COPYRIGHT AND USE OF THIS THESIS

This thesis must be used in accordance with the provisions of the Copyright Act 1968.

Reproduction of material protected by copyright may be an infringement of copyright and copyright owners may be entitled to take legal action against persons who infringe their copyright.

Section 51 (2) of the Copyright Act permits an authorized officer of a university library or archives to provide a copy (by communication or otherwise) of an unpublished thesis kept in the library or archives, to a person who satisfies the authorized officer that he or she requires the reproduction for the purposes of research or study.

The Copyright Act grants the creator of a work a number of moral rights, specifically the right of attribution, the right against false attribution and the right of integrity.

You may infringe the author’s moral rights if you:
- fail to acknowledge the author of this thesis if you quote sections from the work
- attribute this thesis to another author
- subject this thesis to derogatory treatment which may prejudice the author’s reputation

For further information contact the University’s Director of Copyright Services

sydney.edu.au/copyright
Biochemical and structural analysis
of the Nucleosome Remodelling and Deacetylase
(NuRD) complex

Saad Salman M Alqarni

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

School of Molecular Bioscience
Faculty of Science
University of Sydney

2013
Declaration

The work described in this Thesis was performed between March 2009 and June 2013 in the School of Molecular Bioscience (formerly the School of Molecular and Microbial Biosciences) at the University of Sydney. The experiments were carried out by the author unless stated otherwise. This work has not been submitted, in part or in full, for a higher degree at any other institution.

Saad Salman M Alqarni

June 2013
Abstract

This PhD thesis is a biochemical and structural analysis of elements and interactions of the Nucleosome Remodelling and Deacetylase (NuRD) complex.

Chapter 1 presents a general introduction to NuRD. The NuRD complex is a widely conserved transcriptional coregulator that harbours both nucleosome remodelling and histone deacetylase activities. Abnormalities in a number of its constituent proteins are associated with both cancer and ageing. Although we have a fairly complete ‘parts list’ for NuRD, our understanding of what the parts actually look like and, in particular, how they fit together lags a long way behind.

In mammals, the NuRD complex consists of approximately 10 proteins that are consistently observed in purifications of the complex. These components are CHD3/4, HDAC1/2, p66α/p66β, MBD2/3, RbAp46/48 and MTA1-3. So far, there are three-dimensional structures of several subunits: namely RbAp46/48, HDAC2, and structures of a few isolated domains, including the PHD domains and one chromodomain of CHD4, the methyl-DNA binding domain of MBD2 and the SANT domain of MTA3. MTA1, MTA2 and MTA3 have been found in distinct NuRD complexes.

MTA1 is now considered to be one of the most up-regulated proteins in human cancers. However, essentially nothing is known at the molecular level about how MTA1 either mediates its normal functions or acts in cancer. The N-terminal half of MTA-family proteins is well-conserved between family members and species and is predicted to harbour BAH, ELM2, SANT and GATA-type zinc finger (ZnF) domains. Chapter 2 presents efforts to obtain structural information on the N-terminal domains of MTA1 using bacterial and insect cell expression approaches.

Although many inter-subunit interactions have been reported, the data are almost all limited to GST-pulldown and co-immunoprecipitation assays. The single structure that does exist is of a ~40-residue coiled-coil formed between MBD2 and p66α; otherwise we have no information on how this ~1-MDa complex is assembled.

In order to begin to understand the makeup of the NuRD complex, we focused on inter-subunit interactions. RbAp48 and MTA1 are key components of the NuRD complex, and previous studies have reported that MTA1 interacts with the core complex RbAp46/48 and HDAC1/2. However, the molecular basis for this interaction is not
known. Chapter 3 describes work carried out using pulldown assays to probe the interaction between the subunits RbAp46/48 and MTA1/2. Both MTA1 and MTA2 were found to bind to RbAp46/48 using the C-terminal end. Chapters 4 and 5 describe a more detailed analysis of this interaction. The interaction was confirmed by isothermal titration calorimetry (ITC) and the structure of MTA1-RbAp48 complex was determined using X-ray crystallography. The interaction is mediated by the C-terminal motif (KRAARR) of MTA1. It seems that the MTA1-RbAp48 interaction is conserved across species containing a NuRD-like complex. The structure reveals that RbAp48, which is a WD40 protein, interacts with MTA1 specifically via the histone H4 binding site, which is located to the side groove on RbAp48. These data provide a starting point for understanding both the assembly of the NuRD complex and its biochemical mechanism of action. Chapter 6 represents a final discussion that puts our results into context.
Acknowledgements

In the Name of Allah, the Most Beneficent, the Most Merciful.

The prophet Mohammed (Peace be upon him) said, He who would not be thankful to people, he who will not be thankful to ALLAH”.

During this long and challenging journey, several people have helped me with their time, advice, experience, and patience. I would therefore like to offer my sincere thanks to all of them.

First of all, I would like to express my deep thanks to my supervisor, Joel Mackay, for his guidance, encouragement and support throughout the study, and especially for his patience and guidance during the writing process. It has been a privilege to work with and learn from him over this period of intense effort. Thanks also to my associate supervisor, Jacqui Matthews, for her helpful advice, especially in matters of project planning and thesis completion.

I would like to thank our collaborators from the Laue laboratory at the University of Cambridge, particularly Ernest Laue, who have been wonderful to work with. Special thanks go to Ana, who has helped me with all the work related to X-ray crystallography. Ana’s help has been invaluable on the structure-solving front. Thank you to Sandro for teaching me how to use Tecan and Mosquito. Thank you to Philippa for her help with the ITC experiments.

Thanks to Chu Kong and Pep, whose guidance and friendship I will always appreciate, along with their helpful contribution to the work on insect cell expression. A big thank you to Nick, Michael, Philippa, and Cuong for their constructive advice and considerable efforts in proof reading this thesis; and to the occupants of Room 632—Cuong, Ivan and Mitch—who often acted as troubleshooters and helped me in other ways over the years.

Thanks to past and current members of the Mackay, Matthews and Sunde labs (Alexei, Angela, Ann, Dorothy, Eija, Elise, Herman, Jane, Katschi, Krystal, Liza, Lorna, Margie, Marylène, Michael, Mugdha, Neil, Paul, Paula, Robyn, Roland, Sandra, Sarah, Soumya, Tom and Vanessa) for making it such a great environment to work in. All of them provided me with help and advice at various stages of my PhD.
I also appreciate the financial support from King Saud bin Abdulaziz University for Health Sciences.

To أمي وأبي, thank you for all the support you have provided in your own special ways. I am deeply indebted to my parents for their invaluable support and constant encouragement. All I can do is to say thanks for everything, and may Allah give you all the best in return.

My heartfelt gratitude to my wife for all the love, endless support, timely advice, incredible patience and consideration she has shown to me throughout the course of this work. Without her support and encouragement, I could not have finished this work. She was the one who kept our family functioning, and I understand how difficult this was for her.

Lastly, I am immensely thankful to my large family back home in Saudi Arabia. Thank you to my brothers and sisters for all their support. Thank you to all my friends here in Sydney and back in Saudi Arabia.
Contents
Declaration........................................................................................................ii
Abstract .................................................................................................................. iii
Acknowledgment..................................................................................................... v
Contents..................................................................................................................... vii
Abbreviations.......................................................................................................... xiv
Publications ............................................................................................................. xvii

Chapter 1. Introduction: DNA, Histones and Chromatin.................................. 1
1.1 DNA packaging into chromatin ................................................................. 1
1.2 Gene regulation at the chromatin level .................................................... 2
1.3 Histone modifications and chromatin remodelling ................................. 3
  1.3.1 Histone modifications ............................................................................ 3
  1.3.2 Chromatin remodelling ........................................................................ 4
1.4 The NuRD complex ..................................................................................... 4
  1.4.1 Function of NuRD complex ................................................................. 6
  1.4.2 NuRD components ................................................................................ 8
    1.4.2.1 Chromodomain helicase DNA-binding protein CHD3 and CHD4 .... 9
    1.4.2.2 Histone deacetylases 1 and 2 (HDAC1/2) ...................................... 11
    1.4.2.3 Methylated-CpG binding domain containing protein, MBD2/3 ...... 12
    1.4.2.4 p66α and p66β ............................................................................. 14
    1.4.2.5 Retinoblastoma-associated proteins, RbAp46/48 ............................ 16
    1.4.2.6 The Metastasis-associated proteins, MTA1-3 .................................. 18
  1.4.3 MTA1 and RbAp48 in cancer ............................................................... 20
  1.4.4 Interactions between the NuRD components ....................................... 22
1.5 Aims of this thesis ....................................................................................... 25

Chapter 2. Structural studies of the N-terminal domains of MTA1 ............ 26
2.1 Introduction .................................................................................................... 26
2.2 Sequence analysis and initial construct design ........................................ 26
2.3 Overexpression and purification of MTA1/2 constructs in bacteria .......... 31
  2.3.1 Overexpression trials of BAH-ELM2-SANT-ZnF and ELM2-SANT-ZnF of MTA1 ................................................................. 31
  2.3.2 Overexpression trials of BAH-ELM2-SANT and BAH-ELM2 of MTA1 32
2.4 Refolding MTA1 BAH-ELM2 and BAH-ELM2-SANT domains .............. 34
  2.4.1 Refolding MTA1 BAH-ELM2 domains ........................................ 34
  2.4.2 Refolding MTA1 BAH-ELM2-SANT domains ............................. 39
  2.4.3 Estimate of secondary structure of BAH-ELM2-SANT ................. 41
2.5 Overexpression and purification of ELM2 and ELM2-SANT domains ...... 42
  2.5.1 Attempts to crystallise ELM2-SANT domain ............................... 48
2.6 Expression of the BAH domain from MTA1 and MTA2 ....................... 49
2.7 InsectDirect expression of N-terminal domains of MTA1 ...................... 58
2.8 Attempts to express MTA1 BAH and BAH-ELM2 using baculovirus ........ 59
  2.8.1 Principles of Baculovirus Expression Vectors ............................ 59
  2.8.2 Construction and verification of bacmid-MTA1 BAH and BAH-ELM2 60
2.9 Discussion ........................................................................................ 62
  2.9.1 Protein solubility using different expression systems .................... 62
  2.9.2 Future direction ........................................................................... 65

3.1 Introduction ......................................................................................... 66
3.2 Pulldown experiments ......................................................................... 70
3.3 MTA1 and MTA2 interact directly with RbAp46/48 ............................. 71
3.4 Mapping the region(s) of MTA proteins that binds RbAp proteins .......... 72
  3.4.1 Designing the MTA constructs .................................................... 72
  3.4.2 The C-terminal region is important for the binding of MTA1 to both RbAp proteins ............................................................... 75
3.4.3 The C-terminal region is also important for the binding of MTA2 to both RbAp proteins ................................................................. 76

3.5 The RbAp48 interaction domain localises to the C-terminal portion of MTA1 ........................................................................................................... 77

3.5.1 Confirming the importance of the C-terminal region of MTA1 for binding .............................................................................................................. 77

3.5.2 Two regions of MTA1 can contact RbAp48 ........................................... 78

3.6 Narrowing down the RbAp48 binding site to two short regions within the C-terminal end of MTA1 ........................................................................................................ 78

3.6.1 A KRAARR motif mediates an interaction between MTA1/2 and RbAp48 ............................................................................................................... 78

3.6.2 MTA1(448-542) also can bind to RbAp48 .............................................. 79

3.7 The importance of the KRAARR motif for binding .................................. 81

3.7.1 Detailed mapping of the KRAARR motif ............................................. 81

3.7.2 MTA1 and histone H4 appear to bind to the same pocket on RbAp48 .... 83

3.8 MTA1 and FOG1 independently bind to RbAp48 ........................................... 83

3.9 Attempts to express the C-terminal end of MTA1 in bacteria and in insect cells .............................................................................................................. 84

3.9.1 InsectDirect expression of MTA1(448–715) ........................................... 85

3.9.2 Baculovirus expression of MTA1(448-715) ........................................... 86

3.9.2.1 Construction and verification of bacmid-MTA1(448–715) ............... 86

3.9.2.2 Transfection and protein expression ................................................. 87

3.10 Attempts to express RbAp48-binding domains of MTA1 individually ...... 89

3.10.1 Overexpression and purification of MTA1(625–715) ................................ 89

3.10.2 Overexpression and purification of MTA1(656–686) ............................ 91

3.10.3 Overexpression of constructs derived from MTA1(448–542) ................. 93

3.10.3.1 Overexpression and purification of MTA1(465–516 and 465–542) constructs ....................................................................................................... 95

3.11 Discussion ....................................................................................................... 98
3.11.1 Two regions of MTA1 are implicated in the interaction with RbAp48 ... 98
3.11.2 The KRAARR motif is conserved in MTA proteins and mediates the interaction with RbAp proteins ................................................................. 101
3.11.3 MTA1 binds to RbAp48 in the same way that H4 binds to RbAp46 .... 102
3.12 Summary .................................................................................................................. 102

4.1 Baculovirus expression and purification of RbAp48 .......................... 104
4.2 Verification of Bacmid-RbAp48 ................................................................. 104
4.3 Confirming the interaction between MTA1(656–686) peptide and insect cell produced RbAp48 ................................................................. 107
4.4 Determining the affinity of the RbAp48-MTA1(656–686) interaction .... 109
4.4.1 Isothermal titration calorimetry ......................................................... 109
4.5 Crystal structure of the RbAp48–MTA1 subcomplex of NuRD ............ 111
4.5.1 Crystallisation procedure ................................................................. 111
4.5.2 Structure determination of the RbAp48-MTA1(656–686) complex .... 112
4.5.3 Crystal structure of the RbAp48-MTA1(656–686) complex.......... 118
4.6 Summary .................................................................................................................. 124
4.6.1 The biotin tag interfered with the binding of MTA1(671–686) peptide to RbAp48 ...................................................................................... 124
4.6.2 The KRAARR motif in MTA proteins can bind RbAp proteins .... 125
4.6.3 Mutagenesis data corroborates the crystal structure ..................... 125
4.6.4 Comparison of the RbAp48-MTA1 interaction with the RbAp46-H4 interaction .............................................................................. 126

Chapter 5. Crystal structure of the extended MTA1 peptides (670–695) and (670–711) bound to RbAp48 ................................................................. 129
5.1 Additional contacts may be observed using additional C-terminal amino acids ................................................................. 129
5.2 Determining the affinity of the RbAp48-MTA1(670–695 and 670–711) interaction................................................................................ 130
5.2.1 Isothermal titration calorimetry ...................................................... 130
5.3 Crystal screen of RbAp48-MTA1 (670–695 and 670–711) complexes ...... 132
  5.3.1 Crystallisation procedure .............................................................. 132
  5.3.2 Structure determination of the RbAp48–MTA1(670–711) complex ..... 134
  5.3.3 Crystal structure of the RbAp48–MTA1(670–711) complex .......... 138
5.4 Summary .......................................................................................... 142
  5.4.1 Comparison of the crystal structures of RbAp48-MTA1(656–686) and RbAp48-MTA1(670–711) ................................................................. 142

**Chapter 6. Discussion** ........................................................................ 147
  6.1 Chromatin remodelling complexes and WD40 proteins .................... 147
  6.2 WD40-histone interactions ................................................................. 148
  6.3 WD40 repeat proteins exhibit different binding modes with other partners .. 151
  6.4 The RbAp48-MTA1 interaction is highly conserved across species .......... 153
  6.5 Comparison of the RbAp48-MTA1 structure with other similar WD40 proteins ...................................................................................... 157
  6.6 MTA1 and histone H4 compete for the same binding site on RbAp48 ...... 159
  6.7 RbAp46 and RbAp48 might act as ‘docking stations’ in NuRD ............ 161
  6.8 NuRD contacts chromatin through multiple interfaces ...................... 164

**Chapter 7. Materials and Methods** ...................................................... 167
  7.1 Materials .......................................................................................... 167
    7.1.1 Chemicals, reagents and consumables ........................................ 167
    7.1.2 Bacterial strains ........................................................................ 168
    7.1.3 Bacterial culture media and plates ............................................. 169
    7.1.4 Buffers and solution .................................................................. 169
  7.2 Cloning, mutagenesis and plasmids .................................................... 171
    7.2.1 Plasmids ..................................................................................... 171
      7.2.1.1 Plasmids for in vitro transcription/translation ....................... 171
      7.2.1.2 Plasmids for protein overexpression in E. coli ....................... 171
7.2.1.3 Plasmids for transposition of the RbAp48 gene and MTA1(448–715) construct into baculovirus bacmid ............................................. 173

7.2.1.4 Plasmids for expression in Sf9 cells using InsectDirect System .... 174

7.2.2 Polymerase Chain Reaction ......................................................... 174

7.2.3 Restriction enzyme digestion of double-stranded DNA ............... 174

7.2.4 Agarose gel and DNA purification .............................................. 175

7.2.5 Ligation .............................................................................. 175

7.2.6 Transformation of DNA plasmids into E. coli ................................ 175

7.2.6.1 KCM competent cell preparation ........................................... 175

7.2.6.2 Transformation ................................................................. 176

7.2.6.3 Plasmid isolation and sequencing ......................................... 176

7.2.7 Overexpression and purification of recombinant proteins .......... 176

7.3 Overexpression trials of recombinant proteins ......................... 176

7.3.1 Large scale overexpression .................................................... 177

7.3.2 Purification of recombinant proteins .................................... 178

7.3.3 GSH-affinity purification and enzymatic cleavage .................... 178

7.3.4 Amylose-affinity chromatography of MBP-fusion proteins and enzymatic cleavage .............................................................. 178

7.3.5 Ion exchange chromatography .................................................. 179

7.3.6 Size exclusion chromatography ................................................. 179

7.3.7 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis .......... 179

7.4 Refolding of Ub-His-MTA1(BAH-ELM2, 1–284) and (BAH-ELM2-SANT, 1–340) ........................................................................... 180

7.4.1 Preparation of inclusion bodies ............................................... 180

7.4.2 Ni-NTA affinity chromatography ............................................. 180

7.4.3 Refolding the MTA1 fragments and UBP41 cleavage ............... 180

7.5 Expressing MTA1 constructs using the InsectDirect System ......... 181
7.6 Baculovirus expression and purification of RbAp48 and MTA1(448–715) in insect cells .......................................................................................................................... 182

7.6.1 Generation of recombinant virus bacmid by site-specific transposition. 182
7.6.2 Transfection of recombinant bacmid DNA into SF9 insect cells .......... 183
7.6.3 Baculovirus amplification and expression .................................................. 183
7.6.4 Extraction of protein from Sf9 cells ............................................................. 184
7.6.5 Ni-NTA affinity chromatography and thrombin cleavage ....................... 184
7.6.6 Anion exchange chromatography ............................................................... 185
7.6.7 Size exclusion chromatography .................................................................. 185

7.7 Spectrophotometric quantification of protein and nucleic acid ................. 185

7.8 Far UV-Circular Dichroism (CD) spectropolarimetry ............................... 186

7.9 Preparation of MTA1 peptides .................................................................... 186

7.10 One dimensional $^1$H NMR experiments .................................................. 187

7.11 Isothermal titration calorimetry (ITC) ......................................................... 187

7.12 Crystallization and structure determination ................................................ 188

7.13 Pulldown assays ......................................................................................... 189

7.13.1 In vitro transcription/translation of $^{35}$S-Met-labelled protein .......... 189

7.13.2 GST pull-down assay .............................................................................. 190

7.14 Western blot ............................................................................................... 190

References .......................................................................................................... 192

Appendices .......................................................................................................... 207

Appendix A ........................................................................................................ 200

Appendix B ........................................................................................................ 202

Appendix C ........................................................................................................ 203
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>one-dimensional</td>
</tr>
<tr>
<td>6xHis</td>
<td>hexahistidine</td>
</tr>
<tr>
<td>Amp</td>
<td>ampicillin</td>
</tr>
<tr>
<td>Cam</td>
<td>chloramphenicol</td>
</tr>
<tr>
<td>CHD3 and 4</td>
<td>chromodomain and helicase-domain containing proteins 3 and 4</td>
</tr>
<tr>
<td>CR1 and 2</td>
<td>conserved regions 1 and 2</td>
</tr>
<tr>
<td>D₂O</td>
<td>deuterated water</td>
</tr>
<tr>
<td>DSS</td>
<td>2,2-dimethylsilapentane-5-sulfonic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene-diamine-tetra-acetic acid (disodium salt)</td>
</tr>
<tr>
<td>EED</td>
<td>embryonic ectoderm development</td>
</tr>
<tr>
<td>ELM2</td>
<td>Egl-27 and MTA1 homology 2</td>
</tr>
<tr>
<td>EZH2</td>
<td>enhancer of zeste homologue 2</td>
</tr>
<tr>
<td>FOG-1</td>
<td>Friend of GATA-1</td>
</tr>
<tr>
<td>GSH</td>
<td>glutathione (reduced)</td>
</tr>
<tr>
<td>GST</td>
<td>glutathione S-transferase</td>
</tr>
<tr>
<td>H3</td>
<td>histone 3</td>
</tr>
<tr>
<td>H4</td>
<td>histone 4</td>
</tr>
<tr>
<td>HAT</td>
<td>histone acetyltransferase</td>
</tr>
<tr>
<td>H-bond</td>
<td>hydrogen bond</td>
</tr>
<tr>
<td>HDAC1 and 2</td>
<td>histone deacetylases 1 and 2</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl (β)-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>ITC</td>
<td>isothermal titration calorimetry</td>
</tr>
<tr>
<td>Kan</td>
<td>kanamycin</td>
</tr>
<tr>
<td>Kbp</td>
<td>kilo base pair</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Dalton</td>
</tr>
<tr>
<td>K_D</td>
<td>dissociation constant</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani broth</td>
</tr>
<tr>
<td>MBD3</td>
<td>methyl-CpG binding domain containing protein 3</td>
</tr>
<tr>
<td>MLL</td>
<td>mixed-lineage leukaemia</td>
</tr>
<tr>
<td>MQW</td>
<td>Milli-Q® water</td>
</tr>
<tr>
<td>MTA1 and 2</td>
<td>metastasis associated proteins 1 and 2</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>MWCO</td>
<td>molecular weight cut-off</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>nickel-nitrilotriacetic acid</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NuRD</td>
<td>nucleosome remodeling and deacetylase</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>protein data bank</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PHD</td>
<td>plant homeo domain</td>
</tr>
<tr>
<td>ppm</td>
<td>parts per million</td>
</tr>
<tr>
<td>PRC2</td>
<td>polycomb repressor complex 2</td>
</tr>
<tr>
<td>Rb</td>
<td>retinoblastoma</td>
</tr>
</tbody>
</table>
RbAp46 and 48 retinoblastoma associated proteins 46 and 48

RMSD  root mean squared deviation
rpm  revolutions per minute
SA  streptavidin
SANT  SWI3/ADA2/NCoR/TFIIIB domain
SEC  size exclusion chromatography
SDS-PAGE  sodium dodecylsulphate polyacrylamide gel electrophoresis
Sf9  Spodoptera frugiperda insect cell
SUMO  small ubiquitin-like modifier
Ub-His  ubiquitin hexahistidine
WT  wild type
ZnF  zinc finger
Publications

Journal article


Poster Presentations and Abstracts


Oral presentation

Chapter 1. Introduction: DNA, Histones and Chromatin

1.1 DNA packaging into chromatin

The human body is composed of trillions of cells. All human cells (with the exception of mature red blood cells) contain a complete genome stored as DNA sequences within the chromosomes located in the nucleus of the cell and small amount in the mitochondria. The haploid human genome contains about three billion base pairs of DNA packaged into 23 chromosomes, while the diploid genome has double the DNA content. In eukaryotes, DNA is packaged into a highly ordered structure called chromatin (Kornberg and Thomas, 1974, Kornberg, 1974). The basic repeating unit of chromatin is the nucleosome, which consists of about 146 base pairs of DNA that bind and wrap around an octamer of histone core proteins (two copies of each of the histones H2A, H2B, H3 and H4) (Figure 1-1) (Luger et al., 1997). Several structures of the core nucleosome are available in the Protein Databank (Luger et al., 1997, Davey et al., 2002, Richmond and Davey, 2003). Adjacent nucleosomes are connected by the linker histone (H1 or H5), which binds to internucleosomal DNA sequences to form 30-nm chromatin fibres (Fan et al. 2005). Subsequent fibre–fibre interactions contribute to the high degree of compaction observed in the condensed chromosome. The linker histone plays a role in stabilizing higher order chromatin structures and in regulating gene expression (Khorasanizadeh, 2004, Robinson and Rhodes, 2006, Luger et al., 2012).

Besides packaging and condensing the genome to fit into the small space within the cell nucleus, chromatin structure has important implications for key processes such as DNA replication, DNA repair and gene transcription (Fyodorov and Kadonaga, 2001, Tyler et al., 2001, Luger et al., 2012). Based on its degree of condensation state, chromatin can be divided into euchromatin and heterochromatin. The compaction state of chromatin affects the accessibility of genes to regulatory factors. Heterochromatin is a compacted or “closed” form and is not accessible to remodelling or replication machinery, resulting in the silencing of genes in that section of DNA. In contrast, euchromatin is a less-compacted, more open form, and is more accessible, which allows for gene expression, although euchromatic regions of the genome are not necessarily transcriptionally active (reviewed in Downs et al., 2007, Cairns, 2009). Therefore, dynamic alterations in chromatin structure lead to either transcription repression or transcriptional activation.
during normal development and homeostasis. Aberrant changes in chromatin structure are also strongly linked with a variety of human diseases (Portela and Esteller, 2010).

![Figure 1-1: Structure of a single nucleosome showing the flexible histone tails (PDB: 1KX5) (Davey et al., 2002). The histone octamer is shown in ribbon format, and consists of two copies of each histone protein: H4 (blue), H3 (green), H2A (pink) and H2B (yellow). The DNA coiled around the histone octamer is shown in grey.]

1.2 Gene regulation at the chromatin level

Gene expression is a highly organized process which can be regulated at both the DNA and chromatin levels (Khorasanizadeh, 2004). The transition between chromatin states (and therefore its accessibility) is controlled by a range of mechanisms (De La Fuente, 2006, Mellor, 2006, Shaklai et al., 2007), including the remodelling of chromatin structure using energy derived from ATP hydrolysis and the regulation of chromatin accessibility by reversible covalent modifications of both histones (acetylation, methylation, phosphorylation, ubiquitylation and ADP ribosylation) and DNA (De La Fuente, 2006). Other mechanisms have also been shown to be involved in regulating gene expression; these include changes in histone variants and noncoding RNAs (Andreu-Vieyra and Liang, 2013). These mechanisms require the coordinated recruitment and assembly of multiprotein complexes and ultimately lead to changes in the amount of transcription occurring in each section of genome.
1.3 Histone modifications and chromatin remodelling

1.3.1 Histone modifications

In addition to the structured regions of the histone core, each histone contains flexible histone tails which protrude from the nucleosome. The flexible tails are partially visible in X-ray crystal structures, suggesting that they are unstructured, although the flexible tails are likely to adopt some sort of structure upon binding to adjacent nucleosomes or to other regulatory proteins (Luger et al., 2012). The histone tails are rich in lysine and arginine residues, making them extremely basic. These tails are subject to a variety of reversible posttranslational modifications (PTMs) that are proposed to change their charges and therefore alter DNA accessibility and protein-protein interactions within the nucleosome. For example, lysines are subject to acetylation, methylation, ubiquitination or SUMOylation; arginines are subject to methylation; glutamic acids to ADP-ribosylation; and serines and threonines to phosphorylation (Kouzarides, 2007, Bannister and Kouzarides, 2011). These PTMs are thought to act as recruitment signals for chromatin modifying enzymes and other proteins that either repress or activate transcription. For example, lysine acetylation and arginine methylation of histone tails are usually associated with directing the active euchromatin state, whereas lysine methylation and serine phosphorylation are associated with both the activation and repression of transcription, depending on the combinatorial pattern of these modifications (Munshi et al., 2009, Cohen et al., 2011, Bannister and Kouzarides, 2011). Histone modifications may act as specific codes for recognition by transcription factors and have context-dependent effects. These PTMs are often recognized, introduced and removed by specific protein complexes that contain histone-modifying enzymes. Several histone modifying complexes have been identified from different organisms, including histone acetyltransferase (HAT) complexes, which are involved in histone acetylation and gene activation (Carrozza et al., 2003, Pacifico et al., 2007), histone deacetylase (HDAC) complexes such as the nucleosome remodelling and deacetylase (NuRD) complex (Xue et al., 1998) and the Sin3 complex (Silverstein and Ekwall, 2005), which are co-repressors involved in histone deacetylation. These histone modifying enzymes are recruited in a transcription factor-dependent manner to specific locations on the DNA (Hong et al., 2005, Rampalli et al., 2005), which leads to transcriptional activation/repression through acetylation or deacetylation of specific
lysine residues within the histone tails. Ultimately, these complexes direct chromatin into euchromatic or heterochromatic states.

1.3.2 Chromatin remodelling

Chromatin remodelers are complexes that slide, assemble, and remove histones from DNA using energy derived from ATP hydrolysis, but the mechanism underlying this process is poorly understood (Hauk et al., 2010, Clapier and Cairns, 2009). Four different families of chromatin remodelers have been identified: CHD (chromodomain helicase DNA binding, Mi-2), SWI/SNF (switching defective/sucrose nonfermenting), ISWI (imitation switch) and INO80 (inositol requiring 80) (Clapier and Cairns, 2009). They have diverse functions that range from regulating transcriptional activation and repression, tumour suppression, DNA damage repair, to development (Wang et al., 2007, Bao and Shen, 2007, Smeenk et al., 2010). During the last two decades, a number of chromatin-associating protein complexes that combine proteins with related functions have been isolated. One such complex, which couples two chromatin-modifying activities, namely chromatin remodelling and histone modification, is the Nucleosome Remodelling and histone Deacetylase (NuRD) complex.

1.4 The NuRD complex

A macromolecular complex with both histone deacetylase and chromatin remodelling activities has been isolated by different groups as NuRD, NURD, NRD and the Mi-2 complex (Xue et al., 1998, Zhang et al., 1998a, Wade et al., 1998, Tong et al., 1998). NuRD was first identified, independently, by several laboratories by complex, multi-step biochemical purification from mammalian cells (Xue et al., 1998, Zhang et al., 1999, Zhang et al., 1998a, Tong et al., 1998, Le Guezennec et al., 2006) and Xenopus oocytes (Wade et al., 1999). The NuRD complex is widely expressed in most tissues (Denslow and Wade, 2007). Although there were some slight differences, the composition of the NuRD complex identified by the different laboratories was remarkably similar (Table 1-1). The NuRD is an approximately 1.5 MDa multi-subunit complex, and consists of chromodomain-helicase-DNA-binding protein CHD3 (also known as Mi-2α) or CHD4 (also known as Mi-2β), histone deacetylases HDAC1 and HDAC2, retinoblastoma-associated proteins RbAp46 (also known as RBBP7) and RbAp48 (also known as RBBP4), methyl-CpG-binding domain proteins MBD2 or MBD3, metastasis associated proteins 1–3 (MTA1, 2 or 3) and p66α (also known as GATAD2A) and p66β (also known as GATAD2B) (Figure 1-2). All NuRD subunits are
encoded by multiple gene paralogues, which encode similar proteins with distinct functions.

Table 1-1: Protein composition of mammalian NuRD complex preparations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD3/4</td>
<td>CHD4</td>
<td>CHD4</td>
<td>CHD4</td>
<td>CHD3/4</td>
</tr>
<tr>
<td>MTA2*</td>
<td>MTA1</td>
<td>MTA2</td>
<td>MTA1/2</td>
<td>MTA1/2/3</td>
</tr>
<tr>
<td>HDAC1/2</td>
<td>HDAC1/2</td>
<td>HDAC1/2</td>
<td>HDAC1/2</td>
<td>HDAC1/2</td>
</tr>
<tr>
<td></td>
<td>MBD3*</td>
<td>MBD3a, MBD3b</td>
<td>MBD3</td>
<td>MBD2/3</td>
</tr>
<tr>
<td></td>
<td>p66*</td>
<td></td>
<td>p66</td>
<td>p66α/β</td>
</tr>
<tr>
<td></td>
<td>N190*, N170*,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N160*, N140*,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N135*, N130*,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N85*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*proteins not conclusively identified

While NuRD has been purified from mammalian and amphibian (Xenopus) cells, analogous subunits of the complex have been also purified from insect (Drosophila), nematode (C. elegans) and plants (Arabidopsis), suggesting that NuRD is highly conserved across most or all multicellular plants and animals (Bowen et al., 2004, Ahringer, 2000, Passannante et al., 2010). The Drosophila NuRD complex contains distinct homologues for all the mammalian NuRD subunits (Kon et al., 2005). In Arabidopsis, HDA proteins are closely related to the human HDAC1/2 proteins (Pandey et al., 2002), MSI1 is related to the human RbAp46/48 proteins (Ach et al., 1997), and PICKLE is related to CHD3 (Zhang et al., 2008). However, it is still unknown whether HDA, MSI1 and PICKLE in Arabidopsis actually associate to form a plant homologue of the NuRD complex. Table 1-2 lists homologues of the NuRD components found in each of these organisms.
Chapter 1

Figure 1-2: Composition of the human NuRD complex. The NuRD complex consists of CHD3/CHD4 (chromodomain-helicase-DNA-binding proteins), HDAC1/2 (histone deacetylases), RbAp46/48 (retinoblastoma-associated proteins), MBD2/3 (methyl-CpG-binding domains), MTA1/2 (metastasis associated proteins) and p66α and p66β. All components have been found to associate as part of the NuRD complex; however, the nature of the direct protein-protein interactions is still largely unknown.

Table 1-2: Homologues of the NuRD complex components in selected species

<table>
<thead>
<tr>
<th>Species</th>
<th>CHD3/4</th>
<th>HDAC1/2</th>
<th>RbAp46/48</th>
<th>MTA1/2/3</th>
<th>MBD2/3</th>
<th>p66α/β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xenopus</td>
<td>Mi-2</td>
<td>Rpd3</td>
<td>RbAp46/48</td>
<td>MTA1-like</td>
<td>MBD3</td>
<td>p66</td>
</tr>
<tr>
<td>Drosophila</td>
<td>dMi-2</td>
<td>RBD3</td>
<td>p55</td>
<td>MTA1-like</td>
<td>MBD2/3</td>
<td>p66</td>
</tr>
<tr>
<td>C. elegans</td>
<td>CHD3</td>
<td>HDA-1</td>
<td>Lin-53/RBA-1</td>
<td>EGR-1/ EGL-27(Lin-40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>PICKLE</td>
<td>HDA-like</td>
<td>MSI1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1.4.1 Function of NuRD complex

As the name suggests, the NuRD complex has both ATP-dependent chromatin remodelling and histone deacetylase activities. NuRD is the only known complex coupling these two independent enzymatic functions. This dual catalytic activity was proposed to be essential for the formation of heterochromatin with tightly packaged hypoacetylated nucleosomes and the rapid termination of gene transcription (Denslow and Wade, 2007, Wang and Zhang, 2001). Interestingly, although NuRD has mainly been characterized as a co-repressor complex, it has also recently been shown that the NuRD complex can also activate gene expression (Miccio et al., 2010), which could explain the coexistence of two catalytic activities in the NuRD complex. Currently, little is known about the circumstances that determine whether NuRD acts in an activating or repressive fashion.
Chapter 1

NuRD plays an essential role in DNA damage repair and in assembly and maintenance of higher order chromatin structure (Allen et al., 2013, Smeenk et al., 2010). Furthermore, the NuRD complex has functions in normal development, as well as its implication in tumorigenesis, which was revealed by knockout of NuRD complex components (Ramirez and Hagman, 2009). Also, the Drosophila NuRD complex has been shown to be involved in Wnt signalling during wing development (Kon et al., 2005), whereas a NuRD-like complex in C. elegans has been shown to block the Ras signalling pathway during vulval development (Solari and Ahringer, 2000). The NuRD complex is required at various stages of haematopoietic differentiation (Yoshida et al., 2008), and is also involved in the transcriptional regulation of key genes, such as Cd4 gene, that promote the progression and development of T and B lymphocytes (Gao et al., 2009, Williams et al., 2004, Naito et al., 2007). The NuRD complex also plays an essential role in maintaining pluripotency in embryonic stem cells and in the normal differentiation program (Kaji et al., 2006, Zhu et al., 2009).

The unique feature of NuRD, as noted above, is that it contains both histone deacetylase and chromatin remodelling activities and, interestingly, has been implicated in both gene repression and activation. When NuRD was first discovered, it was predicted that it would play a role in gene repression (Tyler and Kadonaga, 1999), either through protein-DNA interactions or through protein-protein interactions between NuRD subunits and transcription factors. Subsequent work has shown that multiple mechanisms are used to recruit the NuRD complex to different genomic locations through either direct or indirect contacts of a specific subunit with DNA or DNA-binding proteins. For example, the MTA and MBD subunits are implicated in targeting the NuRD complex to specific genes through interactions with transcription factors (Fujita et al., 2004) or methylated DNA (Hendrich and Bird, 1998), respectively. Furthermore, several transcription factors can recruit the NuRD complex to repress the expression of key developmental genes, including the lymphoid transcription factor Ikaros (Kim et al., 1999), the co-repressor KAP-1 (Schultz et al., 2001), p53 (Luo et al., 2000) and BCL-6 (Fujita et al., 2004). These interactions are mediated by different components of NuRD: CHD4 for Ikaros/KAP-1, MTA2 for p53 and MTA3 for BCL-6. In addition, through its binding with RbAp48 or MTA1, Friend of GATA1 (FOG1) recruits the NuRD complex to repress multiple genes during haematopoiesis (Hong et al., 2005, Gao et al., 2010, Lejon et al., 2011). Furthermore, it has been shown that
NuRD can suppress essential genes in prostate cancer (Srinivasan et al., 2006), breast cancer (Fujita et al., 2003) and HIV (Cismasii et al., 2008). In Arabidopsis, the CHD3/CHD4 homologue PICKLE suppresses the expression of seed-associated genes during germination (Zhang et al., 2008).

Several posttranslational modifications in NuRD subunits have been reported, and these may alter the function of these proteins. While the exact role of these modifications has not yet been determined, it is likely that they represent another level of functional regulation of the NuRD complex (Lai and Wade, 2011). For example, research on the acetylated form of MTA1 provides evidence that the NuRD complex can function as a transcriptional activator, as well as a transcriptional repressor (Gururaj et al., 2006, Lai and Wade, 2011). Thus, the NuRD complex is potentially implicated in maintaining the balance between repressed and active genes required for proliferation, differentiation and homeostasis.

It has become increasingly clear that the multisubunit NuRD complex is multifunctional and participates in wide range of processes, including histone deacetylation and nucleosome mobilisation, as well as binding to methylated DNA and histones. It also plays important roles in transcription, chromatin assembly, cell cycle and DNA stability (Lai and Wade, 2011, Passannante et al., 2010, Brehm et al., 2000, Denslow and Wade, 2007).

### 1.4.2 NuRD components

The exact composition of the NuRD complex can vary depending on cell type and target genes. In mammalian cells, the NuRD complex appears to consist of six core subunits, two of which have enzymatic activities: CHD3 or CHD4, which utilize the energy of ATP hydrolysis to move histone octamers relative to DNA (a process termed chromatin remodelling); and HDAC1 and HDAC2, which remove acetyl groups from lysine residues of histones and non-histone proteins. Further, NuRD contains non-enzymatic subunits, which include MBD2 or MBD3, p66α and p66β, MTA1, 2 or 3, and RbAp46 and RbAp48. The deacetylase core complex (HDAC1 and HDAC2 together with RbAp46 and RbAp48) is also found in another complex, SIN3 (Zhang et al., 1999).
1.4.2.1 Chromodomain helicase DNA-binding protein CHD3 and CHD4

Human cells have separate genes that encode CHD3 and CHD4, both of which are highly conserved in animals and plants (Woodage et al., 1997). Although CHD3 has been purified from some mammalian cell lines (Xue et al., 1998), CHD4 is the predominant subunit of the NuRD complex and is also the largest subunit (218 kDa). CHD4 was first identified as an autoantigen in the human disease dermatomyositis (Seelig et al., 1995) and is a member of the SWI2/SNF2-related proteins (Eisen et al., 1995). CHD3/4 consist of two tandem plant homeodomain fingers (PHDs), two chromodomains (CHDs) and an ATPase/helicase domain (Figure 1-3). In addition, two domains of unknown function (DUF1/2) have been recognised at the C-terminal end of CHD3/4. The ATPase/helicase domain provides the energy necessary for disrupting histone:DNA interactions, enabling the shifting of histones along the DNA during nucleosome remodelling (Clapier and Cairns, 2009, Hall and Georgel, 2007).

The PHD domains play an important role in the interaction with HDAC1 (Zhang et al., 1998a) and have been shown to bind to the N-terminal tail of histone H3 (Musselman et al., 2009, Mansfield et al., 2011). Nucleosome binding of PHDs promotes the ATPase and chromatin remodelling functions of CHD4 (Watson et al., 2012a). Likewise, the ATPase/helicase domain supports the binding of PHD/CHD domains to histones (Morra et al., 2012). The structure of the second PHD finger of CHD4 in complex with H3K9me3 (Figure 1-4B) provides insight into the mechanism by which CHD4 and the NuRD complex could direct the chromatin remodelling (Mansfield et al., 2011).

In the case of the Drosophila CHD3/4 homologue dMi-2, CHDs have been reported to interact with DNA and nucleosome (Bouazoune et al., 2002). It has been shown that both CHD domains are essential for the repressive function of the NuRD complex (Watson et al., 2012a). The solution structure of a PHD1 finger and CHD2 domain has also been determined (Figure 1-4A and C) (Mansfield et al., 2011).

Small-angle X-ray scattering (SAXS) studies reveal extensive contacts within the individual domains of CHD4, which provide explanation for the regulation of CHD4 activities by intramolecular domain communication (Watson et al., 2012a). In the SAXS models, the CHD domains associate with PHDs, the ATPase/helicase domain and DUF1, and additional contacts are seen between the ATPase/helicase domain and DUF1. The PHDs-CHDs association is found to limit the interaction of CHD4 with
nucleosomes. A structural rearrangement in CHD4 in which CHDs and PHDs release the ATPase/helicase domain to allow access to DNA, would then stimulate the enzymatic activity. It has been proposed that the chromodomains stimulates the ATPase activity, possibly by directly binding DNA and thereby facilitating the movement of CHD4 along DNA (Watson et al., 2012a).

Through CHD3 and/or CHD4, a number of co-repressors, including Tramtrack69, RFP, NAB2 and KAP1, recruit NuRD to DNA sites to repress gene expression (Kehle et al., 1998, Murawsky et al., 2001, Shimono et al., 2003, Srinivasan et al., 2006, Schultz et al., 2001). CHD3/4 proteins are key subunits in NuRD assembly, as well as in the association of NuRD with the nucleosome, and linking non-NuRD proteins to the complex.

Figure 1-3: Domain structure of human CHD3/4. These proteins contain two N-terminal tandem plant homeodomain fingers (PHD1/2), two chromodomains (CHD1/2), an ATPase/helicase domain and two domains of unknown function (DUF1/2).

Through CHD3 and/or CHD4, a number of co-repressors, including Tramtrack69, RFP, NAB2 and KAP1, recruit NuRD to DNA sites to repress gene expression (Kehle et al., 1998, Murawsky et al., 2001, Shimono et al., 2003, Srinivasan et al., 2006, Schultz et al., 2001). CHD3/4 proteins are key subunits in NuRD assembly, as well as in the association of NuRD with the nucleosome, and linking non-NuRD proteins to the complex.

Figure 1-4: The structures of CHD4 domains. A. The solution structure of CHD4 PHD1 finger. Zinc ions are shown as brown spheres (PDB: 2L5U) (Mansfield et al., 2011). B. The solution structure of CHD4 PHD2 finger in complex with histone H3K9me3 peptide. Zinc ions are shown as green spheres (PDB: 2L75) (Mansfield et al., 2011). C. The solution structure of CHD4 CHD2 domain (PDB: 2EE1, unpublished).
1.4.2.2 Histone deacetylases 1 and 2 (HDAC1/2)

Approximately eighteen functionally-distinct human histone deacetylases (HDAC) have been identified. They are classified into three classes depending on sequence homology, catalytic mechanisms and phylogenetics (Yang and Seto, 2008). Most HDAC proteins do not act independently, but function as subunits of multiprotein complexes (Hayakawa and Nakayama, 2011). HDAC1 and HDAC2 are class I histone deacetylases, along with HDAC3 and HDAC8. They are widely expressed in many tissues and conserved in all eukaryotes. In addition to being core subunits of the NuRD complex, HDAC1/HDAC2 have been found in other multiprotein co-repressor complexes, including Sin3 and CoREST, NCOR/SMRT and ES-specific NODE (Zhang et al., 1998b, You et al., 2001a, Ordentlich et al., 1999, Liang et al., 2008). HDAC-complexes can influence transcriptional repression by the deacetylation of specific residues on the histone tails, which can contribute to directing chromatin into its heterochromatic state.

HDAC1 and HDAC2 proteins (also called lysine deacetylases) are highly similar proteins (~85% sequence identity) that consist of a zinc-containing deacetylase catalytic domain that removes the acetyl moiety from specific lysines in both histones and non-histone proteins (Choudhary et al., 2009a), allowing the histones to wrap the DNA more tightly. In addition, they possess a C-terminal Rb-binding domain that might be involved in protein-protein interactions (Figure 1-5) (Thiagalingam et al., 2003). Their action opposes that of histone acetyltransferase enzymes such as CBP/p300 (Shiama, 1997) and TAF(II)250 (Mizzen et al., 1996).

**Figure 1-5: Schematic of HDAC1 and HDAC2.** They each contain a zinc-containing deacetylase catalytic domain and a C-terminal Rb-binding domain.

The crystal structure of the HDAC2 catalytic domain in complex with a small molecule inhibitor (benzamide) has been solved, which provides new insights that will be useful
in the design of HDAC inhibitors (Figure 1-6) (Bressi et al., 2010). Based on the high sequence similarity, we expect that the structure of HDAC1 is similar to HDAC2.

Figure 1-6: Crystal structure of HDAC2 in complex with a small molecule inhibitor N-(4-aminobiphenyl-3-yl) benzamide. HDAC2 is shown in green, the inhibitor compound in grey and the zinc as a yellow spheres (PDB: 3MAX (Bressi et al., 2010)).

1.4.2.3 Methylated-CpG binding domain containing protein, MBD2/3

MBD2 and MBD3 (Brackertz et al., 2002, Feng et al., 2002, Zhang et al., 1999, Wade et al., 1999) are part of the methyl-CpG-binding domain (MBD) protein family, which includes MeCP2 and MBD1-4 (Wade et al., 1999, Zhang et al., 1999). With the exception of MBD3, the other members have been reported to bind methylated DNA and are involved in HDAC-associated gene silencing activity. While MBD2 and MBD3 are highly similar proteins with ~70% identity in humans, MBD3 lacks the N-terminal part of MBD2, which might be responsible for the inability of human MBD3 to bind methylated DNA (Hendrich and Bird, 1998). MBD3 is the smallest subunit of the NuRD complex (30 kDa). MBD2 contains the N-terminal glycine-arginine repeats (GR), an MBD domain, a transcriptional repression domain (TRD) and a C-terminal coiled-coil motif (Figure 1-7). The role of GR repeats has not yet been defined, while the MBD domain is responsible for binding to methylated DNA; the structure of the MBD domain from MBD2 bound to a methylated DNA fragment has been solved.
The TRD of Drosophila MBD2 is implicated in binding to Nurf55 (p55) (Marhold et al., 2004). The coiled-coil motif binds to p66α (Gnanapragasam et al., 2011, Brackertz et al., 2006). MBD3, on the other hand, contains an MBD domain, a TRD domain, a coiled-coil motif and a C-terminal glutamate repeat region (E) (Figure 1-7).

Despite their similarity, MBD2 and MBD3 are related proteins with different functions. MBD3 has been shown to be essential for embryonic development: MBD3 deficient mice die early during embryogenesis (Hendrich et al., 2001). On the other hand, MBD2-deficient mice are viable and show only mild defects (Hendrich et al., 2001). It has been reported that these two proteins are mutually exclusive in NuRD, defining different sub-populations of the complex (Le Guezennec et al., 2006). Surprisingly, it has also been shown that MBD2 and MBD3 can form a heterodimer (Brackertz et al., 2002, Feng et al., 2002, Zhang et al., 1999, Wade et al., 1999). Thus, MBD3, through the binding with MBD2, may link the NuRD complex to methylated DNA sites to mediate gene silencing (Scarsdale et al., 2011). MBD2-binding zinc finger (MIZF) has been shown to associate with MBD2/3 and recruits NuRD complex for transcriptional repression (Sekimata et al., 2001). Focal adhesion kinase (FAK) has also been shown to interact with MBD2 and activate differentiation-associated genes (e.g. myogenin) (Luo et al., 2009).
1.4.2.4 p66α and p66β

p66α and p66β are also core members of the NuRD complex (Brackertz et al., 2002, Feng et al., 2002, Zhang et al., 1999, Wade et al., 1999). They are highly related proteins and widely expressed in various tissues, which suggests that they have similar functions (Brackertz et al., 2002). They share two highly conserved regions, CR1 and CR2, and a C-terminal GATA zinc finger (GATA ZnF) (Figure 1-9). Both of these conserved regions are thought to be responsible for the repressive capacity of the p66 proteins (Brackertz et al., 2002). CR1 has been shown to be required for the interaction with the C-terminal coiled-coil (CC) of MBD2/3 (Brackertz et al., 2006, Feng et al., 2002). Recently, the structure of the p66α–MBD2 coiled-coil complex formed between these domains has been solved (Figure 1-10) (Gnanapragasam et al., 2011), which suggests that this complex contributes to NuRD assembly and might link the NuRD to methylated DNA to mediate gene silencing (Kon et al., 2005). The CR2 region and GATA ZnF are implicated in the binding of p66 proteins to all histone tails (Brackertz et al., 2006). In addition, both p66 proteins interact with MBD2, and a single mutation in p66α disrupts the interaction with MBD2 and interferes with the repression mediated by MBD2 (Brackertz et al., 2006). Silencing the endogenous p66α or p66β reduces MBD2-mediated repression (Brackertz et al., 2006). Both p66 proteins are subject to posttranslational modifications by SUMOylation at K30 and K487 in p66α, and at K33 in p66β (Gong et al., 2006). It has been reported that SUMO modifications of p66
proteins are crucial for the repression activity of p66 and in specific interactions with some NuRD components (HDAC1 and RbAP46) (Gong et al., 2006).

![Figure 1-9: Domain structure of p66α/β. They contain two conserved regions (CR1) and (CR2) and a C-terminal GATA zinc finger (ZF).](image)

Foxp factors, through binding to p66β, recruit the NuRD complex to repress Foxp1/2/4 target gene expression (Chokas et al., 2010). Although p66α and p66β have been demonstrated to be potent transcriptional repressors (Feng et al., 2002, Brackertz et al., 2002), their precise function within the NuRD complex is currently unknown. However, they might act as structural subunits, given that many interactions of p66 with other NuRD subunits have been reported (described in Section 1.4.4) (Gong et al., 2006, Brackertz et al., 2006).

![Figure 1-10: p66α-MBD2 structure. Solution structure of the conserved region (CR1) domain of p66α bound to the coiled-coil (CC) motif of MBD2 (PDB: 2L2L (Gnanapragasam et al., 2011)). p66α is shown in green and MBD2 in blue.](image)

The RbAp46/48 and MTA1/2 proteins are the main subject of this thesis, and are presented in the following sections.
1.4.2.5 Retinoblastoma-associated proteins, RbAp46/48

RbAp46 and RbAp48 are both key subunits of the NuRD complex and are highly conserved across animals and plants. They are very similar, with ~92% identity in amino acid sequence (described in Section 3.1), and were first identified in mammalian cells through their interaction with the tumour-suppressor Rb (Qian et al., 1993). Since then they have been found in several chromatin-modifying complexes, including NuRD and Sin3 (Zhang et al., 1997, Zhang et al., 1999), NURF (nucleosome remodelling factor) (Barak et al., 2003) and PRC2 (Polycomb Repressive Complex 2) (Kuzmichev et al., 2002). Intriguingly, RbAp48 is exclusively a subunit of the chromatin assembly factor 1 (CAF-1) complex, whereas only RbAp46 is part of the histone acetyltransferase (HAT1) complex (Tyler et al., 1996, Verreault et al., 1998, Kaufman et al., 1997, Parthun et al., 1996). Thus, although they are highly similar, RbAp46 and RbAp48 might carry out partially distinct functions. Alternatively, it is also possible that the existence of one RbAp is sufficient for the function of these complexes. RbAp proteins contain a central WD40 repeat flanked by N- and C-terminal helices (Figure 1-11). RbAp46 and RbAp48 together with HDAC1 and HDAC2 are associated to form a histone deacetylase core in two co-repressor complexes, NuRD and Sin3.

![Figure 1-11: Schematic diagram of RbAp46/48.](image)

As noted above (Table 1-2), RbAps orthologues have also been found in *C. elegans*, *S. cerevisiae* and *Drosophila* (Shi and Mello, 1998, Ruggieri et al., 1989, Tyler et al., 1996), and generally, they are correlated with gene repression. For instance, the *Drosophila* p55 protein is essential for the repression of E group genes by dE2F/RBF proteins (Taylor-Harding et al., 2004).

As noted above, RbAp46 and RbAp48 are WD40 domain proteins; WD40 is one of the most abundant domain types in eukaryotic proteomes (Stirnimann et al., 2010). The WD40 domain (also called WD-repeat) was first described in the β subunit of the...
heterotrimeric G protein transducin (Smith et al., 1999). WD40 proteins contain 4–16 tandem repeats of the WD40 domain, which comprises an approximately 44–60 residue stretch ending at the C-terminus with a signature dipeptide, Trp (sometimes Phe or Tyr), then Asp (sometimes Gln). Many WD40 structures have been determined (Xu and Min, 2011).

The crystal structure of the β subunit of transducin shows that the protein forms a highly symmetric, seven-bladed β-propeller structure (Wall et al., 1995). Each WD40 repeat (blade) consists of a four-stranded anti-parallel β-sheet (Figure 1-12). In most cases, the outer strand from the first blade forms a β-sheet with the three strands of the last blade. The seven β-blades are spread around a narrow central cavity and fan outwards such that the diameter of the cavity at one end of the propeller is slightly greater than at the other.

WD40 proteins have been implicated in a wide variety of functions, including transcriptional regulation and chromatin assembly, signal transduction, RNA synthesis and processing, vesicular trafficking and cell cycle regulation (Stirnimann et al., 2010, Xu and Min, 2011). WD40 domains appear to act predominantly as protein-protein recognition modules (examples in Chapter 6) (Stirnimann et al., 2010).

Figure 1-12: Crystal structure of β subunit of G protein transducin (PDB: 1GP2 (Wall et al., 1995)). A. Top view of the seven-bladed propeller structure. Each repeat consists of four β-strands. Strands corresponding to each of the WD40 repeats are colour-coded. The strands of blade 1 are labelled, and each blade is numbered around the propeller. B. Side view of the structure.
1.4.2.6 The Metastasis-associated proteins, MTA1-3

There are three related proteins that make up the MTA family in humans; all of which have been observed to be associated with the NuRD complex (Fujita et al., 2003). These proteins are encoded by three different genes (Kumar et al., 2003). In C. elegans, two homologues to the human MTA1 gene are also present, egl-27 and egr-1 (also known as Lin-40) (Herman et al., 1999, Solari et al., 1999). MTA proteins are conserved across animals. While, MTA1 and MTA2 are predominantly localized to the nucleus, MTA1 can also be located in the cytoplasm (Balasenthil et al., 2007, Bagheri-Yarmand et al., 2004, Moon et al., 2004, Toh et al., 1999), and MTA3 is localized in both the nucleus and the cytoplasm. MTA proteins are involved in many cellular functions, including differentiation, hormonal action, epithelial-to-mesenchymal transitions and cell fate programs, by modifying the acetylation status of crucial target genes through the association with HDAC (Manavathi and Kumar, 2007). It has been reported that the MTA proteins are mutually exclusive and form distinct NuRD complexes (Bowen et al., 2004, Fujita et al., 2003). These associate with different transcription factors and are involved in different signalling pathways. For example, MTA3 can directly associate with the transcriptional repressor BCL-6 to maintain a germinal centre B cell identity in activated B cells (Fujita et al., 2003). In general, it has been shown that MTA proteins associate directly with transcriptional repressors, and with the oestrogen receptor α (ER-α) and its co-regulators MAT1, NRIF3, LMO4 and MICOA, as well as with GATA-1 co-factor FOG1 (Talukder et al., 2003, Talukder et al., 2004, Mishra et al., 2003, Singh et al., 2005, Mazumdar et al., 2001, Lejon et al., 2011). Surprisingly, however, MTA1 has also been found to be a transcriptional activator of Breast Cancer Amplified Sequence 3 (BCAS3), a gene upregulated in breast cancers (Gururaj et al., 2006).

Analysis of the primary structure of MTA proteins reveals 68% and 73% similarity of MTA2 and MTA3, respectively, with MTA1, with the highest similarity is found in the N-terminal part of the proteins, whereas the C-terminal regions are more divergent (Figure 1-13). It is possible that the differences in the C-terminus of the MTA proteins could be an indication of functional diversity of MTA/NuRD complexes. All MTA proteins share four highly conserved domains. The first of these is the BAH (bromo-adjacent homology) domain, which has been found in a wide range of DNA and chromatin-associated proteins, such as the transcription factor Ash1, the DNA-methyltransferases, and the replication origin recognition protein (ORC1). It is likely
that the BAH domain is involved in protein-protein interactions, as suggested by its interaction within Orc1p with the Sir1p (silent information regulator) (Callebaut et al., 1999, Goodwin and Nicolas, 2001).

Next to the BAH domain are the ELM2 (Egl-27 and MTA1 homology) domain and the SANT (SWI, ADA2, N-CoR, TFIIB-B) domain (Manavathi and Kumar, 2007, Grune et al., 2003, Oliver et al., 2005, Humphrey et al., 2001). Interestingly, most ELM2 domain-containing proteins also have a SANT domain, which suggests a functional and/or structural association between these two domains. In addition to MTA proteins, RERE and Atro similarly possess both the ELM2 and SANT domains; these proteins are found in other essential transcriptional repression complexes such as CoREST (REST corepressor) and MIER1 (mesoderm induction early response 1) (Ding et al., 2003, Wang et al., 2008). ELM2 was first described in the C. elegans protein Egl-27, which plays an important role in patterning during embryonic development (Solari et al., 1999). It has been shown that the ELM2 domain directly binds to HDAC1/2 (Wang et al., 2006, Ding et al., 2003). The ELM2 domain (possibly in association with the SANT domain) in the hMIER1 protein recruits the chromatin remodelling enzyme, through binding with HDAC1, for transcriptional repression (Ding et al., 2003). ELM2-SANT domains in MTA proteins might work in a similar way to their counterparts in RERE and Atro by recruiting HDAC1/2 and G9a to sequential deacetylation and then methylation of H3K9 (Wang et al., 2008). Currently, no structures of ELM2 domains are available.

It has been shown that the SANT domain of SWI-SNF complex is involved in DNA binding and protein-protein interactions (Aasland et al., 1996), including interactions with histone deacetylase (HDAC) complexes (Guenther et al., 2001, You et al., 2001b) and histone acetyltransferase (HAT) (Sterner et al., 2002); however, the exact role(s) of the SANT domain in individual proteins remains to be determined. Following the SANT domain is a GATA-type zinc finger domain (ZnF), which is essential for interaction with transcription factors or transcriptional co-regulators such as FOG2 (Roche et al., 2008) and direct interaction with DNA (Nawa et al., 2000). Similar GATA-zinc finger has been shown to recognize the T/AGATAA/G sequence of DNA (Liew et al., 2005).
The C-terminal end of MTA proteins, following the zinc finger domain, is less conserved. However, MTA1 has two proposed SH3 binding motifs, a Pro-rich sequence and a predicted Myb domain towards the C-terminal end (Nawa et al., 2000, Roche et al., 2008, Waksman et al., 2004, Toh and Nicolson, 2009).

Taken together, the presence of a wide range of predicted domains suggests that the MTA proteins are involved in a variety of protein-protein and DNA interactions. In particular, the presence of domains that have been shown to have DNA-binding properties suggests that MTA proteins might be involved in assembly of the NuRD complex on target DNA.

![Figure 1-13: Structural domains of MTA proteins. MTA proteins share four highly conserved domains (BAH, ELM2, SANT and ZnF) at the N-terminus, while the C-terminus is divergent. MTA1 additionally contains predicted Myb-binding and SH3-binding domains.](image)

MTA1 is unique among the MTA proteins due to a number of reasons, such as its ability to suppress or activate transcription in a context-dependent manner, its role in DNA damage repair, inflammation and maintaining an overexpression in a wide variety of human cancers (Li et al., 2012). There have been no structural studies aimed at understanding the molecular mechanism of MTA1, despite its importance in different biological processes.

### 1.4.3 MTA1 and RbAp48 in cancer

The NuRD complex is strongly implicated in cancer, and some NuRD subunits are overexpressed in a number of tumours. In addition, the NuRD complex can directly associate with oncoproteins and tumour suppressors to promote the transcriptional
repression of target genes (Lai and Wade, 2011, Li et al., 2012). Of all the NuRD components, MTA proteins are the best studied in the context of cancer progression. MTA1, the founding member of the MTA family of genes, was first identified as an overexpressed gene in metastatic adenocarcinoma cell lines (Toh et al., 1994), and since then it has been found to be one of the most upregulated proteins in human cancers, in addition to its key role in regulating gene expression (Kumar et al., 2003, Li et al., 2009, Li et al., 2012). Overexpression of MTA1 is associated with invasive growth and poor prognosis in many cancers (Lai and Wade, 2011, Li et al., 2012). MTA1 is upregulated in tumours derived from a variety of tissues, including ovarian, breast, colorectal, gastric, pancreatic, and prostate (Nicolson et al., 2003). In breast cancer, MTA1 and MTA2, but not MTA3, have been shown to suppress oestrogen functions (Toh and Nicolson, 2009). Overexpression of MTA1 in (ER)-positive breast cancer cells is adequate to decrease the levels of ER target genes, which results in aggressive invasive growth in an anchorage-independent manner (Mazumdar et al., 2001). The c-myc is an oncogene, which is widely upregulated in human cancer (Nesbit et al., 1999). The c-MYC oncoprotein associates with the genomic location of MTA1 and regulates the epithelial-to-mesenchymal transition (EMT) by activating MTA1 transcription. Silencing of MTA1 was found to block the ability of MYC to transform mammalian cells (Zhang et al., 2005).

MTA2 has also been found to be overexpressed in different types of tumours, such as ovarian epithelial cancer (Yao and Yang, 2003, Ji et al., 2006), making MTA1/2 important biomarkers and therapy targets in cancer (Hofer et al., 2004).

The NuRD complex, unlike other chromatin remodelling complexes, can promote or repress tumorigenesis depending on the context. For example, an MTA3-containing NuRD complex has been shown to inhibit transcription of SNAIL, a key transcription factor that promotes the EMT that is important in cancer metastasis (Fujita et al., 2003), thereby implicating MTA3 in the suppression of a cancer phenotype. In contrast, the transcriptional regulator TWIST recruits the MTA2-containing NuRD complex to the promoter of a target gene, to repress transcription and promote EMT (Fu et al., 2011).

Furthermore, it has been found that RbAp48 is upregulated in different human carcinomas, such as liver (Song et al., 2004) and lung cancer (Fukuoka et al., 2004), acute myelocytic and lymphoblastic leukemias (Casas et al., 2003), and plays a role in
thyroid cancer. (Pacifico et al., 2007). Other members of the NuRD complex are also implicated in different aspects of cancer, such as CHD4 (Bolotina et al., 1980, Dyson and Wright, 2002) and HDACs (Cress and Seto, 2000). Despite the clear links between NuRD and cancer, few coordinated attempts have been made to understand either the physical context in which these proteins are located, or the full biochemical functions of the constituent domains.

1.4.4 Interactions between the NuRD components

Identifying protein-protein interactions is fundamental to a molecular understanding of the NuRD architecture. Moreover, an understanding of the interactions between NuRD subunits will inform an understanding of the molecular mechanisms of the various functions of NuRD, and will help in understanding disease mechanisms, potentially providing the basis for new therapeutic approaches. Many interactions within the individual subunits of the complex have been reported (Figure 1-14).

Figure 1-14: Schematic diagram indicates the interaction between components of the NuRD complex. Interactions between subunits and binding of the individual subunits to DNA, histones and other proteins or cofactor are shown. Cofactors for MTA: BCL-6, ER-α, NRIF3, MAT1, MiCoA, LMO4 and FOG1; of CHD3/4: hunchback, Tramtrack69, KAP1, NAB2, RFP and Ikaros; of MBD2/3: MIZF and FAK; of RbAp46/48: FOG1 and BRCA1. Arrows indicate a direct interaction between the proteins.
Using GST pulldown, it has been reported that RbAp48 binds to both HDAC1 and HDAC2 (Nicolas et al., 2001; Taplick et al., 2001). In addition, it has been suggested that RbAp46 and RbAp48, together with HDAC1 and HDAC2, form a deacetylase core complex, based on the existence of these proteins in two major complexes, NuRD and Sin3 (Zhang et al., 1999). In addition, they can form a complex \textit{in vitro}, although other subunit(s) are required for full enzymatic activity (Zhang et al., 1999). It has been suggested that MTA1 and/or MTA2 are also essential for the formation of an active histone deacetylase core complex (Yao and Yang, 2003; Zhang et al., 1999).

Using a GST pull down assay, it has been demonstrated that MBD3 directly interacts with HDAC1, RbAp48, RbAp46 and MTA2 (Zhang et al., 1999; Saito and Ishikawa, 2002). Furthermore, MBD2 has been shown to associate with MTA2, HDAC1 and HDAC2, and target NuRD to a set of methylated genes (Ng et al., 1999; Zhang et al., 1999). It has been reported that MTA1 and MTA2 are also essential for the formation of an active histone deacetylase core complex (Yao and Yang, 2003; Zhang et al., 1999).

Using yeast two-hybrid and GST pulldown assays, it has been shown that p66α and p66β proteins bind to MBD2 and MBD3 (Feng et al., 2002; Brackertz et al., 2002; Brackertz et al., 2006; Le Guezenec et al., 2006). The structure of the protein complex formed between p66α and the coiled-coil domain of MBD2 has been solved (Gnanapragasam et al., 2011). Using immunoprecipitation, Gnanapragasam et al. showed that p66α interacts with four components of the NuRD complex \textit{in vivo}: MBD2, MTA2, RbAp48 and HDAC1 (Gnanapragasam et al., 2011). Another research group showed, using a GST pulldown experiment, that \textit{in vitro}, p66 protein interacts with MBD2, MBD3, RbAp46 and RbAp48 (Feng et al., 2002). Both p66 paralogues have been reported to be involved in interactions with the NuRD components MBD2/3, RbAp46/48, HDAC1 and MTA2, and have also been shown to associate with histone tails (Brackertz et al., 2002; Feng et al., 2002; Gong et al., 2006, Brackertz et al., 2006).

Posttranslational modifications also appear to play an important role in some NuRD interactions. SUMOylated p66α and p66β were found to interact with MBD3, RbAp46, RbAp48, and HDAC1 \textit{in vivo} (Gong et al., 2006). Mutating the SUMOylation sites of p66α and p66β disrupted the binding to HDAC1 and reduced the binding to RbAp46, respectively, whereas the other components were not affected by these mutations (Gong...
et al., 2006). This modification is required for their binding to HDAC1 and RbAp46 and is necessary for maximal gene repression activity (Gong et al., 2006).

It has been demonstrated that *in vitro* translated CHD4 protein, through its PHD zinc fingers, interacts with insect cell-expressed HDAC1 but not with the bacterially expressed GST-RbAp46 and RbAp48 (Zhang et al., 1998a). Other data similarly confirms a direct interaction between CHD3/4 and HDAC1 (Shimono et al., 2003).

MTA1 interacts directly with HDAC1 and HDAC2 both *in vitro* and *in vivo* (Mazumdar et al., 2001, Toh et al., 2000). Another example of the importance of posttranslational modifications in interactions within NuRD is that the acetylation of K626 in MTA1 has been shown to be essential for the function and interaction of MTA1–HDAC2 (Ohshiro et al., 2010). Using GST pulldown assay, it has been reported that MTA1 was able to pulldown RbAp46 and RbAp48, but not the other components of the NuRD complex (Roche et al., 2008). More recently, MTA2 was shown to interact with RbAp46 in a GST pulldown assay (Fu et al., 2011). Using immunoprecipitation, it has been demonstrated that MTA3 is associated with CHD4, HDAC1, HDAC2 and MBD3 (Fujita et al., 2003), and that MTA2 directly interacts with MBD3 (Saito and Ishikawa, 2002), whereas MTA1 binds to HDAC1 and HDAC2 (Mazumdar et al., 2001, Toh et al., 2000, Ohshiro et al., 2010).

Using yeast two-hybrid, it has been shown that the *Drosophila* NuRD subunit Nurf55 (p55, RbAp46/48 homologue) is able to bind to all other NuRD components (MTA-like, dMi-2 (CHD3/4 homologue), RPD3 (HDAC homologue) and MBD2/3) and interact directly with core histones (H3 and H4) (Marhold et al., 2004). Also, it has been shown using GST pulldown that MBD2 and MBD3 bind to p55 and Mi-2 (Marhold et al., 2004). In a study using antibodies against either *Drosophila* dMi-2 or p66, Kon et al. were able to immunoprecipitate p66 or dMi-2 *in vivo*, which suggests that dMi-2 and p66 form a complex (Kon et al., 2005). Taken together, the data indicate that RbAp proteins and p66 proteins represent hub molecules that mediate multiple interactions of the NuRD complex.

In spite of many reported interactions within the NuRD complex, less attention has been paid to overall shape of the NuRD complex and the structures of the constituent proteins, or how they interact with target proteins and are recruited to chromatin to carry out their function(s).
1.5 Aims of this thesis

Since its discovery, the biological functions of the NuRD complex and some of its components have been studied extensively. It has become increasingly apparent that the function of the NuRD complex is determined by the exact composition of the complex and on the associations of the individual subunits with other non-NuRD partners. In addition, most individual subunits of the NuRD complex have been shown to be fundamental players in protein interactions in different biological settings, and in many cases specific transcription factors appear to recruit NuRD through a range of interactions with different regions of individual NuRD subunits. Knowledge of the structure of the NuRD complex – that is, what the structures of the individual subunits are and how they interact with each other and with other coregulators – is an important step towards understanding its function. It is the goal of the work described in this thesis to work towards this knowledge.

In Chapter 2, efforts are described to work towards the structure of the N-terminal domains of MTA1. Although these domains are known to be important for protein-protein and/or protein-DNA interactions, none have been structurally characterized to date. In particular, the work focuses on the BAH, ELM2 and SANT domains.

Chapters 3–5 focus on the interaction between RbAp48 and MTA1. In Chapter 3, mapping experiments are carried out to determine which portion(s) of MTA1 are necessary and/or sufficient to mediate the interaction with RbAp48. Having defined this region, Chapters 3 and 4 describe X-ray crystal structures of complexes formed between RbAp48 and the RbAp48-binding region of MTA1.

Chapter 6 comprises a final discussion that puts the preceding experimental work into context.
Chapter 2. Structural studies of the N-terminal domains of MTA1

2.1 Introduction

As discussed in the previous chapter, MTA proteins are one of the defining components of the NuRD complex, and there is a large body of literature that points towards overexpressed MTA1 in particular having an important role in cancer progression. However, essentially nothing is known at the molecular level about how MTA1 either mediates its normal functions or acts in cancer. Many interactions of MTA proteins with other proteins have been reported, although it is unknown which domains of MTA proteins mediate these interactions (Section 1.4.4). The goal of the work in this Chapter was to begin to address this gap in our knowledge by investigating the three-dimensional structure of MTA1.

2.2 Sequence analysis and initial construct design

Analysis of the primary structure of MTA proteins from human shows that they are highly conserved in the N-terminal portion and are equipped with several predicted domains: namely BAH (bromo adjacent homology), ELM2 (Egl-27 and MTA1 homology), SANT (SWI, ADA2, N-CoR, TFIIB-B) and GATA-type ZnF domains (Figure 2-1). Our understanding of the likely functional importance of these domains in the MTA family is derived almost entirely from studies of other proteins containing such domains, which show that they are involved in protein-protein and/or protein-DNA interactions. For example, the BAH domain of the yeast Sir3 is involved in protein-protein interactions, including association with nucleosomes (Oliver et al., 2005, Armache et al., 2011), and SANT domains from SMRT (Silencing Mediator of Retinoid and Thyroid receptors) and Ada2 (Adaptor 2) have been implicated in protein interactions, and might bind to histone tails or DNA (Boyer et al., 2004, Mo et al., 2005, Yu et al., 2003, Aasland et al., 1996). The structure and function of the ELM2 domain is unknown; however, it has also been postulated to have protein or DNA binding ability. For example, it has been shown that the ELM2 domain of MIER1 (Mesoderm induction early response protein 1) is required for HDAC1 binding (Ding et al., 2003). Likewise, GATA-type zinc fingers have been implicated in both DNA and protein (Liew et al., 2005, Omichinski et al., 1993). None of the MTA domains has been characterised structurally, except the SANT domain from mouse MTA3 (PDB: 2CRG, unpublished), for which a solution structure has been determined, but not published.
Chapter 2

**Figure 2-1: Topologies of human MTA proteins.** Domain topologies of human MTA1, MTA2 and MTA3. The N-terminal half of each protein shares BAH (bromo-adjacent homology), ELM2 (Egl-27 and MTA1 homology), SANT (SWI, ADA2, N-CoR, TFIIB-B) and GATA-type zinc finger (ZnF) domains.

In order to begin to explore the structure of the N-terminal region of MTA1, we therefore designed and expressed a number of constructs encoding distinct combinations of the various predicted domains.

We first examined the primary amino acid sequence of MTA1 in order to choose appropriate boundaries for construct creation. The secondary and tertiary structure of MTA1 was predicted using Phyre2 software (Kelley and Sternberg, 2009), and a multiple sequence alignment was constructed from a range of species. The Phyre method uses structural alignments of homologous proteins of similar three-dimensional structure in the PDB to calculate the structural equivalence of residues. Secondary structure prediction for the N-terminal domains of MTA1 shows that more than 50% of MTA1 is predicted to have some sort of secondary structure (Figure 2-2).

Sequence alignment of MTA1 proteins from different species reveals high conservation between these sequences among mammalian proteins, with 96% identity between human, mouse and rat for example (Figure 2-3). Human MTA1 (residues 1–715) also shares 75% and 42% sequence identity with *Xenopus* and *Drosophila* MTA1-like, respectively (Figure 2-3), whereas *Drosophila* MTA1-like is less conserved.
Figure 2-2: Secondary structure prediction of the N-terminal region of MTA1 using the Phyre2 server. Green helices represent α-helices, blue arrows indicate β-strands and faint lines indicate coil. The ‘SS confidence’ line indicates the confidence in the prediction, with red being high confidence and blue low confidence (Kelley and Sternberg, 2009); Seq: amino acid sequence; SS: Secondary Structure prediction; SSC: Secondary Structure Confidence; D: Disorder; DC: Disorder Confidence.
Figure 2-3: Sequence alignment of the N-terminal domains of MTA1 proteins from several species. Protein sequences of human MTA1 (Uniprot: Q13330; residues 1–454), mouse MTA1 (Uniprot: Q8K4B0; residues 1–454), rat MTA1 (Uniprot: Q62599; residues 1–454), frog MTA1 (Uniprot: Q5U4R0; residues 1–459) and fly MTA1-like (Uniprot: Q9VNF6; residues 1–536), respectively, are aligned using ClustalW2. Identical amino acids are black and conserved amino acids are grey. Numbers found immediately to the right of the amino acid sequence correspond to the last residue of the protein listed in that row. The predicted secondary structure of human MTA1 is shown above the sequence.
Several proteins have similar domain arrangements to that seen in the MTA proteins: that is, BAH-ELM2-SANT-GATA-ZF. For example, human RERE (Arginine-glutamic acid dipeptide repeats) protein and C. elegans EGL-27 (Egg-laying defective protein 27) share the BAH-ELM2-SANT-ZnF domains with MTA proteins, and human CoREST (REST corepressor 1) and human MIER1 (Mesoderm induction early response protein 1) share the ELM2-SANT domains (Figure 2-4). It has been shown that EGL-27, a GATA transcription factor, promotes longevity in C. elegans by enhancing stress resistance (Xu and Kim, 2012).

**Figure 2-4: Proteins with domains similar to those in MTA proteins.** Proteins shown contain domains similar to MTA1: C. elegans EGL-27 (Uniprot: Q09228), human RERE (Uniprot: Q9P2R6), human CoREST (Uniprot: Q9UKL0) and human MIER1 (Uniprot: Q8N108).

Based on the observation that the overall domain structure of the N-terminal part of MTA1 is conserved among species, we initially chose to create and express constructs that encompassed two or more of the identified domains, to take account of the possibility that two or more domains might pack together and be required to form a stable protein. These experiments are described in the following Section (and are followed in turn in subsequent sections by more detailed analyses of the individual domains and the generation and testing of a range of additional constructs).
Constructs used for protein expression in this chapter were cloned and inserted into several different plasmids with different tags (pGEX-6P with an N-terminal GST tag, pHUE with an N-terminal His-Ubiquitin-tag, pMAL with an N-terminal MBP-tag and pET28b with a C-terminal His-tag). All of the constructs were cloned from their respective full-length human genes using a standard PCR approach and then ligated into plasmids via BamHI and EcoRI restriction sites. Primers used for PCR are listed in Appendix A.

2.3 Overexpression and purification of MTA1 constructs in bacteria

2.3.1 Overexpression trials of BAH-ELM2-SANT-ZnF and ELM2-SANT-ZnF of MTA1

In order to produce soluble protein, we first attempted to express the four domains of MTA1 together as a GST-fusion protein. Different N-terminal truncations of MTA1 were cloned into the pGEX-6P vector. Overexpression trials of all four domains together (BAH-ELM2-SANT-ZnF, aa 1–437 and aa 87–437) or ELM2-SANT-ZnF (aa 160–437) were conducted at 37 °C and 20 °C and different IPTG concentrations; expression was carried out in the Rosetta2 strain of E. coli (Section 7.3). MTA1 fragments that contained BAH-ELM2-SANT-ZnF or ELM2-SANT-ZnF were highly insoluble (Figure 2-5). Several different lysis buffers were trialled in attempts to increase protein solubility. Buffer changes included varying the salt concentration (150–500 mM NaCl) and changing both the type of buffer (Tris-HCl, MES, HEPES, phosphate) and the pH (Table 7-3). None of these conditions significantly improved the solubility of the protein and we therefore tried to find domain boundaries by designing new shorter constructs.
Figure 2-5: Overexpression of the N-terminal domains of MTA1 using pGEX-6P vector. A. Schematic diagram indicates the constructs used in protein expression. B and C. SDS-PAGE analysis of MTA1 (BAH-ELM2-SANT-ZnF, aa 1–437 and 87–437; and ELM2-SANT-ZnF, aa 160–437) induced at different conditions as indicated. Lane M, protein standards indicated in kDa; lanes Pre, T and S: total cell protein prior to induction, total and soluble fractions, respectively, for each fragment. Right panel: insoluble fractions for each condition. The arrows indicate the expected size of the MTA1 fragments. These figures were created by merging lanes from multiple gels as indicated.

2.3.2 Overexpression trials of BAH-ELM2-SANT and BAH-ELM2 of MTA1
As a first step, constructs were made that encoded BAH-ELM2-SANT (BES, aa 1–340) and BAH-ELM2 (BE, aa 1–284) combinations. These constructs were cloned into (a) the pGEX6P vector, for expression as an N-terminal GST fusion proteins; (b) the pHUE vector, for expression as a fusion protein with an N-terminal ubiquitin-His tag (Ub-His); and (c) the pMAL vector, for expression with an N-terminal MBP. All of these fusion
tags have been suggested to enhance protein solubility in certain cases (Cao and Zhang, 2004). The protein was overexpressed in each vector using Rosetta 2 and BL21 cells under the induction of IPTG either at 37 °C for 3 h or 20 °C overnight. As shown in Figure 2-6A and B, substantial levels of expression were observed in all tested conditions, but no soluble protein was observed. After extensive expression trials, it was found that none of these constructs produced a high level of soluble proteins in E. coli, and purification was not pursued. The same constructs were also expressed with an N-terminal His tag, but no increase in protein solubility was observed (data not shown).

Figure 2-6: Trial expression of BAH-ELM2 (BE) and BAH-ELM2-SANT (BES) domains of MTA1 using different plasmids with different tags. A. Schematic diagram indicates the constructs used in protein expression. B. SDS-PAGE analysis of protein samples induced at 37 °C for 3 h with 0.4 mM IPTG. Lane M, protein standards indicated in kDa; lanes T, S and Ins indicate total, soluble and insoluble fractions, respectively, for each fragment. C. SDS-PAGE analysis of protein samples induced at 20 °C overnight with 0.4 mM IPTG. Lane M, protein standards indicated in kDa; lanes Pre, S and Ins indicate total cell protein prior to induction, soluble and insoluble fractions, respectively, for each fragment. The arrows indicate the expected size of the MTA1 fragments. These figures were created by merging lanes from two gels.
In conclusion, although considerable effort was made to overexpress and purify these different constructs of MTA1 using bacterial expression systems (BL21 and Rosetta 2 cells) with different tags and buffers, no soluble tag-free protein could be isolated for biophysical studies.

2.4 Refolding MTA1 BAH-ELM2 and BAH-ELM2-SANT domains

As described above, extensive efforts were made to express and purify the N-terminal domains (BAH, ELM2, SANT and ZnF) of MTA1 from the soluble fraction of the bacterial extract using different bacteria strains, plasmids with different tags, temperatures, IPTG concentrations and buffers. Overall, many protocols were tested to produce soluble protein, but without success. Thus, we decided to change strategy, switching to purification from inclusion bodies followed by refolding of the protein of interest.

2.4.1 Refolding MTA1 BAH-ELM2 domains

To test this strategy, we chose the plasmid pHUE with an N-terminal His tag and ubiquitin, which can be refolded after denaturation. This plasmid has been used successfully in refolding other proteins (Kwan et al., 2006). We cloned BAH-ELM2(1–284) and BAH-ELM2-SANT(1–340) constructs into the pHUE plasmid. Both constructs were expressed as inclusion bodies in Rosetta2 E. coli and purified under denaturing conditions by using Ni-NTA agarose. The protocol for protein purification started with steps to wash the inclusion bodies, which were collected by centrifugation after cell breakage by sonication. BAH-ELM2 and BAH-ELM2-SANT were purified as outlined in Figure 2-7. After solubilising the protein in a buffer containing 6 M Guanidine-HCl (Gn-HCl) and a reducing agent (β-mercaptoethanol), the denatured soluble protein was purified by affinity chromatography using immobilised Ni-NTA agarose, then eluted from beads at about 2–3 mg/mL in a buffer containing 8 M urea (Section 7.4).
The eluted protein was subjected to a refolding screen to find the best conditions for refolding without aggregation. A set of test solutions were prepared, varying the type of buffer, pH, salt, detergent, glycerol and reducing agent (Table 2-1). The screen solution (190 µl) was added to solubilised protein (10 µl) using a 96-well plate. The plate was shaken for about 60 min, then protein aggregation was monitored by reading the absorbance at 320 nm and 390 nm (Burgess, 2009, Tresaugues et al., 2004). The solution does not absorb the light at these wavelengths, but rather any insoluble protein aggregates scatter the light and reduce the amount of transmitted light measured. A low absorption value at these wavelengths shows the absence of protein aggregation. The conditions that gave the lowest absorbance were used for larger scale refolding experiments.
### Table 2-1: Buffers used in refolding screen test

<table>
<thead>
<tr>
<th>Buffer</th>
<th>pH</th>
<th>NaCl</th>
<th>MgCl₂</th>
<th>Glycerol</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 mM acetate, pH 5.5</td>
<td>20 mM phosphate, pH 7.4</td>
<td>20 mM Tris, pH 8.0</td>
<td>20 mM Tris, pH 8.0</td>
<td></td>
</tr>
<tr>
<td>150 mM NaCl</td>
<td>137 mM NaCl</td>
<td>150mM NaCl</td>
<td>300 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>1 mM DTT</td>
<td>2.7 mM KCl</td>
<td>10mM MgCl₂</td>
<td>10% Glycerol</td>
<td></td>
</tr>
<tr>
<td>20 mM Tris, pH 7.5</td>
<td>20 mM phosphate, pH 7.0</td>
<td>20 mM Tris, pH 7.4</td>
<td>50 mM Tris, pH 7.4</td>
<td></td>
</tr>
<tr>
<td>150 mM NaCl</td>
<td>150 mM NaCl</td>
<td>150 mM NaCl</td>
<td>500 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>1 mM CaCl₂</td>
<td>20 mM phosphate, pH 7.4</td>
<td>20 mM Tris, pH 7.0</td>
<td>10% Glycerol</td>
<td></td>
</tr>
<tr>
<td>20 mM Tris, pH 7.5</td>
<td>20 mM phosphate, pH 7.0</td>
<td>20 mM Tris, pH 7.4</td>
<td>50 mM Tris, pH 7.4</td>
<td></td>
</tr>
<tr>
<td>150mM NaCl</td>
<td>150mM NaCl</td>
<td>150mM NaCl</td>
<td>100 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>7.5mM DTT</td>
<td>5 mM ZnSO₄</td>
<td>50 mM Tris, pH 7.4</td>
<td>150 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>20 mM phosphate, pH 7.4</td>
<td>50 HEPES, pH 7.0</td>
<td>50 HEPES, pH 8.0</td>
<td>50 HEPES, pH 8.0</td>
<td></td>
</tr>
<tr>
<td>100 mM NaCl</td>
<td>200 mM NaCl</td>
<td>100 mM NaCl</td>
<td>150 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>0.5% Triton X-100</td>
<td>0.5% Triton X-100</td>
<td>0.5% Triton X-100</td>
<td>100 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>50 Tris, pH 9.0</td>
<td>100 mM NaCl</td>
<td>100 mM phosphate, pH 8.0</td>
<td>100 mM phosphate, pH 8.0</td>
<td></td>
</tr>
<tr>
<td>300 mM NaCl</td>
<td>100 mM NaCl</td>
<td>100 mM NaCl</td>
<td>100 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>1 mM DTT</td>
<td>0.01% Triton X-100</td>
<td>0.01% Triton X-100</td>
<td>10 mM phosphate, pH 8.0</td>
<td></td>
</tr>
<tr>
<td>0.01% Triton X-100</td>
<td>10 mM phosphate, pH 9.0</td>
<td>100 mM NaCl</td>
<td>10 mM Tris, pH 9.0</td>
<td></td>
</tr>
<tr>
<td>10 mM phosphate</td>
<td>100 mM NaCl</td>
<td>100 mM NaCl</td>
<td>100 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>10 mM Tris, pH 8.5</td>
<td>100 mM NaCl</td>
<td>100 mM NaCl</td>
<td>100 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>10 mM Tris at pH 8.5</td>
<td>10 mM Arg</td>
<td>10 mM Arg</td>
<td>10 mM Arg</td>
<td></td>
</tr>
<tr>
<td>10 mM Tris, 10 mM phosphate at pH 9.0</td>
<td>50mM MES, pH 6.0</td>
<td>300 mM NaCl</td>
<td>300 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>150 mM NaCl</td>
<td>150 mM NaCl</td>
<td>150 mM NaCl</td>
<td>150 mM NaCl</td>
<td></td>
</tr>
</tbody>
</table>

Based on the refolding screen, BAH-ELM2 (~32 kDa, theoretical pI 5.3) was refolded by dialysis against 20 mM Tris or 100 mM phosphate at several pH values (7.0, 8.0, 9.0), 100 mM NaCl, 10% glycerol and 4 M urea; the urea concentration was reduced gradually with each buffer exchange (4 M, 2 M then 0 M). This multi-step dialysis was performed at 4 °C and at room temperature. The sample refolded in Tris buffer at pH...
8.0 recovered more soluble protein with less precipitation (Figure 2-8A). The protein was then dialysed into cleavage buffer (20 mM Tris pH 8.0, 50 mM NaCl, 1 mM CaCl₂) for the removal of His-ubiquitin tag using UBP41 protease, before being subjected to size exclusion chromatography (Figure 2-8). Some of the protein was lost by precipitation after tag cleavage. Protein refolding at pH 8.0 recovered a good yield, but the protein was not clean as can be seen from the size exclusion chromatography fractions (Figure 2-8A). We then tried to re-bind the protein to fresh nickel beads in order to capture the uncleaved protein and elute the BAH-ELM2 protein, but both proteins aggregated on the beads as can be seen in Figure 2-8A (last panel). It could be that the BAH-ELM2 protein had aggregated with the uncleaved protein as they eluted in the same fractions of size exclusion chromatography.

We also tried to refold BAH-ELM2 by dialysis against 10 mM phosphate, 10 mM Tris, 10 mM acetate at pH (7.0 and 9.0) and 150 mM NaCl in three sequential dialysis steps. This multi-step dialysis was performed at 4 °C. The yield of refolded protein was higher at pH 9.0 than at pH 7.0 (Figure 2-8B). Thus, we proceeded with the sample obtained at pH 9.0. The protein was then dialysed into the cleavage buffer for the removal of His-ubiquitin tag using UBP41 protease, before being subjected to size exclusion chromatography. The yield here was less than that in the previous attempt but the protein was more pure (Figure 2-8B).

Biophysical methods like NMR and circular dichroism (CD) may be used to assess the folded state of the fusion protein. Size exclusion chromatography fractions of BAH-ELM2 (Figure 2-8B) were combined, concentrated and used in 1D ¹H NMR analysis. The amide region, which ranges from 10.5 ppm to 6.5 ppm in a 1D ¹H NMR spectrum, is indicative of the degree of folding. The spectrum of a correctly folded protein contains well dispersed and relatively sharp peaks in this region, whereas broader peaks indicate aggregated or unfolded proteins (Rehm et al., 2002). Although some signal intensity is observed at >9 ppm, the appearance of the spectrum of BAH-ELM2 overall suggests that the protein does not take up a well-defined conformation (Figure 2-9).
Figure 2-8: Purification of denatured BAH-ELM2. A. SDS-PAGE analysis of protein samples taken throughout the purification process at pH 8.0. Lane M, protein standards indicated in kDa; lane T, total protein; lane S, soluble protein; lane ds, denatured protein; lane washes, eluted protein during wash steps; lane E, protein eluted from beads; lane bc, protein before cleavage with UBP41 protease; lane ac, after cleavage; lane ppt, precipitated protein after cleavage; SEC fractions, size exclusion fractions; lane be, protein after rebinding to fresh beads; lane eb, protein eluted from beads. B. SDS-PAGE analysis of protein samples taken throughout the purification/refolding process at pH 7.0 and 9.0. Lane washes, protein eluted during wash steps; lane E, eluted protein from beads; R-pH7.0, refolded protein at pH 7.0; R-pH9.0, refolded protein at pH 9.0; lane ac, protein after cleavage with UBP41 protease; lane LS, loaded sample into SEC; SEC fractions, size exclusion fractions. This figure was created by merging lanes from multiple gels (as indicated by dashed lines).
Chapter 2

Figure 2-9: 1D $^1$H NMR spectrum of BAH-ELM2 domain. The amide and aliphatic regions of a 1D $^1$H-NMR spectrum are shown for the MTA1 BAH-ELM2 domain (32 kDa) recorded in 20 mM Tris pH 7.4, 50 mM NaCl, 1 mM DTT, 5% (v/v) D$_2$O and 2 μM DSS at 298 K. The spectrum was recorded on a 600-MHz spectrometer equipped with high-resolution cryogenic TCI probeheads.

2.4.2 Refolding MTA1 BAH-ELM2-SANT domains

Similar steps were undertaken to refold BAH-ELM2-SANT from inclusion bodies using a range of conditions. The general steps used for preparation of the inclusion bodies, solubilisation of the protein and Ni-NTA affinity chromatography were the same as described above for BAH-ELM2. BAH-ELM2-SANT (~39 kDa, theoretical pI 5.3) was refolded by dialysis against 50 mM Tris pH 8.0, 100 mM NaCl, 10% glycerol and 4 M urea; the urea concentration was reduced gradually with each buffer exchange (4 M, 2 M then 0 M). This multi-step dialysis was performed at 4 °C. The protein was then dialysed into cleavage buffer for the removal of His-ubiquitin tag, before being subjected to size exclusion chromatography. The cleavage was not complete and BAH-ELM2-SANT and the fusion protein were eluted in the same fraction of size exclusion chromatography (Figure 2-10A). We then tried to re-bind the protein to fresh nickel beads in order to capture the uncleaved protein and elute the BAH-ELM2-SANT protein, but both proteins aggregated on the beads.
We also tried to refold BAH-ELM2-SANT by dialysis against 10 mM phosphate, 10 mM Tris, 10 mM acetate at pH 9.0 and 150 mM NaCl or 20 mM Tris at pH 8.5, 100 mM NaCl, 10 mM Arg in three sequential dialysis steps (Figure 2-10B). The yield of refolded protein at pH 8.5 was higher than at pH 9.0 (Figure 2-10B). Thus, we proceeded with the sample obtained at pH 8.5. The protein was then dialysed into the cleavage buffer for the removal of His-ubiquitin tag, before being subjected to size exclusion chromatography. Most of the protein was precipitated after the cleavage of the His-ubiquitin tag (Figure 2-10B). The very small amount of BAH-ELM2-SANT obtained from size exclusion chromatography was used in CD analysis to investigate the secondary structure of the protein.

Using Tris buffer at pH 8.5 or Tris/phosphate/acetate mixed buffer at pH 9.0 yielded a soluble protein, in contrast to Tris buffer at pH 8.0 (Figure 2-10A and B). The cleavage of the refolded BAH-ELM2-SANT construct at pH 8.0 was not complete, and the protein aggregated and eluted in the same fractions as the fusion protein from size exclusion chromatography (Figure 2-10A). By increasing the pH to 8.5 or 9.0, we found a lot of precipitation after cleaving the tag. Although we separated the proteins by size exclusion chromatography, the yield was not sufficient for structural characterisation (Figure 2-10B). In addition, we tried to refold the BAH-ELM2 and BAH-ELM2-SANT domains without success using a range of other conditions that are not shown here. In most attempts to refold protein from inclusion bodies, the protein was precipitated during dialysis and at the end the yield was not sufficient to be workable.
2.4.3 Estimate of secondary structure of BAH-ELM2-SANT

Far UV-CD is a valuable method for estimating protein secondary structure as different secondary structural elements yield characteristic far UV-CD spectra. The technique does not require a high concentration of protein (~1–20 μM). In order to check the secondary structure of refolded BAH-ELM2-SANT, sample obtained from size exclusion chromatography was dialysed into 20 mM Tris pH 8.0, 50 mM NaCl and then the far-UV CD spectra was recorded (Figure 2-11A). The spectrum is consistent with a mixture α-helical and β-sheet structure as can be seen from the ideal spectra (Figure 2-11B). This result is consistent with predictions that these domains contain both α and β (Figure 2-3). Unfortunately, the amount of protein obtained from this process (1 L
culture) was very low and not sufficient for additional investigations. Nonetheless, this purification strategy could potentially be used if no better approach can be found.

Figure 2-11: Far-UVCd spectrum of BAH-ELM2-SANT. A. This spectrum was collected using a 12 µM protein sample in 20 mM Tris pH 8.0, 50 mM NaCl at 20 °C. B. Ideal CD spectra for random coil (red), β-sheet (blue) and α-helix (orange) (Bolotina et al., 1980).

It seems that for most MTA1/2 constructs, the protein was expressed in the insoluble fraction. In addition, the attempts to refold these fragments after denaturation were not particularly successful, as the protein either aggregated or was produced in very low yields. Thus, we attempted to obtain soluble, folded protein by screening expression in insect cells using a small scale expression system.

2.5 Overexpression and purification of ELM2 and ELM2-SANT domains

It has been shown that most proteins that contain an ELM2 domain also possess a SANT domain, which might imply a functional and/or structural relationship between these two domains (see for example Figure 2-4) (Ding et al., 2003). Sequence alignment of several human proteins that share ELM2-SANT domains shows some conserved residues in the linker between these two domains (Figure 2-12) which might be an indication of the importance of this region in protein structure and function and therefore that the two domains can act as a single unit. For example, Ding et al. (2003) showed that MIER1 represses gene transcription through the association of the ELM2
domain (and possibly also the SANT domain) with HDAC1. It has also been reported that the ELM2-SANT double domain of RERE (also called atrophin) interacts with HDAC1/2 and mediates the latter’s histone deacetylase activity (Wang et al., 2006).

Figure 2-12: Sequence alignment of some ELM2-SANT-containing proteins. The amino acid sequences from the C-terminal end of ELM2 to the beginning of the SANT domain are shown. Residues from each predicted domain are shaded in grey. Highly conserved residues in the linker between the two domains are highlighted in yellow. The numbers listed above the alignment correspond to amino acid positions in the human MTA1 protein sequence. The Uniprot for hMIER1, hCoREST and hRERE are Q8N108, Q9UKL0 and Q9P2R6, respectively.

The structures of SANT domain from mouse MTA3 (PDB: 2CRG, unpublished), the structure of SANT domain of RERE (PDB: 2YQK, unpublished) and the structure of the deacetylase activation domain (DAD) of SMRT (also known as N-CoR2, Nuclear receptor corepressor 2) alone (Codina et al., 2005) and bound to HDAC3 (Watson et al., 2012b) have been solved (Figure 2-13A). SANT and DAD domains share sequence and structural similarities (Figure 2-13A and D), although the surface of the DAD domain is a mix of basic and acidic patches, whereas the equivalent view of the ISWI SANT domain shows a rather acidic surface (Codina et al., 2005). Upon binding to HDAC3, the N-terminal helix of the DAD domain undergoes structural rearrangement – compared to the NMR structure of DAD alone – creating an N-terminal helix that makes several intermolecular interactions with HDAC3 (Figure 2-13A and B). Interestingly, a molecule of inositol tetraphosphate (Ins(1,4,5,6)P4) is also present at the HDAC3-DAD interface, and makes what are likely to be essential interactions to stabilise the complex (Watson et al., 2012b).

In this context, it is notable that MTA1 has been shown to bind HDAC1/2 (Mazumdar et al., 2001, Toh et al., 2000), and the SANT or ELM2-SANT domains of MTA protein might be involved in this interaction. Sequence and structural alignment of HDAC1/2/3 shows that they are highly conserved and the key residues that mediate the interaction
with DAD are identical (Figure 2-13B and C). Furthermore, sequence alignment of the ELM2-SANT domains of MTA proteins with equivalent region of RERE and SMRT reveals that the HDAC3-binding residues of the DAD domain (SMRT) are conserved in the SANT domains of the RERE and MTA proteins (Figure 2-13D). Taken together, we propose that the SANT or ELM2-SANT domains of MTA proteins could bind to HDAC1/2 in a similar manner. Therefore, we cloned ELM2 alone and both ELM2-SANT domains into pGEX-6P and pMAL plasmids for protein expression.
Figure 2-13: Structure of SANT domain. A. Structural alignment of SANT domain of MTA3 protein (orange, PDB: 2CRG, unpublished) with SANT domain of RERE (salmon, PDB: 2YQK, unpublished) and Deacetylase Activation Domain (DAD) (cyan, PDB: 1xC5) from SMRT (Uniprot: Q9Y618) (Codina et al., 2005). B. Structural alignment of HDAC2 (yellow) (Uniprot: Q92769) bound to N-(2-aminophenyl) benzamide inhibitor (blue, PDB: 3MAX) (Bressi et al., 2010), and HDAC3 (gold) (Uniprot: O15379) bound to Ins(1,4,5,6)P4 (blue) and DAD domain (cyan) (PDB: 4A69) (Watson et al., 2012b). C. Alignments of key regions of human HDAC1/2/3. Residues that mediate interaction with SMRT/DAD are highlighted in green. Identical residues are highlighted in yellow and conserved residues in blue. The numbers listed above the alignment correspond to amino acid
positions in the human HDAC3 protein sequence. D. Sequence alignment of SANT domains of MTA proteins and two other known interaction partners for class I HDACs, namely RERE and SMRT. Residues that mediate the interaction of SMRT/DAD with HDAC3 are highlighted in green. Identical residues in at least four proteins are highlighted in yellow, and conserved residues are highlighted in blue. The underlined sequence of MTA3, RERE and SMRT indicates residues presented in the structures. The known secondary structure of the SANT/DAD domain is indicated on the top of the sequence. The numbers listed above the alignment correspond to amino acid positions in the human MTA1 protein sequence.

Small scale overexpression trials of the ELM2 (161–276, ~13 kDa) and ELM2-SANT (161–340, ~20.8 kDa) domains of MTA1 were conducted at 15 °C and 37 °C, in which Rosetta 2 cells were induced with 0.4 mM IPTG. The GST-ELM2-SANT and GST-ELM2 fusion proteins were insoluble; the only soluble protein under these conditions was MBP-ELM2-SANT (Figure 2-14). Large-scale overexpression of this latter construct was therefore carried out. After primary purification using amylose affinity chromatography and treatment with human rhinovirus 3C (HRV 3C) protease, the protein was further subjected to ion exchange chromatography and size exclusion chromatography (Section 7.3 and Figure 2-15).

![Figure 2-14: Small scale overexpression of ELM2 and ELM2-SANT domains of MTA1. A. Schematic diagram indicates the constructs used in protein expression. B. SDS-PAGE analysis of expression trials at different temperatures in Rosetta 2 cells for ELM2-SANT and ELM2 domains of MTA1 fused to GST or MBP. Lane M, protein standards indicated in kDa; lanes T, S indicate total and soluble fractions, respectively. The arrows indicate the expected size of each construct. C. SDS-PAGE analysis of expression trials of GST-ELM2 at different temperatures in Rosetta 2 cells. Lane M, protein standards indicated in kDa; lanes Pre, T and S indicate total cell protein prior to induction, total and soluble fractions after induction, respectively. The arrows indicate the expected size of each construct.](image-url)
Figure 2-15: Purification of large scale overexpression of ELM2-SANT at 37 °C. A. SDS-PAGE analysis of ELM2-SANT during amylose affinity chromatography and after treatment with HRV 3C protease. Lane M, molecular weight standards indicated in kDa; lane E, elution fractions from amylose affinity chromatography following treatment of amylose-bound MBP-ELM2-SANT with 60 mM Maltose; lane ac, after HRV 3C cleavage; lanes IEC fractions, ion exchange chromatography fractions; SEC fractions, size exclusion chromatography fractions. This figure was created by merging lanes from multiple SDS-PAGE gels (indicated by the dashed lines). B. Elution profile from a SEC of ELM2-SANT.

The sample of ELM2-SANT obtained from SEC displayed a good level of purity (~80–90% judging from SDS-PAGE) and was therefore subjected to crystallisation trials for possible structure determination using X-ray crystallography.
2.5.1 Attempts to crystallise ELM2-SANT domain

Several parameters affect protein crystallisation, which include protein purity and concentration, buffer pH, type and concentration of precipitates, temperature and the period of time needed for protein to crystallise (Chayen, 2004). Thus, crystallisation conditions for each protein should be determined empirically.

Typically, crystallisation conditions for a given protein can be determined by high-throughput screening on a small scale using commercially available kits (Joachimiak, 2009). The successful condition(s) could be used or further optimised to obtain suitable crystals for x-ray diffraction. Therefore, crystallisation trials of ELM2-SANT at ~7 and 14 mg/mL at 18 °C and 22 °C were conducted using several Qiagen NeXtal DWBlock screening kits: namely, Classics Suite, JCSG Core, PACT Suite, PEGs suite and JCSG suite. Initial crystallisation screens were set up using a MOSQUITO crystallization robot (TTP Labtech, USA) as sitting drops with 0.2 µL each of protein solution and crystallisation solution, equilibrated against 75 µL of the crystallization solution in 96-well trays. Crystal trays were kept at 18 °C or 22 °C, and crystal growth was monitored during a period of several months (Section 7.12).

As shown in Figure 2-16, crystalline-like features (including small clusters of needles) were observed in a number of conditions. The crystalline-like particles obtained were either too small for testing or were clusters of needles, thus none were appropriate for data collection. Further optimisation efforts are necessary to improve crystal size and quality. However, these were not attempted due to time constraints.
Crystallisation conditions are:

- **A.** 1.6 M Tris-sodium citrate pH 6.5.
- **B.** 0.1 M SPG buffer pH 7 and 25% (w/v) PEG 1500.
- **C.** 0.2 M sodium chloride, 0.1 M HEPES pH 7 and 20% (w/v) PEG 6000 135731;
- **D.** 0.2 M sodium malonate, 0.1 M Bis-Tris propane pH 6.5 and 20% (w/v) PEG 3350.
- **E.** 0.2 M sodium citrate, 0.1 M Bis-Tris propane pH 7.5 and 20% (w/v) PEG 3350.
- **F.** 0.2 M potassium thiocyanate, 0.1 M Bis Tris propane pH 7.5 and 20% (w/v) PEG 3350.
- **G.** 0.1 M MES pH 6.5 and 20% (w/v) PEG 10000.
- **H.** 0.1 M sodium HEPES pH 7.5 and 25% (w/v) PEG 4000.

2.6 Expression of the BAH domain from MTA1 and MTA2

The BAH domain has been found in a large number of proteins involved in transcriptional regulation; functional evidence available so far indicate that it is involved in protein-protein interactions and, in particular, binds to nucleosomes. Many structures of BAH domains from different proteins, either alone or in complex with a partner, have been determined. For example, the crystal structures of the proximal BAH domain from chicken polybromo (BAH1) (Figure 2-17A) (Oliver et al., 2005) and the BAH domain from *S. cerevisiae* Sir3 (Connelly et al., 2006) in isolation, and the crystal structure of the BAH domain of mouse ORC1 bound to a methylated histone H4 peptide (H4(14–25)K20me2) (Figure 2-17B) (Kuo et al., 2012) have been reported. Sequence
alignment of the BAH domain of MTA1/2 with known BAH domains of polybromo and ORC1 shows relatively low conservation and in fact substantial differences are also observed in the predicted secondary structure of MTA1/2 compared to the other two domains (Figure 2-17C).

Figure 2-17: Comparison of MTA1/2 BAH domain with two BAH domains of known structure. A. Crystal structure of the proximal BAH domain from chicken polybromo (PDB: 1W4S) (Oliver et al., 2005). B. Solution structure of the BAH domain from ORC1 protein bound to histone H4K20me2 (PDB: 4DOW) (Kuo et al., 2012). C. Sequence alignment of MTA1/2 BAH domain with the proximal BAH domain of polybromo (Poly) (Uniprot: Q90941) and the ORC1 BAH domain (Uniprot: Q9Z1N2). The secondary structure prediction of MTA1 BAH domain from Phyre is indicated on the top of the sequence alignment. The known secondary structure of proximal BAH domain is indicated below the alignment. The amino acids highlighted in yellow are identical in MTA1 and MTA2 and are also observed in at least in one of the domains, whereas amino acids highlighted in blue are similar but not identical.
It has also been shown that the BAH domain from Sir3 binds to nucleosomes (Figure 2-18B) (Armache et al., 2011). Sequence alignment of the BAH domain from MTA1 with the Sir3 BAH domain shows some sequence similarity between the two domains (Figure 2-18A). MTA1 seems to lack the H2A binding site residues, and MTA1(9-154) is the region of best conservation with the Sir3 BAH domain.

Thus, overall the secondary structure prediction and sequence alignment of the MTA1 BAH domain with known BAH domains show low sequence conservation, which suggests that the structure might have unique features, but which makes it difficult to predict the domain boundaries for protein expression with confidence.

Several truncations of the BAH domain from MTA1 and MTA2 were therefore cloned into different plasmids with different tags. MTA1 BAH domain (9–154 and 1–179) was cloned into pET28a with a C-terminal His tag; MTA1(1–168) was cloned into pGEX6P, pHUE and pMAL; and MTA2(1–150 and 1–160) were cloned into pGEX-6P. Trial expression was tested in a range of conditions (host cell type: BL21 and Rosetta 2; incubation temperature and length: 37 °C for 3 h, 25 °C, 20 °C and 18 °C overnight; IPTG concentration: 0.1, 0.4 and 1.0 mM).
Figure 2-18: Binding of BAH domain to nucleosome. A. Crystal structure of the yeast Sir3 BAH domain in complex with a nucleosome core particle (PDB: 3TU4) (Armache et al., 2011). B. Sequence alignment of Sir3 BAH domain (Uniprot: P06701) with BAH domain of MTA1. Colours indicate which histone contacts the highlighted residue of BAH domain (key below the alignment). The known secondary structure of Sir3 BAH domain is indicated above the sequence. The Phyre-derived secondary structure prediction of MTA1 BAH domain is indicated below its sequence alignment. Conserved residues are highlighted in grey.
For BAH(1–168) fused to GST and Ub-His, the Ub-His construct was insoluble, whereas a small fraction of the GST construct (perhaps ~10%) was soluble under the same conditions (Figure 2-19A). Efforts were therefore made to purify the GST-BAH(1–168) obtained under the following expression conditions: Rosetta 2 cells, 20 °C for 18 h, 0.4 mM IPTG. The GST-fusion protein was purified from the soluble fraction of the bacterial lysate by affinity chromatography using glutathione-Sepharose (GSH) beads (Section 7.3.4). Although we obtained some soluble protein bound to the beads, the cleavage of the GST tag was incomplete and the yield of cleaved protein was very low (Figure 2-19B). Several different buffer conditions were trialled in attempts to increase protein solubility during lysis and purification. Changes included varying salt concentrations (150–1000 mM NaCl), detergents and changing the type of buffer (Tris-HCl, MES, HEPES, phosphate, acetate) and the pH (5.5, 6, 7, 8 and 9; theoretical pI of MTA1 BAH alone= 5.3) (Table 7-3). None of these conditions significantly improved the solubility of the protein (data not shown).

![Figure 2-19: Overexpression of the BAH domain from MTA1. A. Schematic diagram indicates the construct used in protein expression. B. SDS-PAGE analysis of BAH domain (1–168), in pGEX-6P and pHUE, induced at 20 °C overnight with 0.1, 0.4 and 1.0 mM IPTG. Lane M, protein standards indicated in kDa; lanes T and S, total and soluble fractions, respectively. C. SDS-PAGE analysis of GST-BAH purification using GSH beads. Lane M, protein standards indicated in kDa; lane T, total protein; lane S, soluble fraction; lane Ins, insoluble fraction; lane FT, flow-through (unbound protein) from GSH beads; lane E, eluted protein from beads; lane AC, protein after tag cleavage with HRV 3C protease. This figure was created by merging lanes from two gels (as indicated by the dashed lines).](image-url)
In case of the C-terminal His-tagged MTA1-BAH constructs (9–154 and 1–179), the only condition that produced detectable expression was 37 °C for 3 h and induction with 0.5 mM IPTG, and almost all the protein was insoluble (Figure 2-20A). The expression was not visible on SDS-PAGE under other conditions using a lower temperature and different IPTG concentrations (Figure 2-20A). As above, varying the buffer components and pH did not help to increase protein solubility.

It is notable that a part of the BAH domain of MTA1 that is predicted to be disordered (residues 69–96) is missing from MTA2 and MTA3, which suggests that the BAH domains of these proteins might be better candidates for expression (Figure 2-20B). Thus, two constructs of the MTA2 BAH domain (1–150 and 1–160) were cloned into pGEX6P as GST-fusion proteins. As shown in Figure 2-20C, trial expressions were conducted under different conditions: temperature, IPTG concentration, and length of induction (Table 7-6). Different buffers with different components were also tried to improve protein solubility (Table 7-3); however, under all conditions trialled, the protein was mostly insoluble.
Figure 2-20: Trial expression of different fragments from MTA1/2 BAH domains. 

A. Schematic diagram indicates the MTA1/MTA2 constructs used in protein expression.

B. SDS-PAGE of overexpression of MTA1 BAH (aa 9–154 and 1–179) in BL21 cells at different temperatures and IPTG concentrations as indicated. Lane M, protein standards indicated in kDa; lanes S and Ins, soluble and insoluble fractions, respectively.

C. Sequence alignment of BAH domains from the human MTA protein family showing that they are highly conserved and indicating the missing region of MTA2/3.

D. SDS-PAGE showing overexpression trials of BAH domain from MTA2 using BL21 and Rosetta 2 cells at different conditions as indicated. Lane M, protein standards indicated in kDa; lanes T, S and Ins, total, soluble and insoluble fractions, respectively, for each condition. This figure was created by merging lanes from two gels.

Next an attempt was made to express the BAH domain of MTA1 fused to MBP, which is known to increase protein solubility. Thus, we cloned the BAH domain of MTA1(1–168) into pMAL, with a Factor Xa protease site. High expression level was obtained when MBP-BAH was induced at 20 °C for 18 h with 0.4 mM IPTG in Rosetta 2 cells.
(Figure 2-21). We therefore scaled up the expression using this condition and purified
the MBP-fusion protein from the soluble fraction of the bacterial lysate by affinity
chromatography using immobilised amylose resin. The MBP tag was then removed by
treatment with Factor Xa protease (Figure 2-21A). The cleaved protein was further
purified by ion exchange chromatography and size exclusion chromatography (Figure 2-
21A). However, SDS-PAGE analysis of the protein showed several bands which could
be the result of nonspecific proteolysis at secondary sites by Factor Xa protease (Figure
2-21A). Thus, we cloned the same construct into another pMAL vector with a HRV 3C
cleavage site, which resulted in more specific cleavage, although the tag cleavage was
not complete (Figure 2-21B).
Figure 2-21: MBP-BAH domain overexpression. A. SDS-PAGE analysis of cell and protein samples taken throughout the purification process of MBP-BAH with a Factor Xa cleavage site. Lane M, protein standards indicated in kDa; lane T, total cell protein; lane S, soluble fraction of cell lysate; lane Ins, insoluble fraction; lane FT, amylose-resin flow-through of soluble cell lysate; lanes E, eluted protein from resin; lane bc, eluted protein before cleavage with Factor Xa; lane ac, eluted protein after cleavage; SEC fractions, size exclusion chromatography fractions. B. SDS-PAGE analysis of cell and protein samples taken throughout the purification process of MBP-BAH with a HRV 3C cleavage site. Lane M, protein standards indicated in kDa; lane T: total cell protein; lane S: soluble fraction of cell lysate; lane FT, amylose-resin flow-through of soluble cell lysate; lanes E, eluted protein from resin; lane ac, eluted protein after cleavage with HRV 3C protease; the last two panels indicate ion-exchange chromatography fractions (IEC fractions) and size exclusion fractions (SEC fractions). This figure was created by merging lanes from multiple gels (as indicated by dashed lines). C. SEC elution profile of BAH domain in B.
After subjecting the cleaved proteins to ion exchange chromatography and size exclusion chromatography, the BAH domain eluted in the same fractions as the fusion protein and the MBP, despite its much smaller molecular mass, suggesting the presence of aggregation. Unfortunately, different buffer components (Table 7-2) did not help to separate the proteins (Figure 2-21B). MBP was an excellent solubilising agent for BAH domain, but it seems that the BAH domain exists in a soluble but non-native form that resists aggregation only as long as it remains fused to MBP.

Although bacterial expression systems are one of the most widely used forms for producing recombinant protein, the proteins generally lack the PTMs that take place in eukaryotes. We therefore turned to an insect cell expression system.

### 2.7 InsectDirect expression of N-terminal domains of MTA1

There are several advantages of insect cells over *E. coli* such as improved solubility and ability to integrate post-translational modifications. In recent years, the most commonly used eukaryotic expression system for the large-scale preparation of eukaryotic proteins for structural biology has been the baculovirus directed insect cell approach. However, the use of the baculovirus technique for protein expression can be time consuming and laborious because of the numerous steps involved. The InsectDirect System represents a potential alternative to a baculovirus expression system, at least for the initial screening of constructs. It has been shown that the InsectDirect System, which involves insect cell expression vectors (pIEX), a transfection reagent and *Spodoptera frugiperda* insect cells (Sf9), allows rapid, small-scale expression screening in insect cells (Loomis et al., 2005), and in principle can be scaled up to produce milligram quantities of target protein, although it is very expensive to scale up (because of the price of the transfection reagent).

In an effort to obtain soluble protein corresponding to one or more of the N-terminal domains of MTA1, small-scale expression screening was carried out in Sf9 cells using the pIEX expression vector. Five constructs covering different truncations of the N-terminal domains of MTA1, namely (9–154), (1–287), (155–334), (155–354) and (1–334) (Figure 2-22), were cloned into pIEX plasmid with an N-terminal FLAG tag. Primers used for PCR are listed in Appendix B. We performed transfections of $1.5 \times 10^6$ cells in a 6-well tissue culture dish with 2 µg DNA of the pIEX-MTA1 constructs using Insect Gene Juice transfection reagent, and left the cells to grow at 28 °C for ~60 h. A
Chapter 2

pIEX plasmid harbouring the GFP gene was used as a positive control for protein expression (Section 7.5). We lysed the cells and separated the soluble from the insoluble fraction by centrifugation. A Western blot was carried out using anti-FLAG antibody to check the expression of these constructs of MTA1. This system yielded expression of at least three constructs – MTA1(1–287), (155–354) and (1–334) – and a high proportion of the protein was in the soluble form (Figure 2-22).

![Western blot image](image)

**Figure 2-22: Expression of the N-terminal domains of MTA1 in insect cells.** InsectDirect was employed to express different constructs of BAH, BAH-ELM2 and BAH-ELM2-SANT domains in Sf9 cells. Left: Schematic diagram indicates the constructs used in protein expression and their sizes in kDa. Right: Western blot of soluble and insoluble fractions for each construct using anti-FLAG antibody. Negative control (-ve) indicates uninfected cells. Arrows indicate expected size of protein constructs.

As initial screening using the InsectDirect system indicated that several of the constructs could be expressed successfully in insect cells, and we attempted to scale up expression of two of the constructs using the baculovirus system.

### 2.8 Attempts to express MTA1 BAH and BAH-ELM2 using baculovirus

#### 2.8.1 Principles of Baculovirus Expression Vectors

The Baculovirus Expression Vector System (BEVS) has been widely used for the production of recombinant proteins in cultured insect cells (Hitchman et al., 2009). Baculoviruses are rod-shaped, enveloped viruses with circular, supercoiled double-stranded DNA molecules (~130 kbp) that duplicate in the nuclei of host cells (Summers and Anderson, 1972). The baculoviruses are capable of infecting insect cells and have been adapted to express proteins in insect cells. The Bac-to-Bac system (Invitrogen) is
based on site-specific transposition of an expression cassette into a baculovirus shuttle vector (known as a bacmid).

The foreign gene is first cloned into a donor plasmid under the transcriptional control of a strong promoter, usually that of the polyhedrin gene. Donor plasmid containing the foreign gene is transformed into *E. coli* DH10Bac cells, which contain a baculovirus shuttle vector (bacmid) with a mini-attTn7 target site and a helper plasmid. In the presence of transposition proteins provided by the helper plasmid, transposition occurs between the mini-Tn7 element on the donor plasmid and the mini-attTn7 target site on the bacmid to generate a recombinant bacmid. Once the transposition take place, the high molecular weight recombinant bacmid DNA is isolated and transfected into *Spodoptera frugiperda* Sf9 or Sf21 insect cells to generate recombinant baculovirus particles that can be amplified in consecutive rounds of infection, and ultimately used for protein expression.

### 2.8.2 Construction and verification of bacmid-MTA1 BAH and BAH-ELM2

BAH (1–154) and BAH-ELM2 (1–287) domains of MTA1 were created by introducing a stop codon in the full length MTA1 cloned into a pKL donor plasmid (with an N-terminal 6×His tag and a thrombin cleavage site between the His tag and MTA1). The recombinant plasmid was transformed into DH10BAC competent cells which contain the baculoviral DNA (bacmid). This step was done to allow for transposition of the MTA1 construct into the bacmid. The transformation was done using the Invitrogen Bac-to-Bac system with the exception that the pKL vector was used in place of the pFASTBac vector (Section 7.6). Bacmid DNA was then isolated from the overnight culture of DH10BAC. The transposition of the MTA1 BAH and BAH-ELM2 constructs into the bacmid was confirmed by PCR using an MTA1-specific forward primer and an M13-specific reverse primer from the bacmid. Isolated recombinant bacmid DNA was analysed by PCR followed by 1% agarose gel electrophoresis, in which baculovirus transposition was confirmed by the appearance of a band of ~2800 bp in size (665 bp bacmid +2145 bp MTA1 fragment) (Figure 2-23B).

The recombinant bacmid containing the 6×His-MTA1 BAH or BAH-ELM2 construct was then transfected into the insect *Spodoptera frugiperda* Sf9 cells growing as a monolayer using Fugene transfection reagent (Promega). The cells were left to grow at 27 °C for 48 h to produce recombinant baculovirus particles (P1 viral stock), which
were then used to infect fresh Sf9 cells for further viral amplification. Each amplified generation of virus is assigned a passage (P) number indicating the number of rounds of amplification. The initial recombinant virus is a low titer stock and is amplified to produce high titer baculovirus stocks that can be used to infect cells for protein expression (Section 1.6). After three generations (P1–P3) of growth, harvesting of virus and reinfection, the recombinant baculovirus (P3) harbouring the MTA1 BAH and BAH-ELM2 fragments was collected and used as virus stock to express recombinant protein. Sf9 cells collected from P3 transfection were lysed, and proteins were separated by SDS-PAGE followed by Western blot using monoclonal anti-His antibody. Figure 2-23C showed one band running at a molecular weight consistent with that predicted for MTA1 fragments (BAH, 20.4 kDa and BAH-ELM2, 35.2 kDa).

After we generated a P3 baculoviral stock and confirmed the protein expression, we scaled up the protein expression. To first determine the optimal conditions for expression of MTA1 BAH and BAH-ELM2 fragments, we infected a small population of cells (25 ml at ~1 × 10^6 cells/ml) with the P3 viral stock at varying ratios of virus to cells (multiplicity of infections, MOI: 1, 5 and 10) and assayed for protein expression. We also performed a time course by infecting Sf9 cells with the P3 viral stock at a constant MOI and assayed for protein expression at different times post-infection (24, 48, 72 and 96 h post-infection). Unfortunately, only a small fraction of the expressed protein appeared to be soluble (Figure 2-23C) and it was undetectable by SDS-PAGE (not shown).
Figure 2-23: Confirmation of integration of MTA1 BAH and BAH-ELM2 into the bacmid DNA and Western blot of MTA1 BAH and BAH-ELM2. A. Schematic diagram showing the MTA1 constructs used. B. The PCR product of MTA1 BAH and BAH-ELM2 generated using MTA1 forward and M13 reverse primers was assessed by 1% agarose gel electrophoresis to determine correct recombination and transposition events. Lane Kb, DNA marker indicated in kb; lanes 1, 2, 3 and 4, PCR products from bacmid 665 bp + 2145 bp MTA1 ~2800 bp for BAH and BAH-ELM2 as indicated. Numbers to the left indicate the mass of DNA standards in kb. C. SDS-PAGE of total protein followed by Western blot using monoclonal anti-His antibody. Lane M, protein marker with masses indicated in kDa to the left; lanes T and S, total lysate and soluble fraction of either MTA1 BAH or BAH-ELM2 from P3 viral stock as indicated; positive control (+ve) His-tagged RbAp48 (~49 kDa).

2.9 Discussion

2.9.1 Protein solubility using different expression systems

In this chapter, experiments were described that focused on efforts to obtain structural information on the N-terminal portion of the MTA proteins, MTA1 in particular. Initially, efforts were centred on the use of the standard bacterial expression approach.

Protein solubility can be affected by several factors, including pH, ionic strength, the speed of protein synthesis, media, the amino acid composition of the protein and the cellular location of expression. We tried many different conditions to obtain soluble folded protein, including different plasmids, tags, bacterial strains, low temperatures (15 °C and 18 °C), buffers and pH. The high-level expression of recombinant proteins in E. coli may result in the accumulation of insoluble aggregates as inclusion bodies. Thus, low IPTG concentrations were also tried to reduce the rate of protein synthesis and improve solubility and proper folding, but unfortunately without success. In most cases
in which we obtained high yield, the protein appeared to be not folded properly or aggregated.

We were able to purify the BAH and ELM2-SANT domains fused to MBP, a tag that is known to enhance protein solubility during protein expression; however, we could not subsequently separate the BAH domain from the fusion protein and the MBP following proteolytic cleavage of the fusion protein. This may be because the BAH domain was soluble but not in its native conformation. For the ELM2-SANT fragment, we obtained some crystal hits that were very small or clusters of needles. Unfortunately, they were not appropriate for x-ray diffraction and further optimisation is necessary to improve crystal size and quality. 1D NMR spectroscopy also needs to be used to evaluate the folded state of ELM2-SANT, and determine whether the protein would be suitable for structure determination using NMR rather than crystallographic methods.

As we could not obtain sufficient soluble protein for structural studies and most MTA1 constructs were expressed in inclusion bodies, we attempted to refold both the BAH-ELM2 (32 kDa) and BAH-ELM2-SANT (39 kDa) domains from inclusion bodies. However, it is challenging to correctly refold a protein after denaturation. During refolding of the proteins, we faced problems with precipitation, which lowered the yield of soluble native protein. The refolded BAH-ELM2 did not attain a well-ordered conformation as shown by 1D NMR analysis, although it was at least soluble. In contrast, the refolded BAH-ELM2-SANT fragment showed some secondary structure as judged by CD spectra; however, the yield was very low and not sufficient for structural studies. It is possible that the relatively large size of the domains prevented correct refolding.

Another issue that may have affected the solubility of these proteins is the difficulty in identifying the boundaries of MTA1 domains, due to their low sequence conservation with similar known domains. Thus, it is possible that we did not define the domain boundaries correctly; in addition, all of our MTA1 constructs had N-terminal tags, except the His-tagged ones, and the solubility of a C-terminal tag can make a difference to the overall solubility of the protein.

Although heterologous expression in E. coli has been the cornerstone of biochemistry and structural biology for the last 30 or more years, there are many proteins or protein fragments that are refractory to expression in this system. Many recombinant proteins,
especially those of eukaryotic origin, aggregate or become packaged into inclusion bodies when expressed in bacteria (Frangioni and Neel, 1993, Frankel et al., 1991). Several reasons are commonly cited for this observation, including chaperones and cofactors not found in bacteria, or the proteins may be missing PTMs (Nausch et al., 2013). These post translational modifications can have roles in protein solubility and/or proper folding. Our data and others have shown that MTA1 is subjected to several PTMs (see Section 1.4.4 and Figure 2-24) (Choudhary et al., 2009b). It is likely that the insolubility of our bacterially expressed constructs was due to the lack of post-translational modifications of the recombinant MTA1 domains, resulting in incorrect folding of the proteins.

These issues have led researchers to develop eukaryotic expression systems for the production of ‘difficult’ proteins in order to obtain biologically active protein for biophysical studies. An increasing number of published structural studies have relied on expression in insect or, more recently, mammalian cells. Indeed, RbAp46 and RbAp48 were expressed in insect cells using the baculovirus approach in this Thesis. Other NuRD components have also proven difficult to express in bacteria. For example, there are no reports of structures for HDAC1 and 2, most likely hinting at protein production difficulties. Even expression in eukaryotic systems can be insufficient. The HDAC3 and DAD domain of the SMRT complex were not able to be expressed in a stable form individually, but co-expression in mammalian cells (HEK293) yielded a stable, folded complex (Watson et al., 2012b).

In line with these approaches, the expression screen of MTA constructs carried out in insect cells showed a significant improvement in protein solubility compared to expression using E. coli. However, attempts to scale up the insect cell expression using a baculovirus system were unsuccessful, with minimal expression and negligible soluble protein produced.
Figure 2-24: Predicted modification sites on human MTA1 protein (715 residues). This figure was taken from http://www.phosphosite.org/proteinAction.do?id=2945&showAllSites=true.

2.9.2 Future direction

Our data suggest that insect (or other eukaryotic) cells rather than *E. coli* are the preferred system for overexpression of MTA protein fragments.

At the same time, it is also worth attempting to crystallise the bacterially expressed MTA1 BAH and ELM2-SANT domains without cleaving them from the MBP tag, as there are some examples where proteins have been crystallised while fused to MBP (Smyth et al., 2003). However, the observation that these domains were non-specifically bound to the MBP following cleavage suggests that they were not correctly folded, even as fusion proteins.

Some proteins could aggregate during expression because they are missing cofactors or prosthetic groups such as minerals, trace metals or vitamins, which need to be taken into account during protein expression. Another issue that needs to be addressed is the domain boundaries. To address this issue, several truncations of each domain could be cloned by adding 5–10 residues from both sides considering the tag position. In addition, constructs from MTA2, MTA3 or MTA proteins from different species may be worth trying, as there are some differences in MTA protein sequences.

Finally, it is possible that the N-terminal domains of MTA1 might be well-ordered only in the presence of partners. It is therefore worth trying to find binding partner(s) for these domains and attempting co-expression, which might result in higher expression levels or solubility. It is worth noting in this context that HDAC3 was, as mentioned above, co-expressed and crystallised with the SANT-like DAD domain of SMRT. Thus, co-expression of MTA1 ELM2-SANT domains with HDAC1/2 in a mammalian system may stabilise these domains and produce an intact soluble complex.
Chapter 3. Mapping of the RbAp48-MTA1 interaction

3.1 Introduction

The NuRD complex is composed of approximately 10 subunits, which include RbAp46/48, MTA1/2/3, HDAC1/2, CHD3/4, MBD2/3 and p66α/β. Previous work in our laboratory examined pairwise interactions between the NuRD proteins, as part of an effort to understand how the complex is assembled. In this prior study, each individual protein was divided into multiple overlapping constructs, and all of the resulting polypeptides were screened for their interaction with one another via the yeast two-hybrid system (Sock Thong, unpublished data). Although this system showed some interactions between NuRD protein constructs, the overall number of detected interactions was less than expected, given that the NuRD complex is comprised of an assembly of at least ten different proteins. In addition, previously reported protein-protein interactions within the NuRD complex, such as the interaction between the RbAp46/48 and MTA1/2 (Roche et al., 2008, Fu et al., 2011), were not observed. We hypothesised that problems with expression, folding and post-translational modification (PTM) of human NuRD proteins in yeast might have reduced our chances of seeing the full range of intra-subunit interactions. We consequently decided to change to a mammalian system in an effort to circumvent this possible problem; for experimental simplicity, we chose in vitro translation (IVT) for interaction mapping studies. Several cell-free protein expression systems from numerous sources have been used in recent years for the in vitro expression of proteins, including rabbit reticulocytes, wheat germ and E. coli. We chose the rabbit reticulocyte lysate (RRL) system because it is a mammalian system, which may enable PTMs such as glycosylation, acetylation, and phosphorylation to occur, and because it may also contain chaperones or other factors that can assist with correct folding.

The IVT system was used to test the interaction between RbAp46/48 and MTA1/2. The NuRD components RbAp46 and RbAp48 are, as described in Section 1.4.2.5, members of a family of proteins known as the WD40 repeat proteins, so named because they contain ~40-residue repeats terminating in Trp (sometimes Phe)-Asp (sometimes Gln) (Figure 3-1). These repeats fold together into a β-propeller structure (Murzina et al., 2008, Lejon et al., 2011). RbAp46/48 are highly homologous histone chaperones
(Figure 3-2) that have been found in a number of protein complexes involved in chromatin remodelling, transcriptional repression and chromatin assembly.

**Figure 3-1: Sequence of human RbAp48, highlighting the WD repeats.** RbAp48 is a 425 amino acid residue protein with seven WD40-repeats (Uniprot: Q09028).

**Figure 3-2: RbAp46 and RbAp48 are highly homologous proteins with ~92% similarity at the amino acid sequence (Uniprot for human RbAp46: Q16576).** The identical and similar residues are highlighted on a black and gray background, respectively.

MTA proteins have also been found to be integral members of the NuRD complex. Interestingly, whereas RbAp48 and/or RbAp46 are found in a number of different complexes, MTA proteins are unique to the NuRD complex. MTA proteins are highly...
conserved at the N-terminal end and divergent at the C-terminus (Figure 3-3). The differences in the C-termini of the MTA proteins could be an indication of functional diversity in MTA/NuRD complexes; that is, different versions of NuRD carrying a different MTA protein might have different functions. Although these proteins are ubiquitously expressed in tissues (Manavathi and Kumar, 2007) and have been studied in some detail, the manner in which they interact with other protein(s) in the NuRD complex is not understood at the molecular level.

MTA proteins are equipped with several domains that could be implicated in protein-protein interactions, DNA binding, and signal transduction (Figure 3-4) (Talukder et al., 2004). Currently, however, the functions of these domains are unknown.
Figure 3-3: The human MTA protein family. Protein sequences of MTA1, 2 and 3 (Uniprot: Q13330, O94776 and Q9BTC8, respectively) aligned using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The identical and similar residues are highlighted on a black and gray background, respectively. Predicted MTA1 domains are labelled in colours.
Figure 3-4: Schematic of MTA structural domains. MTA1-3 proteins have several domains that are predicted to be ordered (BAH, ELM2, SANT, ZnF, SH3-binding domain and Myb domains) and could be involved in diverse functions, such as protein-protein interactions, DNA binding, or signal transduction.

The aim of this chapter is to ascertain whether there is an interaction between MTA1 and RbAp48, and if so, to determine the specific region(s) of MTA1 that mediate the interaction with RbAp48.

3.2 Pulldown experiments

Many methods are used to map protein-protein interactions. The pulldown assay is a commonly used technique to either confirm a previously suspected protein-protein interaction or to identify an unknown interaction. The method uses a tagged protein (bait) immobilised on beads by an affinity ligand specific for the tag. The beads are incubated with a protein source that contains putative interaction partners (prey), such as recombinant purified protein, cell lysate or in vitro transcription/translation reactions, and washed to remove non-binding protein(s). Protein-protein interactions (represented by bound prey protein) can be visualised by SDS-PAGE and associated detection methods, including Coomassie or silver staining, Western blotting or [35S] radioactivity detection (Figure 3-5).
Chapter 3

3.3 MTA1 and MTA2 interact directly with RbAp46/48

To investigate the interaction between the MTA and RbAp proteins, MTA1 and MTA2 (full length) were cloned into the pcDNA3 vector while RbAp46 and RbAp48 (also full-length) were cloned into the pcDNA3-FLAG vector using standard molecular biology techniques. All cloned constructs were confirmed by DNA sequencing.

These proteins were then produced using the TnT® Quick Coupled Transcription/Translation System (Promega) in the presence of $[^{35}\text{S}]$ methionine. The pulldowns were carried out by immobilising IVT $^{35}\text{S}$-labelled FLAG-RbAp46/48 on anti-FLAG beads and then incubating the beads with the IVT $^{35}\text{S}$-labelled MTA1/2 proteins (Section 7.13). The pulldown experiments showed that MTA1 interacted efficiently with FLAG-RbAp46 and FLAG-RbAp48 (Figure 3-6A), whereas MTA2 interacted weakly with both RbAp proteins (Figure 3-6B).

**Figure 3-5: Pulldown assay principle.** In a pulldown assay, a tagged bait protein is immobilised on beads using an affinity ligand specific for the tag. The bait protein captures the prey protein via a protein-protein interaction. SDS-PAGE and associated detection methods (such as $[^{35}\text{S}]$ radioactivity detection) can be used to visualise the bound protein.
Figure 3-6: Pulldowns of RbAp46/48 with MTA1/2. A. Pulldown assays were carried out using in vitro translated $^{35}$S-labelled-FLAG-RbAp46/48 immobilised on anti-FLAG beads and in vitro translated $^{35}$S-labelled MTA1. Lanes 1, 2 and 3: 10% input of the lysate used in each pulldown, of RbAp46, RbAp48 and MTA1, respectively; Lane 4: RbAp48/MTA1 pulldown; Lane 5: RbAp46/MTA1 pulldown. B. As in (A) except that the MTA2 was used instead of MTA1. Lanes 1, 2 and 3: 10% input of RbAp48, MTA2 and RbAp46, respectively; Lane 4: RbAp48/MTA2 pulldown; Lane 5: RbAp46/MTA2 pulldown. Reaction mixtures were analysed by SDS-PAGE, and then subjected to autoradiography. This figure was created by merging lanes from multiple SDS-PAGE gels (indicated by the dashed lines).

3.4 Mapping the region(s) of MTA proteins that binds RbAp proteins

3.4.1 Designing the MTA constructs

Having established that RbAp46/48 could interact with MTA1 in our pulldown system we next sought to determine the minimum binding region of MTA1. The structures of RbAp46 and RbAp48 have been determined (Murzina et al., 2008, Lejon et al., 2011), these structures reveal that both proteins form a single folded domain, and we therefore chose to continue to express the full-length protein for interaction studies. In contrast, given that MTA1/2 appear to contain multiple domains, we divided these proteins into shorter fragments to assess which parts mediate interactions with RbAp46/48.

To assist in construct design, the secondary and tertiary structure of MTA1/2 was predicted using online software tools that evaluate the possible structure of a protein based on its amino acid sequence, its conservation across species, and its similarity with other characterised proteins. The secondary structure for MTA1/2 was predicted using Phyre2 software (Kelley and Sternberg, 2009), which is a 3D-structure modelling/prediction program (Figure 3-7). The secondary structure predictions made by Phyre2 are a consensus of three secondary structure prediction methods, Psi-Pred, SSPro and JNet. In addition, we used JPred3 software (Cole et al., 2008), which predicts secondary protein structure from amino acid sequence, based on a consensus from several methods including DSC, NNSSP, PREDATOR, PHD, MULPRED and ZPRED (data not shown). Also, we examined the sequence of the MTA1/2 proteins by
subjecting them to Predictor of Naturally Disordered Regions (PONDR®) analysis, which predicts the likelihood that each part of a protein is disordered or ordered, based on its amino acid sequence. PONDR uses sequence attributes such as fractional composition of amino acid type, hydropathy, and sequence complexity, taken over windows of 9 to 21 amino acids. These attributes are averaged over a particular window to derive a value for the amino acid in the centre of the window, and the predictions are subsequently smoothed over a sliding window of 9 amino acids. The PONDR®VL-XT predictor (Li et al., 1999, Romero et al., 2001, Romero et al., 1997), which we used to predict regions of order and disorder in the MTA1 and MTA2 proteins, outputs fractional scores between 0–1, where 1 is the ideal prediction of disorder and 0 is the ideal prediction of order. A residue with a score that exceeds a threshold of 0.5 is considered disordered (Figure 3-8). Based on these prediction programs, most of the C-terminal portion of both MTA proteins is predicted to be disordered.

MTA1 constructs were then designed to preserve predicted domains; these constructs were cloned using a standard PCR approach from their respective full-length genes and then ligated into pMW172, pcDNA3, or pGEX-6P vectors via BamHI and EcoRI restriction sites. All of the constructs were expressed either in vitro (using the RRL expression system) (see Table 7-4), or in E. coli as GST fusion proteins (see Table 7-5).
Figure 3-7: Secondary structure prediction for MTA1 using Phyre2 server. Green helices represent α-helices, blue arrows indicate β-strands and faint lines indicate coil; Seq: amino acid sequence; SS: Secondary Structure prediction (Kelley and Sternberg, 2009).

Figure 3-8: PONDR VL-XT graphs showing the predicted disorder profiles of MTA1 and MTA2. The thin black line at 0.5 indicates the threshold of disorder. The thick bar on the threshold line indicates a region of 40 or more consecutive residues predicted to be disordered.
To this end, several C- and N-terminal MTA1 deletion constructs were generated and expressed as $^{35}$S-labelled proteins and/or GST-fusion proteins. These were then subjected to pulldown assays with the $^{35}$S-labelled Flag-RbAp46/48 protein by immobilising the Flag-RbAp46/48 or GST-fusion MTA1 constructs on anti-Flag beads or glutathione beads, respectively, and then treating with the partner protein.

### 3.4.2 The C-terminal region is important for the binding of MTA1 to both RbAp proteins

In order to identify the region(s) of MTA1 that binds to RbAp48, we carried out pulldowns, using IVT $^{35}$S-labelled FLAG-RbAp48 as bait to pull down overlapping IVT $^{35}$S-labelled-expressed MTA1 constructs (1–250, 230–550 and 530–715). To our surprise, none of the previously identified structured domains of MTA1 (BAH, ELM2, SANT and ZnF) were required for binding to RbAp48. Instead, the C-terminus of MTA1(530–715) was sufficient to mediate interaction with the RbAp48. While MTA1 230–550, which contains the SANT and GATA-type zinc finger domains, exhibits some RbAp48 binding activity, residues 530–715 bind more strongly (Figure 3-9).

![Figure 3-9: MTA1-RbAp48 pulldowns.](image)

We carried out pulldowns using the same constructs of MTA1 with FLAG-RbAp46, and the results indicate that MTA1 interacts with both RbAp proteins in similar way (Figure 3-10).
3.4.3 The C-terminal region is also important for the binding of MTA2 to both RbAp proteins

In order to test whether MTA2 binds to RbAp46/48 in the same way as MTA1, we performed pulldowns, using either IVT $^{35}$S-labelled-FLAG-RbAp46 or IVT $^{35}$S-labelled-FLAG-Rbap48 as bait, to pull down overlapping IVT-expressed $^{35}$S-labelled-MTA2 constructs (1–250, 220–530 and 520–668). The results show that both C-terminal fragments of MTA2 (220–530 and 520–668), and in particular, the far C-terminal fragment 520–668, demonstrate binding to RbAp46/48 (Figure 3-11).

Figure 3-11: MTA2-RbAp46/48 pulldowns. Left: Schematic diagram of MTA2 showing the constructs used in RbAp46/48 pulldown assays. Right: Pulldowns of IVT $^{35}$S-labelled MTA2 constructs (1–250, 220–530 and 520–668) by either immobilised IVT $^{35}$S-labelled FLAG-RbAp48 or IVT $^{35}$S-labelled FLAG-RbAp46 on anti-FLAG beads. The input is 10% of RbAp46/48 and MTA2 constructs used in each pulldown. The arrows indicate the MTA2 constructs. Reaction mixtures were analysed by SDS-PAGE, and then subjected to autoradiography. This figure was created by merging lanes from two SDS-PAGE gels.
In summary, MTA2 binds to both RbAp46/48 proteins in the same manner as MTA1. We chose to focus our subsequent work on the MTA1-RbAp48 interaction.

3.5 The RbAp48 interaction domain localises to the C-terminal portion of MTA1

3.5.1 Confirming the importance of the C-terminal region of MTA1 for binding

Having already shown a clear interaction of the C-terminus of MTA1(530–715) with both RbAp proteins, we next sought to define further the minimum RbAp48-binding region within the far C-terminal end of MTA1. We therefore split MTA1(530–715) into two smaller fragments – MTA1(530–635) and MTA1(625–715). In addition, MTA1(230–550) showed some interaction with RbAp proteins (Figure 3-9 and 3-10), and so to further investigate this interaction, longer fragments covering all the N-terminal domains (BAH, ELM2, SANT and ZnF; 1–511, 1–550) or the entire C-terminus (449–715 and 512–715) of MTA1 were also generated. Using pulldown assays, we tested the interaction of these MTA1 constructs with RbAp48. Our results indicated that all MTA1 C-terminal constructs, except 530–635, bound to RbAp48 (Figure 3-12). These results also indicate that the C-terminal fragment MTA1(625–715) is sufficient for the interaction. In contrast, MTA1(1–511) and MTA1(1–550) did not interact with RbAp48.

It is not clear why MTA1(230–550) showed some interaction with RbAp46/48 (Figure 3-9 and 3-10), whereas 1-550 and 1-511 did not interact with RbAp48.

Figure 3-12: The RbAp48-binding domain resides within the C-terminal region of MTA1. Left: Schematic diagram of MTA1 showing the constructs used in RbAp48 pulldown assays. Right: Pulldowns of IVT \(^{35}\)S-labelled MTA1 constructs (1–550, 449–715, 530–635, 625–715, 512–715 and 1–511) by immobilised IVT \(^{35}\)S-labelled FLAG-RbAp48 on anti-FLAG beads. The input is 10% of RbAp48 and MTA1 constructs used in each pulldown. Reaction mixtures were analysed by SDS-PAGE, and then subjected to
autoradiography. Note that the smaller constructs have a low signal-to-noise because they contain few sulphur-containing residues. This figure was created by merging lanes from two SDS-PAGE gels (indicated by the dashed lines).

3.5.2 Two regions of MTA1 can contact RbAp48

To further verify the interaction using bacterially expressed MTA1, we created shorter constructs covering residues 448–542, 541–630 and 625–715 of MTA1, which were expressed in E. coli as GST-fusion proteins. GST pulldowns were conducted using 35S-labelled RbAp48 (Figure 3-13). The results show that the GST-MTA1(448–542) and GST-MTA1(625–715) but not GST alone or GST-MTA1(541–630) bound to RbAp48. This indicates that the MTA1 residues (448–541) and (625–715) can both independently interact with RbAp48.

**Figure 3-13: GST pulldowns.** A. Schematic diagram of MTA1 showing the constructs used in RbAp48 pulldown assays. B. Upper panel: GST-MTA1 constructs (448–542, 541–630 and 625–715) immobilised on glutathione beads were used to pulldown IVT 35S-labelled RbAp48. GST alone was used as a control. The input is 10% of 35S-labelled RbAp48 used in each pulldown. Lower panel: a Coomassie blue-stained gel indicating the amount of either GST or GST-MTA1 constructs used.

3.6 Narrowing down the RbAp48 binding site to two short regions within the C-terminal end of MTA1

3.6.1 A KRAARR motif mediates an interaction between MTA1/2 and RbAp48

Examination of the amino acid sequence of MTA1 and MTA2 regions that bind to both RbAp46 and RbAp48 (Figure 3-14) revealed a motif (KRAARR/RRAARR) that
resembles part of the RbAp46-binding motif of histone H4 (RRLARR; see Figure 3-34B for a comparison), suggesting that this region might mediate binding to RbAp48. It is notable that the putative Myb-like domain of MTA1 spans the region 643–696 (Figure 3-3), which might be significant for binding. In addition, although the amino acid sequences of the C-terminal ends of MTA proteins are divergent, they share some sequence similarities (Figure 3-3). To test this hypothesis, we therefore created shorter constructs containing the KRAARR motif fused to GST, namely MTA1(643–695) and MTA1(656–686), and tested their ability to interact with ^35S-labelled RbAp48 using GST pulldown assays. As shown in Figure 3-15B, these constructs interact robustly with RbAp48.

MTA1 663–EEKRIRKLLSSETKRAARRPYKPIALRQSQALPPRPPP–703
MTA2 461–LTRLARRCMRDILOPRRAARRPYAPINANAIAECSTRP–501
MTA2 611–KDTRALKALTHERMAARRRNPLKVIPTLIAVRPPPV–651
MTA3 475–FTKEARQCKNTRLRQAARRFVAINYAAIRAEYADRHA–514

Figure 3-14: Alignment of sequences from human MTA proteins containing the KRAARR motif, highlighting the conserved residues.

3.6.2 MTA1(448–542) also can bind to RbAp48

To further narrow down the RbAp48-binding site of the other region of MTA1(448–542), we created truncations from this region (448–478, 477–501 and 501–542) and expressed them in E. coli as GST-fusion proteins. Their interactions with the IVT ^35S-labelled RbAp48 were tested using GST pulldowns, but none of these MTA1 constructs bound to RbAp48 (Figure 3-15C). Therefore, we created overlapping constructs from the same region of MTA1, namely 448–501, 448–516, 448–526, 465–542, 477–542 and 465–516, and similarly tested their interactions with the IVT ^35S-labelled RbAp48. As shown in Figure 3-15D, all these constructs bound to RbAp48 except MTA1(448–501) and MTA1(477–542). It is notable that the fragments quite prone to proteolytic degradation, and the use of different protease inhibitors and carrying out all purification steps at 4 °C did not help stabilise the proteins. Nevertheless, these results indicate that MTA1(465–516) is the shortest fragment of this region of MTA1 that still binds to RbAp48. Our results therefore suggest that the regions encompassing residues 465–516 and 656–686 of MTA1 are able to bind RbAp48. As shown in Figure 3-3, the MTA1 448–542 region does not contain the KRAARR motif.
Figure 3-15: Narrowing down the region of MTA1 involved in binding RbAp48. A. A schematic diagram of the MTA1 indicates the two regions that are able to bind RbAp48, designated RbAp48-binding domain 1 and 2 (RBD1/2), and the MTA1 constructs used in pulldown experiments. B, C and D: Upper panel: the pulldown experiments of GST-MTA1 constructs (as indicated) with the IVT 35S-labelled RbAp48. Input (i) is 10% of the amount used in each pulldown. GST alone was used as a negative control. Reaction mixtures were analysed by SDS-PAGE, and then subjected to autoradiography. Lower panel: a Coomassie blue-stained gel indicating the amount of either GST or GST-MTA1 fragments used.
3.7 The importance of the KRAARR motif for binding

3.7.1 Detailed mapping of the KRAARR motif

Given that, in the structure of RbAp46 bound to histone H4, the KRAARR motif binds in a helical conformation, we created two double point mutants in which we introduced two alanine residues to each end of the motif: K678A and R679A, such that KRAARR was mutated into AAAARR (M1); and R682A and R683A, such that KRAARR was mutated into KRAAAA (M2) (Section 7.2.2). These mutations were designed to disrupt sidechain interactions without altering the propensity for the sequence to adopt a helical structure. We also introduced two proline residues (K678P and R697P, such that KRAARR was mutated to PPAARR) into both MTA1(643–695; M3) and MTA1(656–686; M4) constructs. Primers used for PCR are listed in Appendix C. We then tested the binding of all four mutants to wild-type RbAp48. As shown in Figure 3-16 B and C, the proline mutations abrogated the binding of the shorter constructs MTA1(643–695) and (656–686), and the alanine mutations reduced the binding substantially in the case of the MTA1(449–715) based construct, supporting the idea that the KRAARR motif is important for the MTA1-RbAp48 interaction.

FOG1 contains a short motif (RRK) that is important for its interaction with both RbAp46/48 and MTA1/2 (Hong et al., 2005, Lejon et al., 2011). The C-terminus of MTA1 also contains another motif (KRRR) adjacent to the KRAARR motif, which is similar to the FOG1 motif (RRK), and might mediate the binding to RbAp48 (Figure 3-17). We therefore introduced a mutation in this motif by replacing two arginines with alanines (R645A and R646A, such that KRRR was mutated to KAAR, M5) in the MTA1(625–715) construct. This mutation did not affect the interaction with RbAp48 using a GST pulldown assay (Figure 3-16D).

To complement the MTA1 mutations, we also mutated residues in the known histone H4 binding site of RbAp46 in either the hydrophobic surface of N-terminal α-helix (L30Y, RM1), the charged PP loop (E356Q + D357N + E359Q + D360N, RM2), or both simultaneously (RM3). The RM3 mutation strongly reduced the binding to GST-MTA1(656–686) in GST pulldown experiments (Figure 3-16E).
Taken together, these data support the idea that the KRAARR motif but not the KRRR motif is essential for binding to RbAp48 and that the nature of the interaction might resemble that seen in the RbAp48-H4 complex.

Figure 3-16: Mapping of the RbAp48-MTA1 interaction using pulldowns. A. Schematic diagram shows the MTA1 fragments used in pulldown assays and the position of mutants. B. Pulldown assays using IVT $^{35}$S-labelled FLAG-RbAp48 and IVT $^{35}$S-labelled MTA1(449–715) wild-type (WT) and mutants M1(K678A/R679A) and M2(R682A/R683A). Beads alone was used as the negative control (-ve). C. GST pulldown assays using IVT $^{35}$S-labelled RbAp48 and GST-MTA1 fragments (643–695 and 656–686) both wild-type and M3/4 mutants (K678P and R679P) of MTA1 are shown. D. GST pulldown assays using IVT $^{35}$S-labelled RbAp48 and a GST-MTA1 mutant in the KRRR motif by substitution of two arginines with alanines (R645A and R646A, M5) in the MTA1(625–715) construct. E. GST pulldown assays using IVT $^{35}$S-labelled RbAp48 and a GST-MTA1 mutant in the H4 binding site on RbAp46: L30Y (RM1); E356Q + D357N + E359Q + D360N (RM2); or RM1 and RM2 combined (RM3), pulled down using GST-MTA1(656–686) as the bait. In C, D and E: the top panel is an autoradiogram showing the amount of $^{35}$S-labelled RbAp46/48 pulled down in each experiment, whereas the lower panel is a Coomassie blue-stained gel showing the amount of either GST or GST-MTA1 constructs used. For all panels the input (i) lane contains 10% of the $^{35}$S-labelled RbAp46/48 protein used in the pulldown assays (p).
3.7.2 MTA1 and histone H4 appear to bind to the same pocket on RbAp48

As noted above, the crystal structure of H4 bound to RbAp46 has been determined (Figure 3-34) (Murzina et al., 2008). In order to further investigate whether the MTA1 recognition site on RbAp48 might overlap with the H4 recognition site, as indicated by the mutagenesis data in the previous section, we expressed GST-H4(1-48), immobilised the protein to glutathione beads and pulled down IVT $^{35}$S-labelled RbAp48 in the absence and presence of full-length IVT $^{35}$S-labelled MTA1. As shown in Figure 3-18A, the presence of MTA1 completely disrupted the histone H4-RbAp48 interaction. In a complementary experiment, we immobilised IVT $^{35}$S-labelled FLAG-RbAp48 on anti-FLAG beads and pulled down IVT $^{35}$S-labelled full length MTA1 in the absence and presence of bacterially expressed and purified histone H4(1–48). Figure 3-18B shows that H4(1–48) significantly impairs the RbAp48-MTA1 interaction, indicating that the MTA1 and H4 binding surfaces of RbAp48 do overlap. Thus, these competition experiments do corroborate the data from Section 3.7.1, providing additional support for the idea that the MTA1 and histone H4 binding sites on RbAp48 overlap.

3.8 MTA1 and FOG1 independently bind to RbAp48

Given the overlap of the MTA1 and histone H4 binding sites on RbAp48, we next asked whether MTA1 also contacts the other known RbAp48 protein-interaction surface. Previously, we solved the crystal structure of a complex formed by RbAp48 and a short peptide (residues 1–15) derived from the transcriptional coregulator FOG1 (Lejon et al., 2011), showing that the peptide bound in an extended conformation to a groove on one end of the RbAp48 barrel. It has also been shown that the N-terminal tail of histone H3 can bind to Nurf55/p55 (the Drosophila melanogaster homologue of RbAp48) at an overlapping site (Schmitges et al., 2011). To explore whether MTA1 uses this site, we performed a competition experiment by adding increasing amounts of FOG1(1-15) peptide to a complex formed from full-length MTA1 and FLAG-RbAp48 immobilised on anti-FLAG beads. Figure 3-18C shows that the FOG peptide had no effect on the
MTA1-RbAp48 interaction. The same result was observed using an MTA1 construct encoding just a C-terminal fragment (530–715) that included the KRAARR motif identified above. Thus, it is likely that the interaction of RbAp48 with MTA1 is distinct from the interactions with FOG and histone H3.

Figure 3-18: Competition experiments probing the FOG1-, MTA1- and H4-binding sites on RbAp48. A. Pulldown showing the binding of IVT 35S-labelled RbAp48 to GST-histone H4(1–48) in the absence and presence of 35S-labeled MTA1. B. Pulldown showing the binding of IVT 35S-labelled MTA1 to FLAG-RbAp48 (immobilised on anti-FLAG beads) in the absence and presence of histone H4(1–48). C. Pulldowns showing the effect of adding increasing amounts of FOG1(1–15) peptide to a complex formed between in vitro translated 35S-labelled RbAp48 and IVT 35S-labelled MTA1 full length or 530–715. The input (i) lane contains 10% of the RbAp48, MTA1 full length and MTA1(530–715) proteins used in pulldown assays. Positive control (+ve) is 35S-labelled RbAp48 and MTA1 without FOG1(1-15) peptide and beads-MTA1 alone was used as the negative control (-ve).

3.9 Attempts to express the C-terminal end of MTA1 in bacteria and in insect cells

The preceding data show that two regions of the C-terminus of MTA1 can independently interact with RbAp48. Our next goal was to determine the structural basis for these interactions. In order to structurally characterise the RbAp48-MTA1 interaction, we required RbAp48 and MTA1 in hundreds of microgram to milligram quantities for biophysical assays and protein crystallography. Previous attempts by one of our laboratory members to express MTA1(448–715) in E. coli were unsuccessful due to instability and low yield. We therefore turned to an insect cell expression system.
3.9.1 InsectDirect expression of MTA1(448-715)

As described in Section 2.5, the InsectDirect system can be used as an alternative to a baculovirus expression system, at least for the initial screening of constructs.

In order to assess the expression of a C terminal fragment of MTA1 that encompassed both RbAp48 binding sites, MTA1(448–715) was cloned into pIEX with an N-terminal His tag. Sf9 cells in 10 ml suspension cultures at 1 x 10^6 cells/ml were transfected with 20 µg of the pIEX-His-MTA1(448–715) plasmid using Insect Gene Juice transfection reagent, and left to grow at 28 °C for 48-60 h. As a control for protein expression, a pIEX-GFP plasmid was transfected into Sf9 cells using the same protocol. As shown in Figure 3-19, no expression of the MTA1(448–715) construct could be detected by either Coomassie Blue staining or by Western blot using an anti-His tag antibody. In contrast, expression of GFP was clearly observed, indicating that the cells and the plasmid system were competent overall for heterologous protein expression. Also, we successfully expressed different constructs of MTA1 using this system (see Section 2.7). Therefore, we changed to the baculovirus-insect cell system, which has been used widely for protein expression in insect cells.

Figure 3-19: Small-scale expression and purification of MTA1(448–715) from Sf9 cells using InsectDirect expression system. A. Coomassie Blue-stained SDS-PAGE showing protein samples taken throughout the purification process. Lane M: molecular mass standards with sizes indicated in kDa; lane 1, total lysate; lane 2, soluble fraction; lane 3, protein bound to Ni-NTA beads before wash; Lanes 4 and 5, beads after washes; lanes 6–8: protein eluted from beads using 0.5 M imidazole. B. Fluorescence microscopy image of Sf9 cells transfected with pIEX-GFP. GFP expression is clearly visible at wavelength 488 nm. C. Western blot analysis of pIEX-His-MTA1(448–715) expression using monoclonal anti-His antibody. Lane 1, negative control of non-transfected Sf9 cells; lane 2, total lysate; lane 3, protein bound to beads; lane 4, positive control, His-tagged RbAp48 ~49 kDa. The arrows on the right indicating the expected size of the His-MTA1(448–715) construct.
3.9.2 Baculovirus expression of MTA1(448-715)

3.9.2.1 Construction and verification of bacmid-MTA1(448–715)

Baculoviruses are rod-shaped, enveloped viruses with circular, supercoiled double-stranded DNA molecules that duplicate in the nuclei of host cells (Summers and Anderson, 1972). The baculoviruses are capable of infecting insect cells and have been adapted to express proteins in insect cells. The codon optimised MTA1(448–715) specific to the insect cells expression was cloned into a pFBDM donor plasmid (as a 6×His-tagged fusion containing a thrombin cleavage site between the His tag and MTA1) between NdeI-BamHI restriction sites, and the recombinant plasmid was transformed into DH10BAC competent cells which contain the baculoviral DNA (known as a bacmid). This step was done to allow for transposition of the MTA1 construct into the bacmid. The transformation was done using the Invitrogen Bac-to-Bac system with the exception that the pFBDM vector was used in place of the pFASTBac vector. Bacmid DNA was then isolated from the overnight culture of DH10BAC. The transposition of the MTA1(448–715) construct into the bacmid was confirmed by PCR using an MTA1-specific forward primer and an M13-specific reverse primer. Isolated recombinant bacmid DNA was analysed by PCR followed by 1% agarose gel electrophoresis, in which baculovirus transposition was confirmed by the appearance of a band of ~1500 bp in size (665 bp bacmid +804 bp MTA1 fragment) (Figure 3-20B).
Figure 3-20: Confirmation of integration of MTA1(448-715) into the bacmid DNA and Western blot of MTA1(448–715). **A.** Schematic diagram indicating the RbAp48-binding domains (RBD1/2) of MTA1 and the construct that was used for insect cell expression. **B.** DNA gel showing the PCR product of MTA1(448–715) generated using MTA1 forward and M13 reverse primers was assessed by 1% agarose gel electrophoresis to determine correct recombination and transposition events. Lane 1, DNA marker indicated in kb; lanes 2–4, PCR product from bacmid 665 bp + 804 bp MTA1(448–715) ~1500 bp. Numbers to the left indicate the mass of DNA standards in kb. **C.** SDS-PAGE of total protein followed by Western blot using monoclonal anti-His antibody. Lane 1, protein marker indicated in kDa; lane 2, the negative control: total lysate of the uninfected Sf9 cells; lane 3, total lysate MTA1(448–715) from P1 viral stock; lane 4, total lysate MTA1(448–715) from P2 viral stock; lane 5, total lysate MTA1(448–715) from P3 viral stock; lane 6, soluble fraction of MTA1(448–715) from P3 viral stock; lane 7, positive control His-tagged RbAp48 (~49 kDa). Numbers to the left indicate the mass of protein standards in kDa. This figure was created by merging lanes from multiple SDS-PAGE gels (indicated by the dashed lines).

3.9.2.2 Transfection and protein expression

The recombinant bacmid containing the 6×His-MTA1(448–715) construct was then transfected into the insect *Spodoptera frugiperda* Sf9 cells growing as a monolayer using Fugene transfection reagent (Promega). The cells were left to grow at 27 °C for 48 h to produce recombinant baculovirus particles (P1 viral stock), which were then used to infect fresh Sf9 cells for further viral amplification. After three generations (P1–P3) of growth, harvesting of virus and reinfection, the recombinant baculovirus harbouring the MTA1(448–715) fragment (P3) was collected and used as virus stock to
express recombinant protein. Sf9 cells from each generation were lysed, and proteins were separated by SDS-PAGE followed by Western blot using monoclonal anti-His antibody. Figure 3-20C showed one band running at a molecular weight consistent with that predicted for MTA1(448-715, 32 kDa). Unfortunately, only a small fraction of the expressed protein appeared to be soluble (Figure 3-20C lane 6).

After we generated a P3 baculoviral stock and confirmed the protein expression, we scaled up the protein expression. To first determine the optimal conditions for expression of MTA1(448–715) construct, we infected a small population of cells (50 ml at ~1 × 10^6 cells/ml) with the P3 viral stock at varying ratio of virus to cells, multiplicity of infections (MOI; 1, 5 and 10) and assayed for protein expression. We also performed a time course by infecting Sf9 cells with the P3 viral stock at a constant MOI and assayed for protein expression at different times post-infection (24, 48, 72 and 96 h post-infection). We then used the MOI and time that provided the recombinant protein expression for large scale expression. To scale up protein expression, a suspension culture of Sf9 cells (500 ml at ~1 × 10^6 cells/ml) was infected with the recombinant baculovirus at MOI 1 and left to grow with shaking at 27 °C for 72 h.

Unfortunately, the yield of MTA1(448–715) was very low from this culture and undetectable by SDS-PAGE (Figure 3-21A). We tried the expression at MOI 5 and 10 but there was not sufficient protein for structural study (Figure 3-21A and B). However, the same construct of MTA1(449–715) appeared to be expressed at very low scale in rabbit reticulocyte lysate using in vitro translation system (Figure 3-12), but not in insect cells.

Due to low yield and instability, it was not possible to pursue the expression of MTA1(448-715). Instead, we focused our attention on expressing both RbAp48-binding regions of MTA1 individually.
Figure 3-21: Expression of MTA1(448–715) construct in Sf9 insect cells. A. SDS-PAGE analysis of the expression of MTA1(448–715) construct in Sf9 cells. Lane M, molecular mass standards indicated in kDa; lanes 1–4, 5–8 and 9–12 indicate total lysate, soluble fraction, protein bound to beads before and after wash at MOI 1, 5 and 10, respectively. B. SDS-PAGE of total protein and protein bound to beads followed by Western blot using monoclonal anti-His antibody. Lane M, molecular mass standards indicated in kDa; lane 1, negative control of uninfected Sf9 cells, lanes 2–3, 4–5, 6–7 are total protein and protein bound to beads at MOIs 1, 5 and 10, respectively; lane 8, positive control of His-tagged RbAp48 (~49 kDa). The arrows on the right indicating the expected size of the His-MTA1(448–715) construct.

3.10 Attempts to express RbAp48-binding domains of MTA1 individually

3.10.1 Overexpression and purification of MTA1(625–715)

As noted above (Section 1.5.1), MTA1(625–715) bind robustly to RbAp48. We therefore attempted to express this fragment individually. MTA1(625–715) was first cloned into the plasmid pGEX-6P, which contains a GST affinity tag and a HRV 3C cleavage site, for protein expression and purification. Large scale overexpression using *E. coli* BL21 cells yielded a high level of expression of GST-MTA1(625–715), although about 50% of the protein was in an insoluble form. Extensive work was done to increase the proportion of soluble protein. The GST-fusion protein was purified from the soluble fraction of the bacterial cell lysate by affinity chromatography using immobilised glutathione (GSH) Sepharose (Section 7.3.4) and treated with HRV 3C protease to remove the GST tag. Further purification was conducted using size exclusion chromatography, which resulted in MTA1(625–715) at a good level of purity (Figure 3-
22). Note that MTA1(625–715) was not stained effectively by Coomassie blue, probably because the peptide does not carry sufficient negatively charged side-chains to interact with copper ions from the stain (Pierce et al., 1999).

![Figure 3-22: Overexpression and purification of MTA1(625–715). A. SDS-PAGE analysis of GST-MTA1(625–715). Lane M: protein standards with their sizes (kDa) indicated; lane 1, total cell extract before induction; lane 2, total cell extract after induction; lane 3, soluble fraction; lane 4, insoluble fraction; lane 5, GST-MTA1(625–715) bound to glutathione Sepharose beads; lanes 6–10, fractions eluted from beads; lane 11, protein remaining on beads after elution; lane 12, combined elution fractions after overnight cleavage with HRV 3C protease; lane 13–18, fractions collected during size exclusion chromatography. Clean MTA1(625–715)-containing fractions (such as those in lanes 16–18) were combined and used for 1D NMR spectroscopy. This figure was created by merging lanes from multiple SDS-PAGE gels (indicated by the dashed lines).]

A one-dimensional $^1$H NMR spectrum can be used to check the folding status of a protein (Rehm et al., 2002). A 1D $^1$H NMR spectrum of MTA1(625–715) is shown in Figure 3-23A. The amide region ranging from 10.5–6.5 ppm is indicative of the degree of folding. The signals are not well dispersed, suggesting that the protein is not folded or is partially folded. For comparison, the 1D $^1$H NMR spectrum of a folded protein is also shown; it can be seen from Figure 3-23B that the 1D NMR spectrum of a folded protein is well-dispersed and sharp, indicating that the protein is folded. The appearance of intensities at chemical shift near 8.3 ppm and no visible signal below approximately 8.5 ppm were an indication of unfolded protein. Also, to the right of the strong methyl peak at 0.8 ppm, no further signals appeared in this region. Thus, the MTA1(625–715) protein appeared to be not properly folded. Moreover, the yield was not sufficient for structural studies.
Figure 3-23: Characterisation of MTA1(625–715). Comparison of the 1D $^1$H NMR spectrum of MTA1(625–715) and a folded protein spectrum. A. 1D $^1$H NMR spectrum of MTA1(625–715). The concentration of the protein was ~100 µM and the spectrum was recorded at 298 K, and referenced to DSS at 0.00 ppm buffered in 20 mM Tris (pH 7.5), 100 mM NaCl, 1 mM DTT, 5% (v/v) D$_2$O and 2 µM DSS. B. A typical 1D $^1$H NMR spectrum of a folded protein.

3.10.2 Overexpression and purification of MTA1(656–686)

MTA1(656–686) is the shortest construct that binds to RbAp48. It was expressed in bacteria as a GST fusion protein from the pGEX-6P vector. Small scale overexpression trials of the MTA1(656–686) construct showed that the optimal conditions for expression were at 25 °C with 0.4 mM IPTG. Thus, these conditions were used further for large scale overexpression of the construct. After primary purification by GSH affinity chromatography and treatment with HRV 3C protease, the complex was further subjected to ion exchange chromatography and size exclusion chromatography (Figure 3-24).
Figure 3-24: Overexpression and purification of MTA1(656–686). SDS-PAGE analysis of GST-MTA1(656–686). Lane M, protein standards with their sizes (kDa) indicated. Lane 1, soluble fraction; lane 2, insoluble fraction; lane 3, GST-MTA1(656–686) bound to glutathione Sepharose beads; lanes 4-7, fractions eluted from beads; lane 8, protein remaining on beads after elution; lane 9, combined elution fractions after overnight cleavage with HRV 3C protease; lane 10–16, fractions collected during ion exchange chromatography; lane 17–21, fractions collected during size exclusion chromatography. Clean MTA1(656–686)-containing fractions (such as those in lanes 18–21) were combined and used in 1D NMR. This figure was created by merging lanes from multiple SDS-PAGE gels (indicated by the dashed lines).

1D $^1$H NMR spectroscopy was used to evaluate the solution conformation of MTA1(656–686). A 1D NMR spectrum of MTA1(656–686) is shown in Figure 3-25A. Despite the broadness, the existence of signals at a chemical shift of >9 ppm indicated that this protein might exist, for at least part of the time, in a well-ordered conformation.

Circular dichroism (CD) spectropolarimetry was used to estimate the secondary structure of MTA1(656-686). It relies on the differential absorption of left and right circularly polarised light by chiral chromophores. For proteins, chromophores that contribute to CD signals include the peptide bond with weak and strong transitions at around 220 and 190 nm, respectively, aromatic side chains around 320–260 nm, disulphide bonds, which give rise to weak and broad signals around 260 nm, and different secondary structural elements (Kelly et al., 2005). Far UV CD (178-230 nm) is a valuable method for estimating protein secondary structure and different secondary structural elements yield characteristic far UV-CD spectra (Figure 3-25C – for standard spectra).

The far-UV CD spectrum of MTA1(656–686) was recorded at 20 °C (Figure 3-25B). A minimum is observed at ~195 nm, which is consistent with largely random coil structure. The data from the CD experiment show that MTA1(656–686) is largely unstructured in solution.
Figure 3-25: Characterisation of MTA1(656–686). A. The 1D \(^1\)H NMR spectrum of MTA1(656–686). The sample was in 20 mM Tris pH 8.0, 100 mM NaCl, 1 mM DTT, and the concentration was ~150 µM. The spectrum were recorded at 298 K, and referenced to DSS at 0.00 ppm. B. Far-UV circular dichroism spectrum of MTA1(656–686). This spectrum was collected using 20 µM protein sample in 20 mM Tris pH 8.0, 150 mM NaF at 20 °C. C. Ideal CD spectra for random coil (red), \(\beta\)-sheet (blue) and \(\alpha\)-helical (orange) polypeptides (Bolotina et al., 1980).

3.10.3 Overexpression of constructs derived from MTA1(448–542)

The second region we have shown to be important for RbAp48 binding is MTA1(448–542), which is also predicted to have secondary structure (Figure 3-32). Overexpression of a GST fusion of MTA1(448–542) produced a significant level of soluble protein. However, a substantial amount of degradation was observed during purification, even though protease inhibitors were used and all steps were performed at 4 °C (Figure 3-26).
Figure 3-26: Overexpression and affinity purification of GST-MTA1(448–542). SDS-PAGE analysis of cell and protein samples taken throughout the purification process. Lane M, protein standards with their sizes (kDa) indicated; lane Pre, total cell protein before induction; T, total cell protein following induction with IPTG; S, soluble fraction of cell lysate; Ins, insoluble fraction of cell lysate; FT, flow-through of soluble cell lysate; GSH-b, proteins bound to GSH-Sepharose; E: GST-MTA1(448–542) eluted from GSH-Sepharose (theoretical MW = 37.5 kDa). This figure was created by merging lanes from multiple SDS-PAGE gels (indicated by the dashed lines).

In an effort to produce a stable polypeptide encompassing the other RbAp48-binding portion of MTA1, four constructs covering the MTA1(448–542) region were cloned into pGEX-6P for protein expression. Overexpression trials of all four fragments (448–516, F1; 448–526, F2; 465–542, F3; and 465–516, F4) were conducted in BL21 and Rosetta2 cells at 25 °C and 20 °C using constructs with an amino terminal GST tag. Although expression was obtained in all cases, all proteins were unfortunately degraded during purification (Figure 3-27 and 3-28). The use of different protease inhibitors and carrying out all steps at 4 °C did not help to stabilise the proteins.

Figure 3-27: Overexpression of MTA1(448–542) constructs in BL21. A. SDS-PAGE showing four overlapping constructs (448–516, F1; 448–526, F2; 465–542, F3; and 465–516, F4) of MTA1(448–542) that were expressed as GST-fusion proteins in BL21. Lane M, protein standards with their sizes (kDa) indicated; lanes 1, 2 and 3 represent total cell extract before
induction, total cell extract after induction and soluble fraction for each fragment (F1–F4), respectively. The arrows show the position of each overexpressed fragment. **B.** Total soluble protein of MTA1 fragments F1-F4 bound to glutathione beads after wash.

![Figure 3-28](image)

**Figure 3-28: Overexpression of MTA1(448–542) constructs in Rosetta2 cells. A.** SDS-PAGE showing four overlapping constructs (448–516, F1; 448–526, F2; 465–542, F3; and 465–516, F4) of MTA1(448–542) that were expressed as GST-fusion protein in Rosetta2. Lane M, protein standards with their sizes (kDa) indicated; lanes 1, 2 and 3 represent total cell extract before induction, total cell extract after induction and soluble fraction for each fragment (F1–F4), respectively. Arrows show the position of each overexpressed fragments. **B.** Total soluble protein of MTA1 fragments F1-F4 bound to GSH beads after wash.

### 3.10.3.1 Overexpression and purification of MTA1(465–516 and 465–542) constructs

In an effort to optimise protein expression, we focused on two constructs that bind to RbAp48 (F3 and F4) for large scale expression. MTA1(465–542) and (465–516) constructs were also cloned into pGEX-4T with an N-terminal GST and thrombin cleavage site. Both constructs (465-542 and 465-516) were expressed in *E. coli* BL21 cells from either the pGEX-4T or pGEX-6P vector, which contain a thrombin or HRV 3C cleavage site, respectively. The proteins were subsequently purified by GSH-affinity chromatography, treated with the appropriate protease to remove the GST tag, and further purified using a final size exclusion chromatography step. The MTA1(465–516) fragment was not cleaved completely from the tag using both proteases, and the protein was later degraded (Figure 3-29). Similar result was obtained with the HRV 3C fragment.
We also tried to express the MTA1(465–542) fragment in E. coli BL21 cells from either the pGEX-4T or pGEX-6P vector. Overexpression of the GST fusion of MTA1(465–542) produced a soluble protein. The soluble fraction was subjected to GSH-affinity chromatography, treated with the appropriate protease to remove the GST tag, then further purified with size exclusion chromatography. The tag cleavage was not complete using both proteases and most of the protein was degraded during purification (Figure 3-30). The final yield was very low, which we used for biophysical analysis. Similar results were obtained with the thrombin fragment.
Figure 3-30: SDS-PAGE showing purification of recombinant GST-MTA1(465–542) from *E. coli* cells. Lane M, protein standards with their sizes (kDa) indicated; lane 1, GSH-GST-MTA1(465–542) after wash; lanes 2–5, elutions of GST-MTA1(465–542) from beads; lane 6, protein post-HRV 3C protease cleavage; lanes 7–12, fractions collected during size exclusion chromatography. MTA1(465–542)-containing fractions (such as those in lanes 10–12) were combined and used in a CD experiment. Lane 13, concentrated MTA1(465–542), theoretical MW = 9.5 kDa. This figure was created by merging lanes from multiple SDS-PAGE gels (indicated by the dashed lines).

CD was used to estimate the secondary structure of MTA1(465–542). The size exclusion chromatography buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM DTT) contained chloride ions, which absorb strongly at wavelengths below 200 nm. Thus, the protein was dialysed into a buffer composed of 20 mM phosphate pH 7.5, 50 mM NaCl. The far-UV CD spectrum of MTA1(465–542) was then recorded (Figure 3-31). A minimum is observed at ~200 nm which is consistent with largely random coil structure. The data from the CD experiment show that MTA1(465–542) is largely unstructured in solution. There is ongoing work in our laboratory to characterise the interaction of this region with RbAp48.
Figure 3-31: A. Far-UV circular dichroism spectrum of MTA1(465–542). The spectrum was collected using a 15 µM protein sample in 20 mM phosphate pH 7.5, 50 mM NaCl at 20 °C. B. Ideal CD spectra for random coil (red), β-sheet (blue) and α-helix (orange) (Bolotina et al., 1980).

Due to time limitations, it was not possible to pursue the expression of MTA1(448-715) or MTA1(448-542). Instead, we focused our attention on one of the RbAp48-binding regions – namely MTA1(656-686) – in an effort to determine the mechanism by which it binds to RbAp48. This is described in Chapter 4.

3.11 Discussion

3.11.1 Two regions of MTA1 are implicated in the interaction with RbAp48

In this chapter, MTA1/2-RbAp46/48 interactions were identified using pulldown assays and site-directed mutagenesis. The MTA proteins are equipped with several domains – BAH, ELM2, SANT and the GATA zinc finger domain at the N-terminal end – that are highly conserved amongst human MTA proteins (Figure 3-3). MTA1 also has additional predicted domains at the C-terminus, including SH3-binding domain and a Myb-like domain. Although the N-terminal domains of MTA proteins may be involved in protein-protein interactions, we found that these domains were not able to interact with RbAp48, but instead found that the C-terminal part of MTA1 and MTA2 mediates the MTA-RbAp protein interactions.

An interaction between MTA1/2 and RbAp46/48 has previously been demonstrated in GST pulldowns. Roche et al. (2008) indicated that the zinc finger domain of
MTA1(392–448) is sufficient for the interaction with RbAp46/48. During the course of the present study, Fu et al. (2011) reported that three non-overlapping constructs of the C-terminal end of MTA2 interact with RbAp46, although the mapping of the MTA2-RbAp46 interaction was not unequivocal as RbAp46 was able to bind to all fragments of MTA2, only one of which contained the zinc finger domain (Fu et al., 2011). Our data show that the GATA zinc finger domain of MTA1 is not involved in the RbAp48 interaction; instead, the region C-terminal to the zinc finger domain is the interacting region.

It has been demonstrated that the C-terminal end of MTA1 also is required for the interaction with FOG1/2 (Roche et al., 2008), NRIF3 (Talukder et al., 2004) and MICOA (Mishra et al., 2003) as well as with H3 (Wu et al., 2013). This indicates the importance of the C-terminus of MTA1 for diverse interactions but without any structural studies.

Attempts to express the whole C-terminal end of MTA1(448–715) in bacteria and insect cells were overall unsuccessful. This may be due to the instability of MTA1(448–715), which tended to undergo degradation during purification. This region is predicted to be largely disordered (Figure 3-32A), which may have prevented the expression of this construct in a soluble form. Therefore, we narrowed down the interacting regions of the C-terminal portion of MTA1 to shorter regions, 465–516 and 656–686, which are each independently able to interact with RbAp proteins. It is notable that these two regions are the two parts of MTA1(448-715) that are predicted to have secondary structure (Figure 3-32B).

Both MTA1(465–516) and MTA1(656–686) were more soluble than the whole C-terminus, suggesting that they can be used in structural studies that require high protein concentrations. However, the yield and stability of both regions of MTA1 were not sufficient for our purposes. Furthermore, the MTA1(625–715) fragment was unfolded as determined by 1D NMR spectroscopy, whereas the MTA1(656-686) might exist in well-ordered conformation (Figure 3-23 and 3-25). It could be that the MTA1 construct folds upon binding to RbAp48 protein. Indeed, it has been reported that many small polypeptides are disordered in solution and probably fold upon binding to their partners (Dyson and Wright, 2002). Also, it has been shown that many eukaryotic proteins contain regions without well-defined structure (intrinsically disordered regions) and
disorder-to-order transition occurs upon binding to their targets. These proteins tend to be associated with specific functions such as cell regulation, signaling and protein or DNA binding, and it has been suggested that disorder in a protein could assist binding to multiple partners (Fong et al., 2009).

Figure 3-32: Secondary structure prediction of the C-terminus of MTA1. A. PONDR VL-XT graphs showing the predicted disorder profiles of MTA1(448–715). The black line in the middle of the graphs indicates the threshold of disorder (score of 0.5). B. Secondary structure prediction using Phyre software. Green helices represent α-helices, blue arrows indicate β-strands and faint lines indicate coil. The 'SS confidence' line indicates the confidence in the prediction, with red being high confidence and blue low confidence (Kelley and Sternberg,
It has been shown that PTMs are important for the function of MTA1. For example, it has been suggested that the acetylation of K626 in MTA1 might play a role in its interaction with HDAC2 and also in the inhibition of Gai2 (Guanine nucleotide-binding protein G(i) subunit alpha-2) expression (Ohshiro et al., 2010). Furthermore, SUMOylation on Lys-509 regulates the co-repressor activity of MTA1 on the PS2 promoter (Cong et al., 2011). In addition, based on Phospho Site an algorithm for the prediction of phosphorylation sites, from Cell Signalling Technology, most of the PTMs in MTA1 occur in the C-terminal portion (Figure 3-33). Taken together, MTA1 protein may be soluble only with correct post-translational modifications, especially glycosylation, which depends on the protein expression system. Therefore, we may require a mammalian system to enable PTMs and proper folding. It is notable in this regard that we were able to produce soluble MTA1(449–715) and other constructs from the C-terminal end of MTA1 using a rabbit reticulocyte lysate system, which enables some modifications such as glycosylation, acetylation, phosphorylation and may also contain chaperones or other factors that can assist with correct folding.

Figure 3-33: Modification sites on human MTA1 protein. This figure was taken from http://www.phosphosite.org/proteinAction.do?id=2945&showAllSites=true.

3.11.2 The KRAARR motif is conserved in MTA proteins and mediates the interaction with RbAp proteins

MTA1 and MTA2 share some sequence similarities in the C-terminus (Figure 3-3). The regions of MTA1 and MTA2 that bind to RbAp46/48 protein contain either a KRAARR or RRAARR motif, which is also conserved in MTA3 (Figure 3-33B). The proline
double point mutants in the KRAARR motif in MTA1(656–686 and 643–695) entirely disrupted the interaction with RbAp48, but not in the MTA1(449–715) construct, which indicates that there is another region(s) that can compensate for the loss of binding in this region. This was confirmed by the finding of another region of MTA1(448–542) which is important for the RbAp48 binding. Our pull downs and mutations showed that the KRAARR motif plays an essential role in the interaction with both RbAp46/48.

3.11.3 MTA1 binds to RbAp48 in the same way that H4 binds to RbAp46

The crystal structure of histone H4 bound to RbAp46 shows that an RRLARR motif in H4 mediates the interaction (Figure 3-34A) (Murzina et al., 2008). Site-directed mutagenesis of the KRAARR motif of MTA1, presented above, indicates the importance of this motif for the binding to RbAp48. Competition experiments provided additional evidence of the overlapping binding site of MTA1 and H4 on RbAp46/48. These data suggest that MTA1 may compete with H4 for binding to a single site of RbAp48, whereas FOG1 uses a unique binding site to bind RbAp48. Sequence alignment of MTA1/2/3 with the histone H4 binding regions to RbAp46 shows that they contain a similar pattern of conserved basic and hydrophobic residues (Figure 3-34B). These data support our hypothesis that the KRAARR motif is crucial for RbAp48 binding and it is likely that all three MTA-family proteins bind to RbAP46/48 using this motif.

3.12 Summary

In this chapter we defined two regions of MTA1, namely 465–516 and 656–686, that are able to interact with RbAp48. Our pulldown data show that the KRAARR motif is essential for the interaction with RbAp48 and that MTA1 and MTA2 bind to both RbAp46/48 proteins in the same manner. Efforts, however, to obtain sufficient expressed protein for biophysical and structural studies were unsuccessful, we therefore decided to use synthesised peptides to further probe the RbAp48-MTA1 interaction. Chapter 3 describes the biophysical and structural analysis of the MTA1-RbAp48 complex.
Figure 3-34: A KRAARR motif mediates binding to RbAp46/48. 

A. Sequence alignment of surrounding region of KRAARR motif of human MTA1 with the corresponding regions from MTA2 and MTA3 and human histone H4(25-47, Uniprot: P62805). The KRAARR motif is underlined and the residues that, in histone H4, make contact with RbAp46 in the crystal structure are shaded. The region of MTA1 predicted to be helical is indicated by a cylinder.

<table>
<thead>
<tr>
<th></th>
<th>hMTA1 (670-689)</th>
<th>hMTA2 (618-637)</th>
<th>hMTA3 (482-501)</th>
<th>hH4 (25-44)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KILSSSETKRAARRPYKPIA</td>
<td>KALTHLEMRAARRPNLPLK</td>
<td>VCKNTLRLQAARRPFVAIN</td>
<td>QGITKPAIRRLARRGGVKRI</td>
</tr>
</tbody>
</table>

B. Crystal structure of RbAp46 bound to histone H4 (PDB: 3CFS, (Murzina et al., 2008)).
Chapter 4. Crystal structure of the RbAp48-MTA1(656–686) complex

Chapter 2 described our first experiments to examine the molecular basis of the specific recognition of MTA1 by RbAp48. It was shown using GST pulldowns that a region of about 30 residues of the C-terminal end of MTA1 (~656–686) is able to bind independently to RbAp48. This chapter describes efforts to structurally characterise the RbAp48-MTA1 interaction.

In 2009, the crystal structure of RbAp48 alone (described in Section 4.4.2) was deposited into the Protein Data Bank (PDB: 3GFC, unpublished) by the Structural Genomics Consortium. The N-terminal helix and the PP loop were not modelled in this structure of RbAp48. More recently, the crystal structure of RbAp48 bound to FOG1 was determined in our laboratory (Lejon et al., 2011). In this structure the N-terminal and PP loop helices of RbAp48 were also not well defined, but overall, it is closely similar to RbAp48 alone. In addition, the crystal structure of RbAp46 bound to histone H4 has been solved (Figure 3-34) (Murzina et al., 2008). Due to the numerous modes that WD-repeat proteins can utilize to interact with their binding partners (discussed in Chapter 6), it was difficult to anticipate how RbAp48 would interact with MTA1 based on its structure alone. Therefore, our goal was to solve the crystal structure of the RbAp48 in complex with the MTA1(656–686) peptide to determine the mechanism that RbAp48 uses to interact with MTA1.

4.1 Baculovirus expression and purification of RbAp48

The most basic requirement for such an analysis is sufficient quantities of each protein that are folded and of a good level of purity. Initial expression trials of RbAp48 in E. coli (not shown) were unsuccessful because the bacterially expressed RbAp48 was insoluble. We therefore turned to insect cell expression, with the assistance of the Laue laboratory in the Department of Biochemistry at University of Cambridge (UK).

4.2 Verification of Bacmid-RbAp48

The donor pFBDM plasmid containing the human RbAp48 gene was a gift from Ernest Laue (University of Cambridge, UK). This plasmid was transformed into DH10BAC competent cells which contain baculoviral DNA. Recombinant baculovirus DNA (bacmid) was first generated by transposition of the RbAp48 gene (as a 6×His-tagged fusion containing a thrombin cleavage site between the His tag and RbAp48) into the
baculoviral DNA using the Invitrogen Bac-to-Bac system (Section 7.6). The transposition of RbAp48 gene into the bacmid was confirmed by PCR using a RbAp48-specific forward primer and M13 reverse primer. Isolated recombinant bacmid DNA was analyzed by PCR followed by 1% agarose gel electrophoresis; baculovirus transposition was confirmed by the appearance of DNA bands at \( \sim 2000 \) bp (665 bp bacmid +1275 bp RbAp48) (Figure 4-1A).

The recombinant bacmid containing the 6×His-RbAp48 was then transfected into a monolayer of Sf9 cells from the insect Spodoptera frugiperda using Fugene 6 transfection reagent (Promega) to produce recombinant baculovirus particles (P1 viral stock), which were then used to infect fresh Sf9 cells (25 mL at \( 1 \times 10^6 \) cells/mL) for further viral amplification. After three generations of this process (P1–P3), the baculovirus-containing supernatant harbouring the RbAp48 was collected and used as viral stock to express recombinant protein by infecting more Sf9 cells. To confirm the protein expression, samples of infected cells from each generation were collected and lysed. SDS-PAGE of total protein followed by western blot using monoclonal anti-His antibody demonstrated one polypeptide band with the predicted molecular weight of RbAp48 (\( \sim 49 \) kDa) for all viral stocks (P1–P3) (Figure 4-1B). The P3 viral stock was used for RbAp48 expression.
After generating a high-titer P3 baculoviral stock, we optimized the RbAp48 protein expression. To determine the optimal conditions for expression of RbAp48, we infected a population of Sf9 cells (25 mL at $1 \times 10^6$ cells/mL) with the P3 viral stock at varying multiplicity of infections (MOI; 0.1, 0.5, 1 and 5) and performed a time course for protein expression post-infection (24, 48, 72, 96 and 120 h post-infection) (Figure 4-2). We selected the MOI and time that provided the optimal level of recombinant protein expression, which was MOI 1 for ~80 h. To scale up protein expression, a suspension culture of Sf9 cells (500 mL at $1 \times 10^6$ cells/mL) was infected with 1.5 mL of P3 viral stock and left to grow with shaking at 27 °C for about 80 h.
107

4.3 Confirming the interaction between MTA1(656–686) peptide and insect cell produced RbAp48

In the previous Chapter, we showed that MTA1(656–686) can interact with RbAp48 in the context of GST pulldown assays carried out using $^{35}$S-labelled RbAp48 expressed in a reticulocyte lysate. In order to confirm that this interaction was direct, we first expressed RbAp48 with an N-terminal His tag in Sf9 insect cells and then purified the protein using nickel affinity chromatography, followed by removal of the His tag through thrombin cleavage. RbAp48 was then subjected to anion exchange then size exclusion chromatography (Figure 4-3). The purity obtained following this protocol was >90%, judging from SDS-PAGE analysis. Then, biotinylated-MTA1(656–686) peptide (chemically synthesized) was immobilised onto high-capacity streptavidin (SA) beads to pull down RbAp48. Figure 4-4A shows that the insect cell produced RbAp48 bound to biotinylated-MTA1(656–686) peptide, whereas it did not bind to SA beads alone.
To further narrow down the region of MTA1(656–686) that binds to RbAp48, and thereby assist in the crystallisation process, we used three short overlapping chemically synthesized peptides (P2–P4) (Figure 4-4) spanning the sequence 656–686 of MTA1: 656–670, 664–679 and 671–686. These biotinylated peptides (P2–P4) were immobilised onto SA beads to pull down RbAp48. It can be seen from Figure 4-4B that none of the three short overlapping peptides covering MTA1(656–686) were able to bind to RbAp48, whereas the positive control, b-MTA1(656–686), interacted with RbAp48. Therefore, we decided to use the longer peptide (656–686) of MTA1 for biophysical and structural studies.

Figure 4-3: RbAp48 purification. A. Elution profile from a SEC purification (using a HiLoad™ 16/60 Superdex™ 75 pg column) of RbAp48. Sample was in 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM DTT. B. SDS-PAGE analysis of RbAp48 purification. Lane M: protein standard in kDa; lanes 1-7: SEC fractions containing purified RbAp48 (theoretical molecular weight: 48.6 kDa).
Figure 4-4: SDS-PAGE showing results of the insect cell produced RbAp48 pulldown experiment using the biotinylated-peptides derived from MTA1(656–686). A. MTA1 peptides used in pulldowns. B. Biotinylated-MTA1(656–686) peptide was bound to streptavidin (SA) beads to pulldown RbAp48. Lane M: molecular mass standards indicated in kDa; Lane 1: negative control pulldown where RbAp48 was incubated with SA beads only; Lane 2: RbAp48 pulled down by SA beads bound to biotinylated-MTA1(656-686) peptide. C. Pulldown experiments of three overlapping peptides spanning the sequence 656–686 of MTA1. Biotinylated MTA1 (b-MTA1) peptides 656–670, 664–679 and 671–686 were bound to SA beads to pulldown RbAp48. Lane M: molecular mass standards indicated in kDa; Lane 1: RbAp48 alone; Lane 2: negative control (RbAp48 incubated with SA beads); Lane 3: positive control (RbAp48 incubated with b-MTA1(656–686)); Lanes 4, 5 and 6: RbAp48 incubated with b-MTA1(656–670), b-MTA1(664–679) and b-MTA1(671–686), respectively.

4.4 Determining the affinity of the RbAp48-MTA1(656–686) interaction
To confirm the interaction using another approach and also determine the affinity of RbAp48 for MTA1(656–686), isothermal titration calorimetry analysis was performed.

4.4.1 Isothermal titration calorimetry
Isothermal titration calorimetry (ITC) is a biophysical technique used to measure the affinity and other thermodynamic parameters of an interaction in solution. The use of ITC to measure the binding of a macromolecule to its partner/ligand relies on the fact that such an interaction is generally accompanied by a heat effect. When a concentrated
solution of protein is titrated into a dilute solution of a partner, the absorption or generation of heat can be monitored by the calorimeter. This heat changes that take place during a titration can be integrated and used to obtain information about the binding constant (Ka), the enthalpy of binding (∆H) and stoichiometry (N); the entropy change (∆S) can also be indirectly calculated (Pierce et al., 1999).

In ITC, a syringe containing a ‘ligand’ solution is titrated into a cell containing a solution of the ‘receptor’ at constant temperature and the heat released or absorbed is measured, which in turn is directly proportion to the amount of binding. As the receptor in the cell becomes saturated with ligand, the heat signal diminishes until only the background heat of dilution is observed.

To determine the binding affinity for the interaction between RbAp48 and MTA1(656–686), the MTA1(656–686) peptide (250 µM) was titrated into a solution of RbAp48 (25 µM). Titrations of MTA1(656–686) into buffer alone were also carried out to measure the background heats of dilution, which were then subtracted from the RbAp48/MTA1 titration. The background-corrected RbAp48-MTA1 titration data were then fitted to a 1:1 binding model using the Microcal Origin software (Figure 4-5). The K_D of the interaction was determined to be 2.3 ± 0.3 µM. Each titration was carried out in duplicate.
Figure 4-5: Isothermal titration calorimetry data for binding of MTA1(656–686) peptide to RbAp48 at 25 °C. The upper panel shows the sequential heat pulses for MTA1-RbAp48 binding and the lower panel shows the integrated heat data, corrected for dilution and fit with a single-site binding model using Origin software.

4.5 Crystal structure of the RbAp48–MTA1 subcomplex of NuRD

The production, purification and crystallisation work described in this section were performed with assistance from Sara Lejon and Andal Murthy in the Laue laboratory at the University of Cambridge, UK.

4.5.1 Crystallisation procedure

To determine the molecular basis of the interaction, we first sought to grow crystals of a complex of RbAp48 and the MTA1(656–686) peptide.
Synthesized MTA1(656–686) peptide was added to purified RbAp48 (10 mg/mL) in a 1:5 protein:peptide molar ratio and used to set up crystallisation trials using the Morpheus protein crystallisation screen (Molecular Dimensions; Gorrec, 2009) and Hampton IndexHT screen in 96-well plates (sitting drop vapour diffusion, 4 °C). Crystals appeared after 1 day in Morpheus screen condition D1 (30% PEG550, MME/PEG20K, 0.1M MES/imidazole pH 6.5 and 10% alcohols), and grew to their final size within 4 days (Figure 4-6A). Crystals grown from condition D1 were harvested into a cryoprotectant solution containing 25% PEG 400 in mother liquid before cryocooling in liquid nitrogen.


4.5.2 Structure determination of the RbAp48-MTA1(656–686) complex

X-ray diffraction data were collected at the Diamond Light Source Synchrotron (Oxfordshire, UK) at Beamline I04 (Figure 4-6B).

Preliminary analysis revealed that the grown crystal belongs to a monoclinic system with space group \( P2_1 \) and unit cell parameters \( a = 81.63 \, \text{Å}, \, b = 59.46 \, \text{Å}, \, c = 104.71 \, \text{Å}, \, \alpha = 90.00^\circ, \, \beta = 89.95^\circ, \, \gamma = 90.00^\circ \) with two complexes per asymmetric unit. The X-ray diffraction data were processed to a maximum resolution of 2.5 Å. Data were processed and scaled using MOSFLM (Leslie and Powell, 2007) and Scala (Evans,
The structure was solved by molecular replacement with the program Phaser (McCoy et al., 2007) using our previously solved structure of RbAp48 (PDB code 2XU7) as a search model. The model was built with Coot (Emsley and Cowtan, 2004) and refined using Refmac5 (Murshudov et al., 1997). Non-crystallographic symmetry (NCS) restraints were used in the refinement process. MolProbity (Davis et al., 2007) was used to validate the structure and the EBI PISA server (Krissinel and Henrick, 2007) was used for interface analysis. Data collection and refinement statistics are presented in Table 4-1. MolProbity validation is shown in Figure 4-7.

\( R \) values reflect on the deviation of structure factor amplitudes calculated from the model to those derived experimentally. In the ideal case, when the structure model is in perfect agreement with the experimental data set, the \( R \) value would be zero. In practice, \( R \) values of 10–15% are typically obtained from refinement of very high resolution diffraction data (~1–1.5 Å) and \( R \) values of 18–22% for data sets collected to ~2.5 Å resolution. The \( R_{\text{free}} \) value is an important check during the process of refinement that can indicate whether the model is being overrefined (Brünger, 1995). For the determination of the \( R_{\text{free}} \) value about 5% of unique reflections are removed from refinement and structure factors calculated from these reflections can be compared to those calculated from the experimental data used for refinement. If the model is being correctly refined then there should be minimal discrepancy between the \( R_{\text{free}} \) and \( R_{\text{work}} \) values. \( R_{\text{free}} \) values should be similar to the \( R_{\text{work}} \) for well refined structures. The reported \( R_{\text{work}} \) and \( R_{\text{free}} \) values after refinement of the RbAp48–MTA1(656–686) complex are 18.7% and 22.9%, respectively (Table 4-1), and represent a well refined model for data collected to 2.5 Å resolution.
### Table 4-1: Table of data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Space Group</th>
<th>$P2_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unit Cell Dimensions</strong></td>
<td></td>
</tr>
<tr>
<td>$a, b, c$ (Å)</td>
<td>81.56, 59.48, 104.63</td>
</tr>
<tr>
<td>$\alpha, \beta, \gamma$ (%)</td>
<td>90.00, 89.95, 90.00</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>52.31-2.5(2.64-2.50)*</td>
</tr>
<tr>
<td>$R_{\text{merge}}$ (%)</td>
<td>12.2 (44.8)</td>
</tr>
<tr>
<td>Total observations</td>
<td>122920(18058)</td>
</tr>
<tr>
<td>Unique observations</td>
<td>34948(5037)</td>
</tr>
<tr>
<td>Mean $(I/\sigma(I))$</td>
<td>10.6 (2.4)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.6 (99.8)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>3.5(3.6)</td>
</tr>
<tr>
<td>Wilson $B$-factor (Å$^2$)</td>
<td>33.0</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20.0-2.50</td>
</tr>
<tr>
<td>No of reflections</td>
<td>33186</td>
</tr>
<tr>
<td>$R_{\text{work}}$/$R_{\text{free}}$</td>
<td>0.180/0.228</td>
</tr>
<tr>
<td><strong>RMSD values</strong></td>
<td></td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>0.010</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.399</td>
</tr>
<tr>
<td>ChirVolume</td>
<td>0.080</td>
</tr>
<tr>
<td>No of atoms</td>
<td>13644</td>
</tr>
<tr>
<td>$B$-factor (Å$^2$)</td>
<td>25.30</td>
</tr>
<tr>
<td>Ramachandran values (%)</td>
<td></td>
</tr>
<tr>
<td>Most favoured</td>
<td>95.0 (766 residues)</td>
</tr>
<tr>
<td>Additionally allowed</td>
<td>4.8 (38 residues)</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.2 (2 residues)</td>
</tr>
</tbody>
</table>

*Values for highest resolution shell are shown in parentheses

---

$a R_{\text{merge}} \ (I) = \frac{\sum_{ijk} \sum | I_{ijk} - (I_{ijk}) |}{\sum_{ijk} \sum | I_{ijk} |}$, where $(I_{ijk})$ is the average intensity of the multiple $I_{ijk}$ observed for symmetry-related reflections.

$b \text{Mean } (I/\sigma(I))$, average of the diffraction intensities, divided by their standard deviation.

c $R_{\text{work}} = \frac{\sum_{ijk} | F_o - F_c |}{\sum | F_o |}$, where $F_o$ and $F_c$ are observed and calculated structure factors, respectively.

d $R_{\text{free}} = \frac{\sum_{ijk} | F_o - F_c |}{\sum | F_o |}$, $R_{\text{free}}$ was calculated with 5% of the reflections not used during refinement.

e Ramachandran plot statistics were calculated by MolProbity (Lovell et al., 2003).
Figure 4-7: MolProbity Ramachandran analysis of the RbAp48-MTA1(656–686) complex. This analysis is for the two molecules in the asymmetric unit. Of all residues 95.0% (766/806) are in favoured regions, 4.8% (38/806) are in additional allowed regions and 0.2% (2/806) are in outliers. Ramachandran plot statistics were calculated with MOLPROBITY, http://molprobity.biochem.duke.edu.

*B* factors, or temperature factors, can also be used to evaluate errors in a structure model and thus reflect on their integrity. *B* factors indicate the degree of fluctuation or the
occupancy of atoms in their positions. Thus, $B$ factors can reflect on protein dynamics. $B$ factors in crystal structures are not only a function of inherent mobility of the atoms in protein, but also depend on experimental conditions used for data collection. Generally, the higher the reported $B$ factors, the greater the degree of displacement of an atom from its equilibrium location in the molecule (Yuan et al., 2005, Parthasarathy and Murthy, 2000). It is typical to obtain $B$ factors of 10–20 Å$^2$ for data collected at very high resolutions and up to 40 for data collected to ~2–2.5 Å resolution. The reported mean protein $B$ factor for the crystal structure of RbAp48–MTA1(656–686) is ~25 Å$^2$ (Table 4–1). This value corresponds to the average $B$ factor for all protein atoms in the PDB file of RbAp48-MTA1(656–686), which consists of two molecules in the asymmetric unit. The structure of RbAp48 is well ordered except for the N-terminal residues 8–15 and some loops, whereas MTA1 peptide shows some flexibility as depicted by a $B$-factor cartoon representation (Figure 4–8).

The MTA1 peptide was modelled using positive difference density maps ($F_o-F_c$) calculated from the initial molecular replacement solution. After modelling RbAp48, the remaining electron density fitted well to residues 674–686 of MTA1 (Figure 4–9). After modelling RbAp48, the m$F_o$-DF$^c$ difference map clearly showed the presence of MTA1 peptide in both molecules. The final model consists of residues 3–411 of RbAp48 and residues 674–686 of MTA1.

![Figure 4-8: A B-factor cartoon representation of RbAp48 shows that the structure is mostly ordered, with N-terminal residues 8–15, the loops and MTA1(674–678) showing significant flexibility. The colours indicate average B factor values for each residue with the lowest values in blue and the highest in red.](image-url)
Figure 4-9: Electron density map of MTA1(656–686). A. The remaining electron density (Fo-Fc) after the RbAp48 protein was modelled. The density map is shown as a blue mesh contoured at 1.0 σ. B. The same as in (A) except that the MTA1 peptide is modelled and shown as yellow stick. In both cases, RbAp48 is shown as gray surface.

The structure of RbAp48 is well defined, except for the first two residues at the N-terminus and the C-terminus (residues 412–425). These residues displayed poor electron density, which is likely to have resulted from motional disorder in those regions in the crystal. Thirteen residues of the MTA1 peptide (sequence NH$_2$-SSETKRAARRPYK-COOH) are defined in the crystal structure, and effectively no density was observed for the remaining 17 residues. The two copies of RbAp48 present in the asymmetric unit were highly similar and superimposed with an RMSD of 0.21 Å over 331 Ca atoms (Figure 4-10A).
Chapter 4

4.5.3 Crystal structure of the RbAp48-MTA1(656–686) complex

As expected, the RbAp48 structure forms a seven-bladed β-propeller. Each blade consists of a four-stranded anti-parallel β-sheet, as in other family members (Figure 4-11). The structure of RbAp48 overall is composed of three distinct regions: an extending N-terminal α-helix (D10 to P29), which sits on the edge of blade seven; the seven blades of the β-propeller (Y32 to A405); and a short C-terminal α-helix (E406 to N410), which sits on top of and extends the N-terminal α-helix (Figure 4-11B). The RbAp48 structure also accommodates a unique negatively charged loop with a short α-helix (residues S348 to E365) protruding from blade six, which ends in P363–P364 (the PP loop).

A surface representation of RbAp48 shows it to have an overall doughnut shape with a hole through the centre (Figure 4-11C and D). An electrostatic surface potential representation shows that RbAp48 is strongly electronegative (Figure 4-12), particularly along the centre hole on each side due to clustering of strongly electronegative residues found in these regions (E41, E126, E179, E231, E275, D379, E395 and D396). This
appears to be a property of a subset of WD40 proteins, since charge distributions vary among other β-propeller structures, indicating that the net negative charge may be present for a specific functional purpose (described in Chapter 6).

In general, the conformation of RbAp48 is not significantly affected by the formation of the complex with MTA1(656–686). An overlay of the RbAp48-MTA1 structure with that of RbAp48–FOG1(1–15) shows that the conformations of the two structures are very similar with an RMSD of 0.25 Å over 315 Ca atoms (Figure 4-10B). The structure reveals that MTA1 binds to a groove at the ‘side’ of the β-propeller of RbAp48. Thus, unlike FOG1, which binds in an extended conformation to the ‘top’ surface of the β-propeller (Figure 4-10B), we find that MTA1 adopts an α-helical conformation that binds to the side pocket of RbAp48 that is located between the N-terminal α-helix, C-helix and the PP loop (Figure 4-13A). The MTA1 binding site on RbAp48 consists of a negatively charged face composed of residues extending primarily from the short helix in the PP loop and a hydrophobic face composed of the N-terminal α-helix and residues from blade six (Figure 4-13B). The protein-protein interaction buries 776 Å² of exposed surface area at the interface of RbAp48 and MTA1.
Figure 4-11: Overview of the RbAp48-MTA1(656–686) complex. A. Top-down view of the structure showing the seven-bladed β-propeller fold, with each blade shown in a different colour; the structure is rotated 90° about the x-axis in (B). The seven blades of RbAp48 are labelled 1–7, starting from the N-terminal end. MTA1(656–686) (yellow) binds on the side edge of RbAp48. C. Space-filling representation of RbAp48-MTA1(656–686) rendered semi-transparent and rotated 90° about the x-axis in (D).
Figure 4-12: Surface characteristics of RbAp48. A. Electrostatic surface potential representation of RbAp48 shows that the core region is strongly electronegative on both sides (B and C, rotated -90° and +90° about the x-axis compared to (A), respectively, to illustrate both sides). The binding site of MTA1(656–686, yellow) is shown in (A). Acidic patches are shown in red, neutral patches in white and basic patches in blue (overall range from −85 to +85 kT).

Close examination of the RbAp48-MTA1(656–686) interface reveals many specific interactions of RbAp48 with MTA1, comprising a network of hydrogen bonds, salt bridges and Van der Waals contacts. The amphipathic KRAARR motif in MTA1 interacts with complementary hydrophobic and charged residues of the RbAp48 binding site. Residues A680 and A681 on one side of the MTA1 helix associate with a hydrophobic face on RbAp48 comprising residues F30, L31, F368, I369, and I408 (Figure 4-14). Through an extensive network of hydrogen bonds and salt bridges the
hydrophilic patch of the MTA1 helix associates with RbAp48. Notably, the positively charged R679, R682 and R683 (the KRAARR motif) of MTA1 make the most contacts with the binding pocket of RbAp48. The side chain of R682 forms a network of hydrogen bonds with the main chain carbonyl groups of RbAp48 (D358, G362, P363 and L366) and R683 forms hydrogen bonds and salt bridges with the side chains of E357 and D361 (Figure 4-14). R679 forms hydrogen bonds and salt bridges with D358. In addition, K678 of MTA1 is another key residue, interacting with the C-terminal α-helix of RbAp48 through side chain hydrogen and ionic bonds. These contacts are summarised in Table 4-2.

![Figure 4-13: The binding pocket of RbAp48–MTA1(656–686).](image)

A. MTA1 binding pocket between N-helix, PP loop in blade 6 and C-helix. RbAp48 is shown in orange and salmon, and MTA1 in yellow. B. Electrostatic surface potential representation of RbAp48 (overall range from −85kT (red) to +85kT (blue)) with the MTA1 peptide shown as a cartoon and stick model in yellow. The MTA1-binding pocket in RbAp48 is mainly formed by the negatively charged PP loop (which terminates in P363 and P364) and a hydrophobic surface on the N-terminal α-helix. The interaction of MTA1 residues (S674, K678, R679, A680, A681, R682, and R683) is shown. Hydrogen bonds are indicated with dotted lines.
The positively charged K678, R679, R682 and R683 residues of MTA1 interact with the backbone carbonyl groups in the PP loop and a cluster of acidic residues (G362, L366, I369, D358, and D361) in RbAp48. In addition, the hydrophobic A680 and A681 residues in MTA1 interact with F30 and L31 in the N-terminal helix of RbAp48. Hydrogen bonds are indicated with dotted lines. RbAp48 residues are shown in orange sticks and MTA1 residues are shown as yellow sticks.

Figure 4-14: Detailed view of the interactions of RbAp48 with MTA1(656–686). The positively charged K678, R679, R682 and R683 residues of MTA1 interact with the backbone carbonyl groups in the PP loop and a cluster of acidic residues (G362, L366, I369, D358, and D361) in RbAp48. In addition, the hydrophobic A680 and A681 residues in MTA1 interact with F30 and L31 in the N-terminal helix of RbAp48. Hydrogen bonds are indicated with dotted lines. RbAp48 residues are shown in orange sticks and MTA1 residues are shown as yellow sticks.
**Table 4-2: Specific interactions between RbAp48 and MTA1**

<table>
<thead>
<tr>
<th>MTA1 residue</th>
<th>RbAp48 residue</th>
<th>Interaction type and distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SER 674</td>
<td>ASN 407</td>
<td>H-bond 3.4</td>
</tr>
<tr>
<td>LYS 678</td>
<td>ASN 410</td>
<td>H-bond 3.7</td>
</tr>
<tr>
<td>LYS 678</td>
<td>ASN 407</td>
<td>Salt bridge 3.0</td>
</tr>
<tr>
<td>ARG 679</td>
<td>ASP 358</td>
<td>Salt bridge 3.3</td>
</tr>
<tr>
<td>ARG 679</td>
<td>ASP 358</td>
<td>H-bond 3.0</td>
</tr>
<tr>
<td>ARG 679</td>
<td>ASP 358</td>
<td>H-bond 3.8</td>
</tr>
<tr>
<td>ARG 679</td>
<td>ASP 358</td>
<td>Salt bridge 3.8</td>
</tr>
<tr>
<td>ARG 679</td>
<td>ASP 358</td>
<td>Salt bridge 3.3</td>
</tr>
<tr>
<td>ALA 681</td>
<td>ILE 369</td>
<td>H-bond 3.1</td>
</tr>
<tr>
<td>ARG 682</td>
<td>PRO 363</td>
<td>H-bond 2.6</td>
</tr>
<tr>
<td>ARG 682</td>
<td>LEU 366</td>
<td>H-bond 3.0</td>
</tr>
<tr>
<td>ARG 682</td>
<td>ASP 358</td>
<td>H-bond 2.5</td>
</tr>
<tr>
<td>ARG 682</td>
<td>GLN 354</td>
<td>H-bond 3.2</td>
</tr>
<tr>
<td>ARG 682</td>
<td>GLY 362</td>
<td>H-bond 2.7</td>
</tr>
<tr>
<td>ARG 683</td>
<td>ASP 361</td>
<td>H-bond 3.1</td>
</tr>
<tr>
<td>ARG 683</td>
<td>ASP 361</td>
<td>H-bond 2.9</td>
</tr>
<tr>
<td>ARG 683</td>
<td>ASP 361</td>
<td>H-bond 3.1</td>
</tr>
<tr>
<td>PRO 684</td>
<td>ASN 27</td>
<td>H-bond 3.1</td>
</tr>
</tbody>
</table>

### 4.6 Summary

#### 4.6.1 The biotin tag interfered with the binding of MTA1(671–686) peptide to RbAp48

Although MTA1(671–686) (LLSSSETKRAARRPYK-Biotin, P4 in Figure 4-4) contains the sequence that we observed (674–686, SSETKRAARRPYK) in the crystal structure (Figure 4-15), it did not show an interaction with RbAp48 in pulldown experiments, whereas the longer peptide (656–686, P1), which included the same residues but an N-terminal biotin, showed binding to RbAp48 (Figure 4-4). This may be due to the C-terminal biotin tag on the MTA1 P4 peptide, which is close to the binding residues and may interfere with the RbAp48 interaction.
Figure 4-15: MTA1 sequence used in pulldown assay and observed in the RbAp48-MTA1 crystal structure. MTA1 peptide (P1) has an N-terminal biotin tag, whereas MTA1 peptide (P4) has a C-terminal biotin tag. The MTA1(674–686) was observed in the crystal structure of RbAp48-MTA1(656–686).

4.6.2 The KRAARR motif in MTA proteins can bind RbAp proteins

Our crystallographic and biochemical data show that a short motif near the C-terminal end of MTA1 is able to independently bind to RbAp48. Specifically, the region including the KRAARR motif of MTA1 takes up a conformation that includes a short α-helix and a hydrophobic cluster, and binds to RbAp48 in a distinct pocket formed by three helical elements in the latter protein.

Although the human MTA1–3 proteins are highly divergent at the C-terminal end, they do share some conserved residues, including the KRAARR/RRAARR motifs (Figure 4-16), which most likely would mediate the binding of MTA2 and MTA3 to both RbAp48 and RbAp46. A more detailed discussion of MTA1-RbAp48 interaction is given in the Chapter 6 (General discussion).

Figure 4-16: Sequence alignment of the human MTA proteins. The surrounding region of the KRAARR motif of human MTA1 is aligned with the corresponding regions from MTA2 and MTA3. The conserved residues are shaded. The helical region of MTA1 is indicated by a cylinder.

4.6.3 Mutagenesis data corroborates the crystal structure

The crystal structure was in good agreement with the data from Chapter 2. As one would predict based on sequence similarity with histone H4, a number of contacts are made by the basic residues in the KRAARR motif. In particular, the side chain of R682 forms a network of hydrogen bonds with the main chain carbonyl groups of D358, G362, P363 and L366, whereas R683 forms hydrogen bonds and salt bridges with the
side chain of D361. The importance of these interactions is demonstrated by our mutagenesis data showing that an R682A/R683A double mutation substantially reduced the binding to RbAp48 (Section 3.7.1). Similarly, R679 makes an electrostatic interaction with the D358 side chain, whereas K678 makes interactions with N407 (backbone) and N410 (side chain). Mutating these two residues to alanines or prolines reduced or disrupted the binding to RbAp48. Conversely, mutations in the histone H4-binding pocket of RbAp46 either in the hydrophobic face (L30Y), the charged residues (E356Q + D357N + E359Q + D360N), or both simultaneously, strongly reduced the binding to MTA1.

4.6.4 Comparison of the RbAp48-MTA1 interaction with the RbAp46-H4 interaction

As anticipated from the sequence alignment, pulldown and mutagenesis results, we find that MTA1 binds to a side groove of RbAp48 which is the same as the binding site for H4 on RbAp46. Preliminary characterisation of the MTA1(656–686) fragment (using CD and 1D NMR) indicated that it is unstructured or partially structured in solution, indicating that this region undergoes a disorder-to-order transition upon binding to RbAp48.

The crystal structure of histone H4 bound to RbAp46 (Murzina et al., 2008) indicates the importance of the IRRLARR motif of H4 for RbAp46 interaction, where arginines separated by hydrophobic residues play a crucial role in the binding. Our data show that a similar motif in MTA1 (TKRAARR) mediates the interaction with RbAp48. The structures of RbAp46 and RbAp48 are identical; therefore, it is likely that they each associate with MTA1/H4 in a similar fashion since the MTA1/H4-interacting residues are identical in both RbAp proteins; this similarity is discussed in more detail in Chapter 6. Both MTA1 and H4 adopt an α-helical conformation to bind their partners (RbAp48 or RbAp46) in an overlapping binding site, which is located on the side of the β-propeller between the N-terminal α-helix and the PP loop (Figure 4-17A). The amphipathic MTA1 and H4 motifs interact with the complementary hydrophobic and charged residues of the RbAp46/48 binding site. Residues N27, Q354, D358, D361, G362, P363, L366, I369, N407 and D411 of RbAp48 and the identical residues on RbAp46 form the hydrogen bonds and ion pairs with MTA1 and H4 (Figure 4-17B and
C), whereas F30, L31 and F368 of RbAp48 and the same residues on RbAp46 make hydrophobic interactions with both MTA1 and H4.

Comparison of the H4 and MTA1 complexes shows that the H4 adopts a longer $\alpha$-helix in the C-terminal direction than does the MTA1 peptide. In this context, it is also notable that electron density was observed all the way to the C-terminal residue of MTA1(656–686). Thus, in order to determine whether we might have truncated the MTA1 peptide too early at the C-terminal end, we decided to investigate the RbAp48-MTA1 interaction using longer peptides. This work is described in Chapter 5.
Figure 4-17: Comparison of the RbAp48-MTA1 and RbAp46-H4 complexes. A. Ribbon diagram showing an overlay of RbAp48-MTA1(656–686) and RbAp46–H4(16–41). Both MTA1 and H4 bind to the corresponding grooves on the side of RbAp46/48, between the N-helix, PP loop and C-helix. B. Detail of the RbAp48-MTA1(656–686) structure, showing interactions made by the basic side chains of MTA1. C. Detail of the RbAp46-H4(16–41) structure, showing interactions made by the basic side chains of H4. Intermolecular hydrogen bonds formed in the complex are indicated as dotted lines. In all parts, RbAp48 is shown in red and MTA1 in magenta, whereas RbAp46 is shown in green and H4 in cyan.
Chapter 5. Crystal structure of the extended MTA1 peptides (670–695) and (670–711) bound to RbAp48

5.1 Additional contacts may be observed using additional C-terminal amino acids

Using pulldown experiments, we demonstrated in Chapter 2 that the residues 656–686 of MTA1 are the minimum region for mediating the interaction with RbAp48. We also used ITC to confirm the interaction and measure the affinity of this fragment to RbAp48, then determined the crystal structure of MTA1 bound to RbAp48 (Chapter 4). We used a 30-residue peptide spanning 656–686 of MTA1 to crystallize with RbAp48, and observed in the structure that we could model the MTA1 peptide until its very C-terminal residue. The high quality of electron density at the C-terminal end of MTA1(656–686) peptide (Figure 5-1) suggested that additional interactions might be observed using a peptide that was extended further in a C-terminal direction. In support of this hypothesis, we noted that the histone H4 α-helix in the H4-RbAp46 structure is longer than the MTA1 helix in our MTA1-RbAp48 structure. Thus, we decided to try extending our MTA1 peptide at the C-terminus to see whether there are any additional contacts made between the two proteins.

Figure 5-1: Electron density map (2|Fo|−|Fc|) contoured at the 1 σ level of the C-terminal residues of MTA1(656-686). C-terminal Tyr-685 and Lys-686 of MTA1 in RbAp48-MTA1(656-686) complex indicate the high quality electron density. Carbon, nitrogen, and oxygen atoms are represented in yellow, blue and red, respectively.
5.2 Determining the affinity of the RbAp48-MTA1(670–695 and 670–711) interaction

5.2.1 Isothermal titration calorimetry

In order to make affinity measurements and collect structural data, we designed two longer peptides MTA1(670–695 and 670–711) (Figure 5-2). To verify the ability of these MTA1 peptides to bind RbAp48 and to determine their binding affinities, ITC analysis was carried out.

To determine the binding affinity of the interaction between RbAp48 and the MTA1 peptides (670–695 and 670–711), a stock solution of each MTA1 peptide (250 µM) was titrated into a solution of RbAp48 (25 µM) in separate experiments. Titrations of MTA1 peptide into buffer alone were also carried out to measure the background heats of dilution, which were then subtracted from data obtained for the RbAp48/MTA1 titration. The background-corrected RbAp48-MTA1 titration data were then fitted with a 1:1 binding model using the Microcal Origin software. In both cases, good quality fits were obtained and the binding affinity (K_D) of the interaction of RbAp48 with MTA1(670–695) was determined to be 0.05 ± 0.007 µM, whereas the binding affinity for the RbAp48-MTA1(670–711) interaction was 0.24 ± 0.16 µM (Figure 5-3). Each titration was carried out in duplicate.
Figure 5-3: Isothermal titration calorimetry data for binding of MTA1(670–695) and MTA1(670-711) peptides to RbAp48 at 25 °C. A. ITC profile for binding of MTA1(670-695) to RbAp48. B. ITC profile for binding of MTA1(670-711) to RbAp48. In both cases, the upper panel shows the sequential heat pulses for MTA1 peptide-RbAp48 binding and the lower panel shows the integrated heat data, corrected for dilution and fit with a single-site binding model using Origin software. The peptide sequence and binding affinity derived from the fit are indicated. Each titration was carried out in duplicate.

As shown in Table 5-1, RbAp48 bound to the MTA1 extended peptides (670–695 and 670–711) with a higher affinity than to the MTA1(656–686) peptide. The MTA1 peptide 670–695 had the highest affinity (~0.05 µM) for RbAp48.

Taken together, the ITC data and the structure of RbAp48-MTA1(656–686) suggested that additional contacts are likely to be seen in these extended peptides. We therefore decided to try to crystallize the new peptides with RbAp48 protein.
Table 5-1: Binding affinities of RbAp48 for MTA1 peptides

<table>
<thead>
<tr>
<th>MTA1 peptide sequence</th>
<th>K_D  (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 656-DVFYMATEETRKIRKLLSSSETKRAARRPYK-686</td>
<td>2.3</td>
</tr>
<tr>
<td>2 670-LLSSSETKRAARRPYKPIALRQSQA-695</td>
<td>0.05</td>
</tr>
<tr>
<td>3 670-LLSSSETKRAARRPYKPIALRQSQALPPPPPAFVNDEPI-711</td>
<td>0.24</td>
</tr>
</tbody>
</table>

5.3 Crystal screen of RbAp48-MTA1 (670–695 and 670–711) complexes

5.3.1 Crystallisation procedure

Preliminary crystallization trials were set up using Qiagen NeXtal DWBlock screening kits: Classics Suite, JCSG Core, PACT Suite, PEGs suite and JCSG suite. Concentrated RbAp48 (~12 mg/mL) was mixed either with MTA1(670–695) or (670–711) peptide at a molar ratio of 1:1.2. Initial crystallization screens were set up using a MOSQUITO crystallization robot as sitting drops with 0.2 µL each of protein-peptide solution and crystallization solution and equilibrated against 75 µl of the crystallization solution in 96-well trays. Crystals were obtained after incubation overnight at 18 °C in several conditions (Figure 5-4).
Figure 5-4: RbAp48-MTA1 peptides (670–695 and 670–711) crystal screen. Crystals obtained from different conditions in the initial screening. Panels A-F are crystals from RbAp48-MTA1(670–695) and panels G-I are crystals from RbAp48-MTA1(670–711). Crystallisation conditions are: A, JCSG Suite solution 42: 0.2 M MgCl₂, 0.1 M Tris pH 8.5 and 20% (w/v) PEG 8000; B-E: PACT Suite solution 16: 0.1 M MIB buffer pH 7 and 25% (w/v) PEG 1500; solution 17: 0.1 M MIB buffer pH 8 and 25% (w/v) PEG 1500; solution 45: 0.2 M LiCl, 0.1 M Tris pH 8 and 20% (w/v) PEG 6000; solution 65: 0.2 M NaNO₃, 0.1 M Bis Tris propane pH 6.5 and 20% (w/v) PEG 3350, respectively; F-H: PEGs Suite solution54: 0.2 M NaCl and 20% PEG 3350; solution 54: 0.2 M NaCl and 20% PEG 3350; solution 61: 0.2 M NaSCN and 20% (w/v) PEG 3350, respectively; I Classics Suite solution 13: 0.2 M Tris pH 8.5 and 20% (v/v) Ethanol.

Single crystals of RbAp48-MTA1(670–711) were observed after 1 day in the PEG screen in condition G4 (0.2 M calcium acetate and 20% (w/v) PEG 3350), which grew to their final size within a week (Figure 5–5) and were used for data collection. These crystals were harvested into a cryoprotectant solution containing 25% PEG 400 in mother liquor before cryocooling in liquid nitrogen.

5.3.2 Structure determination of the RbAp48–MTA1(670–711) complex

X-ray diffraction data were collected on our home diffractometer with the assistance of Dr Ana Silva (Mackay laboratory).

Several crystals from the screen of RbAp48–MTA1(670–695) and RbAp48–MTA1(670–711) were tested for diffraction and the crystal grown in condition G4 of PEG screen of RbAp48–MTA1(670–711) yielded the best diffraction images. The crystal diffracted to 2.5 Å resolution and the data could be indexed to space group \( P2_1 \) with two molecules of the complex per asymmetric unit. Data were processed and scaled using MOSFLM (Battye et al., 2011) and Scala (Evans, 2006). The structure was solved by molecular replacement with the program Phaser (McCoy et al., 2007) using our previous RbAp48 structure (described in Chapter 3) as a search model. The model was built with Coot (Emsley and Cowtan, 2004) and refined using Refmac5 (Murshudov et al., 1997). Non-crystallographic symmetry (NCS) restraints were used in the refinement process. MolProbity (Davis et al., 2007) was used to validate the structure and the EBI PISA server (Krissinel and Henrick, 2007) was used for interface analysis. The mFo-DFc difference map showed clear positive electron density for MTA1 residues 670–711 in both RbAp48 molecules in the crystallographic asymmetric
unit and the structure re-refined to a final $R_{\text{work}}/R_{\text{free}}$ of 0.21/0.25. Table 5-2 shows the crystallographic statistics of the two RbAp48-MTA1 structures, 656–686 and 670–711. MolProbity validation is shown in Figure 5-7. The final model consists of RbAp48 residues 7–411 and MTA1 residues 671–690. The remainder of the RbAp48 and MTA1 residues did not have sufficiently well defined electron density to be modelled with confidence. The lack of clear electron density might arise either from irregularities in the packing of the crystal (static disorder) or from motion of these regions within the crystal (dynamic disorder).

The two copies of RbAp48 in the asymmetric unit are highly similar, and superimpose with an RMSD of 0.16 Å over 379 Cα atoms (Figure 5-6).

![Figure 5-6: Overlay of the two copies of RbAp48-MTA1(670–711) in the asymmetric unit. The two complexes can be seen to be very similar overall. RbAp48 is shown in red and cyan and MTA1 in yellow and magenta.](image)
### Table 5-2: Crystallographic Data and Refinement Statistics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space Group</strong></td>
<td><em>P2₁</em></td>
<td><em>P2₁</em></td>
</tr>
<tr>
<td><strong>Unit Cell Dimensions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>a, b, c</em> (Å)</td>
<td>81.56, 59.48, 104.63</td>
<td>52.25, 123.23, 87.34</td>
</tr>
<tr>
<td><em>α, β, γ</em> (°)</td>
<td>90, 90.12, 90</td>
<td>90, 103.39, 90</td>
</tr>
<tr>
<td><strong>Resolution</strong> (Å)</td>
<td>52.31-2.50(2.64-2.50)*</td>
<td>61.61-2.50 (2.64-2.50)*</td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;merge&lt;/sub&gt; (%)</strong></td>
<td>12.2 (44.8)</td>
<td>9.50 (35.7)</td>
</tr>
<tr>
<td><strong>Total observations</strong></td>
<td>122920 (18058)</td>
<td>234305 (33787)</td>
</tr>
<tr>
<td><strong>Unique observations</strong></td>
<td>34948(5037)</td>
<td>33939(5397)</td>
</tr>
<tr>
<td><strong>Mean ((I)/sd (I))</strong></td>
<td>10.6 (2.4)</td>
<td>11.9 (3.8)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>99.6 (98.8)</td>
<td>91.1 (100.0)</td>
</tr>
<tr>
<td><strong>Multiplicity</strong></td>
<td>3.5(3.6)</td>
<td>6.9 (6.3)</td>
</tr>
<tr>
<td><strong>Wilson B-factor (Å&lt;sup&gt;2&lt;/sup&gt;)</strong></td>
<td>33.0</td>
<td>45.9</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Resolution</strong> (Å)</td>
<td>20.0-2.50</td>
<td>20.00-2.50</td>
</tr>
<tr>
<td><strong>No of reflections</strong></td>
<td>33186</td>
<td>32078</td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;work&lt;/sub&gt;/R&lt;sub&gt;free&lt;/sub&gt;</strong></td>
<td>0.180/0.228</td>
<td>0.212/0.250</td>
</tr>
<tr>
<td><strong>RMSD values</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bond length</strong> (Å)</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>Bond angles</strong> (°)</td>
<td>1.399</td>
<td>0.950</td>
</tr>
<tr>
<td><strong>ChirVolume</strong></td>
<td>0.080</td>
<td>0.058</td>
</tr>
<tr>
<td><strong>No of atoms</strong></td>
<td>13644</td>
<td>14374</td>
</tr>
<tr>
<td><strong>B-factor (Å&lt;sup&gt;2&lt;/sup&gt;)</strong></td>
<td>25.30</td>
<td>34.33</td>
</tr>
<tr>
<td><strong>Ramachandran values (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favoured</td>
<td>95.0 (766 residues)</td>
<td>96.8 (812 residues)</td>
</tr>
<tr>
<td>Additionally allowed</td>
<td>4.8 (38 residues)</td>
<td>3 (26 residues)</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.2 (2 residues)</td>
<td>0.2 (2 residues)</td>
</tr>
</tbody>
</table>

*Values for highest resolution shell are shown in parentheses.

*<sup>a</sup>R<sub>merge</sub> (⟨I⟩) = ∑<sub>k,l</sub> ∑<sub>i</sub> [I<sub>k,l,i</sub> - ⟨I<sub>k,l</sub>⟩] / ∑<sub>k,l</sub> ∑<sub>i</sub> |I<sub>k,l,i</sub>|, where ⟨I<sub>k,l</sub>⟩ is the average intensity of the multiple I<sub>k,l,i</sub> observed for symmetry-related reflections.

*<sup>b</sup>Mean (⟨I⟩/sd (⟨I⟩)), average of the diffraction intensities, divided by their standard deviation.

*<sup>c</sup>R<sub>work</sub> = ∑<sub>k,l</sub> [F<sub>k,l</sub> - F<sub>c,k,l</sub>] / ∑<sub>k,l</sub> |F<sub>k,l</sub>|, where F<sub>k,l</sub> and F<sub>c,k,l</sub> are observed and calculated structure factors, respectively.

*<sup>d</sup>R<sub>free</sub> = ∑<sub>k,l</sub> |F<sub>k,l</sub>| / ∑<sub>k,l</sub> |F<sub>c,k,l</sub>|, R<sub>free</sub> was calculated with 5% of the reflections not used during refinement.

*<sup>e</sup>Ramachandran plot statistics were calculated by MolProbity (Lovell et al., 2003).
Figure 5-7: MolProbity Ramachandran analysis of the RbAp48-MTA1(670–711) complex. This analysis is for the two molecules in the asymmetric unit. Of all residues 96.8% (812/838) are in favoured regions, 3% (26/838) are in additional allowed regions and 0.2% (2/838) are in outliers. Ramachandran plot statistics were calculated with MOLPROBITY, http://molprobity.biochem.duke.edu.
5.3.3 Crystal structure of the RbAp48–MTA1(670–711) complex

The structure shows that the MTA1(670–711) binds in the pocket formed by the PP loop, the N-terminal α-helix and the C-terminal α-helix (Figure 5-8A and B). The surface of RbAp48 that contacts MTA1(670–711) consists of a negatively charged surface formed mainly by residues protruding from the short helix in the PP loop, and a hydrophobic surface formed by the N-terminal helix and residues from blade 6. The two surfaces of the RbAp48 binding pocket match well with the charged surfaces and the hydrophobic moiety of MTA1, where basic and hydrophobic amino acids (TKRAARR) make one side of the helix positively charged and the other side hydrophobic (Figure 5-8C).

Specifically, a glutamate (E357) and two aspartate residues (D358 and D361) of RbAp48 that form hydrogen bonds with arginines (R679, R682 and R683) of MTA1 contribute to the negative surface charge on one side of the RbAp48 binding pocket. In contrast to these ionic interactions, the other surface of MTA1, which is formed by alanine residues (680 and 681), makes hydrophobic interactions with the binding pocket.
Figure 5-8: Crystal structure of the RbAp48–MTA1(670–711) complex.  A. Top view of the RbAp48–MTA1(670–711) structure. MTA1 binds on the side edge of RbAp48; RbAp48 is shown in red and MTA1 peptide in yellow. The seven blades of RbAp48 are labelled 1–7, starting from the N-terminal end. B. MTA1 binding pocket between N-helix, PP loop in blade 6 and C-helix. RbAp48 is indicated in red and salmon and MTA1 in yellow. C. Electrostatic potential mapped on the surface of RbAp48 (turned by 90° around the horizontal axis) indicating the binding pocket of MTA1. Acidic patches are shown in red, neutral patches in white and basic patches in blue (overall range from −85 to +85 kT).

The complex structure of RbAp48–MTA1(670–711) shows that the MTA1(670–711) peptide forms an α-helix that makes numerous intermolecular interactions. Three arginine residues of MTA1 (R679, R682 and R683) form a network of hydrogen bonds with RbAp48. Most prominently, the side chain of R682 of the MTA1 peptide sits snugly in a pocket formed in part by the two proline residues in the PP loop and forms a network of hydrogen bonds involving four main chain carbonyl groups of RbAp48.
(D358, G362, P363 and L366) and the side chain of Q354. The two other arginine residues in the MTA1 peptide, R679 and R683, also play important roles in RbAp48 binding. The side chain of R683 forms two hydrogen bonds and electrostatic interaction with the side chain of D361, whereas the side chain of R679 forms hydrogen bonds and electrostatic interactions with the side chain of D358 (Figure 5-9A). The side chain of residues T677 and Y685 of MTA1 form hydrogen bonds with N407 and G371 of RbAp48, respectively, whereas the main chain of K686 and the side chain of I688 of MTA1 form hydrogen bonds with the side chain of N27 of RbAp48 (Figure 5-9B). This extensive hydrogen bond network seems to be a key contributor to MTA1 binding to RbAp48. A680, A681 and T677 in MTA1 face a hydrophobic patch composed of F30, L31, I369, F368 and I408 from the N-terminal α-helix, the PP-loop, and the C-terminal α-helix of RbAp48, respectively. The sidechain of Y685 sits in a cleft formed by the N-terminal helix and the main body of the propeller, making van der Waals contacts with E20 and W24 (Figure 5-9B). Details of the contacts are shown in Figure 5-9 and Table 5-3.
Figure 5-9: Detailed view showing the interactions of RbAp48-MTA1(670–711). A and B. Close view of contacts between RbAp48 and MTA1 peptide showing the positively charged R679, R682 and R683 MTA1 residues with the backbone carbonyl groups in the PP loop (Q354, G362, P363 and L366) and a cluster of acidic residues (E357, D358 and D361) in RbAp48. B. Close view of contacts between RbAp48 and MTA1 peptide showing the hydrogen bonds for the other residues as indicated. RbAp48 residues are shown in red sticks and MTA1 residues are shown as yellow sticks. C. Y685 in MTA1 sits in a narrow groove on RbAp48, making van der Waals contacts with E20 and W24 of RbAp48. RbAp48 is shown in red and MTA1 in yellow.
Table 5-3: Specific interactions between RbAp48 and MTA1(670–711)

<table>
<thead>
<tr>
<th>MTA1 residue</th>
<th>RbAp48 residue</th>
<th>Interaction type and distance (Å)</th>
<th>H-bond</th>
<th>Salt bridge</th>
</tr>
</thead>
<tbody>
<tr>
<td>SER 674</td>
<td>ASP 411</td>
<td></td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>THR 677</td>
<td>ASN 407</td>
<td></td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>LYS 678</td>
<td>ASN 407</td>
<td></td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>ARG 679</td>
<td>ASP 358</td>
<td></td>
<td>3.4</td>
<td>3.3</td>
</tr>
<tr>
<td>ARG 679</td>
<td>ASP 358</td>
<td></td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>ALA 681</td>
<td>ILE 369</td>
<td></td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>ARG 682</td>
<td>LEU 366</td>
<td></td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>ARG 682</td>
<td>PRO 363</td>
<td></td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>ARG 682</td>
<td>ASP 358</td>
<td></td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>ARG 682</td>
<td>GLN 354</td>
<td></td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>ARG 682</td>
<td>GLY 362</td>
<td></td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>ARG 683</td>
<td>ASP 361</td>
<td></td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>ARG 683</td>
<td>GLU 357</td>
<td></td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>ARG 683</td>
<td>ASP 361</td>
<td></td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>TYR 685</td>
<td>HIS 373</td>
<td></td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>TYR 685</td>
<td>GLY 371</td>
<td></td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>LYS 686</td>
<td>ASN 27</td>
<td></td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>ILE 688</td>
<td>ASN 27</td>
<td></td>
<td>3.0</td>
<td></td>
</tr>
</tbody>
</table>

5.4 Summary

5.4.1 Comparison of the crystal structures of RbAp48-MTA1(656–686) and RbAp48-MTA1(670–711)

The overall structure of RbAp48 in each co-crystal structure is topologically similar to that of previously reported structures: both our structure with MTA1(656–686) (described in Chapter 4) or the one of RbAp48 bound to a FOG1 peptide (Lejon et al., 2011). The RMSD value for RbAp48-MTA1(656–686) with either RbAp48-MTA1(670–711) or RbAp48-FOG1(1–15) is 0.25 Å. This low RMSD value indicates that the backbones of the three complexes are very similar to each other.
Superposition of the structures of both MTA1 peptides (656–686) and (670–711) in complex with RbAp48 demonstrates that the structures of the complexes are very similar (Figure 5-10C), with an RMSD of 0.16 Å over 379 Cα atoms. In the RbAp48-MTA1(656–686) crystal structure, continuous electron density is observed for 13 residues of MTA1, 674–686 (SSETKRAARRPYK), whereas in the RbAp48-MTA1(670–711) structure we observe a longer stretch of 20 amino acids, 671–690 (LLSSSETKRAARRPYKPIAL). The MTA1 peptide adopts an α-helical conformation that is similar in the two structures, except that somewhat more residues are ordered in the RbAp48-MTA1(670–711) structure. It is notable from the structure that the MTA1 peptide (670–711) forms a longer α-helix (by two residues) towards the N-terminus (Figure 5-10A). The binding mode is essentially the same but there are some additional interactions in both directions, especially at the C-terminal end of the 670–711 structure, including a rearrangement of some side chains compared to the 656–686 structure (Figure 5-11). The binding of R679, R682 and R683 of MTA1 is similar between the two complexes. In the RbAp48–MTA1(670–711) structure, the orientation of the side chain of K678 of MTA1 is different from that observed in the RbAp48–MTA1(656–686) structure. In RbAp48–MTA1(670–711) structure, the side chain of K678 points toward Q354 in the PP loop, whereas the same residue in RbAp48–MTA1(656–686) points toward the C-helix, forming contacts with N407 and N410 of RbAp48. The two serines (S674/675) that make the longer helix of the MTA1(670–711) structure might be responsible for the change in the K678 orientation. S674 of MTA1(670–711) forms a hydrogen bond with D411, which we do not observe in the MTA1(656–686) structure. In the MTA1(670–711) structure, the side chain of Y685 forms hydrogen bonds with the main chain of both G371 and H373. I688 also forms a hydrogen bond with N27 in the N-helix of RbAp48, and this residue is not in the MTA1(656–686) structure (Figure 5–11). As can be seen in Figure 5-10A and B, MTA1(670–711) forms a slightly different conformation of the 656–686 peptide that allows the additional C-terminal residues to contribute to the binding.
Figure 5-10: Comparison of RbAp48–MTA1(656–686) and RbAp48–(670–711) structures.  
A. Comparison of MTA1 peptides (656–686) and (670–711), indicating the residues that form a helix in each case. MTA1(656–686) is shown in magenta and MTA1(670–711) in yellow.  
B. Ribbon diagram showing an overlay over all heavy atoms of RbAp48–MTA1(656–686) and RbAp48–MTA1(670–711). RbAp48 is shown in gold and red and the MTA1 peptides are shown in magenta and yellow, respectively.  
C. Surface/cartoon view of RbAp48–MTA1(670–711), indicating the structural elements of the RbAp48 structure that are contacted by MTA1. The structure is rotated 90° in the indicated direction compared to B.
Chapter 5

Figure 5-11: Binding sites of RbAp48-MTA1(656–686) and RbAp48-MTA1(670–711) structures. A. Detail of the RbAp48-MTA1(656–686) structure, showing interactions made by the basic sidechains of MTA1. B. Detail of the RbAp48-MTA1(670–711) structure, showing interactions made by the basic sidechains of MTA1. All residue numbers in the range 670–711 refer to MTA1 and numbers up to 411 refer to RbAp48. C. An overlay of the structures in A and B, shown in wall-eyed stereo view.

The affinity of MTA1 peptides 670–695 and 670–711 for RbAp48 is approximately 40- and 10-fold higher, respectively (K_D=0.05 µM and 0.23 µM) than that of the MTA1(656–686) peptide (K_D=2.2 µM). The structure of RbAp48-MTA1(670–711) provides a rationale for the higher affinity observed for the MTA1(670–711) peptide compared to MTA1(656–686). As shown in Figure 4-12, I688 and L690 form a hydrophobic cluster with L672, and this cluster, which cannot be formed within the MTA1(656–686) peptide, abuts the aromatic sidechain of F30 in RbAp48, presumably providing substantial additional binding affinity. However, it is not clear why MTA1(670–711) showed a weaker affinity than MTA1(670–695). It is possible that the extra residues in the MTA1(670–711) peptide are unstructured and interfere with the binding by blocking the interacting site on RbAp48.

The data from this and previous Chapters are discussed together and in more detail in Chapter 6 (General Discussion).
Figure 5-12: Hydrophobic interaction formed by I688, L690 and L672 from MTA1 (yellow) and F30 from RbAp48 (red) in the RbAp48-MTA1(670–711) structure.
Chapter 6: Discussion

6.1 Chromatin remodelling complexes and WD40 proteins

Chromatin remodelling complexes such as SWI/SNF, ISWI, CHD (e.g. NuRD) and INO80, share a conserved ATPase domain (Clapier and Cairns, 2009) that utilises the energy of ATP to modify histone-DNA contacts and regulate transcription. WD40 proteins are common components of many of these chromatin-modifying complexes, including NuRD, Sin3, NURF, HAT1, CAF-1 and PRC2 (Trievel and Shilatifard, 2009). It has been suggested that some of these complexes play overlapping roles during initiation of transcription (Barbaric et al., 2007). NURF shares both ATPase and the RbAp46/48 WD40 domains with NuRD, whereas Sin3 contains HDAC1/2 and RbAp46/48. However, each complex has unique proteins that adapt it for its specific biological function (Bao and Shen, 2007, Clapier and Cairns, 2009). For example, NuRD is unique in having both ATPase and histone deacetylase domains.

There is a general consensus that the WD40 proteins found in these complexes act in some way as scaffold proteins, given their lack of known enzymatic activity and the observation that they often mediate a range of protein-protein interactions. WD40 proteins contain short ~40 amino acid motifs, often terminating in a Trp-Asp (W-D) dipeptide. Proteins have been found with 4–16 repeated units, and structures have been determined for both 7- and 8-repeat proteins. These proteins form a beta-propeller fold in which each repeat constitutes a single blade (Xu and Min, 2011). WD proteins are a large family, with the majority found in higher eukaryotes, although a few examples have been reported in bacteria (Stirnimann et al., 2010). At least 277 proteins of this type are detected in the human genome (Lander et al., 2001). Many of these proteins are highly conserved from animals to plants. For example, human RbAp46/48 share 92% identity with Nurf55 in fly and 68% identity with MSI1 in Arabidopsis thaliana.

In brief, WD40 domain proteins are able to interact with a variety of proteins, peptides and nucleic acids via diverse surfaces and modes of interactions and are therefore involved in diverse cellular processes (Xu and Min, 2011, Stirnimann et al., 2010). WD40 proteins are implicated in a variety of functions including signal transduction, transcription regulation, cell cycle control and apoptosis (Stirnimann et al., 2010, Xu...
and Min, 2011). WD40 repeat proteins play an essential role in chromatin function and gene regulation (Trievel and Shilatifard, 2009).

6.2 WD40-histone interactions

Several WD40 proteins have been shown to function in the recognition of histone proteins. In all of the cases reported, the interaction is mediated by N-terminal histone tails, which are poorly structured and accessible for binding in the context of the H3-H4 dimer and the nucleosome. Indeed, the recognition of covalently modified isoforms of these histone tails by a host of proteins is thought to be an important aspect of the mechanisms by which genes are switched on and off (Yap and Zhou, 2010).

WDR5 and EED are seven-bladed WD40 proteins that bind histone tails at one end of the β-propeller, which has been defined as the top surface; these two proteins have been studied in some detail. WDR5 is a transcriptional regulator that forms part of the mixed-lineage leukaemia protein (MLL) histone H3 Lys4 (H3K4) methyltransferase complex (Trievel and Shilatifard, 2009). It has been suggested that WDR5 associates with dimethylated H3K4 to catalyse Lys4 trimethylation (Wysocka et al., 2005). In the structure of the WDR5-H3K4me2 complex, the H3 peptide binds in an extended conformation to the top of WDR5 (Figure 6-1A) (Couture et al., 2006, Odho et al., 2010).

EED is a subunit of the Polycomb Repressive Complex PRC2, which consists of four core subunits: Suppressor of zeste 12 (Su(z)12, SUZ12 in mammals), Nurf55 (p55, RbAp46 and RbAp48 in mammals), Enhancer of zeste (E(z), EZH2 in mammals) and Extra-sex combs (ESC, EED in mammals) (Schuettengruber et al., 2007, Wu et al., 2009). E(z) is the catalytic subunit of the PRC2 complex, whereas EED and Nurf55 are WD40 proteins. EED also binds the H3 tail at the top surface of the WD40 propeller (Figure 6-1B) (Margueron et al., 2009) but recognises H3K27 rather than H3K4. The crystal structure of the EED-H3K27me3 complex indicates that the tri-methylated lysine 27 of H3 binds a hydrophobic cage comprising two tyrosines and a phenylalanine on the top surface of EED (Margueron et al., 2009). The binding of EED to H3K27me3 stimulates the activity of the PRC2 complex (Margueron et al., 2009).

PRC2 also contacts nucleosomes via p55. This interaction is mediated by H3 and H4 tails (Furuyama et al., 2006, Nowak et al., 2011) and it has been shown that p55 is essential for high-affinity binding of the PRC2 complex to nucleosomes (Tie et al.,
Recently, the crystal structure of the N-terminal tail of H3 bound to p55 has been determined (Figure 6-1C) (Schmitges et al., 2011). The key residues in Drosophila p55 that are involved in the interaction with H3 are identical in both human RbAp proteins (Figure 6-2), and it is therefore likely that these interactions are preserved in mammalian PRC2. The H3-binding site on RbAp and p55 proteins is analogous to the top surface of both WDR5 and EED.

However, RbAp46/48 and p55 utilise a different mode from that of EED for binding to histones. H4 binds to a groove on one flank of the RbAp46 barrel (Figure 6-1D) (Murzina et al., 2008). Similarly, an independent study has shown the X-ray crystal structure of the histone H4 tail bound to the same surface on p55 (Figure 6-1E) (Song et al., 2008, Nowak et al., 2011). While MTA1 and the H4 tail bind this same side pocket of the WD40 repeat β-propeller in RbAp and p55 proteins, the H1K26me3, H3K27me3 and H4K20me3 peptides bind overlapping sites at the top surface of EED (Figure 6-1B) (Margueron et al., 2009). These modified histone peptides form similar structures and employ the aromatic cage of EED to recognise the trimethyl-lysine residue, whereas RbAp and p55 proteins use different residues for binding H3. RbAp46/48 and p55 therefore use two independent binding sites for interactions with H3 and H4.

The H4-binding pocket of RbAp and p55 proteins, which is made up of N- and C-terminal α-helices, is unique among the WD40 proteins and is totally absent in WDR5/EED proteins (Figure 6-1). These different modes of interaction of WD40-repeat proteins with nucleosomes via H3 and H4 tails might reflect their versatile roles in different chromatin-modifying complexes.

Thus, in nucleosome remodelling complexes, including NuRD, RbAp46/48 and p55 appear to function as histone recognition subunits that associate with H3 and H4. It is interesting to note that, despite the conserved nature of the β-propeller fold, EED and RbAp48 recognise histone H4 via completely different mechanisms.
Figure 6-1: WD40-histone interactions. A. Structure of the WDR5 (gold) in complex with H3K4me2 (orange) (PDB code: 2XL3) (Odho et al., 2010). B. Crystal structure of EED (gold) bound to H1K26me3 (red) (PDB code: 3IIY), H3K27me3 (orange) (PDB code: 3IIW) and H4K20me3 (green) (PDB code: 3IJ1) (Margueron et al., 2009). C. The structure of Nurf55 (p55, gold) bound to histone H3 peptide (orange) (PDB: 2YBA (Schmitges et al., 2011)). D. The structure of RbAp46 bound to histone H4 (cyan) (PDB: 3CFS (Murzina et al., 2008)). E. The structure of p55 bound to histone H4 (blue) (PDB: 2XYI (Nowak et al., 2011)).
Figure 6-2: Sequence alignment of several WD40 proteins (RbAp46/48 from human, mouse and frog and Nurf55 from fly). The secondary structure is indicated by arrows for β strands and coils for α-helices. Residues are numbered at the right of the sequence rows. Green squares, orange triangles, black triangles, red circles and blue stars indicate the key residues involved in binding MTA1, H4, Su(z)12 (Uniprot: Q9NJG9), FOG1 and H3, respectively. Uniprot codes are as follows: RbAp48 mouse, Q60972, and frog, O93377; RbAp46 mouse, Q60973, and frog, Q8AVH1; and Nurf55, Q24572.

6.3 WD40 repeat proteins exhibit different binding modes with other partners

WD40 proteins are not restricted to binding histones; they can also bind other proteins. Crystal structures show that although WD40 proteins form a similar β-propeller fold, they use various binding modes to interact with their ligand partners.

EED has also been shown to interact with the enhancer of zeste homologue 2 (EZH2), which is another subunit in the PRC2 complex, on the opposite face of the EED β-propeller from that shown to interact with histone H3 (Han et al., 2007). The crystal structure of the EED-EZH2 complex shows that EZH2 forms an α-helix that fits into a distinctive groove at the bottom surface of EED (Figure 6-3A) (Han et al., 2007). In
addition to stabilising the PRC2 complex, the interaction of EED and EZH2 has been 
demonstrated to be necessary for the histone methyl transferase activity of PRC2 
(Czermin et al., 2002). Thus, although the top surface of the WD40 repeat domain 
seems a major site for binding of interacting proteins, RBBP5 and EZH2 use the 
opposite side for binding WDR5 and EED, respectively.

Another example of a WD40 protein with a seven-bladed β-propeller is the partner and 
localizer of BRCA2 (PALB2). PALB2 recruits the breast cancer 2 protein (BRCA2) and 
the recombinase RAD51 to the site of DNA damage and promotes homologous 
recombination repair, and functions as molecular platform for the formation of the 
PALB2-BRCA1-BRCA2 complex (Sy et al., 2009). BRCA2 adopts a short α-helix that 
binds to a pocket formed between the fourth and fifth blades on the side edge of the 
PALB2 WD40 β-propeller (Figure 6-3B) (Oliver et al., 2009). This binding mode is 
reminiscent of the binding of MTA1 and H4 to RbAp proteins, albeit without the α- 
helices of the latter proteins (Figure 6-1).

Interestingly, WDR5 interacts with the mixed-lineage leukaemia protein 1 (MLL1), 
another member of the MLL complex, through the histone H3-binding pocket on its top 
surface (Figure 6-3C) (Song and Kingston, 2008). Although it is not clear why MLL1 
and H3 use the same binding surface to associate with WDR5, it has been proposed that 
this may be a mechanism for regulating the activity of the MLL complex (Song and 
Kingston, 2008). Thus, for example, the binding of WDR5 to histone H3 and the 
consequent displacement of MLL1 might induce a conformational change in the 
complex that alters activity. In addition to binding histone H3 peptide and MLL1 on its 
top surface, WDR5 has been shown to interact with retinoblastoma-binding protein-5 
(RBBP5), which is another subunit of the MLL complex, using the opposite surface 
from that shown to interact with histone H3 or MLL1 (Figure 6-3D) (Odho et al., 2010). 
RBBP5 is a WD40 protein that is required by WDR5 for full methyltransferase activity 
(Odho et al., 2010).

As noted above, WD40 proteins use different surfaces/modes for binding their partners. 
While binding sites on opposite sides of the propeller appear to be independent of one 
another, and may serve only to hold the complex together, some binding sites overlap 
with those of histones, suggesting a mechanism for regulation of transcription.
Figure 6-3: WD40 repeat proteins exhibit different binding modes. Crystal structures of WD-repeat protein complexes. A. The structure of EED (gold) in complex with an EZH2 helix (blue, PDB code: 2QXV) (Han et al., 2007) bound to the bottom of the β-propeller. B. Crystal structure of BRCA2 (gold) bound to the side edge of PALB2 (cyan, PDB code: 3EU7) (Oliver et al., 2009). C. Structure of the WDR5 (gold) in complex with MLL1 (green, PDB code: 3EMH) (Song and Kingston, 2008) bound to the top surface. D. Structure of the WDR5 (gold) in complex with RBBP5 (red, PDB code: 2XL3) (Odho et al., 2010).

6.4 The RbAp48-MTA1 interaction is highly conserved across species

Previous studies have used GST pulldowns to map the participating domains in the interaction between MTA1/2 and RbAp46/48. As part of a study focused on investigating the mechanism through which FOG2 mediates gene repression, Roche et al. showed that the predicted GATA-type zinc finger domain of MTA1 (residues 392–
448) is sufficient to mediate an interaction with RbAp46/48 (Roche et al., 2008). More recently, Fu et al. reported that RbAp46 was able to interact with any of three non-overlapping constructs from the C-terminal half of MTA2(308–436, 434–561 and 563–668); only one of these constructs, however, contained the predicted zinc finger domain (Figure 6-4) (Fu et al., 2011). Our pulldowns show that two fragments of MTA2(210–530 and 520–668) were able to interact with RbAp46/48 (Figure 3-12), which agrees with the Fu et al data, although we did not use the same fragments. However, none of our N-terminal fragments containing the ZnF domain of MTA1 showed any interaction with RbAp48, which was unexpected given the previous studies and the high sequence conservation between MTA1 and MTA2 at the N-terminus. Thus, our data show that the ZnF domain did not bind to RbAp46/48; instead two regions of the C-terminal end of MTA1(448–542 and 670–711) are able to independently bind to RbAp48. Specifically, the RbAp48-MTA1 crystal structure indicates that the region MTA1(670–690) takes up a conformation that includes a short α-helix and a hydrophobic cluster and binds to RbAp48 in a distinct pocket formed by three helical elements in the latter protein. The reason for the apparent inconsistency between these different studies is not known. It is worth noting, however, that GST pulldowns can give rise to false positive interactions if the domain fused to GST is not correctly folded (Lee et al., 2007, Mackay et al., 2007).

![Figure 6-4: A and B. Schematic diagram shows the MTA1 and MTA2 fragments used in previous studies to map the interaction with RbAp46 and RbAp48 in GST pulldowns](image-url)
Despite the fact that MTA1 and MTA2 are somewhat divergent in their C-terminal halves (34% sequence similarity for residues 431–715 of human MTA1/2 compared to 74% similarity for residues 1–430), inspection of the amino acid sequences of MTA2/MTA3 and RbAp46 indicates that these proteins should be able to substitute for MTA1 and RbAp48, respectively, to form analogous interactions (Figure 6–2 and Figure 6–5). Further, comparison of the sequences of related proteins from diverse organisms that are predicted to contain a NuRD-like complex (Figure 6–5) suggests that this interaction is broadly conserved and is therefore likely to represent a core element of NuRD architecture. For example, Figure 6–5 shows the sequence of the *D. melanogaster* MTA1-related protein MTA1-like, which carries most of the residues required to make the analogous interaction with Nurf55 (the *D. melanogaster* protein most closely related to RbAp46/48).
Figure 6-5: Sequence alignment of MTA1, MTA2 and MTA3 from different eukaryotes. The well-defined domains are indicated. The KRAARR motif is indicated by a red box in MTA1/2 proteins and a blue box in MTA3 proteins. Uniprot accession numbers are as follows: human MTA1 (Q13330), mouse MTA1 (Q8K4B0), rat MTA1 (Q62599), frog MTA1 (F6YJW8), fly MTA1 (Q9VNF6), human MTA2 (Q94776), mouse MTA2 (Q9R190), rat MTA2 (B2GV01), frog MTA2 (Q4V7T0), hMTA3 (Q9BTC8), mouse MTA3 (Q924K8) and bovine MTA3 (A6QL72). The alignment was produced using ALINE program (Bond and Schuttelkopf, 2009).
6.5 Comparison of the RbAp48-MTA1 structure with other similar WD40 proteins

There are now a number of RbAp46/48 and p55 complexes for which crystal structures have been determined, and together these structures show two binding pockets that can mediate interactions with a variety of partners (Figure 6-6D).

Our data show clearly that MTA1 and histone H4 bind to the same surface of RbAp46/48. The amphipathic MTA1 and H4 motifs interact with the complementary hydrophobic and charged residues of RbAp46/48 and p55 in an overlapping binding site, which is located on the side of the β-propeller between the N-terminal α-helix, the PP loop and C-helix (Figure 6-6A).

The Drosophila protein p55 is a member of the NuRD and PRC2 complexes (Reddy et al., 2010, Tie et al., 2001). Interestingly, the Drosophila protein Su(z)12, which is a member of the PRC2 complex, has been found to bind to p55 in similar way to the MTA1-RbAp48 and H4-RbAp46 interactions (Figure 6-6B) (Schmitges et al., 2011). The crystal structure of p55 bound to a 12-residue peptide of Su(z)12 shows that the peptide occupies the H4/MTA1-binding site of p55, but perhaps surprisingly does not form an α-helix. The backbone instead has an irregular conformation, but nevertheless places several residues into positions that mimic the interactions made by MTA1 with RbAp48. Thus, R85 of Su(z)12 occupies the same pocket at R682 of MTA1, and F82, L87 and Y89 of Su(z)12 form a hydrophobic cluster that interacts with F34 of p55 (the equivalent of F30 in RbAp48) (Figure 6–8A and C).

Sequence alignment shows that the residues of Drosophila Su(z)12 that bind p55 are highly conserved in human, mouse and Xenopus SUZ12 proteins as well (Figure 6-7), and are probably responsible for the interaction between RbAp46/48 and SUZ12 in the PRC2 complex. It has been suggested that SUZ12 also mediates the interaction between EZH2 and nucleosomes and plays an essential role in transcriptional repression of the affected target gene via H3K27 trimethylation (Cao and Zhang, 2004).

Sequence conservation mapped on the surface of RbAp48 exposes a patch of highly conserved residues that are mostly negatively charged, on the top of the β-propeller surrounding the central cavity (Figure 5–8C). This site is usually used for binding to positively charged peptides such as FOG1 and H3 (Schmitges et al., 2011, Lejon et al.,
and the specific interactions made by the two peptides are highly conserved (Figure 6–2 and Figure 6–6C).

Overall, the key residues that are involved in the interactions with MTA1, H3, H4, FOG1 and SUZ12, and in defining the protein fold, are identical in RbAp46, RbAp48, and Drosophila p55 (Figure 6–2), and we conclude that all three of these proteins form the same interactions with their partners.

**Figure 6-6: Comparison of the RbAp48-MTA1 structure with other RbAp46/48 complex structures.** A. Comparison of interactions made by histone H4(28–42) (Murzina et al., 2008, PDB 3CFS, cyan) and MTA1(670–711) (yellow) with RbAp46 and RbAp48, respectively. B. Comparison of interactions made by Su(z)12(79–91) (Schmitges et al., 2011, PDB 3YB8, green) and MTA1(670–711) (yellow) with Nurf55 and RbAp48, respectively. C. Comparison of interactions made by H3(2–20) (Schmitges et al., 2011, PDB 3YB8, orange) and FOG-1(1–15) (Lejon et al., 2011, PDB 2XU7, blue) with Nurf55 and RbAp48, respectively. D. Overlay of RbAp46/48/Nurf55-MTA1/H4/Su(z)12/H3/FOG-1 complex structures. In all parts, RbAp48 is shown as a grey surface.
Figure 6-7: Sequence alignment of SUZ12 from different species. Alignment of the p55-binding region of Su(z)12 with sequences from Drosophila melanogaster (fly, Uniprot: Q9NJG9), human (Q15022), mouse (Q80U70) and Xenopus tropicalis (frog, Q0VA03). The conserved residues are shaded in grey. The region observed in the p55-Su(z)12 structure is underlined. The asterisks indicate the key residues involved in the p55 interaction.

6.6 MTA1 and histone H4 compete for the same binding site on RbAp48

As noted above, MTA1 and histone H4 compete for the same binding pocket on RbAp46/48. Comparison of the RbAp48-MTA1 and RbAp46-H4 structures shows that essentially all of the intermolecular interactions made by histone H4 sidechains are recapitulated by MTA1 in the complex with RbAp48 (see Section 4.5.4 and Figure 6-8A and B). The structures also suggest reasons why the affinity of MTA1 for RbAp48 is substantially higher than that of histone H4 for RbAp46 (K\textsubscript{D} of 0.05 \( \mu \text{M} \) vs 1 \( \mu \text{M} \)).

Firstly, the tyrosine Y685 in MTA1 that sits in a narrow groove on RbAp46, making van der Waals contacts with residues E20 and W24 (Figure 5–9C) that line the groove, is replaced with a glycine in histone H4 (Figure 6–8D), and secondly the residues making up the MTA1 hydrophobic cluster noted above that contacts F30 on RbAp48 are not present in histone H4 (Figure 5–12 and Figure 6–8D).

It is notable that the nature of the interaction between histone H4 and RbAp46/48 requires that the histone take up a conformation that is not consistent with its structure in the nucleosome core particle (Figure 6-9) (Murzina et al., 2008). Taken together, these data suggest several possible explanations for this. On one hand, it is possible that histone H4 binding is a function carried out by RbAp48 that is independent of its role in the NuRD complex; as noted above, RbAp46/48 is found in a number of regulatory complexes, including the Sin3 (Zhang et al., 1998c), PRC2 (Kuzmichev et al., 2002), NURF (Barak et al., 2003), HAT (Verreault et al., 1998) and CAF-1 (Verreault et al.,
complexes. On the other hand, competition between MTA1 and histone H4 might represent an important aspect of NuRD function related to its ability to remodel chromatin: for example, MTA1 binding might modulate the capacity of NuRD to destabilise nucleosome core particles, which it might achieve in part by disrupting H4 structure. MTA1 binding in turn might be regulated by post-translational modifications; databases show that MTA1 in particular is potentially subject to a substantial number of such modifications.

Figure 6-8: Comparison of the MTA1 binding pocket of RbAp48 with other similar WD40 proteins. A. Detail of the RbAp48-MTA1(670–711) structure, showing interactions made by the basic sidechains of MTA1. B. Detail of the RbAp46-H4(28–42) structure, showing interactions made by the basic sidechains of H4. C. Detail of the Nurf55-Su(z)12(79–91) structure, showing interactions made by the basic sidechains of Su(z)12. RbAp48, RbAp46 and Nurf55 are shown in red, green and gold and MTA1, H4 and S(z)12 are shown in yellow, cyan and green, respectively. D. Sequence alignment of MTA1(670–689) with the corresponding regions from MTA2 and MTA3, Drosophila MTA1-like and human histone H4. The KRAARR motif is underlined and the conserved residues with human MTA1 are shaded. The region of MTA1 that forms α helix is indicated.
Figure 6-9: Structure of a single nucleosome showing the RbAp46-binding region of H4 (PDB: 1KX5) (Davey et al., 2002). The histone octamer is shown in ribbon format, and consists of two copies of each histone protein: H4 (blue, except for RbAp46-binding region, orange), H3 (green), H2A (pink) and H2B (yellow). The DNA coiled around the histone octamer is shown in grey.

6.7 RbAp46 and RbAp48 might act as ‘docking stations’ in NuRD

As shown in Figure 6-10, WD40 repeat domains can employ various surfaces for recognising different partners, allowing simultaneous interactions with multiple binding partners. In general, WD40 proteins exploit top, bottom and side surfaces for association with target proteins, particularly in the multiple component complexes. These observations suggest that the WD40 proteins function as a binding platform that supports the formation of protein complexes and promotes their activity by facilitating interactions with other molecules. Thus, WD40 proteins are likely to be fundamental building blocks in the assembly and regulation of a range of gene regulatory complexes.
In addition to the structurally characterised interactions described above, it has been demonstrated that mammalian RbAp46/48 can interact with most NuRD components, including HDAC1 and HDAC2 (Nicolas et al., 2001, Taplick et al., 2001), MBD3 (Zhang et al., 1999), p66α/β (Gnanapragasam et al., 2011, Feng et al., 2002, Gong et al., 2006), and MTA1/2 (Fu et al., 2011, Roche et al., 2008). A similar range of interactions has been reported for the NuRD p55 protein (Marhold et al., 2004). We note (as above) that many of these interactions have been detected using approaches (e.g., GST pulldowns, co-immunoprecipitations) that leave the question open of whether direct interactions in fact exist. For example, in several cases bacterially-expressed fusion proteins of RbAp48 have been used, which are unlikely to be correctly folded, whereas in other cases direct interactions have been claimed from co-immunoprecipitation data, where no effort to distinguish direct from indirect
interactions is made (Mackay et al., 2007). Despite this caveat, it is likely that RbAp46/48 make a number of direct contacts, both within NuRD and with other partners (Figure 6-11).

These observations have lead to the suggestion that RbAp46/48, which carry no known enzymatic activity or nucleic-acid-binding capacity, might act as protein interaction hubs. One question raised by this suggestion is that of how these different interactions are integrated, given the overlap of binding sites. One possibility is that different partners are engaged in different complexes. A second possibility, noted above, is that competition between partners constitutes a part of the mechanism of action of each complex that carries RbAp46/48. A third possibility can also be considered, based on recent mass spectrometry data that probed the stoichiometry of the NuRD complex. Smits et al. concluded from label-free quantitation analysis that the molar ratio of RbAp46/48 to CHD4 is 6:1 (Smits et al., 2013). This result, if generally true, would allow the possibility that different molecules of RbAp46/48 within a single NuRD complex could take on different functions by engaging different partners. It was also notable that they measured a ratio of 2:1 for RbAp46/48 to MTA1/2/3, suggesting that the RbAp48:MTA1 interaction delineated in our current study could be present in up to three copies within NuRD. Discrimination between these different models will only be possible once structural data are available for the intact NuRD complex.

**Figure 6-11: NuRD architecture.** Schematic illustration of protein-protein interactions indicates that RbAp46 and RbAp48 might be able to interact with all other NuRD proteins.
6.8 NuRD contacts chromatin through multiple interfaces

An important aspect of delineating the biochemical mechanism(s) of action of the NuRD complex is to identify the means through which the complex contacts its substrate – namely chromatin. Gel filtration (Hong et al., 2005, Zhang et al., 1999) and mass spectrometry data (Smits et al., 2013) suggest that the NuRD complex is >~1 MDa in size, making it at least four times larger than a single nucleosome. Given that the function of NuRD appears to be, at least in part, to remodel chromatin, it is reasonable to suggest that the complex might make multiple simultaneous interactions with histones and/or DNA during the course of its activity. In addition to the interactions described above between RbAp46/48 and histones H3 and H4, a number of other contacts have been reported. The PHD domains of CHD4 have been shown to bind to the N-terminal tail of histone H3, exhibiting a preference for either acetylated or trimethylated K9 over unmodified K9 (Mansfield et al., 2011, Musselman et al., 2012), with a reduction in binding upon methylation of K4. p66α and p66β have both been shown in GST pulldowns to be able to bind to the N-terminal tails of all four core histones (Brackertz et al., 2006), and similarly, MTA1 and MTA2 can both bind to the N-terminal tail of H3 in a manner that is inhibited by methylation of H3K4 (Wu et al., 2013). It is also notable that the BAH domains of several transcriptional regulators, including ORC1b (Kuo et al., 2012) and CMT3 (Du et al., 2012), have been demonstrated to act as histone recognition modules; MTA1 contains an N-terminal BAH domain.

As well as these histone-focused interactions, several domains in NuRD harbour putative or demonstrated DNA-binding properties. MBD2 can bind selectively to methylated DNA (Scarsdale et al., 2011) and p66α/β contain GATA-type zinc finger domains. Several regions of both MTA1 and CHD4 also have sequence similarity to known DNA-binding domains, including the GATA-type zinc finger of MTA1. It has been further suggested that the chromodomains of CHD4 can interact with DNA (Bouazoune et al., 2002). It is likely that many or all of these direct interactions with DNA will be relatively non-specific and that specificity will be introduced through the involvement of gene-specific transcriptional coregulatory proteins such as FOG1 that target NuRD to particular loci.
NuRD function is influenced by interaction with histones and/or DNA and DNA binding proteins, which target NuRD to genomic regions (Figure 6-12). It is difficult to predict the order of these interactions with NuRD components, but it is likely that the interactions take place at different stages of its functions. Recent data suggest that NuRD action is not limited to gene regulation but affects multiple essential nucleosome-related activities from histone modifications to global chromatin folding (Ramirez and Hagman, 2009).

Many interactions between NuRD components and other proteins have been found (as discussed in Chapter 1). These numerous interactions reflect the diverse functions of NuRD. Currently, it is not fully understood how NuRD acts as a corepressor and coactivator (Miccio et al., 2010, Hong et al., 2005). However, PTMs of NuRD components may affect the NuRD functions: for example, HDAC1 loses its activity when it is subjected to acetylation by histone acetyl transferases (HAT), which leads to gene activation (Yang et al., 2012). Similarly, methylation of MTA1 was recently shown to be important for the NuRD repressor activity, whereas demethylated MTA1 was associated with gene activation (Nair et al., 2013). Numerous other post-translational modifications of NuRD subunits have been discovered (Henk Stunnenberg, personal communication), and it is highly likely that these modifications act as switches that modulate the binding of NuRD to partner proteins, in much the same way as has been extensively documented for histone proteins over the last decade or more.

An appreciation of the interplay of these numerous NuRD-coregulator and NuRD-chromatin interactions with the intra-complex interactions of the type described in the current work will lead us ultimately to an understanding of the biochemical mechanisms underlying the activity of NuRD and other multi-protein regulatory complexes. The work described in this thesis represents a step towards achieving this understanding.
NuRD interacts with histones and nucleosomal DNA via CHD3/4, RbAp46/48, p66α/β and MTA1/2. MBD2 binds methylated DNA. The CHD3/4 remodels chromatin using ATPase activity. HDAC1/2 oppose the action of histone acetyl transferases (CBP/P300) by deacetylating histone to maintain gene repression. FOG1/GATA1 recruits NuRD to target genes where coactivators such as HAT acetylate histones and HDAC1, resulting in gene activation. During gene repression, the histone deacetylation and ATPase remodelling activities of NuRD play the main roles in converting the chromatin into more compact structure that is inaccessible to transcription machinery.

Figure 6-12: Role of NuRD complex. NuRD interacts with histones and nucleosomal DNA via CHD3/4, RbAp46/48, p66α/β and MTA1/2. MBD2 binds methylated DNA. The CHD3/4 remodels chromatin using ATPase activity. HDAC1/2 oppose the action of histone acetyl transferases (CBP/P300) by deacetylating histone to maintain gene repression. FOG1/GATA1 recruits NuRD to target genes where coactivators such as HAT acetylate histones and HDAC1, resulting in gene activation. During gene repression, the histone deacetylation and ATPase remodelling activities of NuRD play the main roles in converting the chromatin into more compact structure that is inaccessible to transcription machinery.
Chapter 7 Materials and Methods

Sequence analyses and alignments of DNA and proteins were carried out using ClustalW (Thompson et al., 1994). Molecular diagrams were produced using PyMOL (DeLano, 2002). All buffers and solutions were made up in Milli-Q® water (MQW), and all chemicals were of analytical grade unless otherwise specified.

7.1 Materials

7.1.1 Chemicals, reagents and consumables

The following is a list of chemicals, reagents and consumables used, along with their suppliers, for the experiments described in this thesis (Table 7-1). All standard chemicals and any reagents not listed were obtained from New England Biolabs (Beverly, MA, USA), Promega (Madison, WI), Sigma (Castle Hill, NSW) or Univar (Auburn, NSW).

Table 7-1: Chemicals, reagents and consumables.

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,2-dimethyl-2-silapentane-5-sulfonate (DSS)</td>
<td>Fluka A. G. (Buchs, Switzerland)</td>
</tr>
<tr>
<td>α-cyano-4-hydroxycinnamic acid</td>
<td>Sigma (Castle Hill, NSW)</td>
</tr>
<tr>
<td>Agar</td>
<td>Amyl Media (Kings Langley, NSW)</td>
</tr>
<tr>
<td>Agarose</td>
<td>Quantum Scientific (Milton, QLD)</td>
</tr>
<tr>
<td>Ampicillin sodium salt (Amp)</td>
<td>Gold Biotechnology (St. Louis, MO, USA)</td>
</tr>
<tr>
<td>Bovine serum albumin (BSA)</td>
<td>Sigma (Castle Hill, NSW)</td>
</tr>
<tr>
<td>Casein peptone</td>
<td>Amyl Media (Kings Langley, NSW)</td>
</tr>
<tr>
<td>CelluSep™H1 1 kDa MWCO dialysis tubing</td>
<td>Membrane Filtration Products (Seguin, TX, USA)</td>
</tr>
<tr>
<td>Chloramphenicol (Cam)</td>
<td>Sigma (Castle Hill, NSW)</td>
</tr>
<tr>
<td>Complete™, EDTA-free protease inhibitor</td>
<td>Roche Applied Science (Mannheim, Germany)</td>
</tr>
<tr>
<td>Sparse matrix crystallization screens</td>
<td>QIAGEN (Doncaster, VIC)</td>
</tr>
<tr>
<td>Diethyl pyrocarbonate (DEPC)</td>
<td>Sigma (Castle Hill, NSW)</td>
</tr>
<tr>
<td>Dithiothreitol (DTT)</td>
<td>Quantum Scientific (Milton, QLD)</td>
</tr>
<tr>
<td>DNA ladders (2-log, 100 bp)</td>
<td>New England Biolabs (Beverly, MA, USA)</td>
</tr>
<tr>
<td>DNA oligonucleotides</td>
<td>Sigma Genosys (Castle Hill, NSW)</td>
</tr>
<tr>
<td>DNase I</td>
<td>Roche Applied Science (Mannheim,</td>
</tr>
<tr>
<td>Chemical/Protocol</td>
<td>Supplier</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>dNTPs</td>
<td>New England Biolabs (Beverly, MA, USA)</td>
</tr>
<tr>
<td>Ethidium bromide</td>
<td>Bio-Rad (Regents Park, NSW)</td>
</tr>
<tr>
<td>Fugene Transfection Reagent</td>
<td>Promega (Annandale, NSW)</td>
</tr>
<tr>
<td>Glutathione-Sepharose® 4B beads</td>
<td>Amersham Biosciences (Castle Hill, NSW)</td>
</tr>
<tr>
<td>Insect GeneJuice Transfection Reagent</td>
<td>Novagen, USA</td>
</tr>
<tr>
<td>Isopropyl β-D-thiogalactopyranoside (IPTG)</td>
<td>Quantum Scientific (Milton, QLD)</td>
</tr>
<tr>
<td>Lysozyme from chicken egg white</td>
<td>Sigma (Castle Hill, NSW)</td>
</tr>
<tr>
<td>Mark12™ protein standards</td>
<td>Invitrogen (Mt Waverley, VIC)</td>
</tr>
<tr>
<td>Mini quick spin™ oligo columns</td>
<td>Roche Applied Science (Mannheim, Germany)</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>Sigma (Castle Hill, NSW)</td>
</tr>
<tr>
<td>NuPAGE® 4x LDS sample buffer</td>
<td>Invitrogen (Mt Waverley, VIC)</td>
</tr>
<tr>
<td>Phenylmethylsulfonyl fluoride (PMSF)</td>
<td>Sigma (Castle Hill, NSW)</td>
</tr>
<tr>
<td>Precast NuPAGE® 4–12% Bis-Tris mini gels</td>
<td>Invitrogen (Mt Waverley, VIC)</td>
</tr>
<tr>
<td>QIAprep® spin miniprep kit</td>
<td>QIAGEN (Doncaster, VIC)</td>
</tr>
<tr>
<td>QIAquick® gel extraction kit</td>
<td>QIAGEN (Doncaster, VIC)</td>
</tr>
<tr>
<td>Restriction enzymes</td>
<td>New England Biolabs (Beverly, MA, USA)</td>
</tr>
<tr>
<td>RNasin® Plus RNase inhibitor</td>
<td>Promega (Annandale, NSW)</td>
</tr>
<tr>
<td>RNA oligonucleotides</td>
<td>Dharmacon, Inc. (Lafayette, CO, USA)</td>
</tr>
<tr>
<td>SnakeSkin™ 3.5 kDa MWCO dialysis tubing</td>
<td>Pierce (Rockford, IL, USA)</td>
</tr>
<tr>
<td>T4 DNA ligase</td>
<td>Fermentas (Ontario, Canada)</td>
</tr>
<tr>
<td>Taq DNA polymerase</td>
<td>New England Biolabs (Beverly, MA, USA)</td>
</tr>
<tr>
<td>Thrombin, from bovine plasma</td>
<td>Sigma (Castle Hill, NSW)</td>
</tr>
<tr>
<td>Trifluoroacetic acid</td>
<td>Sigma (Castle Hill, NSW)</td>
</tr>
<tr>
<td>Triton X-100</td>
<td>Progen (Darra, QLD)</td>
</tr>
<tr>
<td>Vivaspin 6 and 500 sample concentrators</td>
<td>GE Healthcare (Piscataway, NJ, USA)</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>Amyl Media (Kings Langley, NSW)</td>
</tr>
</tbody>
</table>

### 7.1.2 Bacterial strains

- *E. coli* DH5α (genotype: supE44, ΔlacU169, [Φ80lacZAM15], hsdR17, recA1, endA1, gyrA1, thi-1, relA1) was used to produce plasmid DNA.
- *E. coli* BL21 (genotype: F-ompT [lon] hsdSB (rB-mB-: an *E. coli* B strain)), which carries the DE3 λ prophage and contains the T7 RNA polymerase gene (Integrated Sciences, Willoughby, NSW), was used for protein expression.
- *E. coli* Rosetta 2 (genotype: F-ompT hsdSB (rB-mB-) gal dcm) contains the pRARE2 plasmid (CamR) and was used for expression of proteins with *E. coli* rare codons.

- *E. coli* DH10Bac (genotype: mcrA (mrr-hsdRMS-mcrBC) 80lacZM15 lacX74 recA1 endA1 araD139 (ara, leu) 7697 galU galK–rpsL nupG/pMON14272/pMON7124) (Invitrogen, Carlsbad, CA, USA) carries the baculovirus shutter vector (bacmid) and was used for transposition of proteins into baculovirus.

### 7.1.3 Bacterial culture media and plates

All media were made up in Milli-Q® water (MQW) and sterilized by autoclaving prior to use. Antibiotics were added after autoclaving.

- Luria-Bertani (LB) media: 1% (w/v) casein peptone pancreatic digest, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl
- LB-Amp: LB containing 100 µg/mL Amp, used to grow up DH5α and BL21 cells.
- LB-Amp-Cam: LB containing 50 µg/mL Amp and 34 µg/mL Cam, used to culture Rosetta2 cells.
- LB-Amp-Kan: LB containing 100 µg/mL Amp and 50 µg/mL Kan.
- LB agar plates: LB media containing 1.5% (w/v) bacteriological agar.
- LB agar-Amp: LB containing 100 µg/mL Amp, used to culture BL21 cells.
- LB agar-Amp-Cam: LB containing 50 µg/mL Amp and 34 µg/mL Cam, used to culture Rosetta2 cells.
- SOC: 2% (w/v) casein peptone pancreatic digest, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5 mM KCl, 10 mM MgCl₂, 20 mM glucose, pH 7.0.

### 7.1.4 Buffers and solution

All buffers were made in MQW. The following is a list of buffers used for large scale purifications of MBP-MTA1 BAH(1–168) and ELM2-SANT(161–340) and GST-MTA1(448–542), (465–516), (465–542), (625–715) and (656–686) (Table 7–2). The buffers used to increase the protein solubility of MTA1/2 constructs are listed in Table 7–3.
Table 7-2: Buffers used for purification of MBP-MTA1 and GST-MTA1 fragments

<table>
<thead>
<tr>
<th>For MBP-constructs</th>
<th>1-Lysis buffer:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• 20 mM Tris, pH 7.5, 150 mM NaCl, 1 mM DTT, 1 mM PMSF</td>
</tr>
<tr>
<td></td>
<td>• 50 mM Tris, pH 7.5, 500 mM NaCl, 1 mM DTT, 1 mM EDTA, 1 mM PMSF</td>
</tr>
<tr>
<td>Wash buffer:</td>
<td>• 20 mM Tris, pH 7.5, 800 mM NaCl, 10% glycerol, 1 mM DTT, then wash in 50 mM Tris, pH 7.5, 100 mM NaCl, 1 mM DTT (Xa buffer)</td>
</tr>
<tr>
<td>Elution buffer:</td>
<td>• 20 mM Tris, pH 7.5, 150 mM NaCl, 60 mM Maltose</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>For MBP-constructs</th>
<th>2-Lysis buffer:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• 50 mM HEPES, pH 7.5, 500 mM NaCl, 10% glycerol, 1 mM DTT, 1 mM PMSF</td>
</tr>
<tr>
<td>Wash buffer:</td>
<td>• 50 mM HEPES, pH 7.5, 800 mM NaCl, 10% glycerol, 1 mM DTT, then wash in 50 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM DTT</td>
</tr>
<tr>
<td>Elution buffer:</td>
<td>• 20 mM HEPES, pH 7.5, 150 mM NaCl, 60 mM Maltose</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>For GST-constructs</th>
<th>1-Lysis buffer:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• 20 mM Tris, pH 7.5, 150–300 mM NaCl, 1 mM DTT, 1 mM PMSF</td>
</tr>
<tr>
<td></td>
<td>• 50 mM Tris, pH 8.0, 500 mM NaCl, 1 mM DTT, 1 mM EDTA, 1 mM PMSF</td>
</tr>
<tr>
<td>Wash buffer:</td>
<td>• 20 mM Tris, pH 7.5, 500–800 mM NaCl, 10% glycerol, 1 mM DTT</td>
</tr>
<tr>
<td>Elution buffer:</td>
<td>• 20 mM Tris, pH 7.5, 150 mM NaCl, 50 mM GSH (reduced)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>For GST-constructs</th>
<th>2-Lysis buffer:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• 50 mM HEPES, pH 7.5, 500 mM NaCl, 10% glycerol, 1 mM DTT, 1 mM PMSF</td>
</tr>
<tr>
<td>Wash buffer:</td>
<td>• 50 mM HEPES, pH 7.5, 800 mM NaCl, 10% glycerol, 1 mM DTT</td>
</tr>
<tr>
<td>Elution buffer:</td>
<td>• 20 mM HEPES, pH 7.5, 150 mM NaCl, 50 mM GSH (reduced)</td>
</tr>
</tbody>
</table>
Table 7-3: Lysis buffers trialled to increase the solubility of some fragments of MTA1/2

- 100mM acetate, pH 5.5, 150 mM NaCl, 1 mM DTT, 0.01% Tween 20, 1 mM PMSF, cells were lysed by lysozyme
- 100mM acetate, pH 5.5, 500 mM NaCl, 1 mM DTT, 0.01% Tween 20, 1 mM PMSF, cells lysed by sonication
- 50mM MES, pH 6.0, 300 mM NaCl, 1 mM DTT, 0.01% Tween 20, 1 mM PMSF, cells lysed by sonication
- 50mM MES, pH 6.0, 500 mM NaCl, 1 mM DTT, 1 mM PMSF, cells lysed by lysozyme
- 50 mM HEPES, pH 7.0, 500 mM NaCl, 1 mM DTT, 0.01% Tween 20, 1 mM PMSF, cells lysed by sonication
- 50 mM HEPES, pH 7.0, 1000 mM NaCl, 1 mM DTT, 1 mM PMSF, cells lysed by lysozyme.
- 50 mM Tris/HCl, pH 8.0, 500 mM NaCl, 1 mM DTT, 0.01% Tween 20, 1 mM PMSF, cells lysed by lysozyme.
- 50 mM Tris/HCl, pH 9.0, 300 mM NaCl, 1 mM DTT, 0.01% Triton X-100, 1 mM PMSF, cells lysed by lysozyme.
- 50 mM Tris/HCl, pH 9.0, 500 mM NaCl, 1 mM DTT, 0.01% Tween 20, 1 mM PMSF, cells lysed by sonication

7.2 Cloning, mutagenesis and plasmids

7.2.1 Plasmids

7.2.1.1 Plasmids for in vitro transcription/translation
The T7 RNA polymerase-based expression vectors, pcDNA3 and pMW172, were used for in vitro translation (IVT) of RbAp46/48 and MTA1/2 constructs. RbAp46/48 and MTA1/2 genes were used as template for PCR amplification. All constructs used for the in vitro transcription/translation system are listed in Table 7–4.

7.2.1.2 Plasmids for protein overexpression in E. coli
The vectors pGEX-6P and pGEX-4T (with an N-terminal GST-Tag followed by a HRV 3C or thrombin cleavage site, respectively), pMAL (with an N-terminal MBP-Tag), pHUE (with an N-terminal His-Ubiquitin-Tag followed by a UBP41 cleavage site),
pET28a (with a C-terminal His-Tag) were used for overexpression of the MTA1/2 wild-type and mutant constructs. A commercially codon-optimized DNA sequence for bacterial expression of MTA1(448–715) was purchased from GenScript (USA). The MTA1 constructs derived from (448–715) were PCR amplified using this optimized sequence as a template, and then ligated between BamHI and EcoRI sites of the pGEX-6P or pGEX-4T vector. All constructs that were used for protein expression in *E coli* are listed in Table 7-5.

**Table 7-4: Constructs used for IVT expression**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Size (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcDNA3-MTA1 FL</td>
<td>80.7</td>
</tr>
<tr>
<td>pMW172-MTA1(1–250)</td>
<td>28.4</td>
</tr>
<tr>
<td>pMW172-MTA1(230–550)</td>
<td>37.3</td>
</tr>
<tr>
<td>pMW172-MTA1(530–715)</td>
<td>20.7</td>
</tr>
<tr>
<td>pcDNA3-MTA1(1–550)</td>
<td>62.5</td>
</tr>
<tr>
<td>pcDNA3-MTA1(449–715)</td>
<td>29.8</td>
</tr>
<tr>
<td>pcDNA3-MTA1(512–715)</td>
<td>22.6</td>
</tr>
<tr>
<td>pcDNA3-MTA1(1–511)</td>
<td>58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct</th>
<th>Size (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcDNA3-MTA1(530–635)</td>
<td>11.6</td>
</tr>
<tr>
<td>pcDNA3-MTA1(625–715)</td>
<td>10.4</td>
</tr>
<tr>
<td>pcDNA3-MTA2 FL</td>
<td>75</td>
</tr>
<tr>
<td>pcDNA3-MTA2(1–250)</td>
<td>28.5</td>
</tr>
<tr>
<td>pcDNA3-MTA2(220–530)</td>
<td>36</td>
</tr>
<tr>
<td>pcDNA3-MTA2(520–668)</td>
<td>16</td>
</tr>
<tr>
<td>pcDNA3-FLAG-RbAp48</td>
<td>48.7</td>
</tr>
<tr>
<td>pcDNA3-FLAG-RbAp46</td>
<td>49.8</td>
</tr>
</tbody>
</table>
Table 7-5: Constructs used for overexpression in *E. coli*

<table>
<thead>
<tr>
<th>Construct</th>
<th>Size (kDa)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEX-6P-MTA1(1–437)</td>
<td>49.7</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(87–437)</td>
<td>40.1</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(160–437)</td>
<td>31.4</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(1–340)</td>
<td>39.4</td>
</tr>
<tr>
<td>pMAL-MTA1(1–340)</td>
<td>39.3</td>
</tr>
<tr>
<td>pHUE-MTA1(1–340)</td>
<td>39.3</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(1–284)</td>
<td>32.5</td>
</tr>
<tr>
<td>pMAL-MTA1(1–284)</td>
<td>32.5</td>
</tr>
<tr>
<td>pHUE-MTA1(1–284)</td>
<td>32.5</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(1–168)</td>
<td>19.7</td>
</tr>
<tr>
<td>pMAL-MTA1(1–168)</td>
<td>19.7</td>
</tr>
<tr>
<td>pHUE-MTA1(1–168)</td>
<td>19.5</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(161–340)</td>
<td>20.9</td>
</tr>
<tr>
<td>pMAL-MTA1(161–340)</td>
<td>20.9</td>
</tr>
<tr>
<td>pMAL-MTA1(161–340)</td>
<td>20.9</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(161–276)</td>
<td>13.2</td>
</tr>
<tr>
<td>pMAL-MTA1(161–276)</td>
<td>13.2</td>
</tr>
<tr>
<td>GST</td>
<td>26.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct</th>
<th>Size (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEX-6P-MTA1(625–715)</td>
<td>10.4</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(530–635)</td>
<td>11</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(541–630)</td>
<td>10.2</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(656–686)</td>
<td>3.7</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(643–695)</td>
<td>6.8</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(448–542)</td>
<td>11.1</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(448–478)</td>
<td>3.4</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(477–501)</td>
<td>3.1</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(501–542)</td>
<td>4.6</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(448–516)</td>
<td>8</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(448–526)</td>
<td>9</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(477–542)</td>
<td>7.8</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(465–516)</td>
<td>6.2</td>
</tr>
<tr>
<td>pGEX-6P-MTA1(465–542)</td>
<td>9.2</td>
</tr>
<tr>
<td>pGEX-2T-H4(1–48)</td>
<td>5</td>
</tr>
<tr>
<td>MBP</td>
<td>45</td>
</tr>
<tr>
<td>Ub-His</td>
<td>10.5</td>
</tr>
</tbody>
</table>

*Molecular weight indicate the size of MTA1 constructs alone.

### 7.2.1.3 Plasmids for transposition of the RbAp48 gene and MTA1(448–715) construct into baculovirus bacmid

A commercially codon-optimized DNA sequence for insect cell expression of MTA1(448–715) was purchased from GenScript (USA), then cloned into the pFBDM vector. Full-length RbAp48 cloned into the pFBDM vector was provided by the Laue Laboratory (University of Cambridge, UK).
7.2.1.4 Plasmids for expression in Sf9 cells using InsectDirect System

pIEX vectors were used to express MTA1(9–154, 1–287, 155–334, 155–354, 1–334 and 448–715) using InsectDirect expression system in Sf9 cells. The plasmid was provided by the Bardwell laboratory (University of Minnesota, USA).

7.2.2 Polymerase Chain Reaction

Amplification of DNA by PCR was performed in a 50 μL reaction volume containing 1× Pfu buffer (20 mM Tris, pH 8.8, 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1% Triton X-100, 0.1 mg/mL BSA), 2.5% DMSO, 0.1 mM each of dATP, dCTP, dGTP and dTTP, 0.4 μM each primer, 200 ng/μL template plasmid DNA and 2 U Pfu DNA polymerase. The reaction consisted of 28 cycles of denaturation at 94 °C for 30 s, annealing at 40–60 °C for 45 s, and extension at 72 °C for 45 s, followed by a final extension step at 72 °C for 10 min. Wild-type MTA1/2 and RbA p46/48 genes cloned in the pcDNA3 vectors were used as the template. Site-directed mutagenesis was also performed to introduce double mutations to the MTA1 KRAARR motif using one-step site-directed mutagenesis (Liu and Naismith, 2008) with overlapping primers encoding the region of mutation. Upon completion, 1 µL of DpnI was added to the reactions, and incubated at 37 °C for 1 h for digestion of the wild-type template. A 3 µL aliquot of the DpnI-digested reaction mixture was subsequently transformed into E. coli DH5α cells. Completed reactions were stored at −20 °C until further use. Primers used for PCR are listed in Appendix A. All PCR reactions were carried out in an Eppendorf Mastercycler (Eppendorf AG, Hamburg, GER) using T3000 Thermocycler (Biometra, Goettingen, Germany).

7.2.3 Restriction enzyme digestion of double-stranded DNA

In order to subclone PCR products into expression vectors, 1 µL of the appropriate restriction enzymes were added to the purified DNA, along with the recommended buffer and BSA (0.1 mg/mL) in a reaction volume of 30 μL. For plasmid digestion, 15 µL of vector was used in a 30 μL reaction mixture containing 1 µL of each restriction enzyme and the recommended buffer and BSA. Restriction digests were carried out at 37 °C for 2 h. One unit of shrimp alkaline phosphatase was added to the vector digestion reaction and incubated at 37 °C for a further 30 min. Digested products were then gel purified as described in Section 7.2.4 below.
7.2.4 Agarose gel and DNA purification

A 1–2% agarose gel was prepared in 1× TAE buffer (40 mM Tris, pH 8.0, 1 mM EDTA, 0.1% acetic acid, 2 μg/mL ethidium bromide). The gel was allowed to set in a horizontal slab mould with well comb, before immersion in 1×TAE. DNA products were mixed with 5× loading dye (10 mM Tris-HCl, pH 7.6, 0.03% (w/v) bromophenol blue, 0.03% (w/v) xylene cyanel FF, 60% (v/v) glycerol, 60 mM EDTA) before loading into wells, and samples were electrophoresed at 110 V for 30–40 min. A DNA ladder (1 kb) was used as the size marker. DNA bands were visualized under UV light and the bands of interest were excised. DNA was purified from the gel fragments using a QIAquick® Gel Extraction Kit (Qiagen, Valencia, CA, USA) according to manufacturer’s instructions. The DNA was eluted in 30 μL of either the provided elution buffer (10 mM Tris-HCl, pH 8.5) or MQW and stored at −20 °C until required.

7.2.5 Ligation

Ligations were performed in a 10 µL reaction volume, comprising 2 µL purified digested plasmid, 6 µL purified digested insert DNA, 1 µL of commercially supplied 10× T4 DNA ligase buffer (50 mM Tris-Cl pH 7.5, 10 mM MgCl₂, 1 mM ATP and 10 mM DTT) and 1 µL T4 DNA ligase. Ligation reactions were carried out at 16 °C overnight. The entire reaction mixture was transformed into DH5α E. coli cells.

7.2.6 Transformation of DNA plasmids into E. coli

7.2.6.1 KCM competent cell preparation

Competent cells (DH5α, BL21 or Rosetta 2, Section 5.1.4) were streaked out onto LB-agar plates (without antibiotics in the case of DH5α and BL21 cells, or with Cam in the case of Rosetta 2 cells) and incubated overnight at 37 °C. A single colony from the plate was used to inoculate a culture of 10 mL LB medium (with or without antibiotics), which was incubated overnight at 37 °C with shaking at 180 rpm. This culture was then used to inoculate 400 mL of SOC medium, which was grown to A600nm of ~0.4 at 37 °C with shaking at 180 rpm (2–3 h). The culture was cooled on ice, and then centrifuged (4000 rpm, 4 °C, 5 min). The cell pellets were then resuspended in 20 mL of chilled resuspension buffer (LB containing 10% (w/v) PEG-3350, 5% (v/v) DMSO, 10 mM MgCl₂, 10 mM MgSO₄, 10% (v/v) glycerol, pH 6.1). Cells were aliquoted (100 μL) into Eppendorf tubes, snap-frozen using liquid nitrogen and stored at −80 °C until required.
7.2.6.2 Transformation
A ligation reaction mixture (10 μL) or plasmid preparation (1 μL) was added to 50 μL KCM (100 mM KCl, 30 mM CaCl₂, 50 mM MgCl₂), followed by the addition of 50 μL competent E. coli cells, on ice. Cells were left on ice for 30 min, then heat shocked at 42 °C for 90 s, before addition of 200 μL LB medium and incubation at 37 °C for 1 h. All of the culture was plated onto two LB-agar plates containing appropriate antibiotic(s) and incubated at 37 °C overnight.

7.2.6.3 Plasmid isolation and sequencing
A single colony of DH5α cells containing plasmid DNA was used to inoculate 10 mL LB medium containing appropriate antibiotic(s). The inoculated solution was incubated overnight at 37 °C with shaking at 180 rpm. Cells were harvested by centrifugation (6000 rpm, 4 °C, 10 min) and plasmids were extracted from the cell pellet using a QIAprep® spin miniprep kit according to the manufacturer’s instructions. The identity of insert DNA was confirmed by sequencing (SUPAMAC, University of Sydney or AGRF, Westmead Millennium Institute, NSW). See Appendix A for sequencing primers.

7.3 Overexpression and purification of recombinant proteins
All buffers, protein samples and beads were kept on ice unless otherwise specified. Samples were collected throughout the purification process and analysed using precast 4–20% gradient Tris-glycine PAGE gels (Invitrogen, Mt Waverley, VIC) stained in SDS-PAGE staining solution.

7.3.1 Overexpression trials of recombinant proteins
LB medium (10 mL) containing the appropriate antibiotic(s) was inoculated with a single colony of freshly transformed BL21 or Rosetta2 cells containing the relevant plasmid and incubated overnight with shaking (180 rpm, 37 °C). LB media (10, 20 or 50 mL) with the appropriate antibiotic selection was inoculated with the overnight culture to an A600nm of ~0.05 and incubated with shaking (37 °C, 180 rpm) until an A600nm of ~0.6–0.8 was reached. Protein overexpression was induced by the addition of IPTG to a final concentration of 0.1, 0.4 mM or 1.0 mM (0.1 mM ZnSO₄ was added for zinc finger containing constructs: MTA1(1–437, 87–437 and 160–437) and incubation at 15 °C for 24 h, 18 °C for 20 h, 20 °C for 18 h, 25 °C overnight or 37 °C for 3 h (Table 7-6). The cells were harvested by centrifugation (6000 rpm, 4 °C, 10 min)
and stored at −80 °C until required. Culture samples were taken prior to induction and harvest, and expression levels were analyzed by SDS-PAGE (Sambrook et al., 1989)).

Table 7-6: Conditions used for overexpression of recombinant proteins

<table>
<thead>
<tr>
<th>Construct</th>
<th>Plasmid</th>
<th>E. coli strain</th>
<th>Induction temperature (°C)</th>
<th>IPTG (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTA1 1–437, 87–437 and 160–437</td>
<td>pGEX-6P</td>
<td>Rosetta2 and BL21</td>
<td>15, 18, 20, 25 and 37</td>
<td>0.1, 0.4 and 1.0</td>
</tr>
<tr>
<td>MTA1 465–516 and 465–542</td>
<td>pGEX-4T</td>
<td>BL21</td>
<td>25 (overnight)</td>
<td>0.4</td>
</tr>
<tr>
<td>MTA1 1–168, 1–284, 1–340</td>
<td>pGEX6p, pMAL, pHUE</td>
<td>Rosetta2 and BL21</td>
<td>15, 18, 20, 25 and 37</td>
<td>0.1, 0.4 and 1.0</td>
</tr>
<tr>
<td>MTA1 161–340</td>
<td>pMAL and pGEX6P</td>
<td>Rosetta2</td>
<td>37 and 25</td>
<td>0.1, 0.4 and 1.0</td>
</tr>
<tr>
<td>MTA1 9–154 and 1–179</td>
<td>pET28b</td>
<td>Rosetta2 and BL21</td>
<td>18, 25 and 37</td>
<td>0.1, 0.4 and 1.0</td>
</tr>
<tr>
<td>MTA2 1–150 and 1–160</td>
<td>pGEX6P</td>
<td>Rosetta 2</td>
<td>18, 25 and 37</td>
<td>0.1, 0.4 and 1.0</td>
</tr>
</tbody>
</table>

7.3.2 Large scale overexpression

Large scale overexpression was carried out as described in Section 7.3.1, except that 50 mL of LB-media were inoculated and grown overnight, and used to inoculate 1 L of LB medium containing the appropriate antibiotic(s). Cultures were grown at 37 °C until induction, when the temperature, IPTG and the length of grown culture were set to the optimal conditions. The cells were then harvested by centrifugation (6000 rpm, 4 °C, 10
min), resuspended in lysis buffer (1/20 of culture volume), snap frozen and stored at –80 °C until required.

### 7.3.3 Purification of recombinant proteins

Upon thawing at room temperature, cell suspensions (or cell pellets resuspended in appropriate lysis buffer) were incubated with lysozyme (0.1 mg mL\(^{-1}\)) for 30 min with gentle rotation, followed by addition of DNase (0.1 mg mL\(^{-1}\)), RNase (0.1 mg mL\(^{-1}\)) and 2 mM MgCl\(_2\) and further incubation for 30 min at 4 °C. Alternatively, the suspended cells were lysed by sonication four times on ice (microtip power 5, 30 s), following which MgCl\(_2\) (2 mM), DNase (0.1 mg mL\(^{-1}\)) and RNase (0.1 mg mL\(^{-1}\)) were added and incubated for 1 h at 4 °C. Soluble and insoluble fractions were separated by centrifugation (10,000 rpm, 4 °C, 30 min) and the proteins were purified from the soluble fraction or refolded from the inclusion bodies as described below. Samples were retained for SDS-PAGE at each step. Reagents used for the purification are listed in Table 7–2.

### 7.3.4 GSH-affinity purification and enzymatic cleavage

Soluble fractions were applied to glutathione-Sepharose\(^\circ\)\(^\text{4B}\) beads (~5 mL/1L culture) pre-equilibrated with MQW and lysis buffer (without lysozyme, DNase, RNase and MgCl\(_2\)). The beads were then washed with wash buffer (30 column volumes (CV)) and GST-fusion proteins were eluted from the column using elution buffer containing 50 mM reduced glutathione in fractions (2 CV) with a 10 min incubation at 4 °C or room temperature in between each elution.

GST-fusion proteins were used for pulldown assays, or for further investigation the GST tag was removed by incubating the fusion protein with the appropriate protease (Thrombin or HRV 3C) at 4 °C overnight. The proteins were further purified by ion-exchange chromatography and/or size exclusion chromatography (Section 7.3.6 and 7.3.7).

### 7.3.5 Amylose-affinity chromatography of MBP-fusion proteins and enzymatic cleavage

Soluble fractions containing an MBP-tagged MTA1 fragment were applied to an amylose resin (8 mL/1L of culture volume) pre-equilibrated with lysis buffer, followed by 5 rinses with 30 CV of wash buffer. MBP-fusion proteins were eluted from the column using MBP elution buffer in fractions (1 CV) with a 10 min incubation at 4 °C.
in between each elution. The MBP-tag was then cleaved by adding Factor Xa or HRV 3C to the elution fractions and incubation at 4 °C overnight.

7.3.6 Ion exchange chromatography
MTA1 fragments were further purified by ion exchange chromatography. Prior to use, buffers and protein samples were filtered with 0.45 μm and 0.22 μm filters, respectively. The cleaved product was diluted 1:2 in low salt buffer before loading onto an appropriate column (Uno™S-1 or Uno™Q-1 column, BioRad, Hercules, CA, USA) using a Biologic DuoFlow™ HPLC system (BioRad, Hercules, CA, USA). A linear gradient of 0–0.5 M NaCl over 20 mL, followed by a linear gradient of 0.5–1 M NaCl over 30 mL, was used. The flow rate was 1 mL/min.

7.3.7 Size exclusion chromatography
MTA1 fragments were further purified by size exclusion chromatography. Prior to use, buffers and protein samples were filtered with 0.45 μm and 0.22 μm filters, respectively. The cleaved proteins or concentrated fractions obtained from ion exchange chromatography were loaded onto a HiLoad™ 16/60 Superdex™ 75 prep grade column (Amersham Biosciences, Uppsala, Sweden) using a Biologic DuoFlow™ HPLC system (BioRad, Hercules, CA, USA). For the MTA1 constructs (1–340, 1–284, 161–340, 625–715, 656–686, 465–6516 and 564–542) the column was run in 50 mM Tris, pH 7.5, 300 mM NaCl, 1 mM DTT or 50 mM HEPES, pH 7.5, 300 mM NaCl, 1 mM DTT at a flow rate of 1 mL/min. Fractions containing purified proteins were used for further investigation.

7.3.8 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis
Protein samples were mixed with NuPAGE® 4× LDS sample buffer and heated at 80 °C for 10 minutes prior to loading to precast NuPAGE® 4–12% Bis-Tris mini gels. The gels were run in NuPAGE® MES SDS running buffer at 180 V for 40 min. Gels were then fixed in destain solution (30% (v/v) methanol, 10% (v/v) acetic acid) for 10 m, stained in Coomassie stain solution (0.125% (w/v) Coomassie BBR, 40% (v/v) methanol, 7% (v/v) acetic acid) for ~ 1 h, and destained in fixing/destain solution for 1 h.
7.4 Refolding of Ub-His-MTA1(BAH-ELM2, 1–284) and (BAH-ELM2-SANT, 1–340)

7.4.1 Preparation of inclusion bodies
pHUE-MTA1(1–340) and (1–284) constructs were used to express and purify the protein from inclusion bodies. These constructs were expressed in Rosetta2 cells at 25 °C overnight (Section 7.3.2). Cells were lysed in 50 mM Tris 7.5, 300 mM NaCl, 1% Triton X-100, 1 mM DTT, 1 mM PMSF as described in Section 7.3.3. After separating the soluble from insoluble fractions by centrifugation, the pellet was washed in the same buffer with and without 2% Triton X-100, and then centrifuged again at 10,000 rpm for 20 min at 4 °C. The inclusion bodies were solubilized by suspending the pellet in denaturation buffer (100 mM phosphate, 10 mM Tris, pH 8.0, 6 M Guanidine-HCl, 2 mM β-mecaptoethanol (BME)), with stirring at room temperature for ~1 h, and then clarified by centrifugation at 10,000 rpm for 30 min.

7.4.2 Ni-NTA affinity chromatography
Solubilized inclusion bodies containing Ubiquitin-His-tagged (Ub-His) N-terminal MTA1 fragments were applied to nickel-nitrilotriacetic acid (Ni-NTA) beads (~3 mL per 1 L culture) pre-equilibrated with denaturation buffer, followed by 4 rinses with 30 CV of wash buffer (100 mM phosphate, 10 mM Tris, pH 8.0, 8 M urea, 2 mM βME, 10 mM imidazole). Ub-His-tagged MTA1 fragments were eluted from the column using elution buffer (100 mM phosphate, 10 mM Tris, pH 8.0, 6 M urea, 1 mM BME, 500 mM imidazole) in 1 CV fractions.

7.4.3 Refolding the MTA1 fragments and UBP41 cleavage
We performed a refolding screen test using a set of test solutions with different components, including the type of buffer, pH, salt, detergent, glycerol and reducing agent (Table 2-1). The screen solution (190 µl) was added to solubilized protein (10 µl at 2–3 mg/mL) in a 96-well plate. The plate was shaken for about 60 min at room temperature, then protein aggregation was monitored by reading the absorbance at 320 nm and 390 nm (Burgess, 2009, Tresaugues et al., 2004).

The conditions that gave the lowest absorbance were used for protein refolding by multi-step dialysis at 4 °C or at room temperature. Ub-His-MTA1(1–284) was dialysed against 20 mM Tris or 100 mM phosphate at varying pH (7.0, 8.0, 9.0), 100 mM NaCl,
10% glycerol and 4M urea; the urea concentration was reduced gradually with each buffer exchange (4 M, 2 M then 0 M) using a 5000 Da molecular mass cutoff membrane in 2 L of dialysis buffer or against 10 mM phosphate, 10 mM Tris, 10 mM acetate at pH 7.0 or 9.0 and 150 mM NaCl.

Ub-His-MTA1(1–340) was dialysed against 20 mM Tris, pH 8.0, 100 mM NaCl, 10% glycerol and 4M urea; the urea concentration was reduced gradually with each buffer exchange (4 M, 2 M then 0 M). This was followed by dialysis against 10 mM phosphate, 10 mM Tris, 10 mM acetate at pH 9.0 and 150 mM NaCl, or 20 mM Tris at pH 8.5, 100 mM NaCl and 10 mM Arg.

After the above dialysis, the protein was dialyzed into the cleavage buffer for the removal of His-ubiquitin tag using UBP41 protease at 4 °C overnight or room temperature for 3–4 h, before being subjected to size exclusion chromatography (Section 7.3.7).

7.5 Expressing MTA1 constructs using the InsectDirect System

MTA1 constructs (Table 7-8) were cloned using a standard PCR approach from a codon-optimized DNA sequence specific for insect cell expression (purchased from Genscript). MTA1(9–154), (1–287), (155–334), (155–354) and (1–334) were ligated into pIEX with an N-terminal FLAG-tag via BamHI and XhoI restriction sites, whereas MTA1(448–715) was ligated into pIEX vector with an N-terminal His-tag via NcoI and XhoI restriction sites.

Table 7-7: Constructs used for expression in Sf9 insect cells using the InsectDirect system

<table>
<thead>
<tr>
<th>MTA1 constructs</th>
<th>Size (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9–154</td>
<td>18.3</td>
</tr>
<tr>
<td>1–287</td>
<td>34</td>
</tr>
<tr>
<td>155–334</td>
<td>21.9</td>
</tr>
<tr>
<td>155–354</td>
<td>24.2</td>
</tr>
<tr>
<td>1–334</td>
<td>39.7</td>
</tr>
<tr>
<td>448–715</td>
<td>32</td>
</tr>
</tbody>
</table>
Spodoptera frugiperda (Sf9) insect cells (Invitrogen, Carlsbad, CA, USA) were grown in a shake flask in serum free media (SFM: Sf-900™ II, Invitrogen) at 28 °C. Approximately 1.5 × 10^6 cells in 1mL of SFM were transferred to a 6-well tissue culture plate and left for ~1 h at 28 °C to adhere to the plate. DNA (2 µg) of each pIEX-MTA1 construct was mixed with 200 µL SFM and 10 µL of GeneJuice® Transfection Reagent (Novagen) and incubated for 15 min at room temperature. The mixture was then added to the Sf9 cells in the 6-well plate and incubated for 48–60 h at 28 °C (Loomis et al., 2005). Untransfected cells were used as a negative control, whereas pIEX-GFP was used as positive control for protein expression using the same protocol as for the pIEX-MTA1 constructs.

For MTA1(448–715) construct, we scaled up the protocol 10-fold, i.e. 10 mL suspension culture of Sf9 cells at 1 × 10^6 cells/mL, 20 µg DNA of pIEX-MTA1(448–715), and 100 µL of GeneJuice® Transfection Reagent. The mixture was added to the Sf9 cells in a shake flask and then incubated at 28 °C with shaking at 150 rpm for 48 h. Cells were then harvested by centrifugation at 10,000 rpm for 10 min at 4 °C and then lysed using I-PER Insect Cell Protein Extraction Reagent (Thermo Scientific) according to the manufacturer’s instructions. Soluble and insoluble fractions were separated by centrifugation at 12000 rpm for 20 min at 4 °C.

7.6 Baculovirus expression and purification of RbAp48 and MTA1(448–715) in insect cells

7.6.1 Generation of recombinant virus bacmid by site-specific transposition

Recombinant baculovirus containing the RbAp48 gene was generated using the Bac-to-Bac system (Invitrogen, Carlsbad, CA, USA), with the exception that the pFBDM vector was used in place of the pFASTBac vector.

Full-length 6×His-RbAp48 or 6×His-MTA1(448–715) cloned in the pFBDM vector was transformed into E. coli DH10Bac cells (Invitrogen, Carlsbad, CA, USA). Recombinant pFBDM (1 ng) was mixed with 100 µL of DH10Bac cells and incubated on ice for 30 min. The cells were then heat shocked in a 42 °C water bath for 90 s, and chilled on ice for 5 min. A 900 µL aliquot of SOC medium was added to the mixture and then incubated at 37 °C for 4 h with shaking at 200 rpm. Three serial 1 in 10 dilutions were made to the cells, and 100 µL of each dilution was plated on to LB plates containing 50 µg/mL kanamycin, 7 µg/mL gentamycin, 10 µg/mL tetracycline, 100 µg/mL Bluo-gal
and 40 µg/mL IPTG. The plates were incubated for 24–48 h until blue colonies could be discerned from the white colonies.

White colonies, which contained recombinant bacmid, were streaked onto fresh plates and incubated at 37 °C overnight to confirm the white phenotype. Upon determining phenotype, a single white colony was inoculated into 5 mL of LB (containing 50 µg/mL kanamycin, 7 µg/mL gentamycin and 10 µg/mL tetracycline) in a 15 mL polypropylene Falcon tube, and incubated at 37 °C for 24 h with shaking at 200 rpm. The DH10Bac cells were then harvested by centrifugation at 6000 rpm and the recombinant bacmid DNA was extracted from the cells according to the Bac-to-Bac (Invitrogen, Carlsbad, CA, USA) instruction manual. PCR analysis using RbAp48-specific forward primer and M13 reverse primer was carried out to confirm successful transposition to the bacmid. The recombinant bacmid DNA was stored at −20 °C until further use.

7.6.2 Transfection of recombinant bacmid DNA into SF9 insect cells

1 × 10⁶ Spodoptera frugiperda SF9 insect cells (Invitrogen, Carlsbad, CA, USA) in 2 mL Sf-900 II SFM were seeded into 35 mm wells of a 6-well plate. The cells were incubated at 27 °C for 1 h to allow attachment to the wells.

Two separate mixtures were made up: 10 µL bacmid DNA was added to 200 µL Sf-900 II SFM, and 10 µL Fugene® 6 transfection reagent (Promega) was added to 100 µL Sf-900 II SFM. The two mixtures were subsequently combined and added to the adherent SF9 cells. The transfected cells were incubated at 27 °C for 72 h. Uninfected SF9 cells were used as a control.

The baculovirus was harvested by centrifugation of the 72 h post-transfection cell-culture medium at 500 × g for 5 min. The baculovirus-containing supernatant (P1) was transferred to a fresh tube and stored at 4 °C protected from light until further use.

7.6.3 Baculovirus amplification and expression

For amplification of the baculovirus, 500 µL of the P1 baculovirus was added into 1 × 10⁶ SF9 cells/mL in 50 mL Sf-900 II SFM and incubated for 48 h at 27 °C. The baculovirus was harvested by centrifugation of the 48 h post-infection cell-culture medium at 500 × g for 5 min. The baculovirus-containing supernatant (P2) was transferred to a fresh tube and stored at 4 °C protected from light until further use. One more round was performed to produce a high titre P3 baculovirus, which was used for
protein expression. To first determine the optimal conditions for protein expression, the P3 baculoviral stock was used to infect a small population of cells (25 ml at 1 × 10⁶ Sf9 cells/mL) at varying multiplicity of infection (MOI; 1, 5 and 10) and assayed for protein expression. We also performed a time course by infecting Sf9 cells (25 mL at 1 × 10⁶) with the P3 viral stock at a constant MOI of 1 and assayed for protein expression at different times post-infection (24, 48, 72, 96 and 120 h post-infection). We then used the MOI and time that provided the optimal recombinant protein expression for large scale expression.

For large-scale expression of 6×His-RbAp48, 1.5 mL (MOI 1) of the P3 baculovirus was added into 1 × 10⁶ cells/mL in 500 mL Sf-900 II SFM for infection and incubated for 80 h at 27 °C with shaking at 120 rpm. The cells were harvested by centrifugation at 5,000 rpm for 20 min at 4 °C and stored at −80 °C until required.

For large-scale expression of 6×His-MTA1(448-715), different MOI (1, 5 and 10) were used to infect 1–1.5 × 10⁶ cells/mL in 500 mL Sf-900 II SFM, and the cells were incubated for 72–96 h at 27 °C with shaking at 120 rpm.

7.6.4 Extraction of protein from Sf9 cells
Cells from 1 L of baculovirus expression insect cell culture were resuspended in 30 ml of either I-PER Insect Cell Protein Extraction Reagent (Thermo Scientific), according to the manufacturer’s instructions, or in 30 ml of 20 mM Tris, pH 8.0, 150 mM NaCl, 10 mM MgCl₂, 1 mM β-mercaptoethanol, and EDTA-free protease inhibitor. The cells were lysed by sonication. NP-40 (0.1 % v/v) and DNase I were added to the cell lysate before the cell debris was pelleted at 20,000 rpm for 40 min at 4 °C.

7.6.5 Ni-NTA affinity chromatography and thrombin cleavage
The supernatant was transferred to a glass econo-column containing 2.5 mL of Ni-NTA beads and rotated for 2 h at 4 °C to allow protein binding. Beads were then washed with 20 CV of wash buffer (20 mM Tris, pH 8.0, 300 mM NaCl, and 2 mM imidazole), followed by the same buffer containing 150 mM NaCl, and then in the same buffer with stepwise increasing concentrations of imidazole (up to 40 mM). 6×His-RbAp48 was then eluted with 500 mM imidazole in 2 mL fractions. Fractions containing 6×His-RbAp48 were identified by SDS-PAGE, pooled and then dialysed overnight using an 8000 Da molecular mass cutoff membrane in 2 L of dialysis buffer (20 mM Tris, pH
8.0, 150 mM NaCl, and 2.5 mM CaCl2). The 6xHis-tag was cleaved from RbAp48 by adding thrombin protease (10–20 U per 1 mL of elution fractions) to the dialysate.

### 7.6.6 Anion exchange chromatography

Following dialysis and cleavage, RbAp48 was concentrated with a Vivaspin 10,000 Da MWCO filter to a volume of 5–8 mL and diluted in 20 mM Tris (pH 7.5) to bring the NaCl concentration down to 30 mM. DTT (1 mM) and Nonidet P-40 (0.02% v/v) were then added. The sample was then applied to a Uno™Q-1 column (BioRad, Hercules, CA, USA) using a Biologic DuoFlow™ HPLC system (BioRad, Hercules, CA, USA), and unbound protein was washed off with 20 mM Tris, pH 7.5. Bound protein was eluted from the column using a 0–50% gradient of 20 mM Tris (pH 7.5) and 1 M NaCl over 80 mL. Fractions containing RbAp48 were identified by SDS-PAGE.

### 7.6.7 Size exclusion chromatography

Ion exchange fractions containing RbAp48 were pooled and concentrated in a Vivaspin 6 10,000 Da MWCO concentrator and loaded onto a HiLoad™ 16/60 Superdex™ 75 pg column (Amersham Biosciences, Uppsala, Sweden) using a Biologic DuoFlow™ HPLC system (BioRad, Hercules, CA, USA). The column was run in 20 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT at a flow rate of 1.0 mL/min. Fractions containing RbAp48 were identified by SDS-PAGE, pooled and concentrated in a Vivaspin 6 10,000 Da MWCO concentrator to a final concentration of 10 mg/mL.

### 7.7 Spectrophotometric quantification of protein and nucleic acid

Protein and DNA concentrations were determined via a NanoDrop ND-1000 UV-visible spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Samples were diluted where necessary with appropriate buffers that were used as blank solutions.

DNA concentration was determined by absorbance readings at 260 nm with the assumption that 50 μg/mL double-stranded DNA gives an $A_{260} = 1.0$ (Sambrook et al., 1989). Protein concentrations were determined by absorbance at 280 nm according to the Beer-Lambert equation, using extinction coefficients ($\varepsilon$; Table 7–9) calculated from the amino acid sequence using the web-based program ProtParam on the ExPASy proteomics server (Swiss Institute of Bioinformatics, [http://www.isb-sib.ch/](http://www.isb-sib.ch/)). For proteins and peptides with no aromatic residues, the concentration was determined using the following equation (Murphy and Kies, 1960):
\[ C (\mu g/mL) = (A_{215} - A_{225}) \times 144 \]

where \( C \) is the concentration in \( \mu g/mL \) and \( A_{215} \) and \( A_{225} \) are the absorbances at 215 nm and 225 nm, respectively.

### Table 7-8: Molar extinction coefficients of protein constructs

<table>
<thead>
<tr>
<th>Protein</th>
<th>Size (kDa)</th>
<th>( \varepsilon (M^{-1}cm^{-1}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BES</td>
<td>39</td>
<td>46090</td>
</tr>
<tr>
<td>BE</td>
<td>32.1</td>
<td>22140</td>
</tr>
<tr>
<td>ES</td>
<td>20.5</td>
<td>32555</td>
</tr>
<tr>
<td>RbAp48</td>
<td>48.6</td>
<td>83670</td>
</tr>
</tbody>
</table>

### 7.8 Far UV-Circular Dichroism (CD) spectropolarimetry

Protein samples (10–20 \( \mu M \)) for far UV-CD experiments were dialysed into 20 mM phosphate, pH 8.0, 100 mM NaF, 1 mM DTT, or 20 mM Tris, pH 8, 50 mM NaCl, then placed in a 1-mm path length quartz cell seated in a water-jacketed cell holder. CD spectra were recorded at 20 °C on a Jasco J-720 spectropolarimeter equipped with a Neslab RTE-111 temperature controller. CD data were collected over the wavelength range 184–260 nm, with a speed of 20 nm/min, step resolution of 1 nm, bandwidth of 1 nm and a response time of 1 s. Final spectra were the average of two scans, and were baseline corrected. Raw data were converted to mean-residue ellipticity ([\( \theta \]) deg/cm²/dmol) using the following equation:

\[ [\theta]_{M,\lambda} = M \theta_d/10dc \]

where \( M \) is the mean residue weight, \( \theta_d \) is the observed ellipticity (deg) at the wavelength \( \lambda \), \( d \) is the path length (cm), and \( c \) is the concentration (g/mL). Estimates of secondary structure were made using the CDPro suite of programs (Johnson, 1999).

### 7.9 Preparation of MTA1 peptides

Commercially synthesized lyophilized MTA1(656–686, 656–670, 646–679, 671–686, 670–695 and 670–711) peptides (ChinaPeptides, China) were resuspended in 50 mM Tris-HCl at pH 8.0 to a concentration of 5 mM and aliquoted into smaller volumes (50 \( \mu L \)), snap-frozen and stored at −20 °C until further use in either pulldown experiments, ITC or crystallization (Table 7-4).
Table 7-9: List of MTA1 peptides used in this thesis. The ‘b’ indicates the biotin position in peptides (P1–P4).

<table>
<thead>
<tr>
<th>MTA1 peptide</th>
<th>Sequence</th>
<th>MW (Da)</th>
<th>ε_280 (M⁻¹ cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 (656-686)</td>
<td>b-DVFYMATEETRKRKLLSSSETKRAARRPYK</td>
<td>3732.3</td>
<td>2980</td>
</tr>
<tr>
<td>P2 (656-670)</td>
<td>DVFYMATEETRK-b</td>
<td>1887.1</td>
<td>1490</td>
</tr>
<tr>
<td>P3 (646-679)</td>
<td>ETRKIRKLLSSSETK-b</td>
<td>1932.2</td>
<td>-</td>
</tr>
<tr>
<td>P4 (671-686)</td>
<td>LLSSSETKRAARRPYK-b</td>
<td>1863.1</td>
<td>1490</td>
</tr>
<tr>
<td>P5 (670-695)</td>
<td>KLLSSSETKRAARRPYKPIALRQSQA</td>
<td>2956.4</td>
<td>1490</td>
</tr>
<tr>
<td>P6 (670-711)</td>
<td>KLLSSSETKRAARRPYKPIALRQSQALPPRPPPAPVNDPI</td>
<td>4644.4</td>
<td>1490</td>
</tr>
</tbody>
</table>

7.10 One dimensional ¹H NMR experiments

Purified protein samples were concentrated using Vivaspin columns (Sartorius Stedim Biotech, Dandenong, VIC, AUS) with an appropriate MW cut off filter. Deuterium oxide (5% (v/v); D₂O) and 1 μL of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) were added to the protein sample prior to analysis. The sample was placed into a 5 mm or 3 mm susceptibility-matched microcell NMR tube (Shigemi, Tokyo, Japan). 1D ¹H-NMR spectra of samples were recorded at 25 °C on a 600 MHz Bruker DRX-600 spectrometer. NMR spectra were processed using TOPSPIN (Bruker). Spectra were referenced to DSS at 0.00 ppm.

7.11 Isothermal titration calorimetry (ITC)

ITC measurements were carried out at 25 °C using a MicroCal iTC200 titration calorimeter. RbAp48 and MTA1 peptides were dialyzed separately overnight against buffer containing 20 mM Tris, pH 7.5, and 150 mM NaCl. The MTA1 peptide (250 μM, 200 μL) was titrated into RbAp48 (25 μM, 200 μl) in a series of 20 × 2-μl injections, with a 2.5 min interval between each injection. The reference power was set at 2 μcal/s, and the cell was stirred continuously at 1000 rpm. The evolved heats were integrated and normalized for protein concentration. After base-line correction (using data from titration of the MTA1 peptide into buffer), the data were fitted to a single-site binding model in MicroCal Origin 7.0.
MTA1 peptides (656–686, 670–695 and 670–711) were dissolved in 50 mM Tris, pH 8.0. Concentrated RbAp48 (~10–12 mg/ml) was mixed with MTA1(656–686, 670–695 or 670–711) peptide to obtain a protein:peptide molar ratio of 1:5 in the case of MTA1(656-686) peptide and 1:1.2 in the case of MTA1(670–695 or 670–711) peptide.

In the case of RbAp48-MTA1(656–686), crystallization trials were set up in 96-well plates as sitting drops using 0.2 µL each of protein:peptide solution and commercial MORPHEUS crystallization screens (Molecular Dimensions Ltd, MRC Technology) using an Oryx6 protein crystallization robot (Douglas Instruments, Berkshire, UK). Crystals were obtained after incubation overnight at 4 °C in several conditions of MORPHEUS crystallization screens. Single crystals were obtained in condition Morpheus D1 (30% PEG550 MME/PEG20K, 0.1 M MES/imidazole, pH 6.5, 10% alcohols), and grew to a final size within 4 days. These crystals were harvested into a cryo-protectant solution containing 25% PEG 400 in mother liquid before cryocooling in liquid nitrogen. The crystals were sent to the Diamond Light Source Synchrotron (Oxfordshire, UK) for X-ray diffraction data collection.

In the case of RbAp48-MTA1(670–695 and 670–711), crystallization trials were set up in 96-well plates as sitting drops using a Nanodrop Mosquito robot (TTP Labtech, Hertfordshire, UK) and commercial crystallization screens (NeXtal DWBlocks, Qiagen). Crystals grew after incubation overnight at 18 °C in several conditions. X-ray data were collected from a single crystal obtained in 0.2 M calcium acetate and 20% w/v PEG3350 and cryo-protected with the addition of 25% glycerol. Data collection was performed using an in-house rotating copper anode generator on a Mar345 image-plate detector (Mar Research, Norderstedt, Germany) with Cu-Kα X rays from a Rigaku RU-200 rotating anode generator (Rigaku, The Woodlands, Texas, USA).

Data were processed and scaled using MOSFLM (Leslie and Powell, 2007) and Scala (Evans, 2006). The structure was solved by molecular replacement with the program Phaser (McCoy et al., 2007) using RbAp48 as the search model (Protein Data Bank code 2XU7). The model was built with Coot (Emsley and Cowtan, 2004) and refined using Refmac (Murshudov et al., 1997). MolProbity (Davis et al., 2007) was used to
validate the structure and the EBI PISA server (Krissinel and Henrick, 2007) was used
for interface analysis.

7.13 Pulldown assays

7.13.1 In vitro transcription/translation of $^{35}$S-Met-labelled protein

$^{35}$S-Met-labeled MTA1/2 and their constructs (Table 7–5) and $^{35}$S-Met-labeled FLAG-RbAp46/48 (wild-type and mutants) were in vitro transcribed/translated using the TNT® Quick Coupled Transcription/Translation (rabbit reticulocyte lysate, RRL) system according to the manufacturer’s instructions (Promega, NSW, Australia).

For the FLAG pulldowns, rabbit reticulocyte lysate containing the in vitro translated $^{35}$S-labeled FLAG-RbAp46/48 (20 μL) was immobilised onto 20 μL of pre-equilibrated anti-FLAG M2 beads (Sigma) in 300 μL of binding/wash buffer (50 mM Tris, pH 7.5, 300 mM NaCl, 0.5% Triton X-100, 1× Complete EDTA-free protease inhibitor) for 30 min at 4 °C. Rabbit reticulocyte lysate containing the IVT $^{35}$S-labeled MTA1/2 or their fragments (40 μL) was added to the mixture and incubated for an extra 1–2 h. For negative control pulldowns, anti-FLAG beads alone with MTA1/2 were used. Beads were washed five times with wash buffer. 20 μL of SDS-PAGE loading dye was added to the pulldown beads and heated for 85 °C for 10 min. Following electrophoresis the gel was transferred to Whatman 3M paper and exposed overnight on a Phosphoimager™ screen (Molecular Dynamics, Sunnyvale, CA). The Phosphoimager screen was scanned the following day using the Typhoon™ FLA 9000 scanner (Amersham Biosciences, Uppsala, Sweden) and visualized on the ImageQuant™ software.

For competition experiments, the pulldown experiment was carried out as described above, except that increasing amount of FOG1(1–15) was also added during the incubation.

For the streptavidin pulldowns, approximately 10 μg biotinylated MTA1(656-686, 656-670, 664-679, 671-686) peptides were immobilised on 20 μL high capacity streptavidin beads which were then resuspended in 250 μL of binding buffer (Tris pH 7.4, 300 mM NaCl, 0.5% Triton X-100, 1× Complete EDTA-free protease inhibitor). The in vitro translated $^{35}$S-labeled RbAp48 (20 μL) or insect cell produced RbAp48 was incubated with a suspension (250 μL) of streptavidin beads, to which biotinylated-MTA1 peptide
was bound, for 2 h at 4 °C. For negative control pulldowns, streptavidin beads alone were used. Beads were washed five times with binding buffer. 20 µL of SDS-PAGE loading dye was added to the pulldown beads and heated for 85 °C for 10 min. Following electrophoresis, bound proteins were analyzed by SDS-PAGE followed by direct autoradiography.

### 7.13.2 GST pull-down assay

The GST pulldown assays were performed by incubating equal amounts of GST or GST fusion protein immobilised to glutathione-Sepharose beads with IVT \(^{35}\)S-labeled RbAp48 protein. The mixtures were incubated for 2 h at 4°C and washed five times with wash buffer (Tris pH 7.4, 300 mM NaCl, 0.5% Triton X-100, 1× Complete EDTA-free protease inhibitor). SDS-PAGE loading dye (20 µL) was added to the pulldown beads and heated for 85 °C for 10 min, and bound proteins were analyzed by SDS-PAGE followed by direct autoradiography.

For competition experiments, GST-H4(1–48) were expressed in the BL21 strain of *E. coli* and purified by glutathione affinity chromatography as described previously (Murzina et al., 2008). Pulldown experiments were carried out by adding 40 µL each of reticulocyte lysate containing \(^{35}\)S-labeled RbAp48 and \(^{35}\)S-labeled MTA1 during the incubation. Bound proteins were analyzed as described above.

### 7.14 Western blot

Protein samples and Novex® Sharp Pre-stained Protein Standard (Invitrogen, Mt Waverley, VIC) were resolved by SDS-PAGE. The electrophoresed gel and Hybond™ ECL nitrocellulose membrane (GE Healthcare, Castle Hill, NSW)/ polyvinylidene difluoride (PVDF) were both pre-equilibrated (~10 min) with a transfer buffer (1× MES buffer, 20% (v/v) methanol). Proteins were transferred onto a pre-wetted membrane between thick blotting papers (Bio-Rad, Gladesville, NSW) using a semi-dry electroblotting device (Trans-Blot® SD, Bio-Rad Gladesville, NSW). After the transfer (RT, 25 min, 10 V), the membrane was blocked with blocking buffer (5% skim milk (Fonterra™, New Zealand), 0.02% (v/v) Tween-20 in PBS (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4) at room temperature for 30 min. The membrane was washed (5 × 5 ml) in PBS-T (0.01% (v/v) Tween-20 in PBS) with agitation, and then incubated with 1 µg mouse monoclonal anti-His antibody (Amersham Biosciences, Castle Hill, NSW) or 10 µg of anti-FLAG antibody (Sigma)
in 10 mL PBS-T for 1 h at room temperature. The membrane was washed 5 times for 10 min in PBS-T, and then incubated with 0.2 μg goat anti-mouse-HRP secondary antibody (Santa Cruz Biotechnology, CA, USA) in 10 mL PBS-T for 30 min at room temperature. The membrane was washed 5 times in PBS-T again. Proteins were visualized using the Western Lightning® Chemiluminescence reagent (PerkinElmer, Wellesley, MA), X-ray film (Eastman Kodak Company, Rochester, NY) and Kodak reagents.
References


REFERENCES


SOLARI, F., BATEMAN, A. & AHRINGER, J. 1999. The Caenorhabditis elegans genes egl-27 and egr-1 are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. Development, 126, 2483-2494.


References


**Appendices**

**Appendix A**

Oligonucleotides used for generating RbAp46/48 and MTA1/2 constructs. All primers are listed in the 5’→3’ direction.

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>RbAp46 fwd</td>
<td>CGGGATCCATGGCGAGTAAAGCGAGAGGAGT</td>
</tr>
<tr>
<td>RbAp46 rev</td>
<td>GGAATTCTCATCAAGATCCTTTGTCCTCCAGTTTCG</td>
</tr>
<tr>
<td>RbAp48 fwd</td>
<td>CGGGATCCATGGCGACAGGAAGCACAGC</td>
</tr>
<tr>
<td>RbAp48 rev</td>
<td>GGAATTCTCATCAAGGACCCTTTGTCCTCTGGG</td>
</tr>
<tr>
<td>MTA1  fwd</td>
<td>CGGGATCCATGGCGACACATGTACAGG</td>
</tr>
<tr>
<td>MTA1  250 rev</td>
<td>GGAATTCTCATCAAGGTGATGTCTTCGAGGAGCAG</td>
</tr>
<tr>
<td>MTA1  230 fwd</td>
<td>CGGGATCCGTGACGACCTTTCGCACAG</td>
</tr>
<tr>
<td>MTA1  550 rev</td>
<td>GGAATTCTCATCAAGGGCTTGGGGGGGGC</td>
</tr>
<tr>
<td>MTA1  530 fwd</td>
<td>CGGGATCCGTACACGAAGCCGCTGGGAAG</td>
</tr>
<tr>
<td>MTA1  715 rev</td>
<td>GGAATTCTCATACGTCTCGATAGATGGCGG</td>
</tr>
<tr>
<td>MTA1-625 fwd</td>
<td>CGGGATCCATGGCGGACACATGTACAGG</td>
</tr>
<tr>
<td>MTA1-715 rev</td>
<td>CGGAATTCATCAGATCAGATCAGATGACATTG</td>
</tr>
<tr>
<td>MTA1-656 fwd</td>
<td>CGGGATCCATGGCGGACACATGTACAGG</td>
</tr>
<tr>
<td>MTA1-686 rev</td>
<td>CGGAATTCATCAGATCAGATCAGATGACATTG</td>
</tr>
<tr>
<td>MTA1-643 fwd</td>
<td>CGGGATCCATGGCGGACACATGTACAGG</td>
</tr>
<tr>
<td>MTA1-695 rev</td>
<td>CGGAATTCATCAGATCAGATCAGATGACATTG</td>
</tr>
<tr>
<td>MTA1-1 fwd</td>
<td>CGCCTCCATGGCGGACACATGTACAGG</td>
</tr>
<tr>
<td>MTA1-179 rev-XhoI</td>
<td>CGCCTCGAGTCTAGGTGTGATGTCTGCCTGTCAGTAC</td>
</tr>
<tr>
<td>MTA1-9 fwd-NcoI</td>
<td>CGCCTCCATGGCGGACACATGTACAGG</td>
</tr>
<tr>
<td>MTA1-154 rev-XhoI</td>
<td>CGCCTCGAGTCTAGGTGTGATGTCTGCCTGTCAGTAC</td>
</tr>
<tr>
<td>MTA1-448 fwd</td>
<td>CGGGATCCATGGCGGACACATGTACAGG</td>
</tr>
<tr>
<td>Gene</td>
<td>Sequence</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>MTA1-478 rev</td>
<td>GC GAATTC TCA CAGTTCGTTGCTGTCAGG T</td>
</tr>
<tr>
<td>MTA1-477-fwd</td>
<td>CG GGATCC ATG AAATGACGCGCATTGTCAGC C</td>
</tr>
<tr>
<td>MTA1-501 rev</td>
<td>GC GAATTC TCA CAGATACGGGTGACGAGCTGCATGC</td>
</tr>
<tr>
<td>MTA1-501 fwd</td>
<td>CG GGATCC ATG GCGGATATAAAGCAAGCGGC</td>
</tr>
<tr>
<td>MTA1-542 rev</td>
<td>GC GAATTC TCA TTTCCAGATAGCGCAAGACGGC</td>
</tr>
<tr>
<td>MTA1-465 fwd</td>
<td>CG GGATCC ATG AAAACCCGTCAGGCTTTTACC TGC</td>
</tr>
<tr>
<td>MTA1-516 rev</td>
<td>GC GAATTC TCA CAGACGAGCGGTGCATTCTGC</td>
</tr>
<tr>
<td>MTA1-526 rev</td>
<td>GC GAATTC TCA CAGAACCAGCGCGCAGACTGTCGA</td>
</tr>
<tr>
<td>MTA2  fwd</td>
<td>CGGGATCCATGGCGCGCAACATGTACC</td>
</tr>
<tr>
<td>MTA2  250 rev</td>
<td>G GAATTCATCGACATGGCGCTTAGCGAGTC</td>
</tr>
<tr>
<td>MTA2  220 fwd</td>
<td>CGGGATCCAGCAGCGCTCACTCGAGCCAG</td>
</tr>
<tr>
<td>MTA2  530 rev</td>
<td>G GAATTCATCGACATGGCGCTTAGCGAG</td>
</tr>
<tr>
<td>MTA2  520 fwd</td>
<td>CGGGATCCATCGACATGGCGCTTAGCGAG</td>
</tr>
<tr>
<td>MTA2  668 rev</td>
<td>G GAATTCATCGACATGGCGCTTAGCGAG</td>
</tr>
<tr>
<td>MTA2  fwd</td>
<td>CG GGATCC ATG GCGGATATAGAGATCAGCG</td>
</tr>
<tr>
<td>MTA2  150 rev</td>
<td>CG GAATTC TCA ACC AAC TCT GAT CTC CCC TTG ATC AG</td>
</tr>
<tr>
<td>MTA2  160 rev</td>
<td>CG GAATTC TCA ACG ATC TGG GAT CTC AGC TTG GAA TTT G</td>
</tr>
</tbody>
</table>
Appendix B

Oligonucleotides used for generating MTA1 constructs for insect cell expression. All primers are listed in the 5'→3' direction.

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTA1-448 fwd</td>
<td>CGCATATG TCCCCC CACGTT TG C</td>
</tr>
<tr>
<td>MTA1-715 rev</td>
<td>GCGGATCCCTAGTCTTGATGATGTTCCATCGTTG CATGAG</td>
</tr>
<tr>
<td>MTA1-448fwd-NcoI</td>
<td>CGCCATGGGCAGCAGCCATCATCATCATCATCACACAGCG GATCCATGTCCCCACG</td>
</tr>
<tr>
<td>MTA1-715 rev-XhoI</td>
<td>GCCTCGAGTTAGTCTTGATGATGTTCCATCGTTG CATGAG</td>
</tr>
<tr>
<td>MTA1-9fwd-BamHI</td>
<td>CGGGATCCCGCGACTATGTGTATTTCCAAAATAGT TCC TCA AAC</td>
</tr>
<tr>
<td>MTA1-154-rev-XhoI</td>
<td>CGCTCGAGTTAACCTGACGAGCTAGAAGAAGAAGAATCTTCG</td>
</tr>
<tr>
<td>MTA1-155fwd-BamHI</td>
<td>CG GGATCC CCT CAA CAA AAG ACG CTG CTG G</td>
</tr>
<tr>
<td>MTA1-1fwd-BamHI</td>
<td>CGGGATCC ATG GCG GCT AAT ATG TAT CGT GTG GG</td>
</tr>
<tr>
<td>MTA1-448Fwd-NdeI</td>
<td>CGCATATGTCCCCCACGTTG TTG C</td>
</tr>
<tr>
<td>MTA1-715 rev-BamHI</td>
<td>GCCTCGAGTTAGTCTTGATGATGTTCCATCGTTG CATGAG</td>
</tr>
<tr>
<td>mta334 rev-XhoI</td>
<td>CGCTCGAGTTAACCTGACGAGCTAGAAGAAGAATCTTCG</td>
</tr>
<tr>
<td>MTA1-354 rev-XhoI</td>
<td>CGCTCGAGTTAGTACCTGATGATTAGACTCGGCCCTC</td>
</tr>
<tr>
<td>MTA1-287rev-XhoI</td>
<td>CGCTCGAGTTATCCTCCATCTCATCACCGGAC</td>
</tr>
<tr>
<td>MTA1 BAH fwd</td>
<td>GTCGATTAACCAACAAAGACGCTGCTGCCGCATAAG G</td>
</tr>
<tr>
<td>MTA1 BAH rev</td>
<td>GTCTTTTTGTGGTAAATCGTGACACCAGAGTAGAAGAA GAAATCTTCG</td>
</tr>
<tr>
<td>MTA1 BAH-ELM2 fwd</td>
<td>GGAGGAATAATCTGCTGAGGCAGACCTGTTGG</td>
</tr>
<tr>
<td>MTA1 BAH-ELM2 rev</td>
<td>GATGCAGATTATCCTCCATCATTACACCGCACAGGC</td>
</tr>
</tbody>
</table>
Oligonucleotides used for the generation of MTA1 mutants. All primers are listed in the 5’ → 3’ direction.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 (K678A/R679A) fwd</td>
<td>GAAACCGCGCGGCTGCCCGCCGGCCCTACAAG</td>
<td>M1 (K678A/R679A) rev</td>
<td>GGGCAGCCGCCGCGGTTTCCGAGGATGAGAGCAGCCTTGG</td>
</tr>
<tr>
<td>M1 (K678A/R679A) rev</td>
<td>GGGCAGCCGCCGCGGTTTCCGAGGATGAGAGCAGCCTTGG</td>
<td>M2 (R682A/R683A) fwd</td>
<td>CTGCAGCCGCGCTGCAGCCGAGCCCTACAGCAGCCGAG</td>
</tr>
<tr>
<td>M2 (R682A/R683A) rev</td>
<td>CTGCAGCCGCGCTGCAGCCGAGCCCTACCACGAGCCGAG</td>
<td>M3 (K678A/R679A, R682A/R683A) fwd</td>
<td>GAAACCGCGCGGCTGCCCGCCGGCCCTACAACAGCCGAG</td>
</tr>
<tr>
<td>M3 (K678A/R679A, R682A/R683A) rev</td>
<td>GAAACCGCGCGGCTGCCCGCCGGCCCTACAACAGCCGAG</td>
<td>M3/4 (K678P/R679P) fwd</td>
<td>CTGCGTGCGTGCGCGCTGCAGCAGAGCAGCCGAG</td>
</tr>
<tr>
<td>M3/4 (K678P/R679P) rev</td>
<td>CTGCGTGCGTGCGCGCTGCAGCAGAGCAGCCGAG</td>
<td>M3/4 (K678P/R679P) rev</td>
<td>CTGCGTGCGTGCGCGCTGCAGCAGAGCAGCCGAG</td>
</tr>
</tbody>
</table>

Appendix C