) within the C\(\alpha\) and C\(\beta\) planes as indicated by the blue and green lines. The $^{13}$C-HSQC was then assigned simply by navigation from the HCCH-TOCSY. The $^{13}$C-HSQC provides a useful consolidator of the sidechain CH assignments and a reference to the HCCH-TOCSY, much like the $^{15}$N-HSQC constitutes a reference to the amide-correlated spectra.
A Valine residue, Val417, is used to illustrate the assignment of CH groups in longer sidechains (Figure 4.11). Valines have two sidechain methyl groups (CH₃). The three protons in a single methyl group are in free rotation about the Carbon and so have degenerate chemical shifts but the methyl groups are in different chemical environments within the protein and so their ¹³C and protons may have distinct chemical shifts from one another. The methyl group with the lowest proton shift is arbitrarily named γa and the other is γb. In the example shown in Figure 4.11, the methyl groups have indistinguishable ¹³C shifts but their proton shifts can be distinguished in the α and β strips of the HCCH-TOCSY.

Figure 4.10. Assignment of Aliphatic CH Groups in a Short Sidechain. The HNCA (purple), HNCACB (maroon) and HCCH-TOCSY (black) were recorded on 0.9 mM HRITOCA in 20 mM sodium phosphate pH 7.4 with 150 mM NaCl and 10 % D₂O on an Avance AV600 at 25 °C. The ¹⁵N-NOESY (blue) was recorded in the same conditions on a Bruker DRX500. The ¹⁵N-HSQ (orange) was used as a reference to the other spectra as indicated by the orange arrows. The Cα and Cβ peaks for Asn380, shown in the HNCA and HNCACB spectra, were used to navigate to the planes corresponding to the Cα and Cβ shifts in the HCCH-TOCSY as indicated by the purple and maroon arrows. The ¹⁵N-NOESY-HSQC was then used to locate the proton shifts as indicated with the blue and green lines.
The assignment of Val417 was carried out as follows. Firstly, the CaHα and CβHβ strips were located in the HCCH-TOCSY from the HNCA/HNCACB and $^{15}$N-NOESY-HSQC as described for Asn380 above. The cross peaks between CHα and CHγa and between CHα and CHγb, resulting from multiple transfer steps via one-bond scalar couplings ($J_{C\alpha C}$), were next located in the CaHα strip of the HCCH-TOCSY (circled in Figure 4.11). The cross peaks arising from the one-bond scalar couplings between CHβ and the CHγs were also located within the CβHβ strip. These cross peaks encode the Hγ chemical shifts, which were confirmed using the $^{15}$N-NOESY-HSQC.

The HCCH-TOCSY was next viewed in a different orientation, as shown in the ‘orthogonal window’ in Figure 4.11. The Hγa plane was located in the orthogonally viewed HCCH-TOCSY, as illustrated with the red arrow, and return cross peaks to the Hα, Hβ and Hγb were identified as indicated with green lines. The $^{13}$C shift of the first methyl group was thus identified. The CγaHγa strip in the HCCH-TOCSY was then visualised alongside the CaHα and CβHβ strips by navigating from the orthogonal window as indicated with a blue arrow. This was repeated for CγbHγb, which, in this case, has an indistinguishable $^{13}$C shift. The $^{13}$C-HSQC was assigned by navigation from the HCCH-TOCSY (Figure 4.11C). The CH groups of longer aliphatic sidechains with Hδ and Hε groups were assigned in the same way.
Figure 4.11. Assignment of Aliphatic CH Groups in Longer Sidechains. A) A schematic showing useful scalar couplings between CH groups in Valine. B) Regions of the HCCH-TOCSY are shown in two orientations. In the first panel, a $^{13}$C plane is shown, with the x- and y-axes showing proton shifts. In the orthogonal window, a proton plane is shown with $^{13}$C and proton dimensions on the x- and y-axes. The $\alpha$, $\beta$ and $\gamma$ strips shown correspond to Val417. First, the H$\gamma$ plane was located in the orthogonally viewed HCCH-TOCSY as indicated by the red arrow. Next, the return cross peaks to H$\alpha$, H$\beta$ and H$\gamma$ were located, indicated with green lines. The $^{13}$C shift of the first methyl group was then identified and the $\gamma$ strip was visualised in the first TOCSY window next to the $\alpha$ and $\beta$ strips, indicated with the blue arrow. This was repeated for $\gamma$ but the $^{13}$C shift could not be distinguished from that of $\gamma$. A single set of peaks was seen the C$\gamma$/H$\gamma$ strip, representing the two methyl groups. C) Sections of the $^{13}$C-HSQC showing the C$\alpha$H$\alpha$, C$\beta$H$\beta$ and C$\gamma$H$\gamma$ peaks.
4.2.2.3 Assignment of Proline Sidechains

Prolines have no backbone NH group and so are not recorded in the $^{15}$N-HSQC. However, magnetisation transfer occurs from the NH of a non-Proline residue to the Cα and Cβ of a preceding Proline via the carbonyl in the HN(CO)CA and HN(CO)CACB experiments. The Cα and Cβ planes of Proline residues were therefore located in the HCCH-TOCSY via navigation from the strip corresponding to the i+1 amide in the HN(CO)CA and HN(CO)CACB. The Hα and Hβ strips were located within the Cα and Cβ planes using the $^{15}$N-NOESY-HSQC, comparably to the method shown in Figure 4.10 but from the strip corresponding to the i+1 amide. The Cγ and Cδ were assigned using the method shown in Figure 4.11. A $^{13}$C-separated NOESY-HSQC was then used to confirm the assignments (Figure 4.12).

The $^{13}$C-separated NOESY-HSQC is comparable to the $^{15}$N-NOESY-HSQC but instead of the spectrum being separated based on the amide groups, it is separated by a $^{13}$C-HSQC and so NOEs to CH groups are selected. Depending on the structure, the Hδ of Prolines can be close to the NH and/or Hα of the preceding residue and so strong NH$_{i-1}$-Hδ$_i$ and/or Cα$_{i-1}$-Hδ$_i$ sequential NOEs can be observed in the $^{13}$C-NOESY-HSQC spectrum (Figure 4.12). These NOEs are useful in confirming the assignments especially when the HCCH-TOCSY is crowded.
Figure 4.12. Assignment of Proline Sidechains. A) A schematic showing the inter-residue CH-CH NOEs that are useful in the assignment of Proline residues. B) The $^{13}$C-NOESY was recorded on 0.9 mM HR$^{1}$TOCAT in 20 mM sodium phosphate pH 7.4 with 150 mM NaCl and 10 % D$_2$O on an Avance AV600 at 25 °C. The CαHα strip of Asn380 and CδHδ strip of Pro381 are shown. The useful inter-residue NOEs are indicated.
4.2.2.4 Assignment of Sidechain NH$_2$ Groups

The sidechain NH$_2$ groups of Asparagine and Glutamine residues were assigned using intra-residue NOEs between the NH$_2$ and the nearest CH group. These were the NH$_2$-H$\beta$ NOEs for Asparagine and the NH$_2$-H$\gamma$ NOEs for Glutamine. These were visible in the $^{15}$N- and $^{13}$C- NOESY-HSQCs. An example is shown in Figure 4.13.

![Diagram](image)

Figure 4.13. Assignment of Sidechain NH$_2$s. A) A schematic showing the intra-residue NOEs that are useful for assigning the sidechain NH$_2$ groups of Asparagine and Glutamine sidechains. B) The strip of the $^{15}$N-NOESY-HSQC corresponding to the sidechain NH$_2$ of Asn380 is shown in blue. There are two protons with distinguishable shifts attached to a single $^{15}$N, seen as two diagonal peaks at 7.1 and 7.7 ppm in the $^{15}$N plane that is shown. The strip corresponding to C$\beta$H$\beta$ of Asn380 in the $^{13}$C-NOESY-HSQC (pink) and HCCH-TOCSY (black) is shown alongside. There are two protons attached to a single $^{13}$C, seen as two diagonal peaks at 2.6 and 2.9 ppm. The intra-residue NOEs between the NH$_2$ groups and the C$\beta$H$\beta$ protons are circled and labelled.
4.2.2.5 Assignment of Aromatic Sidechains

The aromatic sidechains were also assigned using the $^{15}\text{N}$- and $^{13}\text{C}$-NOESY-HSQC s. Examples of two aromatic residues, Phe337 and His408, are shown in Figure 4.14.

**Figure 4.14. Assignment of Aromatic Sidechains using CH-CH NOEs.** A and B) Schematics showing useful intra-residue NOEs for assignment of Phenylalanine and Histidine sidechains. Coloured arrows indicate CH-CH NOEs that are exemplified in C and D. Black arrows indicate NOEs that are not shown. The two $\delta$ groups in Phe are coloured the same to indicate that they have degenerate chemical shifts, as are the $\varepsilon$ groups. C) The $\text{C}\alpha\text{H}\alpha$, $\text{C}\beta\text{H}\beta$, $\text{C}\varepsilon\text{H}\varepsilon$, $\text{C}\delta\text{H}\delta$, $\text{C}\zeta\text{H}\zeta$ groups of Phe337 are shown in the $^{13}\text{C}$-NOESY-HSQC. The $\text{H}\delta-\text{H}\beta$ NOEs are circled in blue, the $\text{H}\varepsilon-\text{H}\beta$ NOEs in pink, $\text{H}\delta-\text{H}\alpha$ in orange, $\text{H}\delta-\text{H}\varepsilon$ in green and $\text{H}\zeta-\text{H}\varepsilon/\delta$ in purple. D) The $\text{C}\alpha\text{H}\alpha$, $\text{C}\beta\text{H}\beta$, $\text{C}\delta\text{H}\delta$ and $\text{C}\varepsilon\text{H}\varepsilon$ groups of His408 are shown in the $^{13}\text{C}$-NOESY-HSQC. The $\text{H}\delta-\text{H}\beta$ NOEs are circled in blue, $\text{H}\delta-\text{H}\alpha$ in orange and $\text{H}\delta-\text{H}\varepsilon$ in green.
Intra-residue NOEs from $\mathrm{H}\alpha$ and $\mathrm{H}\beta$ to $\mathrm{H}\delta$ were used in both cases to locate $\mathrm{H}\delta$. The $\mathrm{H}\varepsilon$ could then be assigned based on an NOE from $\mathrm{H}\delta$ and, in the case of Phenylalanine, the $\mathrm{H}\zeta$ was assigned based on NOEs from $\mathrm{H}\delta$ and $\mathrm{H}\varepsilon$. The expected NOEs for Phe and His residues are shown in Figure 4.14A and B with example CH-CH NOEs indicated in the $\textsuperscript{13}$C-NOESY-HSQC shown Figure 4.14C and D. The fully assigned aromatic region of the $\textsuperscript{13}$C-HSQC is shown in Figure 4.15.

**Figure 4.15. Aromatic Sidechain Assignments in the $\textsuperscript{13}$C-HSQC.** The aromatic sidechain region of the $\textsuperscript{13}$C-HSQC of 0.9 mM HR1$\textsuperscript{TQCAI}$ in 20 mM sodium phosphate pH 7.4, 150 mM NaCl and 10 % D$_2$O recorded on a Bruker DRX500 at 25 °C. A * indicates chemical shift degeneracy of two $\mathrm{C}\delta\mathrm{H}\delta$ or $\mathrm{C}\varepsilon\mathrm{H}\varepsilon$ groups.
Intra-residue NOEs between Hδ and NH, indicated for Phe and His in Figure 4.14A and B and for Tyr and Trp in Figure 4.16A and B, are also invaluable in the assignment of aromatic sidechains. Example CH-NH NOEs are shown for Tyr and Trp in Figure 4.16C and D. NOEs between Hδ1 and Hζ2 and the sidechain NH (Hε1) of Trp413 were used to complete the sidechain assignment.

Figure 4.16. Assignment of Aromatic Sidechains using NH-CH NOEs. A and B) Schematics showing useful intra-residue NOEs for assignment of Tyrosine and Tryptophan sidechains. Coloured arrows indicate the CH-NH NOEs that are shown in C and D. Black arrows indicate NOEs that are not shown. The two δ groups of Tyrosine are coloured the same to indicate that they have degenerate chemical shifts, as are the ε groups. C) The NH of Tyr377 is shown in the 15N-NOESY-HSQC (blue) and the CδHδ and CeHε of Tyr377 are shown in the 13C-NOESY-HSQC (pink). The Hδ-NH NOEs are circled in blue and the Hε-NH NOEs are circled in green. For Tyr377, the Hδ and Hε chemical shifts are indistinguishable and so the NH-Hδ/Hε NOEs are overlaid. D) The backbone NH and sidechain NHε1 of Trp413 are shown in the 15N-NOESY-HSQC (blue) and the Cδ1Hδ1 and Cζ2Hζ2 of Trp413 are shown in the 13C-NOESY-HSQC (pink). The Hδ-NH NOEs are circled in blue, the Hδ1-NHε1 NOEs are circled in green and the Hζ2-NHε1 NOEs are circled in red.
4.2.2.6 The Completed Assignment

Overall, the $^1$H, $^{15}$N and $^{13}$C assignment was 90% complete. Of the TOCA1 HR1 domain residues (excluding additional residues encoded by the vector), 97.9% of the backbone resonances were assigned, including 100% of the amide protons. The following groups account for the unassigned atoms: the $^{13}$CO of Ser425 and Glu426 were not visible in the HNCO spectrum; the Nitrogen atom of each Proline as it is not correlated with a proton; and the $^{13}$CO of the six residues preceding Prolines. All of the observed sidechain resonances have been fully assigned, equating to 94.3% of all sidechain NH and CH groups in the protein. No Lysine $N_\zeta$, Arginine $N_\eta$ and Histidine $N_\delta$ and $N_\epsilon$ were observed, nor were the NH$\epsilon$ groups of Arginines 348, 358, 402, 404 and 424. These account for the remaining unassigned NH and CH groups.

The $^1$H, $^{15}$N and $^{13}$C chemical shift data was validated using iCING (Doreleijers et al. 2012) prior to deposition in the BioMagResBank (accession number 25945). Shift validation occurs by comparison of the measured chemical shifts with database of reference chemical shifts and any that do not fall within normal ranges are highlighted for manual verification.

4.2.2.7 Assignment of NOEs in the NOESY Spectra

The assignment of intra-residue NOEs in the two NOESY spectra was completed during the sidechain assignment by comparison with the HCCH-TOCSY, as illustrated for the $^{15}$N-NOESY in Figure 4.10. Intra-residue NOEs in the $^{13}$C-NOESY were simply assigned by overlaying the spectrum with the assigned HCCH-TOCSY. Some sequential inter-residue NOEs in both NOESY spectra were also assigned during the sidechain assignment by matching the chemical shifts of the NOEs with the known proton shifts of the neighbouring sidechains. Knowledge of the structures of each amino acid and the sequence of HR1$^{TOCA1}$ was important for this, as has already been illustrated above, for example in Figure 4.12. Here, knowledge of the sequence (Asn380-Pro381) and the amino acid structures predicted the Asn380H$\alpha$-Pro381H$\delta$ inter-residue NOEs, which could therefore be readily located and assigned.

Non-sequential, short range inter-residue NOEs were assigned by matching the chemical shifts of the NOEs with the known proton shifts of other sidechains that are
HR1^{TOCA1} is expected to comprise a pair of α-helices and so NOEs were expected between every fourth residue (corresponding to one turn of an α-helix) within the predicted helices. Some examples of i→i+3 NOEs are shown in Figure 4.17. For example, an NOE from Thr396 Hα to Asn399 NH was seen in the Thr396 CaHα strip of the $^{13}$C-NOESY-HSQC and a return peak was seen in the Asn399 NH strip of the $^{15}$N-NOESY-HSQC (Figure 4.17, indicated with purple circles).

Additional NOEs between the other CH or NH groups within the same two residues were also used to confirm the assignments.

**Figure 4.17. Assignment of i→i+3 NOEs.** The backbone NH and sidechain NH$_2$ strips of Asn399 are shown in the $^{15}$N-NOESY (blue). The CaHα strip of Thr396 and the CβHβ strip of Asn399 are shown in the $^{13}$C-NOESY (pink). i→i+3 NOEs are circled in the Thr396 CaHα strip of the $^{13}$C-NOESY and labelled. The return peaks are circled using the same colours in the Asn399 CβHβ strip of the $^{13}$C-NOESY or backbone N and sidechain NHδ strips of the $^{15}$N-NOESY.
The long range NOEs could not be so readily predicted based on sequence or homology and so the assignments were predominantly made by matching chemical shifts with assigned protons of other residues. This assignment was done iteratively, using the initial structures from structure calculations to complete, confirm and correct the assignments.

Many NOEs could not be unambiguously assigned due to chemical shift degeneracy and spectral overlap. Overall 99.4 % of the cross peaks were assigned, with ~60 % of the assignments fully unambiguous, prior to initial structure calculations.

4.2.3 Structure Calculations
Structure calculations were carried out using Aria 2.3.1 (Rieping et al. 2007) interfaced with CNS (Brünger et al. 1998). The following sections describe the derivation of the necessary restraints and the iterative process of structure determination.

4.2.3.1 Dihedral Angles
Chemical shifts are closely related to local protein structure such that they can be predicted from local geometry using programs such as SPARTA+ (Shen and Bax 2010) and vice versa by programs such as TALOS-N (Shen and Bax 2013). TALOS-N was therefore first used to predict the dihedral angles within HR1 TOCA1 based on the chemical shift data.

TALOS-N creates dihedral angle restraint lists from chemical shifts based on comparisons with two databases of protein structures. The first contains 580 high-resolution X-ray structures collected from the PDB (Berman et al. 2012), for which near complete backbone NMR assignments are available from the BMRB (Markley et al. 2008). The second comprises 9,523 high-resolution X-ray structures, from which chemical shift predictions have been made using SPARTA+. Every residue in the database is given a $\phi/\psi$ distribution code, which relates it to a particular region of the Ramachandran plot, to represent its backbone conformation (as derived from crystal structures). An artificial neural network is then used to correlate this distribution code
with NMR chemical shifts. As such, two databases of relationships between geometry and chemical shift are created.

TALOS-N begins by predicting $\phi/\psi$ distributions for each residue of the query protein from the input chemical shifts using an artificial neural network based upon the smaller database. The query protein is then divided into heptapeptides and the larger database is searched for heptapeptides with $\phi/\psi$ distributions that match the predicted $\phi/\psi$ distributions of each query heptapeptide. The 1000 best matches are then carried forward and their SPARTA+ predicted chemical shifts compared with the experimental chemical shifts of the query heptapeptide. The 25 best matched heptapeptides are kept as the final matched fragments. The average $\phi/\psi$ distribution for the central residue of the 25 matched heptapeptides is considered a prediction of the $\phi/\psi$ distribution of the central residue in the query heptapeptide. In this way, TALOS-N provides average predicted angles for each residue.

To summarise, TALOS-N takes chemical shift data and turns it into dihedral angle data based on a small database of X-ray structures with experimental NMR chemical shifts available, turns it back into chemical shift data using a large database of SPARTA+-predicted chemical shifts and thus selects 25 well matched heptapeptides from a large database of high resolution X-ray structures. It then turns the chemical shift data into dihedral angles, which was the original goal. The purpose of this convoluted method is to allow the much larger database of structures to be used, which in turn allows for a much larger number of peptide matches to be found than when using only the smaller database. Furthermore, it allows for heptapeptides to be compared rather than the tripeptide methods that were used in primitive versions of TALOS.

If the $\phi/\psi$ distributions of the central residue of the 25 matched heptapeptides cluster in the same region of the Ramachandran plot the prediction is classed as ‘strong’. If less than 25 but more than 10 cluster within the same region of the Ramachandran plot TALOS-N will still make a prediction but it will be classed as ‘generous’. If there is no consensus in the database matches, the prediction is marked with ‘warn’ and the
prediction should be excluded. The final class of prediction is ‘dynamic’, indicating that the data is suggestive of a dynamic conformation for that residue.

For HR1^TOCA1, TALOS-N provided a table of dihedral angle predictions for residues 324-425 (Appendix 2). The majority of the predictions at the N-terminal region (324-337) were classed as warnings so were excluded from the table of dihedral restraints for structure calculations. The C-terminal region (419-426) was classed as dynamic so the predictions here were also excluded. The final table of dihedral angle restraints that was used for structure calculations comprised all of the strong predictions ± their standard deviations (for 338-383 and 384-418). No restraints were given for the remaining residues.

4.2.3.2 Distance Restraints
The cross peak lists from the 3D $^{15}$N-NOESY-HSQC and $^{13}$C-NOESY-HSQC were input into Aria 2.3.1 and used to derive distance restraints. The NOE is related to distance according to the following relationship:

$$\text{NOE} \propto D^6$$

but it is also affected by a number of other factors including protein concentration, spin diffusion and internal dynamics (Nilges 1996). Therefore, an unknown constant of proportionality relates the NOE to distance. To account for these effects and to derive an estimate of this constant across the whole spectrum, the NOE lists are subjected to distance calibration protocols throughout the structure calculations in Aria, as described by (Linge et al. 2001). For iteration zero this constant of proportionality, known as the calibration factor $C_{cal}$, is derived from an extended structure with randomized torsion angles and idealized side chain geometries so is dominated by intra-residue distances. In later iterations it is derived from the calculated structures. It is adjusted after each iteration and so is expected to improve in later iterations as the structures improve.

In the most simple case, the calibration after iteration 0 is based directly on distances calculated from the lowest energy structures produced in each iteration, typically the best 35-50%. The arithmetic average across the ensemble is calculated for each proton pair and these averages, known as the ‘distance characteristics’, are then used
to define the calibration factor, $C_{\text{cal}}$, for the spectrum. The calibrated distance restraints, $d^{\text{obs}}$, are extracted for the next iteration, according to the following equation:

$$d^{\text{obs}} = (C_{\text{cal}} V)^{-1/6}$$

where $V$ is provided by the peak volume or intensity (in this case intensity was used).

In fact, better calibrations are actually achieved by indirect use of the ‘distance characteristics’. The distance characteristics derived from the early structures are used to calculate a relaxation matrix, which is in turn used to simulate an NOE spectrum. The calibration factor is derived empirically by comparing the experimental data with the theoretical peak volumes or intensities back calculated from the initial structures.

The ambiguous restraints and partial assignments can be utilised in a similar way but are instead considered in terms of the ‘effective distance’, $D$. This combines contributions from the distances between the proton pairs for all of the potential assignments by summation of the distances over $N_\delta$ possible assignments, where $N_\delta$ is the number of possible assignments within the chemical shift tolerance, $\delta$. This is described by the following equation:

$$D = \left( \sum_{a=1}^{N_\delta} d_a^{-6} \right)^{-1/6}$$

As with the unambiguous NOEs, the ambiguous NOEs depend upon the inverse sixth power of $D$:

$$\text{NOE} \propto D^{-6} = \left( \sum_{a=1}^{N_\delta} d_a^{-6} \right)$$

### 4.2.3.3 Initial Structure Calculations

Iteration 0 was begun with randomization of the $\phi$ and $\psi$ angles of an extended chain of idealized geometry. At this stage no restraints were excluded, i.e. the violation tolerance was set to a very high value (1000 Å), and partial assignments of ambiguous restraints were made based only on chemical shift (assignment parameter, $p$, is set to 1.0). The first ensemble of structures from iteration 0 was used for calibration of the NOESY spectra based on short distances before iteration 1, as described above and this calibration was repeated and improved at each iteration.
The violation tolerance was reduced from 1000 Å to 5 Å in iteration 1, 3 Å in iteration 2, 1 Å in iterations 3-5 and then 0.1 Å in iterations 6-8. The assignment parameter, $p$, was also reduced in subsequent iterations (1.0, 0.9999, 0.999, 0.99, 0.98, 0.96, 0.93, 0.9, 0.8) such that the ambiguous assignments in later iterations are no longer made solely based on chemical shift but instead are made preferentially towards cases where the inter-proton distance is consistently short in the structure ensemble. In this way, Aria uses the structures produced in earlier iterations to improve the calibration of the NOESY spectra and guide the assignments for later iterations.

More than 3000 non-degenerate NOE restraints were used in the initial structure calculations (~ 60 % unambiguous) during which 20 structures were calculated in each of 8 iterations and the 10 lowest energy structures were water refined and analysed. The water refined structures were subjected to violation analysis, producing a list of violated distance restraints for manual inspection.

4.2.3.4 Manual Refinement
The first ensemble of structures contained more than 600 violated distance restraints. One purpose of violation analysis is to identify peaks caused by noise and indeed, many of these violations resulted from spectral artifacts that were simply removed prior to further rounds of structure calculations. The improved structures still contained more than 200 violations, many of which had arisen from incorrect assignments of NOEs. For example, incorrect assignments of ambiguous NOEs were made by Aria to proton pairs at opposite ends of the coiled-coil causing it to curl up. The unambiguous and ambiguous assignments were therefore improved with each round of calculations by manually checking them against the structures produced.

Once the major violations had been rectified, the structural knowledge from later rounds was used to assign many NOEs that were not previously assignable without structural insight and to correct any remaining incorrect assignments. The structure calculations and violation analyses were repeated until no distance restraints were violated. After this manual refinement was complete, 99.8 % of the NOEs in the $^{13}$C-NOESY-HSQC and 100 % of the NOEs in the $^{15}$N-NOESY-HSQC were assigned.
% of the assignments were unambiguous and the automatically assigned peaks had been manually confirmed.

Within the analysis steps of the structure calculations performed in Aria, the structure ensembles were also checked against the dihedral angle restraint lists generated in TALOS-N for dihedral angle violations. Some of the structures produced in early rounds of calculations included dihedral angle violations, mainly in the N-terminal tail and the inter-helical loop region. These were caused by the same incorrect assignments that led to distance restraint violations and so there were no dihedral angle restraint violations once the assignments had been corrected and the distance violations removed.

The structure calculations were repeated with 100 structures calculated in the final iteration. 2,778 non-degenerate NOE restraints (1,791 unambiguous and 987 ambiguous) were used in the first iteration and 1,845 unambiguous NOEs and 757 ambiguous NOEs remained after 8 iterations. 164 dihedral angle restraints were used. The 50 lowest energy structures were water refined and the 35 lowest energy structures following refinement were analysed as described in the following sections.

### 4.2.4 Structure Validation

Firstly, the structure ensemble was assessed using Stride (Heinig and Frishman 2004) in order to define the \( \alpha \)-helical regions based on the dihedral angles (Helix 1: 342-379 and Helix 2: 386-420). The average structure was then calculated based on these ordered regions using CNS and the 35 structures ranked according to their RMSD from the average. The structure closest to the average (i.e. the lowest RMSD) was considered the best representation of the structure.

The RMSDs of the backbone and sidechain atoms were assessed across the ensemble to give a measure of the convergence. For the ordered regions of the protein, the backbone RMSD was 0.67 and sidechain RMSD was 1.1 (Table 4.1). A variety of other analyses, including an assessment of the dihedral angles with reference to the Ramachandran plot, were carried out using iCING (Doreleijers et al. 2012) before the
structure was deposited in the PDB (PDB code: 5FRG) where a full validation report is available. Table 4.1 includes the important validation statistics from the PDB.

These statistics show that the ensemble and the structure closest to the mean are well defined by the NMR data. For example, 99.2% of the dihedral angles across the ensemble were within allowed regions of the Ramachandran plot (Table 4.1 and Figure 4.18) and the 0.8% found to be outliers were all in the unstructured N-terminal region (residues 324-336) except for Ser422, which is in the unstructured C-terminal region. For the ordered residues, all of the dihedral angles were within allowed regions. Furthermore, sidechain analysis showed that 99% of the sidechains within ordered regions were rotomeric (i.e. their torsion angles fall within preferred ranges).

The global validation metrics from the PDB validation report, also shown in Figure 4.18, shows that the structure is high quality relative to all structures deposited in the PDB. The percentile ranks indicate the percentage of structures in the PDB that are better (right of marker) or poorer (left of marker) than this structure in terms of the quality indicator, and so the percentile ranks being towards the right-hand side (blue) indicates a high quality structure.
Table 4.1: Experimental Restraints and Structural Statistics

<table>
<thead>
<tr>
<th>Experimental restraints</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness of chemical shift assignment: 90 %</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Distance restraints:
| Total non-degenerate | 2602 | |
| Unambiguous | 1845 | |
| Ambiguous | 757 | |

Coordinate precision for well-ordered regions

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD of backbone atoms (342-379, 386-419) (Å)</td>
<td>0.67 ± 0.14</td>
<td>0.46</td>
</tr>
<tr>
<td>RMSD of all heavy atoms (342-379, 386-419) (Å)</td>
<td>1.10 ± 0.12</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Ramachandran analysis for all residues

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in most favoured regions (%)</td>
<td>93.1</td>
<td>96.1</td>
</tr>
<tr>
<td>Residues in additionally allowed regions (%)</td>
<td>6.1</td>
<td>3.9</td>
</tr>
<tr>
<td>Residues in disallowed regions (%)</td>
<td>0.8</td>
<td>0</td>
</tr>
</tbody>
</table>

Ramachandran analysis for ordered regions

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in most favoured regions (%)</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>Residues in additionally allowed regions (%)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Residues in disallowed regions (%)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Sidechain analysis for ordered regions

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotomeric sidechains (%)</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>Non-rotomeric sidechains (%)</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

RMS deviations for all residues

<table>
<thead>
<tr>
<th>from the experimental restraints:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NOE distances (Å)</td>
<td>0.011 ± 0.0027</td>
<td>0.011</td>
</tr>
<tr>
<td>Dihedral angles (˚)</td>
<td>0.14 ± 0.067</td>
<td>0.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>from idealised geometry:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond lengths (Å)</td>
<td>0.0027 ± 0.00009</td>
<td>0.00269</td>
</tr>
<tr>
<td>Angles (˚)</td>
<td>0.434 ± 0.012</td>
<td>0.427</td>
</tr>
<tr>
<td>Impropers (˚)</td>
<td>1.025 ± 0.073</td>
<td>1.11</td>
</tr>
</tbody>
</table>

Column A contains information on the ensemble of 35 structures.
Column B contains information for the structure that is closest to the mean.
Figure 4.18. Structure Validation of \textsuperscript{HR1}^{TOCA1}. A) The Ramachandran plot for the complete sequence of \textsuperscript{HR1}^{TOCA1} across the ensemble of 35 structures produced using MolProbity. Outliers are marked as pink circles. B) The global validation metrics from the PDB. The solid black rectangles indicate the percentile rank based on all of the structures in the PDB (up until 30\textsuperscript{th} December 2015) in terms of the validation metric. The open rectangles indicate the percentile rank based on all of the NMR structures. The values given are the percentages of NMR structures better than this structure and so a lower value indicates a better quality structure.
4.2.5 The Structure

The TOCA1 HR1 domain comprises two $\alpha$-helices (residues 342-379 and 390-418) folded into a coiled-coil (Figure 4.19A-C) with a slight left handed twist (Figure 4.19C). The two helices are in good agreement with the helices expected based on the dihedral angle predictions from TALOS-N (residues 342-379 and 386-417). The first helix is longer than the second (38 residues compared to 29) but the average twist (angle of one residue to the next) and average number of residues per turn of the helix are comparable between the two. For helix 1 the twist is 99.5 degrees and the number of residues per turn is 3.62 and for helix 2 the values are 99.8 degrees and 3.61 residues per turn, as determined using HELANAL-Plus (Kumar and Bansal 2012). The two helices come together into an anti-parallel coiled-coil, which is predominantly stabilised by hydrophobic contacts, illustrated in Figure 4.20. Helix 2 is relatively straight and helix 1 twists slightly around it, as is evident in Figure 4.20B.

Figure 4.19. The NMR Structure of HR1$^{TOCA1}$. A) The backbone trace of the 35 lowest energy structures of HR1$^{TOCA1}$ overlaid with the structure closest to the mean. B) A cartoon representation of the structure closest to the mean. C) An end-on view illustrating the slight left-handed twist of the coiled-coil. Flexible regions at the N- and C-termini (330-333 and 421-426) are omitted for clarity in all cases. These images were generated using MacPyMol.
In agreement with the TALOS-N dihedral angle predictions, the C-terminal region following helix 2 (419-426) is not well defined in the ensemble of structures indicating that this region is flexible. The N-terminal region (330-341) is better

Figure 4.20. Inter-helical Contacts in the Coiled-coil. A) A helical wheel representation of the coiled-coil. The hydrophobic residues in the coiled-coil interface are coloured in pairs corresponding to the closest residue on the opposite helix and a polar interaction is indicated with a dotted rectangle. B) A cartoon representation of the HR1 domain is shown along with each separate helix. The inter-helical contacts are shown as sticks and coloured. The twist is evident from the positioning of the contact residues.
defined in the ensemble, indicating a degree of secondary structure that was not predicted by TALOS-N. TALOS-N was not able to provide dihedral angle predictions for the majority of this region but suggested that some of the chemical shifts were indicative of random coil (residue 334 and residues 338-340).

Contrary to this, the $\phi$ and $\psi$ angles of residues 334-340, assessed from the ensemble of structures using Stride (Heinig and Frishman 2004), fall within Ramachandran regions corresponding to turns in 54% of the structures calculated. In the remaining 46% of the structures, the dihedral angles of residues 334-336 are representative of random coil whilst 337-340 are representative of turns. NOEs between residues separated by one (332-334, 333-335 etc.) or two (337-340) residues in the sequence further define the series of turns. For example, there are unambiguous NOEs between Pro332 H$\alpha$ and Leu334 H$\delta$ and between the H$\beta$s of Ala333 and Leu344 and the NH of Asp336.

The N-terminal region reverses direction as a result of these turns, allowing it to contact both helices of the coiled-coil. A network of long range NOEs were observed, linking Leu334, Glu335 and Asp336 with Trp413 of helix 2 and Leu334 with Lys409 of helix 2. For example, there are NOEs between the H$\delta$s of Glu335 and the aromatic sidechain protons of Trp413. Leu340 sits in the turn between this region and helix 1 with NOEs to Phe337, Arg345 and Arg348 of helix 1. For example there are NOEs between the H$\delta$s and Hys of Leu340 and the H$\alpha$s and H$\beta$s of Phe337. The Phe337 sidechain protons also share NOEs with sidechain protons of Leu349, Arg345, and Arg348, which indicate that the sidechain is fixed, pointing away from the solvent and towards the coiled-coil. The important NOEs responsible for the fixed position of this N-terminal region in the ensemble of NMR structures are summarized in Figure 4.21.
The inter-helical loop region was predicted by TALOS-N to encompass residues 380-385 as their chemical shifts were indicative of random coil. The NMR structure, however, reveals a slightly longer loop region of ten residues (380-389) with two short $3_{10}$ helices (residues 381-383 and 386-389) separated by an unstructured region. This loop region is relatively rigid, with sidechains of the region pointing back towards helix 1 and making contacts with residues of the helix. For example, there are a number of distinct NOEs between the sidechains of Asn380 and Met383 of the loop region and Tyr377 and Val376 of helix 1. The backbone NH and CH$_\alpha$ groups of Gly384 and Asp385 also show NOEs with the sidechain of Tyr377. These NOEs are summarised in Figure 4.22.
4.2.6 A Comparison with Other HR1 Domains

The anti-parallel coiled-coil of the TOCA1 HR1 domain is reminiscent of the HR1 domains of CIP4 (Kobashigawa et al. 2009) (PDB code:2KE4) and PRK1 (Maesaki et al. 1999; Owen et al. 2003) (PDB codes:1CXZ, 1URF), as shown in Figure 4.23, but there are notable differences between the structures. For example in the length of the helices and inter-helical loops. A sequence alignment, onto which the secondary structure elements are mapped, is also shown in Figure 4.23 for comparison.

![Figure 4.22. The Important NOEs in the Inter-helical Loop Region.](image)
Figure 4.23. A Comparison of the Structures and Sequences of the TOCA and PRK Family HR1 Domains. A-C) Cartoon representations of the structure closest to the mean for the TOCA1 HR1 (A), CIP4 HR1 (B), PRK1 HR1a (C) and PRK1 HR1b (D) domains. E) A sequence alignment between the TOCA1, CIP4 and PRK1 HR1 domains performed using Expresso (Armougom et al. 2006) with the secondary structure indicated. Differences can be seen in the inter-helical loop, N-terminal regions and the lengths of the helices.
It should be noted that the features of the HR1a domain that are included in the following comparison are derived from its crystal structure in complex with RhoA. The other 3 HR1 domain structures are NMR structures of the free HR1 domains. Some differences may therefore result from the different methods used and from G protein binding. Furthermore, the NMR structure of CIP4 that is included in the comparison is of very low quality. For example, it is in the lowest 31% of PDB structures for Ramachandran outliers with only 86% of the dihedral angles falling within favoured regions and is in the lowest 8% of structures for sidechain rotomers with only 85% rotomeric sidechains. In addition, the ensemble of the 20 lowest energy structures has an all-atom clash score of 47, meaning that 47 out of every 1000 atoms is too close to another atom. This equates to 63 clashes in the best model and 1256 clashes in the complete ensemble. The poor quality of the structure is apparent in the cartoon representation shown in Figure 4.23B. The structural features discussed in relation to CIP4 should therefore be considered with this in mind.

The length of helix 1 is the same (38 residues) in TOCA1, CIP4 and HR1a and shorter (30 residues) in HR1b. Helix 2 is more variable (29, 35, 25 and 32 residues in TOCA1, CIP4, HR1a and HR1b). Helix 1 of all four HR1s and helix 2 of TOCA1, CIP4 and HR1a are similar in terms of twist and number of residues per turn (~99 degrees and 3.6 residues). Helix 2 of HR1b, however, has fewer residues per turn (3.42) and a larger average twist (105 degrees) indicating a slightly looser helix.

As with TOCA1, helix 2 is relatively straight in CIP4, HR1a and HR1b and helix 1 twists around it. The direction of the twist is left handed for all four structures but the degree of twist is not the same due to the variable length ratios of helix 1:helix 2 and variable angles of one helix with respect to the other. For example, HR1a has the biggest difference between the lengths of the two helices (38:25 residues) such that the longer helix protrudes significantly past either end of helix 2. Helix 2 of HR1a and HR1b also tilt away from helix 1 at the open end such that the inter-helical distance is not uniform along the coiled-coils (Figure 4.24). The variable lengths and angles result in different twists, with the two TOCA family HR1 domains exhibiting slightly closer twists than the two PRK family HR1 domains.
The structure of the HR1a domain of PRK1 includes a region N-terminal to the coiled-coil that forms a short α-helix and packs against both helices of the HR1 domain (Maesaki et al. 1999). The corresponding region of the CIP4 HR1 domain forms a series of turns but is not well defined in the family of structures indicating that it is flexible (Kobashigawa et al. 2009). The corresponding region of TOCA1 (residues 334-340) appears to be most similar to CIP4 as it forms a series of turns (Figures 4.19) but unlike CIP4 it is well defined in the ensemble of structures and has many NOEs with helix 1 and 2, as described above and shown in Figure 4.21.

The inter-helical loop regions of the two families of HR1 domains are also different. Firstly, they differ in length, with the PRK1 inter-helical regions being shorter (4 residues for HR1a and 6 residues for HR1b) than the equivalent region in TOCA1 and CIP4 (10 and 9 residues respectively). Secondly, the short loop regions of the PRK1 HR1 domains are unstructured whilst the longer CIP4 and TOCA1 loops exhibit 3_{10} helical structure. Finally, the sidechains of the HR1a and HR1b loops point away from the coiled-coil (Figure 4.25) and, for HR1b, the sidechains are relatively flexible as illustrated by their range of positions in the ensemble. In particular, Lys158 (purple in Figure 4.25B) appears in a large range of orientations. For TOCA1 and CIP4, the

---

**Figure 4.24. A Comparison of the Helices in the TOCA and PRK Family HR1 Domains.** A diagram to represent the differences in helix lengths, separation and angles. The two helices of each HR1 domain are illustrated as rectangles and their length in residue number is given within the rectangle. The approximate distance between the two helices is given between the rectangles in Å.
sidechains of the loop are better defined in the ensemble indicating that the loop is more rigid. In TOCA1, the sidechains point back towards the coiled-coil and the NOEs between the inter-helical region and helix 1 that were shown in Figure 4.22 define a number of contacts, which explain the sidechain orientations and the rigidity.

Figure 4.25. A Comparison of the Loop Regions in the HR1 Domains. Cartoon representations of the inter-helical loop regions in A) the crystal structure of HR1αPRK1 in complex with RhoA and B-D) the NMR structures of B) HR1βPRK1, C) HR1CIP4 and D) HR1TOCA1. The loop sidechains are shown as sticks, coloured and labelled. For the NMR structures, the structure ensemble and the structure closest to the mean are shown.
The structural differences in the N-terminal region or in the inter-helical loops could be linked to the differential affinities of the HR1 domains for their G proteins. The inter-helical loop regions are perhaps the most interesting, as this region of TOCA1 contains the critical MGD motif, mutation of which abolishes Cdc42 binding (Ho et al. 2004). Should these residues represent direct Cdc42 contacts, their occlusion may be related to the lower affinity of this particular G protein-HR1 domain interaction. Without knowing where Cdc42 contacts the HR1 domain of TOCA1, however, such conclusions are speculative. For example, it remains possible that the MGD motif does not lie in the interface but that the mutation instead affects binding by an allosteric mechanism.

The PRK HR1 domains are classed as Type D anti-parallel coiled-coil effectors, interacting via a longer NT-helix and shorter CT-helix with the inter-helical loop pointing towards the GTPase (reviewed in Mott and Owen 2015). Given the similarities between the PRK and TOCA HR1 domains, TOCA1 is also likely to be included in this class of helical pair effector proteins, with the inter-helical loop pointing towards the GTPase. NMR studies of the interaction and structural studies of the complex were next pursued, as described in the following chapter.
5 Structural Investigation of the Cdc42-HR1TOCA1 Interaction

5.1 Introduction

Structural studies of the Cdc42-HR1TOCA1 interaction will be useful for answering some of the questions surrounding G protein signalling pathways in general and in relation to specific Cdc42/TOCA1-dependent actin polymerisation pathways. Firstly, structural information is needed to investigate the molecular basis for the differential binding specificities of different G protein-HR1 interactions. The Cdc42-specific HR1 domains of the TOCA family have not been studied in complex with Cdc42 and so it is not known whether the binding is comparable to the PRK-Rac1/Rho interactions. Rho family members share a high degree of sequence and structural similarity (for example, Rac1 and Cdc42 share 71% sequence identity, RhoA and Cdc42 share 52% identity and Rac1 and RhoA share 57% identity) and so the means by which they achieve specific HR1 interactions remains a prominent question in GTPase signalling.

Secondly, structural insight is needed to investigate the mechanisms that underlie the vastly different affinities of the G protein-HR1 interactions studied so far. The lower affinity of the Cdc42-HR1TOCA1 interaction compared to other Rho-HR1 interactions suggests that the interaction may be remarkably different. Some interesting structural differences can be seen between the TOCA HR1 and the PRK HR1 domains (Chapter 4), which may relate to the lower binding affinity.

Thirdly, structural analysis of the Cdc42-HR1TOCA1 interaction will be useful in determining whether the binding sites of TOCA1 and N-WASP on Cdc42 overlap and so to address the question of whether TOCA1 and N-WASP can bind to a single molecule of Cdc42 simultaneously. Our current understanding of the roles and regulation of the proteins involved in Cdc42-dependent pathways of actin polymerisation is limited in part by this unknown. The following chapter describes a structural investigation into the Cdc42-TOCA1 binding interface using NMR spectroscopy.
5.2 Investigating the Cdc42-HRI<sup>TOCA1</sup> Interaction by NMR

NMR can be used to study protein-ligand and protein-protein interactions in a number of ways (reviewed in Zuiderweg 2002). Typically, an NMR study of a protein-protein interaction will begin with titration experiments, in which an isotope labelled protein is observed (for example using the $^{15}$N-HSQC) in the presence of increasing concentrations of an unlabelled binding partner. If the two proteins interact, the labelled protein will undergo chemical shift perturbations (CSPs) caused by changes in the chemical environment around residues that are affected during the interaction. This method can be used to map the binding interface so is known as chemical shift mapping (reviewed in Williamson 2013). In addition to providing a simple method to define a binding interface, these experiments can be used to determine whether the complex is amenable to further structural studies using NMR.

During a titration experiment, the proteins exchange between free and bound states at a rate which depends in part upon the binding affinity and the sample temperature. Exchange between free and bound states therefore occurs at different rates, which can be categorized into three exchange regimes: fast, intermediate and slow. It is not the absolute rate of exchange that is important, but the rate of exchange ($k_{ex}$) relative to the difference in the chemical shift (or frequency, $\Delta \omega$) between a residue in the free and bound states. When $k_{ex} > \Delta \omega$, the exchange is said to be in the fast regime, when $k_{ex} < \Delta \omega$ the exchange is slow and when $k_{ex} \approx \Delta \omega$ the exchange is intermediate. An increase in temperature at a given magnetic field would increase $k_{ex}$ relative to $\Delta \omega$, whilst an increase in magnetic field at a given temperature would increase $\Delta \omega$ relative to $k_{ex}$. This is illustrated in Figure 5.1 and described in more detail elsewhere (Cavanagh et al. 1998; Williamson 2013).

For a residue exchanging within the fast regime, a single peak is observed, which represents a population-weighted average between the free and bound states. Fast exchange typically occurs for interactions with a $K_d > 3 \mu$M (Williamson 2013) and is recognised in a two-dimensional spectrum such as the $^{15}$N-HSQC when a peak appears to move gradually from one place to another with increasing concentrations of ligand (Figure 5.1B). For a residue exchanging within the slow regime, two peaks are observed, one at the original position and one at the new position. The intensity of the
original peak decreases and of the new peak increases as the population of protein in
the free state decreases and the bound state increases. Intermediate exchange is
classified by chemical shift changes that are accompanied by line broadening,
which can result in the complete disappearance of a peak. For most systems, exchange
on the millisecond timescale will fall into the intermediate limit.

Figure 5.1. Exchange Regimes in NMR. A) The diagram illustrates slow, intermediate and fast exchange of a single resonance between two different conformational states (\( \omega_{\text{free}} \) and \( \omega_{\text{bound}} \)). In the slow regime, two peaks are observed. In the fast regime, a single peak is observed at an average position. \( k_{\text{ex}} \) increases with temperature and \( \Delta \omega \) increases with magnetic field. B) Examples are shown of each regime in a \(^{15}\)N-HSQC.
5.2.1 Mapping the Cdc42 Binding Site on HR1\textsuperscript{TOCA1}

5.2.1.1 HSQC Titrations

Firstly, NMR titrations were used to map the Cdc42 binding interface onto the TOCA1 HR1 domain. A series of \textsuperscript{15}N-HSQCs of HR1\textsuperscript{TOCA1} were recorded in the presence of increasing concentrations of unlabelled Cdc42\textsuperscript{Δ7Q61L}·GMPPNP. The spectra of free and saturated HR1 domain are shown overlain in Figure 5.2. Some HR1 residues were in fast exchange when Cdc42 was added, examples of which are marked on Figure 5.2A and shown in Figure 5.2B. These peaks can be tracked and their changes quantified in order to map the binding interface. The titration was continued until the peaks in fast exchange had stopped changing, up to a ratio of 1:4. No differences were seen between the spectra recorded at 1:3 and 1:4.

In addition to the peaks undergoing fast exchange, the intensities of some peaks decreased in the presence of Cdc42 and some peaks disappeared completely, possibly due to exchange on the intermediate timescale as described in Figure 5.1. Examples of disappearing peaks are labelled ‘intermediate’ in Figure 5.2. Overall, 13 peaks had disappeared from the spectrum when HR1\textsuperscript{TOCA1} was saturated with Cdc42, leaving 106 of the expected 119 peaks. There were no new peaks, suggesting that none of the residues were in the slow exchange limit.
Figure 5.2. Mapping the Cdc42-binding site onto HR1TOCA1 – 15N-HISQCs. A) 15N-HISQC's recorded on 0.2 mM 15N-HR1TOCA1, free (black) or in the presence of 3-fold molar excess of unlabelled Cdc42 (red), in NMR Titration Buffer on a Bruker DRX500 at 25 °C. Two regions of interest containing exchanging peaks are shown enlarged with the assignments. B) The titration from free HR1 (black) up to saturating levels of Cdc42 (red) is shown for one peak (the amide of Gly384), undergoing fast exchange to illustrate how the fast exchanging peaks were mapped. The equation for calculating the chemical shift perturbation (CSP) is shown. The spectrum with 4-fold Cdc42 is shown in grey to show that there were no more changes to the fast exchanging peaks. C) Another region is shown providing examples of an unchanged peak (Gln344), a peak undergoing intermediate exchange (Lys392) and another example of a peak undergoing fast exchange (Leu389).
The intensity decreases observed upon addition of Cdc42 could be explained by exchange or, alternatively, by increased transverse relaxation due to an increased correlation time. The rate of transverse relaxation, which is simply the loss of relaxation from the transverse plane over time, is directly proportional to the correlation time and so the signal intensity is expected to decrease with slower tumbling. The correlation time of $^{15}$N-HR1$^{TOCA1}$ is expected to increase upon Cdc42 binding, as the resulting complex is 33 kDa compared to the 12 kDa of the free HR1 domain. Moreover, were Cdc42 binding to cause oligomerisation or aggregation, the correlation time would be further increased. In cases where line broadening results from binding-, oligomerisation- or aggregation-induced increases in correlation time, the resulting intensity changes are seen for all residues across the spectrum, sometimes with the exception of the termini or highly mobile regions.

A plot of the intensity ratio for each NH peak between free and bound HR1$^{TOCA1}$ against residue number revealed that line broadening occurred for residues across helix 1, the inter-helical loop region and most of helix 2 (Figure 5.3A). This can be explained by the increased correlation time upon complex formation. However, residues in the centre of the coiled-coil, including residues from both helices and the loop, underwent significantly greater line broadening when compared with the majority of residues, indicating additional exchange contributions. The most affected residues mapped to the central region of the coiled-coil.

This additional line broadening could be caused by the simple intermediate exchange between the free and bound state as was shown in Figure 5.1 and so implicate these broadened residues in the binding interface. However, the line broadening could also be caused by other exchange phenomena. For example, conformational changes may occur before or after binding, there may be multiple binding sites or HR1$^{TOCA1}$ could undergo oligomerisation prior to binding (Williamson 2013). If the exchange contribution was simple intermediate exchange as shown in Figure 5.1, the HSQCs would be expected to relate linearly to one another, whilst if there were additional mechanisms at work the relationship between the HSQCs may be non-linear. For example, a fast-exchanging peak may move first in one direction before moving in another – perhaps indicating binding and rearrangement steps or a second binding event at a distinct site. The existence of multiple binding sites typically results in
significant non-linearity whilst more subtle conformational rearrangements may be more difficult to see, especially if the additional states are sparsely populated.

The fast exchanging peaks in Figure 5.2 appear to move linearly and so there is no clear evidence in these HSQCs of additional mechanisms. Therefore, simple exchange between free and bound states on the intermediate timescale provides the most likely explanation. This is also consistent with the Cdc42-HR1 CIP4 AUC data presented in Chapter 3, in which the s-value of the HR1 domain was seen to increase in the presence of Cdc42 suggesting exchange on the millisecond timescale.

Despite the problem of exchange broadening, the overall CSP was calculated for each residue that was visible in the spectra of both free and bound protein. The Nitrogen dimension was first scaled down to account for the large difference in chemical shift range in the two dimensions and then Pythagorus’ theorem was used to calculate the overall CSP (Figure 5.2B). The CSP for each residue is plotted in Figure 5.3B along with the average CSP. It was not possible to determine CSPs for some overlapped residues (marked with asterisks below the chart) nor for residues that had disappeared. Disappearing residues were assigned a CSP of 0.2 for graphical purposes but this value was not included in the calculation of the mean CSP.

The region with the largest CSPs coincided with the region affected by the largest intensity changes that were described above. Although, there were a number of residues with significant CSPs but less significant changes in intensity, indicating that these residues were indeed in fast exchange. Residues with a CSP greater than the mean or a disappearing peak were considered significantly affected.
Figure 5.3. Quantitative Analysis of the $^{15}$N-HSQCs when Cdc42 was Titrated into $^{15}$N-HR1$^{\text{Toca1}}$. A) A bar chart showing the intensity ratio between free and bound HR1$^{\text{Toca1}}$ against residue number. Ratios could not be calculated for peaks with zero intensity in the bound spectrum and so these peaks were assigned an intensity ratio of 10 for graphical purposes and are coloured orange. The mean intensity ratio is shown with a red line. B) A bar chart showing the CSP for each residue. The secondary structure derived from the NMR structure presented in Chapter 4 is illustrated below the chart. The mean CSP is shown with a red line. The CSPs for the sidechain NH$_2$ groups are also shown. A black * indicates a residue for which no CSP could be calculated due to spectral overlap and a green * indicates a Proline residue, for which there is no amide. Orange bars indicate disappeared peaks.
Typically, ligand binding results in a number of chemical shift changes of more than 0.2 ppm (Williamson 2013) but for the HR1-Cdc42 interaction, the CSPs were all less than 0.2 ppm, with the mean CSP falling below 0.05. The small CSPs can be explained as in α-helices the backbone NH groups are buried inside the helices and involved in hydrogen bonds so are mainly inaccessible. The $^{15}$N-HSQC chemical shift mapping experiments report on changes to these buried amide groups so are expected to be small for α-helical proteins. It is also true that larger chemical shift differences would push exchange in the fast regime towards the intermediate regime and so the residues with the largest CSPs may be expected to be those that are no longer visible. The dataset may, therefore, be missing the larger CSPs.

The sidechain NH$_2$s of Gln361, Asn371, Asn382 and Gln399 are above the mean and so these residues may be involved in the interaction, but their CSPs were also small (< 0.2 ppm). Their small CSPs may be indicative of them being buried within the coiled-coil and so not directly involved in the interaction or, if they are solvent exposed, it may indicate that the interface contacts are mainly hydrophobic rather than polar (i.e. not involving the NH$_2$ in a charge-based interaction). Inspection of the HR1 structure revealed that the sidechains of Asn382 and Gln361 are 89 % and 65 % solvent exposed, respectively, whilst Asn371 and Gln399 are 46 % and 39 % solvent exposed and so classed as buried.

To yield further insight into sidechain involvement in the interaction, $^{13}$C-HSCQ experiments were recorded on $^{15}$N/$^{13}$C-labelled HR1$^{TOCA1}$ in the presence of increasing concentrations of unlabelled Cdc42 (Figure 5.4). As with many of the NH groups, most of the affected CH groups underwent significant line broadening and disappeared such that very few CSPs could be calculated. Some peaks underwent fast exchange and there were also some new peaks indicating some slow exchange. Examples of each are marked in Figure 5.4.
As all but a few of the affected peaks were in intermediate exchange, CSPs did not provide a useful way of assessing sidechain involvement across the spectrum. The intensity changes of the peaks undergoing slow or intermediate exchange could not be quantified as the disappearance occurred at the first titration point (1:0.25), indicating that HR1_TOCA1 was exchanging between free and bound conformations even at low concentrations of Cdc42. None of the new peaks arising from slow exchange were assigned as they only appeared in the final titration point so could not be tracked. Therefore, the involvement of a sidechain was assessed qualitatively and all disappeared or shifted peaks were classed as significantly affected.

Figure 5.4. Mapping the Cdc42-binding site onto HR1_TOCA1 – 13C-HSQCs. A spectrum of free HR1_TOCA1, recorded on 13C/15N-labelled protein at 0.2 mM in NMR Titration Buffer on an Avance AV600 at 25 °C, is shown in black. A spectrum of 0.2 mM 13C/15N-HR1_TOCA1 in the presence of 4-fold molar excess of unlabelled Cdc42 is shown in red. Examples of slow, intermediate and fast exchange are indicated.
For most residues, either all or none of the sidechain CH groups were affected. For four Leucine residues (349, 360, 403 and 407), however, only the Cβ was affected, with all of the other sidechain and backbone peaks unaffected. The effect on the Cβ was therefore considered to be insignificant and the sidechains were considered unlikely to be involved in the interaction. For His339, only the CHε group was affected and so this was also considered to be insignificant.

The graph showing the backbone NH CSPs from Figure 5.3 is repeated in Figure 5.5 but it is coloured differently to indicate the sidechain involvement of each residue. Residues with affected sidechains are indicated with red bars and those with unaffected sidechains with black bars. Some residues could not be assessed due to spectral overlap and these are coloured with grey bars. Residues for which no NH information was available are marked with an asterisk coloured according to the same colour scheme to indicate their sidechain involvement (red = affected sidechain, grey = unknown sidechain).

**Figure 5.5. A Graphical Representation of the $^{15}$N and $^{13}$C Mapping Data.** The NH CSPs are shown as in Figure 5.3 but they are coloured to indicate the disappearance or shifting of corresponding sidechain CH groups. Red bars indicate disappeared or shifted CH groups, grey bars indicate a residue for which no sidechain information was available due to overlap and black bars indicate residues for which the sidechain CH groups were unaffected. A red * indicates a residue with affected CHs but for which no NH information was available, a grey * indicates a residue for which no NH or CH information was available and a black * indicates a residue for which the CH groups were unchanged.
5.2.1.2 Analysis of the Binding Surface

The region most affected by Cdc42 binding was localised to the middle of the TOCA1 HR1 sequence, corresponding to a section including the C-terminal half of helix 1, the inter-helical loop region and a few residues at the N-terminus of helix 2 (Figures 5.3 and 5.5). Other regions were not affected, suggesting that there was no widespread conformational change and the coiled-coil stays intact.

Overall, there were 40 residues of the HR1 domain that had significantly affected backbone, sidechain or backbone and sidechain resonances. These significantly affected residues were filtered based on solvent accessibility using NACCESS (Hubbard and Thornton 1993). The backbone and sidechains of 20 of the affected residues were less than 50% solvent accessible and so classed as buried and unlikely to be involved in the interaction. These are coloured dark blue on the NMR structure of HR1$_{TOCA1}$ shown in Figure 5.6. The residues with no significant changes to backbone or sidechain peaks in the presence of Cdc42 are coloured in cyan.

The other 20 of the significantly affected residues were classed as solvent accessible. These were stratified based on whether their backbone, sidechain or both were affected. Asp365, Asp368, Ala369, Lys372, Asp375, Glu378, Pro381, Met383, Asp385, Ser387, Ser388 and Pro391 had both their backbone and sidechain resonances affected so are considered to be the most likely to represent direct contacts. These are coloured red in Figure 5.6. Residues Gln366, Gln382 and His390 had affected sidechains but their backbone resonances were not affected. These are coloured in yellow in Figure 5.6.

The backbone resonances of Gly384 were affected but, as it has no sidechain, less information was available to define it as a likely direct contact. Lys362, Lys379 and Lys392 also had affected backbone resonances but no information was available on their sidechains due to overlap. Gln361 had CSPs slightly above the mean for its backbone NH and sidechain NH$_2$ peaks but its CH groups were not assignable due to overlap. Gly384, Lys362, Lys379, Lys392 and Gln361 are also coloured yellow. Glu395 was found to be solvent accessible but there was no information available for either its backbone or sidechain resonances due to overlap and so it is coloured grey.
The affected residues mapped to one half of the coiled-coil, as previously seen in the graphical representations of the data in Figure 5.3 and 5.5. The majority of the affected sidechains that are buried according to NACCESS (dark blue in Figure 5.6) mapped towards the centre of the coiled coil and are mostly involved in inter-helical contacts. The exchange broadening or significant CSPs of the backbone or sidechain resonances corresponding to these residues are therefore most likely due to secondary effects caused by localised conformational perturbations. The solvent accessible
residues with both backbone and sidechains affected (red in Figure 5.6) are mainly in helix 1 and the inter-helical loop region but there is one in helix 2.

The interface includes some charged residues, for example Asp368 and Lys372, some hydrophobic residues, for example Ala369 and Met383 and some polar residues, for example Gln382. Interestingly, residues involved in the loop-helix 1 contacts shown in Figure 4.22 of Chapter 4, as well as the neighbouring loop residues, have significantly affected backbone NH and sidechain groups in the presence of Cdc42. These include residues of the loop region (Asn380, Gln382, Met383, Gly384, Asp385, Ser387 and Ser388) and of helix 1 (Val376 and Tyr377). It is also noteworthy that the binding region includes the crucial MGD motif (383-385) discussed previously (section 4.2.6).

The chemical shift mapping data indicate that the G protein binding region of the TOCA1 HR1 domain is broadly similar to that of the CIP4 and PRK1 HR1 domains (Figure 5.7). The interfaces of the PRK1 HR1 domains, however, mainly involve residues in $\alpha$-helical structural elements, whilst the CIP4 and TOCA1 HR1 domains also involve residues of the inter-helical loop region comprising $3_{10}$-helix and unstructured elements. The residue types involved are well conserved between TOCA1 and CIP4, which is unsurprising given their overall sequence identity of 68%. There is also some conservation of interacting residues between the PRK1 HR1 domains. The residues are not, however, well conserved between the TOCA and PRK family HR1 domains.
Figure 5.7. A Comparison of the G protein Binding Sites of the TOCA and PRK Family HR1 Domains. A) Cartoon representations of the TOCA1, CIP4 and PRK1 HR1 domain structures with interacting residues marked in red. For CIP4, amide chemical shift mapping data was available (Kobashigawa et al. 2009), which was filtered for solvent accessibility to derive the list of interacting residues. For comparison, all of the solvent accessible residues of TOCA1 with affected NH peaks were including as interacting residues. For the PRK1 HR1 domains, interacting residues were derived from the available structures. B) A sequence alignment, produced using Expresso (Armougom et al. 2006), is shown with the interacting residues marked on. The secondary structure is indicated.
5.2.2 Mapping the HR1<sup>Toca</sup> Binding Site onto Cdc42

5.2.2.1 Protein Purification

<sup>15</sup>N-labelled and <sup>15</sup>N/<sup>13</sup>C-labelled Cdc42Δ7Q61L were expressed in <i>E. coli</i> BL21 grown in M9 minimal media at 20 °C overnight. The proteins were purified using the GST-tag as described for unlabelled Cdc42 in Chapter 3. 13 mg of <sup>15</sup>N-labelled protein and 12 mg of doubly labelled protein were obtained, each from 3 L of <i>E. coli</i>. The proteins were loaded with GMPPNP and purified by gel filtration as described for unlabelled protein in Chapter 3. 3 mg of <sup>15</sup>N-labelled protein and 6 mg of doubly labelled protein were obtained for NMR experiments.

5.2.2.2 HSQC Titrations

Firstly, <sup>15</sup>N-HSQCs were recorded on <sup>15</sup>N-Cdc42Δ7Q61L·GMPPNP in the presence of increasing concentrations of unlabelled HR1 domain. The majority of the peaks were unchanged but some residues were in fast exchange and many disappeared indicating exchange on the unfavourable millisecond timescale. Examples of peaks exchanging on the fast and intermediate timescales are indicated in Figure 5.8. Some regions of interest including exchanging peaks are shown enlarged with assignments and the intermediate titration points are also shown in one of the enlarged regions. The assignments of free Cdc42 were provided by Helen Mott.
C-HSQCs were then recorded on $^{15}$N/$^{13}$C-Cdc42Δ7Q61L·GMPPNP with unlabelled HR1TOCA1 (Figure 5.9). The $^{13}$C assignments for free Cdc42 were provided by Helen Mott. Again, some peaks were in fast exchange and some disappeared. A methyl-selective SOFAST-HMQC (Schanda et al. 2005) was also recorded to achieve better sensitivity in the methyl group region. This is shown within the $^{13}$C-HSQC in Figure 5.9 with assignments. An intermediate titration point (1:0.5) is also shown.

$^{13}$C-HSQCs were then recorded on $^{15}$N/$^{13}$C-Cdc42Δ7Q61L·GMPPNP with unlabelled HR1TOCA1 (Figure 5.9). The $^{13}$C assignments for free Cdc42 were provided by Helen Mott. Again, some peaks were in fast exchange and some disappeared. A methyl-selective SOFAST-HMQC (Schanda et al. 2005) was also recorded to achieve better sensitivity in the methyl group region. This is shown within the $^{13}$C-HSQC in Figure 5.9 with assignments. An intermediate titration point (1:0.5) is also shown.
Figure 5.9. The $^{13}$C-HSQC and Methyl Selective SOFAST-HMQC of Cdc42 in the Presence of HR1$^{TOCA1}$. $^{13}$C-HSQCs of 0.65 mM $^{13}$C/$^{15}$N-Cdc42Δ7Q61L·GMPPNP, free (black) or in the presence of 2-fold HR1$^{TOCA1}$ (red) in NMR Titration Buffer on an Avance AV600 at 25 °C. The inserts show a methyl-selective SOFAST-HMQC with assignments. The top insert includes an intermediate titration point (Cdc42 1:0.5 HR1).
The overlap in the $^{13}$C-HSQC was such that many of the peaks could not be tracked and so constant time (CT) $^{13}$C-HSQCs were also recorded on free and complexed Cdc42. The CT-HSQC is a variation on the standard HSQC, which includes a constant time evolution period. The $^{13}$C-couplings are refocused during the constant time period, resulting in narrower peaks. Figure 5.10 shows the improvement in the spectrum when the CT-HSQC was used.

![Figure 5.10. The Constant Time $^{13}$C-HSQC. A) The methyl region of the $^{13}$C-HSQC and CT-$^{13}$C-HSQC of $^{13}$C/$^{15}$N-Cdc42Δ7Q61L-GMPPNP are shown side by side. The spectra were recorded on 0.65 mM protein in NMR Titration Buffer on an Avance AV600 at 25 °C.](image)

### 5.2.2.3 Analysis of the Titration Data

The chemical shift mapping data for Cdc42 is summarised in Figure 5.11. The $^{15}$N-HSQCs provided information on the backbone NH groups of 136 of the expected 168 residues of Cdc42 (184 residues minus 15 Prolines and the N-terminus). The NH peaks of 32 residues were not assignable in the HSQC of complexed Cdc42 due to spectral overlap or because they were also not visible in the free spectrum (e.g. most of Switch 2). A CSP could not be calculated for these residues nor could the peak be concluded as disappeared. These, along with the N-terminus and the Proline residues, are marked with black asterisks beneath the graphs.

A CSP could be calculated for 92 peaks but the remaining 44 peaks had disappeared. Disappearing peaks were excluded from the calculation of the mean CSP but later
assigned a CSP of 0.1 (a number greater than the maximum calculated CSP) to illustrate on the bar chart that they are affected (orange bars). The large number of dissapeared peaks and hence the lack of CSPs for many of the affected residues resulted in a very low mean CSP (<0.02). For this reason, the mean plus one standard deviation (0.048) was used as the significance threshold to reduce overestimation of the binding surface. 9 of the 92 residues for which CSPs could be determined were therefore classed as significantly affected along with the 44 that had disappeared.

Data from the $^{13}$C-HSQC, $^{13}$C-CT-HSQC and methyl-selective SOFAST-HMQC experiments provided information on the sidechains of 80 residues out of the expected 175 (184 minus 9 Glycines). There was no sidechain information available for the remaining 95 residues due to spectral overlap of some peaks and to the absence of some peaks from the free spectrum due to conformational exchange. 29 of the 80 residues that were assessable were affected by HR$_1$TOCA$_1$ (marked with green asterisks in Figure 5.11) and the remaining 51 were not affected. CSPs could also be calculated for some of the sidechain NH peaks, as shown in Figure 5.12. Two pairs of NH$_2$ peaks, however, were not traceable due to overlap (marked with *) and 3 pairs had disappeared (orange bars).
Figure 5.11. A Graphical Representation of the Titration Data for Cdc42. The bar chart shows the CSP calculated for the backbone NH of each residue. Residues that are not visible in free Cdc42 or not traceable in the complex spectrum due to spectral overlap were not assigned in the complex spectrum and so their CSPs were unknown. These are marked with a black asterisk below the graph. Residues that disappeared are shown with orange bars. Residues with significantly affected sidechains are marked with a green asterisk above the bars. The significance threshold defined as the mean plus one standard deviation (SD) is shown by a red line. The secondary structure, derived from an NMR structure (Mott and Owen, unpublished), is illustrated below the bar chart.
Overall, 67 out of 184 Cdc42 residues had affected backbones, sidechains or both in the presence of HR1TOCA1. This is significantly more than the TOCA1 residues that were affected by Cdc42 (40 out of 104). Assessment of the solvent accessibility revealed that the majority of the affected residues were buried (48 out of 67) suggesting that their effects may be due to small conformational rearrangements. The affected residues are mapped onto an NMR structure of free Cdc42Δ7Q61L·GMPPNP in Figure 5.13. Residues that were not affected are coloured in cyan, affected residues that are not solvent accessible are coloured in dark blue and affected residues that are solvent accessible are coloured yellow if either their backbone or sidechain resonances are affected or red if both are affected.

Figure 5.12. A Graph of the CSP Data for the Sidechain NH Peaks of Cdc42.
The bar chart shows the CSP calculated for each NH peak. The sidechain NH$_2$s of N39 and Q116 were not traceable due to spectral overlap so are marked with *. Peaks that disappeared are shown as orange bars. The mean plus one standard deviation (SD) across the whole $^{15}$N-HSQC is shown by a red line.

5.2.2.4 Analysis of the Binding Surface
Overall, 67 out of 184 Cdc42 residues had affected backbones, sidechains or both in the presence of HR1TOCA1. This is significantly more than the TOCA1 residues that were affected by Cdc42 (40 out of 104). Assessment of the solvent accessibility revealed that the majority of the affected residues were buried (48 out of 67) suggesting that their effects may be due to small conformational rearrangements. The affected residues are mapped onto an NMR structure of free Cdc42Δ7Q61L·GMPPNP in Figure 5.13. Residues that were not affected are coloured in cyan, affected residues that are not solvent accessible are coloured in dark blue and affected residues that are solvent accessible are coloured yellow if either their backbone or sidechain resonances are affected or red if both are affected.
The majority of the switch regions, circled in Figure 5.13, are not visible in NMR spectra of free Cdc42Δ7Q61L·GMPPNP due to conformational exchange on an intermediate timescale. In other G protein-effector complexes (reviewed in Mott and Owen 2015) these switch regions become fixed down upon complex formation and so become visible in NMR spectra. These invisible residues of the switch regions did not become visible in the presence of HR1TOCA1, which could indicate that they are not involved in the interaction and so do not become fixed. However, some of the residues of the switch regions are visible in free Cdc42 (Ser30 of Switch 1 and Leu67 and Arg68 of Switch 2) and these residues disappear from the NMR spectra in the presence of the HR1 domain, indicating their involvement in the interaction. The data is therefore indicative of the involvement of the switch regions in the interaction and, given the reduction in conformational freedom observed for the switches in other G protein-effector complexes, it is likely that the switches of Cdc42 do become fixed down upon binding to HR1TOCA1.

The invisibility of the switches in the complex is therefore better explained by the lifetime of the complex being in the millisecond range such that exchange is occurring on the intermediate timescale leading to line broadening. The widespread line broadening seen for residues of Cdc42 outside of the switch regions and for residues

Figure 5.13. The HR1TOCA1-binding site on Cdc42. A cartoon representation of the NMR structure of Cdc42Δ7Q61L·GMPPNP (Mott and Owen, unpublished) with unaffected residues in cyan, unknowns in grey and affected residues as indicated. The switch regions are circled.
of TOCA1 is consistent with intermediate exchange between free and bound states rather than localised conformational freedom. Exchange on the millisecond timescale has also been previously indicated by the AUC data shown in Chapter 3.

The lifetime of the complex and so the rate of the exchange can be related to the rates of association ($K_{on}$) and dissociation ($K_{off}$) and so to the $K_d (K_d = K_{off}/K_{on})$. Simply, if $K_{on}$ is much faster than $K_{off}$, the complex will be long lived and so undergo slow exchange. This situation equates to a tight binding complex with a low $K_d$. As the difference between $K_{on}$ and $K_{off}$ becomes smaller, the complex becomes weaker ($K_d$ becomes higher) and so the exchange rate increases. Exchange in the fast regime is typically reached when the $K_d$ is greater than ~3 μM (Williamson 2013). The affinity of the Cdc42-TOCA1 complex (~5 μM) is therefore consistent with intermediate to fast exchange.

5.2.3 Optimising the NMR data

Whilst the mapping data presented here provides some insight into the interaction, detailed structural analysis of the complex remains desirable. However, the exchange broadening observed here is not conducive to structural studies because the majority of the interface region is invisible in the NMR spectra. It was therefore necessary to pursue ways of improving the NMR data, specifically focusing on reducing the signal loss due to exchange broadening, prior to performing the experiments needed to derive intermolecular distances.

5.2.3.1 NMR Experiments at Different Temperatures and Field Strengths

As discussed in section 5.2, it is not the absolute rate of exchange that is important, but the rate of exchange relative to the difference in the chemical shift in the free and bound states. The absolute rate of exchange ($k_{ex}$) can be manipulated by altering the temperature and the difference in chemical shift (or frequency, $\Delta \omega$) can be manipulated by altering the magnetic field strength. An increase in temperature at a given magnetic field would increase $k_{ex}$ relative to $\Delta \omega$. An increase in magnetic field at a given temperature would increase $\Delta \omega$ relative to $k_{ex}$ as was shown in Figure 5.1.
Altering the temperature rather than field strength is favourable for a number of reasons. Firstly, there is a wider range of temperatures available compared with magnetic field strengths. Secondly, there was evidence of peaks undergoing both intermediate and fast exchange in the spectra of Cdc42 and HR1\textsuperscript{TOCA}\textsuperscript{1} but little evidence of slow exchanging peaks, suggesting that reaching the fast regime may be the most viable option. Finally, pursuing the fast regime by reducing the magnetic field strength would decrease the resolution in the spectrum whilst increasing the temperature would result in an increased correlation time leading to a decrease in signal loss due to relaxation and so a better resolved spectrum.

A $^{15}$N-HSQC was recorded on $^{15}$N-HR1\textsuperscript{TOCA}\textsuperscript{1} in complex with Cdc42 at 500 MHz at 35 °C and compared to that at 25 °C (Figure 5.14B and C). The spectrum recorded at 35 °C showed more uniform peak intensities across the spectrum, particularly in the centre of the spectrum. The intensities of the weaker signals were generally increased but there were no more peaks than in the spectrum recorded at 25 °C. Both had 106 of the expected 119 peaks when saturated with Cdc42. To further improve the spectrum, a higher temperature or lower field would be required to push the exchange further into the fast regime. Neither option, however, was practicable. Cdc42 is not stable over a significant time period at >35 °C and the Cdc42-HR1\textsuperscript{TOCA}\textsuperscript{1} complex at 33 kDa is too large to study at <500 MHz.

Another spectrum was therefore recorded at 10 °C (Figure 5.14A) to see if slow exchange could be reached, although this risked losing the fast exchanging peaks into the intermediate regime. Lowering the temperature also reduces the tumbling rate of the protein leading to longer correlation times and greater signal loss as a result of increased relaxation. The uniformity of peak intensities across the spectrum was worse than in the spectra recorded at higher temperatures. The exchange had not fallen into the slow limit and many of the previously fast exchanging peaks were now in the intermediate regime. This spectrum contained 87 peaks out of the 106 peaks that were seen at 25 °C and many of these were very weak.

In further pursuit of the slow exchange regime, spectra were recorded at higher magnetic fields (Figure 5.14D-G). Whilst the higher fields gave better resolution, they did not improve on the exchange problems. The number of peaks was not changed but
the peaks subject to exchange broadening were weaker than in the lower fields. The uniformity of peak intensity across the spectrum was therefore worse at higher fields.

![Figure 5.14. Temperature and Magnetic Field Titrations – HR1 domain. $^{15}$N-HSQC's of 0.7 mM HR1$^{15}$C with 1.2-fold excess of unlabelled Cdc42 in NMR Titration Buffer recorded on A-C) a Bruker DRX500, D-E) an Avance AV600 or F-G) an Avance AV800 at 10, 25 or 35 °C as indicated.]

The $^{15}$N-Cdc42 sample in complex with unlabelled HR1 domain was also analysed at different temperatures as shown in Figure 5.15 and again the spectrum recorded at 500 MHz and 35 °C gave the most uniformity across the spectrum. The circled regions highlight peaks that were particularly improved at 35 °C.
Figure 5.15. Temperature Titrations – Cdc42. $^{15}$N-HSQCs of 0.3 mM $^{15}$N-Cdc42ΔQ61L·GMPPNP in the presence of 2.2-fold HR1 in NMR Titration Buffer recorded on a Bruker DRX500 at A) 10 °C, B) 25 °C or C) 35 °C. Some examples of peaks with increased intensities at higher temperatures are circled.
As the HSQCs were improved at 35 °C, a \(^{15}\)N-NOESY-HSQC was recorded on the
\(^{15}\)N-HR1 domain sample in complex with Cdc42 at 35 °C to determine whether any
NH-NH or NH-CH NOEs were visible for resonances in the interface of the complex.
Strips corresponding to the NH groups of six residues are shown as examples in
Figure 5.16, three are in the interface based on the mapping data described above and
three are not. The three residues that are not within the interface have
indistinguishable NOESY strips in the spectra of free (blue) and bound (red) HR1.
The three residues that are within the interface show some NOEs but generally of
lower intensity and some that were seen in the NOESY of free HR1 are no longer
visible in the NOESY of the Cdc42-bound HR1.

The NH resonance for Gly384 is in fast exchange, as seen in the HSQC shown in
Figure 5.16B, and so the NH strips could be located in the NOESYs of both free and
bound HR1. In the NOESY spectrum of the complex, Gly384 has three weak NOEs:
an \(i \rightarrow i+1\) NOE from the NH of Met383 at 8.25 ppm, an \(i \rightarrow i+3\) NOE from the CH\(\alpha\) of
Pro381 at 4.8 ppm and an \(i \rightarrow i+1\) NOE from the CH\(\beta\) of Met383 at 2.0 ppm. These
NOEs are shifted in the N dimension compared with the NOESY spectrum of free
HR1, as expected, but are also of lower intensity. Furthermore, the intra-residue NOEs
from the CH\(\beta\)s that are visible in the NOESY of free HR1 (labelled in Figure 5.16) are
missing in the NOESY of complexed HR1. The NH resonance of Leu389 is also in
fast exchange but, in this case, all of the NOEs from the free spectrum remain visible
and the decrease in intensity is only slight. The NH resonance for Ile407 was subject
to line broadening in the presence of Cdc42 and the NOEs in the NOESY of
complexed HR1 are also weaker. However, more than half of the NOEs are still
visible, including all of the intra-residue NOEs.

Overall, 65 % of the interface residues had lost most or all of their intra-residue NOEs
in the NOESY of the complex. 71 % had lost all of the inter-residue NOEs to adjacent
and long-range residues and 50 % had no NOEs left at all. The lack of information in
the interface region rendered progression towards structural studies impossible.
Intermolecular NOEs within the interface region are likely to be weaker than the
intramolecular NOEs as the NMR experiments (X-filter experiments) required for
recording intermolecular NOEs are particularly insensitive. The poor quality of the
higher sensitivity NOESY experiments therefore suggests that the X-filter experiments would fail to provide any intermolecular distance information at all.

Figure 5.16. The $^{15}$N-NOESY-HSQC of $^{15}$N-HR1$^{TOCA1}$ in Complex with Cdc42 at 35 °C. A) Selected regions of the $^{15}$N-NOESY-HSQC of free HR1 (blue), recorded on 0.7 mM $^{15}$N-HR1$^{TOCA1}$, and HR1 in the presence of 1.2-fold Cdc42 (red) are shown for six residues. The spectra were recorded in NMR Titration Buffer on a Bruker DRX500 at 35 °C. B) The regions of the $^{15}$N-HSQC are shown for each residue that is shown in A.
5.2.4 Modelling the Complex

As no intermolecular distance restraints could be provided for NMR structure calculations, ambiguous interaction restraints (AIRs) were instead used to produce a model of the complex using HADDOCK (de Vries et al. 2010). HADDOCK uses AIRs to drive docking calculations between protein structures using methods based on Aria scripts. Energy minimization and semi-flexible annealing steps are included following the initial rigid body docking and an ensemble of models is produced. The AIRs used to drive the calculations are derived from experimental knowledge of the residues involved in an interaction, for example from mutagenesis data or from chemical shift mapping. For TOCA1 and Cdc42, the AIRs were derived from the NMR mapping data as described in the following section.

5.2.4.1 Ambiguous Interaction Restraints

AIRs are divided into ‘active’ and ‘passive’ residues. Active residues are those with experimental data indicating their involvement in the interaction. Passive residues are other residues that may be involved but for which the experimental data does not suggest involvement or for which no experimental data is available. For chemical shift mapping data, these are usually defined as the surface neighbours of the active residues (those with an average inter-residue distance smaller than 6.5 Å). The ‘active’ residues are constrained to be part of the interface where possible, incurring a penalty score where they are not. Passive residues can be included in the interface but are not constrained to be and do not incur a penalty if they are not.

For TOCA1, there were 40 residues defined as significantly affected in the chemical shift mapping described above. These were filtered for solvent accessibility leaving 20 residues (Gln361, Lys362, Asp365, Gln366, Asp368, Ala369, Lys372, Asp375, Glu378, Lys379, Pro381, Gln382, Met383, Gly384, Asp385, Ser387, Ser388, His390, Pro391 and Lys392). For Cdc42, 67 residues were affected, which when filtered for solvent accessibility were reduced to 19 residues (Gln2, Lys5, Asp11, Lys16, Thr24, Thr25, Ser30, Met45, Thr52, Arg68, Leu70, Asn92, Glu95, Lys96, Leu160, Gln162, Lys163, Lys166 and Glu178). These are the residues shown in yellow and red in Figures 5.6 and 5.13. These lists were manually curated to produce the final list of active residues for the HADDOCK docking as follows.
Of the 20 affected, solvent exposed residues of TOCA1, most had information available for both their backbone and sidechain resonances. Those with both their backbone and sidechains affected by Cdc42 (red in Figure 5.6) were included as active residues. Mutation of Met383/Gly384/Asp385 has previously been shown to abolish Cdc42 binding (Ho et al. 2004) thus further increasing the confidence in including these 3 residues as active. The sidechains of Gln366, Gln382 and His390 were affected in the presence of Cdc42 but their backbone resonances were not (yellow in Figure 5.6). It was therefore concluded that they were less likely to be directly involved in the interaction and so these residues were excluded from the list. These residues were included as passive by using the surface neighbour definition.

5 of the 20 residues were lacking information for either their backbone or sidechain (Gly384, Lys362, Lys379, Lys392 and Gln361). Gly384 does not have a sidechain so the effect on its backbone resonances were considered enough for its inclusion as an active residue. Given the positions of Lys379 and Lys392 within the most affected region (Figure 5.6), the fact that their sidechains are solvent exposed and the large effect seen on their backbone resonances (Figure 5.5), these were also included as active. Lys362 is further from the main interacting region and the effect on its backbone was only slightly above the significance threshold. This residue was therefore excluded from the list of active residues but was instead included as passive by the surface neighbour definition of passive residues. The backbone NH and sidechain NH$_2$s of Gln361 had small CSPs, slightly above the mean for the data set, but there was no information available on its CH groups. This too was included as passive rather than active. Glu395, for which there was no information available due to overlap, was also included as passive.

Met383 and Gly384 were previously shown to be involved in contacts between the inter-helical loop region and helix 1 (Chapter 4, Figure 4.22). Specifically, they contact Val376, Tyr377 and Asn380, which are classed as buried in the structure. As Met383 and Gly384 were classed as active using the chemical shift mapping data and their contact residues were also affected despite being buried, it was assumed that these contacts would be broken upon Cdc42-binding. Val376, Tyr377 and Asn380 were therefore included as active based on the assumption that they become solvent exposed upon breakage of these intramolecular contacts in the presence of Cdc42.
The final list of active residues for TOCA1 was Asp365, Asp368, Ala369, Lys372, Asp375, Val376, Tyr377, Glu378, Lys379, Asn380, Pro381, Met383, Gly384, Asp385, Ser387, Ser388, Pro391 and Lys392.

For Cdc42, Ser30, Met45 and Lys166 were excluded as only their backbone resonances were affected, with their sidechain resonances remaining unchanged. Glu178 was excluded as only its sidechain was affected with its backbone remaining unchanged. Thr25 was excluded as only its CHβ was affected and Asn92 and Gln162 because only their NH2S were affected whilst the remainder of their sidechain and their backbone resonances were unchanged. The final list of active residues was Gln2, Lys5, Asp11, Lys16, Thr24, Thr52, Arg68, Leu70, Glu95, Lys96, Leu160 and Lys163.

The active residues of Cdc42 and HR1TOCA1 are coloured in red in Figure 5.17. The passive residues were automatically defined in HADDOCK as the surface neighbours of the active residues. The active residues of Cdc42 fall into two lines on one face of the protein, separated by the switches. This suggests that some of them are not direct contacts but instead arise from overestimation of the binding interface due to conformational changes. HADDOCK randomly excludes 50 % of the AIRs in each run of the docking protocol, which accounts for this overestimation of the interface. In theory, the best scoring structures will be those containing only the correct active AIRs. It was decided, therefore, to continue with this active residue list and let HADDOCK filter out those resulting from overestimation of the interface.

The active residues of HR1TOCA1 map to one patch on the coiled-coil that encompasses both faces (Figure 5.17B). Again, only one of the faces is likely to make direct contacts and so some of the active residues are likely to represent an overestimation of the binding interface. The inclusion of AIRs that do not in reality make direct contacts may lead to multiple orientations of the HR1 relative to Cdc42 in the HADDOCK results, including orientations that do not represent the real structure. However, scoring of the structures by HADDOCK and manual analysis of the structures covering different orientations should filter out solutions that are less likely to represent the real structure.
5.2.4.2 Deriving Distance Information from AIRs

The AIRs are defined as ambiguous intermolecular distances with a maximum distance of 3 Å between any atom of an active residue of protein A and any atom of any active or passive residue of protein B. 3 Å is chosen as a compromise between

---

**Figure 5.17. Active AIRs for HADDOCK Docking.** Surface representations of A) Cdc42 and B) HR1^{Toca1} are shown in cyan with active residues coloured red. The switch regions of Cdc42 are in grey. The HR1 domain is shown in two orientations rotated by 180° as indicated.
proton-proton and heavy atom-heavy atom van der Waals distances. The effective distance \(d_{\text{eff}}\) for each AIR is given by:

\[
d_{\text{eff}}^{\text{AB}} = \left( \sum_{m=1}^{N_{\text{atoms}}} \sum_{k=1}^{N_{\text{res}}} \frac{1}{d_{m,n}^{6}} \right)^{-1/6}
\]

where \(N_{\text{atoms}}\) is all of the atoms within a residue, \(N_{\text{res}}\) is all of the active and passive residues in protein B, \(m\) is any atom of each active AIR of protein A, \(n\) is any atom of each AIR of protein B and \(k\) is any active or passive residue of protein B. The distances are therefore calculated based on summing across the whole residue and for each active-active or active-passive possibility in the complex. In this way, the passive residues do not have direct ambiguous interaction distances (AIDs) but they are able to satisfy the direct AIDs of the other protein.

This equation is similar to the equation used to calculate ambiguous distance restraints from NOEs in the structure calculations described in Chapter 4 and shown below:

\[
D \equiv \left( \sum_{a=1}^{N_d} d_a^{-6} \right)^{-1/6}
\]

1/\(r^6\) sum averaging is used to derive distances in both equations. In the NOE equation it is used because distance is inversely proportional to \(\text{NOE}^6\). For the AIRs, it is not used to mimic this equation but rather it represents the attractive part of the Lennard-Jones potential, the mathematical model that approximates an interaction between two atoms. Using the 1/\(r^6\) averaging ensures that the AIRs are satisfied immediately when any two atoms of the two proteins come into contact.

The main difference between the two equations is that, in the AIR equation, the pairs of contacts are allowed to be between any atom of any active residue of one protein and any atom of any active or passive residue on the other. This is in contrast to the equation for ambiguous distance restraints, which instead defines distances between only chemical shift matched atoms as described in Chapter 4. As such the AIRs are more ambiguous. Furthermore, ambiguous NOEs contain real distance information whilst the AIRs are uniformly defined as a maximum distance of 3 Å.
Treating the AIRs in this way allows HADDOCK to sample all of the configurations around the binding site defined by the chemical shift data to find the most favourable pairs of intermolecular contacts. The active residues are constrained within the interface. 50% of the AIRs are randomly discarded in each repeat of the docking protocol to account for the overestimation of binding interfaces by chemical shift mapping, resulting in some of the repeats including only the correct contacts. It is assumed that these correct solutions will score better by a range of HADDOCK criteria, including assessment of the buried surface area and the desolvation, van der Waals, electrostatic and violation energies.

5.2.4.3 Rigid Body Docking, Semi-flexible Annealing and Water Refinement

The docking protocol in HADDOCK is based on Aria scripts interfaced to CNS. Firstly, the two structures are separated by 25 Å and randomly rotated. The AIRs are treated as described above and in any given run a random 50% of the active residues are constrained within the interface and any of the active or passive residues can satisfy the active AIDs.

Four cycles of rigid body orientational optimisation are first completed where each protein is allowed to rotate in turn in order to minimize the intermolecular energy function. Next, there are two cycles of rotational and translational rigid body minimization. Following the rigid body docking stages, the structures are scored based on restraint violation, van der Waals, electrostatic and desolvation energies and buried surface area (BSA) as follows:

\[ \text{Score} = 0.01 \ E_{\text{air}} + 0.01 \ E_{\text{vdW}} + 1.0 \ E_{\text{elec}} + 1.0 \ E_{\text{desolv}} - 0.1 \ BSA \]

The van der Waals score is weighted low prior to the later stages of semi-flexible annealing, during which the torsion angles and packing at the interface are optimised. For Cdc42-HR1\text{TOCA1}, 10000 energy minimized structures were produced in these initial stages of the HADDOCK run and the top 400 structures were carried forward to semi-flexible simulated annealing.

In the first stage of simulated annealing, the two proteins are still treated as rigid bodies and 1000 steps of orientational optimization are carried out between 2000 and 50 Kelvin. In the second stage, the sidechains at the interface are allowed to move in
torsion angle space to optimize packing. The interface residues are defined as the active and passive residues plus two residues sequentially either side. This stage involves 4000 steps from 2000 to 50 K. In the third stage, the interface backbone and sidechains are allowed to move over 1000 steps from 500 to 50 K. The resulting structures undergo a further round of energy minimization and are scored according to the following equation:

\[
\text{Score} = 0.1 \ E_{\text{air}} + 1.0 \ E_{\text{vdW}} + 1.0 \ E_{\text{elec}} + 1.0 \ E_{\text{desolv}} - 0.1 \ BSA
\]

The van der Waals energy is now weighted 100-fold higher and the restraint violation energies 10-fold higher. As such, the scoring criteria becomes more stringent to select optimal contacts and packing at the interface. The optimized structures are then water refined and clustered into groups based on common contacts or RMSDs.

The 400 water-refined structures of HR1-Cdc42 were clustered according to the fraction of common contacts. The top 10 clusters contained 94% of the water refined structures. These clusters were then subjected to a number of analyses allowing for a detailed assessment of the models based on all of the individual HADDOCK criteria as well as the combined HADDOCK score calculated according to the following equation:

\[
\text{Score} = 0.1 \ E_{\text{air}} + 1.0 \ E_{\text{vdW}} + 0.2 \ E_{\text{elec}} + 1.0 \ E_{\text{desolv}}
\]

### 5.2.4.4 Analysis of the HADDOCK Results

The top 4 clusters in terms of the HADDOCK score were numbers 9, 3, 7 and 2 (Figure 5.18A), but this score does not consider a number of important factors. For example, cluster 9 has one of the higher backbone RMSDs from the overall lowest energy structure. It also has the highest backbone RMSD from its best scoring structure meaning that the cluster as a whole does not represent its best scoring structure very well compared to other clusters. Furthermore, it scored relatively low on the fraction of common contacts (FCC). Table 5.1 summarises some of the important information for the top scoring clusters.

Cluster 7 claimed the titles of lowest backbone RMSD from the lowest energy structure, lowest backbone RMSD from its best scoring structure and the highest FCC. It also scored better than cluster 9 in the restraint violation energy. In fact, taking
backbone RMSD or FCC as the assessment criteria puts clusters 7, 10, 5 and 8 (Figure 5.18B) in the top 4 positions with clusters 9, 3 and 2 scoring close to bottom. Cluster 7 is the only cluster that resides in the top 4 by all criteria. The orientation of HR₁^{TOCA1} with respect to Cdc42 about the z-axis (in the x-y plane of Figure 5.18) appears relatively similar between the top backbone RMSD and FCC scorers (Figure 5.18B) but is far more variable among the top HADDOCK scorers (Figure 5.18A).

Figure 5.18. The Top 8 Clusters from the HADDOCK Modelling. Cartoon representations of the lowest energy structure in A) each of the four top HADDOCK scoring clusters and B) each of the four clusters with the smallest backbone RMSD from the overall lowest energy structure. The models were aligned on Cdc42 only. A set of axes is defined to aid in describing the rotation of HR₁^{TOCA1} with respect to Cdc42.
Comparing the orientations of the HR1 domain relative to Cdc42 about the x-axis (in the y-z plane of Figure 5.18) we see an interesting difference between the top 4 HADDOCK scorers vs the top 4 RMSD scorers. The HR1 domain in 3 out of 4 of the best HADDOCK scorers (clusters 2, 3 and 9) is in a ‘loop-away’ orientation in which the inter-helical loop (containing the critical MGD motif) points away from Cdc42 (Figure 5.19A). The HR1 domain in the best RMSD scorers (clusters 7, 10, 5 and 8) is rotated 180° along the x-axis and so is in a ‘loop-towards’ orientation (Figure 5.19B).
Multiple orientations along this axis were expected as the active AIRs mapped to both faces of the coiled-coil as shown earlier in Figure 5.17. Clusters 7, 10, 5 and 8 were the only clusters to show the HR1 domain in the ‘loop-towards’ orientation (representing 7.5 % of the clustered structures). Despite the low occupancy of these clusters, this ‘loop-towards’ orientation most closely resembles the structures of the PRK1 HR1 domains in complex with their G protein partners, an example of which is shown in Figure 5.20.

**Figure 5.19. The ‘Loop-away’ and ‘Loop-towards’ Orientations.** Cartoon representations of the lowest energy structure from A) the best HADDOCK scoring cluster (9) and B) the best cluster in terms of backbone RMSD from the overall lowest energy structure (7). The loop residues are shown as sticks in both cases to illustrate the loop-away and loop-towards orientations. A set of axes is defined to aid in the description of the two orientations arising from rotation of the HR1 relative to Cdc42 in the y-z plane. There is also a slight rotation of the HR1 in the x-y plane between the two clusters.
To evaluate the ‘loop-towards’ compared to the ‘loop-away’ orientations, a close investigation of the individual contacts present in each cluster was necessary. For the clusters in the ‘loop-away’ conformation (92.5 % of the structures), the representation of active residues was low (Table 5.1). In the most populous of the top 4 HADDOCK scoring clusters (cluster 2), just one of the Cdc42 active residues is found in a potentially favourable contact. In the second most populous (cluster 3), 12 % are satisfied and in the top scorer (cluster 9) we find just 17 % of the Cdc42 active residues are in favourable positions, e.g. Arg68\text{Cdc42} is in a salt bridge with Asp368\text{TOCA1}. The rest are either simply not in close proximity with HR1 residues or are in unfavourable contacts, for example Leu70\text{Cdc42} is poking into a charged area characterised by Asp368\text{TOCA1} and Lys372\text{TOCA1} of HR1\text{TOCA1}.

In the ‘loop-towards’ cluster 10, Leu70\text{Cdc42} finds itself in a more hydrophobic environment and Arg68\text{Cdc42} makes an alternative salt bridge with Glu395\text{TOCA1} but other residues find themselves in unfavourable positions. For example Lys5\text{Cdc42} is in a rather hydrophobic environment. In total, 25 % of the Cdc42 active residues are satisfied, more so than in any of the ‘loop-away’ clusters but still rather few. In cluster 7, the only cluster to make the top 4 by all criteria, 42 % of the Cdc42 active residues are satisfied. For example, Leu70\text{Cdc42} is packed against the hydrophobic core of the

---

**Figure 5.20. The Best HADDOCK Model of the Cdc42-HR1\text{TOCA1} Complex Compared with the Structure of the Rac1-HR1\text{aPRK1} Complex.** Cartoon representations of A) the lowest energy HADDOCK model from cluster 7 and B) the NMR structure of Rac1 in complex with PRK1 HR1b. The HR1 domains in each are in the ‘loop-towards’ orientation.
HR1 helices. Furthermore, Arg68$^{\text{Cdc42}}$ is in a salt bridge with Glu395$^{\text{TOCA1}}$, Lys5$^{\text{Cdc42}}$ makes electrostatic contacts with Asp365$^{\text{TOCA1}}$ and Asp368$^{\text{TOCA1}}$, Glu2$^{\text{Cdc42}}$ inserts between Asn380$^{\text{TOCA1}}$ and Val376$^{\text{TOCA1}}$ and Thr52$^{\text{Cdc42}}$ contacts Lys372$^{\text{TOCA1}}$. These contacts in the 4 lowest energy structures of cluster 7 are shown in Figure 5.21.

![Figure 5.21. Contacts in Cluster 7.](image)

The residues of Cdc42 involved in these contacts correspond to the line of active residues along the bottom in Figure 5.17A. The remaining 58 % of the Cdc42 active residues (Asp11, Lys16, Thr24, Glu95, Lys96, Leu160 and Lys163) that were not satisfied correspond to the top line. These residues are far from the HR1 domain in all of the HADDOCK clusters, including cluster 7 (Figure 5.22). It is therefore likely that they represent an overestimation of the binding interface due to conformational changes. The residues are behind the two switches, which is consistent with the HR1$^{\text{TOCA1}}$-induced effects on their chemical shifts being caused by conformational change. As these residues were not involved in contacts in any of the HADDOCK clusters, it appears that the exclusion of 50 % of the AIRs in each HADDOCK run has successfully eliminated any structures that were based on these incorrect AIRs.
In addition to the favourable positions of the Cdc42 active residues in cluster 7, 76 % of the HR1 domain active residues are involved in hydrogen bonds, salt bridges or hydrophobic packing and so are satisfied. Importantly, the inter-helical loop region of HR1\textsuperscript{TOCA1} was penetrated by residues of Cdc42. The remaining 24 % of the active residues are generally facing away from Cdc42 and their CSPs could be explained by secondary effects.

Conversely, in the case of the ‘loop-away’ clusters 9, 3 and 2, just 24, 20 and 18 % respectively of the HR1 active residues can be described as satisfied. Many of them are far away from, and pointing away from, Cdc42 and a significant percentage were found in unfavourable contacts. For example, in cluster 2, 12 % are in unfavourable positions such as Glu378\textsuperscript{TOCA1} sitting in a hydrophobic pocket surrounded by Gly54\textsuperscript{Cdc42}, Leu55\textsuperscript{Cdc42}, Phe56\textsuperscript{Cdc42} and Ala59\textsuperscript{Cdc42}. Importantly, the loop region remains in its closed conformation pointing away from Cdc42 (Figure 5.19). None of the inter-helical loop residues, including the critical Met383/Gly384/Asp385 motif, are in close proximity to Cdc42.
To conclude, cluster 7 has the third best HADDOCK score, is the closest to the lowest energy structure, has the best FCC, best represents the experimental data (Table 5.1) and the orientation of the HR1 domain with respect to Cdc42 is similar to that of the PK1 HR1 domains with their G proteins (Figure 5.20). Its outstanding representation of active residues along with its low RMSD from the lowest energy structure and similarity to the other G protein-HR1 complexes suggests that Cluster 7 is a satisfactory model of the complex.

5.2.5 Analysis of the Model: Sequence Specific Contacts

A comparison of the Cdc42-HR1^{TOCA1}, RhoA-HR1^{PRK1} and Rac1-HR1^{PRK1} contacts is shown in Figure 5.23. The residues of RhoA and Rac1 that contact HR1a and HR1b of PRK1 are highlighted in cyan. The residues of Cdc42 that were defined as active based on chemical shift mapping are highlighted as follows; cyan if a conserved contact in RhoA and Rac1 and yellow if not. Residues of Cdc42 that correspond to contacts in RhoA and Rac1 but that were not visible in the chemical shift mapping experiments are coloured grey and Cdc42 contacts derived from the model are boxed.
Two of the residues that are conserved between Rac1, RhoA and Cdc42 appear to be contact residues in all three G protein-HR1 domain interactions: Lys5 and Leu70. Lys5\textsuperscript{Rac1} is in a salt bridge with Asp159\textsuperscript{PRK1} in the Rac1-PRK1 HR1b structure and, similarly, Lys5\textsuperscript{Cdc42} contacts Asp365\textsuperscript{TOCA1} and Asp368\textsuperscript{TOCA1} in the HADDOCK model of Cdc42-HR1\textsuperscript{TOCA1} (Figure 5.24A). Leu70\textsuperscript{Rac1} is near Leu164/Thr165/Ala166\textsuperscript{PRK1} in the Rac1-HR1b structure and Leu70\textsuperscript{Cdc42} contacts Leu370\textsuperscript{TOCA1}, which is within a similarly hydrophobic area comprising Leu370/Ser388/Leu389\textsuperscript{TOCA1}. Leu70\textsuperscript{Cdc42} is also involved in hydrophobic packing in the CRIB effectors, for example it contacts residues of the C-terminal extension to the core GBD of WASP (the β-hairpin and α-helix).

One Rac1/RhoA contact residue (Leu67\textsuperscript{Rac1}/Leu69\textsuperscript{RhoA}) is conserved in Cdc42 but was not defined as active based on chemical shift mapping. Leu67\textsuperscript{Rac1}, which is within Switch 2, pokes into a hydrophobic pocket in HR1b\textsuperscript{PRK1} whilst the sidechain of Leu67\textsuperscript{Cdc42} is buried, pointing away from TOCA1 (Figure 5.24B). Hydrophobic contacts have been identified as key contributors to high affinities and specificities (Owen et al. 2000) and so this missing contact between Cdc42 and HR1\textsuperscript{TOCA1} could be important in terms of affinity and specificity.
Some of the residues that are conserved between the three G proteins are defined as active in the Cdc42-TOCA1 interaction but their equivalents are not defined as contacts in RhoA or Rac1. There are also residues defined as active in Cdc42 that are not conserved in Rac1 or RhoA at all. The majority of these non-conserved contacts (yellow in Figure 5.23) can be attributed to secondary, conformational effects: Asp$^{11}_{\text{Cdc42}}$, Lys$^{16}_{\text{Cdc42}}$, Thr$^{24}_{\text{Cdc42}}$, Glu$^{95}_{\text{Cdc42}}$, Lys$^{96}_{\text{Cdc42}}$, Leu$^{160}_{\text{Cdc42}}$ and Lys$^{163}_{\text{Cdc42}}$ pack behind the two switch regions and are far from the HR1 domain in

Figure 5.24. Conservation of Contacts in Rac1, RhoA and Cdc42. Cartoon representations of Rac1-HR1b$^{\text{PRK1}}$ and Cdc42-HR1$^{\text{TOCA1}}$. Rac1 is shown in green, the PRK1 HR1b domain in blue, Cdc42 in cyan and the HR1$^{\text{TOCA1}}$ in purple. Residues of interest are shown as sticks. A) Lys5 is shown in red, contacting Aspartate residues of either PRK1 or TOCA1. B) Leu$^{67}_{\text{Rac1}}$ is shown in red poking into a hydrophobic pocket of PRK1 HR1b and Leu$^{67}_{\text{Cdc42}}$, also in red, is buried and faces away from TOCA1.
the model, as was shown in Figure 5.22. Furthermore, Lys16\textsuperscript{Cdc42} and its equivalents in other G proteins, contacts the nucleotide so is not free for effector binding. The effects on these residues are therefore indicative of the switch regions being involved in the interaction but these residues are unlikely to be making direct contacts.

The remaining active residues of Cdc42 that do not correspond to contacts in Rac1 or RhoA are defined as contacts in the Cdc42-HR1\textsuperscript{TOCA1} model: Gln2\textsuperscript{Cdc42} contacts Asn380\textsuperscript{TOCA1}, Thr52\textsuperscript{Cdc42} contacts Lys372\textsuperscript{TOCA1} and Arg68\textsuperscript{Cdc42} of Switch 2 contacts Glu395\textsuperscript{TOCA1}, as was shown in Figure 5.21. Gln2\textsuperscript{Cdc42} and Thr52\textsuperscript{Cdc42} have also been identified as contact residues in the Cdc42-ACK complex (Mott et al. 1999; Elliot-Smith et al. 2005). The equivalent of Thr52\textsuperscript{Cdc42} in Rac1 is Asn52\textsuperscript{Rac1} and Asn52Thr is one of a combination of seven mutations found to confer ACK-binding on Rac1 (Elliot-Smith et al. 2005). This may therefore represent a specific Cdc42-effector contact residue. Moreover, the position equivalent to Lys372\textsuperscript{TOCA1} in PRK1 is Glu58\textsuperscript{HR1a} (opposite charge to Lys372\textsuperscript{TOCA1}) or Gln151\textsuperscript{HR1b} and so Thr52\textsuperscript{Cdc42}-Lys372\textsuperscript{TOCA1} may represent a specific Cdc42-HR1\textsuperscript{TOCA1} contact. The Arg68\textsuperscript{Cdc42}-Glu395\textsuperscript{TOCA1} salt bridge may also represent a specific Cdc42-HR1\textsuperscript{TOCA1} contact as the Rac1/RhoA equivalents of Arg68 point away from their HR1 domains. However, the importance of Glu395\textsuperscript{TOCA1} in the Cdc42-TOCA1 interaction remains uncertain as there was no information available from chemical shift mapping on this residue.

It is also necessary to discuss residues that are conserved but that were not visible in the spectra of free Cdc42 or in the titration experiments (grey in Figure 5.23). Many of these residues correspond to contact residues of RhoA and Rac1 and the majority of them are in or adjacent to the switch regions. The involvement of the switch regions of Cdc42 in TOCA1 binding has been suggested by their proximity to HR1\textsuperscript{TOCA1} in the Cdc42-HR1\textsuperscript{TOCA1} model (Figures 5.22) and by the chemical shift perturbations of nearby residues. Furthermore, residues of Switch 2 are defined as contacts in the model: Glu62\textsuperscript{Cdc42} contacts Lys362\textsuperscript{TOCA1}, Asp63\textsuperscript{Cdc42} contacts Gln366\textsuperscript{TOCA1} and Arg66\textsuperscript{Cdc42} is close to Glu395\textsuperscript{TOCA1} (Figure 5.25A).

Phe56\textsuperscript{Cdc42}, which lies just upstream of Switch 2, was also invisible but it is important in WASP and ACK binding (Abdul-Manan et al. 1999; Wang et al. 2006; Elliot-Smith et al. 2007). It is thought, however, to pack behind Switch 1 when Cdc42 interacts.
with ACK, maintaining the switch in a binding-competent orientation (Elliot-Smith et al. 2007), rather than making direct contacts. In the Cdc42-TOCA1 model, this residue is close to the HR1 domain but falls near charged and polar residues, e.g. Asp365\textsuperscript{TOCA1}. It is positioned behind the switches, pointing towards Switch 2 (Figure 5.25B), and so it may be indirectly involved in the interaction by positioning of the switches as seen in Cdc42-ACK. Pro73\textsuperscript{Cdc42}, which lies just downstream of Switch 2, was not visible in the chemical shift mapping and it does not correspond to a contact residue in Rac1 or RhoA but it contacts Ala383\textsuperscript{TOCA1} and Gly384\textsuperscript{TOCA1} in the Cdc42-TOCA1 model, contributing to the disruption in the loop region.

Figure 5.25. The Position of Glu62, Asp63, Arg66 and Phe56 in the Cdc42-HR1 Model. The four lowest energy models of the Cdc42-HR1 complex from cluster 7 are shown overlaid. Cdc42 is shown in cyan and TOCA1 in purple. A) Glu62\textsuperscript{Cdc42}, Asp63\textsuperscript{Cdc42} and Arg66\textsuperscript{Cdc42} are shown as sticks in red and their contacts in TOCA1 are shown as sticks in purple. B) Phe56\textsuperscript{Cdc42} is shown with sticks and coloured red. The switch regions are coloured grey and labelled.
5.3 A Possible Cdc42/N-WASP/TOCA1 Ternary Complex

As TOCA1 and N-WASP both bind to Cdc42 and to each other, it has been suggested that they could bind Cdc42 simultaneously (Ho et al. 2004), which would provide the first example of a single G protein simultaneously binding two effectors. Studies in mammalian cells have indicated that a Cdc42/N-WASP/TOCA1 complex exists (Bu et al. 2010), since FRET was observed between RFP-TOCA1 and GFP-N-WASP and the efficiency was decreased when an N-WASP mutant was used that no longer binds Cdc42. Whether the two effectors were simultaneously in direct contact with Cdc42 could, however, not be concluded as the FRET could also be explained by the TOCA1-N-WASP interaction bringing both effectors into close proximity with Cdc42 while only one was in direct contact.

An overlay of the HADDOCK model of the Cdc42-HR1TOCA1 complex and the structure of Cdc42 in complex with the G protein binding domain (GBD) of the WASP (Abdul-Manan et al. 1999) (PDB code: 1CEE), showed that the HR1TOCA1 and GBDN-WASP binding sites only partly overlap and, therefore, a ternary complex remained possible (Figure 5.26). The basic region N-terminal to the core GBD that has previously been implicated in an electrostatic steering mechanism of binding (Hemsath et al. 2005) would not be sterically hindered by the TOCA1 HR1 domain. The core GBD may also be able to bind in the presence of TOCA1. However, the region C-terminal to the core GBD that is required for high affinity binding of WASP (Rudolph et al. 1998) would sterically clash with the TOCA1 HR1 domain.
Some of the Cdc42 residues identified as important WASP contacts are not found to be contacts in the TOCA1 interaction, consistent with the possibility of a ternary complex. For example, Tyr$^{40}_{\text{Cdc42}}$ forms part of a hydrophobic pocket into which Phe$^{244}_{\text{WASP}}$ inserts and mutation of this Tyrosine decreases the affinity of the Cdc42-WASP interaction (Lamarche et al. 1996) but this residue is not identified as a contact in any of the Rho-HR1 interactions described here (Figure 5.23). Furthermore, Leu$^{67}_{\text{Cdc42/Rac1}}$ is involved in hydrophobic packing with the C-terminal extension of the WASP GBD (Abdul-Manan et al. 1999) but it does not appear to contact HR1$^{\text{TOCA1}}$. Conversely, Leu$^{70}_{\text{Cdc42}}$ has been identified as a contact residue in both the Cdc42-TOCA1 and Cdc42-WASP interactions. The important contacts made by Leu$^{70}_{\text{Cdc42}}$ involve residues of the C-terminal extension, which is consistent with the steric clash of this region with the HR1 domain.

5.3.1 Purification of the N-WASP GBD

In order to investigate whether GBD$^{\text{N-WASP}}$ and HR1$^{\text{TOCA1}}$ bind Cdc42 simultaneously or competitively, an N-WASP GBD construct was produced for use in SPA and NMR.
experiments. The GBD was defined based on the WASP GBD used in previous structural studies (Abdul-Manan et al. 1999) and cloned into pGEX-6P-1 as described in Chapter 2. The protein was then expressed as a GST-fusion in *E. coli* BL21 grown at 20 °C overnight and purified by affinity purification via the GST-tag (Figure 5.27A). The protein was cleaved from the GST-tag (Figure 27B) and further purified by gel filtration (Figure 5.27C and D). 35 mg of protein was obtained from 3 L of *E. coli*, which was concentrated to 6 mg/mL and frozen at -80 °C.

**Figure 5.27. Purification of the N-WASP GBD.** A) A 15 % Laemmli gel showing samples taken from the supernatant (SN), each wash of the glutathione agarose beads and a 50 % bead slurry. B) An 18 % gel showing samples taken from the post-cleavage washes and beads. C) An A_{280} trace from size exclusion chromatography using a Superdex 16/60 S30 run in 50 mM sodium phosphate pH 5.5, 25 mM NaCl, 5 mM MgCl₂, 5 mM DTT. D) An 18 % gel showing samples taken from each 2 mL fraction across Peak 2.

Competition SPAs were carried out to determine the affinity of GBD^{N-WASP} for Cdc42 and to confirm that it was functional prior to NMR experiments. The N-WASP GBD was competed with GST-ACK and GST-WASP GBDs for Cdc42-binding, giving a *K_d* of 37 nM against ACK and 16 nM against WASP (Figure 5.28). This is in good
agreement with the previously reported $K_d$ of 20-30 nM (Abdul-Manan et al. 1999) and the interaction is more than 100x higher affinity than the Cdc42-HR1$^{TOCA1}$ interaction ($TOCA1\:K_d \approx 5 \, \mu M$).

![Figure 5.28](image)

**Figure 5.28. Measuring the Affinity of the Cdc42-GBD$^{N-WASP}$ Interaction.** Competition SPA experiments were carried out with the indicated concentrations of GBD$^{N-WASP}$ titrated into 30 nM GST-ACK or GST-WASP GBD and 30 nM Cdc42Δ7Q61L·[3H]GTP. The $K_d$ values for each experiment, derived from fitting to a competition binding isotherm, are shown on the graphs along with the curve-fitting errors.

Competition SPAs between immobilised HR1$^{TOCA1}$ and free GBD$^{N-WASP}$, and *vice versa*, were desired to provide insight into whether their binding to Cdc42 is competitive. However, competition SPAs with GBD$^{N-WASP}$ against His-tagged HR1$^{TOCA1}$ were not successful due to the low affinity of the HR1-Cdc42 interaction. The low affinity meant that micromolar concentrations of HR1$^{TOCA1}$ would be needed to achieve stable binding prior to titration of the N-WASP GBD. The optimal concentration for an immobilised effector in SPAs is typically 10-50 nM and micromolar concentrations result in artifacts caused by saturation of the SPA beads and non-specific binding. The reciprocal experiment with immobilised GST-tagged GBD$^{N-WASP}$ was also not possible as the GST-tag was removed by contaminating proteases during purification and so no fusion protein was obtained for the experiment. Instead, NMR was used to investigate the formation of a possible ternary complex.
5.3.2 Competitive NMR Titrations

Unlabelled GBD\textsuperscript{\textit{N-WASP}} was titrated into \textsuperscript{15}N-Cdc42Δ7Q61L·GMPPNP and the backbone NH groups monitored using HSQCs (Figure 5.29A). The changes were different to those seen in the Cdc42-HR1 titrations (Figure 5.29B). Unlabelled HR1\textsuperscript{TOCA1} was then added to the Cdc42-N-WASP complex but no changes were seen, suggesting that HR1\textsuperscript{TOCA1} did not bind Cdc42 in the presence of GBD\textsuperscript{N-WASP} (Figure 5.29C). No changes were seen even in the presence of a 5-fold excess of HR1\textsuperscript{TOCA1}, indicating that the GBD\textsuperscript{N-WASP} was not displaced by TOCA1. These experiments were recorded at sufficiently high protein concentrations (145 µM Cdc42, 145 µM N-WASP GBD, 725 µM TOCA1 HR1 domain) to be far in excess of the \(K_d\)s of the individual interactions (TOCA1 \(K_d\approx 5\) µM, N-WASP \(K_d = 37\) nM).

**Figure 5.29. NMR Titrations with Cdc42, HR1\textsuperscript{TOCA1} and GBD\textsuperscript{N-WASP}**. A) Selected regions of the \textsuperscript{15}N-HSQC of 145 µM Cdc42Δ7Q61L·GMPPNP are shown with free Cdc42 (black) and with equimolar GBD\textsuperscript{N-WASP} (red). B) The same region is shown for free Cdc42 (black) and with 2-fold molar excess of HR1\textsuperscript{TOCA1} (blue). C) The same region plus another region of the HSQCs are shown for free Cdc42 (black), with equimolar GBD\textsuperscript{N-WASP} (red) and with equimolar GBD\textsuperscript{N-WASP} and 5-fold molar excess HR1\textsuperscript{TOCA1} (green).
\(^{15}\text{N}-\text{HR1}^{\text{TOCA1}}\) was then monitored in the presence of unlabelled Cdc42Δ7Q61L·GMPPNP (1:1) before and after the addition of 0.25 and 1.0 molar equivalents of unlabelled GBD\(^{\text{N-WASP}}\). Figure 5.30A shows the spectra of free and Cdc42-bound \(^{15}\text{N}-\text{HR1}^{\text{TOCA1}}\). Figure 5.30B shows the same spectrum along with spectra recorded with 0.25 and 1.0 molar equivalents of GBD\(^{\text{N-WASP}}\). The addition of 0.25 molar equivalents resulted in a spectrum that was somewhere between the spectra of free and complexed HR1\(^{\text{TOCA1}}\) and the spectrum when GBD\(^{\text{N-WASP}}\) and HR1\(^{\text{TOCA1}}\) were equimolar was identical to that of the free HR1 domain. This indicated displacement of Cdc42 from HR1\(^{\text{TOCA1}}\) in the presence of GBD\(^{\text{N-WASP}}\). Again, the experiments were recorded on protein samples far in excess of the individual \(K_d\)s (600 µM of each protein).

**Figure 5.30. NMR Titrations with Cdc42, HR1\(^{\text{TOCA1}}\) and GBD\(^{\text{N-WASP}}\).** A) A selected region of the \(^{15}\text{N}-\text{HSQC}\) of 600 µM HR1\(^{\text{TOCA1}}\) (black) shown overlaid with the spectrum of HR1 in complex with unlabelled Cdc42Δ7Q61L·GMPPNP (blue). B) The same region with free HR1 (black), HR1 in complex with Cdc42 (blue) and with 0.25 (green) and 1.0 (red) molar equivalents of GBD\(^{\text{N-WASP}}\).
These data indicate that the HR1 domain is displaced from Cdc42 by N-WASP and that a ternary complex comprising HR1\textsubscript{TOCA1}, GBD\textsuperscript{N-WASP} and Cdc42 is not formed. Taken together, the data in Figures 5.29 and 5.30 indicate unidirectional competition for Cdc42 binding, in which GBD\textsuperscript{N-WASP} displaces HR1\textsubscript{TOCA1} but not \textit{vice versa}. Endogenous N-WASP is present at approximately 100 nM in \textit{Xenopus} extracts, while TOCA1 is present at 10-fold lower concentration than N-WASP (Ho et al. 2004) and so the 5-fold excess of HR1\textsubscript{TOCA1} used in the NMR experiments (i.e. 5:1) far exceeds the \textit{in vivo} TOCA1:N-WASP ratio (0.1:1). In a physiological setting, therefore, GBD\textsuperscript{N-WASP} would be expected to outcompete HR1\textsubscript{TOCA1} for Cdc42-binding, as seen in the NMR experiments.

5.3.3 Pyrene Actin Assays

To extend these studies to a more complex system and to assess the ability of HR1\textsubscript{TOCA1} to compete with full length N-WASP rather than the N-WASP GBD alone, pyrene actin assays were employed. In these assays, described in detail elsewhere (Walrant et al. 2015), \textit{Xenopus} egg extract is first supplemented with pyrene-labelled monomeric actin. The actin remains depolymerised in these conditions but the addition of liposomes containing signalling lipids such as PI(4,5)P\textsubscript{2} causes it to polymerise. Pyrene actin fluoresces as it polymerises and thus the rate and extent of polymerisation can be measured by fluorescence ($\lambda_{\text{ex}} = 365$, $\lambda_{\text{em}} = 407$). To examine the effect of HR1\textsubscript{TOCA1} and GBD\textsuperscript{N-WASP} on actin polymerisation, exogenous HR1\textsubscript{TOCA1} or GBD\textsuperscript{N-WASP} was added to the \textit{Xenopus} lysates prior to the addition of liposomes. In all cases, actin polymerisation was initiated by the addition of PI(4,5)P\textsubscript{2}-containing liposomes, which has previously been shown to depend on TOCA1 and N-WASP (Gallop et al. 2013).

Addition of GBD\textsuperscript{N-WASP} significantly inhibited the polymerisation of actin at concentrations as low as 100 nM and completely abolished polymerisation at higher concentrations (Figure 5.31). The exogenous N-WASP GBD presumably acts as a dominant negative by sequestering endogenous Cdc42 and preventing endogenous full-length N-WASP from binding and becoming activated. The estimated concentration of N-WASP in \textit{Xenopus} extracts is 100 nM (Ho et al. 2004) and, consistent with this, the effect on actin polymerisation was far more pronounced at
concentrations above 100 nM. At 100 nM, the endogenous N-WASP could still effectively compete with the isolated GBD. TOCA1 is present at a 10-fold lower concentration than N-WASP (Ho et al. 2004) and yet the addition of the TOCA1 HR1 domain to concentrations as high as 100 µM had no significant effect on the rate of actin polymerisation or maximum fluorescence. This concentration is in 20-fold excess of the affinity of full length TOCA1 or the isolated HR1 domain and so it would be expected to compete convincingly with endogenous TOCA1 for Cdc42.

This PI(4,5)P₂ induced, and Cdc42-dependent, actin polymerisation pathway has previously been shown to depend on TOCA1 and so redundancy in the pathway does not provide an explanation for the lack of inhibition by exogenous HR1TOCA1 (Gallop et al. 2013). The data is therefore consistent with displacement of the TOCA1 HR1 domain by N-WASP. Given the conclusions from the NMR data, it is expected that endogenous N-WASP would outcompete exogenous HR1TOCA1 for Cdc42-binding and therefore allow N-WASP to become activated and actin polymerisation to proceed despite the presence of excess HR1TOCA1.

![Figure 5.31. Pyrene Actin Assays. A-B) Fluorescence curves showing actin polymerisation in the presence of GBDN-WASP or HR1TOCA1, as indicated. C-D) Maximal rates of actin polymerisation were derived from the linear region of the curves and are represented in bar charts. The number of repeats for each concentration are shown and the error bars show the standard error of the mean.](image-url)
5.4 Conclusions and Perspective

In Chapter 3, we saw that the TOCA1 HR1 domain alone is sufficient for Cdc42-binding \textit{in vitro} and yet the affinity of the TOCA1 HR1 domain for Cdc42 is remarkably low ($K_d \approx 5 \, \mu\text{M}$). This is over 100x lower than that of the N-WASP GBD ($K_d = 37 \, \text{nM}$) and considerably lower than other known G protein-HR1 domain interactions. Despite the remarkably different affinities of the G protein-HR1 interactions, we see in Chapter 4 that the HR1 domains are generally structurally similar. More subtle variations may, however, provide insight into the molecular basis for the differential affinities between the TOCA and PRK1 families.

The best HADDOCK model, based on a number of criteria, represents the NMR mapping data well and also resembles the NMR structures of RhoA and Rac1 in complex with their HR1 domain partners (Maesaki et al. 1999; Modha et al. 2008). It can therefore be used to provide insight into the binding interface and so the differential affinities and G protein-specificities of the HR1 domains. It can also be used in combination with competitive NMR titration experiments with TOCA1 and N-WASP to provide insight into the pathway of Cdc42/N-WASP/TOCA1-dependent actin assembly.

5.4.1 Structural Insight into Differential Affinities and Specificities

The TOCA1 HR1 domain is a left-handed coiled-coil comparable to other known HR1 domains (Maesaki et al. 1999; Owen et al. 2003; Kobashigawa et al. 2009). A single binding interface on both the HR1 domain and Cdc42 can be concluded from the data presented here and the interfaces are comparable to those of other G protein-HR1 interactions. The binding region is well conserved between TOCA1 and CIP4, which is unsurprising given their overall sequence identity of 68 %, but the region is less well conserved between the TOCA and PRK families. Despite this, the biochemical differences and interface properties do not appear to explain the different affinities.

A correlation between the size of a buried protein-protein interface and binding affinity has previously been described (Chen et al. 2013), with the exception of flexible complexes (Kastritis and Bonvin 2013). The G protein-binding interface of
the high affinity HR1b domain of PRK1 is larger than that of the other HR1 domains (2180 Å compared to 1972 Å for TOCA1-Cdc42 and 1640 Å for RhoA-HR1a, Table 5.2) but the highest affinity HR1a domain has the smallest buried surface area. Interface size does not, therefore, correlate with affinity in this case.

The low and high affinity binders can also not be grouped according to the amino acid properties in this region. For example, one of the high affinity binders, HR1a^{PRK1}, has more hydrogen bond donors/acceptors than the lower affinity TOCA1 HR1 domain but the other high affinity binder, HR1b^{PRK1}, has fewer (Table 5.2). Furthermore, the number and percentage of charged residues is different for one of the high affinity binders (HR1b) but is comparable between the highest affinity (HR1a) and the lowest affinity binders (TOCA1 and CIP4). In short, there is no trend between interface composition and affinity.

It should be noted, however, that the interface of Cdc42-TOCA1 has been defined based on the HADDOCK model unlike the interfaces of PRK1-Rac1/RhoA, which have been defined based on NMR and crystal structures and so the composition of the Cdc42-TOCA1 interface is less certain. Furthermore, the Cdc42-binding interface of CIP4 has been assessed only based on chemical shift mapping (Kobashigawa et al. 2009) and solvent accessibility (assessed with NACCESS) so is likely to be overestimated. The data in Table 5.2 should be considered with this in mind.

<table>
<thead>
<tr>
<th></th>
<th>Affinity (µM)</th>
<th>Charged</th>
<th>Polar</th>
<th>Hydrophobic</th>
<th>Specials</th>
<th>Total</th>
<th>Buried Surface Area (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>#</td>
<td>%</td>
<td>#</td>
<td>%</td>
<td>#</td>
<td>%</td>
</tr>
<tr>
<td>TOCA1</td>
<td>5</td>
<td>10</td>
<td>37</td>
<td>6</td>
<td>22</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>CIP4</td>
<td>5</td>
<td>12</td>
<td>39</td>
<td>5</td>
<td>16</td>
<td>10</td>
<td>32</td>
</tr>
<tr>
<td>PRK1</td>
<td>0.06</td>
<td>11</td>
<td>38</td>
<td>4</td>
<td>14</td>
<td>10</td>
<td>34</td>
</tr>
<tr>
<td>PRK1</td>
<td>0.07</td>
<td>9</td>
<td>23</td>
<td>13</td>
<td>33</td>
<td>16</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 5.2. Interface Properties of the TOCA and PRK HR1 Domains. A table showing the affinity of each HR1 domain for its cognate G protein, the number (#) and percentage of each residue type and the buried surface area derived from the structures (PRK1 HR1s), model (TOCA1 HR1) or chemical shift mapping data (CIP4 HR1) of the HR1-G protein complex. ‘Specials’ comprise Hydrogen bond donors and acceptors.
Further insight into the importance of particular contacts in achieving tight binding can be gained from analysis of previously published mutagenesis data for Rac1 and RhoA interactions with the PRK1 HR1 domains (Owen et al. 2003; Hutchinson et al. 2011). Mutation of the Switch 2 residue, Asp65\textsuperscript{RhoA}/Asp63\textsuperscript{Rac1}, significantly decreases the affinities of RhoA and Rac1 for the HR1a and HR1b domains of PRK1. Mutation of Asp65\textsuperscript{RhoA} decreases its affinity 5-fold and mutation of Asp63\textsuperscript{Rac1} decreases the affinity more than 15-fold to >1 \mu M. Asp65\textsuperscript{RhoA} forms a salt bridge with Lys48\textsuperscript{PRK1} and Asp63\textsuperscript{Rac1} contacts Gln134\textsuperscript{PRK1}. If this contact was missing in the Cdc42-TOCA1 complex it could account, in part, for the lower affinity interaction.

Asp63 is conserved in Cdc42 but its involvement in the Cdc42-TOCA1 interaction cannot be concluded based on the shift mapping data as it is missing in the spectrum of free Cdc42. Lys48\textsuperscript{PRK1} is conserved in TOCA1 (Lys362\textsuperscript{TOCA1}, see Figure 5.7B) but the sidechain of this residue in TOCA1 was also not visible in the chemical shift mapping data due to overlap. In the Cdc42-TOCA1 model, Lys362\textsuperscript{TOCA1} is near to Glu62\textsuperscript{Cdc42} and Asp63\textsuperscript{Cdc42} (Figure 5.25B) and it may be involved in an electrostatic contact similar to that seen between Asp65\textsuperscript{RhoA} and Lys48\textsuperscript{PRK1}. Therefore, this is unlikely to provide an explanation for the lower affinity of the Cdc42-TOCA1 interaction.

Mutations of two other switch residues, Phe39\textsuperscript{RhoA} or Tyr66\textsuperscript{RhoA} (equivalent to Phe37 and Tyr64 in Rac1 and Cdc42), reduce the affinity of RhoA for HR1a from 97 nM to >1 \mu M. Phe39\textsuperscript{RhoA} is involved in hydrophobic contacts, for example with Leu59\textsuperscript{PRK1}, and Tyr66\textsuperscript{RhoA} forms electrostatic contacts with Lys51\textsuperscript{PRK1} and hydrophobic contacts with Leu52\textsuperscript{PRK1}. Phe37\textsuperscript{Cdc42} and Tyr64\textsuperscript{Cdc42} are not visible in the Cdc42 shift mapping data, nor are they involved in explicit contacts in the model. They are, however, near to equivalent charged and hydrophobic residues of TOCA1 HR1, for example Tyr64\textsuperscript{Cdc42} is near to Lys362\textsuperscript{TOCA1}, Arg358\textsuperscript{TOCA1}, Leu356\textsuperscript{TOCA1} and Leu359\textsuperscript{TOCA1}. It is likely that these contacts are conserved in Cdc42-TOCA1 but it is not certain. Omission of these contacts could contribute to the lower affinity observed.

Mutation of a final switch residue, Leu67\textsuperscript{Rac1}/Leu69\textsuperscript{RhoA} reduces the affinity of the interaction 5 to 10-fold. In both cases the Leucine residue contacts a hydrophobic
patch on the HR1 domain, but in Cdc42-HR1\textsuperscript{TOCA1}, Leu67 is pointing away from the nearby charged sidechains of HR1 and in towards the core of Cdc42 (as was shown in Figure 5.24). The omission of this hydrophobic interaction could therefore contribute to, but it is unlikely to fully explain, the 100-fold lower affinity of TOCA1 interaction.

Further insight into the different affinities is provided by examining the secondary structures of the HR1 domains. The first differences are seen in a short region N-terminal to the coiled-coil, which in TOCA1 exhibits a series of turns, reverses direction and contacts residues of both helices (Chapter 4, Figures 4.19 and 4.21). The corresponding sequence in CIP4 also includes a series of turns but is flexible (Kobashigawa et al. 2009), while in the HR1a domain of PRK1 the equivalent region adopts an α-helical structure that packs against the coiled-coil. The N-terminal region is distant from the G protein-binding interface in all of the HR1 domains, and so the structural differences may relate to the overall structure and regulation of these proteins rather than their interactions.

The second set of differences are seen in the inter-helical loops of TOCA1 and CIP4, which are within the G protein binding region and which differ from the same region in the HR1 domains of PRK1. In TOCA1 and CIP4 this region is longer and contains two short stretches of $3_{10}$-helix. It is relatively inflexible and, in TOCA1, a convincing number of NOEs are seen between residues of this region and the two α-helices (Chapter 4, Figure 4.22). The loop sidechains therefore appear to be occluded, in contrast to the PRK HR1 domains (Chapter 4, Figure 5.24).

Given the extent to which the inter-helical residues are affected by Cdc42, it is likely that these residues make direct contacts despite their occlusion in the free HR1 domain. Furthermore, this occluded inter-helical region includes the critical residues, Met$^{383}$, Gly$^{384}$ and Asp$^{385}$, mutation of which abolishes TOCA1 binding to Cdc42 (Ho et al. 2004). Therefore, the conformation of this loop may be altered in the Cdc42 complex. The CSPs seen for many of the buried residues around this region and a little further from the key MGD motif (blue in Figure 5.6), are consistent with this hypothesis. The HADDOCK model of the complex is also consistent with this and provides some possible residues involved in disrupting the loop-helix contacts. The
occlusion of these critical residues in the free HR1 domain and subsequent need for structural rearrangements may therefore contribute to the lower affinity interaction.

Some potential sequence specific contacts were derived from the Cdc42-HR1\textsuperscript{TOCA1} model (Figure 5.21) and so the data presented in this chapter have also provided insight into binding specificity. For example, an additional switch $2^{\text{Cdc42}}$-helix $2^{\text{TOCA1}}$ salt bridge (Arg$68^{\text{Cdc42}}$-Glu$395^{\text{TOCA1}}$) does not have an equivalent in the RhoA- or Rac1-HR1 domain complexes. Gln$2^{\text{Cdc42}}$ with Asn$380^{\text{TOCA1}}$ and Thr$52^{\text{Cdc42}}$ with Lys$372^{\text{TOCA1}}$ may also represent specific contacts, the former is consistent with a direct binding role for the inter-helical loop region of TOCA1.

To conclude, the data presented in this chapter has revealed a Cdc42-HR1 interaction that resembles previously studied Rho-HR1 interactions but with some interesting differences. It has identified some specific Cdc42-HR1 contacts and provided insight into the molecular reasons for the different affinities and specificities.

\subsection{5.4.2 Unidirectional Competition between N-WASP and TOCA1}

A comparison of the HADDOCK model presented here with the structure of Cdc42 in complex with the WASP GBD suggested distinct but overlapping binding sites (Figure 5.26). The NMR titration experiments precluded the possibility of a ternary complex and instead suggested unidirectional competition (Figures 5.29 and 5.30). Pyrene actin experiments were consistent with this data (Figure 5.31).

A closer comparison of Cdc42-GBD\textsuperscript{WASP} with Cdc42-HR1\textsuperscript{TOCA1} reveals a potential mechanism by which one effector could be displaced by the other. Firstly, the lysine residues at the N-terminal end of the GBD, thought to be involved in an electrostatic steering mechanism in WASP-Cdc42 binding (Hemsath et al. 2005), are conserved in N-WASP and would be able to interact with Cdc42 even when the TOCA1 HR1 domain is bound. It has been postulated that the initial interactions between this basic region and Cdc42 stabilise the active, extended conformation of WASP (the ‘R’ state), rendering the core GBD free to interact with Cdc42 (Rudolph et al. 1998). The region C-terminal to the core GBD, required for maximal affinity binding (Rudolph et al. 1998), could then bind. This would interfere sterically with binding to HR1\textsuperscript{TOCA1}. 

231
Furthermore, important hydrophobic packing between Leu70\textsuperscript{Cdc42} and Phe258/Leu267/Leu270/Phe271/Ala274/Ile276\textsuperscript{WASP} (within the C-terminal extension to the core GBD) would disrupt the hydrophobic packing between Leu70\textsuperscript{Cdc42} and Leu370/Ser388/Leu389\textsuperscript{TOCA1} (section 5.2.5).

The data presented here have therefore answered a long unanswered question, revealing that TOCA1 and N-WASP do not bind Cdc42 simultaneously. It has also corroborated earlier suggestions that TOCA1 binds to Cdc42 before N-WASP (Lee et al. 2010) and, further to this, it suggests an effector handover by which N-WASP binds to Cdc42 and displaces TOCA1. Moreover, the data provide evidence for an irreversible step in the pathway. A more detailed discussion of the factors affecting and the biological consequences of such an effector handover is presented in Chapter 7.

5.4.3 The Membrane
The pathways of Cdc42/N-WASP/TOCA1-dependent actin assembly occur at the plasma membrane. Moreover, post-translational modification and membrane anchoring of small GTPases, including Cdc42, are central to their biological functions (see Chapter 1, section 1.1.6). Despite this, the studies described here were undertaken in solution. A better understanding of small GTPases and their membrane-associated interacting partners must now be gleaned from structural and biophysical studies in the presence of membranes. The next chapter describes method development and preliminary investigations towards \textit{in vitro} studies of full length, lipid modified Cdc42 in the presence of membrane mimetics.


6 Cdc42: Towards Structural Studies on the Membrane

6.1 Introduction

Membrane anchoring of small GTPases is central to their biological functions (reviewed in Lane 2006; Cherfils and Zeghouf 2011; Wang and Casey 2016). The Ras, Rho and Rab families contain a 4-residue ‘CAAX’ motif at their C-termini and specific lipid modification of the Cysteine is essential for their function. For example, the addition of a geranylgeranyl group (20-carbon) to the CAAX box of Cdc42 is essential for its correct localisation and function in vivo (Roberts et al. 2008). In addition, short polybasic sequences adjacent to the CAAX motif of some GTPases have been identified as additional requirements for selective GTPase-membrane interactions (Roberts et al. 2008; Brunsveld et al. 2009).

Despite the centrality of the C-terminal sequences, lipid modification and membrane binding to the biological functions of the Ras superfamily, the majority of in vitro studies have focused on unmodified, C-terminally truncated forms in the absence of membranes. Historically, lipid modified G proteins have been purified from mammalian expression systems (Porfiri et al. 1995) but such methods are low yield and the modified proteins are difficult to handle. Biophysical studies on modified proteins are therefore sparse. Advancements in chemical methods to produce lipidated proteins have allowed for preliminary structural studies, which showed that membrane anchoring of Ras does not appear to significantly affect the overall structure of the G domain (Güldenhaupt et al. 2008).

Full length, unmodified G proteins also present technical challenges as the C-terminal region is typically flexible and vulnerable to proteolysis and contains one or more Cysteines, which increases aggregation in vitro. The majority of effector interactions do not appear to require the C-terminal region (Mott and Owen 2015), including the Cdc42-TOCA1 interaction (Chapter 3), and so studies on truncated forms are widely accepted. Most biophysical studies of Cdc42 to date have been performed with truncated protein, in which the 7 most C-terminal residues are missing, as omission of these residues facilitates purification of the protein for structural studies. However, in
vitro studies of truncated GTPases have been called into question as there are examples of these C-terminal regions affecting the GTPase cycle and GTPase-effector binding (see Chapter 1, section 1.5.4). Moreover, the C-terminal region of Cdc42 has been implicated in functionally important dimerisation (Zhang and Zheng 1998).

The main aim of this chapter was to move towards in vitro studies of full length, lipid modified Cdc42 in the presence of membrane mimetics using NMR spectroscopy. As most studies of Cdc42 have been on the C-terminally truncated protein, Cdc42Δ7, structural and functional comparisons between full length and truncated proteins formed an integral part of this process. For example, the effect of the C-terminal region on oligomerisation, G domain structure and nucleotide hydrolysis were of interest prior to structural studies of membrane anchored protein. The aims of this chapter were to:

1. determine the oligomeric state of full length Cdc42 in solution, measure the affinity of dimerisation and to investigate the involvement of the C-terminal region in the proposed dimerisation
2. develop methods to produce lipid modified Cdc42 for use in NMR experiments
3. compare the structures and functions of truncated, full length and lipid modified Cdc42 in solution using nucleotide hydrolysis assays, circular dichroism and NMR
4. develop methods to anchor modified Cdc42 to membrane mimetics in order to compare the structures and functions of full length and lipid modified Cdc42 in the presence of membranes

6.2 The Oligomeric State of Cdc42

Basic residues in the C-terminal regions of Cdc42, Rac1, RhoA and Rac2 have been shown to influence dimerisation, nucleotide binding and hydrolysis (Zhang and Zheng 1998; Zhang et al. 2001). The affinity of Cdc42 dimerisation is reported to be in the micromolar range (Zhang and Zheng 1998), which would fit with the tightly regulated pathway of multiple low affinity interactions described in Chapters 3 and 7. It is therefore necessary to begin this chapter with an analysis of the oligomeric state of
Cdc42. Such analysis will have significant implications for our understanding of GTPase function but will also have technical implications when approaching lipid modification of the C-terminal CAAX motif. For example, dimerisation may affect the CAAX motif prior to its lipid modification and so hinder production of lipid modified protein.

6.2.1 Protein Purification

Active mutants (Q61L) of full length Cdc42 (C-terminal region: PKKSRRCVLL) and Cdc42Δ7 (PKK) were expressed and purified as described in Chapter 3. In vivo, Cdc42 is lipid modified on the Cysteine of the CAAX box and further modified by cleavage of the three C-terminal amino acids, as described in the Introduction, and so Cdc42Δ3 (PKKSRRC) better represents physiological full length. Further, the lipid-modified Cysteine residue would be anchored to the membrane so not available to affect Cdc42 dimerisation. This residue can therefore be mutated to Serine to aid in purification of the unmodified protein without any expected functional effects. With this in mind, Cdc42Δ3 was cloned into pGEX-6P-1 and a point mutation, C188S, was introduced by substituting the Cysteine codon for a Serine codon in the reverse PCR cloning primer. Cdc42Δ3Q61L/C188S (PKKSRRS; referred to as Cdc42Δ3 hereafter) was expressed as an N-terminal GST-fusion protein in the same way as Cdc42Δ7 and full length Cdc42. An example purification of Cdc42Δ3 is shown in Figure 6.1A-B.

Nucleotide exchange for the non-hydrolysable GTP analogue, GMPPNP, was carried out with all three proteins and their nucleotide state was assessed using HPLC by comparison with nucleotide standards (Figure 6.1C). A single peak was visible in the trace of each nucleotide sample, representing GMPPNP. The proteins were further purified by size exclusion chromatography (Figure 6.1D) and final gel samples of each protein are shown in Figure 6.1E. Mass spectrometry was used to confirm the purity and completeness of the proteins (PNAC, Appendix 3A).
Figure 6.1. Purification of Cdc42 Proteins. A) A 12 % Laemmli gel showing samples taken from the supernatant and each wash of the glutathione agarose beads (W1-3) and from a 50 % bead slurry (Beads) during purification of Cdc42Δ3. B) A 15 % Laemmli gel showing gel samples taken from the supernatant (SN) after 3C protease cleavage, bead washes (W1-3) and a final bead sample (Beads). C) HPLC traces showing nucleotide standards (GDP in blue, GTP in teal and GMPPNP in red) and nucleotide samples (black) from each protein run on a 1.5 mL Partisphere SAX HPLC column (Whatman) in 0.6 M Ammonium Phosphate pH 4.0. D) A280 traces showing the purification by size exclusion chromatography on a Superdex 16/60 S75 column in 50 mM Tris:HCl pH 7.5 with 150 mM NaCl, 5mM DTT and 5 mM MgCl2. E) A 15 % Laemmli gel showing a sample of each Cdc42 protein.
6.2.2 Analysis of the Oligomeric State

As the reported affinity of dimerisation is ~2 µM (Zhang and Zheng 1998), AUC was used to investigate the oligomeric state of Cdc42 and to determine the affinity of dimerisation more accurately (Balbo and Schuck 2005). Each of the Cdc42 proteins was analysed at increasing concentrations using sedimentation velocity AUC experiments such as those described for CIP4 in Chapter 3.

The optimal range for absorbance optics in AUC is $A_{280} = 0.1$-1.0. The optimal concentration range of Cdc42 to remain within this absorbance range was therefore 4-42 µM, calculated using the Beer-Lambert Law with the theoretical extinction coefficients for each GMPPNP-bound protein. A two-fold dilution series of full length and three-fold dilution series of Cdc42Δ7 and Cdc42Δ3 were prepared with a maximum concentration of 97 µM. Thus, two data points lay within the optimal concentration range and one was higher to ensure a sufficient excess over the proposed affinity ($K_d \approx 2$ µM).

The AUC data were fit with the continuous distribution model, as described in Chapter 3. The data fits were high quality as indicated by the small residuals (<10 % of maximum signal) and an RMSD lower than 0.1 (Figure 6.2). The c(s) distribution contained a single peak with an s-value of 2.1 in all cases (ignoring a buffer mismatch peak with an s-value of ~0). The fitted data were used to derive an estimation of the molecular mass (assuming a uniform frictional ratio of 1.17, derived during the data fitting), which is marked on the distributions in Figure 6.2. The calculated masses were indicative of monomer for each protein. The estimated molecular weights were within 5 % of the expected weights, except for the 10 µM sample of Δ3, which was further from the expected value (24.0 kDa compared to 21.5 kDa). This is probably because the peak in the c(s) distribution is broader in the lower concentration samples and so the s-value is less accurately determined.
Interference optics can be used when assessing protein samples with an absorbance outside the optimal range. However, non-ideality of sedimentation caused by the physical effects of molecular crowding can be a problem (Balbo and Schuck 2005).

Figure 6.2. AUC Analysis of the Oligomeric State of Cdc42. Sedimentation velocity data were recorded on Cdc42 and the data fit with the continuous $c(s)$ distribution model. A) The fringes and the residuals from the fit for the full length samples. B-C) The $c(s)$ distributions are shown for each experiment as indicated. The distributions each show a single species, assuming a uniform frictional ratio of $F_{k,w} = 1.17$, $s_{20,w}^0 = 2.1$ in each experiment. The calculated MW is shown for each peak. The expected MWs for monomeric Cdc42 were as follows: 20.7 kDa for Δ7, 21.5 kDa for Δ3 and 21.7 kDa for full length.
Nonetheless, a high concentration sample of Cdc42Δ3 (330 µM) was analysed (Figure 6.3). This concentration is in line with those used in NMR studies. Despite the observations of non-ideality, seen in the visible diagonal components in the grey plot below the fringes (Figure 6.3), the data fit reasonably well (RMSD = 0.156). The s-value was the same as in the lower concentration samples (2.1), the estimated mass was 19.2 kDa and the data was therefore indicative of monomer. The AUC data showed no indication of dimerisation in solution.

![AUC Analysis of a Higher Concentration of Cdc42](image)

**Figure 6.3. AUC Analysis of a Higher Concentration of Cdc42.** Analytical ultracentrifugation sedimentation velocity data of Cdc42Δ3 at a higher concentration. The data were fit with the continuous c(s) distribution model. The fringes, residuals from the fit and the component sedimentation coefficient distribution are shown. The grey image of the AUC cell is shown below the fringes. This shows a strong diagonal component arising from non-idealities at such high concentrations of protein. The component sedimentation coefficient distribution shows a single species, assuming a uniform frictional ratio of \( F_{k,w} = 1.17 \). \( s^0_{20,w} = 2.1 \). The calculated MW is shown.

Comparable AUC experiments were carried out with full length wild type Cdc42 bound to GDP to ensure that it is also monomeric when in the inactive state. The continuous distributions are shown in Figure 6.4. A single peak with an s-value of 2.1 was observed in the 10 and 32 µM samples, excluding the buffer mismatch peak. An additional peak was seen in the 97 µM sample with an s-value of 3.3. The mass of this
second peak fit to ~40 kDa, which could be consistent with dimeric Cdc42. However, the small percentage of the sample that is accounted for by this peak is indicative of a $K_d$ of dimerisation more than 1 mM, which is unlikely to have any physiological relevance and is most likely due to intermolecular disulphide bonds forming between the free Cysteine residue of the CAAX box at high concentrations. It is also possible that the additional peak comes from noise or artifacts resulting from the molecular crowding and resultant non-ideality of sedimentation at this higher concentration.

**Figure 6.4. AUC Analysis of Inactive Cdc42.** Analytical ultracentrifugation sedimentation velocity data of full length, wild type Cdc42 bound to GDP. The AUC data were fit with the continuous $c(s)$ distribution model and the component sedimentation coefficient distribution are shown here. The calculated MW, assuming a uniform frictional ratio of $F_{k,w} = 1.2$, is shown.
6.3 Lipid Modification of Cdc42

Prenylated GTPases have previously been purified from mammalian expression systems (Porfiri et al. 1995), which are costly and typically result in low yields. This coupled with expensive media means that the cost of producing labelled samples for NMR is prohibitive. Alternative methods to achieve lipid modification have since been developed. For example, a farnesyl group has been chemically linked to the C-terminal Cysteine of Ras via a maleamide group (Bader et al. 2000) and farnesylated proteins have been produced in *E. coli* by co-expression of the GTPase with the appropriate transferase (Fres et al. 2010). *In vitro* farnesylation with Farnesyl Transferase (FTase) purified from *E. coli* has also been successful (Thapar et al. 2004; Kuhlmann et al. 2016).

These newer methods are lower cost, higher yield and more appropriate for NMR labelling schemes but are still low yield relative to desired NMR concentrations, typically producing just a few milligrams of prenylated protein. Furthermore, the prenylated proteins are less stable in solution due to the addition of the hydrophobic lipid group. Alternative methods to achieve membrane localisation without lipid modification have, therefore, also been developed. For example, inclusion of a C-terminal Histidine tag, which can be bound to nickel-lipids has been used to assess the affects of co-localisation of GTPases and GEFs on membranes (Peurois et al. 2017).

The preferred method for achieving membrane-anchoring of Cdc42 for NMR studies is prenylation by co-expression with the appropriate transferase in *E. coli* or *in vitro* prenylation with the appropriate lipid precursor. The alternative method of maleamide linkage relies upon there being a single solvent exposed Cysteine residue, as the maleamide-lipid is reactive towards any exposed Cysteine. Cdc42 has 7 exposed Cysteine residues and so this method would not be suitable unless the other Cysteines were first mutated. The nickel-affinity method is chemically simple but a Histidine tag does not mimic the chemical or physical properties of a lipid modification, nor does it allow for a physiological-like insertion of the protein in the membrane. Therefore, modification of Cdc42 using transferases was attempted first.
In vivo, Cdc42 is geranylgeranylated and this lipid moiety resides either within a hydrophobic pocket of RhoGDI or is inserted into membranes (Hoffman et al. 2000). Geranylgeranyl phosphate is only present at low concentrations in E. coli and so the co-expression method is only applicable if geranylgeranyl synthase is simultaneously expressed (Albermann et al. 2008). Furthermore, the addition of such a long lipid moiety in vitro is expected to reduce the solubility of the protein due to the hydrophobic nature of the long hydrocarbon chain. Conversely, purification of farnesylated GTPases has been successful. For example, the farnesylation of the large GTPase hGBP1 was achieved by co-expression of the G protein with the FTase in E. coli (Fres et al. 2010) and in vitro farnesylation of the small GTPase, Ras, has been achieved with reasonable success (Thapar et al. 2004). The farnesyl group (C<sub>15</sub>H<sub>28</sub>) is significantly shorter than the geranylgeranyl (C<sub>20</sub>H<sub>36</sub>) and so less hydrophobic making it easier to work with in vitro.

The following sections describe the development of a method to farnesylate Cdc42. Mass spectrometry was carried out by the PNAC service at the Department of Biochemistry to assess the farnesylation of Cdc42 throughout this method development. The mass spectrometry data is available in Appendix 3 and the amount of farnesylated protein, derived from this mass spectrometry data, will be stated in the following results sections.

6.3.1 Co-expression of Cdc42 with Farnesyl Transferase

6.3.1.1 Preliminary Expressions

It was decided to begin generating farnesylated Cdc42 by co-expression with the FTase, based on the method used to farnesylate hGBP1 (Fres et al. 2010). However, Cdc42 is not a natural substrate for FTase but rather is a substrate for Geranylgeranyl Transferases (GGTase). The specificities of GGTase and FTase for different GTPases are determined by the C-terminal residue, i.e. the ‘X’ of the CAAX motif (Reid et al. 2004): GGTase accepts Leucine, Phenylalanine, Isoleucine and Valine but the ‘X’-binding pocket of the FTase is more polar so it instead binds better to Methionine, Serine, Glutamine, Alanine, Threonine and Cysteine. The binding and efficiency of different CAAX motifs has been studied and the data suggest that motifs ending in Met or Ala constitute the best substrates for FTase (Roskoski and Ritchie 1998).
The C-terminal residue of Cdc42 was therefore first mutated from a Leucine to an Alanine or Methionine by site directed mutagenesis. The mutagenesis was performed on full length Cdc42Q61L in pGEX-6P-1 expression vector as described in Chapter 2 and confirmed by DNA sequencing. Small scale co-expressions of the L191A mutant with FTase were then carried out in a variety of E. coli strains (Figure 6.5), including the strain which was successfully used for farnesylation of hGBP1, Rosetta2 (Fres et al. 2010). The intense band at ~47 kDa that can be seen across all of the strains and conditions shown in Figure 6.5A-C is consistent with GST-Cdc42. In BL21 CodonPlus RIL and Rosetta2 grown at 37 °C for 5 hours there was an additional band visible just above GST-Cdc42 (marked with a red arrow in Figure 6.5A). This band could be the α and β subunits of the FTase (~48 and ~46 kDa). Given that the band appears distinctly above the band for GST-Cdc42, it could also be prenylated Cdc42 having undergone a gel mobility shift due to the increased size and hydrophobicity.
Figure 6.5. Small Scale Co-expressions of Cdc42 and FTase in E. coli. A-C) 12% Laemmli gels showing samples taken from the indicated bacterial strains transformed with pRSF-Duet-FTase and pGEX-6P-1-Cdc42Q61L/L191A, grown at 37 °C for 5 h or 20 °C for 20 h as indicated. Abbreviations are as follows: U – uninduced, I – induced with 1 mM IPTG, S – soluble, P – pellet. The red arrow in A marks the band that may represent the FTase or modified Cdc42. The condition chosen for large scale expression is marked with a red asterisk.
A large scale co-expression of full length Cdc42Q61L/L191A and FTase was carried out in 3 L of *E. coli* Rosetta2. Cdc42 was purified via the N-terminal GST-tag, similarly to the purification of unmodified Cdc42, but the protein precipitated significantly following cleavage of the GST-tag. Only 0.5 mg of protein was obtained and mass spectrometry analysis indicated a trace of farnesylated protein (Appendix 3B). It was therefore concluded that the FTase was expressed and active on Cdc42Q61L/L191A but that farnesylated Cdc42 was not stable in solution.

### 6.3.1.2 Stabilising Cdc42 with Detergent

The inclusion of 0.1 % octyl glucoside (OG) was found to be necessary for solubilisation of farnesylated H-Ras for NMR studies (Thapar et al. 2004) but before undertaking attempts to stabilise lipidated Cdc42 with this detergent, it was first confirmed that it does not significantly affect the overall structure of Cdc42 using circular dichroism (CD) (Figure 6.6). CD spectra were recorded on 0.1 mg/mL Cdc42Δ7Q61L between 185 and 260 nm with increasing concentrations of OG.

Figure 6.6A shows the spectrum at each concentration of OG following subtraction of the buffer only spectra shown in Figure 6.6B. The peak at 190 nm and two troughs at 215 and 222 nm are indicative of α-helical structure. The peak at 190 nm differed slightly at different OG concentrations but the differences did not correlate with OG concentration. The differences can be explained by increased noise in this region of the spectrum caused by the absorbance by the OG itself (Figure 6.6B). The troughs at 215 and 222 nm were comparable at each concentration of OG. It was concluded, therefore, that the protein structure was not adversely affected by OG at concentrations up to 0.8 % (w/v), which is in excess of the critical micelle concentration (0.67-0.73 %).
Another 3 L expression was performed but this time 0.2 % (w/v) OG, below the critical micelle concentration, was added to one bead set during 3C protease cleavage and the bead washes. Following cleavage of the GST-tag and bead washes, there was some cleaved protein visible in the bead samples (Figure 6.7), which was indicative of precipitation. The bead set with no OG (bead set 1) had a darker band corresponding to Cdc42 in the bead sample and less protein was recovered in the first wash (W1), indicating more precipitation in the absence of OG. Approximately 5 mg was obtained and mass spectrometry analysis showed that ~20 % of the protein was farnesylated (Appendix 3C). It can be concluded from this that OG stabilised the protein but that the level of farnesylation was still low.

Figure 6.6. Circular Dichroism Analysis of Cdc42 in the Presence of Octyl Glucoside. A) CD spectra recorded between 185 and 260 nm on 0.1 mg/mL Cdc42Q61LΔ7·GMPPNP with increasing concentrations of OG are shown following subtraction of blank spectra recorded on buffer only. The concentration of OG is given as percentage w/v. B) The buffer only control data is shown.
Comparing L191A with L191M

Once the FTase was shown to be active and the prenylated Cdc42 stabilised, the farnesylation of the two CAAX mutants, L191A and L191M, were compared. The CVLA and CVLM mutants were each co-expressed with FTase in 1.5 L of E. coli Rosetta2. As the level of farnesylation achieved by co-expression was low, a 1.5 molar excess of Farnesyl Pyrophosphate (FPP) was added to the cell lysates along

Figure 6.7. Co-expression of Cdc42 Q61L/L191A with FTase. A) A 12 % Laemmli gel showing the GST affinity purification of Cdc42 co-expressed with FTase. B) A 15 % Laemmli gel showing samples taken after cleavage of the GST tag with 3C protease. 0.2 % (w/v) OG was added to bead set 2 during and after protease cleavage, as indicated with ‘+’. The final bead sample clearly shows more precipitated Cdc42 in bead set 1.

6.3.1.3 Comparing L191A with L191M

Once the FTase was shown to be active and the prenylated Cdc42 stabilised, the farnesylation of the two CAAX mutants, L191A and L191M, were compared. The CVLA and CVLM mutants were each co-expressed with FTase in 1.5 L of E. coli Rosetta2. As the level of farnesylation achieved by co-expression was low, a 1.5 molar excess of Farnesyl Pyrophosphate (FPP) was added to the cell lysates along
with 20 µM ZnSO₄ (Zn²⁺ is an essential cofactor for FTase) during the incubation with glutathione-agarose beads. The cell lysate, containing FTase and Cdc42, was incubated with the FPP and glutathione beads for 4.5 h at 4 °C before the bead washes and protease cleavage that were carried out as in previous purifications. 6 mg of the CVLA mutant and 1.2 mg of the CVLM mutant were obtained. Mass spectrometry analysis revealed that only ~10-20 % of the CVLM mutant was farnesylated, whereas ~99 % of the CVLA mutant was farnesylated (Appendix 3D).

There are a number of variables in the co-expressions that cannot be controlled and so the reason for the different levels of farnesylation are not easily deduced. A possible reason for the significantly greater farnesylation of the L191A mutant compared with the L191M mutant could be that the FTase was present at higher concentrations relative to the concentration of Cdc42. Alternatively, the L191A mutant may represent a better substrate for FTase. Either way, the L191A mutant was expressed to a higher level and more was farnesylated and so this mutant was used hereafter.

The 6 mg of farnesylated protein that was obtained was frozen at -80 °C until further experiments were carried out (later sections). This is sufficient for preliminary membrane binding experiments, CD and NMR but for NMR, isotope labelled protein would be required. Before expressing labelled protein, the co-expression was repeated with no changes to the protocol to determine whether the yield and level of farnesylation was reproducible. In this repeat, only 3 mg of protein was obtained from 3 L of E. coli and only ~60 % was farnesylated. Further repeats were carried out, resulting in 0-60 % farnesylation.

6.3.2 Separation of Farnesylated and Unmodified Protein
The level of farnesylation was evidently variable between protein preparations and did not exceed 60 % in all but one attempt. Various methods were employed, therefore, to separate the modified and unmodified protein. Firstly, hydrophobic exchange chromatography was employed as the modified protein is expected to be more hydrophobic than the unmodified protein. This method has been used to separate modified hGBP1 from unmodified protein (Fres et al. 2010). The authors found that both species were retained on a butyl sepharose column but eluted at different
concentrations of detergent. The method used by (Fres et al. 2010) was unsuccessful here, as neither Cdc42 nor farnesylated Cdc42 stuck to the column (data not shown). In further attempts, ammonium sulphate was used as a salting out agent, at 1.0, 1.5 and 1.7 M. However, the protein did not stick to the column at 1.0 or 1.5 M ammonium sulphate and it precipitated at 1.7 M (data not shown).

As an alternative, size exclusion chromatography was used. Whilst the size difference between modified and unmodified Cdc42 was not expected to be sufficiently large to allow separation in this way, the size difference between monomeric and dimeric Cdc42 would be. Full length Cdc42 contains an exposed Cysteine (of the CAAX box), which forms intermolecular disulphide bonds in oxidizing conditions, resulting in dimerisation and further aggregation. The modified version has a farnesyl group on this Cysteine and so would be expected to remain monomeric in oxidizing conditions. By running the size exclusion column in the absence of reducing agents, the two species were separated (Figure 6.8). Mass spectrometry showed that the pre sample was ~40 % farnesylated, the void peak contained both the modified and unmodified species and peak 1 contained only farnesylated protein (Appendix 3E). 0.4 mg of farnesylated protein was obtained and frozen at -80 °C. The gel filtration was repeated with 3 more mixed samples to obtain more protein but in each case the void peak and peak 1 both contained a mixture of unmodified and modified protein.

![Figure 6.8. Purification of Farnesylated Protein using Size Exclusion Chromatography.](image)

A) An $A_{380}$ chromatogram showing the purification of modified from unmodified protein by size exclusion chromatography on a Superdex 16/60 S75 column in 50 mM Tris: HCl pH 7.5, 150 mM NaCl.
Finally, ion exchange was also attempted. 1.2 mg of protein (a mixture of modified and unmodified) was loaded onto a MonoQ column in 15 mM NaCl and eluted using a NaCl gradient from 0 to 1M. Unfortunately, no protein was recovered and so it was assumed that all of the protein had precipitated on the column. Due to time constraints, it was decided to make no further attempts at separation but instead to pursue alternative methods to achieve higher levels of farnesylation.

6.3.3 *In vitro* Farnesylation with Purified Proteins

The co-expression has a number of uncontrollable variables and so the FTase was next purified for use in *in vitro* farnesylation reactions, where the concentrations of both proteins and the FPP substrate can be controlled. *In vitro* farnesylation had previously been successful for Ras (Thapar et al. 2004) and, part way through the course of this project, the same method was applied to RhoA and Cdc42 (Kuhlmann et al. 2016).

6.3.3.1 Purification of Farnesyl Transferase from *E. coli*

The FTase was expressed in Rosetta2 and purified by nickel affinity chromatography as described in Chapter 2. The FTase was further purified by gel filtration and both peak 1 and peak 2 contained the FTase (Figure 6.9). Peak 1 contained more contaminants, in particular one of ~70 kDa, so was discarded. The fractions from peak 2, containing the majority of the FTase and less contamination, were pooled and concentrated to 20 mg/mL. In total, 50 mg of FTase was obtained from 3 L of *E. coli*, which was frozen at -80 °C in 500 µL aliquots.
6.3.3.2 *In vitro* Farnesylation

An *in vitro* farnesylation reaction was carried out with a Cdc42 sample produced earlier by co-expression. The Cdc42 sample contained ~30 % farnesylated protein before the *in vitro* reaction and the rest was full length, unmodified protein. The *in vitro* farnesylation was carried out with 200 μM Cdc42, 10 μM FTase and 300 μM FPP for 1 h at 30 °C, based on the method used to produce farnesylated Ras, RhoA and Cdc42 (Thapar et al. 2004; Kuhlmann et al. 2016). Cdc42 was then separated from the FTase using gel filtration and the fractions containing Cdc42 were pooled and concentrated (data not shown). The pre-sample was ~30 % farnesylated and the post-sample was ~60 % farnesylated. The 60 % farnesylated sample was then subjected to another *in vitro* farnesylation under the same conditions but the level of farnesylation did not increase above ~60 % (Appendix 3F). This indicated that some of the protein in the sample was unable to be modified.

A fresh batch of Cdc42Q61L/L191A was prepared for further *in vitro* farnesylation trials and two reactions were carried out with 200 μM Cdc42, 50 μM FTase and 400 μM FPP. One reaction was allowed to continue at 30 °C for 2 h and the other for 3 h.

---

**Figure 6.9. Purification of Farnesyl Transferase.** A) An A280 trace showing the final stage of purification by size exclusion chromatography on a Superdex 16/60 S200 column in 50 mM Tris:Cl pH 7.5 with 150 mM NaCl, 5mM DTT and 20 μM ZnCl₂. B) A 12 % Laemmli gel showing a 1 in 10 dilution of the sample that was loaded onto the gel filtration column (L) and 1 in 2 dilutions of samples taken from peaks 1 and 2. The two subunits can be seen as indicated.
The two samples each contained ~60% farnesylated protein despite the different reaction times.

Several repeats of the in vitro reactions showed that the level of farnesylation varied between 50 and 80% suggesting that there must be another variable that was not controlled that was affecting the reaction. One obvious candidate for this variable was the nucleotide state of Cdc42. The Q61L mutant of Cdc42 is, in theory, unable to hydrolyse GTP and so the samples were each expected to be close to 100% GTP bound. However, this had not been formally shown. The nucleotide state of two Cdc42 samples from separate purifications were therefore tested using HPLC (Figure 6.10).

![HPLC traces showing nucleotide samples from two protein samples run on a 1.5 mL Partisphere SAX HPLC column (Whatman) in 0.6 M Ammonium Phosphate pH 4.0. The peaks were defined manually and integrated automatically in DIONEX Chromeleon (Thermo Fisher).](image)

**Figure 6.10. HPLC Analysis of the Nucleotide State of Cdc42.** HPLC traces showing nucleotide samples from two protein samples run on a 1.5 mL Partisphere SAX HPLC column (Whatman) in 0.6 M Ammonium Phosphate pH 4.0. The peaks were defined manually and integrated automatically in DIONEX Chromeleon (Thermo Fisher).

The two samples were expressed and purified in the same way but sample 1 was left at 4 °C for an extra 48 hours compared to sample 2, which was frozen at -80 °C immediately following gel filtration. The peaks corresponding to GDP and GTP in each HPLC chromatogram were integrated revealing that sample 1 was ~35% GTP bound and sample 2 was more than 90% GTP bound (Figure 6.10). This could be
explained by Cdc42Q61L retaining some hydrolysis activity within *E. coli* or following purification, or by GTP dissociating and undergoing spontaneous hydrolysis in solution. Either way, the nucleotide state may vary between each Cdc42 purification and over time during the farnesylation reactions and purification.

### 6.3.3.3 Farnesylation and the Nucleotide State

In order to control for this additional variable, farnesylation reactions needed to be carried out with a Cdc42 sample that comprised only one nucleotide-bound state that would not change over time. Full length Cdc42Q61L/L191A was therefore exchanged for the non-hydrolysable GTP analogue GMPPNP or for GDP. The two proteins were then used in *in vitro* farnesylation reactions under the same conditions and using the same batch of FTase. Cdc42 was 100 % loaded with GMPPNP (Figure 6.11A) but the protein was not farnesylated to any detectable level (Appendix 3G). The GDP exchange resulted in a mixture of GDP and GTP bound protein (Figure 6.11B). This mixed sample was ~30 % farnesylated (Appendix 3G).

![Figure 6.11. HPLC Analysis of the Nucleotide State of Cdc42 After Nucleotide Exchange.](image)

**Figure 6.11. HPLC Analysis of the Nucleotide State of Cdc42 After Nucleotide Exchange.** A) HPLC traces showing a GMPPNP standard (pink) and two samples of Cdc42 following GMPPNP exchange (blue and black), run on a 1.5 mL Partisphere SAX HPLC column (Whatman) in 0.6 M Ammonium Phosphate pH 4.0. B) A nucleotide sample from the same batch of protein but after GDP exchange, run in the same conditions as in A.
A mutagenesis reaction was carried out on Cdc42Q61L/L191A to revert the Q61L mutation back to wild type and so to produce the fully hydrolysis active protein. This protein was 100 % GDP bound following purification (Figure 6.12A). The protein was exchanged for GMPPNP (Figure 6.12B) or left as GDP bound and further farnesylation reactions carried out. Mass spectrometry revealed that the GMPPNP bound protein was not farnesylated at all whilst the GDP bound protein was ~60 % farnesylated (Appendix 3H).

![HPLC Analysis of the Nucleotide State of Cdc42L191A](image)

**Figure 6.12. HPLC Analysis of the Nucleotide State of Cdc42L191A.** A) An HPLC trace showing a sample from Cdc42L191A run on a 1.5 mL Partisphere SAX HPLC column (Whatman) in 0.6 M Ammonium Phosphate pH 4.0. B) A nucleotide sample from the same protein but after exchange for GMPPNP.

These results suggest that the CAAX motif is available for FTase binding and subsequent farnesylation when Cdc42 is in the GDP-bound state but not when it is in the GMPPNP-bound state. The HPLC data shown in Figure 6.10 suggested that Cdc42Q61L may retain some hydrolysis activity and so hydrolysis of GTP during farnesylation reactions could be contributing to the observed variability. A repeat of the farnesylation reaction with wild-type Cdc42 bound to GDP produced 0.9 mg of ~90 % farnesylated protein.
6.4 The C-terminus in Nucleotide Hydrolysis

Given the effect of the nucleotide state on accessibility of the C-terminus for farnesylation, it is possible that the C-terminus interacts with the G domain in a manner that is affected by the bound nucleotide. It therefore follows that the C-terminus may interact with the G domain near to the nucleotide binding regions and so may have a reciprocal affect on nucleotide binding or hydrolysis. The nucleotide state of Cdc42 can be assessed using HPLC analysis, as already described, and so can report on nucleotide hydrolysis rates when samples are assessed over a time course. Interestingly, the HPLC data shown in Figure 6.10 suggested that full length Cdc42Q61L retains some hydrolysis activity and so the hydrolysis rates of the full length and truncated Q61L proteins were compared using HPLC.

The Q61L mutants of Cdc42Δ7, Δ3 and full length Cdc42 were incubated at 25 °C for 100 hours and samples were taken at 0, 1, 2, 4, 6, 8, 24, 48 and 100 hours. The full length protein used was L191M. A GTP only control was subjected to the same conditions and tested at 0, 24 and 100 hours to control for spontaneous hydrolysis in solution. The nucleotide state of each sample was analysed by comparison with nucleotide standards using HPLC as described earlier. An example of the HPLC data for Cdc42Δ7 is shown in Figure 6.13. The peaks corresponding to GDP and GTP were integrated to determine the percentage of each.
This was repeated for each Cdc42 protein and the data is summarised in a line graph in Figure 6.14A. GTP alone did not hydrolyse at all during the course of the experiment but GTP was hydrolysed by all three Cdc42 proteins. Cdc42Δ3 began with 94% of the protein GTP bound, which decreased to 88% within just 1 hour and by 24 hours only 24% remained GTP bound. Full length protein followed a similar pattern. Conversely, Cdc42Δ7 began with 88% GTP bound and by 24 hours was still 80% GTP bound. These data therefore indicate that the presence of the C-terminus leads to an increase in the rate of nucleotide hydrolysis by the Q61L proteins.

Each experiment was repeated and, in this second experiment, each HPLC sample was run in duplicate to control for variability in the HPLC analysis. The HPLC results were identical between the duplicate samples, as shown in Figure 6.14B, confirming that it is a suitably quantitative method to assess the nucleotide state. The protein samples were each prepared at 0.5 mM to control for any concentration dependent affects. Also, the same buffer was used for all of the proteins, containing 5 mM DTT to prevent disulphide formation and OG, which is required to maintain stability of the full length protein. The full length protein used in this second experiment was L191A
rather than L191M. The hydrolysis seen with Cdc42Δ7 and Cdc42Δ3 was comparable to the first experiment but the hydrolysis observed for full length L191A was slow, resembling that of Cdc42Δ7 rather than Cdc42Δ3. The experiment was repeated with Cdc42Δ3 and farnesylated L191A (Figure 6.14C).

Figure 6.14. HPLC Analysis of Nucleotide Hydrolysis by Cdc42Q61L Proteins. Line graphs showing the % GTP present in protein samples taken at various time points after incubation in 50 mM Tris:HCl, 150 mM NaCl, 5 mM MgCl2, 5 mM DTT and 0.1 % OG at 25 °C. A) The data for GTP (black), Cdc42Δ7Q61L (blue), Cdc42Δ3Q61L (red) and full length Cdc42Q61L/L191M (green) are shown. B) The data for Cdc42Δ7Q61L is shown in dark and light blue, Cdc42Δ3Q61L in dark and light red and full length Cdc42Q61L/L191A in dark and light green. In each case, the two lines are indistinguishable. Each protein was at 0.5 mM during this experiment. C) The data for Cdc42Δ3Q61L (red) and farnesylated Cdc42Q61L/L191A (purple) are shown.

To compare the different proteins quantitatively, the initial rates of hydrolysis for each protein sample were derived from the graphs shown in Figure 6.14. The linear regions of the graphs (1-8 hours) were subjected to linear regression analysis, examples of which are shown in Figure 6.15A. The average initial rates, summarizing the data from all of the experiments, are shown in Figure 6.15B. The initial rate of hydrolysis
by Cdc42Δ7 was the slowest at ~0.3 % per hour. Full length L191A was comparably slow, whilst Cdc42Δ3 hydrolysed GTP more than 17x faster. Full length Cdc42L191M hydrolysed more than 15x faster than Cdc42Δ7. Farnesylated protein hydrolysed at an intermediate rate: 9x faster than Cdc42Δ7 but 0.6x the rate of full length L191M.

Figure 6.15. Initial Rates of GTP Hydrolysis by Cdc42Q61L Proteins. A) The linear region of example graphs. Linear regression analysis was used to derive the slope of each line, which was considered as a measure of the initial rate of hydrolysis. B) A bar chart showing the mean initial rate for each Cdc42 protein as the change in % GTP per hour. The number of repeats for each sample is given above each bar and the error bars represent the standard deviation. GTP only is included as a control, showing that the hydrolysis observed was protein dependent.

The difference between Cdc42Δ7 and Cdc42Δ3 is just the addition of 4 residues (SRRS) at the C-terminus of Δ3. This suggests that these residues alone are responsible for the effect on nucleotide hydrolysis that is seen when comparing these two proteins. The effect of these residues on nucleotide hydrolysis could be direct, by insertion of a catalytic residue. For example, one of the two Arginines could stabilise the phosphate leaving group in the GTP-GDP transition state and so affect hydrolysis in a similar way to the classic Arginine finger of GAPs. Alternatively, the effect could be indirect, for example via structural rearrangements in the G domain.

The difference between the rates of hydrolysis by L191A and L191M is unlikely to be due to a direct influence of the Ala or Met on nucleotide hydrolysis. Rather, the
identity of the terminal residue could affect the interaction of the C-terminal region with the G domain. For example, the residues responsible for the effect on nucleotide hydrolysis that differ between Cdc42Δ7 and Cdc42Δ3 may interact more strongly with the G domain in full length L191M due to a hydrophobic interaction by the Met that is absent in the Ala mutant. The low farnesylation of L191M compared to L191A seen in section 6.3.1.3 is consistent with the L191M C-terminus interacting more convincingly with the G domain and so being less accessible to the FTase.

The farnesylated protein (L191A) hydrolysed more slowly than full length L191M or Cdc42Δ3 suggesting that the farnesyl group hinders the effect of the C-terminal residues slightly. This could be due to steric factors as the large lipid group might be expected to hinder an interaction between the C-terminal region and the G domain. The farnesylated protein did, however, hydrolyse significantly faster than Cdc42Δ7 or full length L191A. The difference between modified and unmodified L191A could be explained by a hydrophobic interaction between the lipid group and the G domain, similar to that suggested for the L191M mutant.

An intramolecular disulphide bond between the C-terminal Cysteine and Cys105 has been observed in a crystal structure of Cdc42 in complex with Cdc42GAP (Nassar et al. 1998). This is unlikely to occur in the oxidizing environment of the cell but nonetheless it does imply a propensity of the C-terminus to interact with the G domain. Such a disulphide might be expected to affect hydrolysis and so the difference between modified and unmodified L191A could relate to the ability of this disulphide to form as the addition of the farnesyl group would abolish the formation of such a disulphide completely. However, the nucleotide hydrolysis assays were carried out in reducing conditions and so the disulphide is unlikely to be a significant factor. Furthermore, the largest difference in hydrolysis rates was seen between Cdc42Δ7 and Cdc42Δ3, neither of which contain the C-terminal Cysteine. Therefore, the effect of the C-terminal residues is evidently independent of this potential intramolecular disulphide.
6.5 An Interaction between the C-terminus and the G Domain

6.5.1 Assessing Changes in Secondary Structure using CD

Given the effect of the nucleotide state on accessibility of the C-terminus for
farnesylation and of the C-terminus on nucleotide hydrolysis, it seemed likely that the
C-terminus interacts with the G domain, even if only transiently. CD was next used to
determine whether the C-terminal region has a significant affect on the overall
secondary structure and thermal stability of Cdc42.

CD spectra were recorded on 0.1 mg/mL Cdc42Δ7Q61L, Cdc42Δ3Q61L and full
length Cdc42Q61L (WT C-terminus), all bound to GMPPNP, between 185 and 260
nm. The data recorded between 190 and 240 nm was fitted using DichroWeb
(Whitmore and Wallace 2008). The CD spectra are shown in Figure 6.16, along with
the reconstructed curves following data fitting. The data were fitted to two different
reference sets containing a mixture of α-helical and β-sheet proteins, as described in
Chapter 2, and the percentage helicity of each protein is given in Table 6.1. The data
fits were imperfect, reflected in high RMSDs between the raw and reconstructed
curves and the calculated helicities were higher than the values derived from the NMR
structure of Cdc42Δ7. The numbers should be considered with this in mind.
Figure 6.16. CD Analysis of Cdc42 Proteins. CD spectra recorded on A) Cdc42Δ7Q61L, B) Δ3Q61L or C) full length Cdc42Q61L. The spectra were recorded on 0.1 mg/mL protein between 185 and 190 nm and buffer only control data was subtracted from the experimental data. The data between 190 and 240 nm was fitted in DichroWeb as described in Chapter 2. The percentage helicity estimated from the data fits are given in Table 6.1, along with the normalised RMSDs as a measure of the fit quality. D) The unfitted data (after subtraction of the buffer only control) for each protein is shown overlaid.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Expected Helicity (%)</th>
<th>Estimated Helicity (%)</th>
<th>Normalised RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>D7</td>
<td>28.4</td>
<td>33.55</td>
<td>0.077</td>
</tr>
<tr>
<td>D3</td>
<td>27.8</td>
<td>35.35</td>
<td>0.081</td>
</tr>
<tr>
<td>FL</td>
<td>27.4</td>
<td>24.75</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 6.1. The Helicity of each Cdc42 Protein. A table showing the percentage helicity estimated in DichroWeb by comparison with reference α-helical datasets (details in Chapter 2). The estimated helicity is given along with the normalised RMSD, which constitutes a measure of the quality of the data fits. The expected helicity was calculated from the NMR structure of Cdc42Δ7, based on the assumption that the C-terminal region is unstructured.
The CD spectra of Cdc42Δ7Q61L and Cdc42Δ3Q61L were indistinguishable, indicating that the difference in hydrolysis between the two proteins is not explained by large changes in secondary structure (Figure 6.16D). However, the spectrum of full length Cdc42 indicated that it was less folded than the truncated proteins. This is reflected in the estimated helicities (Table 6.1). It therefore appears that the structure or stability of the protein is affected by the CAAX motif. The addition of 0.2 % OG to full length Cdc42 had no significant effect on the CD spectrum (Figure 6.17).

![Figure 6.17. The CD Spectrum of Full Length Cdc42 with OG.](image)

Next, the secondary structure was monitored using the CD signal at 222 nm at increasing temperatures up to 86 °C (Figure 6.18). The ellipticity at 222 nm is expected to increase as the structure unfolds and so can be used as a measure of thermal stability. Cdc42Δ3 and full length Cdc42 began to unfold at a lower temperature than Cdc42Δ7 (~56 °C compared to ~58 °C). Cdc42Δ3 underwent the most rapid transition and was completely unfolded at ~68 °C. Cdc42Δ7, however, was still unfolding at 86 °C and full length Cdc42 was not completely unfolded until ~74 °C. The melting temperatures, derived from the first order derivative of the melting curve, were 60 °C for Cdc42Δ3 and 68 °C for Δ7 and full length. The estimate of the melting temperature for Cdc42Δ7 is likely to be an underestimate as the curve did not reach saturation. Together, these data suggest that Cdc42Δ7 is the most stable and the representative of physiological full length (Cdc42Δ3) is the least stable.
Figure 6.18. Thermal Melting Curves for Cdc42 Proteins. The ellipticity at 222 nm was recorded on 0.1 mg/mL Cdc42Δ7Q61L (A), Cdc42Δ3Q61L (B) and full length Q61L (C) at 2 °C increments between 18 and 86 °C.
Altogether, the CD data indicate that the C-terminus does not induce significant unfolding of Cdc42 but that it does affect the thermal stability of the protein. This is indicative of more subtle structural or dynamic changes to the G domain in the presence of the C-terminus that cannot be readily observed using CD.

6.5.2 NMR

NMR was next used to investigate the effects of the C-terminus on the G domain in more detail. Cdc42Δ3Q61L is the most physiologically relevant representation of active Cdc42, and it was Cdc42Δ3Q61L that showed the largest difference in hydrolysis when compared with Δ7. 15N-labelled Cdc42Δ3Q61L was therefore prepared in the same way as 15N-Cdc42Δ7Q61L (Chapter 5). The nucleotide hydrolysis experiments showed that Cdc42Δ3Q61L hydrolysed GTP significantly within just a few hours and so nucleotide exchange for the slowly hydrolysing GTP-analogue, GMPPNP, was necessary to ensure the protein was locked in the active state throughout the NMR experiments. After purification and nucleotide exchange, 13 mg of protein was obtained from 3 L of E. coli.

A 15N-HSQC was recorded and compared with the 15N-HSQC of Cdc42Δ7 (Figure 6.19). Consistent with the CD data, the HSQCs were very similar, indicating no dramatic changes to the protein structure. There were, however, small chemical shift differences for some residues, indicated with arrows in Figure 6.19A. Some examples of shifted peaks are labelled in the enlargement of the central region of the HSQCs (Figure 6.19B).
Figure 6.19. $^{15}$N-HSQCs of Cdc42Δ7 and Cdc42Δ3. The HSQCs were recorded on 0.65 mM $^{15}$N-Cdc42Δ7 Q61L (black) and 0.6 mM Cdc42Δ3Q61L/C188S (orange) in 50 mM Tris:HCl pH 7.5 with 150 mM NaCl, 5 mM DTT, 5 mM MgCl$_2$ and 10 % D$_2$O. A) The full spectra are shown overlaid, with examples of shifted peaks marked with arrows. B) The central region of the spectra, outlined in A, is shown enlarged with examples of shifted peaks labelled with assignments. The C-terminus of Cdc42Δ3 is labelled, as well as two new peaks not seen in Δ7.
The chemical shift differences (CSD) between Cdc42Δ7 and Cdc42Δ3, calculated using the CSP equation from Chapter 5, are shown in Figure 6.20A. A value of zero indicates a residue for which no peak could be assigned and so no CSD calculated. The mean CSD was small (0.009) and 48 residues had a CSD greater than this. These residues are mapped onto the NMR structure of Cdc42Δ7Q61L (Mott & Owen, unpublished) in yellow (Figure 6.20B). 10 of these residues had a CSD greater than the mean plus the standard deviation and are marked in red in Figure 6.20B. 88 residues had CSDs smaller than the mean (cyan in Figure 6.20B) and the remaining residues were not assignable (grey in Figure 6.20B).

Residues with a CSD greater than the mean were found in every structural feature of Cdc42, from β1 through to α5, and both faces of the protein were affected. These differences were, however, very small, with the largest CSDs remaining below 0.09 ppm. The data indicate widespread, albeit subtle, conformational or dynamic differences between the two Cdc42 proteins. Therefore, the C-terminal region is most likely mobile rather than fixed down to a particular region. Were the C-terminus involved in stable contacts with a specific area of the protein, one might expect large CSDs localised to a single face of the protein.

The majority of the switch regions are invisible in both proteins but the residues packed behind and near to the switches are indicative of differences in the switches. For example, residues with CSDs greater than the mean include Leu19, Leu20 and Tyr24 of α1 and Leu160, Gln162 and Gly164 of the loop preceding α5, both of which pack behind Switch 1. Asn92, Val98, Glu100, Ile101, Thr102 and Cys105, which pack behind Switch 2, along with the only residue of Switch 1 that is visible (Ser30) and a residue of Switch 2 (Asp62) also had CSDs greater than the mean.
The largest CSDs were seen for Thr3, Ile4, Lys5, Lys16, Tyr51, Val98, Thr102, Glu140, Gly164 and Glu181. The affects on residues 3, 4 and 5 are unsurprising given the proximity of the N- and C-termini in the structure of Cdc42Δ7 (Figure 6.20B). Tyr51 lies in the loop region connecting the two switches and so may be affected as a
result of perturbations in the switches. Val98 and Thr102 are in α4, behind Switch 2, and Gly164 is packed behind Switch 1, as discussed above.

Perhaps the most interesting residue to be affected is Lys16. This is one of the critical nucleotide binding residues of the P-loop/G1, the sidechain amino group of which contacts the β-phosphate of the nucleotide (see Introduction). Furthermore, the non-bridging Oxygen atoms of the β-phosphate are hydrogen bonded to the backbone amides of Val14, Gly15 and Lys16 of Ras (Pai et al. 1990). These residues are conserved in Cdc42. Lys16 is therefore important in stabilising the negative charge that builds up on the β-phosphate in the GTP/GDP transition state (Figure 6.21). It also maintains contact with the β-phosphate of GDP following hydrolysis, after loss of the contacts between the invariant Thr35/Gly60 and the γ-phosphate (Li and Zhang 2004). The importance of this residue is further illustrated by its repositioning by GEFs, which contributes to nucleotide dissociation.

![Figure 6.21. Stabilisation of Negative Charges During GTP Hydrolysis.](image)

The CSD of this residue is larger than the mean plus one standard deviation but is still relatively small (0.019) and so the difference between this residue in Cdc42Δ7 vs Δ3 is likely to be subtle. Also, given the widespread changes discussed above, it is likely that the effect on this residue is indirect via widespread conformational or dynamic changes.
changes. Nonetheless, a subtle change in position of this residue such that it interacts more tightly with the β-phosphate could result in better stabilization of the developing negative charges on the transition state and increase the reactivity of the β-phosphate-γ-phosphate bond.

Many of the other key nucleotide binding residues were not affected or not visible in the HSQCs. For example, the invariant Thr35 and Gly60 that bind the γ-phosphate are not visible. Ser17, which is responsible for coordination of the magnesium ion and Asp119 and Ala146, which are involved in contacting the Guanine base, are not affected. However, the CSDs of Lys117 and Asp118, which are part of G4 and are involved in stabilising the position of G1, were above the mean. Overall, the data indicates widespread changes, most likely involving changes in dynamics, and implicates the P-loop/G1 as the transducer of the C-terminal affects on nucleotide hydrolysis.

The small, widespread changes are indicative of a dynamic C-terminus sampling multiple conformations, as discussed above. Changing the temperature would be expected to affect the mobility of the C-terminus and so alter its affects on the rest of the protein. Decreasing the temperature such that the C-terminus is less mobile could effectively trap the C-terminus in its most energetically favourable conformation. This conformation may involve a stable interaction between the C-terminus and a localised region of the G domain. If this were the case, the differences would be expected to be relatively large between Δ7 and Δ3 and seen in a more localised area. Conversely, increasing the temperature would increase the rate of conformational sampling by the C-terminus and so any differences between Δ7 and Δ3 might be expected to be smaller and more widespread.

HSQCs were recorded on Δ7 and Δ3 at 5, 10, 15, 30 and 35 °C. The spectra of Δ7 and Δ3 were compared at 5 °C and at 35 °C and the CSDs calculated as with the spectra at 25 °C (Figure 6.22). The CSDs between Δ7 and Δ3 at each temperature were also mapped onto the structure of Cdc42Δ7 for comparison (Figure 6.23). The mapping was generally similar at the different temperatures, with some interesting differences.
Figure 6.22. Chemical Shift Differences at Different Temperatures. Bar charts showing the CSD for each residue at 5 °C (blue), 25 °C (black) and 35 °C (red). A CSD value of zero indicates a residue for which no CSD could be calculated due to spectral overlap or absence of the residue from either spectrum. Regions of particular interest, as described in the text, are circled and labelled.
The CSDs of Thr3 and Ile4 between the two Cdc42 proteins is greatest at 35 °C and least at 5 °C. This implies that the increased mobility of the C-terminus perturbs the N-terminal region more. This is unsurprising given the proximity of the C- and N-termini. This was also the case for Val42 and Val44 (circled in green in Figures 6.22 and 6.23), which lie adjacent to Switch 1, suggesting that increased C-terminal mobility leads to greater perturbations in Switch 1. One possible explanation for this is that the C-terminal region does not affect Switch 1 via a direct interaction but rather its ‘movability’ affects the conformation or dynamics of Switch 1. It is also possible that the C-terminal region affects Switch 1 by increasing its flexibility and motion such that its effect is masked when the motion of Switch 1 is already increased at the higher temperature.

Conversely, for other affected residues the largest effect was seen at 5 °C. For example, the helix behind Switch 2 (α3, circled in red in Figures 6.22 and 6.23) is more affected at the lower temperature. Tyr52, which lies in between Switch 1 and 2, and Phe78, which is packed behind Switch 2, are also most affected at 5 °C. This suggests that a decrease in C-terminal flexibility leads to greater perturbation of Switch 2. Gly15 and Lys16 of the P-loop are also most affected at 5 °C (circled in blue in Figures 6.22 and 6.23).

Figure 6.23. Chemical Shift Changes Between Cdc42Δ7 and Cdc42Δ3 at Different Temperatures. The CSDs at each temperature are mapped onto the NMR structure of Cdc42Δ7. Residues with CSDs smaller than the mean are shown in cyan, greater than the mean in yellow and greater than the mean plus one standard deviation in red. Residues with no calculated CSD are in grey. Regions of particular interest as described in the text are circled and labelled.
One possible explanation for the bigger effects at lower temperatures is that, when the C-terminus is moving less, it interacts with this area of the G domain more stably and so affects Switch 2 and the P-loop more significantly. Gly15/Lys16 and Switch 2 are coloured orange and red, respectively, in Figure 6.24, showing that they fall on one face of Cdc42. It is therefore possible that the C-terminus wraps around the front of Cdc42, contacting Switch 2 and so affecting Switch 2 and the P-loop.

![Figure 6.24. Switch 2 and the P-Loop. A surface representation of the NMR structure of Cdc42Δ7. Switch 2 is shown in red and Gly15 and Lys16 are shown in orange. The C-terminus is shown but its position in the structure is actually not well defined as it is highly flexible.](image)

**6.6 Membrane Anchoring of Cdc42**

The physiological relevance of the findings from these experiments is unknown and so NMR studies of lipid modified, membrane anchored protein is still required. Before NMR experiments could be pursued, it was necessary to optimise membrane anchoring of the lipid modified protein. The farnesylation protocol was not, so far, reproducible and so labelled protein for NMR studies has not been made. Nonetheless, preliminary membrane binding studies could be performed.
The farnesylated protein (made in section 6.3.1.3) was exchanged for GMPPNP prior to membrane binding assays to ensure a fully active sample. Liposomes composed of phosphatidylserine (PS) and phosphatidylcholine (PC), with and without PI(4,5)P₂, were then prepared and incubated with Cdc42Δ7Q61L, Cdc42Δ3Q61L and full length, farnesylated Cdc42Q61L/L191A. The liposomes were pelleted by centrifugation and samples were taken from the supernatant and pellet and analysed by SDS-PAGE, as in the TOCA1 liposome sedimentation assays (Chapter 3). Cdc42Δ7Q61L was included as a negative control and Cdc42Δ3Q61L was included to determine whether the polybasic C-terminal region can interact with the membrane in the absence of lipid modification. For example, the polybasic region could interact with negative lipids such as PS or PI(4,5)P₂. Controls without liposomes were included for each protein.

Figure 6.25 shows the first experiment. For the ‘no liposome’ controls (lanes 3 and 4 in Figure 6.25A and lanes 3, 4, 10 and 11 in 6.25B), the majority of the proteins were in the supernatant with only a trace in the pellet. For Cdc42Δ7Q61L, there was no more protein in the pellet in the presence of liposomes than was seen in the no liposome control (compare lane 4 with lane 6 in Figure 6.25A), indicating that it does not bind to liposomes. The addition of PI(4,5)P₂ to the liposomes gave the same result (lane 8). Cdc42Δ3Q61L also did not pellet with the PC/PS (compare lane 4 with lane 6 in Figure 6.25B) or PC/PS/PI(4,5)P₂ (lane 8) liposomes. For farnesylated Cdc42, there was more protein in the pellet in the presence of PC/PS liposomes than in the no liposome control (compare lane 11 with lanes 13), indicating that some of the protein was localised to liposomes. The amount of protein localised to liposomes was comparable for the PC/PS and PC/PS/PI(4,5)P₂ liposomes (compare lane 13 and 15), indicating that PI(4,5)P₂ has no effect on membrane binding.
Only a small amount of the farnesylated Cdc42 bound to the liposomes. This could be explained by a number of factors: for example, the liposomes may have been of poor quality and/or the liposomes may have been saturated with Cdc42. Further experiments were therefore performed including the TOCA1 F-BAR domain as a positive control for liposome quality and with two concentrations of liposome to assess the affect of liposome quantity on Cdc42 binding (Figure 6.26).

**Figure 6.25. Liposome Binding of Farnesylated Cdc42.** 15 % Laemmli gels stained in InstantBlue (Expedeon), showing samples taken from liposome spin experiments with Cdc42Δ7 (A) and Cdc42Δ3 and full length farnesylated Cdc42 (B). An input sample (In) is shown for each protein followed by samples from the supernatant (SN) and pellet (P) following centrifugation after incubation without liposomes (-) or with PC/PS or PC/PS/Pl(4,5)P₂ liposomes. The final lipid concentration was 0.3 mg/mL and the final protein concentration was 0.2 mg/mL.
The F-BAR domain localised to liposomes (compare the no liposome control in lane 4 with lane 6 in Figure 6.26A), as was seen in Chapter 3. The localisation of the protein

Figure 6.26. Liposome Binding of Farnesylated Cdc42 with Two Liposome Concentrations. 15 % Laemmli gels showing samples taken from liposome spin experiments with the TOCA1 F-BAR domain (A), Cdc42Δ3 (B) and full length farnesylated Cdc42 (C). An input sample (In) is shown for each protein followed by samples from the supernatant (SN) and pellet (P) following centrifugation after incubation without liposomes (-) or with liposomes as indicated. The final lipid concentration was 0.3 mg/mL (1x) or 0.6 mg/mL (2x). The final protein concentration was 0.2 mg/mL. Gels were stained in InstantBlue (Expedeon).
to the liposomes was not increased at the higher concentration of liposomes (compare lanes 6 and 8) but was increased with PI(4,5)P₂ (lane 10). This positive control confirmed that the liposomes were correctly made.

Cdc42Δ3Q61L did not localise to the liposomes, even at higher liposome concentrations (Figure 6.26B). The farnesylated protein was mostly in the supernatant with a small amount visible in the pellet (compare the control in lane 4 with lane 6 in Figure 6.26C) and PI(4,5)P₂ did not increase binding (compare lane 6 with lane 10). Doubling the concentration of liposomes did not increase the amount of farnesylated protein in the pellet (compare lane 6 with lane 8). This indicated that the amount of liposomes is not the factor causing the limited binding.

The limited binding observed could be due to occlusion of the farnesyl group, a requirement for a different liposome composition or size not tested here, or another unknown factor. For example, Rho proteins are delivered to the membrane by regulatory proteins \textit{in vivo}, which may aid in membrane insertion. Alternatively, the nucleotide state may affect membrane association and, given the effect of GMPPNP on farnesylation of the C-terminus, GMPPNP may also have a negative effect on membrane anchoring via the modified C-terminus.

### 6.7 Conclusions

#### 6.7.1 Cdc42 is Monomeric in Solution

The AUC data presented in this chapter indicates that Cdc42 is monomeric in solution at concentrations as high as 0.33 mM. This contradicts previous reports that Cdc42 dimerises with a $K_d$ of ~2 µM (Zhang and Zheng 1998). The reasons for the contradiction are not clear but may relate to the different methods used to assess the oligomeric state. Zhang and Zheng used gel filtration and, as the concentrations of Cdc42 used were low (0.005-5 µM), the peaks were very small. Furthermore, the two peaks were not well resolved and so their quantitation was problematic. Assessment of their chromatograms by eye suggests that both the monomer and dimer peaks vary in size at different concentrations but that the ratio between the two peaks does not significantly change.
In our hands, many small G proteins elute in two peaks in gel filtration similar to the two peaks seen by Zhang and Zheng. However, the relative size of the two peaks does not correlate with concentration. In Figures 6.1D and 6.8, the gel filtration traces contain two large peaks (void and main Cdc42 peak), with a very small peak in between. It is this peak that is reported to represent dimeric Cdc42 but, as the ratio between these peaks does not correlate with protein concentration, we assume that this small peak results from unpredictable protein aggregation in solution rather than functional dimerisation. Consistent with this idea, the concentration and freshness of the reducing agent added affects the aggregation of Cdc42 in our hands, presumably due to intermolecular disulphide formation between exposed Cysteine residues in an oxidizing environment. Cdc42 has a number of exposed Cysteine residues including Cys105 and Cys188 and so the full length protein has an additional exposed Cysteine compared to Cdc42Δ7. This could explain the differences in ‘oligomerisation’ observed between the full length and truncated proteins (Zhang and Zheng 1998).

In addition to the gel filtration experiments, Zhang and Zheng were able to pull down Cdc42 from cell lysates using exogenously expressed, activated GST-Cdc42. This data, however, also does not confirm dimerisation. A pull down from cell lysates with activated Cdc42 would be expected to bring down Cdc42 effector proteins, some of which oligomerize and cluster at membranes. These could act as a bridge between multiple Cdc42 molecules. For example, activated Cdc42 co-localises with clusters of N-WASP and TOCA1 in filopodia signalling (Miki et al. 1998; Lee et al. 2010). Therefore, it is likely that a Cdc42 pull down would result in indirect pull down of endogenous Cdc42 via effector clusters.

Perhaps harder to explain is the observation that the rate of nucleotide hydrolysis by Cdc42 is concentration dependent, which they assume is explained by dimerisation. Furthermore, they saw that the removal of the 7 C-terminal residues, which they claimed to be required for dimerisation, abolished the effect. The effect of the C-terminus on nucleotide hydrolysis shown within this chapter somewhat complicates their results.
Whilst the AUC data presented here does not rule out dimerisation \textit{in vivo} or following membrane localisation, it does appear to directly contradict the in solution data from Zhang and Zheng. The buffer conditions used here match the buffers used by Zhang and Zheng, with the exception of the higher concentration of reducing agent used here (5 mM compared to 1 mM). It is therefore not easy to reconcile the two reports without refuting one or the other. There have been no further reports of Cdc42 dimerisation since and AUC is a highly reliable method for detecting oligomeric species in solution (Balbo and Schuck 2005), and is certainly more trustworthy than gel filtration. Consistent with these data is an extensive array of NMR studies that have been performed on Cdc42 within the Mott/Owen group. These studies have always been consistent with a protein of \~20 kDa and have not indicated exchange between oligomeric forms.

6.7.2 The C-terminus Affects the G Domain

The data presented here showed that the C-terminus of Cdc42 affects the rate of GTP hydrolysis and that farnesylation of the C-terminal Cysteine does not abolish the effect. The difference that was observed between full length L191A and L191M is difficult to explain but, nonetheless, the difference between Cdc42Δ7 and physiological full length (Δ3) was clear, suggesting an interaction between the C-terminus and the G domain. NMR experiments revealed subtle, widespread differences between Cdc42Δ7 and Cdc42Δ3, which implied that the C-terminus interacts with the G domain but that it is dynamic, sampling multiple conformations. Most interestingly, the chemical shift of Lys16 of the P-loop was significantly different in the two proteins, perhaps implicating it in the effect on GTP hydrolysis. The temperature titration experiments were consistent with the idea of a mobile C-terminus and that Switch 2 and the P-loop are involved in communication between the nucleotide and the C-terminus.

6.7.3 Membrane Anchoring of Cdc42 for Structural Studies

The farnesylation by co-expression or \textit{in vitro} reactions were low yield and highly variable. The overriding problem was that the protein was rarely 100 \% farnesylated, but rather \~50-60 \% farnesylation was most commonly seen. The attempts to separate
the two species were not successful and the reaction also showed some nucleotide
dependence, as the GMPPNP bound protein was not farnesylated to any detectable
level whilst the GDP bound protein was farnesylated at 60-90 %. The farnesylated
protein that was successfully purified showed only a weak affinity for liposomes,
which could be due to occlusion of the farnesyl group resulting from interactions
between the C-terminal region and the G domain, the size and/or composition of the
liposomes being incompatible with Cdc42 binding, or a requirement for the
physiological geranylgeranyl modification rather than farnesyl.

The implications of the results presented here are discussed further in the following
chapter, along with possible future directions.
7 Wider Conclusions and Future Directions

The overarching aim of this project was to provide much needed biochemical, structural and biophysical insight into the pathway of Cdc42/N-WASP/TOCA1-dependent actin assembly. Specifically, the project aimed to: measure the affinity of the Cdc42-TOCA1 interaction, determine the minimal region of TOCA1 required for maximal affinity Cdc42 binding and to investigate the possible mechanisms of TOCA1 autoregulation (Chapter 3); to solve the solution NMR structure of the minimal Cdc42 binding region of TOCA1 (Chapter 4); to characterise the Cdc42-TOCA1 interaction structurally in order to address questions surrounding GTPase-effector specificity and to determine whether TOCA1 and N-WASP can bind to a single molecule of Cdc42 simultaneously (Chapter 5); and to work towards structural studies of full length Cdc42 on membrane mimetics (Chapter 6).

These initial aims have been addressed and the conclusions have been discussed within Chapters 3-6. Briefly, Chapter 3 provided an affinity for Cdc42-TOCA1 binding and concluded that the HR1 domain was sufficient for maximal affinity binding. Chapter 4 presented the solution NMR structure of the TOCA1 HR1 domain and highlighted some interesting structural differences between the TOCA and PRK family HR1 domains. Chapter 5 presented a model of the Cdc42-TOCA1 complex and a structural analysis of the Cdc42-TOCA1 binding interface, highlighting some possible sequence specific contacts and offering insight into the differential affinities and specificities of G protein-HR1 binding. Further, it revealed that TOCA1 and N-WASP do not bind to a single molecule of Cdc42 simultaneously but rather that N-WASP outcompetes TOCA1 for Cdc42 binding. Chapter 6 provided insight into how the C-terminus of Cdc42 affects its structure and function and made progress towards structural studies on membrane mimetics.

How these individual conclusions relate to the pathways of actin assembly has not yet been discussed. Therefore, this chapter aims to consolidate the results and relate them to wider biological conclusions about actin pathways, as well as presenting questions that have arisen from the results and possible future directions.
7.1 Biochemical and Structural Insight into the Pathways of TOCA1-dependent Actin Assembly

The low affinity of the Cdc42-HR1\textsuperscript{TOCA1} interaction is consistent with a tightly spatially and temporally regulated pathway, with combinatorial signals that lead to a series of coincident weak interactions as discussed in Chapter 3. The low affinity of the Cdc42-TOCA1 interaction (~5 µM), together with the estimate of the cellular concentration of TOCA1 (5-10 nM) (Ho et al. 2004), indicates that the two proteins are only likely to interact when they are co-localized at the membrane. Previous evidence has indicated that the F-BAR domains of TOCA family proteins can bind to membranes in the absence of other domains (Tsujita et al. 2006; Bu et al. 2009) and the liposome spin assays presented in Chapter 3 also showed binding of the isolated F-BAR domain to liposomes with modest affinity.

The pathway of Cdc42/N-WASP/Arp2/3-dependent actin assembly is evidently complex, with multidirectional, interlinked equilibria relying upon additive, synergistic and cooperative effects (reviewed in Prehoda and Lim 2002; Padrick and Rosen 2010). The data presented within Chapters 3-5 pertaining to TOCA1-Cdc42 binding is consistent with this current understanding. While this is difficult to portray in a simple schematic, by considering the model of the Cdc42-HR1\textsuperscript{TOCA1} complex and the NMR experiments alongside available \textit{in vitro} binding data, it is possible to construct a scheme, which also incorporates the available structural information on WASP/N-WASP (Figure 7.1). This scheme describes how the early stages of Cdc42-dependent actin nucleation may proceed \textit{via} N-WASP and TOCA1 and is explained in more detail in the following section.
Figure 7.1. A Simplified Model of Cdc42-dependent Actin Assembly.

Step 1: TOCA1 is represented by the structure of the CIP4 F-BAR domain (Shimada et al. 2007), the TOCA1 HR1 domain (presented in Chapter 4) and the CIP4 SH3 domain (PDB code: 2CT4, Miyamoto et al. unpublished), connected by flexible linkers. Cell signalling through PI(4,5)P$_2$ pushes the equilibrium in favour of membrane-localized TOCA1 and dimerization via the F-BAR domain.

Step 2: Cdc42-GTP, when coincidently activated in response to extracellular signals, binds to TOCA1. The model of the Cdc42-HR1 complex is shown here. The majority of the TOCA1 F-BAR domain is omitted for clarity.

Step 3: TOCA1 is poised to activate N-WASP. Autoinhibited N-WASP is represented by the structure of the GBD with the cofillin homology region (Kim et al. 2000), with the WH1 (box) and VCA (oval) connected via flexible linkers. The WH1 domain would be bound to WIP but WIP is omitted for clarity. Binding of the TOCA1 SH3 domain to the polyproline region (PP) of N-WASP positions N-WASP for Cdc42 and PI(4,5)P$_2$ binding, allowing initial interactions of the N-terminal extension (blue crosses) and core GBD (red, unstructured) with Cdc42.

Steps 4 and 5: The initial interactions between N-WASP and Cdc42 rapidly push the equilibrium in favour of the fully unfolded, high affinity conformation such that the C-terminal region of the GBD can now bind, displacing the TOCA1 HR1 domain. The Cdc42-GBD complex is represented by the Cdc42-GBD$_{WASP}$ structure (Abdul-Manan et al. 1999). The VCA domains of the clustered N-WASP molecules are now free for robust activation of actin polymerization via the Arp2/3 complex. The steps are linked with single-headed arrows for clarity but are, in reality, expected to be linked via equilibria. Furthermore, single molecules of TOCA1 and N-WASP are shown but it is expected that both cluster at the membrane.
7.1.1 Expansion of the Model of Cdc42-dependent Actin Assembly

The Cdc42/TOCA1/N-WASP pathway is depicted as stepwise in the simplified model presented in Figure 7.1. This is an over simplification and rather a complex network of interconnected equilibria is expected, relying upon an intricate and layered series of low affinity interactions reminiscent of those described in relation to clathrin-dependent endocytosis (Praefcke et al. 2004; Höning et al. 2005). Much like endocytic signaling pathways, this connected interplay of protein-protein and protein-lipid interactions would serve to prevent misfiring of downstream effects, which in this case is actin polymerisation.

The data from this project, combined with previous understanding of TOCA proteins, suggests that recruitment of TOCA1 to membranes in response to PI(4,5)P_2 involves multiple coincident interactions, each of modest affinity. Firstly, the affinity of the F-BAR domain for membranes appears to be ~2 µM (Chapters 3 and 6). Secondly, the affinity of TOCA1 homodimerization via the F-BAR domain is likely to be comparable to that of the FCHo2 F-BAR domain (2.5 µM) (Henne et al. 2007). Step 1 of Figure 7.1 is therefore expected to involve interlinked equilibria between monomeric and dimeric and cytosolic and membrane-bound TOCA1.

The coincident TOCA1-membrane and TOCA1-TOCA1 binding could lead to clustering at the membrane, as has been observed for FBP17 (Tsujita et al. 2006). The clustering of TOCA1 would increase its local concentration in excess of its affinity for Cdc42 and so this interaction would be expected to occur provided Cdc42 has been activated in response to co-incident signals. This interaction may further stabilise TOCA1 at the membrane and localise it to areas rich in active Cdc42. Step 2 of Figure 7.1 is therefore expected to comprise interlinked equilibria between free and Cdc42-bound TOCA1 and inactive and active Cdc42.

The activation of Cdc42 is also expected to be regulated by interlinked equilibria involving membrane binding, nucleotide hydrolysis and regulatory proteins such as GEFs and GAPs. The effect of the C-terminus on nucleotide hydrolysis (Chapter 6), i.e. ‘switching off’ of Cdc42, suggests that it (and in turn its interaction with membranes) may affect the equilibrium between active and inactive Cdc42. Earlier
data suggested that dimerisation of Cdc42 could also represent an additional level of regulation (Zhang and Zheng 1998), but the data presented in Chapter 6 suggest that dimerisation is unlikely.

Following Cdc42-binding, TOCA1 is poised to recruit N-WASP to the membrane (Takano et al. 2008) and near to Cdc42 via the interaction between its SH3 domain and the N-WASP proline-rich region (Ho et al. 2004). The lack of autoinhibitory interactions in TOCA1 (Chapter 3) suggests that Cdc42 and N-WASP do not activate TOCA1 allosterically. Conversely, N-WASP is regulated by intramolecular interactions. TOCA1 may contribute to N-WASP activation by opposing the GBD-VCA interaction, similar to the effect seen with Cdc42. WIP, which is recruited by TOCA1 to membranes along with N-WASP, inhibits the activation of N-WASP by Cdc42 and this effect is reversed by TOCA1 (Ho et al. 2004). It may be envisaged that WIP and TOCA1 exert opposing allosteric effects on N-WASP, with TOCA1 favouring the unfolded, active conformation of N-WASP. This would be expected to increase its affinity for Cdc42 and allow downstream activation of the Arp2/3 complex. Recruitment and allosteric activation of N-WASP by TOCA1 is illustrated in steps 3 and 4 of Figure 7.1.

In addition, TOCA1 clustering at the membrane followed by TOCA1-dependent recruitment of N-WASP would, in turn, be expected to lead to N-WASP clustering. The recruitment of N-WASP alone and of the N-WASP/WIP complex by TOCA1 and FBP17 has been demonstrated (Takano et al. 2008) and N-WASP clustering has been observed on supported lipid bilayers following TOCA1 clustering (Lee et al. 2010). TOCA1 may therefore activate N-WASP by clustering it close to its allosteric activators and also by effective oligomerization, which is important for WASP/N-WASP activation (reviewed in Padrick and Rosen 2010). Steps 1-3 of Figure 7.1 are expected to involve clustering of TOCA1 and N-WASP in addition to the allosteric regulation that is illustrated.

Step 4 illustrates that the positioning of (N-)WASP by TOCA1, along with a potential allosteric effect, allows the (N-)WASP N-terminal extension followed by the core GBD to contact Cdc42. These intermediate states would be short-lived and sparsely populated, as the folded conformation and the unfolded, high affinity conformation of
(N-)WASP would each be highly favoured over the transient states. The high affinity of the unfolded GBD for Cdc42 would ensure that the unfolded state is ultimately favoured. Thus, these initial GBD-Cdc42 contacts rapidly push the equilibrium in favour of unfolded (N-)WASP, allowing the full GBD to contact Cdc42.

It had previously been postulated that TOCA1 and N-WASP bind to Cdc42 simultaneously, and some data exists to back up this idea (Ho et al. 2004; Bu et al. 2010). The data presented in Chapter 5, however, indicate that, instead of forming a ternary complex, N-WASP outcompetes TOCA1 for Cdc42 binding. Full-length TOCA1 and N-WASP have similar affinities for active Cdc42 but in the unfolded, active conformation, the affinity of N-WASP for Cdc42 dramatically increases and the GBD	extsuperscript{N-WASP}-Cdc42 interaction will be favoured.

In such an array of molecules localised to a discrete region of the membrane, it is plausible that WASP could bind to a second Cdc42 molecule, rather than displacing TOCA1 from its associated Cdc42 but the NMR, pyrene actin and affinity data presented in Chapter 5 are consistent with displacement of HR1	extsuperscript{TOCA1} by GBD	extsuperscript{N-WASP}. Furthermore, TOCA1 is required for Cdc42-mediated activation of N-WASP/WIP (Ho et al. 2004), implying that it may not be possible for Cdc42 to bind to and activate N-WASP prior to TOCA1-Cdc42 binding. The commonly used MGD→IST (Cdc42-binding deficient) mutant of TOCA1 has a reduced ability to activate the N-WASP/WIP complex (Takano et al. 2008), further indicating the importance of the Cdc42-HR1	extsuperscript{TOCA1} interaction prior to downstream activation of N-WASP.

In light of this, an ‘effector handover’ scheme is favoured whereby TOCA1 interacts with Cdc42 prior to N-WASP activation, after which N-WASP displaces TOCA1 from its bound Cdc42 in order to be fully activated. There is an advantage to such an ‘effector handover’, since N-WASP would only be robustly activated when F-BAR domains are present. Hence, actin polymerisation cannot occur until F-BAR domains are localised and positioned for membrane distortion. The 100x higher affinity of GBD	extsuperscript{N-WASP} ensures unidirectional competition with HR1	extsuperscript{TOCA1} for Cdc42-GTP, as observed in the NMR experiments with the GBD (in its high affinity state). An irreversible step is therefore implicit in the schematic, ensuring that the pathway is unidirectional. The displacement of the TOCA1 (or the FBP17 or CIP4) HR1 domain
from Cdc42 could precede the displacement of these F-BAR proteins from the membrane. A negative feedback loop has been observed for FBP17, whereby FBP17-induced membrane tension leads to displacement of FBP17 from the leading edge of the cell (Tsujita et al. 2015) and such a process could be dependent upon the earlier disruption of the HR1-Cdc42 interaction.

The comparison between the Cdc42-GBD\textsuperscript{WASP} structure and the model of the Cdc42-HR1\textsuperscript{TOCA1} complex indicated a mechanism by which an effector handover could take place, which was explained in Chapter 5. This handover mechanism is depicted as stepwise in the simplified model presented in Figure 7.1. This is an over simplification in order to illustrate the molecular mechanism of effector handover. Rather, the handover is expected to involve equilibria between folded and extended and free and bound proteins, much like the rest of the pathway.

This project presents the first description of a small GTPase effector handover, which relies upon specific membrane localization, directional recruitment of tightly regulated signalling proteins, interrelated binding equilibria and partially overlapping binding sites. Interactions with multiple effectors is a general feature across the different families of small GTPases (reviewed in Mott and Owen 2015) and so directional effector handovers may also occur in other GTPase signalling pathways.

### 7.2 Future Directions

#### 7.2.1 Determinants of Binding Specificity and Affinity

Some conclusions have been made relating to the binding specificities and affinities of HR1 domains and some potential sequence-specific contacts and affinity determinants have been suggested. Mutagenesis studies to confirm the importance of these residues would be desirable. For example, an assessment of the importance of the possible electrostatic contact between Glu395\textsuperscript{TOCA1} and Arg68/Arg66\textsuperscript{Cdc42} would be interesting. ‘Residue swapping’ experiments to investigate whether the addition of TOCA1-Cdc42 specific contacts to another HR1 domain (or vice versa) would also be of interest.
Competition SPAs were used to study the interaction quantitatively and could be used to assess the affects of mutations in HR1\textsuperscript{TOCA1}. However, each mutant would be required in high concentrations (0.2 mL of at least 0.2 mM of each protein would be required, more for the lower affinity binders). In theory, Cdc42 mutants could also be assessed but only the Cdc42 mutants that do not significantly affect binding to the ACK GBD could be used in the competition SPAs. Therefore, quantitative analysis of an exhaustive set of mutants was not readily achievable using SPAs within the time constraints of this project and, moreover, the interaction could not be quantitatively studied using alternative methods such as SPA, ITC or BLI. Fluorescence polarization assays may be possible using a fluorophore attached to HR1\textsuperscript{TOCA1}.

7.2.2 Confirmation of the Effector Handover
If time would permit, a set of shorter N-WASP GBD fragments would be produced omitting residues from the C-terminal extension of the GBD. These would be assayed for Cdc42 binding in the presence of HR1\textsuperscript{TOCA1} in order to test the effector handover hypothesis. Should the hypothesis hold true, the shorter N-WASP fragments may bind in the presence of the TOCA1 HR1 domain without causing its displacement.

7.2.3 Structure and Function of Cdc42
Cdc42 was shown in Chapter 6 to be monomeric in solution but it would be prudent to carry out further AUC experiments with lipid modified Cdc42 in future, and perhaps also to design NMR experiments to confirm that Cdc42 is monomeric when bound to membranes. It was also shown in Chapter 6 that the C-terminal region of Cdc42, specifically the three residues upstream of the CAAX motif, somehow interacts with the G domain, affecting nucleotide hydrolysis and subtly affecting its structure. It would be interesting to repeat the nucleotide hydrolysis experiments with farnesylated protein in the presence of membrane mimetics, as well as with wild type proteins rather than Q61L. Moreover, NMR studies with lipid modified and membrane anchored protein are desired as the membrane might be expected to affect a C-terminus-G domain interaction. Given the nucleotide-dependence of the farnesylation reaction, it would also be interesting to study the GDP bound proteins in order to assess the effect of the nucleotide on the C-terminus.
7.2.4 The C-terminus and Effector Binding

The C-terminal region of Cdc42 had widespread effects on the G domain (Chapter 6) and among the most affected residues were Lys5 and Tyr51. The NMR data also indicated effects on the switches, in particular Switch 2. Lys5 is an effector binding residue, which was shown to make electrostatic contacts with Asp365 and Asp368 of TOCA1 and Tyr51 is adjacent to Thr52, which forms a possible Cdc42-TOCA1 specific contact with Lys372 TOCA1. A number of Switch 2-HR1 contacts were also identified in the Cdc42-HR1 TOCA1 model (Chapter 5). The C-terminus may, therefore, have subtle effects on effector binding, despite the apparently similar affinities of full length and truncated Cdc42 for TOCA1 (Chapter 3). These data further emphasize the need for structural studies of effectors with full length GTPases and, as the C-terminus is membrane-associated in vivo, with membrane-anchored GTPases.

7.2.5 Membrane-anchoring of Cdc42

Farnesylated Cdc42 was purified during this project but the farnesylation reaction was not reproducible and was rarely complete. The inclusion of membrane mimetics during the farnesylation reaction could stabilise the modified protein and also increase the dissociation of farnesylated protein from the FTase to increase the reaction efficiency. Further attempts to separate modified from unmodified protein could also now be made as, in particular, the ion exchange was not pursued exhaustively. The reaction also appeared to show some nucleotide dependence and so further farnesylation attempts should focus on the GDP bound form and, once farnesylated, the GDP can later be exchanged for other nucleotides. The Q61L protein maintains hydrolysis activity and so a non-hydrolysable GTP-analogue will need to be used to investigate the active form by NMR.

The farnesylated protein that was successfully purified showed only a weak affinity for liposomes, which could be due to occlusion of the farnesyl group. The inclusion of detergents prior to liposome binding may help to expose the lipid for membrane insertion. Alternatively, the low affinity could be due to the size and/or composition of the liposomes being incompatible with Cdc42 binding. NMR studies of membrane proteins with membrane mimetics can require extensive optimisation of the membrane mimetics, something that was not possible during the timeframe of this project. Time
permitting, a range of membrane mimetics would have been trialed, including size filtered liposomes, bicelles and nanodiscs. Nanodiscs better stabilise membrane proteins in the membrane environment (Kaptein and Wagner 2015) and so may also be useful for membrane associated G proteins.

*In vivo*, Cdc42 is geranylgeranylated and so this longer, more hydrophobic lipid could be required to achieve stable membrane anchoring. Farnesylated proteins such as Ras require additional modifications for membrane anchoring (Güldenhaupt et al. 2008; Roberts et al. 2008; Brunsveld et al. 2009) and so farnesylation may not be enough. Nonetheless, farnesylated Cdc42 was purified and showed some propensity to bind membranes and so this project can be built upon to produce stably membrane anchored protein for future NMR studies. For example, the GGTase could be purified and *in vitro* geranylgeranylation reactions carried out based on the methods used for FTase. Detergents may be needed to stabilise the geranylgeranylated protein.

Alternative methods to lipid modify Cdc42 could also be explored. For example, chemical attachment of Cdc42 to membrane mimetics containing maleamide lipids (Bader et al. 2000) could address both the problem of incomplete farnesylation and weak attachment to membranes. However, in order to carry out chemical linkage via the C-terminal Cysteine, all of the other exposed Cysteines in Cdc42 would first need to be mutated to ensure only the C-terminal Cysteine is able to react. With this in mind, other small G proteins containing fewer Cysteine residues may present better substrates for method optimisation and initial experiments.

### 7.2.6 Final Remarks

Overall, this project has identified some interesting effects of the C-terminus on the structure and function of Cdc42 and so highlighted the shortcomings of all *in vitro* studies on truncated GTPases. In the future, structural and functional studies of full length GTPases are desirable and methods to reliably produce lipid modified, membrane anchored GTPases for structural studies will be eagerly awaited. The Cdc42 and TOCA1 studies have also highlighted a need for methods to visualise the mobile regions of small G proteins, in particular Switch 1 and Switch 2, which are typically invisible in NMR spectra.
Appendices
Appendix 1: Mass Spectrometry Analysis of FBP17

A

Maldi analysis. Sample detail: C18+1, elute 50% MeCN/5% TFA. DHAP matrix. Exactive Myo-CyC5 5-25K m/z. 220/cm, 750-820 pv, 2000/1TFA. Initial sample

Peaks represent multiple charge states of component protein(s); number = charge state. Mass (Da) = (m/z value x charge state) – charge state

Different pv settings influence resolution/calibration in different m/z ranges

B

Maldi-ISD analysis. Zt desalt, elute sinapinic to thin layer. 190/lsr, 2080/pv, 2500sdet, extcal PEG

This technique creates a partial sequence ladder close to the N-terminal region and (sometimes) C-terminal region. Useful to check identity and likely start against a known sequence (when it works)
Appendix 2: Dihedral Angle Predictions from TALOS-N
Residue
323 G
324 P
325 L
326 G
327 S
328 H
329 M
330 K
331 G
332 P
333 A
334 L
335 E
336 D
337 F
338 S
339 H
340 L
341 P
342 P
343 E
344 Q
345 R
346 R
347 K
348 R
349 L
350 Q
351 Q
352 R
353 I
354 D
355 E
356 L
357 S
358 R
359 E
360 L
361 Q
362 K
363 E
364 M
365 D
366 Q
367 K
368 D
369 A
370 L
371 N
372 K
373 M
374 K
375 D
376 V
377 Y
378 E
379 K
380 N

Phi (˚)
9999.000
-62.144
-64.145
-100.484
-74.830
-80.358
-75.259
-76.435
130.122
-64.811
86.650
-73.109
-69.667
-81.741
56.363
-66.475
-87.738
-85.137
-60.854
-52.143
-65.654
-71.727
-63.907
-63.088
-65.880
-66.806
-65.027
-64.528
-65.725
-66.685
-62.400
-64.240
-64.534
-64.615
-62.689
-65.993
-65.225
-63.581
-63.894
-66.037
-66.739
-63.854
-64.868
-65.587
-63.593
-64.082
-65.700
-64.492
-65.195
-66.512
-64.929
-61.606
-65.062
-64.462
-65.100
-67.081
-87.601
-129.437

Psi (˚)
9999.000
144.827
138.370
170.783
-23.071
-20.275
-26.845
-27.989
-178.403
147.733
9.819
140.345
-26.845
-15.731
37.870
-23.821
-7.206
144.062
143.735
-39.215
-36.660
-39.621
-42.203
-41.396
-42.592
-38.633
-43.076
-37.824
-40.120
-41.422
-44.158
-42.739
-41.549
-43.333
-40.627
-42.168
-41.812
-43.239
-39.904
-41.306
-41.039
-42.184
-40.878
-40.190
-40.848
-44.042
-40.384
-43.400
-38.690
-40.584
-42.256
-42.332
-40.046
-45.615
-37.701
-33.927
-23.469
74.156

Standard
Deviation
(Phi)
0.000
7.479
8.745
32.986
16.747
18.272
16.370
22.617
39.087
8.593
17.818
11.354
18.044
10.601
8.633
9.566
14.964
16.952
5.493
2.970
5.023
4.901
4.004
4.174
3.635
4.448
2.736
3.680
3.142
4.231
2.442
2.849
3.699
2.986
2.724
3.139
3.107
3.549
3.208
2.450
2.956
2.805
2.788
2.870
3.263
3.881
2.378
2.524
3.094
3.286
2.812
2.860
4.089
3.362
9.668
4.713
6.967
18.765

293

Standard
Deviation
(Psi)
0.000
6.434
12.105
15.653
18.741
18.987
15.630
18.865
21.816
12.519
24.702
14.827
14.270
11.930
14.103
12.210
16.442
16.005
9.763
5.378
4.854
4.800
2.869
4.298
3.067
4.379
4.424
5.326
4.655
3.948
3.120
3.098
4.790
2.565
3.196
3.658
4.663
4.144
4.177
2.955
4.312
3.566
2.762
5.607
5.627
3.949
3.308
3.680
4.328
4.537
3.832
3.580
5.141
4.058
10.493
6.250
13.091
18.245

TALOS
Score
0.000
0.914
0.886
2.296
2.391
4.985
0.031
-0.900
0.001
4.171
2.868
2.495
2.397
2.033
-0.909
0.951
1.012
-0.719
0.204
0.166
0.128
0.123
0.151
0.124
0.119
0.116
0.095
0.075
0.066
0.058
0.064
0.068
0.074
0.080
0.075
0.076
0.073
0.066
0.060
0.051
0.057
0.064
0.068
0.080
0.080
0.063
0.064
0.069
0.072
0.076
0.086
0.119
0.115
0.141
0.205
0.173
0.186
0.246

No. of
Peptides
Used
0
25
25
6
6
6
6
8
5
4
3
25
5
5
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25
25

Class
None
Dyn
Dyn
Dyn
Warn
Warn
Warn
Warn
Warn
Warn
Warn
Strong
Warn
Warn
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong
Strong


<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>Q</th>
<th>R</th>
<th>S</th>
<th>T</th>
<th>U</th>
<th>V</th>
<th>W</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>381</td>
<td>-60.239</td>
<td>-23.912</td>
<td>5.390</td>
<td>7.214</td>
<td>0.367</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>382</td>
<td>-67.619</td>
<td>-20.963</td>
<td>5.096</td>
<td>6.466</td>
<td>0.530</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>384</td>
<td>161.266</td>
<td>-172.591</td>
<td>51.855</td>
<td>18.781</td>
<td>0.669</td>
<td>25</td>
<td>Warn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>385</td>
<td>-87.367</td>
<td>140.136</td>
<td>30.590</td>
<td>17.441</td>
<td>0.558</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>386</td>
<td>-55.018</td>
<td>-36.806</td>
<td>3.772</td>
<td>5.962</td>
<td>0.382</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>387</td>
<td>-65.080</td>
<td>-36.696</td>
<td>2.901</td>
<td>6.653</td>
<td>0.307</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>388</td>
<td>-70.573</td>
<td>-33.579</td>
<td>5.110</td>
<td>6.248</td>
<td>0.189</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>389</td>
<td>-80.371</td>
<td>-21.424</td>
<td>8.917</td>
<td>15.200</td>
<td>0.161</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>390</td>
<td>-57.730</td>
<td>-44.857</td>
<td>6.613</td>
<td>4.576</td>
<td>0.121</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>391</td>
<td>-58.044</td>
<td>-42.061</td>
<td>3.294</td>
<td>4.074</td>
<td>0.108</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>392</td>
<td>-69.017</td>
<td>-38.044</td>
<td>4.308</td>
<td>4.834</td>
<td>0.112</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>393</td>
<td>-65.241</td>
<td>-45.262</td>
<td>3.518</td>
<td>3.593</td>
<td>0.104</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>394</td>
<td>-63.172</td>
<td>-38.073</td>
<td>3.243</td>
<td>4.317</td>
<td>0.086</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>395</td>
<td>-65.063</td>
<td>-41.871</td>
<td>3.220</td>
<td>4.781</td>
<td>0.084</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>396</td>
<td>-65.059</td>
<td>-42.684</td>
<td>4.237</td>
<td>3.676</td>
<td>0.081</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>397</td>
<td>-64.460</td>
<td>-43.034</td>
<td>3.353</td>
<td>3.448</td>
<td>0.076</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>398</td>
<td>-66.126</td>
<td>-39.768</td>
<td>2.740</td>
<td>5.839</td>
<td>0.078</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>399</td>
<td>-66.675</td>
<td>-39.950</td>
<td>4.928</td>
<td>4.146</td>
<td>0.080</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>-63.623</td>
<td>-44.660</td>
<td>2.428</td>
<td>2.681</td>
<td>0.072</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>401</td>
<td>-64.164</td>
<td>-39.775</td>
<td>2.049</td>
<td>3.752</td>
<td>0.059</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>402</td>
<td>-64.790</td>
<td>-42.069</td>
<td>3.014</td>
<td>3.630</td>
<td>0.059</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>403</td>
<td>-63.606</td>
<td>-44.778</td>
<td>2.805</td>
<td>3.906</td>
<td>0.060</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>404</td>
<td>-62.539</td>
<td>-41.971</td>
<td>2.779</td>
<td>2.566</td>
<td>0.070</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>-65.132</td>
<td>-41.760</td>
<td>4.138</td>
<td>4.236</td>
<td>0.073</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>406</td>
<td>-64.697</td>
<td>-42.581</td>
<td>4.187</td>
<td>3.585</td>
<td>0.093</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>407</td>
<td>-62.924</td>
<td>-44.101</td>
<td>2.696</td>
<td>3.029</td>
<td>0.129</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>408</td>
<td>-64.410</td>
<td>-40.485</td>
<td>3.970</td>
<td>5.267</td>
<td>0.119</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>409</td>
<td>-64.127</td>
<td>-42.106</td>
<td>3.306</td>
<td>3.184</td>
<td>0.131</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>410</td>
<td>-64.649</td>
<td>-42.124</td>
<td>4.311</td>
<td>2.174</td>
<td>0.145</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>411</td>
<td>-62.503</td>
<td>-41.520</td>
<td>5.310</td>
<td>4.298</td>
<td>0.111</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>412</td>
<td>-63.705</td>
<td>-40.246</td>
<td>2.742</td>
<td>4.066</td>
<td>0.102</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>413</td>
<td>-66.287</td>
<td>-43.001</td>
<td>3.718</td>
<td>5.950</td>
<td>0.092</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>414</td>
<td>-65.042</td>
<td>-38.305</td>
<td>4.577</td>
<td>3.832</td>
<td>0.073</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>415</td>
<td>-65.480</td>
<td>-37.214</td>
<td>2.988</td>
<td>5.202</td>
<td>0.094</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>416</td>
<td>-67.790</td>
<td>-40.339</td>
<td>4.856</td>
<td>6.922</td>
<td>0.117</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>417</td>
<td>-68.808</td>
<td>-28.707</td>
<td>5.638</td>
<td>5.974</td>
<td>0.149</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>418</td>
<td>-81.506</td>
<td>-13.505</td>
<td>10.022</td>
<td>15.439</td>
<td>0.179</td>
<td>25</td>
<td>Strong</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>419</td>
<td>79.222</td>
<td>11.049</td>
<td>10.292</td>
<td>13.284</td>
<td>0.231</td>
<td>10</td>
<td>Dyn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>420</td>
<td>80.397</td>
<td>145.784</td>
<td>10.776</td>
<td>12.026</td>
<td>0.358</td>
<td>25</td>
<td>Dyn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>421</td>
<td>76.389</td>
<td>138.295</td>
<td>11.767</td>
<td>7.603</td>
<td>0.453</td>
<td>25</td>
<td>Dyn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>422</td>
<td>85.825</td>
<td>145.819</td>
<td>24.052</td>
<td>23.090</td>
<td>0.685</td>
<td>10</td>
<td>Dyn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>423</td>
<td>71.540</td>
<td>143.381</td>
<td>9.118</td>
<td>13.101</td>
<td>0.541</td>
<td>9</td>
<td>Dyn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>424</td>
<td>72.041</td>
<td>142.446</td>
<td>7.660</td>
<td>12.223</td>
<td>0.521</td>
<td>25</td>
<td>Dyn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>425</td>
<td>65.577</td>
<td>142.616</td>
<td>7.109</td>
<td>9.439</td>
<td>0.722</td>
<td>25</td>
<td>Dyn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>426</td>
<td>9999.000</td>
<td>9999.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3: Mass Spectrometry Analysis of Cdc42

A) Cdc42 Δ7, Δ3 and full length proteins for AUC
B) First attempt at Cdc42 farnesylation by co-expression

<table>
<thead>
<tr>
<th>Peak</th>
<th>M^3+ Peak Mass (kDa)</th>
<th>Actual Mass (kDa)</th>
<th>Species</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7205</td>
<td>21613</td>
<td>Cdc42</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>7274</td>
<td>21819</td>
<td>Cdc42 + Farnesyl</td>
<td>10</td>
</tr>
</tbody>
</table>

C) Second attempt by co-expression

<table>
<thead>
<tr>
<th>Peak</th>
<th>M^2+ Peak Mass (kDa)</th>
<th>Actual Mass (kDa)</th>
<th>Species</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10808</td>
<td>21614</td>
<td>Cdc42</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>10910</td>
<td>21818</td>
<td>Cdc42 + Farnesyl</td>
<td>20</td>
</tr>
</tbody>
</table>

D) Farnesylation of L191A vs L191M

<table>
<thead>
<tr>
<th>Peak</th>
<th>M^2+ Peak Mass (kDa)</th>
<th>Actual Mass (kDa)</th>
<th>Species</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10838</td>
<td>21674</td>
<td>Cdc42 L191M</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>10940</td>
<td>21878</td>
<td>Cdc42 + Farnesyl</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Peak</th>
<th>M^2+ Peak Mass (kDa)</th>
<th>Actual Mass (kDa)</th>
<th>Species</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10808</td>
<td>21614</td>
<td>Cdc42 L191A</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10910</td>
<td>21818</td>
<td>Cdc42 + Farnesyl</td>
<td>99</td>
</tr>
</tbody>
</table>
E) Separation of farnesylated and unmodified Cdc42 using gel filtration

<table>
<thead>
<tr>
<th>Sample</th>
<th>% Farnesylated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>40</td>
</tr>
<tr>
<td>Void</td>
<td>40</td>
</tr>
<tr>
<td>Post</td>
<td>99</td>
</tr>
</tbody>
</table>

F) First attempts at *in vitro* farnesylation

Pre-sample 30 %

After one farnesylation reaction 60 %

After two farnesylation reactions 60 %
G) Farnesylation of GMPPNP vs GDP/GTP bound Q61L protein

<table>
<thead>
<tr>
<th>Sample</th>
<th>% Farnesylated</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNP</td>
<td>0</td>
</tr>
<tr>
<td>GDP/GTP</td>
<td>30</td>
</tr>
</tbody>
</table>

H) Farnesylation of GMPPNP vs GDP bound wild type protein

<table>
<thead>
<tr>
<th>Sample</th>
<th>% Farnesylated</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT PNP</td>
<td>0</td>
</tr>
<tr>
<td>WT GDP</td>
<td>60</td>
</tr>
</tbody>
</table>
Bibliography


Cherfils, J., and Zeghouf, M. (2013) Regulation of small GTPases by GEFs, GAPs, and GDIs. Physiol. Rev. 93, 269–309


