Identification of downstream target genes
and analysis of obesity-related variants
of the bHLH/PAS transcription factor
Single-minded 1

Anne Raimondo

B. Science (Molecular Biology) (Honours)

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

Discipline of Biochemistry
School of Molecular and Biomedical Science
University of Adelaide, Australia

June 2011
CONTENTS

CONTENTS ...................................................................................................................... iii

ABSTRACT ....................................................................................................................... ix

DECLARATION .............................................................................................................. xi

ACKNOWLEDGEMENTS ............................................................................................ xii

CHAPTER 1: INTRODUCTION ..................................................................................... 1

1.1. The bHLH/PAS family of transcription factors ................................................... 1
1.2. The PAS domain ................................................................................................. 2
1.3. Drosophila single-minded (sim) ........................................................................ 5
  1.3.1. sim expression and function ....................................................................... 5
  1.3.2. Control of sim expression and activity ...................................................... 6
1.4. The mammalian Sim genes: Sim1 and Sim2 ..................................................... 7
  1.4.1. Sim1 and Sim2 expression patterns .............................................................. 7
  1.4.1.1. Sim1 ........................................................................................................ 7
  1.4.1.2. Sim2 ........................................................................................................ 8
  1.4.2. Control of Sim1 and Sim2 expression and activity ...................................... 13
  1.4.2.1. Sim1 ....................................................................................................... 13
  1.4.2.2. Sim2 ....................................................................................................... 14
  1.4.3. Biochemical properties of SIM1 and SIM2 .............................................. 14
  1.4.4. Sim1 and Sim2 gene knockout/transgenic mouse studies .......................... 15
    1.4.4.1. Sim1 ...................................................................................................... 15
      1.4.4.1.1. Germline Sim1+/− mice .................................................................. 15
      1.4.4.1.2. Germline Sim1+/+ mice ................................................................. 17
      1.4.4.1.3. Conditional Sim1+/− and Sim1+/+ mice ...................................... 18
      1.4.4.1.4. SIM1 overexpressing mice ......................................................... 19
    1.4.4.2. Sim2 .................................................................................................... 20
      1.4.4.2.1. Sim2−/− and Sim2+/− mice ......................................................... 20
      1.4.4.2.2. Sim2 overexpressing mice ......................................................... 21
1.5. The mammalian general partner factors ARNT and ARNT2 .................... 21
  1.5.1. Arnt and Arnt2 expression patterns ............................................................ 22
    1.5.1.1. Arnt ................................................................................................... 22
    1.5.1.2. Arnt2 ................................................................................................. 23
  1.5.2. Biochemical properties of ARNT and ARNT2 ........................................ 23
1.5.3. *Arnt* and *Arnt2* gene knockout studies ............................................................. 24
1.5.3.1. *Arnt*−/− mice ................................................................................................ 24
1.5.3.2. *Arnt2*−/− mice .............................................................................................. 24

1.6. The molecular basis of SIM1 function ................................................................ 25
1.6.1. Hypothalamic control of food intake and weight regulation ......................... 26
    1.6.1.1. The leptin-melanocortin signalling pathway ............................................. 26
    1.6.1.2. *Sim1* and the leptin-melanocortin signalling pathway ....................... 29
1.6.2. Direct SIM1 target genes .................................................................................. 30
    1.6.2.1. Clues from studies of *Sim1* function and activity ................................. 31
    1.6.2.2. Clues from studies of related bHLH/PAS factors ................................. 33
        1.6.2.2.1. SIM ................................................................................................. 33
        1.6.2.2.2. SIM2 ............................................................................................... 34

1.7. *SIM1*: a novel contributor to severe obesity in humans .................................... 39
    1.7.1. The genetic contribution to human obesity ................................................. 39
    1.7.2. *SIM1* mutations in severely obese humans ............................................ 40

1.8. Project Aims and Approaches ............................................................................. 45
    1.8.1. Project Aims ............................................................................................... 45
    1.8.2. Project Approaches .................................................................................... 45

CHAPTER 2: MATERIALS AND METHODS .................................................................. 47

2.1. Abbreviations ...................................................................................................... 47
2.2. Materials ............................................................................................................ 50
    2.2.1. General Materials ...................................................................................... 50
    2.2.2. Chemicals and Reagents .......................................................................... 50
    2.2.3. Kits ............................................................................................................. 52
    2.2.4. Enzymes .................................................................................................... 52
    2.2.5. Antibodies ................................................................................................ 53
        2.2.5.1. Primary antibodies ........................................................................... 53
        2.2.5.2. Secondary antibodies ........................................................................ 53
    2.2.6. Bacterial strains ......................................................................................... 54
    2.2.7. Tissue culture cell lines .............................................................................. 54
    2.2.8. Solutions .................................................................................................... 54
        2.2.8.1. General lab solutions ........................................................................... 54
        2.2.8.2. Bacterial growth media ...................................................................... 56
            2.2.8.2.1. Basic media ............................................................................... 56
            2.2.8.2.2. Final concentrations of antibiotics and other additives .............. 56
        2.2.8.3. Tissue culture solutions ...................................................................... 57
            2.2.8.3.1. Basic media ............................................................................... 57
2.2.9. Plasmids
2.2.9.1. Cloning vectors
2.2.9.2. Expression vectors
2.2.9.2.1. Flp-In\textsuperscript{TM} T-Rex\textsuperscript{TM} vectors
2.2.9.3. pSUPER vectors
2.2.9.4. Luciferase reporter vectors
2.2.9.5. Lentiviral vectors
2.2.9.5.1. Gateway\textsuperscript{®} vectors
2.2.9.5.1.1. Donor and entry vectors
2.2.9.5.1.2. Destination and expression vectors
2.2.9.5.2. Lentivirus production vectors
2.2.10. Primers
2.2.10.1. Sequencing/cloning primers
2.2.10.2. RT-PCR primers
2.2.10.3. QPCR primers
2.2.10.4. EMSA oligos

2.3. Electronic Resources

2.4. Methods
2.4.1. Bacterial culture
2.4.1.1. Preparation of chemically competent DH5\textalpha
2.4.1.2. Heat shock transformation of chemically competent bacteria
2.4.1.3. Growth and maintenance of bacteria
2.4.2. DNA manipulation
2.4.2.1. Preparation of plasmid DNA
2.4.2.2. PCR
2.4.2.2.1. Using Taq polymerase
2.4.2.2.1.1. RT-PCR
2.4.2.2.1.2. Colony PCR
2.4.2.2.2. Using Pfu Turbo
2.4.2.2.3. Using Phusion
2.4.2.2.4. Overlap Extension PCR
2.4.2.2.5. Sequencing
2.4.2.2.6. QPCR
2.4.2.2.6.1. Experimental design
2.4.2.2.6.2. Data processing and statistical analysis
2.4.2.3. Phosphorylation and annealing of dsDNA oligos for subcloning
2.4.2.4. Generation of dsDNA probes for use in EMSAs
2.4.2.4.1. Annealing

2.2.8.3.2. Final concentrations of antibiotics
2.2.9. Plasmids
2.2.9.1. Cloning vectors
2.2.9.2. Expression vectors
2.2.9.2.1. Flp-In\textsuperscript{TM} T-Rex\textsuperscript{TM} vectors
2.2.9.3. pSUPER vectors
2.2.9.4. Luciferase reporter vectors
2.2.9.5. Lentiviral vectors
2.2.9.5.1. Gateway\textsuperscript{®} vectors
2.2.9.5.1.1. Donor and entry vectors
2.2.9.5.1.2. Destination and expression vectors
2.2.9.5.2. Lentivirus production vectors
2.2.10. Primers
2.2.10.1. Sequencing/cloning primers
2.2.10.2. RT-PCR primers
2.2.10.3. QPCR primers
2.2.10.4. EMSA oligos

2.3. Electronic Resources

2.4. Methods
2.4.1. Bacterial culture
2.4.1.1. Preparation of chemically competent DH5\textalpha
2.4.1.2. Heat shock transformation of chemically competent bacteria
2.4.1.3. Growth and maintenance of bacteria
2.4.2. DNA manipulation
2.4.2.1. Preparation of plasmid DNA
2.4.2.2. PCR
2.4.2.2.1. Using Taq polymerase
2.4.2.2.1.1. RT-PCR
2.4.2.2.1.2. Colony PCR
2.4.2.2.2. Using Pfu Turbo
2.4.2.2.3. Using Phusion
2.4.2.2.4. Overlap Extension PCR
2.4.2.2.5. Sequencing
2.4.2.2.6. QPCR
2.4.2.2.6.1. Experimental design
2.4.2.2.6.2. Data processing and statistical analysis
2.4.2.3. Phosphorylation and annealing of dsDNA oligos for subcloning
2.4.2.4. Generation of dsDNA probes for use in EMSAs
2.4.2.4.1. Annealing
2.4.2.4.2. $^{32}$P labelling ................................................................................................................ 80
2.4.2.5. Restriction enzyme digestion ...................................................................................... 80
2.4.2.6. Dephosphorylation of plasmid DNA ......................................................................... 80
2.4.2.7. Generation of blunt-ended DNA using Klenow Large Fragment ......................... 80
2.4.2.8. Ligations ...................................................................................................................... 81
  2.4.2.8.1. Ligation of purified DNA fragments ........................................................................ 81
  2.4.2.8.2. Ligation of phosphorylated and annealed dsDNA oligos ........................................ 81
  2.4.2.8.3. pGEM-T Easy cloning ............................................................................................ 82
    2.4.2.8.3.1. A-tailing .............................................................................................................. 82
    2.4.2.8.3.2. Ligation into pGEM-T Easy ............................................................................... 82
  2.4.2.8.4. Gateway® cloning .................................................................................................. 82
    2.4.2.8.4.1. Generation of entry vectors ................................................................................. 82
    2.4.2.8.4.2. Generation of expression vectors ....................................................................... 82
2.4.3. RNA manipulation ......................................................................................................... 83
  2.4.3.1. Isolation of total RNA ............................................................................................... 83
  2.4.3.2. cDNA synthesis from RNA ...................................................................................... 83
  2.4.3.3. Microarray studies ...................................................................................................... 84
    2.4.3.3.1. Sample preparation and validation ......................................................................... 84
    2.4.3.3.2. Microarray and statistical analyses ........................................................................ 84
2.4.4. Mammalian cell culture and protein analysis .............................................................. 85
  2.4.4.1. Thawing frozen cell line stocks ............................................................................... 85
  2.4.4.2. Freezing cell line stocks .......................................................................................... 85
  2.4.4.3. Production of stable cell lines .................................................................................. 85
    2.4.4.3.1. 293 Flp-In® T-Rex® lines ....................................................................................... 85
    2.4.4.3.2. Production of stable cell lines via lentiviral infection ........................................... 86
  2.4.4.4. Routine maintenance of cell lines ............................................................................... 88
  2.4.4.5. Transient transfections .............................................................................................. 89
  2.4.4.6. Harvesting cells with TEN buffer ............................................................................ 89
  2.4.4.7. Whole cell protein extract preparation and quantification ........................................... 89
  2.4.4.8. Detergent-free nuclear and cytoplasmic protein extract preparation and quantification .......................................................................................................................... 90
  2.4.4.9. In vitro coupled transcription/translation of mammalian proteins using $^{35}$S-Met ......................................................................................................................................................... 90
  2.4.4.10. EMSA .................................................................................................................... 91
    2.4.4.10.1. FAM labelling experiments .................................................................................. 91
    2.4.4.10.2. $^{32}$P labelling experiments .................................................................................. 91
  2.4.4.11. Reporter gene studies in 24 well tray format ............................................................ 92
    2.4.4.11.1. Transfection ....................................................................................................... 92
    2.4.4.11.2. Dual luciferase activity assay .............................................................................. 93
    2.4.4.11.3. Statistical analysis .............................................................................................. 94
2.4.4.12. Western blotting

2.4.4.12.1. Protein transfer to nitrocellulose membrane

2.4.4.12.2. Ponceau staining

2.4.4.12.3. Blocking and antibody incubations

2.4.4.12.4. Chemiluminescent detection of proteins

2.4.4.13. Immunohistochemistry of monolayer cultures

2.4.4.13.1. Fixing and permeabilisation

2.4.4.13.2. Blocking and antibody incubations

2.4.4.13.3. Nuclear staining, mounting and photographing

2.4.4.14. Immunoprecipitation

2.4.5. Gel electrophoresis

2.4.5.1. Agarose gel electrophoresis

2.4.5.2. Non-denaturing PAGE

2.4.5.3. Denaturing SDS-PAGE

CHAPTER 3: RESULTS

Identification and characterisation of obesity-associated SIM1 variants

3.1. Screening for disease-associated SIM1 variants

3.1.1. The UK study: a population based screen for SIM1 variants associated with severe, early onset obesity

3.1.2. The French study: a familial screen for SIM1 variants associated with a Prader-Willi like phenotype and severe obesity

3.2. Physiological data

3.3. Functional characterisation of disease-associated SIM1 variants

3.3.1. Several variants display altered activity in a CME-driven reporter gene assay system

3.3.2. SIM1 T292A displays an altered subcellular localisation pattern

3.3.3. SIM1 T292A is reduced in its ability to dimerise with ARNT and ARNT2

3.3.4. Attempts to develop a SIM1 DNA binding assay

3.3.4.1. Experimental approach

3.3.4.2. Attempts to optimise an EMSA protocol using probe sequences derived from the regulatory regions of several SIM2 and HIF-1α direct target genes

3.3.4.3. Further optimisation of probe sequences and binding conditions

CHAPTER 4: RESULTS

Identification of novel downstream SIM1 target genes

4.1. Choice of experimental strategy

4.2. 293 T-Rex cells can be successfully induced to express SIM1 and ARNT2

4.3. Microarray experiments
4.3.1. Experimental design ................................................................. 153
4.3.2. Data processing, analysis and initial verification ................. 154

4.4. Validation of differential expression in independent samples by QPCR ...... 162

4.5. Analysis of differential expression in independent cell systems ........... 167
   4.5.1. Analysis of differential expression upon endogenous SIM1 knockdown in 293T cells ............................................................ 167
   4.5.2. Analysis of differential expression in hypothalamus derived cell lines ...... 168
      4.5.2.1. Marker gene analysis of immortalised, embryonic hypothalamus derived cell lines N4, N7 and N39 ................................. 173
      4.5.2.2. Analysis of differential expression in N39 cells stably overexpressing SIM1 and ARNT2 ....................................................... 174
      4.5.2.3. Additional analyses of candidate hypothalamic downstream target genes ................................................................. 183
      4.5.2.4. Analysis of differential expression in N4, N7 and N39 cells stably overexpressing SIM1 ........................................................... 183
      4.5.2.5. Analysis of Myom2 transcript levels upon Sim1 knockdown in N39 cells ................................................................. 189

4.6. Bioinformatic analysis of the upstream regulatory regions of the Myom2 and Ss genes ............................................. 190

CHAPTER 5: DISCUSSION .................................................................. 195

APPENDIX ......................................................................................... 211

REFERENCES .................................................................................... 215
Single-minded 1 (SIM1) is a basic Helix-Loop-Helix/PER-ARNT-SIM (bHLH/PAS) transcription factor essential for survival in mice. The early post-natal lethality exhibited by Sim1−/− mice is believed to be the consequence of severely compromised hypothalamus development, although the contribution of reduced SIM1 signalling in the numerous other tissues in which it is expressed has never been formally investigated. The presence of a single Sim1 allele is sufficient to avoid this perinatal lethality, and instead confers an early onset, hyperphagic obesity phenotype, potentially via disruption of critical intracellular signalling pathways that are activated in Sim1-expressing hypothalamic neurons in response to food intake. Similar correlations between reduced SIM1 gene dosage and severe obesity have also been documented in humans. Alterations in SIM1 expression and/or function therefore have important implications in health and disease, and warrant a detailed investigation into the downstream target genes and regulatory behaviours of this critical transcription factor, which are thus far almost entirely lacking in the literature.

The studies presented in this thesis describe a twofold approach to dissecting the gene regulatory properties of the SIM1 protein. Firstly, we optimised and performed a range of functional assays, including a cell-based luciferase reporter gene assay, a subcellular localisation assay, a co-immunoprecipitation assay, and an electrophoretic mobility shift assay, which were designed to assess the contribution of nineteen unique point mutations within the SIM1 protein sequence to altered SIM1 expression and behaviour. These nineteen mutations were identified in multiple cohorts of severely obese humans, and therefore represent potentially pathogenic alterations in the SIM1 sequence. Indeed, we observed a significant loss of function for many of these variants in luciferase reporter gene assays relative to wild type SIM1. The severe loss of function observed for one of these variants, SIM1 T292A, could be further attributed to altered subcellular localisation, thus impacting on its ability to form a stable heterodimer with ARNT2 in co-immunoprecipitation experiments. Secondly, we performed microarray studies on cultured kidney-derived cells inducibly overexpressing Myc-tagged SIM1 and its obligate partner factor ARNT2, and subsequently identified several genes that selectively responded to SIM/ARNT2 overexpression in this context. Further validation in hypothalamus-derived cultured cells highlighted Myomesin 2 (Myom2) as a potentially
genuine downstream SIM1 target gene in both kidney and hypothalamus. We also present data that are the first to indicate Somatostatin (Ss) as a hypothalamic target gene regulated by SIM1 in a cell-autonomous manner.

These data are among the first to dissect the downstream target genes and regulatory properties of the SIM1 protein, and therefore make an important contribution to our understanding of the molecular basis to the hyperphagic obesity exhibited by Sim1+/− mice. They are also the first to link reduced activities of mis-sense mutations in the SIM1 coding sequence to increased weight gain in humans, and give further credence to the possibility that SIM1 represents a novel genetic contributor to obesity disorders in the wider population. This knowledge may inform future attempts to develop therapies for obese phenotypes in humans, and broaden our understanding of the molecular events that underpin Sim1-mediated survival and maintenance of homeostasis.
DECLARATION

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Anne Raimondo, and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Anne Raimondo

June 2011
ACKNOWLEDGEMENTS

I have benefited from the intelligence, insight, enthusiasm and friendship of countless people during my time at this University. I would particularly like to acknowledge the following people:

Associate Professor Murray Whitelaw – an unfailingly supportive, enthusiastic and encouraging supervisor who has given me an extraordinary amount of freedom to develop independently as a scientist. I am very appreciative of his co-operative and stimulating approach to scientific research and collaboration, and value his contribution to my scientific development over the last six years.

Dr. Anne Chapman-Smith – a source of continued inspiration, from whose fastidious, rational and methodical approach to science I have constantly benefited, and for whose contribution to my professional and personal development I am extremely grateful.

Dr. Simon Koblar and Dr. Dan Peet – official and unofficial co-supervisors respectively, who have been so generous in their contributions to my work, and who have always been available for advice and support.

Colleen Bindloss – a wonderful scientist and human being, without whose assistance this thesis would not have been completed.

Dr. Shwetha Ramachandrappa – a truly inspiring person whose friendship and support I value highly, and whose hard work and determination have made an invaluable contribution to this work. I hope you enjoy this, the companion volume to "Planes, Trains and Automobiles".

I am indebted to Mark Van der Hoek and Rosalie Kenyon at the Adelaide Microarray Facility, and Dr. Steven Pederson at the Women's and Children's Hospital, for their contributions to the microarray studies presented in this thesis, and to Associate Professor Gary Glonek for advice on statistical analysis.

I am grateful to our collaborators Dr. Sadaf Farooqi, Dr. Phillippe Froguel and Dr. Fanny Stutzmann, and all other scientists at the Metabolic Research Laboratories at the University of Cambridge and the Institut de Biologie de Lille at the Institut Pasteur de
Lille, for giving me the opportunity to contribute to their projects, and their patience in waiting for results.

To the many other members of the Whitelaw and Peet labs that I have known over the years, particularly Jodi, Susi, Jo, Fiona, Alix, Scott, Margo, Andrew, Dave, Matt, Adrienne, Veronica, Anthony, Cameron, Sarah, Bec, Karolina, Rachel, Sam, Teresa, Natalia and Sarah – I value your friendship, and the contributions you have made to my personal and professional development, very much.

To the professional staff who make a critical contribution to the daily maintenance of this building and its facilities. I am indebted particularly to John Mackrill for his skill in maintaining the tissue culture facility in the Department of Biochemistry, upon which I was absolutely reliant for my studies, and Serge Volgin and Shirley Coad for their dedication in running the Store.

To all other members of the Thomas and Jensen labs and the Departments of Biochemistry and Genetics, amongst whom I count many friends and admirable scientists.

I must also acknowledge the assistance of the Healthy Development Adelaide Research Cluster in financially supporting my studies.

Finally, deepest thanks must to go my parents, brothers and sister, extended family and friends for their unfailing support and encouragement, which I am not always very good at acknowledging but appreciate nonetheless.