

Detailed characterisation of the Opitz syndrome protein MID1, its homologues and molecular interactions

By

Kieran M. Short BSc. (Hons.)



**THE UNIVERSITY
OF ADELAIDE
AUSTRALIA**

**The University of Adelaide
School of Molecular and Biomedical Science
Discipline of Genetics**

October, 2006

TABLE OF CONTENTS

Abstract.....	v
Declaration.....	viii
Acknowledgements.....	ix
Author Contribution Statements	x
Chapter One: Introduction and Literature Review	1
1.1 Human Congenital Disorders.....	1
1.1.1 <i>X-linked Disorders</i>	1
1.1.1.1 X chromosome anomalies	3
1.1.1.2 Opitz syndrome, an X-linked disorder	5
1.2 Opitz syndrome.....	5
1.2.1 <i>Delineation of Opitz syndrome</i>	6
1.2.1.1 BBB syndrome	6
1.2.1.2 G syndrome	7
1.2.1.3 Opitz G/BBB syndrome	8
1.2.2 <i>The Opitz syndrome phenotype</i>	9
1.2.2.1 The potential causes of the tissue malformations underlying Opitz syndrome	9
1.2.3 <i>The genetics of Opitz syndrome</i>	15
1.2.3.1a Autosomal OS	17
1.2.3.1b 22q11.2 critical region	18
1.2.3.1c Human disorders within the 22q11.2 TDR	19
1.2.3.2 X-linked OS	20
1.2.4 <i>Other syndromes with common phenotypic features</i>	21
1.3 <i>MID1</i>	22
1.3.1 <i>Genomic location and organisation</i>	22
1.3.1.1 MID1 promoter	23
1.3.2 <i>MID1 mRNA expression pattern</i>	24
1.3.3 <i>MID1 mutations in Opitz syndrome</i>	25
1.4 MID1 protein	27
1.4.1 <i>Homologous domains within MID1</i>	28

1.4.1.1 RING Finger.....	28
1.4.1.2 B-box2.....	29
1.4.1.3 B-box1.....	30
1.4.1.4 Coiled coil.....	31
1.4.1.5 Fibronectin type III domain.....	32
1.4.1.6 B30.2-like domain.....	33
1.4.2 Homologous RBCC motif-containing proteins.....	34
1.4.3 MID1 cellular localisation.....	37
1.4.3.1 Mutations in MID1 alter its cellular localisation.....	38
1.4.4 MID2 - a homologue of MID1.....	39
1.4.4.1 The evolutionary relationship between MID1 and MID2.....	40
1.4.4.2 Subcellular localisation of MID2.....	40
1.4.5 MID1 protein-protein interaction.....	41
1.4.5.1 Yeast TAP42 and TOR.....	43
1.4.5.2 Alpha 4 and PP2Ac.....	45
1.4.5.3 Ubiquitin E3 Ligase activity of MID1 and its target.....	49
1.4.5.4 Other MID1 interacting proteins.....	52
1.5 Molecular pathways and Opitz syndrome.....	52
1.5.1 Epithelial-mesenchymal transition (EMT) and Apoptosis.....	53
1.6 Significance and aims of the project.....	54
Chapter Two: Short <i>et al</i> 2002.....	56
Chapter Three: Short <i>et al</i> 2002 to Short and Cox 2006.....	70
3.1 Significant findings relating to MID1 and Alpha 4.....	70
3.1.1 Liu <i>et al.</i> , 2001.....	70
3.1.2 Trokenbacher <i>et al.</i> , 2001.....	71
3.1.3 Implications of the findings presented by Liu 2001, Trokenbacher 2001 and Short 2002.....	71
3.1.4 Alpha 4 as a candidate for FG syndrome.....	73
3.1.5 A working model for MID1-Alpha 4 function.....	74
3.2 Moving forward to Short and Cox 2006.....	75
3.2.1 Approach and methodological detail.....	75

Chapter Four: Short and Cox 2006	78
Chapter Five: Discussion and Conclusions	89
5.1 Review of Project Aims and Significance	89
5.2 Identification of the C-I subfamily and similarities in structure/function	91
5.3 Significance of the MID1 Alpha 4 interaction.....	96
5.4 Epithelial-mesenchymal transition (EMT)	97
5.5 Apoptosis	104
5.6 Model for apoptosis-based pathogenesis of Opitz syndrome	107
5.7 Future Directions	113
Bibliography	114

Appendices

1. Perry *et al.*, 1999
2. Graham *et al.*, 2003
3. Massiah *et al.*, 2006

Abstract

Opitz G/BBB syndrome (OS) is a genetically heterogeneous congenital disorder with both X-linked and autosomal forms. Although there are established diagnostic criteria, the phenotypic variability among patients can be considerable. Most commonly, OS patients present with craniofacial, laryngotracheoesophageal, cardiac, and genital anomalies. OS has been linked to Xp22.3 and 22q11.2; and additional patients with unique chromosomal anomalies have allowed the tentative assignment of three other autosomal loci at 14q11.2, 13q32.3ter and 5p12-13. Yet despite differing genetic bases, the clinical presentation in either form is essentially indistinguishable, despite marked variability in phenotypic presentation even among related family members.

Our laboratory has previously shown that approximately 50% of patients with a diagnosis of OS harbour mutations in the X-linked *MID1* gene. *MID1* encodes a 667 amino acid RING finger, B-box, coiled-coil (RBCC) microtubule binding protein that is expressed widely throughout embryonic development and congruent with tissues affected in OS. A highly related protein, *MID2*, is the closest homologue to *MID1* and has been found to similarly associate with the microtubule network either as homomultimers, or as heteromultimers with *MID1*. In this thesis, several approaches were undertaken in order to better understand the sub-cellular and molecular aspects of OS. Firstly, a search for *MID1*-interacting proteins was undertaken, and secondly an analysis of homologous proteins was completed with a view to providing insight based on the function of similar proteins.

To identify protein partners of *MID1* that might be components of the large MID complexes, a yeast two-hybrid screen was performed. Using this approach, Alpha 4, a regulatory subunit of PP2-type phosphatases was identified as a strong interactor of *MID1* and additionally through separate testing, *MID2*. Cell localisation studies showed that both *MID1* and *MID2* tethered Alpha 4 to the microtubules. Co-expression of domain-deleted


and mutant forms of MID1 with Alpha 4 localised the interaction domain of MID1 to the B-boxes, which was confirmed by yeast two-hybrid assays. In addition to MID1 and MID2 interacting with Alpha 4, it was also found that MID1 and MID2 interacted with one-another through their coiled-coil domains. This further implicated MID2 as a potential modifier of the OS phenotype. Therefore, in order to better understand the basic functional aspects of MID1 and MID2, a detailed search for other homologues was carried out.

The RBCC tripartite (TRIM) motif skewed protein BLAST searches for N-terminal matches for MID1/MID2 homologues, so combinations of N-terminal and C-terminal domains were also used to identify structurally related proteins. BLAST and PSIBLAST searches identified the homologous proteins, TRIM9 and TRIM36 as well as partial transcripts of two new proteins TRIFIC and TNL (TRIM9-like). Subsequently, full sequences were determined for both of these genes. This group of six related RBCC proteins were designated as the RBCC C-I subfamily. To further identify sequence identities within the C-I subfamily and the rest of the RBCC protein superfamily, a relatively novel technique using HMMER was employed. This showed that the most highly scoring region across the C-I subfamily was not a known domain but a region between the coiled-coil and Fibronectin type-III domains. This new region, called the COS box, was shown to be restricted to a select group of RBCC and non-RBCC proteins, all which associate with microtubules. Mutation of particular amino acids in the COS box completely abolished microtubule association, and addition of a MID1 C-terminal fragment containing the COS box to a non-microtubule associating RBCC protein directed the hybrid protein to microtubules. These findings completely change the previous understanding that the coiled-coil or C-terminus of MID1 and other RBCC proteins were responsible for microtubule association.

The work presented in this thesis shows that normal association of MID1 to microtubules is dependent on the COS box. In addition, the B-box domain is necessary for tethering Alpha 4, a protein phosphatase 2A regulatory protein, to the microtubules. The homology of MID1 and MID2, which extends functionally to homo and hetero-dimerisation, microtubule association, and interaction with Alpha 4, has led to the suggestion of potential heterogeneity in Opitz syndrome. In this light, Alpha 4 was subsequently shown to be mutated in a phenotypically similar syndrome, further supporting the importance of the molecular interaction between these proteins. Potential molecular pathways in which these proteins act are discussed, as are the effects of disruption of these pathways, and the impact such changes may have on the development of key tissues affected in the diseases.

The work contained in this thesis contains no material that has been accepted for the award of any other degree or diploma in any University or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published except where due reference has been made.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and copying.


Kieran M. Short, BSc. (Hons.)

Acknowledgements

I am eternally grateful to my supervisor Timothy Cox, without whom I would not have had the opportunity to partake in The University of Adelaide PhD programme. Tim, your enthusiasm for science and thirst for knowledge was always a constant source of inspiration for me, and the trials and tribulations of grant successes and failures have introduced me to the sometimes harsh world of scientific research. Thank you very much for always being there when I needed advice or help. The time I have spent in your lab, I feel, has seen me transform from a rather immature unconfident scientist, to one that has the ability to drive his own research, critically assess other research, and importantly publish.

Thank you to all the co-authors of the papers with whom I have published. Scientific research with a Genetic or Molecular approach is extremely time consuming, and such publications have a long lead-time. Teamwork makes such endeavours happen faster.

When I joined the lab, I was unaware of the future adventures that would lie in my path. One of those has been the interaction with our co-workers, including people from Germany, France, Philippines, Malaysia and China. Learning about other cultures and religions without travelling to those countries has been wonderful. There have been many people through the laboratory during my stay and I have had friendships with many of you, but the people who I am most indebted to for various reasons are:

Dr. Quenten Schwarz. Quenten, you are one of my best friends, and were my 'best man' at my wedding. You love to chat about science, and were always there at the pub when we needed a beer during our PhD's to chat about work or extracurricular activities.

Dr. Blair Hopwood. Blair, you were like a 'big brother', a wonderful source of knowledge for all things worldly, and a 'grand master' of Molecular Biology. Much of my confidence at the bench comes from your teachings and observing your extraordinary skill. Thank you Blair!

Dr. Jörg Drenckhahn. While my PhD did take longer than expected for various reasons, Jörg, you are one reason that I am glad my research took longer than expected. You joining the lab during the twilight of my research was fortuitous, and you became a great house mate and were always interested in seeing more of Australia while you were here. Your approach to research was a great inspiration for me, it was different - decidedly *German*. You certainly taught me a few things about *serious* organisation.

I also thank all the support staff that have kept the departments running smoothly. Thank you especially to Paul Moir for your help with building gadgets (without which my Yeast two-hybrid would have been impossible), IT support, and your ongoing friendship.

The preparation of this thesis was aided by Liza Cox, thank you very much for your help. Tim, thank you also for the detail with which you have criticised my work. Critical review is the only way we can learn! Belinda Thomas, thank you greatly for helping me once more with the finishing touches.

Thank you to my wonderful, beautiful, loving family. My parents, Jenny and Rob, are a constant inspiration to me. They provide unconditional love, support and without their encouragement I would not have gotten to this level of academic achievement. I also thank the rest of my family, Liam and Emma, Brendan, Michelle, James and Taylor and my extended 'Thomas' family, Mike and Chris, Jade and Trevor, Rebecca and Dave for their encouragement. Dandy, Nan, Granny and Grandpa, thank you eternally for your love and support.

I also thank my beautiful wife, Dr. Belinda Thomas. Thank you for your endless patience, encouragement, support and love. You have helped me emotionally and kept me motivated through some tough times with me contracting arthritis during this PhD, and you have also helped me scientifically on many occasions. I am forever grateful for you being in my life. Thank you Bel.

Statement of Authorship:

MID1 AND MID2 HOMO- AND HETERODIMERISE TO TETHER THE RAPAMYCIN-SENSITIVE PP2A REGULATORY SUBUNIT, ALPHA 4, TO MICROTUBULES: IMPLICATIONS FOR THE CLINICAL VARIABILITY OF X-LINKED OPITZ GBBB SYNDROME AND OTHER DEVELOPMENTAL DISORDERS.

BioMed Central Cell Biology: 2002, 3, 1.

KIERAN M SHORT (Candidate)

Conception and design of experiments. Performed: yeast two-hybrid screen of MID1 and identified interactors, immunofluorescence which identified interaction domains between MID1 and Alpha 4, yeast two-hybrid interaction testing to complement the immunofluorescence work, parts of the construct cloning for use in the experiments, computational analysis and phosphorylation site prediction. Interpretation and evaluation of data. Writing of the manuscript, evaluation and preparation of all figures for publication, and final approval of completed manuscript.

Signed: .

.....

Date: 9/2/2006...

BLAIR HOPWOOD

Conception and design of experiments. Performed: yeast two-hybrid work for publication, western and in vitro protein experiments including the phosphorylation study, construct cloning for experiments (approximately 90%), and domain subtraction cloning and yeast two-hybrid identification of MID1/Alpha 4 binding domain by positive interaction, and interpretation/data evaluation of these experiments. Manuscript critical evaluation and revision. Final approval of completed manuscript.

Signed:

.....

Date: 15/2/06.....

Statement of Authorship (*BioMed Central Cell Biology: 2002, 3, 1*) continued:

ZOU YI

MID2 yeast two hybrid screen (not specifically outlined in detail in this manuscript) which also identified Alpha 4 as an interactor. This result was used as tertiary confirmation to the main data presented in the manuscript. Final approval of completed manuscript.

Signed:

.....

Date: 11/2/06.....

TIMOTHY C COX

Conception and design of experiments. Supervised development of work, helped in data interpretation, manuscript critical evaluation, editing and revision. Corresponding author. Final approval of completed manuscript.

Signed:

Date: 9/2/06.....

Statement of Authorship:

SUB-CLASSIFICATION OF THE RBCC/TRIM SUPERFAMILY REVEALS A NOVEL MOTIF NECESSARY FOR MICROTUBULE BINDING.

The Journal of Biological Chemistry: 2006. Published online ahead of print Jan 23.

KIERAN M SHORT (Candidate)

Conception and design of experiments. Performed all experiments presented in the manuscript, including yeast two-hybrid, immunofluorescence, western, co-immunoprecipitation, cloning, directed mutagenesis, identification and determination of new gene sequences and structures. Also performed HMMer analysis, Perl script debugging (for solaris, i386 and Darwin (MacG5) platforms) and modification for GenBank searching, online bioinformatic analyses and generation of multiple sequence alignments. Evaluated and interpreted data, devised experiments, wrote, critically evaluated, and approved the final manuscript.

Signed: ..

.....

Date: 24/2/06.....

TIMOTHY C COX

Supervised development of the work, advice of design of some experiments, manuscript critical evaluation, editing and revision. Corresponding author. Final approval of completed manuscript.

Signed:

.....

Date: 24/2/06.....

Chapter One

Introduction and Literature Review

Chapter One: Introduction and Literature Review

1.1 Human Congenital Disorders

Inherited disorders are one of the most prevalent problems affecting newborns. Congenital abnormalities resulted in 493,000 deaths in the World Health Organisation member states in 2002, which equates to almost 1 in 100 births (World Health Organisation, 2004 statistics). However, this statistic does not include the proportion of newborns burdened with inherited disorders at birth that survive. For these individuals, the health, psychological, and financial impact on the child, family and hospital system is significant. These early life complications are confounded by the general lack of knowledge of the processes underlying the disease symptoms in the patient, hindering effective treatment of many disorders of this kind. To further understand and work towards effective health care for such individuals, it is necessary to understand the basic processes that occur at the molecular level within the cells of these patients. A basic understanding of the nature and progression of disease during embryonic development should assist management of patient care, improve genetic counselling services to families, provide an understanding of the interactions between genotype and environment which account for the disease phenotype, and eventually the hope of minimisation or prevention of malformations through early intervention strategies.

1.1.1 X-linked Disorders

Genetic disorders of X-linked origin constitute the highest proportion of congenital diseases attributed to a single chromosome (954 of 17088 total OMIM entries, September 26 2006) (<http://www.ncbi.nih.gov/Omim/mimstats.html>). A unique feature of X-linked disorders is that the severity of the phenotype is dependent on the gender of the individual

with the mutation. The hemizyosity of the X chromosome effectively guarantees that males are significantly affected by any X-linked mutation since no normal allele is available for functional gene expression. In many respects females have a similar situation because only one copy of most X-linked genes is expressed in somatic cells. This occurs as a result of the process of X chromosome inactivation, whereby one of the two X chromosomes is transcriptionally silenced to ensure an equivalent gene dosage to males. However, X-inactivation is, in most cases, a random process, such that any particular female tissue is mosaic for X-linked gene expression. Specifically, some cells will express the alleles from one X chromosome, while other cells (which might have the other X chromosome remaining active), will express the other alleles. This process ensures that the effect of most X-linked mutations in females will be dependent on which cells express a mutated copy of the gene (i.e. are affected), and which cells have the mutated X inactivated (i.e. cells expressing the normal copy of the gene). The process of X-inactivation, however, can appear non-random or skewed in several disorders, with one X chromosome inactivated in preference to the other (Hedera and Gorski, 2003). This phenomenon alters the phenotype of the patient, making it either more severe (akin to male presentation, if the normal X chromosome is preferentially inactivated) or less severe (if the mutated X chromosome is preferentially inactivated). Independent of any mutation or deletion actively causing skewed inactivation, skewing has also been associated with increasing age of the mother at the time of conception. As such, maternal age may affect the severity of the phenotype of X-linked diseases (Hatakeyama *et al.*, 2004). However, in most cases the phenotypic variability in females carrying X-linked disease-causing mutations does indeed suggest that X-inactivation is essentially a random process. In this state, females with X-linked congenital disorders generally present with a mosaic pattern of affected cells nestled amongst cells which are normal, reflecting the clonal origin of the cells themselves. The

patterning of affected and normal cells may allow the tissue to compensate for dysfunction caused by the affected cells or maintain sufficient function, providing that the proportion (or positioning) of such cells does not adversely affect function of the tissue or organ.

1.1.1.1 X chromosome anomalies

The diagnosis of individuals presenting with particular phenotypic features characteristic of congenital disease can sometimes be assisted through the observation of karyotype, which may identify deleted and/or rearranged chromosomal segments. The short arm of the X chromosome (Xp, Figure 1.1) is one such region of the genome that shows a higher than usual frequency of deletions, segmental rearrangements and translocations that can be readily detected by standard cytogenetic techniques. In fact, it has been reported that this area of the X chromosome is 20 fold more likely to undergo recombination (and hence abnormalities of recombination: rearrangement, translocation or deletion) compared with the genomic average (Filatov and Gerrard, 2003). This segmental instability is due to the position of a region of identity with the Y chromosome. This region is known as the pseudo-autosomal region and is used for pairing of the sex chromosomes during male meioses. A large number of locally repeated sequences present on the X chromosome are also believed to contribute to genomic rearrangement (Lutskiy *et al.*, 2002).

The study of human diseases resulting from such X chromosome mutations has been assisted by positional mapping of chromosomal breakpoints. The disorders caused by genomic rearrangement or deletion of regions from the telomeric end of the short arm of the X chromosome are phenotypically distinct in males and generally correspond to the size of the anomaly. Such correlations have also been critical in determining disease loci within the region by physical mapping (Ballabio and Andria, 1992).

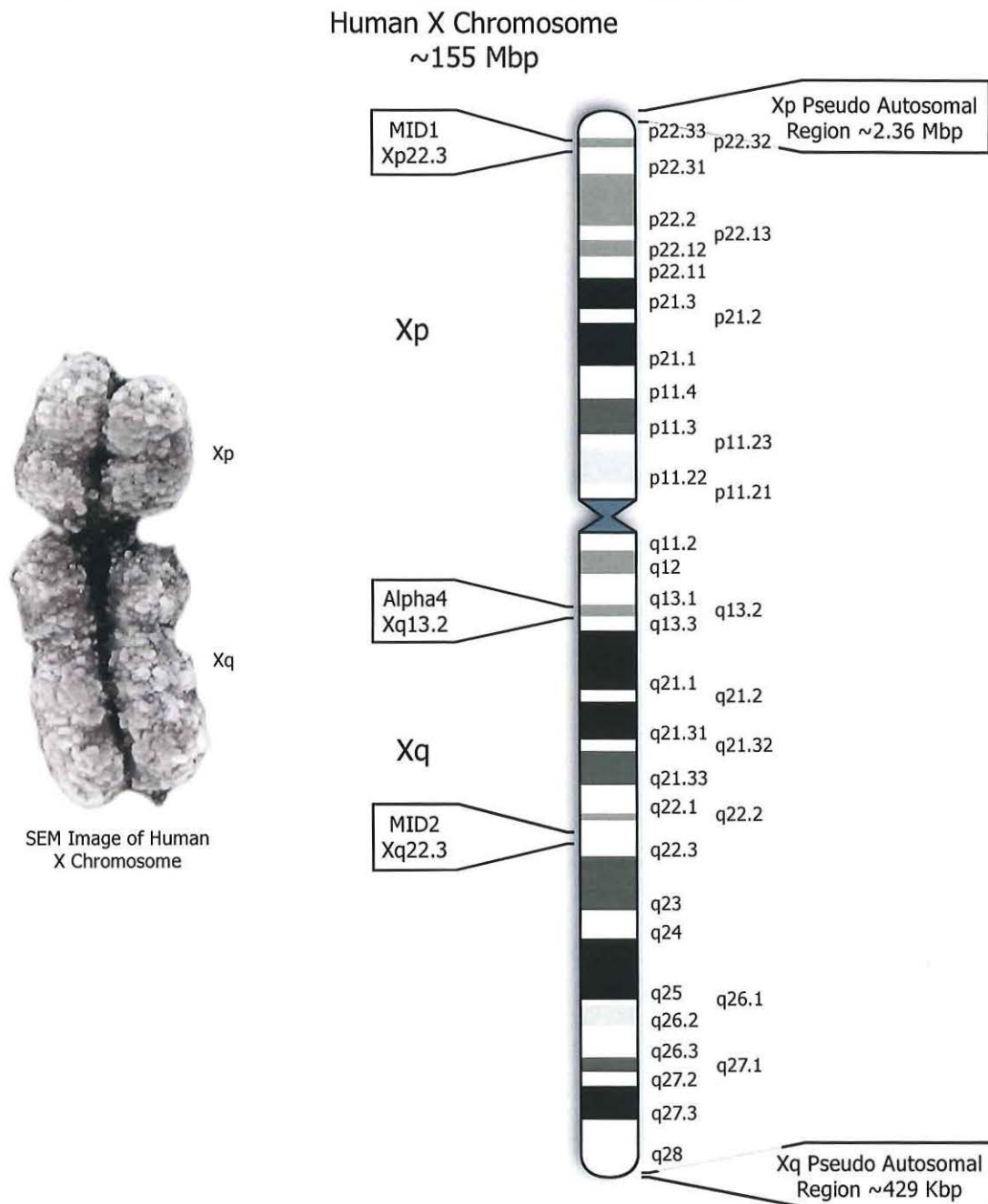


Figure 1.1 The Human X Chromosome.

The X Chromosome, seen from the Scanning Electron Microscopy (SEM) image (*left*), consists of a short ‘p’ arm and a long ‘q’ arm. Genes of interest in this project are located at various positions along the X chromosome and are marked on the ideogram of a G-banded X chromosome (*right*), indicating their relative positions. The pseudoautosomal regions (PAR), which sit at the tips of Xp and Xq, are the points of Y chromosome pairing during male meiosis. Note the relative close proximity between *MID1* and the Xp PAR, a region that is 20 fold more likely than the genomic average to undergo genetic recombination, thought to be one cause of genetic instability in the area.

1.1.1.2 Opitz syndrome, an X-linked disorder

Opitz syndrome¹ (OS) is a particularly interesting congenital disorder with a complex genetic history. The main causative gene for OS was first pinpointed as a result of detection and positioning of a specific X chromosome rearrangement in a large kindred presenting with the OS phenotype. Affected members of this family carried a pericentric inversion of the X chromosome: inv(x)(p22.3q26) (Verloes *et al.*, 1995). However, Opitz syndrome has also been associated with autosomal inheritance patterns and a number of different chromosome aberrations which have provided supporting evidence for numerous autosomal OS loci (Fryburg *et al.*, 1996; Kurczynski *et al.*, 1998; Leichtman *et al.*, 1991; McDonald-McGinn *et al.*, 1995; Urioste *et al.*, 1995). These data suggest that although OS is classically a monogenic disorder, there may be numerous genes that when defective can give rise to a similar or indistinguishable clinical phenotype or contribute to the marked phenotypic variability seen in this disorder. No specific mutations in autosomal genes have yet been identified in OS patients, however the X-linked gene mutated in this syndrome has been identified, and forms the foundation for the work in this thesis.

1.2 Opitz syndrome

OS is an inherited genetic disorder typically associated with ventral midline abnormalities that arise during the process of embryonic development. The severity of OS ranges from perinatal lethality (stillborn, neonatal or infantile death), to less severe forms with treatment sometimes vital depending on the type of malformations (and severity of each) present, to minimal clinical features. The high level of variability in severity of phenotypic presentation of OS (Opitz, 1987) makes estimation of the true prevalence

¹ In various papers, reports and texts, Opitz syndrome is also referred to as hypertelorism-hypospadias syndrome, Opitz oculo-genital-laryngeal syndrome, telecanthus-hypospadias syndrome, Opitz-Friás syndrome, hypospadias-dysphagia syndrome, Opitz G syndrome, BBB syndrome and Opitz G/BBB syndrome.

within the general population difficult. The exact nature of the aberrant tissue morphogenesis is not known, however the malformations provide insight into the possible developmental mechanisms behind the phenotypes observed.

1.2.1 Delineation of Opitz syndrome

Three reports of what were initially classified as three separate disorders were published in the second issue of the fifth volume of Birth Defects: Original Article Series (Christian *et al.*, 1969; Opitz *et al.*, 1969b; Opitz *et al.*, 1969a). Many groups noticed the striking similarity between the constellation of malformations shared between the diseases, but only after eighteen years were they known together as Opitz, or Opitz G/BBB syndrome (referred to as Opitz syndrome in this thesis).

1.2.1.1 BBB syndrome

The earliest report of OS was an abstract by Opitz *et al.* (Opitz *et al.*, 1965) in which the authors described a new hereditary malformation called '*hypertelorism and hypospadias syndrome*', based on the primary phenotype of patients with the disorder. Subsequently, Opitz *et al.* described the syndrome in detail with the designation BBB syndrome (Opitz *et al.*, 1969a). BBB syndrome was characterised by the phenotypes of eight males from three families which shared the malformations of hypertelorism with variable hypospadias. Male members of the different families had combinations of mild to severe mental retardation and bilateral cleft upper lip and palate, coarctation of the aorta with hypoplasia of the aortic arch or other congenital heart defects, with transmission in all cases being from females with telecanthus/hypertelorism (Opitz *et al.*, 1969a). At the same time, Christian *et al.* described a family with congenital anomalies strikingly similar to BBB syndrome; hypertelorism, hypospadias, cryptorchidism and cleft lip and palate as well as imperforate anus and urethrocolic fistula (abnormal connection between the urethra

and colon) (Christian *et al.*, 1969). The family presented by Christian *et al.* (1969) was later classified as exhibiting BBB syndrome (Cordero and Holmes, 1978) and represents the varied spectrum of malformations that occurs even between affected male siblings with BBB syndrome (Cavallo *et al.*, 1988; da Silva, 1983; Halal and Farsky, 1981; Hogdall *et al.*, 1989; Horak and Smahel, 1987; Michaelis and Mortier, 1972; Miller *et al.*, 1977; Opitz *et al.*, 1965; Opitz *et al.*, 1969a; Parisian and Toomey, 1978; Reed *et al.*, 1975; Say and Carpenter, 1987; Stevens and Wilroy, 1988; Stoll *et al.*, 1985).

1.2.1.2 G syndrome

The malformations described for BBB syndrome were also characteristic of another inherited disease, G syndrome (Cordero and Holmes, 1978; Parisian and Toomey, 1978), the report of which was published at the same time as the BBB and Christian reports (Opitz *et al.*, 1969b). G syndrome was defined by the clinical observations of four brothers with a similar pattern of inherited physical malformations including hypospadias, oesophageal anomalies, hoarse cry, minor anomalies of the ear, and imperforate anus (Opitz *et al.*, 1969b). Subsequent reports suggested a wider prevalence of G syndrome, with diagnostic features expanding to cater for the increased phenotypic variation in presenting individuals (Arya *et al.*, 1980; Buckley *et al.*, 1988; Coburn, 1970; Cote *et al.*, 1981; Gilbert *et al.*, 1972; Greenberg and Schraufnagel, 1979; Howell and Smith, 1989; Kimmelman and Denny, 1982; Little and Opitz, 1971; Opitz, 1987; Parisian and Toomey, 1978; Pedersen *et al.*, 1976; Tolmie *et al.*, 1987; van Biervliet and van Hemel, 1975; Williams and Frias, 1987; Wilson and Oliver, 1988). As a result of these observations, the full phenotypic spectrum was revised to include hypertelorism, hypospadias, prominent forehead, cleft lip and/or palate, narrow palpebral fissures with epicanthal folds, dysphagia, aspiration, stridor and laryngotracheoesophageal (LTE) clefts (Opitz, 1987).

1.2.1.3 Opitz G/BBB syndrome

Initially, the BBB and G syndromes were thought to be distinct, with G syndrome showing an inheritance pattern consistent with autosomal dominance, whilst BBB syndrome showed X-linked inheritance. Historically each syndrome had specific defining features that further distinguished them as separate entities; however multiple reports over ensuing years presenting patient data inconsistent with the notion of BBB and G syndromes being distinct entities changed this opinion (Allanson, 1988; Cappa *et al.*, 1987; Cordero and Holmes, 1978; Farndon and Donnai, 1983; Verloes *et al.*, 1989). The significance of the phenotypic similarity between BBB and G syndromes was first suggested during a report of two half-siblings possessing clinically distinct, but partially overlapping phenotypes which were characteristic of both the BBB and G syndromes (Parisian and Toomey, 1978). Ten years later a report by Allanson described a proband which had typical LTE manifestations characteristic of the G syndrome, however the father and all four of his siblings displayed strong manifestations of the BBB syndrome phenotype (Allanson, 1988). These, as well as many reports of families in which the BBB and G syndrome segregated within a single kindred has reinforced the notion that BBB and G syndromes represent the same clinical disorder (Allanson, 1988; Cappa *et al.*, 1987; Cordero and Holmes, 1978; Farndon and Donnai, 1983; Verloes *et al.*, 1989). These studies, along with reports of an original G syndrome patient having a nephew with BBB syndrome features (Opitz, 1987), and a larger family with two members having features of both G and BBB syndromes (Verloes *et al.*, 1989) prompted a re-classification of the disorder as Opitz G/BBB syndrome, or simply Opitz syndrome, to include the spectrum of the more frequent phenotypes presented by all patients (Cappa *et al.*, 1987; Sedano and Gorlin, 1988).

1.2.2 The Opitz syndrome phenotype

Many years of study of OS have provided a description of the most common developmental defects associated with the disease. The OS phenotype is unified from the descriptions of G and BBB syndromes with a diverse array of midline defects including hypertelorism/telecanthus (widely spaced eyes) (Figure 1.2A,B), hypospadias (Figure 1.4A) or splayed labia majora (defects of the male and female genitalia, respectively), cryptorchidism (undescended testes), clefts of the lip and/or palate (Figure 1.3A), oesophageal abnormalities often presenting as gastric reflux and/or difficulty swallowing (dysphagia), congenital heart defects (Figure 1.3B) and imperforate anus (Figure 1.4B,C) (Cox *et al.*, 2000; Opitz, 1987; Opitz *et al.*, 1969b; Opitz *et al.*, 1969a; Quaderi *et al.*, 1997; So *et al.*, 2005; Verloes *et al.*, 1995).

1.2.2.1 The potential causes of the tissue malformations underlying Opitz syndrome

Tissue malformations such as those presenting in OS patients have been hypothesised to result from a delay or impediment of normal fusion events in early embryonic development (Cox, 2004; Hynes and Fraher, 2004b). While the cellular aspects of organ development are not fully understood for all tissues involved, the normal development of midfacial, tracheoesophageal, cardiac and genital tissues relies on a similar mode of fusion to connect extruding cell masses together (Cox, 2004; Hynes and Fraher, 2004a; Hynes and Fraher, 2004b; Hynes and Fraher, 2004c; Nievelstein *et al.*, 1998). Failure or incomplete fusion of these tissues has been hypothesised to be the key underlying cellular cause of cleft lip/palate, oesophageal closure, heart septation defects, hypospadias and anorectal fistulae (see Table 1). It is thought that many of the affected tissues require the processes of apoptosis and epithelial to mesenchyme transition in order to fuse epithelial cell layers, suggesting that a failure to initiate changes in cell

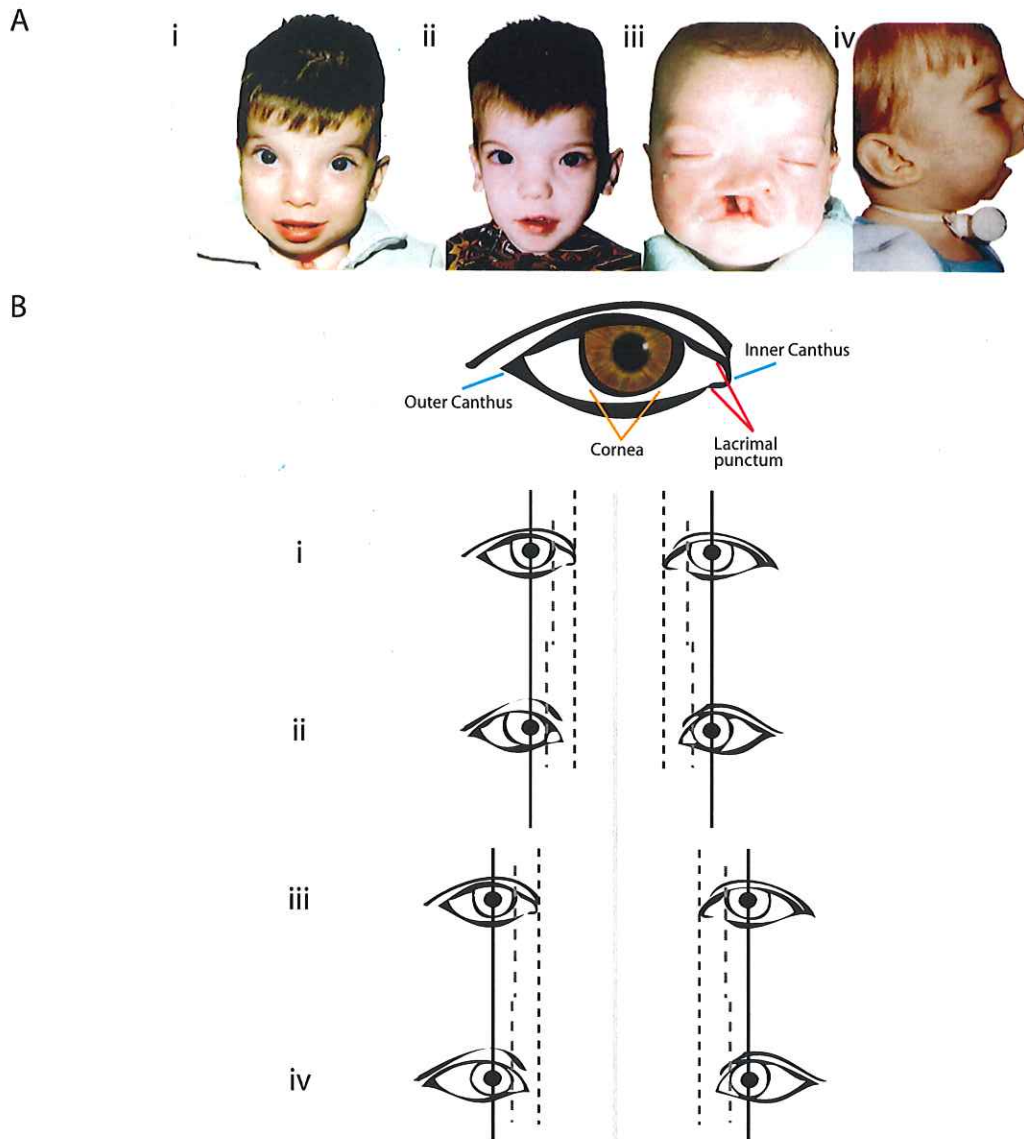
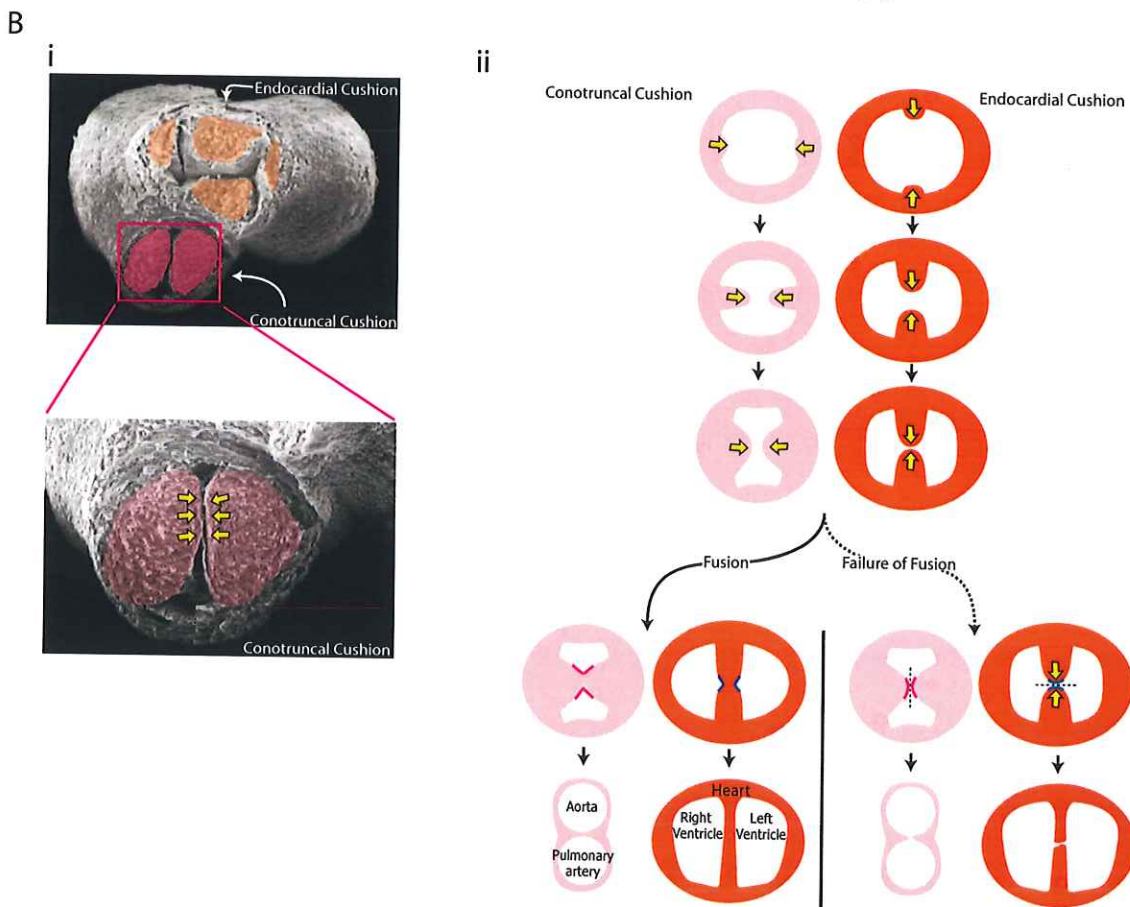
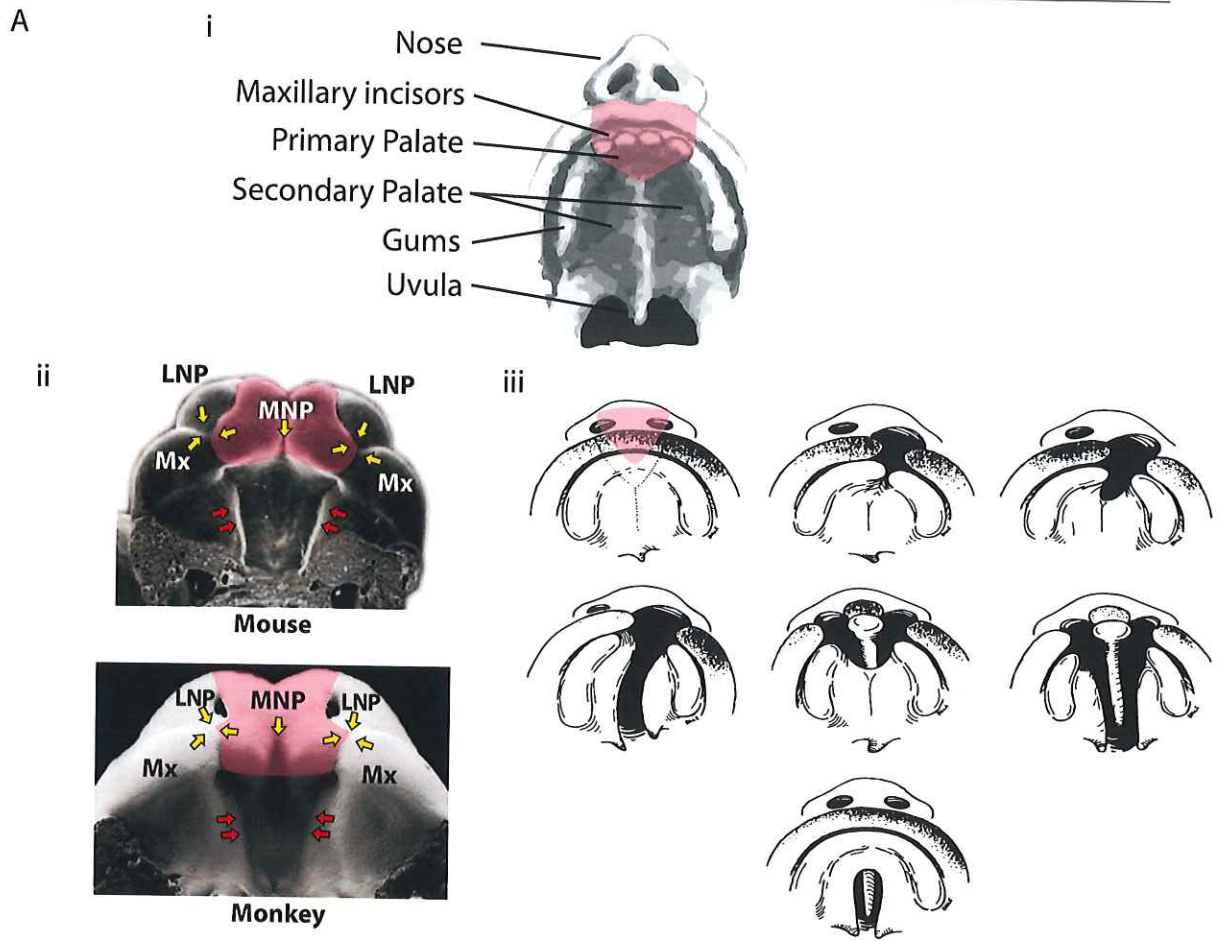


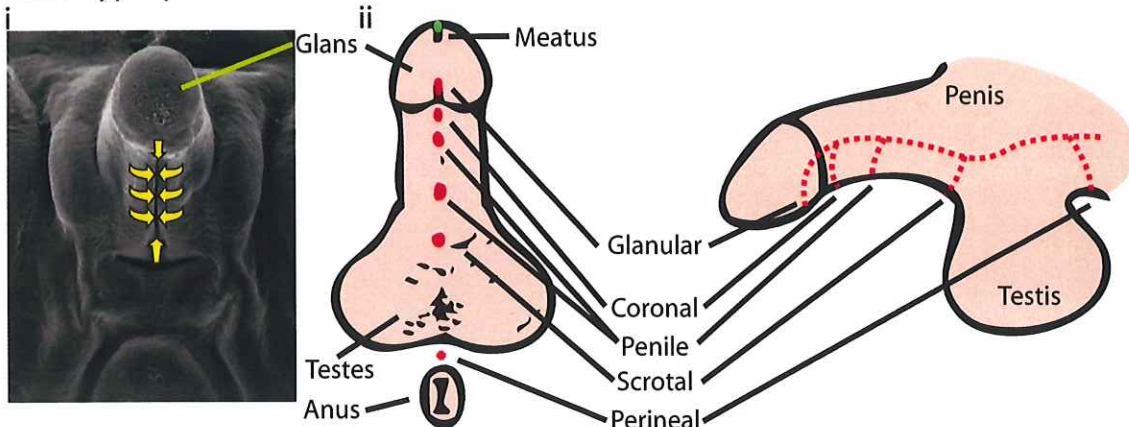
Figure 1.2 Opitz syndrome facial appearance

(A) The craniofacial anomalies seen in Opitz syndrome patients are variable, but include hypertelorism (i-iv), broad nasal bridge (i-iv), depressed nasal root (ii,iii), cleft lip and/or palate (iii), posteriorly rotated ears (iv) and widow's peak (ii) (Images from Dr. T.C. Cox, Monash University).

(B) Normal interocular distance (i) as compared with primary telecanthus (ii), ocular hypertelorism (iii), and ocular hypertelorism with secondary telecanthus (iv). Telecanthus refers to a broadened positioning of the inner canthus only, causing the lacrimal punctum to align with the cornea (forcing the cornea 'inward'). In hypertelorism, the entire eye is shifted laterally such that both inner and outer canthi are moved proportionately. Images modified from (Cohen, 1980).



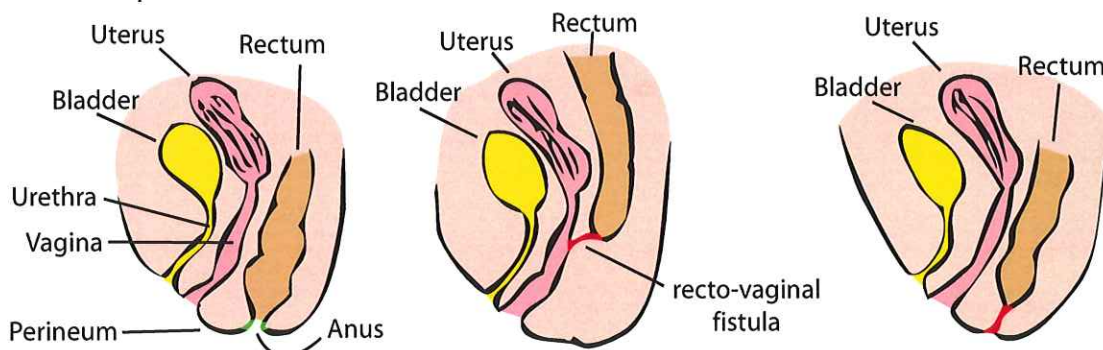
A Male Hypospadias



Urethral groove fusion in the developing human genitalia. Normal tissue movement shown in yellow.

Incomplete urethral groove fusion can result in one or more external urethral openings (hypospadias) along the line of the fusion 'seam' shown here in red (typical urethral opening, meatus, in green).

B Female Imperforate Anus



Normal

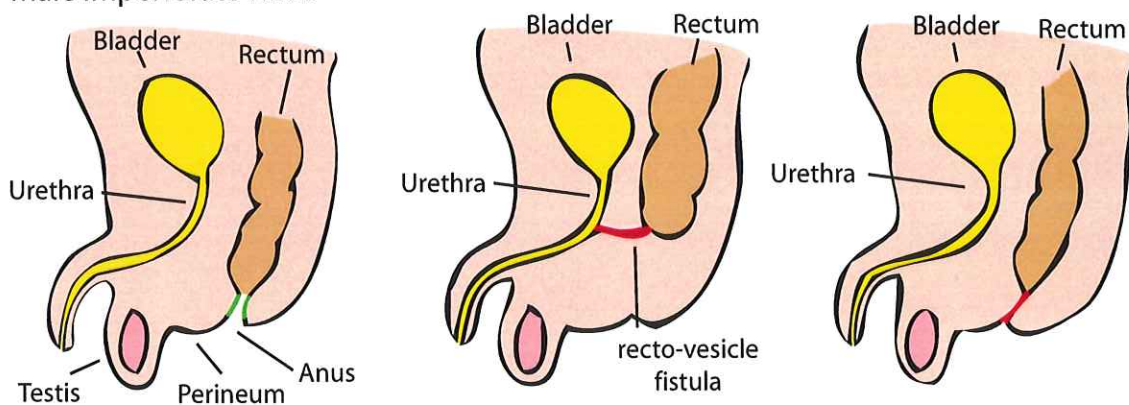
Imperforate anus

Underdevelopment of cloaca during lower gut genesis, resulting in an abnormal communication (fistula) linking the rectum with the vagina.

Imperforate anus

Resulting from a failure of cloacal membrane/plug rupture. The rectum is covered by skin or blocked by the epithelial plug.

C Male Imperforate Anus



Normal

Imperforate anus

Underdevelopment of cloaca during lower gut genesis, resulting in an abnormal communication (fistula) linking the rectum with the urethra.

Imperforate anus

Caused by a failure of cloacal membrane/plug rupture. The rectum is covered by skin or blocked by the epithelial plug.

programming is important in the development of fusion defects (Figure 1.5). Supporting this notion are animal models, such as *Gli3* heterozygous mutant mice which have a susceptibility to form urogenital defects with imperforate anus and tracheoesophageal fistulas, which indicates shared molecular genetic mechanisms during development of these tissues (Mo *et al.*, 2001). Since human tissue defects caused by failure or incomplete fusion also occur with high prevalence as isolated anomalies, the study of the tissue defects in OS may provide an increased understanding of these common defects.

Defect	Process	Prevalence	References
Cleft lip/palate	Failure of nasal and maxillary processes to fuse	~1:600 to 1:1000	(Cox, 2004)
Cardiac septation defects	Failure of endocardial or conotruncal cushions to fuse	~1:2500	(Webb <i>et al.</i> , 2003)
Hypospadias	Failure of labioscrotal urethral folds to fuse	Typically ~1:1800 to 1:3700 As high as ~1:283 (World Health Organisation)	(Canning, 1999; Mastroiacovo <i>et al.</i> , 2003)
Imperforate anus	Failure of the dorsal cloacal membrane to fuse with the urorectal septum tip epithelia	~1:3350	(Bai <i>et al.</i> , 2004; Yang <i>et al.</i> , 1994)

Table 1 – Prevalence of the most common forms of tissue malformations (both syndromal and non-syndromal). These malformations are also exhibited by OS patients.

1.2.3 The genetics of Opitz syndrome

Opitz syndrome is genetically heterogeneous, having been mapped to several loci through linkage analysis of large families (Xp22.3, 22q11.2) and the association with various chromosomal abnormalities in several isolated cases (14q11.2, 13q32.3ter and 5p12-13) (Fryburg *et al.*, 1996; Kurczynski *et al.*, 1998; Leichtman *et al.*, 1991; Urioste *et al.*, 1995; Verloes *et al.*, 1995). To date, data gained from the majority of patients has indicated that Xp22.3 and 22q11.2 are the two most common regions involved with OS. Importantly, studies over many years have indicated that individuals with linkage to either Xp22.3 or 22q11.2 have clinically indistinguishable OS features. It is of note, however, that although gene disruption through mutation is the primary causative factor in Opitz

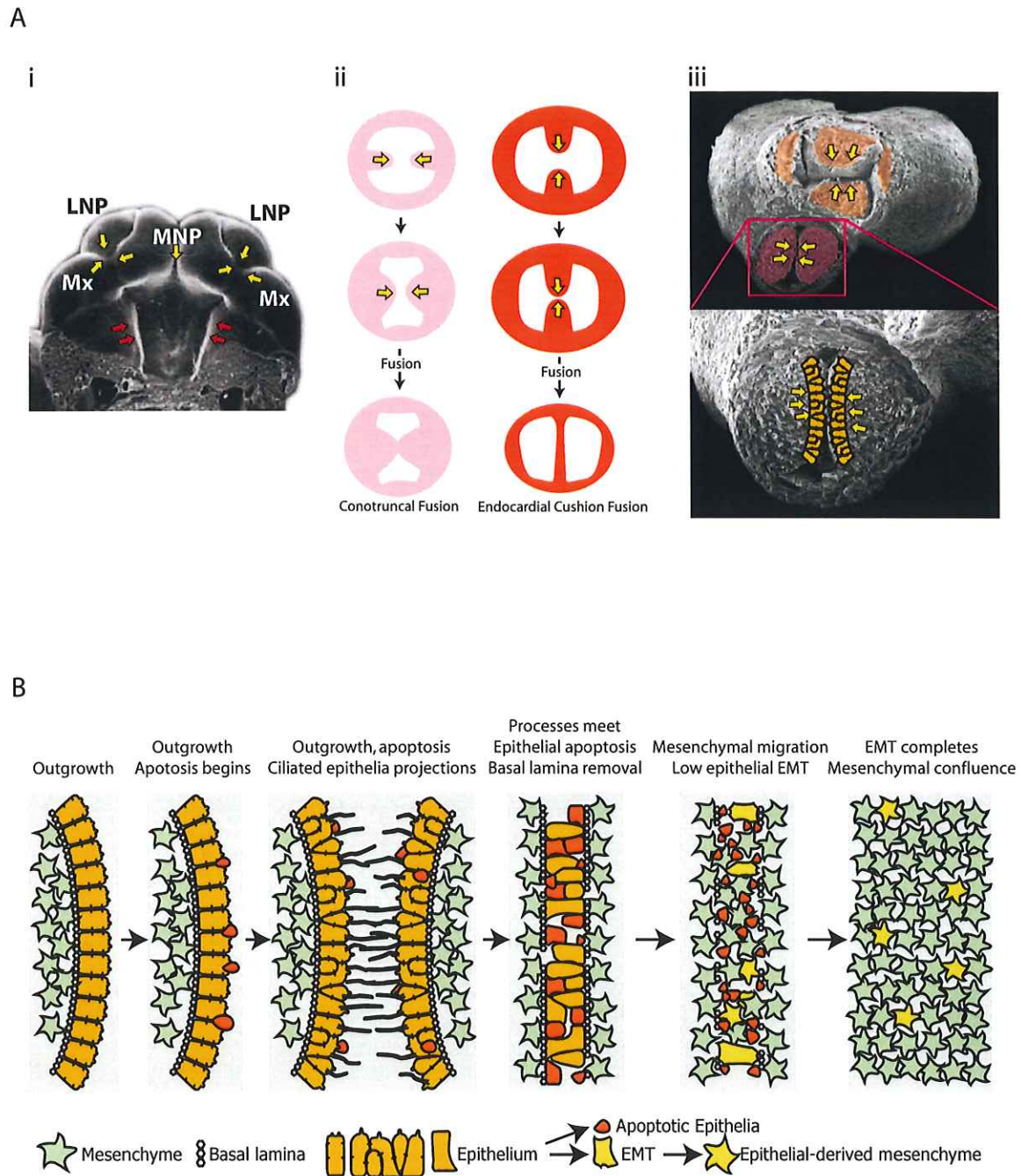


Figure 1.5 Epithelial fusion

Fusion of the craniofacial palate (A,i) and heart cushions (A,ii,iii) share common tissue transformation features. The primary goal of the fusion process is the removal of the outer epithelial layer with concurrent basal lamina breakdown in order to permit the formation of a confluent inner mesenchymal bridge (B). This is hypothesised to result from a tightly coordinated balance of cell death through varying degrees of apoptosis, cellular transformation using EMT and cell migration (figure modified from Cox, 2004).

syndrome, the extensive intrafamilial variability even among males with the same genetic lesion (Cox *et al.*, 2000) and reports of monozygotic twins exhibiting different phenotypic severity (Young *et al.*, 1988) indicates that additional environment or epigenetic phenomena may also influence the presentation and severity of OS.

1.2.3.1a Autosomal OS

Autosomal inheritance has been observed through male to male transmission of the disease (Chemke *et al.*, 1984; Farndon and Donnai, 1983; Stoll *et al.*, 1985) and also through direct observation of autosome aberrations in karyotypes of individuals with an OS phenotype (Christodoulou *et al.*, 1990; Fryburg *et al.*, 1996; Kurczynski *et al.*, 1998; Leichtman *et al.*, 1991; McDonald-McGinn *et al.*, 1995; Urioste *et al.*, 1995). Although these reports have clearly identified an autosomal contribution to OS, identification of mutations in autosomal genes in OS-linked regions has proven difficult. The most common autosomal region disrupted in Opitz syndrome is 22q11.2 (Christodoulou *et al.*, 1990; Fryburg *et al.*, 1996, Lacassie, 1996 #382; McDonald-McGinn *et al.*, 1995; Robin *et al.*, 1995). Some phenotypic overlap between OS and other diseases associated with haploinsufficiency of 22q11 indicates that genes within the affected region are important for development of the tissues commonly affected by these disorders. Interstitial deletion of a 2-3 Mb typical deleted region (TDR) within the proximal long arm of chromosome 22 (Fu *et al.*, 1976) is the cause of a group of disorders known collectively as '22q11 deletion syndrome' (or *del22q11*-syndrome) with patients displaying what has been termed the CATCH22 phenotype. CATCH22 is the acronym used to describe the range of patient phenotypes (including Cardiac defect, Abnormal facies, Thymic hypoplasia, Cleft palate, Hypocalcaemia, and 22q11 deletions) attributed to hemizyosity of a critical region encompassing 22q11 (Wilson *et al.*, 1993). The minimal critical region has been continually refined by comparison of reports describing patients with smaller or partial

deletion of the TDR (Amati *et al.*, 1999). Congenital malformation syndromes associated with these deletions include DiGeorge syndrome (DGS), Velo-Cardio-Facial syndrome (VCFS), conotruncal anomaly face syndrome and Opitz syndrome (Fryburg *et al.*, 1996; Lacassie and Arriaza, 1996; McDonald-McGinn *et al.*, 1995). The estimated incidence of a 22q11.2 microdeletion is 1:3000 to 1:4000 live births. This relatively high incidence of chromosomal rearrangement is caused by low copy repeats flanking the 3 Mb region (as well as within the region), which may promote homologous recombination resulting in deletions and duplications of 22q11 (Uddin *et al.*, 2006).

1.2.3.1b 22q11.2 critical region

The molecular consequence of removal of 22q11 is not fully understood, and while many deletions remove some of the 24 protein-encoding genes within a smaller 1.5 Mb critical region (which are expressed in affected tissues), some microdeletions exist which do not remove functional genes (Sutherland *et al.*, 1996), indicating that regulatory elements also exist within the region. Disruption (or the effects of complete absence) of 18 of these 24 protein-encoding genes, including *HIRA*, *CDC45-L*, *UFD1L* and *TBX1*, have been studied in mouse models (for a review, see (Lindsay, 2001). Murine deletions of segments of the 22q11.2 syntenic chromosomal region within mouse chromosome 16B replicate some of the CATCH22 phenotypes including congenital heart disease, schizophrenia and craniofacial anomalies (Jerome and Papaioannou, 2001; Kimber *et al.*, 1999; Lindsay *et al.*, 1999; Lindsay *et al.*, 2001; Merscher *et al.*, 2001). Deletion of different syntenic regions has both excluded and implicated particular causative genes, including those potentially involved in schizophrenia (Kimber *et al.*, 1999) and congenital heart disease (Lindsay *et al.*, 1999; Lindsay *et al.*, 2001; Merscher *et al.*, 2001; Puech *et al.*, 2000). These studies led to the identification of *Tbx1* as the gene primarily responsible for congenital heart defects in mouse models of *del22q11*-syndrome (Jerome and

Papaiouannou, 2001; Lindsay *et al.*, 2001; Merscher *et al.*, 2001) and also indicate that *Tbx1* is one of the contributory factors in the presentation of craniofacial defects (Lindsay, 2001). However, other genes within the TDR are also likely to contribute. For example, the role of the gene *HIRA* has been studied in chickens using gene knockdown (Farrell *et al.*, 1999). In these studies, knockdown of chick *Hira* in the cardiac neural crest cells (a migratory cell type that differentiates and contributes to many heart tissues) results in a significant increase in truncus arteriosus, one of the phenotypes sometimes displayed in patients with DiGeorge syndrome (Farrell *et al.*, 1999).

1.2.3.1c Human disorders within the 22q11.2 TDR

Whilst animal models provide vital clues to the genetic contribution of particular genes to the phenotype, information gathered from deletions in patients presenting with some, but not all, CATCH22 phenotypes is vital to determining the 22q11 genetic contribution to phenotype. One of the best examples comes from a patient presenting with atypical features of VCFS/DGS including craniofacial and cardiac defects. This patient was shown to have a 20 kb microdeletion of the region disrupting exons 1-5 of *CDC45L* and exons 1-3 of *UFD1L* (Yamagishi *et al.*, 1999). However, other VCFS/DGS patients with similar cardiac and craniofacial defects have deletions that exclude *CDC45L/UFD1L* (Saitta *et al.*, 1999) suggesting heterogeneity within the region for these phenotypes (Edelmann *et al.*, 1999).

Since there is overlap in the types of heart and craniofacial defects between VCFS/DGS and OS, mapping of such deletions even in other disorders can provide important information when determining which genes to screen for mutations in autosomal OS patients. Our laboratory has sequenced some 22q11.2 genes (such as *UBCH7*) in patients with autosomal inheritance patterns without finding mutations in exons (Blair Hopwood, unpublished observations). In the greater field of 22q11.2-related research, and

especially 22q11.2 OS-gene research, such work is necessary to find a significant proportion of remaining OS causative loci. To highlight the difficulty of finding such mutations, approximately 50% of patients presenting with OS, including some with demonstrable X-linked inheritance, have no identified mutations in *MID1* (So *et al.*, 2005).

1.2.3.2 X-linked OS

The strongest evidence for an X-linked form of Opitz syndrome was exhibited by a large twenty member family in which Opitz syndrome co-segregated with a pericentric inversion of the X chromosome (inv (X)(p22.3q26)) (Verloes *et al.*, 1995). This indicated that the causative gene resided within the vicinity of one of the breakpoints, although supplementary data pointed towards the Xp22 region (Cox *et al.*, 1998). Fine mapping of this Xp22 inversion breakpoint was facilitated by the use of YAC probes, with subsequent exon trapping experiments successfully used to clone the gene, which was named *MID1* (*Midline 1*) (Quaderi *et al.*, 1997). Ensuing analysis of patient genomic DNA revealed mutations within this gene in other patients with a definitive diagnosis of OS (Cox *et al.*, 2000; Gaudenz *et al.*, 1998).

Palmer *et al.* had previously cloned the murine homologue of *MID1* which was found to lie across the pseudoautosomal boundary on the mouse X chromosome. This unique situation means that the first 3 exons of the gene are X chromosome specific, while the remainder are found on both the X and Y chromosomes. Using this as a descriptor, the gene was named *Fxy* (Finger on X and Y) (Palmer *et al.*, 1997). This same group also subsequently cloned the human homologue, naming it *FXY* (Perry *et al.*, 1998). Although being the first group to discover the gene, *MID1*, not *FXY*, was the name subsequently accepted by the Human Genome Nomenclature Committee and has been used as the name for all orthologues discovered.

The most recent review of Opitz syndrome cases with mutations in *MID1* reports

the proportion of male individuals with specific malformations: hypertelorism/telecanthus (98%), hypospadias (72%), laryngotracheoesophageal (LTE) defects including LTE clefting (50%), cleft lip and palate (48%), defects of the ear (45%), developmental delay (42%), anal atresia or anteriorly displaced anus (24%) and cardiac structural defects (22%) (So *et al.*, 2005). The early distinction between BBB and G syndromes was that of inheritance (where known) and phenotype (Bershof *et al.*, 1992; Chemke *et al.*, 1984; Farndon and Donnai, 1983; Kasner *et al.*, 1974; Opitz, 1987). Although the notion of separate etiologies has since been disproved, a study of *MIDI* X-linked OS patients by So *et al.* has indicated a potential difference with relation to specifically associating phenotypes (So *et al.*, 2005). It was shown that there was a significantly increased rate of cleft lip and palate in males with the *MIDI* mutation (versus mutation-negative OS males) as well as an increased proportion of hypospadias and anal defects in males with *MIDI* mutations. Furthermore, male OS patients without *MIDI* mutations were shown to have a significantly increased incidence of developmental delay, mental retardation, dysphagia, aspiration, gastrointestinal reflux and isolated cleft palate. Both males and females without *MIDI* mutations also have a significantly increased incidence of neurological abnormalities (So *et al.*, 2005). The finding of a marginally increased incidence of cardiac defects present in OS patients without *MIDI* mutations (So *et al.*, 2005) could potentially reflect the contribution of the main autosomal OS locus of 22q11.2, a chromosomal region that has been directly associated with 5% of all congenital cardiac defects through the 22q11.2 deletion syndrome (Glover, 1995).

1.2.4 Other syndromes with common phenotypic features

Disorders that display multiple common phenotypic similarities in affected body systems are generally thought to reflect different abnormalities in a common or interacting

genetic pathway and thus identification of the causative genes underlying such syndromes may assist in dissecting such developmental pathways. Indeed, information of this nature assisted the study of the G and BBB syndromes (previously discussed) and led to the unification of the syndromes into a single disorder. Another syndrome which has similar clinical features to Opitz syndrome is FG syndrome, an X-linked disorder with affected individuals having some, but not necessarily all, of a broad spectrum of clinical features such as: congenital hypotonia (low muscle tone), mental retardation, telecanthus, hypertelorism, cowlicks, large head, tall and broad forehead, characteristic facial appearance, imperforate anus, congenital heart defects and small ears (Graham *et al.*, 1999). Five X-linked loci have been identified by linkage analysis and observation of chromosome breakpoints: Xq12-q21.31 (FGS1) (Briault *et al.*, 1997), inv(X)(q11q28) (FGS2) (Briault *et al.*, 2000), Xp22.3 (FGS3) (Dessay *et al.*, 2002), Xp11.4-p11.3 (FGS4) (Piluso *et al.*, 2003) and Xq22.3 (FGS5) (Jehee *et al.*, 2005). Interestingly, the position of the FGS3 locus marker (DXS1060) is 5 Mb downstream from the *MIDI* gene, and the FGS5 locus is also located near *MID2* (the closest *MIDI* homologue) and has been shown to be within an Xq22.3 region duplicated in a patient with FG syndrome (Jehee *et al.*, 2005). This is important for understanding the potential heterogeneity of these disorders (to be discussed later), since partial phenotypic overlap between FG and Opitz syndromes also hints at the notion of a genetic interaction between the genes that are mutated in patients with these diseases.

1.3 *MIDI*

1.3.1 Genomic location and organisation

MIDI is positioned at the distal end of the long arm of the human X chromosome at Xp22.3 (10.399M-10.225M). The coding region of *MIDI* spans approximately 120kb and

encompasses 9 exons (Figure 1.6) (Cox *et al.*, 2000). Several human *MIDI* isoforms are transcribed from five differentially incorporated untranslated first exons (Cox *et al.*, 2000), with each isoform splicing into the first coding exon at the same position. Thus each alternate mRNA isoform is predicted to encode the same protein (Landry and Mager, 2002). The non-coding alternative primary exons are positioned up to 156 kb upstream of exon 1, and several of the non-coding exons have their own promoters, indicating that the largest possible *MIDI* transcript is at least 275 kb in length (Cox *et al.*, 2000; Landry and Mager, 2002; Ensembl v39 Homo sapiens Gene report).

1.3.1.1 *MIDI* promoter

Several alternative promoters (known as B, C and E) have been identified as transcriptionally active for endogenous *MIDI* expression, as well as two promoters (A and D) which permit tissue-specific expression in adipose and placental tissues, respectively (Landry and Mager, 2002). The mRNA isoforms produced by transcription from these alternative promoters have the same translation initiation site and encode the same protein. Surprisingly high levels of DNA sequence conservation of common alternative promoter regions between species such as human, marmoset, dog, mouse, rat, pig and cow demonstrates that mutations in the region have been heavily selected against during evolution of animals (Landry and Mager, 2002). This indicates that the presence of the alternative promoters and their function in regulating *MIDI* expression are important for development pathways that are common to many animals. In this respect, a retroviral origin of promoter D (which is only present in higher primates) is conceivable since the positioning of the promoter and its ability to permit transcription of *MIDI* in low amounts would likely not affect the activity or specificity of the other promoters (Landry *et al.*, 2002). The function or purpose of these promoters is largely unknown, since only two provide tissue specificity. However the *MIDI* isoforms may have different translational

properties conferred by the alternate 5' UTRs (untranslated regions). For example, any additional AUG codons and/or secondary structures within these 5'UTRs could contribute to such variation in translational efficiency between the different isoforms (Landry and Mager, 2002).

1.3.2 *MID1* mRNA expression pattern

In chick, embryonic expression of *MID1* begins at a time approximately equivalent to embryonic day 15 of human development, with expression around Hensen's node (where Left-Right body axis asymmetry is established). At this time, it is initially expressed bilaterally around the midline axis, but subsequently becomes restricted to the right side of the node: the structure with global 'organising' properties that coordinates patterning and layout of the gastrulating embryo (Granata and Quaderi, 2003). During embryonic development, significant *MID1* expression pattern similarities have been found between chicks, mice, humans and other vertebrate species (Dal Zotto *et al.*, 1998; Gaudenz *et al.*, 1998; Granata and Quaderi, 2003; Landry and Mager, 2002; Pinson *et al.*, 2004; Richman *et al.*, 2002) supporting the evolutionary conservation of function of *MID1* during development. While much of the body expresses low levels of *MID1*, high levels are found in neural tissue, and in epithelia of the developing face, heart, gastro-intestinal (GI) tract, urethral folds and anal tubercle. The overlap between the tissues affected in Opitz syndrome, including (but not restricted to) brain, trachea, heart, lung and kidney (Gaudenz *et al.*, 1998; Pinson *et al.*, 2004) and the expression patterns in vertebrate species is striking.

To appreciate the specificity of the expression pattern, the human *MID1* expression data obtained by *in situ* hybridisation of embryonic and foetal tissue is summarised hereafter (Pinson *et al.*, 2004). Transcripts were detected in human 5 week old embryos,

with concentration in the nervous system, foregut, head and neck (pharyngeal arches, otic vesicle), kidney (mesonephros) and limb buds. At 6 weeks, embryos show restricted expression in the nervous system (telencephalic vesicle, dorsal root ganglion and spinal cord), head and neck (nasal, oropharynx and oesophageal epithelia), lung epithelia and at the top of the interventricular septum of the heart. Expression continues in 7 week embryos in the developing nervous system (telencephalon, cerebellum bud, spinal cord), as well as head, neck, respiratory and digestive tissues (neural retina, nasopharyngeal epithelia, oesophagolaryngeal junction, respiratory and digestive tract epithelia), limb mesenchyme and muscle cells in proximal limbs, renal epithelia, anal folds, and genital tubercle. In the heart, expression at the tip of the interventricular septum (the inward-growing tissue that later divides the ventricles) is still detectable at 8.5 weeks of human embryonic development (Pinson *et al.*, 2004). The expression profile shown by Pinson *et al.* provide position-specific evidence for the expression of *MIDI* within the tissues affected in Opitz syndrome, and provide insight into the cell types which are compromised during the development of the disease.

1.3.3 *MIDI* mutations in Opitz syndrome

As briefly mentioned earlier, mutational analysis has provided ample support for *MIDI* being the X-linked Opitz syndrome gene. More than 40 different mutations have now been reported in both sporadic and familial cases of the disease. The majority of mutations in *MIDI* are located in (but not restricted to) exons, and include insertions, duplications, truncations, missense and nonsense mutations, deletions and splice-acceptor site mutations (summarised in Fig 1.6) (Cox *et al.*, 2000; Gaudenz *et al.*, 1998; Pinson *et al.*, 2004; Quaderi *et al.*, 1997; Schweiger *et al.*, 1999; So *et al.*, 2005; Winter *et al.*, 2003). The relative position of the mutations in *MIDI* does not appear to correlate with severity of

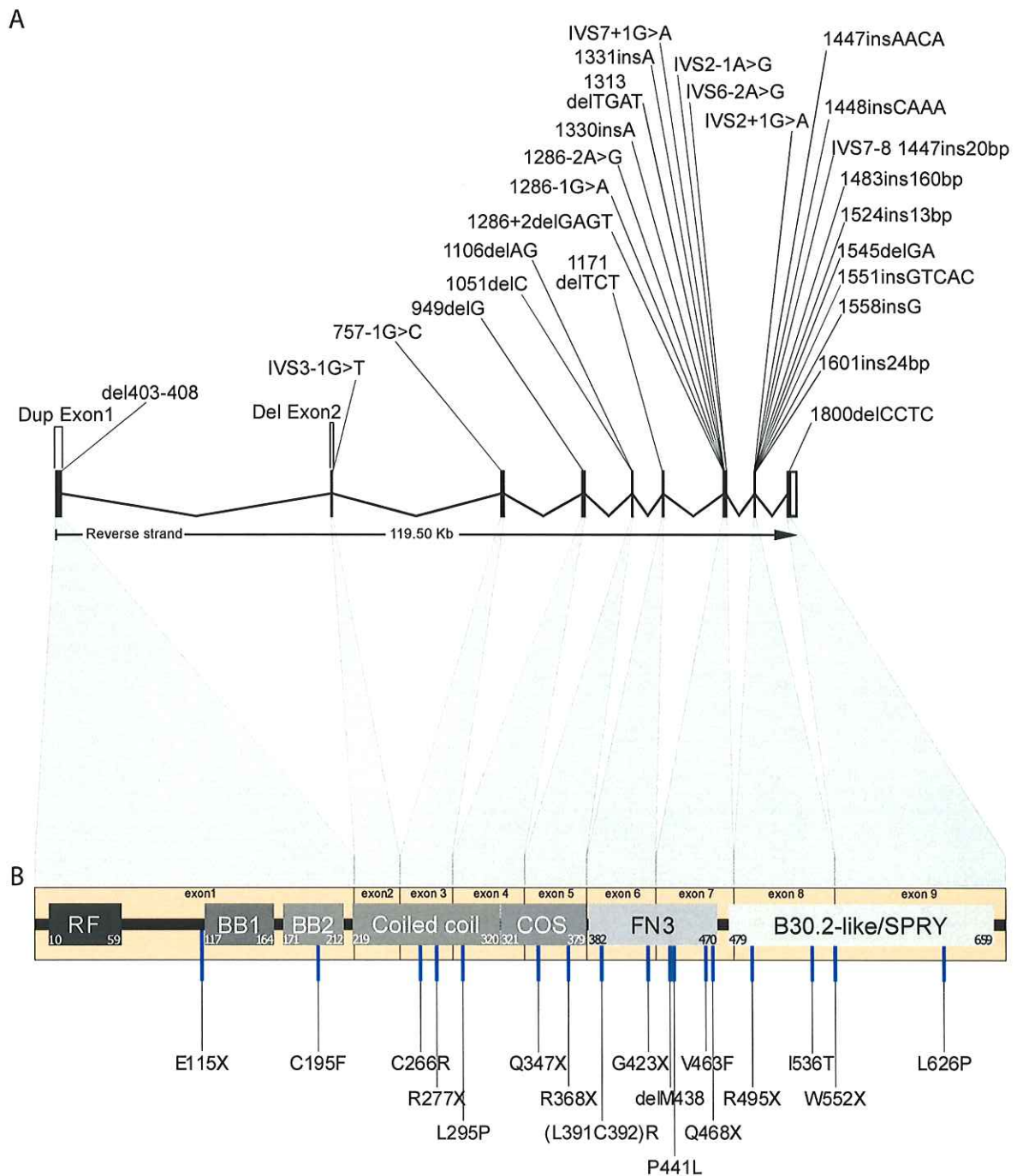


Figure 1.6 *MID1* genetic mutations and their positions within the protein.

Mutations in *MID1* that cause nonsense mutations, duplications, splicing errors or other complex mutations are shown with their relative positions marked on the exons (Figure 1.6A). The protein-coding exons of *MID1* are represented by vertical black boxes/lines, with spliced introns represented by 'v' shaped lines. Any mutations which cause truncation or missense changes to amino acids are represented on the protein schematic, with relative positions of domains and exon origins shown (Figure 1.6B).

the disease phenotype, although an observation has been made in which multiple patients within particular families with inherited *MID1* mutations consistently show laryngotracheoesophageal (LTE) malformations compared with other families with *MID1* mutations (So *et al.*, 2005). Such observations have been rare, since considerable variation within a single family is normally observed and thought to be due to the influence of environmental conditions during embryonic development. Even identical twins have been reported with discordant phenotypes (Young *et al.*, 1988) (although it should be noted that it has not been determined if the individuals reported by Young *et al.* have *MID1* mutations). The careful analysis of Opitz syndrome-causing mutations in *MID1* indicates that most mutations result in a loss-of-function of the encoded protein (Cox *et al.*, 2000).

1.4 MID1 protein

MID1 has an open reading frame of approximately 2 kb that encodes a protein of 667 amino acids (Figure 1.6) (Quaderi *et al.*, 1997). Initial investigation showed that the predicted protein had homology to several previously identified domains: at the N-terminus a zinc-binding RING finger motif, followed by two zinc-binding B-box motifs, and a coiled-coil domain were noted. The RING finger, B-box and coiled-coil domains together are known as an RBCC or tripartite motif. The observation that many proteins with N-terminal RBCC domains are found in multi-protein macromolecular assemblies within cells (Borden and Freemont, 1996) hinted that MID1 may function similarly in large multi-protein complexes (Quaderi *et al.*, 1997). In addition, the carboxy (C)-terminus of the protein shows homology to a region variably referred to as an RFP domain (based on homology to the C-terminus of Ret Finger Protein), a B30.2-like domain and PRY/SPRY domains. With more thorough analysis of the MID1 amino acid sequence we identified homology to a Fibronectin Type III domain located between the coiled-coil and C-terminal

domains (Perry *et al.*, 1999) (Appendix 1). Typically, Fibronectin type III domains act as scaffold sequences within proteins (Mulhern *et al.*, 1998) yet there are many that have been shown to directly mediate protein-protein interactions, particularly within cell signalling molecules (Koide *et al.*, 1998). Although there is some evidence (presented later) to support a role for MID1 in regulating intracellular signalling, the diverse structures and overall functions of proteins with such domains precludes predictions of any specific cellular function for the full MID1 protein. Nevertheless, to this end, some insight can be gained from our recent findings of a small group of proteins with the same domain architecture as MID1 (Short and Cox, 2006).

1.4.1 Homologous domains within MID1

1.4.1.1 RING Finger

The RING finger is a C3HC4-type specialised zinc finger domain which has been found in many proteins and is not restricted to the RBCC protein family. In general, zinc fingers are small domains that normally fold independently of the remaining protein structure and require zinc metal ions to stabilise their tertiary structure. Various types of zinc fingers are involved in DNA or RNA binding, protein-protein interaction and membrane association. The RING finger has a structure containing coordinating Cysteine (Cys, C) and Histidine (His, H) groups with the arrangement: C-x₍₂₎-C-x₍₉₋₃₉₎-**C**-x₍₁₋₃₎-**H**-x₍₂₋₃₎-C-x₍₂₎-C-x₍₄₋₄₈₎-**C**-x₍₂₎-**C** where x can be any amino acid and the bracketed numbers refer to the number of residues in the intervening sequences. The coordination of zinc involves the side-chains of Cys pairs one and three (underlined) arranging to bind one zinc ion. A second zinc ion is bound by Cys/(His/Cys) pairs two and four (bold), and together the two zincs are bound in what has been referred to as a 'cross-brace motif' (Borden and Freemont, 1996). The naming of this zinc-binding motif was derived from the protein where the novel fold was first identified, namely in the *RING1* gene (Really Interesting

New Gene) (Freemont *et al.*, 1991). Functional studies of C3HC4-type RING fingers have shown that it is the domain primarily responsible for labelling targeted proteins for degradation through the covalent addition of small ubiquitin modules to proteins which are in turn recognised by the proteasome (Joazeiro *et al.*, 1999). Proteins with such domains therefore function as E3 ubiquitin ligases, enzymes of the ubiquitin protein degradation system that usually present substrates to the ubiquitin conjugating (E2) enzymes and facilitate the subsequent attachment of ubiquitin. Thus, E3 ubiquitin ligases provide the substrate specificity for the degradation process by interacting with target proteins and using their RING fingers to transfer ubiquitin groups from E2 ubiquitin conjugation enzymes to their substrate targets. Examples of demonstrated E3 ubiquitin ligases include tumour suppressors such as c-Cbl and BRCA1/BARD1 (Joazeiro *et al.*, 1999; Mallery *et al.*, 2002), as well as a growing number of RBCC proteins such as EFP, TRIM11, ARD1 and MUL (Kallijarvi *et al.*, 2005; Niikura *et al.*, 2003; Urano *et al.*, 2002; Vichi *et al.*, 2005).

1.4.1.2 B-box2

MID1 contains two B-box domains positioned immediately C-terminal to the RING domain. Although the Cys-His arrangement in each of the B-boxes are distinct (now referred to as type 1 and type 2 B-boxes), their similar nomenclature has nevertheless persisted in the literature since the original report of this domain in the *Xenopus laevis* XNF7 protein (Reddy and Etkin, 1991). In XNF7, an RBCC protein, two Cys-His zinc finger domains were noted and called the A-box and B-box (Reddy and Etkin, 1991). The A-box had already been referred to as the RING finger nine months earlier (Freemont *et al.*, 1991) and as the B-box had not previously been identified, its name remained.

The first tertiary structure reported for a B-box was for the single B-box (B-box type 2) from XNF7 (Reddy *et al.*, 1991). Nuclear magnetic resonance (NMR) analysis of a

folded synthetic B-box peptide revealed that the B-box2 solution structure was made up of two beta-strands, two helical turns and three extended loop regions different from any other zinc binding motif and structurally distinct from the RING finger domain (Reddy *et al.*, 1991). The B-box2 domain consensus sequence is $\underline{C}\text{-X}_{(2-4)\text{-}}\underline{H}\text{-X}_{(7-10)\text{-}}\underline{C}\text{-X}_{(2-4)\text{-}}\underline{D}(\underline{C})\text{-X}_{(4-7)\text{-}}\underline{C}\text{-X}_{(2)\text{-}}\underline{C}\text{-X}_{(4-7)\text{-}}\underline{H}\text{-X}_{(2-5)\text{-}}\underline{H}$ where ‘x’ can be any amino acid, underlined residues coordinate zinc, and the bracketed numbers refer to the number of residues in the intervening sequences. The structure of the XNF7 B-box2 and the subsequent NMR solution structure of the B-box2 domain from MID1 provide evidence that the B-box2 domain only binds a single zinc ion yet strikingly the MID1 B-box2 structure shows little resemblance to the XNF7 B-box2 and utilises different coordinating residues: the final Cys-Cys and His-His pairs in MID1 versus the first Cys-His pair and Cys5 and His6 in XNF7. Interestingly, the MID1 B-box2 structure is more similar to that of a RING domain despite coordinating one less zinc (Massiah *et al.*, *submitted*). Ongoing work is aimed at understanding the basis for these structural differences, if indeed they are *bona fide*.

1.4.1.3 B-box1

The focus of early RING and B-box research on XNF7 may account for the proliferation of general RBCC/TRIM literature reporting only a single type of B-box. However, like MID1, around half of the RBCC protein family have an additional ‘B-box’ between the N-terminal RING finger and their B-box2 domains. As inferred above, this additional B-box sequence (a type 1 B-box, B-box1) has the spatial arrangement of coordinating residues ($\underline{C}\text{-X}_{(2)\text{-}}\underline{C}\text{-X}_{(7-18)\text{-}}\underline{C}\text{-X}_{(2)\text{-}}\underline{C}\text{-X}_{(2)\text{-}}\underline{C}\text{-X}_{(2-3)\text{-}}\underline{C}(\underline{H})\text{-X}_{(3-4)\text{-}}\underline{H}\text{-X}_{(1-11)\text{-}}\underline{H}$) that is different from B-box2 sequences (italic residues bind one zinc, bold residues bind an additional zinc). This lack of understanding or confusion about the differences between B-box domains in the literature has meant that even domain databases have pooled consensus sequences from the two distinct B-box variants.

Studies conducted during the period of candidature by K.M. Short and T.C. Cox in collaboration with Dr. Michael Massiah at Oklahoma State University (Stillwater, Oklahoma, USA) have revealed the NMR solution structure for bacterially expressed MID1 B-box1 (Massiah *et al.*, 2006) (Appendix 3). The B-box1 domain coordinates two zinc ions, with Cys pair one and Cys/(Cys or His) pair three (underlined and italicised) binding one zinc ion, and Cys pair two and His pair four (underlined and bold) coordinating the second zinc ion in a ‘cross-brace motif’ similar to that observed with the RING finger (Massiah *et al.*, 2006). The B-box1 structure also therefore grossly resembles the topology of the MID1 RING finger structure. Three dimensional structural similarity searches found the B-box1 structure also to be very similar to ubiquitin-ligation motifs other than the RING, in particular the ZZ and U-box domains. The striking structural similarity between B-box1 and these ubiquitylation motifs has raised the possibility that the B-box1 domain may in fact possess some ubiquitylation activity (Massiah *et al.*, 2006). While no evidence is available to directly support any B-box1 ubiquitylation activity, this finding highlights some future areas of interest aimed at dissecting the role of this domain in RBCC proteins such as MID1.

1.4.1.4 Coiled coil

Coiled coil domains are frequently found to mediate homo-interactions and are present in many thousands of proteins either as an isolated motif or part of large multi-domain proteins. In fact, the domain has also been shown to function as a homo-interaction interface in a growing number of RBCC proteins, thus facilitating their multimerisation (Reymond *et al.*, 2001). Structurally, the coiled-coil domain is composed of leucine-rich right-handed alpha helices wrapped around each other with a slight left-handed super-helical twist (Yadav *et al.*, 2006). Two-stranded, three-stranded, and four-stranded coiled coils are the most common forms of coil found. Two-stranded coiled coils have drawn

particular interest because of the “leucine-zipper” motif, found in many DNA-binding proteins (O’Shea *et al.*, 1989). Coiled coils consist of a repeating pattern of hydrophobic residues forming a heptad repeat with the first and fourth residues being hydrophobic and fifth and seventh amino acids being predominantly charged residues, although variation in such repeats is apparent in some proteins (Hicks *et al.*, 1997). The interactions between amino acids at these heptad positions are essential to the formation of oligomeric forms of the coiled-coil structure. The specificity of such interactions and the amount of packing space required for the residues at each position vary depending on the oligomeric state of the coiled coil (Yadav *et al.*, 2006).

As work conducted during the period of this research project has verified, a coiled coil domain is always preceded by a B-box2 domain within every RBCC/TRIM superfamily member (Short and Cox, 2006). This observation raises the question as to whether there is some level of co-operation or concerted function between the B-box2 and coiled coil. If indeed RBCC domains provide a relatively conserved spectrum of molecular functions, divergence of function within the superfamily of RBCC proteins would then largely be dependent on the remaining C-terminal domains.

1.4.1.5 Fibronectin type III domain

The Fibronectin type III domain (FN3) was first identified in Fibronectin proteins, which are multi-domain glycoproteins found both in soluble forms in plasma and insoluble forms in connective tissue and basement membranes (Skorstengaard *et al.*, 1986). The repeats in Fibronectins are able to associate with several proteins including heparin, collagen, actin, fibrin and fibronectin receptors (Skorstengaard *et al.*, 1986). The FN3 domain is approximately 100 amino acids in length, and proteins containing the domain are involved in a vast array of cellular processes including wound healing, cell adhesion, cell differentiation/migration, maintenance of the cytoskeleton and metastasis (Dean *et al.*,

1987). The presence of FN3 domains in bacteria, archaea, and single or multicellular eukaryotes indicates that the domain is fundamental to processes common to all life forms, and is highly flexible in the respect that it may be used for multiple purposes by minor modifications to the primary sequence of the domain (Craig *et al.*, 2004).

1.4.1.6 B30.2-like domain

The B30.2 domain was originally named after an exon identified in gene trapping experiments. The exon encoded a 166 amino acid peptide that subsequently showed high identity to other predicted proteins whose genes mapped to the same chromosomal region around the major histocompatibility complex (MHC) (Vernet *et al.*, 1993). Homology to the domain has since been identified in many additional proteins and secreted factors (hence known as B30.2-like domains). Proteins containing B30.2-like domains contain three highly conserved amino acid motifs (LDP, WEVE and LDYE) within the domain. The most probable fold for the B30.2-like domain has been predicted to consist of two distinct beta-domains (PRY and SPRY) involving 15 beta-strands, with the 5th beta-strand corresponding to the WEVE motif (Seto *et al.*, 1999). Interestingly, the region corresponding to the second beta domain is found in many other proteins (where it is known as the SPRY domain). Thus, the two distinct beta domains of the B30.2 fold have been hypothesised to work together as a larger unit although this is still to be functionally confirmed (Seto *et al.*, 1999). Although many proteins only possess the SPRY domain portion of the B30.2-like domain, analysis of data available in genome databases provides evidence of the presence of whole B30.2-like domains in simple metazoans such as *Drosophila melanogaster* and *Caenorhabditis elegans*, indicating that both the PRY and SPRY beta-domains making up the B30.2 have co-existed for a long time through animal evolution. However, the presence of SPRY and absence of PRY or B30.2-like domains in plants and fungi suggests that the PRY and B30.2-like domain likely evolved later in

eukaryotic evolution when multicellularity gave rise to animals. Confirming this, a B30.2 X-ray crystal structure of the *Drosophila* GUSTAVUS protein has recently been solved (Woo *et al.*, 2006). This shows that the PRY beta-domain is an alpha-helical and beta-stranded looping structure positioned next to the C-terminal SPRY sub-domain which independently forms a two-layered beta sandwich core structure that acts as a ligand-binding surface (Figure 1.7). Additional supportive evidence from an NMR structure of the SPRY domain of the murine SSB-2 protein confirms the presence of a beta sandwich inner core structure (Figure 1.7)(Masters *et al.*, 2006) that folds similarly to the GUSTAVUS SPRY beta-domain of the larger B30.2-like domain (Figure 1.7).

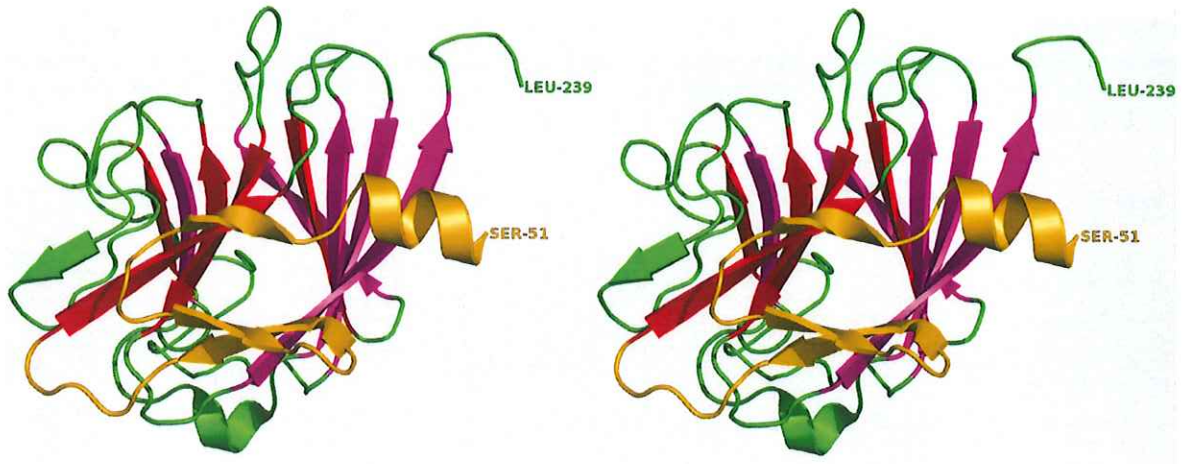
Recently, the B30.2-like domain has become a major focus of research, owing to the unique properties of the TRIM5 α B30.2-like domain being able to suppress retroviral infection (Stremlau *et al.*, 2004). In primates, specific orthologues of TRIM5 α are able to restrict HIV-1 infection and introduction of primate-specific differences in the B30.2-like domain of monkey TRIM5 α into human TRIM5 α confer HIV-1 resistance to human cells (Yap *et al.*, 2005), providing a key insight into the infection cycle of the virus and a possible target for preventative medicines.

1.4.2 Homologous RBCC motif-containing proteins

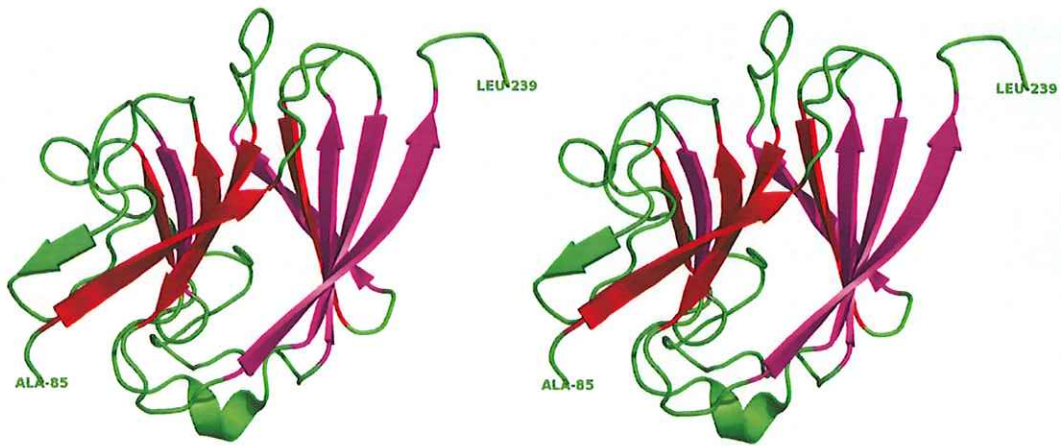
As introduced above, the family of proteins which have N-terminal RING Finger, B-box and Coiled-coil motifs is known as the RBCC or Tripartite motif (TRIM) superfamily. The RBCC protein family is large, comprising around 50 members in humans alone, with the different proteins having a wide variety of functions dependent on the combination of RBCC domains with a diverse accompaniment of C-terminal domains. The significance of these intriguing proteins in animal development is underscored by the number of developmental disorders and diseases associated with mutations in such

Figure 1.7 B30.2 and SPRY domain structure

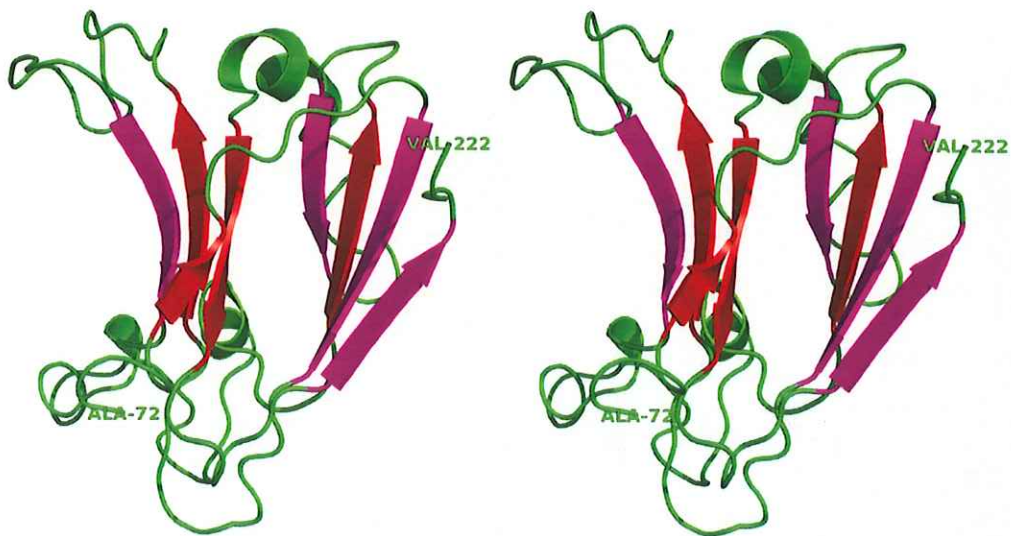
The B30.2-like domain of GUSTAVUS (A) (Woo *et al.*, 2006) and SPRY domain of SSB-2 (C) (Masters *et al.*, 2006) were both recently solved, uncovering their complex beta-stranded sandwich structures that support a core considered to be necessary for protein-protein interaction. With the PRY domain (orange) removed, the similarity between the GUSTAVUS SPRY (B) and SSB-2 SPRY is apparent. Importantly, the MID1 B30.2-like domain is the site of the many of Opitz syndrome patient mutations, where both single amino acid substitutions as well as truncations cause severe patient phenotypes. These are hypothesised to disrupt the folding of the domain which is a complex interlaced network of beta strands (alternatively colour coded purple and red). The image is a stereo view, produced for 3d representation using the “crossed eye” technique.



GUSTAVUS B30.2-like domain



GUSTAVUS SPRY domain



SSB-2 SPRY domain

proteins: e.g. MUL/TRIM37 - Mulbrey Autism; MID1 - Opitz syndrome; Ro/SSA - Sjogren's syndrome; PML - acute promyelocytic leukaemia; and TIF1 and RFP - other forms of tumourigenesis. Although some RBCC proteins are well characterised, the molecular function of most remain poorly understood. The properties of the tripartite motif are a combination of the domains making up the RBCC “module” (in general), and include ubiquitin ligase activity, protein-protein interaction, and self-interaction (homodimerisation). Some of these functions are highlighted in detail in Chapters 2 and 4 (Short and Cox, 2006; Short *et al.*, 2002). Expression of RBCC proteins during development occurs at widely variant time points with different proteins expressed in different tissues of the human body. Despite each superfamily member having essentially homologous N-terminal domains, the sub-cellular localisation and function of these proteins can be vastly different. Early investigation of MID1 function based on random members from the whole RBCC superfamily did not provide any specific evidence for structure/function relationships. This led to the search for new homologues of MID1 and MID2 to see if they could provide any further knowledge of structure/function relationships. This search and its findings are described in detail in the published manuscript in Chapter 4 (Short and Cox, 2006).

1.4.3 MID1 cellular localisation

Direct observation of the cellular localisation of endogenous and over-expressed MID1 protein has shown that it associates with the microtubule cytoskeleton throughout the cell cycle, extending from the microtubule organising centre (MTOC) to the cell membrane (Cainarca *et al.*, 1999; Cox *et al.*, 2000; Schweiger *et al.*, 1999). Furthermore, biochemical evidence has also proven the nascent ability of MID1 to associate with microtubules (MT) both *in vivo* and *in vitro* (Schweiger *et al.*, 1999). Although association

with the MT cytoskeleton has been observed, endogenous levels of MID1 are unlikely to alter MT dynamics since depolymerisation of microtubules is not affected by over-expressed MID1 and these levels of MID1 do not alter the normal distribution of MTs in embryonic fibroblasts, Cos-7, Cos-1 or 293T cells (Cainarca *et al.*, 1999). Similar to the observations from other RBCC family members, MID1 is able to homodimerise and this was initially thought to be primarily responsible for MT association (Short *et al.*, 2002). However, while not excluding a potential role for dimerisation, it was later determined that MT association was actually dependent on a newly identified motif found in MID1 and its closest homologues, and that this domain could even drive a non-microtubule associating RBCC protein to microtubules, as outlined in Chapter 3 (Short and Cox, 2006).

1.4.3.1 Mutations in MID1 alter its cellular localisation

Opitz syndrome-causing mutations in MID1 have been shown to disrupt the normal pattern of association of endogenous and over-expressed MID1 with the MT network, resulting in a dissociated or speckled cytoplasmic appearance (Cox *et al.*, 2000; Quaderi *et al.*, 1997; Schweiger *et al.*, 1999). These observations suggest that complete coverage of the MT cytoskeleton is necessary for MID1 function and thus all mutations result in a loss of function (Cox *et al.*, 2000). Similar to the sub-cellular distribution patterns of some patient mutant MID1 proteins, directed deletion of the B-boxes, FN3, or B30.2 domains also affects MID1 MT localisation, giving a speckled appearance. In many cases, these speckles may still be loosely associated with the MT cytoskeleton, although the nature of the speckled appearance is unknown (Short *et al.*, 2002). Interestingly, an OS patient mutation resulting in the deletion of a single methionine residue in the FN3 domain was found to retain some minor microtubule binding capacity, although the coverage of the microtubules was notably reduced (Schweiger *et al.*, 1999). In comparison with these speckled appearances, deletion of the RING finger does not significantly alter the MT

association of MID1, and only restricts localisation from the cytoskeleton at the cell periphery (Short *et al.*, 2002). In contrast, deletion of a segment of the protein containing the coiled-coil domain (and subsequently found to include part of the newly identified COS box) results in a completely diffuse cytoplasmic pattern, with no evidence of association with the microtubules (Short *et al.*, 2002). This is consistent with an OS patient mutation which results in the deletion of MID1 exon 2, and loss of a large portion of the coiled-coil domain (Cox *et al.*, 2000). Although all mutations tested thus far affect cell localisation, only those affecting the coiled-coil domain disrupt dimerisation of the protein, as shown in the yeast two-hybrid system (Short *et al.*, 2002).

1.4.4 MID2 – a homologue of MID1

MID2 is an X-linked gene that was identified as a homologue of *MID1* shortly after the identification of *MID1* as the X-linked OS gene (Buchner *et al.*, 1999; Perry *et al.*, 1999). The close similarity between *MID1* and *MID2* extends to both genes having an equivalent 9 coding exons each, and the open reading frames having 69% nucleic acid identity, with the proteins encoded having 85% amino acid similarity. Like *MID1*, *MID2* also has homology to RING finger, B-box, Coiled coil, Fibronectin Type III, and B30.2–like domains. The expression pattern of *Mid2* RNA in the developing mouse embryo matches closely with that of *Mid1*, but includes unique features such as a higher level of expression in the heart (Buchner *et al.*, 1999). This was originally thought to be somewhat different to the pattern shown for human *MID1*, with expression restricted to the interventricular septum of the heart as shown in human embryos (Pinson *et al.*, 2004). However, *Mid1* expression in the mouse heart has since been found by the presence of mild X-gal staining in the ventricles and atria of 12 dpc *Mid1*^{TM1(LacZ)AA} knock-out mouse embryos (Saidi Jaafar and Kieran Short, unpublished data). Note, the transgenic nomenclature reads as **T**argeted **M**utation number **1** generated by insertion of the **LacZ**

gene (into *Mid1*), made by Alan Ashworth (Breakthrough Breast Cancer Research Centre, London, U.K.).

1.4.4.1 The evolutionary relationship between MID1 and MID2

Homology between the genes flanking *MID1* at Xp22.3 and the Xq22.3 genes surrounding *MID2* indicates that an ancient intrachromosomal duplication of the Xq sequence may have resulted in the Xp region containing *MID1* and flanking Xq22 homologues (Buchner *et al.*, 1999; Perry *et al.*, 1998). The presence of avian and mammalian *MID1* and *MID2* genes indicates that the duplication of an ancestral *MID* locus likely pre-dated the divergence of mammals and avifauna approximately 310 million years ago (Glazko *et al.*, 2005). While fish *MID* sequences have also been identified, the presence of both *MID1* and *MID2* genes is difficult to identify from current genome resources. Only a single *MID2* gene on Chromosome 9 and a partial gene missing the RING Finger and B-boxes on Chromosome 14 have been found in *Danio rerio* (Zebrafish). In addition, the presence of a single *MID* gene in *Takifugu rubripes* also suggests that either the duplication occurred after the evolution of fish from a common avian/mammalian ancestor or alternatively, that duplication occurred in a common fish/terrestrial animal ancestor, but was subsequently lost in fish. The question of which gene is the result of a duplication of the other is perhaps answered by the higher level of similarity of *MID2* sequences with the more distant RBCC-FN3-B30.2-like homologues, which indicates that *MID1* is more likely the diverged product of an ancestral *MID2* duplication.

1.4.4.2 Subcellular localisation of MID2

The homology between *MID2* and *MID1* at the amino acid level also translates to functional homology. Like *MID1*, *MID2* associates with the microtubule network

throughout the cell cycle (Buchner *et al.*, 1999; Perry *et al.*, 1999), and it has been shown that MID2 can also homodimerise through its coiled-coil domain (Short *et al.*, 2002). Strikingly, the MID1 and MID2 proteins can also interact with one another and this interaction is also mediated by their coiled-coil domains (Short *et al.*, 2002). This intriguing finding raises many interesting questions for a potential role of MID2 in Opitz syndrome, and whether the overlapping expression of the two genes translates to a functional biological context during development, with both proteins potentially required for functional output of a “signalling complex”.

1.4.5 MID1 protein-protein interaction

The MID1 protein has domains which allow it to homodimerise, as well as heterodimerise with MID2 (Short *et al.*, 2002). The trimeric coiled-coil which has been shown to mediate this interaction has also been implicated in association with the MT cytoskeleton (Short *et al.*, 2002). However, the identification of the COS box, a new motif partly deleted in the coiled-coil mutant analysed in Short *et al.* 2002, indicates a reduced role for the coiled-coil in MT localisation (Short and Cox, 2006). There is also significant evidence that MID1 exists in multi-protein complexes, which has been shown by multiple groups using a sucrose gradient technique to isolate large protein complexes specifically associated with over-expressed MID1 protein (Schweiger *et al.*, 1999; Yi Zou, unpublished data). At the onset of the work presented in this thesis, no information was known as to the identity of any proteins which may directly interact with MID1 (other than MID1 itself) (Cainarca *et al.*, 1999). Therefore, an important step in acquiring further information to build an improved understanding of MID1 function was to conduct a protein-protein interaction screen using the MID1 protein as “bait” against a cDNA expression library derived from tissues relevant to Opitz syndrome (tissues where MID1 may be biologically

active). Yeast two-hybrid screens were conducted by several groups, including our own, and the screens identified multiple MID1 protein interactors (Liu *et al.*, 2001; Short *et al.*, 2002; Trockenbacher *et al.*, 2001). In the screen presented within this thesis, a human MID1 cDNA was used as “bait” against 1×10^6 clones from a mouse 10.5 dpc whole embryo cDNA library. This screen identified multiple interacting clones which represented full length or partial transcripts of the *Alpha 4* gene, also known as *IGBP-1* (Short *et al.*, 2002). Trockenbacher and co-workers also reported the MID1-Alpha 4 interaction, after using mammalian MID1 to screen a human foetal brain cDNA library (Trockenbacher *et al.*, 2001). Around the same time as the Trockenbacher and Short publications, Liu and colleagues reported the identification of MID1 in the ‘reverse’ yeast two-hybrid screen using a murine Alpha 4 cDNA as bait with a 9 dpc whole mouse embryonic cDNA library (Liu *et al.*, 2001). The MID1 motif forming the interaction interface with Alpha 4 was identified to be the B-boxes of MID1 (Short *et al.*, 2002), specifically the first B-box motif (Trockenbacher *et al.*, 2001). Consistent with the high amino acid sequence identity between MID1 and MID2 (including their ability to heterodimerise through their coiled-coil motifs) we showed that MID2 also strongly interacted with Alpha 4 (Short *et al.*, 2002).

Analysis of the embryonic expression pattern of Alpha 4 in mice revealed significant overlap with that reported for Mid1 (Everett and Brautigan, 2002). These observations provided additional support that an interaction between Alpha 4 and MID1/MID2 may occur *in vivo* during early embryonic tissue patterning and formation. Furthermore, our group subsequently identified a mutation in the 5’UTR of Alpha 4 in a family with a unique clinical presentation that had similar tissue involvement as seen in many Opitz syndrome cases, although the phenotype was more mild. Cells isolated from the patient showed a decreased level of expression of Alpha 4 (IGBP1) compared with

control fibroblasts, likely as a result of the mutation impeding efficient translation of the *Alpha 4* mRNA (Graham *et al.*, 2003) (Appendix 2). As discussed below, the association of Alpha 4 with both MID1 and MID2 provided the first insight into a potential function for these RBCC proteins.

1.4.5.1 Yeast TAP42 and TOR

The most comprehensive study of Alpha 4 function has been on its yeast orthologue, TAP42, which has approximately 37% amino acid similarity to mammalian Alpha 4. TAP42 is a target of the Target of Rapamycin (TOR, also called FRAP and RAFT) kinase, the activity of which is inhibited by the anti-fungal compound rapamycin, isolated from *Streptomyces hydropscopicus* (Arndt *et al.*, 1999) Also known commercially as Sirolimus, RAPA and Rapamune, the action of rapamycin has received much attention in recent years because, in addition to its anti-fungal properties, it has been found to also have powerful anti-inflammatory, immunosuppressive, and anti-tumour properties (Arndt *et al.*, 1999; Attur *et al.*, 2000; Chen and Fang, 2002; Huang and Houghton, 2001; Law, 2005). Briefly, it has been shown to block mammalian T-cell activation and proliferation, activate p70 S6 kinase and exhibit strong binding to FK-506 binding proteins (Brown *et al.*, 1994; Brown *et al.*, 1995; Sabatini *et al.*, 1994). Notably, rapamycin binds covalently to the FKBP12 prolyl isomerase (Figure 1.8), creating a molecule which competitively interacts with TOR (Brown *et al.*, 1994), in a ‘pocket’ or interface that is necessary for TAP42 phosphorylation and activation (Jiang and Broach, 1999). Under normal conditions, phosphorylation of TAP42 by TOR greatly enhances its association with Sit4/Pph21/Pph22 protein phosphatases, which are homologues of the PP2-type phosphatases such as PP2A in mammals (Di Como and Arndt, 1996). Unlike in stationary phase, yeast in active growth phase exhibit elevated TOR activity which leads to the preferential association between TAP42 and Pph21/22 (Di Como and Arndt, 1996). This

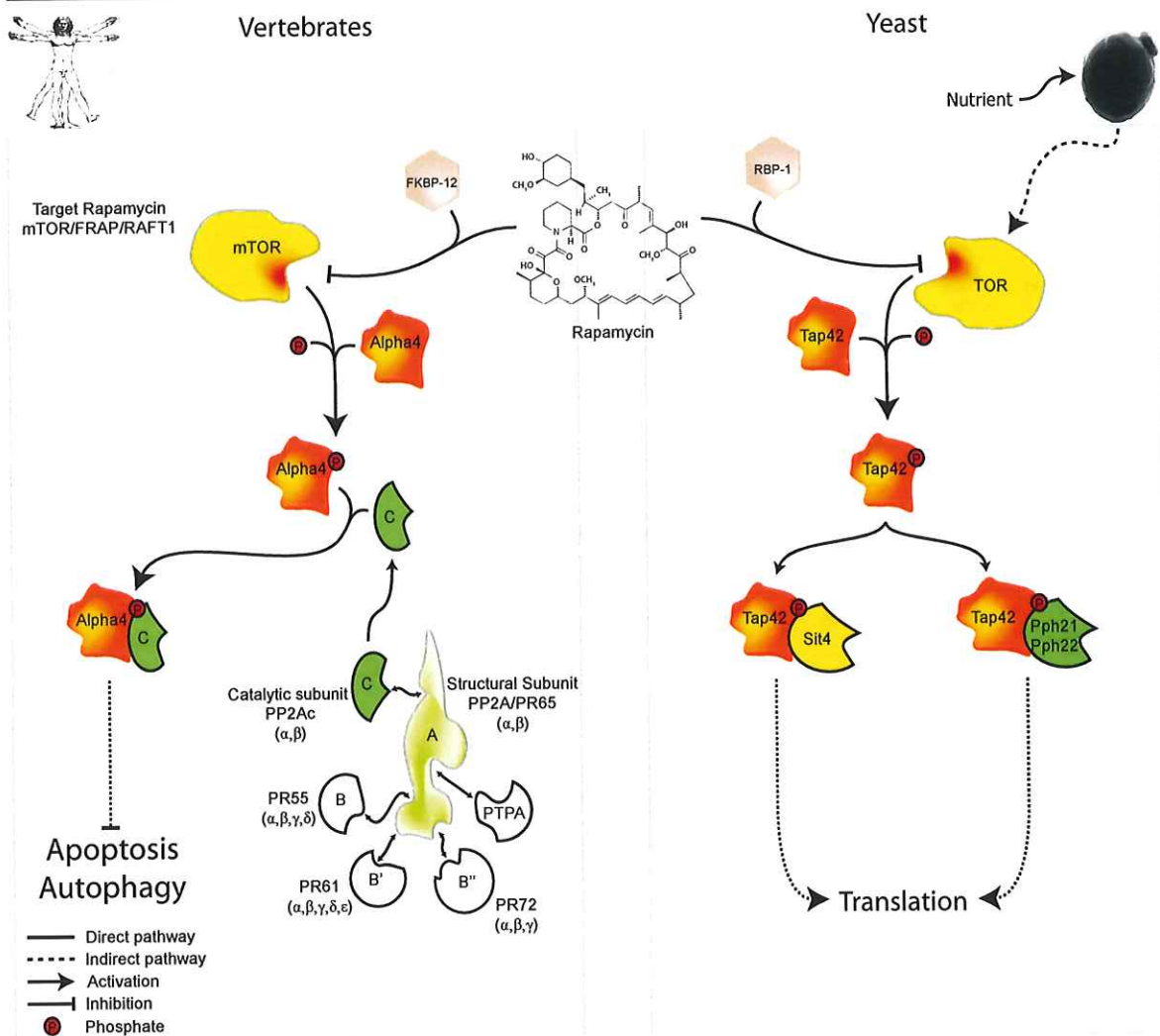


Figure 1.8 Mammalian and yeast Alpha 4/Tap42 rapamycin sensitivity.

Although some groups have indicated that mammalian Alpha 4 displays distinct differences to Tap42 in regard to global affect of the protein (apoptosis versus translation control), there are similarities in initial rapamycin activity. The binding of rapamycin to FKBP12 and RBP1 (homologous rapamycin binding proteins) blocks the activity of the TOR pocket to bind Alpha 4 and Tap42. The reduction in phosphorylation of these proteins as a result is chiefly thought to be the main driving force behind rapamycin sensitivity. Phosphorylation of Alpha 4 and Tap42 is necessary for their interaction with PP2Ac, Pph21/22 and Sit4 homologues where activity of the paired proteins is necessary for translation initiation (yeast) or possibly apoptosis (mammals). It should be noted that the rapamycin pathway *per se* does not control apoptosis. No observations of apoptotic effects post treatment of mammalian cells have been published, but the findings of Kong *et al.* (2004) strongly suggest that a core mammalian Alpha 4 activity is the inhibition of apoptosis, however this does not exclude other activities which could be impacted by rapamycin.

interaction is thought to limit the activity of the phosphatases to specific targets due to preferential binding between TAP42 and the type-2 phosphatase units (Cooper, 2002). However, TAP42-Pph21/22 activity has also been proposed to direct phosphatase activity to targets in yeast which regulate translation in a positive way, promoted by phosphorylation of TAP42 by TOR (Jiang and Broach, 1999). These protein phosphatases control a diverse array of cellular processes, such as the cell cycle and translation initiation – so a lack of control of these phosphatases is thought to be the end-effect of mutations in the rapamycin sensitivity pathway in yeast.

Although TAP42, TOR and Pph21/22 all have homologues in mammals (Alpha 4, mTOR, and PP2A), a distinct interrelationship between Alpha 4 and mTOR indicates that there are differences which set the mammalian and yeast systems apart (Nanahoshi *et al.*, 1998). Recent evidence suggests either lack of conservation between rapamycin-associated Alpha 4 signalling between yeast and metazoan (multicellular) eukaryotes, or multiple pathways of activity diverged from yeast (Cygner *et al.*, 2005; Kong *et al.*, 2004). For example, it has recently been found that *Drosophila melanogaster* Tap42 functions independently of TOR signalling with mutations in Alpha 4 instead showing increases in intracellular JNK signalling, caspase activation and cell death (apoptosis) (Cygner *et al.*, 2005). This demarcates a distinct difference in adaptive function since programmed cell death pathways are very different in filamentous and unicellular fungi (Fedorova *et al.*, 2005) and likely represent a key difference between fungi and other metazoan eukaryotes.

1.4.5.2 Alpha 4 and PP2Ac

Many of the known homologues of Alpha 4 in different eukaryotic kingdoms, such as yeast *Saccharomyces cerevisiae* TAP42 and plant *Arabidopsis thaliana* TAP46, interact with homologues of the protein phosphatase 2A (PP2A) catalytic subunit in their respective systems (Harris *et al.*, 1999). The association of Alpha 4 is not only restricted to

the one phosphatase, as Alpha 4 also associates with protein phosphatases 4 and 6 (Chen *et al.*, 1998; Nanahoshi *et al.*, 1999).

Protein phosphatase 2A exists as a holoenzyme, with a core dimer made up of an A subunit (known as PR65) and a catalytic (C) subunit (PP2Ac), with phosphatase substrate specificity directed by interchangeable regulatory B subunit isoforms (making an ABC holoenzyme, Figure 1.8). While the main AC core dimer is found primarily with B regulatory subunits, activity of the core can be directed by alternative proteins. Such proteins include the phosphotyrosine phosphatase activator (PTPA) protein (Zolnierowicz, 2000) and, in small amounts, Alpha 4 (Murata *et al.*, 1997). This latter interaction is likely a momentary one between an alternative state of PP2Ac association with the rest of the PP2A holoenzyme and a specific interaction between Alpha 4 and PP2Ac (Murata *et al.*, 1997). Alpha 4 is the only non-PP2A component found to associate and function directly with PP2Ac (Murata *et al.*, 1997). The interaction between Alpha 4 and PP2Ac (and other phosphatases) has been shown by some groups to be an inhibitory interaction *in vitro*, with a reduction of phosphatase activity of PP2A, 4 and 6 on pNPP (a non-specific, small organic substrate used as a general tool to measure phosphatase activity) and 4E-BP1 (Nanahoshi *et al.*, 1998; Nanahoshi *et al.*, 1999). Interestingly, Alpha 4 has also been shown to augment the activity of PP2Ac within some cells, increasing phosphatase activity to several PP2Ac targets such as phospho-MBP, -Histone, -Casein and -eEF2 (Chung and Brautigan, 1999; Inui *et al.*, 1998). These conflicting reports for a general function of the Alpha 4-PP2Ac interaction could make analysis of the interaction in signal transduction pathways difficult. However, the true nature of the PP2A-Alpha 4 interaction is not simply for inhibition or activation of PP2Ac activity, but rather is substrate-specific, with some substrates selectively targeted for dephosphorylation, while the phosphatase action toward other PP2Ac substrates is specifically inhibited (Prickett and Brautigan, 2004).

Rapamycin sensitivity of the Alpha 4-PP2Ac interaction in mammalian cells has been another area where experimental conclusions are seemingly contradictory. Multiple groups have noted both rapamycin sensitive and insensitive Alpha 4-PP2Ac interactions. For instance, some cell culture lines (Cos-7 and Jurkat) exhibit Alpha 4-PP2Ac interaction sensitivity to rapamycin (Inui *et al.*, 1998; Murata *et al.*, 1997). However others have shown that particular cell lines (for example, Raji, 293 (complete and serum starved media), HEK293 and Cos-M6) indicate that the Alpha 4-PP2Ac interaction is not sensitive to rapamycin treatment (Chen *et al.*, 1998; Inui *et al.*, 1998; Kloeker *et al.*, 2003; Nanahoshi *et al.*, 1998). Furthermore, a yeast three-hybrid assay has been performed within our lab to show an interaction between MID1, Alpha 4 and PP2Ac – in an environment without mTOR (although phosphorylation of mammalian Alpha 4 by the 57% identical yeast TOR proteins has not been tested), also indicating that an Alpha 4-PP2Ac interaction may occur without phosphorylation of Alpha 4 (Blair Hopwood, unpublished data). The ability of yeast TOR proteins to actively phosphorylate yeast TAP42 and induce interaction with Sit4/Pph21/Pph22 is therefore likely to differ from that occurring in the mammalian system. Indeed, there has been little evidence to show an *in vitro* nascent ability of mTOR to phosphorylate Alpha 4, which would likely be the homologous point in the rapamycin sensitivity pathway in yeast where rapamycin-FKBP12 blocks phosphorylation of TAP42. Demonstrating that there are indeed other pathways activating Alpha 4, phosphorylation of Alpha 4 in mammalian systems has been shown to be stimulated through Protein Kinase C (PKC) pathways since the tumour promoter Phorbol-12-myristate-13-acetate (PMA/TPA, which is an activator of PKC) can stimulate phosphorylation of murine Alpha 4 (Kuwahara *et al.*, 1994).

Curiously, while increasing evidence may suggest that the Alpha 4 – PP2Ac interaction may be rapamycin insensitive, the interaction is inhibited *in vitro* by the serine-

threonine phosphatase inhibitors, Okadaic acid and microcystin (Kloeker *et al.*, 2003). This indicates that phosphatase activity may either regulate the interaction of Alpha 4 and PP2Ac through targeting the activity of regulatory kinases, or alternatively by directly altering the activity of PP2Ac through displacement or disruption of the PP2Ac structure, therefore blocking a PP2Ac-Alpha 4 interaction (Kloeker *et al.*, 2003).

Regardless of the nature of the interaction between the two proteins, whether Alpha 4 requires direct phosphorylation by mTOR or not, the downstream targets of Alpha 4-PP2Ac in mammalian systems are likely to be very different to yeast, due to the vast differences and complexities in signalling systems. The most compelling evidence for intracellular function of effectors downstream of mammalian Alpha 4-PP2Ac interaction, as distinct from its yeast homologue TAP42, has come from recent research highlighting a role for Alpha 4 and PP2Ac in the regulation of phosphorylation of transcription factors c-Jun and p53 and puts Alpha 4-PP2Ac interaction at the forefront of regulating apoptosis (Kong *et al.*, 2004). These findings, together with the observation that *Drosophila melanogaster* TAP42 deletion increases JNK signalling, caspase activation, and apoptosis provides strong evidence that animal Tap42/Alpha 4 has a protective role against apoptosis (Cygner *et al.*, 2005; Kong *et al.*, 2004).

In yeast, new evidence for complex formation between yeast homologues of PTPA (Rrd1 and Rrd2) with TAP42/PP2A-like phosphatases has been shown, with rapamycin treatment dissociating TAP42 from PP2Ac/Rrd(1/2) dimers. In addition to this, mutations in Rrd2 and Rrd1 render yeast rapamycin insensitive (Zheng and Jiang, 2005). This indicates that in yeast, the PTPA proteins are integral to the PP2A-TAP42 signalling pathway, and are likely key mediators. While conservation of this interaction from yeast to higher eukaryotes is possible, no such interaction between PPP2R4 (PP2A regulatory subunit B' PR53 - the human orthologue of yeast PTPA) and Alpha 4-PP2Ac subunits has

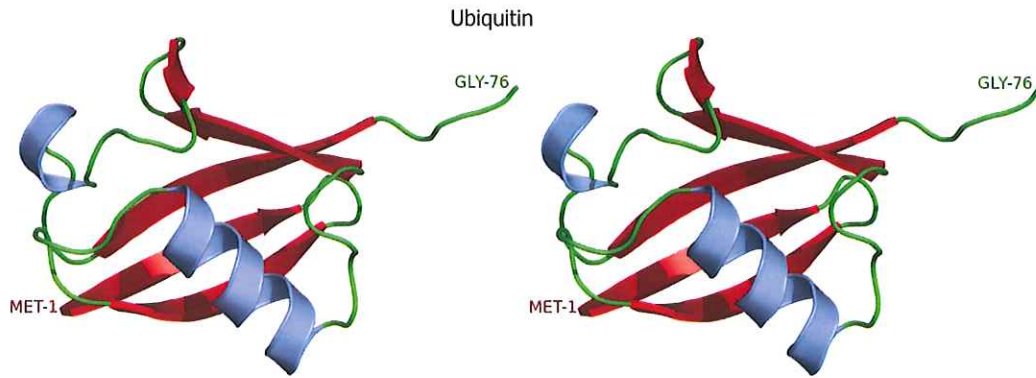
yet been identified.

1.4.5.3 Ubiquitin E3 Ligase activity of MID1 and its target

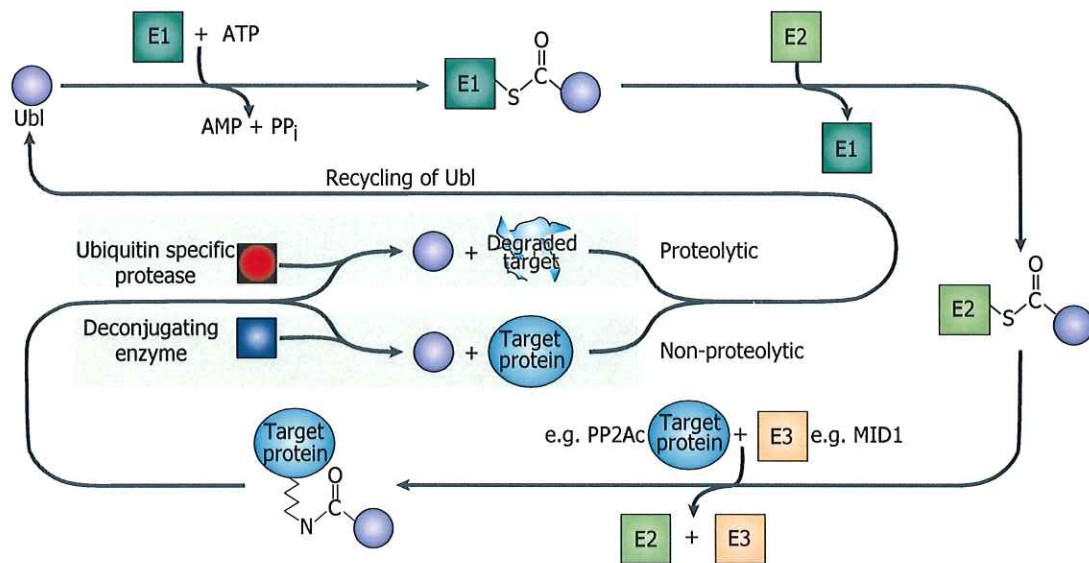
Posttranslational modification by ubiquitylation is a flexible mechanism for modification of the structure, function and localisation of target proteins, and is the primary mode of directed proteasome-mediated proteolysis. Several proteins are necessary for the process of ubiquitylation (Figure 1.9), including ubiquitin itself, the ubiquitin activating enzyme (E1), the ubiquitin conjugating enzyme (E2), and a ubiquitin ligase (E3). Additionally, the activity of an E4 enzyme in yeast and mammals has also been identified which assists ubiquitin chain assembly for multiubiquitin ligation necessary for proteasomal degradation of some proteins (Koegl *et al.*, 1999).

Ubiquitin was the first moiety discovered to be a post-translational modifier of protein function (Hershko *et al.*, 1980; Hershko *et al.*, 1981), however, since then multiple other Ubiquitin-like units (called 'ubiquitons') which similarly affect the function of proteins have been identified. The uses for the additional ubiquitin-like moieties is varied; NEDD8 is involved in the regulation of E3 enzymes (Xirodimas *et al.*, 2004); SUMO1, 2 and 3 are all key modification enzymes in the processes of nuclear localisation, transcriptional regulation and antagonising ubiquitylation (Eaton and Sealy, 2003; Gill, 2005); ISG15/UCRP is active in immune responses and interferon signal transduction (Malakhova *et al.*, 2003); the FAT10 moiety is involved in apoptosis and cytokine signal transduction (Raasi *et al.*, 2001); FUB1 has been shown to be required for T-cell activation (Nakamura *et al.*, 1995); UBL5 is used in pre-mRNA splicing (Dittmar *et al.*, 2002); Urm1 is involved in the oxidative stress response and nutrient sensing through the TOR signalling pathway (Goehring *et al.*, 2003); and both Atg8 (Ichimura *et al.*, 2000) and Atg12 (Mizushima *et al.*, 1998) are important for autophagy and cytoplasm-to-vacuole targeting. It has been proposed that ubiquitons may have evolved over millions of years

A



B General ubiquitin-like-protein conjugation cascade*



*The E3 usually facilitates direct Ubl transfer from the E2 to the target protein without forming a thioester bond itself, although, in a minority of cases, E3s directly form thioesters with the Ubl before the Ubl is conjugated to the target protein. In addition, transfer sometimes occurs without the involvement of an E3.

Figure 1.9 Ubiquitin and Ubiquitylation by E1, E2 and E3 enzymes.

The Ubiquitin moiety is highly conserved throughout all eukaryotes, and is a 76 amino acid compact structure (A; rendered from 1UBQ (Vijay-Kumar *et al.*, 1987)). The covalent attachment of ubiquitin is typically used to mark proteins for proteolytic degradation, but there are also non-proteolytic functions ascribed to covalent ubiquitin attachment. Ubiquitylation occurs through a progressive recruitment from the activating enzyme E1 to E2 conjugating enzymes through to E3 ligation enzymes (although some E3 enzymes do not come into direct contact with ubiquitin), where ubiquitin is covalently linked from one of its seven lysine residues to a target protein (B; modified from (Welchman *et al.*, 2005)). In the case of Opitz syndrome, E3 ubiquitin ligase activity of MID1 is thought to be reduced or absent, such that PP2Ac is no longer targeted for degradation by the proteasome.

from a simple molecular tagging mechanism in cyanobacteria to a wide-ranging system regulating multiple signalling pathways in eukaryotes (Hochstrasser, 2000). Ubiquitons are such long standing components of vertebrate cellular signalling that many viruses actually co-opt them during viral infection (Shackelford and Pagano, 2005).

RING finger proteins like MID1, such as Siah-1, BRCA1, Praja1, TRC8, NF-X and kf-1, have been implicated as E3 ubiquitin ligases that facilitate E2-dependent ubiquitylation of multiple protein targets (Lorick *et al.*, 1999). The RING finger of the proto-oncogene c-Cbl has also been shown to be the domain responsible for the E3 ubiquitin ligase activity of this protein towards most tyrosine kinase receptors (Thien and Langdon, 2001). Although compelling evidence indicating a likely role for MID1 in E3 ubiquitin ligase activity on PP2Ac has been provided (Troockenbacher *et al.*, 2001), ubiquitin ligase activity of MID1 has not been shown directly by *in vitro* ubiquitylation assays. The observation of an increased amount of localised PP2Ac in microtubule fractions in Opitz syndrome-derived cell lines (compared with controls), along with hypophosphorylation of microtubule proteins and a reduction in the amount of polyubiquitylated PP2Ac in the same patient cells has been the evidence supporting the involvement of MID1 as an E3 ubiquitin ligase targeting PP2Ac (Troockenbacher *et al.*, 2001). However, at this stage no published data is available to indicate a direct interaction with E2 ubiquitin conjugating enzymes, although work within our laboratory indicates a likely interaction between UbcH7 (an E2 conjugating enzyme) and MID1 by yeast two-hybrid and cell localisation studies (Blair Hopwood, unpublished observations). This evidence, with further studies using *in vitro* ubiquitylation, will likely resolve the exact nature of any MID1 ligase activity, although *in vitro* expression of MID1 is difficult due to the relatively insoluble nature of the protein and the complexity of additional chaperones required for proper tertiary folding of the protein structure (M.A. Massiah, personal

communication).

1.4.5.4 Other MID1 interacting proteins

Other confirmed interacting proteins of MID1 identified by us and others include Calpain1, Ubiquitin B, HoxA11 (Kieran Short, unpublished observations), and Mig12 (Berti *et al.*, 2004). Whilst all are not necessarily physiologically-proven interactions, they have been tested using mammalian cells *in vivo* and *in vitro*. Expressed in the neuroepithelial midline, urogenital region and digits during development, Mig12 has also been shown to interact with MID1 through the coiled coil domain, and is recruited to microtubules when co-expressed in cell culture (Berti *et al.*, 2004).

1.5 Molecular pathways and Opitz syndrome

Clues as to the nature of the cellular defect(s) underlying the malformations in Opitz syndrome can be gleaned by appreciating the complex embryological processes leading to development of each tissue. For example, fusion of the lip and primary palate, enclosure of the male urethra, septation of the cardiac tissues through fusion of endocardial cushions, and formation of the rectum and anal openings all require the epithelia lining the early tissue primordia to be removed, remodelled or transformed into mesenchyme (Cox, 2004; Hynes and Fraher, 2004c; Nievelstein *et al.*, 1998). In all cases, this break down, dissolution, conversion, or death of the surface epithelium is an absolute requirement in order for proper morphogenesis of the tissue to occur; failure to eliminate the epithelia can result in a physical barrier to subsequent mesenchymal cell movements (Cox, 2004; Hynes and Fraher, 2004c; Nievelstein *et al.*, 1998; Person *et al.*, 2005).

1.5.1 Epithelial-mesenchymal transition (EMT) and Apoptosis

Epithelial-mesenchymal transition (EMT) and apoptosis (programmed cell death) are two of the principal methods to eliminate or move epithelia and both have been proposed to act in the removal of epithelial barriers during tissue fusion processes, such as in the fusion of the palate (Kang and Svoboda, 2005). EMT is an important developmental mechanism that describes the conversion of relatively fixed, immotile epithelial cells to highly motile (mesenchyme) cells that have reduced cell-cell contacts. The process of EMT is critical for many development events in vertebrates including gastrulation (Hay, 2005) and heart morphogenesis (Person *et al.*, 2005). In contrast, apoptosis is a fundamental developmental and homeostatic mechanism used to enact the controlled removal of cells that are functionally redundant or are otherwise detrimental to the organism. Developmentally, apoptosis is well known for its role in neural tube fusion during the neurulation stage of gastrulation (Harris and Juriloff, 1999) as well as its role in organogenesis and tissue remodelling (Penalosa *et al.*, 2006). Examples of organs and complex vertebrate body shapes which require apoptosis include the cardiovascular system (Poelmann and Gittenberger-de Groot, 2005), the eye lens (Yan *et al.*, 2006) and the interdigital tissue between the digits (Mirkes, 2002).

In the context of Opitz syndrome, affected tissues such as the heart, genitalia, face and anus that at first glance might seem developmentally unconnected, do in fact share similar requirements for dramatic changes to the properties of particular epithelia. These similar embryological or developmental events and the association with expression of MID1 in such epithelia suggest an underlying unified molecular pathway in the processes. Therefore, it is attractive to consider that the cellular effect of mutations in MID1 in these epithelia might directly impact on the process of EMT or apoptosis in the morphogenesis of the affected tissues. This is particularly intriguing given the important role microtubules

and PP2A play in the EMT process and the recent association of Alpha 4 in regulating cellular apoptosis (Kong *et al.*, 2004). These areas will be the focus of future work on MID1 and were not directly investigated as part of this thesis work.

1.6 Significance and aims of the project

Opitz syndrome is a complex multi-organ phenotype linked by potentially similar anomalous cellular transformations during development. Many of the individual malformations seen in OS are also commonly seen as isolated malformations in the population, and are representative of some of the most common developmental anomalies in humans. For example, hypospadias (present in up to 70% of male OS patients) and cleft lip with or without cleft primary palate (exhibited by up to 50% of OS patients) are the third and fourth most common birth defects and affect up to 1 in every 350 or 600 live births worldwide, respectively. Therefore, understanding the molecular interactions and pathways behind Opitz syndrome will not only provide a better appreciation of OS disease pathogenesis, but should also assist in understanding the molecular events behind isolated developmental malformations in non-syndromal contexts.

This project investigated the molecular interactions of the X-linked Opitz syndrome gene product, MID1, and used the information to identify potential molecular pathways where it might be acting during development. In addition, a detailed understanding of the structure/function relationship of MID1 was undertaken by searching for homologues and studying them to reveal common functional properties as well as use of a computational approach to identify conserved regions within the proteins. This analysis has provided the basis for a more detailed understanding of how MID1 (and its homologues) may have evolved and what functions it has, particularly those which may be relevant to the pathogenesis of Opitz syndrome.

The aims of this project were:

1. to identify developmentally relevant protein interactors of MID1 that might provide clues as to the cellular processes in which it is involved. This was to be achieved using a powerful genetic approach, the yeast two-hybrid assay.
2. to study the effect of mutations on MID1 cell localisation and protein-protein interaction using immunofluorescence, co-immunoprecipitation, and yeast two-hybrid analysis.
3. using bioinformatic means, to identify homologues of MID1 and use computational analysis of protein conservation as a platform to study structure/function relationships across the subfamily.

Due to the publication of the work carried out during the progress of this study, this thesis is presented as a portfolio of publications. Publications related to the body of the thesis work to which the candidate significantly contributed but was not primary author are included in the appendices.

Chapter Two
Short *et al.*, 2002

Research article

MIDI and MID2 homo- and heterodimerise to tether the rapamycin-sensitive PP2A regulatory subunit, Alpha 4, to microtubules: implications for the clinical variability of X-linked Opitz GBBB syndrome and other developmental disorders

Kieran M Short¹, Blair Hopwood¹, Zou Yi¹ and Timothy C Cox*^{1,2}

Address: ¹Department of Molecular Biosciences & ARC Special Research Centre for the Molecular Genetics of Development, University of Adelaide, Adelaide, South Australia, Australia 5005 and ²South Australian Clinical Genetics Service, Women's & Children's Hospital, North Adelaide, South Australia, Australia 5006

E-mail: Kieran M Short - kieran.short@adelaide.edu.au; Blair Hopwood - blair.hopwood@adelaide.edu.au; Zou Yi - julie.zou@adelaide.edu.au; Timothy C Cox* - timothy.cox@adelaide.edu.au

*Corresponding author

Published: 4 January 2002

Received: 5 November 2001

BMC Cell Biology 2002, 3:1

Accepted: 4 January 2002

This article is available from: <http://www.biomedcentral.com/1471-2121/3/1>

© 2002 Short et al; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any non-commercial purpose, provided this notice is preserved along with the article's original URL. For commercial use, contact info@biomedcentral.com

Abstract

Background: Patients with Opitz GBBB syndrome present with a variable array of developmental defects including craniofacial, cardiac, and genital anomalies. Mutations in the X-linked *MIDI* gene, which encodes a microtubule-binding protein, have been found in ~50% of Opitz GBBB syndrome patients consistent with the genetically heterogeneous nature of the disorder. A protein highly related to *MIDI*, called *MID2*, has also been described that similarly associates with microtubules.

Results: To identify protein partners of *MIDI* and *MID2* we undertook two separate yeast two-hybrid screens. Using this system we identified Alpha 4, a regulatory subunit of PP2-type phosphatases and a key component of the rapamycin-sensitive signaling pathway, as a strong interactor of both proteins. Analysis of domain-specific deletions has shown that the B-boxes of both *MIDI* and *MID2* mediate the interaction with Alpha 4, the first demonstration in an RBCC protein of a specific role for the B-box region. In addition, we show that the *MIDI/2* coiled-coil motifs mediate both homo- and hetero-dimerisation, and that dimerisation is a prerequisite for association of the *MID*-Alpha 4 complex with microtubules.

Conclusions: Our findings not only implicate Alpha 4 in the pathogenesis of Opitz GBBB syndrome but also support our earlier hypothesis that *MID2* is a modifier of the X-linked phenotype. Of further note is the observation that Alpha 4 maps to Xq13 within the region showing linkage to FG (Opitz-Kaveggia) syndrome. Overlap in the clinical features of FG and Opitz GBBB syndromes warrants investigation of Alpha 4 as a candidate for causing FG syndrome.

Background

Opitz GBBB syndrome (OS; Opitz syndrome) is a genetically and phenotypically complex disorder defined by characteristic facial anomalies (hypertelorism and varia-

bly labiopalatine and laryngotracheo-esophageal (LTE) clefting), structural heart defects, as well as anal and genital anomalies [1,2]. Recently, we and others identified the *MIDI* gene (also called *FXY*) as the underlying cause of

the X-linked form of the disease [3–5]. Defects in *MID1* have been found in ~50% of OS cases consistent with evidence from genetic linkage and cytogenetic studies that at least one autosomal form of the disorder, at chromosome position 22q11.2, also exists [6–8]. The deletion of the same interval produces the 22q11 deletion syndrome, which encompasses a group of disorders (eg. DiGeorge and velocardiofacial syndromes) showing some phenotypic overlap with OS [6,9]. Collectively, 22q11 anomalies represent one of the most common genetic causes of malformations (estimated 1 in 5000 live births) [10]. Although progress has recently been made towards elucidating the genes contributing to the 22q11 deletion phenotype [11–14], a specific autosomal OS gene has not yet been identified.

The *MID1* gene encodes the defining member of a new subclass of the RBCC (RING, B-box, Coiled-Coil) family of proteins. This subclass is characterised by the combination of both a fibronectin type III motif and a B30.2-like (or SPRY) domain positioned C-terminal to the RBCC domain [15,16]. As in other members of the RBCC protein family, *MID1* forms multiprotein complexes of between 250 and 450 kDa [17]. The *MID1* protein, presumably as part of these complexes, has been shown to associate with cytoplasmic microtubules along their length and throughout the cell cycle using immunofluorescence detection of endogenous *MID1*, transient expression of GFP-*MID1* fusion proteins, and cellular fractionation [5,18]. Most mutations in *MID1* that cause OS are truncating mutations with many, but not all, directly affecting the C-terminal half of the protein [5]. Interestingly, all examined *MID1* mutations disrupt the normal microtubule-associated distribution and this has been demonstrated to occur *in vivo* with endogenous mutant protein and in transient transfection studies using GFP fusion proteins [5,17,18].

An intriguing aspect of OS is the marked intrafamilial variability seen even among related males with the same X-linked mutation. This observation may be explained by the fact that other proteins, such as the highly related *MID2* protein that is expressed in some of the same tissues and also associates with the microtubule network [15,19], could at least partially compensate for the loss of *MID1* [5]. Such a mechanism has also been proposed for other microtubule-associated proteins, such as tau [20]. Alternatively, variations in the level or action of other factors in the same molecular pathway (for example, a factor encoded by the 22q11.2 locus and/or another component of the *MID1* macromolecular complex) could contribute to this variability. Both explanations are consistent with the conclusion that OS is caused by loss of function of *MID1* [5].

In this paper we report the identification of Alpha 4 as an interacting partner of both the Opitz syndrome protein,

MID1, and the highly related *MID2* protein. Alpha 4 is a unique and highly conserved regulatory subunit of PP2-type phosphatases, such as PP2A [21–24], and an integral component of the rapamycin-sensitive signaling pathway [25]. Our finding that both *MID1* and *MID2*, either as homo- or hetero-dimers, are able to tether a key regulator of intracellular signaling to microtubules has significant implications for our understanding of the pathophysiological basis of OS and provides additional support for the hypothesis that *MID2* could act as a modifier of the OS phenotype. Furthermore, as the *Alpha 4* gene maps to Xq13 [26], our results identify it as a candidate for other X-linked disorders such as FG (Opitz-Kaveggia) syndrome that overlaps clinically with Opitz GBBB syndrome.

Results

Yeast two-hybrid screens identify Alpha 4 as an interacting partner of both *MID1* and *MID2*

In order to begin to determine the identity of potential interacting factors and the processes with which *MID1* is involved, we performed a yeast two-hybrid assay utilising the full length *MID1* protein (which shares 94.9% identity, 99.9% similarity with mouse *Mid1*) as bait in a screen of a mouse 10.5 dpc (days post-coitum) whole embryo cDNA library. At 10.5 dpc, expression of *Mid1* is seen essentially throughout the embryo although strongest levels of expression are evident in highly proliferating tissues such as the developing craniofacial region [27]. The use of the mouse 10.5 dpc whole embryo library was therefore chosen to maximise the likelihood of detecting functionally relevant interactors.

Approximately 1×10^6 cDNAs were initially screened using *MID1* as bait. Potential interacting clones were selected based on the activation of the three endogenous reporter genes (*His*, *lacZ*, and *Leu*). Recovery and sequencing of the positives from the library screen demonstrated nine cDNA clones with insert sizes ranging from 1.0 kb to 1.3 kb but representing the same gene. Five additional, singly represented putative interacting clones were also identified but do not constitute part of this report. Database searches identified the nine similar clones as encoding Alpha 4, a rapamycin-sensitive regulatory subunit of protein phosphatases 2A (PP2A) and other PP2-type phosphatases. By comparison with the published murine *Alpha 4* sequence, all clones were judged as being essentially full-length. Notably, the nine in-frame cDNAs represented a minimum of seven independent clones as 5' fusion to the GAL4 activation domain occurred at either nucleotide +4, +7, +10 or +13 and most clones containing different length polyadenylated tails. To confirm that this interaction was not an artefact of the independent GAL4 activation domain and DNA binding domain fusion events, the full-length cDNAs were interchanged such that Alpha 4 was fused to the GAL4-DBD and *MID1* to GAL4-

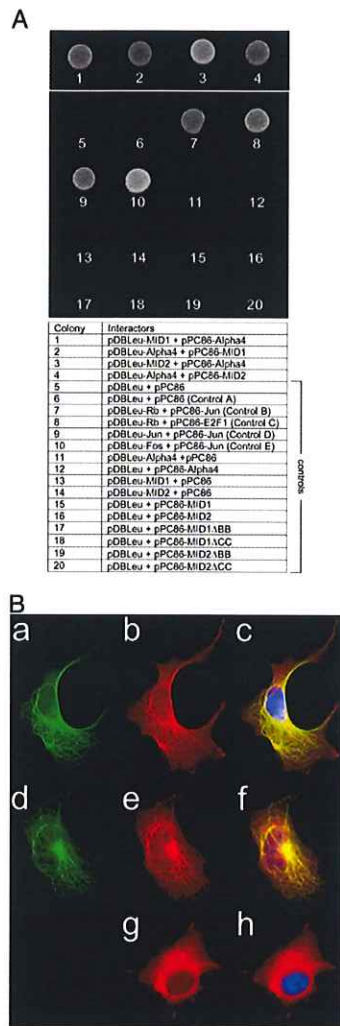


Figure 1
Alpha 4 interacts with MID1 and MID2. (A) Yeast two-hybrid analysis of the interaction between MID1 and Alpha 4 as well as MID2 and Alpha 4. Yeast agar plate (leu⁻ trp⁻ his⁻, 75 mM 3-AT) showing growth for MID1/Alpha 4 and MID2/Alpha 4 interactions as well as positive control two-hybrid combinations and no growth for negative controls. (B) Detection of full-length myc tagged-Alpha 4 when co-expressed with GFP-MID1 and GFP-MID2 fusion proteins in transiently transfected Cos1 cells. (a) GFP-MID1 fluorescence (green), (b) anti-myc antibody detecting myc-Alpha 4 localisation (red), (c) overlay of (a), (b) showing co-localisation of the myc-Alpha 4 fusion protein and GFP-MID1, with DAPI stain for DNA (blue). (d) GFP-MID2 fluorescence (green), (e) myc-Alpha 4 localisation (using same detection as b) (red), (f) overlay of (d), (e) with DAPI (blue) showing co-localisation of the myc-Alpha 4 fusion protein and GFP-MID2. (g) Detection of transiently expressed myc-Alpha 4 fusion protein in Cos1 cells, (h) overlay of (g) and DAPI stain showing cytoplasmic distribution of myc-Alpha 4 fusion protein.

AD. Co-transformation of these two constructs into MaV203 gave similar levels of growth on 3AT (Fig 1A), ie. a level of activation comparable to the strongest interacting proteins, Fos and Jun (Fig 1A, Control E – colony 10).

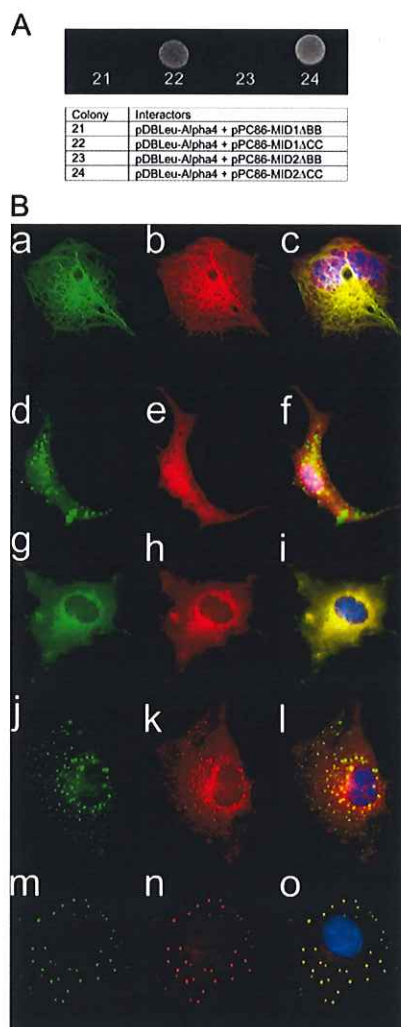
Given that MID1 and MID2 share 77% overall amino acid identity (92% similarity), we wished to investigate whether MID2 was capable of binding to the same proteins, such as Alpha 4, or whether this interaction was specific for MID1. This was investigated using two approaches: a direct yeast two-hybrid test of the ability of MID2 to bind Alpha 4, and a full two-hybrid screen of the 10.5 dpc whole mouse embryo cDNA library this time using MID2 as bait. The direct test demonstrated that MID2 did indeed interact with Alpha 4. Comparison of the growth of the MID2/Alpha 4 transformed yeast on plates (SC -Leu, -Trp, -His) containing 75 mM 3AT indicated the interaction between MID2 and Alpha 4 was as strong, or stronger, than that observed for MID1 (Fig 1A). A similar result was obtained when assessing the conversion of Xgal (data not shown). In the full yeast two-hybrid screen, some potentially novel interactors were identified for MID2. However, the majority of the putative interacting clones, and again the only sequences represented more than once in the isolates, represented Alpha 4.

MID1 and MID2 tether Alpha 4 to the microtubules

To investigate whether Alpha 4 also associated with MID1 and MID2 in a homologous (mammalian) system, we co-transfected both GFP-MID1 (or MID2) with myc epitope-tagged Alpha 4. Transfection of the myc-tagged Alpha 4 expression construct alone resulted in a diffuse distribution of the myc-Alpha 4 fusion protein throughout the cytoplasm in most cells (Fig 1B; g, h). However, in some cells, a faint filamentous distribution that resembled the appearance of microtubules could be seen along with the cytoplasmic protein (data not shown). In contrast, in all cells co-transfected with GFP-MID1 and myc-tagged Alpha 4, there was negligible diffuse cytoplasmic staining. Instead, essentially all the myc-tagged Alpha 4 protein displayed a filamentous meshwork of staining that completely co-localised with MID1 along the length of the microtubules (Fig 1B; a-c). This result also suggested that MID1 was a limiting factor in the tethering of Alpha 4 to the microtubule network, a conclusion supported by western analysis that shows low levels of MID1 in these cells (data not shown). Co-expression of GFP-MID2 and myc-tagged Alpha 4 similarly resulted in co-localisation along the microtubules (Fig 1B; d-f).

Alpha 4 does not co-localise with MID1 or MID2 proteins harboring in-frame B-box deletions

Both endogenous mutant MID1 protein in OS patient cells [18] and various transiently expressed mutant MID1-GFP fusion proteins form cytoplasmic clumps

**Figure 2**

MID1/Alpha 4 and MID2/Alpha 4 interactions are maintained in all MID domain-specific deletions except for those involving the B-boxes. (A) Yeast two hybrid analysis shows that the MID B-boxes are required for interaction with Alpha 4. The interaction of Alpha 4 with MID1 Δ BB (21), or MID2 Δ BB (23), is compared to its interaction with MID1 Δ CC (22), or MID2 Δ CC (24). (B) Subcellular localisation of myc tagged-Alpha 4 when co-expressed in Cos-1 cells with MID1 domain-specific deletions as GFP fusion proteins. Fluorescence detection of GFP-MID1 Δ RF (a), GFP-MID1 Δ BB (d), GFP-MID1 Δ CC (g), GFP-MID1 Δ FNIII (j), and GFP-MID1 Δ CTD (m). Anti-myc antibody detection of myc-Alpha 4 in the same cells as expressing the various MID1 domain-specific deletions (b,e,h,k,n). Overlay of the GFP and anti-myc images of the same cells merged with DAPI stain of nuclei (c,f,i,l,o). All merged images, with the exception of (f) show co-localisation of myc-Alpha 4 with the various MID1 domain deletions. In (f), myc-Alpha 4 fails to co-localise with GFP-MID1 Δ BB in small cytoplasmic aggregates.

[5,15,17,18]. We chose to exploit this previous observation by co-transfecting GFP-tagged MID1 Δ CTD (or MID2 Δ CTD) with a construct expressing a myc-tagged Alpha 4 protein in order to investigate whether Alpha 4 still remained bound to MID1 within such aggregates. The results clearly showed a distribution of myc-Alpha 4 that was indistinguishable from the clumped MID1 Δ CTD and MID2 Δ CTD truncated proteins, indicating that Alpha 4 indeed aggregates with the truncated MID1 and MID2 proteins (Fig 2B; m-o).

To further define the motif in MID1 (and MID2) responsible for the interaction with Alpha 4, we generated in-frame deletions of all other motifs (Table 1) and fused the resulting clones to GFP. We initially transfected each construct alone to examine the effect of each deletion on the intracellular localisation of the proteins. The results showed that each motif, or at least their conserved spacing or arrangement, was essential for the distribution of MID1 (and MID2) along the length of the microtubules. The distribution of the individual domain deleted MID proteins when transfected alone was indistinguishable from their distributions when transfected along with myc-Alpha 4 (see below). Consequently, the GFP fluorescence images in Fig 2B can also be considered a representation of each distribution pattern in the absence of co-transfected myc-Alpha 4. Like the Δ CTD constructs, deletion of either the B-boxes or the FNIII domain also resulted in cytoplasmic clumps or speckles although these usually appeared smaller and greater in number and the speckles still appeared to co-localise with microtubules (data not shown). Both Δ RING proteins, in contrast, showed variability in their distribution with most transfected cells still showing association along the length of the microtubules. However, the microtubule association of these Δ RING proteins often did not extend to the cell periphery. Notably, deletion of the coiled-coil motif in each protein resulted a diffuse cytoplasmic distribution, suggesting that these proteins had lost their ability to associate with microtubules (see Fig 2B; g-i).

We then individually co-transfected each MID1 (and MID2) deletion construct with myc-Alpha 4 in an attempt to define the domain responsible for the interaction with Alpha 4. Co-transfection of either the Δ RING or Δ FNIII proteins with myc-Alpha 4 (Fig 2B; a-c and j-l, respectively) resulted in co-localisation of Alpha 4 with the abnormally distributed MID1 and MID2 proteins, as seen with the Δ CTD proteins. Co-expression of either Δ CC construct with Alpha 4 resulted in both proteins exhibiting a diffuse cytoplasmic distribution although the pattern of each was still suggestive of the two proteins being able to interact (Fig 2B; g-i). Strikingly, when either of the Δ BB constructs was co-expressed with Alpha 4, the mutant MID1 and MID2 proteins still formed cytoplasmic clumps but, in

Table 1: The MID1 and MID2 deletion constructs used in pEGFP-C2 for cellular co-localisation analysis, co-immunoprecipitation and in pPC86/pDBLeu for interaction analysis with Alpha 4. Terms: RF denotes a C3HC4 RING finger; BB denotes C2H2 B-Boxes; CC denotes a Coiled-coil motif; FNIII denotes a Fibronectin Type III domain and CTD denotes a C-terminal domain (encompassing a SPRY domain).

Construct	Amino Acids deleted
Mid1 Δ RF	1-69
Mid1 Δ BB	71-213
Mid1 Δ CC	214-349
Mid1 Δ FNIII	370-473
Mid1 Δ CTD	483-667
Mid2 Δ RF	1-69
Mid2 Δ BB	71-213
Mid2 Δ CC	214-349
Mid2 Δ FNIII	370-473
Mid2 Δ CTD	483-686
BB(Mid1)	1-70 and 214-667
CC(Mid1)	1-213 and 350-667
BBCC(Mid1)	1-70 and 350-667

both cases, Alpha 4 remained diffuse in the cytoplasm (Fig 2B; d-f). These results imply that the B-boxes (and/or the linker residues between the B-boxes and RING motifs) are primarily responsible for the interaction with Alpha 4.

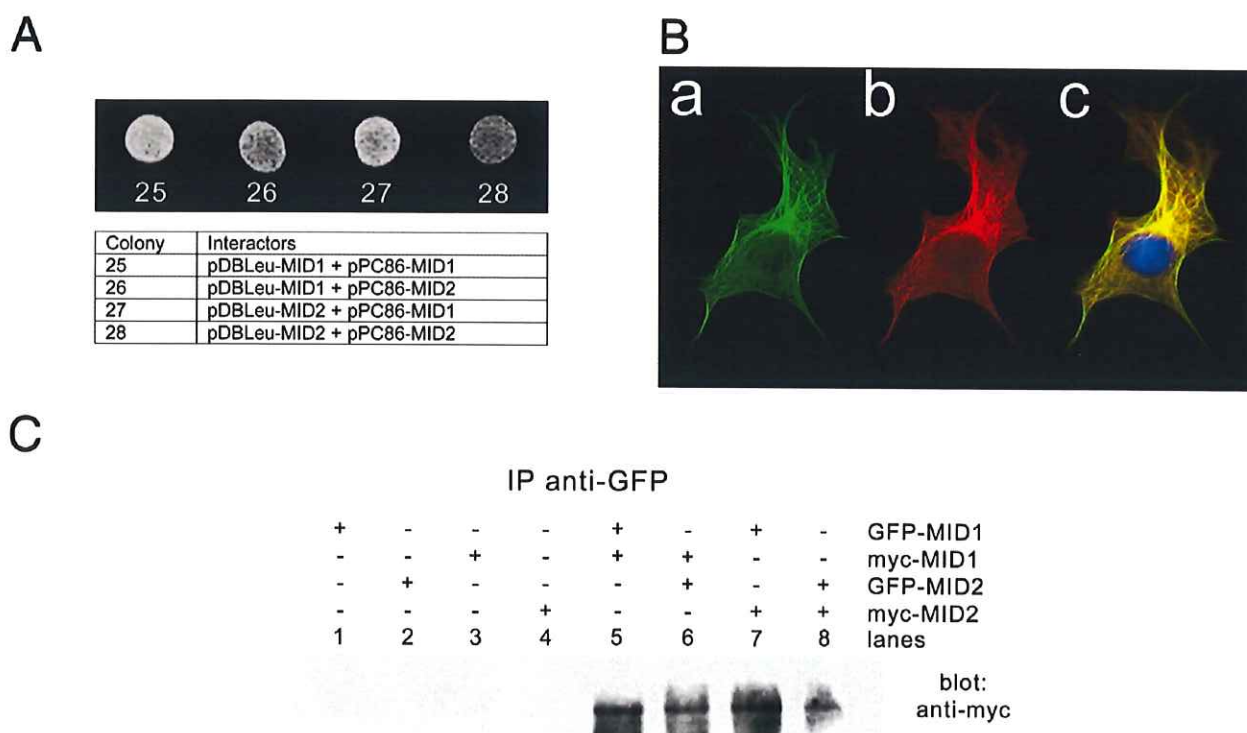
To confirm these results, we cloned the MID1 and MID2 domain-deletion constructs in-frame into pPC86 and co-transformed into MaV203 with pDBLeu-Alpha 4. As expected from the immunofluorescence data, activation of all two-hybrid reporter genes, at a level similar to that seen with the full-length MID proteins, was observed with the Δ RING, Δ FNIII and Δ CTD proteins (results not shown), confirming that neither of these domains mediated binding to Alpha 4. Notably, both Δ CC proteins also interacted with Alpha 4 (Fig 2A) suggesting that the two proteins were indeed still interacting in the cytoplasm in the immunofluorescence experiments despite not being associated with microtubules. However, the strength of the interaction between the Δ CC proteins and Alpha 4 appeared to be reduced compared to the full-length, Δ RING, Δ FNIII and Δ CTD (data not shown). In contrast, but again confirming the immunofluorescence result, the Δ BB constructs did not activate reporter gene expression indicating the interaction between Alpha 4 and either MID protein was abolished by removal of this motif (Fig 2A).

The coiled-coil domains of MID1 & MID2 are required for homodimerisation and microtubule binding but not Alpha 4 interaction

Immunoprecipitation experiments have previously shown that MID1 can form homodimers/homomultimers [17]. To investigate whether homodimerisation was a prerequisite for binding Alpha 4 and/or association with the microtubules, we first tested whether MID1 can interact with itself in the context of the yeast two-hybrid system. The results clearly indicated that the MID1-MID1 interaction is strong (Fig 3A). Domain-deletion constructs (in pDBLeu and pPC86 vectors) were then introduced into the yeast two-hybrid system. Unlike the proteins harboring a deletion of either the RING, B-boxes or FNIII motifs, the MID1 protein lacking the coiled-coil motif had significantly reduced capacity to bind the wild-type MID1 protein, inferring that the coiled-coil domain is critical for efficient homodimerisation (result not shown). Assessment of MID1 using the MultiCoil algorithm [28] supports the conclusion of dimer formation (not trimer) and that the first of the two coiled-coils in this domain is largely responsible for this property. Interestingly, the MID1 protein in which the CTD motif was removed also demonstrated a reduced ability to bind the wild-type MID1 in this system, although this effect was not as marked as that seen for the Δ CC proteins. Notably, the yeast two-hybrid and immunofluorescence experiments demonstrate that Alpha 4 can still interact with the MID1 and MID2 Δ CC proteins. This observation indicates that Alpha 4 must be able to interact with MID monomers although its tethering to the microtubule network is dependent on MID dimerisation, that is; the MID proteins only associate with microtubules as dimers.

MID1 & MID2 can form heterodimers on microtubules

Given their high level of identity, we investigated whether MID1 and MID2 can also form heterodimers using the yeast two-hybrid system, immunofluorescence of transiently transfected Cos1 cells and co-immunoprecipitation. Co-transformation of the yeast MaV203 strain with both pDBLeu-MID1 and pPC86-MID2 (or in the reverse vector combination) resulted in high level activation of all reporter genes indicating a strong interaction that was comparable to the strength of the MID1-MID1 homo-interaction (Fig 3A). These findings are contradictory to initial reports from a study by Cainarca et al [17] but have been confirmed in experiments involving transient transfection of GFP-MID1 and myc-MID2 fusion constructs (Fig 3B) as well as by co-immunoprecipitation (Fig 3C). Introduction of individual domain-deletions of MID1 together with full-length or domain-deletion MID2, and vice versa, into Cos1 cells and the yeast strain MaV203 using the relevant constructs showed, as expected, that the coiled-coil motif was largely responsible for mediating this heterodimerisation (data not shown).

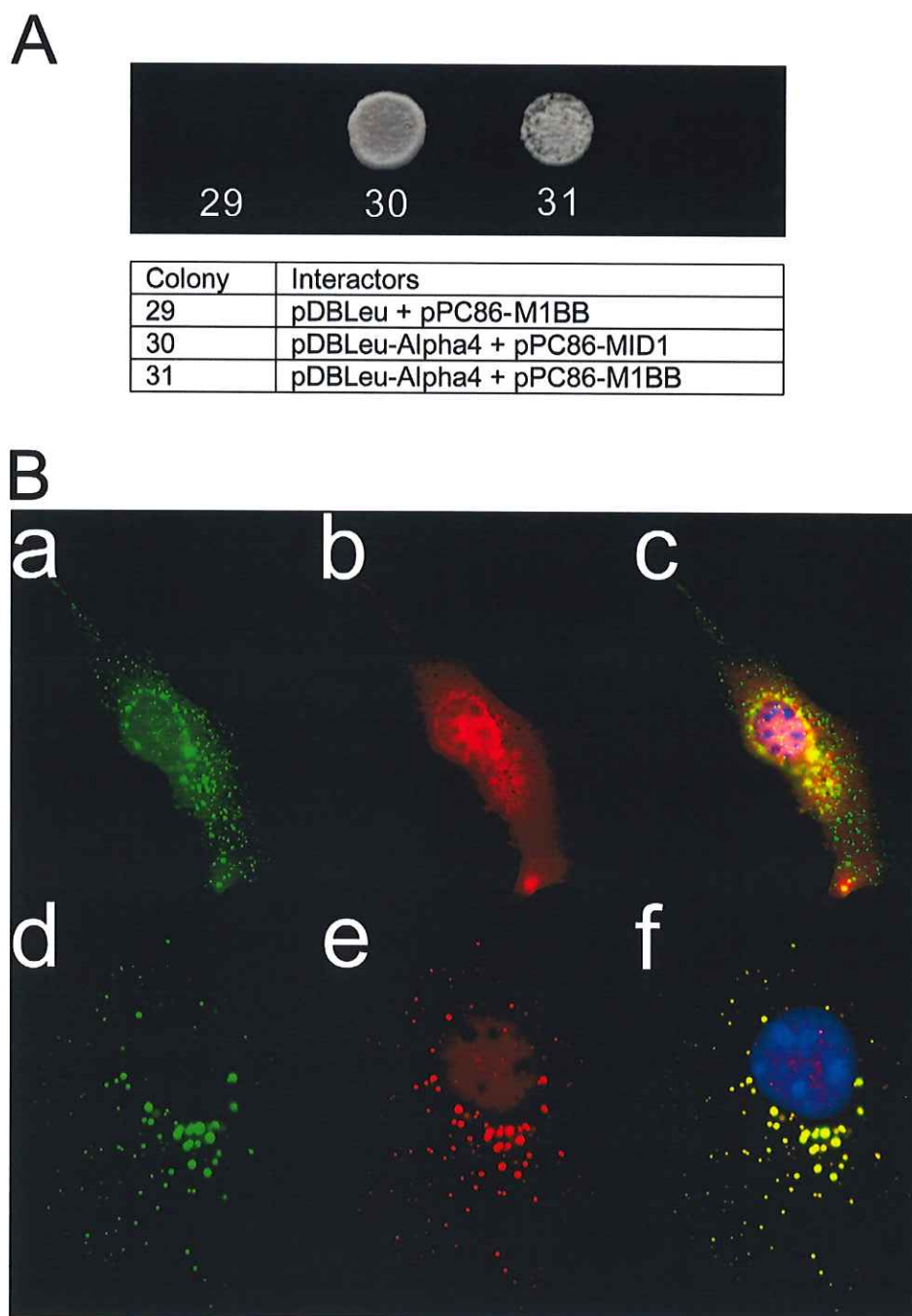
**Figure 3**

MID1 and MID2 can homo- and heterodimerise with one another. (A) Yeast two-hybrid assay for MID1 and MID2 multimerisation. Yeast agar plate (leu⁻ trp⁻ his⁻, 75 mM 3-AT) showing growth for MID1/MID1 (25), MID1/MID2 (26 and 27) and MID2/MID2 (28). (B) MID1 and MID2 co-localise to the microtubules. Co-expression of GFP-MID1 (a) and myc-MID2 (b) in transiently transfected Cos1 cells showing co-localisation to the microtubular cytoskeleton in an overlay (c) with a DAPI stained nucleus (blue). (C) Co-immunoprecipitation of MID1 and MID2 homo- and heterodimers. Shown are extracts from Cos1 cells, transfected with GFP-MID1 (lane 1), GFP-MID2 (lane 2), myc-MID1 (lane 3), myc-MID2 (lane 4), GFP-MID1 and myc-MID1 (lane 5), GFP-MID2 and myc-MID1 (lane 6), GFP-MID1 and myc-MID2 (lane 7) and GFP-MID2 and myc-MID2 (lane 8). Samples were immunoprecipitated with anti-GFP antibody/protein-A sepharose beads, separated on a 8% SDS polyacrylamide gel, transferred to a nitrocellulose membrane and blotted with anti-c-myc antibody to detect co-precipitate protein.

The B-boxes of MID1 & MID2 are sufficient to bind Alpha 4

To verify that the Alpha 4 interaction and the association of the complex with microtubules were indeed dependent on the B-boxes and coiled-coil, respectively, and not just an artefact of altering the relative spacing of remaining domains, we fused the MID1 B-boxes and coiled-coil motif, or the coiled-coil motif alone, in-frame to GFP and co-transfected with the myc-Alpha 4 construct. Of interest was the observation that the MID1 coiled-coil domain alone fused to GFP (GFP-M1CC) resulted in cytoplasmic clumping (Fig 4B; a), supporting the notion that the coiled-coil domain is required for MID1 dimerisation. Consistent with this is that the phenomenon of cytoplasmic clumping seen in most OS patients has only been observed in those cases where the expressed mutant MID1 protein harbours a mutation outside the coiled coil motif [5]. Notably, however, in cells co-transfected with the

GFP-M1CC and myc-Alpha 4 constructs, Alpha 4 did not co-localise with these M1CC clumps (Fig 4B; a-c). Like the M1CC protein, the construct expressing the fusion between GFP and the MID1 B-boxes plus coiled-coil domains (GFP-M1BBCC) resulted in clumps within the cytoplasm. Importantly, and in contrast to the co-transfection of M1CC, Alpha 4 was found to co-localise with the GFP-M1BBCC fusion protein (see Fig 4B; d-f). We did not undertake the generation of a GFP-M1BB (B-boxes alone) construct as it was predicted that the resultant GFP fusion protein would show a diffuse cytoplasmic distribution largely indistinguishable from Alpha 4. Instead, the MID1 B-boxes were cloned into pDBLeu and tested directly for their ability to interact with Alpha 4 in the two-hybrid system. The result (Fig 4A) clearly supports the conclusion that the B-boxes are sufficient to bind Alpha 4.

**Figure 4**

The B-boxes of MID1 are sufficient to bind Alpha 4. (A) Yeast two-hybrid analysis shows that the MID1 B-boxes (pPC86-M1BB) alone can interact with Alpha 4 (pDBLeu-Alpha 4) (31). The wild-type MID1-Alpha 4 interaction (30) and a control with the pPC86-M1BB and no interaction partner (29) were included for comparison. (B) Immunofluorescence assay highlights the importance of the B-boxes for Alpha 4 binding. Subcellular distribution of the GFP-fused MID1 coiled coil domain, GFP-M1CC (a) and myc-Alpha 4 (b) in the same cell shows that the two proteins do not co-localise, as seen in the merged image (c). However, the MID1 fragment, GFP-M1BBCC (d) and the myc-Alpha 4 (e) do co-localise in cytoplasmic speckles as seen in the merged image (f). Both merged images, (c) and (f), also show DAPI stain (blue), which indicates the position of the nucleus.

MID1 is phosphorylated on serine and threonine residues

As Alpha 4 is a regulator of PP2-type serine/threonine phosphatases, we investigated whether MID1 and MID2 might themselves be phosphorylated and hence possible targets of Alpha 4/phosphatase action. Due to the low levels of endogenous MID1 and MID2 in all tested cultured cell lines, we performed western analysis of immunoprecipitated MID-GFP proteins using extracts of Cos1 cells that had been transiently transfected with the various expression constructs. Immunoprecipitation with anti-GFP antibodies followed by western analysis using anti-phosphoserine and anti-phosphothreonine antibodies showed that MID1- and MID2-GFP fusion proteins were phosphorylated on both serine and threonine residues (Fig 5A). Similar analysis using anti-phosphotyrosine antibodies failed to demonstrate phosphorylation of tyrosine residues on either protein (result not shown).

In an attempt to define the location of the sites in MID1 that were phosphorylated, we performed similar immunoprecipitation and western analysis but this time using extracts of Cos1 cells that had been transfected with the individual domain-specific deletion constructs (Fig 5B). That the overall phosphorylation of the MID1 fusion protein was not significantly affected by deletion of any individual domain may suggest that MID1 is phosphorylated at multiple threonine residues along the protein. However, using the anti-phosphoserine antibody, no serine phosphorylation of the Δ BB protein was detected suggesting that most serine phosphorylation in MID1 occurs at or near the B-boxes.

Identification of potential sites of phosphorylation in MID1 and MID2

Computer prediction of potential target residues for phosphorylation by serine/threonine kinases not surprisingly identified numerous consensus sites throughout both human MID1 and MID2. Given the conservation of the rapamycin-sensitive pathway from yeast to mammals and the fact that Alpha 4 binds to both MID1 and MID2 which are only 77% identical, we reasoned that any functionally relevant phosphorylation site should be conserved across species and in both MID1 and MID2 proteins. Examination of available orthologous MID1 and MID2 sequences (from seven and three species, respectively) showed that sixteen of these sites (6 threonines and 10 serines) were fully conserved between all MID proteins (Fig 5C). Interestingly, two of these sites (Ser92 and Ser96), which represent consensus phosphorylation sites for GSK3 and MAPK/CKI/CKII respectively, are the only two conserved serine residues in the amino-terminal half of the protein and, as both are located in the region deleted in the Δ BB constructs of MID1 and MID2, are likely to be the primary sites of serine phosphorylation in these proteins.

Discussion

We have previously shown that the X-linked form of Opitz GBBB syndrome (OS) results from loss of function mutations in *MID1*, a gene that encodes a member of a new class of microtubule-associated proteins [5]. A highly related factor, termed MID2, has also been identified which shares 77% amino acid identity (92% similarity) with MID1 and is expressed in many of the same embryological tissues, albeit at a lower level in most [15,19]. Both MID1 and MID2 are members of the RBCC family of proteins, a group of proteins with diverse intracellular localisations and presumably varied functions. Despite these differences, an ability to function as part of large multi-protein complexes is common to RBCC proteins [29]. Consistent with this is the observation that the 67 kDa MID1 is mostly found in complexes of between 250 kDa and 450 kDa [17]. However, no function has been definitively ascribed to any of the motifs in the MID proteins or indeed either protein as a whole. In this study, we have undertaken yeast two-hybrid screens to identify interacting partners that were likely to constitute part of these MID-complexes as well as performed precise deletion analysis of the MID proteins to begin to elucidate the role of individual motifs.

The only apparent similarity between all the *MID1* mutations identified in patients with OS is that the resultant mutant protein is no longer able to completely decorate the microtubules. This observation implies that the function of one or more proteins within the MID1 complex is absolutely required along the length of the microtubules. Identification of such *bona fide* interactors might therefore be expected to shed light on the possible role of the OS protein and thus provide a better understanding of the molecular pathogenesis of the disorder. To this end, we performed yeast two-hybrid screens using MID1 and MID2 as bait and a 10.5 dpc mouse embryo cDNA library as prey. Significantly, we identified the phosphoprotein, Alpha 4 (also known as IGBP1) as an interactor of both MID proteins. This interaction between the MID proteins and Alpha 4 was confirmed by both vector swapping in the two-hybrid system as well as by co-localisation of differentially tagged proteins to the microtubule network. The specificity of this interaction is also supported by the fact that mouse Mid1 (but not mouse Mid2 presumably because of its lower abundance) was recently identified in the reverse two-hybrid screen of a 9 dpc murine embryonic cDNA library where the mouse Alpha 4 was used as the bait [30]. To characterise this interaction further, we chose to exploit the observation that C-terminal mutations of MID1 form cytoplasmic clumps both *in vivo* [18] and in transfected cell lines over-expressing GFP-mutant MID1 fusion proteins [5,15,17,18] by co-transfecting constructs expressing the Δ CTD proteins (and the other domain-specific deletions of MID1 and MID2) as GFP fusions with

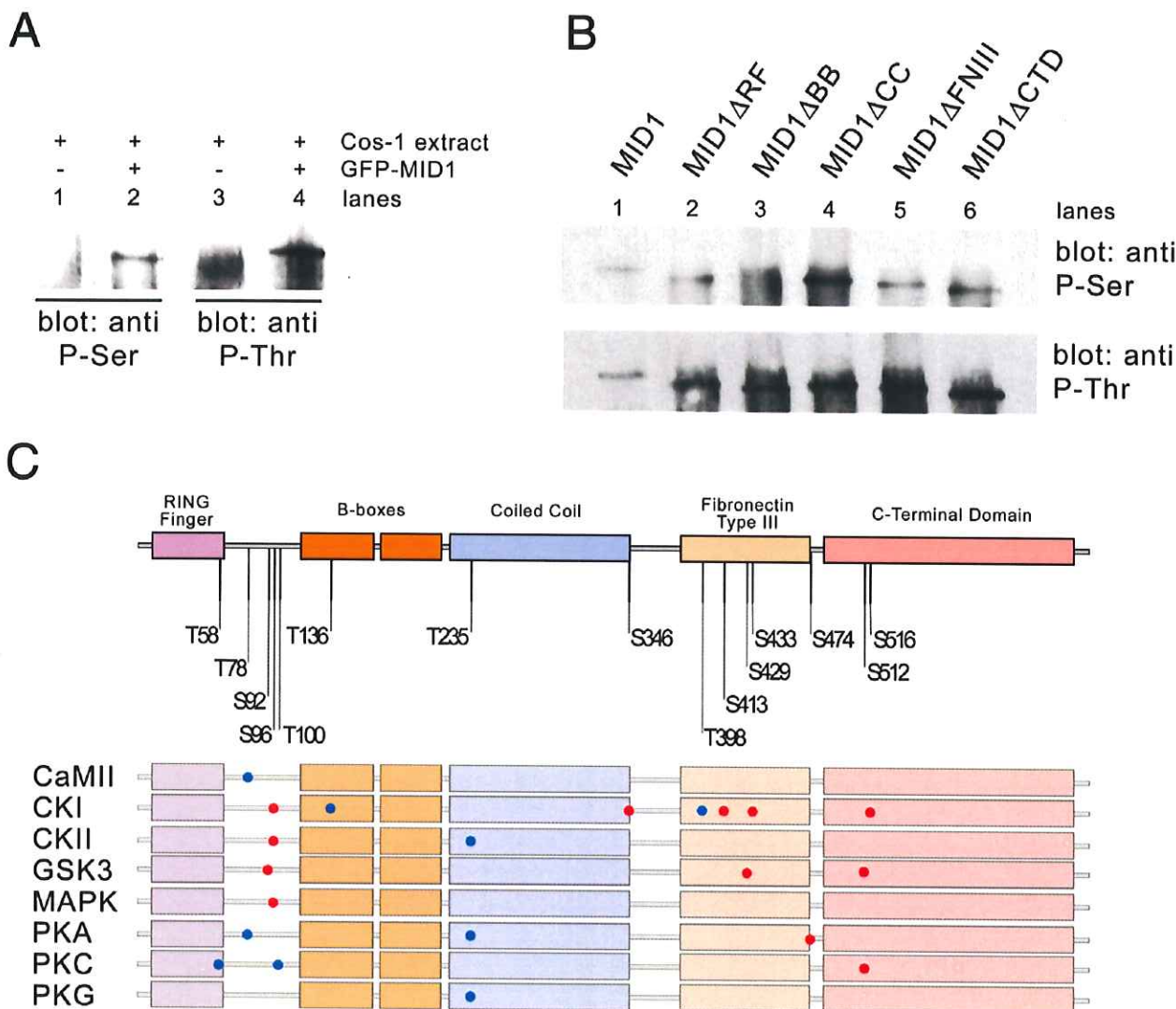


Figure 5
MID1 contains phosphorylated serine and threonine residues. (A) Extracts from untransfected Cos1 cells (lanes 1 & 3) or Cos1 cells transfected with GFP-MID1 (lanes 2 & 4) were immunoprecipitated with anti-GFP antibody/protein-A sepharose beads and analysed by western blot analysis using either an anti-phosphoserine antibody (lanes 1 & 2) or anti-phosphothreonine antibody (lanes 3 & 4). Protein bands in lanes 2 and 4 indicated that GFP-MID1 contains both phosphoserine and phosphothreonine residues. (B) Domain-specific deletions of MID1 were used in an attempt to crudely map locations of the phosphorylated serine and threonine residues. Shown are extracts from Cos1 cells transfected with full length GFP-MID1 (lane 1), GFP-MID1ΔRF (lane 2), GFP-MID1ΔBB (lane 3), GFP-MID1ΔCC (lane 4), GFP-MID1ΔFNIII (lane 5), and GFP-MID1ΔCTD (lane 6). The samples were immunoprecipitated with anti-GFP antibody/protein-A sepharose beads, separated on 8% SDS polyacrylamide gels, transferred to nitrocellulose membranes and blotted with anti-phosphoserine antibody (top panel) or anti-phosphothreonine antibody (bottom panel). (C) Computer assisted detection of potential serine/threonine phosphorylation sites in MID1 were identified using NetPhos 2.0 software [38]. Examination of a multiple alignment of available MID1 and MID2 sequences was carried out to identify fully conserved putative phosphorylation sites. A diagrammatic representation of this analysis shows 16 conserved sites depicted as dots (red for serine and blue for threonine) along the length of a representative MID1 protein (numbered residue positions are also depicted at the top of diagram). The actual kinases that recognise the residues are listed on the left of the figure; CaMII (R-X-X-S/T-X), CKI (Sp/Tp-X₂₋₃-S/T-X), CKII (X-S/T-X-X-D/E), GSK3 (X-S/T-X-X-X-Sp), MAPK (P-X-S/T-P), PKA (R-X₁₋₂-S/T-X), PKC (X-S/T-X-R/K), PKG ((R/K)₂₋₃-X-S/T-X).

the myc-tagged Alpha 4 construct. These experiments were complemented by yeast two-hybrid analysis of the same combinations of constructs.

Through a series of in-frame deletions of both MID1 and MID2 employed in both immunofluorescence and yeast two-hybrid assays, we have demonstrated functions for three domains of these RBCC proteins. We have shown a direct role for an RBCC B-box region in protein-protein interactions, that being the binding of Alpha 4, a regulatory subunit of the PP2-type phosphatases including the principal cellular phosphatase, protein phosphatase 2A (PP2A; [21–24]). In addition to the identified role for the B-boxes, we have shown that the coiled-coil domain not only mediates the homodimerisation of these proteins but also their ability to heterodimerise. Finally, our data, together with previous observations by others [18], support a role for the C-terminal domain in microtubule binding and to a lesser extent in dimerisation. Collectively, the immunofluorescence and yeast two-hybrid data indicate that MID dimerisation is a prerequisite for association of each MID-Alpha 4 complex with microtubules.

The interaction of MID1 and MID2 with Alpha 4 raised the possibility that these RBCC proteins, like Xnf7, are phosphoproteins. Indeed, western blot analysis of the transiently expressed MID-GFP fusion proteins using anti-phosphoserine and anti-phosphothreonine antibodies has confirmed these suspicions. Analysis of the domain-specific deletion proteins using these same antibodies showed that most serine phosphorylation of MID1 seems to occur at, or immediately adjacent to, the B-boxes, whereas threonine phosphorylation was likely to occur at residues in more than one domain. As a preliminary step towards identifying those residues that are phosphorylated, computer-based searches identified numerous potential sites for phosphorylation by serine/threonine kinases. However, only sixteen of these (see Fig 5C) are completely conserved across all MID species isolated to date. Of note are the potential phosphorylation sites at serine 96 (S96; [30]) and serine 92 (S92) that are located immediately amino-terminal to the B-boxes of MID1 and MID2 and are deleted in the Δ BB constructs. S92 is a consensus site for GSK3 that would be dependent on prior phosphorylation of S96. S96 falls within a consensus phosphorylation sequence for the MAP-kinase (MAPK), ERK2 (P-N-S/T-P; [31]), as well as casein kinases I and II. Therefore, either of these kinases could conceivably play a priming role for GSK3 phosphorylation of S92. Similar phosphorylation mechanisms have been observed for other proteins including some microtubule-associated proteins, eg. tau and MAP2, where it has been implicated in the regulation of specific functions of those proteins [32,33]. Notably, in the case of both eIF2Bepsilon (at serine residues 535 &

539) and tau (at serine residues 208 & 212), DYRK, a MAPK-related kinase, plays such a priming role for GSK3 [34]. However, DYRK is unlikely to be involved in phosphorylation of MID1 S96 because of a single amino acid difference in its surrounding sequence required for recognition of this site for phosphorylation. In fact, Liu et al [30] have provided evidence to support the involvement of a MAPK, although the target of this activity was not shown. Irrespective of the identity of the kinase(s) responsible, it is conceivable that phosphorylation at these or other sites in MID1 and MID2 could play an important role in the overall function of the MID proteins, for example: regulation of the Alpha 4-MID interaction and hence regulation of PP2-type phosphatase activity. However, our preliminary analysis of a MID1 Ser92/Ser96 double mutant has suggested that MID dimerisation and the MID-Alpha 4 interaction are not dependent on phosphorylation at these residues (unpublished observations).

Regardless of the role of MID phosphorylation, the implication of both MID1 and MID2 in the Alpha 4-mediated regulation of phosphatase activity may provide valuable clues as to the pathophysiological consequences of MID1 mutations that underlie Opitz syndrome. It can be envisaged that, in tethering Alpha 4 to the microtubules, MID1 (and MID2) could be affecting the activity of the PP2-type phosphatases and thereby, in-turn, modulating the rapamycin-sensitive signaling pathway. This could conceivably occur by one of a number of mechanisms. Firstly, the MID proteins may control the availability of Alpha 4 to the phosphatases either by its 1) sequestration or 2) turnover, facilitated by the possible role of the MID RING finger motif in ubiquitination. Secondly, the Alpha 4-MID interaction may direct PP2A (and PP2-related) phosphatase activity to specific targets on the microtubules, which may include MID1 and MID2 themselves. The fact that both endogenous MID1 and transiently expressed GFP-MID1 and GFP-MID2 decorate microtubules throughout the cell cycle ([17]; unpublished observations) suggests that binding of MID1 and MID2 to the microtubule network may not itself be regulated by targeted Alpha 4-dependent PP2A activity. However, it remains possible that dynamic regulation of MID1 microtubule binding may escape detection by immunofluorescence as endogenous levels of MID1 and Alpha 4 in examined cell lines are both low (unpublished observations). The fact that we have not been able to demonstrate PP2A(C) co-localisation with the MID-Alpha 4 complexes on the microtubules, despite a known microtubule-associated pool of PP2A [35], would perhaps support the former of these mechanisms. However, we cannot exclude the possibility that recognition of PP2A(C) by this antibody is blocked by the interaction of Alpha 4 and MID1/2. If this is indeed the case, then it could be envisaged that PP2A(C) is also a target of MID RING-mediated degradation through its in-

dependent interaction with Alpha 4. An alternate hypothesis is that MID function is indeed controlled by Alpha 4-PP2-type phosphatases but through the regulated binding of additional factors that might be components of the MID1 macromolecular complexes. The characterisation of additional interacting partners will therefore be important to assess this possibility.

Numerous studies in mice have demonstrated that Alpha 4, like its yeast homologue Tap42, plays an essential regulatory role within the cell through its regulated binding to the catalytic (C) subunit of PP2-type serine/threonine protein phosphatases [21–24]. The interaction of Alpha 4 with PP2A(C) has been most extensively studied and shown to be dependent on phosphorylation of Alpha 4 by the mTOR (target of rapamycin) kinase. Although PP2A has a wide range of biological functions, Alpha 4 regulates a distinct subset of events in a rapamycin-sensitive manner, including progression through the cell cycle and the regulation of protein biosynthesis (for a review see [36]). The implication that disruption to some aspects of this rapamycin-sensitive pathway might be associated with the pathogenesis of the developmental disorder, Opitz syndrome, raises two intriguing possibilities. Firstly, it can be envisaged that mutations in other components of the pathway may give rise to similar clinical phenotypes. It is therefore perhaps worthy to note that the Alpha 4 gene maps to Xq13 [26] in the vicinity of the critical interval for FG syndrome, a malformation disorder with some clinical overlap with that of Opitz GBBB syndrome. We are currently investigating whether molecular defects in Alpha 4 indeed underlie FG syndrome or other developmental disorders mapping to the proximal long arm of the X chromosome. Secondly, it is also feasible that genetic polymorphisms that reflect a variation in the level of expression or activity of one or more components of the rapamycin-sensitive pathway might also contribute to the clinical variability of OS. In this regard, the indication from our immunofluorescence studies that the microtubular localisation of Alpha 4 is likely to be limited by the level of the MID1 and MID2 proteins within a particular cell type provides indirect support for our earlier hypothesis that MID2 may be able to compensate, at least partially, for the loss of MID1 in Opitz syndrome [5].

Conclusions

The finding that Alpha 4, a rapamycin-sensitive regulatory subunit of PP2-type phosphatases, interacts strongly with the RBCC proteins, MID1 and MID2, implicates this signalling pathway in the pathogenesis of the X-linked form of Opitz GBBB syndrome and provides a possible explanation for the intrafamilial variability in clinical presentation of the disorder. Other components of the rapamycin-sensitive pathway should be considered as candidates for similar malformation disorders.

Materials & Methods

Miscellaneous enzymes and chemicals

All restriction endonucleases were purchased from New England Biolabs (Genesearch Pty Ltd, Arundel, Queensland), Klenow fragment from GeneWorks Pty Ltd (Thebarton, South Australia), and both T4 ligase and T4 DNA polymerase from Roche Diagnostics Australia (Castle Hill, New South Wales). 4',6-diamidino-2-phenylindole dihydrochloride (DAPI) and 3-Amino Triazole (3AT) were obtained from Sigma-Aldrich (Castle Hill, New South Wales).

The yeast two-hybrid screen

The ProQuest™ yeast two-hybrid system (Invitrogen, Mulgrave, NSW) was employed to screen for potential interacting partners of MID1. In order to generate the "bait" construct, the full-length human MID1 cDNA was subcloned from pBSMID1 [5] using Sall and NcoI into the similarly restricted pDBLeu vector. Subsequent digestion with Sall, end-filling with Klenow fragment and religation generated the full-length MID1 cDNA in-frame with the GAL4 DNA binding domain (Gal4DBD). The construct, pDBLeu-MID1, was verified by sequencing.

The selection of an appropriate cDNA library was deemed critical to maximise the chance of detecting *bona fide* interacting factors. The expression of the murine *Mid1* gene during embryological development is consistent with the clinical presentation of OS [27]. Given the very high level of primary sequence identity between all vertebrate MID1 proteins (unpublished data), a murine 10.5 dpc whole embryo cDNA library (Invitrogen) directionally cloned in the GAL4 activation domain (GAL4-AD) plasmid, pPC86, was selected for use as the "prey" in the two-hybrid screen.

The pDBLeu-MID1 construct was transformed into MaV203 strain (Genotype: *MAT α* , *leu2-3, 112*, *trp1-901*, *his3 Δ 200*, *ade2-101*, *gal4 Δ* , *gal80 Δ* , *SPAL10::URA3*, *GAL1::lacZ*, *HIS3_{UAS} GAL1::HIS3@LYS2*, *can1^R*, *cyh2^R*) along with the parental pPC86 plasmid as per the manufacturers instructions (Invitrogen). The HIS3 reporter was used to determine the level of self-activation of the MID1-GAL4DBD fusion based on the level of 3-amino triazole resistance (3AT^R) of the fusion protein. A concentration of 50 mM 3AT was found to be an adequate level for the assay, although subsequent analyses were performed on 75 mM 3AT plates to further reduce background transactivation of the reporter genes. To screen for potential interacting proteins, the pDBLeu-MID1 fusion construct was co-transformed with the 10.5 dpc mouse embryo cDNA-pPC86 library according to standard protocols (ProQuest™ Yeast Two-Hybrid Manual, Invitrogen) and plated on a synthetic complete medium (SC-Leu-Trp-His) containing either 50 mM or 75 mM 3AT.

The full-length MID2 (FX2) cDNA was cloned from pEGFP-FXY2. ORF [15] into pDBLeu using the same strategy as used for MID1. As *Mid2* is generally expressed in many of the same tissues during embryological development as *Mid1* albeit at considerably lower levels [19], the resultant clone, pDBLeu-MID2, was consequently used as "bait" in a similar screen of the 10.5 dpc embryo cDNA library as well as directly against the identified MID1 interacting clones.

For confirmation of putative interacting clones, constructs and library clone isolates (in both parental vectors and swapped vectors) were re-transformed into the MaV203 yeast strain using the LiAc method [37]. Transformed yeast cultures were incubated for 24 hours at 30°C in selective media. Cultures (10 µl - 0.1 OD₆₀₀) were then spotted onto selective plates and incubated at 30°C for 48 hours. Replica cleaning of plates was performed as required.

Generation of GFP- and myc-full-length cDNA fusion constructs for immunofluorescence

The generation of full-length MID1-GFP and MID2-GFP fusions in pEGFP have previously been reported [15,5]. A vector for the production of myc-tagged fusion proteins was generated by modification of the pEGFP-N2 vector (Clontech, Palo Alto, CA). Briefly, the GFP coding region was excised from pEGFP-N2 with NotI and BamHI, the 5' overhangs filled using T4 DNA polymerase and the vector religated to give pCMV-N2. Six copies of the myc epitope containing a start codon was amplified from pGEM-6mycT (gift from M. Whitelaw, University of Adelaide), and cloned into the HindIII site of pCMV-N2. To facilitate in-frame insertion of cDNAs directly from the pPC86 library vector, the plasmid was linearised with EcoRI, end-filled and re-ligated to create pCMV-6myc-ΔE. In order to clone Alpha 4 into pCMV-6myc-ΔE, pPC86-Alpha 4 was digested with Sall/AatII and the full-length cDNA insert ligated into Sall/SmaI restricted pCMV-6Myc-ΔE vector. The reading frame of the construct was confirmed by automated sequencing.

Generation of MID1 and MID2 domain-specific deletion constructs for immunofluorescence and yeast two-hybrid analyses

The FNIII domain-specific deletions in both MID1 and MID2 were generated by precise deletion of the domains by a two-step PCR strategy. The other domain-specific deletion constructs were generated using QuickChange™ site-directed mutagenesis (Stratagene, La Jolla, CA) to introduce unique restriction sites, as required, at the start and end of each domain within MID1 and MID2. Exceptions to this were the RING and B-box deletions of MID1 where a native XbaI site located at nucleotides +207-212 was used as the 3' and 5' excision point in the respective

constructs, and the CTD deletions in both MID1 and MID2 where a native BamHI site located at nucleotides +1464-1469 was used as the 5' excision point for these constructs. Restriction sites were chosen so that, where possible, the encoded amino acid sequence remained unaltered or only resulted in conservative substitutions. Furthermore, the sites were positioned such that excision of individual domains did not alter the reading frame of the encoded protein (Table 1). In any one construct, a maximum of two restriction sites were introduced. Details of the strategies and primers (for both PCR and site-directed mutagenesis) used in the generation of these deletion constructs will be forwarded upon request to the corresponding author. Each introduced restriction site was confirmed by digestion and sequencing and then independently tested for its effect on the microtubule binding capacity of the respective proteins. In each case, this was determined by direct visualisation of fluorescence of the created GFP fusion protein. No introduced change had any appreciable effect on the microtubular distribution of the proteins.

To further test the function of some of the regions of the MID1 protein, selected separate motifs were generated using the appropriate existing, and/or inserted, restriction sites and ligated in-frame and C-terminal to either EGFP (in pEGFP) or Gal4DBD (in pDBLeu). The following constructs were generated: the MID1 B-box fusion (pDBLeu-M1BB) containing residues 71-213, the MID1 coiled-coil fusion (pEGFP-M1CC) containing residues 214-349, and the MID1 B-boxes plus coiled-coil fusion (pEGFP-M1BBCC) containing residues 71-349.

Transfection and immunofluorescence analysis of GFP-MID1 constructs

Preparations of the various GFP- and myc-tagged expression constructs were made using the Qiagen Midi kit (Qiagen, Clifton Hill, Victoria). Two picomoles (approximately 1 microgram) of each construct were transfected into cultured cell lines (Cos1, HeLa, NIH3T3) using FuGene transfection reagent (Roche Diagnostics Australia). Transfected cells were grown on coverslips in DMEM plus 10% FBS and fixed 24 hours post-transfection as previously described [5].

In test transfections, where only a single GFP expression construct was introduced into cells, control microtubule staining was performed post-fixation using an anti-α tubulin antibody plus an anti-mouse Texas Red-conjugated secondary antibody (Jackson Laboratories, Bar Harbor, Maine). In cells transfected with myc-tagged expression constructs (either alone or in combination with a GFP-tagged expression construct), anti-α tubulin staining was not performed. Instead the Texas Red-conjugated secondary antibody was used in combination with an anti-myc

monoclonal antibody (9E10) to detect the expression of the myc-tagged protein. In all cases, nuclei were stained using the DNA-specific stain, DAPI. GFP and Texas Red fluorescence were visualised under appropriate wavelength light on an Olympus AX70 microscope. Images were captured using a Photometrics CE200A Camera Electronics Unit and processed using Photoshop 6.01 software (Adobe Systems Incorporated, San Jose, California).

Immunoprecipitation and western analysis

Preparations of the various GFP- and myc-tagged expression constructs were made using a DNA plasmid Midi kit (Qiagen). Six picomoles (approximately 3 micrograms) of each construct were transfected into approximately 10^7 Cos1 cells using FuGene transfection reagent (Roche Diagnostics Australia). After 24 hours incubation, cells were scraped from the culture dish and lysed on ice for 30 minutes in 1 ml lysis buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 5 mM EDTA, 1.0 % Triton X-100). Cell lysates were cleared by centrifugation at 4°C (15 minutes, $16 \times g$), and protein extract recovered as supernatant. After pre-clearing 200 μ l of protein extract with 10 μ l of 50% protein-A sepharose bead slurry, extracts were incubated with 1 μ g of antibody for 2 hours at 4°C and then for another 2 hours with 20 μ l of fresh 50% protein-A sepharose bead slurry. The beads were washed four times with wash buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 5 mM EDTA, 0.1% Triton X-100) and protein eluted from the beads by boiling in 2 \times SDS load buffer. Proteins were separated by 8% SDS PAGE and blotted onto Hybond-C membranes (Amersham Pharmacia) using a semi-dry transfer apparatus (BioRad). Membranes were blocked, incubated with the appropriate primary antibody, washed, incubated with the appropriate HRP-conjugated secondary antibody and washed again according to established method described in Current Protocols in Cell Biology. Detection was carried out using an enhanced chemiluminescence reagents (ECL) kit (Amersham Pharmacia) as per the manufacturer's instructions. Antibodies used in immunoprecipitation and western blot analysis included; rabbit polyclonal anti-GFP antibody (gift from Pam Silver, Dana-Farber Cancer Institute, Boston), mouse anti-myc monoclonal antibody (gift from Stephen Dalton, University of Adelaide), rabbit polyclonal anti-phosphoserine and anti-phosphothreonine antibodies (Zymed) and HRP conjugated anti-rabbit and anti-mouse secondary antibodies (Amersham Pharmacia).

Computer-assisted detection of serine/threonine phosphorylation sites

Analysis of putative consensus serine/threonine phosphorylation sites in MID1 was performed using NetPhos version 2.0 [<http://www.cbs.dtu.dk/services/NetPhos/>] software [38]. Examination of a multiple sequence alignment of all available MID1 and MID2 sequences was per-

formed in order to determine if any putative phosphorylation site was fully conserved. For this analysis, orthologous MID1 sequences were assessed from a variety of species, including human, mouse, rat, chick, tammar wallaby, zebrafish (partial sequence only) and fugu, and MID2 orthologous sequences from human, mouse and rat (partial sequence only).

Note Added In Proof

During the review of this manuscript, a paper by Trockenbacher *et al* [Trockenbacher A, Suckow V, Foerster J, Winter J, Krauß S, Roper H-H, Schneider R. and Schweiger S: **MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation.** *Nature Genetics* 2001, 29:287–294] independently reported the interaction of Alpha 4 and MID1. These investigators also showed that MID1, possibly through Alpha 4, regulates the turnover of the microtubule-associated fraction of PP2AC and hence may represent a possible pathological mechanism for the Opitz syndrome phenotype.

Acknowledgements

We would like to thank Professor David Brautigan for kindly sharing data prior to publication. This work was supported by project grant #157958 and in part by an R. Douglas Wright Award (#997706) (to T.C.C.) from the National Health and Medical Research Council of Australia.

References

1. Opitz JM: **G syndrome (hypertelorism with esophageal abnormality and hypospadias, or hypospadias-dysphagia, or "Opitz-Frias" or "Opitz-G" syndrome) – perspective in 1987 and bibliography.** *American Journal of Medical Genetics* 1987, 28:275-285
2. Robin NH, Opitz JM, Muenke M: **Opitz G/BBB syndrome: clinical comparisons of families linked to Xp22 and 22q, and a review of the literature.** *American Journal of Medical Genetics* 1996, 62:305-317
3. Quaderi NA, Schweiger S, Gaudenz K, Franco B, Rugarli EI, Berger W, Feldman GJ, Volta M, Andolfi G, Gilgenkrantz S, Marion RW, Hennekam RC, Opitz JM, Muenke M, Ropers HH, Ballabio A: **Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22.** *Nature Genetics* 1997, 17:285-291
4. Gaudenz K, Roessler E, Quaderi N, Franco B, Feldman G, Gasser DL, Wittwer B, Horst J, Montini E, Opitz JM, Ballabio A, Muenke M: **Opitz G/BBB syndrome in Xp22: mutations in the MID1 gene cluster in the carboxy-terminal domain.** *American Journal of Human Genetics* 1998, 63:703-710
5. Cox TC, Allen LR, Cox LL, Hopwood B, Goodwin B, Haan E, Suthers GK: **New mutations in MID1 provide support for loss of function as the cause of X-linked Opitz syndrome.** *Human Molecular Genetics* 2000, 9:2553-2562
6. Robin NH, Feldman GJ, Aronson AL, Mitchell HF, Weksberg R, Leonard CO, Burton BK, Josephson KD, Laxova R, Aleck KA: **Opitz syndrome is genetically heterogeneous, with one locus on Xp22, and a second locus on 22q11.2.** *Nature Genetics* 1995, 11:459-461
7. McDonald-McGinn DM, Driscoll DA, Bason L, Christensen K, Lynch D, Sullivan K, Canning D, Zavod W, Quinn N, Rome J: **Autosomal dominant "Opitz" GBBB syndrome due to a 22q11.2 deletion.** *American Journal of Medical Genetics* 1995, 59:103-113
8. Fryburg JS, Lin KY, Golden WL: **Chromosome 22q11.2 deletion in a boy with Opitz (G/BBB) syndrome.** *American Journal of Medical Genetics* 1996, 62:274-275
9. McDonald-McGinn DM, Tonnesen MK, Laufer Cahana A, Finucane B, Driscoll DA, Emanuel BS, Zackai EH: **Phenotype of the 22q11.2 deletion in individuals identified through an affected relative: cast a wide FISHing net!** *Genet Med* 2001, 3:23-29

10. Glover TW: **CATCHing a break on 22.** *Nature Genetics* 1995, **10**:257-258
11. Guris DL, Fantes J, Tara D, Druker BJ, Imamoto A: **Mice lacking the homologue of the human 22q11.2 gene CRKL phenocopy neurocristopathies of DiGeorge syndrome.** *Nature Genetics* 2001, **27**:293-298
12. Jerome LA, Papaioannou VE: **DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1.** *Nature Genetics* 2001, **27**:286-291
13. Lindsay EA, Vitelli F, Su H, Morishima M, Huynh T, Pramparo T, Jurecic V, Ogunrinu G, Sutherland HF, Scambler PJ, Bradley A, Baldini A: **Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice.** *Nature* 2001, **410**:97-101
14. Merscher S, Funke B, Epstein JA, Heyer J, Puech A, Lu MM, Xavier RJ, Demay MB, Russell RG, Factor S, Tokooya K, Jore BS, Lopez M, Pandita RK, Lia M, Carrion D, Xu H, Schorle H, Kobler JB, Scambler P, Wynshaw-Boris A, Skoutchki AI, Morrow BE, Kucherlapati R: **TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome.** *Cell* 2001, **104**:619-629
15. Perry J, Short KM, Romer JT, Swift S, Cox TC, Ashworth A: **FXY2/MID2, a gene related to the X-linked Opitz syndrome gene FXY/MID1, maps to Xq22 and encodes a FNIII domain-containing protein that associates with microtubules.** *Genomics* 1999, **62**:385-394
16. Short KM, Hopwood B, Cox TC: **Further subdivision of the RBCC family of proteins defined by the presence of a Fibronectin type III motif.** *in preparation*
17. Cainarca S, Messali S, Ballabio A, Meroni G: **Functional characterization of the Opitz syndrome gene product (midin): evidence for homodimerization and association with microtubules throughout the cell cycle.** *Human Molecular Genetics* 1999, **8**:1387-1396
18. Schweiger S, Foerster J, Lehmann T, Suckow V, Muller YA, Walter G, Davies T, Porter H, van Bokhoven H, Lunt PV, Traub P, Ropers HH: **The Opitz syndrome gene product, MID1, associates with microtubules.** *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**:2794-2799
19. Buchner G, Montini E, Andolfi G, Quaderi N, Cainarca S, Messali S, Bassi MT, Ballabio A, Meroni G, Franco B: **MID2, a homologue of the Opitz syndrome gene MID1: similarities in subcellular localization and differences in expression during development.** *Human Molecular Genetics* 1999, **8**:1397-1407
20. Takei Y, Teng J, Harada A, Hirokawa N: **Defects in axonal elongation and neuronal migration in mice with disrupted tau and map1b genes.** *Journal of Cell Biology* 2000, **150**:989-1000
21. Chen J, Peterson RT, Schreiber SL: **Alpha 4 associates with protein phosphatases 2A, 4, and 6.** *Biochemical and Biophysical Research Communications* 1998, **247**:827-832
22. Inui S, Sanjo H, Maeda K, Yamamoto H, Miyamoto E, Sakaguchi N: **Ig receptor binding protein 1 (alpha4) is associated with a rapamycin-sensitive signal transduction in lymphocytes through direct binding to the catalytic subunit of protein phosphatase 2A.** *Blood* 1998, **92**:539-546
23. Murata K, Wu J, Brautigan DL: **B cell receptor-associated protein alpha4 displays rapamycin-sensitive binding directly to the catalytic subunit of protein phosphatase 2A.** *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**:10624-10629
24. Nanahoshi M, Tsujishita Y, Tokunaga C, Inui S, Sakaguchi N, Hara K, Yonezawa K: **Alpha4 protein as a common regulator of type 2A-related serine/threonine protein phosphatases.** *FEBS Letters* 1999, **446**:108-112
25. Schmelzle T, Hall MN: **TOR, a central controller of cell growth.** *Cell* 2000, **103**:253-262
26. Onda M, Inui S, Maeda K, Suzuki M, Takahashi E, Sakaguchi N: **Expression and chromosomal localization of the human alpha 4/IGBP1 gene, the structure of which is closely related to the yeast TAP42 protein of the rapamycin-sensitive signal transduction pathway.** *Genomics* 1997, **46**:373-378
27. Dal Zotto L, Quaderi NA, Elliott R, Lingerfelter PA, Carrel L, Valsecchi V, Montini E, Yen CH, Chapman V, Kalcheva I, Arrigo G, Zuffardi O, Thomas S, Willard HF, Ballabio A, Disteche CM, Rugarli EI: **The mouse Mid1 gene: implications for the pathogenesis of Opitz syndrome and the evolution of the mammalian pseudoautosomal region.** *Human Molecular Genetics* 1998, **7**:489-499
28. Wolf E, Kim PS, Berger B: **a program for predicting two- and three-stranded coiled coils.** *Protein Science* 1997, **6**:1179-1189
29. Borden KL: **RING fingers and B-boxes: zinc-binding protein-protein interaction domains.** *Biochemistry and Cell Biology* 1998, **76**:351-358
30. Liu J, Prickett TD, Elliott E, Meroni G, Brautigan DL: **Phosphorylation and microtubule association of the Opitz syndrome protein mid-1 is regulated by protein phosphatase 2A via binding to the regulatory subunit alpha 4.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**:6650-6655
31. Himpel S, Tegge W, Frank R, Leder S, Joost HG, Becker W: **Specificity determinants of substrate recognition by the protein kinase DYRK1A.** *Journal of Biological Chemistry* 2000, **275**:2431-2438
32. Chu B, Soncin F, Price BD, Stevenson MA, Calderwood SK: **Sequential phosphorylation by mitogen-activated protein kinase and glycogen synthase kinase 3 represses transcriptional activation by heat shock factor-1.** *Journal of Biological Chemistry* 1996, **271**:30847-30857
33. Sanchez C, Perez M, Avila J: **GSK3beta-mediated phosphorylation of the microtubule-associated protein 2C (MAP2C) prevents microtubule bundling.** *European Journal of Cell Biology* 2000, **79**:252-260
34. Woods YL, Cohen P, Becker W, Jakes R, Goedert M, Wang X, Proud CG: **The kinase DYRK phosphorylates protein-synthesis initiation factor eIF2Bepsilon on Ser539 and the microtubule-associated protein tau at Thr212: potential role for DYRK as a glycogen synthase kinase 3-priming kinase.** *Biochemical Journal* 2001, **355**:609-615
35. Sontag E, Numbhakdi Craig V, Bloom GS, Mummy MC: **A novel pool of protein phosphatase 2A is associated with microtubules and is regulated during the cell cycle.** *Journal of Cell Biology* 1995, **128**:1131-1144
36. Goldberg Y: **Protein phosphatase 2A: who shall regulate the regulator?** *Biochemical Pharmacology* 1999, **57**:321-328
37. Dunn B, Wobbe RC: **Introduction of DNA into yeast cells.** *In: Current Protocols in Molecular Biology* Edited by Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, vol. 13.7.1. New York: John Wiley & Sons, Inc.; 1993
38. Blom N, Gammeltoft S, Brunak S: **Sequence and structure-based prediction of eukaryotic protein phosphorylation sites.** *Journal of Molecular Biology* 1999, **294**:1351-1362

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright



Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>

editorial@biomedcentral.com

Chapter Three
Short *et al.*, 2002 to Short and Cox, 2006

Chapter Three: Short *et al.*, 2002 to Short and Cox, 2006

The purpose of this chapter is to provide information regarding the transition of research findings since those reported in Short *et al.*, 2002 that led to the research findings of Short and Cox, 2006.

3.1 Significant findings relating to MID1 and Alpha 4

Around the time of Short *et al.*, 2002, two other groups (Liu *et al.*, 2001 and Trockenbacher *et al.*, 2001) also reported the interaction between Alpha 4 and MID1. Although the primary findings in each paper were similar, all three publications provided additional unique information which together further improved the understanding of the nature of the interaction at the cellular level.

3.1.1 Liu *et al.*, 2001

Liu *et al.* showed using an immunoprecipitation technique that a three-way interaction between MID1, Alpha 4 and PP2Ac can occur *in vivo*. By deletion mapping, of Alpha 4, the authors found that amino acids 220-340 were sufficient to bind MID1, whereas amino acids 111-249 bound PP2Ac (Liu *et al.*, 2001). They noted that the MID1-interacting region at the Alpha 4 C-terminus is the most conserved region between various Alpha 4 orthologues. While this might indicate a long period of co-evolution of the function of these proteins, alternative explanations include adaptation of MID1 to bind the Alpha 4 C-terminus which has another specific conserved function. For example, the recognition and phosphorylation of the yeast Alpha 4 homologue by Target of Rapamycin (TOR) is essential for yeast cell survival (Jiang and Broach, 1999) and is a process that also occurs in humans (Inui *et al.*, 1998). In this regard, Alpha 4 orthologues are present in all eukaryotes, while MID1 orthologues are only found in vertebrates, consistent with

MID1 having adapted to interact with Alpha 4. However, this doesn't exclude the MID interaction having placed further selective pressure on the Alpha 4 C-terminus.

3.1.2 Trockenbacher *et al.*, 2001

A ubiquitylation-related function for MID1 was hypothesised by both Short *et al.* and Trockenbacher *et al.* based on the presence of the RING finger domains in many E3 ligase proteins. Supporting this was the finding in both yeast two hybrid screens identified a direct association of MID1 with polyubiquitin (data not included in Short *et al.*, 2002). Following on from their findings of an interaction with Alpha 4 and polyubiquitin, Trockenbacher *et al.* also reported that a primary fibroblast cell line from an Opitz syndrome patient showed an increase in the microtubular pool of PP2Ac compared with controls (Trockenbacher *et al.*, 2001). Addition of a proteasome inhibitor to inhibit degradation or turnover of polyubiquitylated proteins did not affect the levels of PP2Ac ubiquitylation in this patient cell line, whereas in control cells the amount of polyubiquitylated PP2Ac increased (Trockenbacher *et al.*, 2001). Thus it was concluded that the polyubiquitylation, and hence turnover, of PP2Ac was dependent on MID1. These experiments therefore provided the best evidence that MID1 has E3 ubiquitin ligase activity towards PP2Ac as predicted by the direct interaction of MID1 with Alpha 4.

3.1.3 Implications of the findings presented by Liu 2001, Trockenbacher 2001 and Short 2002

As presented in detail in Short *et al.*, 2002 and alluded to in Trockenbacher *et al.*, 2001, MID2 also independently interacts with Alpha 4 (and presumably PP2Ac) through its B-box domains (Short *et al.*, 2002) and can also dimerise with MID1. This ability is reflected in the high level of sequence similarity between MID1 and MID2 and the fact that the genomic regions in which the genes lie appear to have arisen from a regional

duplication of the X chromosome (Perry *et al.*, 1999). In this light, sequence similarities support the notion that *MID2* is the more ancient of the two *MID* genes, since *MID2* is more closely related to its paralogues than *MID1* is to its paralogues (Short and Cox, 2006). These additional findings have potentially important implications for understanding the pathogenesis of Opitz syndrome. For example, our laboratory has subsequently shown that endogenous levels of mutant *MID1* in an Opitz syndrome cell line is sufficient to perturb the localisation of exogenously expressed *MID2*. Therefore, such disruption of function of the microtubule bound *MID2*, which exhibits considerable overlap in temporal and spatial expression with *MID1*, could conceivably contribute to the presentation of Opitz syndrome. This apparent ‘redundancy’ between the *MID* proteins, whether full or partial, may also explain the lack of gross phenotype in *Mid1* knockout mice (unpublished observations) and highlights the importance of understanding the historic function of *MID2*.

So have any *MID2* mutations been found in patients with Opitz syndrome? Currently, there are no reported cases of Opitz syndrome in which a causative *MID2* alteration has been identified although an Alanine358 to Aspartate substitution in *MID2* (that was not seen in ~100 controls) has been identified in at least eight patients with Opitz syndrome that did not have demonstrable mutations in *MID1*. However, the current SNP database indicates this as a natural polymorphism. Alternative explanations for the absence of causative changes in *MID2* include the possibility that disruptions in *MID2* function are deleterious to embryonic development (such that no mutations in surviving newborns could ever be identified) or that such mutations result in a notably different phenotype. The former might be expected if indeed a large part of the control of PP2Ac is through *MID2*, while the latter might be true if *MID2* also has some unique functions or is more essential in tissues different to those in which *MID1* function is critical (i.e. dependent on the

difference in the level of expression of the two genes in any given tissue). However, further experimentation will be required to determine the level of impediment to normal MID2 function at endogenous levels and whether it indeed possesses E3 ligase activity.

3.1.4 Alpha 4 as a candidate for FG syndrome

Like *MID1* and *MID2*, the *Alpha 4* gene was found to reside on the X chromosome, at Xq13.1. At the time this region of the X chromosome was of interest because of the mapping of a number of clinical disorders to that chromosomal interval, including an FG syndrome locus, an X-linked Mental Retardation (XLMR) locus, and a CHARGE-like locus (CHARGE is an acronym for the most common features seen in patients with mutations within this region, including: Coloboma (eye defect), Hear defects, Atresia of the choanae (narrow nasal passages), Retardation of growth and developmental delay, Genital abnormalities, and Ear abnormalities). The partial overlap in phenotype between these disorders and Opitz syndrome, and the evidence of strong expression of MID1 and Alpha 4 in the developing central nervous system, prompted a more detailed investigation of the Alpha 4 sequence in these cases. Yet despite the phenotypic similarities, no sequence abnormalities were found in the numerous FG syndrome and XLMR cases screened. However, an additional case was included in the screening because of some uncertainty as to whether it represented a *bona fide* case of FG syndrome. Specifically, two brothers were identified who exhibited a range of malformations, including (but not limited to) mental retardation, prominent forehead, low set ears, broad nasal bridge, heart septation defects, and agenesis of the corpus callosum (ACC). Sometime later, this unique presentation was assigned as a new syndrome. Regardless, sequencing of the boys' DNAs revealed an unusual nucleotide abnormality: a single base pair deletion (del -57T) and a substitution (-55T>A) both within the 5' untranslated region that was associated with a

decrease in the levels of Alpha 4 protein (Graham *et al.*, 2003). Nevertheless, the overlap of features with those seen in cases of Opitz syndrome strengthens the evidence for the role of heterogeneity being a key effector in the pathogenesis of these diseases. Further supporting this was the findings in mice of significant similarities in the pattern of expression of *Alpha 4* and *Mid1* (Everett and Brautigan, 2002).

3.1.5 A working model for MID1-Alpha 4 function

The findings from yeast two hybrid screens, gene expression studies and cell biological experiments have enabled the synthesis of a working model for how these components might interact. The site of activity of the pathway is on the microtubule cytoskeleton and under the model is likely for switching microtubule protein activity (through an overriding control of phosphorylation). In this scenario, at the core of the pathway would be a microtubule-bound protein, MID1, acting as a microtubule-specific ubiquitin ligase that interacts with Alpha 4 that in-turn acts as “bait” to attract PP2Ac for the purpose of its destruction, resulting in an increase in general microtubule protein phosphorylation levels. If mutations or decreased levels of Alpha 4 or MID1 protein occurred, it is conceivable that the resulting change in microtubule phosphorylation state might be such that specific further “downstream” signalling is either not triggered, or there is a lack of signalling to promote a change in pathway activation. The effective action would require both Alpha 4 and MID1 proteins to be microtubule associated at the same time in order to degrade microtubular PP2Ac, thus reduce microtubule-directed phosphatase activity. If mutated MID1 is no longer capable of normal microtubule association, this would be predicted to result in microtubule protein hypophosphorylation. Additionally, the same would hold true for Alpha 4: reduction in the amount of Alpha 4 protein (as observed in the cells of the patient in Graham *et al.*, 2003), the ability of MID1

to target PP2Ac for proteolysis would theoretically be compromised. Although such a scenario may not result in complete inhibition of MID1 activity towards microtubular PP2Ac, it would likely result in a modification overall microtubule-directed phosphatase activity with less (but not all) PP2Ac degraded. This could account for a similar, but distinct phenotype in the FG-like patient.

3.2 Moving forward to Short and Cox, 2006

Together, the MID2-MID1 and MID1/2-Alpha 4/PP2Ac interactions suggest a complex interplay of factors that is likely to impact on the phenotypic presentation of Opitz and FG-like syndromes. With a view to dissecting these relationships further, and particularly with the aim of better understanding basic MID1 and MID2 function, a bioinformatics approach was taken to identify and characterise any other homologues. It was envisaged that this approach would provide a framework to facilitate definition of new sequence/structure/functional relationships for MID1 and MID2, that would improve understanding of the basic functional processes which determine MID1 and MID2 specific functions.

To this end, an approach was taken to investigate domain organisation and conservation in the MID proteins and how they might have evolved basic functional properties (such as microtubule association).

3.2.1 Approach and methodological detail

In order to determine functionally important amino acids within MID1 and MID2 (through conservation of sequence/structure), a search for all MID1/MID2 homologues was conducted. The definitive search for domain homologues re-assessed the domains within entire RBCC domain superfamily. The process of identifying the domains made subclassification of the RBCC superfamily into C-terminal domain specific subfamilies

possible. Once subfamilies were identified, careful alignment of the MID1 subfamily was used to generate phylogenies to infer relatedness and additionally, allow analysis of MID1 homologous sequence conservation across the remaining RBCC proteins. This was done using a profile hidden Markov model (HMM) approach to provide homology matches between individual query amino acid sequences against a special database generated from the aligned MID1 homologues. This approach differs from maximum parsimony or maximum likelihood approaches in determining identity. Regular molecular phylogeny programs query homology between two (or multiple) peptide sequences on a *position independent* basis, so they sometimes fail to accurately compare moderately diverged sequences. The approach taken with a suite of profile modeling programmes, called HMMER, used dissection of a subject *alignment* into short aligned fragments of a preset size (a 'window'). Specifically, each individual alignment window progressively advances one amino acid forward such that each alignment window overlaps the previous window by n-1 amino acids (for an 'n' amino acid window size). HMMER then processes each of the windowed alignments to generate statistical model *profiles* of the alignments by analysing the conservation of each position of the alignment and generating a model of the probability of expected residues. These are then placed together as a single library of profiles representing the entire original alignment. This means that (in subsequent analyses) individual scores of homology are received for a scan of a subject sequence against a *library* of window *profiles* (originally generated from the multiple sequence alignment fragments). The power of this method of measuring sequence homology is that the profile captures *position-dependent* information about amino acids conserved between homologues. This is important information from a structure/function standpoint because strong conservation between specific residues (for example, Cysteine, Histidine, Lysine, Valine, Isoleucine) within proteins which otherwise have relatively low (20-40%) amino

acid similarity are important for basic low-level properties of many proteins (such as domain homology). Therefore, during the process of comparing a subject amino acid sequence against a profile HMM database, scores for profile matches preceding regions of high homology progressively increase and scores for matches to successive lower-homology profiles decrease, when surrounded by weakly conserved flanking sequences. The significant amount of data generated by the scan (with over 24,000 results) required a histogram plot to represent the scores (and expectancy values for each) associated with matches.

The technique used to process and plot the data was one adapted from a proof-of-concept study using the cadherin and EF-hand protein superfamilies as a measure of the accuracy of the technique to blindly (and accurately) detect regions of protein structural significance (called *signatures*) within subfamilies (Truong and Ikura, 2002). The detailed output provides amino-acid scale homology scores, and allows the definition of discrete and broad homologies. The groundbreaking study by Truong and Ikura was used as the basis for Short and Cox, 2006 because early tests showed that traditional homology methods could not account for the vast variation across the RBCC superfamily (for reasons outlined previously). Use of the technique was successful in determining a MID1 subfamily specific signature, which formed the later focus of the study.

Chapter Four
Short and Cox, 2006

Subclassification of the RBCC/TRIM Superfamily Reveals a Novel Motif Necessary for Microtubule Binding*^[5]

Received for publication, November 29, 2005 Published, JBC Papers in Press, January 23, 2006, DOI 10.1074/jbc.M512755200

Kieran M. Short and Timothy C. Cox¹

From the School of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia 5005, Australia and the Department of Anatomy and Cell Biology, Monash University, Clayton, Victoria 3800, Australia

The biological significance of RBCC (N-terminal RING finger/B-box/coiled coil) proteins is increasingly being appreciated following demonstrated roles in disease pathogenesis, tumorigenesis, and retroviral protective activity. Found in all multicellular eukaryotes, RBCC proteins are involved in a vast array of intracellular functions; but as a general rule, they appear to function as part of large protein complexes and possess ubiquitin-protein isopeptide ligase activity. Those members characterized to date have diverse C-terminal domain compositions and equally diverse subcellular localizations and functions. Using a bioinformatics approach, we have identified some new RBCC proteins that help define a subfamily that shares an identical domain arrangement (MID1, MID2, TRIM9, TNL, TRIM36, and TRIFIC). Significantly, we show that all analyzed members of this subfamily associate with the microtubule cytoskeleton, suggesting that subcellular compartmentalization is determined by the unique domain architecture, which may in turn reflect basic functional similarities. We also report a new motif called the COS box, which is found within these proteins, the MURF family, and a distantly related non-RBCC microtubule-binding protein. Notably, we demonstrate that mutations in the COS box abolish microtubule binding ability, whereas its incorporation into a non-microtubule-binding RBCC protein redirects it to microtubule structures. Further bioinformatics investigation permitted subclassification of the entire human RBCC complement into nine subfamilies based on their varied C-terminal domain compositions. This classification schema may aid the understanding of the molecular function of members of each subgroup and their potential involvement in both basic cellular processes and human disease.

Members of the RBCC (N-terminal RING finger/B-box/coiled coil) or TRIM (tripartite motif) family of proteins perform a diverse array of cellular roles, yet are believed to share some functional properties: 1) act as a scaffold for the assembly of larger multiprotein complexes and 2) possess RING-dependent ubiquitin ligase activity (1, 2). The RBCC domain can be found in isolation or in combination with a variety of

other C-terminal domains, including the NHL (NCL-1/HT2A/LIN-41 repeat), immunoglobulin, MATH (meprin and tumor necrosis factor receptor-associated factor homology), B30.2-like/RFP (Ret finger protein)/SPRY (SplA and ryanodine receptor) (the largest subgroup in humans), ARF (ADP-ribosylation factor), PHD (plant homeodomain finger), and BROMO domains (1, 2). As a family of proteins, their biological significance is perhaps best highlighted by the growing number that have a demonstrated role in disease pathogenesis, including immunological and developmental disorders, tumorigenesis, and retroviral protective activity (3–7).

We have previously identified and characterized two RBCC proteins, MID1 and its closely related homolog, MID2. MID1 and MID2 contain a B30.2-like domain at their C terminus and a single fibronectin type III (FN3)² motif between it and their N-terminal RBCC domain (8). Both proteins have been shown to associate with microtubules (8, 9) and do so through both homo- and heterodimerization (5). Consistent with their high level of sequence and structural similarity, MID1 and MID2 can even interact with some of the same proteins (5), and both likely possess ubiquitin-protein isopeptide ligase activity (10). Mutations in MID1 have been identified in patients with Opitz syndrome, a disorder recognized by a combination of congenital anomalies that includes cleft lip and palate as well as heart and anogenital defects. Although mutations have been found in most domains of MID1, mutations in the C-terminal domain predominate and result in significantly altered cellular localization of the protein (4, 11) as well as disrupted targeting of its ubiquitin ligase activity (10).

At the time of their identification, MID1 and MID2 represented the only known RBCC superfamily members with an FN3 motif. A third mammalian protein (Spring/TRIM9) that shares the same domain organization as MID1 and MID2 was subsequently discovered (12, 13). This protein is specifically expressed in neural tissue and appears to play a regulatory role in synaptic vesicle exocytosis (12). To identify additional proteins with a similar domain arrangement, we undertook an analysis of public expressed sequence tag (EST) and genomic data, which indicated that other related proteins may indeed be encoded by the human genome. Support for this notion came recently from the report of Haprin, a new testis-specific FN3 motif- and B30.2-like domain-containing RBCC protein that is clearly the murine ortholog of a sequence identified in our screen. Functional data suggest that Haprin plays a regulatory role in exocytosis of the sperm vesicle (14).

In this study, we report the identification of two novel proteins, TRIFIC and TNL, which share the same overall domain organization as MID1, MID2, TRIM9 (the human ortholog of rat Spring), and TRIM36 (the human ortholog of mouse Haprin), and show for the first time that all analyzed members of this RBCC subfamily associate with the microtubule

* This work was supported in part by Project Grant 157958 from the National Health and Medical Research Council of Australia. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

^[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Fig. 1.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY253917, AY251386, AY251388, AY251389, NM_070994, and AY251387.

The amino acid sequences of these proteins can be accessed through NCBI Protein Database under NCBI accession numbers AAP69949, AAP51206, AAP51208, AAP51209, NP_503395, and AAP51207.

¹ To whom correspondence should be addressed: Dept. of Anatomy and Cell Biology, Monash University, Bldg. 13c, Rm. c180, Clayton, Victoria 3800, Australia. Tel.: 61-3-9905-2712; Fax: 61-3-9905-2766; E-mail: timothy.cox@med.monash.edu.au.

² The abbreviations used are: FN3, fibronectin type III; EST, expressed sequence tag; HMM, hidden Markov model; MSAs, multiple sequence alignments; GFP, green fluorescent protein; X-gal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside.

A New Subfamily of Microtubule-associated RBCC Proteins

cytoskeleton. We also describe a novel amino acid signature adjacent to the coiled coil (named the COS box) that, when mutated at key conserved residues, results in a complete loss of microtubule localization, but not dimerization ability. Additionally, a fusion protein generated from a C terminus containing a COS box from a microtubule-associated protein and an RBCC/TRIM domain from a non-cytoskeleton-associated protein redirects localization to microtubule structures. We subsequently reanalyzed the entire human RBCC family, enabling classification into nine subgroups based on a consensus C-terminal domain organization provided by Pfam, SMART, PRINTS, PROFILE, and PATTERN analyses. The implications of this subclassification in understanding the broader function of RBCC proteins are discussed.

EXPERIMENTAL PROCEDURES

Alignment and Phylogenetic and Hidden Markov Model (HMM) Analysis—General amino acid and nucleotide sequence handling was aided by *Vector NTI Suite* (version 7.01, Invitrogen, Mulgrave, New South Wales, Australia), and multiple sequence alignments (MSAs) were post-processed for publication using CHROMA, removing regions of low homology across the family and producing an 85% consensus according to convention (15). MSAs were generated using a combination of manual, ClustalX (16) with Blosum62 (17) and Gonnet (18) matrices, and T-COFFEE (19) alignments using a best fit approach. Phylogenetic analysis was performed on MSA data using PHYLIP Version 3.62 (20). Human sequences were used together with *Drosophila melanogaster* TRIM9 as a representative of an ancestral RBCC/FN3/B30.2-like subfamily member. A maximum likelihood approach was used to infer phylogeny using the SEQBOOT, ProML, CONSENSE, and DRAWTREE programs, with 100 bootstrap resample data sets being used to test inferred phylogeny vigor. Bootstrap consensus figures were placed on the tree with post-processing using Adobe Illustrator Version 10 (Adobe Systems Inc., San Jose, CA). HMM analysis used pre- and post-processing Perl scripts³ for the handling of MSA data and HMMER execution (21). Two Sun Blade 100 workstations and an Apple G5 dual CPU workstation were used for data generation, and all HMMs were created using Sun Solaris Version 6.0 and G5 Altivec optimized versions of HMMER Version 2.3.2 (22). Subsequently, Microsoft Excel was used for converting post-processed HMMER data to a graphical output format using in-built pivot table functions. CHROMA-processed MSAs, the phylogenetic tree, and HMMER graphical output was presented using Adobe Illustrator Version 10.

Plasmid Constructions—MID1 and MID2 cDNAs and fusions were already cloned and available (4, 5, 8). Human full-length cDNA clones from the IMAGE collection were identified for TRIM9 (IMAGE clone 5243288) and TRIM36 (IMAGE clone 4826462), but not for TRIFIC. Instead, a murine full-length *Trifc* cDNA (IMAGE clone 3813814) that encodes a protein with 98% amino acid identity to human TRIFIC was identified. A full-length cDNA clone was not available for either human *TNL* or mouse *Tnl* at the time of this study. Full-length cDNAs were amplified for the purpose of recloning using the following sequence-specific primer pairs: hTRIM9-dTop (5'-CACCATGGAGGAGATG-GAAGAGGAG-3') and hTRIM9-rEco (5'-GGAATTCTTAGGCT-ATTGATGCTCTGCTG-3'); hTRIM36-dTop (5'-CACCATGGAGGCGATGGCTCAGATTCCG-3') and hTRIM36-rNot (5'-ATAA-GAATGCGGCCCTACATGTCTCTTGGTATTCCAG-3'); and mTrifc-dTop (5'-CACCATGGGTGGTGCCTTGAATAAATGC-3') and mTrifc-rEco (5'-GGAATTCTCAGTCCAGTTTGGCA-

AAGCCC-3'). All cDNA sequences were verified by alignment with publicly available genomic data. Of note, IMAGE clone 5243288 contains a point mutation resulting in a single amino acid alteration in the B30.2-like domain. This change was corrected by site-directed mutagenesis (Stratagene, La Jolla, CA) using primers fTRIM9-1959C→A (5'-GGGTCCTCCTCGACTTAAATAGAAAAAC-3') and rTrim9-1959C→A (5'-GTTTTTCTATTTAAGTCGAGGAG-GACCC-3'). IMAGE clone 4826462 contains a large insertion of bacterial sequence within the cDNA, which was corrected using a high fidelity PCR approach with cloning primers in combination with specifically designed primers T36EST-fixFor (5'-CCACCGTGT-AACCACCGTGTAACTATGAGCAGTGCCTACAAAACC-TTAAAGGAAAAG-3') and T36EST-fixRev (5'-GCTTTCCTT-TAAGGTTTTGTAGGCACTGCTCATAGTGGTTACACGGTG-3'). Sequence-verified cDNAs were cloned into the mammalian green fluorescent protein (GFP) expression vector pcDNA-DEST53 and the yeast two-hybrid Gal4 activation domain (pDEST-22) and DNA-binding domain (pDEST-32) vectors using the Gateway cloning system (Invitrogen). Mutations were introduced into the GFP-MID1 construct to study COS box function and were generated by site-directed mutagenesis (Stratagene) using primers fMIDcosFLQ→AAA (5'-GGAGAATGATCATGCGCGTGC GGCCGCTACTGCTAAGA-ATATCACCGAG-3'), rMidcosFLQ→AAA (5'-CTCGGTGATA-TTCTTAGCAGTAGCGGCCGCACGCGCATGATCATTCTCC-3'), fMIDcosLDY→AAA (5'-GAAGAACTGCTAGAATGTGC-GGCCGCCCTTACAGTCCCAACCCTC-3'), and rMidcosLDY→AAA (5'-GAGGGTTGGGAGCTGTAAGGGCGGCCGCACATTC-TAGCAGTTTCTTC-3'). All introduced mutations were confirmed by automated sequencing.

Fusions of the TRIM37 tripartite domain with the COS box or C-terminal domains (including the COS box) were generated by PCR amplification of *TRIM37D* cDNA (kindly supplied by Anna-Elina Lehesjoki, Neuroscience Center and Folkhälsan Institute of Genetics, Biomedicum Helsinki, University of Helsinki, Finland) using primers TRIMD-F (5'-CACCATGGATGAACAGAGCGTGGAGAGC-3') and TRIMD-R (5'-CCGCTTGGGAGATGAGTGATGCATGAACTTGCTGAAAC-ATCATAAGGATCTCTG-3'), which contain extensions that overlap the fusion region from the MID1 C-terminal fragments. The MID1 fragments were amplified with forward primer COS-F (5'-CCTTATG-ATGTTTCAGCAAGTTCATGCATCACTCATCTCCAAGCGG-3') and reverse primer M1FN3-R (5'-GTTTGTCTTCAACTTCCCA-GG-3'), from a mutated MID1 cDNA template (containing an introduced TGA codon at position 380) for the COS box fusion, or with M1SPRY-R (5'-TCACGGCAGCTGCTCTGTGCAGTC-3'), from a wild-type MID1 cDNA for the full C-terminal domain fusion. Products from the TRIMD reaction were pooled with either COS box or C-terminal reactions, and extension of the annealed overlapping regions was completed with a secondary round of PCR. The full-length fragments were purified and cloned into pENTR-D/TOPO (Invitrogen), and positives were subcloned into vector pcDNA-DEST53 using the Gateway cloning system.

Cell Culture, Transfection, and Immunofluorescence—COS-1 cells were grown in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with qualified heat-inactivated 10% fetal calf serum (Invitrogen) at 37 °C in 5% CO₂. Transfection of plasmid DNA was performed using FuGENE 6 transfection reagent (Roche Applied Science, Castle Hill, New South Wales). Transfected cells were grown on coverslips for 18 h in Dulbecco's modified Eagle's medium plus 10% fetal calf serum and then fixed as described previously (4). Microtubule staining was performed post-fixation using an anti- α -

³ Modified Perl scripts used for large-scale GenBank™ HMM scanning (which conserve disk space) can be requested from the authors, as can a Cygwin-packaged version to run the analysis on a Windows PC running Cygwin (Redhat) with Perl Version 5.8.

A New Subfamily of Microtubule-associated RBCC Proteins

tubulin primary antibody and Texas Red-conjugated anti-mouse secondary antibody (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA). Fused Myc tag was detected using mouse anti-Myc antibody 9E10, and the same mouse secondary antibody was used as per tubulin. Nuclei were visualized using the DNA-specific Hoechst stain (Sigma, Sydney, New South Wales). GFP and Texas Red fluorescence was detected at appropriate wavelength light using an Olympus AX70 microscope. Images were captured using a Photometrics CE200A camera electronics unit and processed using Photoshop Version 7 software (Adobe Systems Inc.).

Yeast Two-hybrid Analysis—The ProQuest yeast two-hybrid system (Invitrogen) was employed to assess interactions between MID1, MID2, and other subfamily members. All yeast two-hybrid fusion constructs were transfected into the *Saccharomyces cerevisiae* MaV203 yeast strain (*MAT α* , *leu2-3,112*, *trp1-901*, *his3 Δ 200*, *ade2-101*, *gal4 Δ* , *gal80 Δ* , *SPAL10::URA3*, *GAL1::lacZ*, *LYS2::P_{HIS3}UAS_{GAL1}-HIS3*, *can1^{rk}*, *cyh2^{rk}*) by polyethylene glycol/LiAc/Tris-EDTA transfection (23). The *HIS3* reporter was used to determine the level of self-activation of the fusions based on the level of 3-aminotriazole resistance of the transformed yeast. 50 mM 3-aminotriazole was found to be an adequate level for the assay of all generated fusions. Transfections were plated on selective synthetic complete medium lacking Leu and Trp, from which colonies were picked, and cultures were established and grown for 24 h at 30 °C in selective medium. Normalized diluted cultures (10 μ l at 0.1 *A*₆₀₀) were then spotted onto interaction-selective synthetic complete medium lacking Leu, Trp, and His and containing 50 mM 3-aminotriazole and incubated at 30 °C for 48 h. For X-gal assays, spots of selectively cultured transfected yeast were placed on nitrocellulose Hybond-N⁺ membrane (Amersham Biosciences, Castle Hill) overlaid on complete medium and allowed to grow at 30 °C for 48 h. X-gal assays were performed as reported previously (24), and all post-assay plates and filters were scanned using a Hewlett-Packard Scanjet scanner with Adobe Photoshop.

Immunoprecipitation and Western Blot Analysis—Preparations of the various GFP- and Myc-tagged expression constructs were made using a DNA plasmid midi kit (Promega Corp., Annandale, New South Wales). 6 pmol (~3 μ g) of each construct were transfected into 1 \times 10⁷ COS-1 cells using FuGENE 6 transfection reagent. After a 24-h incubation, cells were scraped from the culture dish and lysed on ice for 30 min in 1 ml of lysis buffer (50 mM Tris-HCl (pH 7.4), 300 mM NaCl, 5 mM EDTA, and 1.0% Triton X-100). Cell lysates were cleared by centrifugation at 4 °C for 15 min at 16,000 \times g, and protein extract was recovered as the supernatant. After preclearing 200 μ l of protein extract with 10 μ l of 50% protein G-Sepharose bead slurry (Sigma), extracts were incubated with 1 μ g of antibody for 2 h at 4 °C and then for another 2 h with 20 μ l of fresh 50% protein G-Sepharose bead slurry. The beads were washed four times with wash buffer (50 mM Tris-HCl (pH 7.4), 300 mM NaCl, 5 mM EDTA, and 0.1% Triton X-100), and protein was eluted from the beads by boiling in 2 \times SDS loading buffer. Proteins were separated by 10% SDS-PAGE and blotted onto Hybond-C membranes (Amersham Biosciences) using a semidry transfer apparatus (Owl Separation Systems, Portsmouth, NH). Membranes were blocked, incubated with the appropriate primary antibody, washed, incubated with the appropriate horseradish peroxidase-conjugated secondary antibody, and washed again according to established methods (43). Detection was carried out using an enhanced chemiluminescence kit (ECL, Amersham Biosciences) following the manufacturer's instructions. The antibodies used in immunoprecipitation and Western blot analysis included rabbit anti-GFP polyclonal antibody (gift from Pam Silver, Dana-Farber Cancer Institute, Boston), mouse anti-Myc monoclonal anti-

body 9E10 (gift from Stephen Dalton, University of Adelaide), and horseradish peroxidase-conjugated anti-rabbit and anti-mouse secondary antibodies (Amersham Biosciences).

RESULTS

Identification of New RBCC/FN3/B30.2 Proteins—To identify additional RBCC/FN3/B30.2-like proteins, we performed separate BlastX and PSI-BLAST searches on both the publicly available human genome and EST data bases using the full-length sequences of the three reported human FN3 motif-containing RBCC proteins (MID1, MID2, and TRIM9) as well as their individual RBCC, FN3, and B30.2-like sequences. Included in this analysis were all previously identified human full-length RBCC sequences (Fig. 1A), excluding those TRIM proteins that were not of human origin or that were missing part or all of the RBCC domain components (see Fig. 1 legend). All RBCC sequences to be analyzed were retrieved and further scrutinized using InterPro (25), SMART (26), Pfam (27), PROSITE (28), PRINTS (29), and MultiCoil (30). These programs were chosen because they use a range of analyses to determine sequence domain similarity by comparison with domain data bases from HMM to PROFILE and pattern matching methods as well as primary sequence analysis. The data provided by these searches allowed the thorough identification of particular subgroups of RBCC proteins based on the presence or absence of currently identified domain families in the Pfam Database (Version 16), PROSITE PROFILE Database (Version 18.37), and MultiCoil when no coil was detected using domain search tools (Fig. 1, A and B).

These analyses identified two human RBCC proteins with significant matches to the FN3 consensus: TRIM36 (14, 31) and FLJ23229 (TRIM46). Although Pfam and SMART Database scores were not significant for a SPRY domain (a submotif of the B30.2-like domain) match over the C terminus of each of these proteins, both were shown to have significant scores for B30.2-like domains using a PROFILE search (available at PROSITE), which is the only tool available to detect complete B30.2-like domains. To provide a similar comparison, the B30.2-like profile matrix (PS50188) was converted to a HMMER data base using ptoh (32). Results of scans performed against the identified RBCC/FN3/B30.2-like proteins were also significant matches, similar to the PROFILE scan. Using the domains from homologous proteins as a guide, available EST data and genomic sequence representing *FLJ23229/TRIM46* were assembled to build a predicted cDNA, which was assisted with predictions available from the GenBankTM Data Bank. From this, we identified that the putative protein contained the same domain architecture as the other identified proteins, and we hereafter refer to this novel protein as TRIFIC (tripartite, fibronectin type III, and C-terminal B30.2-like motif; human GenBankTM accession number AY251386 (nucleotide) and NCBI accession number AAP51206 (protein) and mouse GenBankTM accession number AY251388 (nucleotide) and NCBI accession number AAP51208 (protein).

Alignment and Phylogenetic Analysis of the RBCC/FN3/B30.2 Proteins—Protein MSA indicated a similar motif arrangement in each of the human FN3 motif-containing RBCC sequences (supplemental Fig. 1), with the most notable differences being within the core of the C3HC4 RING finger domain, which is the most variable part of all RING fingers (refer to Pfam accession number zf-C3HC4 and SMART accession number SM00184). The alignment also reiterated the similarity across the B30.2-like region for the entire subfamily and significantly identified regions of high homology not previously associated with domains or protein function.

MSA was used as the basis to perform a maximum likelihood phylogenetic analysis to determine the relative interrelationships between

A New Subfamily of Microtubule-associated RBCC Proteins

A

Consensus Name	Other Names	Location	Subfamily
MID2	FXY2,RNF60,TRIM1	Xq22.3	C-I
TRIM9	Spring	14q22.1	C-I
MID1	TRIM18,FXY,RNF59	Xp22	C-I
TRIM36	HAPRIN,RBCC728,RNF98	5q22	C-I
TRIFIC	TRIFIC,FLJ23229,TRIM46	1q22	C-I
TNL	TNL,TRIM67	1q42	C-I
MURF1	RNF28,IRF,SMR2,TRIM63	1p36.11	C-II
MURF2	RNF29,TRIM55	8q12	C-II
MURF3	RNF30,TRIM54	2p23.3	C-II
TRIM42	TRIM42	3q23	C-III
TRIM4	RNF87	7q22.1	C-IV
TRIM5	RNF88	11p15	C-IV
TRIM6	RNF89	11p15	C-IV
HERF1	RFB30,TRIM10	6p21.3	C-IV
TRIM11	RNF92	1q42.13	C-IV
TRIM15	ZNFB7,RNF93	6p21	C-IV
TERF	RBCC,RNF16,TRIM17	1q42	C-IV
SS-A/Ro	RNF81,SSA1,TRIM21	11p15.5	C-IV
STAF50	GPSTAF50,RNF94,TRIM22	11p15	C-IV
EFF	TRIM25,RNF147,ZNF147	17q23.1	C-IV
TRIM26	AFP,RNF95,ZNF173	6p21.3	C-IV
RFP	RNF76,TRIM27	6p22	C-IV
RNF21	TRIM34	11p15	C-IV
HLS5	MAIR,TRIM35	8p21	C-IV
RoRet	RNF15,TRIM38	6p21.3	C-IV
TRIM39	RNF23,TFP	6p21.3	C-IV
TRIM41	MGC1127	5q35.3	C-IV
TRIM43	-	2q11	C-IV
TRIM47	GOA,RNF100	17q25	C-IV
TRIM50A	FLJ32804,TRIM50	7q11.23	C-IV
TRIM60	RNF33,RNF129,FLJ35882	4q32.3	C-IV
TRIM62	FLJ10759,FLJ16558	1p35.1	C-IV
TRIM64	C11orf28	11q14.3	C-IV
TRIM68	SS-56,RNF137,FLJ10369	11p15.4	C-IV
TRIM7	RNF90	5q35.3	C-V
TRIM8	GERP,RNF27	10q24.3	C-V
RFP2	LEU5,RNF77,TRIM13	13q14	C-V
PML	MYL,RNF71,TRIM19	15q22	C-V
TRIM31	RING,RNF,HCGI,C6orf13	6p21.3	C-V
RNF35	TRIM40	6p21.32	C-V
TRIM56	RNF109	7q11.2	C-V
TRIM59	MRF1,TSBF1,MGC26631	7q11.2	C-V
TIF1-alpha	TIF1,hTIF1,Tif1a,RNF82,TRIM24	7q33	C-VI
TIF1-beta	KAP1,TF1B,RNF96,TRIM28	19q13.4	C-VI
TIF1gamma	TRIM33	1p13.1	C-VI
NARF	RNF86,TRIM2	4q31.3	C-VII
BERP	HAC1,RNF97,RNF22,TRIM3	11p15.5	C-VII
TRIM32	HT2A,TATIP	9q33.2	C-VII
TRIM45	FLJ13181,RNF99	1p13.1	C-VII
MUL	TRIM37	17q23	C-VIII
ARD1	ARFD1,TRIM23	5q12	C-IX

B

Family	N-terminus	C-terminus	No.
C-I	R-B1-B2-CC	COS-FN3-B30.2-like	6
C-II	R-B2-CC	COS-AR	3
C-III	Cys-R-B1-B2-CC	COS-FN3	1
C-IV	R-B1-B2-CC	B30.2-like	24
C-V	R-B1-B2-CC		8
C-VI	R-B1-B2-CC	PHD-Bromo	3
C-VII	R-B2-CC	IG-NHL repeats	4
C-VIII	R-B2-CC	MATH	1
C-IX	R-B1-B2-CC	ARF	1

FIGURE 1. Summary of the human RBCC superfamily and their domain-based subfamily classification. A, classification relied on the identification of domains within the amino acid sequences of the proteins listed. The identity, alternative names, human chromosome location, and subfamily allocation for each gene are indicated. B, shown is a schematic representation of each subfamily, with classification of subfamilies based on the differences in C-terminal domain composition: C-I (COS/FN3/B30.2-like domains),

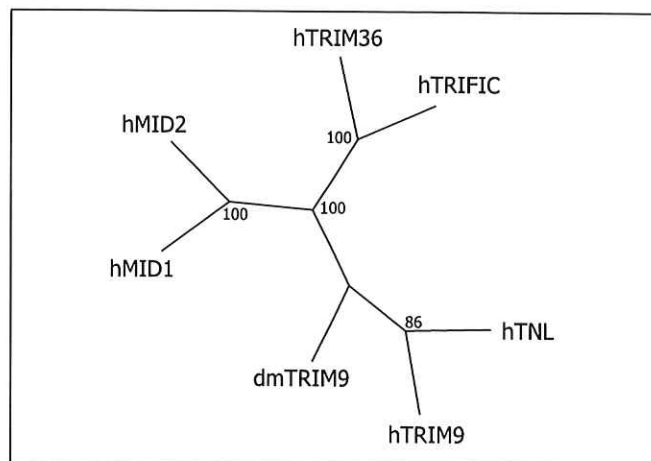


FIGURE 2. Phylogenetic tree of the C-I subfamily members. An unrooted tree was plotted using data obtained from a bootstrapped maximum likelihood analysis of a multiple sequence alignment. Bootstrap consensus figures (from 100 data sets) are plotted at tree branch points, indicating confidence of the inferred phylogeny. The tree shows the radiation of C-I members from an "ancestral" protein through gene duplication events. Of all human C-I subfamily members, human (h) TRIM9 is the closest ortholog to the ancestral *D. melanogaster* (dm) TRIM9 protein.

MID1, MID2, TRIM36, TRIFIC, TRIM9, and a novel TRIM9 homolog (TNL; discussed below). All human sequences and the *D. melanogaster* TRIM9 sequence were used as the basis for a maximum likelihood analysis, and an unrooted phylogenetic tree was plotted (Fig. 2). The tree shows the relationship between the RBCC/FN3/B30.2-like human sequences and MID1 and MID2, TRIFIC and TRIM36, and TRIM9 and TNL forming paired groups, indicating they are the result of progressive gene duplication events. When murine and chick sequences were added to the analysis (data not shown), the orthologous sequences formed monophyletic groups to the exclusion of paralogs, indicating that the gene duplication events predated the speciation of humans, mice, and chicks.

HMMER-based Analysis of the RBCC Superfamily—Verification that we had identified all members containing the FN3 motif came with the reanalysis of each of the 51 human full-length RBCC proteins using HMMER (22). A concatenated HMM data base was generated (as per Ref. 21) using 698 50-amino acid "window" alignments derived from an MSA of four of the five human FN3 motif-containing RBCC proteins (and their orthologs, based on the MSA depicted in supplemental Fig. 1). The calibrated data base containing 698 HMMs representing significant sequence identity was then used for scanning all human RBCC proteins in an attempt to identify sequence window homology. We chose to omit the TRIFIC sequence from the HMM data base so it could be used as a "positive" reference sequence in these analyses, such that each recognized motif would be identifiable within the histogram output to aid interpretation and visual presentation of the data for the entire protein family (Fig. 3A). The sensitivity and visual presentation were aided using 50-amino acid wide "sliding windows" of the alignment so that a score for an HMM search (22) identi-

C-II (COS/acid-rich (AR) region), C-III (COS/FN3 domains), C-IV (B30.2-like domains), C-V (as yet no C-terminal motifs identified), C-VI (PHD/BROMO domains), C-VII (immunoglobulin/NHL repeats), C-VIII (MATH domain), and C-IX (ADP-ribosylation factor (ARF) domain). Domains in *dashed boxes* may be variably presented. The TRIM proteins excluded from the study do not follow the TRIM nomenclature and lack the characteristic tripartite (RING (R), B-box, and coiled coil (CC)) RBCC motif: TRIM14, TRIM16 (EBBP), TRIM20 (PYRIN/MEFV), TRIM29 (ATDC), TRIM44 (DIPB), and TRIM66, all of which do not have RING fingers; TRIM51 and TRIM58, which have no RING or B-boxes; TRIM48, TRIM49, TRIM52, and TRIM61, which have no coiled coil; and TRIM65, which has no RING, B-boxes, or coiled coil. TRIM12 and TRIM30 were also excluded because they represent murine proteins with no obvious human ortholog.

A New Subfamily of Microtubule-associated RBCC Proteins

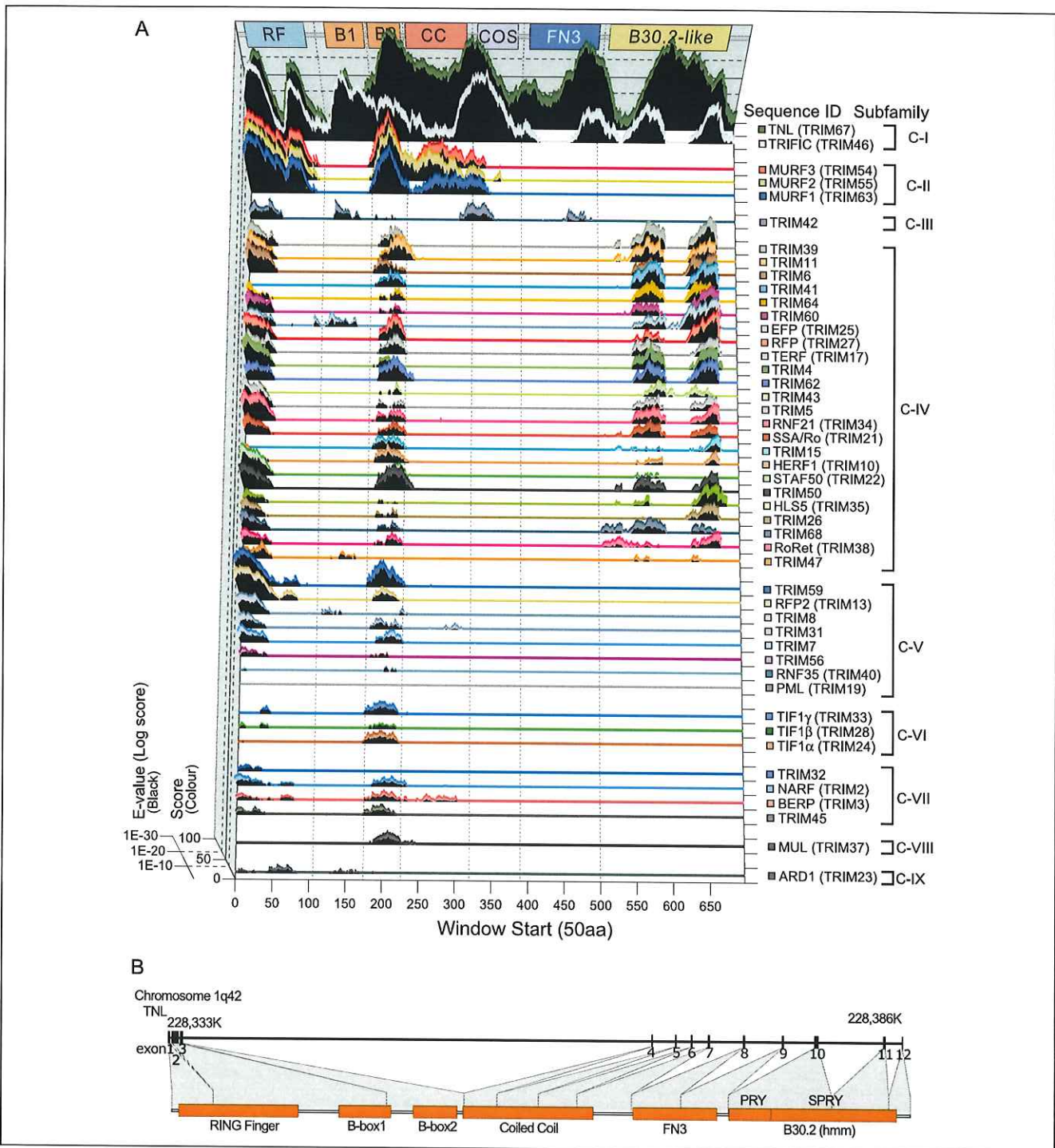
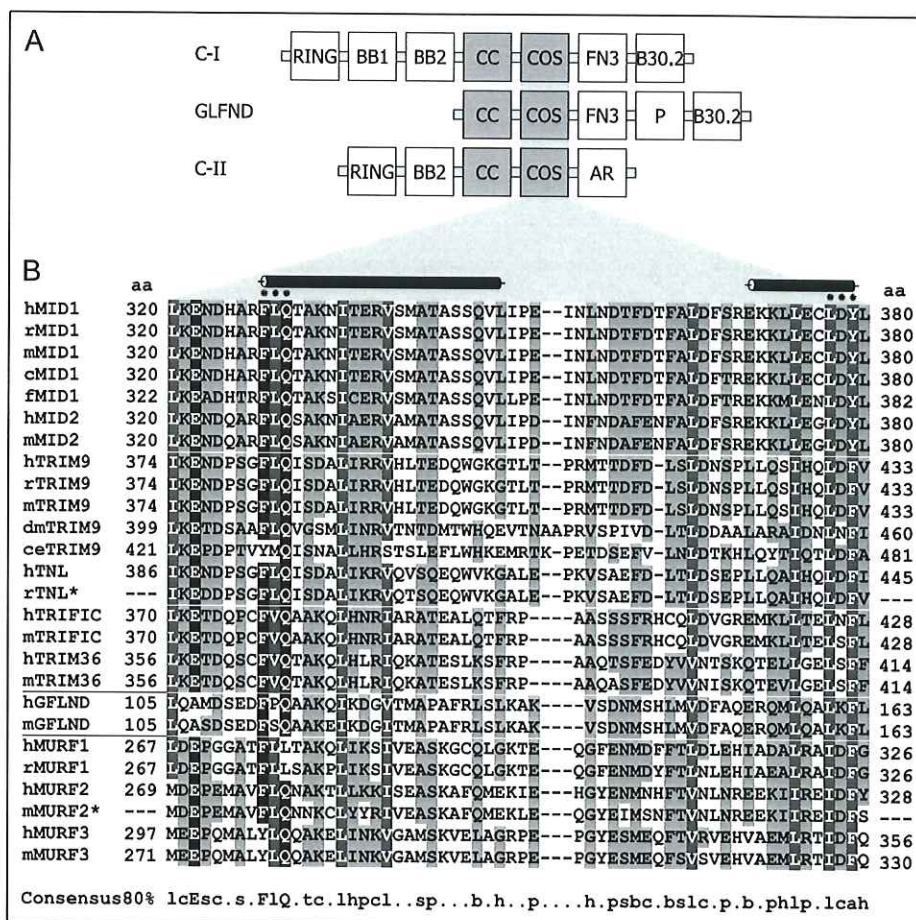


FIGURE 3. HMMER analysis of the RBCC superfamily using a C-I subfamily window data base. **A**, shown are the results from HMMER analysis of human full-length RBCC proteins, which assisted delineation into nine RBCC subgroups. The HMM data base utilized 50-amino acid windows from the established alignment (supplemental Fig. 1). The scores (in color) indicating a significantly matching region (*E*-values in black) are shifted relative to the sequence due to the window size and similarity of surrounding residues (21). The position relating to the COS box in the C-I subfamily (highlighted on the window axis) is shown; the COS box region scored higher than almost all previously established motifs for the C-I subfamily (represented here by TRIFIC and TNL) and is seen as a large peak between windows 310 and 360 at the C-terminal end of the coiled coil (CC). Likewise, C-terminal peaks at the end of the coiled coil of the MURF proteins and TRIM42 also produced significant, albeit weaker scores for COS box sequences. It should be noted that, due to the progressive windows of the C-I subfamily MSA used to build the HMM data base, scores for non-C-I subfamily members where sequence identity is lower are weighted toward the origin of higher identity. *RF*, RING finger. **B**, Ensembl Version 11 was used to annotate the human gene structure homologous to a partial rat clone encoding a TRIM9-like (TNL) sequence that was identified by a COS box HMM search of the GenBank™ non-redundant protein sequence data base. We predict that the human *TNL* gene (GenBank™ accession number AY253917) consists of 12 exons, spanning ~53 kb of chromosome 1q42, and encodes a previously undiscovered RBCC/COS/FN3/B30.2 protein.

A New Subfamily of Microtubule-associated RBCC Proteins

FIGURE 4. The COS box motif. The domains of the known microtubule-associated proteins with COS box homology are represented (A), and an alignment of the peptide sequences of representative proteins was performed using the following sequences (B): MID1 (human (*h*), rat (*r*), mouse (*m*), chicken (*c*), and fugu (*f*)), MID2 (human and mouse), TRIM9 (human, rat, mouse, *D. melanogaster* (*dm*), and *C. elegans* (*ce*)), TNL (human and predicted rat), TRIFIC (human and mouse), TRIM36 (human and mouse), GFLND (human and mouse), MURF1 (human and rat), MURF2 (human and predicted mouse), and MURF3 (human and mouse). Asterisks represent regions that are mutated as shown in Fig. 6. Cylinders above the sequences represent the approximate consensus positions of JPred/PSIPRED-predicted α -helical coils across the COS boxes of the proteins. aa, amino acids; h, hydrophobic; p, polar; a, aromatic; l, aliphatic; s, small; b, big; c, charged. The accession numbers for other sequences used here are available: human *MID1*, AF269101 and AAG33130; mouse *Mid1*, NM_010797 and NP_034927; chicken *Mid1*, AF269102 and AAK58598; rat *Mid1*, NM_022927 and NP_075216; human *MID2*, AF196481 and AAF07341; mouse *Mid2*, NM_011845 and NP_035975; human *TRIM9*, AF220037 and AAG53491; mouse *Trim9*, NM_053167 and NP_444397; rat *Trim9*, AF350422 and AAL27988; rat *Tnl*, XM_226563 and XP_226563; human *TRIM36*, NM_018700 and NP_061170; human *MURF1*, NM_032588 and NP_115977; rat *Murf1*, NM_080903 and NP_543179; human *MURF2*, AJ431704 and CAD24432; mouse *Murf2*, XM_355438 and XP_355438; human *MURF3*, AJ291714 and CAC32842; mouse *Murf3*, AF294790 and AAG03076; human *GFLND*, AY032617 and AAK51145; and mouse *Glnf*, NM_183178 and NP_899001. The fugu *MID1* sequence was obtained from A. Ashworth (personal communication).



fyng the level of similarity between every progressive window against each sequence from the RBCC superfamily could be plotted in a fashion similar to that used by Truong and Ikura (21). This approach also revealed a greater level of sequence similarity (and therefore likely evolutionary relatedness) between various RBCC members that would not have been discernible using phylogenetic algorithms on similarly large numbers of highly variant sequences. Together with domain data base and coil prediction programs, these analyses have permitted delineation of nine subfamilies based on their C-terminal domain composition (called the C-I through C-IX subfamilies) (Figs. 1 and 3). Notably, this analysis categorizes the MURF (muscle-specific RING finger) proteins (MURF1–3) as a distinct subfamily of RBCC/TRIM proteins, a grouping suggested previously (33), but not identified in the original TRIM classification (2).

As expected for a member of the first RBCC subfamily (C-I), the TRIFIC sequence scored significantly across each known motif (Fig. 3A). However, it was unexpected to find the second highest score (79.3) and *E*-value (9.4×10^{-22}) for a region of 67 amino acids extending from the end of the TRIFIC coiled coil bridging through to the beginning of the FN3 domain (Fig. 3A, large peak at the end of the coiled coil). Alignment of this region from the C-I subfamily revealed a high level of sequence conservation, comparable with the level among known motifs discovered by domain search engines (Fig. 4B). Members of the C-II subfamily, consisting of MURF1–3, revealed scores above cutoff levels (Fig. 3A) over the same region (scores of 15.9–18.8 and *E*-values of 0.015–0.0031, corresponding to amino acids 255–311, 254–310, and 256–312 within the proteins, respectively), as did TRIM42 from the C-III subfamily (score of 20.4 and *E*-value of 0.00075, corresponding to

amino acids 435–491). The weighted nature of the HMM analysis resulted in peaks with high scores and significance being positioned at the end of the coiled coil in the MURF proteins and TRIM42 and even a higher peak in TRIFIC, corresponding to sequence identity C-terminal to the coiled coil. Like MID1 and MID2 from the C-I subfamily, the MURF proteins have an ability to form both homo- and heterodimers with each other and to associate with the microtubule cytoskeleton (33, 34). Wider heterodimerization between these subfamilies has not been investigated but is unlikely in light of the findings presented in this study. Although loss or mutation of the MID1/MID2 B30.2-like domain is associated with loss of cytoskeletal binding (4, 8, 12, 35), the structural and localization similarities with the MURF proteins (C-II subfamily), which do not possess the B30.2-like domain, suggest that another region of the protein is directly responsible for microtubule binding and that the impact of mutation in the B30.2-like domain is indirect. In this regard, partitioning of the coiled coil from the rest of the RBCC domain in members of these subfamilies has shown that it is an essential domain required for both dimerization and cytoskeletal association (5, 34). Reanalysis of all C-I subfamily members using the MultiCoil program (30) aided delineation of the coiled coil from the conserved 67-amino acid region adjacent to the coiled coil. We have called this novel signature the COS (C-terminal subgroup one signature) box because we first identified it in the C-I subfamily. Notably, the MURF proteins and TRIM42 also show some domain similarity to the C-I subfamily across the N-terminal RBCC domain (Fig. 1B), but have little to no similarity C-terminal to the 67-amino acid signature. TRIM42 is the only other RBCC protein that we also identified as possessing an FN3 domain.

A New Subfamily of Microtubule-associated RBCC Proteins

However, TRIM42 has no C-terminal B30.2-like domain, a novel cysteine-rich motif N-terminal to the RBCC domain, an expanded region between the coiled-coil and FN3 domains (Fig. 1B), and also has low overall sequence similarity to the C-I subfamily. We have detected mouse (GenBankTM accession number NP_084495), rat (accession number AAH87151), and predicted chicken (accession number XP_422632) orthologs of human TRIM42, indicating that this indeed represents a *bona fide* unique and separate member of the RBCC superfamily.

A GenBankTM-wide Scan for Human COS Box-containing Proteins—To further investigate the specificity of the 67-amino acid COS box, we generated an HMM data base of these sequences from the five human C-I subfamily members and available orthologs (based on the MSA depicted in Fig. 4B), and this was scanned against the GenBankTM non-redundant protein sequence data base containing all entries from GenPept, Swiss-Prot, Protein Information Resource, PDF, Protein Data Bank, and RefSeq (<ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz>) using HMMpfam (22). As expected, all previously identified members of the C-I, C-II, and C-III subfamilies produced significant high scores. In addition, this analysis identified an incomplete rat sequence highly related to, but distinct from, TRIM9 (score of 108.5 and *E*-value of 1.3×10^{-28}) as well as a recently deposited Trim9 sequence from the mosquito *Anopheles gambiae* (score of 121.6 and *E*-value of 1.5×10^{-32}). Although no human sequence orthologous to the novel rat sequence was found in GenBankTM, a search of Ensemble Version 11 identified a predicted human transcript with high identity. Based on subsequent alignments with and homology to the C-I subfamily, we have re-annotated this sequence and called this gene *TNL* (TRIM nine-like; GenBankTM accession number AY253917) (Fig. 3B). Strikingly, the predicted human TNL protein shares high identity with human TRIM9 across its entire length (76% similarity and 65% identity) and as such represents an additional member of the C-I subfamily. Of interest, an additional sequence gave a lower but above cutoff score in the HMM analysis: GLFND (score of 14.9 and *E*-value of 0.0035, corresponding to amino acids 105–162). The microtubule-associated protein GLFND shares C-terminal similarity with the C-I subfamily (including coiled-coil, COS box, FN3, and B30.2-like domains), but lacks RING and B-box motifs, so is not an RBCC/TRIM protein (36–38). An HMM data base generated from the alignment of all available COS box sequences was used to perform an additional search for COS box-containing proteins in the most current human GenBankTM peptide sequence data base using the current NCBI 35 assembly available from Ensembl. This search detected all known human COS box-containing proteins, although no additional high scoring matches were revealed.

Subcellular Localization of the C-I Subfamily Members—Early analysis of MID1 and MID2 showed that both proteins associate with the microtubule network and have the ability to form homo- and heteromultimers (5, 8, 9, 35). To investigate the subcellular distribution of the remaining C-I subfamily members, we transiently transfected either GFP- or Myc-tagged expression constructs of each available full-length protein (see “Experimental Procedures”). The localization patterns of GFP-TRIM36, GFP-Trific, and Myc-TRIM9 in COS-1 cells resembled those of GFP-MID1 and GFP-MID2 observed in simultaneous experiments, and this was confirmed by immunohistological detection of α -tubulin. These data indicate that all C-I subfamily proteins associate with the microtubule cytoskeleton (Fig. 5). Myc-TRIM9 was used because N-terminally GFP-tagged TRIM9 exhibited cellular localization similar to that of GFP alone when compared with immunohistochemical detection (12), indicating a disruption of normal localization (data not shown). Similar disruption of MID1 and MID2 has been seen with some tags such as FLAG.⁴ Although a full-length TNL sequence

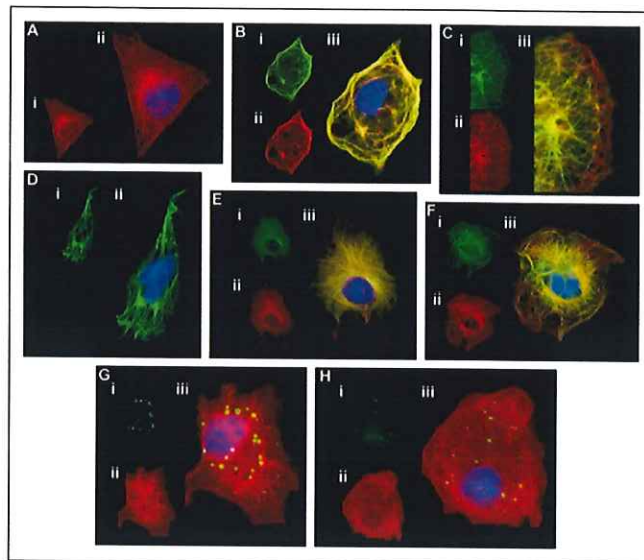


FIGURE 5. Transient expression of tagged wild-type and mutant C-I subfamily members in COS-1 cells. All proteins were expressed as N-terminally GFP-tagged fusions, with the exception of TRIM9 (D), for which an N-terminal Myc tag was used. The α -tubulin microtubule staining (A-i, B-ii, C-iii, E-ii, and F-iii) is congruent with the distribution of each of the tagged C-I subfamily members, GFP-TRIM36 (E), GFP-Trific (F), and Myc-TRIM9 (D) (B-i-F-i), like the previously reported GFP-MID1 (B) and GFP-MID2 (C) proteins (B-iii, C-iii, E-iii, and F-iii). C-terminal (B30.2) mutations in TRIM9 (GFP-TRIM9) (H-i), like a similarly mutated MID1 (GFP-MID1) (G-i), lose their ability to tether to microtubules (G-ii and H-ii, α -tubulin immunofluorescence). G-iii and H-iii are merged images of G-i and G-ii and H-i and H-ii, respectively. Texas Red Myc-TRIM9 detection (D) was artificially colored green to avoid confusion with microtubule staining.

was not available, given the extensive sequence similarity to TRIM9, it too would be expected to have a microtubule distribution in line with that reported here for other members of the C-I subfamily. Of note, the original *TRIM9* EST harboring the L653F missense change (see “Experimental Procedures”) in the B30.2 domain localized in cytoplasmic clumps (Fig. 5H). This localization is reminiscent of MID1 B30.2 domain mutants (Fig. 5G), which are associated with Opitz syndrome, and suggests that other C-I subfamily members associate with microtubules in a manner that is similarly disrupted by B30.2 domain mutation.

Functional Analysis of the C-I Subfamily: Protein-Protein Interaction—In addition to their ability to homo- and heteromultimerize, microtubule-bound MID1 and MID2 also bind Alpha 4, a negative regulatory subunit of protein phosphatase 2A (5). Given the similar domain architecture and degree of sequence identity over specific domains such as the B-boxes, we chose to use the ProQuest yeast two-hybrid system to determine whether other members of this subfamily also show an ability to homo- and heteromultimerize and to bind Alpha 4. The C-I subfamily members TRIM9 and TRIM36 both showed strong yeast two-hybrid interaction for homodimerization (Table 1). However, no evidence of heterodimerization or interaction with Alpha 4 could be found for any of the C-I subfamily members except MID1 and MID2 (Table 1). Interestingly, Trific was the only C-I subfamily member that did not show any self-interaction, although we cannot exclude the possibility that the particular N-terminal fusion interferes specifically with Trific protein folding and or dimerization capacity in this system, similar to that mentioned previously with the N-terminal GFP fusion to TRIM9. Although not tested, TNL is predicted to be able to both homo- and heterodimerize with TRIM9 because their level of amino acid similarity, especially across their coiled-coil domains (91% similarity and 78% identity), is comparable with that of MID1 and MID2 (83% similarity and 72% identity). The results indicate that het-

⁴ K. M. Short and T. C. Cox, unpublished data.

TABLE 1
Protein-protein interaction between the C-I subfamily and associated proteins

Yeast two-hybrid analysis showed that heterodimerization was restricted to MID1 and MID2. Homodimerization was shown for all tested C-I members in both yeast two-hybrid orientations, except Trific (see "Results"). Two independent reporters were used to confirm interactions: *HIS3* on 3-aminotriazole (first marker) and *lacZ* analyzed by an X-gal assay (second marker). DBD, DNA-binding domain; AD, activation domain; ++, positive interaction; -, no interaction.

	Gal4 DBD	MID1	MID2	TRIM9	TRIM36	Trific	Alpha 4
Gal4 AD	-	-	-	-	-	-	-
MID1	-	++	++	-	-	-	++
MID2	-	++	++	-	-	-	++
TRIM9	-	-	-	++	-	-	-
TRIM36	-	-	-	-	++	-	-
Trific	-	-	-	-	-	-	-
Alpha 4	-	++	++	-	-	-	-

erodimerization within the C-I subfamily is restricted to those members that, through more recent gene duplication events, have enough similarity within the respective coiled-coil regions.

Analysis of a Potential Microtubule-associated Function for the COS Box—Homology to the COS box is present in the microtubule-associated MID1 and MID2 proteins (C-I subfamily), MURF1-3 (C-II), and GLFND (10, 39). Although these proteins show regions of similarity to each other, the coiled coil and COS box are the only motifs shared by all (Fig. 4A). We thus hypothesized that the new 67-amino acid COS box may, at least in part, play a role in the cytoskeletal association or, at a minimum, be a signature for microtubule-associated RBCC proteins. In this respect, microtubule association of several associated C-I and C-II proteins has been shown to be dependent on dimerization mediated by the coiled coil (2, 5, 33). In nearly all studies of homodimerization, the coiled coil has been expressed to capture full-length proteins *in vitro* or in yeast two-hybrid assays; yet, in all instances, we have revealed that part or all of the COS box has also been removed. This indicates that dimerization in the absence of an intact COS box is still strong (5, 34) and suggests that the COS box is not necessary for dimerization.

However, a number of studies have not disrupted COS boxes. Spencer *et al.* (34) generated a C-terminal truncation that removed 31 amino acids from the C terminus of MURF3, including the acid-rich domain. The mutation (amino acids 354–384 deleted) did not significantly impact on the predicted COS box (amino acids 298–354) or coiled coil (positioned N-terminal to the COS box), and it was noted that both microtubule association and homodimerization were not significantly affected by this deletion. However, removal of both the coiled coil and predicted COS box with a 154-amino acid C-terminal deletion of MURF3 abolished microtubule association and homodimerization (34). Similar studies of microtubule association in GLFND have indicated that removal of N-terminal portions of the protein (including the coiled coil, but with COS boxes intact) permits microtubule association, albeit with a lower affinity and of a whorl-like appearance (36, 37). Furthermore, GLFND without the coiled coil loses the ability to stabilize microtubules after treatment with nocodazole, which is thought to be the result of a loss of dimerization ability (36). Additionally, in these studies, fortuitous removal of regions containing the COS box motif severely inhibited the ability of the protein to associate with microtubules (36, 37). Therefore, to test our hypothesis, COS box mutations were introduced into the well characterized MID1 protein, and cellular localization patterns and homodimerization were tested in the presence of these mutations. To minimize effects on tertiary structure, two individual point mutations were generated in the COS box that replaced conserved amino acid triplets (FLQ and LDY) (highlighted in Fig. 4B and supplemental Fig. 1) with alanines. Replacement with triple alanine residues was anticipated to mildly disturb the local folding of the motif by shortening the predicted α -helical

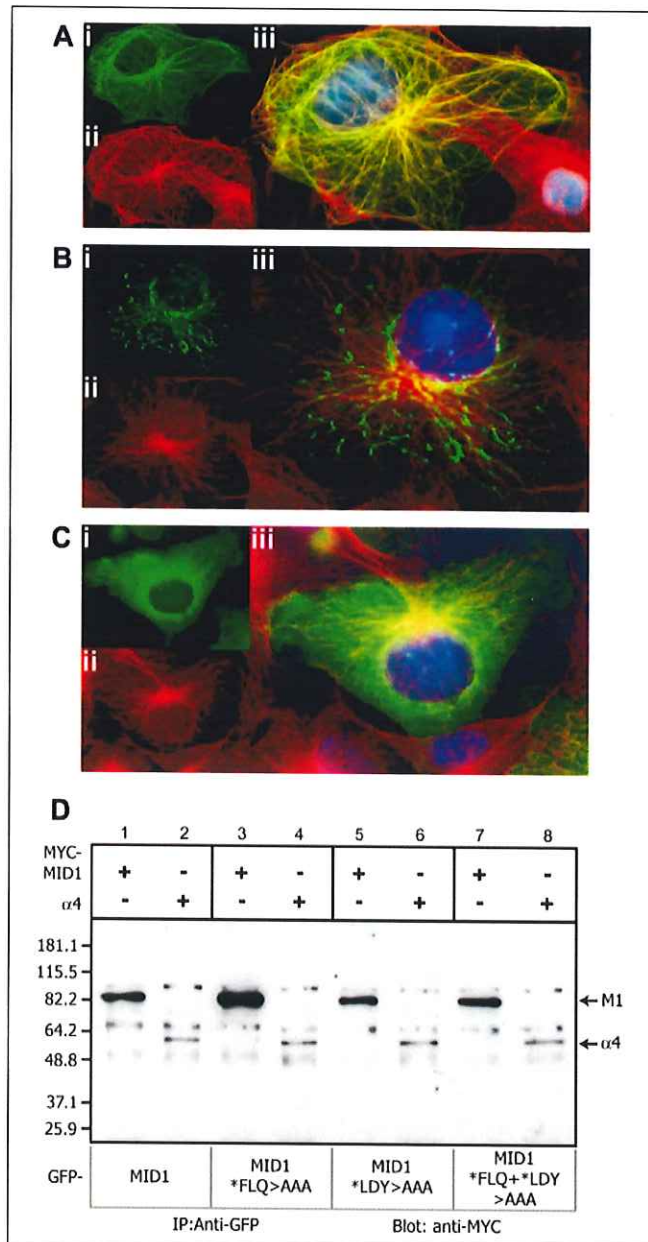


FIGURE 6. Mutations within the COS box of MID1 affect microtubule localization, but not dimerization. Replacement of residues FLQ (A), LDY (B), and both FLQ and LDY (C) with AAA residues in the COS box affected the localization of GFP-MID1 (A–C) from a normal microtubule distribution (A–ii–C–ii) to one that was progressively disrupted (A–iii–C–iii show the overlay, with Hoechst nuclear staining in blue). Co-immunoprecipitation (IP; D) revealed that the same mutations (FLQ → AAA (lanes 3 and 4), LDY → AAA (lanes 5 and 6), and both FLQ → AAA and LDY → AAA (lanes 7 and 8)) did not affect self-association (lanes 1, 3, 5, and 7) or interaction with Alpha 4 ($\alpha 4$; lanes 2, 4, 6, and 8).

coil (data not shown). Mutation of FLQ to AAA had little effect on cellular localization patterns, with prominent microtubule association seen (Fig. 6A). In contrast, the LDY → AAA mutation resulted in a punctate speckled cytoplasmic appearance, with some short "tracks" of microtubule staining observed in a small percentage of cells (Fig. 6B). Both mutations were then introduced into MID1, and the resultant protein completely lost its ability to interact with microtubules (Fig. 6C). Consistent with predictions, the double mutant was still able to efficiently dimerize and also to bind Alpha 4 (Fig. 6D), indicating that the tertiary structure was not grossly disrupted.

A New Subfamily of Microtubule-associated RBCC Proteins

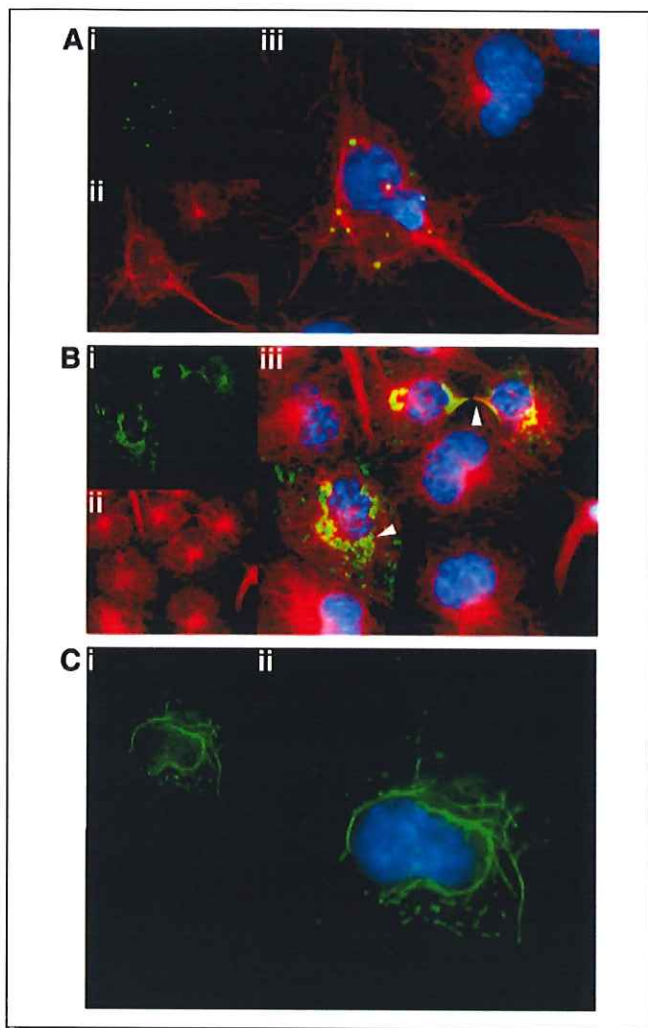


FIGURE 7. Localization of the non-microtubule-associated TRIM37 protein is altered by the presence of a COS box. Shown is the expression of GFP-tagged fusions of the TRIM37 RBCC domain with the MID1 COS box (A) or the COS box plus the MID1 C-terminal domains (B and C). Transiently transfected GFP fusions (A-i-C-i) in fixed and anti- α -tubulin-stained COS-1 cells (A-ii and B-ii) indicate a punctate localization when overlaid (A-iii and B-iii, with Hoechst nuclear stain in blue), although some minimal microtubule association can also be observed in fixed cells (B-iii, upper arrowhead), which improves in unfixed cells (C-ii). Aggresomal localization can also be observed (lower left of B-i and B-iii, lower arrowhead), which has been reported previously for the overexpressed TRIM37 RBCC domain and which is thought to be intrinsic to the function of the protein (42).

These data support a specific role for the COS box in directing microtubule association.

Modular Role for COS Box-assisted Microtubule Association—To test whether the COS box is sufficient to confer microtubule association properties to RBCC/TRIM proteins normally not associated with microtubules, chimeric proteins were generated with the RBCC domain (amino acids 1–252) from TRIM37/MUL and a COS box (Fig. 7A) or a COS box with the C-terminal domain from MID1 (Fig. 7B and C). Endogenous TRIM37 has been observed to natively localize to peroxisomes (40), with overexpressed full-length TRIM37 or the TRIM domain found in cytosolic bodies and aggresomes (41, 42). Cellular localization by direct fluorescence revealed that the fusion of the COS box alone to the TRIM domain of TRIM37 redirected the localization to a small number of large punctate speckles within the cytoplasm (Fig. 7A). The disruption of cytoskeletal organization, which is similar to the effect seen with overexpressed full-length TRIM37 (42), and the low

viability of transfected cells indicated that the C-terminal fusion was harmful to cells (data not shown). However, partial microtubule overlap could be seen in fixed cells (Fig. 7B, upper arrowhead), with more definitive microtubule localization observed in living cells (where microtubules could not be counterstained) (Fig. 7C). In addition, some perinuclear aggresomal localization could also be noted (Fig. 7B, lower arrowhead), which is similar to the overexpressed full-length protein (data not shown) (42). Assessment of dimerization by co-immunoprecipitation revealed that dimerization did not occur between both the fusions and either full-length MID1 or the RBCC domain of TRIM37 (data not shown), consistent with this property being conferred by the coiled-coil domain. The inability of the chimeric TRIM-COS fusion to associate with microtubules may be explained by the observation that, in all COS box-containing proteins, domains C-terminal to the COS box are present, and these may assist in folding, placement, or protection of a microtubule-interfacing structure.

DISCUSSION

In a recent large study, Reymond *et al.* (2) identified numerous new RBCC proteins encoded in the human genome. Based on initial findings using N-terminal GFP tags of some of these factors, these researchers suggested that this protein family identifies a unique array of subcellular compartments, although few structures were identified. Here, we sought to determine the relationship of all human RBCC proteins using a bioinformatics approach with a view to better understand the functions of this important family. Underpinning this approach was our previous observation of an FN3 domain in the Opitz syndrome protein, MID1, and its homolog, MID2 (8), and the subsequent finding of the same domain architecture in an additional RBCC protein, TRIM9 (12). Like the microtubule-associated MID1 and MID2 proteins, TRIM9 had been reported to associate with the cytoskeleton, yet the actual cytoskeletal elements with which it associates had not been determined (12). We therefore surmised that the domain architecture might reflect or determine a common subcellular localization with perhaps even similarities in function. As the presence of FN3 domains was not noted in the structural representations of the various RBCC/TRIM proteins described by Reymond *et al.* (2), we first sought to determine whether any additional RBCC family members possess an FN3 domain. Using both available and newly developed bioinformatics resources, including PSI-BLAST-, PROFILE-, and HMMER-based analyses, we undertook a sensitive screen of both the known RBCC family members and the translated current releases of the human genomic sequence. This approach uncovered three additional RBCC proteins that share the same domain organization as MID1, MID2, and TRIM9. Two of these represent products of previously unidentified genes that we have called TRIFIC and TNL. The third, TRIM36, had previously been reported by Reymond *et al.* (2), but was represented as only possessing an N-terminal RBCC/TRIM domain. Validating our findings, two other groups identified and reported TRIM36 during the course of this study and also determined that it possesses a MID1/MID2-like domain architecture that includes both C-terminal FN3 and B30.2-like domains (14, 31).

Significantly, we have also shown in this study that the five tested RBCC proteins that share the same domain architecture all localize to the microtubule cytoskeleton, representing the first report of a family of structurally related B30.2-like proteins functioning in the same subcellular compartment. In addition, we have shown that most of the members form dimers with the exception of murine Trifric, although this may be due to issues with the system used (as described). Based on the findings of the structural and functional similarities to these homologous proteins, we re-evaluated the domain architecture of all remaining RBCC proteins and performed similar groupings on the basis of their

A New Subfamily of Microtubule-associated RBCC Proteins

C-terminal domain composition. This classification revealed nine subgroups with distinct C-terminal architectures: COS/FN3/B30.2-like domains (C-I subfamily), COS/acid-rich region (C-II), COS/FN3 domains (C-III), B30.2-like or SPRY-containing domains (C-IV), no identifiable C-terminal domain homology (C-V), PHD/BROMO domains (C-VI), immunoglobulin/NHL repeats (C-VII), MATH domain (C-VIII), and ADP-ribosylation factor domain (C-IX). Representative proteins of six of the nine subgroups have been identified in *D. melanogaster* and *Caenorhabditis elegans* (data not shown), with exceptions being the C-II, C-III, and C-VIII subfamilies. In the latter, which is defined by a C-terminal MATH domain, the domain is nevertheless present and found C-terminal to RING and coiled-coil motifs (e.g. *C. elegans* TRF1). Interestingly, there is no evidence from EST and genomic sequences to date to support the existence of the RBCC tripartite domain in fungi and plants despite the presence of RING, B-box, and coiled-coil motifs individually in many proteins. This also indicates that rearrangement of the separate domains into a tripartite cluster may have occurred only during or after metazoan evolution in animals. Consistent with this, TRIM9 is the only invertebrate protein with the C-I (RBCC/FN3/B30.2) motif architecture, and as such, the likely role of this protein was perhaps not necessary until multicellularization in animals. It is therefore likely that these model organisms will be useful for investigating basic functions of the different RBCC subgroups that, in turn, might aid our understanding of various human disease processes in which RBCC proteins have been implicated.

We used a HMMER-based approach in part of the human RBCC superfamily analysis. Graphical depiction of the data generated from 50-amino acid wide sliding windows of the partial C-I subfamily MSA clearly highlights the similarities between the members of the C-I subfamily and the other subfamilies. Truong and Ikura (21) have previously highlighted the significance of high scoring HMM matches from a sliding window data base to regions of structural identity within the cadherin superfamily of proteins, linking statistical analysis findings to actual structural identities. A similar unbiased approach was used in this study, highlighting areas of significant conservation within the subfamily as well as throughout the RBCC superfamily. The analysis resolved highly significant matches for TRIFIC and TNL throughout the RING finger, B-boxes, and coiled-coil regions, but also aided the definition of a high scoring region immediately adjacent to the C-terminal end of the coiled coil in a number of the RBCC subfamilies. We have called this region the COS box, and it is a region consisting of two α -helical coils (Fig. 4). To determine whether this motif supports more than just conserved positioning of the neighboring coiled-coil and FN3 domains in the C-I proteins, an HMM data base representing the COS box was used to scan the GenBank™ FASTA Protein Database. This scan identified COS boxes in the non-FN3 domain-containing MURF proteins and in the non-TRIM domain-containing protein GLFND, although all these proteins also have in common a coiled-coil domain. Previous analyses suggested that the coiled-coil domains of the RBCC proteins (MID1, MID2 and MURF3) are required for dimerization and microtubule association (5, 34). Inadvertently, in these studies, the regions encoding the COS boxes were also retained when attempts were made to isolate coiled-coil function. The other COS box-containing protein, GLFND, has also been found to associate with microtubules and has a C terminus similar to that of the C-I subfamily with coiled-coil, FN3, and B30.2 domains (36), yet it does not possess a RING or B-box. We therefore considered whether the COS motif simply provides an adjunct to the coiled-coil domain or whether it has a more direct functional activity. A role for the coiled-coil/COS box region in dimerization and microtubule association is supported by the observation that mutations in

MURF3 that remove amino acids either N-terminal to the coiled coil or C-terminal to the COS box maintain their microtubule association and dimerization properties (34). However, the presence of coiled-coil domains throughout the generally non-microtubule RBCC superfamily would suggest that a coiled coil alone is not enough for microtubule association and gives support to the notion that the COS box confers microtubule binding capacity assisted by the coiled coil. To investigate this, we introduced mutations (FLQ \rightarrow AAA and LDY \rightarrow AAA) into two conserved regions of the COS box in MID1 (or consensus sites *FlQ* and *lca* in Fig. 4). Whereas the individual mutations alone had relatively moderate effects on microtubule localization of the protein (Fig. 6, A and B), when introduced together, both mutations completely abolished microtubule association (Fig. 6C). Notably, the mutations did not affect coiled-coil function because dimerization with wild-type MID1 was not affected (Fig. 6D). Furthermore, these mutants were still capable of interacting with Alpha 4 (which binds B-box 1), indicating that the overall RBCC tertiary structure was likely to be intact. We therefore wanted to investigate whether the addition of a COS box to a non-microtubule-localized RBCC protein would direct the chimeric protein to the cytoskeleton. Fusion of the COS box alone to the RBCC/TRIM domain from the Mulibrey nanism protein, TRIM37, altered the localization of overexpressed TRIM37 to cytoplasmic clumps. Notably, however, fusion of a larger COS box C-terminal fragment permitted the chimeric protein to associate with cytoskeletal structures (Fig. 7C). These data indicate that the COS box, which is found only in the C-I, C-II, and C-III RBCC/TRIM proteins and GLFND, acts independently of the other general dimerization properties of the coiled coil to mediate direction of these proteins to the microtubule cytoskeleton. However, the association of the coiled coil with the COS box in all these proteins (including the chimeric TRIM37-COS protein) indicates a potential need of these two domains together for normal microtubule association and that these proteins all bind microtubules through a similar mechanism. Our bioinformatics-based approach used to subclassify the entire human RBCC complement (based on C-terminal domain composition) may thus assist in understanding the compartmentalization of the various RBCC subfamilies and in turn may better reflect the evolutionary and functional relatedness of the diverse members of this large family of proteins.

Acknowledgments—We thank Belinda Thomas for invaluable assistance with the TRIM9 EST mutation correction, Kevin Truong for general assistance with the HMMER package and for providing Perl scripts to aid data processing, Dan Kortschak for assistance with debugging Perl scripts for use with Perl Version 5.8 on Darwin, Alan Ashworth for the contribution of Takifugu rubripes MID1 sequence, and Philipp Bucher for assistance with the PROSITE PROFILE to HMMER Database conversion.

REFERENCES

1. Henry, J., Mather, I. H., McDermott, M. F., and Pontarotti, P. (1998) *Mol. Biol. Evol.* 15, 1696–1705
2. Raymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzi, L., Riganelli, D., Zanaria, E., Messali, S., Cainarca, S., Guffanti, A., Minucci, S., Pelicci, P. G., and Ballabio, A. (2001) *EMBO J.* 20, 2140–2151
3. Jensen, K., Shiels, C., and Freemont, P. S. (2001) *Oncogene* 20, 7223–7233
4. Cox, T. C., Allen, L. R., Cox, L. L., Hopwood, B., Goodwin, B., Haan, E., and Suthers, G. K. (2000) *Hum. Mol. Genet.* 9, 2553–2562
5. Short, K. M., Hopwood, B., Yi, Z., and Cox, T. C. (2002) *BMC Cell Biol.* 3, 1
6. Rhodes, D. A., Ihrke, G., Reinicke, A. T., Malcherek, G., Towey, M., Isenberg, D. A., and Trowsdale, J. (2002) *Immunology* 106, 246–256
7. Stremlau, M., Owens, C. M., Perron, M. J., Kiessling, M., Autissier, P., Sodroski, J., and Goff, S. P. (2004) *Nature* 427, 848–853
8. Perry, J., Short, K. M., Romer, J. T., Swift, S., Cox, T. C., and Ashworth, A. (1999) *Genomics* 62, 385–394

A New Subfamily of Microtubule-associated RBCC Proteins

9. Cainarca, S., Messali, S., Ballabio, A., and Meroni, G. (1999) *Hum. Mol. Genet.* **8**, 1387–1396
10. Trockenbacher, A., Suckow, V., Foerster, J., Winter, J., Krauss, S., Ropers, H. H., Schneider, R., and Schweiger, S. (2001) *Nat. Genet.* **29**, 287–294
11. Quaderi, N. A., Schweiger, S., Gaudenz, K., Franco, B., Rugarli, E. I., Berger, W., Feldman, G. J., Volta, M., Andolfi, G., Gilgenkrantz, S., Marion, R. W., Hennekam, R. C., Opitz, J. M., Muenke, M., Ropers, H. H., and Ballabio, A. (1997) *Nat. Genet.* **17**, 285–291
12. Li, Y., Chin, L. S., Weigel, C., and Li, L. (2001) *J. Biol. Chem.* **276**, 40824–40833
13. Berti, C., Messali, S., Ballabio, A., Reymond, A., and Meroni, G. (2002) *Mech. Dev.* **113**, 159–162
14. Kitamura, K., Tanaka, H., and Nishimune, Y. (2003) *J. Biol. Chem.* **278**, 44417–44423
15. Goodstadt, L., and Ponting, C. P. (2001) *Bioinformatics* **17**, 845–846
16. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998) *Trends Biochem. Sci.* **23**, 403–405
17. Henikoff, S., and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10915–10919
18. Benner, S. A., Cohen, M. A., and Gonnet, G. H. (1994) *Protein Eng.* **7**, 1323–1332
19. Notredame, C., Higgins, D. G., and Heringa, J. (2000) *J. Mol. Biol.* **302**, 205–217
20. Felsenstein, J. (2004) *PHYMLIP Version 3.62*, Department of Genome Sciences, University of Washington, Seattle, WA
21. Truong, K., and Ikura, M. (2002) *BMC Bioinformatics* **3**, 1
22. Eddy, S. R. (1998) *Bioinformatics* **14**, 755–763
23. Elble, R. (1992) *BioTechniques* **13**, 18–20
24. Myers, A. M., Tzagoloff, A., Kinney, D. M., and Lusty, C. J. (1986) *Gene (Amst.)* **45**, 299–310
25. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R. R., Courcelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-Jones, S., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lopez, R., Letunic, I., Lonsdale, D., Silventoinen, V., Orchard, S. E., Pagni, M., Peyruc, D., Ponting, C. P., Selengut, J. D., Servant, F., Sigrist, C. J., Vaughan, R., and Zdobnov, E. M. (2003) *Nucleic Acids Res.* **31**, 315–318
26. Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Doerks, T., Schultz, J., Ponting, C. P., and Bork, P. (2004) *Nucleic Acids Res.* **32**, D142–D144
27. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) *Nucleic Acids Res.* **32**, D138–D141
28. Hulo, N., Sigrist, C. J., Le Saux, V., Langendijk-Genevaux, P. S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P., and Bairoch, A. (2004) *Nucleic Acids Res.* **32**, D134–D137
29. Attwood, T. K., Bradley, P., Flower, D. R., Gaulton, A., Maudling, N., Mitchell, A. L., Moulton, G., Nordle, A., Paine, K., Taylor, P., Uddin, A., and Zygouri, C. (2003) *Nucleic Acids Res.* **31**, 400–402
30. Wolf, E., Kim, P. S., and Berger, B. (1997) *Protein Sci.* **6**, 1179–1189
31. Balint, I., Muller, A., Nagy, A., and Kovacs, G. (2004) *Gene (Amst.)* **332**, 45–50
32. Bucher, P., Karplus, K., Moeri, N., and Hofmann, K. (1996) *Comput. Chem.* **20**, 3–23
33. Centner, T., Yano, J., Kimura, E., McElhinny, A. S., Pelin, K., Witt, C. C., Bang, M. L., Trombitas, K., Granzier, H., Gregorio, C. C., Sorimachi, H., and Labeit, S. (2001) *J. Mol. Biol.* **306**, 717–726
34. Spencer, J. A., Eliazer, S., Ilaria, R. L., Jr., Richardson, J. A., and Olson, E. N. (2000) *J. Cell Biol.* **150**, 771–784
35. Schweiger, S., Foerster, J., Lehmann, T., Suckow, V., Muller, Y. A., Walter, G., Davies, T., Porter, H., van Bokhoven, H., Lunt, P. W., Traub, P., and Ropers, H. H. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 2794–2799
36. Manabe, R., Whitmore, L., Weiss, J. M., and Horwitz, A. R. (2002) *Curr. Biol.* **12**, 1946–1951
37. Stein, P. A., Toret, C. P., Salic, A. N., Rolls, M. M., and Rapoport, T. A. (2002) *J. Cell Sci.* **115**, 3389–3402
38. Carim-Todd, L., Escarceller, M., Estivill, X., and Sumoy, L. (2001) *Biochim. Biophys. Acta* **1518**, 200–203
39. Witt, S. H., Granzier, H., Witt, C. C., and Labeit, S. (2005) *J. Mol. Biol.* **350**, 713–722
40. Kallijarvi, J., Avela, K., Lipsanen-Nyman, M., Ulmanen, I., and Lehesjoki, A.-E. (2002) *Am. J. Hum. Genet.* **70**, 1215–1228
41. Zapata, J. M., Pawlowski, K., Haas, E., Ware, C. F., Godzik, A., and Reed, J. C. (2001) *J. Biol. Chem.* **276**, 24242–24252
42. Kallijarvi, J., Lahtinen, U., Hamalainen, R., Lipsanen-Nyman, M., Palvimo, J. J., and Lehesjoki, A.-E. (2005) *Exp. Cell Res.* **308**, 146–155
43. Morgan, K. (ed) (2005) *Current Protocols in Cell Biology*, pp. 6.0.1–6.1.34, John Wiley & Sons, Inc., New York

Chapter Five

General Discussion and Conclusions

Chapter Five: Discussion and Conclusions

5.1 Review of Project Aims and Significance

Genetic mutation influenced by as yet unknown 'environmental' factors are the cause of the Opitz syndrome (OS) phenotype which may include hypertelorism, hypospadias, laryngo-tracheoesophageal (LTE) defects, clefting of the lip and/or primary palate, congenital heart defects and anorectal anomalies. Observations of many patients has led clinicians and researchers to believe that the nature of the genetic lesion underlying the disease has no bearing on the constellation of phenotypes presented by patients (Robin *et al.*, 1996; So *et al.*, 2005). The obvious involvement of multiple genetic loci raises interesting questions regarding the molecular pathogenesis of the disease, but the lack of knowledge of most genes involved makes it difficult to directly investigate these genetic interactions. However, the phenotypic similarity between patients with mutations originating from different genetic loci indicates that the causative genes may act in a similar pathway, and may even directly interact during development (such that a failure of any one gene results in a fundamental change in the regulation of another gene or pathway). Therefore a primary aim of this study was to use the one gene, *MID1*, known to be mutated in Opitz syndrome as a tool to find interacting proteins using a yeast two-hybrid approach, which is a genetic screen used to identify protein-protein interactions.

Prior to the completion of the human genome project, the identification of a *MID1* homologue by two groups (Buchner *et al.*, 1999; Perry *et al.*, 1999) was the first indication that a family of MID1-like proteins may exist with the potential to have an equally important function during embryological development. The primary sequence identity between MID1 and the new homologue, MID2, was suggestive of the potential for

functional redundancy because they also have similar tissue expression patterns throughout development. This is important because functional homology between MID1 and MID2 could raise the possibility that MID2 may itself have a role in the aetiology of OS. One of the classical hallmarks of MID1 cellular expression is microtubule localisation (Schweiger *et al.*, 1999), and the first suggestion of functional homology between the proteins was the similar pattern of microtubule localisation exhibited by MID2 (Buchner *et al.*, 1999; Perry *et al.*, 1999). Therefore, additional aims of the study were to: (1) identify further functional homology between MID1 and MID2 which may support the notion of MID2 involvement in OS; and (2) to take advantage of the sequence deposited by mammalian genome projects to search for more MID1 and MID2 homologues, and use sequence identity between them as a platform to study structure/function relationships. The broad approach taken to identify homologues was expected to identify a large number of proteins that possess similarity specifically through the RBCC (RING finger, B-box, coiled coil) tripartite motif, which forms the basis of the classification of a large protein superfamily, the RBCC/TRIM superfamily. Early classification of this superfamily by Reymond *et al.* indicated that this family of proteins identified particular cell compartments (Reymond *et al.*, 2001), although little rationale for this diversity in location was presented. In this thesis, I have employed a combination of bioinformatics approaches to assess the entire human RBCC/TRIM complement, subclassifying them by their vastly different C-terminal domains. In total, nine different C-terminal domain compositions and arrangements make up the superfamily, many of which are found in lower species such as *Drosophila melanogaster*. Thus, the diversity in subcellular compartmentalisation and function of the proteins have likely evolved independently and over considerable evolutionary time and allude to the diversity, at least in part, being the result of differences in the C-terminal regions of the proteins (Short and Cox, 2006). Therefore, in order to provide a better understanding of the

relationship between MID1 and MID2, my bioinformatics analysis was subsequently used to define homologues with the same C-terminal domain architecture. More broadly, the knowledge gleaned from the greater RBCC/TRIM subclassification is expected to assist the study of structure/function relationships of other homologous subgroups in a fashion similar to that used to determine the relationships of the MID1/MID2 subfamily.

5.2 Identification of the C-I subfamily and similarities in structure/function

The MID1 protein subfamily study began with an analysis of the relationships between MID1 and MID2, and further branched out to include TRIM9, TNL, TRIM36 and TRIFIC. These proteins, together known as the C-I (carboxy terminus I) subfamily, share the same domain ‘building blocks’ to form a protein family with a structural architecture comprising of RING finger, B-box 1, B-box 2, coiled coil, COS box, Fibronectin type III and B30.2-like domains. In order to determine general functional characteristics within the subfamily, the degree of conservation of both specific and broad-based protein behaviours was first identified. The identification of the MID1 and MID2 homologues TRIM9, TRIM36, TRIFIC and TNL allowed the determination of functional homology across the subfamily. Representative cDNAs for five of the six members were cloned into suitable fusion expression vectors for analysis of cell localisation and protein-protein interactions. Dimerisation was shown to be a property of nearly all C-I proteins with homodimerisation being the predominant form over heterodimerisation. While the majority of C-I proteins only homodimerise, it is thought the heterodimerisation ability shared between MID1 and MID2 is probably due to the high degree of identity between the coiled coil domains of the protein – which act in a pseudo ‘homodimerisation’ mode (Short and Cox, 2006). Although a full length TNL cDNA was not available, it was expected that it too would

heterodimerise with TRIM9 given their identity approached that seen between MID1 and MID2.

The analysis of cell localisation additionally indicated that all C-I subfamily proteins were localised to the microtubules. Since microtubule association has never been attributed to any specific domains, analysis of the similarities and differences between the C-I subfamily proteins and other RBCC proteins was used to determine the domains or sequences that might be responsible for microtubule localisation. The initial analysis of the C-I subfamily multiple sequence alignment used techniques to generate hidden Markov model (HMM) profiles of the subfamily alignment, and used these to identify regions of conservation amongst the RBCC protein superfamily. This analysis identified sequence identity outside of the regions defined as domains by online domain database tools such as SMART (Letunic *et al.*, 2004), PFAM (Bateman *et al.*, 2004), PRINTS (Attwood *et al.*, 2003), Prosite PROFILE and PATTERN analyses (Hulo *et al.*, 2004). Further analysis using HMMER hinted at a structurally significant region between the coiled coil domain and the FN3 domains of the C-I subfamily. This new motif, present in all MID1 homologues, was called the COS box and sequences for the motif from each homologue were extracted and used to generate a specific COS box profile HMM. HMMsearch, a programme used to report matches between profile HMMs and FASTA amino acid sequences (Eddy, 1998) was used to identify other COS box containing proteins by scanning the complete GenBank protein sequence database. Interestingly, most proteins identified as containing significant matches to the COS box motif had also been independently found to associate with the microtubular cytoskeleton. In all, the proteins shown to possess COS boxes and be microtubule associated are MID1, MID2, TRIM36, TRIFIC, TRIM9 (and presumably TNL), MURF1, MURF2, MURF3 and GFLND. Comparisons of the domain architecture of all COS box containing proteins indicated that

the only domains that all the proteins had in common were a coiled coil and the COS box motif. Coiled coil domains have been relatively well characterised as being a dimerisation domain (Peng *et al.*, 2002; Short *et al.*, 2002; Spencer *et al.*, 2000), for example, the classical “leucine zipper” transcription factors are characterised by coiled coils which are necessary for the dimerisation of these proteins (O’Shea *et al.*, 1989). While coiled coil domains are present in all the proteins identified to have COS boxes (which are microtubule-localised), the absence of microtubule localisation in the greater RBCC superfamily indicates that the coiled coil domain alone is likely not responsible for specifying the localisation. Additionally, the absence of RING finger and B-box domains in the microtubule-associated protein GFLND excludes these domains (which are present in the remainder of the COS box-containing proteins) as being required for microtubule localisation. In addition, the MURF proteins have completely different C-termini to the microtubule associating C-I subfamily, which excludes domains C-terminal to the COS box motifs in microtubule association. Notably, Spencer and colleagues had previously searched for similarities between MID1, MID2 and MURF3, but failed to identify any specific domain that might mediate their common binding to the microtubule cytoskeleton, hypothesising that each of the proteins uses a unique *modus operandi* to associate with the microtubule network (Spencer *et al.*, 2000). While this may still be possible, the identification of the COS box would now argue against this. Indeed, the “exclusion” approach described above that provided the initial support for the involvement of the COS box with microtubule association was subsequently validated through functional testing by mutation of two regions predicted to be involved in alpha-helical secondary structure formation within the COS box sequence. These two mutations were sufficient to completely disrupt microtubule association, with no apparent effect on dimerisation through the coiled coil domain. Thus, it can be concluded that the COS box is responsible

for the localisation of those RBCC proteins that are microtubule associated, and is necessary for the correct positioning of MID1 and MID2 within cells (Short and Cox, 2006).

The localisation exhibited by MID1 and MID2 is also required to tether Alpha 4 to the microtubules (Short *et al.*, 2002) which brings PP2Ac into a three-protein complex (Liu *et al.*, 2001), to enable the subsequent degradation of the microtubule localised PP2Ac pool (Troddenbacher *et al.*, 2001). Although, this process of localised PP2Ac turnover on the microtubules appears to be commonly affected among Opitz syndrome patients, there have been no Opitz syndrome causing MID1 mutations identified to date which would directly alter COS box amino acids (albeit two truncating mutations have been found which terminate in the COS box, see Figure 1.6B). However, our group has identified a COS box sequence variant (A358D) in MID2 that from large scale genome sequencing efforts appears to result from a low frequency polymorphism (~5%) that is restricted to people of European ancestry. Interestingly, we have recently noted the presence of this variant in ~20-25% of our Opitz syndrome cohort (but less than 2% of controls), curiously all of which do not have any detectable MID1 alteration. Although the ethnicity of many of these patients is still unknown, there is unfortunately no direct functional data as yet to support a role for this change in susceptibility to disease presentation although the data are suggestive of such (TC Cox unpublished data).

Despite the lack of COS box abnormalities, the preponderance of mutations outside this region that disrupt the normal microtubule localisation indicates that point mutations in the surrounding domains or truncations after it must interfere with the function of this microtubule interaction motif. For example, during the work presented as part of this thesis, a mutation was identified in the C-terminus of TRIM9 which led to a cytoplasmic “clumping” appearance (Short and Cox, 2006) reminiscent of the cell localisation pattern

caused by mutation of either MID1 or MID2 C-terminal domains (Short *et al.*, 2002). Although the identity between MID1/MID2 and TRIM9 is approximately 38% at the amino acid level, all proteins are members of the C-I subfamily and thus may indicate a common mode of disruption of the microtubule localising apparatus, the COS box. Since the C-terminal B30.2-like domain and indeed also Fibronectin type III domains have been shown to be globular (Koide *et al.*, 1998; Seto *et al.*, 1999), mutations causing the misfolding of these complex domain structures could result in hydrophobic amino acids from core regions of the domains becoming displaced. Any non-specific attraction of exposed hydrophobic regions of the protein could conceivably disrupt any protein interaction interfaces that require specific placement within the protein tertiary structure for normal activity. The discovery of an interaction between MID1 and the protein chaperone Heat shock protein 90 (M.A. Massiah and T.C. Cox, unpublished data) indirectly supports the idea that MID1 has a complex tertiary structure (since it requires chaperones for folding).

While study of the wider C-I subfamily has helped determine the basic functional properties of MID1 such as microtubule association and dimerisation, a more specific investigation of MID1/MID2 has provided greater detail about the roles unique to these two proteins. In comparison with other homologous proteins, MID1 and MID2 share the greatest similarity within the C-I subfamily. Indeed, their ability to heterodimerise on the microtubule cytoskeleton and interact with Alpha 4 (Perry *et al.*, 1999; Short and Cox, 2006; Short *et al.*, 2002) together with their similar tissue expression patterns during development (Pinson *et al.*, 2004) supports the notion of functional redundancy and the possibility of co-operativity in the pathogenesis of Opitz syndrome. From an evolutionary perspective, this level of identity is indicative of a relatively recent duplication of an ancestral MID1/MID2 gene. Similar but earlier gene duplications can also be envisaged to

have generated the TRIM36/TRIFIC and TRIM9/TNL gene pairs. Notably, the TRIM9/TNL ancestral gene and not the MID1/MID2 ancestral gene is likely to have given rise to all members of the C-I subfamily given invertebrates such as *Drosophila melanogaster* and *Caenorhabditis elegans* only have a single C-I family member that most closely resembles TRIM9, despite both having Alpha 4 orthologues (Short and Cox, 2006). As has been demonstrated in this thesis, however, human TRIM9 does not physically interact with Alpha 4 (Short and Cox, 2006) and recently Cygnar *et al.* (Cygnar *et al.*, 2005) have also provided genetic evidence against an interaction between *Drosophila* Trim9 and Alpha 4/Tap42 orthologue.

5.3 Significance of the MID1 Alpha 4 interaction

The discovery of the physical interaction between Alpha 4 and MID1, and the known function of Alpha 4 as a regulator of PP2Ac activity in cellular phosphatase signalling is intriguing for several reasons. Firstly, there was previously no linkage between any intracellular phosphatase protein activity and the genes known to be involved (or implicated) in craniofacial development or other tissue morphogenetic processes requiring fusion. While studies of tissue morphogenesis have detailed the types and transformations of many cell types during development of the face, the molecular interactions controlling the known cell movements and fates are not well characterised. It is now known that the microtubule (MT) pool of a protein phosphatase (PP2Ac) is most likely controlled through a direct interaction between MID1 and Alpha 4 and degradation of the phosphatase mediated by a process of ubiquitylation.

There are multiple discrete pools of PP2Ac present in cells, and those pools are independently regulated in order to maintain balances in phosphorylation of substrates in particular sub-cellular compartments (Sontag *et al.*, 1995). This is important due to the

relatively small number of phosphatases that control phosphorylation of proteins in comparison with the vast number of highly specific kinases which act in many discrete pathways. Compartmentalisation of PP2Ac ensures that the regulation of phosphorylation within one specific area (e.g. microtubules) does not affect the regulation of it in other areas (e.g. at the cell membrane) which would non-specifically imbalance the phosphorylation status of proteins at other sites. As implied by the spectrum of clinical features that define the Opitz syndrome phenotype, such a loss of the elegant balance of PP2Ac activity within sub-cellular compartments is likely to be catastrophic for the localised independent control of protein function. Considering the potential for apoptosis, epithelial migration or epithelial-mesenchymal transition (EMT) being the unifying process(es) in the breakdown of normal tissue patterning events which underlie the Opitz syndrome phenotype (as discussed in the introduction), the effect of MID1 mutation on microtubular PP2Ac pools and how subsequent PP2Ac dysregulation may affect pathways involved in both epithelial behaviour (e.g. EMT) and apoptosis will be discussed here.

5.4 Epithelial-mesenchymal transition (EMT)

Epithelial to mesenchymal transition is a cellular transformation employed throughout development to differentiate epithelial cells into mesenchymal cells with fibroblastic migratory properties in order to enact programmed cellular changes critical to tissue morphogenesis. Key aspects of development are reliant on EMT, for example, the generation of mesoderm during gastrulation (Vincan, 2004), and the production of neural crest cells from neural tube epithelia (Pla *et al.*, 2001). The process of EMT is activated through the coordination of several molecular pathways, with one of the pathways being controlled by canonical Wnt signalling.

The Wnts are a family of conserved morphogenic ligands present in animals which have been shown to activate a well established 'canonical' pathway, or alternative newly identified signalling cascades (known as non-canonical pathways) (Widelitz, 2005). The term 'non-canonical' loosely describes the alternative pathways in which particular members of the well defined canonical pathway may additionally act. Non-canonical pathways include the calcium pathway (controlling cell adhesion and promotion of ventralised fate during embryonic axis determination (Kohn and Moon, 2005)), convergent extension pathway (establishment of cell polarity (Tada *et al.*, 2002)), steroid receptor binding pathway (binding of Androgen receptor by β -catenin (Mulholland *et al.*, 2005)) and the atypical receptor tyrosine kinase (RTK) pathway (the binding of RTKs directly to Wnt ligands (Cheyette, 2004)). While these pathways control significant cell and developmental signalling pathways of their own, the canonical pathway in which the Wnt proteins are involved primarily controls the induction of EMT. The canonical pathway triggers the formation of a nuclear transcription factor signalling complex which results in the expression of transcription factors known to be critical for homeostasis and development (e.g. *Slug* and *Snail*). Expression of the *Slug* and *Snail* transcription factors results in the down regulation of E-cadherin, an epithelial cell marker necessary for cell-cell contact (Cano *et al.*, 2000; Conacci-Sorrell *et al.*, 2003). Down regulation of E-cadherin results in the loss of characteristic epithelial cell-cell contacts, and begins the differentiation pathway known as epithelial-mesenchymal transition (EMT) (for a review of EMT in development, see (Shook and Keller, 2003)). Due to the ability for Wnt signalling to change epithelial cells into a migratory cell type (mesenchyme), the consequences of misregulation of the Wnt pathway are severe. The normal function of the pathway controls cell plasticity and transdifferentiation, and many proteins involved in the pathway have been shown to be either upregulated or mutated in multiple forms of cancer

such as colorectal, thyroid and prostate cancers (Abbosh and Nephew, 2005; Bos *et al.*, 2006; Mulholland *et al.*, 2006; Suraweera *et al.*, 2006). As such, the Wnt signalling cascade is coordinated by a highly complex series of interactions between many proteins required to finely tune activation to minimise the possibility of aberrant activity. The coverage of all proteins and co-activators within Wnt signalling and the EMT pathway is beyond the scope of this discussion; however the key proteins necessary for basic activity of the pathway will be covered in order to provide an understanding of the implications of the effect of potential MID1 activity on the pathway during development.

Normally, Wnt signalling is only activated in the presence of Wnt ligands. In naïve cells (i.e. those not receiving a Wnt stimulus), the main proteins acting within the Wnt pathway form a “destruction complex” with the core protein Axin docking to the microtubules through an interaction with Adenomatous Polyposis Coli (APC), with available β -catenin, Glycogen Synthase Kinase 3 (GSK3) and Casein Kinase I (CKI) also indirectly anchored to the microtubules through Axin. The two kinases GSK3 β and CKI α act on available β -catenin, the phosphorylated form of which is subsequently recognised by an E3 ubiquitin ligase β -TrCP, and degraded by the proteasome (Jia *et al.*, 2005). Therefore, no significant amount of β -catenin accumulates in the cytoplasm and little or no β -catenin is transported into the nucleus to upregulate the expression of Wnt target genes (such as *Snail*) (Figure 5.1A). The activity of GSK3 β , in a fashion similar to its role in the demise of β -catenin, also selectively phosphorylates the transcription factor Snail, resulting in its eventual ubiquitylation by β -TrCP and subsequent degradation (Yook *et al.*, 2005). Thus, in naïve epithelia GSK3 β directly inhibits the further transcription of *Snail* through the degradation of β -catenin, and actively degrades Snail protein leading to continued maintenance of E-cadherin cell-cell contacts, and inhibition of EMT.

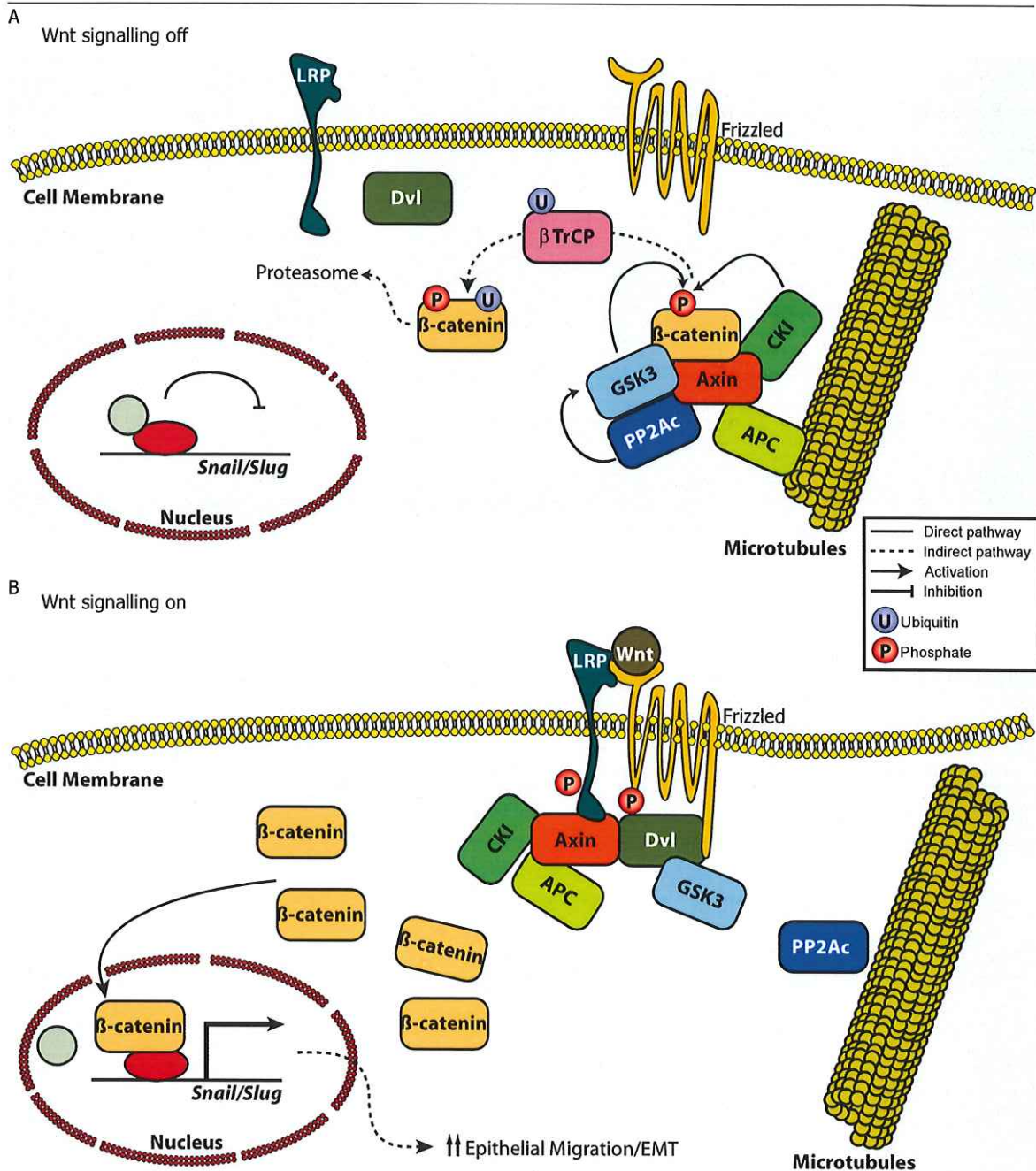


Figure 5.1 Simplified model of canonical Wnt signalling.

In the absence of a Wnt ligand, β -catenin is rapidly turned over by ubiquitylation. β -catenin is associated with a microtubule bound complex, and its eventual ubiquitylation begins with phosphorylation by protein kinases within the complex (A). When Wnt ligands bind Frizzled, Axin and Dvl are recruited to the cell membrane and phosphorylated which causes a redistribution of GSK3 β . Thus, β -catenin is no longer degraded, and acts as a transcription co-activator to up-regulate the expression of genes involved in the induction of EMT (B).

Activation of Wnt signalling occurs through a relay of messages from the cell membrane, where the soluble Wnt ligands bind Frizzled (Fz) transmembrane receptors coupled with low density lipoprotein (LDL) receptor-related protein (LRP) transmembrane receptor complexes. Several models explain the flow of protein activation from this point to nuclear β -catenin accumulation and enhancement of gene transcription. One recent model by Cadigan and Liu has been refined by a myriad of the most recent Wnt pathway investigations and, in simple terms, explains the observations of an accumulation of membrane-bound Axin and Dishevelled (Dvl) proteins followed by activation of Wnt target genes (Cadigan and Liu, 2006).

The activation model proposed follows binding of Wnt to Fz/LRP at the cell membrane, at which point two simultaneous signals are transmitted. One of the transmitted signals begins with the phosphorylation of the intracellular component of LRP by GSK3 and then by CKI, which rapidly recruits Axin to the cell membrane (Davidson *et al.*, 2005). This results in a fundamental change in the activity of the β -catenin “destruction complex” since Axin, the main anchor protein for the complex, is no longer able to bring all the remaining proteins into close proximity with one-another on the microtubules (where they together coordinate the degradation of β -catenin). This process increases levels of nuclear localised β -catenin which enhances the expression of *Snail*. The resulting down regulation of E-cadherin by Snail causes a decrease in cell-cell contacts, which is a fundamental trigger for the induction and commitment to EMT. The derepression of Snail during Wnt induction is similar to that seen for β -catenin, where available cytoplasmic Snail accumulates and translocates to the nucleus (Yook *et al.*, 2005). Repression of the degradation of both β -catenin and Snail through Wnt signalling has been proposed as a conserved tandem-acting cellular mechanism to activate transcription of genes that control epithelial-mesenchymal transition (Yook *et al.*, 2005).

The second signal transmitted from the cell membrane upon Wnt binding of Fz begins with the phosphorylation of Dishevelled (Dvl) (Takada *et al.*, 2005) which becomes plasma membrane associated (Cliffe *et al.*, 2003). The phosphorylation of Dvl has been shown to be independent of LRP activation (Gonzalez-Sancho *et al.*, 2004) and is thought to down regulate the activity of GSK3 (which would promote both β -catenin and Snail activity). There has also been a suggestion that Dvl may bind to Axin at the cell membrane along with LRP, or may be involved in the active recruitment of Axin to the cell membrane (Cadigan and Liu, 2006). While Dvl has been shown to be critical for Wnt signalling, and inhibits GSK3 activity at the membrane, much of the finer detail of Dvl activation and activity is still unknown. Thus the Cadigan and Liu model of the Wnt signalling pathway explains experimental evidence of Wnt activation of Dvl and LRP, which causes recruitment of Axin and GSK3 to the membrane. At the membrane, GSK3 is inhibited by Dvl and therefore is unable to start the cascade of events that leads to the degradation of both β -catenin and Snail (Cadigan and Liu, 2006).

In essence, the Wnt pathway is controlled by a phosphorylation cascade using multiple kinases to activate downstream effector proteins that ultimately pass an extracellular signal through to the nucleus to enact a change in gene expression. In order to maintain control of pathway activity, most kinases are counteracted by phosphatases, and specificity for dephosphorylation of discrete proteins is generated by variations in the regulatory, catalytic and core subunits of the phosphatases, such as described for Protein Phosphatase 2A (Li *et al.*, 2001). Since the phosphorylation of some proteins can have either positive or inhibitory effects on Wnt signalling, the act of phosphatase activity can also promote or inhibit Wnt signalling depending on the target proteins involved. This sometimes confusing paradigm has been described by Li and colleagues upon the first discovery of opposing roles of phosphatase activity through the B56 PP2A regulatory

subunit on Wnt signalling as distinct from earlier findings of the role of PP2A catalytic subunit activity (Hsu *et al.*, 1999; Li *et al.*, 2001).

A major component of Wnt signalling relies on the accumulation of a complex primarily anchored to the microtubules by APC (Zumbrunn *et al.*, 2001). This is one point where other signalling pathways may also modulate activity of several proteins essential to Wnt signalling, including those controlled by the microtubular PP2Ac pool. While Axin, GSK3 β , APC and β -catenin proteins are main components of the pathway, they are additionally controlled by other signalling cascades under the influence of alternative extra- and intra-cellular signals. This means that pathways under the control of other factors can impact the activation state of Wnt signalling components, and therefore the activity of Wnt in modulating cell behaviour. For example, activation of the Akt pathway generally inhibits the activity of GSK3 β (which phosphorylates β -catenin, leading to its degradation) (Liang and Slingerland, 2003), while independent E-cadherin regulation controls the release of β -catenin from the cell membrane and its availability to signal within the Wnt pathway through the cytoplasm and nucleus (Stockinger *et al.*, 2001).

While these pathways are important for Wnt activity, they are relatively ancillary to the main pathway, where microtubular PP2Ac activity on Wnt signalling is highly significant. PP2Ac has been shown to directly interact with a region of Axin which would permit concurrent interaction with APC, GSK3 β and β -catenin as well as either dimerise or interact with Dvl (Hsu *et al.*, 1999). As has been previously highlighted, control of the activity of the microtubule associated Wnt signalling components (APC, Axin, GSK3 β and β -catenin) is through kinase and phosphatase activity. In relation to Opitz syndrome, MID1 mutation has been shown to result in a reduced ability to degrade microtubular PP2Ac, leading to hypophosphorylation of numerous microtubule proteins (Troockenbacher *et al.*, 2001). Therefore, if PP2Ac is a component of the microtubule bound Axin signalling

complex, a breakdown in PP2Ac regulation (as is seen in Opitz syndrome) could potentially result in changes in the regulation of Axin phosphorylation and Wnt signalling. Additionally, PP2Ac can positively regulate Wnt signalling downstream of β -catenin, although the target of this PP2Ac activity and site of action are unknown (Ratcliffe *et al.*, 2000).

Together, these findings provide a mechanism in which MID1, MID2, Alpha 4 and PP2Ac may act within the bounds of Wnt signalling such that any misregulation of the microtubule localised fraction of PP2Ac could feasibly impact on the microtubule-linked components of the Wnt signalling cascade to effect changes E-cadherin levels and ultimately epithelial behaviour. For example, modulation of the cell adhesive properties of epithelia, a precursor to both epithelial movement and EMT, is a requirement for embryonic fusion events and one that is known to be mediated by canonical Wnt signalling. It is therefore possible that misregulation of Wnt signalling in specific tissues could give rise to the fusion defects and thus provide a causative link for the consequences of MID1 mutation and the phenotypic anomalies observed in Opitz syndrome, such as hypospadias, cleft lip with or without cleft palate, heart septation defects and ano-rectal fistulae/imperforate anus.

5.5 Apoptosis

The complex patterning of organs and appendages from primordial buds requires coordinated growth and differentiation, counterbalanced by a reduction of cells in certain areas of the body. The process of growth and differentiation is highly regulated and exquisitely timed, as is the control of cell death pathways. The precise control of apoptosis is important for most developmental processes, but perhaps best recognised in the process of interdigitation necessary to correctly pattern the digits of hands and feet (Mirkes, 2002).

It is also thought to play an important role in the removal of epithelial cell layers during the fusion of tissue masses, for example during formation of the heart septum (Poelmann and Gittenberger-de Groot, 2005).

The proposition that apoptosis is a necessary process for normal fusion events (such as those in the mammalian lip/palate) has been around for some time (Mori *et al.*, 1994). The model for the involvement of apoptosis in this process relates to the removal of external epithelial cell layers which allows the underlying mesenchyme from opposing tissue masses to converge to form a solid inner confluent mesenchymal tissue mass (Cox, 2004). In this model, a failure in epithelial apoptosis leaves a barrier, inhibiting the underlying mesenchyme from joining and forming a confluent mesenchymal bridge between the adjacent tissue masses. Therefore, any disruption to the normal regulatory mechanisms governing apoptosis could conceivably also affect fusion events in the developing embryo and provide an alternative explanation for the spectrum of features in patients with Opitz syndrome (Pinson *et al.*, 2004).

Recently, evidence from conditional inactivation of the murine *Alpha 4* gene (Kong *et al.*, 2004) has provided support for a role for Alpha 4 in the regulation of apoptosis and thus, indirectly implicating MID1 in the regulation of programmed cell death. The Cre-lox conditional inactivation approach used by Kong *et al.* permitted avoidance of an early embryonic lethal phenotype by restricting *Alpha 4* deletion to particular cell lineages (Kong *et al.*, 2004). Floxed *Alpha 4* animals, which have a normal functioning gene (but with *loxP* sites inserted by homologous recombination flanking a region spanning several *Alpha 4* exons), were mated with transgenic mice engineered to express the Cre recombinase in a tissue-specific manner. The pups generated from the mating were *Alpha 4* deficient in cells expressing Cre recombinase, that is the promoter/enhancers driving Cre expression defined the cells and tissues in which *Alpha 4* was deleted.

It is known that Alpha 4 and PP2Ac interact and target several proteins for dephosphorylation (Chen *et al.*, 1998; Murata *et al.*, 1997), and even MID1 has been postulated to be a possible target of this activity (Liu *et al.*, 2001). Thus, the effect of complete *Alpha 4* removal from cells is likely to have broader effects, independent from MID1/MID2 related activity. However, the evidence that MID1/2 interact with Alpha 4 to potentially drive the degradation of PP2Ac (Short *et al.*, 2002; Trockenbacher *et al.*, 2001) indicates that MID1/2 may also be involved in the same Alpha 4 pathways discovered by Kong *et al.* (Kong *et al.*, 2004). Considering this, the observation of early embryonic lethality by Kong *et al.* in null *Alpha 4* mice is perhaps not surprising (Kong *et al.*, 2004). Due to the lethality observed, mouse embryonic fibroblasts were isolated from floxed *Alpha 4* mice, and cells were stably transfected with Cre-expression constructs to specifically knock out *Alpha 4* in these cells. A rapid decline in cell viability with only 20% cell survival 120 hours after *Alpha 4* deletion was observed, with cells displaying increased levels of phosphorylated cJun(Ser63) and p53(Ser18) which are potent inducers of apoptosis. Notably, several intrinsic mitochondrial apoptotic pathway genes were also upregulated in these cells, indicating that the apoptotic response to *Alpha 4* deficiency was broad and not specifically related to phospho-p53 induction. Caspase-3 assays, TUNEL staining, and a reduction in apoptosis when an intrinsic apoptosis pathway inhibitor (Bcl-x_L) was overexpressed confirmed that apoptosis was a direct consequence of the absence of *Alpha 4*. In addition, apoptosis driven by *Alpha 4* deficiency occurred in both terminally differentiated non-proliferating cells/tissues, as well as developing cells (Kong *et al.*, 2004).

5.6 Model for apoptosis-based pathogenesis of Opitz syndrome

Together, the findings of many groups allow the formulation of a unifying model based on apoptosis and induction of migratory properties to explain the molecular pathogenesis of fusion defects in Opitz syndrome. For simplicity, the model has been related here to the cellular events involved in fusion of the lip and primary palate and, as a result of mutations in MID1, the disruption to normal midfacial development (seen as a cleft of the lip and/or primary palate).

1. In the normal state (Figure 5.2A-B):
 - a. The expression of MID1 and MID2 is spatially and temporarily regulated, with notably stronger expression in epithelia covering the tissue that is destined to fuse (Everett and Brautigan, 2002; Pinson *et al.*, 2004; Richman *et al.*, 2002).
 - b. At the subcellular level, MID1 and MID2 are localised to the microtubules (Perry *et al.*, 1999; Schweiger *et al.*, 1999) through their COS box motifs (Short and Cox, 2006), leaving other domains available for other functions.
 - c. MID1 and MID2 interact with Alpha 4 through their B-box type 1 domains (Short *et al.*, 2002), and indirectly with PP2Ac (through Alpha 4) (Liu *et al.*, 2001), leading to the eventual ubiquitylation of PP2Ac (Troockenbacher *et al.*, 2001) and its subsequent targeting for destruction by the proteasome.
 - d. Since Alpha 4 provides substrate specificity for PP2Ac phosphatase action, degradation of PP2Ac by MID1/2 provides a mechanism by which to regulate the level of active Alpha 4/PP2Ac dimers (Kong *et al.*, 2004) or PP2Ac alone: this phosphatase activity of Alpha 4/PP2Ac is probably acting to counter signals from a variety of extracellular stimuli including, for

example, BMP4 and Wnt, that control the balance between apoptosis, epithelial movement and EMT. Crosstalk between various signalling molecules and pathways in the early oral epithelia may facilitate the observed restriction of apoptosis to the epithelia of the pre-fusion zone (the region of the facial primordia that contact during fusion) (Cox, 2004).

- e. As the merging tissue masses of the facial primordia meet, epithelial apoptosis, movement and EMT continues to occur to remove any remaining pockets of the epithelial seam (Cox, 2004; Hay, 2005; Kang and Svoboda, 2005). With all epithelia removed or converted to mesenchyme, thickening of the mesenchymal bridge occurs, providing strong resistance to torsional stresses imposed on the primary palate during the remainder of facial development (tearing of the bridge by these stresses would cause a cleft) (Cox, 2004).

2. In the disease state (Figure 5.2C):

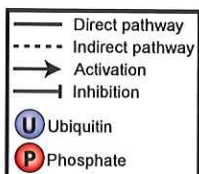
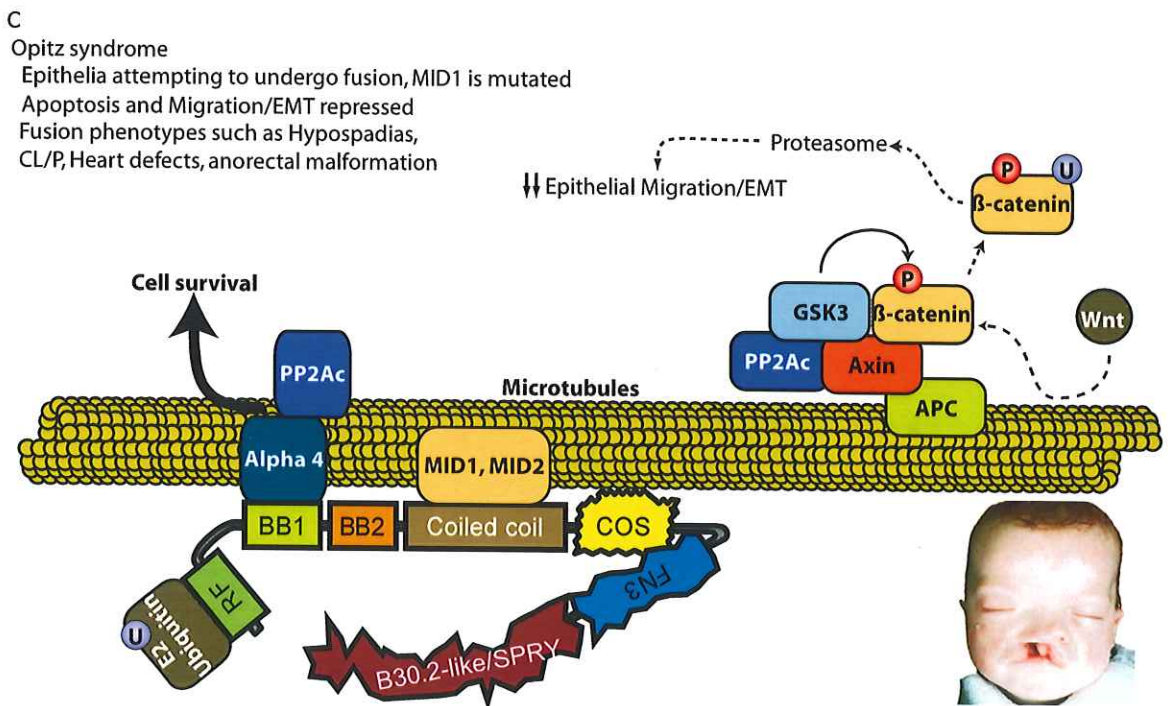
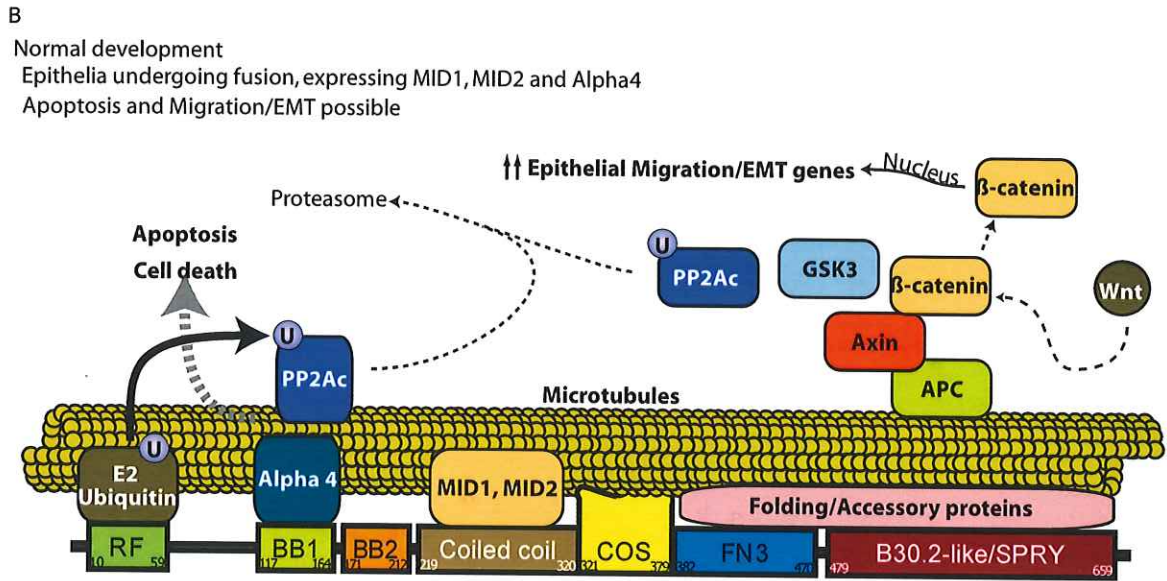
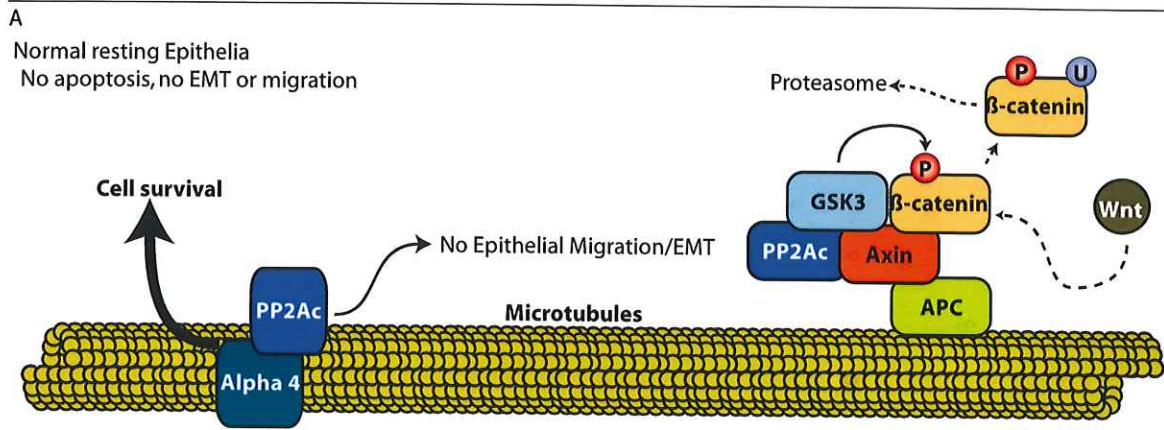
- a. Mutation of MID1 disrupts its ability to associate with the microtubule cytoskeleton (Cox *et al.*, 2000). The majority of MID1 mutations affect the C-terminal domains leaving the coiled coil domain of MID1 potentially still available to facilitate dimerisation with MID2, and consequently also causing MID2 to become mislocalised, i.e. a 'dominant negative' effect on MID2.
- b. MID1/2 may still be able to interact with Alpha 4 (Short *et al.*, 2002) and possibly even cause ubiquitylation of PP2Ac. However, the lack of sufficient microtubular coverage means that the effects of PP2Ac turnover are not localised.

- c. In the developing face, the inability to regulate levels of Alpha 4/PP2Ac dimers (or PP2Ac alone) in the epithelia in response to the appropriate developmental signals results in sustained repression of apoptosis, movement and EMT. Inhibition of these processes would result in a sustained epithelial state, thus causing an inability to completely remove the epithelia trapped between the fusing facial primordia, resulting in a cellular barrier to the formation of a strong mesenchymal bridge (Cox, 2004).
- d. If some of the MID2 escapes the inactivation by the mutated MID1 (or even if some residual MID1 ubiquitylation activity remains) some degradation of PP2Ac may still occur. The decreased regulatory capacity could lead to variable levels of PP2Ac or Alpha 4/PP2Ac phosphatase dimers and thus incomplete or mild derepression of apoptosis (Kong *et al.*, 2004) or a small amount of EMT or movement, which may influence whether or not the patient will present with a cleft. In this scenario, variable levels of expression of MID2 or signalling from various growth factors (perhaps as a result of epigenetic phenomena) could be determining factors for which individuals present with the abnormality and thus account for the observation that only ~50% of patients with X-linked OS show cleft of the lip and/or primary palate (So *et al.*, 2005).

Consistent with the involvement of MID2 in the model and the likelihood of some degree of functional redundancy between MID1 and MID2, a *Mid1* knock out mouse (on a 129Sv/MF1 genetic background) has been generated by Alan Ashworth (Breakthrough Breast Cancer Research Centre, London, UK; personal communication). Unlike the situation in humans, these mice show no gross developmental anomalies, however more

Figure 5.2 Model of Opitz syndrome molecular pathogenesis.

Normally, Alpha 4 and PP2Ac act to maintain cell viability and keep epithelial homeostasis balanced to incoming signals, with PP2Ac acting in concert with other proteins to regulate the phosphorylation of β -catenin by GSK3 β (A). During development, constant expression of MID1 (and MID2) maintains moderate levels of microtubular PP2Ac. With a transient increase in expression of the MID proteins, for example in facial epithelia prior to fusion, microtubule PP2Ac levels would be expected to decrease. Such a reduction in PP2Ac could feasibly alter the balance of the Axin-bound complex (part of the Wnt signalling pathway) and lead to phosphorylation of GSK3 β , inhibiting the phosphorylation of β -catenin (which in-turn inhibits its ubiquitylation and proteolysis). As a result, β -catenin would be stabilised, increasing its transport to the nucleus and thus transcription of genes that can lead to changes in epithelial behaviour such as cell movement or epithelial-mesenchyme transition (B). When MID1 is mutated, PP2Ac levels are not altered by ubiquitylation (C), thus promoting the maintenance of an epithelial state (since PP2Ac levels are maintained, GSK3 β remains active), and repression of apoptosis by Alpha 4/PP2Ac. If apoptosis and cell migratory signals are repressed, then areas of the developing body reliant on MID-controlled apoptosis and cell migration such as the primary palate, heart, urethra and ano-rectal area will not form correctly, as is seen in Opitz syndrome, with many patients having defects in these tissues.



detailed analyses are still required. Although it is well documented that phenotypic presentation can vary dramatically in mice (from no phenotype to full disease states) depending on the genetic background of the strain, a heightened susceptibility in a state of corticosteroid or ‘nutritional’ restriction is also feasible since these are known factors influencing facial clefting in humans and some strains of mice (Schubert *et al.*, 1990).

Ongoing studies will begin to investigate these possibilities. In addition, breeding programs are underway in our laboratory to move the *Mid1* knockout locus onto different genetic backgrounds to investigate the possibility of any strain-specific sensitivities. The model outlined above would account for the observation of a normal phenotype in *Mid1* knockout mice because the complete deletion would not alter the subcellular localisation of Mid2 (as demonstrated with some OS-causing MID1 mutations). The Mid2 protein similarly associates with microtubules (Perry *et al.*, 1999), interacts even more strongly with Alpha 4 than Mid1 (Short *et al.*, 2002), and likely has similar E3 ubiquitin ligase activity due to RING finger sequence identity. Therefore, there is a strong possibility that this functional redundancy may still facilitate sufficient regulation of Alpha 4/PP2Ac levels in times of Mid1 deficiency.

The model also accounts for the observation of similar developmental anomalies between Opitz syndrome and a disease with an FG syndrome-like phenotype that was associated with a 5’UTR mutation in Alpha 4 (Graham *et al.*, 2003). Any genetic intervention which affects the formation of Alpha 4/PP2Ac dimers, such as a reduction in the amount of Alpha 4 being translated (Graham *et al.*, 2003), could conceivably alter Alpha 4’s ability to repress apoptosis. In this case, the phosphatase activity of the total pool of available Alpha 4 (with PP2Ac) may be insufficient to inhibit the transcription of apoptosis-inducing proteins (Kong *et al.*, 2004).

5.7 Future Directions

The model proposed provides a framework to fit current observations of patient phenotype with modifications to the molecular function of several controller pathways that can affect tissue and cellular changes. While it represents a plausible explanation for the current observations, there are areas which need to be investigated further. Particular attention should be paid to the interaction between MID1 and Alpha 4, due to the compelling evidence that Alpha 4 has a key role in maintaining cell viability. The suggestion that MID1 activity may negate the Alpha 4-PP2Ac cell survival signal is important for understanding the cellular changes during many MID-involved fusion events during development. If apoptosis is a key process in such events, it is important to determine the role that MID1 and MID2 may play in apoptosis in these tissues. In addition, since the microtubular PP2Ac-Axin interaction is critical for Wnt signalling, reduction of PP2Ac levels by MID1 may therefore be important for Wnt signalling. Thus characterising any changes in behaviour of epithelia in the tissues affected by Opitz syndrome is essential in any investigation of the role of MID1 in the Wnt pathway. Studies aimed at targeted knock down of Mid2, both by itself and together with Mid1 in mice, will also be important for confirming any functional redundancy between these closely related proteins and illuminating any role they play in modulating the activity of signalling pathways in development.

Bibliography

Bibliography

Abbosh, P. H., and Nephew, K. P. (2005). Multiple signaling pathways converge on beta-catenin in thyroid cancer. *Thyroid* 15, 551-561.

Allanson, J. E. (1988). G syndrome: an unusual family. *American Journal of Medical Genetics* 31, 637-642.

Amati, F., Conti, E., Novelli, A., Bengala, M., Diglio, M. C., Marino, B., Giannotti, A., Gabrielli, O., Novelli, G., and Dallapiccola, B. (1999). Atypical deletions suggest five 22q11.2 critical regions related to the DiGeorge/velo-cardio-facial syndrome. *European Journal of Human Genetics* 7, 903-909.

Arndt, C., Cruz, M., Cardenas, M., and Heitman, J. (1999). Secretion of FK506/FK520 and rapamycin by *Streptomyces* inhibits the growth of competing *Saccharomyces cerevisiae* and *Cryptococcus neoformans*. *Microbiology* 145, 1989-2000.

Arya, S., Viseskul, C., and Gilbert, E. F. (1980). The G syndrome - additional observations. *American Journal of Medical Genetics* 5, 321-324.

Attur, M. G., Patel, R., Thakker, G., Vyas, P., Levartovsky, D., Patel, P., Naqvi, S., Raza, R., Patel, K., Abramson, D., *et al.* (2000). Differential anti-inflammatory effects of immunosuppressive drugs: cyclosporin, rapamycin and FK-506 on inducible nitric oxide synthase, nitric oxide, cyclooxygenase-2 and PGE2 production. *Inflammation Research* 49, 20-26.

Attwood, T. K., Bradley, P., Flower, D. R., Gaulton, A., Maudling, N., Mitchell, A. L., Moulton, G., Nordle, A., Paine, K., Taylor, P., *et al.* (2003). PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Research* 31, 400-402.

Bai, Y., Chen, H., Yuan, Z. W., and Wang, W. (2004). Normal and abnormal embryonic development of the anorectum in rats. *Journal of Pediatric Surgery* 39, 587-590.

Ballabio, A., and Andria, G. (1992). Deletions and translocations involving the distal short arm of the human X chromosome: review and hypotheses. *Human Molecular Genetics* 1, 221-227.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., *et al.* (2004). The Pfam protein families database. *Nucleic Acids Research* 32 *Database issue*, D138-141.

Bershof, J. F., Guyuron, B., and Olsen, M. M. (1992). G syndrome: a review of the literature and a case report. *Journal of Cranio-Maxillo-Facial Surgery* 20, 24-27.

Berti, C., Fontanella, B., Ferrentino, R., and Meroni, G. (2004). Mig12, a novel Opitz syndrome gene product partner, is expressed in the embryonic ventral midline and co-

- operates with Mid1 to bundle and stabilize microtubules. *BMC Cell Biology* 5, 9.
- Borden, K. L., and Freemont, P. S. (1996). The RING finger domain: a recent example of a sequence-structure family. *Current Opinion in Structural Biology* 6, 395-401.
- Bos, C. L., Diks, S. H., Hardwick, J. C., Walburg, K. V., Peppelenbosch, M. P., and Richel, D. J. (2006). Protein phosphatase 2A is required for mesalazine-dependent inhibition of Wnt/ β -catenin pathway activity. *Carcinogenesis*, [Epub ahead of print].
- Briault, S., Hill, R., Shrimpton, A., Zhu, D., Till, M., Ronce, N., Margaritte-Jeannin, P., Baraitser, M., Middleton-Price, H., Malcolm, S., *et al.* (1997). A gene for FG syndrome maps in the Xq12-q21.31 region. *American Journal of Medical Genetics* 73, 87-90.
- Briault, S., Villard, L., Rogner, U., Coy, J., Odent, S., Lucas, J., Passage, E., Zhu, D., Shrimpton, A., Pembrey, M., *et al.* (2000). Mapping of X chromosome inversion breakpoints [inv(X)(q11q28)] associated with FG syndrome: a second FG locus [FGS2]? *American Journal of Medical Genetics* 95, 178-181.
- Brown, E. J., Albers, M. W., Shin, T. B., Ichikawa, K., Keith, C. T., Lane, W. S., and Schreiber, S. L. (1994). A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature* 369, 756-758.
- Brown, E. J., Beal, P. A., Keith, C. T., Chen, J., Shin, T. B., and Schreiber, S. L. (1995). Control of p70 s6 kinase by kinase activity of FRAP in vivo. *Nature* 377, 441-446.
- Buchner, G., Montini, E., Andolfi, G., Quaderi, N., Cainarca, S., Messali, S., Bassi, M. T., Ballabio, A., Meroni, G., and Franco, B. (1999). MID2, a homologue of the Opitz syndrome gene MID1: similarities in subcellular localization and differences in expression during development. *Human Molecular Genetics* 8, 1397-1407.
- Buckley, J. G., Hinton, A. E., Penter, G., and Farndon, P. A. (1988). Total laryngotracheal hypoplasia in a case of G syndrome. *Journal of Laryngology and Otology* 102, 1056-1059.
- Cadigan, K. M., and Liu, Y. I. (2006). Wnt signaling: complexity at the surface. *Journal of Cell Science* 119, 395-402.
- Cainarca, S., Messali, S., Ballabio, A., and Meroni, G. (1999). Functional characterization of the Opitz syndrome gene product (midin): evidence for homodimerization and association with microtubules throughout the cell cycle. *Human Molecular Genetics* 8, 1387-1396.
- Canning, D. A. (1999). Hypospadias trends in two US surveillance systems. Rise in prevalence of hypospadias. *The Journal of Urology* 161, 366.
- Cano, A., Perez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., Portillo, F., and Nieto, M. A. (2000). The transcription factor snail controls epithelial-

mesenchymal transitions by repressing E-cadherin expression. *Nature Cell Biology* 2, 76-83.

Cappa, M., Borrelli, P., Marini, R., and Neri, G. (1987). The Opitz syndrome: a new designation for the clinically indistinguishable BBB and G syndromes. *American Journal of Medical Genetics* 28, 303-309.

Cavallo, L., Acquafredda, A., and Laforgia, N. (1988). Endocrinological studies in the hypertelorism-hypospadias (BBB) syndrome. *European Journal of Pediatrics* 148, 89.

Chemke, J., Shor, E., Ankori-Cohen, H., and Kazuni, E. (1984). Male to male transmission of the G syndrome. *Clinical Genetics* 26, 164.

Chen, J., and Fang, Y. (2002). A novel pathway regulating the mammalian target of rapamycin (mTOR) signaling. *Biochemical Pharmacology* 64, 1071-1077.

Chen, J., Peterson, R. T., and Schreiber, S. L. (1998). Alpha 4 associates with protein phosphatases 2A, 4, and 6. *Biochemical and Biophysical Research Communications* 247, 827-832.

Cheyette, B. N. (2004). Ryk: another heretical Wnt receptor defies the canon. *Science's STKE [electronic resource] : signal transduction knowledge environment* 2004, pe54.

Christian, J. C., Bixler, D., Blythe, S. C., and Merritt, A. D. (1969). Familial telecanthus with associated congenital anomalies. *Birth Defects Original Article Series* V, 82-85.

Christodoulou, J., Bankier, A., and Loughnan, P. (1990). Ring chromosome 22 karyotype in a patient with Opitz (BBBG) syndrome. *American Journal of Medical Genetics* 37, 422-424.

Chung, H., and Brautigan, D. L. (1999). Protein phosphatase 2A suppresses MAP kinase signalling and ectopic protein expression. *Cellular Signalling* 11, 575-580.

Cliffe, A., Hamada, F., and Bienz, M. (2003). A role of Dishevelled in relocating Axin to the plasma membrane during wingless signaling. *Current Biology* 13, 960-966.

Coburn, T. P. (1970). G syndrome. *American Journal of Diseases of Children* 120, 466.

Cohen, M. (1980). Malformation syndromes, In *Surgical Correction of Dentofacial Deformities*, W. Bell, W. Proffit, and R. White, eds. (Philadelphia: W.B. Saunders), pp. 7-44.

Conacci-Sorrell, M., Simcha, I., Ben-Yedidia, T., Blechman, J., Savagner, P., and Ben-Ze'ev, A. (2003). Autoregulation of E-cadherin expression by cadherin-cadherin interactions: the roles of beta-catenin signaling, Slug, and MAPK. *Journal of Cell Biology* 163, 847-857.

-
- Cooper, T. G. (2002). Transmitting the signal of excess nitrogen in *Saccharomyces cerevisiae* from the Tor proteins to the GATA factors: connecting the dots. *FEMS Microbiology Reviews* 26, 223-238.
- Cordero, J. F., and Holmes, L. B. (1978). Phenotypic overlap of the BBB and G syndromes. *American Journal of Medical Genetics* 2, 145-152.
- Cote, G. B., Katsantoni, A., Papadakou-Lagoyanni, S., Costalos, C., Timotheou, T., Skordalakis, A., Deligeorgis, D., and Pantelakis, S. (1981). The G syndrome of dysphagia, ocular hypertelorism and hypospadias. *Clinical Genetics* 19, 473-478.
- Cox, T. C. (2004). Taking it to the max: the genetic and developmental mechanisms coordinating midfacial morphogenesis and dysmorphology. *Clinical Genetics* 65, 163-176.
- Cox, T. C., Allen, L. R., Cox, L. L., Hopwood, B., Goodwin, B., Haan, E., and Suthers, G. K. (2000). New mutations in MID1 provide support for loss of function as the cause of X-linked Opitz syndrome. *Human Molecular Genetics* 9, 2553-2562.
- Cox, T. C., Cox, L. L., and Ballabio, A. (1998). A very high density microsatellite map (1 STR/41 kb) of 1.7 Mb on Xp22 spanning the microphthalmia with linear skin defects (MLS) syndrome critical region. *European Journal of Human Genetics* 6, 406-412.
- Craig, D., Gao, M., Schulten, K., and Vogel, V. (2004). Tuning the mechanical stability of fibronectin type III modules through sequence variations. *Structure* 12, 21-30.
- Cygnar, K. D., Gao, X., Pan, D., and Neufeld, T. P. (2005). The phosphatase subunit tap42 functions independently of target of rapamycin to regulate cell division and survival in *Drosophila*. *Genetics* 170, 733-740.
- da Silva, E. O. (1983). The hypertelorism-hypospadias syndrome. *Clinical Genetics* 23, 30-34.
- Dal Zotto, L., Quaderi, N. A., Elliott, R., Lingerfelter, P. A., Carrel, L., Valsecchi, V., Montini, E., Yen, C. H., Chapman, V., Kalcheva, I., *et al.* (1998). The mouse Mid1 gene: implications for the pathogenesis of Opitz syndrome and the evolution of the mammalian pseudoautosomal region. *Human Molecular Genetics* 7, 489-499.
- Davidson, G., Wu, W., Shen, J., Bilic, J., Fenger, U., Stanek, P., Glinka, A., and Niehrs, C. (2005). Casein kinase 1 gamma couples Wnt receptor activation to cytoplasmic signal transduction. *Nature* 438, 867-872.
- Dean, D. C., Bowlus, C. L., and Bourgeois, S. (1987). Cloning and analysis of the promoter region of the human fibronectin gene. *Proceedings of the National Academy of Sciences of the United States of America* 84, 1876-1880.
- Dessay, S., Moizard, M. P., Gilardi, J. L., Opitz, J. M., Middleton-Price, H., Pembrey, M.,
-

Moraine, C., and Briault, S. (2002). FG syndrome: linkage analysis in two families supporting a new gene localization at Xp22.3 [FGS3]. *American Journal of Medical Genetics* 112, 6-11.

Di Como, C. J., and Arndt, K. T. (1996). Nutrients, via the Tor proteins, stimulate the association of Tap42 with type 2A phosphatases. *Genes and Development* 10, 1904-1916.

Dittmar, G. A., Wilkinson, C. R., Jedrzejewski, P. T., and Finley, D. (2002). Role of a ubiquitin-like modification in polarized morphogenesis. *Science* 295, 2442-2446.

Eaton, E. M., and Sealy, L. (2003). Modification of CCAAT/enhancer-binding protein-beta by the small ubiquitin-like modifier (SUMO) family members, SUMO-2 and SUMO-3. *The Journal of Biological Chemistry* 278, 33416-33421.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* 14, 755-763.

Edelmann, L., Pandita, R. K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R. S., Magenis, E., Shprintzen, R. J., and Morrow, B. E. (1999). A common molecular basis for rearrangement disorders on chromosome 22q11. *Human Molecular Genetics* 8, 1157-1167.

Everett, A. D., and Brautigan, D. L. (2002). Developmental expression of alpha4 protein phosphatase regulatory subunit in tissues affected by Opitz syndrome. *Developmental Dynamics* 224, 461-464.

Farndon, P. A., and Donnai, D. (1983). Male to male transmission of the G syndrome. *Clinical Genetics* 24, 446-448.

Farrell, M. J., Stadt, H., Wallis, K. T., Scambler, P., Hixon, R. L., Wolfe, R., Leatherbury, L., and Kirby, M. L. (1999). HIRA, a DiGeorge syndrome candidate gene, is required for cardiac outflow tract septation. *Circulation Research* 84, 127-135.

Fedorova, N. D., Badger, J. H., Robson, G. D., Wortman, J. R., and Nierman, W. C. (2005). Comparative analysis of programmed cell death pathways in filamentous fungi. *BMC Genomics* 6, 177.

Filatov, D. A., and Gerrard, D. T. (2003). High mutation rates in human and ape pseudoautosomal genes. *Gene* 317, 67-77.

Freemont, P. S., Hanson, I. M., and Trowsdale, J. (1991). A novel cysteine-rich sequence motif. *Cell* 64, 483-484.

Fryburg, J. S., Lin, K. Y., and Golden, W. L. (1996). Chromosome 22q11.2 deletion in a boy with Opitz (G/BBB) syndrome. *American Journal of Medical Genetics* 62, 274-275.

Fu, W., Borgaonkar, D. S., Ladewig, P. P., Weaver, J., and Pomerance, H. H. (1976).

Structural aberrations of the long arm of chromosome no. 22. Report of a family with translocation t(11;22) (q25;q11). *Clinical Genetics* 10, 329-336.

Gaudenz, K., Roessler, E., Quaderi, N., Franco, B., Feldman, G., Gasser, D. L., Wittwer, B., Montini, E., Opitz, J. M., Ballabio, A., and Muenke, M. (1998). Opitz G/BBB syndrome in Xp22: mutations in the MID1 gene cluster in the carboxy-terminal domain. *American Journal of Human Genetics* 63, 703-710.

Gilbert, E. F., Viseskul, C., Mossman, H. W., and Opitz, J. M. (1972). The pathologic anatomy of the G-syndrome. *Zeitschrift für Kinderheilkunde* 111, 290-298.

Gill, G. (2005). Something about SUMO inhibits transcription. *Current Opinion in Genetics and Development* 15, 536-541.

Glazko, G. V., Koonin, E. V., and Rogozin, I. B. (2005). Molecular dating: ape bones agree with chicken entrails. *Trends in Genetics* 21, 89-92.

Glover, T. W. (1995). CATCHing a break on 22. *Nature Genetics* 10, 257-258.

Goehring, A. S., Rivers, D. M., and Sprague, G. F., Jr. (2003). Attachment of the ubiquitin-related protein Urm1p to the antioxidant protein Ahp1p. *Eukaryotic Cell* 2, 930-936.

Gonzalez-Sancho, J. M., Brennan, K. R., Castelo-Soccio, L. A., and Brown, A. M. (2004). Wnt proteins induce dishevelled phosphorylation via an LRP5/6- independent mechanism, irrespective of their ability to stabilize beta-catenin. *Molecular and Cellular Biology* 24, 4757-4768.

Graham, J. M., Jr., Wheeler, P., Tackels-Horne, D., Lin, A. E., Hall, B. D., May, M., Short, K. M., Schwartz, C. E., and Cox, T. C. (2003). A new X-linked syndrome with agenesis of the corpus callosum, mental retardation, coloboma, micrognathia, and a mutation in the Alpha 4 gene at Xq13. *American Journal of Medical Genetics Part A* 123, 37-44.

Graham, J. M. J., Superneau, D., Rogers, R. C., Corning, K., Schwartz, C. E., and Dykens, E. M. (1999). Clinical and behavioral characteristics in FG syndrome. *American Journal of Medical Genetics* 85, 470-475.

Granata, A., and Quaderi, N. A. (2003). The Opitz syndrome gene MID1 is essential for establishing asymmetric gene expression in Hensen's node. *Developmental Biology* 258, 397-405.

Greenberg, C. R., and Schraufnagel, D. (1979). The G syndrome: a case report. *American Journal of Medical Genetics* 3, 59-64.

Halal, F., and Farsky, K. (1981). Brief Clinical Report: coloboma hypospadias. *American Journal of Medical Genetics* 8, 53-58.

-
- Harris, D. M., Myrick, T. L., and Rundle, S. J. (1999). The Arabidopsis homolog of yeast TAP42 and mammalian alpha4 binds to the catalytic subunit of protein phosphatase 2A and is induced by chilling. *Plant Physiology* 121, 609-617.
- Harris, M. J., and Juriloff, D. M. (1999). Mini-review: toward understanding mechanisms of genetic neural tube defects in mice. *Teratology* 60, 292-305.
- Hatakeyama, C., Anderson, C. L., Beever, C. L., Penaherrera, M. S., Brown, C. J., and Robinson, W. P. (2004). The dynamics of X-inactivation skewing as women age. *Clinical Genetics* 66, 327-332.
- Hay, E. D. (2005). The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it. *Developmental Dynamics* 233, 706-720.
- Hedera, P., and Gorski, J. L. (2003). Oculo-facio-cardio-dental syndrome: skewed X chromosome inactivation in mother and daughter suggest X-linked dominant inheritance. *American Journal of Medical Genetics Part A* 123, 261-266.
- Hershko, A., Ciechanover, A., Heller, H., Haas, A. L., and Rose, I. A. (1980). Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proceedings of the National Academy of Sciences of the United States of America* 77, 1783-1786.
- Hershko, A., Ciechanover, A., and Rose, I. A. (1981). Identification of the active amino acid residue of the polypeptide of ATP-dependent protein breakdown. *The Journal of Biological Chemistry* 256, 1525-1528.
- Hicks, M. R., Holberton, D. V., Kowalczyk, C., and Woolfson, D. N. (1997). Coiled-coil assembly by peptides with non-heptad sequence motifs. *Folding and Design* 2, 149-158.
- Hochstrasser, M. (2000). Evolution and function of ubiquitin-like protein-conjugation systems. *Nature Cell Biology* 2, E153-157.
- Hogdall, C., Siegel-Bartelt, J., Toi, A., and Ritchie, S. (1989). Prenatal diagnosis of Opitz (BBB) syndrome in the second trimester by ultrasound detection of hypospadias and hypertelorism. *Prenatal Diagnosis* 9, 783-793.
- Horak, I., and Smahel, Z. (1987). The hypertelorism-hypospadias (BBB) syndrome in members of two families. *Acta chirurgiae plasticae* 29, 61-76.
- Howell, L., and Smith, J. D. (1989). G syndrome and its otolaryngologic manifestations. *The Annals of otology, rhinology, and laryngology* 98, 185-190.
- Hsu, W., Zeng, L., and Costantini, F. (1999). Identification of a domain of Axin that binds to the serine/threonine protein phosphatase 2A and a self-binding domain. *The Journal of Biological Chemistry* 274, 3439-3445.
-

-
- Huang, S., and Houghton, P. J. (2001). Mechanisms of resistance to rapamycins. *Drug resistance updates* 4, 378-391.
- Hulo, N., Sigrist, C. J., Le Saux, V., Langendijk-Genevaux, P. S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P., and Bairoch, A. (2004). Recent improvements to the PROSITE database. *Nucleic Acids Research* 32 *Database issue*, D134-137.
- Hynes, P. J., and Fraher, J. P. (2004a). The development of the male genitourinary system. I. The origin of the urorectal septum and the formation of the perineum. *British Journal of Plastic Surgery* 57, 27-36.
- Hynes, P. J., and Fraher, J. P. (2004b). The development of the male genitourinary system: II. The origin and formation of the urethral plate. *British Journal of Plastic Surgery* 57, 112-121.
- Hynes, P. J., and Fraher, J. P. (2004c). The development of the male genitourinary system: III. The formation of the spongiose and glandular urethra. *British Journal of Plastic Surgery* 57, 203-214.
- Ichimura, Y., Kirisako, T., Takao, T., Satomi, Y., Shimonishi, Y., Ishihara, N., Mizushima, N., Tanida, I., Kominami, E., Ohsumi, M., *et al.* (2000). A ubiquitin-like system mediates protein lipidation. *Nature* 408, 488-492.
- Inui, S., Sanjo, H., Maeda, K., Yamamoto, H., Miyamoto, E., and Sakaguchi, N. (1998). Ig receptor binding protein 1 (alpha4) is associated with a rapamycin-sensitive signal transduction in lymphocytes through direct binding to the catalytic subunit of protein phosphatase 2A. *Blood* 92, 539-546.
- Jehee, F. S., Rosenberg, C., Krepischi-Santos, A. C., Kok, F., Knijnenburg, J., Froyen, G., Vianna-Morgante, A. M., Opitz, J. M., and Passos-Bueno, M. R. (2005). An Xq22.3 duplication detected by comparative genomic hybridization microarray (Array-CGH) defines a new locus (FGS5) for FG syndrome. *American Journal of Medical Genetics Part A* 139, 221-226.
- Jerome, L. A., and Papaioannou, V. E. (2001). DiGeorge syndrome phenotype in mice mutant for the T-box gene, *Tbx1*. *Nature Genetics* 27, 286-291.
- Jia, J., Zhang, L., Zhang, Q., Tong, C., Wang, B., Hou, F., Amanai, K., and Jiang, J. (2005). Phosphorylation by double-time/CKIepsilon and CKIalpha targets cubitus interruptus for Slimb/beta-TRCP-mediated proteolytic processing. *Developmental Cell* 9, 819-830.
- Jiang, Y., and Broach, J. R. (1999). Tor proteins and protein phosphatase 2A reciprocally regulate Tap42 in controlling cell growth in yeast. *The EMBO journal* 18, 2782-2792.
- Joazeiro, C. A., Wing, S. S., Huang, H., Levenson, J. D., Hunter, T., and Liu, Y. C. (1999). The tyrosine kinase negative regulator c-Cbl as a RING-type, E2-dependent ubiquitin-
-

protein ligase. *Science* 286, 309-312.

Kallijarvi, J., Lahtinen, U., Hamalainen, R., Lipsanen-Nyman, M., Palvimo, J. J., and Lehesjoki, A. E. (2005). TRIM37 defective in mulibrey nanism is a novel RING finger ubiquitin E3 ligase. *Experimental Cell Research* 308, 146-155.

Kang, P., and Svoboda, K. K. (2005). Epithelial-Mesenchymal Transformation during Craniofacial Development. *Journal of Dental Research* 84, 678-690.

Kasner, J., Gilbert, E. F., and Viseskul, C. (1974). Studies of malformation syndromes VID: the G syndrome. Further observations. *Zeitschrift für Kinderheilkunde* 118, 81-85.

Kimber, W. L., Hsieh, P., Hirotsune, S., Yuva-Paylor, L., Sutherland, H. F., Chen, A., Ruiz-Lozano, P., Hoogstraten-Miller, S. L., Chien, K. R., Paylor, R., *et al.* (1999). Deletion of 150 kb in the minimal DiGeorge/velocardiofacial syndrome critical region in mouse. *Human Molecular Genetics* 8, 2229-2237.

Kimmelman, C. P., and Denny, J. C. (1982). Opitz (G) syndrome. *International Journal Of Pediatric Otorhinolaryngology* 4, 343-347.

Kloeker, S., Reed, R., McConnell, J. L., Chang, D., Tran, K., Westphal, R. S., Law, B. K., Colbran, R. J., Kamoun, M., Campbell, K. S., and Wadzinski, B. E. (2003). Parallel purification of three catalytic subunits of the protein serine/threonine phosphatase 2A family (PP2A(C), PP4(C), and PP6(C)) and analysis of the interaction of PP2A(C) with alpha4 protein. *Protein Expression and Purification* 31, 19-33.

Koegl, M., Hoppe, T., Schlenker, S., Ulrich, H. D., Mayer, T. U., and Jentsch, S. (1999). A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell* 96, 635-644.

Kohn, A. D., and Moon, R. T. (2005). Wnt and calcium signaling: beta-catenin-independent pathways. *Cell Calcium* 38, 439-446.

Koide, A., Bailey, C. W., Huang, X., and Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins. *Journal of Molecular Biology* 284, 1141-1151.

Kong, M., Fox, C. J., Mu, J., Solt, L., Xu, A., Cinalli, R. M., Birnbaum, M. J., Lindsten, T., and Thompson, C. B. (2004). The PP2A-associated protein alpha4 is an essential inhibitor of apoptosis. *Science* 306, 695-698.

Kurczynski, T. W., Micale, M. A., Assad, S., Gaba, C. G., and French, B. F. (1998). Opitz G/BBB syndrome in a male child with an unbalanced der(3)t(3;14)(q29;q11.2),-14 karyotype. *American Journal of Human Genetics* 63 (Supplement), A110.

Kuwahara, K., Matsuo, T., Nomura, J., Igarashi, H., Kimoto, M., Inui, S., and Sakaguchi, N. (1994). Identification of a 52-kDa molecule (p52) coprecipitated with the Ig receptor-

related MB-1 protein that is inducibly phosphorylated by the stimulation with phorbol myristate acetate. *Journal of Immunology* 152, 2742-2752.

Lacassie, Y., and Arriaza, M. I. (1996). Opitz GBBB syndrome and the 22q11.2 deletion [letter]. *American Journal of Medical Genetics* 62, 318.

Landry, J. R., and Mager, D. L. (2002). Widely spaced alternative promoters, conserved between human and rodent, control expression of the Opitz syndrome gene MID1. *Genomics* 80, 499-508.

Landry, J. R., Rouhi, A., Medstrand, P., and Mager, D. L. (2002). The opitz syndrome gene *mid1* is transcribed from a human endogenous retroviral promoter. *Molecular Biology and Evolution* 19, 1934-1942.

Law, B. K. (2005). Rapamycin: An anti-cancer immunosuppressant? *Critical Reviews in Oncology/Hematology* 56, 47-60.

Leichtman, L. G., Werner, A., Bass, W. T., Smith, D., and Brothman, A. R. (1991). Apparent Opitz BBBG syndrome with a partial duplication of 5p. *American Journal of Medical Genetics* 40, 173-176.

Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Doerks, T., Schultz, J., Ponting, C. P., and Bork, P. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Research* 32 Database issue, D142-144.

Li, X., Yost, H. J., Virshup, D. M., and Seeling, J. M. (2001). Protein phosphatase 2A and its B56 regulatory subunit inhibit Wnt signaling in *Xenopus*. *The EMBO journal* 20, 4122-4131.

Liang, J., and Slingerland, J. M. (2003). Multiple roles of the PI3K/PKB (Akt) pathway in cell cycle progression. *Cell Cycle* 2, 339-345.

Lindsay, E. A. (2001). Chromosomal microdeletions: dissecting del22q11 syndrome. *Nature Reviews in Genetics* 2, 858-868.

Lindsay, E. A., Botta, A., Jurecic, V., Carattini-Rivera, S., Cheah, Y. C., Rosenblatt, H. M., Bradley, A., and Baldini, A. (1999). Congenital heart disease in mice deficient for the DiGeorge syndrome region. *Nature* 401, 379-383.

Lindsay, E. A., Vitelli, F., Su, H., Morishima, M., Huynh, T., Pramparo, T., Jurecic, V., Ogunrinu, G., Sutherland, H. F., Scambler, P. J., *et al.* (2001). *Tbx1* haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature* 410, 97-101.

Little, J. R., and Opitz, J. M. (1971). The G syndrome. *American Journal of Diseases of Children* 121, 505-507.

-
- Liu, J., Prickett, T. D., Elliott, E., Meroni, G., and Brautigan, D. L. (2001). Phosphorylation and microtubule association of the Opitz syndrome protein mid-1 is regulated by protein phosphatase 2A via binding to the regulatory subunit alpha 4. *Proceedings of the National Academy of Sciences of the United States of America* *98*, 6650-6655.
- Lorick, K. L., Jensen, J. P., Fang, S., Ong, A. M., Hatakeyama, S., and Weissman, A. M. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proceedings of the National Academy of Sciences of the United States of America* *96*, 11364-11369.
- Lutskiy, M. I., Jones, L. N., Rosen, F. S., and Remold-O'Donnell, E. (2002). An Alu-mediated deletion at Xp11.23 leading to Wiskott-Aldrich syndrome. *Human Genetics* *110*, 515-519.
- Malakhova, O. A., Yan, M., Malakhov, M. P., Yuan, Y., Ritchie, K. J., Kim, K. I., Peterson, L. F., Shuai, K., and Zhang, D. E. (2003). Protein ISGylation modulates the JAK-STAT signaling pathway. *Genes and Development* *17*, 455-460.
- Mallery, D. L., Vandenberg, C. J., and Hiom, K. (2002). Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. *The EMBO journal* *21*, 6755-6762.
- Massiah, M. A., Simmons, B. N., Short, K. M., and Cox, T. C. (2006). Solution structure of the RBCC/TRIM B-box1 domain of human MID1: B-box with a RING. *Journal of Molecular Biology* *352*, 532-545.
- Masters, S. L., Yao, S., Willson, T. A., Zhang, J. G., Palmer, K. R., Smith, B. J., Babon, J. J., Nicola, N. A., Norton, R. S., and Nicholson, S. E. (2006). The SPRY domain of SSB-2 adopts a novel fold that presents conserved Par-4-binding residues. *Nature Structural and Molecular Biology* *13*, 77-84.
- Mastroiacovo, P., Dolk, H., and Siffel, C. (2003). *World Atlas of Birth Defects: Hypospadias, Vol 21: Hypospadias, 2nd edn* (Geneva, Switzerland: World Health Organization).
- McDonald-McGinn, D. M., Driscoll, D. A., Bason, L., Christensen, K., Lynch, D., Sullivan, K., Canning, D., Zavod, W., Quinn, N., and Rome, J. (1995). Autosomal dominant "Opitz" GBBB syndrome due to a 22q11.2 deletion. *American Journal of Medical Genetics* *59*, 103-113.
- Merscher, S., Funke, B., Epstein, J. A., Heyer, J., Puech, A., Lu, M. M., Xavier, R. J., Demay, M. B., Russell, R. G., Factor, S., *et al.* (2001). TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. *Cell* *104*, 619-629.
- Michaelis, E., and Mortier, W. (1972). Association of hypertelorism and hypospadias - the BBB-syndrome. *Helvetica paediatrica acta* *27*, 575-581.
-

- Miller, P. R., Bernstein, R. M., Pathak, A., and Decancq, H. G., Jr. (1977). Hypertelorism-hypospadias syndrome with a laryngotracheoesophageal cleft. *The Journal of Pediatrics* *90*, 157-158.
- Mirkes, P. E. (2002). 2001 Warkany lecture: to die or not to die, the role of apoptosis in normal and abnormal mammalian development. *Teratology* *65*, 228-239.
- Mizushima, N., Noda, T., Yoshimori, T., Tanaka, Y., Ishii, T., George, M. D., Klionsky, D. J., Ohsumi, M., and Ohsumi, Y. (1998). A protein conjugation system essential for autophagy. *Nature* *395*, 395-398.
- Mo, R., Kim, J. H., Zhang, J., Chiang, C., Hui, C. C., and Kim, P. C. (2001). Anorectal malformations caused by defects in sonic hedgehog signaling. *The American Journal of Pathology* *159*, 765-774.
- Mori, C., Nakamura, N., Okamoto, Y., Osawa, M., and Shiota, K. (1994). Cytochemical identification of programmed cell death in the fusing fetal mouse palate by specific labelling of DNA fragmentation. *Anatomy and Embryology* *190*, 21-28.
- Mulhern, T. D., Booker, G. W., and Cosgrove, L. (1998). A third fibronectin-type-III domain in the insulin-family receptors. *Trends in Biochemical Sciences* *23*, 465-466.
- Mulholland, D. J., Dedhar, S., Coetzee, G. A., and Nelson, C. C. (2005). Interaction of nuclear receptors with the Wnt/beta-catenin/Tcf signaling axis: Wnt you like to know? *Endocrine Reviews* *26*, 898-915.
- Mulholland, D. J., Dedhar, S., Wu, H., and Nelson, C. C. (2006). PTEN and GSK3beta: key regulators of progression to androgen-independent prostate cancer. *Oncogene* *25*, 329-337.
- Murata, K., Wu, J., and Brautigan, D. L. (1997). B cell receptor-associated protein alpha 4 displays rapamycin-sensitive binding directly to the catalytic subunit of protein phosphatase 2A. *Proceedings of the National Academy of Sciences of the United States of America* *94*, 10624-10629.
- Nakamura, M., Xavier, R. M., Tsunematsu, T., and Tanigawa, Y. (1995). Molecular cloning and characterization of a cDNA encoding monoclonal nonspecific suppressor factor. *Proceedings of the National Academy of Sciences of the United States of America* *92*, 3463-3467.
- Nanahoshi, M., Nishiuma, T., Tsujishita, Y., Hara, K., Inui, S., Sakaguchi, N., and Yonezawa, K. (1998). Regulation of protein phosphatase 2A catalytic activity by alpha4 protein and its yeast homolog Tap42. *Biochemical and Biophysical Research Communications* *251*, 520-526.
- Nanahoshi, M., Tsujishita, Y., Tokunaga, C., Inui, S., Sakaguchi, N., Hara, K., and Yonezawa, K. (1999). Alpha4 protein as a common regulator of type 2A-related

serine/threonine protein phosphatases. *Febs Letters* 446, 108-112.

Nanci, A. (2003). *Ten Cate's Oral Histology: Development, Structure, and Function*, 6th edn (Missouri: Mosby).

Nievelstein, R. A., van der Werff, J. F., Verbeek, F. J., Valk, J., and Vermeij-Keers, C. (1998). Normal and abnormal embryonic development of the anorectum in human embryos. *Teratology* 57, 70-78.

Niikura, T., Hashimoto, Y., Tajima, H., Ishizaka, M., Yamagishi, Y., Kawasumi, M., Nawa, M., Terashita, K., Aiso, S., and Nishimoto, I. (2003). A tripartite motif protein TRIM11 binds and destabilizes Humanin, a neuroprotective peptide against Alzheimer's disease-relevant insults. *The European Journal of Neuroscience* 17, 1150-1158.

O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989). Evidence that the leucine zipper is a coiled coil. *Science* 243, 538-542.

Opitz, J. M. (1987). G syndrome (hypertelorism with esophageal abnormality and hypospadias, or hypospadias-dysphagia, or Opitz-Frias or Opitz-G syndrome)--perspective in 1987 and bibliography. *American Journal of Medical Genetics* 28, 275-285.

Opitz, J. M., Frias, J. L., Gutenberger, J. E., and Pellet, J. R. (1969b). The G syndrome of multiple congenital anomalies. *Birth Defects Original Article Series* V, 95-101.

Opitz, J. M., Smith, D. W., and Summit, R. L. (1965). Hypertelorism and hypospadias: a newly recognized hereditary malformation syndrome. *Journal of Pediatrics* 67 (Suppl), 968.

Opitz, J. M., Summit, R. L., and Smith, D. W. (1969a). The BBB syndrome: Familial telecanthus with associated congenital anomalies. *Birth Defects Original Article Series* V, 86-94.

Palmer, S., Perry, J., Kipling, D., and Ashworth, A. (1997). A gene spans the pseudoautosomal boundary in mice. *Proceedings of the National Academy of Sciences of the United States of America* 94, 12030-12035.

Parisian, S., and Toomey, K. (1978). Features of the G (Opitz-Frias) and BBB (hypospadias hypertelorism) syndrome in one family - are they a single disorder? (Abstract). *American Journal of Human Genetics Supplementary*, 62A.

Pedersen, I. L., Mikkelsen, M., and Oster, J. (1976). The G syndrome. A four-generation family study. *Human Heredity* 26, 66-71.

Penaloza, C., Lin, L., Lockshin, R. A., and Zakeri, Z. (2006). Cell death in development: shaping the embryo. *Histochemistry and Cell Biology* 126, 149-158.

Peng, H., Feldman, I., and Rauscher, F. J., 3rd (2002). Hetero-oligomerization among the TIF family of RBCC/TRIM domain-containing nuclear cofactors: a potential mechanism for regulating the switch between coactivation and corepression. *Journal of Molecular Biology* 320, 629-644.

Perry, J., Feather, S., Smith, A., Palmer, S., and Ashworth, A. (1998). The human FXY gene is located within Xp22.3: implications for evolution of the mammalian X chromosome. *Human Molecular Genetics* 7, 299-305.

Perry, J., Short, K. M., Romer, J. T., Swift, S., Cox, T. C., and Ashworth, A. (1999). FXY2/MID2, a gene related to the X-linked Opitz syndrome gene FXY/MID1, maps to Xq22 and encodes a FNIII domain-containing protein that associates with microtubules. *Genomics* 62, 385-394.

Person, A. D., Klewer, S. E., and Runyan, R. B. (2005). Cell biology of cardiac cushion development. *International Review of Cytology* 243, 287-335.

Piluso, G., Carella, M., D'Avanzo, M., Santinelli, R., Carrano, E. M., D'Avanzo, A., D'Adamo, A. P., Gasparini, P., and Nigro, V. (2003). Genetic heterogeneity of FG syndrome: a fourth locus (FGS4) maps to Xp11.4-p11.3 in an Italian family. *Human Genetics* 112, 124-130.

Pinson, L., Auge, J., Audollent, S., Mattei, G., Etchevers, H., Gigarel, N., Razavi, F., Lacombe, D., Odent, S., Le Merrer, M., *et al.* (2004). Embryonic expression of the human MID1 gene and its mutations in Opitz syndrome. *Journal of Medical Genetics* 41, 381-386.

Pla, P., Moore, R., Morali, O. G., Grille, S., Martinozzi, S., Delmas, V., and Larue, L. (2001). Cadherins in neural crest cell development and transformation. *Journal of Cellular Physiology* 189, 121-132.

Poelmann, R. E., and Gittenberger-de Groot, A. C. (2005). Apoptosis as an instrument in cardiovascular development. *Birth defects research Part C, Embryo today* 75, 305-313.

Prickett, T. D., and Brautigan, D. L. (2004). Overlapping binding sites in protein phosphatase 2A for association with regulatory A and alpha-4 (mTap42) subunits. *The Journal of Biological Chemistry* 279, 38912-38920.

Puech, A., Saint-Jore, B., Merscher, S., Russell, R. G., Cherif, D., Sirotkin, H., Xu, H., Factor, S., Kucherlapati, R., Skoultschi, A. I., *et al.* (2000). Normal cardiovascular development in mice deficient for 16 genes in 550 kb of the velocardiofacial/DiGeorge syndrome region. *Proceedings of the National Academy of Sciences of the United States of America* 97, 10090-10095.

Quaderi, N. A., Schweiger, S., Gaudenz, K., Franco, B., Rugarli, E. I., Berger, W., Feldman, G. J., Volta, M., Andolfi, G., Gilgenkrantz, S., *et al.* (1997). Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. *Nature Genetics* 17, 285-291.

Raasi, S., Schmidtke, G., and Groettrup, M. (2001). The ubiquitin-like protein FAT10 forms covalent conjugates and induces apoptosis. *The Journal of Biological Chemistry* 276, 35334-35343.

Ratcliffe, M. J., Itoh, K., and Sokol, S. Y. (2000). A positive role for the PP2A catalytic subunit in Wnt signal transduction. *The Journal of Biological Chemistry* 275, 35680-35683.

Reddy, B. A., and Etkin, L. D. (1991). A unique bipartite cysteine-histidine motif defines a subfamily of potential zinc-finger proteins. *Nucleic Acids Research* 19, 6330.

Reddy, B. A., Kloc, M., and Etkin, L. (1991). The cloning and characterization of a maternally expressed novel zinc finger nuclear phosphoprotein (xnf7) in *Xenopus laevis*. *Developmental Biology* 148, 107-116.

Reed, M. H., Shokeir, M. H., and Macpherson, R. I. (1975). The hypertelorism-hypospadias syndrome. *Journal of the Canadian Association of Radiologists* 26, 240-248.

Reymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzi, L., Riganelli, D., Zanaria, E., Messali, S., Cainarca, S., *et al.* (2001). The tripartite motif family identifies cell compartments. *The EMBO journal* 20, 2140-2151.

Richman, J. M., Fu, K. K., Cox, L. L., Sibbons, J. P., and Cox, T. C. (2002). Isolation and characterisation of the chick orthologue of the Opitz syndrome gene, *Mid1*, supports a conserved role in vertebrate development. *International Journal of Developmental Biology* 46, 441-448.

Robin, N. H., Feldman, G. J., Aronson, A. L., Mitchell, H. F., Weksberg, R., Leonard, C. O., Burton, B. K., Josephson, K. D., Laxova, R., and Aleck, K. A. (1995). Opitz syndrome is genetically heterogeneous, with one locus on Xp22, and a second locus on 22q11.2. *Nature Genetics* 11, 459-461.

Robin, N. H., Opitz, J. M., and Muenke, M. (1996). Opitz G/BBB syndrome: clinical comparisons of families linked to Xp22 and 22q, and a review of the literature. *American Journal of Medical Genetics* 62, 305-317.

Sabatini, D. M., Erdjument-Bromage, H., Lui, M., Tempst, P., and Snyder, S. H. (1994). RAFT1: a mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs. *Cell* 78, 35-43.

Saitta, S. C., McGrath, J. M., Mensch, H., Shaikh, T. H., Zackai, E. H., and Emanuel, B. S. (1999). A 22q11.2 deletion that excludes UFD1L and CDC45L in a patient with conotruncal and craniofacial defects. *American Journal of Human Genetics* 65, 562-566.

Say, B., and Carpenter, N. J. (1987). Skeletal dysplasia in an infant with hypertelorism, hypospadias, developmental delay, and a complex chromosomal translocation. *The Southern Medical Journal* 80, 1190-1192.

- Schubert, J., Schmidt, R., and Raupach, H. W. (1990). New findings explaining the mode of action in prevention of facial clefting and first clinical experience. *Journal of Cranio-Maxillo-Facial Surgery* 18, 343-347.
- Schweiger, S., Foerster, J., Lehmann, T., Suckow, V., Muller, Y. A., Walter, G., Davies, T., Porter, H., van Bokhoven, H., Lunt, P. W., *et al.* (1999). The Opitz syndrome gene product, MID1, associates with microtubules. *Proceedings of the National Academy of Sciences of the United States of America* 96, 2794-2799.
- Sedano, H. O., and Gorlin, R. J. (1988). Opitz oculo-genital-laryngeal syndrome (Opitz BBB/G compound syndrome) [letter]. *American Journal of Medical Genetics* 30, 847-849, 851.
- Seto, M. H., Liu, H. L., Zajchowski, D. A., and Whitlow, M. (1999). Protein fold analysis of the B30.2-like domain. *Proteins* 35, 235-249.
- Shackelford, J., and Pagano, J. S. (2005). Targeting of host-cell ubiquitin pathways by viruses. *Essays in Biochemistry* 41, 139-156.
- Shook, D., and Keller, R. (2003). Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development. *Mechanisms of Development* 120, 1351-1383.
- Short, K. M., and Cox, T. C. (2006). Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. *The Journal of Biological Chemistry* 281, 8970-8980.
- Short, K. M., Hopwood, B., Yi, Z., and Cox, T. C. (2002). MID1 and MID2 homo- and heterodimerise to tether the rapamycin-sensitive PP2A regulatory subunit, alpha 4, to microtubules: implications for the clinical variability of X-linked Opitz GBBB syndrome and other developmental disorders. *BMC Cell Biology* 3, 1.
- Skorstengaard, K., Jensen, M. S., Sahl, P., Petersen, T. E., and Magnusson, S. (1986). Complete primary structure of bovine plasma fibronectin. *European Journal of Biochemistry* 161, 441-453.
- So, J., Suckow, V., Kijas, Z., Kalscheuer, V., Moser, B., Winter, J., Baars, M., Firth, H., Lunt, P. W., Hamel, B., *et al.* (2005). Mild phenotypes in a series of patients with Opitz GBBB syndrome with MID1 mutations. *American Journal of Medical Genetics* 132, 1-7.
- Sontag, E., Nunbhakdi Craig, V., Bloom, G. S., and Mumby, M. C. (1995). A novel pool of protein phosphatase 2A is associated with microtubules and is regulated during the cell cycle. *Journal of Cell Biology* 128, 1131-1144.
- Spencer, J. A., Eliazer, S., Ilaria, R. L., Jr., Richardson, J. A., and Olson, E. N. (2000). Regulation of microtubule dynamics and myogenic differentiation by MURF, a striated muscle RING-finger protein. *Journal of Cell Biology* 150, 771-784.

-
- Stevens, C. A., and Wilroy, R. S., Jr. (1988). The telecanthus-hypospadias syndrome. *Journal of Medical Genetics* 25, 536-542.
- Stockinger, A., Eger, A., Wolf, J., Beug, H., and Foisner, R. (2001). E-cadherin regulates cell growth by modulating proliferation-dependent beta-catenin transcriptional activity. *Journal of Cell Biology* 154, 1185-1196.
- Stoll, C., Geraudel, A., Berland, H., Roth, M. P., and Dott, B. (1985). Male-to-male transmission of the hypertelorism-hypospadias (BBB) syndrome. *American Journal of Medical Genetics* 20, 221-225.
- Stremlau, M., Owens, C. M., Perron, M. J., Kiessling, M., Autissier, P., Sodroski, J., and Goff, S. P. (2004). The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* 427, 848-853.
- Suraweera, N., Robinson, J., Volikos, E., Guenther, T., Talbot, I., Tomlinson, I., and Silver, A. (2006). Mutations within Wnt pathway genes in sporadic colorectal cancers and cell lines. *International Journal of Cancer* 199, 1837-1842.
- Sutherland, H. F., Wadey, R., McKie, J. M., Taylor, C., Atif, U., Johnstone, K. A., Halford, S., Kim, U. J., Goodship, J., Baldini, A., and Scambler, P. J. (1996). Identification of a novel transcript disrupted by a balanced translocation associated with DiGeorge syndrome. *American Journal of Human Genetics* 59, 23-31.
- Tada, M., Concha, M. L., and Heisenberg, C. P. (2002). Non-canonical Wnt signalling and regulation of gastrulation movements. *Seminars in Cell and Developmental Biology* 13, 251-260.
- Takada, R., Hijikata, H., Kondoh, H., and Takada, S. (2005). Analysis of combinatorial effects of Wnts and Frizzleds on beta-catenin/armadillo stabilization and Dishevelled phosphorylation. *Genes to Cells* 10, 919-928.
- Thien, C. B., and Langdon, W. Y. (2001). Cbl: many adaptations to regulate protein tyrosine kinases. *Nature Reviews in Molecular Cell Biology* 2, 294-307.
- Tolmie, J. L., Coutts, N., and Drainer, I. K. (1987). Congenital anal anomalies in two families with the Opitz G syndrome. *Journal of Medical Genetics* 24, 688-691.
- Trockenbacher, A., Suckow, V., Foerster, J., Winter, J., Krauss, S., Ropers, H. H., Schneider, R., and Schweiger, S. (2001). MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation. *Nature Genetics* 29, 287-294.
- Truong, K., and Ikura, M. (2002). Identification and characterization of subfamily-specific signatures in a large protein superfamily by a hidden Markov model approach. *BMC Bioinformatics* 3, 1.
-

- Uddin, R. K., Zhang, Y., Siu, V. M., Fan, Y. S., O'Reilly, R. L., Rao, J., and Singh, S. M. (2006). Breakpoint Associated with a novel 2.3 Mb deletion in the VCFS region of 22q11 and the role of Alu (SINE) in recurring microdeletions. *BMC Medical Genetics* 7, 18.
- Urano, T., Saito, T., Tsukui, T., Fujita, M., Hosoi, T., Muramatsu, M., Ouchi, Y., and Inoue, S. (2002). Efp targets 14-3-3 sigma for proteolysis and promotes breast tumour growth. *Nature* 417, 871-875.
- Urioste, M., Arroyo, I., Villa, A., Lorda-Sanchez, I., Barrio, R., Lopez-Cuesta, M. J., and Rueda, J. (1995). Distal deletion of chromosome 13 in a child with the "Opitz" GBBB syndrome. *American Journal of Medical Genetics* 59, 114-122.
- van Biervliet, J. P., and van Hemel, J. O. (1975). Familial occurrence of the G syndrome. *Clinical Genetics* 7, 238-244.
- Verloes, A., David, A., Odent, S., Toutain, A., André, M. J., Lucas, J., and Le Marec, B. (1995). Opitz GBBB syndrome: chromosomal evidence of an X-linked form. *American Journal of Medical Genetics* 59, 123-128.
- Verloes, A., Le Merrer, M., and Briard, M. L. (1989). BBBG syndrome or Opitz syndrome: new family. *American Journal of Medical Genetics* 34, 313-316.
- Vernet, C., Boretto, J., Mattei, M. G., Takahashi, M., Jack, L. J., Mather, I. H., Rouquier, S., and Pontarotti, P. (1993). Evolutionary study of multigenic families mapping close to the human MHC class I region. *Journal of Molecular Evolution* 37, 600-612.
- Vichi, A., Payne, D. M., Pacheco-Rodriguez, G., Moss, J., and Vaughan, M. (2005). E3 ubiquitin ligase activity of the trifunctional ARD1 (ADP-ribosylation factor domain protein 1). *Proceedings of the National Academy of Sciences of the United States of America* 102, 1945-1950.
- Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 Å resolution. *Journal of Molecular Biology* 194, 531-544.
- Vincan, E. (2004). Frizzled/WNT signalling: the insidious promoter of tumour growth and progression. *Frontiers in Bioscience* 9, 1023-1034.
- Webb, S., Qayyum, S. R., Anderson, R. H., Lamers, W. H., and Richardson, M. K. (2003). Septation and separation within the outflow tract of the developing heart. *Journal of Anatomy* 202, 327-342.
- Wei, X., Senders, C., Owiti, G. O., Liu, X., Wei, Z. N., Dillard-Telm, L., McClure, H. M., and Hendrickx, A. G. (2000). The origin and development of the upper lateral incisor and premaxilla in normal and cleft lip/palate monkeys induced with cyclophosphamide. *The Cleft Palate-Craniofacial Journal* 37, 571-583.

- Welchman, R. L., Gordon, C., and Mayer, R. J. (2005). Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nature Reviews in Molecular Cell Biology* 6, 599-609.
- Widelitz, R. (2005). Wnt signaling through canonical and non-canonical pathways: recent progress. *Growth Factors* 23, 111-116.
- Williams, C. A., and Frias, J. L. (1987). Apparent G syndrome presenting as neck and upper limb dystonia and severe gastroesophageal reflux. *American Journal of Medical Genetics* 28, 297-302.
- Wilson, D. I., Burn, J., Scambler, P., and Goodship, J. (1993). DiGeorge syndrome: part of CATCH 22. *Journal of Medical Genetics* 30, 852-856.
- Wilson, G. N., and Oliver, W. J. (1988). Further delineation of the G syndrome: a manageable genetic cause of infantile dysphagia. *Journal of Medical Genetics* 25, 157-163.
- Winter, J., Lehmann, T., Suckow, V., Kijas, Z., Kulozik, A., Kalscheuer, V., Hamel, B., Devriendt, K., Opitz, J., Lenzner, S., *et al.* (2003). Duplication of the MID1 first exon in a patient with Opitz G/BBB syndrome. *Human Genetics* 24, 24.
- Woo, J. S., Imm, J. H., Min, C. K., Kim, K. J., Cha, S. S., and Oh, B. H. (2006). Structural and functional insights into the B30.2/SPRY domain. *The EMBO journal* 25, 1353-1363.
- Xirodimas, D. P., Saville, M. K., Bourdon, J. C., Hay, R. T., and Lane, D. P. (2004). Mdm2-mediated NEDD8 conjugation of p53 inhibits its transcriptional activity. *Cell* 118, 83-97.
- Yadav, M. K., Leman, L. J., Price, D. J., Brooks Iii, C. L., Stout, C. D., and Ghadiri, M. R. (2006). Coiled coils at the edge of configurational heterogeneity. Structural analyses of parallel and antiparallel homotetrameric coiled coils reveal configurational sensitivity to a single solvent-exposed amino Acid substitution. *Biochemistry* 45, 4463-4473.
- Yamagishi, H., Garg, V., Matsuoka, R., Thomas, T., and Srivastava, D. (1999). A molecular pathway revealing a genetic basis for human cardiac and craniofacial defects. *Science* 283, 1158-1161.
- Yan, Q., Liu, J. P., and Li, D. W. (2006). Apoptosis in lens development and pathology. *Differentiation* 74, 195-211.
- Yang, P., Khoury, M. J., Stewart, W. F., Beaty, T. H., Chee, E., Beatty, J. C., Diamond, E. L., and Gordis, L. (1994). Comparative epidemiology of selected midline congenital abnormalities. *Genetic Epidemiology* 11, 141-154.
- Yap, M. W., Nisole, S., and Stoye, J. P. (2005). A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. *Current Biology* 15, 73-78.

Yook, J. I., Li, X. Y., Ota, I., Fearon, E. R., and Weiss, S. J. (2005). Wnt-dependent regulation of the E-cadherin repressor snail. *The Journal of Biological Chemistry* *280*, 11740-11748.

Young, I. D., Dalgleish, R., MacKay, E. H., and MacFadyen, U. M. (1988). Discordant expression of the G syndrome in monozygotic twins. *American Journal of Medical Genetics* *29*, 863-869.

Zheng, Y., and Jiang, Y. (2005). The yeast phosphotyrosyl phosphatase activator is part of the Tap42-phosphatase complexes. *Molecular Biology of the Cell* *16*, 2119-2127.

Zolnierowicz, S. (2000). Type 2A protein phosphatase, the complex regulator of numerous signaling pathways. *Biochemical Pharmacology* *60*, 1225-1235.

Zumbrunn, J., Kinoshita, K., Hyman, A. A., and Nathke, I. S. (2001). Binding of the adenomatous polyposis coli protein to microtubules increases microtubule stability and is regulated by GSK3 beta phosphorylation. *Current Biology* *11*, 44-49.

Appendices

***FXY2/MID2*, a Gene Related to the X-Linked Opitz Syndrome Gene *FXY/MID1*, Maps to Xq22 and Encodes a FNIII Domain-Containing Protein That Associates with Microtubules**

Jo Perry,* Kieran M. Short,† Justyna T. Romer,* Sally Swift,* Timothy C. Cox,† and Alan Ashworth*¹

*Section of Gene Function and Regulation, Chester Beatty Laboratories, The Institute of Cancer Research, Fulham Road, London SW3 6JB, United Kingdom; and †Department of Genetics, University of Adelaide, North Terrace, Adelaide, South Australia 5005, Australia

Received August 25, 1999; accepted October 25, 1999

Opitz G/BBB syndrome (OS) is a genetically heterogeneous disorder with an X-linked locus and an autosomal locus linked to 22q11.2. OS affects multiple organ systems with often variable severity even between siblings. The clinical features, which include hypertelorism, cleft lip and palate, defects of cardiac septation, hypospadias, and anorectal anomalies, indicate an underlying disturbance of the developing ventral midline of the embryo. The gene responsible for X-linked OS, *FXY/MID1*, is located on the short arm of the human X chromosome within Xp22.3 and encodes a protein with both an RBCC (RING finger, B-box, coiled coil) and a B30.2 domain. The *Fxy* gene in mice is also located on the X chromosome but spans the pseudoautosomal boundary in this species. Here we describe a gene closely related to *FXY/MID1*, called *FXY2*, which also maps to the X chromosome within Xq22. The mouse *Fxy2* gene is located on the distal part of the mouse X chromosome within a region syntenic to Xq22. Analysis of genes flanking both *FXY/MID1* and *FXY2* (as well as their counterparts in mouse) suggests that these regions may have arisen as a result of an intrachromosomal duplication on an ancestral X chromosome. We have also identified in both *FXY2* and *FXY/MID1* proteins a conserved fibronectin type III domain located between the RBCC and B30.2 domains that has implications for understanding protein function. The *FXY/MID1* protein has previously been shown to colocalize with microtubules, and here we show that the *FXY2* protein similarly associates with microtubules in a manner that is dependent on the carboxy-terminal B30.2 domain. © 1999 Academic Press

INTRODUCTION

Members of the burgeoning RING finger family of proteins are involved in a wide range of cellular pro-

cesses including signal transduction, apoptosis, oncogenesis, and development and are found in plants, bacteria, animals, and viruses (Freemont, 1993; Reddy *et al.*, 1992; Saurin *et al.*, 1996). The family is characterized by the presence of a zinc-binding RING finger domain, often in association with other domains that are also thought to be involved in mediating protein-protein interactions (reviewed in Borden, 1998; Freemont, 1993; Saurin *et al.*, 1996). We, and others, have previously described a novel RING finger gene, *FXY* (also known as *MID1*), which is located on the human X chromosome within Xp22.3 (Quaderi *et al.*, 1997; Perry *et al.*, 1998). In mice, *Fxy* spans the pseudoautosomal boundary on the X chromosome (Palmer *et al.*, 1997). The pseudoautosomal region is a region of sequence identity between the X and the Y chromosomes of eutherian mammals required for correct chromosomal pairing and segregation during male meiosis (Burgoyne, 1982). Hence the 5' portion of the gene is X-unique while the 3' portion is located on the X and Y chromosomes (Palmer *et al.*, 1997). We have shown recently that, as a consequence of its pseudoautosomal location, the 3' portion of the mouse gene is evolving at an accelerated rate (Perry and Ashworth, 1999).

The RING finger of *FXY/MID1* forms part of a larger tripartite motif that also comprises two zinc-binding B-box domains and a leucine-rich coiled coil, which characterizes members of the RBCC (RING-B-box-coiled coil) subfamily of the RING finger proteins (Reddy *et al.*, 1992). In addition, a B30.2 domain is present at the C-terminus of *FXY/MID1* (Henry *et al.*, 1998). This domain is a conserved region of approximately 170 amino acids associated with several different protein domains, including the immunoglobulin folds of butyrophilin (Jack and Mather, 1990) and other RBCC domains in proteins such as RFP (Ret finger protein) (Takahashi and Cooper, 1987) and Xnf7 (Reddy *et al.*, 1991).

Mutations in the *FXY/MID1* gene have recently been found in approximately 50% of patients with the diagnosis of Opitz G/BBB syndrome (OS), a disorder of the primary midline developmental field (Quaderi *et al.*,

1991).

¹ To whom correspondence should be addressed. Telephone: 0171-352-8133. Fax: 0171-352-3299. E-mail: alana@icr.ac.uk.



1997; Gaudenz *et al.*, 1998; Allen *et al.*, in preparation). The genetics of OS has had a somewhat unsettled history. The disorder was originally reported as two separate syndromes, BBB syndrome (Opitz *et al.*, 1969b) and G syndrome (Opitz *et al.*, 1969a), each suggested to follow an X-linked pattern of inheritance. However, reports of both syndromes segregating in one family and of male to male transmission led to its somewhat controversial reassignment as a single autosomal entity with a "male preponderance" (MIM 145410) (McKusick *et al.*, 1995). Definitive evidence of genetic heterogeneity was finally provided by way of linkage data for both autosomal (chromosome 22q11.2) and X-linked (Xp22) forms of OS (Robin *et al.*, 1995). Notably, all of the identified mutations in *FXY/MID1* result in disruption of, or alteration in, the carboxy-terminal region of the *FXY/MID1* protein (Quaderi *et al.*, 1997; Gaudenz *et al.*, 1998; Allen *et al.*, in preparation). Recently, the *FXY/MID1* protein has been shown to colocalize with microtubules, suggesting that *FXY/MID1* may be a microtubule-associated protein (Schweiger *et al.*, 1999). Intriguingly, many tested mutant *FXY/MID1* proteins fail to associate with microtubules both *in vivo* and following overexpression in transfected cell lines; these mutant proteins instead form cytoplasmic clumps (Schweiger *et al.*, 1999; Allen *et al.*, in preparation).

Here we report the identification of a gene showing strong sequence similarity to *FXY/MID1*, which we have named *FXY2*. Unlike *FXY/MID1*, which maps to Xp22.3, *FXY2* is located on the long arm of the human X chromosome, within Xq22. Analysis of other genes within Xp22 and Xq22 and their syntenic regions in the mouse suggests that there may have been a duplication event in the recent history of the mammalian X chromosome. In addition, we have identified a novel FNIII domain located between the RBCC and the B30.2 domains in both *FXY2* and *FXY/MID1* and show that *FXY2*, like *FXY/MID1*, associates with microtubules.

MATERIALS AND METHODS

Cloning the human and mouse *FXY2* genes. A PAC containing part of the human *FXY2* gene was initially isolated by a sequence similarity search of the National Center for Biotechnology Information (NCBI) Htgs (high-throughput genomic sequence) database using BLASTN. The coding portion of the gene was used in further database searches to identify two IMAGE cDNA clones, one of which (IMAGE Clone 360565) contained the entire open reading frame and part of the 3' untranslated region (UTR) of the gene, which was then completely sequenced.

To isolate the mouse *Fxy2* cDNA, oligos derived from the rat and human *Fxy2* coding sequences (5'-CCTCTGCGTTCTTAATGCC-3' and 5'-AGAGTTCTTCTCGGATAGAT-3') were used to amplify a 1.22-kb fragment from mouse (C57BL/6) kidney cDNA. RACE was used to isolate the 3' portion of the gene; nested oligos derived from the mouse cDNA sequence (5'-CATCAACCAAGCTGAG-CATATCC-3' and 5'-CTAGAGAAAAGAACTATTGGAGG-3') as well as a SMART PCR oligo (Clontech) were used to amplify a 1.75-kb fragment from mouse kidney oligo(dT)-primed cDNA.

Automated DNA sequencing was performed by cycle sequencing with Applied Biosystems Big Dye chemistry and analyzed on an Applied Biosystems 377 sequencer. Sequence analysis and searches

were performed using either the GCG suite of programs at the HGMP (Hinxton, UK) (<http://www.hgmp.mrc.ac.uk/>) or BLAST at NCBI (<http://www.ncbi.nlm.nih.gov/>). The amino acid sequence of *FXY2* was also analyzed using a variety of Web-based programs including Pfam, ProDom, and Blocks.

Genetic mapping. The mouse *Fxy2* gene was mapped using a PCR-based assay to type mice from the European Collaborative Interspecific Backcross (EUCIB) (The European Backcross Collaborative Group, 1994). Oligos derived from exon 6 and exon 7 of the human *FXY2* cDNA were used in a PCR to amplify a 455-bp band from both B6 and *Mus spretus* genomic DNA. Digestion of the PCR product from B6 mice with *RsaI* resulted in three bands of 245, 150, and 60 bp, while digestion of the *M. spretus* product resulted in two bands of 305 and 150 bp. Animals obtained from the backcross of (C57BL/6 × *M. spretus*)F₁ to *M. spretus* were typed for either the B6 or the *M. spretus* variant and mapped using the Mbx database at the HGMP (Hinxton, UK) (<http://www.hgmp.mrc.ac.uk/>) (The European Backcross Collaborative Group, 1994).

Northern analysis. A commercial Northern filter containing mRNA from a variety of adult mouse tissues (Clontech) was hybridized using as the probe a full-length human *FXY2* cDNA, radiolabeled using a Gigaprime labeling kit (Geneworks, Thebarton, South Australia). Following washing according to the manufacturer's specifications, the filter was exposed to Hyperfilm (Amersham) for 4 days with two intensifier screens.

Generation of green fluorescent protein (GFP)-*FXY2* fusion constructs. The complete reading frame of *FXY2* was fused to the carboxy-terminus of the green fluorescent protein, using the following strategy: The *FXY2* cDNA was enzymatically amplified using *Pfu* polymerase (Stratagene) and the primers 5'*FXY*fusion (5'-GTGAATTCCTGAAGATGGAAACACTGGAGTC-3') and *FXY2*ex10R(Ec) (5'-GTGAATTCGGAAGAAGACTGCTGAAGTACTAGAG-3'); the product was digested with *EcoRI* and ligated into appropriately digested pEGFP-C2 (Clontech). The correct orientation of the insert was determined by direct sequencing. The resultant construct was named pEGFP-*FXY2*.ORF. The GFP-*FXY/MID1* control construct was similarly generated using the 5'*FXY*fusion and G232-2 (5'-GTGAATTCGGGACACTTCTGGTGAG-3') primers. To generate a GFP-*FXY2* construct that was devoid of the B30.2 domain, pEGFP-*FXY2*.ORF was digested with *BamHI* to excise the 3' ~640 bp, and the remaining *FXY2* sequence was religated with the cut vector end. The resultant clone (called pEGFP-*FXY2*delCTD) terminated the *FXY2* reading frame at amino acid residue 490.

Transfection and image analysis of GFP-*FXY2* constructs. Two picomoles of each plasmid DNA prepared using Qiagen purification systems was transfected by electroporation into Cos1 cells and grown on coverslips in Dulbecco's modified Eagle's medium supplemented with 10% FBS. Cells were fixed with 3.5% paraformaldehyde 24 to 36 h posttransfection. A standard immunofluorescence protocol was followed with the exception that 1× PEM buffer (0.1 M Pipes, 5 mM EGTA, 2 mM MgCl₂, pH 7.0) was used in place of PBS buffer to protect the microtubules. Cells were then visualized under appropriate wavelength light on an Olympus AX70 microscope, and images were captured using a cooled CCD (Photometrics CE200A) camera. Microtubule staining was achieved using an anti- α -tubulin antibody (Sigma) and a rhodamine-labeled secondary anti-mouse antibody (Boehringer Mannheim). Nuclear staining was achieved using the DNA-specific stain, 4',6-diamidino-2'-phenylindole dihydrochloride (DAPI; Sigma).

RESULTS

Isolation and Chromosomal Localization of the Human *FXY2* Gene

Sequence similarity searches with the human *FXY/MID1* nucleotide sequence using BLASTN identified a P1 artificial chromosome (PAC), 191P20, in the Htgs

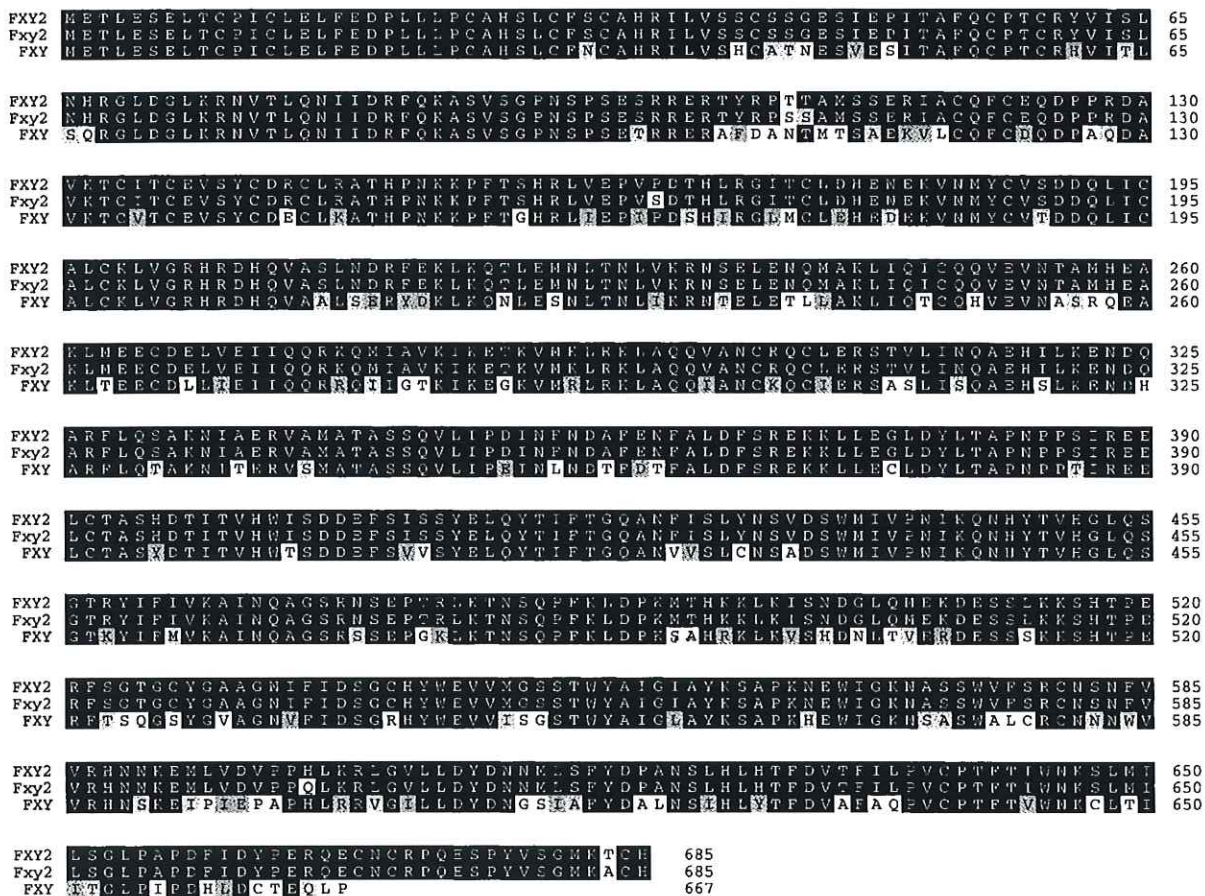


FIG. 1. Amino acid alignment of the human FXY/MID1 and FXY2 and mouse Fxy2 proteins. Sequences were aligned using the GCG program Pileup at the HGMP (<http://www.hgmp.mrc.ac.uk>). Black boxes indicate identical amino acid residues, dark gray boxes indicate amino acid residues whose comparison value is greater than or equal to the average nonidentical comparison value in the blosum62.cmp scoring matrix; light gray boxes indicate similar amino acid residues whose comparison scores are less than this value but are ≥ 1 .

database (<http://www.ncbi.nlm.nih.gov/BLAST/>) that had been partially sequenced and that contained exons of a gene similar to *FXY/MID1*. Closer examination of the PAC DNA sequence revealed that it contained 7 exons showing a high level of identity to exons 4 to 10 of the human *FXY/MID1* gene (Quaderi *et al.*, 1997; Perry *et al.*, 1998). The sequence of these new exons was used to search the GenBank expressed sequence database (dbEST) using BLASTN, and two IMAGE clones were identified (IMAGE Clones 360565 and 1977538). One of these clones (Clone 360565, GenBank Accession No. AA016125) from a human cDNA library, when sequenced, was found to contain the entire coding region and part of the 3'UTR of the gene, which we have called *FXY2*. The *FXY2* cDNA is 2249 bp long with an open reading frame of 2055 bp coding for a 685-amino-acid protein, 18 amino acids longer than the FXY/MID1 protein, largely due to a C-terminal extension (Fig. 1). The human FXY2 and FXY/MID1 proteins show very strong sequence similarity (Fig. 1) with 76% identity/83% similarity seen at the amino acid level and 71% identity at the nucleotide level. This full-

length cDNA clone was then used to determine the remainder of the *FXY2* gene structure. Overall, the *FXY2* gene consists of 9 coding exons like that of *FXY/MID1*, with the positions of splice junctions and exon size similar in both genes (Fig. 2A).

The predicted amino acid sequence of FXY2 was analyzed using a variety of programs including Pfam (<http://www.sanger.ac.uk/Software/Pfam/>), ProDom (<http://protein.toulouse.inra.fr/prodom.html>), and Motif (<http://www.motif.genome.ad.jp>). This analysis predicts, as expected based on the extensive identity with FXY/MID1, that FXY2 also encodes an RBCC or tripartite domain at its amino-terminus (consisting of a RING finger motif, two B-box motifs, and a coiled coil motif) (Reddy *et al.*, 1992; Saurin *et al.*, 1996) and a B30.2 domain at its carboxy-terminus (Fig. 2B). Notably, however, this analysis also identified a sequence located between the RBCC and the B30.2 domains with significant identity to a fibronectin type III repeat (FNIII). An FNIII signature is characterized by four elements, and the presence of significant matches to each of these elements is considered a strong indication

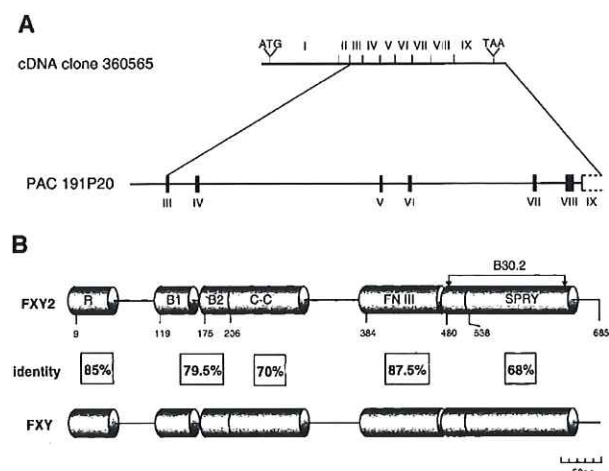


FIG. 2. (A) Genomic structure of *FXY2*. Schematic representation of the human *FXY2* cDNA and the position of the seven exons located on the PAC clone 191P20. Exons are indicated on the genomic DNA by black boxes and Roman numerals. (B) Comparative structure of the *FXY2* and *FXY* proteins. The constituent domains and their individual percentage of amino acid identity are indicated. R, RING finger domain; B1/B2, Bbox domains; C-C, coiled coil domain; FNIII, fibronectin type III domain; SPRY, SPRY domain; B30.2, B30.2 domain. The amino acid positions at which these domains start are marked.

of a functional FNIII motif (<http://www.biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS/PRINTS.html>). Interestingly, these analyses also identified the same domain in *FXY/MID1*, which was not described in previous reports (Quaderi *et al.*, 1997; Perry *et al.*, 1998; Gaudenz *et al.*, 1998; Schweiger *et al.*, 1999). Additional indirect support for the existence of this FNIII domain is the 87.5% amino acid identity in this region between *FXY2* and the corresponding *FXY/MID1* sequence (Fig. 2B); this is higher than any other recognized domain within the two proteins. Furthermore, one Opitz syndrome patient harbors a single amino acid deletion (methionine 438) that resides within this predicted FNIII domain, and this deletion disrupts the normal intracellular distribution of *FXY/MID1* (Schweiger *et al.*, 1999). This methionine residue is conserved in *FXY2*. Overall, the *FXY2* protein is most similar to two other RING finger proteins, a nuclear phosphoprotein from *Xenopus laevis* (*xnf7*) involved in dorsal-ventral pattern formation during embryo development (El-Hodiri *et al.*, 1997; Reddy *et al.*, 1991) and a protein similar to human estrogen responsive finger protein (EFP) (GenBank Accession No. BAA13398).

The PAC clone containing the 3' end of the *FXY2* gene (Clone 191P20; GenBank Accession No. AL034399) has subsequently been completely sequenced, and details of the clone can be found on the Sanger Centre Web page (<http://www.sanger.ac.uk/HGP/ChrX>). Clone 191P20 has been linked to a PAC contig (chromosome Xctg215) by the Sanger Centre and has been mapped to the long arm of the human X chromosome near Xq22.3 by fluorescence *In situ* hybridization. Analysis of other markers on the PAC

clone 191P20 identified an EST (GenBank Accession No. R43678) that has been mapped on the GeneBridge 4 radiation hybrid mapping panel (Gyapay *et al.*, 1996). This places R43678 267.38 cR from the top of the X chromosome linkage group (http://carbon.wi.mit.edu:8000/cgi-bin/contig/phys_map). R43678 has also been placed on the GeneMap '98 map of the X chromosome by the International Radiation Hybrid Mapping Consortium at NCBI (<http://www.ncbi.nlm.nih.gov/genemap98/>), and R43678 is located between two genes, thyroxine-binding globulin (TGB) and collagen type IV α -6 chain (COL4A6). In addition, a sequence-tagged site (STS) (Whitehead Accession No. WI-14337, GenBank Accession No. R43664) also located on the PAC clone 191P20 close to *FXY2* has been mapped on the GeneBridge 4 radiation hybrid mapping panel. R43664 is located 267.79 cR from the top of the X chromosome linkage group and colocalizes with a second STS (Whitehead Accession No. WI-7912, GenBank Accession No. G06158). G06158 is located within Xq22 between DXS7891 and DXS1210 on an integrated STS and YAC physical, genetic, and transcript map spanning Xq21.3 to Xq23-q24, placing *FXY2* within this interval (Srivastava *et al.*, 1999) (Fig. 3A).

Isolation and Localization of the Mouse *Fxy2* Gene

A search of the EST database (dbEST) with the entire human *FXY2* cDNA identified an EST from *Rattus norvegicus* (GenBank Accession No. AA925587). The clone contained sequence that showed 92% identity with the first 283 bp of the human *FXY2* cDNA, indicating that this was likely to be a partial cDNA of the rat *FXY2* gene. Oligonucleotides were designed to the rat sequence in an attempt to clone the mouse *Fxy2* gene. Using a combination of RT-PCR on mouse kidney poly(A)⁺ RNA and rapid amplification of cDNA ends, a full-length mouse *Fxy2* cDNA was isolated and then sequenced. The entire *Fxy2* cDNA is 2622 bp and, like the human cDNA, codes for a 685-amino-acid protein. Sequence analysis shows that the mouse and human *Fxy2* genes are highly conserved with 94% identity at the nucleotide level over the open reading frame and 99% identity at the amino acid level. A comparison of the mouse *Fxy* and *Fxy2* genes in laboratory mouse strains (e.g., C57BL/6) is not appropriate due to the fact that the 3' end of the *Fxy* gene is subject to accelerated divergence as a result of its pseudoautosomal location (Perry and Ashworth, 1999). However, the *M. spretus* *Fxy* gene is entirely X-unique and therefore provides a better comparison. The C57BL/6 (B6) *Fxy2* gene and the *Fxy* gene from *M. spretus* are very similar with 70% identity at the nucleotide level and 82% similarity and 75% identity at the amino acid level.

We mapped the mouse *Fxy2* gene by taking advantage of a restriction fragment length variant between C57BL/6 (B6) and *M. spretus* within intron 6 of the gene. The PCR-based assay was used to type individual animals from the EUCIB. Seventy-four animals from a backcross of the (B6 X *M. spretus*)F₁ generation to *M.*

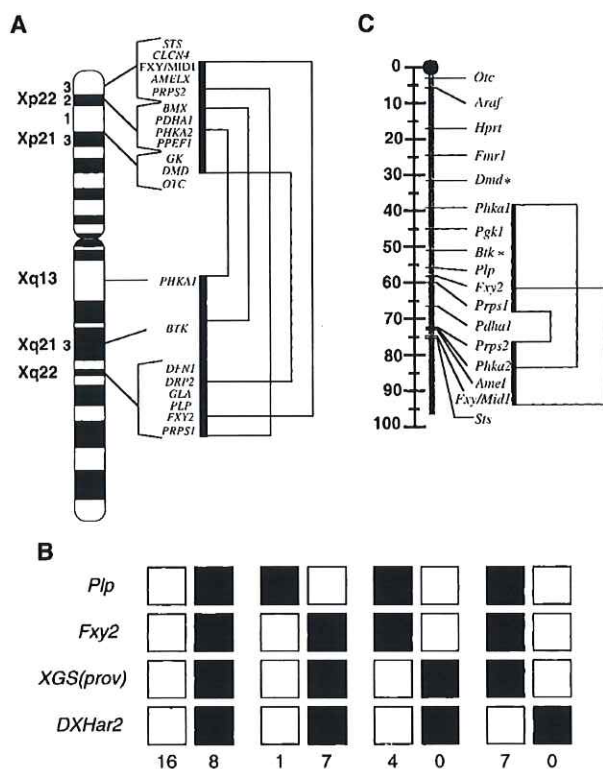


FIG. 3. Mapping of *FXY/MID1* and *FXY2* genes in humans and mice suggests an ancient intrachromosomal duplication of the X chromosome. **(A)** Schematic representation of the positions of *FXY/MID1* and *FXY2* on the human X chromosome. *FXY2* is located within Xq22 while the related gene, *FXY/MID1*, is located in Xp22. A potential duplication between these two regions is indicated by a thick black line, and duplicated genes within these regions are linked by a thin black line. **(B)** Genetic mapping of *Fxy2* in mice. Haplotype analysis of the *Fxy2* gene in the EUCIB interspecific backcross. Parental and recombinant X chromosomes are shown for 43 of the mice analyzed. Shaded boxes indicate the presence of the B6 variant, and open boxes indicate the *Mus spretus* variant. The markers tested are listed to the left side of the figure. **(C)** Location of the *Fxy2* gene on the mouse X chromosome. Schematic representation of the position of *Fxy2* on the mouse X chromosome. Genes similarly duplicated on the mouse X chromosome are indicated and linked by a thin black line. The genes related to *Btk* and *Dmd* (indicated by an asterisk), *Bmx* and *Drp2*, respectively, have been cloned but have not been mapped.

spretus were typed for either the B6 or the *M. spretus* variant, and the results were analyzed using the MbX database at the HGMP. Haplotype analysis showed that *Fxy2* is located on the mouse X chromosome between the myelin proteolipid protein gene (*Plp*) and a secondary marker, XGS(prov), which colocalizes with the microsatellite marker *DXMit196* (Fig. 3B).

Fxy2 Expression

The pattern of *Fxy2* expression in adult mouse tissue was examined by Northern analysis (MTN blot; Clontech) using a radiolabeled full-length human *FXY2* cDNA as a probe. This analysis detected an approximately 8-kb *Fxy2* transcript in low abundance in brain

and lung, with even lower levels in heart, liver, and kidney (data not shown). A relatively low level of expression could have perhaps been expected considering that only five ESTs for human *FXY2* (one each from adult retina, adult brain, and thyroid and two from germ cell tumors) and no mouse *Fxy2* ESTs were present in dbEST, at the time of submission of this article.

Subcellular Localization of *FXY2*

Recently, MID1 was shown to be associated with the microtubule network by colocalization of endogenous *FXY/MID1* as well as GFP-*FXY/MID1* fusion proteins with β -tubulin. To investigate the subcellular distribution of *FXY2*, we have similarly fused it to the carboxy-terminus of GFP (pEGFP-*FXY2*.ORF) and examined its localization in transfected Cos1 cells. The GFP-*FXY2* fusion protein showed a subcellular localization that resembled the GFP-*FXY/MID1* fibrous pattern observed in parallel experiments (Fig. 4A). This similarity was confirmed following staining of transfected cells with a monoclonal anti- α -tubulin antibody (Sigma), revealing that GFP fluorescence colocalized with the tubulin fibers. However, in most instances, GFP-*FXY2* fluorescence was not observed over all tubulin-positive fibers and frequently not to fibers near the periphery of the cell (see Figs. 4A and 4B). This was never observed in the parallel experiments using GFP-*FXY/MID1*.

Given that microtubule association of *FXY/MID1* is disrupted by mutation or deletion of the B30.2 domain (Schweiger *et al.*, 1999), we chose to investigate whether deletion of the B30.2 domain of *FXY2* would result in a similar dissociation from the microtubules with the formation of cytoplasmic clumps. A plasmid, pEGFP-*FXYdelCTD*, was constructed to express a fragment of *FXY2*, amino acid residues 1–490, fused to GFP. Following transfection into Cos1 cells, a distinctive "clumping" of the protein was observed that was reminiscent of that seen with MID1 mutant isoforms that resulted in disease presentation (Schweiger *et al.*, 1999) (Fig. 4C).

DISCUSSION

We have isolated a novel gene, *FXY2*, that shares 71% identity with the Xp22-linked Opitz syndrome gene, *FXY/MID1*, at the nucleotide level and which is located on the X chromosome at Xq22 (Quaderi *et al.*, 1997; Perry *et al.*, 1998). During the preparation of this article, a gene apparently identical to *FXY2*, named *MID2*, was described (Buchner *et al.*, 1999). The human *MID2* clone contains an additional 30 amino acids inserted at a splice junction. This may represent an alternatively spliced form of *FXY2/MID2* but was not present in any of our human or mouse PCR products or cDNA clones.

Several studies have recently made progress toward an integrated physical, genetic, and transcript map of

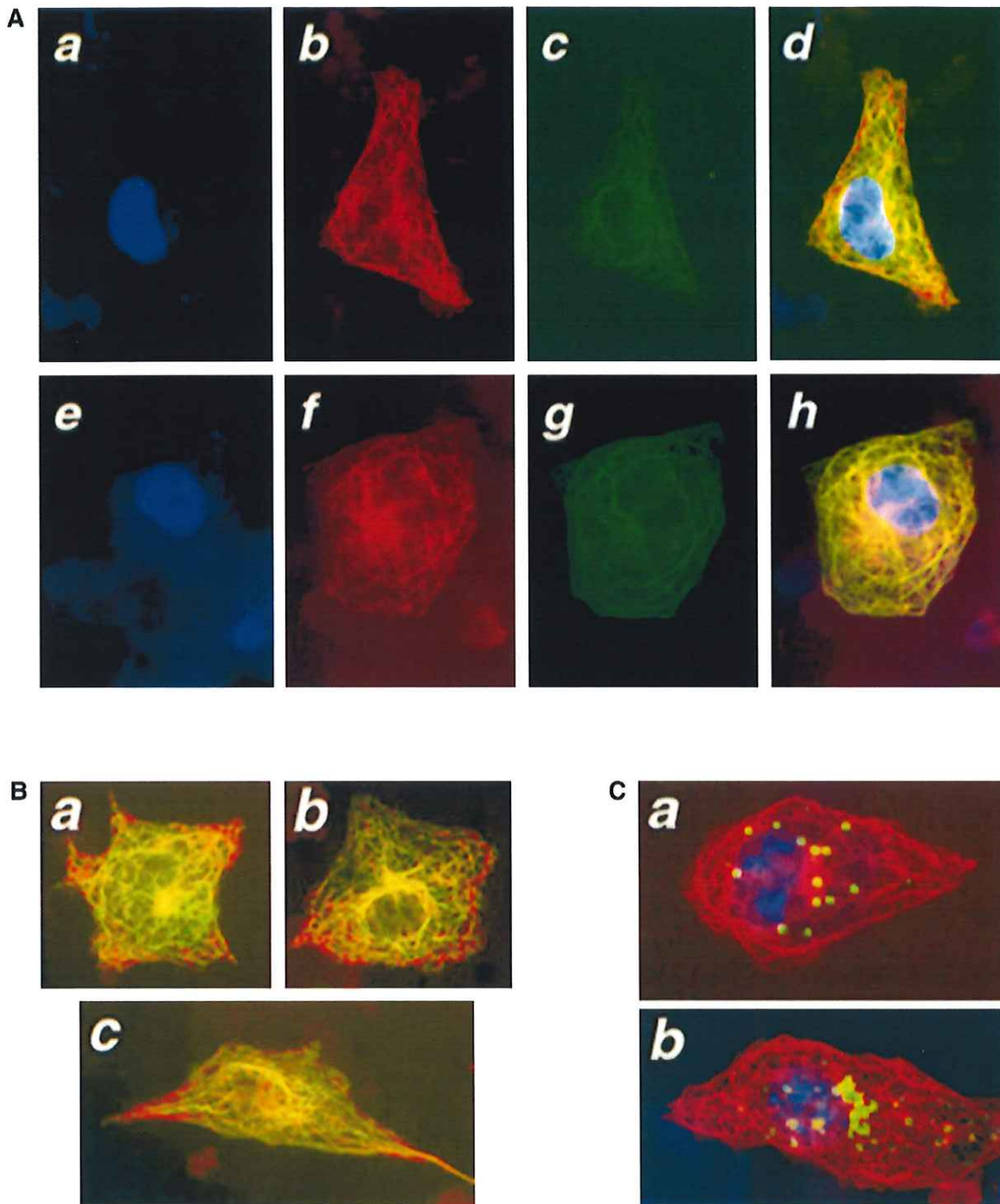


FIG. 4. Subcellular localization of FXY2. (A) Detection of full-length GFP-FXY2 and GFP-FXY/MID1 fusion proteins in transiently transfected Cos1 cells. (a and e) DAPI stain for DNA (blue); (b and f) anti- α -tubulin staining (red); (c and g) GFP fluorescence (green). (d) Overlay of a, b, and c showing the partial colocalization of the GFP-FXY2 fusion protein with α -tubulin (colocalization indicated by yellow color). The yellow color (i.e., colocalization) is less clearly seen in this image as the GFP fluorescence has been enhanced and the intensity of tubulin staining reduced to allow easier visualization of each. (h) Overlay of e, f, and g showing the complete colocalization of the GFP-FXY/MID1 fusion protein with α -tubulin (colocalization indicated by yellow color). These experiments were done in parallel and under identical conditions. (B) (a, b,) and c) Further examples of near-complete colocalization of GFP-FXY2 fusion protein and α -tubulin in Cos1 cells. Note the poor coverage of GFP-FXY2 near the periphery of cells. Yellow color represents the overlay of red (α -tubulin) and green (GFP-FXY2) colors. (C) (a and b) Distribution of the GFP-FXY2delCTD fusion protein in Cos1 cells. Blue, red, and green colors indicate staining for DAPI, α -tubulin, and GFP, respectively, as described above for A.

an approximately 20-Mb region from Xq21.3 to q24 (Srivastava *et al.*, 1999, and references therein), which has facilitated our fine-mapping of the *FXY2* gene. Several human genetic disorders map to this region of the human X chromosome including congenital perceptive deafness DFN2 (OMIM 304500), X-linked megalocornea (OMIM 309300), Arts syndrome, which is characterized by X-linked mental retardation with deafness, loss of vision, and ataxia (OMIM 301835) (<http://www.ncbi.nlm.nih.gov/Omim/>), and several nonspecific X-linked mental retardation syndromes (MRX23,35,47,54) (des Portes *et al.*, 1997; Gregg *et al.*, 1996; Gu *et al.*, 1996; Jemaa *et al.*, 1999). There are also several complex mental retardation disorders mapping to this chromosomal region. These are associated with seizures and hypogammaglobulinemia (Chudley *et al.*, 1999), with microcephaly and short stature (MRXS9) (Shrimpton *et al.*, 1999), and with growth hormone deficiency (OMIM 300123), as well as a female-restricted epilepsy associated with mental retardation, EFMR (OMIM 300088) (<http://www.ncbi.nlm.nih.gov/Omim/>). The involvement of *FXY/MID1* in disease and its striking level of sequence similarity with *FXY2* support the consideration of *FXY2* as a candidate for all of these genetic disorders.

The *FXY2* gene described here and *FXY1/MID1* (Quaderi *et al.*, 1997; Perry *et al.*, 1998) map to the two arms of the X chromosome, in the regions Xq22 and Xp22, respectively. Interestingly, using a BLAST search to compare gene sequences from the two arms of the X chromosome, we have identified several genes that map to the Xp22 region and have related genes on the long arm of the X chromosome, within or close to Xq22 (Fig. 3A). These are the phosphoribosyl pyrophosphate synthetase-2 (*PRPS2*) gene on Xp22.2–p22.3 and *PRPS1* on Xq22–q24 (Becker *et al.*, 1990); the Duchenne muscular dystrophy (*DMD*) gene on Xp21.2 and dystrophin-related protein 2 (*DRP2*) on Xq22 (Koenig *et al.*, 1987; Roberts *et al.*, 1996); the nonreceptor tyrosine kinase (*BMX*) gene on Xp22.2 and the related Bruton's tyrosine kinase (*BTK*) gene on Xq21.3–q22 (Tamagnone *et al.*, 1994; Vorechovsky *et al.*, 1994); the phosphorylase kinase α -2 (*PHKA2*) gene on Xp22.2–p22.1; and the phosphorylase kinase α -1 (*PHKA1*) gene on Xq12–q13 (Davidson *et al.*, 1992; Gossen *et al.*, 1995). These similarities between the Xq22 and the Xp22 regions suggest that an intrachromosomal duplication, followed by changes in the gene order, may have taken place between the long and the short arms of the X chromosome. An alternate explanation is that each of these related genes arose by separate events of gene duplication, followed by gene divergence. This latter possibility seems unlikely. The X chromosome harbors several examples of gene clusters, which usually arise from gene duplications; however, most occupy discrete chromosomal regions, with members separated by less than a few megabases: the *GABRE*, *GABRA3*, and *GABRA4* genes on Xq28 (Wilke *et al.*, 1997); *MTM1* and *MTMR1* on Xq28 (Kioschis *et al.*, 1998); *SCML1* and *SCML2* on Xp22 (Montini *et al.*, 1999); *GPC3* and

GPC4 on Xq26.1 (Huber *et al.*, 1998); *RCP* and *GCP* on Xq28 (Reyniers *et al.*, 1995); and *COL4A5* and *COL4A6* on Xq21.3–q22 (Srivastava *et al.*, 1995). Although other gene families were found with members separated by larger distances (*MYCL2* on Xq22–q23 and *MYCL3* on Xq27.2 (Redolfi *et al.*, 1999), and one gene family was present in three separate clusters, on Xq28, q26, and p21.3 (De Plaen *et al.*, 1999), these findings support our conjecture that the similarities between the Xp22 and the Xq22 regions are more likely to have arisen by a large-scale intrachromosomal duplication between X chromosome arms.

If such a rearrangement occurred before the mouse and human lineages separated, the human genes within the duplicated region should have mouse orthologs present in two copies on the mouse X chromosome. We therefore searched the NCBI Human/Mouse Homology Maps to find the chromosomal positions of mouse orthologs of the human genes present on both Xp22 and Xq22 (Fig. 3C). In addition to the *Fxy1/Fxy2* gene pair, the *Phka1/Phka2* and the *Prps1/Prps2* pairs are present on the mouse X chromosome. This observation indicates that the suggested intrachromosomal duplication would have taken place before divergence of the mouse and human lineages. The *Dmd* and *Btk* genes are also on the mouse X chromosome, but the related genes *Drp2* and *Bmx*, even though they have been cloned in mice, have not yet been mapped. The intrachromosomal duplication hypothesis described above, together with Ohno's (1967) law, predicts that these two genes will also map to the mouse X chromosome.

The *FXY2* and *FXY/MID1* proteins are highly conserved with 83% similarity seen at the amino acid level, and the proteins are particularly conserved throughout the RBCC, FNIII, and B30.2 domains (Fig. 2B). The biological significance of RBCC domains has not been fully elucidated but some insight into their function has come from mutation analysis. The Ret finger protein (RFP) has been shown to homomultimerize through its coiled-coil domain, and the B-box domains are necessary for this interaction. Furthermore, mutations in the RING and B-box domains of this protein affect subcellular compartmentalization (Cao *et al.*, 1997). Interestingly, RFP and a second RBCC protein, PML, which is disrupted in acute promyelocytic leukemia as a result of a chromosomal translocation, have been shown to interact and contribute to a larger nuclear multiprotein complex *in vivo*, the formation of which is dependent on the presence of their RBCC domains (Boddy *et al.*, 1997). In contrast to the nuclear location of PML and RFP (and many other RBCC proteins), our GFP fusion studies demonstrate that *FXY2* colocalizes with the microtubule network, which is consistent with the finding for *FXY/MID1*. However, although in some cells near-complete colocalization with α -tubulin was observed for GFP-*FXY2*, in many other cells, the GFP fusion protein was visible only over some tubulin fibers and frequently not on fibers near the periphery of the cell (see Figs. 4A and

4B). This was never observed in parallel experiments using GFP-FXY/MID1 and may suggest subtle differences in the microtubule association capabilities between the two proteins. However, this will require further clarification before firm conclusions can be drawn.

A second type of domain, the B30.2 domain, is located in the C-terminal portion of the FXY2 protein. B30.2 domains can be found in many RING finger proteins as well as apparently unrelated proteins such as the milk-fat globule membrane glycoprotein, butyrophilin, and the secreted cytolytic toxins, stonustoxin α , stonustoxin β , and verrucotoxin α (reviewed in Henry *et al.*, 1998). Currently, no function has been assigned to the B30.2 domain. However, its importance is apparent by the concentration of mutations in this region of FXY/MID1 in OS (Quaderi *et al.*, 1997; Gaudenz *et al.*, 1998; Allen *et al.*, in preparation). Moreover, mutations of the B30.2 domain in another protein, pyrin/marenostrin, have been correlated with the development of the autosomal recessive disease, familial Mediterranean fever (MIM 249100) (The French FMF Consortium, 1997; The International FMF Consortium, 1997). The absence or mutation of this domain in FXY/MID1 has been shown to disrupt the normal subcellular distribution of the FXY/MID1 protein, both in cells from OS patients and in transfection experiments with GFP fusion constructs, such that clumps are seen within the cytoplasm. To investigate whether the FXY2 C-terminal domain may be similarly important for correct subcellular localization, we generated a construct in which this domain was removed. As with FXY/MID1, deletion of this domain in FXY2 totally disrupted the protein's normal fibrous pattern, whereas the tubulin network was not disrupted, giving a speckled or clumped appearance that was reminiscent of that seen in Opitz syndrome patients with FXY/MID1 truncations (Schweiger *et al.*, 1999; Allen *et al.*, in preparation). These data support the conclusion that the B30.2 domain is essential for correct subcellular localization and function of this subfamily of RBCC proteins. Although the biological functions of FXY2 and FXY/MID1 are still unknown, recent data from the analysis of B30.2 domains may provide clues to a possible function. These data suggest that B30.2 domains represent a subclass of the SPRY domain family, a recently characterized domain found primarily in proteins involved in the regulation of intracellular signaling (Schultz *et al.*, 1998; Seto *et al.*, 1999). Both FXY2 and FXY/MID1 have recognizable SPRY domains.

In addition to the RBCC and B30.2 (and SPRY) domains, both FXY and FXY2 encode a single FNIII domain. FNIII domains are independently folding protein modules of approximately 100 amino acids in length. These domains are thought to mediate specific protein-protein interactions in both intracellular and extracellular compartments. FNIII domains have been identified in hundreds of different proteins including cell adhesion molecules, cell surface hormone and cytokine receptors, and chaperonins (Hunkapiller and

Hood, 1989; Campbell and Spitzfaden, 1994; Plaxco *et al.*, 1996; Spitzfaden *et al.*, 1997). The FNIII domain, while not well conserved at the amino acid level (8–20% identity) among the multitude of proteins in which it is found (Bork and Doolittle, 1992), does form very similar tertiary structures best described as a β -sandwich formed by two antiparallel β -sheets (Goll *et al.*, 1998). The FNIII domains of FXY/MID1 and FXY2 are most similar to those of the striated muscle structural protein, M-protein, found in fast-muscle fibers (Carlsson *et al.*, 1990), which belongs to a family of large muscle proteins that also include Titin (the largest known protein at 3.7 MDa) and Twitchin (a *Caenorhabditis elegans* protein with a relative molecular mass of 750 kDa). Further analysis of the function of the RBCC, FNIII, and B30.2 domains in FXY/MID1 and FXY2, along with the identification of interacting proteins, will hopefully shed light on their exact cellular roles.

ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust and by an R. Douglas Wright Award (997706) (to T.C.C.) from the National Health and Medical Research Council of Australia. We thank the HGMP, UK, for provision of DNA and computing resources, and Quenten Schwarz, for assistance with the GFP analysis.

REFERENCES

- Allen, L. R., Suthers, G. K., Cox, L. L., Goodwin, B., Haan, E., and Cox, T. C. Predominance of nonsense mutations in familial and isolated Australian patients with Opitz GBBB syndrome. In preparation.
- Becker, M. A., Heidler, S. A., Bell, G. I., Seino, S., Le Beau, M. M., Westbrook, C. A., Neuman, W., Shapiro, L. J., Mohandas, T. K., Roessler, B. J., *et al.* (1990). Cloning of cDNAs for human phosphoribosylpyrophosphate synthetases 1 and 2 and X chromosome localization of PRPS1 and PRPS2 genes. *Genomics* **8**: 555–561.
- Boddy, M. N., Duprez, E., Borden, K. L., and Freemont, P. S. (1997). Surface residue mutations of the PML RING finger domain alter the formation of nuclear matrix-associated PML bodies. *J. Cell Sci.* **110**: 2197–2205.
- Borden, K. L. (1998). RING fingers and B-boxes: Zinc-binding protein-protein interaction domains. *Biochem. Cell. Biol.* **76**: 351–358.
- Bork, P., and Doolittle, R. F. (1992). Proposed acquisition of an animal protein domain by bacteria. *Proc. Natl. Acad. Sci. USA* **89**: 8990–8994.
- Buchner, G., Montini, E., Andolfi, G., Quaderi, N., Cainarca, S., Messali, S., Bassi, M. T., Ballabio, A., Meroni, G., and Franco, B. (1999). MID2, a homologue of the Opitz syndrome gene MID1: Similarities in subcellular localization and differences in expression during development. *Hum. Mol. Genet.* **8**: 1397–1407.
- Burgoyne, P. S. (1982). Genetic homology and crossing over in the X and Y chromosomes of mammals. *Hum. Genet.* **61**: 85–90.
- Campbell, I. D., and Spitzfaden, C. (1994). Building proteins with fibronectin type III modules. *Structure* **2**: 333–337.
- Cao, T., Borden, K. L., Freemont, P. S., and Etkin, L. D. (1997). Involvement of the rfp tripartite motif in protein-protein interactions and subcellular distribution. *J. Cell Sci.* **110**: 1563–1571.
- Carlsson, E., Grove, B. K., Wallimann, T., Eppenberger, H. M., and Thornell, L. E. (1990). Myofibrillar M-band proteins in rat skeletal muscles during development. *Histochemistry* **95**: 27–35.

- Chudley, A. E., Tackels, D. C., Lubs, H. A., Arena, J. F., Stoerber, W. P., Kovnats, S., Stevenson, R. E., and Schwartz, C. E. (1999). X-linked mental retardation syndrome with seizures, hypogammaglobulinemia, and progressive gait disturbance is regionally mapped between Xq21.33 and Xq23. *Am. J. Med. Genet.* **85**: 255–262.
- Davidson, J. J., Ozelik, T., Hamacher, C., Willems, P. J., Francke, U., and Kilimann, M. W. (1992). cDNA cloning of a liver isoform of the phosphorylase kinase alpha subunit and mapping of the gene to Xp22.2–p22.1, the region of human X-linked liver glycogenosis. *Proc. Natl. Acad. Sci. USA* **89**: 2096–2100.
- De Plaen, E., De Backer, O., Arnaud, D., Bonjean, B., Chomez, P., Martelange, V., Avner, P., Baldacci, P., Babinet, C., Hwang, S. Y., Knowles, B., and Boon, T. (1999). A new family of mouse genes homologous to the human MAGE genes. *Genomics* **55**: 176–184.
- des Portes, V., Soufir, N., Carrie, A., Billuart, P., Bienvenu, T., Vinet, M. C., Beldjord, C., Ponsot, G., Kahn, A., Boue, J., and Chelly, J. (1997). Gene for nonspecific X-linked mental retardation (MRX 47) is located in Xq22.3–q24. *Am. J. Med. Genet.* **72**: 324–328.
- El-Hodiri, H. M., Shou, W., and Etkin, L. D. (1997). *xnf7* functions in dorsal–ventral patterning of the *Xenopus* embryo. *Dev. Biol.* **190**: 1–17.
- The European Backcross Collaborative Group (1994). Towards high resolution maps of the mouse and human genomes—A facility for ordering markers to 0.1 cM resolution. *Hum. Mol. Genet.* **3**: 621–627.
- Freemont, P. S. (1993). The RING finger. A novel protein sequence motif related to the zinc finger. *Ann. N.Y. Acad. Sci.* **684**: 174–192.
- The French FMF Consortium (1997). A candidate gene for familial Mediterranean fever. *Nat. Genet.* **17**: 25–31.
- Gaudenz, K., Roessler, E., Quaderi, N., Franco, B., Feldman, G., Gasser, D. L., Wittwer, B., Horst, J., Montini, E., Opitz, J. M., Ballabio, A., and Muenke, M. (1998). Opitz G/BBB syndrome in Xp22: Mutations in the MID1 gene cluster in the carboxy-terminal domain. *Am. J. Hum. Genet.* **63**: 703–710.
- Goll, C. M., Pastore, A., and Nilges, M. (1998). The three-dimensional structure of a type I module from titin: A prototype of intracellular fibronectin type III domains. *Structure* **6**: 1291–1302.
- Gossen, M., Wullrich, A., and Kilimann, M. W. (1995). Dinucleotide repeat polymorphism within the PHKA1 gene at Xq12–q13. *Hum. Genet.* **95**: 469–470.
- Gregg, R. G., Palmer, C., Kirkpatrick, S., and Simantel, A. (1996). Localization of a non-specific X-linked mental retardation gene, MRX23, to Xq23–q24. *Hum. Mol. Genet.* **5**: 411–414.
- Gu, X. X., Decorte, R., Marynen, P., Fryns, J. P., Cassiman, J. J., and Raeymaekers, P. (1996). Localisation of a new gene for non-specific mental retardation to Xq22–q26 (MRX35). *J. Med. Genet.* **33**: 52–55.
- Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillet, D., Muselet, D., Prud'Homme, J. F., Dib, C., Auffray, C., Morissette, J., Weissenbach, J., and Goodfellow, P. N. (1996). A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5**: 339–346.
- Henry, J., Mather, I. H., McDermott, M. F., and Pontarotti, P. (1998). B30.2-like domain proteins: Update and new insights into a rapidly expanding family of proteins. *Mol. Biol. Evol.* **15**: 1696–1705.
- Huber, R., Mazarella, R., Chen, C. N., Chen, E., Ireland, M., Lindsay, S., Pilia, G., and Crisponi, L. (1998). Glypican 3 and glypican 4 are juxtaposed in Xq26.1. *Gene* **225**: 9–16.
- Hunkapiller, T., and Hood, L. (1989). Diversity of the immunoglobulin gene superfamily. *Adv. Immunol.* **44**: 1–63.
- The International FMF Consortium (1997). Ancient missense mutations in a new member of the RoRet gene family are likely to cause familial Mediterranean fever. *Cell* **90**: 797–807.
- Jack, L. J., and Mather, I. H. (1990). Cloning and analysis of cDNA encoding bovine butyrophilin, an apical glycoprotein expressed in mammary tissue and secreted in association with the milk-fat globule membrane during lactation. *J. Biol. Chem.* **265**: 14481–14486.
- Jemaa, L. B., des Portes, V., Zemni, R., Mrad, R., Maazoul, F., Beldjord, C., Chaabouni, H., and Chelly, J. (1999). Refined 2.7 centimorgan locus in Xp21.3–22.1 for a nonspecific X-linked mental retardation gene (MRX54). *Am. J. Med. Genet.* **85**: 276–282.
- Kioschis, P., Wiemann, S., Heiss, N. S., Francis, F., Gotz, C., Poustka, A., Taudien, S., Platzer, M., Wiehe, T., Beckmann, G., Weber, J., Nordsiek, G., and Rosenthal, A. (1998). Genomic organization of a 225-kb region in Xq28 containing the gene for X-linked myotubular myopathy (MTM1) and a related gene (MTMR1). *Genomics* **54**: 256–266.
- Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C., and Kunkel, L. M. (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* **50**: 509–517.
- McKusick, V. A., Francomano, C., and Antonarakis, S. E. (1995). "Mendelian Inheritance in Man," Johns Hopkins Univ. Press, Baltimore.
- Montini, E., Buchner, G., Spalluto, C., Andolfi, G., Caruso, A., den Dunnen, J. T., Trump, D., Rocchi, M., Ballabio, A., and Franco, B. (1999). Identification of SCML2, a second human gene homologous to the *Drosophila* sex comb on midleg (Scm): A new gene cluster on Xp22. *Genomics* **58**: 65–72.
- Ohno, S. (1967). "Sex Chromosomes and Sex-Linked Genes," Springer-Verlag, Berlin.
- Opitz, J. M., Frias, J. L., Guttenberger, J. E., and Pellet, J. R. (1969a). The G syndrome of multiple congenital anomalies. *Birth Defects Orig. Art. Ser.* **V2**: 95–102.
- Opitz, J. M., Summit, R. L., and Smith, D. W. (1969b). The BBB syndrome familial telecanthus with associated congenital anomalies. *Birth Defects Orig. Art. Ser.* **V2**: 86–94.
- Palmer, S., Perry, J., Kipling, D., and Ashworth, A. (1997). A gene spans the pseudoautosomal boundary in mice. *Proc. Natl. Acad. Sci. USA* **94**: 12030–12035.
- Perry, J., and Ashworth, A. (1999). Evolutionary rate of a gene affected by chromosomal position. *Curr. Biol.* **9**: 987–989.
- Perry, J., Feather, S., Smith, A., Palmer, S., and Ashworth, A. (1998). The human *FXY* gene is located within Xp22.3: Implications for evolution of the mammalian X chromosome. *Hum. Mol. Genet.* **7**: 299–305.
- Plaxco, K. W., Spitzfaden, C., Campbell, I. D., and Dobson, C. M. (1996). Rapid refolding of a proline-rich all-beta-sheet fibronectin type III module. *Proc. Natl. Acad. Sci. USA* **93**: 10703–10706.
- Quaderi, N. A., Schweiger, S., Gaudenz, K., Franco, B., Rugarli, E. I., Berger, W., Feldman, G. J., Volta, M., Andolfi, G., Gilgenkrantz, S., Marion, R. W., Hennekam, R. C., Opitz, J. M., Muenke, M., Ropers, H. H., and Ballabio, A. (1997). Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. *Nat. Genet.* **17**: 285–291.
- Reddy, B. A., Etkin, L. D., and Freemont, P. S. (1992). A novel zinc finger coiled-coil domain in a family of nuclear proteins. *Trends Biochem. Sci.* **17**: 344–345.
- Reddy, B. A., Kloc, M., and Etkin, L. (1991). The cloning and characterization of a maternally expressed novel zinc finger nuclear phosphoprotein (*xnf7*) in *Xenopus laevis*. *Dev. Biol.* **148**: 107–116.
- Redolfi, A., Pizzuti, A., Di Bacco, A., Susani, L., Labella, T., Affer, M., Montagna, C., Reinbold, R., Mumm, S., Vezzoni, P., and Zucchi, I. (1999). Mapping of the MYCL2 processed gene to Xq22–23 and identification of an additional L MYC-related sequence in Xq27.2. *FEBS Lett.* **446**: 273–277.
- Reyniers, E., Van Thienen, M. N., Meire, F., De Boule, K., Devries, K., Kestelijn, P., and Willems, P. J. (1995). Gene conversion between red and defective green opsin gene in blue cone monochromacy. *Genomics* **29**: 323–328.
- Roberts, R. G., Freeman, T. C., Kendall, E., Vetrie, D. L., Dixon, A. K., Shaw-Smith, C., Bone, Q., and Bobrow, M. (1996). Charac-

- terization of DRP2, a novel human dystrophin homologue. *Nat. Genet.* **13**: 223-226.
- Robin, N. H., Feldman, G. J., Aronson, A. L., Mitchell, H. F., Weksberg, R., Leonard, C. O., Burton, B. K., Josephson, K. D., Laxova, R., Aleck, K. A., *et al.* (1995). Opitz syndrome is genetically heterogeneous, with one locus on Xp22, and a second locus on 22q11.2. *Nat. Genet.* **11**: 459-461.
- Saurin, A. J., Borden, K. L., Boddy, M. N., and Freemont, P. S. (1996). Does this have a familiar RING? *Trends Biochem. Sci.* **21**: 208-214.
- Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **95**: 5857-5864
- Schweiger, S., Foerster, J., Lehmann, T., Suckow, V., Muller, Y. A., Walter, G., Davies, T., Porter, H., van Bokhoven, H., Lunt, P. W., Traub, P., and Ropers, H. H. (1999). The Opitz syndrome gene product, MID1, associates with microtubules. *Proc. Natl. Acad. Sci. USA* **96**: 2794-2799.
- Seto, M. H., Liu, H. L., Zajchowski, D. A., and Whitlow, M. (1999). Protein fold analysis of the B30.2-like domain. *Proteins* **35**: 235-249.
- Shrimpton, A. E., Daly, K. M., and Hoo, J. J. (1999). Mapping of a gene (MRXS9) for X-linked mental retardation, microcephaly, and variably short stature to Xq12-q21.31. *Am. J. Med. Genet.* **84**: 293-299.
- Spitzfaden, C., Grant, R. P., Mardon, H. J., and Campbell, I. D. (1997). Module-module interactions in the cell binding region of fibronectin: Stability, flexibility and specificity. *J. Mol. Biol.* **265**: 565-579.
- Srivastava, A. K., Featherstone, T., Wein, K., and Schlessinger, D. (1995). YAC contigs mapping the human COL4A5 and COL4A6 genes and DXS118 within Xq21.3-q22. *Genomics* **26**: 502-509.
- Srivastava, A. K., McMillan, S., Jermak, C., Shomaker, M., Copeland-Yates, S. A., Sossey-Alaoui, K., Mumm, S., Schlessinger, D., and Nagaraja, R. (1999). Integrated STS/YAC physical, genetic, and transcript map of human Xq21.3 to q23/q24 (DXS1203-DXS1059). *Genomics* **58**: 188-201.
- Takahashi, M., and Cooper, G. M. (1987). ret transforming gene encodes a fusion protein homologous to tyrosine kinases. *Mol. Cell. Biol.* **7**: 1378-1385.
- Tamagnone, L., Lahtinen, I., Mustonen, T., Virtaneva, K., Francis, F., Muscatelli, F., Alitalo, R., Smith, C. I., Larsson, C., and Alitalo, K. (1994). BMX, a novel nonreceptor tyrosine kinase gene of the BTK/ITK/TEC/TKX family located in chromosome Xp22.2. *Oncogene* **9**: 3683-3688.
- Vorechovsky, I., Vetrie, D., Holland, J., Bentley, D. R., Thomas, K., Zhou, J. N., Notarangelo, L. D., Plebani, A., Fontan, G., Ochs, H. D., *et al.* (1994). Isolation of cosmid and cDNA clones in the region surrounding the BTK gene at Xq21.3-q22. *Genomics* **21**: 517-524.
- Wilke, K., Gaul, R., Klauck, S. M., and Poustka, A. (1997). A gene in human chromosome band Xq28 (GABRE) defines a putative new subunit class of the GABA_A neurotransmitter receptor. *Genomics* **45**: 1-10.

A New X-Linked Syndrome With Agenesis of the Corpus Callosum, Mental Retardation, Coloboma, Micrognathia, and a Mutation in the *Alpha 4* Gene at Xq13

John M. Graham, Jr.,^{1*} Patricia Wheeler,² Darci Tackels-Horne,³ Angela E. Lin,⁴ Bryan D. Hall,⁵ Melanie May,³ Kieran M. Short,⁶ Charles E. Schwartz,³ and Timothy C. Cox^{6,7}

¹Medical Genetics Birth Defects Center, Ahmanson Department of Pediatrics, Cedars-Sinai Medical Center, UCLA School of Medicine, Los Angeles, California

²Tufts-New England Medical Center, Boston, Massachusetts

³Greenwood Genetic Center, Greenwood, South Carolina

⁴Genetics and Teratology Unit, MassGeneral Children's Hospital, Boston, Massachusetts

⁵Department of Pediatrics, University of Kentucky Medical Center, Lexington, Kentucky

⁶School of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia

⁷Australian Craniofacial Institute, North Adelaide, South Australia

We describe two brothers with a unique pattern of malformations that includes coloboma (iris, optic nerve), high forehead, severe retrognathia, mental retardation, and agenesis of the corpus callosum (ACC). Both boys have low-set cupped ears with sensorineural hearing loss, normal phallus, pectus excavatum, scoliosis, and short stature. One brother had choanal atresia and cardiac defects consisting of ventricular septal defect (VSD) and patent ductus arteriosus (PDA) which resolved spontaneously. Differential diagnosis between a number of clinical entities was considered, however, because ACC and the distinctive facial fea-

tures were reminiscent of FG syndrome, DNA was analyzed for markers linked to the FGS1 locus at Xq13-q21. Notably, the brothers were concordant for markers spanning this presumed FG region, and in both we have identified adjacent alterations (-57delT and T-55A) in the *Alpha 4* gene located within this interval. *Alpha 4* is a regulatory subunit of the major cellular phosphatase, PP2A, that has recently been shown to interact with MID1, the product of the gene mutated in X-linked Opitz GBBB syndrome. The double nucleotide change identified in this family was not observed in 410 control chromosomes, suggesting that it may be a pathogenetic change. Altered expression of *Alpha 4*, through either a change in translational efficiency, mRNA stability or splicing, could explain the clinical phenotype in these boys and the phenotypic overlap with Opitz GBBB syndrome.

© 2003 Wiley-Liss, Inc.

KEY WORDS: agenesis of corpus callosum; X-linked mental retardation; *Alpha 4*; CHARGE syndrome; Temtamy syndrome; FG syndrome; Opitz GBBB syndrome; Toriello-Carey syndrome

This article celebrates the Festschrift of Dr. Bryan Hall (Lake Arrowhead Conference Center, Lake Arrowhead, CA, March 11 and 12, 2003).

Grant sponsor: UCLA Intercampus NIH/NIGMS Medical Genetics Training Program; Grant number: GM08243; Grant sponsor: NIH/NICHD (from the U.S. Department of Health and Human Services, Public Health Service); Grant number: HD22657; Grant sponsor: NIH/NICHD (to C.E.S.); Grant number: HD26202-12; Grant sponsor: South Carolina Department of Disabilities and Special Needs (SCDDSN); Grant sponsor: RD Wright Award (to T.C.C.); Grant number: 997706; Grant sponsor: NHMRC of Australia; Grant number: 157958.

*Correspondence to: John M. Graham, Jr., M.D., Sc.D., Director of Clinical Genetics and Dysmorphology, Cedars Sinai Medical Center, 444 South San Vicente Blvd #1001, Los Angeles, CA 90048. E-mail: john.graham@cshs.org

Received 1 May 2003; Accepted 14 July 2003

DOI 10.1002/ajmg.a.20504

INTRODUCTION

Phenotypic variability has been a confounding problem in establishing a clinical diagnosis in many genetic diseases and is particularly pertinent to dysmorphic

syndromes. Advances in gene identification and mutational screening technologies have only begun to reveal the extent of this problem, with recent molecular studies disclosing considerably more genetic heterogeneity than perhaps would have been thought a decade ago. Consequently, it is now generally appreciated that mutations in different genes within a pathway may cause indistinguishable or at least overlapping clinical phenotypes. Here, we report two brothers with an intriguing and distinct clinical presentation, whose differential diagnosis includes FG syndrome (among other possibilities, Table I) and a unique double mutation in the *Alpha 4* gene which is located within the FGS1 interval.

These two brothers were encountered at a CHARGE syndrome support group meeting (Boston, MA; July 25–27, 1997) where a call had been made for families with possible recurrent CHARGE syndrome. Members of the CHARGE Medical Advisory Board who examined both brothers did not feel that they met diagnostic criteria for CHARGE syndrome [Blake et al., 1998; Graham, 2001]. The possibility of X-linked CHARGE-like syndrome (MIM # 302905) was also considered, but this condition with cleft palate, coloboma, coronal hypospadias, deafness, short stature, and radial-ulnar synostosis with normal IQ was excluded because of their mental retardation and agenesis of the corpus callosum (ACC). Their facial features resembled a syndrome described by Temtamy et al. [1996] which included mental retardation, ACC, micrognathia, low-set ears, and coloboma, but the absence of brachydactyly, bulbous thumbs, genu vara, and pes planus differentiated these brothers from Temtamy syndrome. Because of the ACC and marked retrognathia, there was thought to be some similarity to Toriello–Carey syndrome [Toriello and Carey, 1988], but the facial features were different. These two brothers were felt to have a new syndrome consisting of coloboma (iris, optic nerve), high forehead, severe retrognathia, mental retardation, ACC, low-set cupped ears with sensorineural hearing loss, cardiac septal defects, pectus excavatum, scoliosis, and short stature (Fig. 1).

CLINICAL REPORTS

Patient 1

This boy was delivered to a 19-year-old primigravida woman with no reported drug or alcohol use during the pregnancy. Maternal eclampsia developed at 35 weeks gestation, resulting in early delivery via emergency C-section with Apgar scores of 5 and 8 at 1 and 5 min. Birth weight and length were at the 10th centile and occipito-frontal circumference (OCF) was at the 25th–50th centile. He required treatment for bilateral cryptorchidism, and left testicular torsion required orchiectomy, but gonadotropins were normal at age 16 years. During most of childhood, his height followed the 5th centile, with final adult height measured at 155.5 cm (<5th centile). Although he was thought to have CHARGE association during this time, his facial features are not consistent with this diagnosis (Fig. 2). He has mild mental retardation (IQ 62) with good social skills and with

TABLE I. Initial Differential Diagnosis for Clinical Features in Both Brothers

Features	Patient 1	Patient 2	CHARGE association	XLR CHARGE	FG syndrome	Temtamy et al. [1996]
Coloboma	Left iris	Optic nerve	85%	Iris 2/3	None	Iris, retina 3/3
Heart defect	None	VSD	60%	ASD 1/3	VSD, ASD ~30%	Dilated aorta, 2/3
Atresia choanae	None	Both nares	55%	None	Rare	None
Short stature	Ht <5th centile	Ht <5th centile	70%	Ht <5th centile 3/3	~50–60%	None
Low IQ	Mild MR	Mild-mod MR	>90%	None	Mod MR ~97%	Mild MR 2/3
GU hypoplasia	None	None	75%	Hypospadias 3/3	Hypospadias ~25%	None
Hearing loss	Mod mixed	Mild mixed	90%	Mild-mod 3/3	SNHL ~35%	None
Ear anomalies	Low-set, cupped	Low-set	Cupped 90%	Large, soft 3/3	Small, low-set ~60%	Low-set, cupped 3/3
CNS anomalies	ACC, macrocephaly	ACC, macrocephaly	Rare ACC, ventricular dilation	None	ACC ~25%, seizures 70%	ACC with ventricular dilation
Cranial nerve abnormalities	Aspiration pneumonia	None	80%	None	None	None
Micrognathia	Marked	Marked	Mild	None	Mild/mod	Moderate 3/3
Other	Pectus excavatum, broad forehead	Pectus excavatum, broad forehead	Cleft lip, facial palsy	Radio-ulnar synostosis, cleft palate	Anal anomalies constipation	Brachydactyly, dislocated lens

ACC, agenesis of the corpus callosum; ASD, atrial septal defect; CNS, central nervous system; GU, genito-urinary; MR, mental retardation; VSD, ventricular septal defect.



Fig. 1. Patient 1 (left) at age 17 years (a right lop-ear surgically reconstructed), and Patient 2 (right) at age 16 years. Note relative macrocephaly, broad forehead, downslanting palpebral fissures, left iris coloboma, and small low-set ears in Patient 1, retrognathia (surgically corrected in Patient 2), and short and broad neck.

progressive mixed hearing loss, which required bilateral hearing aids by his teenage years. MRI revealed ACC. He had a history of recurrent pneumonia due to aspiration, no difficulties with constipation, and self-limiting left grade 3-vesicoureteral reflux, which resolved with normal renal ultrasound evaluation.

Examination at age 17 years revealed macrocephaly (OFC 60 cm), broad forehead, bilateral downslanting



Fig. 3. Patient 1 at age 17 years, showing front and side facial views (after surgical correction of right lop ear) as well as hands.



Fig. 2. Patient 1 during childhood at various ages showing uncorrected right lop ears (top figures and bottom left), compared with adult features at age 17 years.

palpebral fissures, high-arched palate, bifid uvula, and retrognathia, and short and broad neck (Fig. 3). He had a left iris coloboma noted at birth, and decreased vision in his left eye resulted in nystagmus when he was not wearing glasses. External ears were low-set and small, and a right lop-ear had been surgically reconstructed. He had a normal echocardiogram, with no known heart defects. Genitalia were Tanner stage V with hypoplasia of the foreskin and a normally developed penis. He had a distal pectus excavatum, right thoracolumbar scoliosis requiring bracing and mildly decreased supination at the elbows bilaterally.

Patient 2

The younger brother was born at 32 weeks gestation to a 20-year-old G2P2 via an emergency C-section for preeclampsia to a mother who was on Phenobarbital for the last several weeks of the pregnancy. There was no reported drug or alcohol use during the pregnancy. Birth size was unknown, but during most of his childhood, his height was at the 10th centile, with height at 17 years below the 5th centile. He had chronic constipation with mild-moderate mental retardation (IQ 50) with minimal social interaction, as well as significantly impaired vision with mild sensorineural hearing loss. Cardiac defects included a ventricular septal defect (VSD) and patent ductus arteriosus (PDA) which resolved spontaneously without treatment. Occasional pneumonias



Fig. 4. Patient 2 at age 16 years after correction of retrognathia with surgical advancement.

improved after choanal atresia revision. Cranial MRI revealed ACC.

Examination at age 16 years (Fig. 4) revealed macrocephaly (OFC 60.25 cm), broad forehead, prominent nasal bridge and cheekbones, bilateral downslanting palpebral fissures, short broad neck, and slightly eccentric irises, with bilateral optic nerve colobomas. He had significantly decreased peripheral vision and horizontal nystagmus when he was not wearing his glasses. Bilateral choanal atresia treated initially at birth necessitated multiple surgeries, complicated by a high arched palate, and short oropharyngeal space, with severe retrognathia requiring mandibular advancement. Ears were well formed, with normal position. He had normal penile development, descended testes with Tanner stage V, normal pubertal development, and age-

appropriate testosterone and FSH levels. He had distal pectus excavatum and progressive thoracic scoliosis requiring bracing. He had a normal 46,XY karyotype, with negative FISH for 22q11.2 deletion.

MOLECULAR STUDIES

Blood was collected from the two boys, their mother, mother's half brother, and mother's maternal grandmother (family K8795) and genomic DNA isolated using a high salt method [Schwartz et al., 1990]. Purified DNA was stored in TE (10 mM Tris-HCl pH 7.6, 1 mM EDTA) at a concentration of 105 mg/ml at 4°C. Because of similarities to FG syndrome, microsatellite analysis was conducted with 11 markers spanning Xq12-q22 using an Automated Laser Fluorescent sequencer (ALF) (Amersham-Pharmacia) as previously reported [Stevenson et al., 1998]. The boys were concordant for all of these markers, four of which were informative in their mother (Fig. 5). These findings are consistent with the causative gene being located in the FGS1 region of Xq13-q21 [Briault et al., 1997; Graham et al., 1998]. This region contains 14 known genes, including the *Alpha 4* gene. As Alpha 4 has recently been shown to physically interact with MID1 which is mutated in X-linked Opitz GBBB syndrome [Liu et al., 2001; Trockenbacher et al., 2001; Short et al., 2002], this gene was considered a candidate gene for FG syndrome [Trockenbacher et al., 2001; Short et al., 2002] as well as for the clinical phenotype in these boys.

Amplification and direct sequencing of the six common coding exons (and immediately flanking intronic sequences) of the *Alpha 4* gene was performed. Primer sequences are available upon request to the corresponding author, TCC or CES. Amplification products across each exon were of the expected size, and sequences

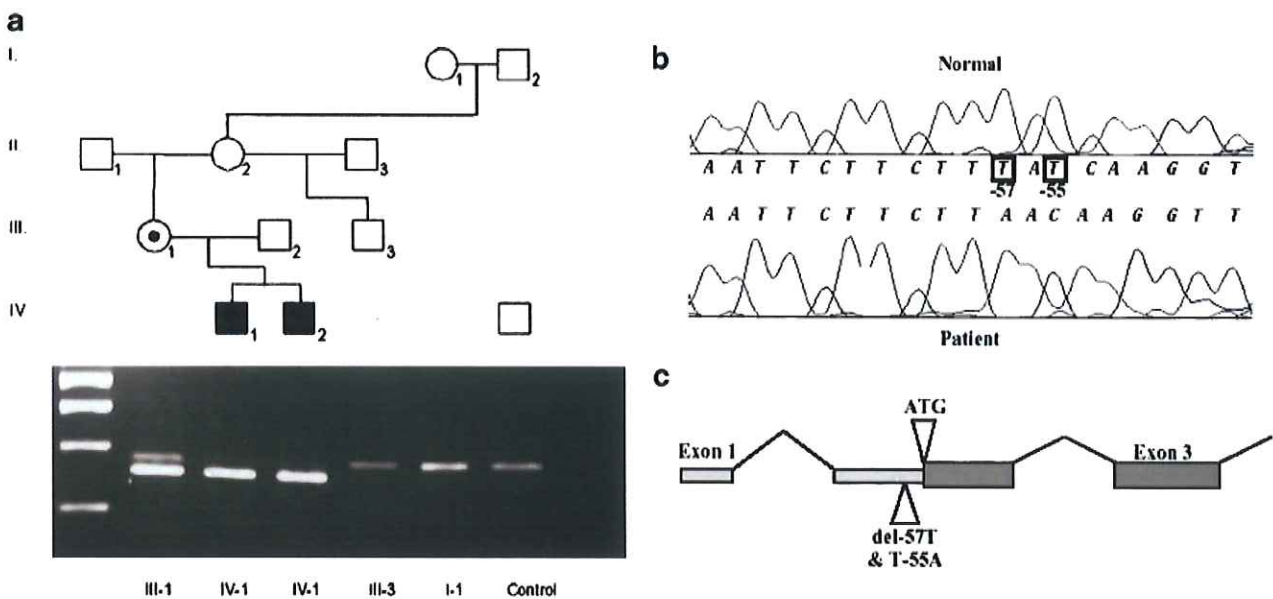


Fig. 5. a: Mutation analysis illustration. b: Patient chromatogram showing mutation. c: Schematic showing location of the changes.

generated from each were compared against available 'normal' *Alpha 4* gene sequence from the human genome draft sequence. The coding region of the *Alpha 4* gene was found to be normal in both brothers. However, both were found to have two nucleotide changes just 5' of the AUG translation initiation codon: -57delT and T-55A (Fig. 5b). Their mother was shown to be a carrier of both changes. The structure of the *Alpha 4* gene (www.ncbi.nlm.nih.gov/AceView; January 2003) indicates that these two changes are located in the 5'UTR of this gene (Fig. 5c). The presence of the changes was independently confirmed in both patients and their mother using primers designed such that a unique restriction enzyme site for Mse I was introduced when bound to and amplified from template containing the altered sequence(s) [Haliassos et al., 1989]. The Mse I analysis was utilized to show that the altered sequence was not present in the mother's half brother nor her maternal grandmother (Fig. 5a).

Notably, using this method, neither change was found in DNA from 410 control chromosomes (from 314 normal males and 48 normal females) suggesting that the difference was unlikely to be a common polymorphism and therefore, potentially causative of the clinical phenotype. In addition, we screened DNA from 9 cases of FG syndrome, 1 case of Toriello-Carey syndrome, 30 cases of X-linked mental retardation (XLMR) which had previously been linked to this same region, and 5 cases of Opitz GBBB syndrome that had no detectable mutation in MID1. No sequence alterations were found in the *Alpha 4* gene sequence in any of these cases.

In order to investigate whether these unique changes might affect *Alpha 4* expression or activity, we performed both Western blot analysis and immunocytochemistry on cultured skin fibroblasts from Patient 2 (IV-1, Fig. 5d) and skin fibroblast cultures from two unaffected control individuals. Using an anti-*Alpha 4* antibody in immunocytochemistry, endogenous *Alpha 4* localization showed no gross differences when compared to control cells. However, quantitation by Western blot analysis following normalization to histone deacetylase 1 (HDAC1), revealed a ~20% increase in *Alpha 4* levels over that seen in controls in the patient's fibroblasts (data not shown).

DISCUSSION

We describe two brothers with a unique pattern of malformations that includes coloboma (iris, optic nerve), high forehead, severe retrognathia, mental retardation, and ACC. Both boys have low-set cupped ears with sensorineural hearing loss, normal phallus, pectus excavatum, scoliosis, and short stature. The brothers have a striking similarity in physical features and medical problems with one brother in addition having choanal atresia and atrial and ventricular cardiac septal defects. The photographs document that they share the same condition with slight variation in expressivity. For much of their lives, they were considered to have CHARGE association, but at an International CHARGE Syndrome Support Group meeting several years ago, the CHARGE Syndrome Medical Advisory Board

rejected this diagnosis. This decision was made because they lacked typical ear findings and both had atypical facial features, ACC, and no hypogonadotropic hypogonadism. Updating the original criteria proposed by Pagon et al. [1981], new criteria for diagnosing CHARGE association were described by Blake et al. [1998]. Within this broad diagnostic category of CHARGE association patients, it has been proposed that there is a subgroup designated as the Hall-Hittner syndrome (a recognizable pattern of anomalies, with a probable genetic basis), a distinction advocated by Graham [2001]. The boys discussed in this report do not have the Hall-Hittner syndrome, but, interestingly, both fulfill classic CHARGE association criteria [Pagon et al., 1981].

Other possible diagnoses have been considered for these brothers and are included in the table and discussed below. However, the finding of a sequence abnormality in the *Alpha 4* gene in this family but not in a significant number of cases with any of the differential diagnoses, supports the notion that these boys represent a newly delineated X-linked syndrome that is related at the molecular level to Opitz GBBB syndrome. The additional diagnoses considered for these brothers are discussed in further detail below.

Abruzzo-Erickson Syndrome (X-Linked CHARGE Association)

Abruzzo and Erickson [1977] reported an apparently "new" syndrome of cleft palate, coloboma, hypospadias, deafness, short stature, and radial synostosis expressed variably in two brothers, their mother, and a maternal uncle. Davenport et al. [1986] and Metlay et al. [1987] cited this family as an instance of familial CHARGE syndrome, and Abruzzo and Erickson [1989] provided follow-up on the two brothers of the family who were children at the time of the first report. Like their mother and maternal uncle, neither had choanal atresia. Both had coronal hypospadias but genital development was otherwise normal. The ears were large and protruding, and hearing impairment required a hearing aid. Wide spacing between the second and third fingers as well as unilateral or bilateral radioulnar synostosis was noted in several members of the family, and mental retardation was not present. The manifestations in the mother, who according to the proposed inheritance as an X-linked disorder would be heterozygous, consisted of large ears and flat malar configuration, like her sons and brother, with wide spacing between the second and third digits as well as unusual rugosity of the palate. The brothers were below the 5th centile for height at age 19 and 16, respectively.

Temtamy Syndrome

Among the six children of a first-cousin marriage, Temtamy et al. [1996] observed a boy and two girls with a syndrome consisting of dysmorphic craniofacial features ACC and iris coloboma. Two of the three sibs had aortic dilatation with aortic regurgitation, and one had moderate mental retardation. The craniofacial

anomalies consisted of a large, dolichocephalic head, arched eyebrows, antimongoloid slant of the eyes, beaked nose, low-set and simple lop ears, long philtrum, short upper lip, and micrognathia. In conjunction with the 'keyhole' coloboma of the iris, retina, and choroid, the lenses were dislocated upward, and there was myopia and hypertelorism. Electron microscopy of the gingiva showed widening of the intercellular space, thickness of collagen fibers, and lack of periodicity. The sister, who died at age 22 years from heart failure, had moderate dilatation of the aorta, myocardial impairment, and bulbous thumbs.

FG Syndrome (Opitz-Kaveggia Syndrome)

FG syndrome is a rare X-linked recessive form of mental retardation, first described by Opitz and Kaveggia [1974]. Based on over 50 reported cases, FG syndrome is associated with ACC, minor facial anomalies (high, broad forehead with frontal cowlick, ocular hypertelorism, downslanted palpebral fissures, and small cupped auricles), relative macrocephaly, broad thumbs and halluces, partial toe syndactyly, and prominent fetal fingertip pads [Graham et al., 1999]. Affected individuals manifest neonatal hypotonia and severe constipation, with or without anal anomalies. Congenital hypotonia with joint hyperlaxity evolves into spasticity with joint contractures in later life [Romano et al., 1994]. Affability, hyperactivity, and excessive talkativeness are noted frequently in patients with FG syndrome [Graham et al., 1999], and 80% meet DSM-VI criteria for attention-deficit hyperactivity disorder, but not autism [Ozonoff et al., 2000].

Complete or partial ACC has been reported in a number of cases [Opitz et al., 1988; Kato et al., 1994; Sorge et al., 1996; Graham et al., 1999]. Opitz et al. [1982] suggested that the diagnosis of FG syndrome be considered in boys with congenital hypotonia and anorectal anomalies, and in older males with mental retardation, congenital hypotonia, joint contractures, chronic constipation, and characteristic facial appearance and personality. Constipation, with or without anal anomalies, is a distinctive major finding of FG syndrome, which may not resolve until middle childhood, and which is occasionally noted in female carriers [Romano et al., 1994].

Mental retardation is usually moderately severe, though milder cognitive effects have been reported [Ozonoff et al., 2000]. Sensorineural hearing loss has also been reported in several cases [Neri et al., 1984; Romano et al., 1994]. Thompson and Baraitser [1987] noted mental retardation in all 35 FG syndrome cases in their review, with hyperactive behavior in 18 of these 35 cases (51%). Opitz et al. [1988] reported a characteristic personality in 54% of cases. Congenital hypotonia secondary to CNS involvement also appears to be a very important distinguishing feature of this syndrome. Other distinctive findings include abnormal pinnae (small, rounded, and protruding), epilepsy, craniosynostosis, pectus excavatum, scoliosis, pyloric stenosis, cryptorchidism, hypospadias, hernia, single transverse palmar creases, syndactyly, and campto-

dactyly [Opitz et al., 1982, 1988; Bianchi, 1984; Neri et al., 1984; Richieri-Costa et al., 1986; Thompson and Baraitser, 1987; Kato et al., 1994; Romano et al., 1994; Sorge et al., 1996; Graham et al., 1999].

Neuropathology findings have included megalencephaly, midline fusion of mammillary bodies, and neuroglial heterotopias with cerebral cortical dysgenesis and pachygyria suggesting a defective neuronal cell migration [Opitz et al., 1982; Thompson et al., 1986]. Congenital hypotonia with joint hyperlaxity usually progresses to contractures with spasticity and unsteady gait in later life [Romano et al., 1994]. Death during infancy has been reported in as many as a third of patients as a consequence of bronchopulmonary problems and/or heart defects, but once patients have survived infancy, early death in childhood is rare [Opitz et al., 1982; Thompson and Baraitser, 1987; Sorge et al., 1996].

Opitz GBBB Syndrome (Hypospadias-Dysphagia Syndrome)

The Opitz GBBB syndrome, also known as the hypospadias-dysphagia syndrome or telecanthus with associated abnormalities, is characterized by hypertelorism or telecanthus, labiopalatine defects, congenital heart disease (commonly ASD/VSD), mild CNS anomalies, and genitourinary defects such as hypospadias in males or splayed labia majora in females. However, there are numerous additional abnormalities that have been reported to be associated with this diagnosis. Opitz syndrome is a genetically heterogeneous syndrome with both X-linked and autosomal forms [Robin et al., 1995]. Interestingly, the clinical presentation of Opitz GBBB syndrome, even in males in a family showing X-linked inheritance, can be highly variable, even in males in a family showing X-linked inheritance [Cox et al., 2000]. Robin et al. [1996] compared the phenotypic features of the X-linked and autosomal forms of the Opitz syndrome. They found that anteverted nares and posterior pharyngeal cleft were seen only in the X-linked form, but this distinction was questioned by Cox et al. [2000]. All other manifestations of the syndrome, such as hypertelorism, swallowing difficulties, hypospadias, and developmental delay, can be seen in both forms.

Molecular Genetics of Opitz GBBB and FG Syndromes

Molecular investigations of the X-linked form of Opitz syndrome have shown that loss-of-function mutations in the *MID1* gene underlie this condition [Quaderi et al., 1997; Cox et al., 2000]. More recent studies have demonstrated that the Alpha 4 protein, a novel regulatory subunit of the central cellular protein phosphatase, PP2A, interacts with MID1 on microtubules [Liu et al., 2001; Trockenbacher et al., 2001; Short et al., 2002]. Using an embryonic fibroblast cell line derived from a patient harboring a MID1 mutation, Trockenbacher et al. [2001] showed that mutations in MID1 lead to a marked accumulation of the catalytic subunit of PP2A

(known as PP2Ac). This accumulation is thought to result from impairment of the presumed E3 ubiquitin ligase activity of the MID1 protein that normally facilitates ubiquitylation and thus, marking of PP2Ac for degradation through binding to its Alpha 4 regulatory subunit [Trockenbacher et al., 2001; Short et al., 2002]. One outcome of elevated PP2Ac levels was shown to be hypophosphorylation of proteins in the microtubule-fraction of cell extracts, a pathologic mechanism viewed as being consistent with a pathogenetic role in the Opitz syndrome phenotype [Trockenbacher et al., 2001].

Both Short et al. [2002] and Trockenbacher et al. [2001] pointed out that the *Alpha 4* gene (also known as *IGBP1*) maps to Xq13, within an interval showing linkage to FG syndrome [Briault et al., 1997; Graham et al., 1998; Lossi et al., 2000]. Similar to Opitz GBBB syndrome, FG syndrome is characterized by mental retardation combined with imperforate anus, congenital heart defects, and characteristic face. Consequently, it was speculated that Alpha 4 might therefore have a role in the pathogenesis of FG syndrome [Trockenbacher et al., 2001; Short et al., 2002] or alternatively, phenotypically similar entities that also map to this chromosomal region [Short et al., 2002]. In support of this proposition is the observation in mice that Alpha 4, similar to that of Mid1, is expressed during embryogenesis in tissues such as the developing craniofacial complex, skeletal muscle, brain, and heart [Everett and Brautigam, 2002; Jaafar and Cox, unpublished observations].

The two brothers described in this report have a unique spectrum of clinical features and are concordant for markers spanning this presumed FG region in Xq13-q21. Notably, both brothers have an unusual double mutation (-57delT and T-55A) in the 5'UTR of the *Alpha 4* gene that they inherited from their carrier mother. In contrast, there was no detectable change in the *Alpha 4* gene in 9 patients with classic FG syndrome, 1 case of Toriello-Carey syndrome, 30 case of XLMR which linked to Xq13, or 5 cases of Opitz GBBB syndrome which did not have a demonstrable mutation in MID1.

In conclusion, we propose that the unique clinical presentation in these boys represents a new XLMR syndrome due to a mutation in the *Alpha 4* gene. Although a causative role for the Alpha 4 nucleotide changes has not been definitively proven, one can envisage their location affecting expression of Alpha 4 through one of a number of possible mechanisms including: (1) altered translational efficiency, (2) altered stability, or (3) perturbed splicing of the Alpha 4 mRNA. These possibilities are currently being investigated.

In preliminary studies aimed at investigating the consequences of the nucleotide changes, we have observed a ~20% increase in Alpha 4 protein levels in cultured skin fibroblasts from Patient 2 in comparison to control fibroblasts. It is tempting to speculate that this difference is significant and causative of the clinical phenotype in this family. At this stage, however, we cannot exclude that the difference simply reflects some of the normal variation in Alpha 4 protein levels,

although we view this as less likely given the nucleotide changes have not been found in 410 control chromosomes and 45 cases presenting with overlapping clinical features. Consequently, we propose that the unique clinical presentation in these boys represents a new XLMR syndrome, which is likely due to a mutation in the *Alpha 4* gene. Ongoing studies aimed at quantitating Alpha 4 levels in additional control fibroblast lines and assessing the functional consequences of elevated Alpha 4 protein should help clarify the precise contribution of these alterations to the clinical presentation in this family.

ACKNOWLEDGMENTS

We thank the ongoing support of the CHARGE Syndrome Support group who facilitated our introduction to this family. We also thank the patients and their family for their unwavering willingness to assist in the research and without whom this research would not be possible. We appreciate the generous support of SHARE's Child Disability Center and the Steven Spielberg Pediatric Research Center.

REFERENCES

- Abruzzo MA, Erickson RP. 1977. A new syndrome of cleft palate associated with coloboma, hypospadias, deafness, short stature, and radial synostosis. *J Med Genet* 14:76-80.
- Abruzzo MA, Erickson RP. 1989. Re-evaluation of new X-linked syndrome for evidence of CHARGE syndrome or association. *Am J Med Genet* 34:397-400.
- Bianchi DW. 1984. Brief clinical report: FG syndrome in a premature male. *Am J Med Genet* 19:383-386.
- Blake K, Davenport SH, Hall BD, Hefner MA, Pagon R, Williams MS, Lin AE, Graham JM Jr. 1998. CHARGE association—An update and review for the primary pediatrician. *Clin Pediatr* 37:159-174.
- Briault S, Hill R, Shrimpton A, Zhu D, Till M, Ronce N, Marguerite-Jeannin P, Baraitser M, Middleton-Price H, Malcolm S, Thompson E, Hoo J, Wilson G, Romano C, Guichet A, Pembrey M, Fontes M, Poustka A, Moraine C. 1997. A gene for FG syndrome maps in the Xq12-q21.31 region. *Am J Med Genet* 73:87-90.
- Cox TC, Allen LR, Cox LL, Hopwood B, Goodwin B, Haan E, Suthers GK. 2000. New mutations in MID1 provide support for loss of function as the cause of X-linked Opitz syndrome. *Hum Mol Genet* 9:2553-2562.
- Davenport SLH, Hefner MA, Thelin JW. 1986. CHARGE syndrome: Part I. External ear anomalies. *Int J Pediatr Otorhinolaryng* 12:137-143.
- Everett AD, Brautigam DL. 2002. Developmental expression of alpha4 protein phosphatase regulatory subunit in tissues affected by Opitz syndrome. *Dev Dyn* 224:461-464.
- Graham JM Jr. 2001. Editorial comment: A recognizable syndrome within CHARGE association: Hall-Hittner syndrome. *Am J Med Genet* 99:120-123.
- Graham JM Jr, Tackels D, Dibbern K, Superneau D, Rogers C, Corning K, Schwartz CE. 1998. FG syndrome: Report of three new families with linkage to Xq12-q21.1. *Am J Med Genet* 80:145-156.
- Graham JM Jr, Superneau D, Rogers RC, Corning K, Schwartz CE, Dykens EM. 1999. Clinical and behavioral characteristics in FG syndrome. *Am J Med Genet* 85:470-475.
- Haliassos A, Chomel JC, Grandjouan S, Kruh J, Kaplan JC, Kitzis A. 1989. Detection of minority point mutations by modified PCR technique: A new approach for a sensitive diagnosis of tumor-progression markers. *Nucleic Acids Res* 17:8093-8099.
- Kato R, Niikawa N, Nagai T, Fukushima Y. 1994. Japanese kindred with FG syndrome (Letter). *Am J Med Genet* 52:242-243.
- Liu J, Prickett TD, Elliott E, Meroni G, Brautigam DL. 2001. Phosphorylation and microtubule associations of the Opitz syndrome protein mid-1 is regulated by protein phosphatase 2A via binding to the regulatory subunit alpha 4. *Proc Natl Acad Sci* 98:6650-6655.

- Lossi AM, Colleaux L, Chiaroni P, Fontes M, Villard L, Abidi F, Schwartz C, Briault S, Moraine C. 2000. Exclusion of nine candidate genes for their involvement in X-linked FG syndrome (FGS1) in three families (Letter). *Am J Med Genet* 94:386–388.
- Metlay LA, Smythe PS, Miller ME. 1987. Familial CHARGE syndrome: Clinical report with autopsy findings. *Am J Med Genet* 26:577–581.
- Neri G, Blumberg B, Miles PV, Opitz JM. 1984. Sensorineural deafness in the FG syndrome: Report on four new cases. *Am J Med Genet* 19:369–377.
- Opitz JM, Kaveggia EG. 1974. The FG syndrome: An X-linked recessive syndrome of multiple congenital anomalies and mental retardation. *Z Kinderheilk* 117:1–18.
- Opitz JM, Kaveggia EG, Adkins WN Jr, Gilbert EF, Viseskul C, Pettersen JC, Blumberg B. 1982. Studies of malformation syndromes of humans. XXXIII: The FG syndrome—further studies on three affected individuals from the FG family. *Am J Med Genet* 12:147–154.
- Opitz JM, Richieri-Costa A, Aase JM, Benke PJ. 1988. FG syndrome update 1988: Note of 5 new patients and bibliography. *Am J Med Genet* 30:309–328.
- Ozonoff S, Williams BJ, Rauch AM, Opitz JM. 2000. Behavior phenotype of FG syndrome: Cognition, personality, and behavior in eleven affected boys. *Am J Med Genet* 97:112–118.
- Pagon RA, Graham JM Jr, Zonana J, Yong SL. 1981. Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: CHARGE association. *J Pediatr* 99:223–227.
- Quaderi NA, Schwiger S, Gaudenz K, Francon B, Rugarli EI, Berger W, Feldman GJ, Volta M, Andolfi G, Gilgenkrantz S, Marion RW, Hennekam RC, Opitz JM, Muenke M, Ropers HH, Ballabio A. 1997. Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. *Nat Genet* 17:285–291.
- Richieri-Costa VM, Hassler E, Lubinsky MS. 1986. The FG syndrome: Further characterization, report of a third family, and of a sporadic case. *Am J Med Genet* 1:47–58.
- Robin NH, Feldman GJ, Aronson AL, Mitchell HF, Weksberg R, Leonard CO, Burton BK, Josephson KD, Laxova R, Aleck KA, Allanson JE, Guion-Almeida ML, Martin RA, Leichtman LG, Price RA, Opitz JM, Muenke M. 1995. Opitz syndrome is genetically heterogeneous, with one locus on Xp22, and a second locus on 22q11.2. *Nat Genet* 11:459–461.
- Robin NH, Opitz JM, Muenke M. 1996. Opitz G/BBB syndrome: Clinical comparisons of families linked to Xp22 and 22q, and a review of the literature. *Am J Med Genet* 62:305–317.
- Romano C, Baraitser M, Thompson E. 1994. A clinical follow-up of British patients with FG syndrome. *Clin Dysmorph* 3:104–114.
- Schwartz CE, Ulmer J, Brown A, Panscot I, Goodman HO, Stevenson RE. 1990. Allan–Herndon syndrome II. Linkage to DNA markers in Xq21. *Am J Hum Genet* 47:454–458.
- Short KM, Hopwood B, Yi Z, Cox TC. 2002. MID1 and MID2 homo- and heterodimerise to tether the rapamycin-sensitive PP2A regulatory subunit, Alpha 4, to microtubules: Implications for the clinical variability of X-linked Opitz GBBB syndrome and other developmental disorders. *BMC Cell Biol* 3:1.
- Sorge G, Polizzi A, Ruggieri M, Smilari P, Mauceri L. 1996. Early fatal course in three brothers with FG syndrome. *Clin Pediatr* 35:365–367.
- Stevenson RE, Arena JF, Ouzts E, Gibson A, Shokeir MH, Vnencak-Jones C, Lubs HA, May M, Schwartz CE. 1998. Renpenning syndrome maps to Xp11. *Am J Hum Genet* 62:1092–1101.
- Temtamy SA, Salam MA, Aboul-Ezz EHA, Hussein HA, Helmy SAH, Shalash BA. 1996. New autosomal recessive multiple congenital abnormalities/mental retardation syndrome with craniofacial dysmorphism absent corpus callosum, iris colobomas, and connective tissue dysplasia. *Clin Dysmorph* 5:231–240.
- Thompson E, Baraitser M. 1987. FG syndrome. *J Med Genet* 24:139–143.
- Thompson EM, Harding BN, Lake BD, Smith SC. 1986. Necropsy findings in a child with FG syndrome. *J Med Genet* 23:372–373.
- Toriello HV, Carey JC. 1988. Corpus callosum agenesis, facial anomalies, Robin sequence, and other anomalies: A new autosomal recessive syndrome? *Am J Med Genet* 31:17–23.
- Trockenbacher A, Suckow V, Foerster J, Winter J, Krauss S, Ropers HH, Schneider R, Schweiger S. 2001. MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation. *Nat Genet* 29:287–294.

Solution Structure of the RBCC/TRIM B-box1 Domain of Human MID1: B-box with a RING

Michael A. Massiah^{1*}, Brandi N. Simmons¹, Kieran M. Short²
and Timothy C. Cox²

¹Department of Biochemistry
and Molecular Biology
Oklahoma State University
Stillwater, OK 74075 USA

²Department of Anatomy and
Cell Biology, Monash
University, Clayton, Vic.
Australia 3800

B-box domains are a defining feature of the tripartite RBCC (RING, B-box, coiled-coil) or TRIM proteins, many of which are E3 ubiquitin ligases. However, little is known about the biological function of B-boxes. In some RBCC/TRIM proteins there is only a single B-box (type 2) domain, while others have both type 1 and type 2 B-box domains in tandem adjacent to their RING domain. These two types of B-boxes share little sequence similarity, except the presence of cysteine and histidine residues: eight in most B-box1 domains and seven in B-box2 domains. We report here the high-resolution solution structure of the first B-box1 domain (from the human RBCC protein, MID1) based on 670 nuclear Overhauser effect (NOE)-derived distance restraints, 12 hydrogen bonds, and 44 dihedral angles. The domain consists of a three-turn α -helix, two short β -strands, and three β -turns, encompassing Val117 to Pro164, which binds two zinc atoms. One zinc atom is coordinated by cysteine residues 119, 122, 142, 145, while cysteine 134, 137 and histidine 150, 159 coordinate the other. This topology is markedly different from the only other B-box structure reported; that of a type 2 B-box from *Xenopus* XNF7, which binds a single zinc atom. Of note, the B-box1 structure closely resembles the folds of the RING, ZZ and U-box domains of E3 and E4 ubiquitin enzymes, raising the possibility that the B-box1 domain either has E3 activity itself or enhances the activity of RING type E3 ligases (i.e. confers E4 enzyme activity). The structure of the MID1 B-box1 also reveals two potential protein interaction surfaces. One of these is likely to provide the binding interface for Alpha 4 that is required for the localized turnover of the catalytic subunit of PP2A, the major Ser/Thr phosphatase.

© 2006 Elsevier Ltd. All rights reserved.

*Corresponding author

Keywords: MID1; B-box; ubiquitin E3/E4 ligase; NMR; TRIM

Introduction

The tripartite RBCC (or TRIM) motif consists of an N-terminal RING domain (originally termed an "A box"), one or two "B-boxes" and a coiled-coil domain.¹ This RBCC domain arrangement is highly conserved, being found in all multi-cellular organ-

isms. In humans, the RBCC domain characterizes a family of over 50 proteins, and although few have been characterized in detail, their importance is underscored by the fact that some are oncoproteins (e.g. PML, RFP and TIF1a) while others, when mutated (e.g. MID1), give rise to various congenital abnormalities. More recently, members of this large family have also been found to play regulatory roles in a variety of specific cellular processes, including sperm vesicle exocytosis and intracellular release of HIV.^{2,3} Unlike the highly conserved arrangement of their N termini, RBCC proteins can have quite diverse C-terminal domain arrangements,⁴ the most common being a single SPRY or B30.2 domain. Despite the growing importance of this RBCC family of proteins, the overall function of the highly conserved tripartite structure remains to be elucidated. Evidence from a number of studies has implicated the RING domain in ubiquitin

Present address: T. C. Cox, Division of Craniofacial Medicine, Department of Pediatrics, University of Washington, Seattle, WA 98125, USA.

Abbreviations used: NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; RBCC, RING, B-box, coiled-coil; XNF7, *Xenopus* nuclear factor-7; GST, glutathione S-transferase; HSQC, heteronuclear single quantum coherence; TOCSY, total correlated spectroscopy; DTNB, 5,5'-dithio bis-2-nitrobenzoic acid; TRIM, tripartite motif.

E-mail address of the corresponding author:
massiah@biochem.okstate.edu

ligase function and the coiled-coil domain in multimerization, yet little is known about the function of B-box domains.⁵

B-boxes can have one of two different zinc-binding motifs. Type 1 B-boxes (B-box1) contain the following zinc-binding consensus sequence: C-X(2)-C-X(7-10)-C-X(2)-C-X(4-5)-C-X(2)-C(H)-X(3-6)-H-X(2-8)-H (C5(C/H)H2), while the consensus for type 2 B-boxes (B-box2) is: C-X(2)-4-C/H-X(7-10)-C-X(7)-C-X(2)-C-X(3)-6-H-X(2)-8-H (C(C/H)C3H2).^{6,7} B-box1 domains are larger with their eight zinc-binding residues spaced differently than those found in B-box2 domains that have seven classical zinc-binding residues. Thus, these two domains essentially show no primary sequence homology that belies the similarity in the nomenclature. In RBCC proteins, a type 2 B-box is a defining feature. However, around one-quarter of all human RBCC proteins possess tandem B-boxes where the type 2 B-box is accompanied by an N-terminally positioned type 1 B-box (B1B2 arrangement).⁴ B-boxes can also be found in isolation in some proteins and these are predominantly but not exclusively type 2 B-boxes.

To date, the only reported structure of a B-box was determined for the B-box2 domain from *Xenopus* nuclear factor-7 (XNF-7)⁸ for which there is no known orthologue in humans. XNF-7 is a factor required for the establishment of dorsal-ventral patterning in this species. The XNF-7 B-box, which contains four cysteine and three histidine residues, adopts a compact structure with two β -strands positioned perpendicular to each other, a helical turn and a number of small turns. It reportedly binds one zinc atom employing the Cys1, His1, Cys4 and His2 residues.⁸ The two uncoordinated cysteine residues (C2, C3) and a conserved aspartic acid (D) residue that is substituted with a cysteine in many TRIM proteins, are located ~ 3.5 Å from each other and oriented to bind a metal ion. The unliganded His37 (H3) is located on a loop that is close to the zinc binding cluster and which could easily be rotated to be in close proximity to the uncoordinated cysteine residues. However, as a result of this structure, B-box domains have been assumed to all exist as single zinc-binding domains.⁸

One of the better-characterized RBCC proteins is MID1, a 667 amino acid protein that contains two so-called B-boxes.^{4,9} At its C-terminal end, MID1 shares a distinct domain arrangement with five other RBCC proteins; the C terminus consisting of a single fibronectin type III (FNIII) motif and both a PRY (pre-SPRY) and SPRY domain.⁴ Mutations in MID1 have been found to underlie the X-linked form of Opitz G/BBB syndrome (OS), a developmental disorder characterized by numerous abnormalities including clefts of the lip and primary palate, cardiac structural defects, and genital anomalies.^{10,11} Recent studies have revealed that MID1 predominantly associates with the microtubule (MT) cytoskeleton and this association persists throughout the cell cycle.^{10,12,13} Work by three groups has provided

evidence that MID1 functions as an E3 ubiquitin ligase. In this role, MID1 reportedly facilitates the ubiquitylation and subsequent degradation of the catalytic subunit of the most abundant Ser/Thr protein phosphatase, PP2A (PP2Ac), which regulates a vast array of intracellular processes.^{9,14-16} This function of MID1 is likely mediated by the recruitment of Alpha 4, a PP2Ac-binding protein and negative regulator of PP2A activity, by the B-box1 domain and an as yet unidentified E2 ubiquitin-conjugating enzyme by the RING domain.^{9,15,16} Interestingly, the activity of MID1 itself may be regulated by PP2A, since it has recently been shown to be the target of a Ser/Thr MAP kinase, presumably at the consensus recognition site (P-N-S96-P) located between the RING and B-box1 domains.¹⁶ The direct interaction between Alpha 4 and the MID1 B-box1 domain was the first demonstration of an independent function for a B-box1 motif.^{9,15}

Here we report the solution structure of the MID1 B-box1 domain, the first such B-box1 structure, and show that, unlike the reported XNF-7 B-box2, this B-box type coordinates two zinc atoms. Strikingly, we find that the type 1 B-box adopts a classical RING-like fold with conserved structural features, despite the lack of conservation of coordinating zinc residue positioning, raising the possibility of this domain possessing ubiquitin ligase activity.

Results

B-box1 expression and purification

To simplify purification, the MID1 B-box1 was expressed as an N-terminal glutathione *S*-transferase (GST)-fusion protein. While a small percentage of GST-B-box1 was found to be soluble when the bacteria were grown in Luria-Broth (LB), greater than 95% of the protein formed inclusion bodies in cultures grown in M9 minimal media, which is required to obtain ¹⁵N and ¹³C-labeled protein for NMR studies. Most attempts to increase solubility by varying growth conditions were unsuccessful. The most successful purification protocol was obtained by lysing cells in the presence of an alkyl anionic detergent, *N*-laurylsarcosine (Sarkosyl), which effectively solubilized all GST-B-box1.¹⁷ Success of this protocol was aided by supplementing the M9 media with 440 mM sorbitol and 1.5 mM betaine. These osmolytes are hypothesized to increase the water content in the cells, minimizing aberrant folding of protein that might expose more hydrophobic surfaces and thus promote their aggregation. Despite its solubility, GST-B-box1 only bound GSH Sepharose in the presence of 4% Triton-X100 and 40 mM Chaps. While the efficiency of binding to the GSH Sepharose varied between 10–30%, we serendipitously discovered that both GST and GST-B-box1 bound benzamidine Sepharose 4FF (normally employed to bind and capture thrombin only) with ~ 75 –90% efficiency (B.N.S. & M.A.M., unpublished results). After the resin was

thoroughly washed of most detergent, resin-bound GST-B-box1 was treated with thrombin for 12 h at 22 °C. Greater than 95% of GST-B-box1 was cleaved before thrombin became bound to the same resin, suggesting that the thrombin had a higher affinity for the cleavage site than for the benzamidine resin. Cleavage released almost pure B-box1, while residually released GST was captured with GSH Sepharose. Even with thorough washing on the benzamidine Sepharose column, concentrating the B-box1 protein also resulted in concentration of residual detergents to ~0.015%, as monitored by UV-VIS spectroscopy. Nevertheless, purification of a sufficient quantity of B-box1 for NMR experimentation was possible with this method, and allowed for stable storage for ~8–12 months. Additionally, B-box1 was also stable in 300 mM NaCl. However, B-box1 purified without Sarkosyl, Triton-X 100 and Chaps was only stable for about one week, which was not optimal for extensive NMR experiments.

The purified B-box1 was 83 amino acid residues in length, representing 78 amino acids of native MID1 sequence and five amino acids (GSPGI) representing residues derived from the cloning and thrombin cleavage sites of the expression vector. The 78 native amino acid residues comprised the 48 amino acid region defining the B-box1 motif and an additional 30 residues from the N-terminal linker that joined the RING and B-box1 sequences (see Materials and Methods). For cross-referencing to the full-length MID1 sequence, native residues are labeled Gln87 to Pro164 as per their position in MID1 and the non-native amino acids (GSPGI) labeled residues 1 through 5.

NMR assignments

Since B-box1 was purified in the presence of detergent, we wished to determine whether this

would affect the quality of NMR signals acquired. To investigate this, ^1H NMR spectra of B-box1 purified with and without detergent were collected. The narrow line-width and the sharp dispersed signals of ^1H spectra were consistent with a folded 10 kDa protein that did not appear to be embedded in any detergent micelle or bicelle structure. The ^1H NMR spectra of B-box1 from the two purification protocols were the same. Sharp signals observed in the ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) spectrum were also consistent with a folded 10 kDa protein (Figure 1).

To elucidate the tertiary structure of the MID1 B-box1, two and three-dimensional NMR experiments were performed using ~0.75 mM $^{15}\text{N}/^{13}\text{C}$ -uniformly labeled B-box1 in ST buffer (50 mM Tris-HCl, 300 mM NaCl, 5 mM ZnCl_2 , 50 mM β -mercaptoethanol (pH 7.5)) with 2% (w/v) sodium azide. The 2D ^1H - ^{15}N HSQC spectrum (Figure 1) showed sharp NH signals for 95% of all non-proline residues characteristic of a folded protein. The NH signals for the first 30 amino acid residues were less dispersed and located centrally within the spectrum with less than 1 ppm and 5 ppm signal dispersion in the ^1H and ^{15}N resonances, respectively. This indicated that these residues were most likely unstructured. The NH signals for the first two amino acid residues (Gly1-Ser2) as well as for Ser98 and Glu103 could not be assigned. However, side-chain ^1H and ^{13}C atoms for the latter two amino acid residues were assigned using the ^{13}C -edited data. Furthermore, as the Ser98 NH proton was not observed, the side-chain ^1H and ^{13}C resonances of Pro97 could also not be assigned.

The ^1H , ^{15}N , $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ resonances were assigned using the HNCA, HNCACB, and the CBCACONH triple resonance NMR data. The HNCA data showed excellent correlations between the NH proton of the intra-residue N(*i*) and C $^\alpha$ (*i*)

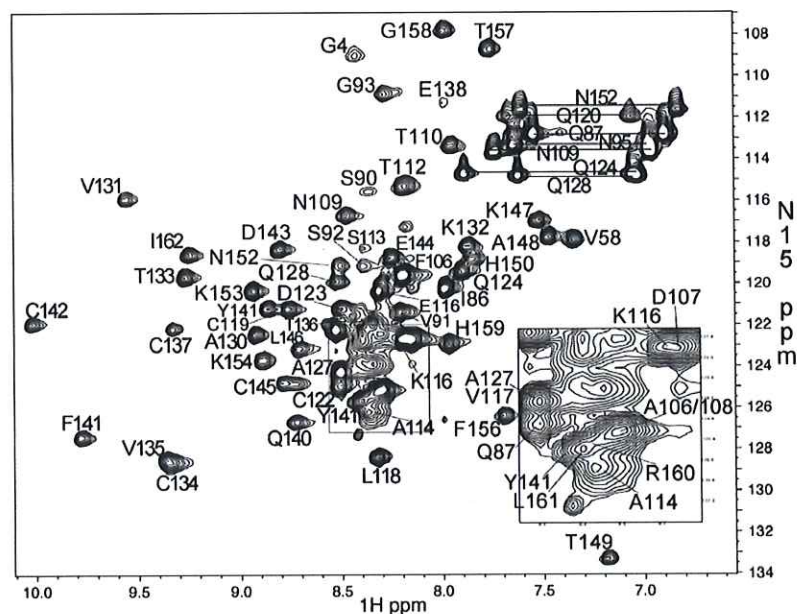


Figure 1. Two-dimensional ^1H - ^{15}N HSQC spectrum of B-box1 performed at 600 MHz. Amino acid assignments of each peak, corresponding to ^1H - ^{15}N atoms of an amino acid, are shown. Horizontal lines identify side-chain NH_2 groups of glutamine and asparagine residues. The inset is a close-up of a crowded region in the center of the spectrum.

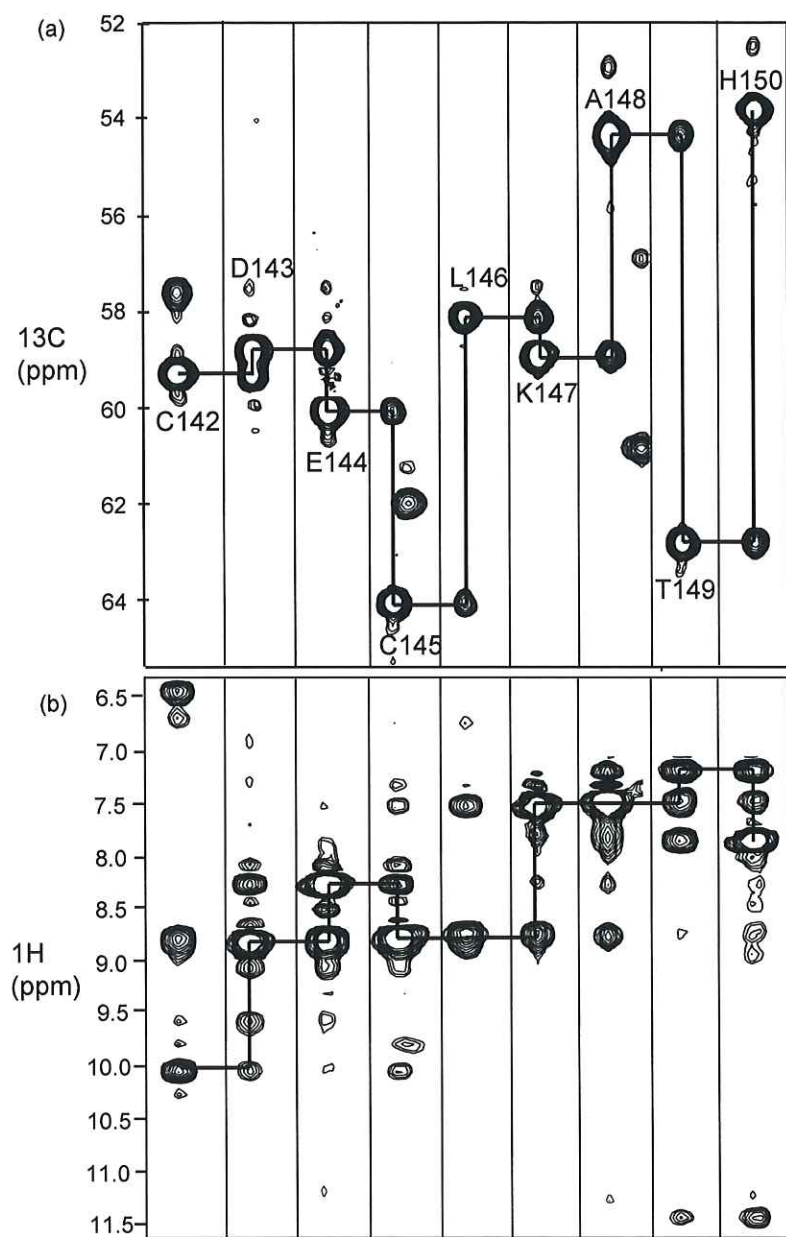


Figure 2. (a) ^1H - ^{15}N strips taken from the three-dimensional HNCA spectrum showing NH to intra- and preceding residue C^α atom connectivities for residues C142-D-E-C-L-K-A-T-H150. Of the two peaks in each plane, the weaker intensity ones correspond to the preceding residue C^α atoms, aligning with the stronger intensity peaks of the preceding amino acids, best depicted going from the plane showing His150 to Cys142. (b) ^1H - ^{15}N strips depicting strong intensity backbone NH-NH NOE correlations for residues C142-D-E-C-L-K-A-T-H150 from the 3D ^1H - ^{15}N NOESY-HSQC spectrum. These residues form the three-turn α -helix that includes three zinc-binding residues (Cys142, 145, and His150). Vertical lines separate the slices of each amino acid taken from different ^{15}N -planes from these two three-dimensional spectra.

atoms and the $\text{C}^\alpha(i-1)$ atom from the preceding residue (Figure 2(a)). The intensities of the $\text{C}^\alpha(i)$ peaks were stronger than the $\text{C}^\alpha(i-1)$ peaks, allowing straight-forward assignment of sequential amino acids. Overlapping C^α resonances were resolved by C^β connectivities from the HNCACB and CBCACONH spectra. The C^α and C^β resonances were assigned for all but three amino acids (Gly1, Ser90, and Pro97). The first seven N-terminal amino acid residues showed multiple conformations due to the slow *cis-trans* conformational changes of the Pro3 residue, resulting in additional ^1H - ^{15}N peaks in the HSQC spectrum.

Sequence specific backbone HN-NH($i,i+1-4$) and H^α -NH($i,i+1$) correlations were made with the 3D ^1H - ^{15}N HSQC-NOESY spectrum (Figure 2(b)).

Nuclear Overhauser effects (NOEs) for >5 residues showing HN-NH($i,i+1-4$) correlations not only identify sequential amino acids but also establish those residues that are part of an α -helix. As shown in Figure 2(b), residues Cys142 to His150 showed strong HN-NH($i,i+1$) NOEs, confirming that they were part of a three turn α -helix. The assignment of the NH-NH proton NOEs utilized the ^1H - ^{15}N - ^{15}N HSQC-NOESY-HSQC spectrum, which revealed the ^{15}N planes between the NH protons. The combination of these two spectra facilitated rapid identification of amino acids showing NH-NH proton NOE correlations. Strong sequential H^α -NH($i,i+1$) NOE correlations in the ^1H - ^{15}N HSQC-NOESY spectrum were used to identify extended secondary structural elements.

Assignments of the side-chain ^1H and ^{13}C resonances were made using the HCCH-TOCSY spectrum, which revealed intra-residue ^1H - ^{13}C correlations resolved by their directly attached carbon atoms. Together with the C^α and C^β chemical shifts obtained from the triple resonance spectra and the chemical shift patterns observed in the HCCH-TOCSY spectrum, positive assignments of B-box1 amino acids were established. Two of the five proline residues (Pro126 and 155) were determined to be in a *cis* conformation based on the observations of intermediate strength NOE correlations between preceding residue H^α and NH proton and the H^α proton of these proline residues. In addition, the C^α and C^β shifts of these proline residues were ~ 63 and 34 ppm, respectively, consistent with a *cis* conformation.¹⁸ The C^α and C^β chemical shifts of *trans*-proline residues are ~ 63.7 and 32 ppm, respectively. Proline 126 and 155 are conserved in MID1 and MID2 B-box1 domains from all species, although not in all B-box1 domains (see Figure 4(c)). Side-chain atoms of the C-terminal proline (Pro164) were assigned using the NOE patterns between Glu163 and Pro164 to establish the proton resonances in the HCCH-TOCSY spectrum.

Secondary structure

The secondary structural elements were established from backbone sequential proton NOE correlations and $^1\text{H}^\alpha$ and $^{13}\text{C}^\alpha$ chemical shift deviations (chemical shift index).^{19,20} Based on these criteria, a three-turn α -helix between residues Cys142 and His150, and two β -strands between residues Val131 to Cys134 and Glu138 to Tyr141 were determined. Three helical turns were observed between residues Cys119 and Asp123, Cys134 and Cys137, and Lys154 and Thr157. The two CXXC turns containing the zinc-binding cysteine residues (Cys119 to Cys122, and Cys134 to Cys137) formed classical rubredoxin (Rd) knuckles.^{21,22} The Rd knuckle orients the two cysteine residues to tetrahedrally coordinate a zinc atom. Three short extended regions were also identified based on strong H^α -NH($i, i+1$) and H^α and C^α chemical shift indices between Val117 and Cys119, Ala127 and Ala130, and His159 and Ile162. Interestingly, secondary structure prediction programs, such as PHD, were unsuccessful in correctly predicting any of the observed elements in B-box1. The core of the B-box1 domain is therefore defined from Val117 to Pro164.

The first 30 amino acid residues were unstructured based on the lack of sequential and long-range NOE correlations in the ^{15}N or ^{13}C -edited NOESY spectra. The C^α and H^α chemical shift values of these residues were similar to random coil values. The exception was Ser96 with C^α and H^α resonances 2.0 ppm downfield and 0.4 ppm upfield shifted, respectively. Using TALOS²³, the ϕ and ψ angles of Ser96 were estimated to be $-86(\pm 23)^\circ$ and $146(\pm 17)^\circ$, respectively, consistent with Ser96

being in an extended conformation. Interestingly, this serine is part of the P-N-S96-P motif that may be the site for MAP kinase phosphorylation.^{9,16} A search of the PNSP motif in the Protein Data Bank (PDB) revealed that the ϕ and ψ values of 70% of such serine residues were found in an extended conformation. This conformation, possibly induced or stabilized by adjacent proline residues, may orient the serine in an optimal orientation for the active site of a MAP kinase.

Structure calculation

Tertiary structures of B-box1 were calculated using CYANA 2.1²⁴ based on 670 NOE restraints, 12 hydrogen bonds, 21 Zn-protein restraints and 44 ϕ (ϕ) and ψ (ψ) dihedral angles estimated using TALOS.²² The restraints from the cysteine (2.30 – 2.35 Å) and histidine (2.0 – 2.05 Å) residues to zinc were based on results from EXAFS studies.^{25,26} Instead of employing the automated structure calculation protocol of CYANA, NOE-derived restraints were manually determined. As a result, many NOEs deemed the result of potential spin diffusion were omitted. For instance, protons that exhibit NOEs to a specific prochiral proton or methyl group often give rise to a weaker intensity NOE cross-peak to the neighboring geminal proton or methyl group. Since spin diffusion likely contributes to the weaker signal, and restraints derived from the weaker of two NOEs are redundant in the structure calculation, they were not employed. Consequently, $\sim 25\%$ of NOEs were omitted. Only NOE restraints effective in defining side-chain orientations, as well as secondary and tertiary structural elements, were employed in the calculations.

A total of 200 structures were generated from random initial positions using 20,000 steps of torsion angle dynamics. Twenty structures with the lowest target function values (<1.0), NOE violations <0.1 Å, dihedral angles $<2^\circ$, and van der Waals violations <0.2 Å were selected. Input data and statistics of the calculated structures are summarized in Table 1. The root-mean-square deviation (RMSD) for the superposition of the backbone atoms (N,C $^\alpha$,C) for residues Val117 to Pro164 was $0.78(\pm 0.16)$ Å (Figure 3(a)). Residues Lys153 to Gly158, which formed a turn based on strong sequential NH-NH($i, i+1-2$) NOEs, did not show extensive long range NOEs with the hydrophobic core of the protein and therefore their relative positions could not be properly defined. This is depicted by greater dispersion in their positions of backbone atoms in Figure 3. Excluding residues Lys153 to Gly158 and the C-terminal Pro164 from the superposition, the N,C $^\alpha$,C backbone superposition yielded a RMSD value of $0.45(\pm 0.1)$ Å. The RMSD value for superposition of all heavy atoms for the well-defined regions was $0.99(\pm 0.2)$ Å. Evaluation of the structures with Procheck revealed 95% of the residues in the most-favored and favored regions of the

Table 1. Statistics of MID1 B-box1 structure calculations performed with and without two zinc atoms

NOE distance restraints		
Intra-residue ($ i-j =0$)	303	
Sequential ($ i-j =1$)	177	
Medium range ($ i-j =2-4$)	68	
Long range ($ i-j >4$)	122	
Total NOEs	670	
Dihedral angle restraints^a		
ϕ	22	
ψ	22	
Hydrogen bonds	12	
Zinc-protein restraints	21	
Total restraints	747	
Violations	w/Zn ²⁺	w/o Zn ²⁺
Distance restraint >0.2 Å	0	0
VDW >0.2 Å	6 ^b	3
Dihedral >5°	0	0
CYANA2.1 Target function (20 structures)		
B-box 1 with zinc ^c	0.65±0.1	
B-box 1 without zinc	0.5±0.13	
Ramachandran plot (%) ^d	w/Zn ²⁺	w/o Zn ²⁺
Most favored	61.3±4	55.3±5
Additionally allowed	33.0±4	38.3±4
Generously allowed	4.0±3	5.0±3
Disallowed	1.5±1.5	1.6±1.2
Overall precision of an ensemble of 20 structures calculated with and without zinc (Å)		
Ordered residues ^e	w/Zn ²⁺	w/o Zn ²⁺
Backbone atoms ^f	0.45±0.1	0.77±0.40
Heavy atoms ^g	0.99±0.2	1.34±0.40
Residues Val116-Pro164		
Backbone atoms	0.78±0.16	0.98±0.43
Heavy atoms	1.31±0.17	1.53±0.41

^a Dihedral angles estimated using TALOS based on NH,¹⁵N, H^α, C^α, and C^β chemical shifts of each amino acid.

^b All van der Waals (VDW) violations arise from S-Zn restraints, due to larger radius of zinc in CYANA 2.1. Additional VDW violations arise from the two *cis*-proline bonds.

^c Target function for zinc-bound B-box was calculated by subtracting violations due to Zn-S and two *cis*-proline VDW violations. Subtracting TF calculated with just 21 zinc restraints resulted in an average TF value of 0.95±0.1.

^d All residues in B-box1 were used in Procheck.

^e Lys153 to Gly158, and Pro164 were omitted from the superposition because of limited long range NOEs between residues and the core the protein. Root-mean-square deviation (RMSD) values were calculated with CYANA 2.1.

^f Backbone atoms include backbone N, C^α, CO.

^g Heavy atoms include both backbone and side-chain non-hydrogen heavy atoms.

Ramachandran plot (Table 1). Only ~2% of residues were found in the disallowed region; all of these residues were located in the unstructured N terminus.

Zinc coordination

Canonical zinc binding domains bind zinc ions with high affinities (K_d) in the range of 10^{-8} – 10^{-13} M with ~ 10^{2-3} -fold greater preference for zinc over other divalent metals.²⁷ The requirement for 5–10 mM ZnCl₂ during protein purification and NMR sample preparation allowed us to conclude that the MID1 B-box1 is coordinated to zinc. Addition of tenfold molar excess EDTA resulted in complete collapse of the signals in the NH proton region (7–10 ppm), similar to the observed NMR

spectra of denatured proteins and apo-zinc finger proteins.²⁸ This confirms the requirement of B-box1 for zinc to stably fold this domain.

Three lines of evidence support the conclusion that the MID1 B-box1 coordinates two zinc atoms. The first is based on the use of Ellman's reagent²⁹ such that if the six cysteine residues are involved in coordinating zinc then their thiols would not react with 5,5'-dithio *bis*-2-nitrobenzoic acid (DTNB). The number of free thiols was determined by thiol-DTNB adducts monitored by UV-VIS absorbance at 412 nm divided by the extinction coefficient of DTNB ($13,500 \text{ M}^{-1} \text{ cm}^{-1}$) and the concentration of the protein. Folded B-box1 was first confirmed by 1D ¹H NMR spectroscopy. With native folded protein, no cysteine residues reacted with DTNB, suggesting that the cysteine residues were coordinated to zinc. When both 1 M urea and a 10× molar excess of EDTA were added, the formation of six thiol adducts was observed. Only two to four adducts were observed after 30 min with either EDTA or urea alone. Zinc coordination by the six cysteine residues was also confirmed by chemical shifts of their C^β atoms that were found to be between 30.5 to 32.5 ppm, consistent with the sulfur groups coordinating zinc ions. The C^β shifts for free and oxidized cysteine residues are ~28.5 ppm and ~41 ppm, respectively.^{20,25} Second, coupling patterns between the carbon-bound protons to the nitrogen atoms of the histidine side-chains in the ²J-NH ¹H-¹⁵N HSQC spectrum revealed the N^ε and N^δ nitrogen atoms of His150 and His159 coordinated zinc, respectively. Third, structures calculated with just NOE-based restraints revealed that Cys119, 122, 142 and 145 were clustered and oriented to coordinate a zinc atom, while Cys134, 137 and His149, 159 were in close proximity to coordinate another zinc (Figure 3(b)). The zinc-coordinating atoms of these residues were 3.5–4.5 Å from each other. Key NOEs between these zinc-binding residues were essential in defining their proximity. For instance, NOE correlations between Cys119 H^α and both Tyr141 H^α and Cys142 NH placed the Cys119-X-X-Cys122 Rd knuckle in close proximity to Cys142 and Cys145. A NOE between His150 H^α and His159 proR-H^β placed the two histidine residues in close proximity. NOE correlations between Cys134 and Leu161 backbone H^α and NH protons as well as Leu146 side-chain protons placed the Cys134-X-X-Cys137 Rd knuckle in close proximity to histidine residues 150 and 159. The precision of the ensemble of 15 structures calculated without the 21 protein-zinc restraints was very good. The RMSD value of C^α,C^β,N atom superposition of the structured regions (Table 1) was 0.77(±0.4) Å. Superposition of backbone atoms for residues Val117 to Pro164 resulted in a RMSD value of 0.98(±0.4) Å. The quality of the backbone angles was also similar to structures calculated with the 21 protein-zinc restraints. An average of 94% of residues had phi and psi angles in the most-favored and favored regions of the Ramachandran plot. In addition, no significant structural violations were

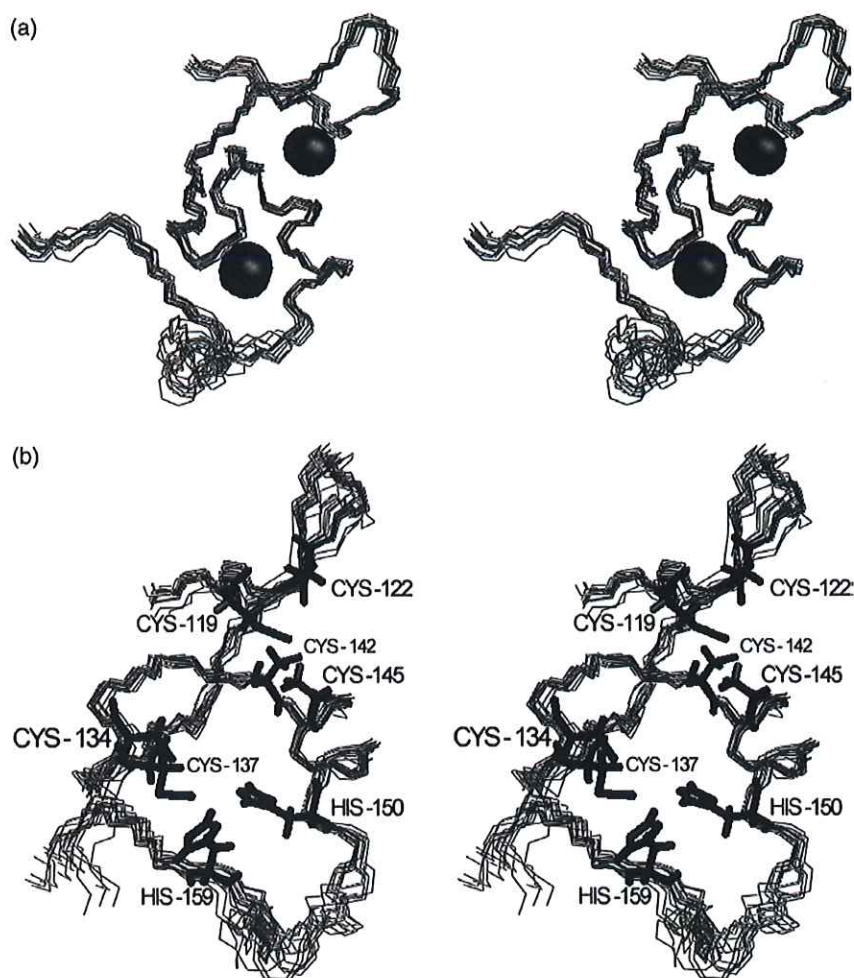


Figure 3. (a) Solution structure of B-box1. Stereo view of the superposition of backbone C α atoms of 20 structures calculated with CYANA 2.1. Only residues Val117 to Pro164, which formed the structure core, were included in the superposition. Two zinc atoms are shown. (b) Stereo view of the superposition of backbone C α atoms of 15 structures calculated with NOE-derived restraints, hydrogen bonds and dihedral angles but without protein–zinc restraints. The proximities of the zinc-binding residues are shown. One zinc is coordinated by Cys119, 122, 142, 145, while Cys134, 137 and His150, 159 coordinate the other zinc atom.

observed when the 21 restraints between B-box1 and the two zinc atoms were implemented in the structure calculations, confirming that the zinc restraints were consistent with the NOE-derived structures. Finally, zinc coordination by Cys137 is also consistent with deletion of Val135–Thr136–Cys137 observed in a patient with Opitz syndrome.³⁰ Located on the turn separating the antiparallel β -strands, these residues stabilize the CCHH zinc-binding cluster.

Tertiary structure

The MID1 B-box1 binds two zinc atoms with its six cysteine and two histidine residues in a cross-brace scheme similar to RING and PHD domains. The two zinc atoms are located ~ 12.8 Å apart. The first zinc atom is coordinated with two cysteine residues located on the first turn of the α -helix and the two cysteine residues located between residues Val117 to Val131 (Figure 4(a) and (b)). These

residues form a “lasso”-like loop structure positioning residues Cys119 to Cys122 on top of the N-terminal end of the helix, turning away and then looping back to lead into the antiparallel β -strands. A hydrogen bond (H-bond) between the Cys119 carbonyl and Cys122 NH proton stabilizes the Rd knuckle. Two additional H-bonds between the Cys119 NH proton and Gln128 carbonyl and the Ala130 NH proton and Val117 carbonyl stabilize the lasso-like structure. Hydrogen bonds were identified from NOE cross-peak patterns, which place residues Cys119, Gln120 and Phe121 close to Val139 and Tyr141.

The antiparallel β -strands between residues Val131 and Tyr141 are stabilized by four H-bonds between Lys132 and Tyr141, and between Cys137 and Val139. The two strands pack against the helix making key hydrophobic contacts with the buried side-chain of Leu146. Residues Thr133, Cys134, and Val135 are in close proximity with Leu161 and Ile162, placing Cys134 and Cys137 near His159.

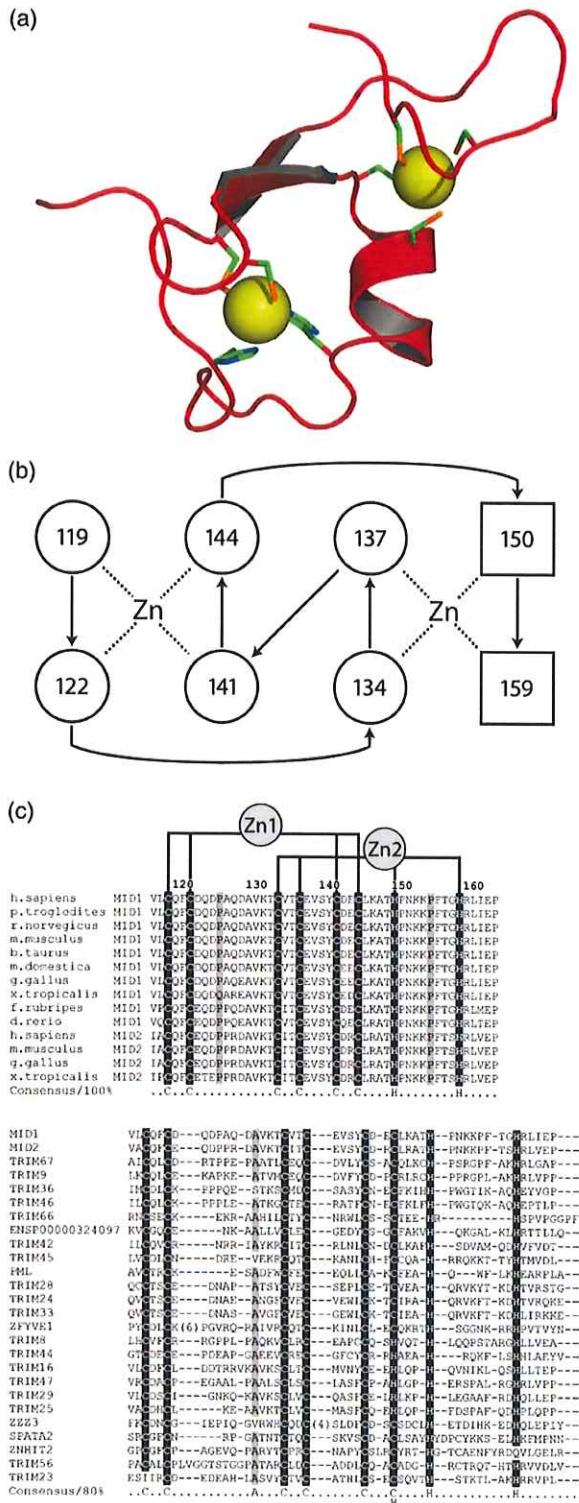


Figure 4. (a) Ribbon representation of B-box1. Two zinc atoms are shown as yellow spheres. Atom types of zinc-coordinating residues are color coded: orange, sulfur; green, carbon; blue, nitrogen; red, backbone atoms. (b) Zinc-binding topology within B-box1. One zinc atom is coordinated by four cysteine residues, while two cysteine and two histidine residues coordinate the other. The amino acid numbering is consistent with that of the full-length MID1 sequence. (c) (Top) Sequence alignment of

Residues between His159 and Pro164 extend away from the core of the protein leading into B-box2.

Two large turns, formed between residues Gln124 to Gln128 and Lys153 to Gly158, contain the two conserved proline residues (Pro126 and 155). These proline residues were determined to be in the *cis* conformation based on sequential NOEs and C α and C β shifts. Pro126 was found on the turn that twisted the lasso-like structure back towards the core of the protein, while Pro155 is located on the turn preceding the zinc-coordinating His159. Both proline residues are adjacent to electrostatic residues, with Pro126 and Pro155 surrounded by acidic and basic residues, respectively. The locations and conserved nature of these proline residues suggest that they are essential for structural stability and function of the MID1 and MID2 B-box1 domains.

Interestingly, the tertiary structure contained a cleft on the surface opposite the helix that is lined with the hydrophobic residues Val131, Leu132, Phe156, Leu161, and Ile162. A minor hydrophobic patch is formed by Phe121, Val139, and Tyr141. Flanking the hydrophobic cleft are two acidic residues, Glu138 and Glu163 (Figure 5). On the opposite side of the cleft, on the outer aspects of the helix, two charged surfaces are observed: a basic surface formed by residues Lys132, Lys147, Lys153, and Lys154, all clustered within ~5 Å of each other, and an acidic surface formed by residues Asp123, 125, and 129, as well as Asp143 and Glu144 (Figure 5). In the case of both MID1 and MID2, these surfaces may be essential for mediating the interaction with Alpha 4.

Discussion

The B-box was first identified along with the RING motif (originally referred to as the A-box) as zinc binding domains.^{6,7,31} Since this time, much effort has been placed on understanding the role of the RING domain in the hundreds of proteins that contain this motif. In contrast, very few studies have focused on determining the role of B-boxes. RING and B-box motifs are predominantly found together in RBCC/TRIM proteins, a superfamily characterized by an N-terminal RING domain followed by either one or two B-boxes and a coiled-coil domain. The importance of the RBCC proteins is increasingly

B-box1 domains from MID1 and the highly homologous MID2 from seven vertebrate species. Conserved zinc-binding residues are highlighted with black. The conserved, structurally important Pro126 and 155 are highlighted with gray. (Bottom) Alignment of all human B-box1 domains identified by an HMMer analysis of the Genbank database. In almost all cases, the C5(C/H)H2 zinc-binding motif is conserved. Note the high conservation of primary sequence between the final two histidine residues particularly for MID1 and MID2, TRIM67 and TRIM9, TRIM36 and TRIM46, TRIM28, TRIM24 and TRIM33.

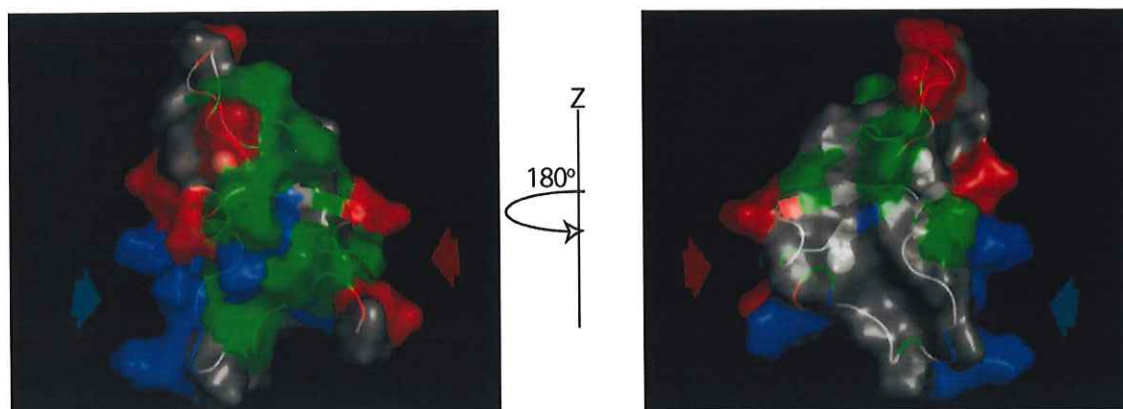


Figure 5. Surface representation of B-box1 showing basic patches (colored blue) comprised of lysine and arginine residues, acidic patches of glutamic and aspartic acids (colored red), and hydrophobic surfaces (colored green) formed by leucine, isoleucine, alanine, phenylalanine, and valine residues. The red arrow indicates a cleft formed by two acidic residues flanking a hydrophobic pocket. The blue arrow indicates a large cleft of basic residues that may be important for protein–protein interaction (possibly with Alpha 4).

being appreciated with many of these proteins recently shown to play regulatory roles in key intracellular processes.⁵ In humans, there are more than 50 members of the RBCC/TRIM superfamily with the order and spacing of these domains highly conserved throughout the family.⁴

In RBCC/TRIM proteins that possess two B-boxes, a type 1 B-box is always nearest the RING domain and is separated by about ten amino acid residues from a type 2 form that lies adjacent the coiled-coil region.^{4,5} However, B-box domains can also be found in proteins that do not have RING or coiled-coil domains. In such cases, and like RBCC proteins that have only a single B-box domain, they are predominantly but not exclusively type 2 B-boxes. Although the B-box name suggests that these domains would show primary sequence similarity, they actually have different zinc-binding motifs. The slightly larger B-box1 (C5(C/H)H2) contains eight classical zinc-binding residues (Cys/His) that show a different spacing to the seven zinc-binding residues of B-box2.^{5,6}

Distinct functions for the type 1 and type 2 B-boxes were first demonstrated with the finding that the B-box1 of MID1 specifically bound Alpha 4, a factor previously shown to also bind, and act as a negative regulator of, the catalytic subunit of protein phosphatase 2A. The binding of Alpha 4 by MID1 tethers it to the microtubule cytoskeleton and in turn facilitates recruitment of PP2Ac to the microtubules, where it is marked by ubiquitin for proteasome-mediated degradation.^{9,15,16} Here, we have presented the solution structure of the B-box1 domain of MID1, revealing that it adopts a tertiary structure reminiscent of the “cross-brace” fold of RING domains.

B-box1 adopts a “cross-brace” structure similar to RING, ZZ and U-box folds

Of the 12 RING structures currently in the protein database, ten adopt a conserved structural topology

consisting of an α -helix and two short antiparallel β -strands ($\beta\beta\alpha$ motif) positioned approximately perpendicular to each other (Figure 6(a)). The two exceptions are the PML RING and the second RING domain from the RING-IBR-RING Triad motif of HHARI (the human homologue of *Drosophila* ariadne).³² Notably, the MID1 B-box1 structure shares this $\beta\beta\alpha$ topology and is most similar to the E3 ubiquitin ligase RING domain folds from BRCA1, BARD1,³³ TFIIF,³⁴ (Figure 6(a)) RAG1,³⁵ and NOT4.³⁶ In such E3 ligases, the RING tertiary structure, which binds two zinc atoms, has been shown to be essential for presenting specific target proteins to their cognate E2-conjugating enzymes.³⁷ As shown with the RING consensus sequence (Figure 6(b)), there are notable differences between RING structures in two principal regions. These variable regions, where the number of residues can vary between 9–39 and 4–48, respectively, may assist in determining the nature of any protein–protein interaction and contribute to the substrate specificity in the ubiquitylation reaction. In the case of B-box1, the number of amino acids in each of these two regions lies at the lower end of the range seen in RING domains, with seven to ten and three to six residues, respectively (see Figure 6(b)). Yet despite this difference, the relative proximities of the zinc atoms with respect to the α -helix are essentially the same. The average distance between the two zinc atoms in RING domains is $\sim 14.6(\pm 1.1)$ Å, with the largest and smallest separations of 17.5 Å and 13.5 Å measured for the RING domains of cNOT (PDB: 1UR6)³³ and BARD1 (PDB: 1JM7),³⁸ respectively. The separation of the zinc atoms in the MID1 B-box1 is slightly smaller than in the RING domains at ~ 12.8 Å. Whether the distance between the two zinc atoms is important for function is not clear because a correlation between zinc distances and E3 ligase activity has not been established. Nevertheless, given the presence of a similar zinc coordinated “cross-brace” scheme to that of E3 RING domains, this raises the possibility that

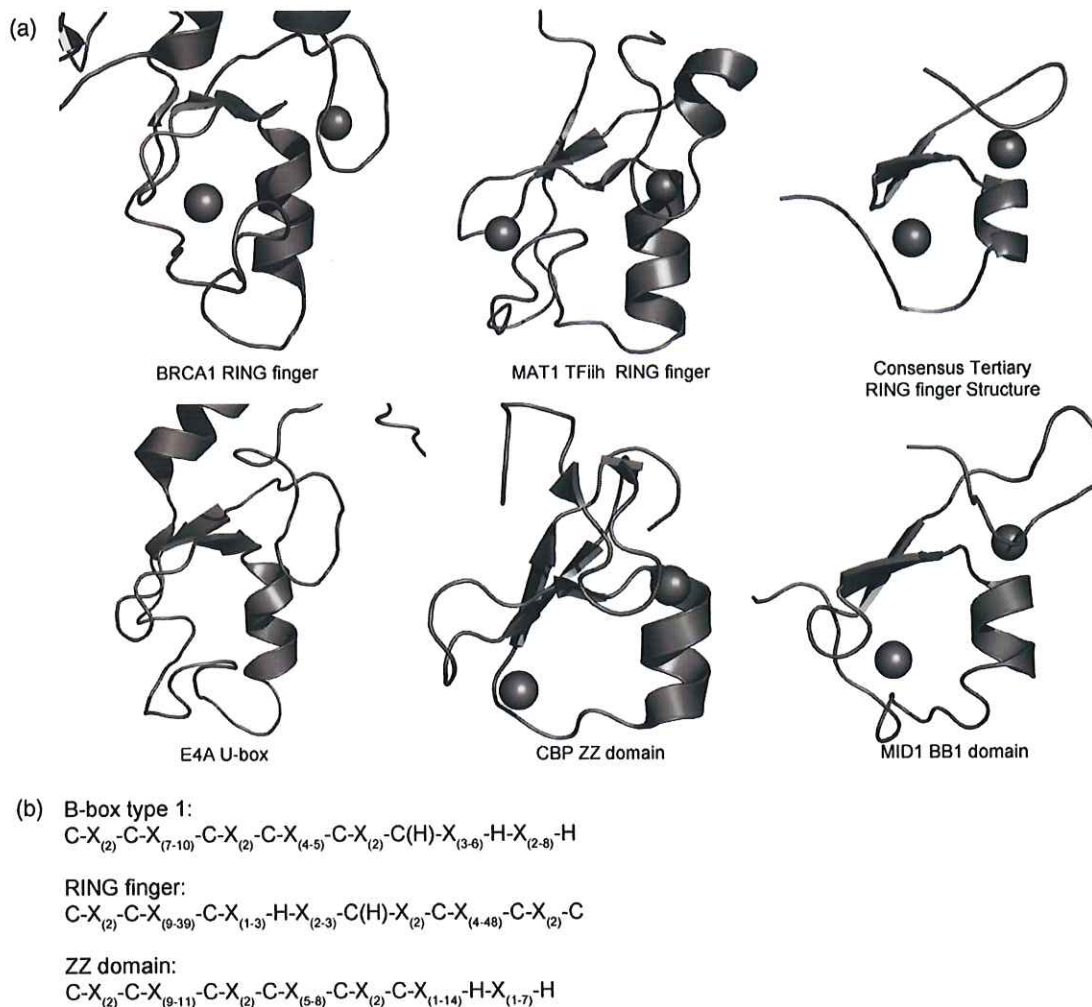


Figure 6. Structural similarities between B-box1, RING, U-box and ZZ domains. (a) Ribbon representation of RING domains from the breast cancer susceptibility protein, BRCA1 (1JM7), CDK-activating kinase assembly factor, MAT1 (1G25), a schematic representation of a canonical RING structure, the U-box domain from E4A, the ZZ domain from CBP/p300, and the B-box1 domain from MID1. (b) Consensus zinc-binding motifs of RING, ZZ, U-box and B-box1 domains; X=any amino acid, C and H are cysteine and histidine, zinc-binding residues.

B-box1 domains also have E3 ligase activity at least *in vitro*, and thus perhaps adding greater substrate diversity to RBCC proteins that possess a B-box1 domain.

In consideration of possible E3 ligase activity, we compared the structure of B-box1 with the determined RING domain structure of cCbl when bound to the UbcH7 E2-conjugating enzyme (PDB: 1FBV) (data not shown).³⁹ A groove (or cleft) between the helix and the two loops involved in zinc coordination formed the surface of cCbl that makes contact with UbcH7. Hydrophobic residues line this groove and are flanked by two basic residues. The complementary binding surface of UbcH7 is localized to the tips of two loops and the C-terminal end of the helix, with both hydrophobic and charged residues making contacts. These loops are inserted within the groove of the RING domain. Notably, a similar hydrophobic cleft is found in

the MID1 B-box1 (see Figure 5). However, unlike in the cCbl RING domain the hydrophobic residues of the MID1 B-box1 are flanked by two acidic residues. These residues could easily form salt-bridges with any two of the six surface lysine residues of UbcH7, in particular Lys1020 and 1040 (numbering scheme in 1FBV), which directly make contact with the groove in the case of the cCbl RING. A similar model for the interaction between the NOT4 RING domain and UbcH5b (PDB: 1UR6) has also recently been described.⁴⁰

Given the gross similarity in the distance between the coordinated zinc atoms and the presence of a $\beta\alpha$ motif in a similar cross-brace scheme, it was initially surprising that a tertiary structure search using DALI⁴¹ resulted in only two significant matches: E4A (PDB: 1WGM) and CBP/p300 (PDB: 1TOT). Neither of these proteins has classical RING domains yet they have been shown *in vitro* to have

E3 ligase activity. However, *in vivo* data indicate that they only enhance endogenous E3 ligases by promoting polyubiquitylation of mono- or oligo-ubiquitylated proteins, i.e. act as E4 ubiquitylation enzymes.⁴²

E4A is a protein that contains a U-box motif at its C terminus. The U-box motif has been recognized as having evolved from the RING motif to the point where many of the zinc-coordinating cysteine and histidine residues have been lost but the same $\beta\alpha$ tertiary structure is maintained by numerous salt bridges.⁴² Although also possessing E4 enzyme activity, CBP/p300 does not have a recognizable U-box but instead has a ZZ domain that is also found in many other proteins including some RING-type E3 ligases.⁴³ The recent determination of the solution structure of the ZZ domain has shown that it also adopts a RING-like fold. Notably, the zinc coordinating Cys-His arrangement in the ZZ domains is the same as that seen in the majority of B-box1 domains: CCCC, CCHH.⁴³ The striking tertiary structure similarities with the U-box and ZZ domain of these two E4 enzymes (see Figure 6(a)) has led us to hypothesize that the B-box1 domain of MID1 (and presumably other B-box1-containing RBCC and non-RBCC proteins) confers E4 activity. Such activity, in the case of the B-box1-containing RBCC proteins, may enhance the activity of its own RING E3 function or that of dimerization partner ligases, similar to the BARD1-BRCA1 heterodimer.⁴⁴ Ongoing studies are aimed at investigating this intriguing possibility for the MID1 B-box1.

Biological implications

The solution structure of the MID1 B-box1 demonstrates that B-box1 is quite distinct from the only other determined B-box structure, that of B-box2 from XNF-7⁸, in that it binds two zinc atoms and adopts a RING-like fold. The B-box1 sequence is highly conserved across both MID1 and MID2 proteins from all vertebrate species in which they have been identified (see Figure 4(c)). The core zinc-coordinating residues of B-box1 are also conserved in all other human RBCC/TRIM proteins that contain tandem B-boxes as well as in those non-RBCC proteins in which the domain has been noted. However, the differences in sequence and size within the variable regions across the B-box1-containing proteins suggest that these regions, like in RING domains, may confer the specificity to substrate or partner ligase interactions. Of note, the alignments reveal some unexpected similarity in these regions between a number of B-box1-containing proteins suggesting perhaps they bind similar substrates or conjugating factors.

Our data suggesting that MID1 and other B-box1-containing RBCC/TRIM proteins have two domains with RING-like folds (RING and B-box1) raises some interesting possibilities about the regulation or control of their E3 ligase functions. Based on structure-function studies with the

BARD1-BRCA1 complex,^{33,38} Xia *et al.*⁴⁴ reported that the BARD1 significantly enhanced BRCA1 ubiquitin E3 ligase activities and that the enhancement is directly correlated with their interaction. The two RING domains of BARD1 and BRCA1 form a heterodimer,³³ the interfaces of which localize to the antiparallel β -strands and comprise both charged and hydrophobic residues. This interaction still exposes the hydrophobic clefts that are essential for E2-conjugating enzyme recognition. It is feasible that the U-box and ZZ domains, which enhance E3 ligase activity of RING domain-containing proteins, could indirectly do so by similarly enhancing E2-conjugating enzyme recognition. Although in the case of MID1, we and others have shown that B-box1 binds Alpha 4 and subsequently recruits and targets PP2Ac for degradation,^{9,15,16} we describe here two opposing interfaces on B-box1 which may differentially interact with Alpha 4 and an E2 enzyme. Further structure-function studies employing site-directed mutagenesis and binding assays will help clarify this and determine whether an interaction (either direct or indirect) between the RING and B-box1 domains may be regulating overall E3 ligase activity levels.

Materials and Methods

B-box1 domain search

To identify the full complement of B-box1 domains within the human proteome, an alignment of available B-box1 domains from all RBCC proteins was used to generate an HMM database using HMMbuild from the HMMer suite of programs.⁴⁵ Perl scripts were used for the execution of programs as well as parsing of program output. The B-box1 HMM database was then used to scan all Genbank human amino acid sequences† using HMMpfam, and any proteins reported to be significantly high scoring were aligned with the remainder of the known B-box1 domains, as shown in Figure 4(c).

Construct

A native MID1 cDNA was used as the template to amplify the region encoding residues Gln87 to Pro164. Amplification was carried out using the following primers: M1BB1-GEXF2, 5'-CCAGGAATTCAGAAAGCATCAGTGAGCGGCCCC-3' and M1BB1-GEXR, CTGAATTCATGGCTC AATCAGACGATGGCC (EcoR1 restriction sites are indicated in bold; the created stop codon is underlined). The PCR product was sub-cloned into appropriately restricted pGEX-4T2 plasmid (Amersham Biosciences), with orientation and sequence verified by DNA sequencing. For expression of the domain, the construct was transformed into BL21(DE3) cells. Even though B-box1 is 48 amino acid residues in length, 30 additional amino acids from the linker region between the RING and B-box1 were included for two reasons: (i) part of this region (Glu98 to Ala105) was predicted to contain an α -helix that we considered might stabilize the B-box1 structure; and (ii) to determine whether

† ftp://ftp.ensembl.org/pub/current_homo_sapiens/data/fasta/pep/Homo_sapiens.NCBI35.nov.pep.fa.gz

Ser96, a proposed site of MAP kinase phosphorylation,^{9,16} was part of a secondary structural element. The position of the EcoRI linker on the forward primer was such that after cloning, expression and cleavage of the GST-B-box1, five amino acid residues from the GST remained (GSPGI). Thus, the purified peptide containing B-box1 was 83 amino acid residues in length, with the first five residues representing non-native amino acids.

Expression and purification of the MID1 B-box1 domain

To obtain ¹⁵N or ¹⁵N/¹³C-labeled proteins, cells were grown under antibiotic selection in M9 minimal media containing either 1 g/l ¹⁵NH₄Cl and 4 g/l glucose or 1 g/l ¹⁵NH₄Cl and 2 g/l ¹³C-uniformly labeled glucose (Cambridge Isotope Laboratories, Inc., Boston, MA), respectively. To increase protein solubility and reduce inclusion body formation, the media was supplemented with 440 mM sorbitol and 1.5 mM betaine. Cells were initially grown at 37 °C to an optical density at 600 nm (*A*₆₀₀) of 0.5, induced with 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) with the temperature reduced to 30 °C. Cells were harvested after 4 h and stored at -80 °C until lysis. Sorbitol, betaine, IPTG, reduced glutathione (cat. No. G4251), and vitamin cocktail solution (cat. no. B6891) were purchased from Sigma Chemical Co. (St. Louis, MO, USA).

To obtain soluble GST-B-box1, cells were lysed at 22 °C with ST buffer (50 mM Tris-HCl, 300 mM NaCl, 5 mM ZnCl₂, 50 mM β-mercaptoethanol (pH 8.0)), lysozyme (1.5 mg/g cells), 2% (v/v) *N*-Sarkosyl, 0.2 mM phenylmethanesulfonyl fluoride (PMSF), 400 mg/ml deoxycholic acid, 1 μg/ml DNase, pulse sonicated and centrifuged at 20,000g for 30 min. We serendipitously determined that GST and GST-B-box1 bound at least five times better to benzamidine Sepharose 4FF than to GSH Sepharose. Binding was aided by the addition of 4% (v/v) Triton-X 100 and 40 mM 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (Chaps). To remove detergents, the resin was washed with ST buffer until the *A*₂₀₀₋₄₀₀ of the eluate matched that of the wash buffer. The resin with bound protein was then incubated for 12 h with thrombin at 22 °C. Even though thrombin can bind to the benzamidine Sepharose, complete cleavage of the GST-B-box1 was observed. The majority of cleaved GST remained bound, while residual GST was removed from B-box1 with GSH Sepharose. Thrombin, glutathione- and benzamidine Sepharoses, and the GST Detection Kit (cat. no 27-4590-1) were purchased from Amersham Biosciences (Piscataway, NJ, USA). Chaps was purchased from Sigma-Aldrich.

NMR sample

The NMR sample contained ~0.75 mM ¹⁵N/¹³C-labeled B-box1 in ST buffer (pH 7.5), 2% (w/v) sodium azide, 90% H₂O/10% ²H₂O. The protonated buffer and relatively high salt concentration did not significantly affect the quality of the NMR data, but the high salt did stabilize the purified B-box1. Using UV-VIS spectroscopy, it was estimated that ~0.015% of Triton-X 100 was present in the NMR sample. B-box1 was stable for ~12 months when purified in the presence of detergents, but stable for only one to two weeks when detergents were omitted.

NMR experiments

NMR experiments were performed at 21 °C (probe temperature) on a Varian Inova 600 MHz Spectrometer

equipped with a 5 mm triple resonance (¹H, ¹³C, ¹⁵N) probe with z-axis gradient. All pulse sequences were included in the Varian BioPack with the Vnmr 6.1C operating system. To assign backbone atoms, the ¹⁵N-¹H HSQC, HNCA, CBCACONH, HNCACB, were performed. Side-chain protons, secondary structural elements and tertiary structure were established with the 3D ¹H-¹⁵N-¹⁵N-HSQC-NOESY-HSQC (200 ms mixing time), ¹H-¹⁵N-NOESY-HSQC (150 ms mixing time), ¹H-¹³C-NOESY-HSQC (200 ms mixing time) and ¹H-¹³C-HCCH-TOCSY spectrum (14 ms spin lock, 9 MHz field-strength). The tautomeric states of the two histidine residues were established using ¹⁵N-¹H HSQC adjusted to ²J-NH coupling constant of 22 Hz.⁴⁶ Water suppression was accomplished with WATERGATE⁴⁷ or Gradient Enhancement.⁴⁸ All NMR data were processed with NMRPipe⁴⁹ and analyzed with SPARKY 3† using PowerPC G4 Macintosh computers operating with OSX 10.3.7.

Structural restraints

All NOEs were grouped into three categories of distance restraints based on their intensities in the 3D ¹⁵N- and ¹³C-edited NOESY spectra: 1.8 Å to 2.8 Å, 1.8 Å to 3.3 Å and 1.8 Å to 5.0 Å for strong, medium and weak NOEs, respectively. For NOEs between methyl groups (CH₃) and ambiguo geminal protons (CH₂), an additional 1.0 Å and 0.5 Å, respectively, were added to the above distance ranges. For NOEs involving aromatic side-chain protons, 2.3 Å were added to account for the pseudo-atom positions for the δ and ε-protons. Two distance restraints between the O-H (1.8 Å to 2.5 Å) and O-N (2.7 Å to 3.5 Å) of the carbonyl and amide groups were used to define each hydrogen bond (H-bond). H-bonds were established from NOE patterns and the proximities of donor and acceptor groups in initial structures calculated primarily on NOE-derived restraints. Based on the chemical shifts of ¹H, ¹⁵N and ¹³C atoms for each amino acid, dihedral angles were estimated for 22 amino acids using TALOS19.²³ Distance restraints between the zinc atom and cysteine side-chain atoms were rigidly fixed to 2.30 Å to 2.35 Å (Zn-S^γ) and 3.5 Å (Zn-H^β). Distances between the zinc atom and histidine side-chain nitrogen were fixed at 1.95 Å to 2.05 Å based on EXAFS data.^{25,26} To maintain tetrahedral geometry around the zinc, distances between the S^γ, N^ε and N^δ atoms were set to 3.85 Å (M. Summers, personal communication). Four cysteine residues (Cys119, 122, 142, 145) coordinated one zinc and two cysteine (Cys134, 137) and two histidine residues (His150, 159) coordinated the second (Figure 4(b)).

Structure calculations

A total of 670 NOE distance restraints, 12 hydrogen-bond restraints, 44 dihedral angles and 21 zinc restraints were employed in structure calculations. The hydrogen bonds were identified by NOE patterns and the proximity of carbonyl and NH groups. The 44 phi and psi angles were estimated from amino acid chemical shifts using TALOS. All atoms of the proteins including two zinc atoms were included in the calculations. The structures were computed using CYANA 2.1 on PowerPC G4 Macintosh computers operating with OSX 10.3.7. From a total of 200 randomly calculated structures, 20 were selected based on low target functions (TF) less than one. A total of 20,000 steps for torsion-angle simulated

† <http://www.cgl.ucsf.edu/home/sparky/>

annealing were employed. For structures calculated without zinc atoms, the 21 restraints to zinc were omitted in the calculation. Two hundred structures were randomly calculated and 15 with low target functions and low RMSD values were selected.

Ellman's assay

5,5'-Dithio bis-2-nitrobenzoic acid (DTNB) and cysteine solutions were freshly prepared in Tris-HCl and divided equally for spectroscopic analyses. A standard curve using 0 to 1.5 mM cysteine was first determined by adding small aliquots of DTNB and incubating for 15 min before measuring the absorbance at 412 nm. Additional standard curves in 50 mM Tris-HCl, 300 mM NaCl, 0.1% (w/v) NaN₃, with or without EDTA and urea were established. All standard curves yielded *R*-values >0.999. DTNB titration with approximately 0.15 mM B-box1 in ST buffer alone, as well as ST buffer with urea, EDTA, and both urea and EDTA were performed. Free thiols were determined by dividing the absorbance at 412 nm (indicating thiol-DTNB adducts) by the extinction coefficient of DTNB (13,500 M⁻¹ cm⁻¹) and the concentration of the protein.

Protein Data Bank accession codes

The coordinates and restraints used in CYANA for 15 structures of the MID1 B-box1 were submitted to the Protein Data Bank with the accession code 2FFW. Restraints and chemical shifts of B-box1 were deposited in the BMRB database under the accession code 6920.

Acknowledgements

We thank Drs Michael F. Summers (University of Maryland Baltimore County) and Roberto DeGuzman (University of Kansas) for helpful discussions with CYANA and discussion regarding the manuscript. Our appreciation is also extended to Shashak Bhide for his computer assistance. The work was supported in part by the Oklahoma State University Agricultural Experimental Station (project # 2527, to M.A.M.), establishment funds from the Department of Anatomy and Cell Biology, Monash University, and by a grant from the Oklahoma Center for the Advancement of Science and Technology (OCAS, HR-196, to M.A.M.).

References

1. Reymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzi, L. *et al.* (2001). The tripartite motif family identifies cell compartments. *EMBO J.* **20**, 2140–2151.
2. Kitamura, K., Tanaka, H. & Nishimune, Y. (2003). Haprin, a novel haploid germ cell-specific RING finger protein involved in the acrosome reaction. *J. Biol. Chem.* **278**, 44417–44423.
3. Yap, M. W., Nisole, S., Lynch, C. & Stoye, J. P. (2004). Trim5alpha protein restricts both HIV-1 and murine leukemia virus. *Proc. Natl Acad. Sci. USA*, **101**, 10786–10791.
4. Short, K. M. & Cox, T. C. (2006). An RBCC subfamily that associates with the cytoskeleton via a novel microtubule-binding signature. *J. Biol. Chem.* Jan 23; [Epub ahead of print].
5. Meroni, G. & Diez-Roux, G. (2005). TRIM/RBCC, a novel class of "single protein RING finger" E3 ubiquitin ligases. *Bioessays*, **27**, 1147–1157.
6. Borden, K. L. (1998). RING fingers and B-boxes: zinc-binding protein-protein interaction domains. *Biochem. Cell. Biol.* **76**, 351–358.
7. Torok, M. & Etkin, L. D. (2001). Two B or not two B? Overview of the rapidly expanding B-box family of proteins. *Differentiation*, **67**, 63–71.
8. Borden, K. L., Lally, J. M., Martin, S. R., O'Reilly, N. J., Etkin, L. D. & Freemont, P. S. (1995). Novel topology of a zinc-binding domain from a protein involved in regulating early *Xenopus* development. *EMBO J.* **14**, 5947–5956.
9. Short, K. M., Hopwood, B., Yi, Z. & Cox, T. C. (2002). MID1 and MID2 homo- and heterodimerise to tether the rapamycin-sensitive PP2A regulatory subunit, alpha 4, to microtubules: implications for the clinical variability of X-linked Opitz GBBB syndrome and other developmental disorders. *BMC Cell. Biol.* **3**, 1.
10. Cox, T. C., Allen, L. R., Cox, L. L., Hopwood, B., Goodwin, B., Haan, E. & Suthers, G. K. (2000). New mutations in MID1 provide support for loss of function as the cause of X-linked Opitz syndrome. *Hum. Mol. Genet.* **9**, 2553–2562.
11. So, J., Suckow, V., Kijas, Z., Kalscheuer, V., Moser, B., Winter, J. *et al.* (2005). Mild phenotypes in a series of patients with Opitz GBBB syndrome with MID1 mutations. *Am. J. Med. Genet. A*, **132**, 1–7.
12. Cainarca, S., Messali, S., Ballabio, A. & Meroni, G. (1999). Functional characterization of the Opitz syndrome gene product (midin): evidence for homodimerization and association with microtubules throughout the cell cycle. *Hum. Mol. Genet.* **8**, 1387–1396.
13. Schweiger, S., Foerster, J., Lehmann, T., Suckow, V., Muller, Y. A., Walter, G. *et al.* (1999). The Opitz syndrome gene product, MID1, associates with microtubules. *Proc. Natl Acad. Sci. USA*, **96**, 2794–2799.
14. Van Hoof, C. & Goris, J. (2003). Phosphatases in apoptosis: to be or not to be, PP2A is in the heart of the question. *Biochim. Biophys. Acta*, **1640**, 97–104.
15. Trockenbacher, A., Suckow, V., Foerster, J., Winter, J., Krauss, S., Ropers, H. H. *et al.* (2001). MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation. *Nature Genet.* **29**, 287–294.
16. Liu, J., Prickett, T. D., Elliott, E., Meroni, G. & Brautigan, D. L. (2001). Phosphorylation and microtubule association of the Opitz syndrome protein mid-1 is regulated by protein phosphatase 2A via binding to the regulatory subunit alpha 4. *Proc. Natl Acad. Sci. USA*, **98**, 6650–6655.
17. Frangioni, J. V. & Neel, B. G. (1993). Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. *Anal. Biochem.* **210**, 179–187.
18. Schwarzinger, S., Kroon, G. J., Foss, T. R., Wright, P. E. & Dyson, H. J. (2000). Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J. Biomol. NMR*, **18**, 43–48.
19. Wishart, D. S. & Sykes, B. D. (1994). The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using the ¹³C chemical-shift data. *J. Biomol. NMR*, **222**, 311–333.
20. Lee, M. S., Palmer, A. G., 3rd & Wright, P. E. (1992). Relationship between ¹H and ¹³C NMR chemical

- shifts and the secondary and tertiary structure of a zinc finger peptide. *J. Biomol. NMR*, **2**, 307–322.
21. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
 22. Blake, P. R., Lee, B., Summers, M. F., Adams, M. W., Park, J. B., Zhou, Z. H. & Bax, A. (1992). Quantitative measurement of small through-hydrogen-bond and "through-space" ^1H - ^{113}Cd and ^1H - ^{199}Hg J couplings in metal-substituted rubredoxin from *Pyrococcus furiosus*. *J. Biomol. NMR*, **2**, 527–533.
 23. Massiah, M. A., Blake, P. R. & Summers, M. F. (1998). Nucleic Acid Interactive Protein Domains That Require Zinc. In *Bioorganic Chemistry: Peptides and Proteins*, pp. 258–278, Oxford University Press, Oxford.
 24. Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378.
 25. Wang, B., Alam, S. L., Meyer, H. H., Payne, M., Stemmler, T. L., Davis, D. R. & Sundquist, W. I. (2003). Structure and ubiquitin interactions of the conserved zinc finger domain of Npl4. *J. Biol. Chem.* **278**, 20225–20234.
 26. Summers, M. F., Henderson, L. E., Chance, M. R., Bess, J. W., Jr, South, T. L., Blake, P. R. *et al.* (1992). Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Sci.* **1**, 563–574.
 27. Lachenmann, M. J., Ladbury, J. E., Dong, J., Huang, K., Carey, P. & Weiss, M. A. (2004). Why zinc fingers prefer zinc: ligand-field symmetry and the hidden thermodynamics of metal ion selectivity. *Biochemistry*, **43**, 13910–13925.
 28. Berkovits-Cymet, H. J., Amann, B. T. & Berg, J. M. (2004). Solution structure of a CCHHC domain of neural zinc finger factor-1 and its implications for DNA binding. *Biochemistry*, **43**, 898–903.
 29. Bulaj, G., Kortemme, T. & Goldenberg, D. P. (1998). Ionization-reactivity relationships for cysteine thiols in polypeptides. *Biochemistry*, **37**, 8965–8972.
 30. Pinson, L., Auge, J., Audollent, S., Mattei, G., Etchevers, H., Gigarel, N. *et al.* (2004). Embryonic expression of the human MID1 gene and its mutations in Opitz syndrome. *J. Med. Genet.* **41**, 381–386.
 31. Reddy, B. A. & Etkin, L. D. (1991). A unique bipartite cysteine-histidine motif defines a subfamily of potential zinc-finger proteins. *Nucl. Acids Res.* **19**, 6330.
 32. Capili, A. D., Edghill, E. L., Wu, K. & Borden, K. L. (2004). Structure of the C-terminal RING finger from a RING-IBR-RING/TRIAD motif reveals a novel zinc-binding domain distinct from a RING. *J. Mol. Biol.* **340**, 1117–1129.
 33. Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M. C. & Kleit, R. E. (2001). Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. *Nature Struct. Biol.* **8**, 833–837.
 34. Gervais, V., Busso, D., Wasielewski, E., Poterszman, A., Egly, J. M., Thierry, J. C. & Kieffer, B. (2001). Solution structure of the N-terminal domain of the human TFIID MAT1 subunit: new insights into the RING finger family. *J. Biol. Chem.* **276**, 7457–7464.
 35. Bellon, S. F., Rodgers, K. K., Schatz, D. G., Coleman, J. E. & Steitz, T. A. (1997). Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster. *Nature Struct. Biol.* **4**, 586–589.
 36. Hanzawa, H., de Ruwe, M. J., Albert, T. K., van Der Vliet, P. C., Timmers, H. T. & Boelens, R. (2001). The structure of the C4C4 ring finger of human NOT4 reveals features distinct from those of C3HC4 RING fingers. *J. Biol. Chem.* **276**, 10185–10190.
 37. Fang, S., Lorick, K. L., Jensen, J. P. & Weissman, A. M. (2003). RING finger ubiquitin protein ligases: implications for tumorigenesis, metastasis and for molecular targets in cancer. *Semin. Cancer Biol.* **13**, 5–14.
 38. Mallery, D. L., Vandenberg, C. J. & Hiom, K. (2002). Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. *EMBO J.* **21**, 6755–6762.
 39. Zheng, N., Wang, P., Jeffrey, P. D. & Pavletich, N. P. (2000). Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell*, **102**, 533–539.
 40. Dominguez, C., Bonvin, A. M., Winkler, G. S., van Schaik, F. M., Timmers, H. T. & Boelens, R. (2004). Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis, and docking approaches. *Structure (Camb)*, **12**, 633–644.
 41. Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480.
 42. Kuhlbrodt, K., Mouysset, J. & Hoppe, T. (2005). Orchestra for assembly and fate of polyubiquitin chains. *Essays Biochem.* **41**, 1–14.
 43. Legge, G. B., Martinez-Yamout, M. A., Hambly, D. M., Trinh, T., Lee, B. M., Dyson, H. J. & Wright, P. E. (2004). ZZ domain of CBP: an unusual zinc finger fold in a protein interaction module. *J. Mol. Biol.* **343**, 1081–1093.
 44. Xia, Y., Pao, G. M., Chen, H. W., Verma, I. M. & Hunter, T. (2003). Enhancement of BRCA1 E3 ubiquitin ligase activity through direct interaction with the BARD1 protein. *J. Biol. Chem.* **278**, 5255–5263.
 45. Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
 46. Pelton, J. G., Torchia, D. A., Meadow, N. D. & Roseman, S. (1993). Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIIGlc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. *Protein Sci.* **2**, 543–558.
 47. Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
 48. Zhang, O., Kay, L. E., Olivier, J. P. & Forman-Kay, J. D. (1994). Backbone ^1H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR*, **4**, 845–858.
 49. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multi-dimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

Edited by M. F. Summers

(Received 21 December 2005; received in revised form 30 January 2006; accepted 1 February 2006)
Available online 20 February 2006

To be inserted on line 11 on page -X- in place of existing candidate statement

KIERAN M SHORT (Candidate)

Conception and design of experiments. Performed: the full yeast two hybrid screen of $\geq 1 \times 10^6$ clones, directly resulting in the verified MID1-Alpha 4 interaction as presented in Figure 1A. Confirmation of the yeast two-hybrid interaction was provided by co-localisation studies in mammalian cells, showing an *in vivo* interaction between MID1 and Alpha 4, and additionally MID2 and Alpha 4 (Figure 1B). Domain-based interaction analysis was performed using deleted forms of MID1 (Figure 2B) (using constructs already available in the lab, mainly cloned by Blair Hopwood). The domain-deletion studies were confirmed using yeast two-hybrid analysis with vectors cloned by Kieran M. Short (Figure 2A). Dimerisation between MID1 and MID2 was also identified by yeast two-hybrid and immunofluorescence (Figures 3A and 3B, respectively). Identification of the B-box domains as the main MID1-Alpha 4 interacting site were highlighted by expression of the MID1 B-boxes and coiled-coil together (BBCC) or coiled-coil alone (Kieran M. Short; Figure 4B), which was verified by yeast two-hybrid testing by Blair Hopwood (Figure 4A). An improved yeast spotting technique (in methods section, Chapter Two) was discovered toward the latter stages of manuscript preparation (it was better than the previously used 'streaking' technique), at which time Blair Hopwood gave assistance to improve the data quality for publication. Computational analysis and phosphorylation site prediction (Figure 5C) was also performed by Kieran M. Short. Other duties included interpretation and evaluation of data, writing of the manuscript, evaluation and preparation of all figures for publication, and final approval of completed manuscript.

Signed: .

..

Date: 29/1/2007