



**“Investigation of the role
of the *MID1* and *MID2* genes in the
Opitz Syndrome”**

By

Lillian Rose Allen B.Sc.(Hons)

2003

A thesis submitted in fulfillment
for the requirements of the award of,
Master of Science,
University of Adelaide.

Contents

	Page
Contents list	i
Abstract	vi
Acknowledgements	ix
Declaration	x
Abbreviations	xi
List of Figures	xiv
List of Tables	xvii
Chapter One	1
Literature Review	
<u>1.1 Introduction</u>	2
<u>1.2 Congenital Syndromes</u>	3
<u>1.3 Description of OS</u>	5
1.3.1 Inheritance of OS	8
<i>1.3.1.1 Cytogenetics and Mapping of OS</i>	9
1.3.2 Diagnosis and Clinical Features of OS	10
<i>1.3.2.1 Phenotypic Comparison of Xp22 with 22q OS Forms</i>	10
<u>1.4 Other Xp22 and 22q11 Syndromes</u>	12
1.4.1 Comparison of Autosomal OS to del(22)(q11) Syndromes	12
<i>1.4.1.1 The CATCH22 Region</i>	14
1.4.2 Comparison of Xp22 Form to Other Xp22 Syndromes	15
<u>1.5 MID1 Gene Responsible for the OS Phenotype in the X-linked Form</u>	15
1.5.1 The MID1 Gene	15
1.5.2 Midin, the MID1 Protein	16
<i>1.5.2.1 The Tripartite Domain</i>	18
<i>1.5.2.2 The Fibronectin Type III Repeat</i>	19
<i>1.5.2.3 The B30.2-like Domain</i>	19
<i>1.5.2.4 SPRY Domain</i>	20

1.5.3 Mouse <i>Fxy1/Mid1</i> Gene	21
<u>1.6 <i>MID2</i> Gene</u>	22
<u>1.7 Research Objectives</u>	25
Chapter Two Materials and Methods	26
<u>2.1 Restriction Enzyme Digests</u>	27
2.1.1 Agarose Gel Electrophoresis	27
2.1.2 Preparation of Cloning Vectors	27
2.1.3 Removal of 5' Phosphates From Vectors	28
<u>2.2 Ligation Conditions</u>	28
<u>2.3 Transformation</u>	28
2.3.1 Mini and Midi Preparation of Plasmid DNA	
29	
<u>2.4 End Filling/Blunting</u>	29
<u>2.5 DNA Precipitation</u>	30
2.5.1 Ethanol Precipitation	30
2.5.2 Isopropanol Precipitation	30
<u>2.6 Computer Programmes</u>	30
<u>2.7 <i>MID</i> Gene Sequences</u>	31
2.7.1 Evolutionary Analysis of <i>MID</i> Genes	31
<u>2.8 OS Patients</u>	31
<u>2.9 Mutation Analysis</u>	32
<u>2.10 Polymerase Chain Reaction</u>	32
<u>2.11 Automated Sequencing</u>	37
<u>2.12 SSCP Analysis</u>	37

2.12.1 Radiolabelling of PCR Products	37
2.12.2 Loading of Samples	38
<u>2.13 Generation of GFP-MID1 Mutations</u>	38
2.13.1 Transfection of GFP-MID1 Constructs	39
<u>2.14 Isolation of a Zebrafish <i>MID</i> Homologue</u>	39
2.14.1 Isolation of RNA from Zebrafish Embryos	40
2.14.2 RT-PCR	40
<u>2.15 Wholemount <i>in situ</i> Hybridisation of Zebrafish Embryos</u>	41
<u>2.16 Genomic Southern</u>	41
2.16.1 Genomic DNA Extraction from Adult Zebrafish	41
2.16.2 Digests	42
2.16.3 Southern Transfer	42
2.16.4 Probes	43
2.16.5 Hybridisation	43
2.16.6 Washes	44
<u>2.17 Buffers</u>	44
Chapter Three Mutation Analysis in OS patients	50
<u>3.1 Clinical and Cytogenetic Findings of OS Probands</u>	52
<u>3.2 <i>MID1</i> Mutation Analysis</u>	
54	
3.2.1 Truncating Mutations	59
3.2.1.1 Nonsense Mutations	59
3.2.1.2 Frameshift Mutations	64
3.2.2 Exon Deletion	68
3.2.3 Missense Mutation	70

3.2.4 OS Probands With No <i>MID1</i> Mutations	73	
3.2.5 Screening the Un-translated Regions of <i>MID1</i>	73	
<u>3.3 <i>MID2</i> Mutation Analysis</u>		
75		
<u>3.4 Discussion</u>	79	
Chapter Four	Investigation of <i>MID1</i> Mutations	84
<u>4.1 Constructing the GFP/<i>MID1</i>-Mutated Vectors</u>		86
<u>4.2 Intracellular Localisation of Mutant <i>MID1</i> Protein</u>		90
4.2.1 GFP-E115X Analysis		90
4.2.2 GFP- Δ exon2 Analysis		90
4.2.3 GFP-R368X Analysis		93
4.2.4 GFP-L626P Analysis		93
<u>4.3 Discussion</u>		93
Chapter Five	Zebrafish <i>MID</i> Homologue	98
<u>5.1 Zebrafish <i>MID</i> Homologue</u>		101
<u>5.2 Zebrafish Genomic Southern Analysis</u>		104
<u>5.3 z<i>MID</i> Expression During Embryogenesis</u>		108
5.3.1 RT-PCR		108
5.3.2 Wholemount Zebrafish <i>in situ</i> Hybridisation		
110		
<u>5.4 Phylogenetic Analysis of the <i>MID</i> Genes</u>		113
5.4.1 Sequence Alignments of <i>MID</i> Homologues		114
5.4.2 Rooted and Un-rooted Phylogenetic Trees		119
5.4.3 Relative Rates Test		123

<u>5.5 Discussion</u>	124	
Chapter Six	Final Discussion	130
<u>6.1 The Role of MID1 in the X-Linked Form of OS</u>	131	
<u>6.2 The Role of MID2 in the X-Linked Form of OS</u>	133	
<u>6.3 Effects of <i>MID1</i> Mutations</u>	136	
<u>6.4 Correlation Between Genotype and Phenotype</u>	139	
<u>6.5 The Xp22 and Xq22 Regions</u>	140	
<u>6.6 The Zebrafish OS Model</u>	141	
<u>6.7 Conclusion</u>	143	
References	144	
Appendices	162	
Publication	191	

Abstract

To determine whether or not the *MID1* and *MID2* genes contribute to the OS phenotype, a mutation screen was undertaken. Fifteen Australasian and one British patient diagnosed as having Opitz Syndrome were screened for mutations in the *MID1* and *MID2* genes. In total, 8 unique *MID1* mutations (E115X, Δ exon2, F/S 1051delC, R368X, F/S 1330insA, Q468X, R495X & L626P) were identified. The majority of these mutations resulted in the truncation of the MID1 protein, midin (6 of the 8 mutations), disrupting the C-terminal domain (CTD). E115X was the most N-terminal mutation identified, resulting in a midin protein with only the RING finger domain. Two OS patients had nucleotide changes producing a frameshift, one was a deletion of a C nucleotide (1051delC) while the other was an insertion of a single A nucleotide (1330insA). One OS patient had a small in-frame deletion encompassing exon 2 (Δ exon2) and another unique mutation resulted in an amino acid change from a leucine to a proline (L626P), representing the most C-terminal mutation identified so far. In the *MID2* gene, there was a missense mutation caused by a change 1073C>A. This mutation was present in OSP#7, 9 and 13 probands, which suggested that the *MID2* gene may also have a role in the OS phenotype. The OS phenotype varies between individuals but there was no correlation between the OS phenotype and genotype in the mutations examined.

The *MID1* mutations, E115X, Δ exon2, R495X and L626P, were chosen to investigate the effect of the mutations on the intracellular localisation of the MID1 protein (midin). A Green Fluorescent Protein tag (GFP) was used to visualise the intracellular localization of the wild-type and mutated midin. Wild-type midin was located only in the cytoplasm of the cell and was associated with the microtubule network, while the mutated forms of midin were found to have an altered intracellular localisation. The E115X mutated protein lost its ability to localise in the cytoplasm. The Δ exon2 mutated protein remained in the cytoplasm but lost

its ability to associate with the microtubules. Intracellular localisation of R368X and L626P mutated proteins showed cytoplasmic clumping. The R368X and L626P mutated proteins also had a reduced ability to associate with microtubules. This suggests that the mutations result in a loss of function of the ability of midin to bind to the microtubules, overall resulting in the OS phenotype.

An antibody to human midin, MID1 antibody, was raised in rabbits and characterised using Western analysis. Experiments were undertaken to determine the specificity of the MID1 antibody. The MID1 antibody was found to be specific to midin and was unable to interact with the MID2 protein. In addition, the MID1 antibody was unable to interact with midin that had mutations. The MID1 antibody also was unable to interact with tissue samples from zebrafish embryos and chicken embryos although these species have the *MID1* gene present. This suggested that the C-Terminal domain appeared to facilitate the interaction between MID1 antibody and midin.

The zebrafish (*Danio rerio*) was used as a model system in which to find a *MID1* homologue in order to further analyse the function of *MID1* and how it may cause the OS phenotype. From the zebrafish genome a homologue, *zMID*, was isolated and cloned. *In situ* hybridisation experiments revealed that the *zMID* was expressed only in the retinal neuroepithelium cells of the developing eye in zebrafish, highlighting the boundary of the optic stalk before it differentiated into the optic nerve. This expression was very different to the MID1 pattern of expression in mice, which has been shown to be ubiquitous. The multiple banding patterns observed in Southern analysis, when various probes were hybridised to zebrafish genomic DNA, indicated that there were multiple copies of the *MID1* gene in the zebrafish genome. In addition, the banding pattern suggested there might be multiple copies of *MID2* or even the presence of *MID-like* genes in the zebrafish genome. A phylogenetic analysis using the *MID1* and *MID2* homologues revealed that the *zMID* gene and the fugu *MID* gene were a monophyletic group that excluded the other homologues. A

protein alignment of all MID sequences revealed these homologues had been highly conserved across species. The lowest conservation seen across species was in the most carboxy terminus of the MID1 protein.

In the future, all multiple copies of the *MID1* gene present in the zebrafish genome need to be isolated and sequenced to enable transgenic experiments to be carried out in the zebrafish. Transgenic zebrafish could be used to model the mutations found in OS patients and used to determine the mechanisms involved in the loss of function of the *MID1* gene.

Acknowledgements

I would like to thank my supervisors Dr Timothy Cox for the opportunity to undertake this study in his laboratory.

Dr Joan Kelly for the support, guidance and patience through out the last stages of the thesis, I really appreciated your positive approach. And thank you for not giving up on me.

Dr Cindy Bottema for her advice throughout the editing of the thesis, you comments and direction were valuable.

Dr Rory Hope for assistance with the phylogenetic analysis.

Thanks also to everyone in the Cox laboratory, for their assistance with various techniques and experiments.

Thanks to my dear friend Esther for helping me with the zebrafish experiments, I enjoyed working with you.

My dearest friend Dr Philip Shields (or should that be Professor!), thank you for the support you have provided for me, you're a gem.

Thanks also to my family who are always there for me.

Declaration

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

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

Lillian Allen

Abbreviations

A	adenine
bp	base pair
BSA	bovine serum albumin
C	cytosine
°C	degree Celcius
cDNA	complementary deoxyribonucleic acid
CTD	C-terminal domain
dATP	2'-deoxycytosine 5'-triphosphate
dCTP	2'-deoxyguanosine 5'-triphosphate
dH ₂ O	distilled water
ddH ₂ O	de-ionised distilled water
DNA	deoxyribonucleic acid
dpc	days post-copulation
DTT	dithiothreitol
DTTP	2'-deoxythymidine 5'-triphosphate
EDTA	ethylenediaminetetraactetic acid
EtBr	ethidium bromide
FITC	fluorescein isothiocyanate
FNIII	fibronectin type III repeat
Fxy	mouse homologue of MID1
g	gram
G	guanine
GFP	green fluoescent protein

h	human
hpf	hours post-fertilisation
dpf	days post-fertilisation
IPTG	isopropyl-1- β -D-thiogalactopyranoside
kb	kilobase
kV	kilovolts
LB	Luria broth
MID1	midline 1 gene
MID2	midline 1 gene
M	molar
m	mouse
min or '	minutes
ml	millilitre
mM	millimolar
mRNA	messenger ribonucleic acid
NaCl	sodium chloride
NaOH	sodium hydroxide
ng	nanograms
ORF	open reading frame
OS	Opitz syndrome
OSP#	Opitz syndrome patient number
PBS	phosphate buffer saline
PBS	plasmid blue script
PCR	polymerase chain reaction
RE	restriction enzyme

RNA	ribonucleic acid
RT-PCR	reverse transcription-polymerase chain reaction
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate – Polyacrylamide Gel Electrophoresis
SSC	sodium chloride sodium citrate
SSCP	single stranded conformation polymorphism
T	thymine
T _A	annealing temperature
TAE	tris acetate ethylenediaminetetraacetic acid
TBE	tris-borate-EDTA
TE	tris ethylenediaminetetraacetic acid
TEMED	N, N, N', N'-tetramethylethylenediamine
Tris	trizma base (tris(hydroxymethyl) amino methane)
µg	microgram
µl	microliter
µM	micromolar
UV	ultraviolet
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

List of Figures

	Page
Figure 1.1: An Opitz syndrome patient.	6
Figure 1.2: The domain structure of midin.	17
Figure 1.3: Comparison of the MID1 and MID2 proteins.	23
Figure 3.1: The <i>MID1</i> ORF sequence.	55
Figure 3.2: The <i>MID1</i> mutations identified as part of this study.	58
Figure 3.3: The chromatogram showing the OSP#12 mutation E115X in exon 1 of <i>MID1</i> .	61
Figure 3.4: The chromatogram showing the OSP#3 mutation R368X, in exon 5 of <i>MID1</i> .	62
Figure 3.5: The chromatogram showing the OSP#10 mutation Q468X in exon 7 of <i>MID1</i> .	63
Figure 3.6: The chromatogram showing the OSP#9 mutation R495X in exon 8 of <i>MID1</i> .	65
Figure 3.7: The chromatogram showing the OSP#6 frameshift mutation 1051delC in exon 5 of <i>MID1</i> .	66
Figure 3.8: The OSP#6 family members.	67
Figure 3.9: A schematic diagram of the STSs in relation to the <i>MID1</i> gene.	70
Figure 3.10: The <i>MID1</i> deletion of the exon 2 in OSP#11.	71
Figure 3.11: The chromatogram showing the OSP#5 mutation L626P in exon 9 of <i>MID1</i> .	72
Figure 3.12: The <i>MID1</i> gene structure showing the untranslated exons.	74
Figure 3.13: The ORF <i>MID2</i> sequence.	76
Figure 3.14: The <i>MID2</i> open reading frame and protein sequence.	78
Figure 3.15: The exon/intron boundary where the <i>MID2</i> variant was found.	79

Figure 4.1: The The generation of the 343G>T mutation in <i>MIDI</i>	87
Figure 4.2: Construction of the Δ exon2 mutation into the ORF of the <i>MIDI</i> cDNA.	89
Figure 4.3: Intracellular localisation of the wild-type and mutated forms of midin in CosI cells.	92
Figure 5.1: The <i>zMID</i> protein structure.	103
Figure 5.2: The position of the <i>zMID</i> probes used in Southern analysis.	105
Figure 5.3: Zebrafish genomic Southern analysis using the human <i>MID1</i> and <i>MID2</i> probes.	106
Figure 5.4: Zebrafish genomic Southern analysis using the <i>zMID</i> probe.	107
Figure 5.5: <i>zMID</i> homologue expression.	109
Figure 5.6: The position of the <i>zMID</i> probe used for <i>in situ</i> hybridisation.	110
Figure 5.7: <i>In situ</i> hybridization experiments using the antisense <i>zMID</i> .	112
Figure 5.8: Control <i>in situ</i> hybridization experiments using the sense <i>zMID</i> .	113
Figure 5.9: Alignment of the protein sequences of the <i>MID1</i> and <i>MID2</i> homologues.	116
Figure 5.10: Maximum parsimony phylogenetic tree.	121
Figure 5.11: Un-rooted maximum parsimony phylogenetic tree of the <i>MID</i> homologues.	122
Figure 5.12: Sections of the zebrafish eye during development.	146
Figure 5.13: Combined species and gene tree assuming the root is defined by the <i>fMID</i> .	129
Figure 6.1: An overview of the possible functions of the midin domains.	138

List of Tables

	Page
Table 1.1: The major, minor and associated phenotypes seen in the Opitz syndrome.	7
Table 1.2: Percentage occurrence of phenotypes in OS from 23 published cases and 5 probands from 3 families.	8
Table 1.3: Comparison of phenotype frequencies between the X-linked and autosomal forms of OS.	12
Table 1.4: Comparison between the <i>MID1</i> and <i>MID2</i> genes.	24
Table 2.1: Primers used in PCR to screen for mutations in the <i>MID1</i> gene.	34
Table 2.2: Primers used in PCR to screen for mutations in the <i>MID2</i> gene.	34
Table 2.3: <i>MID1</i> primers used in PCR to generate inserts for the various constructs in further experiments.	35
Table 2.4: Primers used to isolate zebrafish <i>MID</i> nucleotide sequence.	36
Table 2.5: The vector primers used in this study.	36
Table 3.1: Summary of the Opitz patient phenotypes and molecular findings involved in this study.	53
Table 3.2: Summary of the mutations found in <i>MID1</i> in this study.	57
Table 5.1: The percentages of identity between the aligned <i>MID</i> homologue protein sequences.	119
Table 5.2: Genes that evolve at significantly different rates are revealed by the relative rates tests.	124
Table 6.1: Summary of the <i>MID1</i> and <i>MID2</i> mutations and <i>MID1</i> polymorphisms found to date.	134

Chapter One:

Literature Review

Literature Review

1.1 Introduction

There is limited knowledge available on the molecular processes involved during mammalian development. Congenital malformations provide a means to study the mechanisms involved in development. Multiple congenital abnormalities are often grouped into syndromes according to characteristic phenotypes affecting major organs during development. Particular congenital abnormalities may include, defects involving limb, craniofacial, heart and urogenital development.

One such multiple congenital syndrome is Opitz Syndrome (OS), which is a heterogeneous disorder characterised by defects of the primary midline developmental field. The phenotype of OS encompasses a broad range of clinical manifestations that usually include craniofacial, heart and genitourinary malformations. Specific features of the OS phenotype include hypertelorism (widely spaced eyes), clefting of the lip/palate, laryngotracheal abnormalities, heart defects (atrial septal defects, patent ductus arteriosus, coarctation of aorta) and genital defects such as hypospadias (Robin *et al*, 1996; Opitz *et al*, 1969a and b). These abnormal phenotypes are the result of disruptions in midline development during embryogenesis.

The heterogeneous nature of OS caused confusion for some time in its classification. Initially two separate entities were described, the G and BBB syndromes, which had overlapping phenotypes and conflicting reports of inheritance patterns (Opitz *et al*, 1969a and b). Recently, it was reported that both of these syndromes co-segregated in a single family, indicating that OS (Opitz GBBB syndrome) is a single entity (McKusick *et al*, 1995) but with an X-linked form (Xp22) and at least one autosomal form (22q11.2) (Robin *et al*, 1995; McDonald-McGinn *et al*, 1995). In addition, cytogenetic analysis of OS patients has suggested chromosomes 5, 13 and 14, may be involved in the OS genotype (Robin *et al*, 1996; Urioste *et al*, 1995; Leichtman *et al*, 1991; Kurczynski *et al*, 1998).

A gene, *MID1*, has recently been cloned and shown to be responsible for the X-linked form of OS (Quaderi *et al*, 1997; Gaudenz *et al*, 1998; Perry *et al*, 1998; Cox *et al*, 2000). *MID1* consists of nine exons spanning a DNA region of ~400 kb (Cox *et al*, 2000) and encodes a predicted protein sequence of 667 amino acids (Quaderi *et al*, 1997; Gaudenz *et al*, 1998; Perry *et al*, 1998). Homology searches with the predicted amino acid sequence have revealed the presence of a tripartite domain, a fibronectin type III repeat (FNIII) and a B30.2 like domain (Cox *et al*, 2000). The functions of these domains in *MID1* are unknown. A mouse homologue of *MID1*, *Mid1/Fxy*, has been isolated, and shown to be expressed ubiquitously in the embryo, except in the heart. The highest level of expression was reported in the pharyngeal arches (Quaderi *et al*, 1997; Palmer *et al*, 1997; Dal Zotto *et al*, 1998; Perry *et al*, 1998). *Fxy* maps to the pseudoautosomal region (PAR) in mice (Palmer *et al*, 1997; Dal Zotto *et al*, 1998). *MID1* has been shown to be associated with microtubules in the cytoplasm of cells (Schweiger *et al*, 1999; Perry *et al*, 1999). In addition, computer based searches have identified another gene, *MID2*, that has a high ~95% DNA homology with to *MID1* (Perry *et al*, 1999; Buchner *et al*, 1999). *MID1* has been located in the Xp22 region, while *MID2* was located on Xq22 (Quaderi *et al*, 1997; Buchner *et al*, 1999). It is unknown whether *MID2* has a function in OS.

1.2 Congenital Syndromes

There are numerous genetic disorders leading to a broad range of genetic defects. Variability within genetic disorders depends upon how the disease genes are inherited and the developmental processes that are consequently perturbed. Certain abnormalities can be corrected by surgery after birth. However, some are too complex and serious for the neonate to survive, for instance, multiple birth defects (e.g. from chromosome abnormalities) affecting major organs such as the heart, liver, kidney, intestines and brain. In some disorders, a single gene may be responsible for the birth defects (e.g. sickle cell anaemia and cystic fibrosis),

while many other disorders have multiple genes affected, making it more complicated to determine the actual molecular process involved.

Many genetic syndromes display characteristic craniofacial defects in addition to other malformations. These syndromes provide a model in which the genetics of normal and abnormal development of the craniofacial region can be researched. For example, achondroplasia (ACH), Apert (AS), Pfeiffer (PS) and Crouzon (CS) syndromes are some of the craniosynostosis syndromes which display an underdevelopment of the midface in addition to other specific craniofacial defects (Muenke and Schell, 1995). In these syndromes (autosomal dominant craniosynostosis syndromes), mutations have been discovered in the fibroblast growth factor receptors (FGFRs #1-3) (Webster and Donoghue, 1997; DeMoerlooze and Dickson, 1997; Muenke and Schell, 1995; Richman, 1995), and investigation of these mutations has provided clues as to the role of the FGFRs. The investigation of genetic syndromes could be used to establish the roles of the relevant genes in normal mammalian development (Webster and Donoghue, 1997).

The autosomal region, 22q11, has had a number of similar phenotypic syndromes assigned to it. This region is referred to as the CATCH 22 region. Phenotypic characteristics include craniofacial, heart and limb defects (Section 1.4.1.1).

Similarly, the Opitz syndrome (Section 1.3) is a multiple congenital disorder with craniofacial abnormalities for which the underlying molecular defect has recently been determined. If genes that are expressed along the midline are perturbed, then the primary midline developmental field may also be affected and consequently affect the midline organs (Opitz and Gilbert, 1982). The midline refers to an imaginary bilateral plane or field through the embryo. Organs such as the heart and central nervous system (CNS) develop from this region and are termed midline organs. The *MIDI* gene that underlies OS is expressed in the midline of embryos. Investigation of the role of the *MIDI* gene in embryo development could

be useful in providing clues to the process of development and ways to better diagnose and prevent OS.

1.3 Description of OS

There is a broad range of phenotypic defects in OS patients that differ in severity between individuals and within the same family (Table 1.1, Figure 1.1). As mentioned, the distinct array of OS clinical features includes hypertelorism, clefting of the lip/palate, laryngotracheal abnormalities, various cardiovascular malformations (eg. atrial septal defects, patent ductus arteriosus, coarctation of aorta), swallowing difficulties, respiratory problems and genital defects (eg. hypospadias) (McKusick *et al*, 1995; Opitz, 1987). The frequency of occurrence of the OS features observed have been summarised in studies by Wilson and Oliver (1988) and Robin *et al* (1995, 1996) (Table 1.2). The information was gathered from 5 probands in 3 families and 23 previously published cases. The most widespread phenotype observed was hypospadias in males, followed by hypertelorism, while mental retardation was not a common manifestation. Originally, OS was referred to as the Hypertelorism-Hypospadias syndrome because of the frequency with which these phenotypes had been observed (Stevens and Wilroy, 1998).



Figure 1.1: An Opitz syndrome patient. This child has some typical characteristics of the OS phenotype, such as hypertelorism, a prominent forehead and clefting of the lip. The nasal bridge is broad and ears are low set and posteriorly rotated.

Table 1.1: The major, minor and associated phenotypes seen in the Opitz syndrome (modified from Robin *et al*, 1996).

	Major Features	Minor Features	Possible Features
Facial	Hypertelorism Epicanthic folds Slanted, narrow palpebral fissures Flat broad nasal bridge Anteverted nares Low set posteriorly rotated ears Widows peak	Strabismus Cleft lip/palate High arched palate Short lingual frenulum Cleft uvula Bifid tongue Grooved tip of nose Philtrum flattened, elongated Abnormally positioned teeth hypodontia Web neck	Agensis of nasal cartilage Anosmia
Gastrointestinal	Swallowing abnormalities/dysphagia Oesophageal dysmotility Gastro-oesophageal reflux	Imperforate/ectopic anus Umbilical hernia Hiatus hernia Tracheo-oesophageal fistula Diastasis recti Laryngeal cleft Short trachea/high carina Laryngeal hypoplasia Bronchial stenosis Pulmonary hypoplasia/agenesis	Gall bladder agensis Duodenal stricture Failed fusion of mesocolon
Respiratory	Hoarse cry/voice Stridor Respiratory distress Aspiration	Cryptorchidism Renal/ureteric abnormalities Bifid scrotum Splayed labia majora Dermatoglyphic changes	Scalp defects
Genito-urinary	Hypospadias +/- chordae	Frontal/parietal/occipital bone prominence Micrognathia Long fingers/toes Clinodactyly Pectus excavatum/carinatum Brachycephaly ASD/VSD PDA Coarctation of aorta PS Systemic venous drainage anomalies	Craniosynostosis
Skin			
Skeletal	Wide fontanelles/sutures		
Cardiovascular			
Central Nervous System		Mild mental retardation Cerebellar vermal hypoplasia Dandy Walker anomaly Agenesis/hypoplasia of corpus callosum Cerebral atrophy/ventriculomegally Obstructive hydrocephalus Generalised reduction in white matter Wide cavum septum pellucidum Macro cisterna magna Enlarged IV ventricle Hyperreflexia EEG abnormalities	Cholinergic dysfunction
General	Failure to thrive		Immune mediated platelet dysfunction
Perinatal		Mortality Polyhydramnios Hydrops	

PS: Pulmonary stenosis; VSD: Ventriculoseptal defects; ASD: Atrial septal defect; PDA: Patent ductus arteriosus

(Tolmie *et al*, 1987; Stevens and Wilroy, 1988; Wilson and Oliver, 1988; Neri *et al*, 1987; Dunder *et al*, 1995; Guion-Almeida and Richieri-Costa, 1992; MacDonald *et al*, 1993; Opitz, 1969a and b, 1987; Williams and Frias, 1987; Einfeld *et al*, 1987; Farndon and Donnai, 1983; Chemke *et al*, 1984; Stoll *et al*, 1985; Cappa *et al*, 1987; Verloes *et al*, 1989; Christodoulou and Loughnan, 1990; Fryburg *et al*, 1996; McDonald-McGinn *et al*, 1995; Urioste *et al*, 1995; Neri *et al*, 1987; Kasner *et al*, 1974; Robinson and Hilger, 1989; Pedersen *et al*, 1976; Brooks *et al*, 1992; Coburn, 1970; Van Biervliet and Van Hemel, 1975; Cote *et al*, 1981; Christian *et al*, 1969; Fryns *et al*, 1992; Verloes *et al*, 1995).

Table 1.2: Percentage occurrence of phenotypes in OS from 23 published cases and 5 probands from 3 families (modified from Wilson and Oliver, 1988; Robin *et al*, 1995, 1996).

Phenotype	Observed Percentage (%)
Hypertelorism	89
Esophageal dysmotility	69
Laryngotracheal clefts	44
Cleft palate or bifid uvula	34
Heart defects	29
Hypospadias	100 in males
Renal or ureteral anomalies	42
Mental retardation	38

1.3.1 Inheritance of OS

The inheritance of OS was originally suggested to be autosomal dominant as described by Kasner *et al* (1974) where a female suffering from OS had been born into the family initially reported by Opitz *et al* (1969a). At this time OS was known as the G Syndrome, and the BBB syndrome was considered a separate entity (Opitz *et al*, 1969 a and b). Additional cases described over the ensuing years continued to support the theory that the inheritance of the G syndrome was autosomal dominant (Frias and Rosenbloom, 1975; Pedersen *et al*, 1976; Cordero and Holmes, 1978; Arya *et al*, 1980; Chemke *et al*, 1984). During this time male-to-male transmission was reported, providing further evidence of autosomal dominant inheritance (Farndon and Donnai, 1983; Stoll *et al*, 1985; Chemke *et al*, 1984). There were also cases where males were more severely affected with OS than females (Cappa *et al*, 1987; Christian *et al* 1969, Opitz *et al*, 1969a and b). It was noted that the OS phenotype varied among males of the same family, adding to the complexity of this syndrome (Opitz *et al*, 1987; Robin *et al*, 1996).

McDonald-McGinn *et al* (1995) reported a family in which OS was paternally inherited. In two of the family members, a vascular ring was present. Vascular rings are often found when there are also deletions of the 22q11.2 region (Zackai *et al*, 1995). Upon

further investigation McDonald-McGinn *et al* (1995) found a deletion in the 22q11.2 region of these OS patients, and in one other affected family member. They noted that there was a significant overlap of phenotypes between OS and the 22q11.2 microdeletion syndrome (see Section 1.4.1.1). Based on this overlap of manifestations and 22q11.2 deletions in OS, they showed that in some cases OS was due to a 22q11.2 deletion. This was further confirmed by a report of a 22q11 deletion in an OS patient by Lacassie and Arriaza (1996).

As more cases became available and follow up studies were carried out, it was concluded that there was an X-linked inheritance pattern localised to the Xp region (Verloes *et al*, 1995; May *et al*, 1997), as well as an autosomal form localised to the 22q chromosomal region in OS (Robin *et al*, 1995; Fryburg *et al*, 1996; Lacassie and Arriaza, 1996; McDonald and McGinn *et al*, 1995). A large pedigree was reported by Verloes *et al* (1995) where OS co-segregated with a pericentric inversion of the X chromosome (inv(X)(p22.3q26). The recent identification of the *MIDI* gene on Xp22 has prompted investigation of the function of this gene in the X-linked inheritance of OS (Quaderi *et al*, 1997; Palmer *et al*, 1997; Dal Zotto *et al*, 1998; Gaudenz *et al*, 1998).

1.3.1.1 Cytogenetics and Mapping of OS

Cytogenetic analysis has identified additional autosomal chromosomes (5, 13, and 14) that could potentially contain genes responsible for OS. Leichtman *et al* (1991) described a proband with OS who had a partial duplication of 5p. They suggested that a candidate gene for OS may also be located in this duplicated region 5p13-p12. A terminal deletion of chromosome 13, with a breakpoint at 13q32.3, in an OS patient has also been described by Urioste *et al* (1995). Recently, an OS proband was reported as having an unbalanced der(3)t(3;14)(q29;q11.2).14 karyotype, implicating chromosome 14q11.2 as possibly containing a candidate gene(s) for OS (Kurczynski *et al*, 1998). The gene(s) that may

underlie OS in these regions (5, 13 and 14) further extend the genetic heterogeneity of the Opitz syndrome.

1.3.2 Diagnosis and Clinical Features of OS

When OS was considered to be two separate entities, the G and BBB syndromes, the presence of laryngotracheoesophageal anomalies, respiratory problems, imperforated anus and swallowing difficulties indicated the diagnosis of the G syndrome (Opitz *et al*, 1969a and b; Allanson, 1988; Robin *et al*, 1995). A broad, flat nasal bridge with nares anteverted was also considered suggestive of G syndrome, as patients with BBB syndrome frequently had a high and broad nasal bridge (Parisian and Toomey, 1978). However, Cappa *et al* (1987) suggested that there were actually no qualitative differences to discriminate between the G and BBB syndromes and that they should be considered as one entity. In support of this was the observation that G and BBB syndromes segregated together in the same family (Cordero and Holmes, 1978; Cappa *et al*, 1987; Sedano and Gorlin, 1988).

Consequently, OS is now recognised as a single entity, but genetically heterogeneous (Verloes *et al*, 1989; Cappa *et al*, 1987; Robin *et al*, 1996). Diagnosis is generally made on the basis of the presence of a combination of the characteristic features such as hypertelorism, clefting of the lip/palate, and heart and genital defects (Hogdall *et al*, 1989; Schrandt *et al*, 1995; Opitz, 1987). In addition, cytogenetic and genetic linkage analysis can be utilised to determine if the form is X-linked or autosomal (Quaderi *et al*, 1997; Robin *et al*, 1996).

1.3.2.1 Phenotypic Comparison of Xp22 with 22q OS Forms

Robin *et al* (1996) compared the phenotypes of OS affected individuals between the Xp22 and 22q forms reported in the literature. They found that the most common abnormality in the X-linked and autosomal forms was hypertelorism. There were also a number of anomalies that appeared to be more prominent in one form than the other (Table

1.3). For example, hypospadias were more frequently observed in males of the X-linked form (9/10) as opposed to those families showing autosomal linkage (5/16). Even though in the study by Robin *et al* (1996) no imperforated anus phenotype was seen in 22q forms, it has previously been observed in families with male-to-male transmission, and therefore, may not be an exclusive phenotype of the Xp22 form of OS (Tolmie *et al*, 1987). Laryngotracheal manifestations were found at similar frequencies in both forms. However, it was observed that two minor manifestations were seen only in X-linked OS patients: anteverted nares and posterior pharyngeal clefting (Opitz *et al*, 1969a). However, further evidence is required to confirm that these phenotypes are exclusive to the X-linked form. This need has been partially addressed in this study (see Chapter 3). There were other minor manifestations, such as a prominent forehead, which were not seen exclusively in either of the forms. Females from the X-linked form had a lower frequency of defects, such as dysphagia, than males. Furthermore, females who carried the OS allele also had a lower number of anomalies than the males (Robin *et al*, 1996). This was likely the result of females having two copies of the X chromosome (and thus two copies of the gene responsible for the OS phenotype). If one copy of the allele was normal while the other was an abnormal copy, the OS abnormalities would not be as severe. Males only have one X chromosome and since the gene responsible for the OS phenotype is on the X chromosome, if the abnormal allele is present, the number and severity of the abnormalities is increased. For the OS autosomal form, differences in phenotypic severity between males and females were not as distinct as those observed with the X-linked counterpart (eg. cleft lip/palate) (Robin *et al*, 1996).

Table 1.3: Comparison of phenotype frequencies between the X-linked and autosomal forms of OS (Robin *et al*, 1996).

Manifestation	X-linked	Autosomal
Hypospadias	9/10	5/16
Imperforated anus	4/10 males	not present
Dysphasia	7/9 males 2/13 females	3/13 males 1/8 females
Anteverted nares	4/6 males 4/13 females	not present
Broad prominent forehead	7/8 males 4/12 females	11/13 males 7/9 females
Cleft lip/palate	3/7 males 0/14 females	3/15 males 2/7 females

1.4 Other Xp22 and 22q11 Syndromes

The Xp22 and 22q11 forms are of particular interest because other syndromes with similar phenotypes to OS have also been assigned to these regions. These include Craniofrontonasal Syndrome (CFNS) (Feldman *et al*, 1997) and CATCH 22 (Glover, 1995). The main clinical manifestations for CFNS include coronal synostosis, widely spaced eyes, clefting of the nasal tip and various skeletal anomalies. In the CATCH 22 group, the DiGeorge Syndrome is the most common genetic cause of birth defects (Glover, 1995; Demczuk *et al*, 1996).

1.4.1 Comparison of Autosomal OS to del(22)(q11) Syndromes

In half of the families studied by Robin *et al* (1996) OS was linked to markers that localised to 22q. The 22q region has a number of syndromes which also have 22q11.2 deletions: velocardiofacial (VCF) (OMIM #192430), DiGeorge Syndrome (DGS) (OMIM #188400) (Glover, 1995; Demczuk *et al*, 1996) and conotruncal anomaly/face syndrome (CFS) (Driscoll *et al*, 1992 a and b; Burn *et al*, 1993; Emmanuel *et al*, 1998). The main

characteristics seen in the DGS include hypocalcaemia, thymic hypoplasia and defects in the outflow tract of the heart (e.g. tetralogy of Fallot, truncus arteriosus and interrupted aortic arch) (OMIM #188400). DGS is also heterogeneous, like OS (Chisaka and Capecchi, 1991). The majority of the DGS cases have a deletion of 22q11.2 resulting in the loss of a number of genes. The region deleted is typically 3Mb in size (OMIM #188400). For example, TUPLE1, a putative transcription factor, is often deleted, placing it as a possible candidate for this syndrome. VCF has the following characteristics: cleft palate, cardiac anomalies, typical facies and learning problems, in addition to less common phenotypic features such as microcephaly (OMIM #192430). The 22q11 deletion seen in DGS has also been reported in VCF patients (OMIM #188400). It is clear that there is a large phenotypic overlap between the DGS and VCF, as well as an etiologic connection due to the deletion of the same chromosomal region, 22q11.2 (OMIM #19430 and OMIM 188400).

The VCF and DGS syndromes overlap in the type of congenital heart defects seen as well as in their craniofacial phenotypes. The DGS/VCF syndromes have conoventricular congenital heart defects (specifically interrupted aortic arch, coarctation of the aorta (CoA), ventriculoseptal defects (VSD), tetralogy of Fallot (TOF), vascular rings (Goldmutz *et al*, 1993) and pulmonary stenosis (PS) (Lipson *et al*, 1991; Seaver *et al*, 1994). Similar cardiac disorders have been observed in OS patients (Table 1.1), such as PS and VSD (Stoll *et al*, 1988), aortic stenosis, PS and CoA (Farndon and Donnai, 1983). The most common congenital heart disease in OS is patent ductus arteriosus (PDA), which is rarely seen in DGS/VCF (Robin *et al*, 1996). Developmental delay is observed in almost every DGS/VCF patient; however, in autosomal OS probands it is not common (eg. 6/25) (Robin *et al*, 1996; Goldberg *et al*, 1993). These syndromes are part of what has been referred to as the CATCH22 region.

1.4.1.1 The CATCH22 Region

CATCH22 (Cardiac defect, Abnormal facies, Thymic hypoplasia, Cleft palate, Hypocalcaemia and 22q11 anomalies) encompasses a number of distinct syndromes, which have overlapping phenotypes and deletions in the chromosome 22q11 region (Glover, 1995). As the acronym implies, heart, craniofacial, cleft palate, thymic hypoplasia and hypocalcaemia are the main phenotypes commonly seen across these syndromes. A number of these phenotypes, such as heart and craniofacial abnormalities, are also seen in Opitz syndrome. CATCH 22 syndromes are usually more complex as a result of a deletion, thus producing an array of different malformations not seen in OS, including hypocalcaemia and thymic hypoplasia (Wilson *et al*, 1994; Glover, 1995). Syndromes in the CATCH22 region are believed to be the major source of birth defects at an incidence estimated at >1 in 5,000 (Glover, 1995). The 22q11 region is large (~1.5 Megabases) making it difficult to find the responsible gene(s). This is also the most common deletion syndrome seen in humans (Wilson *et al*, 1994). It is believed that the deletion results in a reduced number of genes in this region, thus producing the CATCH22 phenotype (Kelly *et al*, 1993; Scambler *et al*, 1992).

The CATCH22 syndromes are of particular interest because not only do they have a similar phenotype to OS, but the autosomal form of OS has also been assigned to the 22q11 region (Robin *et al*, 1995, 1996). Therefore, the CATCH22 critical region is likely to contain the gene(s) responsible for the autosomal form of OS. Within the CATCH22 critical region, there are a number of candidate genes for the CATCH22 phenotype (or even the autosomal form of OS). Candidate genes include HIRA/TUPLE1 (Halford *et al*, 1993; Lamour *et al*, 1993), IDD (Demczuk *et al*, 1995; Wadey *et al*, 1995), ES2 (Lindsay *et al*, 1998) and goosecoid-like (Gottlieb *et al*, 1997).

1.4.2 Comparison of the Xp22 Form of OS to Other Xp22 Syndromes

The X-linked form of OS is the third craniofacial disorder to be localised to Xp22. Other syndromes that have been localised in this region include CFNS (McPherson *et al*, 1991) and a novel multiple congenital anomaly (MCA) syndrome (Wittwer *et al*, 1994). CFNS has manifestations that are also observed in OS patients. For example, CFNS and OS patients both have craniofacial asymmetry and a grooved nasal tip (Opitz, 1987). Linkage analysis on MCA patients localised a candidate gene to Xp22 (Wittwer *et al*, 1994), and a CFNS patient has been reported as having an Xp22 deletion (McPherson *et al*, 1991). Even though OS and CFNS have these similarities (phenotype and localisation to Xp22), it is believed that separate loci are responsible for each disorder (Muenke *et al*, 1997 and unpublished data).

1.5 *MID1* Gene Responsible for the OS Phenotype in the X-linked Form

Genetic linkage studies have revealed that OS is a heterogeneous disorder linked to the marker DXS987 in Xp22 in three families and to the marker D22S345 from the 22q11.2 region in five families (Robin *et al*, 1995). An OS locus was further localised by genetic linkage to DXS7104 in Xp22 (May *et al*, 1997), in a candidate region of 4cM. Verloes *et al* (1995) described an X-linked OS family where there was co-segregation with a pericentric inversion of the X-chromosome $\text{inv}(X)(p22.3q26)$. The *midline 1* (*MID1*) gene has been localised in the Xp22 region (Quaderi *et al*, 1997; DeConcilliis *et al*, 1998; Van de Veyver *et al*, 1998). Mutations in this gene from OS probands further confirm *MID1* to be responsible for the X-linked form (Quaderi *et al*, 1997; Gaudenz *et al*, 1998; Cox *et al*, 2000).

1.5.1 The *MID1* Gene

There are nine coding exons in the *MID1* gene, as well as an alternatively spliced 5' region covering a ~400kb region of DNA (Cox *et al*, 2000; Van den Veyver *et al*, 1998). The

gene encodes a protein sequence of 667 amino acids. *MID1* does not escape X-inactivation in humans (Dal Zotto *et al*, 1998). Homology searches have indicated that the MID1 protein belongs to a group that includes transcriptional regulators. The transcriptional regulators contain protein-protein interacting domains, which have been implicated in fundamental processes such as body axis patterning (Shou *et al*, 1996) and cell transformation (Grignani *et al*, 1994). Northern blots have shown that MID1 is ubiquitously expressed in foetal and adult human tissue (Quaderi *et al*, 1997). In foetal tissue, MID1 had highest expression in the kidney, then the brain and lung. Adult human tissue showed that MID1 expression was highest in the heart, placenta and brain (Quaderi *et al*, 1997). A mouse homologue of the *MID1* gene, *Fxy*, has been isolated (see Section 1.6.3) (Palmer *et al*, 1997; Dal Zotto *et al*, 1998; Van den Veyver *et al*, 1998).

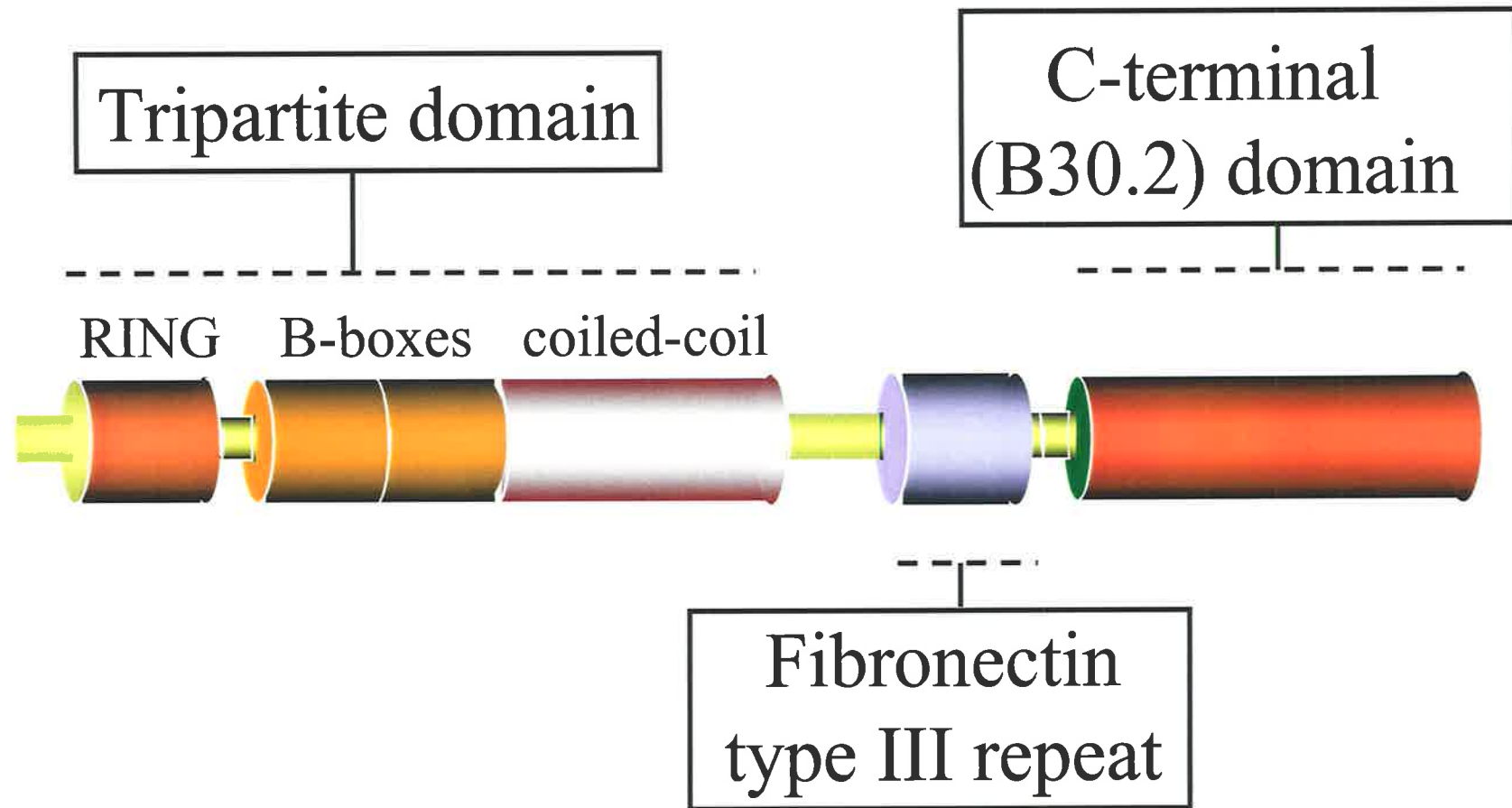
1.5.2 Midin, the MID1 Protein

Based on amino acid homology searches, the MID1 protein, midin, can be divided into the following domains; the tripartite domain, the fibronectin type III repeat (FNIII) and the C-terminal or B30.2, domain (Figure 1.2) (Quaderi *et al*, 1997; Gaudenz *et al*, 1998; DeConcilliis *et al*, 1998). In midin, the tripartite motif also known as RBCC consists of a RING finger, two B-boxes and a coiled-coil domain. The FNIII region was not originally identified and the sub-domain known as SPRY in the B30.2 domain was overlooked (Perry *et al*, 1999; Cox *et al*, 2000). Midin has been found to be highly conserved across species with amino acid similarity nearly 100% between human, mouse and chicken. The exact amino acid homology between these species is slightly lower, ranging between 91 and 95 percent.

The exact functions of the tripartite, FNIII and CTD motifs in midin are yet to be determined. However, the presence of these domains in other developmental genes enables speculation as to the role(s) of the motifs in MID1.

Figure 1.2: The domain structure of midin. There are three main domains in midin: the tripartite domain which is further divided into RING-Finger, B-Boxes and coiled-coil motifs, a fibronectin type III repeat and a C-terminal domain also known as the B30.2-like domain.

MID1 Domains



1.5.2.1 The Tripartite Domain

Midin has a characteristic RING-finger motif which consists of a conserved Cys/His cluster, responsible for joining the zinc atoms in a tetrahedral form (Gaudenz *et al*, 1998; Borden and Freemont, 1996). Diverse cellular roles are associated with this RING-finger family, such as putative transcriptional activators, proto-oncogenes, developmental and viral replication genes (Freemont *et al*, 1991 & 1993; Saurin *et al*, 1996). This RING finger family is further divided based on a second zinc-binding, B-box motif (of which MID1 contains two) (Borden *et al*, 1996; Reddy *et al*, 1992; Freemont, 1993; Saurin *et al*, 1996). Genes that require these motifs for normal function include proto-oncogenes such as promyelocytic leukaemia protein (PML) and proteins with developmental functions (e.g. in the regulation of cell growth and differentiation in the embryo). The main group of developmental proteins includes PwA33, Xnf7, RPT-1 and efp. *Pleurodeles waltli* PwA33 is required for normal function of oocytes of a newt (Bellini *et al*, 1995), and the *Xenopus laevis* Xnf7 is involved in dorsal ventral patterning (Borden *et al*, 1995; Shou *et al* 1996; Reddy *et al*, 1991). The efp (estrogen-responsive transcription factor) is a transcription factor, RPT-1 (Inoue *et al*, 1990) and rfp (ret finger protein) have expression at specific stages of spermatogenesis (Borden *et al*, 1996; Cao *et al*, 1997). All of these have a basic tripartite domain, with one RING-finger, one B-box and one coiled-coil motif (RBCC).

By analysing developmental genes that have a similar structure to MID1, the roles could provide clues on how MID1 functions. For example, the functions of the rfp tripartite domains have been predicted to be involved at least in part in protein-protein interactions or structure. This information could be used to provide clues on the process occurring with the MID1 protein. Cao *et al* (1997) have shown that rfp homomultimerisation occurs through the coiled-coil domain. The B-box was found not to be responsible for this interaction; however, its structural integrity was required for homomultimerisation to occur normally. This

suggested that the B-box motif and coiled-coil domains are involved in the regulation of protein-protein interactions.

1.5.2.2 The Fibronectin Type III Repeat

The FNIII domain can be found in proteins with various functions. These include proteins involved in cell adhesion, cell surface hormone and cytokine receptors and chaperonins (Hunkapiller and Hood, 1989; Campbell and Spitzfaden, 1994; Plaxco *et al*, 1996; Spitzfaden *et al*, 1997). It is believed that the FNIII domain may mediate specific protein-protein interactions. As with the tripartite domain, its exact function has not been determined. The FNIII domain in midin has closest structural similarity to the FNIII motif in the striated muscle structural protein, M-protein. M-protein is found in fast-muscle fibres (Carlsson *et al*, 1990) and belongs to a large family of proteins found in muscles. Data from proteins that have the FNIII repeat may provide information, such as likely protein structure this motif in MID1.

1.5.2.3 The B30.2-like Domain

The B30.2-like domain (or the C-terminal domain, CTD) of MID1 can be found in a variety of unrelated proteins. The B30.2-like domain has been reported to associate with two different types of N-terminal domains. In the case of butyrophilin and BT2 proteins, association is with an immunoglobulin domain (Jack and Mather, 1990; Henry *et al*, 1997); while for rfp and a number of other RBCC proteins, interaction occurs with a tripartite domain (Takahashi *et al*, 1988). However, in all proteins, the function of the B30.2-like domain remains to be determined.

The B.30.2 domain is present in transmembrane proteins (eg. butyrophilin and BT2) intracellular proteins (eg. RFP and Ro/Ssa) and secreted proteins (eg. Stonustoxin a and b, which do not have the RBCC domain). Henry *et al* (1997) conducted database searches in

relation to the B30.2 domain using three consensus motifs, LDP, WEVE and LDYE. These three highly conserved motifs in the B30.2 domains could have a role in the binding of ligand(s) and or domain folding.

Mutations in the B30.2 regions of a number of genes that result in pathological conditions, suggest a possible function for this domain (Henry *et al*, 1998). These conditions include the autosomal recessive disorder, familial Mediterranean fever with mutations in the B30.2 region of pyrin/marenostrin (French FMF Consortium, 1997; International FMF Consortium, 1997) and the heterogeneous disorder Opitz syndrome with novel mutations perturbing the B30.2 region (Quaderi *et al*, 1997, Gaudenz *et al*, 1998, work generated from this thesis).

1.5.2.4 SPRY Domain

Within the B30.2 domain there is a subdomain known as the SPRY domain (Seto *et al*, 1999), which has been identified in a number of proteins including efp, PwA33, Xnf7, butyrophilin and MID1/MID2. The SPRY motif was identified recently using the alignment of numerous RBCC proteins, butyrophilin, and stonustoxin a&b and then identifying the conserved amino acids (Schultz *et al*, 1998; Seto *et al*, 1999). Its role remains unknown. There have been some suggestions of potential functions for the SPRY domain-containing proteins, which include RNA-binding, signal transduction, and cell growth, suppressor of cytokine signaling and differentiation (Seto *et al*, 1999; Ponting *et al*, 1997). For example, the SOCS family of proteins have the SPRY domain. The SOCS (suppressor of cytokine signaling) proteins are a recently identified family of proteins (SOCS #1-7). SOCS #1 & 3 have been shown to negatively regulate the cytokine signaling pathway (Nicholson *et al*, 1999; Hilton, 1999). Analysis of the folding of the SPRY domain indicates that the C-terminal region may adopt a single Ig-like fold, suggesting proteins with SPRY domains are members of the immunoglobulin super-family (Seto *et al*, 1999).

1.5.3 Mouse *Fxy1/Mid1* Gene

The *Fxy1* (or *Mid1*) gene is the mouse homologue of the human *MID1* gene (Palmer *et al*, 1997; Dal Zotto *et al*, 1998; Van den Veyver *et al*, 1998). The *MID1* and *Fxy1* homologues contain identical gene structures (nine coding exons) and protein structure (tripartite, FNIII and B30.2 domains). Whilst these two genes are highly conserved, there are a number of slight differences. Dal Zotto *et al* (1998) showed that, unlike *MID1*, *Fxy1* was not subjected to X-inactivation. The *MID1* gene is the first example of a gene subject to X-inactivation in humans but not in mice. Northern blot experiments showed that *Fxy1* was ubiquitously expressed, except in the heart, during adulthood (Quaderi *et al*, 1997; Palmer *et al*, 1997; Dal Zotto *et al*, 1998; Perry *et al*, 1998). Quaderi *et al* (1997) conducted *in situ* hybridisation experiments and found that *Fxy1* had higher expression within the frontonasal processes, the branchial arches and the central nervous system (CNS) when using E10.5 embryos. Dal Zotto *et al* (1998) used older embryos (E12.5 to E16.5) and found that *Fxy1* expression was present in the CNS, the mucosa of the oropharynx, oesophagus, trachea and larynx, the stomach, gut and urogenital system. These tissues are often abnormal in OS patients. Generally, *Fxy1* expression was confined to proliferating cells. From these data, it has been proposed that the mutated form of *MID1* may affect cell proliferation, which in turn disrupts organ development along the midline, producing the OS phenotype (Dal Zotto *et al*, 1998).

Another difference between the *MID1* and *Fxy1* homologues was the localisation of *Fxy1* on the pseudoautosomal region (PAR) in mice hence escaping X-inactivation (Palmer *et al*, 1997; Dal Zotto *et al*, 1998). Fluorescence *in situ* hybridisation (FISH) and RT-PCR experiments have shown that in *Mus musculus* the first three exons (5') of *Fxy1* were found on the X chromosome, while the remaining 3' region was present on both the X and Y chromosomes (Palmer *et al*, 1997; Dal Zotto *et al*, 1998). When these experiments were

repeated using *Mus spretus*, a line of mice that is not inbred like *M. musculus*, *Fxy1* was found to be located entirely on the X chromosome and was subject to X-inactivation (Perry *et al*, 1998; Dal Zotto *et al*, 1998). There have been numerous genetic differences identified between *M. spretus* and *M. musculus*, suggesting that the chromosomal rearrangement at the PAR has been important in mammalian evolution and speciation (Perry *et al*, 1999; Ellison *et al*, 1996).

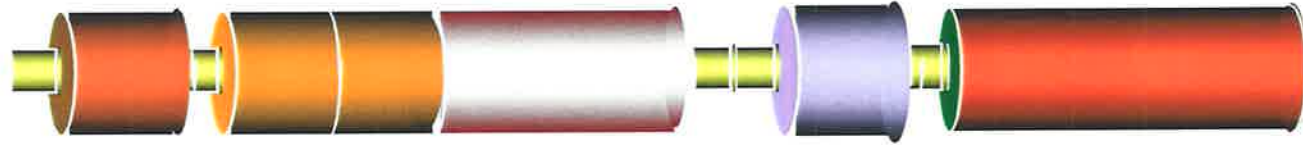
1.6 MID2 Gene

The *MID2* gene was initially identified from computer database searches (Perry *et al*, 1999; Buchner *et al*, 1999) and is of interest because of its high homology and overall similarity to *MID1* (Table 1.4). Similar to *MID1*, the *MID2* gene has nine coding exons, producing a ~2.1kb mRNA which encodes a slightly longer amino acid sequence than *MID1* of 685. Each of the tripartite, FNIII, B30.2 and SPRY domains are present (Perry *et al*, 1999). Individual domains have similarities of 85% for the RING finger, 79% for B-boxes, 70% for coiled-coil, 87% for FNIII and 68% for B30.2 (Figure 1.3) (Perry *et al*, 1999). Overall, the *MID1* and *MID2* proteins are highly similar, with 76% identity and 82% similarity at the amino acid level and there is 71% identity at the nucleotide level (Perry *et al*, 1999).

Northern blots were used to investigate *MID2* expression in human foetal and adult tissues (Buchner *et al*, 1999; Perry *et al*, 1999). Foetal tissues showed low *MID2* expression in the kidney and lung. For the adult tissues there was low *MID2* expression in the prostate, ovary and small intestine. Overall, *MID2* was ubiquitously expressed, but at lower levels than *MID1* (Buchner *et al*, 1999). A mouse homologue, *Fxy2*, has also been identified. Mouse *in situ* hybridisation experiments showed the *MID2* level of expression to be low overall and confined to the CNS and developing heart. During organogenesis, *MID2*

Figure 1.3: Comparison of the MID1 and MID2 proteins. The identity between individual MID1 and MID2 domains was relatively high, with the lowest percentage seen in the CTD. The similarity between each domain was very high, close to 100% in most cases again the CTD had the lowest percentage. Overall, the amino acid identity between MID1 and MID2 was 75% and the amino acid similarity was 95.2%.

MID1



% identity

85

79.5

70

87.5

68

% similarity

98.5

97

96

97

94

MID2



Overall: 76.0% identity; 95.2% similarity

expression level was higher in the mucosal lining of the stomach, the thymus, thyroid, kidney, nasal and oral cavity epithelia and eye (Buchner *et al*, 1999).

Table 1.4: Comparison between the *MID1* and *MID2* genes.

Characteristic	<i>MID1</i>	<i>MID2</i>
Localisation	Xp22	Xq22
Pathogenic of OS	Yes	Unsure
Length of Protein	667 amino acids	685 amino acids
Expression Pattern	Ubiquitously expressed, except in the heart	Ubiquitously expressed, lower levels of expression compared with <i>MID1</i>
Mouse Homologue	<i>Fxy</i> , in the PAR region	<i>Fxy2</i> , syntenic region of mouse X chromosome
Protein	Localised in the cytoplasm and associates with microtubules	Localised in the cytoplasm and associates with microtubules

The main differences between these two genes are the localisation and expression pattern. *MID2* was localised on the long arm of the X chromosome, Xq22, in humans and to the syntenic region of the mouse X chromosome (Perry *et al*, 1999), as opposed to the human Xp22 and pseudoautosomal region (PAR) in mice for *MID1* (Quaderi, *et al*, 1997; Palmer *et al*, 1997; Dal Zotto *et al*, 1998; Perry *et al*, 1998). Like *MID1*, *MID2* is ubiquitously expressed in the mouse embryo. However, *MID2* is also expressed in the heart (Buchner *et al*, 1999). The *MID2* protein associates with microtubules. Further studies are required to confirm or exclude *MID2* in a role for the pathogenesis of OS (see Chapter 3).

1.7 Research Objectives

The objective of this project was to investigate the location and regulation of *MID1* expression during embryogenesis and in a cell culture system. The focus was to better understand the role of *MID1* during normal mammalian development. To generate more information on how the *MID1* and *MID2* genes function, a detailed mutation screen of both genes was carried out for the 15 OS patients available in this study. Mutations found in the *MID1* and *MID2* genes may enable a correlation to be made between the patient phenotypes and their genotypes. Mutations of interest were chosen for further experiments, to investigate the intracellular localisation of disrupted forms of midin, as well as wild-type midin. The mutations were chosen based on where the disruption was located in the protein structure of midin. A polyclonal antibody to the human *MID1* gene was generated in rabbits to examine the midin motifs and their possible roles in the function of midin by using Western analysis. The second part of this project was to isolate a *MID1* homologue in the zebrafish species. Isolation of *MID1* homologues in other species will enable the study of the role and regulation of *MID1*. The zebrafish was chosen because it is quicker and easier to work with than mice. Advantages of using zebrafish as a model include the availability of a large number of transparent eggs (which are fertilised externally) and a relatively short the generation time 2-3 months. In addition, a number of *MID1* and *MID2* homologues have already been isolated from different species. These homologues were utilised to carry out a phylogenetic study to investigate how well conserved these homologues are between species and between the *MID1* and *MID2* genes. Establishing the molecular mechanisms involved in OS should lead to an improved understanding of the embryological processes responsible for heart and craniofacial development, as well as more accurate methods to diagnose and prevent Opitz syndrome.

Chapter Two:

Materials and Methods

Materials and Methods

2.1 Restriction Enzyme Digests

Restriction enzyme (RE) digests were used for analytical purposes as well as in the preparation of both vectors and inserts for ligations. Restriction enzymes were either from New England BioLabs or Boehringer Mannheim. The recommended buffers were used for single enzyme digests, while 10x Super Duper buffer (see Section 2.17) was used for multiple enzyme digests. Generally, 1 unit of enzyme was used to digest 1 μ g of DNA at 37°C for at least 1h.

2.1.1 Agarose Gel Electrophoresis

The agarose gel was made up to 1%, 1.5% or 2% agarose in 1x TAE buffer depending on the size of the DNA products to be electrophoresed. Electrophoresis was generally at 65 volts for 40 min. All DNA fragments (eg. RE digests, PCR) were analysed by gel electrophoresis using the 1kb ladder as a size marker (100ng/ μ l, Gibco). Agarose gels were stained with ethidium bromide at the concentration of 0.5 μ l/ml in order to visualise the DNA. Pictures of agarose gels were taken under UV light (long wavelength).

2.1.2 Preparation of Cloning Vectors

The pBS (KS+) (Stratagene), pGEX-4T2 (Pharmacia), pEGFP-C2 (Clontech) and pDBleu (Gibco BRL products) cloning vectors were all prepared in a similar fashion as follows. The vectors were digested with the appropriate restriction enzymes, 1 unit of enzyme for 1 μ g DNA to be digested for at 37° for 1h, in a minimum volume. Once digested, one tenth of the volume was electrophoresed on a 1% agarose gel as a check to ensure complete digestion. The remaining digested vector was then electrophoresed, cut and excised from the gel using the QIAquick gel extraction kit (QIAGEN Cat #28706).

2.1.3 Removal of 5' Phosphates From Vectors

When vectors had been digested with a single restriction enzyme, they were also subjected to calf intestine alkaline phosphatase (CIAP) (Roche) treatment to remove the 5' phosphates and prevent the vector re-circularising during ligation. The digest was heated to 70°C for 15 min to inactivate the restriction enzyme. Then 1µl CIAP was added to the reaction and incubated at 37°C for 30 min. The vector was purified using the gel extraction kit (QIAGEN) to remove the CIAP enzyme. Once purified, 1/10th of the vector volume was electrophoresed on an agarose gel to determine its concentration.

2.2 Ligation Conditions

The ligations were generally set up with a vector:insert ratio of 1:3 (eg. 100ng/µl:300ng/µl). For cloning with the pDBleu vector, the ratio was 4:1 (eg. 400ng/µl:100ng/µl) because of its large size (~10kb). A final concentration of 1x ligation buffer, 1 unit T₄ ligase (1µl), template to the total of 1µg and dH₂O was used to make up the lowest possible volume (eg. 15µl). Ligation was carried out at room temperature for 1hr. The ligation reaction was extracted with phenol/chloroform followed by an ethanol precipitation with 1µg glycogen (DNA carrier) and a 70% ethanol wash. The pellet was resuspended in 10µl of water and 4µl of this was transformed by electroporation into competent DH5α cells.

2.3 Transformation

Generally, 4µl of the cleaned ligation was added to 40µl of competent DH5α cells and electroporated at 2.5 KV. Once electoporated, the cells were transferred to 1ml SOC media and incubated at 37°C for 45 min. The cells were then centrifuged at 2500xg for 5 min, the

pellet was resuspended in 100µl of SOC and plated on to L-agar containing the appropriate antibiotic. Plates were incubated overnight at 37°C. The final concentration was 100µg/ml, for ampicillin (Amp) and 20µg/ml for kanamycin (Kan).

2.3.1 Mini and Midi Preparation of Plasmid DNA

Following the overnight growth on L-agar plates, colonies were selected, and further grown overnight at 37°C with agitation in an appropriate volume of L-broth (eg. 2ml or 25-50ml) with an antibiotic for positive selection purposes. The DNA was extracted using QIAprep Spin Miniprep kit (QIAGEN Cat # 27106) as outlined by the manufacturer. Larger plasmid preparations were completed using QIAGEN plasmid midi kit (QIAGEN Cat # 12145).

2.4 End Filling/Blunting

The T₄ DNA polymerase reaction fills in the bases for 5' overhangs and removes bases from the 3' overhangs, thus altering the reading frame as required. For example, the *Sall* RE site was cut in the pBS(KS⁺). To 1µg of purified digested vector DNA, the following were combined, dNTPs to a final concentration of 0.2mM, 1x T₄ buffer, 5 units of T₄ DNA polymerase enzyme and the volume was made up to 20µl. This mixture was incubated at 37°C for 1h, 10 units of T₄ DNA polymerase was added and incubated for a further 1hr at 37°C. The enzyme was inactivated by heating the reaction to 68°C for 10 min. Reactions were purified by a standard phenol/chloroform extraction and then ethanol precipitation. The pellet was resuspended in 10µl sterile water. This vector fragment was then blunt end ligated together and transformed into DH5α cells to allow large quantities of this vector to be produced. Once vector DNA had been collected, RE digestions and sequencing reactions were completed to ensure the RE site had been removed.

2.5 DNA Precipitation

Throughout this study, two methods were used to precipitate DNA, the ethanol or the isopropanol protocols. DNA was precipitated after a phenol/chloroform extraction to concentrate the DNA.

2.5.1 Ethanol Precipitation

DNA was precipitated by the addition of 1/10th of the volume of 3M Na-acetate (pH5.2) and between 2-3x volumes of cold absolute ethanol. The solution was mixed and left at -80°C for ~15min. To recover the DNA, the solution was microcentrifuged at 14000xg for 10min. The pellet was washed with cold 70% ethanol (~250µl), microcentrifuged for 5min at 14000xg, then dried and resuspended in an appropriate volume of sterile ddH₂O.

2.5.2 Isopropanol Precipitation

Instead of using ethanol, the DNA was also precipitated using isopropanol. To precipitate the DNA, 1/10th of the volume of 3M Na-acetate (pH5.2) and equal volumes of isopropanol was added and mixed thoroughly. The DNA solution was left at -80°C for ~15min and then microcentrifuged at 14000xg for 10min to recover the DNA. The pellet was washed with 70% ethanol (250µl), microcentrifuged for 5min at 14000xg, dried and resuspended in an appropriate volume of sterile ddH₂O.

2.6 Computer Programmes

Nucleotide sequences were aligned and compared in the DNasis v2.0 (Hitachi) programme. A number of searches were done using the BLAST programme (<http://www.ncbi.nlm.nih.gov/>). Sequencing Chromatograms were viewed in Chromas v1.4.

2.7 MID Gene Sequences

There have been various *MID1* gene sequences analysed in this study in addition to the *MID1* sequences generated in the Cox Laboratory. A number of these were obtained from the GenBank database. These included the mouse (Fxy, AF026565; Mid1, Y14848), human (*MID1*, Y13667; FXY1, AF035360; AF269101) sequences, *Rattus norvegicus* (#AF186461), and *Mus spretus* (AF186460). Professor Alan Ashworth (Institute for Cancer Research, London) supplied the fugu sequence for *MID1* and the chick *MID1* sequence was provided by Jane Sibbons (Adelaide University, Cox Laboratory). The *MID2* sequences were taken from GenBank. These included the mouse (Mid2, Y18881) and human (*MID2*, Y18880; FXY2, AF196481).

2.7.1 Evolutionary Analysis of *MID* Genes

Sequences were edited to remove missing base symbols (-----), aligned using CLUSTAL-W (Thompson *et al*, 1994) and displayed using GeneDoc (Nicholas *et al*, 1997a and b). Aligned sequences were subjected to phylogenetic analysis using the following programs available in the PHYLIP package (Felsenstein, 1993). The computer programs used were DNAPARS (maximum parsimony), DNAML (maximum likelihood), SEQBOOT (bootstrapping) and CONSENSE (obtaining consensus tree). Trees were displayed using TREEVIEW (Page, 1996). Sequence identity/similarity percentages were obtained using GeneDoc (Nicholas *et al*, 1997 a and b). Relative rates tests (Wu and Li, 1985) were carried out using the program K2WuLi (version 1.0) (Jermin, 1996).

2.8 OS Patients

There were 17 probands (OSP# 1-17) who had been diagnosed by clinicians at the Women's and Children's Hospital, Adelaide, Sydney Children's Hospital, Randwick, and Northern Regional Genetic Services, NZ to have Opitz Syndrome. All patients had given consent to the collection of blood samples and molecular testing. In addition, when there was a confirmed OS family history (OSP #6, 8, 10 & 11) samples were gathered from parents and relatives (when available) whether they were affected with OS or not. The clinical phenotypes of the OS patients used in this study were provided (Table 3.2).

2.9 Mutation Analysis

Genomic DNA was extracted from the OS probands using the method described by John *et al*, 1991. For both the *MID1* and *MID2* genes, primers were designed to the intron/exon boundary sequences (Table 2.1). Professor Alan Ashworth (Institute for Cancer Research, London) provided the boundary sequence around the exons of the *MID1* gene. The primers were used in PCR to amplify each exon and resultant products were subjected to automated sequencing (see Section 2.10). The sequence data were analysed manually for mutations by comparing the nucleotide sequence between the patient and 'normal' sequence in the DNasis programme. For the larger exons (# 1 and 9), internal primers were designed and used to ensure the exon was sequenced in its entirety.

2.10 Polymerase Chain Reaction

The primers used throughout this study and their respective annealing temperatures are outlined in tables 2.1, 2.2, 2.3, 2.4, 2.5. The annealing temperatures for primers were determined by the percentage of the GC content using the following formula; $T_A = 2(A+T) + 4(G+C) - 5^\circ\text{C}$. Primers were ordered from a variety of sources including Pacific Oligos, Geneworks, Operon, and Sigma/Genosys.

The final concentrations of the PCR components were 1x PCR buffer, 2.5mM MgCl₂, 200μM dNTP's mix, 33ng/μl of each primer, 1 unit of Taq DNA polymerase and the volume made up to 50μl with water. Human genomic DNA was used as the template at a final concentration of ~100ng/μl. The basic PCR cycle conditions consisted of an initial step at 94°C for 3 min, then 35 cycles of 94°C for 1 min to denature the DNA, the respective annealing temperature for 1 min to anneal the primers and 72°C for various times to complete the DNA extension. There was a final extension time at 72°C for 10 min. The extension time varied depending on the length of DNA to be amplified. For the purpose of screening the *MID1* and *MID2* genes for mutations, the PCR products were purified by running samples through a G₅₀ sepharose column (equilibrated in 1xTE) and then concentrated by isopropanol precipitation (see Section 2.5.2). The pellets were resuspended in a volume that depended upon the initial concentration of the PCR samples and ranged from 11 to 20μl of water.

Table 2.1: Primers used in PCR to screen for mutations in the *MIDI* gene.

Exon #	Forward Primer 5'→3'	Reverse Primer 5'→3'	Product Size bp	Ann Temp °C
U2	GATTCCGAGCTGGACAGAGC	TGTGGGGTTAGAGGCTGAGC	277	60
U4	GAGCAGAGTGC GTTAGCAA	TGCTAACCAGCAAGCTCTC	289	60
1	TGACTTCCCTGTGCATAATC	CGATGGCCTGTAAAGGGCTTC	460	55
1b	CCTTTGACGCCAACACCATG	GAAAACCTTCACCTGCCATG	350	55
2	CTGTAAAGTATCCCTGAGTTAGTG	AGGATATCAAAACCTTGATCTGGC	416	57
3	CAGAGATTACTAGAAACATCCTAG	AAAGTCCCTGGAGTGTATTAGG	278	60
4	GAAGGATCCAGCCTGATGCC	CCAGCAGCCGAGAAATTCCC	266	60
5	GCAGTGATGTTGAAAGTCATC	AACTGATCTGGTGGCAAAGTG	257	55
6	TAGGATAGTAGTCCTGATGG	GAAGAAAGCCACCAAAGACAG	311	55
7	TTGTGCCAAAGAAGACTGCAC	GAGCAGATAAGACATGACAG	262	55
8	TTGAAAATATGCTATTGGTTATG	GTATTACTCAGCTCTTGCCAGG	361	55
9	ATGTGCAACATGGCTCATTG	AAACAGCTGCCAAATGCGAC	320	56
9b	GATGCTTTGAACTCCATCCAC	ACAGTAAAAGGAGAGGAATC	330	56

Primers for exons 7, 8, and 9 were designed with *EcoRI* linkers (GAATTC) at the 5' end

Table 2.2: Primers used in PCR to screen for mutations in the *MID2* gene.

Exon #	Forward Primer 5'→3'	Reverse Primer 5'→3'	Product Size bp	Ann Temp °C
1	CAGCCTCCGTGGTTCTTCTTAATG	CTTGAGTTTCTCAAATCGATCATT	661	57
2	CAAACCTCTGGAGATGAACCTCAC	CTCAACCTGCTGGCAGATCTG	100	57
3	GTCCAGACCTTTTAGAGCCAG	TGTAAAGTCAACCCTTCCTTGG	226	57
4	TCACTGCAGCATGCCTCCTTC	CCTCTTTTCAACAAGAGCAGGC	288	60
5	TCAAACACTGGGTGAAAGTTGC	CTGTCTGAAAAGACACATGCATC	343	60
6	CCAGAAGAGAAAGTATGCTTTCC	CCTTACATTTCTTATCTCTGGC	246	60
7	CATTCTGTCTCTTACCTGCTG	TACTCCCCTGGCATGGAGAAG	214	57
8	GATGATGCTGGTGGAAAAGGTAG	GAATCTCCTAGGAACCTAATGAC	348	57
9	TATGGTGTCTTACACAACCTCAG	GGAAGAACTGCTGAACTAGAG	521	57

Table 2.3: *MID1* and *MID2* primers used in PCR to generate inserts for the various constructs in further experiments.

Primer name	5'→3'	Ann Temp °C	Location & Purpose
5'MID-fusion	GTGAATTCCTGAAGATGGAAACACTGGAGTC	57	over the ATG
HM-MID#1	GAAAGCTTTGCCTGAAGATGGAAACACTG	55	over the ATG also used in ZF
HM-MID#2	CGATGGCCTGTAAAGGGCTTC	55	exon 1 also used in ZF also internal primer R
HM-MID#4	GAGGATGCAGTTGCCATGGAGAC	55	start of exon 5 also used in ZF
G232-2	GTGAATTCGGGACACTTCTGGTGAG	57	after TGA in UTR used to generate Δexon2 in OSP#11 R
MID-CTD	CTGCTCGAGCCCGCCTAGTTGATGGCCTTCACC	55	exon 7 removes the CTD R
MIDOS#12r	CGTGATGCATTGACCTCAATTTGTCATAGCA	63	used to generate Δexon2 in OSP#11 R
MIDOS#12f	GCAGCTTTGAGTGAGGTCAATGCATCACGT	63	used to generate Δexon2 in OSP#11
PE115STOP	GTGAATTCTAGGCGGAGGTCATGGTC	67	generate OSP#12 R
MID2#1	CTTGAGTTTCTCAAATCGATCATTC	57	used in Southern Analysis to generate MID2 probe R

Regions highlighted in bold show the *EcoRI* restriction sites. HM-MID#1 introduces a *HindIII* (AAGCTT) site just 5' of the ATG codon (bold italics). The R indicates this is a 'reverse' primer. MID-CTD has a *XhoI* linker shown in italics (CTCGAG). ZF refers to zebrafish PCR experiments. In the primer name, the H and M stand for human and mouse respectively, and indicate that these species nucleotide sequence were used to design the conserved primers. The underlined bases indicate where the premature stop codon has been incorporated into the PE115STOP primer.

Table 2.4: Primers used to isolate zebrafish *MID* nucleotide sequence.

Primer name	5'→3'	Ann Temp °C	Location
ZFMID#1	GTCACCTGCGAGGTGTCGTA	55	exon 1
ZFMID#2	GTCTGGAGCTGTTTCGAGGATC	55	exon 1
ZFMID#3	ATGCTTGACCGCCTCCTGCG	55	exon 1 R
ZFMID#4	CTGCTTGCAGTTGGCGATCT	55	exon 4 R
ZFMID#5	ATCGTCCTCGCTCATCATTC	55	exon 4
HF-MID1#2	GTGTAGTGGTTCTGCTTGATGT	55	exon 7 R

The H and F denote that human and fugu nucleotide sequence was used to design the primer. The R indicates a 'reverse' primer.

Table 2.5: The vector primers used in this study.

Primer name	5'→3'	Ann Temp °C	Vector
Sp6	ATTTAGGTGACACTATAG	55	pBS pGEM (Sp6 promoter)
T ₇	AATACGACTCACTATAGGG	55	pBS PGEM
T ₃	ATTAAACCCTCACTAAAGGG	55	pBS
RSP - M13	CACACAGGAAACAGCTATGACCATC	55	pBS
USP - M13	GTAAAACGACGGCCAGT	53	pBS
pGEX5'	GGGCTGGCAAGCCACGTTTGGTG	60	pGEX
pGEX3'	CCGGGAGCTGCATGTGTCAGAGG	60	pGEX
BSLK - For	TACCGGGCCCCCTCGAGGTCGA	58	pBS
BSLK - Rev	CGGTGGCGCCGCTCTAGAACTAG	58	pBS

2.11 Automated Sequencing

Automated sequencing was carried out at the Institute of Medical and Veterinary Science (IMVS) Sequencing Facility, Adelaide, on Applied Biosystems 373 and 377 sequencing machines. The automated sequencing reaction was set up using the PCR product as template total of ~10-30ng/ μ l, 33ng primer, 8 μ l of either Terminator ready reaction mix or BIG Dye terminator mix (ABI) and the volume made up to 20 μ l with water. The reaction was covered with mineral oil and then subjected to thermal cycling as follows; rapid thermal ramp to 96°C, 96°C for 30 sec, rapid thermal ramp to 50°C, 50°C for 15 sec, and 60°C for 4 min. These cycles were repeated 25 times. The samples were ethanol precipitated to clean up the reaction (Section 2.5.1). When sequencing directly from a plasmid, the only change was the concentration of template: a total of 0.2 μ g/ μ l of double stranded DNA.

2.12 SSCP Analysis

Acrylamide gels were set up using a 29:1 acrylamide:bisacrylamide (Sequagel) solution to generate a 6% gel, containing 5% glycerol, 1xTBE in a 60ml volume. The glass plates used were 13x16 inches and 2mm thick. 480 μ l of 10% ammonium persulfate and 24 μ l TEMED were added to the acrylamide solution to polymerise the gel.

2.12.1 Radiolabelling of PCR Products

Initially, the target area was amplified in PCR from 50 to 100 control patients (Red Cross Blood Band, Adelaide). To radioactively label the PCR product, 2 μ l was taken from the 50 μ l reaction and used to set up the secondary PCR in a 20 μ l volume, with 2.5mM dATP, 5mM dCTP, dTTP, and dGTP and ~0.2 μ l α -dATP³²P per reaction. The remaining solutions were the same concentrations as in section 2.10. When radioactively labelling the PCR product, the same cycling conditions were used. However, only 10 cycles of amplification

was performed to enable incorporation of the label. Once the PCR cycles had finished, 9µl of stop buffer (see Section 2.17) was added to each reaction.

2.12.2 Loading of Samples

Radioactively labelled PCR products were denatured by heating to 95°C for 3-5 min and 1.5µl loaded onto the acrylamide gel. The SSCP gel was run at between 4-6 watts overnight at 4°C to ensure that the PCR products were separated as much as possible.

2.13 Generation of GFP-MID1 Mutations

Some of the patient mutations that were identified in *MID1* were introduced into the *MID1* cDNA and then cloned into the pEGFP-C2 vector (see Table 3.1). Mutations cloned into pEGFP-C2 were the R368X and L626P (generated by site-directed mutagenesis by Jane Sibbons, Adelaide University, Cox Laboratory), 343G>T and Δexon 2. The OSP#12 mutation consisted of a premature stop codon that was designed into a primer, PE115STOP, with an *EcoRI* linker (Table 2.3). This primer was paired with the 5'MID1 fusion primer that extended over the start codon and also had an *EcoRI* linker region (Table 2.3). Once the OSP#12 mutation had been amplified (as outlined in Section 2.10) the *EcoRI* linkers were utilised to enable in frame cloning into the pEGFP-C2 vector. Similarly, the OSP#11 mutation was generated using internal primers (MIDOS#12f & r, Table 2.3), both of which spanned from exon 1 to 3 resulting in the deletion of exon 2 when *MID1* was amplified in PCR. This was achieved using two PCR amplification steps, first by pairing the MIDOS#12f & r with 5' MID fusion and G232-2 respectively. The second step used this PCR product as the template and the 5' MID fusion and G232-2 primers were paired together. Restriction digests were used to determine the correct orientation and that pEGFP-C2 was fused to the 3'

end of *MIDI*. Automated sequencing was additionally carried out to ensure the inserts were in frame.

2.13.1 Transfection of GFP-MID1 Constructs

Once the constructs had been generated, 2 μ g of each DNA was transfected into CosI cells using FuGene transfection reagent (Koche) following the protocol outlined by the manufacturer. The CosI cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal calf serum on a coverslip. After ~24h growth post transfection, the cells were fixed with 3.5% paraformaldehyde in 1xPEM buffer pH7.0 (100mM PIPES, 5mM EDTA, 2mM MgCl₂), washed 3x in 1xPEM, then permeabilised in 0.2%NP-40 in 1xPEM for 2 mins at room temperature. Microtubules were stained post-fixation using anti- α tubulin antibody (mouse)(Sigma, 1/2000 dilution) plus a rhodamine-labelled secondary antibody (Boehringer Mannheim, 1/50 dilution). The nuclei of the CosI cells were stained using the DNA-specific stain, 4', 6-diamidino-2' phenylindole dihydrochloride (DAPI, sigma). GFP (green fluorescent protein), DAPI and rhodamine fluorescence were visualised under appropriate wavelength light on an Olympus AX70 microscope. The images taken were analysed using a Photometrics CE200A camera electronics unit and processed using Photoshop v5.0 software (Adobe).

2.14 Isolation of a Zebrafish *MID* Homologue

To determine if a *MIDI* homologue existed in zebrafish, the primers HM-MID#1 & 2 (Table 2.3) which had been designed from human and mouse nucleotide sequence were used in PCR with zebrafish genomic DNA as template. When these two primers were paired together in PCR, the exon 1 region of *MIDI* was amplified. Hot start PCR was used when attempting to amplify zebrafish *MID* sequences. The conditions per reaction were as follows;

2 μ l of 10xPCR buffer, 2 μ l of 25mM MgCl₂, 1 μ l 10mM dNTPs, 1 μ l 100ng/ μ l each primer and the volume made up to 20 μ l with water. Each PCR reaction was heated to 94°C in the thermal cycler for 30 seconds and then a Taq polymerase mixture (0.5 μ l of 10xPCR buffer and 3.5 μ l of water for 1 μ l of Taq in a total of 5 μ l per reaction) was added. The PCR cycle continued as outlined in Section 2.10.

The carboxy region of the MID homologues had a lower level of conservation especially when compared to the N-terminal. This made it difficult to design PCR primers to amplify the remaining *zMID* sequence. The 11 amino acids just before the #480 are identical between all of the homologues. Consequently, this region was used to design the HF-MID1#2 primer.

2.14.1 Isolation of RNA from Zebrafish Embryos

RNA was prepared from 24h, 48h, 3 day, 5 day embryos and adult zebrafish using the QIAGEN RNeasy Midi kit (Cat#75142). The isolation of total RNA from animal tissue provided with the kit was followed exactly. RNA isolated from these stages was electrophoresed on an agarose gel under Rnase-free conditions to ensure RNA was present. An aliquot was also used in the spectrophotometer to determine the concentration. The RNA from each stage was used as the template for RT-PCR to generate cDNA.

2.14.2 RT-PCR

The two-step RT-PCR kit from Gibco BRL (Cat#18064-04) was used to synthesize cDNA. From each stage, 1 μ g RNA was added to 1 μ l hexamer primers (40ng/ μ l) and the volume made up to 12 μ l with water. This mixture was incubated at 70°C for 10 min. A master mix containing, 4 μ l of 5x first strand buffer, 2 μ l 0.1 M DTT and 1 μ l 10mM dNTP was made up per reaction. From this master mix, 7 μ l was added to the RNA mix and incubated at

37°C for 2 min. Next, 1µl of Superscript™ enzyme (1 unit/µl) was added and the mixture incubated at 37°C for 1hr, then at 70°C for 15 min to inactivate the enzyme. This produced 20µl of cDNA (~50ng/µl) and 1µl was used as the template in PCR. The cDNA was also made using oligo (dT) primers.

2.15 Wholemount *in situ* Hybridisation of Zebrafish Embryos

The zebrafish *in situ* hybridisation experiments were carried out following the protocol from the Lardelli lab, as outlined on their web site <http://www.science.adelaide.edu.au/genetics/lardelli/mawpt.html>. The buffer and solutions were also used as outlined from this site (Appendix One).

2.16 Genomic Southern

2.16.1 Genomic DNA Extraction from Adult Zebrafish

The protocol for the extraction of genomic DNA was adapted from Westerfield (1994). An adult zebrafish was sacrificed in ice water, weighed and gutted. A hand held homogeniser was used to grind up the zebrafish and 10ml extraction buffer (see Section 2.17) was added per 1g of fish. This mixture was incubated at 50°C for a minimum of 3h and maximum of overnight, occasionally mixing the solution. The solution was cooled to room temperature and extracted twice with equilibrated phenol, mixed gently until an emulsion formed, and the layers separated by centrifugation at 3000-5000xg for 10min. The aqueous layer was carefully removed and extracted with phenol:chloroform:isoamyl alcohol (25:24:1). The supernatant was transferred to a fresh tube, NaCl to a final concentration of 200mM was added with 2x volumes of ethanol. Solutions were gently mixed. A pasteur pipette which had a sealed hook was used to pick up the DNA and transferred into a tube of

70% ethanol. The DNA was allowed to stand for ~5mins, occasionally mixing the solution. DNA was removed and placed into a clean tube and air dried. The DNA was resuspended in 5-10µl per gram of starting material in 10mM Tris pH8, 5mM EDTA and 100µg/ml RNase. To facilitate genomic DNA resuspension, the DNA solution was placed at 37°C and mixed occasionally. Alternatively, to assist in genomic DNA resuspension the DNA was incubated at 4°C overnight.

To further purify the genomic DNA, a phenol:chloroform:isoamyl alcohol extraction was completed and centrifugation was carried out at 3000-5000xg for 10 min. The supernatant was transferred to a fresh tube, 0.1x volume of 7.5M ammonium chloride was added followed by 2x volume of ethanol. The tube was inverted slowly until solution was mixed to precipitate the genomic DNA. As previously described a glass pasteur pipette was used to transfer DNA to a 70% ethanol wash, then to a new tube and air dried. Genomic DNA was resuspended in 1-2ml TE buffer per gram of starting material. The tube was incubated at 50°C to facilitate the resuspension of genomic DNA. Spectrophotometry (A_{260}) was used to determine the concentration of the genomic DNA.

2.16.2 Digests

The zebrafish genomic DNA (10µg) was separately digested using ~50 units of *PstI*, *BamHI*, *EcoRI* (New England BioLab and Boehringer Mannheim) and double combinations of the three enzymes in 50µl reaction volumes. Digestions were completed overnight at 37°C. Combination digests were done in 10x Super Duper buffer (see Section 2.17), whilst single enzyme digestions were done in the buffer recommended by the manufacturer.

2.16.3 Southern Transfer

After overnight digestion, DNA samples were loaded onto a 20cm x 20.5cm 0.8% agarose gel that had been made up in 1x TAE buffer. Ethidium bromide had been incorporated into the agarose at a concentration of 0.5µl/ml. The DNA was electrophoresed overnight at 15 volts (~20h) and then the voltage was increased to 25 volts for 4 hours to ensure the bands had separated. The gel was analysed and photographed under UV light (long wavelength).

The gel was first treated in 0.25 M HCl with agitation for 10-20 min, then in 0.4 M NaOH with agitation for 15-20 min to denature the genomic DNA in the agarose matrix. The gel was rinsed briefly in 2xSSC and placed upside down on plastic wrap. A dry Hybond (Amersham) filter was placed on top of the gel, followed by two sheets of Whatman paper, which have been soaked in 2xSSC, and two sheets of dry Whatman paper. The filter and Whatman paper were exactly the same size as the gel. A stack of paper towels (also the same size of the gel) was placed on top of the gel followed by a large weight. A few mls of 2xSSC was placed at the bottom of the gel (without touching the paper) to assist the transfer of genomic DNA to the filter. The genomic DNA was allowed to transfer to the nylon filter overnight. The next day, the weights and Whatmann paper were removed, the gel together with the filter paper were turned over. The wells of the gel were marked onto the filter. The gel was removed and the DNA was fixed onto the filter by a UV cross-linker (starting at 1200, UVC-508, ultraviolet crosslinker Ultra-Lum).

2.16.4 Probes

The radiolabelled probes for Southern analysis were generated using the Gigaprime labelling kit (Geneworks Cat#GPK-1). 25ng of DNA was used in the labelling reaction. The DNA probes had a specific activity greater than 2×10^9 cpm/µg. This corresponded to at least 50% incorporation of the radioactive label (^{32}P -dATP). Once the probe had been made, it was

spun through a G₅₀ Sepharose column (described in Section 2.10) before being added to the pre-hybridisation solution and the filter.

2.16.5 Hybridisation

Zebrafish genomic DNA filters were pre-hybridised in sealed plastic bags with agitation in the hybridisation solution (see Section 2.17) for at least 1hr at a temperature according to the probe that was to be used. The probes from human homologues MID1 & MID2 were hybridised to the genomic filter at 42°C. The zebrafish MID probe was hybridised at 65°C. These probes were added to the filters with ~10-20ml of hybridisation solution. The probes were hybridised to the filter overnight with agitation.

2.16.6 Washes

Washing of filters was carried out using fresh hybridisation solution (see Section 2.17) at various temperatures starting with a low stringency and increasing the stringency at each wash. The first wash was carried out at room temperature for 20mins with agitation and the filters were then exposed to X-ray film. After exposure, the filters were washed again using a fresh solution of hybridisation solution at 42°C, for ~20 mins with agitation. Again the filters were exposed to X-ray film, before the last wash using a fresh solution of hybridisation solution at 65°C, for ~20 mins with agitation. After the last wash, the filters were stripped of radioactivity to ensure that there were no false positives. The filters were stripped by pouring boiling 0.1% SDS and leaving for up to 15 min. This process was repeated until no radioactive signal could be detected. Between each wash (including after the filters had been stripped) the filters were exposed to X-ray film at -80°C for an appropriate time depending on the radioactivity level.

2.17 Buffers

Solutions used for Western Analysis

6xSDS loading buffer

0.55M Tris-HCl (pH6.8)

36% (v/v) Glycerol

10.28% (w/v) SDS

5% β -mercaptoethanol

0.012% (w/v) bromophenol blue

The β -mercaptoethanol was added prior to use.

Coomassie stain

0.1% coomassie blue

10% acetic acid

40% ethanol

De-staining solution

40% ethanol

10% acetic acid

Non-denaturing lysis buffer

1% (w/v) Triton-x-100

50mM Tris-Cl pH7.4

300mM NaCl, 5mM EDTA

water up to 10ml

Before use, to 1ml of buffer add, 10mM DTT and 1mM PMSF

General use solutions

Hybridisation Solution

Used for the hybridisation and washing of Southern Analysis filters.

6xSSC

0.1%SDS

6x Stop Buffer

10mM Tris-HCl pH7.5

1mM EDTA pH 8.0

0.2% (w/v) SDS

6x Loading Dye

Taken from Sambrook *et al*, 1989 for loading samples onto agarose gels.

Final concentration

Bromophenol blue	0.25%
Xylene cyanol FF	0.25%
Ficoll (type 400)	15%

Make up to 10ml using sterile water

1x TE Buffer

10mM Tris-HCl pH8.0

1mM EDTA pH 8.0

10x PCR Buffer

100mM Tris (pH 8.5-9.0)

500mM KCl

1%TritonX-100

10x TAE Buffer

Ingredient

Final Concentration

Tris

0.04M

Boric Acid 0.5g

Di-sodium EDTA 5g

Make up to one litre using ddH₂O and autoclave.

LB Broth

<u>Ingredient</u>	<u>Final Concentration</u>
Bactotryptone	1%
Yeast Extract	0.5%
NaCl	1.5%

Adjust pH to 7.5 with NaOH. Make up to one litre with ddH₂O and autoclave. For LB agar add 15g of agar to give a final concentration of 15% before autoclaving. Ampicillin (Amp) stock was made up in ddH₂O and the final concentration was 100µg/ml. For kanamycin (Kan) stock was made up in ddH₂O to 200nl/ml and when used in the culture the final concentration was 20µg/ml

10x TBE Buffer

<u>Ingredient</u>	<u>Final Concentration</u>
Tris	0.45M
Boric acid	0.45M
0.5M EDTA pH5	10mM

Make up to 1 litre with dH₂O and then filter. The pH should be about 8.3 with no adjustment required.

20x SSC

<u>Ingredient</u>	<u>Final Concentration</u>
-------------------	----------------------------

NaCl 3.0M

NaCitrate 3.3M

Adjust the pH to 7.0 with NaOH and then autoclave.

SOC Medium

<u>Ingredient</u>	<u>Final Concentration</u>
Bactotryptone	2%
Bactoyeast extract	0.5%
NaCl	10mM
Potassium chloride	2.5mM
MgSO ₄ /MgCl ₂	20mM
ddH ₂ O	up to 490ml

Adjust pH to 7.0 and autoclave.

Immediately before use add:

Glucose 20mM

10x 'Super Duper' buffer

330mM TrisHAc pH 7.8

625mM KAc

100mM MgAc

40mM spermidine

5mM DTT

EDTA (0.5M, pH8.0)

<u>Ingredient</u>	<u>Final Concentration</u>
-------------------	----------------------------

EDTAdi-sodium salt	0.5M
NaOH	20g of pellets

Adjust the pH to 8.0 with NaOH and then autoclave.

PBS

NaCl	8g
KCl	0.2g
Na ₂ HPO ₄	1.44g
KH ₂ PO ₄	0.24g

800ml of H₂O and adjust to pH 7.4 with HCl.

Add H₂O to one litre and autoclave.

Zebrafish genomic DNA isolation

10mM Tris pH8.0
100mM EDTA pH8.0
0.5% SDS
200µg/ml Proteinase K

Solutions used for Southern Analysis

Denaturing Solution

<u>Ingredient</u>	<u>Final Concentration</u>
NaCl	1.5M
NaOH	0.5M

Make up to one litre with ddH₂O.

Neutralising Solution

<u>Ingredient</u>	<u>Final Concentration</u>
NaCl	1.5M
Tris-HCl pH 7.2	0.5M
EDTA	0.001M

Make up to one litre with ddH₂O.

Chapter Three:

Mutational Analysis in OI Patients

Mutational Analysis in OS Patients

The presence of mutations in a particular gene can confirm the presence of a genetic disorder, or whether or not the patient is a carrier of a genetic disease. In addition, the positioning and the types of mutations in a gene can also affect the translated protein and result in the phenotype of the disorder. Therefore, determining the presence of mutations in a gene is useful as a diagnostic tool as well as providing information on how the gene is inherited (e.g. dominant *vs* recessive) and how the mutation could affect its function (e.g. developmental, structural).

The *MID1* gene was identified prior to the commencement of this project concurrently in the Cox laboratory as well as by Quaderi *et al*, 1997. It was found that *MID1* consists of nine coding exons encoding a protein of 667 amino acids. The following motifs have been identified in the MID1 protein; RING finger, B-boxes, coiled-coil, FNIII and CTD. Exon 5 was in the area initially considered to be the linker region between the coiled-coil and C-terminal domains. More recently, protein alignments have shown that this region has similarity to the FNIII repeat (Perry *et al*, 1999). Therefore, this region has been re-classified as a FNIII repeat. The majority of the mutations found in the *MID1* gene from OS patients, reside in the latter half of the protein and truncate the protein (Quaderi *et al*, 1997; Gaudenz *et al*, 1998; Cainarca *et al*, 1999; Schweiger *et al*, 1999). An additional gene named *MID2* (Perry *et al*, 1999; Buchner *et al*, 1999) which has a high similarity to *MID1* at both the nucleotide and protein level has also been identified. At this stage the involvement of *MID2* in OS is uncertain.

In this study, a mutational screen of the *MID1* and *MID2* genes was carried out in fifteen Australasian and one British patient diagnosed as having Opitz Syndrome (OS) to determine novel changes in both genes and to investigate whether or not *MID2* is involved in

the OS. Automated sequencing was used to analyse the nucleotide sequence of the nine exons in *MID1* and *MID2* for each OS patient. While undertaking the initial stages of the mutational screen of *MID1*, Quaderi et al, (1997) published three novel mutations in the *MID1* gene. Subsequently, six additional *MID1* mutations were reported, clustered in the Carboxy-terminal domain (Gaudenz et al, 1998). The mutations described hereafter were conducted largely prior to the publication of Gaudenz et al, (1998) and the results reported as part of a larger body of work (Cox et al, 2000).

3.1 Clinical and Cytogenetic Features of OS Probands

The clinical phenotypic and cytogenetic findings in each OS proband screened for *MID1* and *MID2* mutations were provided, with the exception of OSP#16 due to a lack of supplied clinical information (Table 3.1). All of the OS probands studied here had hypertelorism. Of the 15 OS probands for which clinical reports were provided, 10 patients displayed anteverted nares, a feature suggested as possibly distinguishing the X-linked and autosomal forms of OS (Robin et al, 1996). However, patients studied here that had been confirmed to have the X-linked form of OS did not necessarily have anteverted nares, indicating that this theory was not entirely true. Eleven of the probands had developmental problems such as motor and language delays, ten had laryngotracheal abnormalities, and 12 displayed dysphagia/aspiration/GER defects. Eight patients had lip/palate clefting, and 4 of these also displayed a highly arched palate. Only 4 of the OS patients (OSP# 5, 6, 9 and 10) have congenital heart disease such as atrial and ventricular septal defects, coarctation of the aorta and patent ductus arteriosus. There were 10 probands with hypospadias. In addition, 3 of these had an imperforate/ectopic anus and a different three patients also showed urinary tract abnormalities. None of these OS patients displayed cryptorchidism.

Table 3.1: Summary of the Opitz patient phenotypes and molecular findings involved in this study. The OSP#16 was not included in the table. Taken from Cox *et al*, 2000.

OSP#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Feature															
hypertelorism/telecanthus	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
prominent forehead	Y	-	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
widow's peak	N	Y	Y	N	Y	Y	N	Y	Y	Y	Y	N	-	N	N
broad nasal bridge	Y	-	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
anteverted nares	Y	-	Y	Y	Y	n/a	Y	Y	Y	Y	Y	-	-	-	Y
cleft lip/palate	N	Y	Y	Y	Y	Y	N	N	N	N	N	Y	Y	Y	N
high arched palate	-	n/a	Y	Y	N	n/a	N	Y	-	-	-	n/a	-	n/a	Y
grooved nasal tip	N	-	N	N	Y	N	N	N	N	N	N	N	N	Y	N
flat philtrum	Y	n/a	n/a	N	n/a	n/a	Y	Y	Y	Y	Y	n/a	-	n/a	N
abnorm/post. rotated auricles	Y	-	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	-	Y	Y
laryngotracheal abnormality	Y	N	N	Y	N	N	Y	Y	Y	Y	Y	Y	N	Y	Y
dysphagia/aspiration/GER	N	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
congenital heart disease	N	N	N	N	Y	Y	N	N	Y	Y	N	N	N	-	N
					(PDA; ASD)	(VSD; ASD)			(VSD)	(VSD; PS)					
imperforate/anterior anus	N	N	N	Y	N	N	N	N	Y	N	N	Y	N	-	N
hypospadias	Y	Y	Y	n/a	Y	Y	Y	Y	Y	N	Y	N	Y	n/a	N
urinary tract abnormalities	-	-	-	N	N	Y	Y	N	Y	N	N	N	N	N	N
cryptorchidism	N	N	N	n/a	N	N	N	N	N	N	N	N	N	N	N
developmental delay	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
corpus callosum agenesis	N	N	N	N	N	N	-	-	N	N	N	N	N	N	N
other features	COL	LDP		BPA		STR	DPH	MGN	IVA	HYC	AUT			LDS	INH
	DOD									LCM				CLN	UMH
	CAT									MSK				SYN	GRT
														MAL	MAC
															LDS
sex	M	M	M	F	M	M	M	M	M	M	M	M	M	F	M
family history	N	N	N	N	N	Y	N	Y	N	Y	Y	N	N	N	N
22q11 region by FISH	not deleted	-	not deleted	not deleted	not deleted	not deleted	not deleted	-	not deleted	-	-	not deleted	not deleted	not deleted	not deleted

Y, yes; N, no; -, not known; N/A, not applicable (eg. features of philtrum after surgery for cleft lip); CAT, cortical atrophy; COL, coloboma of right macula; DOD, dysplastic optic discs; MGN, micrognathia; IVA, inferior vernal agenesis; LCM, large cisterna magna; HYC, hypoplastic cerebellum; BPA, Bell's palsy; MSK, mottled skin; LDP, multiple linear depigmented patches; vWill, von Willebrands; ASD, atrial septal defect; VSD, ventricular septal defect; PDA, patent ductus arteriosus; CoA, coarctation of the aorta; PS, pulmonary stenosis; +, other cardiac anomalies; LDS, limb / digit shortening; CLN, clinodactyly; SYN, syndactyly; INH, inguinal hernia; UMH, umbilical hernia; GRT, grooved tongue; DPH, diaphragmatic hernia; AUT, autistic features; MAC, macrocephaly; STR, strabismus; MAL, malrotation.

Additional features included two probands with limb and/or digit shortening (OSP#14 and 15), a grooved tongue (OSP#15) and a diaphragmatic hernia (OSP#7). There were also phenotypes which had previously been unreported with Opitz syndrome, such as mottled skin, inferior vermal agenesis with large cisterna magna (OSP#9) and multiple linear depigmented patches (OSP#2).

Only two of the 15 probands were female (OSP# 4 and 14). Family histories were available for OSP# 6, 8, 10 and 11 that enabled a more extensive investigation to be carried out. In these instances pedigree information was used to determine X-linked inheritance and if consistent with X-linkage, family members of probands were screened for mutations in *MIDI* to determine if they had the same mutation. For example, the pedigree of the OSP#8 family was compatible with X-linked inheritance of OS. Here the mutation can be traced over 4 generations (see Figure 3.8).

The 22q11 region was analysed for anomalies by fluorescence *in situ* hybridisation (FISH) in the Cytogenetic Units of the referring Institution using routine 22q11.2 standard probes (e.g. probe D22575(N25)). Eleven probands were tested and none were found to have detectable 22q11 deletions. No other chromosomal abnormalities were found. However, these findings do not exclude the possibility that there was a deletion in the 22q11 region outside of the standard probe used for these FISH experiments or that there was a smaller molecular change within the probed region.

3.2 *MIDI* Mutation Analysis

To screen *MIDI* for mutations, each of the 9 exons was amplified by PCR (Tables 2.1 and 2.2) and then sequenced directly. The sequence data were analysed for mutations by comparison to the normal *MIDI* sequence (AF035360; AF269101; see Section 2.7) (Figure

3.1). Initially, forward primers from the PCR amplification were used to sequence each exon. Potential DNA changes were verified by sequencing in the reverse direction.

Figure 3.1: The *MIDI* ORF sequence. Normal (wild type) sequence used to align the automated sequencing from the OS patients to find mutations. The domains found in MID1 are highlighted in the protein sequence. The RING finger motif is highlighted in red, the B-boxes in green, the coiled-coil domain in blue, the FNIII domain in pink and the CTD in yellow. The different colours in the nucleotide sequence show the exons found in MID1. Exon 1 is shown in dark blue, exon 2 in purple, exon 3 in yellow, exon 4 in dark green, exon 5 in orange, exon 6 in light blue, exon 7 in light green, exon 8 in red, and exon 9 in black.

MID1 ORF

1 **M E T L E S E L T C P I C L E L F E D R** 20
 1 ATGGAAACACTGGAGTCAGAACTGACCTGCCCTATTTGTCTGGAGCTCTTTGAGGACCCCT 60
 21 **L L L P C A H S L C F N C A H R I L V S** 40
 61 CTTCTACTGCCCTGCGCACACAGCCTCTGCTTCAACTGCGCCACCGCATCCTAGTATCA 120
 41 **H C A T N E S V E S I T A F O C P T C R** 60
 121 CACTGTGCCACCAACGAGTCTGTGGAGTCCATCACCGCCTTCCAGTGCCCCACCTGCCGG 180
 61 **H V I T L S** Q R G L D G L K R N V T L Q 80
 181 CATGTCATACCCTCAGCCAGCGAGGTCTAGACGGGCTCAAGCGCAACGTACCCTACAG 240
 81 N I I D R F Q K A S V S G P N S P S E T 100
 241 AACATCATCGACAGTTCCAGAAAGCATCAGTGAGCGGGCCCAACTCTCCAGCGGAGACC 300
 101 R R E R A F D A N T M T S A **E K V L C Q** 120
 301 CGTCGGGAGCGGGCCTTTGACGCCAACACCATGACCTCCGCCGAGAAGTCTCTGCCAG 360
 121 **F C D Q D P A Q D A V K T C V T C E V S** 140
 361 TTTTGTGACCAGGATCCTGCCAGGACGCTGTGAAGACCTGTGTCACTTGTGAAGTATCC 420
 141 **Y C D E C L K A T H P N K K P F T G H R** 160
 421 TACTGTGACGAGTGCCTGAAAGCCACTCACCCGAATAAGAAGCCCTTTACAGGCCATCGT 480
 161 **L I E P T P D S H I R G L M C L E H E D** 180
 481 CTGATTGAGCCAATTCCGGACTCTCACATCCGGGGGCTGATGTGCTTGGAGCATGAGGAT 540
 181 **E K V N M Y C V T D D Q L I C A L C K L** 200
 541 GAGAAGGTGAATATGTACTGTGTGACCGATGACCAGTTAATCTGTGCCTTGTGTAAACTG 600
 201 **V G R H R D H** Q V A A L S E R **Y D K L K** 220
 601 GTTGGGCGGCACCGGATCATCAGGTGGCAGCTTTGAGTGAGCGCTATGACAAATTGAAG 660
 221 **Q N L E S N L T N L I K R N T E L E T L** 240
 661 CAAAACCTTAGAGAGTAACCTCACCAACCTTATTAAGAGGAACACAGAACTGGAGACCCTT 720
 241 **L A K L I Q T C Q H V E V N A S R Q E A** 260
 721 TTGGCTAAACTCATCAAACCTGTCAACATGTTGAA**YTCIATGICATCACGTCAGAAADCC** 780
 261 **K L T E E C D L L I E I I Q Q R R Q I I** 280
 781 **AKATTGACAGAGAGATGATGATCTTTCATTGRRDLCATTCAGCAAAACCCACAGATTATT** 840
 281 **G T K I K E G K V M R L R K L A Q Q I A** 300
 841 **QIAACCAAGCTCAAGACGATG**GTGATGAGGCTTCGCAAACCTGGCTCAGCAGATTGCA 900
 301 **N C K Q C I E R S A S L I S Q A E H S L** 320
 901 AACTGCAAACAGTGCATTGAGCGGTCAGCATCACTCATCTCCCAAGCGGAACACTCTCTG 960
 321 **K E N D H A R** F L Q T A K N I T E R V S 340
 961 AAGGAGAATGATCATGCGCGTTTCCCTACAGACTGCTAAGAATATCACCGAGAG**AGTCTCC** 1020
 341 M A T A S S Q V L I P E I N L N D T F D 360
 1021 ATGGCAACTGCATCCTCCCAGGTTCTAATTCCTGAAATCAACCTCAATGACACATTTGAC 1080
 361 T F A L D F S R E **K K L L E C L D Y L T** 380
 1081 ACCTTTGCCTTAGATTTTTCCCGAGAGAAGAAACTGCTAGAATGTCTGGATTACCTTACA 1140
 381 **A P N P P T I R E E L C T A S Y D T I T** 400
 1141 GCTCCCAACCCTCCCACAATTAGAGAAGAGCTCTGCACAGCTTCATATGACACCATCACT 1200
 401 **V H W T S D D E F S V V S Y E L Q Y T I** 420
 1201 GTGCATTGGACCTCCGATGATGAGTTTACGCGTGGTCTCCTACGAGCTCCAGTACACCATA 1260
 421 **F T G Q A N V V S L C N S A D S W M I V** 440

1261 TTCACCGGACAAGCCAACGTCGTTAGTCTGTGTAATTCGGCTGATAGCTGGATGATAGTA 1320
 441 P N I K Q N H Y T V H G L O S G T K Y I 460
 1321 CCCAACATCAAGCAGAACCCTACACGGTGCACGGTCTGCAGAGCGGCACCAAGTACATC 1380
 461 F M V K A I N Q A G S R S S E P G K L K 480
 1381 TTCATGGTCAAGGCCATCAACCAGGCGGGCAGCCGCAGCAGTGAGCCTGGGAAGTTGAAG 1440
 481 T N S Q P F K L D P K S A H R K L K V S 500
 1441 ACAAACAGCCAACCATTTAAACTGGATCCCAAATCTGCTCATCGAAAACCTGAAGGTGTCC 1500
 501 H D N L T V E R D E S S S K K S H T P E 520
 1501 CATGATAACTTGACAGTAGAACGTGATGAGTCATCATCCAAGAAGAGTACACACCTGAA 1560
 521 R F T S Q G S Y G V A G N V F I D S G R 540
 1561 CGCTTCACCAGCCAGGGGAGCTATGGAGTAGCTGGAAATGTGTTTATTGATAGTGGCCGG 1620
 541 H Y W E V V I S G S T W Y A I G L A Y K 560
 1621 CATTATTGGGAAGTGGTCATAAGTGAAGCACATGGTATGCCATTGGTCTTGCTTACAAA 1680
 561 S A P K H E W I G K N S A S W A L C R C 580
 1681 TCAGCCCCGAAGCATGAATGGATTGGGAAGAACTCTGCTTCCTGGGCGCTCTGCCGCTGC 1740
 581 N N N W V V R H N S K E I P I E P A P H 600
 1741 AACATAACTGGGTGGTGAGACACAATAGCAAGGAAATCCCCATTGAGCCTGCCCCCAC 1800
 601 L R R V G I L L D Y D N G S I A F Y D A 620
 1801 CTCCGGCGCGTGGGCATCCTGCTGGACTATGATAACGGCTCTATCGCCTTTTATGATGCT 1860
 621 L N S I H L Y T F D V A F A Q P V C P T 640
 1861 TTGAACTCCATCCACCTCTACACCTTCGACGTGCGATTTGCGCAGCCTGTTTGCCCCACC 1920
 641 F T V W N K C L T I I T G L P I P D H L 660
 1921 TTCACCGTGTGGAACAAGTGTCTGACGATTATCACTGGGCTCCCTATCCAGACCATTG 1980
 661 D C T E Q L P * 668
 1981 GACTGCACAGAGCAGCTGCCGTGA 2004

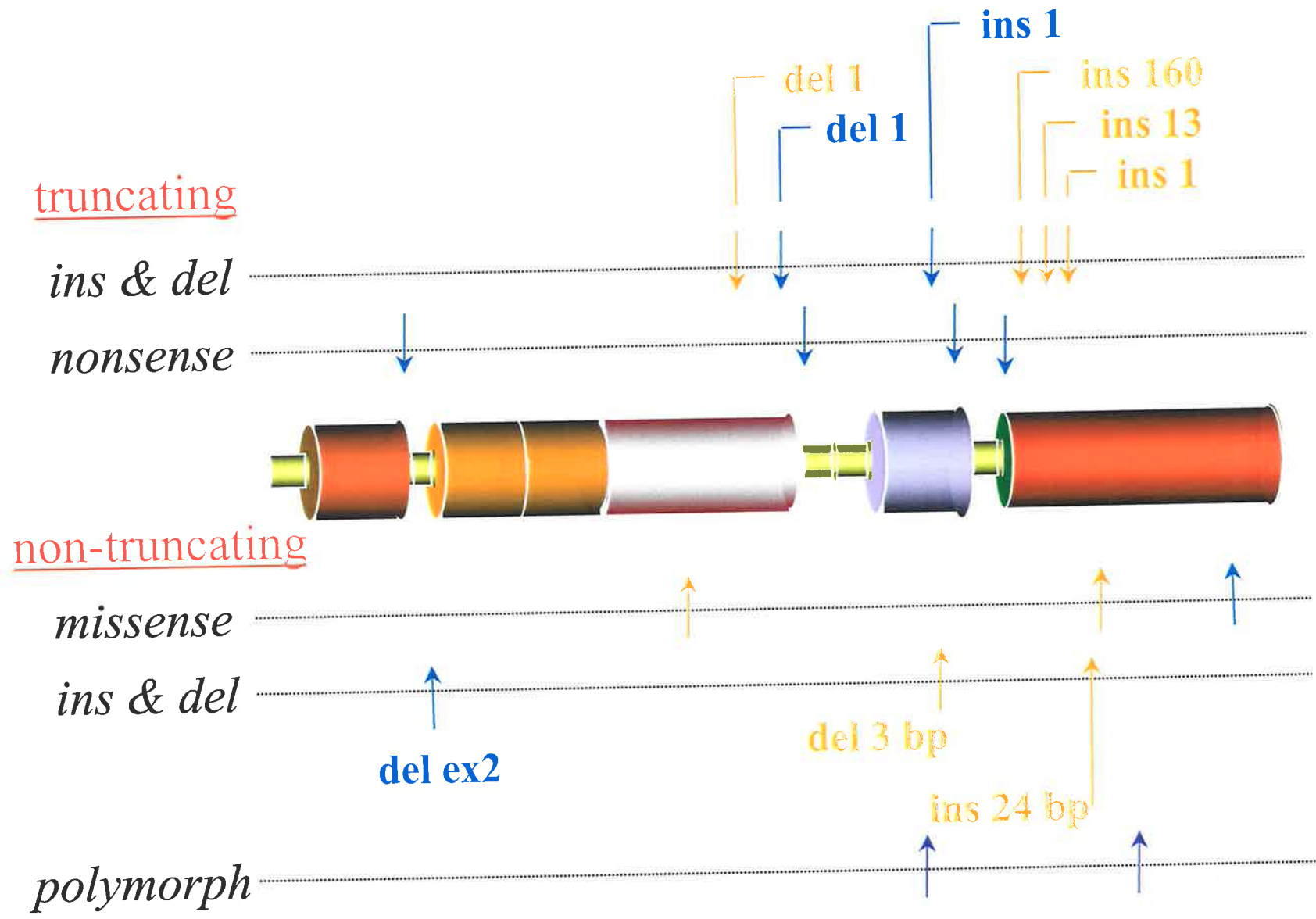
The types of nucleotide sequence changes found in the *MIDI* gene from this study included nonsense mutations, frameshifts, an in-frame deletion and a missense mutation (Table 3.2, Figure 3.2). However, some OS patients did show a mutation in the coding region of the *MIDI* gene. Single strand conformational polymorphism (SSCP) was also carried out in some mutations to confirm the presence of the mutation in family members, to determine whether or not family members were carriers and to confirm other effectors such as polymorphisms or additional mutations.

Table 3.2: Summary of the mutations found in *MIDI* in this study.

OSP#	Nucleotide Change	Amino acid Change	Exons Affected	Domain Affected
12	343G>T	E115X	1-9	before B-box
11	Δexon 2	deletion of 1 st third of coiled-coil	2	coiled-coil
6	1051delC	frameshift	5-9	before FNIII
3	1102C>T	R368X	5-9	before FNIII
16	1330insA*	frameshift	7-9	end of FNIII
10	1402C>T	Q468X	7-9	FNIII
9	1483C>T	R495X	8-9	CTD
5	1877T>C	L626P	9	end of CTD

*A sample from OSP#16 was also made available to another laboratory and subsequently reported by Cainarca *et al* (1999).

Figure 3.2: The *MID1* mutations identified as part of this study. E115X and ? exon2 are the most N-terminal mutations found to date and L626P the most C-terminal mutation found to date.



3.2.1 Truncating Mutations

Nucleotide changes in the *MID1* gene that resulted in the truncation of the MID1 protein (midin) were identified in six of the eight OS patients in this study. Truncating mutations included nonsense mutations (OSP#3, 9, 10 and 12) and frameshift mutations (OSP# 6 and 16).

3.2.1.1 Nonsense Mutations

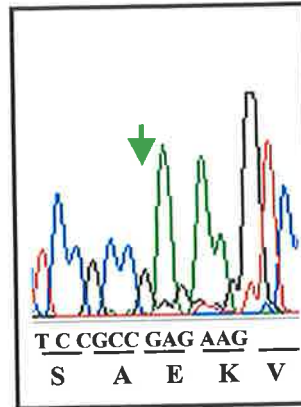
The analysis of automated sequencing of OSP#12 revealed a G>T mutation at position 343 resulting in a change from a glutamate to a premature stop codon at amino acid number 115 (E115X) (Figure 3.3). This mutation in exon 1 occurs after the RING-finger, resulting in truncation of the B-boxes, coiled-coil, fibronectin type III repeat (FNIII) and C-terminal domain (CTD). The proband OSP#12 has no siblings and the family history does not suggest the presence of OS, although in the paternal lineage von Willebrands disease was present which is unrelated to OS. Consistent with her normal appearance the mother of this proband was found not to have the E115X change, indicating that the mutation in this proband arose *de novo*. The E115X is the most N-terminal mutation in the *MID1* gene recorded so far.

In exon 5, there were two mutations found, one of these was a nonsense mutation (Figure 3.4) in OSP#3, the other was a frameshift in OSP#6 (see Section 3.2.1.2). For OSP#3 the mutation, a C>T change was found at nucleotide position 1102. At the amino acid level, the nonsense mutation resulted in a premature stop codon (R368X). Therefore, in OSP#3, the nonsense mutation resulted in the deletion of the FNIII and CTD domains. No OSP#3 family members were available to further test for the presence of R368X.

Two unique mutations were also found in exon 7 of the *MID1* gene upon screening the probands. One was a frameshift in OSP#16 (see Section 3.2.1.2), while in OSP#10, a nonsense mutation, a C>T nucleotide change at position 1402 was identified (Figure 3.5).

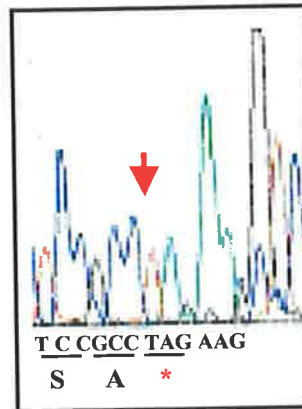
The mutation found in OSP#10 resulted in a premature stop codon in the FNIII region (Q468X). The mother of OSP#10 was also sequenced revealing the same disruption of the *MID1* gene in exon 7. The mother had hypertelorism, while the proband's brother had a normal phenotype suggesting he did not have the mutation. However, this could not be confirmed, as only the mother was available for mutation screening.

Wild

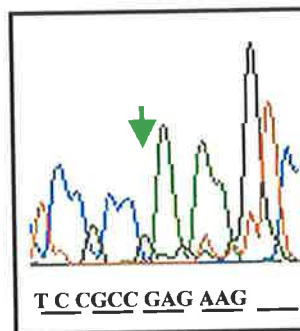


Proban

OSP#1



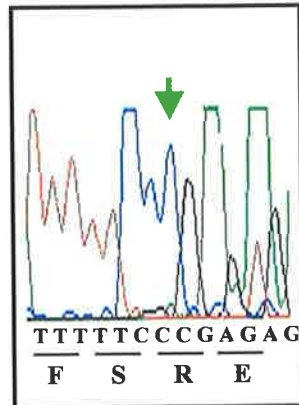
Mothe



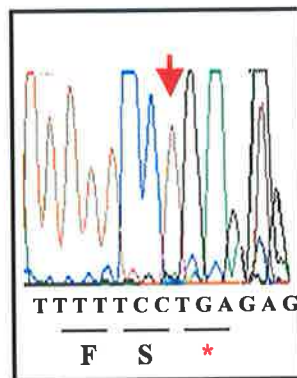
Exon

Figure 3.3: The chromatogram showing the OSP#12 mutation E115X in exon 1 of *MID1*. The G nucleotide was replaced with a T causing a premature stop codon just after the RING finger in OSP#12. The mother does not carry the mutated E115X form of *MID1*.

Wild



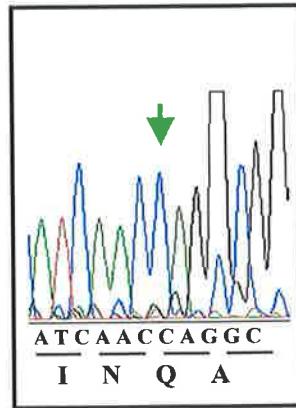
**Proban
OSP#**



Exon

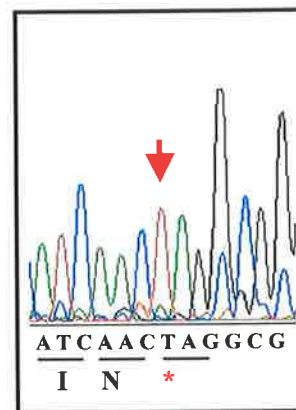
Figure 3.4: The chromatogram showing the OSP#3 mutation R368X, in exon 5 of *MID1*. The C nucleotide has been replaced by a T resulting in a premature stop codon.

Wild

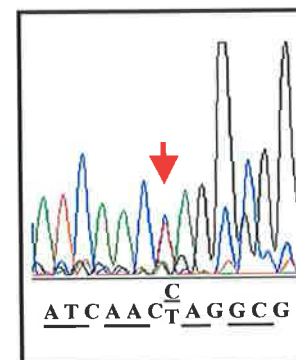


Proban

OSP#1



Mothe



Exon

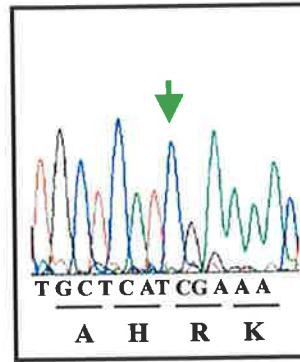
Figure 3.5: The chromatogram showing the OSP#10 mutation Q468X in exon 7 of *MID1*. The C nucleotide was replaced with a T resulting in a premature stop codon. The mother of OSP#10 carried both the wild type and the Q468X mutated form of *MID1*.

The final nonsense mutation found was a nucleotide change 1483C>T in OSP# 9 producing a premature stop codon at the amino acid level (R495X) (Figure 3.6). There were no family members available for mutation analysis. This mutation resulted in the deletion of the CTD in the MID1 protein (midin).

3.2.1.2 Frameshift Mutations

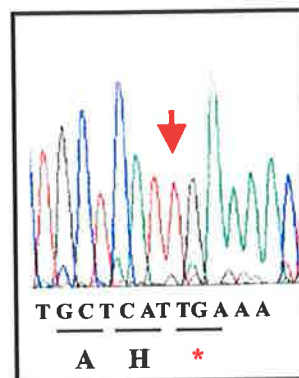
There were two frameshift mutations found in the *MID1* gene, in exon 5 and 7. These frameshifts resulted in the truncation of midin, disrupting the FNII and CTD regions. The frameshift in exon 5 was found in OSP#6 and was caused by a deletion of a single C nucleotide at position 1051 (Figure 3.7). The computer program DNasis revealed that at the protein level, this mutation resulted in the disruption of the FNIII domain and CTD due to the presence to premature stop codons. Genomic DNA samples were available from the family members of proband OSP#6 who were screened for the frameshift (F/S) mutation. SSCP was used to confirm the presence of the mutation in family members and confirm those affected, determine those who were carriers and those unaffected. Mutations were observable in family members because the size of the band was one base pair smaller in exon 5 due to the single nucleotide deletion. This caused a downward shift on the SSCP gel (Figure 3.8). The control samples used in SSCP did not show the downward shift. Family members who were carriers (a wild type *MID1* copy and a mutated copy of *MID1*) had the downward shift as well as the normal band (Figure 3.8).

Wild



Proban

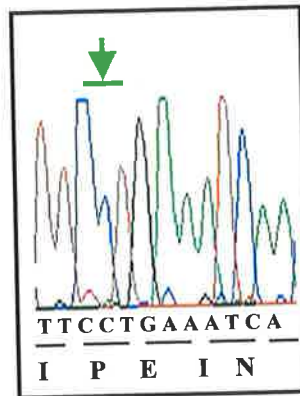
OSP#



Exon

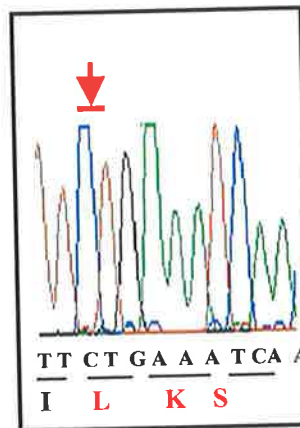
Figure 3.6: The chromatogram showing the OSP#9 mutation R495X in exon 8 of *MIDI*. The C nucleotide was replaced with a T resulting in a premature stop codon.

Wild



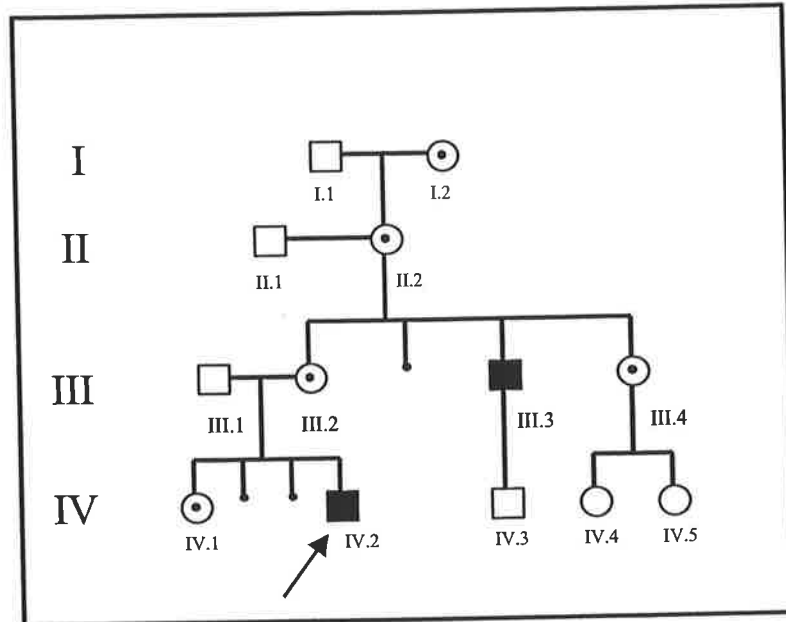
Proban

OSP#



Exon

Figure 3.7: The chromatogram showing the OSP#6 frameshift mutation 1051delC in exon 5 of *MID1*. The single C nucleotide was deleted in the proband, causing a frame shift in the remaining 3' *MID1* sequence.

A**B**

I.2 II.2 III.1 III.4 III.3 IV.3 IV.1 III.2 IV.2 C

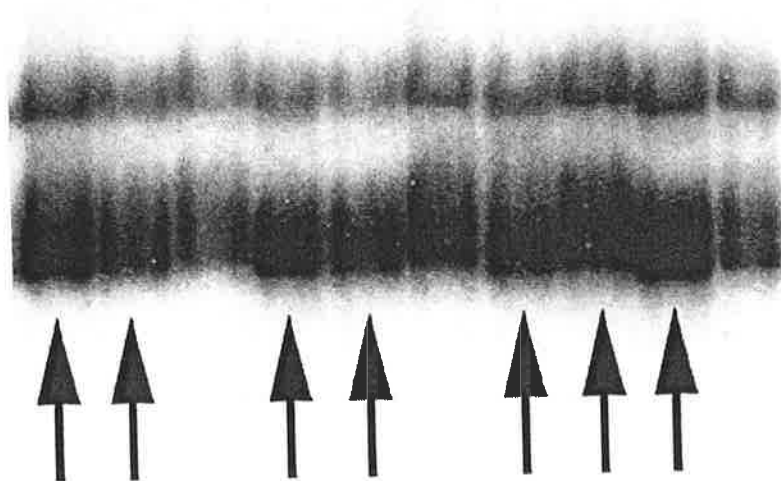


Figure 3.8: The OSP#6 family members. (A) Pedigree of OSP#6. Additional family members of OSP#6 that have the frameshift mutation are I.2, II.2, III.2, III.3, III.4, and IV.1.

The arrow shows the proband OSP#6. (B) The SSCP gel showing the OSP#6 family members. Lane C was a normal sample. The arrows show where there is a downward shift.

Results from the SSCP analysis revealed that a number of these family members did, in fact have the mutant form of the *MIDI* gene. The sister of OSP#6 was found to have the same mutation in the *MIDI* gene, as well as his mother, uncle, aunt, grandmother and great grandmother as indicated on the pedigree (Figure 3.8). Each of these family members exhibited features of OS but had reduced severity of the phenotype. The proband had the most severe phenotype that included congenital heart defects, bilateral cleft lip/palate, hypospadias, hypertelorism, palpebral fissures and posteriorly rotated ears. His sister was not as severely affected but had congenital heart defects and hypertelorism. The proband's mother only displayed telecanthus and the father's phenotype was normal. The uncle had a relatively mild OS phenotype which included hypospadias, marked hypertelorism and feeding difficulties as a child. He had a son who displayed no OS phenotype. The aunt and grandparents showed no characteristics of the OS phenotype. However, the grandmother did have a copy of the mutated form of *MIDI* present. Finally, the great grandmother of the proband was diagnosed with mild OS because of the presence of hypertelorism. There were also a few miscarriages throughout the pedigree, raising the possibility that these miscarriages may have been caused by severe expression of the OS phenotype (Figure 3.8). However, the miscarriages were not investigated and they could have been caused by other factors.

The other frameshift (1330insA) mutation was found in the only British patient in this study, OSP#16, and has since been reported (Gaudenz *et al*, 1998). The mutation was the result of an insertion of a single A nucleotide at position 1330. This resulted in the latter part of the FNIII and CTD domains being completely disrupted, which would likely result in a change in the midin function.

3.2.2 Exon Deletion

To sequence the *MIDI* gene in the OS probands, each exon was individually amplified by PCR. Despite numerous attempts, exon 2 of *MIDI* in OSP#11 could not be amplified. DNA was available from the probands mother and exon 2 was successfully amplified by PCR and sequenced. The automated sequencing results revealed that exon 2 was normal in the probands mother. However, it was possible that the mother may be a carrier. Therefore, the mother would have both a normal *MIDI* copy (detected in PCR) and a copy of *MIDI* with exon 2 deleted (not detected in PCR). To investigate this mutation further, the first and third exons of OSP#11 were amplified by PCR from genomic DNA, indicating that the deletion occurred in the intron region between exons 1/2 extending to the intron between exons 2/3. To further define the size of the deletion in OSP#11, primers of a sequenced tagged site (STS239A) located in the intron region between exon 1 and 2 were amplified by PCR (Figure 3.9 and 3.10). PCR amplification was possible from the STS239A in both the mother and proband (Figure 3.10). This indicated that the deletion in this region of *MIDI* for OSP#11 was not large. However, even though the size of the deletion was not determined in the genomic sequence, exon 2 appears to be eliminated in the cDNA. There is a possibility that the primers used to amplify exon 2 may not have bound due to a rare polymorphism or alteration within the intron region surrounding exon 2. In case the primers were not binding during PCR, the annealing temperature was altered numerous times. However, there was still no amplification of exon 2. In the future, the Δ exon2 mutation could be further confirmed by designing internal primers to exon 2 as well as primers designed to a different region in the intron. Alternatively, Southern analysis could have been used to confirm the Δ exon 2 mutation. However, there was insufficient genomic DNA available from the proband and the mother to complete Southern analysis.

The deletion of exon 2 would cause the very first 32 amino acids of the coiled-coil motif to be eliminated and result in an in-frame deletion. If the MID1 domains interact with one another for normal function and part of a motif was missing (as seen here for Δ exon2), midin protein may not function properly. Since the deletion of exon 2 results in severe OS, the coiled-coil domain encoded by exon 2 must be important in the normal function of midin. The proband's mother displayed the mild OS characteristic of hypertelorism and PCR amplification of exon 2 was always less intense compared to the controls (Figure 3.10) indicating that she may be a carrier.

3.2.3 Missense Mutation

There was one missense mutation found in exon 9 of *MID1* for OSP#5 (Figure 3.11). The change, T > C, was at position 1877 altering the amino acid at 626 from a leucine to a proline (L626P) which would result in the midin protein structure being altered since the proline causes a change in the orientation of the proteins structure. There were no family members available for further mutation screening. The L626P was the most C-terminal mutation reported, with only a single amino acid change, suggesting that this particular region of the CTD is also important for midin to function normally.

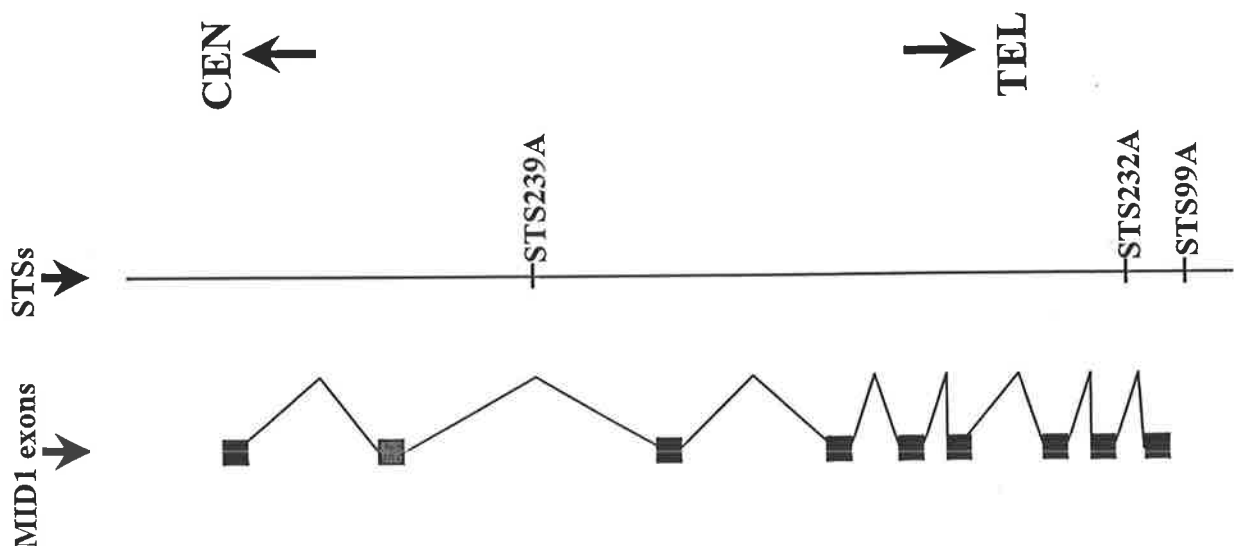


Figure 3.9: A schematic diagram of the STSs in relation to the *MID1* gene. The position of the STS239A (highlighted in blue) in relation to exon 2 can be seen in this diagram. The exon highlighted in red shows the exon 2 of *MID1*.

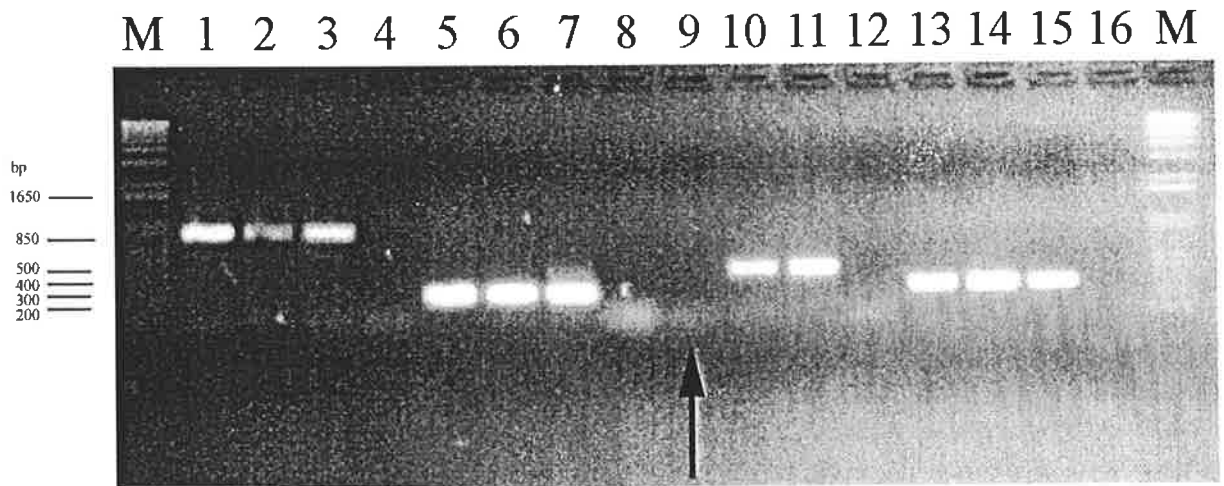
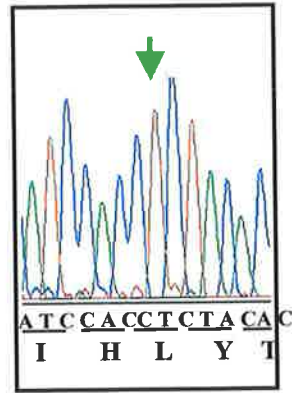


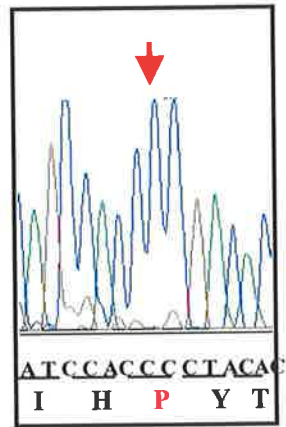
Figure 3.10: The *MID1* deletion of the exon 2 in OSP#11. Lane M represents the 1kb marker used. Lanes #1, 5, 9 and 13 represent OSP#11, lanes #2, 6, 10 and 14 the mother, lanes #3, 7, 11 and 15 a genomic control (unaffected individual) and lanes #4, 8, 12 and 16 a negative control (no DNA). Lanes #1-4 show the PCR amplification of exon 1 from *MID1*, lanes #5-8 the microsatellite STS239A, lanes #9-12 the exon 2 from *MID1* and lanes #13-16 the exon 3 of *MID1*. The arrow shows where the amplification of exon 2 should have been if it was present in OSP#11.

Wild



Proban

OSP#



Exon

Figure 3.11: The chromatogram showing the OSP#5 mutation L626P in exon 9 of *MID1*. The T nucleotide was replaced with a C resulting in an amino acid change from leucine to a proline.

3.2.4 OS Probands With No *MIDI* Mutations

It has been shown that mutations in *MIDI* are responsible for the X-linked form of OS (Quaderi *et al*, 1997; Gaudenz *et al*, 1998, and this study). The mutational analysis of *MIDI* in the 16 patients did not show any changes for OSP#1, 2, 4, 7, 8, 13, 14, and 15. However, these OS probands may have mutations in regions not sequenced (e.g. introns and promoter region). For example, a mutation in the promoter region could affect the expression levels of *MIDI* and a mutation in the intron region close to the intron/exon boundary could affect the splice junctions.

Also, OS is a heterogeneous disorder, and there may be mutations in genes that have not yet been identified. There have been previous linkage studies (Robin *et al*, 1995) and abnormalities (Urioste *et al*, 1995; Verloes *et al*, 1995; McDonald-McGinn *et al*, 1996; Lacassie and Arriaza, 1996) suggesting an association of OS with chromosomes 13 and 22, in addition to the X chromosome indicating that these chromosomes may also harbour OS loci. Since OS is heterogeneous, the autosomal form may be responsible for the OS phenotype in these other probands or there may be other OS loci on the X chromosome that have gone undetected. By genetic linkage analysis, OSP#8 inheritance at least was confirmed to be the X-linked form of OS using highly polymorphic microsatellite markers from chromosome X, 22 and 13 (Cox *et al*, 2000). The markers (DXS9994, DXS10006 and CxM40) surrounding the *MIDI* gene at Xp22, were co-inherited with the OS phenotype, further suggesting that the *MIDI* is involved in producing the OS phenotype. At this stage, there have been no large disruptions discovered on chromosome 22 in the above probands. However, there may be small disruptions present on chromosome 22 that have gone undetected.

3.2.5 Screening the Un-translated Regions of *MIDI*

When the *MIDI* cDNA sequence was isolated, two un-translated exons (U2 and U4) in the 5' region were identified (AF035360; AF269101) which had previously been unreported. To determine if there were any upstream deletions in the *MIDI* gene these un-translated exons, were amplified in PCR using genomic DNA from the 16 OS patients as a template. Subsequent to the mutational analysis, all 16 OS patients were screened for the presence of these untranslated exons. The two un-translated exons were clearly amplified by PCR indicating that there were no deletions in the 16 probands (data not shown). As these exons were untranslated and not coding exons, they were not sequenced. A computer-assisted analysis (Blast search, see Section 2.6) revealed that there are at least 5 untranslated exons present in the *MIDI* gene, with a possible sixth untranslated exon (U1) (Figure 3.12). This indicates that there may be different splicing patterns occurring when *MIDI* is transcribed. RNA samples from the OS patients used in this study would enable these splicing patterns to be investigated further. Unfortunately, at this stage of the investigation RNA samples were not available from the OS patients. Since these two untranslated exons were amplified using PCR, the other untranslated exons were not amplified, since it was considered more beneficial to investigate the different splicing patterns within this region. The sixth untranslated exon is not in the genome database and matches the first 38bp of the further cDNA sequence isolated in this study. This suggests that this exon may be located in the gap in the genomic sequence or be an artefact from the construction of the cDNA library.

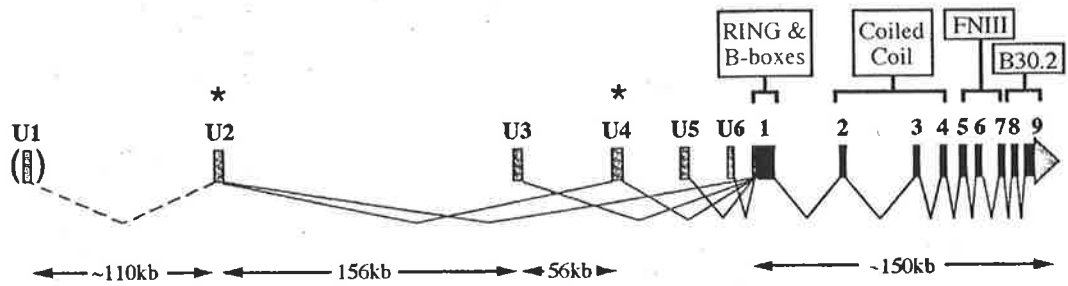


Figure 3.12: The MID1 gene structure showing the untranslated exons. The hatched boxes show the untranslated exons while the black boxes show the coding exons of *MID1*. The asterisk shows the untranslated exons (U2 and U4) identified in this study (modified from Cox *et al*, 2000).

3.3 MID2 Mutation Analysis

Since there were 50% of the OS cases that did not have a MID1 mutation identified, the *MID2* gene (Figure 3.13) was chosen as a possible candidate based on the sequence and structural similarity and possible functional redundancy between these two genes (Buchner *et al*, 1999; Perry *et al*, 1999). Consequently, *MID2* was screened for potential mutations in the OS probands. To screen the *MID2* gene for mutations, a direct sequencing approach was taken, as for the mutation analysis of *MID1*. A missense mutation was discovered that caused a change 1073C>A, in OSP#7, OSP# 9 and OSP#13 (Figure 3.14). This altered the amino acid sequence by changing an alanine to an aspartic acid, (A358D), between the tripatite motif and FNIII. Alanine is a neutral and hydrophobic amino acid, while aspartic acid is acidic. Therefore, this change could be enough to cause the mutated MID2 protein to interact differently. A t-test was performed to further confirm that A358D was a mutation. The results showed that the A358D mutation was greater than one percent suggesting that it is not a polymorphism.

SSCP was used to further determine if the mutation was likely to cause disease or was a polymorphism. OSP#7 had family members available for screening the mother (half-brother and –sister). In addition, 100 control samples were used (Red Cross, Adelaide). The OSP#7 change in the sequence resulted in a downward shift on the SSCP gel (data not shown). The control samples did not have the A358D *MID2* change present. The mother of the proband had the normal allele and the A358D mutation, the brother had the nucleotide change and the sister did not have the mutation. These results were confirmed by automated sequencing. Unfortunately, it was revealed that the brother and sister of the proband OSP#7 had a different father and there was no paternal information available.

Figure 3.13: The ORF *MID2* sequence. Normal (wild type) sequence used to align the automated sequencing from the OS patients to find mutations. The domains found in *MID2* are highlighted in the protein sequence. The RING finger motif is highlighted in red, the B-boxes in green, the coiled-coil domain in blue, the FNIII domain in pink and the CTD in yellow. The different colours in the nucleotide sequence show the exons found in *MID1*. Exon 1 is shown in dark blue, exon 2 in purple, exon 3 in yellow, exon 4 in dark green, exon 5 in orange, exon 6 in light blue, exon 7 in light green, exon 8 in red, and exon 9 in black.

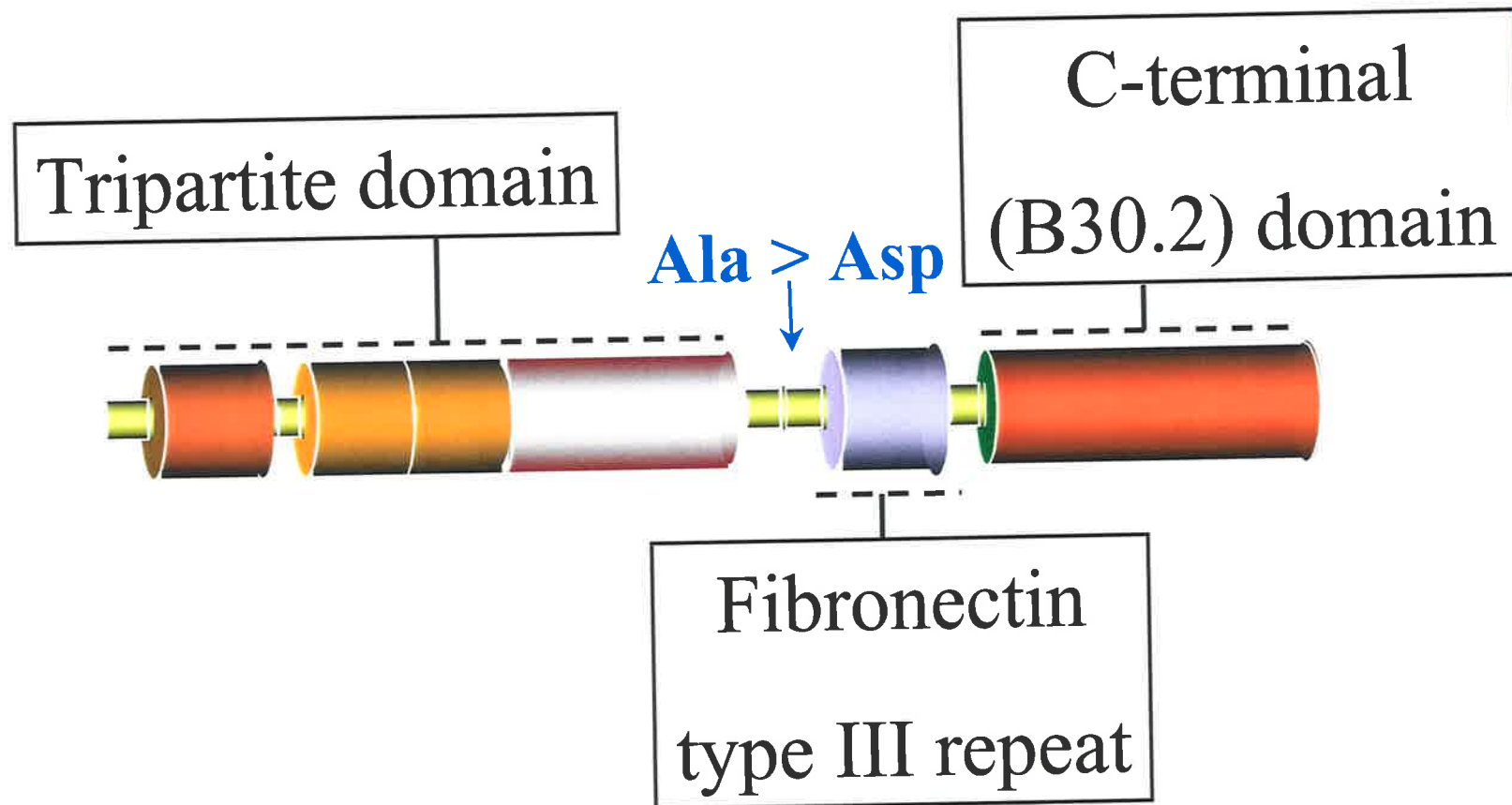
MID2 ORF

1 **M E T L E S E L T C P I C L E L F E D F** 20
 1 ATGGAACACTGGAGTCTGAATTGACCTGTCCAATCTGCCTAGAGTTGTTTGAAGACCCC 60
 21 **L L L P C A H S L C F S C A H R I L V S** 40
 61 CTTCTGCTCCCTTGTGCTCACAGCCTCTGCTTCAGCTGTGCCCATCGCATTTTGGTATCA 120
 41 **S C S S G E S I E P I T A F Q C P T C R** 60
 121 AGCTGCAGCTCTGGTGAATCCATTGAACCCATTACTGCTTCCAGTGTCCACCTGCAGG 180
 61 **Y V I S L N H R G L D G L K R N V T L Q** 80
 181 TATGTTATCTCGCTGAACCACCGGGGCTGGATGGCCTCAAGAGGAATGTGACTCTGCAG 240
 81 N I I D R F Q K A S V S G P N S P S E S 100
 241 AACATTATTGATCGCTTCCAGAAGGCTTCAGTCAGTGGGCCCAATTCCCCTAGTGAGAGC 300
 101 R R E R T Y R P T T A M S S **E R I A C Q** 120
 301 CGCCGGAAAGGACTTACAGGCCACCACCTGCCATGCTTAGCGAGCGAATTGCTTGCCAA 360
 121 **F C E Q D P P R D A V K T C I T C E V S** 140
 361 TTCTGTGAGCAGGACCCGCAAGGGATGCAGTAAAAACATGCATCACCTGTGAGGTCTCC 420
 141 **Y C D R C L R A T H P N K K P F T S H R** 160
 421 TACTGTGACCGTTGCCTGCGGGCCACGCACCCCAACAAGAACCTTTCACCAGCCACCGC 480
 161 **L V E P V P D T H L R G I T C L D H E N** 180
 481 CTGGTGAACAGTGCAGACACACATCTTCGAGGGATCACCTGCCTGGACCATGAGAAT 540
 181 **E K V N M Y C V S D D Q L I C A L C K L** 200
 541 GAGAAAGTGAACATGTACTGTGTATCTGATGACCAATTGATCTGTGCCTTATGCAAACCTG 600
 201 **V G R H R D H Q V A S L N D R F E K L K** 220
 601 GTGGGTCGTCACCGAGACCATCAGGTCGCATCCCTGAATGATCGATTTGAGAAACTCAAG 660
 221 **Q T L E M N L T N L V K R N S E L E N Q** 240
 661 CAAACTCTGGAGATGAACCTCACCAACCTGGTTAAGCGCAACAGCGAACTAGAAAATCAA 720
 241 **M A K L I Q I C Q Q V E V N T A M H E A** 260
 721 ATGGCCAAACTAATACAGATCTGCCAGCAGGTTGAG **TTGAATACTTCTATTCATGACCA** 780
 261 **K L M E E C D E L V E I I Q Q R K Q M I** 280
 781 **AAACTTATGGAAGAAATGTTGACCAATTTGCTAGAGATCATCCACAGAGDGAACCAATGATC** 840
 281 **A V K I K E T K V M K L R K L A Q Q V A** 300
 841 **GCTDTCAAAATCAAAATGAAACAAAG** GTTATGAAACTGAGAAAGTTGGCACAGCAGGTTGCT 900
 301 **N C R Q C L E R S T V L I N Q A E H I L** 320
 901 AATTGCGCCAGTGTCTTGAACGGTCAACAGTCTCATCAACCAAGCTGAGCATATCCTG 960
 321 **K E N D Q A R F L Q S A K N I A E R V A** 340
 961 AAAGAAAATGACCAGGCACGGTTTCTACAGTCTGCAAAAAATATTGCTGAGAG**GGTTCGCT** 1020
 341 M A T A S S Q V L I P D I N F N D A F E 360
 1021 ATGGCAACTGCATCTTCTCAAGTTCTGATTCCAGACATCAATTTTAATGATGCCTTTGAA 1080
 361 N F A L D F S R E **K K L L E G L D Y L T** 380
 1081 AACTTTGCTTTAGATTTTTCCAGAGAAAAGAAACTGCTAGAGGGGTTAGATTATTTAACA 1140
 381 **A P N P P S I R E E L C T A S H D T I T** 400
 1141 GCCCCAAACCCACCATCTATCCGAGAAGAACTCTGTACTGCCTCCCATGACACCATTACA 1200
 401 **V H W I S D D E F S I S S Y E L Q Y T I** 420
 1201 GTCCACTGGATCTCGGATGATGAGTTCAGCATCAGCTCCTATGAGCTTCAGTACACCATA 1260

421 **F T G Q A N F I S K S W C S W G L W P E** 440
 1261 **TTC**ACTGGCCAGGCTAACTTCATCAGTAAGTCATGGTGTAGTTGGGGCCTGTGGCCAGAG 1320
 441 **I R K C K E A V S C S R L A G A P R G L** 460
 1321 **ATA**AGGAAATGTAAGGAAGCAGTAAGCTGCTCAAGATTGGCCGGGGCGCCACGAGGCCTG 1380
 461 **Y N S V D S W M I V P N I K Q N H Y T V** 480
 1381 **TATA**ATTCACTAGACAGCTGGATGATTGTTCCCAACATTAAACAGAACCATTACACAGTG 1440
 481 **H G L O S G T R Y I F I V K A I N Q A G** 500
 1441 **CAT**GGACTCCAGAGCGGGACTCGCTACATCTTCATCGTTAAAGCCATAAACCAAGCCGGC 1500
 501 **S R N S E P T R L K T N S Q P F K L D P** 520
 1501 **AG**CCGGAACAGTGAACCTACCCGACTAAAAACAAACAGCCAACCCTTTAAATTGGATCCC 1560
 521 **K M T H K K L K I S N D G L Q M E K D E** 540
 1561 **AAA**ATGACTCACAAGAAGTTGAAGATCTCCAATGATGGATTGCAGATGGAGAAGGATGAA 1620
 541 **S S L K K S H T P E R F S G T G C Y G A** 560
 1621 **AG**CTCTCTAAAGAAGAGCCACACCCAGAGAGGTTTAGTGGCACAGGGTGTCTATGGGGCA 1680
 561 **A G N I F I D S G C H Y W E V V M G S S** 580
 1681 **GC**AGGAAATATATTCATTGACAGTGGCTGCCACTATTGGGAGGTGGTTCATGGGTTCTCTCA 1740
 581 **T W Y A I G I A Y K S A P K N E W I G K** 600
 1741 **AC**ATGGTATGCAATTGGCATTGCCTACAAATCAGCTCCAAAGAATGAATGGATTGGCAAG 1800
 601 **N A S S W V F S R C N S N F V V R H N N** 620
 1801 **AAT**GCCTCCTCATGGGTCTTCTCTCGCTGCAATAGTAACTTCGTGGTGAGACACAACAAC 1860
 621 **K E M L V D V P P H L K R L G V L L D Y** 640
 1861 **AAG**GAAATGCTGGTGGATGTGCCCCACACCTGAAGCGTCTGGGTGTCTCTCTGGATTAT 1920
 641 **D N N M L S F Y D P A N S L H L H T F D** 660
 1921 **GACA**ACAATATGCTGTCTTTCTATGACCCAGCTAACTCTCTCCATCTTCATACTTTTGAT 1980
 661 **V T F I L P V C P T F T I W N K S L M I** 680
 1981 **GT**GACCTTCATTCTTCCAGTTTGTCCAACATTTACAATCTGGAACAAATCCCTAATGATC 2040
 681 **L S G L P A P D F I D Y P E R Q E C N C** 700
 2041 **CT**GTCTGGCTTGCCCTGCCCCAGATTTTATTGATTACCCTGAGCGGCAGGAATGCAACTGC 2100
 701 **R P Q E S P Y V S G M K T C H *** 716
 2101 **AG**GCCTCAAGAATCCCCTTATGTTTCTGGGATGAAAACCTGTTCATTAA 2148

Figure 3.14: The *MID2* mutation identified from as part of this study. The A358D mutation in *MID2* occurred between the end of the tripartite domain and the beginning of the FNIII domain.

A353D in MID2



When screening the control samples for MID2 A358D using SSCP, there was an upward shift seen in one of the male samples. Automated sequencing of this sample revealed a 1 base pair deletion of an A nucleotide in the intron region of exon 5 and 6 (Figure 3.15). This variant was not present in any of the OS patients or other normal samples.

Of the 3 probands (OSP#7, 9 and 13) with the *MID2* mutation, OSP#9 also had a *MID1* mutation. OSP#9 had heart defects as well as a few minor characteristics such as mottled skin (Table 3.2), which were absent in OSP# 7 and 13. OSP#7 had diaphragmatic hernia that was not present in OSP#9 and 13. This does not clearly indicate whether or not the *MID2* mutation underlies the OS phenotype. From these data there is no clear evidence that the *MID2* mutation and *MID1* mutations interact to produce the OS phenotype.

GCG ACC↓CTG AAA AAA AAA AAA
Exon 6

Figure 3.15: The exon/intron boundary where the *MID2* variant was found. The nucleotides underlined are part of exon 6 and the arrow marks the start of the intron sequence between exons 6 and 7. One of the A nucleotides in the (A)₁₂ repeat sequence was deleted in the control sample.

3.4 Discussion

In order to determine the role of *MID1* in the pathogenesis of OS, a detailed mutational analysis was conducted using 16 probands (15 Australian and one British) diagnosed with Opitz Syndrome. A direct sequencing approach was undertaken to screen the *MID1* coding

exons in each of these probands. In addition, the isolation of the X-linked *MID2*, a homologue of *MID1*, led to the mutational analysis being extended to include this gene and determine whether or not it also was responsible for an X-linked form of Opitz syndrome.

The proband OSP#5 carrying the L626P mutation had the most severe OS phenotype in this study. The L626P mutation found in OSP#5 was a single amino acid change from a leucine to a proline. Leucine and proline are in the same amino acid group, as they both have a neutral charge and are hydrophobic. However, the change of an amino acid to a proline can be highly detrimental to the structure of the protein. The proline has the nitrogen atom of the amino group incorporated into its ring, causing the proline to disrupt the normal organisation of a polypeptide with a sharp change in direction of the chain. When a mutation leads to the incorporation of a proline, the result is a 'turn' in the protein structure that would not normally be there, thus changing the conformation of the protein and potentially altering the function of the protein (Lewin, 1994; Lehninger *et al* 1993). In *MID1*, the change to a proline would change the structural orientation of midin and therefore may have a large affect on the function of midin. This in turn could potentially affect how midin interacted with other proteins. The amino acid is in a conserved region within the CTD across all species from which *MID1* has been isolated (human, mouse, rat, chick, and fugu, unpublished data) and in *MID2* (Perry *et al*, 1999).

A lack of correlation between the genotype and phenotype in OS can be observed in the OSP#6 family, where the *MID1* mutation can be traced over four generations. Family members that have the *MID1* mutation display a varied severity of the OS phenotype. In this study, the only 4 OS probands with reported structural heart anomalies were found to have *MID1* mutations. However, there was no other similarity between the *MID1* genotype and particular OS phenotypes described in the probands.

The mutations found in the *MID1* gene are dramatic to its structure (e.g. nonsense, deletions), suggesting that the function of MID1 is more likely to be structural. The diverse effects of the *MID1* mutations and the lack of correlation between genotype and phenotype would suggest that the OS phenotype may be caused by a change of midin interactions with other proteins, thus effecting numerous developmental pathways. The change in protein interactions of midin in the OS phenotype can be further supported by the observation of OS features in females with MLS (MIDAS) syndrome, which is a male lethal disorder associated with large Xp22 deletions over the *MID1* gene (Kayserili *et al*, 2001).

The high variability of the OS phenotype could, therefore, be the result of the ability of other factors' (eg. other microtubule associated proteins, MAPs or MID2) to compensate for the loss of MID1 function. Genetic polymorphism in such compensatory factors would assist in explaining the variation of the OS phenotype. One compensatory factor could be the *MID2* gene, since it has a high similarity to *MID1* and its expression pattern is similar (Perry *et al*, 1999; Buchner *et al*, 1999). For instance, a mutation in MID1 may affect the expression level, however this could be compensated by MID2, which may reduce the abnormalities produced during embryogenesis.

The loss of function of MID1 would suggest that the interactions of midin are significant in the OS phenotype. During embryogenesis, if midin was unable to interact with proteins for normal development, then this would result in a disruption of a number of processes and ultimately the malformation of organs.

In this study and the reports by Quaderi *et al* (1997) and Gaudenz *et al* (1998), there have been mutations found throughout *MID1* affecting all the domains, except the RING finger. The B-box motif has been affected as a result of the E115X mutation, however, there has been no mutations reported in the B-box. Overall, there have been 18 mutations reported in *MID1* (Quaderi *et al*, 1997; Gaudenz *et al*, 1998; Cox *et al*, 2000). Other laboratories have

reported ten of these mutations: 6 of these 18 mutations were located in the CTD, 2 in the FNIII and 2 in the coiled-coil domain. Of the 8 mutations identified in this study, 4 specifically disrupt the CTD (R495X, 1330insA, 1052delC and L626P). The Q468X resulted in the disruption of part of FNIII and all of CTD, and R368X caused both FNIII and CTD to be lost. The deletion in exon 2 (Δ exon2) in proband OPS#11 only affected the first part of the coiled-coil domain, leaving the remaining motifs intact. Finally, the E115X change occurred prior to the B-box resulting in a truncated protein that lacked all the domains except the RING finger. This highlights that while a number of mutations perturb the CTD, other domains have been disrupted, indicating their importance also for normal midin function, and supports overall loss of function as the underlying cause of X-linked OS.

The isolation of the *MID2* gene (Buchner *et al*, 1999; Perry *et al*, 1999), which has a very high level of similarity to *MID1*, led to the mutational screen of this 'second' *MID* gene in the 16 probands available. There was a single missense mutation discovered in exon 5 of *MID2* in three of the probands (OSP#7, 9, and 13). To ensure that this alteration of an alanine to an aspartic acid was a mutation and not a polymorphism, SSCP analysis was utilised. In the 100 control samples screened by SSCP, there were 50 females and 42 male samples resulting in the testing of 142 X chromosomes for the 1073C>A (A358D) change. The A358D change was not detected in these control samples, which statistically indicates that it was in fact a mutation since the percentage is less than one and generally polymorphisms are found in greater than one percent of the population. A t-test further suggested that statistically A358D was a mutation. However, it is important to note that the number of OS patients is not large enough to make a firm conclusion.

Of the three probands with the A358D mutation in *MID2*, only OSP#9 had a mutation in *MID1* (R495X). Again there was no correlation between the genotypes of *MID1* and *MID2* and the particular OS phenotype. The single amino acid change for the *MID2* mutation

resulted in the neutral and hydrophobic alanine being change to an acidic amino acid (aspartic acid). This change may be sufficient to alter the structure of the MID2 protein and its ability to interact with other proteins, therefore influencing the clinical severity of Opitz Syndrome. However, since there were no mutations discovered in the MID1 gene in many of these patients it remains possible that either MID1 or MID2 mutations are capable of producing the OS phenotype.

There were family members from the OSP#7 proband that were available for mutational analysis. The presence of the mutation was confirmed in this family be by SSCP and sequencing results, the mother was a carrier, the brother had the affected allele, while the sister was normal. This suggests that the A358D mutation was inherited from the mother of the proband. However, there could be no complete interpretation of these results due to the brother and sister having a different father to the proband and there was no paternal information. The brother and sister did not display the OS phenotype and the mother showed a mild OS phenotype.

While attempting to confirm A358D, a novel variant was detected for *MID2*. The novel variant was found in the intron region between exons 5 and 6 caused by a deletion of a single A nucleotide. This variant was not present in any of the OS probands or in the other normal individuals (142 X chromosomes were examined by SSCP analysis). The deleted A nucleotide was part of a (A)₁₂ repeat sequence (Figure 3.15). Generally, repeated sequences of greater than least 11 nucleotides have an increased possibility of polymorphism (Weber and May, 1989) and therefore while statistically this alteration is below one percent of the population, it was considered to be a rare variant. The deletion of a single A may arise due to the DNA polymerase mis-reading the number of repeats. While this variant is close to the splice junction it is unlikely to affect the function of MID2 because it is part of a poly (A)

string. If the variant was within 2 bases of the splice junction and resulted in a change of base, it may be more likely to affect the function of MID2.

To determine the function of *MID1* mutations, E115X, Δ exon2, R368X and L626P, were selected for further investigation on their sub-cellular localisation and protein-protein interactions (see Chapter Four).

Chapter Four:

Investigation of MFD₁ Mutations

Investigation of *MIDI* Mutations

Microtubules consist of the proteins α and β tubulin producing a robust polymer that has a high resistance to bending and compression. In non-dividing cells microtubules are key components in the organisation of the cytoplasm. Together with actin filaments, microtubules have a role in organising the spatial distribution of organelles within the cell (Wade and Hyman, 1997). In dividing cells, a large number of microtubules comprise the mitotic spindle that orientates the plane of cleavage and separates chromosomes during mitosis (Desai and Mitchison, 1997). Microtubules are arranged with the negative end associated with the microtubule organising centre (MTOC) whilst the positive end is responsible for the elongation of the microtubule network within a cell (Gundersen and Cook, 1999).

In cell culture experiments, endogenous and over-expressed MID1 protein (midin) have been shown to associate with microtubules (Schweiger *et al*, 1999; Perry *et al*, 1999). Since the endogenous levels of midin are low and over-expression of normal midin does not significantly change its ability to associate with the microtubule network, over-expression of the mutant midin was used for the intracellular localisation experiments.

Various mutations in genes can result in an alteration in the translated protein. For instance, the location and type of the mutations in the *MIDI* gene could result in the alteration in part of a domain (e.g. L626P), in an entire domain (e.g. Δ exon2), or in a number of domains (e.g. E115X). The mutated gene may affect either function or the intracellular localisation of the protein. As a consequence of the mutation, the alteration in protein function can range from a slight modification in activity to a total loss of function.

To determine if the mutated forms of midin had different intracellular localisations to wild-type midin, four mutations, E115X, Δ exon2, R368X and L626P were selected. The E115X, Δ exon2, R368X and L626P mutations were chosen because of the location of the

change and the consequent domains that were then affected. Green fluorescent protein (GFP) as an in-frame tag was used to investigate the intracellular localisation of these mutated forms of midin. Each of these mutated sequences were inserted into a pEGFP-C2 expression construct, transfected into a CosI cell line and compared to the intracellular localisation of the wild-type midin.

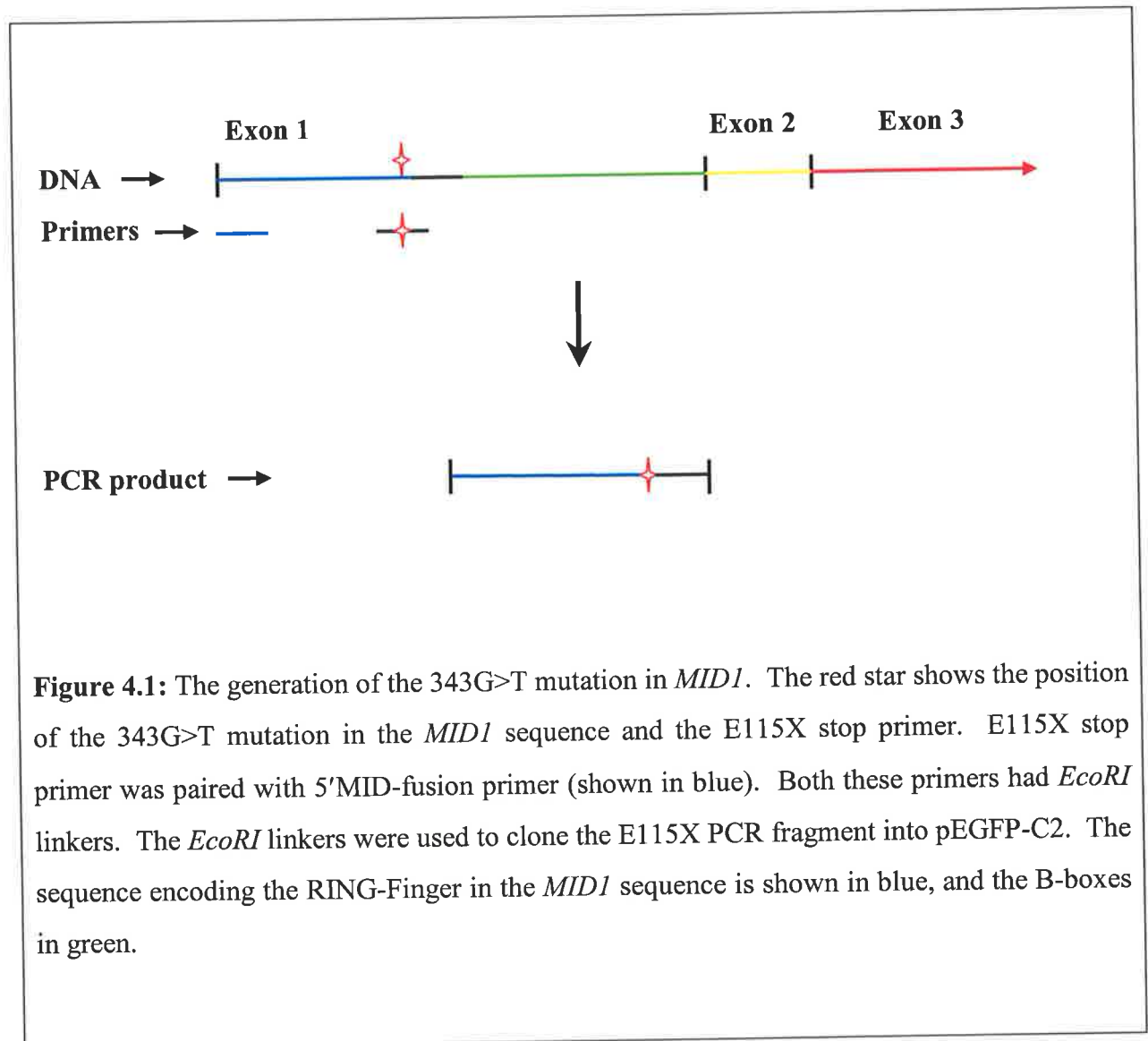
4.1 Constructing the GFP/MID1-Mutated Vectors

Four of the eight mutations (E115X, Δ exon2, R368X and L626P) were selected because these mutations represented a variety of the types of novel mutations in *MID1* found in this study. The E115X mutated protein only had the RING-finger motif present; the remaining motifs had been deleted. Δ exon2 mutated protein had an in-frame deletion of exon 2 affecting the first part of the coiled-coil domain. The majority of mutations found in *MID1* have been truncating resulting in the deletion of the CTD region; the R368X is an example of such a mutation with the FNIII and CTD domains deleted. L626P mutated protein is unique as it is the only missense mutation found to date that affects the CTD. These mutated forms of midin were suspected of having their intracellular localisation altered due to the mutation, and therefore, provided an opportunity to investigate the midin motifs using GFP analysis.

Sequences carrying the E115X, Δ exon2, R368X and L626P mutations were transferred into the pEGFP-C2 vector by a combination of PCR, cloning or site directed mutagenesis techniques (see Section 2.13). These *MID1* mutations were then fused in-frame into the pEGFP-C2 vector.

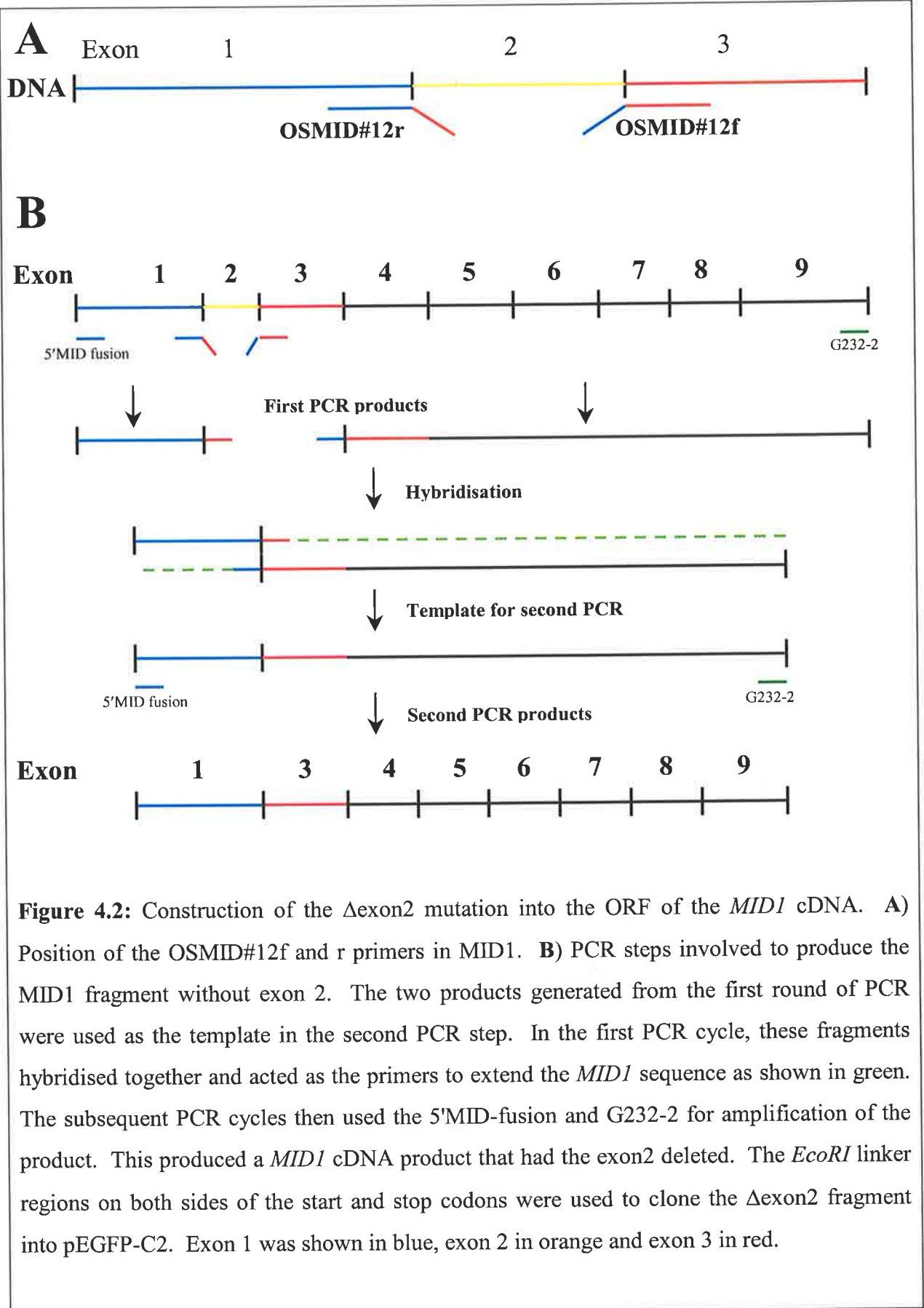
The generation of the E115X mutation involved PCR. A 'reverse' primer was designed over the region of the mutation such that the premature stop codon of the E115X mutation was incorporated into this primer. The E115X stop primer was paired with the 5' MID-fusion primer to generate the E115X in the PCR product. Both of these primers had

EcoRI linker regions that were used to clone the E115X PCR fragment into the pEGFP-C2 vector in-frame (Figure 4.1). Automated sequencing confirmed the insert was in-frame with the vector and the mutation had been introduced successfully.



The non-truncating Δ exon2 mutation provided the opportunity to investigate how the intracellular localisation of midin was altered when the first third of the coiled-coil was specifically deleted. The generation of the Δ exon2 mutation involved two PCR steps. Two primers were designed (MIDOS#12f and r) in opposite directions over the end of exon 1 and the start of exon 3. The first PCR step involved two separate PCR reactions, 5'MID-fusion and MIDOS#12r primers, MIDOS#12f and G232-2 primers with the full length *MID1* cDNA was used as the template. This produced two PCR fragments. The most 5' fragment contained the start codon in exon 1 and the first 18 nucleotide bases from exon 3 (from MIDOS#12r). The other PCR fragment contained the first 15 bases of exon 1 (from MIDOS#12f) followed by exon 3 and the remaining exons including exon 9, which had the stop codon. The second PCR step used these two fragments as a template with the primers 5'MID-fusion and G232-2. There was a small overlap (~15 nucleotides) between these two fragments that would have hybridised during the first PCR cycle and then acted as the primer to fill in the nucleotides. The subsequent PCR cycles used the 5'MID-fusion and G232-2 as the primers ensuring that an entire length of MID1 minus exon 2 was produced (Figure 4.2). These two primers also had *EcoRI* linkers that were used to clone the Δ exon2 PCR fragment into the EGFP vector in-frame with the pEGFP-C2 coding region. Automated sequencing confirmed that exon 2 was deleted and the insert was in-frame.

The comparison of the intracellular localisations of R368X and L626P enabled a comparison of a truncating and a non-truncating mutation affecting the carboxy terminus. pEGFP-C2 vectors with the R368X and L626P mutations were previously constructed (see Section 2.13) by J.Sibbons (Adelaide University) and transfected by Q.Schwarz (Adelaide University) using site-directed mutagenesis, PCR and cloning techniques. If the single amino acid changed the intracellular localisation, then this would further suggest that L626P was a mutation, and not a common polymorphism.



4.2 Intracellular Localisation of Mutant MID1 Protein

The pEGFP-C2/MID1 mutation constructs were transfected into CosI cells and analysed under an Olympus AX70 fluorescence microscope (Olympus Australia, Mount Waverly, Victoria) to determine if their cellular localisation was different to the wild-type midin. When transfected into CosI cells, signal from the wild type midin (normal) was observed to be associated with the microtubule network as seen by the artificial fluorescent yellow generated by the overlay of the photographs of the GFP-MID1 (green fluorescence) co-localising with the α -tubulin (red fluorescence) (Figure 4.3a-d). Similar observations have been reported from other studies (Gaudenz *et al*, 1998; Schweiger *et al*, 1999; Cainarca *et al*, 1999).

4.2.1 GFP-E115X Analysis

The E115X mutated protein contained the RING finger motif of the tripartite domain and was chosen as it represented the most N-terminal mutation identified to date. When transfected into CosI cells, this mutated form of midin was clearly localised throughout the cytoplasm and in the nucleus (Figure 4.3e-h). Since the wild-type midin localised only in the cytoplasm, it appeared that without the entire tripartite domain, the FNIII and CTD, midin had lost its ability to remain only in the cytoplasm as shown by the green fluorescence still present in the overlay of the blue, green and red fluorescence pictures (Figure 4.3h). When the GFP and α -tubulin staining patterns were overlaid, it was difficult to determine co-localisation of the midin and microtubules. This mutant midin appeared to either no longer associate, or had a reduced ability to associate, with the microtubule network.

4.2.2 GFP- Δ exon2 Analysis

The Δ exon2 mutant only had the first third of the coiled-coil domain deleted, with the remaining midin motifs intact and assumed functional. When the Δ exon2 mutation was transfected into CosI cells, the mutant midin displayed a completely different intracellular localisation to the wild-type midin. The Δ exon2 midin was found distributed throughout the cytoplasm and was absent from the nucleus of the cell. However, it was not associated along the length of the microtubules as displayed in the overlay by green fluorescence (Figure 4.3i-1). This suggested that the coiled-coil domain may have a role in the association of midin to the microtubules.

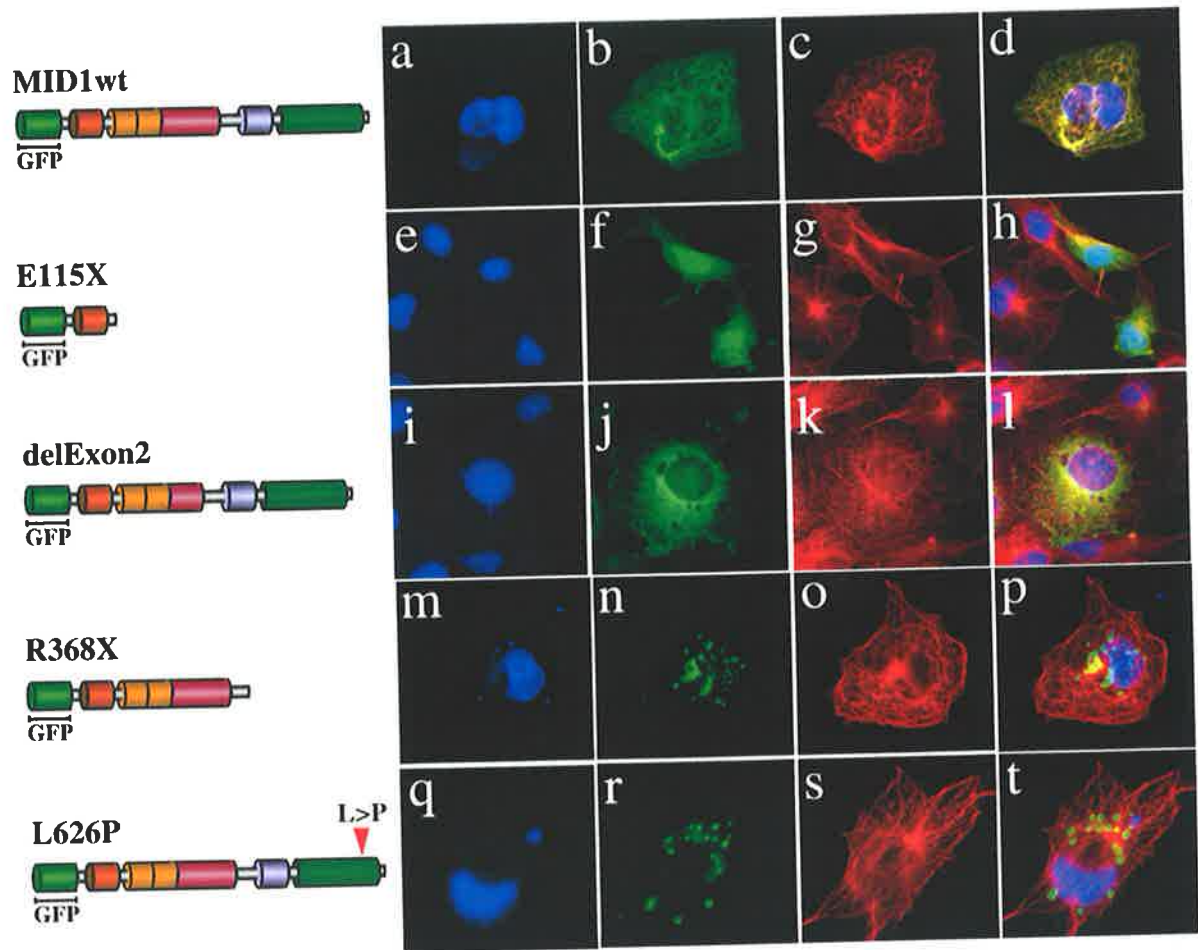


Figure 4.3: Intracellular localisation of the wild-type and mutated forms of midin in CosI cells. The normal wild-type MID1 (a-d) shows cytoplasmic localisation and microtubule association. The mutated forms of midin, E115X (e-h), Δ exon2 (i-l), R368X (m-p) and L626P (q-t) have different intracellular localisation in the CosI cells. The E115X and Δ exon2 have reduced association with the microtubules. E115X was localised in the nucleus and cytoplasm while the Δ exon2 was only localised in the cytoplasm. The R368X and L626P mutations display cytoplasmic aggregates. The pictures from left to right show DAPI staining, the GFP fluorescence, α -tubulin staining and then these three pictures are merged together in the fourth picture. The α -tubulin antibody in these experiments highlighted the position of the microtubules. The DAPI staining (blue) showed the nucleus of the cell. On the left side are diagrams representing the type of midin protein expressed (taken from Cox *et al*, 2000). I completed the experiments and photographs for e-h and i-l, the others were completed by Q.Shwarz (Adelaide University).

4.2.3 GFP-R368X Analysis

The GFP-R368X truncated protein lacking the FNIII and CTD domains showed cytoplasmic aggregation (when transfection into CosI cells) (Figure 4.3m-p). The aggregation was the result of the formation of high molecular weight cytoplasmic complexes as seen by the large individual artificial yellow aggregates present in the overlaid photograph. A prior study has shown these aggregates have a high molecular weight (Schweiger *et al*, 1999). These cytoplasmic complexes appeared not to associate with the microtubule network since there was no artificial fluorescent yellow produced along the lengths of the microtubules in the overlaid photographs as seen for the wild-type midin. Since this particular mutation had both the FNIII and the CTD missing, it is possible that these domains also have a role in the association of midin to the microtubules.

4.2.4 GFP-L626P Analysis

When the GFP-L626P fusion construct was transfected into CosI cells, a cytoplasmic distribution of fluorescence was observed that varied in appearance between cells even in the same experiment (Figure 4.3q-t). Similar to the GFP-R368X, cytoplasmic complexes were apparent. The cytoplasmic complexes seen here were consistent with C-terminal truncations, described in reports by other laboratories (Gaudenz *et al*, 1998; Schweiger *et al*, 1999; Cainarca *et al*, 1999).

4.3 Discussion

Four mutations, E115X, Δ exon2, R368X and L626P, were chosen for GFP-tag analysis due to the position of the alteration in the *MIDI* sequence. These mutations were transferred

into the wild-type *MID1* nucleotide sequence using PCR (E115X, Δ exon2 and R368X) or site-directed mutagenesis (L626P). Once the mutations had been generated, they were cloned in-frame into the pEGFP-C2 vector, transfected into CosI cells and the intracellular localisation determined for comparison to that of wild-type midin. The four mutated proteins displayed different localisation to the wild type midin suggesting the domains affected by the mutation can result in a change of function of the affected motif and consequently the overall function of midin.

The two mutations, Δ exon2 and E115X, were the most N-terminal changes identified and showed a different localisation to all described C-terminal changes, including L626P and R368X midin. There were no cytoplasmic complexes produced from the over-expression of these two N-terminal changes. The Δ exon2 was found diffused throughout the cytoplasm with no microtubule association, while E115X was localised in both the cytoplasm and nucleus. These intracellular localisation results indicate that the coiled-coil domain could be important for the association of midin to the microtubules and the tripartite domain may be necessary for midin to reside in the cytoplasm.

These results were compatible with other findings in the laboratory, where the individual motifs of MID1 had been systematically deleted and fused to the GFP vector (Short *et al*, submitted). For example, the tripartite domain has been completely deleted in experiments undertaken by others in the laboratory and resulted in midin losing its ability to reside only in the cytoplasm (Short *et al*, submitted). This was analysed further by deleting each domain of the tripartite motif and it was found that the cytoplasmic retention was controlled by the B-box (B Hopwood, unpublished data). These observations show the variable results possible from the N- or C- terminal mutated proteins in relation to midin cellular localisation. Yeast two-hybrid experiments have shown that in-frame deletions of MID1 motifs do not have a significant change on the conformation of midin. For example, the

interaction of the B-box domain with the rapamycin-sensitive PP2A regulatory subunit, Alpha 4 protein is unchanged when the individual motifs such as coiled-coil and CTD are deleted (Short *et al*, submitted).

In addition, the cytoplasmic complexes indicated the midin mutated proteins were likely to interfere with the site of function for midin. Of the mutations studied here, when the coiled-coil domain was deleted (E115X and Δ exon2), there was no aggregation observed. However, if the coiled-coil domain was intact and the mutation resided in the carboxyl terminus (R368X and L626P), then cytoplasmic aggregation was seen. This suggests that the coiled-coil domain and not the CTD controls the cytoplasmic aggregation. From yeast two-hybrid analysis, one theory is that the coiled-coil domain is responsible for homo-dimerisation of midin and thus allows the CTD to interact along the length of the microtubules (Short *et al*, submitted). If there is a mutation in the CTD region, then the resultant protein can still dimerise because of the functional coiled-coil domain. However, the mutated protein can no longer associate with the entire length of microtubules, and therefore, the mutated protein aggregates.

In addition to this study, concurrent intracellular localisation studies (Schweiger *et al*, 1999, Cainarca *et al*, 1999 and Perry *et al*, 1999) have also shown the wild type MID1-GFP fusion is associated with microtubules and when a truncating mutation is present in the CTD, aggregation of mutant midin occurs and no longer associates with the microtubules. The CTD mutation, L626P, identified in my study was also consistent with these the studies, displaying aggregation of the perturbed midin. This missense mutation could have such an affect due to the position of the single amino acid change near the end of the CTD in a region conserved across *MID1* and *MID2* homologues. Alternatively, it could be the type of amino acid change. The mutation changes a leucine to a proline. For the protein the addition of a proline can be

highly detrimental to the tertiary structure by introducing a turn and is more likely to prevent normal interaction with other proteins.

Ultimately, these mutations and the change in their cellular localisation were responsible for the OS phenotype. Results from the GFP analysis show the intracellular localisation of these mutations further confirmed that whilst the CTD is crucial for midin to function normally, the other domains also play a role. Each motif, except the RING finger and the FNIII domains appeared to have specific role in the overall function of midin.

The diverse effects of the *MIDI* mutations and the lack of correlation between genotype and phenotype would suggest that the OS phenotype may be caused by a general loss of function of *MIDI* as shown by the loss of ability for mutated midin to associate with the entire length of microtubules. The high variability of the OS phenotype could be the result of the ability of other factors (e.g. other microtubule associated proteins, MAPs and MID2) to compensate for MID1 function. This would suggest that cells with high levels of MID1 expression normally may be more likely to be affected. The loss of function of MID1 would suggest that the proteins interacting with midin may have an indirect role in the OS phenotype. This led to current investigation of how some of the *MIDI* mutations interact with MID1 and binding partners of *MIDI* in yeast two-hybrid systems in the laboratory. For example, one compensatory factor could be the MID2 protein since it has high similarity to midin and its expression pattern is similar (Perry *et al*, 1999; Buchner *et al*, 1999). Genetic polymorphism in such compensatory factors that affect either levels of expression or their overall activity would assist in explaining the variation of the OS phenotype under this hypothesis.

The microtubule structure produces a large network, providing a surface for protein-protein interactions (Gundersen and Cook, 1999). The binding of microtubule associated proteins (MAPs) stabilises and facilitates assembly of microtubules (Hirokawa, 1994). Examples of MAPs include the kinesin and dynein molecular motors, plus MAP1, 2, 3, 4 and

tau. Tau, MAP 2 and 4 have common C-terminal microtubule-binding domains (Hirokawa, 1994; Maccioni and Cambiazo, 1995) and a nearby proline-rich region that influences microtubule binding (Ulitzur *et al*, 1997). When these MAPs are transfected into a cell line, microtubule bundling is observed similar to the bundling observed with overexpression of the GFP/MID1 fusion protein in Cos1 cells (this project, Schweiger *et al*, 1999, Cox *et al*, 2000). Further investigation of the microtubular bundling processes for various MAPs may provide possible clues on how midin interacts with microtubules.

The experiments using various midin mutations delineate how MID1 underlies OS indicating the possible role of the coiled-coil domain in homo-dimerisation and the CTD in the association with the microtubule network. Future experiments in the laboratory involve using immuno-precipitation and yeast two-hybrid experiments to investigate the types of protein(s) that interact with midin. Results from such experiments could be used to determine which pathways are disrupted during embryogenesis and the consequences of the *MID1* mutations with regard to cellular responses.

Chapter Five:

Zebrafish MIPD Homologue

Zebrafish MID homologue

The zebrafish has emerged as an organism useful for studies in vertebrate development. Other species, such as *Drosophila melanogaster* and *Caenorhabditis elegans*, have been used successfully to obtain information about early developmental processes. For example, the twinfilin protein in *Drosophila* is required for actin-developmental processes (Wahlstrom *et al*, 2001), and several Rac proteins and alternative Rac regulators in *C. elegans* are responsible for controlling axon guidance, cell migration and apoptotic cell phagocytosis during development (Lundquist *et al*, 2001). However, there are limits to these models. For example, in the mouse, the developmental processes are difficult to study because the embryo is within the mother's uterus. Therefore, to harvest the mouse embryos the mother must be sacrificed. Vertebrates such as mouse and humans have a more complex genome and a more complex embryonic development. *Drosophila melanogaster* and *C. elegans* as models are unable to address some of the developmental processes in vertebrates such as the kidney, a multi-chambered heart, multi-lineage hematopoiesis, the notochord and neural crest cells (Weinstein *et al*, 1995; Dooley and Zon, 2000).

There are numerous advantages in using the zebrafish as a model to study early vertebrate development, including the production of large numbers of eggs, which are fertilised and developed externally. The generation time is relatively short, ~2-3 months, and all developmental stages are easy to access (i.e. there is no need to kill the mother in order to access the embryos) (reviewed in Kimmel, 1989; Driver *et al*, 1994; URL: <http://www.science.adelaide.edu.au/genetics/lardelli/zebrafishusefulcharacters>). Transparency of the embryo enables cells to be identified, and therefore, allows manipulation during experiments. Overall, these advantages mean that the zebrafish model is often quite useful in the study of human diseases (Weinstein *et al*, 1995; Zhang *et al*, 1998). Like most systems, there are also disadvantages. Firstly, it is difficult to generate zebrafish embryonic stem cells

for homologous recombination to specifically knockout a target gene (Detrich *et al*, 1999). Secondly, during the evolution of the zebrafish, there was a duplication of the genome following the divergence of fish and mammals. This has resulted in additional copies of genes that may complicate some studies, as the gene in mammals could be related to numerous homologues in the zebrafish (Detrich *et al*, 1999).

There are a number of ways the zebrafish model can be utilised to examine embryogenesis. These include analysis of gene and protein expression *in situ* and the disruption of development using either physical or chemical treatments (Detrich *et al*, 1999). Numerous large-scale genetic screens in zebrafish have generated over 500 mutant phenotypes in early development (Haffter *et al*, 1996; Driever *et al*, 1996). For example, mutations affecting the heart (Stainier *et al*, 1996), the nervous system (Woo and Fraser, 1995), retina (Malicki *et al*, 1996), craniofacial development (Neuhauss *et al*, 1996) and hematopoiesis (Amatruda and Zon, 1999) have been isolated. Thus, a large resource of information has been generated that can be utilised to analyse vertebrate organogenesis and human diseases. As an example, there is the *one-eyed pinhead* mutation in zebrafish that perturbs an EGF signalling pathway. This mutant phenotype is seen as holoprosencephaly in humans (Zhang *et al*, 1998).

To further investigate the function of *MIDI* during embryogenesis, the zebrafish (*Danio rerio*) was utilised as a model. The zebrafish model was chosen because of its short gestation time and the ease with which the embryos can be manipulated to create transgenic fish. The zebrafish homologue of *MIDI* (*zMID*) was identified and its expression pattern determined using *in situ* hybridisation. The aim was to use the zebrafish to produce a phenotype from the over-expression of *MIDI*. If a distinct phenotype was observed, then the equivalent *MIDI* mutations could be generated in the zebrafish homologue. The phenotype of the *MIDI* mutations could then be compared to the wild-type to provide a model for midin function. Thus, determining the function of *zMID* in zebrafish could be correlated to the abnormalities seen in the Opitz syndrome patients.

5.1 Zebrafish MID Homologue

The cDNA zebrafish libraries available were only at stages of early development when *MIDI* is not expressed (see Section 6.3.1). So to determine if there is a *MIDI* homologue in zebrafish, primers designed to the human sequence (HM-MID#1 and 2 primers) from the exon 1 region, were used in PCR (~500bp, Figure 5.1) on genomic zebrafish DNA. A PCR product was generated, cloned into pGEMT, and sequenced (see Section 2.10). From this zebrafish PCR product sequence, primers were designed (ZFMID#2and3) and used to generate more 3' sequence (see Table 2.4). The human primer HM-MID4, when paired with ZFMID#1, (see Table 1.3) was also utilised to identify a further ~500bp of the *zMID* sequence. Since there are 9 exons in the *MIDI* gene and large intron sequences, zebrafish cDNA was used as the template in the subsequent PCR reactions.

Numerous combinations of the human and mouse *MIDI* primers (see Table 2.3) were paired together in order to amplify further *zMID* sequence from 48 hours post-fertilisation (hpf). No other pairs of human and mouse *MIDI* primers produced a PCR product that corresponded to the *zMID* sequence. However, there have been a number of *MIDI* genes that have been isolated from various species (human, mouse, rat and, during the course of this project, chick (Appendix Five) and fugu (Appendix Six)). These *MIDI* nucleotide sequences were aligned and regions of high homology between the species were identified in order to design additional primers (see Table 2.4) that could be used to amplify the entire *zMID* sequence (Appendix Three and Four). The HF-MID#2 primer, paired with a zebrafish specific primer (ZF-MID#5) generated a further ~300bp fragment, producing a total of ~1.3kb of the *zMID* homologue. The exons/intron boundaries were identified in the *zMID* nucleotide sequence by computer comparisons (DNasis) with the *MID* homologues from other species. Exons 1 through to 7 were present in the *zMID* gene, however, only the first part of exon 7 has been isolated. The *zMID* predicted protein sequence contained the RING finger domain, B-

boxes, coiled-coil domain and the first part of the FNIII motif as seen in the human MID1 and MID2 proteins (Figure 5.1). Once enough zebrafish nucleotide sequence had been generated, specific primers were designed for use in PCR to ensure there was no contamination such as human *MIDI* amplification instead of the zebrafish sequences. There was no human contamination revealed in any of the PCR products.

Multiple attempts were made using the PCR technique to isolate the 3' region of the *zMID* gene (estimated at ~700bp). However, between the species known to have the *MIDI* gene, the 3' region is not as highly conserved as the 5' region. This hampered attempts to design primers within the 3' region that would be expected to produce a PCR product from zebrafish cDNA. Degenerate primers within this region were designed using the human, mouse, rat, chick and fugu sequences and used in PCR to try and amplify more *zMID* sequence. However, no observable PCR products were produced, again suggesting that maybe the *zMID* sequence is less conserved in the 3' region. RACE-PCR experiments (3' and 5' prime) were conducted in order to sequence the entire *zMID* gene. However, the *zMID* gene was not pulled out using this PCR technique.

Similar to the strategy adopted for *MIDI*, human primers from *MIDI2* were tried in PCR using zebrafish cDNA (48hpf) to see if *MIDI2* could be detected. No PCR product was produced. However, further evidence by Southern analysis (see below) was performed to conclude whether or not *MIDI2* was present in the zebrafish genome.

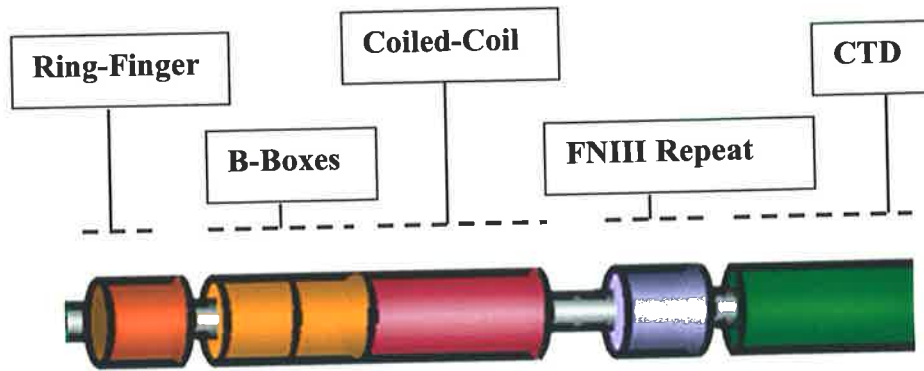


Figure 5.1: The *zMID* protein structure. Not all of the CTD was identified. The RING finger motif is highlighted in red, the B-boxes in orange, the coiled-coil domain in pink, and the FNIII domain in purple.

5.2 Zebrafish Genomic Southern Analysis

During the evolution of the zebrafish, their genome underwent duplication. Therefore, it is possible that there are multiple copies of the *zMID* gene. To determine if there was more than one *MID1* or a *MID2* gene present in the zebrafish genome, Southern analysis was conducted. A number of probes were made from the exon 1 of zebrafish and human genes. Exon 1 was amplified because a higher level of conservation was found in this region between the homologues (Figure 5.9). The zebrafish probe was amplified using ZFMID#2 and #3 primers with zebrafish genomic DNA (Figure 5.2; see Section 2.10). Human *MID1* was amplified with HM-MID1 and HM-MID2 primers, while human *MID2* amplification used 5'MID fusion and MID2#1 primers. The 5'MID fusion primer sequence was designed to amplify both the *MID1* and *MID2* genes. Since the same region from these three genes was amplified (exon 1) the PCR products used to make the probes were similar in size. Two zebrafish genomic Southern filters were probed with human *MID1* and *MID2*, stripped and re-probed with the zebrafish probe. Before probing with the *zMID*, the filters were exposed to film to ensure that the previous probe had been removed. The filters showed no radioactivity before they were re-probed.

When the zebrafish genomic Southern blot was probed with the human *MID1* and *MID2* genes, a similar banding pattern was observed (Figures 5.3). This suggests that the multiple copies of the *MID1* in zebrafish are similar enough to be identified using human *MID1* and *MID2* probes. However, the filter probed with *MID1* had some bands not seen on the filter probed with *MID2* (Figure 5.3). This indicates that there is at least one *MID1* homologue that cannot be identified using *MID2* probe, presumably because *zMID* is more closely related in sequence to *MID1* than *MID2*. There were more bands on the genomic southern than expected. This suggests that there could be multiple copies of the related *MID1* and *MID2* homologues in the zebrafish genome. Alternatively, the extra bands could have been due to restriction sites present in the intron regions. Since the genomic sequence of the zebrafish genes is not available, these possibilities have not been resolved.

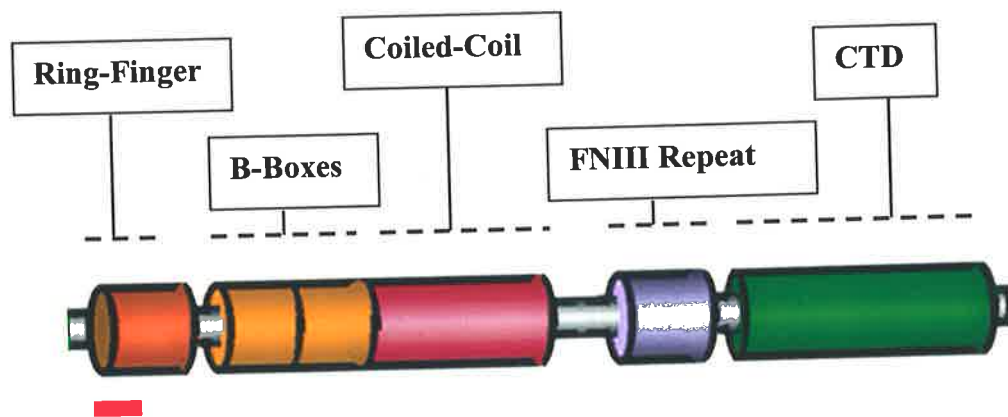


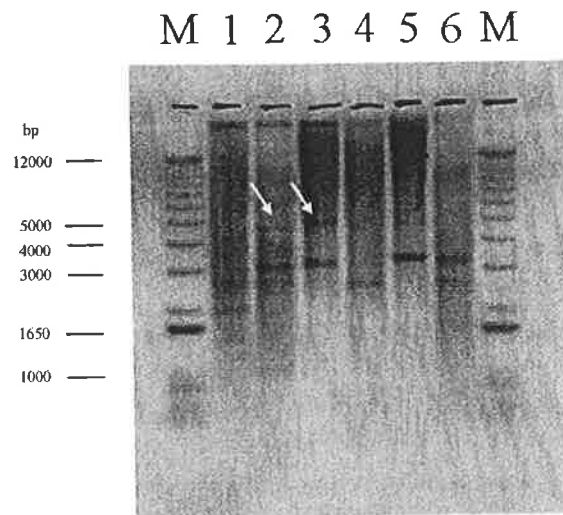
Figure 5.2: The position of the zMID probe used in Southern analysis. The probe is highlighted in red (~500bp). The position of the human *MID1* and *MID2* probes used for Southern analysis were at the same position as the zMID probe.

When the filters were re-probed with *zMID*, a different banding pattern was seen compared to the *MID1* and *MID2* probe, however the banding pattern was the same between these two filters. There were bands present that had not been observed when the filters were probed with the human *MID1* and *MID2*. In addition, all 3 probes detected the same bands

(Figure 5.3 and 5.4). This suggests that while in general these homologues in zebrafish have a high similarity, some of these homologues are less similar to each other.

It is difficult to conclude whether or not there are multiple copies of the *MID1* or *MID2* genes or if there is a family of *MID-like* genes present in the zebrafish genome.

A



B

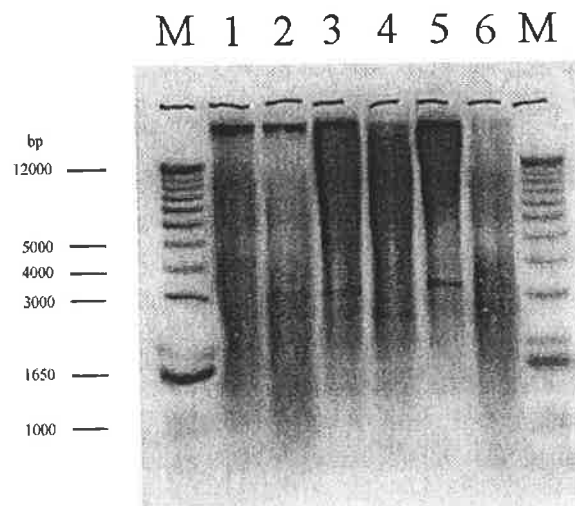


Figure 5.3: Zebrafish genomic Southern analysis using the human MID1 and MID2 probes. Lane M represents the marker lane. Lane #1 was the *PstI* digest, lane #2 the *PstI/EcoRI* digest, lane #3 the *EcoRI* digest, lane #4 the *BamHI/EcoRI* digest, lane #5 the *BamHI* and lane #6 the *PstI/BamHI* digest. MID1 (A) and MID2 (B) probes were hybridised at low stringency (42°C). The (A) and (B) diagrams show similar banding patterns. The white arrows in (A) represent the bands that were not observed in filter (B).

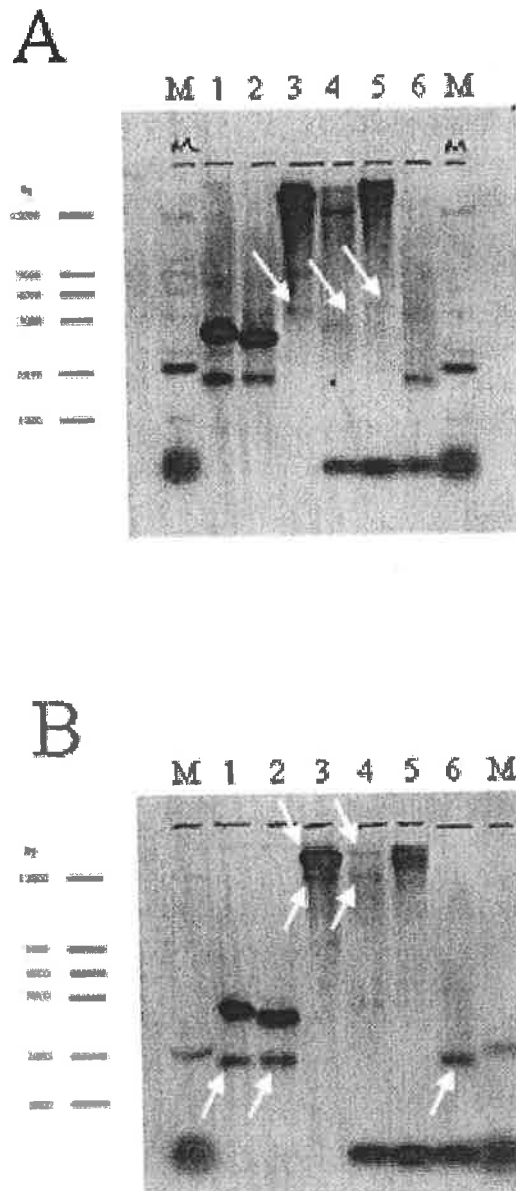


Figure 5.4: Zebrafish genomic Southern analysis using the zMID probe. Lane M represents the marker lane. Lane #1 was the *PstI* digest, lane #2 the *PstI/EcoRI* digest, lane #3 the *EcoRI* digest, lane #4 the *BamHI/EcoRI* digest, lane #5 the *BamHI* and lane #6 the *PstI/BamHI*

digest. The filters were probed with MID1 and MID2 (Figure 5.3) stripped and re-probed with zMID. (A) Represents the filter that had been previously probed with human MID1. (B) The filter that had previously been probed with human MID2. The white arrows in (A) represent bands that were observed when the filters were probed with MID1 and MID2. The white arrows in (B) represent bands that were not observed when the filters were probed with MID1 and MID2 (Figure 5.3).

5.3 zMID Expression During Embryogenesis

5.3.1 RT-PCR

Once the first ~500 bp of the *zMID1* homologue had been isolated, it was decided to use this sequence to determine when the gene was expressed during embryogenesis since a region of the exon could be amplified from cDNA. Zebrafish embryos were collected from various stages, 24hpf, 48hpf, 3 day, 5 day and adult. The cDNA was produced and PCR was used to see if the original 500bp could be amplified by RT-PCR (Figure 5.5). The zebrafish homologue, *zMID*, is not expressed at 24hpf, but it is from 48hpf through to adulthood, with the highest expression observed at 48hpf. In addition to these whole embryo stages, cDNA from the zebrafish adult head was used for PCR and it was shown that the MID1 homologue was expressed in this tissue (Figure 5.5).

The *notch5* gene was also amplified as a genomic control, to ensure that there was no genomic contamination in the cDNA. This gene was chosen because it is known to be expressed at the developmental stages selected. The primers were designed over an intron/exon region. Therefore, if there was amplification from genomic DNA, a larger PCR product (~800bp) would be produced compared to using the cDNA as the template (~200bp) (Figure 5.5). There was no genomic contamination observed in any of the different stages used in this experiment. The presence of gene expression shown using RT-PCR determined which of the embryonic stages were required for the *in situ* hybridisation experiments.

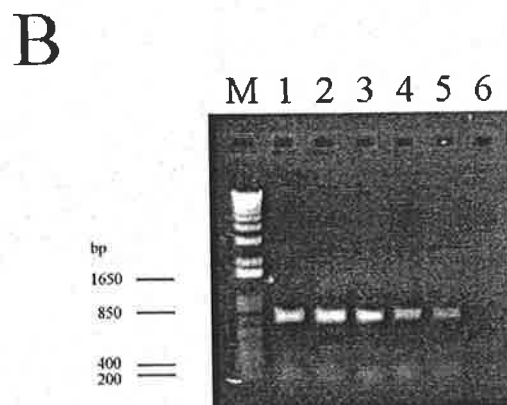
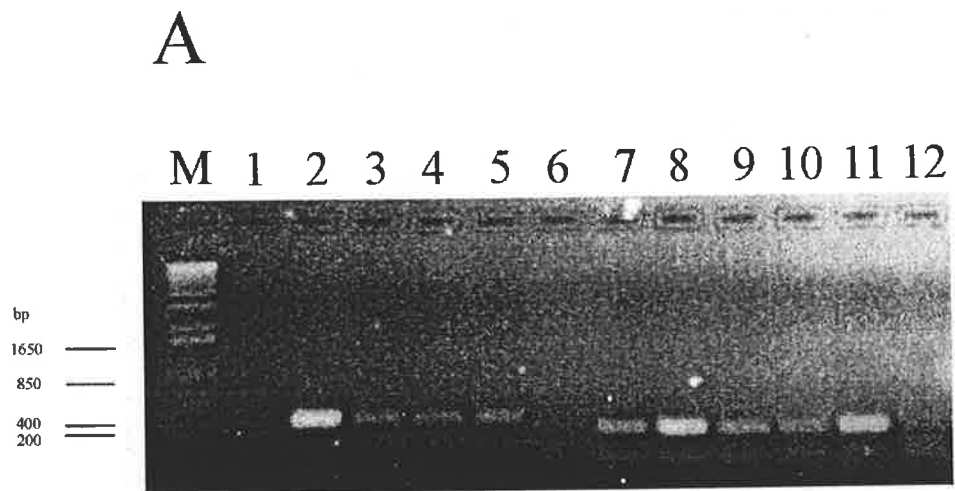


Figure 5.5: *zMID* homologue expression. **(A)** RT-PCR was carried out using cDNA from various stages of zebrafish development. Lanes #1-6 show amplification using the *zMID* primers, while lanes #7-12 are a control PCR using the *Notch5* primers. Lanes # 1 and 7 represent 24h embryos, lanes #2 and 8 the 48h embryos, #3 and 9 the 3day embryos, #4 and 10 the 5day embryos and #5 and 11 the adult head, while lanes # 6 and 12 are negative controls. The *zMID* does not appear to be expressed at 24h (lane #1). At 48h (lane #2) was the highest expression for *zMID* homologue. The expected PCR product was ~500bp. **(B)** Amplification of the *notch5* gene from genomic zebrafish DNA. The product produced was ~800bp, which was a larger sized product than when using cDNA (~200bp). This confirms that there was no genomic contamination present in any of the zebrafish cDNA stages used in this experiment. Lane #1 represents 24h embryos, #2 the 48h embryos, #3 the 3 day embryos,

#4 the 5 day embryos, #5 the adult stage and #6 was a negative control. M represents the marker lane, with the 1kb ladder marker.

5.3.2 Wholemount Zebrafish *in situ* Hybridisation

Initially *in situ* hybridisation was carried out using the ~500bp product of the *zMID* gene (Figure 5.1) generated by the HM-MID#1 and 2 primers. The minimum size for an effective expression pattern in zebrafish *in situ* hybridisation is ~400bp. However, a ~500bp probe used on whole 24hpf embryos through to 3 day old embryos did not produce a staining pattern. To ensure that there was, in fact, expression of the *zMID* gene, a larger product of ~800bp was generated and used as a probe (Figure 5.6). Zebrafish *in situ* hybridisation was carried out on whole 24h to 3 day old embryos to determine where the zebrafish *MID1* homologue was expressed during embryogenesis. The embryos were stained for 3 days before an observable expression pattern developed. The length of time before an expression pattern is seen depends on the level of expression of the candidate gene (Jowett, 1997). This suggests that *zMID* is expressed at a very low level from 48hpf through to adulthood. Other examples of genes which are expressed at low levels and require three days of staining to show an expression pattern are three zebrafish *spalt* like homologues, *spalt1a*, *spalt1b* and *spalt3* (E. Camp, personal communication).

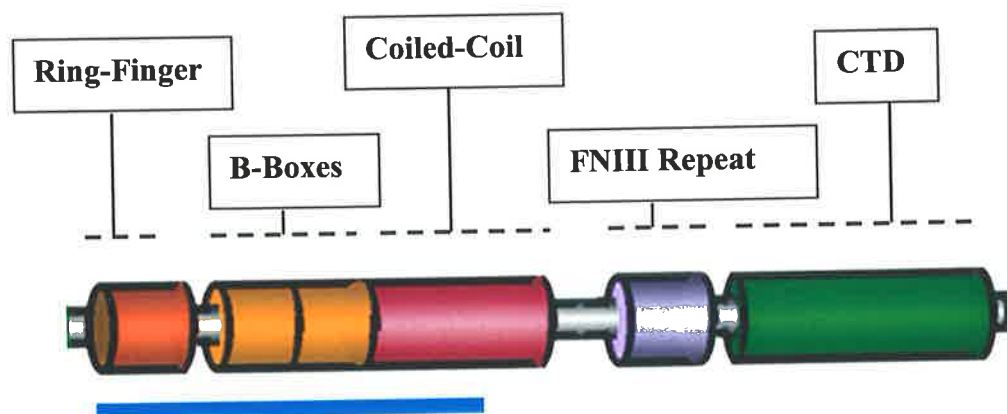
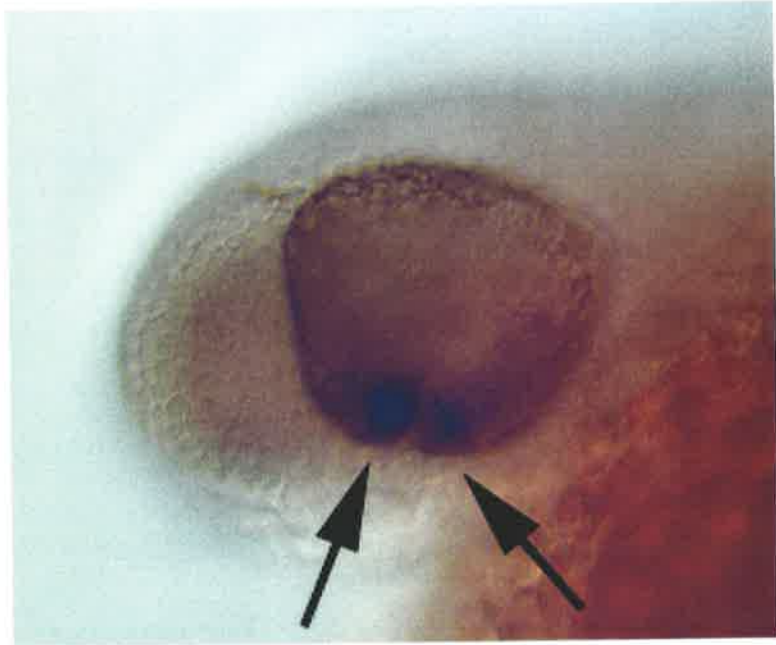


Figure 5.6: The position of the zMID probe used for *in situ* hybridisation. The probe is highlighted in blue (~800bp).

The zMID antisense probe showed that there was expression in the eye of the embryos (Figure 5.7). Expression of zMID was restricted and began around 30 days post-fertilisation (dpf). In the older embryos, ~48dpf-72hpf, the zMID staining was still present. However, viewing the stain was difficult due to the pigment in the eye at these stages. The zMID was expressed in the retina neuroepithelium cells of the developing embryo eye and marks the area in the eye surrounding the optic stalk. A sense probe was also used on the same stages of zebrafish embryos to act as a negative control (Figure 5.8). There was no background staining observed, indicating that the expression seen for the antisense probe was a positive result. The expression pattern observed here in zebrafish was unexpected since in mouse and human the MID1 expression is found in numerous tissues (Quaderi *et al*, 1997; Palmer *et al*, 1997; Dal Zotto *et al*, 1998; Perry *et al*, 1998). This suggests that there may be more than one MID1 homologue in zebrafish, as was also indicated from the Southern analysis (see Section 5.2).

A



B

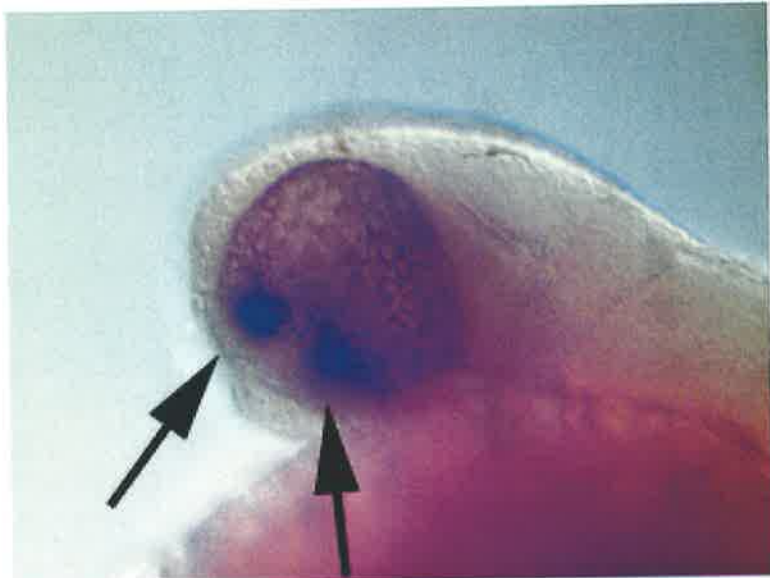


Figure 5.7: *In situ* hybridization experiments using the antisense zMID. The zMID was expressed in the eye of the zebrafish embryos. **A** and **B** images are different zebrafish embryos of ~30hpf. The arrows indicate where the zMID was expressed in the eye. It can be seen that the expression borders the area around the optic stalk. These images were taken at 200x.



Figure 5.8: Control *in situ* hybridization experiments using the sense zMID. There was clearly no staining in the embryos, suggesting there was no background. This image was taken on 100x.

5.4 Phylogenetic Analysis of the MID Genes

A phylogenetic analysis was completed using the various homologues of the *MID1* and *MID2* nucleotide sequences to determine their evolutionary relationships. In particular, a phylogenetic analysis will provide information on how the *zMID* homologue relates to the *MID1* and *MID2* genes isolated from various species. It was important to obtain such phylogenetic information because the Southern analysis (see Section 5.2) and *in situ* hybridisation experiments (see Section 5.3.2) strongly suggested that there were multiple copies of *MID* genes in the zebrafish genome.

Figure 5.9: Alignment of the protein sequences of the MID1 and MID2 homologues. The shading depends on the degree of similarity between the protein sequences. Black shading indicates a highest similarity while no shading shows lowest similarity. The amino acids shaded a dark or light gray represent residues conserved between most MID species identified to date, or residue positions showing greater variability across species.

```

*          20          *          40          *          60          *          80          *          100
cMID1      : METLESELTOPICLELFEDPLLLPCAHSLCFNCAHRILVSHCASN-EPVESTTAFQCPTCRYVITLSORGLEGLKRNVTLQNIIDRFOKASVSGPNSPSE : 99
fMID       : METLESELTOPICLELFEDPLLLPCAHSLCFNCAHRILVSHCTPS-EPIQSI SAFQCPTCRYVITLNORGLEGLKRNVTLQNIIDRYOKASVSGPNSPSE : 99
hMID2     : METLESELTOPICLELFEDPLLLPCAHSLCFNCAHRILVSSCSSG-ESIEPTTAFQCPTCRYVITLSLNHRGLDGLKRNVTLQNIIDRFOKASVSGPNSPSE : 99
hMID1     : METLESELTOPICLELFEDPLLLPCAHSLCFNCAHRILVSHCATN-ESVESTTAFQCPTCRYVITLSORGLDGLKRNVTLQNIIDRFOKASVSGPNSPSE : 99
mMID2     : METLESELTOPICLELFEDPLLLPCAHSLCFNCAHRILVSSCSSG-ESIEPTTAFQCPTCRYVITLSLNHRGLDGLKRNVTLQNIIDRFOKASVSGPNSPSE : 99
mMID1     : METLESELTOPICLELFEDPLLLPCAHSLCFNCAHRILVSHCATN-EPVESINAFQCPTCRYVITLSORGLDGLKRNVTLQNIIDRFOKASVSGPNSPSE : 99
spretusMID : METLESELTOPICLELFEDPLLLPCAHSLCFNCAHRILVSHCATN-EPVESINAFQCPTCRYVITLSORGLDGLKRNVTLQNIIDRFOKASVSGPNSPSE : 99
rMID1     : METLESELTOPICLELFEDPLLLPCAHSLCFNCAHRILVSHCATN-EPVESINAFQCPTCRYVITLSORGLDGLKRNVTLQNIIDRFOKASVSGPNSPSE : 99
zMID      : METLESELTRPICLELFEDPLLLPCAHSLCFNCAHRILVSHCAAFDESSKPPSAFQCPTCRYVITLDOGLDGLKRNVTLQNIIDRFOKASVSGPNSPSE : 100

```

```

*          120         *          140         *          160         *          180         *          200
cMID1      : TRRE-RAFDNSMS-CEKVLCOFCQDPAQDAVKTCTVCEVSYCECLKATHPNKKPFTGHRLTEPIPDSHIRGLMCLHEHEDEKVNMYCVTDDQLICAL : 197
fMID       : TRREGAVPDSRAMTSPGDRVPCQFCQDPPQDAVKTCTVCEVSYCECLKATHPNKKPFTGHRLMEPLDSHLRGLMCAEHEDEKVNMYCVTDEQLICAL : 199
hMID2     : SRRERTYRPTTAMSS--ERLACQFCQDPPRDAVKTCTVCEVSYCDRLRATHPNKKPFTSHRLMEPVDPDTHLRGITCLDHEHEKVNMYCVSDDQLICAL : 197
hMID1     : TRRE-RAFDANTMSS-AEKVLCOFCQDPAQDAVKTCTVCEVSYCECLKATHPNKKPFTGHRLTEPIPDSHIRGLMCLHEHEDEKVNMYCVTDDQLICAL : 197
mMID2     : SRRERTYRPTTAMSS--ERLACQFCQDPPRDAVKTCTVCEVSYCDRLRATHPNKKPFTSHRLMEPVDPDTHLRGITCLDHEHEKVNMYCVSDDQLICAL : 197
mMID1     : TRRE-RAFDANTMSS-AEKVLCOFCQDPAQDAVKTCTVCEVSYCECLKATHPNKKPFTGHRLTEPIPDSHIRGLMCLHEHEDEKVNMYCVTDDQLICAL : 197
spretusMID : TRRE-RAFDANTMSS-AEKVLCOFCQDPAQDAVKTCTVCEVSYCECLKATHPNKKPFTGHRLTEPIPDSHIRGLMCLHEHEDEKVNMYCVTDDQLICAL : 197
rMID1     : TRRE-RAFDANTMSS-AEKVLCOFCQDPAQDAVKTCTVCEVSYCECLKATHPNKKPFTGHRLTEPIPDSHIRGLMCLHEHEDEKVNMYCVTDDQLICAL : 197
zMID      : TRKQ--ISPSAMSSACEAVQCQFCQDPPQDAVKTCTVCEVSYCECLRATHPNKKPFTGHRLTEPVDPDARLRGLACPEHGEKVNMYCITDEQLICAL : 197

```

```

*          220         *          240         *          260         *          280         *          300
cMID1      : CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLAKLIQTQCOHVEVNASRQETKLMEECDQLIEIIQORRQIIGTKIKEGKVVRRLKLAQ : 297
fMID       : CKLVGRHRDHQVAALGDRYDKLKESLESNLTNLIKRNTSDESIMGKLIQTQCOHVEVNASRQENKLEECDLILINIIQORRQIITTKIKEGKAVRLRKLAAQ : 299
hMID2     : CKLVGRHRDHQVAALSNDRFKQKOTLEMNLTNLVKNRSELENQMAKLIQTQCOVEVNTAMHEAKLMEECDDELVEIIQORKOMIAVKIKETKVMKLRKLAQ : 297
hMID1     : CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLAKLIQTQCOHVEVNASRQEAKLTEECDLIEIIQORRQIIGTKIKEGKVVRRLKLAQ : 297
mMID2     : CKLVGRHRDHQVAALSNDRFKQKOTLEMNLTNLVKNRSELENQMAKLIQTQCOVEVNTAMHEAKLMEECDDELVEIIQORKOMIAVKIKETKVMKLRKLAQ : 297
mMID1     : CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLAKLIQTQCOHVEVNASRQEAKLTEECDLIEIIQORRQIIGTKIKEGKVVRRLKLAQ : 297
spretusMID : CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLAKLIQTQCOHVEVNASRQEAKLTEECDLIEIIQORRQIIGTKIKEGKVVRRLKLAQ : 297
rMID1     : CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLAKLIQTQCOHVEVNASRQEAKLTEECDLIEIIQORRQIIGTKIKEGKVVRRLKLAQ : 297
zMID      : CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLAKLIQTQCOHVEVNASRQEAKLTEECDLIEIIQORRQIIGTKIKEGKVVRRLKLAQ : 297

```

```

      *           320           *           340           *           360           *           380           *           400
cMID1 : QIANCKQCIERSSTSLISQAEQSLKENDHARFLOTAKNI TERVSMATASSQVLIPEINLNDTFDFALDFTREKKLLECLDYLTAPNPPPTIREELCTASMD : 397
fMID : QIAGCKQCIERSSSSLITQADQALKEADHTRFLOTAKSICERVSMATASSQVLLPEINLNDTFDFALDFTREKKMLENLDYLTAPNPPVIREELCTASMD : 399
hMID2 : QVANCROCLERSIVLINQAEHILKENDQARFLOSAKNIAERVAMATASSQVLIPEINLNDTFDFALDFSREKKLLECLDYLTAPNPPPTIREELCTASMD : 397
hMID1 : QIANCKQCIERSASLISQAEHSLKENDHARFLOTAKNI TERVSMATASSQVLIPEINLNDTFDFALDFSREKKLLECLDYLTAPNPPPTIREELCTASMD : 397
mMID2 : QVANCROCLERSIVLINQAEHILKENDQARFLOSAKNIAERVAMATASSQVLIPEINLNDTFDFALDFSREKKLLECLDYLTAPNPPPTIREELCTASMD : 397
mMID1 : QIANCKQCIERSASLISQAEHSLKENDHARFLOTAKNI TERVSMATASSQVLIPEINLNDTFDFALDFSREKKLLECLDYLTAPNPPPTIREELCTASMD : 397
spretusMID : QIANCKQCIERSASLISQAEHSLKENDHARFLOTAKNI TERVSMATASSQVLIPEINLNDTFDFALDFSREKKLLECLDYLTAPNPPPTIREELCTASMD : 397
rMID1 : QIANCKQCIERSASLISQAEHSLKENDHARFLOTAKNI TERVSMATASSQVLIPEINLNDTFDFALDFSREKKMLECLDYLTAPNPPPTIREELCTASMD : 397
zMID : QIANCKQEVVERSSSLITQADQALKETD HARFLOMAKSI SERVSMATASSQVLIPEINLNDTFDFALDFSREKKMLECLDYLTAPNPPPTIREELCTASMD : 397

```

```

      *           420           *           440           *           460           *           480           *           500
cMID1 : TITVHWTSDDEFSVVSYELOYTI FTGQANVV ----- SLNSADSWMIVPNIKONHYTVHGLQSGTRYIFIVKAIN : 467
fMID : TITVHWTSDDEFTVVSYELOYAIFTCQSNVVS ----- LNSADSWMIVPNIKONHYTVHGLQCGTRYIFIVKAIN : 469
hMID2 : TITVHWTSDDEFSISSYELOYTI FTGQANFISKSWCSWGLWPEIRKCKEAVSCSRLAGAPRGLYNSVDSWMI VPNIKONHYTVHGLQSGTRYIFIVKAIN : 497
hMID1 : TITVHWTSDDEFSVVSYELOYTI FTGQANVV ----- LNSADSWMIVPNIKONHYTVHGLQSGTRYIFIVKAIN : 467
mMID2 : TITVHWTSDDEFSISSYELOYTI FTGQANFIS ----- LNSADSWMIVPNIKONHYTVHGLQSGTRYIFIVKAIN : 467
mMID1 : TITVHWTSEDEFVVSYELOYTI FTGQANVVNVACDGTCLL ----- GSAGLNSADSWMIVPNIKONHYTVHGLQSGTRYIFIVKAIN : 480
spretusMID : TITVHWTSDDEFSVVSYELOYTI FTGQANVV ----- LNSADSWMIVPNIKONHYTVHGLQSGTRYIFIVKAIN : 467
rMID1 : TITVHWTSDDEFSVVSYELOYTI FTGQANVV ----- LNSADSWMIVPNIKONHYTVHGLQSGTRYIFIVKAIN : 467
zMID : TITVHWTSEDEFVVSYELOYTI FTGQANVV ----- LNSMDSWMI VPNIKO ----- : 445

```

```

      *           520           *           540           *           560           *           580           *           600
cMID1 : QAGSRSESEPCKLKTNSQPFLDPKSAHRKLVSHDNL TVERDETSKKSHSHPERFSTQGSYGVAGNVFIDSGRHYWEVVTISGSTWYAIGI SYKSAPKHEW : 567
fMID : QAGNRSESEPAKLVKNSQPFRDPKSAHRKLVSHDNL TVERDETSAKKSHSQDRFTSHSSYGVTCNVYIDSGRHYWEAL ICGSTWFAVGVAYKSAPRHEW : 569
hMID2 : QAGSRNSEPTRLKTNSQPFLDPKMTTHKKLKI SNDGLOMEKDESSLKKSHTPERFSGTCYGAAGNIFIDSGCHYWEVVMGSSSTWYAIGIAYKSAPKNEW : 597
hMID1 : QAGSRSESEPCKLKTNSQPFLDPKSAHRKLVSHDNL TVERDESSKKSHSHPERFSTQGSYGVAGNVFIDSGRHYWEVVTISGSTWYAIGLAYSAPKHEW : 567
mMID2 : QAGSRNSEPTRLKTNSQPFLDPKMTTHKKLKI SNDGLOMEKDESSLKKSHTPERFSGTCYGAAGNIFIDSGCHYWEVVMGSSSTWYAIGIAYKSAPKNEW : 567
mMID1 : QAGSRSESEPCKLKTNSQPFRDPKSAHRKLVSHDNL TVERDSSKKSHSHPERFAGQGSYGVAGNVFIDSGRHYWEVVTISGSTWYAIGLAYSAPKHEW : 580
spretusMID : QAGSRSESEPCKLKTNSQPFLDPKSAHRKLVSHDNL TVERDESSKKSHSHPERFSTQGSYGVAGNVFIDSGRHYWEVVTISGSTWYAIGLAYSAPKHEW : 567
rMID1 : QAGSRSESEPCKLKTNSQPFLDPKSAHRKLVSHDNL TVERDESSKKSHSHPERFSTQGSYGVAGNVFIDSGRHYWEVVTISGSTWYAIGLAYSAPKHEW : 567
zMID : ----- : -

```

```

          *      620      *      640      *      660      *      680      *      700
cMID1    : IGKNSASWVLCRCNNWVVRHNSKEIPTEPAPHLRRVGLLDYDNGSTAFYDALNSIHLHTFDVTFGQPVCPFTFWWVKCLTIITGLPIPDHLDSEOLA : 667
fMID     : VGKNSASWVLSRCNNSWVVRHNSKEMPEEPPHLRRLGILLDYDNGSLSFYDAVSSQHLHTFDVAFAHPVCPFTFWWNRCLTIITGLPIPDHLEGTDYNK : 669
hMID2    : IGKNASSWVFSRCNSWVVRHNNKEMLDVDPVPHLKRLGVLLDYDNNMLSFYDPANSEHLHTFDVTFILPVCPTFTIWNKSLMILSGLPAPDFIDYPERQE : 697
hMID1    : IGKNSASWALCRCNNWVVRHNSKEIPTEPAPHLRRVGLLDYDNGSTAFYDALNSIHLHTFDVTFILPVCPTFTIWNKSLMILSGLPAPDFIDYPERQE : 667
mMID2    : IGKNASSWVFSRCNSWVVRHNNKEMLDVDPVQLKRLGVLLDYDNNMLSFYDPANSEHLHTFDVTFILPVCPTFTIWNKSLMILSGLPAPDFIDYPERQE : 667
mMID1    : IGKNAASWALCRCHNHVVRHDGKETPTAPAPHLRRVGLLDYDNGSTAFYDALNSVHHLHTHAALACPVCPTFTIWNKCLTIITGLPIPDHLDCTEORP : 680
spretusMID : IGKNSASWALCRCNNWVVRHNSKEIPTEPAPHLRRVGLLDYDNGSTAFYDALNSIHLHTFDVALACPVCPTFTIWNKCLTIITGLPIPDHLDCTEQLP : 667
rMID1    : IGKNSASWALCRCNNWVVRHNSKEIPTEPAPHPRRVGLLDYDNGSTAFYDALNSIHLHTFDVALACPVCPTFTIWNKCLTIITGLPIPDHLDCTEQLP : 667
zMID     : -----

```


Table 5.1: The percentages of identity between the aligned MID homologue protein sequences.

	zMID	fMID	hMID1	mMID1	rMID	cMID1	<i>Spretus</i>	hMID2	mMID2
zMID		78.8	81.0	78.5	80.8	81.2	80.2	71.1	72.9
fMID	78.8		82.6	78.1	82.1	82.9	82.3	71.8	70.7
hMID1	81.0	82.6		93.2	99.1	95.4	99.1	77.3	75.8
mMID1	78.5	78.1	93.2		93.5	90.3	93.8	78.1	72.5
rMID	80.8	82.1	99.1	93.5		95.1	99.4	76.6	75.0
cMID1	81.2	82.9	95.4	90.3	95.1		95.2	76.2	75.6
<i>Spretus</i>	77.2	82.3	99.1	93.8	99.4	95.2		99.1	75.0
hMID2	71.1	71.8	77.3	78.1	76.6	76.2	99.1		96.9
mMID2	72.9	70.7	75.8	72.5	75.0	75.6	75.0	96.9	

5.4.2 Rooted and Un-rooted Phylogenetic Trees

The *MID1* and *MID2* nucleotide sequences were compiled into appropriate databases and used to construct phylogenetic trees (see Section 2.6). To construct a rooted tree, the fMID1 homologue was chosen as the outgroup on the assumption that it was the most anciently diverged sequence. Maximum parsimony phylogenetic rooted and un-rooted trees were constructed (Figures 5.10 and 5.11). Maximum likelihood phylogenetic rooted and un-rooted trees were also constructed (data not shown). There was little difference between the maximum likelihood and parsimony phylogenetic rooted trees. The maximum likelihood phylogenetic trees were constructed without the bootstrapped values and using the default parameters of a simple evolutionary model that may not have been appropriate. Therefore, the maximum parsimony phylogenetic trees were chosen, as they would likely be more correct in the presentation of the data. In both types of trees, the *mMID2* and *hMID2* genes were grouped together, while the *cMID1* and *hMID1* were each placed in separate phylogenetic

groups. The bootstrapped values for the maximum parsimony trees were close to 100%, which indicates that the phylogenetic groupings of the *MID* genes are reliable.

Maximum likelihood rooted and unrooted phylogenetic trees are also constructed using the *MID* homologues (data not shown). The un-rooted phylogenetic tree shows the relationship without the identification of a root to the tree. Like the rooted trees, bootstrap values were high, suggesting the results of the different phylogenetic grouping were reliable.

There were no discrepancies in the results between un-rooted and rooted phylogenetic trees. As observed with the rooted phylogenetic trees, the *rMID1* gene was placed in different phylogenetic groups depending upon whether maximum likelihood or parsimony trees were constructed. The un-rooted maximum parsimony tree grouped together the *zMID* and *fMID1* genes and the *mMID2* and *hMID2* genes into another group. The remaining species (chick, human and *M. musculus*) were grouped separately. Similarly to the rooted tree, the *spretusMID1* and *rMID1* were placed in the same phylogenetic group. In the un-rooted maximum likelihood tree, the *mMID1* and *spretusMID1* genes had been placed in the same phylogenetic group and the *rMID1* was placed in a separate phylogenetic group. This indicates that the phylogenetic grouping of the mouse *MID1* gene depends on the particular breeding of the mouse. The *M. spretus* mouse line is not as inbred as the *M. musculus* mouse line. In addition, between these two species the localisation of the *MID1* gene is different. In *M. spretus* the *MID1* gene is not localised in the PAR region, while in *M. musculus* the *MID1* gene is localised in the PAR region (Perry and Ashworth, 1999). The *rMID1* gene is also located entirely on the X chromosome. The chromosomal localisation of a gene can influence the different rates of the evolution of genes. Therefore, the differences seen here in the groupings of the mouse *MID1* genes could be a result of these genes evolving at different rates according to their chromosomal localisation (Perry and Ashworth, 1999).

Maximum Parsimony Tree

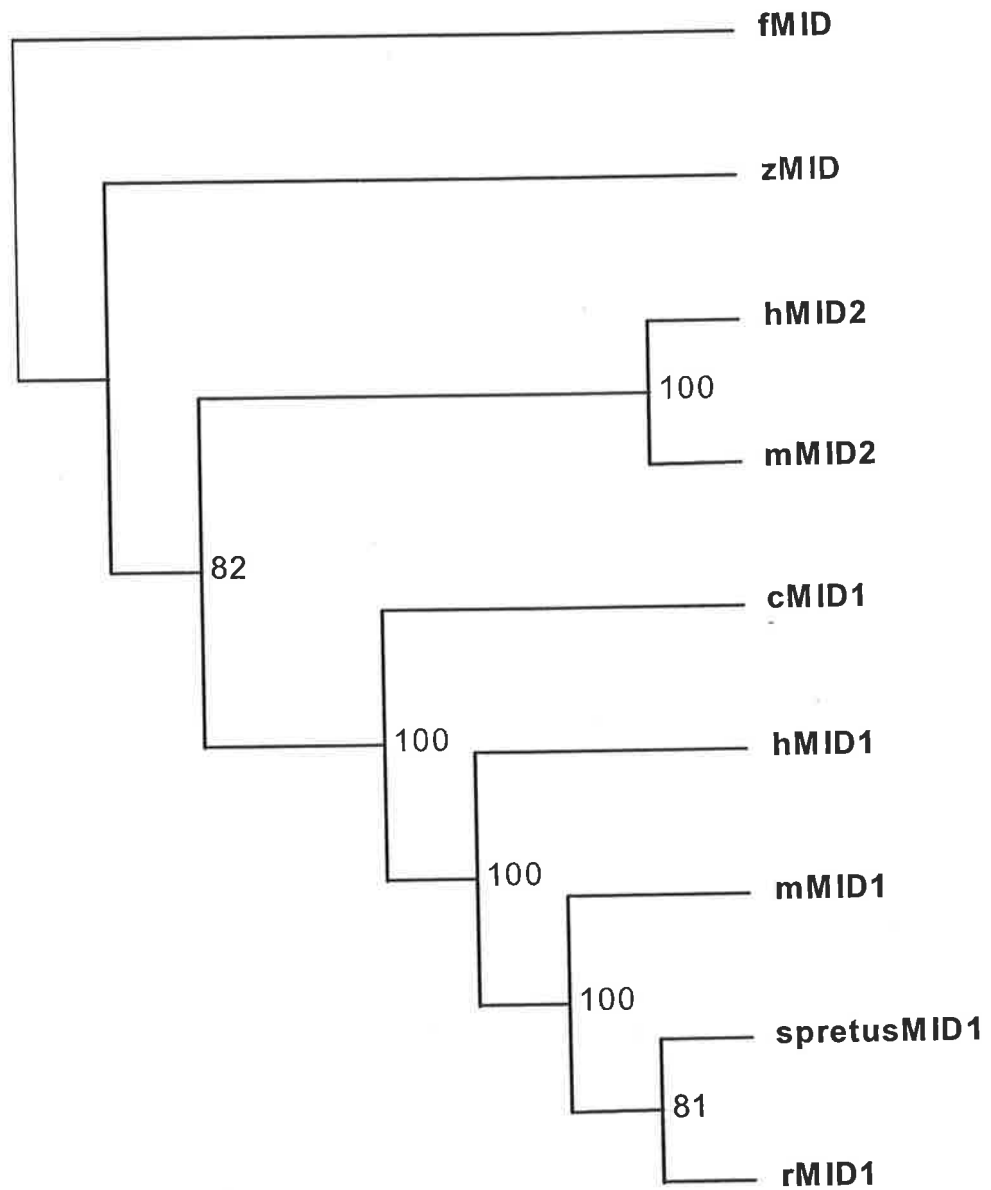


Figure 5.10: Maximum parsimony phylogenetic tree. The numbers on the phylogenetic tree represent the bootstrapped values.

Un-rooted Maximum Parsimony Tree

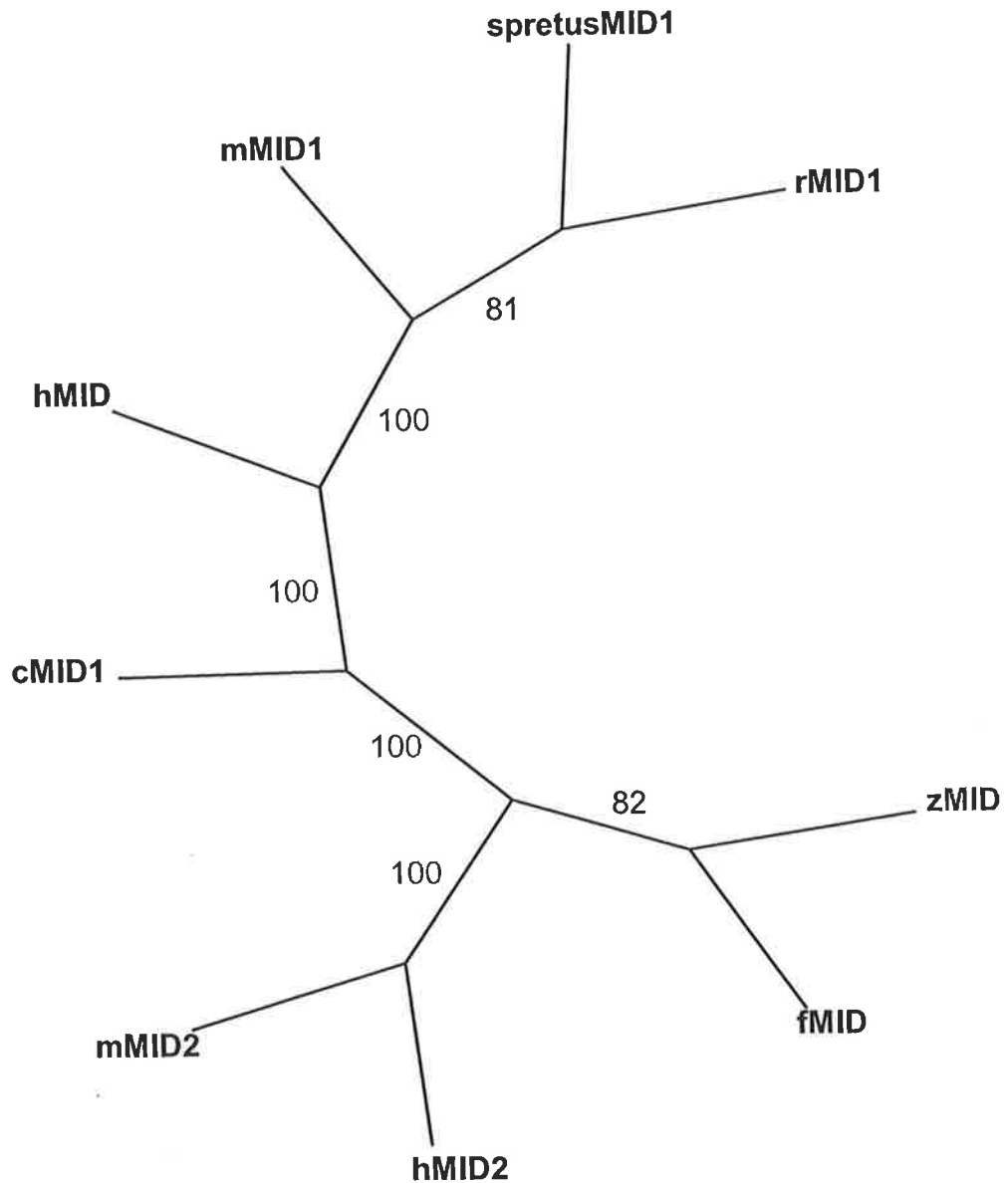


Figure 5.11: Un-rooted maximum parsimony phylogenetic tree of the *MID* homologues. The *rMID1* and *spretusMID1* genes have been placed in the same phylogenetic group and *mMID1* in a different group. The numbers on the phylogenetic tree represent the bootstrapped values.

6.4.3 Relative Rates Test

Relative rates tests were carried out to determine if the rates of nucleotide substitution were constant in different *MID* gene lineages (Wu and Li, 1985). The nucleotide sequences of all the *MID* homologues were used in these tests. The fugu *MID1* gene was again selected as the outgroup. The z-score has to be either $> +1.96$ or < -1.96 for there to be a significant difference in the rates of nucleotide substitution between two test sequences, 1 and 2, since they shared a common ancestor. A significant negative value for the z-score indicated that *MID* of the second sequence had evolved more rapidly than the first. For example, *hMID2* vs *mMID2* produced a z-score of -6.19 (Table 5.2). This implies that *mMID2* evolved more rapidly than *hMID1*.

A number of conclusions can be drawn from the relative rates tests. The *MID* homologues, *hMID1*, *spretus MID1*, *rMID1*, *zMID*, and *mMID2*, are evolving more rapidly than the *cMID1* gene (Table 5.2). The *hMID2* gene is evolving more rapidly than all the other *MID* homologues. Similarly, the *mMID2* is also evolving more rapidly than the *mMID1*, *spretusMID1*, *rMID1* and *zMID* homologues. These conclusions are based on the assumptions that fMID was an appropriate outgroup.

Table 5.2: Genes that evolve at significantly different rates are revealed by the relative rates tests.

Species (1)	vs	Species (2)	z-score
<i>cMID1</i>	vs	<i>hMID1</i>	-7.19
	vs	<i>mMID2</i>	-7.18
	vs	<i>spretusMID1</i>	-2.32
	vs	<i>rMID1</i>	-2.36
	vs	<i>zMID</i>	-2.33
<i>hMID2</i>	vs	<i>hMID1</i>	+6.30
	vs	<i>mMID1</i>	+6.26
	vs	<i>spretusMID1</i>	+5.63
	vs	<i>rMID1</i>	+5.49
	vs	<i>zMID1</i>	+3.16
	vs	<i>mMID2</i>	-6.19
<i>mMID2</i>	vs	<i>mMID1</i>	+6.23
	vs	<i>spretusMID1</i>	+5.53
	vs	<i>rMID1</i>	+5.39
	vs	<i>zMID1</i>	+3.04

5.5 Discussion

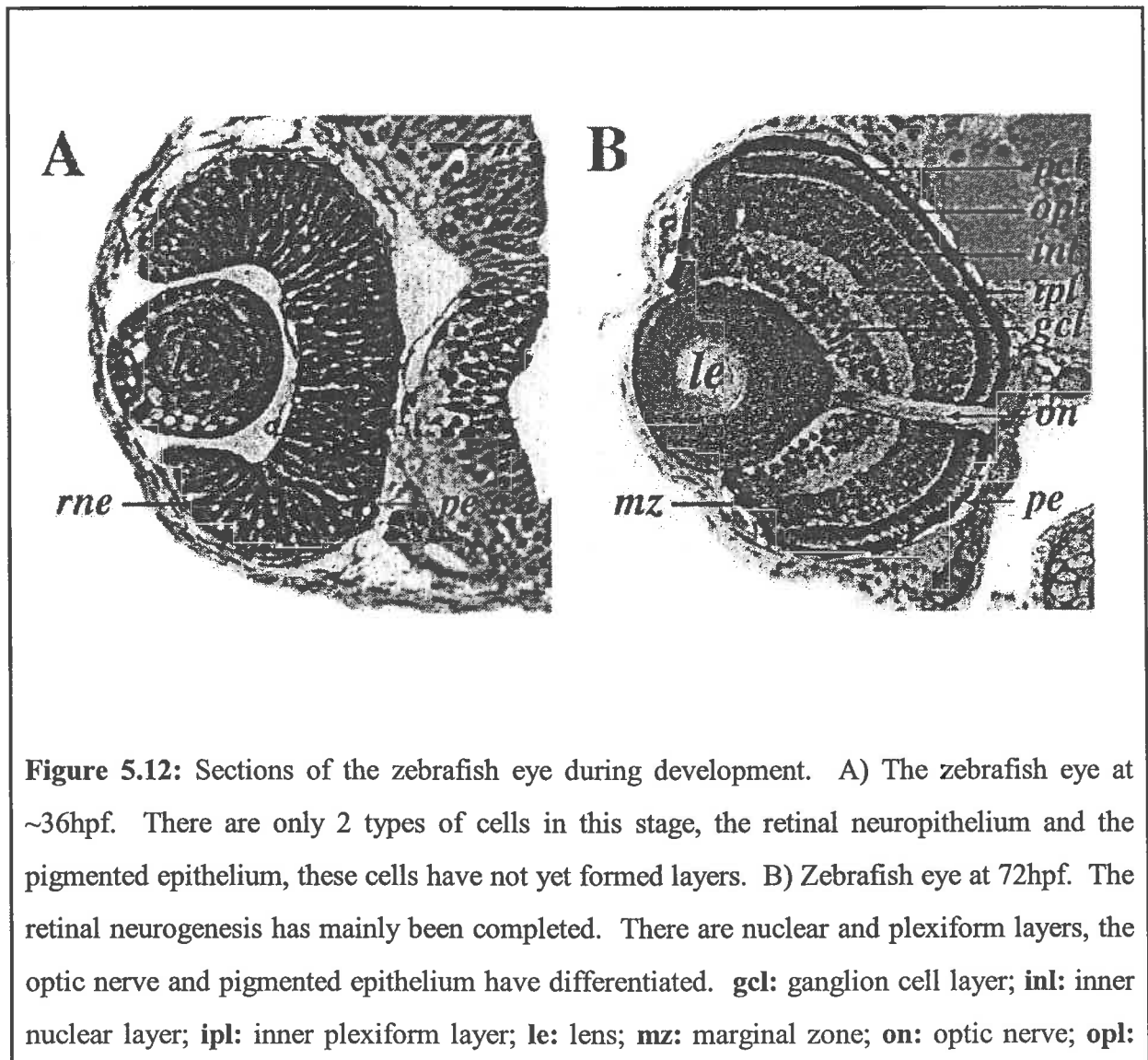
The zebrafish was chosen as a simpler vertebrate model in which to further investigate the function of the *zMID1* gene during embryogenesis. PCR confirmed that one or more *MID1* homologues are present in the zebrafish genome, *zMID*. So far ~1.3kb (of potentially 2kb) of *zMID* nucleotide sequence has been determined. Since the carboxy terminus is not as highly conserved, it has been difficult to obtain the remaining nucleotide sequence by using PCR. The exon/intron boundaries for exons 1-7 have been identified in the *zMID* nucleotide sequence. Numerous techniques were used (eg. RT-PCR), however, the remaining *zMID* nucleotide sequence was not identified and other techniques such as the screening of a

zebrafish cDNA library needed to be used. However, at the time of this study there was not an appropriate cDNA library (ie corresponding to maximum zMID expression) to be screened. Consequently, the remaining sequence was not able to be presented in the confines of this body of work and therefore the original aim of producing a zebrafish model of OS was not pursued.

The Ring finger, B-box, coiled-coil and part of the FNIII domains have also been identified. The *zMID* gene is found to be expressed at 48hpf, 3d, 5d, and adulthood. There was no expression seen at 24hpf, and the highest expression was observed at 48hpf. *In situ* hybridisation experiments showed that the *zMID* is expressed specifically in the eye. The expression of *zMID* observed by *in situ* hybridisation began to appear around ~30hpf. Due to the pigment in the zebrafish eye, it was difficult to see the expression clearly in older embryos (>48hpf). *zMID* is expressed in the retina neuroepithelium cells of the developing zebrafish eye, surrounding the position of the optic stalk. As the zebrafish eye develops, the optic nerve replaces the optic stalk at ~48hpf (Detrich *et al*, 1999). At 36hpf, the retina neuroepithelium cells have not formed layers, as this occurs later around 60hpf when the three nuclear and two plexiform layers become visible (Figure 5.12; Malicki *et al*, 1996; Detrich *et al*, 1999). The area of *zMID* expression was also the point where the first ganglion cell axons extend out from the eye into the brain. These axons extend along the optic stalk to the ventral area of the brain and then towards the midline at ~34-36hpf (Detrich *et al*, 1999). *zMID* can thus be used as a 'marker' to highlight the retina neuroepithelium cells in a double staining *in situ* hybridisation to determine if candidate genes are also expressed in the region where the first ganglion cell axons extend. This indicates that the *zMID* may be involved in the initial extension of the axons from the optic stalk region to the brain.

The expression pattern of the *zMID* gene is different to the observations with other *MID1* homologues such as mouse. As mentioned previously, the *MID1* gene is ubiquitously expressed in mouse embryos, except the adult mouse heart (Quaderi *et al*, 1997; Palmer *et al*,

1997; Dal Zotto *et al*, 1998; Perry *et al*, 1998). However, the *MID1* gene is highly expressed in the retina neuroepithelium cells, which is where *zMID* expression was observed (Dal Zotto *et al*, 1998). This difference in the expression pattern between these two species could be because the zebrafish has a simpler expression pattern compared to the mouse. Alternatively, zebrafish may have multiple copies of *MID1* and *zMID* may not be the equivalent homologue to *mMID1* (*Fxy1*). The latter is more likely to be the case.



outer plexiform layer; **pcl**: photoreceptor cell layer; **pe**: pigmented epithelium; **rne**: retinal neuroepithelium. This photograph was taken from Dietrich *et al*, 1999.

Indeed, the results from the Southern analysis indicate that there may be multiple copies of *MID1* homologues. However, a number of interpretations can be made. For instance, there may also be a *MID2* homologue present and with more than one copy. Alternatively, there may be other *MID-like* genes present in the zebrafish genome or the additional bands may have been due to restriction sites between the intron regions which would suggest that there was no multiple *MID* genes.

It would be expected that these multiple copies of the *MID1* homologues could 'act together' to produce an equivalent pattern of expression as seen with hMID1 and mMID1. Therefore, future experiments would involve screening the zebrafish genome for these additional copies. This would at least provide information on how many *MID* genes are present. In determining these additional *MID* expression patterns and undertaking phylogenetic analysis, the roles of these genes should be determined and the proteins they interact with (e.g by using Yeast two-hybrid technique). It should be noted though that some members of the MID family may be pseudogenes and not functional. However, *zMID* was isolated from cDNA, suggesting that this gene at least is transcribed during zebrafish embryogenesis. The expression pattern further indicated that *zMID* is functional.

If the localisation of *zMID* in the zebrafish genome was determined, then comparative mapping could be used to determine if *zMID* maps to a region syntenic to human and mouse. This would provide further information on how *zMID* was related to the other *MID* homologues. Alternatively, computer database searches will provide more information once the zebrafish genome has been completely sequenced.

The fact that there are multiple copies of *MID* in zebrafish, complicates using the zebrafish as a model to study the function of *MID1* during embryogenesis. However, this can

be overcome by creating transgenic zebrafish for individual copies of the *MID* genes, once they have been identified. This would also determine if the *MID* genes in zebrafish alter the zebrafish phenotype and if the phenotype produced mimics the OS phenotype. The homologues may or may not mimic the OS phenotype in zebrafish unless all of the zebrafish homologues are over-expressed. Alternatively, there may only be one homologue responsible for producing a phenotype since some *MID* copies in zebrafish may have evolved particular functions that are different to the human *MID1* homologue. Therefore, these *MID* copies would not be expected to produce a phenotype consistent with OS.

The protein alignment provided a visual analysis of the levels of similarity between the *MID* homologues (see Figure 5.9). This alignment shows that the *MID* homologues had been highly conserved. However, the carboxy terminus of the CTD is shown to have a lower level of conservation across species.

The relative rates tests indicated that the *MID* homologues have not evolved at the same rate (ie. the rate of nucleotide substitution differs significantly along different evolutionary branches). The test revealed that the human and mouse *MID2* genes have evolved more rapidly than the other *MID* homologues.

A phylogenetic analysis of the *MID1* and *MID2* homologues using maximum parsimony (Figure 5.10 and 5.11) determined their evolutionary relationships. The high bootstrapped values produced in the phylogenetic trees indicated that the data are likely to be correct. The rooted parsimony tree indicated that gene duplication occurred after *zMID*, giving rise to the *MID1* and *MID2* paralogues. *zMID* diverged from an ancient *MID* lineage before the *MID1* and *MID2* sequences (Figure 5.13).

Both maximum parsimony and maximum likelihood phylogenetic trees gave similar results. However, the maximum likelihood phylogenetic trees were not bootstrapped and used the default parameters plus a simple evolutionary model that may not have been valid.

Therefore, the maximum parsimony trees were considered to be more appropriate for the data provided.

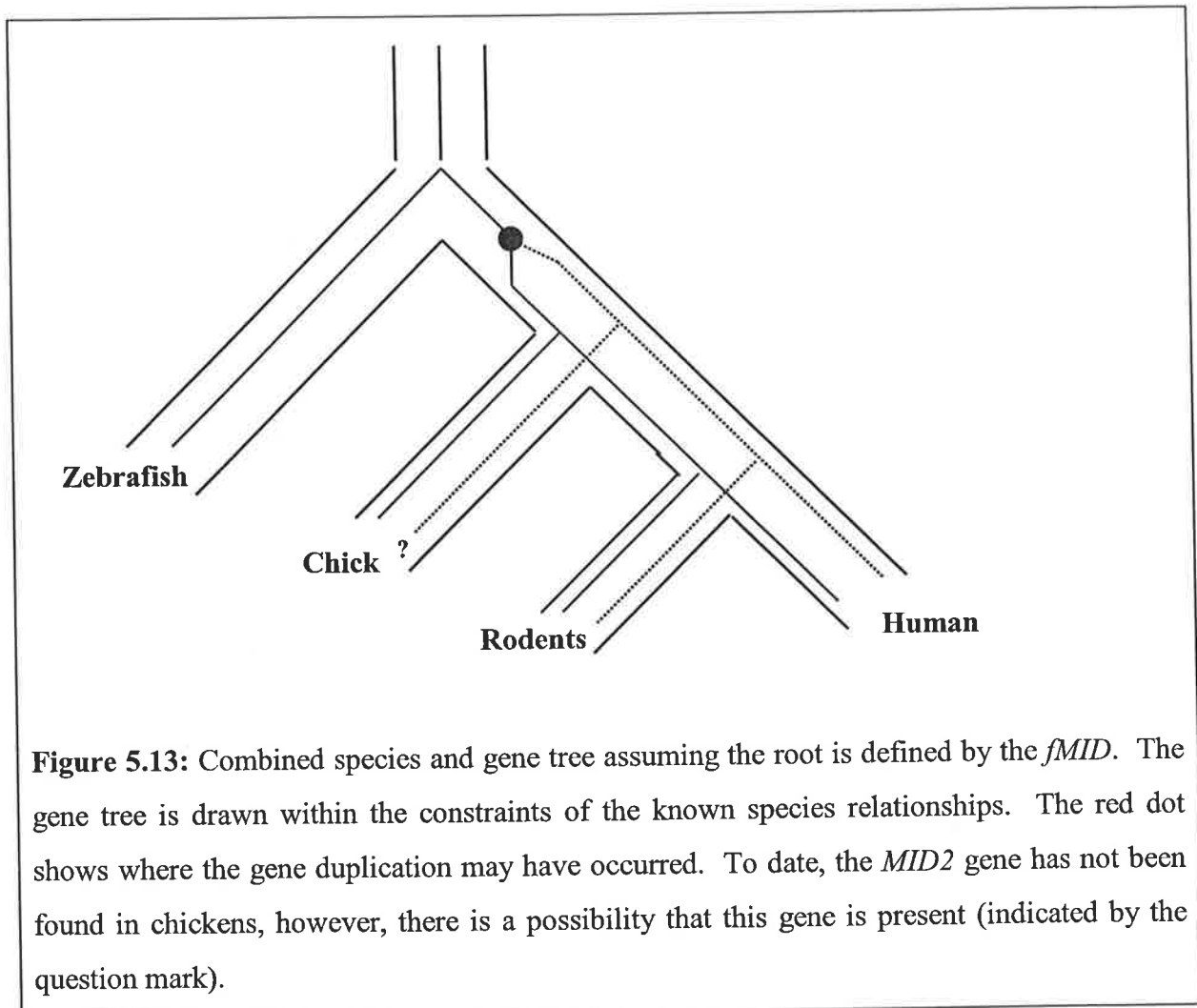


Figure 5.13: Combined species and gene tree assuming the root is defined by the *fMID*. The gene tree is drawn within the constraints of the known species relationships. The red dot shows where the gene duplication may have occurred. To date, the *MID2* gene has not been found in chickens, however, there is a possibility that this gene is present (indicated by the question mark).

Chapter Six:

Final Discussion

Final Discussion

To determine whether the *MID1* gene is underlying of the OS phenotype, a mutation screen was completed on the coding region of *MID1* as well as its *MID2* homologue. The location and regulation of *MID1* expression during embryogenesis and in a cell culture system was investigated by the intracellular localisation of the wild-type *MID1* protein (midin) and mutated forms of midin (found in this study), the generation of a *MID1* antibody and the examination of a zebrafish model system in which to create transgenic zebrafish.

6.1 The Role of *MID1* in the X-Linked Form of OS

In the screening of *MID1* for mutations, eight novel changes were identified from 16 OS probands (50%), five from 12 sporadic cases and three from 4 familial cases (Figure 6.2). Previous mutational screens of *MID1* in OS probands identified mutations in only 22.5% of 40 OS patients where 36% were familial cases and 6% were sporadic cases (Quaderi *et al*, 1997; Gaudenz *et al*, 1998) (Table 6.1). Overall, 77.5% of the probands in those studies did not have *MID1* mutations. This could suggest that the autosomal form of OS was predominant in their patients.

The variation between the mutation detection rate in the previous work and the results herein may be due to the method utilised for mutation analysis: direct sequencing opposed to SSCP analysis. Alternatively, the criteria for OS diagnosis may have differed. The 16 probands in this study had been brought to the clinicians attention because of their clinical severity. A comprehensive clinical description of the patients used in the study by Quaderi *et al*, 1997 and Gaudenz *et al*, 1998 was unavailable, preventing a direct comparison of the probands between these two independent studies.

The mutations identified in concurrent studies (Quaderi *et al*, 1997; Gaudenz *et al*, 1998; Schweiger *et al*, 1999; Cainarca *et al*, 1999) were mainly located in the C-terminal area

of the protein. The majority of the mutations resulted in the premature truncation of the protein. Similarly, in this study, the mutations were mainly identified in the C-terminal region, highlighting its importance for the normal function of the protein. However, there were mutations identified throughout the length of *MID1* showing that all motifs are required for the overall normal function of midin.

There is a possibility that alterations in the nucleotide sequence in *MID1* may have been missed as the sequencing was restricted to the coding regions of *MID1* (and similarly *MID2*) and mutations could reside in the introns or promoter region. For example, a mutation residing in a regulatory region of *MID1* could affect the level of midin. This may be the case for OSP#8, where genetic linkage analysis has shown that this family has the X-linked form of OS, as in this family no *MID1* ORF mutations were identified. If indeed the family of OSP#8 contains a *MID1* regulatory mutation, then other OS probands without ORF *MID1* mutations could also be in the same situation, making it difficult to know the true incidence of the X-linked and autosomal form of OS. Alternatively, there may be more than one gene responsible for the X-linked form of OS (e.g. *MID2* gene). Once the promoter region of *MID1* has been identified, those OS patients with X-linked inheritance and no mutations in the coding region (e.g. OSP#8) could have the *MID1* promoter region sequenced to find possible mutations. A mutation in the promoter region may or may not have a significant effect on the expression levels of *MID1*, depending upon whether *MID2* acts as a compensatory factor. For example, a mutation in the promoter of *MID1* may alter the expression levels significantly. However, *MID2* may then act as a compensatory factor making it difficult to predict the change in the *MID1* expression.

Anteverted nares and posterior pharyngeal clefting have only been reported in the X-linked form of OS (Robin *et al*, 1995 and 1996). In this study, the OS patients with confirmed X-linked inheritance were not always found to have these two features (see Table 3.1).





















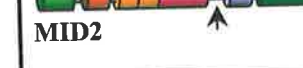
However, this may be due to the variability of the X-linked OS phenotype. In addition, the OS patients in this study have not been confirmed to have the autosomal form of OS. Therefore, this theory may hold true once more information becomes available on the distinguishing features of the autosomal form of OS versus the X-linked form.

6.2 The Role of MID2 in the X-Linked Form of OS

One missense variation, A358D, was found in the *MID2* sequence in three of the OS patients, OSP#7, 9 and 13 (Table 6.1). The incidence of the A358D change was less than one percent. This indicated that the A358D was more likely a mutation than a polymorphism. There was one proband, OSP#9, that also had a mutation in the *MID1* gene (1483C>T) as well as *MID2*. The proband had a variety of the OS characteristics. This suggests that, with the alteration in the *MID2* gene, there may be a reduced ability for compensation that results the presence of more OS characteristics.

DNA samples from family members for OSP#7 were available to screen for the *MID2* mutation. The half-brother of the proband had the *MID2* mutation, the mother was a carrier for the *MID2* mutation and the half-sister was normal. OS characteristics of the proband were mild, as was also the case for the half-brother and mother. There was no *MID1* mutation found in this family, indicating that a disruption in the *MID2* gene alone can affect the OS phenotype. However, it is important to note that the proband had a different father which could complicate the interpretation of the results. For instance, complications in the correct interpretation of results may arise if genetic linkage analysis was completed using the family members of OSP#7. If there are differences observed in how the genetic markers are inherited, these may be due to the proband having a different father. No DNA sample was available from the father of the proband.

Table 6.1: Summary of the *MID1* and *MID2* mutations and *MID1* polymorphisms found to date. Arrows show the position of the change.

Location of Change	Nucleotide Change	Amino Acid Change	Domains Affected	Reference
	343G>T	E115X	before B-box	This study
	Δexon 2	deletion of 1 st third of coiled-coil	coiled-coil	This study
	1051delC	frameshift	before FNIII	This study
	1102C>T	R368X	before FNIII	This study
	1330insA	frameshift	end of FNIII	This study and Cainarca et al, 1999
	1402C>T	Q468X	FNIII	This study
	1483C>T	R495X	CTD	This study
	1877T>C	L626P	end of CTD	This study
	796T>C	C266R	Coiled-coil	Gaudenz et al, 1998
	948delG	frameshift	Coiled-coil	Gaudenz et al, 1998
	IVS6-2A>G*	Skipping of final third of FNIII	Before FNIII	Gaudenz et al, 1998
	1312delATG	Deletion of M438	FNIII	Quaderi et al, 1997
	1483ins160bp	frameshift	CTD	Gaudenz et al, 1998
	1527ins 13bp	frameshift	CTD	Quaderi et al, 1997
	1558insG	frameshift	CTD	Quaderi et al, 1997
	1601ins24bp	Insertion of 8 amino acids	CTD	Gaudenz et al, 1998
	1607T>C	I536T	CTD	Gaudenz et al, 1998
	1800delCCTC	frameshift	CTD	Schweiger et al, 1999
	C1452T	No change	Polymorphism	Gaudenz et al, 1998
	A1884G	No change	Polymorphism	Gaudenz et al, 1998
	1073C>A	A358D	Linking region between FNIII and CTD.	This study

*This change needs to be confirmed experimentally.

Yeast two-hybrid and co-immunoprecipitation experiments have shown that there is hetero-dimerisation between MID1 and MID2 (Short *et al*, submitted). This suggests the possibility that MID2 may be a modifier of the X-linked phenotype. Therefore, it is possible that the A358D mutation could disrupt the ability of MID2 to dimerise with MID1 and thus modify the OS phenotype that is produced. Investigation of the intracellular localisation of the MID2 mutation A358D has revealed that there is no significant difference in the intracellular localisation compared with the wild-type MID2. The MID2 mutation could be further investigated using yeast two-hybrid and co-immunoprecipitation experiments to determine the affect the A358D alteration has on MID2 homo-dimerisation and hetero-dimerisation with MID1.

There have been two rare polymorphisms found in the coding region of the *MID1* gene of OS patients and one polymorphism identified in five control patients in the intron region before exon 3 (Gaudenz *et al*, 1998). The polymorphisms found in the OS patients did not alter the amino acid. This suggests that the MID2 A358D alteration is more likely to be a mutation than a polymorphism since an amino acid is changed in the coding region.

It is still unclear what the exact role of *MID2* may be in OS and, like *MID1*, there was no correlation between the genotype and phenotype of the OS patients. To date, the A358D is the only mutation found in the *MID2* gene and polymorphism have been identified. There are a number of human genetic disorders that have been mapped to the Xq22 region. These include, the Arts syndrome (OMIM 301835), X-linked megalocornea (OMIM 309300) and a locus for congenital deafness (OMIM 304500). It is possible that the *MID2* gene may also be

a positional candidate gene for these genetic syndromes. However, this seems less likely with the isolation of a *MID2* mutation in an OS patient identified in this study and the observed hetero-dimerisation of MID1 and MID2 (Short *et al*, submitted), which suggests that *MID2* has some role in OS.

MID2 could be further investigated by determining the regulation and protein-protein interactions of *MID2*. These novel *MID1* mutations and the novel *MID2* mutation could be used in yeast two-hybrid experiments to determine how the protein-protein interactions are altered by the mutations. This would help to provide information on the roles of the midin domains in the protein-protein interactions and the impact of the mutated forms on OS patients during embryogenesis.

A rare variant (deletion of a single A nucleotide) in a *MID2* intron was discovered in one of the control samples and confirmed not to be present in the OS patients. Given the location of the rare variant and its presence in a normal individual, it is unlikely to affect *MID2* function.

6.3 Effects of *MID1* Mutations

Four mutations (E115X, Δ exon2, R368X and L626P) were chosen for further investigation based on the position of the alteration in the *MID1* sequence. GFP experiments showed the intracellular localisation of mutant midin depended on the position of the mutation in the MID1 structure. If there was a mutation in the CTD region (and the coiled-coil domain remain unaffected by the mutation) the mutated form of midin was no longer able to associate with microtubules. This suggests that the mutations inability to associate with microtubules is a result in a loss of function with the microtubules which then produces the OS phenotype. MID1 has been shown to homo-dimerise (Cainarca *et al*, 1999). The loss of function of

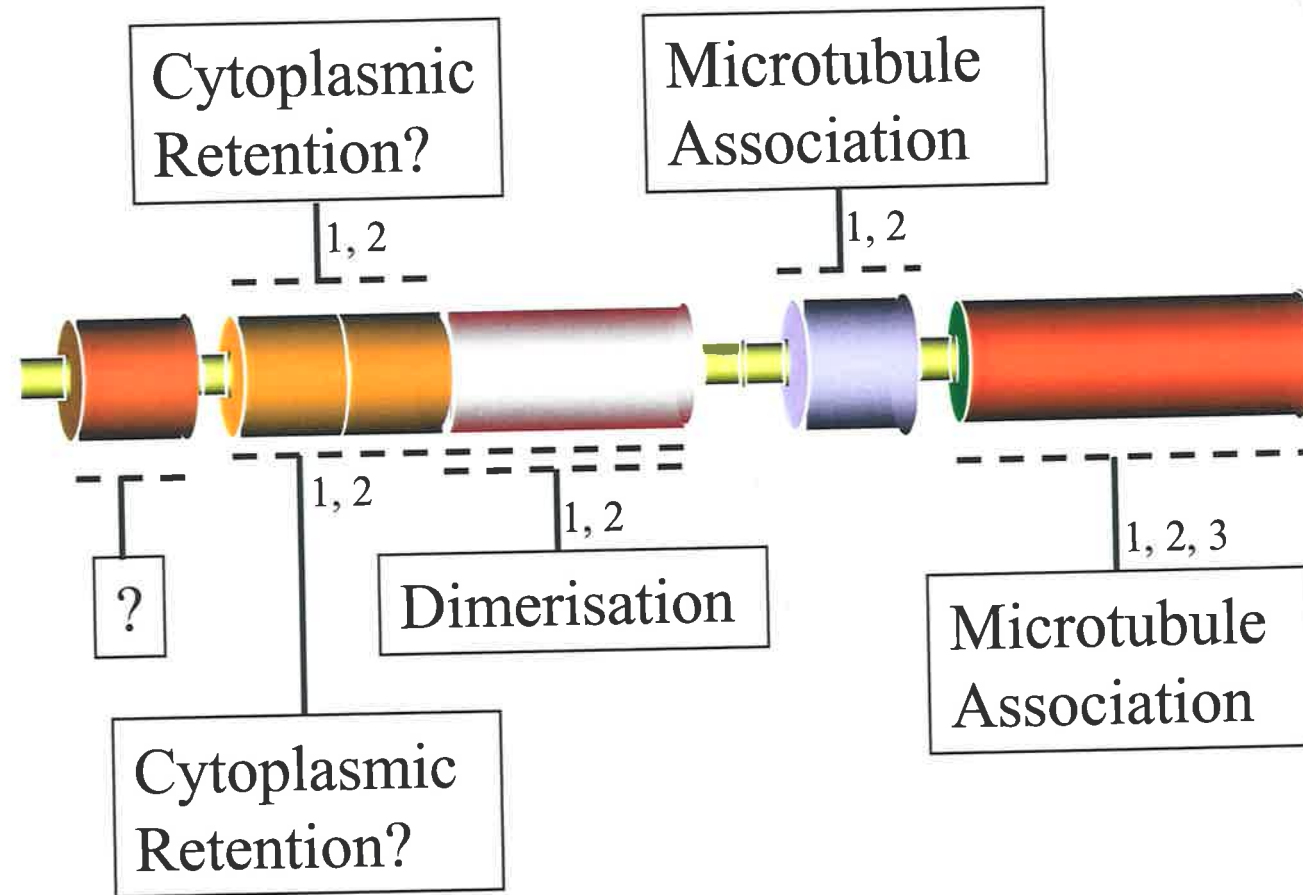
mutant midin to bind to microtubules could be caused by the inability for MID1 to homo-dimerise, suggesting that homo-dimerisation is necessary for microtubule binding.

From these results, it can be concluded that the coiled-coil motif may assist the CTD to associate with microtubules (Figure 6.1). The tripartite domain, in particular the B-Boxes, was required in its entirety to retain midin in the cytoplasm (Figure 6.1). In concurrent GFP experiments (Schweiger *et al*, 1999; Cainarca *et al*, 1999; Perry *et al*, 1999), the CTD was postulated to be responsible for the microtubule interaction. Results from the study herein imply that the CTD being responsible for the midin's ability to bind to the microtubule network (Figure 6.1), the coiled-coil domain may also be important for microtubule association. Yeast two-hybrid and co-immunoprecipitation experiments have shown that the coiled-coil domain could be responsible for the ability of midin to homo-dimerise, based on data from other RBCC proteins (Short *et al*, submitted).

Future experiments could involve further investigation of the midin protein structure and how the mutations affect the structure. Protein modelling computer programs could be used to determine the tertiary structure of midin as well as the mutated midin protein. Of particular interest would be the effect of the L626P missense on the tertiary structure of midin since it is a single amino acid change that results in cytoplasmic bundling similar to mutations that have had the FNIII and/or CTD regions deleted.

Figure 6.1: An overview of the possible functions of the midin domains. 1) This study (Cox *et al*, 2000), 2) Cox Laboratory, personal communication. 3) Gaudenz *et al*, 1998; Perry *et al*, 1999; Schweiger *et al*, 1999 and Cainarca, 1999.

Possible Functions of Midin Domains



6.4 Correlation Between Genotype and Phenotype

Analysis of the mutations and corresponding symptoms in OS probands did not show any correlation between genotype and phenotype. For example, the L626P was a single non-truncating amino acid change at the most carboxy region of the CTD. The phenotype of the proband was the most severe of all the probands available in the mutation screen. In addition, the mutation found in the proband OPS#6 could be identified over four generations of the family. However, OPS#6 family members carrying the mutation displayed varying severity of the OS phenotype. Therefore, the location and the type of mutations in the *MIDI* gene do not enable a prediction to be made of the severity of the OS phenotype that would be produced.

The fact that the *MIDI* mutations were varied in location and effect and that there was no correlation between genotype and phenotype indicates that the OS phenotype may be caused by an inability of mutated *MIDI* to interact with the microtubule network. The variability seen in the OS phenotype could be the result of the ability of other factors to compensate the midin function. Currently, experiments are underway to identify the *MIDI* interactors using yeast two-hybrid systems and their role in producing the OS phenotype. Once interactors have been identified, they can be investigated during embryogenesis and in a cell culture systems.

There are other genetic disorders that, like OS, do not have a correlation between the genotype and phenotype. For example, muscular dystrophy (Duchenne and Becker syndromes) is heterogeneous and the phenotype varies depending upon the type of mutation in the *dystrophin* gene. While muscular dystrophy (310200 OMIM) and OS are phenotypically different, they are similar in the fact that both disorders show no correlation between the extent of the deletion and the severity of the physical characteristics of the disease (Beggs *et al*, 1991; Cox, Allen *et al*, 2001). This suggests that the epigenetic and environmental factors may have a significant role in determining the severity of a disease, as seen in this study with

the OS patients. There may also be compensatory factors such as MID2 which may be responsible for the varying severity of the OS phenotype. The presence of genetic variations in the *MID2* gene could result in its ability to act as a compensatory factor, and consequently affect the severity of the OS phenotype. The latter is a more likely explanation given the evidence from this study and subsequent experiments (Short *et al*, submitted) that indicate there may be compensatory factors involved. However, epigenetic and environmental factors may also have an affect on the OS phenotype.

6.5 The Xp22 and Xq22 Regions

There are a number of different genes localised in the Xp22 region that have a closely related homologue localised to the Xq22 region. For instance, *MID1* is localised at Xp22 while the closely related *MID2* is localised at Xq22. Other examples include, phosphoribosyl pyrophosphate synthetase-2 (*PRPS2*) localised to Xp22, and its homologue *PRPS1* found at Xq22 (Becker *et al*, 1990), the nonreceptor tyrosine kinase gene (*BMX*) from Xp22 and the related Bruton's tyrosine kinase gene (*BMX*) from Xq21.3-q22 (Tamagnone *et al*, 1994; Vorechovsky *et al*, 1994) and the Duchenne muscular dystrophy gene (*DMD*) located at Xp21.2 and the related homologue dystrophin-related protein 2 (*DRP2*) at Xq22 (Koenig *et al*, 1987; Roberts *et al*, 1996). This suggests that there is a region of homology between these two regions of the X chromosome, Xp21 and Xp22, which may have arisen due to an intrachromosomal duplication between the long and short arms (Buchner *et al*, 1999; Perry *et al*, 1999). There are other examples of intrachromosomal and interchromosomal duplications along the X chromosome, such as between the Xq28 and 1611.1 regions (Lundin, 1993; Eichler *et al*, 1996) indicating that an intrachromosomal duplication could be feasible in the case of Xp21 and Xq22. Alternatively, these homologues may have arisen due to separate

gene duplications. However, due to the large number of different homologues, including the *MID1* and *MID2* genes, found within these regions, this theory would be unlikely.

There are a number of the above homologues (e.g. *MID1/MID2* and *Prps1/Prps2*) that have mouse orthologues present on the mouse X chromosome. This indicates that the duplication occurred before the divergence of the human and mouse lineages (Perry *et al*, 1999), a conclusion that is further supported by the phylogenetic analysis carried out in this study.

Phylogenetic analysis carried out in this study suggests that the *zMID* and *fMID* genes had evolved before the gene duplication event that led to the chick, rodent and human *MID1* and *MID2* genes. While results from the Southern analysis were not conclusive of whether or not a *MID2* homologue was present in the zebrafish, the results from the phylogenetic analysis indicate that *MID2* would not be present in zebrafish or fugu since the duplication event occurred after *MID* genes had evolved in these species. The isolation of *MID* homologues in other species would further assist in understanding the evolution of the *MID* homologues.

6.6 The Zebrafish OS Model

Mouse knockouts are currently being generated in the laboratory, but the zebrafish model was initially chosen as a quicker alternative to investigate the pathogenesis of OS. However, due to the difficulties outlined in this thesis this was not pursued. It was expected that the investigation using the zebrafish transgenesis would complement the long term studies using the mice. In cell lines, there is initial evidence of the over-expression of mutant midin binding to the normal midin, resulting in the normal midin no longer having the ability to bind to the microtubules, provided that the coiled-coil domain is intact in the mutant midin to enable homo-dimerisation (Short *et al*, submitted). Therefore, the over-expression of mutant midin in zebrafish should result in an OS phenotype, due to the homo-dimerisation between

the normal and mutant forms of the protein. This may also overcome the complexity of the presence of more than one *MID* homologue in the zebrafish. Since there is a high similarity between the *MIDI* homologues across species, another possibility is to over-express the human *MIDI* gene in the zebrafish to determine if a phenotype is produced. If no phenotype is produced, then the OS mutations found in the *MIDI* gene could then be over-expressed in zebrafish to determine if a phenotype is produced and if it mimics the OS phenotype observed in humans.

zMID is likely to be a functional gene because it has been isolated from cDNA and there is an expression pattern observed in the zebrafish embryos. However, the isolation of the remaining sequence from a cDNA library would further confirm that there are no stop codons in the coding region and that *zMID* is functional. Screening a zebrafish library should enable other homologues to be isolated. Homologues can be aligned at the nucleotide and amino acid sequence level, and expression patterns compared to the known *MIDI* and *MID2* homologues. This would provide information on how the *MID* homologues in zebrafish relate to their counterparts in other species.

In situ hybridisation experiments showed that the *zMID* has a specific expression pattern, only in the eye, from ~30days post-fertilisation. The *zMID* was found expressed in the retinal neuroepithelium surrounding the region that differentiated into the optic nerve. The optic nerve highlights the region where the first axons extend from the eye to the brain. Consequently, *zMID* could potentially be used as a marker to highlight this region before the optic nerve has differentiated. Alternatively, this expression pattern in the eye indicates that *zMID* may be involved in the development of the extension of axons from the eye to the brain during embryogenesis.

This expression pattern in zebrafish was very different to what was observed for *MIDI* in other species such as mouse and human where near ubiquitous expression of midin is

observed. Therefore, there is initial evidence indicating that the *zMID* may not be the equivalent *MID1* homologue. However, the chicken *MID1* gene is expressed in the retinal neuroepithelium of the chicken embryo (T.Cox, unpublished observation) and the *Fxy* is highly expressed in the retinal neuroepithelium of the mouse embryo (Dal Zotto *et al*, 1998). In light of this and the fact that there may be two *MID1* homologues in zebrafish, the *zMID* gene may be responsible for expression in the eye only, while the other unidentified duplicate maybe responsible for expression in the remaining tissues of the zebrafish embryo. Alternativley, the *zMID* gene may be expressed at low levels in other tissues during embryogenesis that was not detectable using *in situ* hybridisation.

The alignment of the *MID* homologues protein sequences showed that the carboxy terminus was not highly conserved, especially at the CTD. There was an 81% level of identity between the *zMID* and human *MID1* amino acid sequences. When the *zMID* and human *MID2* amino acid sequences were compared there was a lower identity of 67% which suggests that *zMID* was more closely related to *MID1*.

6.7 Conclusion

In the long term, the investigation of the *MID1* and *MID2* genes, and the production of a zebrafish model could help to provide information on understanding the pathogenesis of OS. With an improved understanding of the pathogenesis, OS could be better diagnosed. The dissection of the molecular pathology of OS by further investigating the function of the *MID1* and *MID2* genes, may eventually lead to prediction of the OS phenotype. By establishing the mechanisms underlying OS, then the embryological processes responsible for heart and craniofacial development may also be understood and related to other genetic syndromes.

References

- Allanson, J. E. (1988). G syndrome: an unusual family. *Am. J. Hum. Genet.* **31**, 637-642.
- Amatruda, J., and Zon, L. I. (1999). Dissecting hematopoiesis and disease using the zebrafish. *Dev. Biol.* **216**, 1-15.
- Arya, S., Viseskul, C., and Gilbert, E. F. (1980). The G syndrome – addition observations. *Am. J. Hum. Genet.* **5**, 321-324.
- Becker, M.A., Heidler, S.A., Bell, G.I., Seino, S., LeBeau, 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 localisation of PRPS1 and PRPS2 genes. *Genomics* **8**, 555-561.
- Bellini, M., Lacroix, J.-C., and Gall, J. G. (1995). A zinc-binding domain is required for targeting the maternal nuclear protein PwA33 to lampbrush chromosome loops. *J. Cell Biol.* **131**, 563-570.
- Borden, K. L. B., and Freemont, P. S. (1996). The RING finger domain: a recent example of a sequence-structure family. *Curr. Opin. Struct. Bio.* **6**, 395-401.
- Borden, K. L. B., Lally, J. M., Martin, S. R., O'Reilly, N. J., Etkin, L. D., and 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.
- Borden, K. L. B., Lally, J. M., Martin, S. R., O'Reilly, N. J., Solomon, E., and Freemont, P. S. (1996). *In vivo* and *in vitro* characterisation of the B1 and B2 zinc-binding domains from the acute promyelocytic leukemia protooncogene PML. *Proc. Natl. Acad. Sci. USA* **93**, 1601-1606.
- Brooks, J. K., Leonard, C. O., and Jr, P. J. C. (1992). Opitz (BBB/G) syndrome: oral manifestations. *Am. J. Med. Genet.* **43**, 595-601.

- Buchner, G., Montini, E., Andolfi, G., Quaderi, N., Cainarca, S., Messali, S., Bassi, M., Ballabio, A., Meroni, G., and Franco, B. (1999). MID2, a homologue of Opitz syndrome gene MID1: similarities in subcellular localisation and differences in expression during development. *Hum. Mol. Genet.* **8**, 1397-1407.
- Burn, J., Takao, A., Wilson, D., Cross, I., Momma, K., Wadey, R., Scambler, P., and Goodship, J. (1993). Conotruncal anomaly face syndrome is associated with a deletion within chromosome 22q11. *J. Med. Genet.* **30**, 822-824.
- Cainarca, S., Messali, S., Ballabio, A., and Meroni, G. (1999). Functional characterisation of the Opitz syndrome gene product (midin): evidence for homodimerisation and association with microtubules throughout the cell cycle. *Hum. Mol. Genet.* **8**, 1387-1396.
- Campbell, I.D., and Spitzfaden, C. (1994). Building proteins with fibronectin type III modules. *Structure* **15**, 333-337.
- Cao, T., Borden, K. L. B., Freeman, P. S., and Etkin, L. D. (1997). Involvement of the rfp tripartite motif in protein-protein interactions and subcellular distribution. *J Cell Science* **110**, 1563-1571.
- Cappa, M., Borrelli, P., Marini, R., and Neri, G. (1987). The opitz syndrome: a new designation for the clinically indistinguishable BBB and G syndrome. *Am. J. Med. Genet.* **28**, 303-309.
- Carlsson, I., 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.
- Chemke, J., Shor, E., Ankori-Cohen, H., and Kazuni, E. (1984). (Letter) Male to male transmission of the G syndrome. *Clin. Genet.* **26**, 164-167.

- Chisaka, O., and Capecchi, M. R. (1991). Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *hox-1.5*. *Nature* **350**, 473-479.
- Christian, J. C., Bixler, D., Blythe, S. C., and Merritt, A. D. (1969). Familial telecanthus with associated congenital anomalies. *Birth Defects Orig. Art. Ser.* **V(2)**, 82-85.
- Christodoulou, J., Bankier, A., and Loughnan, P. (1990). (Letter) Ring chromosome 22 karyotype in a patient with optiz (BBG) syndrome. *Am. J. Med. Genet.* **37**, 422-424.
- Coburn, T. P. (1970). G syndrome. *Am. J. Dis. Child.* **120**, 466.
- Cote, G. B., Katsantoni, A., Papadakou-Lagoyanni, S., Costalos, C., Timotheous, T., Skordalakis, A., Deligeorgis, D., and Pantelakis, S. (1981). The G syndrome of dysphagia, ocular hypertelorism and hypospadias. *Clin. Genet.* **19**, 473-478.
- Cordero, J. F., and Holmes, L. B. (1978). Phenotypic overlap of the BBB and G syndromes. *Am. J. Med. Genet.* **2**, 145-152.
- 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 Optiz Syndrome, *Hum. Mol. Gen.* **9**, 2553-2562.
- Dal Zotto, L. D., Quaderi, N. A., Elliott, R., Lingerfelter, P. A., Carrel, L., Valsecchi, V., Montini, E., Yen, C.-H., Chapman, V., Kalcheva, I., Arrigo, G., Zuffardi, O., Thomas, S., Willard, H. F., Ballabio, A., Disteche, C. M., and Rugarli, E. I. (1998). The mouse *Mid1* gene: implications for the pathogenesis of optiz syndrome and the evolution of the mammalian pseudoautosomal region. *Hum. Mol. Genet.* **7**, 489-499.
- Demczuk, S., Aledo, R., Zucman, J., Delattre, O., Desmaze, D., Dauphinot, L., Jalbert, P., Rouleau, G., Thomas, G., and Aurias, A. (1995). Cloning of a balanced translocation breakpoint in the DiGeorge syndrome critical region and isolation of a novel potential adhesion receptor gene in its vicinity. *Hum. Mol. Genet.* **4**, 551-558.

- Demczuk, S., Thomas, G., and Aurias, A. (1996). Isolation of a novel gene from the DiGeorge syndrome critical region with homology to *Drosophila gdl* and to human *LAMC1* genes. *Hum. Mol. Genet.* **5**, 633-638.
- DeMoerlooze, L. D., and Dickson, C. (1997). Skeletal disorders associated with fibroblast growth factor receptor mutations. *Curr. Opin. Geneti. Dev.* **7**, 378-385.
- Desai, A., and Mitchison, T. J. (1997). Microtubule Dynamics. *Annu. Rev. Cell Biol.* **13**, 83-117.
- Detrich, W. H., Westerfield, M., and Zon, L. I. (Editors) (1999). *Methods in cell biology: The zebrafish biology*. Vol 59, Academic Press San Diego.
- Dooley, K., and Zon, L. I. (2000). Zebrafish: a model system for the study of human disease. *Curr. Opin. Genet. Dev.* **10**, 252-256.
- Driver, W., Stemple, D., Schier, A., Solnica-Krezel, L. (1994). Zebrafish: genetic tools for studying vertebrate development. *Trends in Genet.* **10**, 152-159.
- Driever, W., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C., Malicki, J., Stemple, D. L., Stainier, D. Y., Zwartkruis, F., Abdelilah, S., rangini, Z., Belak, J., and Boggs, C, (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Dev.* **123**, 37-46.
- Driscoll, D. A., Budarf, M. L., and Emanuel, B. S. (1992a). A genetic etiology for DiGeorge syndrome: consistent deletions and microdeletions of 22q11. *Am. J. Hum. Genet.* **50**, 924-933.
- Driscoll, D. A., Spinner, N. B., Budarf, M. L., and McDonald-McGinn, D. M. (1992b). Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. *Am. J. Hum. Genet.* **49**, 261-268.

- Dundar, M., Erkilic, K., Demiryilmaz, F., Kucukaydin, M., Kendirci, M., Okur, H., and Kazez, A. (1995). Congenital alacrima in a patient with G (opitz frias) syndrome. *Hum. Genet.* **97**, 540-542.
- Eichler, E.E., Lu, F., Shen, Y., Antonacci, R., Jurecic, V., Doggett, N.A., Moyzis, R.K., Baldini, A., Gibbs, R.A., and Nelson, D.L. (1996). Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum. Mol. Genet.* **5**, 899-912.
- Einfeld, S. L., Fairley, M. J., Green, B. F., and Opitz, J. M. (1987). Brief clinical report: sudden death in childhood in a case of the G syndrome. *Am. J. Med. Genet.* **28**, 293-296.
- Ellison, J.W., Li, X., Francke, U., and Shapiro, L.J. (1996). Rapid evolution of human pseudoautosomal genes and their mouse homologs. *Mammalian Genome* **7**, 25-30.
- Emanuel, B.S., Buarf, B.S., and Scambler, P.J. (1998). The genetic basis of conotruncal heart defects: the chromosome 22q11.2 deletion. In: Rosenthal, N., and Harvey, R. (Ed). *Heart development*, pp463-478. Academic Press, San Diego.
- Farndon, P.A., and Donnai, D. (1983). Male to male transmission of the G syndrome. *Clin. Genet.* **24**, 446-448.
- Feldman, G.J., Ward, D.E., Lajeunie-Renier, E., Saavedra, D., Robin, N.H., Proud, V., Robb, L.J., Kaloustain, V.D., Carey, J.C., Jr, M.M. C., Cormier, V., Munnich, A., Zackai, E.H., Wilkie, A.O.M., Price, R.A., and Muenke, M. (1997). A novel phenotypic pattern in X-linked inheritance: craniofrontonasal syndrome maps to Xp22. *Human Molecular Genet.* **6**, 1937-1941.
- Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.5c, Department of Genetics, University of Washington, Seattle. Distributed by the author.
- Freemont, P.S. (1993). The RING finger: a novel protein sequence motif related to the zinc finger. *Ann. NY. Acad. Sci.* **684**, 174-192.

- Freemont, P.S., Hanson, I. M., and Trowsdale, J. (1991). A novel cysteine-rich sequence motif. *Cell*. **64**, 483-484.
- Frias, J.L., and Rosenbloom, A.L. (1975). Two new familial cases of the G syndrome. In Bergsma, D. *Malformation Syndromes*. New York, p54-57.
- The French FMF Consortium (1997). A candidate gene for familial Mediterranean fever. *Nat. Genet.* **17**, 25-31.
- Fryburg, J. S., Lin, K. Y., and Golden, W. L. (1996). Chromosome 22q11.2 deletion in a boy with opitz (G/BBB) syndrome. *Am. J. Med. Genet.* **62**, 274-275.
- Fryns, J. P., Delooz, J., and Van den Berghe, H. (1992). Posterior scalp defects in Opitz syndrome: another symptom related to a defect in midline development. *Clin. Genet.* **42**, 314-316.
- 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. *Am. J. Med. Genet.* **63**, 703-710.
- Goldmutz, E., Driscoll, D. A., Budarf, M. L., Zackai, E. H., McDonald-McGinn, D. M., Biegel, J. A., and Emanu, B. S. (1993). Microdeletions of chromosomal region 22q11 in patients with congenital conotruncal cardiac defects. *J. Med. Genet.* **30**, 807-812.
- Glover, T. W. (1995). CATCHing a break on 22. *Nature Genet.* **10**, 257-258.
- Gottlieb, S., Emanuel, B., Driscoll, D., Sellinger, B., Wang, Z., Roe, B., and Budarf, M. (1997). The DiGeorge syndrome minimal critical region contains a *Gooseoid*-like (GSCL) homeobox gene that is expressed early in human development. *Am J Hum Genet.* **60**, 1194-1201.

- Grignani, F., Fagioli, M., Alcalay, M., Longo, L., Pandolfi, P. P., Donti, E., Biondi, A., Lococo, F., Grignani, F., and Pelicci, P. G. (1994) Acute promyelocytic leukemia: from genetics to treatment. *Blood*, **83**, 10-25.
- Guion-Almeida, M. L., and Richieri-Costa, A. (1992). CNS midline anomalies in the opitz G/BBB syndrome: report on 12 brazilian patients. *Am. J. Med. Genet.* **43**, 918-928.
- Gundersen, G. G. and Tiffani, A. C. (1999). Micotubules and signal transduction. *Curr. Op. Cell Biol.* **11**, 81-94.
- Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J., Jiang, Y. J., Heisenberg, C. P., Kelsh, R. N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., and Nusslein-Volhard, C. (1996). The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Dev.* **123**, 1-36.
- Halford, S., Wadey, R., Roberts, C., Daw, S., Whiting, J., O'Donnell, H., Dunham, I., Bentley, D., Lindsay, E., Baldini, A., Francis, F., Lehrach, H., Williamson, R., Wilson, D., Goodship, J., Cross, I., Burn, J., and Scambler, P. (1993). Isolation of a putative transcriptional regulator factor from the region of 22q11 deleted in DiGeorge syndrome, Shprintzen syndrome and familial congenital heart disease. *Hum. Mol. Genet.* **2**, 2099-2107.
- 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.
- Henry, J., Ribouchon, M.-T., Offer, C., and Pontarotti, P. (1997). B30.2-like domain proteins: a growing family. *Biochemical and Biophysical Research Communications* **235**, 162-165.
- Hilton, D.J. (1999). Negative regulators of cytokine signal transduction. *Cell Mol Life Sci.* **55**, 1568-77.

- Hirokawa, N. (1994). Microtubule organisation and dynamics dependent on microtubule-associated proteins. *Curr. Op. Cell Biol.* **6**, 74-81.
- 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 Diag.* **9**, 783-793.
- Hunkapiller, T., and Hood, L. (1989). Diversity of the immunoglobulin gene superfamily. *Adv. Immunol.* **44**, 1-63.
- Inoue, S., Urano, T., Ogawa, S., Saito, T., Orimo, A., Hosoi, T., Ouchi, Y., and Muramatsu, M. (1999). Molecular cloning of rat efp: expression and regulation in primary osteoblasts. *Biochemical and Biophysical Research Communications* **261**, 412-418.
- 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. W., and Mather, I. H. (1990). Cloning and analysis of cDNA encoding bovine butyrophilin, and apical glycoprotein expressed in mammary tissue and secreted in association with the milk-fat globule membrane during lactation. *J Biol. Chem.* **265**, 14481-14486.
- John, S. W. M., Weitzner, G., Rozen, R., and Sriver, C. R. (1991). A rapid procedure for extracting genomic DNA from leukocytes. *Nucl Acids Res.* **19**, 408.
- Jowett, T. (1997). Tissue *in situ* hybridisation: methods in animal development. John Wiley and Sons Inc., NY.
- Jermiin, L. S. (1996). K2WuLi: a programme for estimating evolutionary distances and performing relative rates tests. © Lars S. Jermiin, The Australian National University. Distributed by the author.
- Kasner, J., Gilbert, F. F., Viseskul, C., Deacon, J., Herrmann, J. P. R., and Opitz, J. M. (1974). Studies of malformation syndromes VII. The G syndrome. *Z. Kinderheilk* **118**, 81-86.

- Kastner, P., Grondona, J. M., Mark, M., Gansmuller, A., LeMeur, M., Decimo, D., Vonesch, J.-L., Dolle, P., and Chambon, P. (1994). Genetic analysis of RXRa developmental function: convergence of RXR and RAR signaling pathways in heart and eye morphogenesis. *Cell* **78**, 987-1003.
- Kayserili, H., Cox, T.C., Cox, L.L., Basaran, S., Kilic, G., Ballabio, A., and Yuksel-Apak, M. (2001). Molecular characterisation of a new case of microphthalmia with linear skin defects (MLS). *J. Med. Genet.* **38**, 411-7.
- Kelly, D., Goldberg, R., Wilson, D., Lindsay, E., Carey, A., Goodship, J., Burn, J., Cross, I., Shprintzen, R. J., and Scambler, P. J. (1993). Confirmation that the velo-cardio-facial syndrome is associated with haplo-insufficiency of genes at chromosome 22q11. *Am. J. Med. Genet.* **45**, 308-312.
- Kimmel, C.B. (1989). Genetic and early development of zebrafish. *Trends in Genet.* **5**, 283-288.
- 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 organisation of the DMD gene in normal and affected individuals. *Cell* **50**, 509-517.
- Kurczynski, T. W., Micale, M. A., Assad, S., Gaba, C. G., and French, B. F. (1998). Opitz G/BBB syndrome in male child with an unbalanced der (3)t(3;14)(q29;q11.2)-14karyotype. *Am. J. Hum. Genet.* **63**, A110.
- Lacassie, Y., and Arriaza, M. I. (1996). (Letter) Opitz GBBB syndrome and the 22q11.2 deletion. *Am. J. Med. Genet.* **62**, 318.
- Lamour, V., Levy, N., Desmaze, C., Baud, V., Lecluse, Y., Delattre, O., Bernheim, A., Thomas, G., Aurias, A., and Lipinski, M. (1993). Isolation of cosmids and fetal brain cDNAs from the proximal long arm of human chromosome 22. *Hum. Mol. Genet.* **2**, 535-540.

- Lehninger, A.L., Nelson, D.L., and Cox, M.M. (1993). *Principles of Biochemistry*, 2nd Ed, Worth Pub, NY.
- 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. *Am. J. Med. Genet.* **40**, 173-176.
- Lewin, B. (1994). *Genes V*, Oxford University Press, Oxford.
- Lindsay, E. A., Botta, A., Jurecic, V., Carattini-Rivera, S., Cheah, Y-C., Rosenblatts, H. M., Bradley, A., and Baldini, A. (1999). Congenital heart disease in mice deficient for the Digeorge syndrome region. *Nature* **401**, 379-383.
- Lipson, A. H., Yuille, D., Angel, M., Thompson, P. G., Vandervoord, J. G., and Beckenham, E. J. (1991). Velocardiofacial (shprintzen) syndrome: an important syndrome for the dysmorphologist to recongnise. *J. Med. Genet.* **28**, 596-604.
- Lundin, L. G. (1993). Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* **16**, 1-19.
- Lundquist, E.A., Reddien, P.W., Hartweg, E., Horvitz, H.R., and Bargmann, C.I. (2001). Three *C.elegans* Rac proteins and several alternative Rac regulators control axon guidance, cell migration and apoptotic cell phagocytosis. *Deve.* **15**, 4475-4488.
- Maccioni, R. B., and Cambiazo, V. (1996). Role of microtubule-associated proteins in the control of microtubule assembly. *Physiol. Rev.* **75**, 835-864.
- MacDonald, M. R., Schaefer, G. B., Olney, A. H., Tamayo, M., and Frias, J. L. (1993). Brain magnetic resonance imaging findings in the opitz G/BBB syndrome: extension of the spectrum of midline brain anomalies. *Am. J. Med. Genet.* **46**, 706-711.

- May, M., Huston, S., Wilroy, R. S., and Schwartz, C. (1997). Linkage analysis in a family with the opitz GBBB syndrome refines the location of the gene in xp22 to a 4 cM region. *Am. J. Med. Genet.* **68**, 244-248.
- Malicki, J., Neuhauss, S. C. F., Schier, A. F., Solnica-Krezel, L., Stemple, D. L., Stainier, D. Y. R., Adbelilah, S., Zwartkruis, F., Rangini, Z., and Driever, W. (1996). Mutations affecting development of the zebrafish retina. *Devl.* **123**, 263-273.
- McDonald-McGinn, D. M., Driscoll, D. A., Bason, L., Christensen, K., Lynch, D., Sullivan, K., Canning, D., Zavod, W., Quinn, N., Rome, J., Paris, Y., Weinberg, P., Clark, B. J., Emanuel, B. S., and Zackai, E. H. (1995). Autosomal dominant opitz GBBB syndrome due to a 22q11.2 deletion. *Am. J. Med. Genet.* **59**, 103-113.
- McKusick, V. A., Francomano, C., and Antonarakis, S. E. (1995). Mendelian Inheritance in Man, 11th edn. Johns Hopkins University Press, Baltimore, MD.
- McPherson, E., Estop, A., and Paulus-Thomas, J. (1991). Craniofrontonasal dysplasia in a girl with der(x)(p22.2). *Am. J. Hum. Genet.* **49(s)**, A774.
- Muenke, M., Feldman, G. J., Ward, D. E., Lajeunie-Renier, E., Saavedra, D., Robin, N. H., Proud, V., Robb, L. J., Der Kaloustian, V., Carey, J. C., Cohen, M. M., Cormier, V., Munnich, A., Zackai, E. H., Wilkie, A. O. M., and Price, R. A. (1997). A novel phenotypic pattern in X-linked inheritance: craniofrontonasal syndrome maps Xp22. *Am. J. Hum. Genet.* **61(suppl.)**, A49.
- Muenke, M., and Schell, U. (1995). Fibroblast-growth-factor receptor mutations in human skeletal disorders. *TIG* **11**, 308-313.
- Neri, G., Genuardi, M., Natoli, G., Costa, P., and Maggioni, G. (1987). A girl with G syndrome and agenesis of the corpus callosum. *Am. J. Med. Genet.* **28**, 287-291.

- Neuhauss, S. C. F., Solnica-Krezel, L., Schier, A. F., Zwartkruis, F., Stemple, D. L., Malicki, J., Abdelilah, S., Stainier, D. Y. R., and Driever, W. (1996). Mutations affecting craniofacial development in zebrafish. *Devl*, **123**, 357-367.
- Nicholas, K. B., Nicholas, H. B. Jr., and Deerfield, D. W., (1997a). GeneDoc: analysis and visualisation of genetic variation. *EMBNEW NEWS* **4**, 14.
- Nicholas, K. B., and Nicholas, H. B. Jr., (1997b). GeneDoc: analysis and visualisation of genetic variation. <http://www.cris.com/~Ketchup/genedoc.shtml>
- Nicholson, S. E., Willson, T. A., Farley, A., Starr, R., Zhang, J. G., Baca, M., Alexander, W. S., Metcalf, D., Hilton, D. J., and Nicola, N. A. (1999). Mutational analyses of the SOCS proteins suggest a dual domain requirement but distinct mechanisms for inhibition of LIF and IL-6 signal transduction. *EMBO J.* **15**; 375-385.
- Opitz, J. M. (1987). G syndrome (hypertelorism with esophageal abnormality and hypospadias, or hypospadias-dysphagia, or "opitz-frias" or "optiz-G" syndrome)- perspective in 1987 and bibliography. *Am. J. Med. Genet.* **28**, 275-285.
- Opitz, J. M., and Gilbert, E. F. (1982). CNS anomalies and the midline as a "development field". *Am. J. Med. Genet.* **12**, 443-455.
- Opitz, J. M., Frias, J. L., Gutenberger, J. E., and Pellett, J. R. (1969a). The G syndrome of multiple congenital anomalies. *Birth Defects Orig. Art. Ser.* **2**, 95-101.
- Opitz, J. M., Summitt, R. L., and Smith, D. W. (1969b). The BBB syndrome: familial telecanthus with associated congenital anomalies. *Birth Defects Orig. Art. Ser.* **2**, 86-94.
- Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357-358.
- 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.

Parisian, S., and Toomey, K. E. (1978). Features of the G(opitz-Frias) and BBB (hypospadias hypertelorism) syndrome in the one family – are they a single disorder? (Abstract) *Am. J. Hum. Genet.* **30**, 62A.

Perry, J., and Ashworth, A. (1999). Evolutionary rate of a gene affected by chromosomal position. *Curr.Biol.* **9**, 987-989.

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.

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- β -sheet fibronectin type III module. *Proc. Natl. Acad. Sci. USA* **93**, 10702-10706.

Ponting, C., Schultz, J., and Bork, P. (1997). SPRY domains in ryanodine receptors (Ca^{2+} -release channels). *TIBS* **22**, 193-194.

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. M., 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. *Nature Geneti.* **17**, 285-291.

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

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.

Richman, J. M. (1995). Craniofacial genetics makes headway. *Current Biology* **5**, 345-348.

Roberts, R.G., Freeman, T.C., Kendall, E., Vetrie, D.L., Dixon, A.K., Shaw-Smith, C., Bone, Q., and Bobrow, M. (1996). Characterisation 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., Allanson, J. E., Guion-Almeida, M. L., Martin, R. A., Leichtman, L. G., Price, R. A., Opitz, J. M., and Muenke, M. (1995). Opitz syndrome is genetically heterogeneous with one locus on xp22, and a second locus on 22q11.2. *Nature Genet.* **11**, 460-461.

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. *Am. J. Med. Genet.* **62**, 305-317.

Robinson, B., and Hilger, P. (1989). Hereditary agenesis of nasal cartilage: surgical implications. *Arch.Otolaryng.* **115**, 985-988.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning: a laboratory manual* (2nd edition). Cold Spring Harbour Press, USA.

Saurin, A.J., Borden, K.L., Boody, M.N., and Freemont, P.S. (1996). Does this have a familiar RING? *Trends Biochem. Sci.* **21**, 208-214.

Scambler, P. J., Kelly, D., Lindsay, E., Williamson, R., Goldberg, R., Shprintzen, R., Wilson, D. I., Goodship, J. A., Cross, I. E., and Burn, J. (1992). Velo-cardio-facial syndrome associated with chromosome 22 deletions encompassing the DiGeorge locus. *Lancet* **330**, 1138-1139.

- Schrandner, J., Schrandner-Stumpel, C., Berg, J., and Frias, J. L. (1995). Opitz BBBG syndrome: new family with late-onset, serious complication. *Clin. Genet.* **48**, 76-79.
- 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.
- Seaver, L. H., Pierpont, J. W., Erickson, R. P., Donnerstein, R. L., and Cassidy, S. B. (1994). Pulmonary atresia associated with maternal 22q11.2 deletion: possible parent of origin effect in the conotruncal anomaly face syndrome. *J. Med. Genet.* **31**, 830-834.
- Sedano, H. O., and Gorlin, R. J. (1988). Opitz oculo-genital-laryngeal syndrome (Opitz BBB/G compound syndrome). *Am. J. Med. Genet.* **30**, 847-849.
- Seto, M. H., Liu, H-L. C., Zajchowski, D. A., and Whitlow, M. (1999). Protein fold-analysis of the B30.2-like domain *Proteins: structure, function and genetics* **35**, 235-249.
- Short, K.M., Hopwood, B., Yi, Z and Cox, T.C. (2001). MID1 and MID2 homo- and hetero-dimerise 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 disorder. *Submitted*.
- Shou, W., Li, X., Wu, C. F., T.Cao, Kuang, J., Che, S., and Etkin, L. D. (1996). Finely tuned regulation of cytoplasmic retention of *Xenopus* nuclear factor 7 by phosphorylation of individual threonine residues. *Mol. Cell Biol.* **16**, 990-997.

- 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.
- Stainier, D. Y. R., Fouquet, B., Chen, J-N., Warren, K. S., Weinstein, B. M., Meiler, S. E., Mohideen, M-A. P. K., Neuhauss, S. C. F., Solnica-Krezel, L., Schier, A. F., Zwartkruis, F., Stemple, D. L., Malicki, J., Driever, W., and Fishman, M. C. (1996). Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Devl.* **123**, 285-292.
- Stevens, C. A., and Wilroy, R. S. W. (1988). The telecanthus-hypospadias syndrome. *J. Med. Genet.* **25**, 536-542.
- Stoll, C., Geraudel, A., Berland, H., Roth, M.-P., and Dott, B. (1985). Male-to-male transmission of the hypertelorism-hypospadias (BBB) syndrome. *Am. J. Med. Genet.* **20**, 221-225.
- Takahashi, M., Inaguma, Y., Hiai, H., and Hirose, F. (1988). Developmentally regulated expression of a human "finger"-containing gene encoded by the 5' half of the *ret* transforming gene. *Mol. Cell Biol.* **8**, 1853-1856.
- 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/TXK family located in chromosome Xp22.2. *Oncogene* **9**, 3683-3688.
- Tolmie, J. L., Coutts, N., and I.K.Drainer. (1987). Congenital anal anomalies in two families with the opitz G syndrome. *J. Med. Genet.* **24**, 688-691.
- Thompson, J. D., Higgins, D. G., and Gibson, T. L. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weigh matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

- Ulitzur, N., Humbert, M., and Pfeffer, S. R. (1997). Mapmodulin: A possible modulator of the interaction of microtubule-associated proteins with microtubules. *Proc. Natl. Acad. Sci. USA*, **94**, 5084-5089.
- 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. *Am. J. Med. Genet.* **59**, 114-122.
- Van Biervliet, J. P. G. M., and Van Hemel, J. O. (1975). Familial occurrence of the G syndrome. *Clin. Genet.* **7**, 238-244.
- Van den Veyver, I.B., Cormier, T.A., Jurecic, V., Baldini, A., and Zoghbi, H.S. (1998) Characterisation and physical mapping in human and mouse of a novel RING finger gene in Xp22. *Genomics* **51**, 251-262.
- Verloes, A., David, A., Odent, S., Toutain, A., Andre, M.-J., Lucas, J., and Marec, B. L. (1995). Opitz GBBB syndrome: chromosomal evidence of an x-linked form. *Am. J. Med. Genet.* **59**, 123-128.
- Verloes, A., LeMerrer, M., and Briard, M.-L. (1989). BBBG syndrome or opitz syndrome: a new family. *Am. J. Med. Genet.* **34**, 313-316.
- 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.
- Wade, R. H., and Hyman, A. A. (1997). Microtubule structure and dynamics. *Curr. Op. Cell Biol.* **9**, 12-17.
- Wadey, R., Daw, S., Taylor, C., Atif, U., Kamath, S., Halford, S., O'Donnell, H., Wilson, D., Goodship, J., Burn, J., and Scambler, P. (1995). Isolation of a gene encoding an integral membrane protein from the vicinity of a balanced translocation breakpoint associated with DiGeorge syndrome. *Hum. Mol. Genet.* **4**, 1027-1033.

- Wahlstrom, G., Vartiainen, M., Yamamoto, L., Mattila, P.K., Lappalainen, P., and Heino, T.I. (2001). Twinfilin is required for actin-dependent developmental processes in *Drosophila*. *J. Cell Biol.* **26**, 787-796.
- Weber, J.L., and May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388-396.
- Webster, M. K., and Donoghue, D. J. (1997). FGFR activation in skeletal disorders: too much of a good thing. *TIG* **13**, 178-182.
- Weinstein, B. M., Stemple, D. L., Driever, W., and Fishman, M. C. (1995). Gridlock, a localised heritable vascular patterning defect in the zebrafish. *Nat. Med.* **1**, 1143-1147.
- Westerfield (1994). The zebrafish book. 2nd edition, Eugene, OR: Univ. Oregon Press.
- Williams, C. A., and Frias, J. L. (1987). Brief clinical report: apparent G syndrome presenting as neck and upper limb dystonia and severe gastroesophageal reflux. *Am. J. Med. Genet.* **28**, 297-302.
- Wilson, D. J., Cross, I. E., Wren, C., Scambler, P. J., Burn, J., and Goodship, J. (1994). Minimum prevalence of chromosome 22q11 deletions. *Am. J. Hum. Genet.* (Suppl) **55**, A975.
- Wilson, G. N., and Oliver, W. J. (1988). Further delineation of the G syndrome: a manageable genetic cause of infantile dysphagia. *J. Med. Genet.* **25**, 157-163.
- Wittwer, B., Kircheisen, R., Leutelt, J., and Gal, A. (1994). Gene of a new X-linked syndrome with multiple congenital anomalies and severe mental retardation maps to Xp22-pter. *Am. J. Hum. Genet.* **55(S)**, A1207.
- Woo, K., and Fraser, S. E. (1995). Order and coherence in the fate map of the zebrafish nervous system. *Dev.* **121**, 2595-2609.

Wu, C-I., and Li, W-H. (1985). Evidence for higher rates of nucleotide substitution in rodents that in man. *Proc. Natl. Acad. Sci. USA*. **82**, 1741-1745.

Zackai, E. H., McDonald-McGinn, D. M., Driscoll, D. A., Emanuel, B. S., Christensen. K. M., Chein, P., Mahboubi, S., Hubbard, A. M., Weinberg, P., and Clark, B. J. (1995). Respiratory symptoms may be the first presenting sign of a 22q11.2 deletion: A study of vascular rings. Platform presentation, 16th David W Smith, workshop on malformations and morphogenesis, Big Sky, MT.

Zhang, J., Talbot, W. S., and Schier, A. F. (1998). Positional cloning identifies zebrafish *one-eyed pinhead* as a permissive EGF-related ligand required during gastrulation. *Cell* **92**, 241-251.

Appendix One

Whole embryo (whole mount) in situ hybridisation (Quick and dirty in tubes)

Taken from the Internet from the site:

<http://www.science.adelaide.edu.au/genetics/lardelli/mawpt.html>

The web sit was constructed by:

Michael Lardelli

michael.lardelli@adelaide.edu.au

Richard Tamme

rtamme@genetics.adelaide.edu.au

Notes:

"Rinse" means resuspend the embryos in a solution and then immediately remove the solution.

"Wash" means incubate the embryos in the solution, with gentle agitation, for the specified time.

Before hybridisation begins, avoid RNAase contamination by wearing gloves, using untouched plasticware and using autoclaved (where practical) solution etc.. You do not need to worry about RNAase contamination when the embryos are in fixation or proteinase K solutions.

There is no need to use paraformaldehyde. Formaldehyde works just fine.

Use cut-off yellow or blue tips to transfer embryos when necessary. Flame a pair of scissors to cut off the tips!

Unless stated otherwise, operations are carried out at room temperature.

cRNA probe preparation

The cDNA for antisense RNA (cRNA) production must be cloned into a vector that provides an SP6, T3 or T7 promoter site that drives in vitro transcription of the clone in the antisense direction. We use the BlueScript II KS+ vector (Stratagene) for this purpose. This is especially advantageous since it contains T3 and T7 promoters flanking a multiple restriction site and these promoters are flanked by M13 primer binding sites (see below).

- 1) Clone your cDNA fragment into BlueScript such that you know the orientation of the insert.
- 2) To obtain linear, RNAase-free cDNA, amplify the insert in a 25 uL PCR reaction using M13 primers. To do this use 1-2 uL of a colony lysate or 1 uL of a 1 in 10,000 dilution of purified plasmid.
- 3) Electrophorese 5 uL of the PCR reaction on a gel beside DNA standards of known concentration. Estimate the concentration of the PCR product.

4) Mix the following reagents:

- 100-200 ng of amplified cDNA insert
- 0.5 uL of DIG labelling mix (see below*)
- 1 uL of 10X transcription buffer
- 1 uL of 20 u/uL T3 or T7 RNA polymerase (ie. for antisense transcription)
- dH₂O to a total reaction volume of 10 uL.

Note: This synthesis is one half the volume and uses one quarter of the DIG-11-UTP recommended by the manufacturer.

*Digoxigenin (DIG) labelling mix:

- 3.6 uL of 100 mM ATP
- 3.6 uL of 100 mM CTP
- 3.6 uL of 100 mM GTP
- 2.3 uL of 100 mM UTP
- 12.5 uL of 100 mM DIG-11-UTP
- 10.1 uL of dH₂O

35.7 uL TOTAL VOLUME

5) Incubate at 37°C for 2 hours.

6) Confirm cRNA synthesis by electrophoresing a 1 uL sample on a 1% agarose gel in TBE - 100V for approx. 30 min. (ie. Add 2 uL autoclaved 5X loading buffer and 7 uL autoclaved TE, pH 7.0 to the sample). Estimate the concentration of the cRNA by running DNA standards of known concentration on the gel. Hint: Wash the gel tank with water and SDS, use only autoclaved TBE and do not cover the gel with buffer (ie. fill the tank up so that the buffer touches the sides of the gel). Fill the wells with autoclaved TBE before loading the samples. This will minimise the risk of RNAase contamination breaking down the sample.

7) Store cRNA at -70°C.

Fixation, storage and preparation of embryos

1) Fix zebrafish embryos (with or without chorions) in PBS + 4% formaldehyde overnight at 4°C. (Embryos can be stored in PBS + 4% formaldehyde at 4°C for months.). To remove chorions after fixing, rinse and resuspend embryos in PBT (PBS + 0.1% Tween 20) in a dish (eg. sterile 10 cm petri dish) and use flamed watchmakers forceps - wear gloves and avoid dipping fingers in the PBT! Alternatively, dissect in PBT + 4% formaldehyde in a fume cupboard.

2) Rinse 1X and then wash 1X in 100% methanol (MeOH).

3) Place at -20°C for at least one hour.

4) Wash for 5 min.:

1X in 75% MeOH + 25% PBS then

1X in 50% MeOH + 50% PBS then

1X in 25% MeOH + 75% PBS then

4X in PBT.

5) Treat embryos with proteinase K in PBT (eg. 1 mL per tube of a 1:2000 diln. in PBT of a 20 mg/mL stock of proteinase K). Do not treat 0-4 hpf embryos. Treat 4-8 hpf embryos for 1 min., 8-16 hpf embryos for 3 min., 16-24 hpf embryos for 5 min., 48+ hpf embryos for 10 min.

6) Fix for 20 min. in PBS + 4% formaldehyde.

7) Wash 4 X 5 min. in PBT.

8) Transfer embryos to 1.5 mL microfuge tubes with locking caps (eg. Eppendorf "Safe-Lock" Reorder No. 0030 120.086). Rinse with 1 mL per tube of *prehybridisation solution then replace this with another mL of prehybridisation solution. (Caution, the formamide in prehybridisation is teratogenic - it can cause birth defects - so pregnant women should be especially careful in its presence. All such work should be performed in a well ventilated area or in a fume hood.)

*prehybridisation solution:

50% formamide (deionised)

5 X SSC

2% Blocking reagent (Boehringer Mannheim Cat. No. 1096176)

0.1% Tween-20

0.5% CHAPS (Sigma)

50 ug/mL yeast RNA

5 mM EDTA

50 ug/mL heparin

(Store as 50 mL aliquots at -20°C)

9) Incubate at least one hour at 70°C. The tubes should be rocked slowly and gently in a horizontal position. We place the tubes in a Grant BT1

Block Thermostat and then place this on its side on a rocking platform. After the prehybridisation the embryos may be stored for up to a month in

prehybridisation solution at -20°C. Note: Microfuge tubes without locking caps will pop open during the incubation!

Hybridisation and post-hybridisation washes

Note: embryos in prehybridisation solution can be very difficult to see. Take care not to suck them up and throw them out when removing solutions from the tube!

1) Place embryos in 1.5 mL microfuge tubes with locking caps with 1 mL of prehybridisation solution. Preheat to 70°C in the heating block.

2) Add 200-300 ng of probe to the top of the solution. Close the tube and invert rapidly a couple of times. The idea is to mix the concentrated probe into the prehybridisation solution before it can contact the embryos. Concentrated probe must not touch the embryos!

3) Hybridise overnight at 70°C with gentle rocking as previously.

4) Rinse embryos 3X with 1 mL of prewarmed prehybridisation solution (ie. it should be at least at room temperature when you add it to the embryos - not straight out of the freezer!). To do this remove almost all of the hybridisation solution from the tubes and place them upright in the heating block. As you add prehybridisation solution to the tubes, close the caps, invert the tubes a few times to thoroughly (but gently!) resuspend the embryos and then place them upright in the heating block at 70°C again for the embryos to settle to the bottom. Make sure the rinsing solution contacts all parts of the inside of the tube and remove as much as possible each time without leaving the embryos in the bottom of the tube uncovered.

5) Wash 2 X 15 min. in prehybridisation solution at 70°C.

6) Wash 1 X 30 min. in 50% prehyb. + 50% 2 X SSC at 70°C.

7) Rinse 1 X and wash 1 X 15 min. in 2 X SSC + 0.1% CHAPS at 70°C.

8) Rinse 1 X and wash 2 X 30 min. in 0.2 X SSC + 0.1% CHAPS at 70°C.

9) Rinse 1 X and wash 2 X 10 min. then 1 X 5 min. in PBT at room temperature.

10) Rinse 1 X and wash 2-3 hours in PBT + 1% BSA at 4°C.

Antibody application, washing and staining

1) Dilute anti-DIG Fab fragments conjugated to alkaline phosphatase 1:4000 in PBT + 1% BSA. Mix gently (ie. by inversion - avoid foaming!!) and then place at 4°C for at least 1 hour.

2) Remove PBT + 1% BSA solution from the embryos and replace with the diluted antibody. Rock very slowly and gently overnight at 4°C.

3) Rinse embryos 3 X wash 5 X 1 hour in PBT + 0.1% BSA at room temperature with gentle rocking.

4) Rinse 1 X and wash 3 X 5 min. in NTMT buffer at room temperature. Note: the NTMT buffer should be fresh (no more than one day old).

*NTMT buffer (50 mL):

100 mM NaCl
100 mM Tris-HCl (pH 9.5)
50 mM MgCl₂
0.1% Tween-20

Prepare this fresh from stocks of 1M Tris-HCl (pH 9.5) and 0.5M MgCl₂ + 1M NaCl.
(The Tris and [MgCl₂+NaCl] are separate stocks since they cannot be stored together at high concentration).

5) Replace the NTMT with 1 mL of Staining Solution* per tube.
*Staining Solution (per mL)

1 mL NTMT
3.4 uL of 100 mg/mL NBT
3.5 uL of 50 mg/mL BCIP

Place tubes on their side, cover with aluminium foil to protect from light and rock gently at room temperature for 30 min. Inspect the embryos
(look straight down the opened tube with a dissection microscope or remove into a glass staining dish to do this. Note!: NEVER place the staining solution in contact with POLYSTYRENE containers - black crystals will form). If only weak staining is evident then rock the embryos overnight at 4°C. If strong staining is evident then fix the embryos.

Note: Depending upon the number of embryos in each tube and on the level of expression of the gene being studied the staining solution may become exhausted. If the staining solution takes on a brownish tinge instead of appearing yellowish-green then replace it with fresh staining solution.

If you intend to section the embryos then "overstain" them until they appear rather dark. The apparent background staining will not be evident in the sections.

Post-stain fixation

- 1) Rinse 1X and wash embryos 1X in NTMT for 10 minutes.
- 2) Rinse 1X and wash 1X in PBT for 10 minutes.
- 3) Fix in PBT + 4% formaldehyde for 30 min. to overnight at 4°C

4) Remove fixative solution and replace with 1 mL 80% glycerol. Do not attempt to resuspend the embryos. Just let them rise to the surface of the glycerol. Cap the tube and place upright at 4°C. When the embryos have sunk to the bottom of the glycerol then they have equilibrated with the glycerol and can be examined microscopically. Be aware when handling the embryos that some formaldehyde is still present.

5) Store embryos at -20°C or -80°C until use.

Appendix Two: Alignment of the protein sequences of the MID homologues. The zebrafish protein sequence has been excluded.

```

*           20           *           40           *           60           *           80           *           100
cMID1      : METLESELTCPICLELFEDPLLLPCAHSLCFNCAHRILVSHCAIN-EPVESITAFQCPTCRYVITLSORGLEGLKRNVTLQNIIDRFQKASVSGPNSPSE : 99
fMID       : METLESELTCPICLELFEDPLLLPCAHSLCFNCAHRILVSHCTPS-EPIQSISAFQCPTCRYVITLNORGLEGLKRNVTLQNIIDRYOKASQSGPNSPKE : 99
hMID2      : METLESELTCPICLELFEDPLLLPCAHSLCFNCAHRILVSSCSSG-ESIEPTAFQCPTCRYVITSLNHRGLDGLKRNVTLQNIIDRFQKASVSGPNSPSE : 99
hMID1      : METLESELTCPICLELFEDPLLLPCAHSLCFNCAHRILVSHCAIN-EPVESITAFQCPTCRHVITLSORGLDGLKRNVTLQNIIDRFQKASVSGPNSPSE : 99
mMID2      : METLESELTCPICLELFEDPLLLPCAHSLCFNCAHRILVSSCSSG-ESIEPTAFQCPTCRYVITSLNHRGLDGLKRNVTLQNIIDRFQKASVSGPNSPSE : 99
mMID1      : METLESELTCPICLELFEDPLLLPCAHSLCFNCAHRILVSHCAIN-EPVESINAFQCPTCRHVITLSORGLDGLKRNVTLQNIIDRFQKASVSGPNSPSE : 99
spretusMID: METLESELTCPICLELFEDPLLLPCAHSLCFNCAHRILVSHCAIN-EPVESINAFQCPTCRHVITLSORGLDGLKRNVTLQNIIDRFQKASVSGPNSPSE : 99
rMID1      : METLESELTCPICLELFEDPLLLPCAHSLCFNCAHRILVSHCAIN-EPVESINAFQCPTCRHVITLSORGLDGLKRNVTLQNIIDRFQKASVSGPNSPSE : 99

*           120          *           140          *           160          *           180          *           200
cMID1      : TRRE-RAFDNSMS-CEKVLQFCQDPAQDAVKTCTVTCVSYCECLKATHPNKKPFTGHRLIEPIPDShIRGLMCLHEHEDEKVNMYCVTDQOLICAL : 197
fMID       : TRREGAVPDSRAMTSPGDRVPCQFCQDPPQDAVKTCTITCEVSYCECLKATHPNKKPFTGHRLMEPLDShLRGIMCAEHEDEKVNMYCVTDEQOLICAL : 199
hMID2      : SRRERTYRPTTAMSS--ERTACQFCQDPPRDAVKTCTITCEVSYCDRCLRATHPNKKPFTSHRLVEPVDPThLRGITCLDHENEKVNMYCVSDDQOLICAL : 197
hMID1      : TRRE-RAFDANTMTS-AEKVLQFCQDPAQDAVKTCTVTCVSYCECLKATHPNKKPFTGHRLIEPIPDShIRGLMCLHEHEDEKVNMYCVTDQOLICAL : 197
mMID2      : SRRERTYRPSAMSS--ERTACQFCQDPPRDAVKTCTITCEVSYCDRCLRATHPNKKPFTSHRLVEPVSDThLRGITCLDHENEKVNMYCVSDDQOLICAL : 197
mMID1      : TRRE-RAFDANTMSS-AEKVLQFCQDPAQDAVKTCTVTCVSYCECLKATHPNKKPFTGHRLIEPIPDShIRGLMCLHEHEDEKVNMYCVTDQOLICAL : 197
spretusMID: TRRE-RAFDANTMSS-AEKVLQFCQDPAQDAVKTCTVTCVSYCECLKATHPNKKPFTGHRLIEPIPDShIRGLMCLHEHEDEKVNMYCVTDQOLICAL : 197
rMID1      : TRRE-RAFDANTMSS-AEKVLQFCQDPAQDAVKTCTVTCVSYCECLKATHPNKKPFTGHRLIEPIPDShIRGLMCLHEHEDEKVNMYCVTDQOLICAL : 197

*           220          *           240          *           260          *           280          *           300
cMID1      : CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLLAKLIQTCQHVENVASRQETKLMEECDQLEIIQORRQIIGTKIKEGKVMRLRKLAAQ : 297
fMID       : CKLVGRHRDHQVAALGDRYDKLKESLESNLTNLIKRTSDLESIMGKLIQTCQHVENVASRQENKLEECDLIIIIQORRQIITTKIKEGKAVRLRKLAAQ : 299
hMID2      : CKLVGRHRDHQVASLNDRFKELKOTLEMNLTNLVKRNSLENQMAKLIQTCQOQVEVNTAMHEAKLMEECDLVEIIQORRQIMAVKIKETKVMKLRKLAAQ : 297
hMID1      : CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLLAKLIQTCQHVENVASRQEAKLTEECDLIEIIQORRQIIGTKIKEGKVMRLRKLAAQ : 297
mMID2      : CKLVGRHRDHQVASLNDRFKELKOTLEMNLTNLVKRNSLENQMAKLIQTCQOQVEVNTAMHEAKLMEECDLVEIIQORRQIMAVKIKETKVMKLRKLAAQ : 297
mMID1      : CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLLAKLIQTCQHVENVASRQEAKLTEECDLIEIIQORRQIIGTKIKEGKVMRLRKLAAQ : 297
spretusMID: CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLLAKLIQTCQHVENVASRQEAKLTEECDLIEIIQORRQIIGTKIKEGKVMRLRKLAAQ : 297
rMID1      : CKLVGRHRDHQVAALSERYDKLKONLESNLTNLIKRNTELETLLAKLIQTCQHVENVASRQEAKLTEECDLIEIIQORRQIIGTKIKEGKVMRLRKLAAQ : 297

```

```

      *      320      *      340      *      360      *      380      *      400
cMID1 : QIANKCKCIERSSTSLISQAEQS LKENDHARFLOTAKNI I E RVSMATASSQVLIPE INLNDTFDTFALDF TREKKLLE LDYLTAPNPPPT IREELCTASYD : 397
fMID : QIAGCKCIERSSSSLITQADQALKEADHTRFLOTAKSIC E RVSMATASSQVLLPE INLNDTFDTFALDF TREKKMLEN LDYLTAPNPPV IREELCTASYD : 399
hMID2 : QVANCRCLERS TVLINQAEHT LKENDQARFLOS AKNI I E RVSMATASSQVLIPE INLNDTFDTFALDF SREKKLLE LDYLTAPNPPS IREELCTASHD : 397
hMID1 : QIANKCKCIERSASLISQAEHS LKENDHARFLOTAKNI I E RVSMATASSQVLIPE INLNDTFDTFALDF SREKKLLE LDYLTAPNPPS IREELCTASHD : 397
mMID2 : QVANCRCLERS TVLINQAEHT LKENDQARFLOS AKNI I E RVSMATASSQVLIPE INLNDTFDTFALDF SREKKLLE LDYLTAPNPPS IREELCTASHD : 397
mMID1 : QIANKCKCIERSASLISQAEHS LKENDHARFLOTAKNI I E RVSMATASSQVLIPE INLNDTFDTFALDF SREKKLLE LDYLTAPNPPS IREELCTASYD : 397
spretusMID : QIANKCKCIERSASLISQAEHS LKENDHARFLOTAKNI I E RVSMATASSQVLIPE INLNDTFDTFALDF SREKKLLE LDYLTAPNPPPT IREELCTASYD : 397
rMID1 : QIANKCKCIERSASLISQAEHS LKENDHARFLOTAKNI I E RVSMATASSQVLIPE INLNDTFDTFALDF SREKKLLE LDYLTAPNPPPT IREELCTASYD : 397

```

```

      *      420      *      440      *      460      *      480      *      500
cMID1 : TITVHWISDDEFSVMSYELQYITIFTGOANVV -----SLONSADSWMIVPNIKQNHYTVHGLQSGTKYIFIVKAIN : 467
fMID : TITVHWISDDEFVMSYELQYALFTCOSNVVS -----LONSADSWMIVPNIKQNHYTVHGLQCGTKYIFIVKAIN : 469
hMID2 : TITVHWISDDEFSISSYELQYITIFTGOANFISKSWCSWGLWPEIRKCKEAVSCSRLAGAPRGLYNSVDSWMI VPNIKQNHYTVHGLQSGTRYIFIVKAIN : 497
hMID1 : TITVHWISDDEFSVMSYELQYITIFTGOANVVS -----LONSADSWMIVPNIKQNHYTVHGLQSGTKYIFIVKAIN : 467
mMID2 : TITVHWISDDEFSISSYELQYITIFTGOANFIS -----LYNSVDSWMI VPNIKQNHYTVHGLQSGTRYIFIVKAIN : 467
mMID1 : TITVHWISDDEFSVMSYELQYITIFTGOANVVNVACDGTCLL -----GSAGLONSADSWMIVPNIKQNHYTVHGLQSGTKYIFIVKAIN : 480
spretusMID : TITVHWISDDEFSVMSYELQYITIFTGOANVVS -----LONSADSWMIVPNIKQNHYTVHGLQSGTKYIFIVKAIN : 467
rMID1 : TITVHWISDDEFVMSYELQYITIFTGOANVV -----SLONSADSWMIVPNIKQNHYTVHGLQSGTKYIFIVKAIN : 467

```

```

      *      520      *      540      *      560      *      580      *      600
cMID1 : QAGSRSEPECKLKTNSQPFKLDPKSAHRKLVSHDNLITVERDETS SKKSHTPERFITSQGSYGVAGNVFIDSGRHYWEVWISGSTWYAIGI SYKSAPKHEW : 567
fMID : QAGNRSEPEAKLKTNSQPFRLDPKSAHRKLRVSHDNLITVERDETS AKKSHSODRFITSHSSYGVITGNVYIDSGRHYWEALIGGSTWFVAVGVAYKSAPRHEW : 569
hMID2 : QAGSRNSEPETRLKKTNSQPFKLDPKMTHKKLKI SNDGLOMEKDESSLKKSHTPERFSGTCYGAAGNIFIDSGCHYWEVWVGSTWYAIGIAYKSAPKNEW : 597
hMID1 : QAGSRSEPECKLKTNSQPFKLDPKSAHRKLVSHDNLITVERDESSSKKSHTPERFITSQGSYGVAGNVFIDSGRHYWEVWISGSTWYAIGIAYKSAPKHEW : 567
mMID2 : QAGSRNSEPETRLKKTNSQPFKLDPKMTHKKLKI SNDGLOMEKDESSLKKSHTPERFSGTCYGAAGNIFIDSGCHYWEVWVGSTWYAIGIAYKSAPKNEW : 567
mMID1 : QAGSRSEPECKLKTNSQPFRLDPKSAHRKLVSHDNLITVERDESSSKKSHPERFAGCGSYGVAGNVFIDSGRHYWEVWISGSTWYAIGLAYRSAPKHEW : 580
spretusMID : QAGSRSEPECKLKTNSQPFKLDPKSAHRKLVSHDNLITVERDESSSKKSHTPERFITSQGSYGVAGNVFIDSGRHYWEVWISGSTWYAIGLAYRSAPKHEW : 567
rMID1 : QAGSRSEPECKLKTNSQPFKLDPKSAHRKLVSHDNLITVERDESSSKKSHTPERFITSQGSYGVAGNVFIDSGRHYWEVWISGSTWYAIGLAYRSAPKHEW : 567

```

```

          *           620           *           640           *           660           *           680           *           700
cMID1    : IGKNSASWVLCRCNNWVVRHNSKETPIEPAPHLRRVGI LLDDYDNGSLAFYDALNSIHLMTFDVAFAPVCPFTFVWNKCLTIITGLPTPDHLLSSEQLA : 667
fMID     : VGKNSASWVLSRCNNSWVVRHNSKEMPVEPPPHLRRLGILLDYDSGSLSFYDAVSSOHLMTFDVAFAPVCPFTFVWNKCLTIITGLPTPDHLLSSEQLA : 669
hMID2    : IGKNASSWVFSRCNSNFVVRHNNKEMLVDPVPHLKRLGVLLDYDNNMLSFYDPANSTHLLHMTFDVAFAPVCPFTFVWNKCLTIITGLPTPDHLLSSEQLA : 697
hMID1    : IGKNSASWALCRCNNWVVRHNSKETPIEPAPHLRRVGI LLDDYDNGSLAFYDALNSIHLMTFDVAFAPVCPFTFVWNKCLTIITGLPTPDHLLDCTEQLP : 667
mMID2    : IGKNASSWVFSRCNSNFVVRHNNKEMLVDPVPHLKRLGVLLDYDNNMLSFYDPANSTHLLHMTFDVAFAPVCPFTFVWNKCLTIITGLPTPDHLLSSEQLA : 667
mMID1    : IGKNASSWALCRCNNWVVRHNSKETPIEPAPHLRRVGI LLDDYDNGSLAFYDALNSIHLMTFDVAFAPVCPFTFVWNKCLTIITGLPTPDHLLDCTEQLP : 680
spretusMID: IGKNSASWALCRCNNWVVRHNSKETPIEPAPHLRRVGI LLDDYDNGSLAFYDALNSIHLMTFDVAFAPVCPFTFVWNKCLTIITGLPTPDHLLDCTEQLP : 667
rMID1    : IGKNSASWALCRCNNWVVRHNSKETPIEPAPHLRRVGI LLDDYDNGSLAFYDALNSIHLMTFDVAFAPVCPFTFVWNKCLTIITGLPTPDHLLDCTEQLP : 667

```

```

          *
cMID1    : ----- : -
fMID     : ----- : -
hMID2    : CNCRPQESPYPVSGMKTCH : 715
hMID1    : ----- : -
mMID2    : CNCRPQESPYPVSGMKACH : 685
mMID1    : ----- : -
spretusMID: ----- : -
rMID1    : ----- : -

```

Appendix Three: Alignment of the nucleotide sequence of the *MID* homologues.

	*	20	*	40	*	60	*	80	*	100		
cMID1	:	ATGGAACA CTGGAGTC GGA ACTGACCTGCCCTATCTGTCTCGAGCTGTTTGAGGACCCGCTGCTGCTGCCCTGCCCTCACAGTCTCTGCTTCAACTGCC										: 100
fMID	:	ATGGAACG CTGGAGTC GGAGCTGACCTGTCCAATCTGTCTGGAGCTATTTGAGGACCCGCTGCTCTGCCCTGCCCTCACAGCCTTTGTTTCAACTGCC										: 100
hMID2	:	ATGGAACA CTGGAGTCTGAATTGACCTGTCCAATCTGCCCTAGAGTTGTTTGAAGACCCCTTCTGCTCCCTTGTGCTCACAGCCTCTGCTTCAACTGCC										: 100
hMID1	:	ATGGAACA CTGGAGTCAGA ACTGACCTGCCCTATTTGTCTCGAGCTCTTTGAGGACCCCTTCTACTGCCCTGCCACACAGCCTCTGCTTCAACTGCC										: 100
mMID2	:	ATGGAACA CTGGAGTCAGA ATTGACCTGTCCAATCTGCCCTCGAGTTGTTTGAAGACCCCTTCTGCTTCCCTTGTGCTCACAGCCTATGTTTCAACTGCC										: 100
mMID1	:	ATGGAACA CTGGAGTCGGAGCTGACCTGTCCATTTGTCTCGAGCTTTTGAAGACCCCTCTCCTGCTTCCCTGCCACACATAGCCTCTGCTTCAACTGCC										: 100
spretusMID	:	ATGGAACA CTGGAGTCGGAGCTGACCTGTCCATTTGTCTCGAGCTTTTGAAGACCCCTCTCCTGCTTCCCTGTGCACATAGCCTCTGTTTCAACTGCC										: 100
rMID1	:	ATGGAACA CTGGAGTCGGAGCTGACCTGTCCATTTGTCTCGAGCTTTTGAAGACCCCTCTCCTGCTTCCCTGTGCACATAGCCTCTGTTTCAACTGCC										: 100
zMID	:	ATGGAACA CTGGAGTCAGA ACTGACTGCCCAATCTGTCTCGAGCTGTTTGAAGACCCCTCTCCTGCTGCCCTGCCACAGCCTGTGCTTCAACTGCC										: 100
	*	120	*	140	*	160	*	180	*	200		
cMID1	:	GGCACCGAATCCTGGTCTCCACTGCGCCTCC---AAGGAGCCGGTGGAGTCTATCAACGGCCTTCCAGTGC---CCACCTGCCGCTATGTCATCACCTCAG										: 197
fMID	:	CCACCGAATCCTGGTGTCAACTGCAAGCC---AGCGAGCCCATCCAATCCATCAGCGCCTTCCAGTGC---CCGACCTGCCGCTATGTCATCACCTCAA										: 197
hMID2	:	CCACCGAATTTGGTATCAAGCTGAGCTCT---GGTGAATCCATTGAACCCATTACTGCTTTCCAGTGTCTTACCTGCAGGATGTTATCTCGCTGAA										: 197
hMID1	:	CCACCGAATCCTAGTATCAACTGTGCCACC---AAGGAGTCTGTGGAGTCCATCAACGGCCTTCCAGTGC---CCACCTGCCGGCATGTCATCACCTCAG										: 197
mMID2	:	CCACCGAATCCTGGTCTCAAGCTGAGCTCT---GGTGAATCCATTGAACCCATTACTGCTTTCCAGTGTCTTACATGCAGGATGTTATCTCGCTGAA										: 197
mMID1	:	CCACCGAATCCTGGTCTCTCACTGTGCCACC---AAGGAGCCTGTGGAGTCCATCAACGGCCTTCCAGTGC---CCACCTGCCGGCATGTCATCACCTCAG										: 197
spretusMID	:	CCACCGAATCCTGGTCTCTCACTGTGCCACC---AAGGAGCCTGTGGAGTCCATCAACGGCCTTCCAGTGC---CCACCTGCCGGCATGTCATCACCTCAG										: 197
rMID1	:	CCACCGAATCCTGGTCTCTCACTGTGCCACC---AAGGAGCCTGTGGAGTCCATCAACGGCCTTCCAGTGC---CCACCTGCCGGCATGTCATCACCTCAG										: 197
zMID	:	CCACCGAATCCTGACATCCACTGCGCAGCCTTCGACCACTCCAGCAAAACCACCCTCCGCCTTCCAGTGTCCACCTGCCGCTATGTCATCACCTGGA										: 200
	*	220	*	240	*	260	*	280	*	300		
cMID1	:	CCAGCGGGTCTAGAGGGCTCAAGCGAACGTCACCTGCAAAACATCAGACCGCTTTTTCAGAAAGCCTCGGTGAGCGGGCCAACTCCCCAGCGAG										: 297
fMID	:	CCAGAGGGGCTAGAGGGCTCAAGCGAACGTTAGCTGCAAAACATCATTGACCGTTTACAGAAAGCCTCGCAAAGCGGACCCAACTCTCCCAAGAG										: 297
hMID2	:	CCACCGGGCTTGGATGGCTCAAGAGGAATGTGACTCTGCAGAACATTATTGATCGCTTTCAGAAAGCCTTCACTCAGTGGGCCCCAACTCTCCAGCGAG										: 297
hMID1	:	CCAGCGGGTCTAGACGGCTCAAGCGAACGTCACCTTACAAACATCAGACAGGTTTCAGAAAGCATCAGTGCAGCGGGCCCCAACTCTCCAGCGAG										: 297
mMID2	:	TCACCGGGGCTTGGATGGCTCAAGAGGAATGTGACTCTGCAGAACATTATTGATCGCTTTCAGAAAGCCTTCACTCAGTGGGCCCCAACTCTCCAGCGAG										: 297
mMID1	:	CCAGCGGGTCTAGACGGCTCAAGCGAACGTCACCTTCAAAACATCATTGACAGATTTTCAGAAAGCATCTGTGAGCGGGCCCCAACTCTCCAGCGAG										: 297
spretusMID	:	CCAGCGGGTCTAGACGGCTCAAGCGAACGTCACCTTACAAACATCATTGACAGATTTTCAGAAAGCATCTGTGAGCGGGCCCCAACTCTCCAGCGAG										: 297
rMID1	:	CCAGCGGGTCTAGACGGCTCAAGCGAACGTCACCTTACAAACATCATTGACAGGTTTTCAGAAAGCATCTGTCAGTGGGCCCCAACTCTCCAGCGAG										: 297
zMID	:	CCAGCAGGGTCTAGACGGCTGAAACCGAACGTCACCTTACAAACATCAGACCGCTTTCAGAAAGCCTCTCTCAGCGGGCCCCAACTCTCCAGCGAG										: 300


```

*          320          *          340          *          360          *          380          *          400
cMID1      : ACCCGTCCGGGACCGGGCAATTCCAGCAGCAAC-----AGCATGTCCTCTGCTCCGAGAAGGTCCTCTGCAGATTCTGCGAACAGGACCCCTGCCAGGAGGCAG : 391
fMID       : ACCCGCGCCCGAGCGGAGCCGTCCCGACAGCAGAGCCATGACGTCAACCCGCGACAGCCTGCCGTGTCACTTCTGTGAGCAGGATCCCTCCCTCAGGATGCCG : 397
hMID2     : AGCCGCGGGGAAAGGACTTACAGGCCACC-----ACTGCCATGCTTACCAGCGAATGTCTTGCATTTCTGTGAGCAGGACCCGCCAAGGATGCAG : 391
hMID1     : ACCCGTCCGGGACCGGGCCTTTGACCCCAAC-----ACCATGACCTCCCGGAGAGGTCCTCTGCAGTTTCTGTGAGCAGGATCCCTCCGAGAGATGCTG : 391
mMID2     : AGCCGCGGGGAGGGACTTACAGGCCTAGC-----TCCGCCATGTGACTGAGAGAATTCATGCTCAATTTCTGTGAGCAGGACCCCTCCGAGAGATGCTG : 391
mMID1     : ACCCGCAGGGAAAGGGCCTTTGACCCCAAC-----ACCATGTCCTCTGCCGAGAAGGTCCTCTGCAGTTTCTGTGAGCAGGATCCCTCCGAGGATGCCG : 391
spretusMID : ACCCGCAGGGAAAGGGCCTTTGACCCCAAC-----ACCATGTCCTCTGCCGAGAAGGTCCTCTGCAGTTTCTGTGAGCAGGATCCCTCCGAGGATGCCG : 391
rMID1     : ACCCGCAGGGAAAGGGCCTTTGACCCCAAC-----ACCATGTCCTCTGCCGAGAAGGTCCTCTGCAGTTTCTGTGAGCAGGATCCCTCCGAGGATGCCG : 391
zMID      : ACACGT-----AAACAGATTTCTCCCTCCG-----CATGAGTTCTGCATCCGAGGCCGTAAGAGTGTCACTTCTGCCAGCAGGATCCGCCGAGGAGCCG : 391

```

```

*          420          *          440          *          460          *          480          *          500
cMID1      : TCAAGACCTGCCCTGACCTGCCAGGTCCTCTACTGCGACGAGTGCCTGAAAGCCACGCCCAACAAGAAGCCCTTTACCCGGCCATCCGCTGATCGAGCC : 491
fMID       : TCAAGACCTGCCATCACCTGCCAGGTCCTCTACTGCGACGAGTGTCTCAAGGCCACCCCAACCACAAGAAGCCGTTTACCCGGCTCACCCCTCATGGAGCC : 497
hMID2     : TAAAAACATGCATCACCTGTGAGGTCCTCTACTGCGACCGCTTGCCTCCGGCCACGCCCAACAAGAAGCCCTTTACCCAGCCACCCGCTGGTGGACC : 491
hMID1     : TCAAGACCTGCTTACCTGTGAGGATCCTACTGTCGACGAGTGCCTGAAAGCCACTCACCCGAATAAGAAGCCCTTTACCCAGCCATCCCTCGTGGACC : 491
mMID2     : TAAAGACCTCCATCACCTGTGAGGTCCTCTACTGTCGACCGTGTGCTCCGGCCACACAACCACAAGAAGCCCTTTACAGGCCATCCGCTGATGAGCC : 491
mMID1     : TCAAGACTTGCCTTACTGTGAGGTCCTCTACTGTCGATGAGTGCCTGAAAGCCACTCATCCGAACAAGAAGCCCTTTACAGGCCATCCGCTGATGAGCC : 491
spretusMID : TCAAGACTTGCCTTACTGTGAGGTCCTCTACTGTCGATGAGTGCCTGAAAGCCACTCATCCGAACAAGAAGCCCTTTACAGGCCATCCGCTGATGAGCC : 491
rMID1     : TCAAGACCTGCTTACTTGCAGGTCCTCTACTGTCGATGAGTGCCTGAAAGCCACTCATCCGAATAAGAAGCCCTTTACAGGCCATCCGCTGATGAGCC : 491
zMID      : TCAAGACTTGCCTTACTTGCAGGTCCTCTACTGTCGATGAGTGCCTTCCGGCCGACGCCCAATAAGAAGCCGTTTACAGGCACAGACTGACCGAGCC : 491

```

```

*          520          *          540          *          560          *          580          *          600
cMID1      : CATTCCGGACTCCACATCAGGGGATTAATGTGCTTGGACCAAGAGGACGAGAAGGTGAACATGTACTTCCGTCAGAGATGACCACTGATCTGCCCTTG : 591
fMID       : TCTGCTGACTCCCATCTGCCCGGGATAATGTGCGCGGAGCAAGAGGACGAGAAGGTCAAATGTACTTGTGTACCGACCAACAATTGATCTGCCGATTG : 597
hMID2     : AGTGCAGACACACATCTTCGAGGGATCACCTGCCCTGGACCAATGAGAAAGAGAAGGTGAACATGTACTTGTGTATCTGATGACCAATTGATCTGTGCCTTA : 591
hMID1     : AATTCCGGACTCCACATCCGGGGGTGATGTGCTTGGACCAATGAGAAAGAGAAGGTGAATATGTACTTGTGTACCGATGACCAAGTTAATCTGTGCCTTG : 591
mMID2     : AGTTTCAGACACACATCTTCGAGGGATTAACCTGCCCTGGACCAAGAGAAAGAGAAGGTGAATATGTACTTGTGTACCGATGACCAAGTTAATCTGTGCCTTG : 591
mMID1     : AATCCGGACTCCGACATCCGGGGGTGATGTGCCCTGGACCAAGAGAAAGAGAAGGTGAATATGTACTTGTGTACCGATGACCAAGTTAATCTGTGCCTTG : 591
spretusMID : AATCCGGACTCCGACATCCGGGGGTGATGTGCCCTGGACCAAGAGAAAGAGAAGGTGAATATGTACTTGTGTACCGATGACCAAGTTAATCTGTGCCTTG : 591
rMID1     : AATCCGGACTCCGACATCCGGGGGTGATGTGCCCTGGACCAAGAGAAAGAGAAGGTGAATATGTACTTGTGTACCGATGACCAAGTTAATCTGTGCCTTG : 591
zMID      : AGTCCCGGATGCAAGACTTAGAGGACTCGCCCTGTCGCCAACACGGAGAGGAGAAGGTCAAATGTACTTGCATCACAGACGAAACAGCTCATTTGTTCCGCTG : 591

```

```

*          620          *          640          *          660          *          680          *          700
cMID1      : TGCAAGCTGGTCCGGCCGCCACCGGGACCATCAAGTGGCAGCTCTAAGCCGAGCGCTACGACAAAGCTCAAGCAAAATTTGGAGAGTAACTCACCAACCTTA : 691
fMID       : TGCAAGCTGGTGGTCCGACACAGGGACCATCACGTGGCGGCCCTCCGCCGATCGATACGACAAACTCAAGGAATCTTCCAATCTAACTCAACAATCTAA : 697
hMID2     : TGCAAACTGGTGGGTCCTCACCGAGACCATCAGGTCCGATCCCTGAATGATCGAATTTAGAAACTCAAGCAAACTCGAGATGAACCTCACCAACCTTGG : 691

```

```

hMID1 : TGAAACTGGTGGCCGCCACCCGATCATCAGGTGGCAGCTTTGAGTGAACCGTATGACAAATGAAGCAAACTTGGAGAGTAATCTCACCAACTTA : 691
mMID2 : TGCAAACTGGTGGGTGGTCCGTCACCCGAGACCATCAGGTGGCTTCTTTGAAATGATCGATTGAGAAAATAAAGCAAACTCTCGAGATGAACTCACCAACTGG : 691
mMID1 : TGAAACTGGTGGCCGCCACCCGATCATCAGGTGGCAGCTTTGAGTGAACCGTATGACAAATGAAGCAAACTTGGAGAGTAATCTCACCAACTTA : 691
spretusMID : TGAAACTGGTGGCCGCCACCCGATCATCAGGTGGCAGCTTTGAGTGAACCGTATGACAAATGAAGCAAACTTGGAGAGTAATCTCACCAACTTA : 691
rMID1 : TGAAACTGGTGGCCGCCACCCGATCATCAGGTGGCAGCTTTGAGTGAACCGTATGACAAATGAAGCAAACTTGGAGAGTAATCTCACCAACTTA : 691
zMID : TGAAACTCGTAGGACATCATCGTGAACACAGGTGGCGTTCGCTCGGGGACCCGATACGACAACTGAAGCAAACTTGGATTCCAACCTCAGCAATTGA : 691

```

```

* 720 * 740 * 760 * 780 * 800
cMID1 : TTAAGAGGAATACTGAACTGGAAACTCTCTGGCAAAACTCATTCAGACCTGTCAACATGTAGAAGTAAATGCATCTCGCCAAAGAAACCAAGCTGATGGA : 791
fMID : TCAAGAGGACCAAGTGAATTTGGAAAGTCTGATGGCTAAACTTATTCAAACCTGCCAGCACGTCGAGGTAATGCATCCAGACAGGAAACCAAGCTCCTCGA : 797
hMID2 : TTAAGCGCAACAGCGAACTAGAAAATCAAATGGCCAAACTAATACAGATCTCCAGCAGGTTGACGTGAATACTGCTATGCATGAGCCAAACTTATGGA : 791
hMID1 : TTAAGAGGAACACAGAACTGGAGACCTTTGGCTAAACTCATCCAAACCTGTCAACATGTTGAAGTCAATGCATCCCGTCAAGAAGCCAAACTGACAGA : 791
mMID2 : TTAAGCGCAACAGTGAACTAGAAAATCAAATGGCCAAACTAATACAGATCTCCAGCAAGTTGACGTGAATACTGCTATGCATGAGCCAAACTTATGGA : 791
mMID1 : TTAAGAGAAACACAGAACTGGAGACTCTTTGGCTAAACTCATCCAAACCTGTCAACATGTTGAAGTCAATGCATCCCGTCAAGAAGCCAAACTGACAGA : 791
spretusMID : TTAAGAGAAACACAGAACTGGAGACTCTTTGGCTAAACTCATCCAAACCTGTCAACATGTTGAAGTCAATGCATCCCGTCAAGAAGCCAAACTGACAGA : 791
rMID1 : TTAAGAGAAACACAGAACTGGAGACCTTTGGCTAAACTCATCCAAACCTGTCAACATGTTGAAGTCAATGCATCCCGTCAAGAAGCCAAAGTTGACAGA : 791
zMID : TCAAGAGGAACAGCGACTGGAAACCTGATGGCAAAAGTTGATCCAGACCTGTCAACATGTTGAGCACCAATGCGGCTCATCAGGAGAGTAAAGCTGACCGA : 791

```

```

* 820 * 840 * 860 * 880 * 900
cMID1 : AGAATGTGACCAGCTCATGAAATAATCAACACAGACGCAAAATAATCGAACCAAAATCAAGGAGGCAAGGTTGGTGGAGTTGAGAAAACCTGGCTCAG : 891
fMID : GGAGTGTGACCTGCTCATCAACATATACAGCAGCGAAGAATAATAACCACAAAATAAAGAAAGGCAAGGCCGTGGCCCTGAGGAACTCGCCCCAG : 897
hMID2 : AGAATGTGACGAGTTGGTAGAGATATCCAGCAGGAGCAAAATGATCGCTGTCAAAATCAAGAGACAAAGGTTATCAAACTGAGAAAATTGACACAG : 891
hMID1 : GGAGTGTGATCTTCTCATGAGATCATTCAGCAAAAGACGACAGATATCGAACCAAGATCAAGAAAGGCAAGGTTGATGAGCCTTCGCAAACTCGCTCAG : 891
mMID2 : AGAATGTGACGAGTTGGTAGAGATATCCAGCAGGAGAAATAATGATTCCTGTCAAAATCAAGAGACAAAGGTTATCAAACTGAGAAAATTGACACAG : 891
mMID1 : AGAATGTGATCTTCTCATGAAATCATTCAGCAAAAGACAAATATTCGAACAAAGATCAAGAAAGGCAAGGTTGATGAGCCTTCGCAAACTGAGCTCAG : 891
spretusMID : AGAATGTGATCTTCTCATGAAATCATTCAGCAAAAGACAAATATTCGAACAAAGATCAAGAAAGGCAAGGTTGATGAGCCTTCGCAAACTGAGCTCAG : 891
rMID1 : GGAATGTGATCTTCTCATGAAATTATTCAGCAAAAGACAAATATTCGAACAAAGATCAAGAAAGGCAAGGTTGATGAGCCTTCGCAAACTGAGCTCAG : 891
zMID : GGAGTGGACACACTGATCAACATCTCCAGCAGAGGGACAGATCATCGCAACCAAGATCAAGAGGGCAAGGTTGGTTCCTTCAGAAAACCTGGCTCAG : 891

```

```

* 920 * 940 * 960 * 980 * 1000
cMID1 : CAGATTGCCAACTGTAACAGTGCATTCAGCCCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGACAACGATCAAGCTCGCTTCTGCAAA : 991
fMID : CAGATAGCCCGGCTGAAGCAGTGCATTCAGAGGTCCTCTTCCCTCATCAACAGCCGACCAAGGCTCAAGGAGGCGGACCAACTCGTTCCTCCTCAA : 997
hMID2 : CAGATTGCCAACTGTAACAGTGCATTCAGCCCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGACAACGATCAAGCTCGCTTCTGCAAA : 991
hMID1 : CAGATTGCCAACTGTAACAGTGCATTCAGCCCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGACAACGATCAAGCTCGCTTCTGCAAA : 991
mMID2 : CAGATTGCCAACTGTAACAGTGCATTCAGCCCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGACAACGATCAAGCTCGCTTCTGCAAA : 991
mMID1 : CAGATTGCCAACTGTAACAGTGCATTCAGCCCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGACAACGATCAAGCTCGCTTCTGCAAA : 991
spretusMID : CAGATTGCCAACTGTAACAGTGCATTCAGCCCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGACAACGATCAAGCTCGCTTCTGCAAA : 991
rMID1 : CAGATTGCCAACTGTAACAGTGCATTCAGCCCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGACAACGATCAAGCTCGCTTCTGCAAA : 991
zMID : CAGATCGCCAACTGTAACAGTGCATTCAGCCCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGACAACGATCAAGCTCGCTTCTGCAAA : 991

```

```

*      1020      *      1040      *      1060      *      1080      *      1100
cMID1 : CTGCTAAAACATCACCGAAAGGCTTTCATGGCAACTGCATCTCCAGGTTCTAATTCCTGAAATTAATCTCAACGATACTTTTGATACCTTTGGCACT : 1091
fMID  : CTGCTAAAAGCATCTGTGAGAGGGTTTTCATGGCAACAGCATCTTCAAGAGTGTGTACCAGAGATAAAGTTGAATGACACATTTTGATACCTTTTGGCTCT : 1097
hMID2 : CTGCAAAAATATTTGCAGAGAGGTTTCATGGCAACTGCATCTTCTCAAGTTCTGATTCAGACATCAATTTTAATGATGCCTTTGAAAACCTTTGGCTTT : 1091
hMID1 : CTGCTAACAAATACACCGAGAGAGTCTCCATGGCAACTGCATCTCCAGGTTCTAATTCCTGAAATCAACCTCAATGACACATTTTGAACCTTTTGGCTTT : 1091
mMID2 : CTGCCAAAACATTTGCCGAGAGGGTCTGCTATGGCAACTGCATCTTCCAGGTTCTGATTCAGACATCAATTTTAATGATGCCTTTGAAAACCTTTGGCTTT : 1091
mMID1 : CAGCAAAACAATATCACAGAGAGTCTCCATGGCAACTGCATCTCCAGGTTCTAATTCCTGAAATCAACCTCAATGACACATTTTGAACCTTTTGGCTTT : 1091
spretusMID : CAGCAAAACAATATCACAGAGAGTCTCCATGGCAACTGCATCTCCAGGTTCTAATTCCTGAAATCAACCTCAATGACACATTTTGAACCTTTTGGCTTT : 1091
rMID1 : CAGCAAAACAATATCACAGAGAGTCTCCATGGCAACTGCATCTCCAGGTTCTAATTCCTGAAATCAACCTCAATGACACATTTTGAACCTTTTGGCTTT : 1091
zMID  : TGGCCAAAGAGCATCTCCGAGAGAGTCTCCATGGCAACTGCATCTCCAGGTTCTGATTCCTGAGATAAAGCTGAAAGACAGCTTTTGAACCTTTGGCACT : 1091

```

```

*      1120      *      1140      *      1160      *      1180      *      1200
cMID1 : TGATTTTACCAGGGAGAAGAAATGTGTGAAATGCTTTGATTTATCTTACAGCTCCCAACCTCTCCCAACATTCAGAGAAGAGCTCTGTACAGCTTCTTAAGAT : 1191
fMID  : GGATTTTACACGGGAGAAGAAATGTGTTGAAATTTAGATTAAGCTTACAGCAACCGAATCTCCAGTAATCTCCGAGGAGCTGTGACAGCTTCTGACAGAC : 1197
hMID2 : AGATTTTTCACAGAGAAGAAATGTGCTAGAGGGTTAGATTTATTTACAGCCCAACCTCCCAACATCTATCTCCAGAGAAGACTCTGTACTGCCCTCCATGAC : 1191
hMID1 : AGATTTTTCACAGAGAAGAAATGTGCTAGAAATGCTTGATTAAGCTTACAGCTCCCAACCTCTCCCAACATTCAGAGAAGAGCTCTGTACAGCTTCTTAAGAT : 1191
mMID2 : AGATTTTTCACAGAGAAGAAATGATTTGAGGGGGTAGAGTATTTACAGCTCCCAACCTCTCCCAACATTCAGAGAAGAGCTCTGTACAGCTTCTTAAGAT : 1191
mMID1 : GGATTTTTCACGGGAGAAGAAATGTGCTAGAAATGCTTGATTAAGCTTACAGCTCCCAACCTCTCCCAACATTCAGAGAAGAGCTCTGTACAGCTTCTTAAGAT : 1191
spretusMID : GGATTTTTCACGGGAGAAGAAATGTGCTAGAAATGCTTGATTAAGCTTACAGCTCCCAACCTCTCCCAACATTCAGAGAAGAGCTCTGTACAGCTTCTTAAGAT : 1191
rMID1 : AGATTTTTCACGGGAGAAGAAATGTGCTAGAAATGCTTGATTAAGCTTACAGCTCCCAACCTCTCCCAACATTCAGAGAAGAGCTCTGTACAGCTTCTTAAGAT : 1191
zMID  : GGATTTTTCACAGAGAAGAAATGTGCTGAAAGGCTTGATTAAGCTTACAGCTCCCAACCTCTCCCAACATTCAGAGAAGAGCTCTGTACAGCTTCTTAAGAT : 1191

```

```

*      1220      *      1240      *      1260      *      1280      *      1300
cMID1 : ACCATACAGTTCACCTGGACATCCAGACGATGAGTTTCAGCGTGGTCTCTTACGAGCTGCAGTACACCATCTTCACAGGACAGCTAACGTTGTTA----- : 1285
fMID  : ACCATACAGTTCACCTGGACATCCAGACGATGAGTTTCAGCGTGGTCTCTTACGAGCTGCAGTACACCATCTTCACAGGACAGCTAACGTTGTTA----- : 1291
hMID2 : ACCATACAGTTCACCTGGACATCCAGACGATGAGTTTCAGCGTGGTCTCTTACGAGCTGCAGTACACCATCTTCACAGGACAGCTAACGTTGTTA----- : 1291
hMID1 : ACCATACAGTTCACCTGGACATCCAGACGATGAGTTTCAGCGTGGTCTCTTACGAGCTGCAGTACACCATCTTCACAGGACAGCTAACGTTGTTA----- : 1286
mMID2 : ACCATACAGTTCACCTGGACATCCAGACGATGAGTTTCAGCGTGGTCTCTTACGAGCTGCAGTACACCATCTTCACAGGACAGCTAACGTTGTTA----- : 1291
nMID1 : ACCATACAGTTCACCTGGACATCCAGACGATGAGTTTCAGCGTGGTCTCTTACGAGCTGCAGTACACCATCTTCACAGGACAGCTAACGTTGTTA----- : 1285
spretusMID : ACCATACAGTTCACCTGGACATCCAGACGATGAGTTTCAGCGTGGTCTCTTACGAGCTGCAGTACACCATCTTCACAGGACAGCTAACGTTGTTA----- : 1285
rMID1 : ACCATACAGTTCACCTGGACATCCAGACGATGAGTTTCAGCGTGGTCTCTTACGAGCTGCAGTACACCATCTTCACAGGACAGCTAACGTTGTTA----- : 1285
zMID  : ACCATACAGTTCACCTGGACATCCAGACGATGAGTTTCAGCGTGGTCTCTTACGAGCTGCAGTACACCATCTTCACAGGACAGCTAACGTTGTTA----- : 1285

```

```

*      1320      *      1340      *      1360      *      1380      *      1400
cMID1 : -----GTTTATGTAACCTCAGC : 1301
fMID  : -----GTTTGTGCAACTCTGC : 1307
hMID2 : CATGGTGTAGTTGGGGCTGTGGCCAGAGATAAGGAAATGTAAGGAAGCAGTAAGCTGCTCAAGATTGGCCGGGGCGCCACGAGGCTGTATAAATTCAGT : 1391

```

```

hMID1 : -----GTCTCTGTAAATTCGGC : 1301
mMID2 : -----CCTGTACAATTCAGT : 1301
mMID1 : -----CCTCGGCTCCGCAGGTCTGTGTAATTCGGC : 1340
spretusMID : -----GTCTATGTAATTCAGC : 1301
rMID1 : -----GTCTTTGTAAATTCGGC : 1301
zMID : -----GTCTCTGTAAATTCGAT : 1301

```

```

* 1420 * 1440 * 1460 * 1480 * 1500
cMID1 : CGACAGCTGGATGATGTTCCCAAATACAAAACAAACCACATACACCGTGCATGGGTACACAGAGGCACAAAGTACATCTTTCATTTGTTAAAGCCATTAAT : 1401
fMID : CGACAGCTGGATGATGTAACCAACATCAAGCAGAACCACTACACTGTGCACGGCCTCCAGTCCGGAACCAAGTACATCTTTATTTGTGAAAGCCATAAAC : 1407
hMID2 : AGACAGCTGGATGATGTTCCCAACATTAAGCAGAACCATTACACAGTGCATGGACTCCACAGCGGGACTCGCTACATCTTTCATCGTTAAAGCCATAAAC : 1491
hMID1 : TGATAGCTGGATGATGATACCAACATCAAGCAGAACCACATACACCGTGCACGGTCTGCAGAGCGGGACCAAGTACATCTTTCATCGTTAAAGCCATAAAC : 1401
mMID2 : GGATAGCTGGATGATGTTCCCAACATTAAGCAGAACCATTACACCGTGCACGGCCTGCAGAGCGGGACCAAGTACATCTTTCATCGTTAAAGCCATAAAC : 1440
mMID1 : GGAAGCTGGATGATGTTCCCAACATCAAGCAGAACCATTACACTGTGCATGGCCTACAAAGTGGACCAAGTATATCTTTCATGGTGAAGGCCATCAAC : 1401
spretusMID : GGAAGCTGGATGATGTTCCCAACATCAAGCAGAACCATTACACTGTGCATGGCCTACAAAGTGGACCAAGTATATCTTTCATGGTGAAGGCCATCAAC : 1401
rMID1 : AGATAGCTGGATGATGTTCCCAACATCAAGCAGAACCATTACTGTGCATGGCCTACAAAGTGGACCAAGTATATCTTTCATGGTGAAGGCCATCAAC : 1401
zMID : GGAAGCTGGATGATGTTCCCAACATCAAGCAG----- : 1335

```

```

* 1520 * 1540 * 1560 * 1580 * 1600
cMID1 : CAGCCTGGCAGCAGAGCAGCGAGCCAGCCAACTCAAGACAAACAGTCAGCCATTTAAACTGGACCCCAAAATCTGCTCATAGAAAGCTGAAAGTCTCTC : 1501
fMID : CAGCGGGGAAACCGCAGCAGCGAACCAGCAAACTCAAAACTAACAGTCAGCCCTTCAGGTGGACCCAAAGTCAGCTCACGAAAGCTGAGCCTCTCTC : 1507
hMID2 : CAAGCCGGCAGCCCGAACAGTGAACTACCCGACTAAAAACAAACAGCCAACTTTAAATGGATCCCAAAATGACTCACAAGAACTGAAGATCTCCA : 1591
hMID1 : CAGCGGGGAGCCGACAGTGACCTGGGAGTTGAAACAAACAGCCAACTTTAAACTGGATCCCAAAATCTGCTCATAGAAAGCTGAAAGTCTCTC : 1501
mMID2 : CAAGCAGCAGCCGAAACAGCGAACCAGCAAACTCAAAACTAACAGTCAGCCCTTCAGGTGGACCCAAAGTCAGCTCACGAAAGCTGAGCCTCTCTC : 1501
mMID1 : CAGCGGGGAGCCGACAGTGACCTGGGAGTTGAAACAAACAGTCAGCCCTTCAGGTGGATCCCAAAATGACTCACAAGAACTGAAGATCTCCA : 1540
spretusMID : CAGCCAGCAGCCGACAGTGACCTGGGAGTTGAAACAAACAGTCAGCCATTTAAACTGGATCCCAAAATCTGCTCATAGAAAGCTGAAAGTCTCTC : 1501
rMID1 : CAGCGAGCAGCCGACAGTGACCTGGGAGTTGAAACAAACAGTCAGCCATTTAAACTGGATCCCAAAATCTGCTCATAGAAAGCTGAAGTCTCTC : 1501
zMID : ----- :

```

```

* 1620 * 1640 * 1660 * 1680 * 1700
cMID1 : ATGACAACCTGACAGTGGAACTGATCAAACTCTCTCCAAAAGAGTCATACACCAGAGCCATTCAGGCGCCACGGAGCTACCGAGTAGCTGGCAATGT : 1601
fMID : ATGACAACCTGACAGTGGAACTGATCAAACTCTCTCCAAAAGAGTCATACACCAGAGCCATTCAGGCGCCACGGAGCTACCGAGTAGCTGGCAATGT : 1607
hMID2 : ATGATGGATTCAGATGGAGAAGGATCAAGCTCTCTAAAGAAAGAGCCACACCCAGAGAGGTTAGTGGACAGGGTGTATGGGCGACAGGAAATAT : 1691
hMID1 : ATGATGGATTCAGATGGAGAAGGATCAAGCTCTCTAAAGAAAGAGCCACACCCAGAGAGGTTAGTGGACAGGGTGTATGGGCGACAGGAAATAT : 1601
mMID2 : ATGATGGATTCAGATGGAGAAGGATCAAGCTCTCTAAAGAAAGAGCCACACCCAGAGAGGTTAGTGGACAGGGTGTATGGGCGACAGGAAATAT : 1601
mMID1 : ATGACAACCTGACAGTGGAACTGATCAAACTCTCTCCAAAAGAGTCATACACCAGAGCCATTCAGGCGCCACGGAGCTACCGAGTAGCTGGCAATGT : 1640
spretusMID : ATGACAACCTGACAGTGGAACTGATCAAACTCTCTCCAAAAGAGTCATACACCAGAGCCATTCAGGCGCCACGGAGCTACCGAGTAGCTGGCAATGT : 1601
rMID1 : ATGACAACCTGACAGTGGAACTGATCAAACTCTCTCCAAAAGAGTCATACACCAGAGCCATTCAGGCGCCACGGAGCTACCGAGTAGCTGGCAATGT : 1601
zMID : ----- :

```

```

*      1720      *      1740      *      1760      *      1780      *      1800
cMID1 : CTTTCATGACAGCGGACGGCAATTAATGGGAAGTGGTATATAAGTGGGAGTACATGGTATGCCAATGGTATTTTCATACAAAGTCAGGACCGAAGCAATGAGTGG : 1701
fMID  : CTTACATCGACAGCGGGGCGCCATTAATGGGAAGCTCTGATGTGAGGAAGCACATGGTTTGGAGTGGGAGTTGCCATACAACTCGGCCCGGAGAACCCAGTGG : 1707
hMID2 : ATTTCATGACAGTGGTGGCCACTAATGGGAGGTGGTCAATGGCTTCCTCAACATGGTATGGCAATGGCAATTGCCATACAAATCAAGCTCCAAAGAAATGAATGG : 1791
hMID1 : CTTTATGATAGTGGCCGGCAATTAATGGGAAGTGGTCAATAAGTGGGAAGCACATGGTATGCCAATGGTCTTTGGCTTACAAATCAAGCTCCAAAGAAATGAGTGG : 1701
mMID2 : ATTTCATCGACAGTGGTGGCCACTAATGGGAGGTGGTCAATGGCTTCCTCAACATGGTATGGCAATGGCAATTGCCATACAAATCAAGCTCCAAAGAAATGAGTGG : 1740
spretusMID : CTTTCATGACAGCGGCGGTCACTAATGGGAAGTGGTCAATAAGTGGGAAGTACATGGTATGCCAATGGCCTTGGCATACAAATCAAGCTCCAAAGAAATGAATGG : 1701
rMID1 : CTTTATGACAGTGGCCGTCACTAATGGGAAGTGGTCAATAAGTGGGAAGCACATGGTATGCCAATGGCCTTGGCTTACAAATCAAGCTCCAAAGAAATGAATGG : 1701
zMID  : -----

```

```

*      1820      *      1840      *      1860      *      1880      *      1900
cMID1 : AATCGGGAAGAATCTAGCTTCCTGGGTGCTGTGCCGCTGCAACAACACATGGGTGGTGAGACAACAACAACAAGAAATCCCAATAGAGCCCTGCGCCTCATC : 1801
fMID  : GTGGCCAAAACCTCGCCCTCCTGGGTGCTGTGCCAGATGCAACAACACTCGTGGGTGGCCGCCCAACAACAACAAGGAGATGCCCGTGGAGCCCGCCGACACC : 1807
hMID2 : AATGGCAAGAATGCTCTCATGGGTCTTCTCTCGCTGCAATAGTAACTTCCTGGTGAGACAACAACAAGGAAATGTGGTGGATGTGCCCCACACC : 1891
hMID1 : AATGGCAAGAATCTGCTTCCTGGGCGCTGTGCCGTGCAACAATAACTGGGTGGTGAGACAACAACAAGGAAATGTGGTGGATGTGCCCCACAGT : 1801
mMID2 : AATGGCAAGAATGCTCTCATGGGTCTTCTCTCGAATGCAACAGTAACTTCCTGGTTCGGCAACAACAAGGAAATGTGGTGGATGTGCCCCACAGT : 1801
nMID1 : AATCGGGAAGAATGCTGCTTCCTGGGCGCTGTGCCGCTGCAACAACCACATGGGTGGTGAGACAACAACAAGGAGACCCCAATCGCGCGGCGCCCTCACC : 1840
spretusMID : AATGGGAAGAATCTGCTTCCTGGGCGCTGTGCCGTGCAACAATAACTGGGTGGTGAGACAATAACAACAAGGAAATCCCAATCGAGCCAGCCCGCTCACC : 1801
rMID1 : AATGGGAAGAATCTGCTTCCTGGGCGCTGTGCCGTGCAACAATAACTGGGTGGTGAGACAATAACAACAAGGAAATCCCAATCGAGCCAGCCCGCCACC : 1801
zMID  : -----

```

```

*      1920      *      1940      *      1960      *      1980      *      2000
cMID1 : TCCGCGCTGTCGGCAATTTGCTGGAACTACGCAAAATGCTTCCCTGGCTTTTATGATGGCTTGAACCTCCCTGCAACTTTACACCTTTGACATTACGTTGG : 1901
fMID  : TGGGACCCCTGGGAATATTCTTGGAACTACGACTCCGGACTCTGTCTTTCTACGACCGCTGTGAGCTCGCAGCAATGTACAGGTTCCGACCTCGCTTTGG : 1907
hMID2 : TGAAGCGTCTGGGTGTCTCTCTGGATTATGACAAATAATATGCTGTCTTTCTATGACCCAGCTAACTCTCTCATCTTCATCTTCATACCTTTGATGTGACCTTCAT : 1991
hMID1 : TCCGCGCTGTCGGCAATCTCTGGAACTAATGATAACGGCTCTATCGCTTTTATGATGCTTTGAACCTCCATCCACCTCTACACCTTTGACGTTGCAATTTGG : 1901
mMID2 : TGAAGCGTCTAGGTGTCTCTCTGGATTATGACAAATAATATGCTGTCTTTCTATGACCCAGCTAACTCTCTCATCTTCATCTTCATACCTTTGATGTGACCTTCAT : 1901
nMID1 : TCAAGCGCTGTCGGGTGTCTCTGGAACTACGCAAAATGCTTCCCTGGCTTTTCTACGACCGCTGTGAGCTCCGTCACCTCCACACCTTTCCACGCGGCGCTCGC : 1940
spretusMID : TCAAGCGCTGTCGGCAATCTCTGGAACTAATGATAATGGCTCAATCGCTTTTATGATGCTTTGAACCTCCATCCACCTCTACACCTTTGATGTGCGCTTGC : 1901
rMID1 : CCCGCGCTGTCGGCAATCTCTGGAACTAATGATAATGGCTCAATCGCTTTTATGATGCTTTGAACCTCCATCCACCTCTACACCTTTGATGTGCGCTTGC : 1901
zMID  : -----

```

```

*      2020      *      2040      *      2060      *      2080      *      2100
cMID1 : GCAGCCCGTCTGCCCCACATTCACCTGTGTGGAATAACTCTTTGACCATTATAACAAGCTTTGCCATACCTCATCACTTACACTCCTCTGAGCAAGCTGGGA : 2001
fMID  : TCACCCGCTCTGTCTGTGTTTAACTGTGTGGAACAGATGTCTGCAAACTCTCACCCGACTCCCAATCCCAAGACCACTTGAAGGGGACAGACTACAACAAG : 2007
hMID2 : TCTTCCAGTTTGTCCAACATTTACAATCTGGAACAAATCCCTAATGATCTGTCTGGCTTGGCTTGGCCCAAGATTTTATTGATTACCCTGAGGGCCAGGAA : 2091

```

```

hMID1 : GCAGCCTGTGTGCCCCACCTTCACCGTCTGGAACAAGTGTCTGACGATTATCAGTGGGCTCCCTATCCGAGCCATTGGACTGCACAGACAGCTGCCG : 2001
mMID2 : TCTTCCAGTTTGTCCAAGATTGAGAACTGGAACAAGTCCCATAATGATTCCTTCTGGCTTGGCTGGCCCTGATTTTATTGATTATCCTGACCGGCAGGAA : 2001
mMID1 : GCAGCCCCTGTGTGCCCCACCTTCACCGTCTGGAACAAGTGTCTGACGATTATCAGTGGGCTCCCTATCCGAGCCATTGGACTGCACAGACAGCTGCCG : 2040
spretusMID : GCAGCCTGTGTGTGCCCCACCTTTACTGCTGGAACAAGTGTCTGACGATTATAAAGTGGTCTTCCATCCCAGACCATTGGACTGTACAGAACAGCTACCT : 2001
rMID1 : GCAGCCTGTGTGTGCCCCACCTTTACTGCTGGAACAAGTGTCTGACGATTATAAAGTGGTCTTCCATCCCAGACCATTGGACTGTACAGAACAGCTACCT : 2001
zMID : -----

```

```

*          2120          *          2140          *
cMID1 : TGA----- : 2004
fMID  : TGA----- : 2010
hMID2 : TCCAAGTGCAGGCCCTCAAGAATCCCCTTATGTTTCTGGGATGAAAACCTGTCATTAA : 2148
hMID1 : TGA----- : 2004
mMID2 : TCCAAGTGCAGGCCCTCAAGAATCCCCTTATGTTGTCAGGGATGAAAAGCTTGCCATTAA : 2058
mMID1 : TGA----- : 2043
spretusMID : TGA----- : 2004
rMID1 : TGA----- : 2004
zMID  : ----- : -

```

Appendix Four: Alignment of the nucleotide sequence of the *MID* homologues. The zebrafish sequence has not been included in this alignment to provide a more accurate visual of the similarity between these *MID* homologues at the nucleotide level.

		*	20	*	40	*	60	*	80	*	100		
cMID1	:	ATGGAACA CTGGAGTC CGAAC TGACC TGCC ATCTGTCTC GAGCTG TTTGAGGACCC GCTGCTGCTG CCGTGC TCAAGTCTCTG TTCAACTGCC										:	100
fMID	:	ATGGAACG CTGGAGTC CGAGCTGACC TGCCAATCTGTCTC GAGCTA TTTGAGGACCC GCTGCTCTTG CCGTGC TCAAGCCTTTGT TTCAACTGCC										:	100
hMID2	:	ATGGAACA CTGGAGTC TGAAT TGACC TGCCAATCTGCC TTAGAGTTG TTTGAGGACCC CTTCTGCTCC CTGTGCTCA CAGCCTCTGC TTCAAGCTGTG										:	100
hMID1	:	ATGGAACA CTGGAGTC AGAAC TGACC TGCC ATTTGTCTC GAGCTC TTTGAGGACCC CTTCTACTGCC TGCGACA CAGCCTCTGC TTCAACTGCC										:	100
mMID2	:	ATGGAACA CTGGAGTC CAGAAT TGACC TGCCAATCTGCC CTGAGTTG TTTGAGGACCC CTTCTGCTCC CTGTGCTCA CAGCCTATGT TTTAGCTGTG										:	100
mMID1	:	ATGGAACA CTGGAGTC CGAGCTGACC TGCCAATTTGTCTC GAGCTT TTTGAGGACCC TCTCCTGCTCC CTGTGCGACA TTAGCCTCTGC TTCAACTGCC										:	100
spretusMID	:	ATGGAACA CTGGAGTC CGAGCTGACC TGCC ATTTGTCTC GAGCTT TTTGAGGACCC TCTCCTGCTCC CTGTGCGACA TTAGCCTCTGC TTCAACTGCC										:	100
rMID1	:	ATGGAACA CTGGAGTC CGAGCTGACC TGCCAATTTGTCTC GAGCTT TTTGAGGACCC TCTCCTGCTCC CTGTGCGACA TTAGCCTCTGC TTCAACTGCC										:	100
		*	120	*	140	*	160	*	180	*	200		
cMID1	:	CCACCCGATCCTGGT TCCACTG CCGCTCC --- AACGAGCC GGTGGAGTCTATCAC CGCCTTCCAGTGCCC ACCTGCCGCTATGTCATCACCCTCAG										:	197
fMID	:	CCACCCGATCCTGGT TCAACTG CAGGCC --- AGCGAGCC CATCCAACTCATCAG CGCTTCCAGTGCCC ACCTGCCGCTATGTCATCACCCTCAA										:	197
hMID2	:	CCCATCGATTTGGTATCAAGCTGCAGCTCT --- GGTGAATCCATTTGAACCAATTA CTGCTTCCAGTGCTCTACCTGCAGGTATGTATCTCCCTCAA										:	197
hMID1	:	CCACCCGATCCTGGT TCAACTG CCGACC --- AACGAGCTCTGTGGAGTCCATCAC CGCCTTCCAGTGCCC ACCTGCCGCTATGTCATCACCCTCAG										:	197
mMID2	:	CCACCCGATCCTGGT TCAAGCTGCAGCTCT --- GGTGAATCCATTTGAACCAATTA CTGCTTCCAGTGCTCTACCTGCAGGTATGTATCTCCCTCAA										:	197
mMID1	:	CCACCGAATCCTGGT TCTCACTGTCCACC --- AACGAGCCCTGTGGAGTCCATCAC CGCCTTCCAGTGCCC ACCTGCCGCTATGTCATCACCCTCAG										:	197
spretusMID	:	CCACCGAATCCTGGT TCTCACTGTCCACC --- AACGAGCCCTGTGGAGTCCATCAC CGCCTTCCAGTGCCC ACCTGCCGCTATGTCATCACCCTCAG										:	197
rMID1	:	CCACCCGATCCTGGT TCTCACTGTCCACC --- AACGAGCCCTGTGGAGTCCATCAC CGCCTTCCAGTGCCC ACCTGCCGCTATGTCATCACCCTCAG										:	197
		*	220	*	240	*	260	*	280	*	300		
cMID1	:	CCAGCGGGCTAGAGGGCTCAAGCGCAACGTACCCTGCAGAACATCATC GACCGCTTT CAGAAAGCCTCGGTGAGCGGGCCCAACTCCCCAGCGAG										:	297
fMID	:	CCAGAGGGCTAGAGGGACTCAAACGCAACGTTACCTGCAGAACATCATGAC CGTTAC CAGAAAGCCTCGCAAGCGGACCCAACTCTCCCAAGAG										:	297
hMID2	:	CCACCGGGCTTGGATGGCTCAAGCGGAATGTGACTCTGCAGAACATCATTTGATCGCTTC CAGAAAGCCTCAGTCAAGTGGGCCCAATTCCTCCAGT										:	297
hMID1	:	CCAGCGGGCTAGAGGGCTCAAGCGCAACGTACCCTGCAGAACATCATGACCGCTTT CAGAAAGCCTCAGTCAAGTGGGCCCAATTCCTCCAGT										:	297
mMID2	:	TCACCGGGCTGGATGGCTCAAGCGGAATGTGACCTGCAGAACATCATTTGATCGCTTC CAGAAAGCCTCAGTCAAGTGGGCCCAATTCCTCCAGT										:	297
mMID1	:	CCAGCGGGCTAGAGGGCTCAAGCGCAACGTACCCTGCAGAACATCATTTGACAGATTT CAGAAAGCCTCTGTGAGCGGGCCCAACTCTCCAGT										:	297
spretusMID	:	CCAGCGGGCTAGAGGGCTCAAGCGCAACGTACCCTGCAGAACATCATTTGACAGATTT CAGAAAGCCTCTGTGAGCGGGCCCAACTCTCCAGT										:	297
rMID1	:	CCAGCGGGCTAGAGGGCTCAAGCGCAACGTACCCTGCAGAACATCATTTGACAGGTTT CAGAAAGCCTCCTGAGTGGGCCCAATTCCTCCAGT										:	297


```

hMID2 : TGC AAACTGGTGGGTCGTCACCGAGACCATCAGGTGCGATCCCTGAATGATCGATTGAGAAAACCTCAAGCAAACCTCGGAGATGAACCTCACCAACCTGG : 691
hMID1 : TGTAAACTGGTGGGCGGCACCCGCGATCATCAGGTGGCAGCTTTGAGTGAGCGCTATGACAAATTCAGGCAAACCTTAGAGACTAACCTCACCAACCTTA : 691
mMID2 : TGC AAACTGGTGGTTCGTCACCGAGACCATCAGGTGCGCTTCTGAAATGATCCATTTGAGAAAACCTCAAGCAAACCTCGGAGATGAACCTCACCAACCTGG : 691
mMID1 : TGTAAACTGGTGGGCGGCACCCGCGATCATCAGGTGGCAGCTTTGAGTGAGCGCTATGACAAATTCAGGCAAACCTTAGGAGACTAATCTCACCAATCTTA : 691
spretusMID : TGTAAACTGGTGGGCGGCACCCGCGATCATCAGGTGGCAGCTTTGAGTGAGCGCTATGACAAATTCAGGCAAACCTTAGGAGACTAATCTCACCAATCTTA : 691
rMID1 : TGTAAACTGGTGGGCGGCACCCGCGATCATCAGGTGGCAGCTTTGAGTGAGCGCTATGACAAATTCAGGCAAACCTTAGGAGACTAATCTCACCAATCTTA : 691

```

```

* 720 * 740 * 760 * 780 * 800
cMID1 : TTAAGAGCAATACCTGAACCTGGAAACTCTTTCTGGCAAAAACCTCATTCAGACCTGTCAACAAGTAGAAGTAATGCATCTCCCAAGAAACCAAGCTTGATGGA : 791
fMID : TCAAGAGGACCAAGTGATTTGGAAAGCTGTGATGGGTAACCTTATTCAAACCTGCCAGCACGTCGAGGTAATGCATCCACACAGGAAACCAAGCTGCTGGA : 797
hMID2 : TTAAGCGCAACAGCGAACTAGAAAATCAAATGGCCAAACTAATAACAGATCTGCCAGCAGGTTGAGGTGAATGCTGTATGCATGAGGCCAAAACCTTATGGA : 791
hMID1 : TTAAGAGCAACACAGAACTGGAGACCTTTTGGCTAAAACCTATCCAAACCTGTCAACAAGTTGAAGTCAATGCATCAAGCTCAAGAAGCCAAATTGACAGA : 791
mMID2 : TTAAGCGCAACAGCTGAACCTAGAAAATCAAATGGCCAAACTAATAACAGATCTGCCAGCAAGTTGAGGTGAATGCTGTATGCATGAGGCCAAAACCTTATGGA : 791
mMID1 : TTAAGAGAAACACAGAACTGGAGACTCTTTTGGCTAAAACCTATCCAAACCTGTCAACAAGTTGAAGTCAATGCATCCCGTCAAGAAGCCAAACTGACAGA : 791
spretusMID : TTAAGAGAAACACAGAACTGGAGACTCTTTTGGCTAAAACCTATCCAAACCTGTCAACAAGTTGAAGTCAATGCATCCCGTCAAGAAGCCAAACTGACAGA : 791
rMID1 : TTAAGAGAAACACAGAACTGGAGACTCTTTTGGCTAAAACCTATCCAAACCTGTCAACAAGTTGAGGTCAATGCATCCCGTCAAGAAGCCAAAGTTGACAGA : 791

```

```

* 820 * 840 * 860 * 880 * 900
cMID1 : AGAATGTGACCAGCTCATTGAAATAATACAAAGAGACGACAAATAAATGGAAACCAAAATCAAGGAAGGAAAGGTGGTGAGCTTGACAAAACCTGGCTCAG : 891
fMID : GGAGTGTGACCCTGCTCATCAACATCATACAGCAGCGAAGACAAATAAATTAACCACCAAAATAAAAAGAGGAAAGGCCGTCCGGCTGAGGAAGCTCGCCACAG : 897
hMID2 : AGAATGTGACCAGTGGTAGAGATCATCCAGCAGAGGAAGCAAATGATCGCTGTCAAAATCAAAGAGACAAAGGTTATGAAACTGAGAAAGCTGGCCACAG : 891
hMID1 : GGAGTGTGATCTTCTCATGAGATCATTCAGCAAAGACGACAGATTATTGGTAACCAAGATCAAAGAAGGCAAGGTGATGAGGCTTCGCAAACTGGCTCAG : 891
mMID2 : AGAATGTGACCAGTGGTAGAGATCATCCAGCAGAGGAACCAATGATTTGCTGTCAAAATCAAAGAGACAAAGGTTATGAAACTGAGAAAGCTGGCCACAG : 891
mMID1 : AGAATGTGATCTTCTCATGAAATCATTCAGCAACGAAGACAAATTTATGGTAACAAGATTAAAGAAGGCAAGGTGATGAGGCTCCGCAAACTAGCTCAG : 891
spretusMID : AGAATGTGATCTTCTCATGAAATCATTCAGCAACGAAGACAAATTTATGGTAACAAGATTAAAGAAGGCAAGGTGATGAGGCTCCGCAAACTAGCTCAG : 891
rMID1 : GAAATGTGATCTTCTCATGAAATTATTCAGGAAAGGAAGACAAATTTATGGTAACAAGATTAAAGAAGGCAAGGTGATGAGGCTCCGCAAACTAGCTCAG : 891

```

```

* 920 * 940 * 960 * 980 * 1000
cMID1 : CAGATTGCCAAACTGCAAAACAGTGCATTCGAGCGCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGAGAACCATCACCTCTCGCTTCTCCTGCAAA : 991
fMID : CAGATAGCCGGCTGCAAGCAGTGCATCGAGAGGTCCTCTTCCCTCATCAACAAGCCGACCAAGGCTCAAGGAGGCGGACCACTCGTTTCCCTCCAAA : 997
hMID2 : CAGGTTGCTAATTTGCCGCAAGTGTCTTGAACGGTCAACAGTCTCATCAACCAAGCTGAGCATATCCTGAAGGAAAATGACCAGGCCACGGTTTCTACAGT : 991
hMID1 : CAGATTGCCAAACTGCAAAACAGTGCATTCGAGCGCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGAGAACCATCACCTCTCGCTTCTCCTACAGA : 991
mMID2 : CAGGTTGCTAATTTGCCGCAAGTGTCTTGAACGGTCAACAGTCTCATCAACCAAGCTGAGCATATCCTGAAGGAAAATGACCAGGCCACGGTTTCTGCACT : 991
mMID1 : CAGATTGCCAAACTGTAACAAGTGCATTCGAGCGCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGAGAACCATCACCTCTCGCTTCTCCTACAGA : 991
spretusMID : CAGATTGCCAAACTGTAACAAGTGCATTCGAGCGCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGAAAATGACCAGGCCCGTTTCTACAGA : 991
rMID1 : CAGATTGCCAAACTGTAACAAGTGCATTCGAGCGCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTGAAGGAGAACCATCACCTCTCGCTTCTCCTACAGA : 991

```

```

*      1020      *      1040      *      1060      *      1080      *      1100
cMID1 : CTGCTAAAAACATCACCGAAAGGGTTCCATGGCAACTGCATCTCCAGGTTCTAATTCCCTGAAATTAATCTCAACGATACTTTTGATACTTTTCGCACT : 1091
fMID  : CTGCTAAAAGCATCTGTGAGAGGGTTTCCATGGCAACAGCATCTTCACAAGTGTGGTTACCAGAGATAAACTTGAATGACACATTTTGATACTTTTGCTCT : 1097
hMID2 : CTGCTAAAAAATTTCGAGAGGGTCCGCTATGGCAACTGCATCTTCTCAAGTTCTGATTCCAGACATCAATTTTAATGATGCCTTTGAAAACTTTGCTTT : 1091
hMID1 : CTGCTAAGAAATCACCGAGAGAGTCTCCATGGCAACTGCATCTCCAGGTTCTAATTCCCTGAAATCAAACTCAATGACACATTTTGACACTTTTGCCCT : 1091
mMID2 : CTGCTAAAAACATTCGCGAGAGGGTCCGCTATGGCAACTGCATCTCCAGGTTCTAATTCCCGAAATCAAACTCAATGACACATTTTGACACTTTTGCCCT : 1091
mMID1 : CAGCAAAGAAATCACAGAGAGTCTCCATGGCAACTGCATCTCCAGGTTCTAATTCCCTGAAATCAAACTCAATGACACATTTTGACACTTTTGCCCT : 1091
spretusMID : CAGCAAAGAAATCACAGAGAGTCTCCATGGCAACTGCATCTCCAGGTTCTAATTCCCTGAAATCAAACTCAATGACACATTTTGACACTTTTGCCCT : 1091
rMID1 : CAGCAAAGAAATCACAGAGAGTCTCCATGGCAACTGCATCTCCAGGTTCTAATTCCCTGAAATCAAACTCAATGACACATTTTGACACTTTTGCCCT : 1091

```

```

*      1120      *      1140      *      1160      *      1180      *      1200
cMID1 : TGATTTTACCAGCGGAGAAGAAAATGTTGGAAATGCCCTTGATTATCTTACAGCTCCCAACCCCTCCACCATTCCGAGAAGAGCTCTGTACAGCTTCTTATGAT : 1191
fMID  : GGATTTTACACCGGAGAAGAAAATGCTTGAAAATTTAGATTACCTCACAGCACCGAATCCTCCAGTAATCCGCGAGGAGCTGTGCACAGCTTCGTACGAC : 1197
hMID2 : AGATTTTCCACAGAAAAGAAAAGAACTGCTAGAGGGGTAGATTATTTAACAGCCCAAAACCCACCATTCTATCCGAGAAGAACTCTGTACTGCCCTCCATGAC : 1191
hMID1 : AGATTTTCCCGAGAGAAGAAAAGAACTGCTAGAAATGCTTGGAATACCTTACAGCTCCCAACCCCTCCACAATTAGAGAAGAGCTCTGTACAGCTTCATATGAC : 1191
mMID2 : AGATTTTCTAGAGAAAAGAAAAGAACTATTTGGAGGGGCTAGACTATTTAACAGCTCCCAAAACCCACCATTCTATCCGAGAAGAGCTCTGTACTGCTTCCATGAC : 1191
mMID1 : GGATTTTCCCGGAGAAGAAAAGAACTGCTAGAAATGCTTGGAATACCTTACAGCTCCCAACCCCTCCACAATTAGAGAAGAGCTCTGTACAGCTTCCATGAC : 1191
spretusMID : GGATTTTCCCGGAGAAGAAAAGAACTGCTAGAAATGCTTGGAATACCTTACAGCTCCCAACCCCTCCACAATTAGAGAAGAGCTCTGTACAGCTTCCATGAC : 1191
rMID1 : AGATTTTCCCGGAGAAGAAAAGAACTGCTAGAAATGCTTGGAATACCTTACAGCTCCCAACCCCTCCACAATTAGAGAAGAGCTCTGTACAGCTTCCATGAC : 1191

```

```

*      1220      *      1240      *      1260      *      1280      *      1300
cMID1 : ACTAATTAAGTTCACACTGGACATCAGACGAGAGTTTCAGCGTGGTCTCTTACGACTTCAGTACACCATTCTCACAGGACAGCTAAACGTTGTTA----- : 1285
fMID  : ACCATCACACTTCACTGGACCTCGGACGACGAGTTTCACCGTGGTGTCTATAAGACTTCAGTACCCATCTTCACCTGCCAGTCCAAACCTCTCA----- : 1291
hMID2 : ACCATTAAGTTCACACTGGACATCAGACGAGAGTTTCAGCGTGGTCTCTTACGACTTCAGTACACCATTCTCACAGGACAGCTAAACGTTGTTA----- : 1285
hMID1 : ACCATCACTGTGCATTTGGACCTCCGATGAGAGTTTCAGCGTGGTCTCTTACGACTTCAGTACACCATTCTCACAGGACAGCTAAACGTTGTTA----- : 1286
mMID2 : ACCATTAAGTTCACACTGGACATCAGACGAGAGTTTCAGCGTGGTCTCTTACGACTTCAGTACACCATTCTCACAGGACAGCTAAACGTTGTTA----- : 1291
mMID1 : ACCATCACCTCCACTGGACCTCAGAGGACGAGTTTCAGCGTGGTCTCTTACGACTTCAGTACACCATTCTCACAGGACAGCTAAACGTTGTTA----- : 1285
spretusMID : ACCATCACTGTGCATTTGGACCTCCGATGAGAGTTTCAGCGTGGTCTCTTACGACTTCAGTACACCATTCTCACAGGACAGCTAAACGTTGTTA----- : 1285
rMID1 : ACCATCACTGTCCACTGGACTTCGATGAGAGTTTCAGTGTGGTCTCTGATGAACTTCAGTACACCATTCTCACAGGACAGCTAAACGTTGTTA----- : 1285

```

```

*      1320      *      1340      *      1360      *      1380      *      1400
cMID1 : -----GTTTATGTAACTCAGC : 1301
fMID  : -----GTTTGTGCAACTCTGC : 1307
hMID2 : CATGGTGTAGTTGGGGCCTGTGGCCAGAGATAAGGAAATGTAAGGAAGCAGTAAGCTGCTCAAGATTGGCCGGGGCGCCACGAGGCTGATAAATTCAGT : 1391

```

```

hMID1      : -----GTCTGTGTAATTCGGC : 1301
mMID2      : -----CCTGTACAATTCAGT : 1301
mMID1      : CCTGTGACGGCACCTGTCT-----CCTCGGCTCCGCAGTCTGTGTAATTCGGC : 1340
spretusMID : -----GTCTATGTAATTCAGC : 1301
rMID1      : -----GTCTTTGTAATTCGGC : 1301

```

```

          *      1420          *      1440          *      1460          *      1480          *      1500
cMID1      : CGACAGCTGGATGATTGTTCCCAATATCAAAACAAAACCACTACACCGTGCATGGGTTACACAGTGGCACAAAGTACATCTTTCATTTGTTAAAGGCTATTAAT : 1401
fMID       : CGACAGCTGGATGATTGTACCAACATCAAGCAGAACCACTACACTGTGCACGGGCTCCAGTGCAGAACCAAGTACATCTTTTATTGTGAAAGCCATAAAT : 1407
hMID2      : AGACAGCTGGATGATTGTTCCCAACATTAAGCAGAACCACTACACAGTGCATGGACTCCAAGCGGGACTCGCTACATCTTTCATCGTTAAAGCCATAAAT : 1491
hMID1      : TGATAGCTGGATGATAGTACCAACATCAAGCAGAACCACTACACCGTGCACGGTGTGCACAGCGGACCAAGTACATCTTTCATCGTTAAAGCCATAAAT : 1401
mMID2      : GGATAGCTGGATGATTGTGCCAACATTAAGCAGAACCACTACACAGTTCATGGACTCCAGAGTGGGACACGCTACATCTTTTATCGTGAAAGCCATAAAT : 1440
mMID1      : GGAAGCTGGATGATCGTGCACCAACATCAAGCAGAACCACTACACTGTGCATGGCTTACAAAGTGGCACCAAGTATATCTTTCATCGTTAAAGCCATCAAC : 1401
spretusMID : GGAAGCTGGATGATTGTGCCAACATCAAGCAGAACCACTACACTGTGCATGGCTTACAAAGTGGCACCAAGTATATCTTTCATCGTTAAAGCCATCAAC : 1401
rMID1      : AGATAGCTGGATGATTGTGCCAACATCAAGCAGAACCACTATACTGTGCATGGCTTACAAAGTGGCACCAAGTATATCTTTCATCGTTAAAGCCATCAAC : 1401

```

```

          *      1520          *      1540          *      1560          *      1580          *      1600
cMID1      : CAGGCTGGCAGCAGAAGCAGCGAGCCAGGCAAGCTCAAGACAAACAGTCAACCAATTTAAACTGGACCCCAAAATCTGCTCATAGAAAAGCTGAAAGTGTCTC : 1501
fMID       : CAGGCGGGAAACCGCAGCAGCGAACCCAGCAAACTCAAAACTAACAGTCAACCCCTTCAGGTGGACCCCAAGTCAGCTCACCGAAAGCTGAGGGTGTCTC : 1507
hMID2      : CAAGCCGGCAGCCGGAAACAGTGAACCTACCCGACTAAAAACAACAGCCAAACCCCTTTAAATTTGGATCCCAAAATGACTCAAAAGAGTTGAAGATCTCCA : 1591
hMID1      : CAGGCGGGCAGCCCGACAGTGAACCTGGGAACTTGAACACAAACAGCCAAACCAATTTAAACTGGATCCCAAAATCTGCTCATAGAAAAGCTGAAAGTGTCCC : 1501
mMID2      : CAAGCAGGCAGCCGGAAACAGCGAACCCACCAGACTAAAAACGAACAGCCAAACCAATTTAAACTAGATCCCAAAATGACTCAAAAGAGTTGAAGATCTCCA : 1501
mMID1      : CAGGCGGGCAGCCGTAGCAGCGAACCCGAAACCTGAAACCAACAGTCAACCCCTTTAGACTGGATCCCAAAATCGGCTCATAGCAAAGCTGAAAGTGTCCC : 1540
spretusMID : CAGGCGAGCCAGCCGTAGCAGTGAACCTGGGAACTGAAACCAACAGTCAACCAATTTAAACTGGATCCCAAAATCTGCTCATAGAAAAGCTGAAAGTGTCCC : 1501
rMID1      : CAGGCGAGCCAGCCCGACAGTGAACCTGGGAACTGAAACCAACAGTCAACCAATTTAAACTGGATCCCAAAATCTGCTCATAGAAAAGCTGAAAGTGTCCC : 1501

```

```

          *      1620          *      1640          *      1660          *      1680          *      1700
cMID1      : ATGACAACCTGACAGTGGAACTGATGAAACCTCTCTCAAAAAGAGTCATACACAGAGCGGATTCACAGCCCAAGGGAGCTACGGAGTACCTGGCAATGT : 1601
fMID       : ACGACAACCTGACAGTGGAGAGGAGCAGAGCTCGGCCAAAAAGACCCACAGTCAGGACCGATTCACAGGCCACAGAGCTACGGCGTGAAGGCAAACTG : 1607
hMID2      : ATGATGGATTCAGATGGAGAAGGATGAAAGCTCTCTAAAGAAGAGCCACACCCAGAGAGGTTAGTGGCACAGGGTGTATGGGGCAGCAGGAAATAT : 1691
hMID1      : ATGATAACTTGACACTAGAACTGATGAGTCAATCAAGAAGAGTCACACACCTGAACCGTTCAACCCAGGGGACCTATGGAGTACCTGGAAATGT : 1601
mMID2      : ATGATGGATTCAGATGGAGAAGGATGAGAGCTCTCTCAACAAGAGTCACAGCCCGGAGCGCTTCGCTGGTCAAGGGGAGCTACGGAGTGGCTGGCAACCT : 1640
mMID1      : ACGACAACCTGACTGTTGAACTGATGAATCATCACTAAAAAGAGTCACACACCTGAACCGTTCACTAGTCAAGGGAGCTATGGAGTACCTGGAAATGT : 1601
spretusMID : ATGACAACCTGACTGTTGAACTGATGAATCATCACTAAAAAGAGTCACACACCTGAACCGTTCACTAGTCAAGGGAGCTATGGAGTACCTGGAAATGT : 1601
rMID1      : ATGACAACCTGACTGTTGAACTGATGAGTCAATCACTAAAAAGAGTCACACACCTGAACCGTTCACTAGTCAAGGGAGCTATGGAGTACCTGGAAATGT : 1601

```

```

*      1720      *      1740      *      1760      *      1780      *      1800
cMID1 : GTTCATGACAGCGGACGCCATFACGGGAAGTGGTTATAAGGGGAGTACATGGTATGCCATTGGTATTTCATACAAGTCAGCACCGAAGCATGAGTGG : 1701
fMID  : CTACATCGACAGCGGGCGCCATFACGGGAAGCTCTGATGGAGCAACACATGGTTTGGCAGTGGGAGTTGCCATACAAGTCGGCCCGGAGACACGAGTGG : 1707
hMID2 : ATTTCATGACAGTGGCTGCCACTATTGGGAAGTGGTCAATGGGTTCCCTCAACATGGTATGCCAATTTGGCATTGCCATACAAATCAGCTCCAAGAATGAATGG : 1791
hMID1 : GTTTATGATAGTGGCCGCCATFATGGGAAGTGGTCAATAGGGGAAGCACATGGTATGCCATTGGCTTTGGTTACAAATCAGCCCGGAAGCATGAATGG : 1701
mMID2 : ATTTCATCGACAGTGGCTGCCACTACTGGGAAGTGGTCAATGGGTTCTTCAACATGGTATGGCTATTGGTATTGCCATACAAATCAGCTCCAAGAATGAGTGG : 1701
mMID1 : GTTCATCGACAGCGGCCGTCACTACTGGGAAGTGGTCAATAGGGGAAGCACATGGTATGCCATTGGCCCTTGGCATACAAATCAGCTCCAAAACATGAATGG : 1740
spretusMID : GTTCATGATAGTGGCCGTCACTACTGGGAAGTGGTCAATAGGGGAAGCACATGGTATGCCATTGGCCCTTGGCATACAAATCAGCTCCAAAACATGAATGG : 1701
rMID1 : GTTTATGACAGTGGCCGTCACTACTGGGAAGTGGTCAATAGGGGAAGCACATGGTATGCCATTGGCCCTTGGCATACAAATCAGCTCCAAAACATGAATGG : 1701

```

```

*      1820      *      1840      *      1860      *      1880      *      1900
cMID1 : ATCGGCAAGAACHCAGCCTCCCTGGGTGCTGTCCCGCTGCAACAACCAATGGGTGGTGAGACACAAACAGCAAAAGAAATCCCAATAGAGCCCTCCGCTCATC : 1801
fMID  : GTGGGCAAGAACHCGCCCTCCCTGGGTGCTGTCCAGATGCCAACAACCTCGTGGGTGGCCGCCACAAACAGCAAGGAGATGCCCGTGGAGCCCCGCCACACC : 1807
hMID2 : ATGGCAAGAATGCCCTCCCTGAGGTCTTCTCTGGCTGCAATAGTAACCTTCGTTGGTGAGACACAACAACAAGGAAATGCTGGTGGATGTGCCCCCCACACC : 1891
hMID1 : ATGGCAAGAACHCTGCTTCCCTGGGCGCTCTCCCGTGCACAACAACCTGGGTGGTGAGACACAATAGCAAGGAAATCCCAATGAGCCCTCCGCTCATC : 1801
mMID2 : ATGGCAAGAATGCCCTCTCCCTGGGTCTTCTCTCGATGCAACAGTAACCTTCGTTGGTTCGGCACAACAACAAGGAAATGCTGGTGGATGTGCCCCCCACAGT : 1801
mMID1 : ATCGGCAAGAACCGGCCCTCCCTGGGCGCTCTCCCGTGCACAACAACCTGGGCGGTCCGACACGACGGCAAGGACACCCCAATCCGCGCGGCCCTCACC : 1840
spretusMID : ATGGCAAGAATCTGCTTCCCTGGGCGCTCTCCCGTGCACAACAACCTGGGTGGTGAGACATAACAGCAAGGAAATCCCAATGAGCCAGCCCCCTCACC : 1801
rMID1 : ATGGCAAGAATCTGCTTCCCTGGGCGCTTCTCCCGTGCACAACAACCTGGGTGGTGAGACATAACAGCAAGGAAATCCCAATGAGCCAGCCCCCACC : 1801

```

```

*      1920      *      1940      *      1960      *      1980      *      2000
cMID1 : TCCGCGTGTCGGCAATTTTGGCTGGACTACGACAATGGTTCCCTTCCCTTTTATGATGCCTTGAACTCCTTGACCTTTACAGCTTTTGACATTACGTTTGG : 1901
fMID  : TGCCAGCCCTGGGAATATTGTTGGACTACGATCCGGATCTCTGTCTTTCTACGACCCTGTGAGCTCGCAGCACTGTACAGCTTCGACCTCGCTTTTGG : 1907
hMID2 : TGAAGCGTCTGGGTGTCCCTCCCTGGATTATGACAAACAATATGCTGTCTTTCTATGACCCAGCTAACTCTCTCCATCTTCATACCTTTGATGTGACCTTCAT : 1991
hMID1 : TCCGCGCGTGGGCAATCCCTGGCTGGACTATGATAAAGCTCTATCCCTTTTATGATECTTTGAACCTCATCCACCTCTAACCTTCGACCTCGCATTTGG : 1901
mMID2 : TGAAGCGTCTAGGTGTCCCTCTGGATTATGACAAACAATGCTGTCTTTCTATGACCCAGCTAACTCTCTCCATCTTCATACCTTTTATGATGTGACCTTCAT : 1901
mMID1 : TGAAGCGCGTGGGCTCCCTGCTGGACTACGACAACAAGCTCCATCCCTTTCTACGACCTCTGAGCTCTGTCACCTCCAGACCTTCCACCGCGCGCTCCG : 1940
spretusMID : TGAAGCGCTAGGCAATCCCTGGCTGGACTATGATAAAGCTCAATCCCTTTTATGATGCTTTGAACCTCATCCACCTCTAACCTTTGATGTGGCGCTTGG : 1901
rMID1 : CCCGCGTGTCGGCAATTTTGGCTGGACTATGATAAAGCTCAATCCCTTTTATGATGCTTTGAACCTCATCCACCTCTATACCTTCGATGTGGCGCTTGG : 1901

```

```

*      2020      *      2040      *      2060      *      2080      *      2100
cMID1 : GCAGCCCGTCTGCCCACATTCACCTGTGTGGAAATAGTGTTTGACCATTATAACAAGCTTGCCCATACCTGATCACCTAGACTCCTCTGAGCAGCTGGCA : 2001
fMID  : TCACCCGGTCTGTCTCTGTGTTTAAAGCTGTGGAACAGATCTCTGACAATCCTCACCAGACTCCCATCCAGACCCACTTAGAGGGGACAGACTACAACAAG : 2007
hMID2 : TCTTCCAGTTTGTCCAACATTTACAATCTGGAACAATTCCTAATGATCCTGTCTGGCTTGGCTGCCCCAGATTTTATTTGATTACCCTGAGCCGGCAGGAA : 2091
hMID1 : GCAGCCTGTTTGGCCCACCTTCACCGTGTGGAAACAAGTGTCTGACCATTATCACCTGGCTCCCTATCCCAAGCCATTTGGAATGCACAGAGCAGCTGGCG : 2001

```

```

mMID2 : TCTTCCAGTTTGTCCACATTCACAATCTGGAAOAAATCCCTAATGATTCCTTCTGGCTTGCCTGCCCTGATTTTATTGATTATCCTGACCGCCAGGAA : 2001
mMID1 : CCACCCCGTGTGCCCACCTTCACCGTGTGGAAOAAAGTGTCTGACCATCGTCCGGGTCTGCCATCCCGGACCATCTGGACTGTACGGAGCAGCGACCT : 2040
spretusMID : CCACCGTGTGTGCCCACCTTACTGTGTGGAAOAAAGTGTCTGACCATTATAACTGGTCTTCCATCCAGACCATTTGGACTGTACAGAACAGCTACCT : 2001
rMID1 : CCACCCGTGTGTGCCCACCTTACTGTGTGGAAOAAAGTGTCTGACCATTATAACTGGTCTTCCATCCAGACCATTTGGACTGTACAGAACAGCTACCT : 2001

```

```

                *      2120          *      2140          *
cMID1 : TGA----- : 2004
fMID : TGA----- : 2010
hMID2 : TCCAACTGCAGGCCTCAAGAATCCCCTTATGTTTCTGGGATGAAAACCTGTCATTAA : 2148
hMID1 : TGA----- : 2004
mMID2 : TCCAACTGCAGGCCTCAAGAATCCCCTTATGTGTGTCAGGGATGAAAGCTTGCCATTAA : 2058
mMID1 : TGA----- : 2043
spretusMID : TGA----- : 2004
rMID1 : TGA----- : 2004

```

Appendix 5: The nucleotide and protein sequence for chick MID1.

1	M E T L E S E L T C P I C L E L F E D P	20
1	ATGGAACACTGGAGTCGGAACGACCTGCCCTATCTGTCTCGAGCTGTTTGAGGACCCG	60
21	L L L P C A H S L C F N C A H R I L V S	40
61	CTGCTGCTGCCCTGCGCTCACAGTCTCTGCTTCAACTGCGCGCACCCGCATCCTGGTCTCC	120
41	H C A S N E P V E S I T A F Q C P T C R	60
121	CACTGCGCCTCCAACGAGCCGGTGGAGTCTATCACCGCCTTCCAGTGCCCCACCTGCCGC	180
61	Y V I T L S Q R G L E G L K R N V T L Q	80
181	TATGTCATCACCCCTCAGCCAGCGCGTCTAGAGGGGCTCAAGCGCAACGTCAACCCCTGCAG	240
81	N I I D R F Q K A S V S G P N S P S E T	100
241	AACATCATCGACCGCTTTCAGAAAGCCTCGGTGAGCGGGCCCAACTCCCCAGCGAGACC	300
101	R R E R A F D S N S M S S C E K V L C Q	120
301	CGTCGGGAGCGGGCATTTCGACAGCAACAGCATGTCTCCTGCGAGAAGGTCTCTGCCAG	360
121	F C D Q D P A Q E A V K T C V T C E V S	140
361	TTCTGCGACCAGGACCCTGCCAGGAGGCAGTGAAGACGTGCGTGACCTGCGAGGTCTCC	420
141	Y C E E C L K A T H P N K K P F T G H R	160
421	TACTGCGAGGAGTGCCTGAAAGCCACGCACCCCAACAAGAAGCCTTTCACCGGCCATCGG	480
161	L I E P I P D S H I R G L M C L E H E D	180
481	CTGATCGAGCCCATCCCGGACTCTCACATCAGGGGATTAATGTGCTTGGAGCACGAGGAC	540
181	E K V N M Y C V T D D Q L I C A L C K L	200
541	GAGAAGGTGAACATGTACTGCGTGACAGATGACCAGCTGATCTGCGCCCTGTGCAAGCTG	600
201	V G R H R D H Q V A A L S E R Y D K L K	220
601	GTCGGGCGGCACCGGGACCATCAAGTGGCAGCTCTAAGCGAGCGCTACGACAAGCTGAAG	660
221	Q N L E S N L T N L I K R N T E L E T L	240
661	CAAAATTTGGAGAGTAACCTCACCAACCTTATTAAGAGGAATACTGAACTGGAAACTCTT	720
241	L A K L I Q T C Q H V E V N A S R Q E T	260
721	CTGGCAAAACTCATTCAGACCTGTCAACATGTAGAAGTAAATGCATCTCGCCAAGAAACC	780
261	K L M E E C D Q L I E I I Q Q R R Q I I	280
781	AAGCTGATGGAAGAATGTGACCAGCTCATTGAAATAATACAACAGAGACGACAAATAAT	840
281	G T K I K E G K V V R L R K L A Q Q I A	300
841	GGAACCAAAATCAAGGAAGGAAAGGTGGTGGAGTTGAGAAAACCTGGCTCAGCAGATTGCA	900
301	N C K Q C I E R S T S L I S Q A E Q S L	320
901	AACTGCAACAGTGCATCGAGCGCTCGACATCCCTCATCTCTCAGGCTGAGCAGTCACTG	960
321	K E N D H A R F L Q T A K N I T E R V S	340
961	AAGGAGAACGATCACGCTCGCTTCTGCAAACTGCTAAAAACATCACCGAAAGGGTTTCC	1020
341	M A T A S S Q V L I P E I N L N D T F D	360
1021	ATGGCAACTGCATCCTCCCAGGTTCTAATTCCTGAAATTAATCTCAACGATACTTTTGAT	1080
361	T F A L D F T R E K K L L E C L D Y L T	380
1081	ACTTTCGCACTTGATTTTACCAGGGAGAAGAAATTGTTGGAATGCCTTGATTATCTTACA	1140
381	A P N P P T I R E E L C T A S Y D T I T	400
1141	GCTCCCAACCTCCCACCATTTCGAGAAGAGCTCTGTACAGCTTCTTATGATACTATTACT	1200
401	V H W T S D D E F S V V S Y E L Q Y T I	420
1201	GTCCACTGGACATCAGACGATGAGTTCAGCGTGGTCTCTTACGAGCTGCAGTACACCATC	1260

421	F T G Q A N V V S L C N S A D S W M I V	440
1261	TTCACTGGACAAGCTAACGTTGTTAGTTTATGTAACCTCAGCCGACAGCTGGATGATTGTT	1320
441	P N I K Q N H Y T V H G L Q S G T K Y I	460
1321	CCCAATATCAAACAAAACCACTACACCGTGCATGGGTTACAGAGTGGCACAAGTACATC	1380
461	F I V K A I N Q A G S R S S E P G K L K	480
1381	TTCATTGTTAAGGCTATTAATCAGGCTGGCAGCAGAAGCAGCGAGCCAGGCAAGCTCAAG	1440
481	T N S Q P F K L D P K S A H R K L K V S	500
1441	ACAAACAGTCAGCCATTTAAACTGGACCCCAAATCTGCTCATAGAAAGCTGAAAGTGTCT	1500
501	H D N L T V E R D E T S S K K S H T P E	520
1501	CATGACAACCTTGACAGTGGAACGTGATGAAACCTCCTCCAAAAAGAGTCATACACCAGAG	1560
521	R F T S Q G S Y G V A G N V F I D S G R	540
1561	CGATTACAGAGCCAAGGGAGCTACGGAGTAGCTGGCAATGTGTTTATTGACAGCGGACGG	1620
541	H Y W E V V I S G S T W Y A I G I S Y K	560
1621	CATTACTGGGAAGTGGTTATAAGTGGGAGTACATGGTATGCCATTGGTATTTTCATACAAG	1680
561	S A P K H E W I G K N S A S W V L C R C	580
1681	TCAGCACCGAAGCATGAGTGGATCGGGAAGAACTCAGCCTCCTGGGTGCTGTGCCGCTGC	1740
581	N N T W V V R H N S K E I P I E P A P H	600
1741	AACAACACATGGGTGGTGGAGACACAACAGCAAAGAAATCCCAATAGAGCCTGCGCCTCAT	1800
601	L R R V G I L L D Y D N G S L A F Y D A	620
1801	CTCCGCCGTGTCGGCATTGCTGGACTACGACAATGGTTCCCTTGCCTTTATGATGCC	1860
621	L N S L H L Y T F D I T F G Q P V C P T	640
1861	TTGAACTCCCTGCACCTTTACACCTTTGACATTACGTTTGGGCAGCCCGTCTGCCCCACA	1920
641	F T V W N K C L T I I T G L P I P D H L	660
1921	TTCAGTGTGTTGGAATAAGTGTGTTGACATTATAACAGGCTTGCCCATACCTGATCACTTA	1980
661	D S S E Q L A *	668
1981	GACTCCTCTGAGCAGCTGGCATGA	2004

Appendix 6: The nucleotide and protein sequence for fugu MID1.

1	M E T L E S E L T C P I C L E L F E D P	20
1	ATGGAACGCTGGAGTCGGAGCTGACCTGTCCAATCTGTCTGGAGCTATTTGAGGACCCG	60
21	L L L P C A H S L C F N C A H R I L V S	40
61	CTGCTCTTGCCCTGCGCTCACAGCCTTTGTTTCAACTGCGCCACCGCATCCTGGTGTCA	120
41	H C T P S E P I Q S I S A F Q C P T C R	60
121	CACTGCACGCCCAGCGAGCCCATCCAATCCATCAGCGCCTTCCAGTGCCCGACCTGCCGC	180
61	Y V I T L N Q R G L E G L K R N V T L Q	80
181	TATGTCATCACCCCTCAACCAGAGGGGCTAGAGGGACTCAAACGCAACGTTACGCTGCAG	240
81	N I I D R Y Q K A S Q S G P N S P K E T	100
241	AACATCATTGACCGTTACCAGAAGGCATCGCAAAGCGGACCCAACTCTCCCAAAGAGACG	300
101	R R E G A V P D S R A M T S P G D R V P	120
301	CGGCGCGAGGGAGCCGTCCCCGACAGCAGAGCCATGACGTACCCGGCGACAGGGTGCCG	360
121	C Q F C E Q D P P Q D A V K T C I T C E	140
361	TGTCAGTTCTGTGAGCAGGATCCTCCTCAGGATGCCGTGAAGACCTGCATCACCTGCGAG	420
141	V S Y C D E C L K A T H P N K K P F T G	160
421	GTGTCCTACTGCGACGAGTGTCTCAAGGCCACCCACCCCAACAAGAAGCCGTTACGGGT	480
161	H R L M E P L L D S H L R G I M C A E H	180
481	CACCGCCTCATGGAGCCTCTGCTGGACTCCCATCTGCGGGGATAATGTGCGCGGAGCAC	540
181	E D E K V N M Y C V T D E Q L I C A L C	200
541	GAGGACGAGAAGGTCAACATGTACTGTGTGACCGACGAACAATTGATCTGCGCATTGTGC	600
201	K L V G R H R D H H V A A L G D R Y D K	220
601	AAGCTGGTTGGTCGACACAGGGACCATCACGTGGCGGCCCTCGCGGATCGATACGACAAA	660
221	L K E S L E S N L N N L I K R T S D L E	240
661	CTCAAGGAATCCTTGAATCTAACCTCAACAATCTAATCAAGAGGACCAGTGATTTGGAA	720
241	S L M G K L I Q T C Q H V E V N A S R Q	260
721	AGTCTGATGGGTAAACTTATTCAAACCTGCCAGCACGTGAGGTAATGCATCCAGACAG	780
261	E N K L L E E C D L L I N I I Q Q R R Q	280
781	GAAAACAAGCTGCTGGAGGAGTGTGACCTGCTCATCAACATCATACAGCAGCGAAGACAA	840
281	I I T T K I K E G K A V R L R K L A Q Q	300
841	ATAATAACCACCAAAATAAAAGAAGGAAAGGCCGTGCGGCTGAGGAAGCTCGCCAGCAG	900
301	I A G C K Q C I E R S S S L I T Q A D Q	320
901	ATAGCCGGCTGCAAGCAGTGCATCGAGAGTCTCTTCCCTCATCACACAAGCCGACCAG	960
321	A L K E A D H T R F L Q T A K S I C E R	340
961	GCGCTCAAGGAGGCGGACCACACTCGTTTCCCTCCAAACTGCTAAAAGCATCTGTGAGAGG	1020
341	V S M A T A S S Q V L L P E I N L N D T	360
1021	GTTTCCATGGCAACAGCATCTTCACAAGTGTTGTTACCAGAGATAAACTTGAATGACACA	1080
361	F D T F A L D F T R E K K M L E N L D Y	380
1081	TTTGATACTTTTGCTCTGGATTTACACGGGAGAAGAAAATGCTTGAAAATTTAGATTAC	1140
381	L T A P N P P V I R E E L C T A S Y D T	400
1141	CTCACAGCACCGAATCCTCCAGTAATCCGCGAGGAGCTGTGCACAGCTTCGTACGACACC	1200
401	I T V H W T S D D E F T V V S Y E L Q Y	420
1201	ATCACAGTTCCTGACGTCGGACGACGAGTTCACCGTGGTGTCTATGAACCTCAGTAC	1260

421	A I F T C Q S N V V S L C N S A D S W M	440
1261	GCCATCTTTCACCTGCCAGTCCAACGTCGTACGTTTGTGCAACTCTGCCGACAGCTGGATG	1320
441	I V P N I K Q N H Y T V H G L Q C G T K	460
1321	ATTGTACCAAACATCAAGCAGAACCACTACACTGTGCACGGGCTCCAGTGCGGAACCAAG	1380
461	Y I F I V K A I N Q A G N R S S E P A K	480
1381	TACATCTTTATTGTGAAAGCCATAAACCAGGCGGGAAACCGCAGCAGCGAACCAGCAAAA	1440
481	L K T N S Q P F R L D P K S A H R K L R	500
1441	CTCAAAACTAACAGTCAGCCCTTCAGGTTGGACCCAAAGTCAGCTACCGAAAGCTGAGG	1500
501	V S H D N L T V E R D E T S A K K S H S	520
1501	GTGTCTCAGACAACCTGACGGTGGAGAGGGACGAGACGTTCGGCCAAAAAGAGCCACAGT	1560
521	Q D R F T S H S S Y G V T G N V Y I D S	540
1561	CAGGACCGATTCACGAGCCACAGCAGCTACGGCGTGACGGGAAACGTCTACATCGACAGC	1620
541	G R H Y W E A L I G G S T W F A V G V A	560
1621	GGGCGCCATTACTGGGAAGCTCTGATTGGAGGAAGCACATGGTTTTGCAGTGGGAGTTGCA	1680
561	Y K S A P R H E W V G K N S A S W V L S	580
1681	TACAAGTCGGCGCCGAGACACGAGTGGGTGGGCAAAAACCTCGGCCTCCTGGGTGCTGTCC	1740
581	R C N N S W V A R H N S K E M P V E P P	600
1741	AGATGCAACAACCTCGTGGGTGGCGGCCACAACAGCAAGGAGATGCCCCGTGGAGCCCCCG	1800
601	P H L R R L G I L L D Y D S G S L S F Y	620
1801	CCACACCTGCGACGCCTGGGAATATTGTTGGACTACGACTCCGGATCTCTGTCTTTCTAC	1860
621	D A V S S Q H L Y T F D V A F A H P V C	640
1861	GACGCTGTGAGCTCGCAGCACTTGTACACGTTTCGACGTCGCTTTTGCTCACCCGGTCTGT	1920
641	P V F N V W N R C L T I L T G L P I P D	660
1921	CCTGTGTTTAACTGTGGAACAGATGTCTGACAATCCTCACCGGACTCCCCATCCCAGAC	1980
661	H L E G T D Y N K *	670
1981	CACTTAGAGGGGACAGACTACAACAAGTGA	2010