# STRUCTURE FUNCTION ANALYSIS OF THE DEUBIQUITYLATING ENZYME FAM

**Poon-Yu Khut, B.Sc**

**School of Molecular & Biomedical Science (Biochemistry)**
**University of Adelaide**
**Adelaide, South Australia 5005**

**October 2006**

# CHAPTER 1

# INTRODUCTION

## *Ubiquitin: a Multipurpose Tag*

Ubiquitin is a highly conserved 76 amino acid protein that serves in a complex, post-translational modification system that is largely used by the cell to either target a protein for destruction at the proteasome or affect its intracellular trafficking. Although exclusively found in eukaryotes, prokaryotes (that do not have any form of signalling system based on covalent protein-protein attachments) appear to possess clear ancestors of the ubiquitin fold (reviewed in Hochstrasser 2000). Ubiquitylation is the process by which ubiquitin is covalently attached to a specific target protein via an isopeptide bond between the C-terminal glycine of ubiquitin to a ε-amino group of a substrate's lysine residue. The protein can remain with only one attached ubiquityl moiety (monoubiquitylation), have multiple single ubiquityl moieties attached to several substrate lysine residues (multiubiquitylation), or through successive ubiquitylation reactions form a chain of ubiquitins (polyubiquitylation). As ubiquitin itself contains several lysine residues, it is possible to form a variety of polyubiquityl chains based on which lysine residue the subsequent ubiquitin is attached to. It has become apparent that the number and nature of these ubiquitin attachments is paramount in affecting the appropriate, and very different cellular outcomes.

### UBIQUITIN AND PROTEIN DEGRADATION

Ubiquitin's cellular role has been best characterised in protein turnover where addition of a polyubiquityl chain serves as a destruction tag, targeting the protein to the proteasome where it is rapidly and irreversibly degraded (*figure 1.1*). To form an effective substrate for proteasome recognition, proteins must have multiple ubiquitin moieties attached to it (polyubiquitylation) via gly76-lys48 linkages (reviewed in Coux *et al.* 1996). *In vitro* binding studies with Rpn10, a polyubiquityl receptor subunit of the proteasome, showed that only ubiquitin chains of at least four moieties were efficiently recognised, and binding affinity increased with length. Indeed, binding affinity to proteasomes increases more than

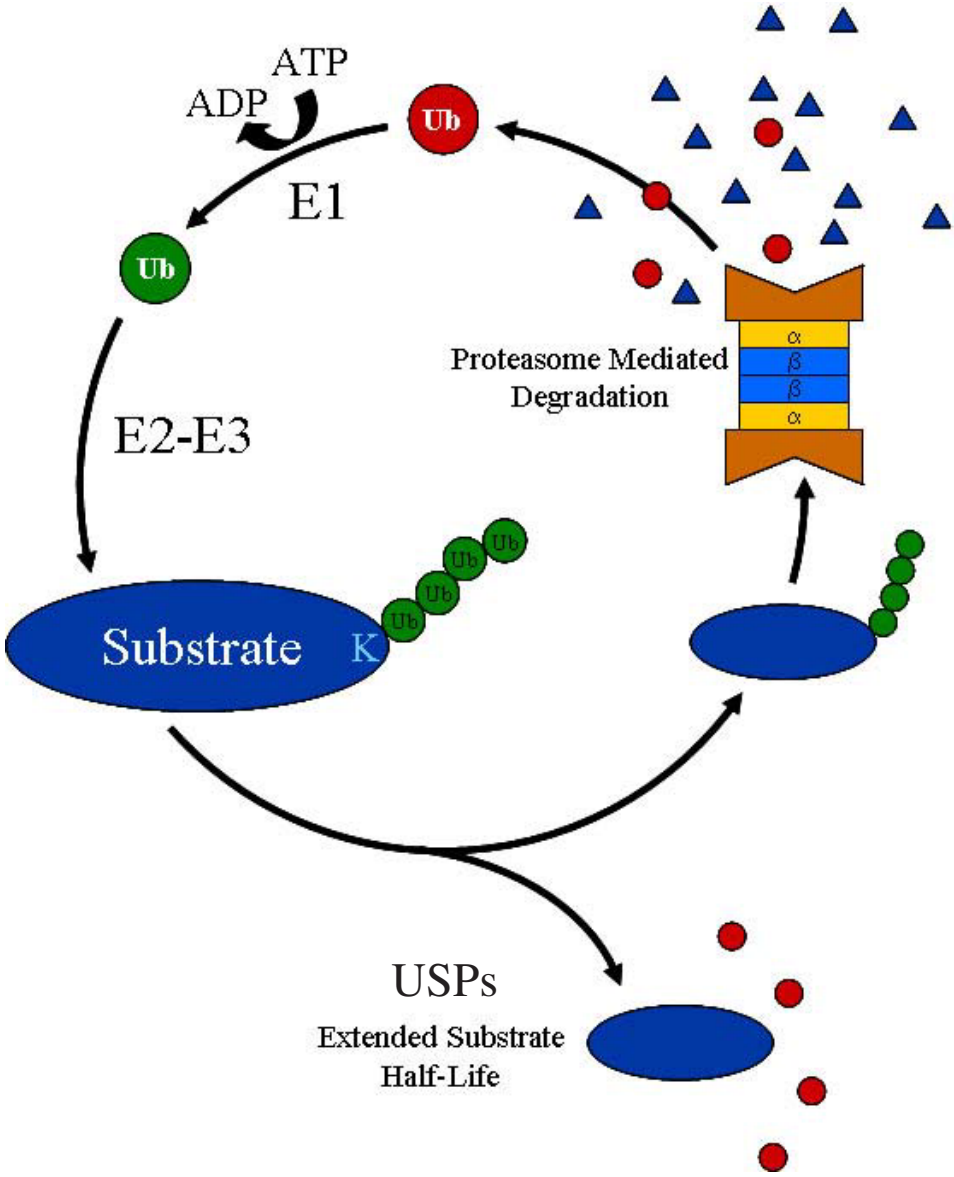



**Figure 1.1**

**The Ubiquitin Proteasome Pathway**

Ubiquitin (red) is activated by the ubiquitin-activating enzyme, E1 coupling the hydrolysis of ATP to generate a high energy thiol-ester bond between Gly76 of ubiquitin and a specific cysteine residue of the E1. The activated ubiquitin (green) is specifically attached to an internal lysine (K) residue of substrate in a series of transesterification reactions mediated by E2 and E3 thiol-ester intermediates. Through successive reactions or with the involvement of E4s (not shown), a polyubiquityl chain is formed that targets the protein to the proteasome where it is rapidly and irreversibly degraded. Ubiquitin is recycled during this process. Alternatively, substrates' half-lives can be extended by the intervention of ubiquitin specific proteases (USPs). USPs act upstream of the proteasome and preserve the substrate by the removal of the polyubiquityl destruction tag.

100-fold when the chain is lengthened from two to four ubiquitins, but only 10-fold more when the chain is increased to eight (Deveraux *et al.* 1994).

It is estimated that over 30% of newly synthesised cellular proteins are disposed of without being properly folded i.e., misfolded and/or unassembled, despite lacking mutational or translational error (Schubert *et al.* 2000). Additionally, properly folded proteins often are damaged through various external stresses such as heat, oxidation and ultraviolet damage. Where the chaperone system has either not rescued or acted upon these proteins, the ubiquitin system is largely responsible for their removal and its role in "cellular garbage disposal" has been long established. Indeed failure to remove such proteins can lead to protein aggregation and formation of toxic inclusion bodies. Within the nervous system, defects in the ubiquitin-dependent removal of certain proteins is closely associated with the pathogenesis of several major human neurodegenerative diseases such as Alzheimer's and Parkinson's disease (reviewed in Layfield *et al.* 2001, Tanaka *et al.* 2004)

Extending from initial views that the ubiquitin pathway served only a housekeeping function in the destruction of unwanted proteins, it has since been revealed through various molecular, biochemical, cellular, genetic and clinical studies, that it plays a significant role in a broad array of cellular processes (reviewed in Ciechanover *et al.* 2000). Examples include the regulation of cell cycle, differentiation and development, the cellular response to extracellular effectors and stress, modulation of cell surface receptors and ion channels, DNA repair, regulation of the immune and inflammatory responses and biogenesis of organelles. Due to its speed, permanency and capacity for specificity, ubiquitin mediated protein degradation is almost always featured in regulatory mechanisms involving timing control (reviewed in Hochstrasser 1995). Substrates thus include cell cycle regulators, tumour suppressors and growth modulators, transcriptional activators and inhibitors, cell surface receptors and endoplasmic reticulum proteins (reviewed in Ciechanover *et al.* 2000).

## UBIQUITIN AND TRAFFICKING

In recent years, the significance of the ubiquitin system in cellular functions other than that of pure protein half-life regulation has become better understood. Ubiquitin K63-linked chains are used to regulate processes such as, DNA repair (Spence *et al.* 1995), activation

of certain protein kinases (Wang *et al.* 2001), endocytosis of some plasma membrane proteins (Galan and Haguenauer-Tsapis 1997), as well as being implicated in stress response, mitochondrial DNA inheritance and ribosomal function (reviewed in Pickart 2000). Monoubiquitylation on one or more lysine residues, is known to be involved in histone regulation and retroviral budding, but it is its role in intracellular trafficking and endocytosis that has attracted the most attention (reviewed in Hicke 2001a) (*figure 1.2*).

Ubiquitin can regulate protein transport by either acting as a sorting signal when attached to transmembrane proteins (*cis*-regulation), or by modifying the activity of protein transport machinery (*trans*-regulation). In the endocytic pathway, ubiquitin is responsible for signalling the entry of endocytic cargo into vesicles from the plasma membrane (although not the sole signal in yeast) and/or the late endosome (reviewed in Hicke 2001b). Cell surface proteins that enter the cell at the plasma membrane, bud off into endocytic vesicles and are delivered to the early endosome, where they can either be recycled to the cell surface, or carry on to the late endosome. At the late endosome, proteins arrive from both the endocytic and biosynthetic pathways *en route* to the lysosome where they are degraded. As the late endosome matures, multiple regions of its membrane invaginate and bud off into the lumen of the organelle forming the multivesicular body (MVB). Finally the MVB fuses with the lysosome, unleashing lysosomal proteases in extreme acidity on the MVB's lipids and proteins, resulting in their degradation (reviewed in Gruenberg 2001). Signal transduction receptors represent a major group of proteins that rely on ubiquitin for internalisation and lysosome degradation for signal down-regulation. In yeast, a G protein-coupled receptor (GPCR) involved in determining mating type is rapidly ubiquitylated on its cytoplasmic tail upon pheromone ligand binding. In this case, ubiquitylation serves to trigger its initial internalisation, and also to sort the receptor into the inner vesicles of the MVB for delivery and down-regulation in the lysosome-like vacuole (reviewed in Hicke 1999). Direct ubiquitylation of the target receptor is not always required for its internalisation. In the developing nervous system of drosophila, Roundabout (Robo) is a cell surface receptor that transmits a repulsive signal in axon guidance. Ubiquitylation is responsible for both its internalisation and also the diversion of the newly synthesised Robo from the secretory pathway to the lysosome as a mechanism to decrease the activity of the receptor. Although the receptor is responsive to ubiquitylation, Robo itself is not the direct target. Robo is negatively regulated by the integral membrane

**Figure 1.2**

**Ubiquitin and Protein Transport Regulation**

a) Monoubiquitin and Lys63-linked di-ubiquitin chains serve as internalization signals that can be attached to a plasma membrane protein to trigger its entry into primary endocytic vesicles budding in from the plasma membrane.
b) Monoubiquitin serves as a signal for the entry of transmembrane proteins into multivesicular body (MVB) vesicles.
c) Ubiquitin, ubiquitin-binding proteins, and ubiquitin ligases are important for the budding of enveloped viruses.
d) Polyubiquitylation regulates protein sorting at the trans-Golgi network to the lysosome/vacuole.
e) Ubiquitin modifies and regulates components of the endocytic machinery.

(Adapted from Hicke and Dunn 2003)

protein Commissureless (Comm), which when ubiquitylated, supplies the trafficking signals to the Robo/Comm complex in *trans* (reviewed in Hicke and Dunn 2003).

Clearly the ubiquitin signal has important and diverse roles within the cell and so must be tightly regulated. As with virtually all biological systems, ubiquitylation can be reversed, this process is known as deubiquitylation. Regulation of certain proteins at the fundamental level of ubiquitylation and deubiquitylation can either directly or indirectly, affect cellular processes.

## *Ubiquitylation*

Conjugation of ubiquitin to a substrate involves a hierarchal three-step process requiring the activity of at least three different classes of enzymes. The first step involves activation of ubiquitin by a single dedicated ubiquitin-activating (E1) enzyme. In eukaryotes, activation comprises two steps: the initial coupling of ATP to form an ubiquitin-adenylate intermediate, followed by the formation of a high energy thiol-ester bond between the carboxyl-terminal glycine (Gly76) of the intermediate with a specific cysteine residue of the E1. One of several members of the ubiquitin-conjugating enzyme family (E2) then moves the activated ubiquitin in a trans-esterification reaction via an E2 ubiquitin thiol-ester intermediate, to the substrate which has been specifically bound to an ubiquitin-protein ligase (E3). The E3-assisted transfer of ubiquitin to substrate occurs either directly or via an additional E3 ubiquitin thiol-ester intermediate (reviewed in Varshavsky, 1997) (*figure 1.3 A*). Attachment of the activated carboxyl-terminal of ubiquitin is directed to the ε-amino groups of internal lysine residues of the substrate to form an isopeptide bond, which in successive reactions or with the involvement of the recently described E4 class of enzymes (Keog *et al.* 1999) generates a polyubiquityl chain (*figure 1.1*).

### REGULATION AND SPECIFICITY OF UBIQUITYLATION

There exists no single conserved peptide motif that targets substrates for ubiquitylation. Specificity is imparted by protein-protein interactions with the ubiquitylation machinery, specifically the E2s and E3s that recognise their substrates via specific motifs. The human

NOTE: This image is included on page 4a in the print copy of the thesis held in the University of Adelaide Library.

**Figure 1.3**

**Ubiquitin Conjugation**

a) Basic steps in substrate modification by the ubiquitylation machinery consisting of ubiquitin-activating (E1), ubiquitin-conjugating (E2) and ubiquitin-protein ligase (E3) classes of enzymes
b) The ubiquitin conjugation cascade. Humans are predicted to contain genes for one E1, over 40 E2s and more than 500 different E3s
c) Modes of E3 Recognition of Protein Substrates: 1) Constitutive recognition by the E3 via a primary motif, 2) Recognition of the substrate following its post-translational modification (eg. phosphorylation), 3) Recognition of the substrate following the post-translational modification of the E3, 4) Recognition of the substrate following its association with an ancillary protein.

(A, B adapted from Pickart and Eddins 2004, and C adapted from Ciechanover et al. 2000)

genome project has revealed the presence of 1 E1 (although alternative translation initiation sites gives rise to nuclear and cytoplasmic isoforms), >40 different E2s and >500 different E3s (Wong *et al.* 2003). This large repertoire of enzymes undoubtedly reflects the requirement for regulation and specificity of the ubiquitin system. The hierarchical layout of these enzymes is such that one E1 interacts with all the E2s, and each different E2 species may be able to form an E2-E3 ligase complex with one or more E3s (*figure 1.3 B*). This type of combinatorial expansion is a way to potentially increase the ubiquitin systems' substrate specificity repertoire. Furthermore, each class of enzyme in this cascade affords an extra level of control over the ubiquitylation process and are regulated both temporally and spatially. For example, the intracellular localisation of some E2 enzymes has recently been shown to be regulated. A group of E2's, murine UbcM2, and human UbcH6 and UBE2E2 are imported into the nucleus upon transfer of the activated ubiquitin intermediate to the E2's active site cysteine (Plafker *et al.* 2004). Although these enzymes are theoretically small enough to freely diffuse into the nucleus, their importation relies on their recognition by importin-11, a member of the karyopherin family of nuclear transport receptors that mediates translocation through nuclear pore complexes. Importin-11 however only has affinity for ubiquitin charged forms of these E2s and selectively transports them into the nucleus to perform their function. Previous observations that UbcM2 shuttles continuously between the nucleus and cytoplasm has led to the proposition that these E2s are first activated by cytoplasmic E1s, imported into the nucleus and then returned to the cytosol once they have completed their function. It has been further proposed that the E2 delivery maybe directed to a particular sub-nuclear compartment as another member of the karyopherin family, Kap104 in yeast has been shown to release their cargo only when delivered to RNA. Importin-11 may deliver their activated E2 cargo to specific E3's and thus regulate the formation of specific E2-E3 complexes within the nucleus (reviewed in Zhang and Matunis 2005).

The ubiquitin system gains another level of exquisite specificity from the E3s, which act as molecular scaffolds, binding substrates for presentation to the various E2s. E3s employ a number of strategies to recognise and bind substrates. In most cases, an E3 recognises a subset of protein substrates that share a particular structural motif. Some substrates contain more than one distinct recognition motif, allowing for recognition by different E3s. In other cases, the mode of recognition can be dependent on loss of DNA binding (as in the case of transcription factors which require dissociation from DNA binding sites in order to

be recognised), post-translational modification (such as phosphorylation of either substrate or E3) or association with ancillary proteins (eg. molecular chaperones acting as recognition elements in *trans*, reviewed in Ciechanover *et al.* 2000) (*figure 1.3 C*). Several classes of ubiquitin ligases have been described based on the presence of certain domains: HECT (homologous to the E6-AP C-terminus) domain, RING (really interesting new gene) finger, the related PHD (pleckstrin homology domain) finger and U-box (UFD2 homology) domain (reviewed in Pickart and Eddins 2004). Additionally, certain E3s are regulated by tissue specific expression. For example, expression of MAFbx, a muscle-specific E3 ubiquitin ligase is dramatically up-regulated in atrophying muscle. Through the ubiquitin-proteasome pathway, MAFbx ubiquitylates and regulates the protein levels of MyoD, a helix-loop-helix transcription factor that directs myogenic differentiation. In this way, selective expression of MAFbx and its affect on MyoD in muscle is thought to regulate the overall balance of differentiated cells and undifferentiated quiescent cells. Thus upregulation of MAFbx leads to MyoD degradation resulting in muscle atrophy while downregulation increases MyoD half-life, leading to muscle fibre regeneration (Tintignac *et al.* 2005)

Use of a polyubiquitylating class of enzymes could theoretically provide the ubiquitin system an extra level of regulatory control once a protein has been monoubiquitylated. The cell would then have the options of either to degrade the protein by creating the K48 polyubiquityl tag (a step E4s have been implicated in), to direct it to another fate by creation of different polyubiquityl tag (such as a K63-linked di-ubiquitin tag implicated in DNA repair) or have its intracellular localisation regulated by leaving it monoubiquitylated.

The ubiquitin system is regulated on yet another level by the localisation of its enzymes. It has been suggested that the more enzymatically promiscuous deubiquitylating enzymes are prevented from indiscriminate activity if tethered to the proteasome (Wilkinson 1997). Additionally, certain ubiquitin conjugases involved in the polyubiquitylation of misfolded proteins may localise tightly to the endoplasmic reticulum, where quality control of membrane and secreted proteins occur (Lord *et al.* 2000).

## *Deubiquitylation*

Deubiquitylation is mediated by deubiquitylating enzymes (Dubs), a large family of proteins that the human genome project has estimated more than 90 members (reviewed in Baek 2003). Dubs function to specifically cleave ubiquitin-linked moieties after the last C-terminal residue of ubiquitin (Gly76) and their cellular roles vary from housekeeping to modification of protein fate/localisation (*figure 1.4*).

Dubs can be classified into several distinct sub-families. The first family termed ubiquitin carboxyl-terminal hydrolases (UCHs) consist of relatively small sized proteins of around 20-30 kDa. Their catalytic core domain spans around 230 amino acids and contains a catalytic triad of spatially conserved Cys, His and Asp residues, a geometry similar to that of cysteine proteases (*figure 1.5 A*). UCHs primarily serve housekeeping functions by preferentially cleaving small adducts from ubiquitin, such as peptides degraded by the proteasome. Beyond ubiquitin recycling, UCHs are also required in processing newly synthesised ubiquitin molecules that are translated as linear "head" to "tail" polyubiquityl precursors, and also releasing some ribosomal proteins that are translated with an ubiquityl group at their N-terminus that targets them to the ribosome (reviewed in Ciechanover, 2000).

A second group of Dubs belongs to the ubiquitin specific peptidase (USP) family (otherwise known as the ubiquitin-specific proteases (UBP) family). They exhibit substrate specificity, display various tissue-specific expression patterns (Baker *et al.* 1999, reviewed in Hochstrasser 1996) and are involved in a broad number of cellular processes such as control of growth, differentiation, oncogenesis and genome integrity (reviewed in Wilkinson, 1997). Of the 90 odd deubiquitylating enzymes identified by the human genome project, around 80% of them are USPs. The molecular weights of USPs are quite large compared to UCHs ranging from 50 to 300 kDa, which may be attributed to their requirement to recognise multiple substrates. The USP catalytic core is roughly delimited by two highly conserved motifs, the Cys and His boxes however, unlike UCHs the USP catalytic core displays considerable sequence and size variance, ranging from 300 to 600 amino acids (*figure 1.5 B*). Within the catalytic core there are several other highly conserved motifs including the Asp box and KRF box (*figure 1.5 C*). The remaining

**Figure 1.4**

**Roles of Deubiquitylating Enzymes**

- **Proprotein processing:** Ubiquitin and several ubiquitin like proteins are synthesised as fusion proteins, requiring cleavage of the ubiquityl group to be functionally active
- **Salvage:** recycling of ubiquitin from degraded proteins
- **Editing of ubiquitylated proteins:** Either as a proofreading function or to reverse ubiquitylation to alter protein fate
- **Disassembly of degradation intermediates:** disassembly of polyubiquitin chains to regenerate free ubiquitin. If not removed, these chains may act as competitive inhibitors to the proteasome

(Adapted from Wilkinson 2000)

NOTE: This image is included on page 7b in the print copy of the thesis held in the University of Adelaide Library.

**Figure 1.5**

**Conservation of Deubiquitylating Enzyme Catalytic Cores.**

Alignment of the conserved Cys and His box motifs surrounding catalytically active amino acid residues (asterisks) for **(A)** UCHs and **(B)** USPs. **(C)** Schematic representation of various USPs highlighting the relative positions of several conserved motifs. Conserved motifs are labelled C (Cys box that contains the catalytic C), D (Asp box), Zn (Zinc binding region), K (KRF box), and H (His box that contains the catalytic H and D). Proteins are to scale, except for the N- and C-terminal regions of FAM, which are given in amino acids.

(A, B Adapted from Amerik and Hochstrasser 2004)

regions of these proteins consist of a variety of N-terminal extensions, occasional C-terminal extensions and insertions within the catalytic core. The sequences of these extensions are not conserved amongst the family and it has been proposed that these regions of diversity may function in substrate recognition, subcellular localisation and protein-protein interactions (reviewed in Kim *et al.* 2003). Indeed substrate specificity has been shown for the USP, herpesvirus-associated ubiquitin-specific protease (HAUSP). Through mammalian cell culture and biochemical assays, HAUSP has been demonstrated to stabilise p53 but not p27 (both of which are polyubiquitylated), while the unrelated human USP11 neither binds or deubiquitylates p53 (Li *et al.* 2002).

Recently, three other subfamilies of Dubs have been identified. They include the OTU (*o*varian *tu*mour)-related proteases, ataxin-3 proteases and the JAMM/MPN+ proteases. It is known that these families catalyse the disassembly of ubiquitin conjugates, but much of their biological function and significance as a separate subfamily from UCHs and USPs remains unclear. Few members have been characterised for these subfamilies but one OTU-related protease, A20 has been shown to inhibit nuclear factor-κB (NF-κB) activation by its surprising ability to both deubiquitylate and ubiquitylate (reviewed in Heyninck and Beyaert 2005). NF-κB is a transcriptional regulator involved in innate and adaptive immunity, inflammation, development, cell proliferation and survival. NF-κB is activated by the tumour necrosis factor (TNF) signal transduction pathway, which requires the recruitment of several proteins to the signalling complex to mediate signal transduction. Receptor-interacting protein (RIP) is a key signalling protein that is recruited to the activated TNF-receptor complex when polyubiquitylated via Lys63 linkages. Signal termination is mediated by A20, which acts to remove RIP from the complex. Interestingly, A20 directs the removal of RIP by first removing the Lys63 polyubiquityl tag via its OTU domain (deubiquitylation), and then finally conjugating a new Lys48 linked polyubiquityl tag (ubiquitylation) that targets the protein for destruction at the proteasome. A20 expression is induced by NF-κB activation thus completing the negative feedback loop. The manner by which the two opposing activities of A20 are regulated still remain unclear but does represent a new paradigm in the ubiquitin field.

## CATALYTIC CORE STRUCTURE OF DUBS

Structural data for representative members of the UCH, USP, MPN+/JAMM and OTU classes of Dubs have been determined through X-ray crystallography. Of particular significance, crystal structures of UCH and USP enzymes in both the free enzyme form, and also covalently complexed to an ubiquitin derivative, have revealed the mechanisms employed in deubiquitylation. This was made possible by use of ubiquitin aldehyde (Ubal), a form of ubiquitin that has had its C-terminal carboxylate reduced. When the cysteine of a Dubs catalytic triad attacks Ubal, a relatively stable hemi-thioacetal intermediate is trapped, locking the active site into a functional conformation. Mimicking this reaction intermediate allows sufficient quantities of Dub-Ubal to be crystallised. Studies into the structures of UCHs UCH-L3 (Johnston *et al.* 1997) and Yuh1 (Johnston *et al.* 1999), and the USP catalytic core of HAUSP (Hu *et al.* 2002) revealed remarkable similarities to the active site geometry of the classical papain family of cysteine proteases such as cathepsin B, particularly in the region including the catalytic triad (reviewed in Amerik and Hochstrasser 2004) (*figure 1.6 A, B*). In these regions, the UCH and USP proteins have nearly indistinguishable three-dimensional folds despite their sequence divergence, and the conformations of the catalytic triad residues are superimposable (*figure 1.6 C*). In their respective substrate free forms, the active sites are not in a catalytically competent conformation. Only upon an apparent ubiquitin-induced conformational rearrangement, do these enzymes become catalytically active by elimination of steric obstructions in the active cleft (UCH-L3, Yuh1) or by the bringing together of the catalytic residues into their proper relative positions (HAUSP).

In order to cleave ubiquitin from conjugates, the catalytic triad residues of cysteine, histidine and aspartic acid must be properly aligned. The catalytic cysteine undergoes deprotonation, unleashing a nucleophilic attack on the carbonyl carbon atom of the ubiquitin gly76 at the scissile peptide bond, forming an initial tetrahedral intermediate, followed by a more stable acyl intermediate through the expulsion of the C-terminal leaving group. Attack by a water molecule generates a carboxylate on the ubiquitin product (its negative potential dissipated by the proteases' oxyanion hole) and this simultaneously regenerates a free thiol on the enzyme. The charged groups on the histidine and aspartic acid side chains interact to assist and stabilise the reaction, rendering the thiol group on the

**Figure 1.6**

**The UCH Catalytic Core**

A. Ribbon diagram comparison of the UCH family DUB, UCH-L3 (upper) and cysteine papain family protease, cathepsin B (lower). Equivalent residues are shown in the cyan ribbon with catalytic triad residues in red.
B. Topology diagram of secondary structure for structurally equivalent segments of the UCH-L3 (upper) and cathepsin B (lower). The main topological difference is in the location of the helix. In the papain-like enzymes, the helix is the first of these secondary structural elements while for UCH-L3, it is located between strands 2 and 3. A long disordered loop is indicated by a dotted line.
C. Structural superimposition of the catalytic cores of the UCH, Yuh1 (magenta) and USP, HAUSP (blue) when both are bound to Ubal. The residues that form the catalytic triad and oxyanion hole are depicted.
D. Superimposition of the active sites of Yuh1 (green) UCH-L3 (blue) with Ubal (purple). The active site of UCH-L3 is occluded by Leu9 in the space taken by Gly75 of Ubal in the Yuh1 complex. Ubal binding results in ~4Å displacement of the equivalent residue Ile11 of Yuh1.
E. The carbonyl oxygen of UCH-L3 Ser92 blocks the oxyanion hole in the unliganded structure. In the Yuh1-Ubal complex, the equivalent residue Lys87 is rotated by ~180° to allow formation of hydrogen bonds. Hydrogen bonds are marked between Ubal Gly76 O and Yuh1 (green) and UCH-L3 (black).

(A, B adapted from Johnston et al. 1997, C adapted from Hu et al. 2002, and D, E adapted from Johnston et al. 1999)

cysteine protease an effective nucleophile for attack on the peptide bond (reviewed in Wing 2003).

In an inactive conformation, the UCH active site and oxyanion hole is inaccessible due to the presence of a disordered active site-crossover loop (not present in classical cysteine proteases), which becomes ordered upon ubiquitin binding (*figure 1.6 D, E*). However, even at its most open state, the loop diameter is no greater than 15 Å, much smaller than the majority of folded proteins and thus limits UCHs to the cleavage of small adducts or unfolded polypeptides from the C-terminus of ubiquitin. Additionally, the active site cysteine is located at the bottom of a narrow groove in the surface of the enzyme that restricts access to large side-chain residues. Ubiquitin, which terminates with a pair of glycines, is small enough to be accommodated in this groove. These factors (along with the extensive binding interactions between ubiquitin and the UCH) determine the specificity of UCH enzymes.

Crystal structures of the catalytic core of HAUSP in both its free and Ubal-complexed form reveal a three globular domain configuration for USPs that resembles an extended right hand comprised of Fingers, Palm and Thumb (Hu *et al.* 2002) (*figure 1.7 A*). The general conservation of the residues that constitute the secondary structural elements within the three domains appears conserved among other USP enzymes, and more specifically, residues that contribute to the structural integrity of the Fingers-Palm-Thumb architecture are invariant. Work into predicting the 3D-structure of putative domains for human USP9Y (Ginalski *et al.* 2004) found that four cysteine residues in the Fingers domain of the catalytic core that may coordinate a zinc ion. These cysteines form a putative zinc ribbon-like structure, which was recently confirmed in HAUSP by comparative sequence and structural analysis, as a zinc ribbon that has lost its zinc-binding ability (Krishna and Grishin, 2004).

Using this hand analogy, the highly conserved Cys and His boxes are positioned on opposite sides of a deep, inter-domain catalytic cleft created by the Palm and Thumb. Correct positioning of the ubiquitin moiety is coordinated by the Fingers, with the C-terminal glycines of ubiquitin placed in the active site between the Palm and the Thumb (*figure 1.7 B*). In the free form of HAUSP, the catalytic triad is misaligned with the catalytic Cys and His residues too far apart to allow any meaningful interactions. Binding

NOTE: This image is included on page 10a in the print copy of the thesis held in the University of Adelaide Library.

**Figure 1.7**

**The USP Catalytic Core**

- **A.** Ribbon diagram depicting the overall topology of the 40 kDa catalytic core domain of HAUSP comprising of Fingers (green), Palm (blue) and Thumb (gold). The active site (black oval) comprising the Cys box (cyan) and His box (purple) is located between the Palm and Thumb.
- **B.** HAUSP catalytic core (blue) when covalently bound to Ubal (green). Previously disordered loops in the free HAUSP are highlighted in red. Catalytic triad residues shown as yellow sticks.
- **C.** Superimposition of HAUSP in isolation (purple) and in complex with Ubal (blue). Regions become ordered (red arrows) and undergo drastic conformational changes (gold arrows) upon binding of Ubal.
- **D.** Arrangement of active site residues in the free form of HAUSP is misaligned, Cys box (cyan) Cys223 and His box (magenta) His464 are too far away for a productive interaction.
- **E.** Superimposition of HAUSP in both isolation (purple) and in complex (blue) to Ubal. Ubal binding (green) induces localised conformational changes bringing the residues of the active site together to enable catalysis.

(Adapted from Hu et al. 2002)

of Ubal caused a dramatic, highly localised conformational change in the catalytic cleft, bringing together the relevant residues of the catalytic triad and causing the ordering of two previously flexible surface loops, switching the enzyme into an active conformation (*figure 1.7 C-E*). HAUSP makes extensive contacts with ubiquitin involving interactions with both the Fingers and the Palm-Thumb scaffold resulting in the burial of ~3600 $Å^2$ of solvent accessible surface area. Binding of Ubal by HAUSP can be visualised as grabbing Ubal with the tip of the Fingers and the catalytic cleft between the Palm and Thumb, while a cushion of water mediates the contact between Ubal and the middle portion of the Fingers.

While the three domain structure of USP allows for recognition of larger substrates, such as polyubiquitylated proteins and free polyubiquityl chains, the HAUSP catalytic core by itself is not sufficient for recognition of ubiquitylated substrates *in vivo* as gauged by high dissociation constants. HAUSP instead recognises its substrate p53, via a domain N-terminal to the catalytic core that contains the elements required for its specific interaction with p53 (Hu *et al.* 2002). Interestingly, HAUSP only requires a short 26 amino acid C-terminal segment of p53 for this interaction, and this p53 element includes several lysine residues thought to be involved in p53 ubiquitylation. This data supports a model whereby an N-terminal extension of HAUSP directly binds p53 proximal to the ubiquitin linkage sites to enable the catalytic core, which normally has weak affinity for ubiquitin, to bind the conjugate and become activated for proteolysis (reviewed in Lima 2003). Such a model has implications for other USPs and lends strength to the argument that the sequence divergent extensions of the various USPs comprise in part, substrate recognition sites.

Although UCHs and USPs share nearly identical active site geometries, the overall structure and topology are vastly different. UCHs lack the Fingers domain and have a shortened thumb. Although the Palm domain is present with the same overall fold, large deviations are apparent throughout the structure. UCHs make contact with ubiquitin largely from one surface due to a lack of Fingers, and the catalytic triad within the active site does not undergo any significant conformational change during catalysis. Crystal structures of both human otubain2 (OTU family of Dubs) and an even more distantly related cysteine protease that acts on the ubiquitin-like molecule SUMO, are also found to have the same active site configuration as UCHs and USPs demonstrating a conserved catalytic mechanism for deubiquitylation (reviewed in Amerik and Hochstrasser 2004).

## *Drosophila Fat Facets (faf), a Developmentally Regulated USP*

*Fat facets* (*faf*) is a developmentally regulated, 2747 amino acid USP that was identified in a mutation screen for genes that affect drosophila eye development. It is essential in at least two key developmental events: regulation of photoreceptor number in compound-eye formation, and nuclear migration during syncytial stage oocyte development (Fischer-Vize *et al.* 1992). Using a β-galactosidase fusion to the first 392 amino acids of FAF, it was also shown that FAF protein is spatially regulated. A complex localisation pattern was observed in the imaginal eye disc and localisation within the syncytial oocyte was detected at the posterior pole, indicating a requirement of specifically localised FAF activity for proper eye and oocyte development in drosophila (Fischer-Vize *et al.* 1992).

### FAF AND COMPOUND-EYE DEVELOPMENT

In drosophila, the compound eye consists of a hexagonal array of around 800 ommatidia eye units, or facets that develop from a monolayer of cells termed the eye imaginal disc. A depression called the morphogenetic furrow, moves across the disc of undifferentiated cells as a wave, posterior to anterior and in the process of this movement, facet preclusters emerge, mature and recruit additional cells in a precise series of inductions. Contained within these initial facet preclusters are seven cells, five of which will become photoreceptors that have a key role in directing the differentiation of the rest of the facet, while the other two, termed mystery cells, will detach from the precluster and re-enter the surrounding pool of undifferentiated cells (reviewed in Ready 1989) (*figure 1.8 A-C, I*). Although *faf* null mutant flies were viable, they showed abnormal "rough" eye morphology characterised by the appearance of extra photoreceptors in addition to the normal complement of eight in each facet (*figure 1.8 D-F*). Ectopic photoreceptor formation is the result of inappropriate differentiation of mystery cells that would otherwise have retreated into the imaginal disc. Analysis of eye cells mosaic for $faf^{+}/faf^{-}$, showed that the cell communication pathway that negatively regulates the neural cell fate in the developing eye requires FAF in cells near to, but outside the eight photoreceptors in wild-type facets. Additionally, mutant alleles of a 20S proteasome subunit strongly suppress the *faf* mutant phenotype, suggesting that FAF acts to limit the ubiquitylation and therefore degradation of one or more regulators of eye development (Huang *et al.* 1995, Wu *et al.* 1999).

**Figure 1.8**

**Rescue of the faf Mutant Eye Phenotype by Fam Transgenes**

Genotypes: wild-type **(A-C),** fafFO8/fafBX4 - a strong mutant combination **(D-F),** P{w+, ro- Fam-}3; fafFO8/fafBX4 - Fam rescue of a faf mutant background **(G-H)** restores photoreceptor number. Arrow in **H** indicates a mutant facet with an extra photoreceptor. Scanning electron micrographs **(A, D, G),** apical tangential sections of adult eyes **(B, E, H)** and cobalt sulphide stained pupal eyes **(C, F). I** shows wild-type photoreceptor cell arrangement in apical tangential sections (left) and cone and pigment cells in cobalt-sulphide stained pupal retinas (right). The numbers in the left diagram indicate photoreceptor cells (R1-R7) visible in the apical plane section shown in **B, E, H**. R8 lies beneath R7 and is not visible at this plane of section. Symbols in right diagram: bristle cell (b), cone cell (c), primary, secondary and tertiary pigment cells (1°, 2°, 3°). **J** The rescuing activities of two copies of three independent P{w+, ro-Fam-} transgene insertions were quantified by scoring the numbers of mutant and wild-type facets in apical tangential sections. For each line, 150-250 facets were scored in three different eyes.

(Adapted from Chen et al. 2000)

A genetic screen to uncover candidates for the critical substrate of FAF in the eye identified *liquid facets* (*lqf*), an endocytic protein orthologous to vertebrate Epsin1 (Cadavid *et al.* 2000). Four genetic observations were made that suggested a role for FAF in preventing Lqf degradation: (1) *lqf* loss-of-function mutants are strong dominant enhancers of the *faf* mutant eye phenotype, (2) *faf* and *lqf* loss-of–function mutations have similar mutant eye phenotypes, (3) the $faf^{+}$ and $lqf^{+}$ genes are required in the same group of cells in the eye, and (4) one extra copy of the $lqf^{+}$ gene overcomes the need for the $faf^{+}$ gene in the eye. These genetic observations were confirmed with biochemical data that showed: (1) there is less Lqf protein in the developing drosophila eye in the absence of functional FAF protein, (2) Lqf is ubiquitylated in the developing eye and is deubiquitylated by FAF, and (3) Lqf and FAF interact physically (Chen *et al.* 2002). Therefore it was concluded that FAF regulates the levels of Lqf by deubiquitylating it, thus preventing its degradation at the proteasome. The fact that Lqf is involved in endocytosis carries added significance given the recent understanding that the ubiquitin pathway and its Dubs can influence endocytosis and intracellular trafficking. Indeed, a role for Lqf in endocytosis is supported by genetic and co-localisation evidence that shows that *lqf* interacts with several critical endocytosis genes (Cadavid *et al.* 2000, Chen *et al.* 2002).

Recent genetic experiments have provided a molecular model for how FAF, Lqf, and endocytosis relate to cell signalling to inhibit photoreceptor development by the surrounding cells. The Notch pathway participates in a wide range of cell communication events, either inhibiting or promoting a variety of cell fates, and has been implicated in the FAF/Lqf control of photoreceptor development (Overstreet *et al.* 2003). Notch activation by Delta/Serrate/Lag2 (DSL) ligands requires endocytosis in both signalling and signal-receiving cells and it has been proposed that the extracellular domain of the Notch receptor (on the signal-receiving cell) when bound to Delta, is trans-endocytosed into the Delta-expressing (signalling) cell. Processing of the intracellular domain of the cleaved Notch receptor in the signal-receiving cell leads to signal transduction (and in this case, inhibition of photoreceptor development). It has been shown that Lqf is required by signalling cells to transmit DSL signals (Wang and Struhl 2004). Through immunoctyochemistry and genetic experiments, it was deduced that FAF through its substrate Lqf, promotes Delta internalisation (a process dependent on its ubiquitylation by the E3, Neuralised) and subsequent Delta signalling by signalling cells. The signalling cells then activate Notch in surrounding undifferentiated cells, preventing ectopic photoreceptor differentiation

(Overstreet *et al.* 2003, 2004). It has also been proposed that endocytosis of monoubiquitylated DSL signalling ligands by Lqf, may target them to an endocytic recycling compartment, converting them from inactive 'pro-ligands' into active ligands (Wang and Struhl 2004).

FAF has also been shown to genetically interact with *Rap1* and *Ras1*, members of the mitogen-activated protein kinase (MAPK) pathway. Analysis of these interactions has revealed that *faf* has an additional function later in eye development involving these proteins, in the continued influence of facet assembly from the adjacent signalling cells (Li *et al.* 1997).

## FAF AND OOCYTE DEVELOPMENT

In addition to abnormal eye development, *faf* mutations have a maternal effect phenotype. Homozygous mutant females are sterile despite possessing seemingly normal ovaries, and their eggs are unable to reach the syncytial blastoderm stage (Fischer-Vize *et al.* 1992). Normally in drosophila development, a fertilised egg will undergo a series of 14 synchronous divisions to form a multi-nucleated cell termed the syncytium. By division 10, the nuclei migrate to the periphery of the cell forming the "syncytial blastoderm", and the primordial germ cells (pole cells) then form at the posterior end of the embryo. The nuclei continue to the $14^{th}$ synchronous division when cellularisation occurs around each nucleus to form the cellular blastoderm. In *faf* mutant embryos, it was observed through nuclei and cell membrane staining, that syncytial blastoderm development had been disrupted in all cases. Embryos were able to reach at least the $10^{th}$ division as determined by the presence of the pole cells, however most nuclei (except patches of asynchronously divided nuclei) did not migrate to the periphery. With the exception of the pole cells (which were fewer in number and more spread out) cellularisation failed to occur (Fischer-Vize *et al.* 1992).

Although no direct critical substrate of FAF has been identified for oogenesis, FAF has been shown to interact with Vasa, an RNA helicase that is a component of polar granules and maternally essential for posterior patterning and germ cell specification (Liu *et al.* 2003). FAF has been shown to reverse Vasa ubiquitylation and stabilise it in the pole plasm by inference, but has not been biochemically proven to be a bona fide, nor critical substrate of FAF in oogenesis.

## *Fat Facets in Mouse (Fam)*

The mouse orthologue of *faf, Fat Facets in Mouse (Fam)* was identified in a gene trap screen in embryonic stem cells for genes expressed during gastrulation and neurulation (Wood *et al.* 1997). One particular clone identified by this screen displayed a developmentally restricted β-galactosidase expression pattern, beginning only at the late primitive streak stage. After isolating the complete cDNA, sequence analysis revealed an open reading frame encoding a 2554 amino acid USP bearing strong sequence similarity to *faf*. FAM and FAF are collinear over nearly the entire length and show approximately 50% amino acid identity and 70% similarity, which increase considerably over the conserved Cys and His boxes. Furthermore, over-expression of *Fam* in *faf* mutants is able to rescue the phenotype which suggests that *Fam* and *faf* are true orthologues (Chen *et al.* 2000) (*figure 1.8 G, H, J*).

FAM is critical for early mouse development as anti-sense oligodeoxynucleotides knockdowns of FAM protein levels in pre-implantation mouse embryos inhibits their progression from the two-cell to morulae or blastocyst stages. These embryos exhibit an apparent diminution in cell-cell adhesion (*figure 1.9*). Depletion of FAM also corresponded with decreased protein levels of two of its substrates, β-catenin and AF-6. However, following an initial decrease, after 48 hours AF-6 levels returned to normal but the nascent protein was mislocalised to the apical surface of blastomeres (Pantaleon *et al.* 2001).

There are several other known isoforms, homologues and orthologues of FAM. Northern analysis of *Fam* transcripts performed on post-implantation embryos revealed the existence of three different length transcripts: 8.5 kb, 10.0 kb and 11.5 kb of which the 10.0 kb transcript was the most abundant. All the transcripts encode the same protein and the variance in size is due to differences in the length of the 3' untranslated region however, their functional significance remains unclear (Wood *et al.* 1997). *Fam* also exists as two isoforms caused by alternate splicing of exons. This splice variation results in the frame shift insertion or exclusion of 15 bps, which encode the amino acids SSSRF (at position 589-590 aa). Thus the isoforms are named SSSRF+ (2559aa) and SSSRF- (2554aa) however again, the functional significance is unknown.

**Figure 1.9**

**FAM is Essential for Pre-Implantation Mouse Development**

Immuno-localisation of FAM following treatment with Fam sense **(A, C)** and anti-sense **(B, D)** oligodeoxynucleotides (ODN). Two-cell stage embryos were cultured in the respective ODNs for 24 h **(A, B)** or for 72 h to the late blastocyst stage **(C, D)** and then probed with anti-FAM antibodies. The perinuclear and cytoplasmic positive FAM immuno-reactivity present in the sense treated embryos **(A, C)** is reduced in the 24 h anti-sense ODN treated embryos **(B)** and nearly abolished by 72 h **(D).** Colour wedge indicates highest intensity of immunofluorescence as white. Bar = 25µm.

(Adapted from Pantaleon et al. 2001)

*Fam* is located on the sex chromosomes and so two homologues exist, an X and Y copy. The Y homologue is exclusively expressed in the testes and maps to the $Sxr^b$ deletion, which is associated with an early post-natal blockage of spermatogonial proliferation and differentiation and results in the adult testis being almost totally devoid of germ cells (Brown *et al.* 1998). The X homologue which will be referred to as *Fam*, displays temporally and spatially restricted expression throughout embryogenesis. The remainder of this thesis describes research carried out on the SSSRF+, X-chromosome homologue unless otherwise indicated.

The only described orthologue of *Fam* other than drosophila *faf*, are the human DFFRX and DFFRY (drosophila fat-facets related X or Y) genes, otherwise known as USP9X or USP9Y according to new nomenclature conventions. USP9X escapes X-inactivation and both the X and Y copies are expressed in a wide and similar range of tissues (Brown *et al.* 1998) however, USP9Y is additionally expressed in the testes and has been shown to be a functional USP (Lee *et al.* 2003). USP9X shares 97% identity and 99% similarity to FAM while USP9X and USP9Y share 89 % identity and 98% similarity (Brown *et al.* 1998).

## TEMPORAL AND SPATIAL REGULATION OF *FAM* DURING DEVELOPMENT

Whole-mount *in situ* analysis of post-implantation mouse embryos revealed that *Fam* is expressed in a complex temporal and spatial pattern throughout development (Wood *et al.* 1997). Transcripts were first detected at the mid-streak stage (E7.5) but switches from relatively ubiquitous expression to more tissue specific after E10.5, where expression became progressively restricted to the developing central nervous system (CNS), limb buds and branchial arches. By E13.5 *Fam* transcripts were undetectable by whole-mount *in situ* analysis but were still expressed strongly as shown by *in situ* analysis (*figure 1.10 A, B, C*).

Within the CNS, *Fam* is expressed in a complex fashion. *Fam* expression in the nervous system is strongest in the third, fourth and lateral ventricles, and in the developing spinal cord. At E14.5, *Fam* expression is obvious in the olfactory lobe and in certain ganglia: trigeminal, inferior glossopharyngeal and dorsal root ganglia. Within the developing cortex, *Fam* is expressed by the neuroblasts of the ventricular zone, as well as in migrating and differentiating neurons of the intermediate zone and cortical plate (Friocourt *et al.* 2005).

**Figure 1.10**

**Fam Expression in Wholemount Mouse Embryos**

Wholemount in situ hybridisation analysis of E9.5 **(A)** E11.5 **(B)** and E12.5 **(C)** with a Fam anti-sense probe. Ubiquitous expression at E9.5 is increasingly restricted to the E12.5 stage. Expression is lost from the body wall while being maintained longest in the distal halves of the limb buds, branchial arches as well as the eye and central nervous system. At E12.5 only the eye, mesencephalon, telencephalon and the apoptotic regions between the digits remain positive for Fam. Bar represents 500μm in each panel. Sections of E12.5 midbrain **(D)** reveal strong Fam expression in the diencephalon (Di) and metencephalon (Met) which is progressively lost laterally. Sagittal sections of an E14.5 eye **(E)** show strong expression in the outer nuclear layer (on) but expression is lost from the inner nuclear layer (in) of the retina. Sections of the outer root sheath of the vibrissae **(F)** show strong expression around the hair follicle.

(Adapted from Wood et al. 1997)

Strong expression was detected in the eyes, ubiquitously at first, but later becoming more neurally confined until retinal differentiation when expression weakened (*figure 1.10 E*). *Fam* expression throughout eye development has been well characterised and found to co-localise extensively with one of its substrates AF-6 (discussed later), an epithelial tight-junction protein (Kanai-Azuma *et al.* 2000). At E10.5, FAM and AF-6 are both expressed in the outer layer of the optic cup and a single layer of cells immediately adjacent in the inner layer (*figure 1.11 A, B*). The outer layer goes on to form the retinal pigment epithelium (RPE) and continues to maintain FAM and AF-6 expression. During lens development, FAM and AF-6 were initially expressed at the apical surface of cells lining the cavity of the lens vesicle and later (E13.5), present at the contact zone between the differentiated lens fibres and sub-capsular epithelium (*figure 1.11 C*). FAM and AF-6 are also co-expressed throughout corneal epithelium development. In late embryonic development, both FAM and AF-6 are expressed in the inner layer of the neural retina and in the adult, localise to the outer plexiform layers. *Fam* is expressed in the outer nuclear layer by E18.5 (*figure 1.11 D, E*).

Interestingly, expression within the neural tube and brain was seen to become increasingly polarised (Wood *et al.* 1997). Expression within the diencephalon and metencephalon at E12.5 is strong in the ependymal layer, but is laterally lost by some cells through the mantle and even fewer in the marginal layer giving the appearance of a gradient of FAM expression (*figure 1.10 D*). Polarised expression was also observed during limb bud development where it was initially ubiquitous, but gradually was confined to only the distal halves at E11.5, and only present in the apoptotic regions between the digits twenty four hours later.

*Fam* transcripts were additionally detected in many internal organs, including the liver and other epithelial tissues such as the bronchi of the lungs, gut and nasal vibrissae (*figure 1.10 F*). Such diverse and specific expression patterns during early mouse development imply that *Fam* is associated with multiple developmental or cellular events (Wood *et al.* 1997). *Fam* has also been characterised during gonadal development and gametogenesis where it was found to be stage-dependently expressed in the germ cells and supporting cells (Noma *et al.* 2002). FAM further colocalises with AF-6 in Sertoli and granulosa cells of the testis and ovary in a stage-dependent, synchronous manner going from a diffuse cytoplasmic

**Figure 1.11**

**Fam Expression in the Developing Mouse Eye**

Immunofluorescence micrographs showing FAM localisation during eye development in E10.5 **(A, B)**, E13.5 **(C)** and E18.5 **(D, E)** mouse embryos. **(B)** is an enlargement of **(A)** while **(E)** is an enlargement of **(D)**. At E10.5, the eye primordium is composed of lens (l) and inner and outer layer of optic cup, which are precursors of the nervous layer of the retina (nr) and retinal pigmented epithelium (rpe) respectively. FAM is restricted to the outer layer and cells of the inner layer immediately adjacent (Arrows in **A** and **B**). FAM is strongly expressed in the retinal pigment epithelial cells (arrows in **C-E**) and the lens subcapsular epithelia cells (arrowheads in **C**). In the retina, the inner portion of the ventricular layer (in) increases expression from E13.5 to E18.5 while there also appears to be some FAM expression in the outer nuclear layer (on) at 18.5.

(Adapted from Kanai-Azuma et al. 2000)

distribution, to the Sertoli-Sertoli and Sertoli-spermatid junctions at later stages (Sato *et al.* 2004).

*Fam* has also been detected in other adult tissues. FAM was detected by immunohistochemistry in adult mouse uterus, where the epithelial lining stained strongly positive and to a lesser extent in surrounding muscle. Furthermore, in a micro-array screen comparing cDNA from mid-gestation placenta and embryo, it was found that *Fam* was expressed 124 times higher in placenta than in E12.5 embryos (Tanaka *et al.* 2000). Although *Fam* expression is restricted to a few areas in E12.5 embryos, this level of placental expression is the highest ever reported.

## FAM SUBSTRATES AND BINDING PROTEINS

Two bona fide FAM substrates have been identified, Acute Lymphoblastic Leukemia-1 (ALL-1) Fusion partner from chromosome 6 (AF-6) and  -catenin. Both are peripheral components of cell adhesion complexes and can also act as signalling molecules. These proteins bind FAM in a region termed *Fam*-CAT (1476-1918 aa), which includes the conserved catalytic Cys/His boxes of the catalytic core (Taya *et al.* 1998, 1999). FAM was identified as an AF-6-interacting protein in a glutathione-s-transferase (GST) pull-down assay using bovine brain extracts and the C-terminal third of AF-6 as bait (Taya *et al.* 1998). AF-6 is a 182 kDa protein which is thought to participate in cell-cell adhesion regulation as a downstream target of Ras (Kuriyama *et al.* 1996), and has an apparent role in regulation and maintenance of epithelial cell polarity (Zhadanov *et al.* 1999). Consistent with the model that FAM regulates AF-6, FAM partially co-localised with AF-6 at cell-cell contact sites in epithelial cells and various stages of eye and gonadal development, interacted *in vivo* and *in vitro*, was able to be ubiquitylated and FAM prevented its ubiquitylation (Taya *et al.* 1998, Kanai-Azuma *et al.* 2000, Sato *et al.* 2004).

Given the ubiquitin-mediated regulation of AF-6 at cell-cell contact sites, investigation was conducted into the possible role of *Fam* in stabilising  -catenin, a cadherin binding and signalling protein which is localised to cell adhesion sites as well as the cytoplasm.  -catenin is an 88 kDa protein that has two distinct roles within the cell. The first role is in cell-cell adhesion where it forms a complex with cadherins and α-catenin to mediate a link between the cell contact and the actin cytoskeleton. The second role of β-catenin is in Wnt

signalling where activation of the pathway leads to the accumulation of β-catenin in the cytoplasm followed by translocation to the nucleus where it activates Wnt target genes (cell cycle regulators) after heterodimerisation. FAM was shown to interact with β-catenin *in vivo* and *in vitro*. In mouse L cells, over-expression of the carboxyl-terminal half of *Fam* including the catalytic domain, leads to elevation of β-catenin levels and extension of its half-life. Immunofluorescence studies in these mouse L cells also show partial co-localisation of *Fam*-CAT with β-catenin at dot-like structures in the cytoplasm (Taya *et al.* 1999). Interestingly, β-catenin binds FAM via a series of armadillo repeats, a region that bears homology to the ENTH (Epsin N-Terminal Homology) domain of drosophila Lqf (Chen *et al.* 2000). This raises the possibility that FAM may recognise a conserved binding motif contained on a subset of its substrates.

*Fam* has also been shown to regulate Epsin1 (mammalian orthologue of Lqf). FAM and Epsin1 display overlapping immunostaining in synapses of rat brain sections, FAM coprecipitates with anti-Epsin1 immunoprecipitates, and Epsin1 is immobilised in a GST pull-down from rat-brain cytosol with a C-terminal portion of FAM (1554-1953aa) as bait (Chen *et al.* 2003). Further, when *Fam* expression in HeLa cells is suppressed by short interfering RNA (siRNA), deubiquitylation of Epsin1 was specifically inhibited. These results are consistent with the biochemical results obtained in drosophila that show that Lqf is a direct substrate of FAF (Chen *et al.* 2002).

*Fam* was also recently identified in a yeast two-hybrid screen, searching for interactors with Doublecortin (DCX), a microtubule-associated protein involved in neuronal migration (Friocourt *et al.* 2005). Although this interaction was confirmed by targeted mutagenesis, colocalisation, and immunoprecipitation studies, DCX does not appear to be ubiquitylated and is unlikely to be a FAM substrate. DCX was found to bind to human FAM (USP9X) in a region outside of the catalytic core, a novel C-terminal recognition domain comprising the last 256 amino acids (USP9X 2292-2547 aa). Interestingly, DCX also interacts with the μ subunits of the clathrin adaptor complexes AP-1 and AP-2 (Friocourt *et al.* 2001), known to be involved in vesicle trafficking, and associated with the trans-golgi network and plasma membrane respectively. Given that DCX is a microtubule-associated protein, it has been suggested that it may act to localise and/or regulate FAM in specialised compartments of neuronal cells (Friocourt *et al.* 2005). Interestingly these investigators

also report that DCLK1C, a protein similar to DCX which is expressed in astroglial and neuronal cells also interacts with FAM.

## SUBCELLULAR LOCALISATION OF FAM

Through immunohistochemistry and confocal microscopy, analysis of *Fam* expression in pre-implantation mouse embryos revealed complex subcellular localisation patterns (Pantaleon *et al.* 2001). In the unfertilised ovulated egg, FAM was strongly associated with the spindle and chromosomes, as well as displaying low level punctate staining in the cytoplasm. However upon fertilisation, FAM was observed to vacate to the cytoplasm where it displayed very strong punctate staining while being completely absent from the female and male pronuclei. In the following cleavage stages, FAM was still observed as puncta in the cytoplasm but also highly localised in the perinuclear region, but not to the chromosomes and spindle during mitosis. This pattern of localisation was maintained following the initiation of differentiation at compaction and also in the blastocyst (*figure 1.12*). It is noteworthy that a cytoplasmic punctate distribution has also been observed for FAF in drosophila S2 cells (Wu *et al.* 1999).

Investigation of FAM's subcellular localisation in non-differentiated proliferating PC12 cells (derived from rat adrenal gland) has revealed different localisations depending of the stage of cell cycle (Friocourt *et al.* 2005). FAM is located at microtubule-containing structures such as the midbody during cytokinesis, the microtubule-organising centre, and mitotic spindle in dividing cells. In confluent glial cells and MDCKII cells, FAM colocalises at cell-cell contacts with β-catenin (Friocourt *et al.* 2005, Taya *et al.* 1999) and AF-6 in MDCKII cells (Taya *et al.* 1998).

Recent immunofluorescence microscopy has further analysed FAM's subcellular localisation in polarised epithelial cells. FAM was found to localise to various puncta throughout the basolateral but not the apical cytoplasm (Murray *et al.* 2004). These puncta represented sites of protein sorting and trafficking as FAM partially colocalised with the Golgi apparatus, late endosomes and the lysosome. Through immunoprecipitation and gel filtration assays, FAM also was found to associate with nascent β-catenin and E-cadherin complexes in the cytoplasm, but not at the plasma membrane in sub-confluent cells. These observations suggest a role for FAM in facilitating the vesicular transport of nascent β-

**Figure 1.12**

**Subcellular Localisation of FAM During Mouse Pre-Implantation Development**

Confocal immunofluorescent optical sections of mouse embryos. **(A)** Fertilised egg and **(B)** blastocyst incubated with preimmune rabbit serum IgG. In unfertilised oocytes **(C),** FAM is strongly associated with the spindle and chromosomes, and also detected at lower levels in the cytoplasm. Fertilised oocytes **(D)** displayed positive FAM immunoreactivity in the cytoplasm, while both the male and female pronuclei are devoid of FAM. Two-cell **(E)**, four-cell **(F),** morula **(G),** and blastocyst **(H)** embryos all show positive staining for FAM, mainly associated with the perinuclear membrane but is also present in cytoplasmic puncta. Colour wedge indicates highest intensity of immunofluorescence as white. Bar = 25µm.

(Adapted from Pantaleon et al. 2001)

catenin E-cadherin complexes from the trans-golgi network to the basolateral plasma membrane of sub-confluent epithelial cells that are establishing cell-cell contacts (Murray *et al.* 2004).

It is possible that these observations made in epithelial cells may also explain the punctate staining pattern observed during pre-implantation development, and may also provide a mechanistic reason as to why a lack of FAM leads to a reduction in cellular adhesion. FAM-depleted embryos may not be able to establish strong cell-cell contacts because adhesion complexes are not able to be trafficked to the plasma membrane in the absence of FAM. FAM depletion correlates with an initial reduction in AF-6 protein levels which later return to normal; however the nascent protein is mislocalised to the apical surface of blastomeres. Assuming that FAM normally localises and sequesters AF-6 to the baso-lateral compartment (as observed in polarised epithelia), it is conceivable then that a lack of FAM results in the mislocalisation of AF-6 to the opposite compartment.

The findings that FAM is localised to multiple points of protein trafficking, coupled with the role of FAF in regulating the endocytic factor Lqf, implies that FAM and its orthologues may have a role beyond that of mere half-life extension of specific substrates. Their additional roles in modifying specific proteins for intracellular transportation, imparts to them a far more active role in regulating cellular events, the consequences of which have dramatic effects on development.

## *Structure and Function of USPs*

It is well established that the E2-E3 ubiquitin ligase family of proteins supply the ubiquitin machinery a high level of specificity. By deduction, the reverse process, deubiquitylation must also have the same level of specificity in order to avoid futile cycles of ubiquitylation and deubiquitylation. However, little is known of the structure-function relationship of USPs such as FAM. Given that USPs display significant sequence diversity (excluding the conserved catalytic boxes) in the form of insertions within the catalytic core and a variety of different N and/or C-terminal extensions, it has been proposed that these unique regions

may serve (amongst other things) as substrate-binding sites, conferring substrate specificity to USPs (Baker *et al.* 1992). This proposition makes two predictions. The first is that these unique regions specifically bind substrates and secondly, these regions preclude inappropriate enzymatic activity. Indeed, a USP has been identified where these extensions have been shown to specifically bind proteins. The mouse deubiquitylating enzyme mUBPy binds the SH3 domain of Hrs-binding protein (Hbp) via two novel SH3-binding motifs located in its N-terminal extension (Kato *et al.* 2000). A clear example where the second prediction of the proposition is demonstrated comes from testis USP. Testis USP exists as two isoforms, USP-t1 and USP-t2, each containing the same catalytic core regions, but with distinct N-termini. Not only do these N-termini target the two isoforms to different subcellular locations (Lin *et al.* 2000) but they also inhibit the ability of the enzyme to indiscriminately cleave ubiquitin from an artificial substrate as compared to the catalytic core alone, without significantly affecting catalytic function (Lin *et al.* 2001).

This modular model of function also has precedence in a member of the E3 HECT ligase family, *RSP5*. Like USPs, Hect E3 ligases are diverse in size (from 92 to over 500 kDa) and generally only have the HECT domain in common and so the same model, whereby the highly variable regions define substrate specificity, has been proposed. In the case of RSP5, it has been shown that the HECT domain by itself is enzymatically active, while its interaction with the binding site of its substrate Rpb1 is independent of this (Wang *et al.* 1999).

However, the modular model of USP structure/function may prove to be an oversimplification. In a study to determine the regions of FAF that are essential in its role in drosophila eye development, a two interesting pieces of genetic evidence were raised that do not support this model. Firstly, all protein domains along the entire length of the FAF protein are seemingly required for full activity as a series of deletion constructs fail to rescue the mutant eye phenotype (Chen and Fischer, 2000). Furthermore, analysis of 14 point mutant *faf* alleles with a maternal affect lethal phenotype show a distribution over the full length of FAF. Secondly the functionally active catalytic core is also insufficient to rescue the eye phenotype (Chen and Fischer, 2000) whereas the expression of *Fam* and the yeast USPs, USP2 and USP3 can substitute for endogenous FAF more effectively than most of the prementioned deletions (Chen *et al.* 2000, Wu *et al.* 1999).

It should also be noted that although protein binding can occur in these extension regions outside the catalytic core of USPs as in the case of mUBPy, several examples are documented where protein binding regions exist within the catalytic core, presumably binding to the unique sequences interspersed between the Cys, His, Asp and KRF boxes. A region of FAM containing the catalytic core (*Fam*-CAT 1476-1918 aa) binds both AF-6 and -catenin and its over-expression leads to their stabilisation (Taya *et al.* 1998, 1999). The catalytic core of yet another USP, Unp, binds Retinoblastoma tumour suppressor protein (Rb) and Rb family members p107 and p130 (Blanchette *et al.* 2001, Desalle *et al.* 2001). These examples raise the possibility that these catalytic cores and other regions may consist of multi-substrate binding domains, as is the organization of proteins such as CBP/p300 (Reviewed in Goodman and Smolik 2000).

## *Aims*

Of the 2554 amino acids of FAM, only a region that spans the catalytic core has been directly characterised. This region has been shown to bind its substrates -catenin and AF-6 (*Fam*-CAT: 1476-1918 aa) (Taya *et al.* 1998, 1999), and also Epsin (1554-1953aa) (Chen *et al.* 2003). By extrapolation from human USP9X (97% identity and 99% similarity to FAM), Doublecortin binds FAM in the last 256 amino acids of its C-terminal extension (2299-2554 aa) (Friocourt *et al.* 2005). The identity and function of the remainder of the FAM, some 1800 odd amino acids remain unclear largely due to the lack of significant homology to any other protein. Some recent work has predicted the 3D structure of some putative human USP9Y domains (Ginalski *et al.* 2004). They found that four cysteine residues (Cys-1726, Cys-1729, Cys-1773, and Cys-1776) in the Fingers domain of the catalytic core (1553-1996 aa) may coordinate a zinc ion. These cysteines form a putative zinc ribbon-like structure, absent from the crystal structure of HAUSP. Three previously uncharacterised long α-helical regions were also predicted in the N- and C-terminal extensions (71-868, 1008-1532, 2004-2476 aa), the most C-terminal presumably the Doublecortin interacting domain. Lastly a domain located in the N-terminus between two presumptive α-helical regions, appears to have a -grasp fold (884-971 aa) characteristic of ubiquitin-like proteins (*figure 1.13, includes the corresponding residue positions of*

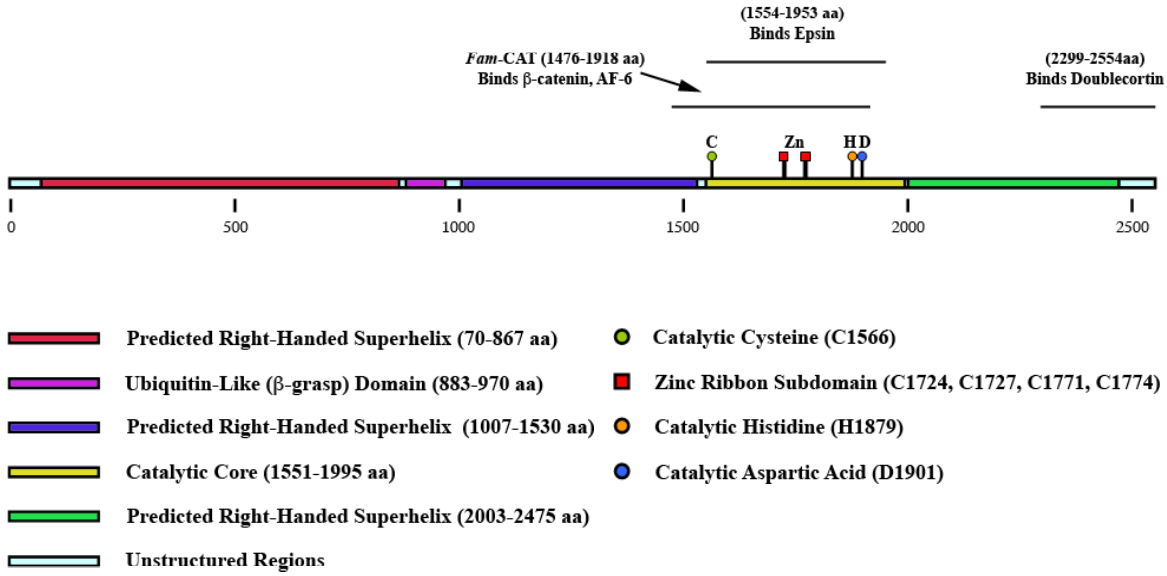



**Figure 1.13**

**Predicted Domain Structure of FAM**

Scale map of FAM (2554 aa) incorporating available structural and functional information. Vertical bars and numbers under the diagrammatic protein denote blocks of 500 amino acids. Functionally characterised regions (bars) include areas that bind -catenin and AF-6 (Taya *et al.* 1998, 1999), Epsin (Chen *et al.* 2003) and Doublecortin (Friocourt *et al.* 2005). Recent domain predictions are extrapolated from human USP9Y (Ginalski *et al.* 2004) by alignment to FAM.

*these putative domains for FAM*). It has been postulated that this ubiquitin-like domain corresponds to a distant homologue of other ubiquitin-like proteins and that it functions to target USP9Y to its specific cellular localisation (Ginalski *et al.* 2004).

Without any experimental structural data, the boundaries of these putative domains remain purely theoretical. Several attempts at expressing regions of FAM in various bacterial and mammalian systems have either failed to express or have been found in the insoluble fraction, presumably due to the disruption of a folding domain. Precise knowledge of FAM's domain structure would not only provide insight into regions of amino acids that hold no homology to any other known domains, but would also provide a valuable experimental tool for the design of FAM domain constructs to dissect FAM's functions. Use of individual domains would aid in the discovery of novel binding proteins, be it substrate or non-substrate proteins such as regulatory, localisation or ancillary factors. To this effect, expression of full-length FAM and subsequent partial proteolysis were undertaken to identify protease resistant regions of FAM that would theoretically correspond to folding domains.

Given the complex temporal and spatial expression of *Fam* and its orthologues throughout development, it has been proposed that FAM maybe involved in several developmental events. One way to investigate FAM's role in development is to search for developmentally relevant binding partners. Identification of such proteins would not only gain insight into the developmental events that FAM may regulate, but also shed light on the molecular mechanism. This thesis will also detail the establishment of a developmental screen that sought to identify dominant-negative developmental defects from the ectopic expression of regions of FAM in *Danio rerio* (zebrafish).

# CHAPTER 2

## MATERIALS AND METHODS

## *Abbreviations*

| | |
|---|---|
| Aa | Amino Acid |
| AF-6 | ALL-1 Fusion partner from chromosome 6 |
| ALL-1 | Acute Lymphoblastic Leukemia-1 |
| Amp | Ampicillin |
| APC | Adenomatous Polyposis Coli |
| ARM | Armadillo |
| BES | N,N-Bis-(2-hydroxyethyl)-2-aminoethanesulfonic acid |
| BCIP | 5-bromo-4-chloro-3-indolyl-phosphate |
| Bp | base pair |
| °C | Degrees Celsius |
| CAM | Calmodulin homology |
| CBC | Cap Binding Complex |
| cDNA | Complementary Deoxyribonucleic Acid |
| CNS | Central Nervous System |
| Comm | Commissureless |
| *Cul1* | Cullin homolog 1 |
| DCX | Doublecortin |
| Dub | Deubiquitylating enzyme |
| DFFRX | Drosophila Fat-Facets Related X |
| DFFRY | Drosophila Fat-Facets Related Y |
| DMEM | Dulbecco's Modified Eagles Medium |
| DMSO | Dimethyl Sulfoxide |
| Dpf | Days Post-fertilisation |
| DSL | Delta/Serrate/Lag2 |
| DTT | Dithiothreitol |
| Dub | Deubiquitylating enzymes |
| eGFP | Enhanced Green Fluorescent Protein |
| *E. coli* | *Escherichia coli* |
| ECL | Enhanced Chemiluminescence |
| EDTA | Ethylene Diamine Tetra Acetic acid |
| EGTA | Ethyleneglycol-bis-(2-aminoethyl)-N,N,N,N'-tetraacetic acid |
| ENTH | Epsin N-Terminal Homology |
| EPL cells | Early Primitive Ectoderm-Like Cells |
| ER | Endoplasmic Reticulum |
| ES cells | Embryonic Stem Cells |
| *Faf* | Drosophila *Fat Facets* |
| *Fam* | *Fat Facets in Mouse* (chromosome X) |
| *FamY* | *Fat Facets in Mouse* (chromosome Y) |
| *Fam*-CAT | *Fam* catalytic core (1476 – 1918aa) |

| | |
|---|---|
| FCS | Fetal Calf Serum |
| G | Gravitational forces |
| GPCR | G Protein-Coupled Receptor |
| GST | Glutathione-S-Transferase |
| HAUSP | Herpesvirus-Associated Ubiquitin-Specific Protease |
| HBP | Hrs-Binding Protein |
| HCL | Hydrochloric Acid |
| HCRs | Highly Conserved Regions |
| HEAT | Huntingtin-Elongation-A subunit - TOR |
| HECT | Homologous to the E6-AP C-Terminus |
| HEK293T | Human Embryonic Kidney 293T fibroblasts |
| Hepes | N-2-hydroxyethylpiperazine-N'-2-ethane sulphonic acid |
| His tag | Histidine tag |
| Hpf | Hours Post-Fertilisation |
| Hr | Hour |
| HRP | Horseradish Peroxidase |
| I Bα | Inhibitor of nuclear factor Bα |
| IBB | Importin- Binding |
| IPTG | Isopropyl- -D-Thiogalactopyranoside |
| kDa | Kilodalton |
| LOOPP | Learning, Observing and Outputting Protein Patterns |
| Lqf | Liquid facets (drosophila *epsin1*) |
| MALDI | Matrix-Assisted Laser Desorption/Ionisation |
| MAPK | Mitogen-Activated Protein Kinase |
| MDCKII | Madin-Darby Canine Kidney II |
| Min / ' | Minutes |
| mL | Millilitres |
| mM | Millimoles |
| M | Molar concentration |
| MQ | Milli-Q |
| mRNA | messenger RNA |
| MVB | Multivesicular Body |
| NBT | 4-nitro blue tetrazolium chloride |
| NF B | Nuclear Factor- B |
| NMR | Nuclear Magnetic Resonance |
| NP-40 | NonidetP-40 |
| NPC | Nuclear Pore Complex |
| NTMT buffer | NaCl, Tris, $MgCl_2$, Tween-20 buffer |
| OD | Optical Density |
| ODN | Oligodeoxynucleotides |
| OUT | Ovarian Tumour |
| PAGE | Polyacrylamide Gel Electorphoresis |
| PBS | Phosphate Buffered Saline |
| PBT | PBS + 0.1% Tween 20 |
| PCR | Polymerase Chain Reaction |
| PHD | Pleckstrin Homology Domain |
| PIP2 | Phosphatidylinositol 4,5-bisphosphate |
| PIP3 | Phosphatidylinositol 3,4,5-trisphosphate |
| PMSF | Phenyl Methyl Sulfonyl Fluoride |
| Ppm | Parts per million |

| | |
|---|---|
| PTC | Phenylthiocarbamide |
| QTOF2 | Quadrupole/time-of-flight |
| Rb | Retinoblastoma |
| RING | Really Interesting New Gene |
| RIP | Receptor-Interacting Protein |
| Robo | Roundabout |
| RPE | Retinal Pigment Epithelium |
| Rpm | Revolutions per minute |
| SCF | Skp1-Cullin-F-box protein |
| SDS | Sodium Dodecyl Sulphate |
| Sec | Seconds |
| siRNA | Short Interfering RNA |
| SNARE | Soluble NSF attachment protein receptors |
| snRNA | small nuclear RNA |
| TBS | Tris-buffered saline |
| TCA | Trichloroacetic Acid |
| TNF | Tumour Necrosis Factor |
| TTBS | Tris-buffered saline with 1% Tween-20 |
| Ubl | Ubiquitin-like |
| U-box | UFD2 homology |
| UBP | Ubiquitin specific proteases |
| UCH | Ubiquitin Carboxyl-terminal Hydrolases |
| µl | microliter |
| µm | micrometer |
| USP | Ubiquitin Specific Peptidase |
| *Usp9X* | Ubiquitin Specific Protease 9 on Chromosome X (Human) |
| *Usp9Y* | Ubiquitin Specific Protease 9 on Chromosome Y (Human) |
| *usp9* | Zebrafish Orthologue of *Fam(X)* |
| UTP | Uracil triphosphate |
| UV | Ultraviolet |

# *Materials*

Platinum Pfx DNA polymerase was obtained from Invitrogen (Mount Waverly, VIC), Expand High Fidelity PCR System and dNTPs from Roche Applied Sciences (Castle Hill, NSW), and custom primers from Geneworks (Adelaide SA). All Gateway related materials including enzymes and vectors were obtained through Invitrogen. All *Fam* cDNA templates were obtained from Dr. Stephen Wood (Child Health Research Institute, Adelaide SA). All restriction enzymes were obtained from New England Biolabs Inc. (Ipswich, MA, USA) and used according to the manufacturer's specifications. Benchmark Molecular Weight Markers and RPN800 Rainbow Molecular Weight Markers were obtained through Invitrogen and Amersham Biosciences (Castle Hill, NSW) respectively. Foetal Calf Serum (FCS), Dulbecco's Modified Eagle Medium (DMEM), Phosphate Buffered Saline (PBS) and antibiotics for tissue culture were obtained from CSL Biosciences (Parkville, NSW) and Gibco BRL (Invitrogen). All Zebrafish related materials were kindly and generously provided by Dr. Michael Lardelli (Department of Molecular Bioscience, University of Adelaide). All other materials were obtained from Sigma-Aldrich (Sydney, NSW) unless otherwise stated.

### ANTIBODIES

Monoclonal anti-V5 antibody was obtained from Invitrogen. Monoclonal anti-Myc antibody was a kind gift from Dr. Michael Lardelli and originally sourced from Sigma-Aldrich. Polyclonal anti-GST serum raised in rabbit was kindly donated by Dr. Stephen Rodda, (Department of Molecular Bioscience, University of Adelaide). Rabbit polyclonal anti-N1 FAM antibodies raised against a synthetic peptide corresponding to the first 20aa of murine FAM (TATTRGSPVGGNDNQGQAPC), denoted N1, which was produced by Susan Millard (Department of Molecular Bioscience, University of Adelaide). Goat anti-rabbit and rabbit anti-mouse secondary antibodies conjugated with HRP were obtained from DAKO (Carpinteria, CA, USA).

### BACTERIAL STRAINS

DH5α: supE44Δlac U169 (ϕ80 lacZ ΔM15) hsdR17 recA 1 endA 1 gyrA96 thi 1 relA1

BL21 (DE3): F- ompT hsdSB (rB-mB-) gal dcm lon- λ(DE3)

DB3.1: F- gyrA462 endA1 Δ(sr1-recA) mcrB mrr hsdS20(rB-, mB-) supE44 ara14 galK2 lacY1 proA2 rpsL20(Smr) xyl5 Δleu mtl1

DH10Bac: F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K λ⁻ *rps*L *nup*G /pMON14272 / pMON7124

## *Plasmids*

Much of the cloning presented in this thesis was performed using gateway technology (Invitrogen). Gateway technology allows the transfer of genes of interest in the context of an "entry clone" into a variety of "destination vectors" using site-specific recombination borrowed from λ phage. Destination vectors are available commercially and can be constructed by insertion of a cassette (in frame with any fusion tags) that contains the attR recombination sites, along with a *ccd*B gene for negative selection against non-recombinants in standard bacterial lines. The entry clones used in this research were constructed by PCR of the gene of interest with flanking attB sites to facilitate the "BP" reaction into the attP sites of the entry vector pDONR201 (*figure 2.1 A, B*). The resultant "entry clone" now contains attL sites for "LR" reaction transfer of the gene of interest into the attR sites of the destination vector. All gateway reactions were performed as directed by the manufacturer's protocols. In all instances, constructed entry clones used for experiments were verified by restriction digests and then sequenced with the pDONR sequencing primers (*table 2.1*) uncovering no errors.

Two different PCR kits were used according to the manufacturer's instructions. The Expand High Fidelity PCR System (Roche) and the Platinum Pfx DNA polymerase (Invitrogen) kit, which polymerise at 72°C and 68°C respectively. A general PCR program

**Figure 2.1**

**Entry Vector and the BP Reaction**

**A.** Vector map of entry vector pDONR201.
**B.** Diagram detailing the BP reaction to transfer a PCR product with attB sites into the attP sites of pDONR201 to form an entry clone.

(Adapted from the Invitrogen website (http://invitrogen.com))

**Vector Sequencing Primers**

| | |
|---|---|
| pDONR201 F | 5'-tcgcgttaacgctagcatggatctc-3' |
| pDONR201 R | 5'-gtaacatcagagattttgagacac-3' |
| M13/pUC F | 5'-gttttcccagtcacgac-3' |
| M13/pUC R | 5'-cgccagggttttcccagtcacgac-3' |

**Sense/Antisense Transcription**

| | |
|---|---|
| T7 | 5'-taatacgactcactataggg-3' |
| SP6 | 5'-cgatttaggtgacactatag-3' |

**Fam-Specific Primers Used for Sequencing Entry Clones**

| Primer | Region (bps) | Primer Sequence | Fam-N | Fam-C | Fam± |
|---|---|---|---|---|---|
| Fam1192-1175 | 1192-1175 | 5'-tcatcttgcaggattttt-3' | | | ✓ |
| SW67 | 722-741 | 5'-gggctaacgatctcatttac-3' | | | ✓ |
| SW11 | 1155-1174 | 5'-ctcttcacacagtgaaaaag-3' | ✓ | | ✓ |
| SW76 | 1903-1922 | 5'-cagtgatgatgtgcctgtag-3' | ✓ | | ✓ |
| SW13 | 2441-2461 | 5'-gcctgttttaagtggtattcc-3' | ✓ | | ✓ |
| Fam2901-2918 | 2901-2918 | 5'-gaatggtccgagtgttaa-3' | ✓ | | ✓ |
| Fam3487-3508 | 3487-3508 | 5'-gaaaacttatgccgccagatag-3' | ✓ | | ✓ |
| Fam4154-471 | 4154-471 | 5'-atttatgagaagaccaat-3' | ✓ | | ✓ |
| Fam4663-4680 | 4663-4680 | 5'-gctactggcttttcagac-3' | | | ✓ |
| SW46 | 5147-5167 | 5'-gatatgtctggggatgagaag-3' | | | ✓ |
| Fam5709-5729 | 5709-5729 | 5'-taattaaaaaattgcctcctg-3' | | | ✓ |
| Fam6280-6300 | 6280-6300 | 5'-cacaaggccccatcagattgt-3' | | | ✓ |
| Fam6938-6958 | 6938-6958 | 5'-cagctgctcaaactgagtgta-3' | | ✓ | ✓ |
| Fam7240-7260 | 7240-7260 | 5'-ctctgatgagaccgtcaaatt-3' | | | ✓ |
| SW86 | 7672-7691 | 5'-caacaattggtctcctccag-3' | | | ✓ |

*Primers were obtained from Dr. Stephen Wood and have been previously used (Diam, 1999)*

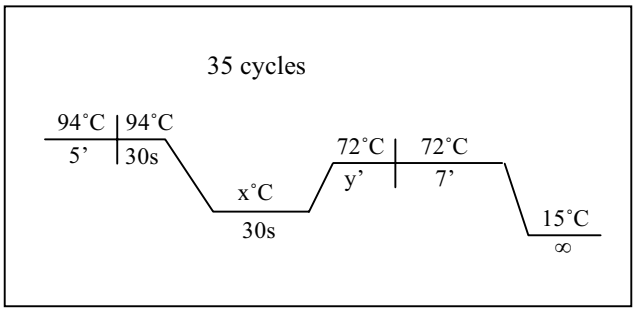
**Fam-Specific primers Used for PCR analysis of Fam-C-GST transposition**

| | | |
|---|---|---|
| F7 R | 6568-6549 | 5'-ggaattccataccaatcactggcagatc-3' |
| F9 F | 7421-7443 | 5'-cgcggatccgcgcataatgcccttaaaggaattcc-3' |

**Table 2.1: Primers for Sequencing and Transcription**

Table of primers used for sequencing and transcription. Primers used for sequencing of the large N, C-terminal and full length FAM constructs are indicted with a ✓.

used is as follows where x°C is the primers annealing temperature and y' is the extension time in minutes. The polymerisation temperature of the Expand High Fidelity PCR System is given as an example. Primer sequences and annealing temperatures used for cloning are summarised in *table 2.2*.




### USP9 SPECIFIC REGION FOR IN SITU HYBRIDISATION

A *usp9*-specific *in situ* probe was designed to hybridise to an 955 bp region of the gene and contains 161 bps of the 5' untranslated region and 794 bps of the coding region (corresponding to 196-1150 bps in relation to *Fam*'s Genbank entry U67874). Primers that are specific to both the *Fam* and *usp9* sequence were identified from a series of pre-existing *Fam* primers and subsequently used in PCR reactions (*table 2.2*) and were chosen based on sequence data obtained from blast searches of the zebrafish genome. The PCR product was then cloned into pGEMTeasy (Promega, Madison, WI, USA) and sequenced to check fidelity, orientation *usp9*-specificity. The Expand High Fidelity PCR System was used with a template of zebrafish cDNA library (a kind gift from Dr. Michael Lardelli). An annealing temperature of 52°C and an extension time of 2.5 mins were used.

| Fragment | Primer | Primer Sequence | Length (bp) | Annealing T°C | Region (bps) | Size (bps) | Region (aa) | Size (aa) |
|---|---|---|---|---|---|---|---|---|
| *usp9 In situ* Probe | Famz196-214 | ggttatgcaatggtctctg | 19 | 51 | 196-1150 | 955 | 1-265 | 265 |
| | Famz1150-1133 | gaaactcatagcactgtac | 19 | 49 | | | | |
| TDE | TDEE F | GW5'-ct-actgatgaagagctgaagaaag | 53 | 57 | 1223-1999 | 777 | 290-548 | 259 |
| | TDEE R | GW3"-t-cttttgggtgtcccggtc | 48 | 55 | | | | |
| ERL | ERL F | GW5'-ca-gaacggcttaactttcttagg | 52 | 57 | 2321-2827 | 507 | 656-824 | 169 |
| | ERL R | GW3'-c-tttcaaacgatcaaagcaagac | 52 | 59 | | | | |
| RKE | RKE F | GW5'-at-aggaaagagtacaatattggtg | 53 | 59 | 5261-6226 | 966 | 1636-1957 | 322 |
| | RKE R | GW3'-t-gtccattcgttcataaaaaagt | 52 | 55 | | | | |
| NPF | NPF F | GW5'-ca-aatccttttggtgatcctaac | 52 | 59 | 7124-7786 | 663 | 2257-2477 | 221 |
| | NPF R | GW3'-c-ctcctcaggacagagttcacaag | 53 | 61 | | | | |
| eGFP | eGFP F | GW5'-cc-atggtgagcaagggc | 46 | 51 | 613-1410 | 830 | 1-266 | 266 |
| | eGFP R | GW3'-a-gttatctagatccggtgg | 48 | 51 | | | | |
| Fam-DUB | Fam-DUB F | GW5'-cc-cgtccaccaaaaggatttg | 50 | 59 | 5006-6355 | 1350 | 1551-2000 | 450 |
| | Fam-DUB R | GW3'-t-gtgcatgaactgaacattctgc | 52 | 61 | | | | |
| HAUSP-DUB (USP7) | HAUSP-DUB F | GW5'-ca-aagaagcacacaggctacg | 50 | 61 | 820-1878 | 1059 | 208-560 | 353 |
| | HAUSP DUB R | GW3'-c-ttcctgccgctccttccg | 48 | 59 | | | | |
| Fam-N | Fam F | GW5'-gt-atgacagccacgactcgt | 49 | 57 | 356-5020 | 4665 | 1-1555 | 1555 |
| | Fam-N R | GW3'-g-gggtccaacgggtggc | 46 | 55 | | | | |
| Fam-C | Fam-C F | GW5'-ac-aaccgaatgcagtacagtc | 50 | 57 | 6371-8035 | 1665 | 2006-2559 | 554 |
| | Fam R | GW3'-a-ctgatccttggtctgagga | 49 | 55 | | | | |
| Fam (SSSRF+) | Fam F | GW5'-gt-atgacagccacgactcgt | 49 | 57 | 356-8035 | 7680 | 1-2559 | 2559 |
| | Fam R | GW3'-a-ctgatccttggtctgagga | 49 | 55 | | | | |
| Fam (SSSRF-) | Fam F | GW5'-gt-atgacagccacgactcgt | 49 | 57 | 356-8020 | 7665 | 1-2554 | 2554 |
| | Fam R | GW3'-a-ctgatccttggtctgagga | 49 | 55 | | | | |

GW5' sequence: ggggacaagtttgtacaaaaaagcaggct
GW3' sequence: ggggaccactttgtacaagaaagctgggt

**Table 2.2: Primers for PCR**

Table of primers used for cloning. Region in base pairs (bp) are given in reference to genbank entries; *Fam* (acc. No. U67874), HAUSP (acc. No. NM003470). Region in amino acid residues (aa) are given in respect to the translated protein.

## *Entry Clones*

### HIGHLY CONSERVED REGIONS (HCRS)

Entry clone construction of FAM's four highly conserved regions; TDEE, ERL, RKE and NPF (see pg. 56 for a description) was achieved by amplifying regions of full-length *Fam* SSSRF+ cDNA (290-548, 356-824, 1636-1957, 2257-2477 respectively) by PCR with the Expand High Fidelity PCR System according to the manufactures instructions. An annealing temperature of 55°C and an extension time of 1.5 mins were used.

### EGFP

A plasmid containing eGFP was kindly donated from Dr. Gavin Chapman (Karolinska Institute, Sweden). This plasmid, pEGFP-C3FC is a destination vector, a modified version of pEGFP (Clontech Laboratories, Inc., Palo Alto, CA, USA) that has had Gateway cassette frame C inserted into the SmaI site of the multiple cloning site. The gateway cassette was removed by digestion with SacII and BamHI, blunted and religated. This strategy allowed retention of the 3' multiple cloning site with three frames of stop codons. Not only was the removal of the cassette necessary to provide a stop codon for eGFP, but it also allowed flexibility for any future cloning into the multiple cloning sites. As such, this plasmid can only be used as a C terminal fusion when combined with destination vectors containing N terminal tags, such as GW-pCS2+MT. A Gateway compatible eGFP entry clone was constructed by amplification of the eGFP gene from the pre-mentioned pEGFP derived plasmid with similar PCR reaction conditions used for the pre-mentioned HCRs and then inserted into pDONR201 via a BP reaction and verified by restriction digests and sequenced.

### USP CATALYTIC CORES

Entry clone construction of *Fam*-DUB was achieved by PCR off full-length *Fam* SSSRF-cDNA template with the Platinum Pfx DNA polymerase PCR system according to the manufactures instructions. An annealing temperature of 59°C and an extension time of 2 mins were used.

Entry clone construction of HAUSP-DUB was achieved by PCR off *HAUSP* (*USP7*) cDNA template (accession number Z72499, kindly donated by Dr. RD Everett, Medical Research Council Virology Unit, Scotland) with the Expand High Fidelity PCR system according to the manufactures instructions. An annealing temperature of 55°C and an extension time of 3 mins were used.

### *FAM* N AND C-TERMINAL EXTENSIONS

Entry clone construction of *Fam* N and C terminal extensions was achieved by PCR off full-length *Fam* SSSRF+ cDNA template using the Platinum Pfx DNA polymerase PCR system according to the manufactures instructions. For the *Fam* N-terminal extension, an annealing temperature of 55°C and an extension time of 5.5 mins was used. For the *Fam* C-terminal extension, an annealing temperature of 55°C and an extension time of 2.5 mins was used. In addition to verification by restriction digests and sequencing with pDONR sequencing primers, several other *Fam*-specific primers were used for full sequence coverage (*table 2.1*). No errors were detected.

### FULL-LENGTH *FAM* (SSSRF±)

Entry clone construction of both isoforms of full-length *Fam* was achieved by PCR off *Fam* SSSRF+ or *Fam* SSSRF- cDNA template with the Expand High Fidelity PCR system according to the manufactures instructions. An annealing temperature of 57°C and an extension time of 8.5 mins were used. In addition to verification by restriction digests and sequencing with pDONR sequencing primers, several other *Fam*-specific primers were used for full sequence coverage (*table 2.1*). Conservative errors were detected (figure 4.7).

### *FAM*-CYS MUTANT (C1566S)

An entry clone containing full-length *Fam* (SSSRF+) with the critical cysteine of the catalytic triad mutated to a serine was obtained from Susan Millard (Department of Molecular Bioscience, University of Adelaide). Its construction involved site directed mutagenesis on the *Fam* (SSSRF+) entry clone described above.

## *Destination Vectors*

### GWPCS2$^{+}$MT

A vector for *in vitro* transcription pCS2+MT was kindly obtained by Dr. Michael Lardelli and allows mRNA transcription and *in vivo* translation with an N terminal 6-myc tag. This vector was made Gateway compatible by insertion of a Gateway cassette (reading frame A) into blunted StuI and XbaI sites within the multiple cloning region (*figure 2.2*).

### COMMERCIALLY AVAILABLE DESTINATION VECTORS

The pDEST15 vector (which generates a bacterially expressed N-terminal GST fusion), the pEF-DEST51 vector (which adds a C-terminal V5 6x His tag fusion for mammalian expression) and the pDEST20 vector (which adds an N-terminal GST fusion for baculoviral expression) were obtained through Invitrogen (*figure 2.3*)

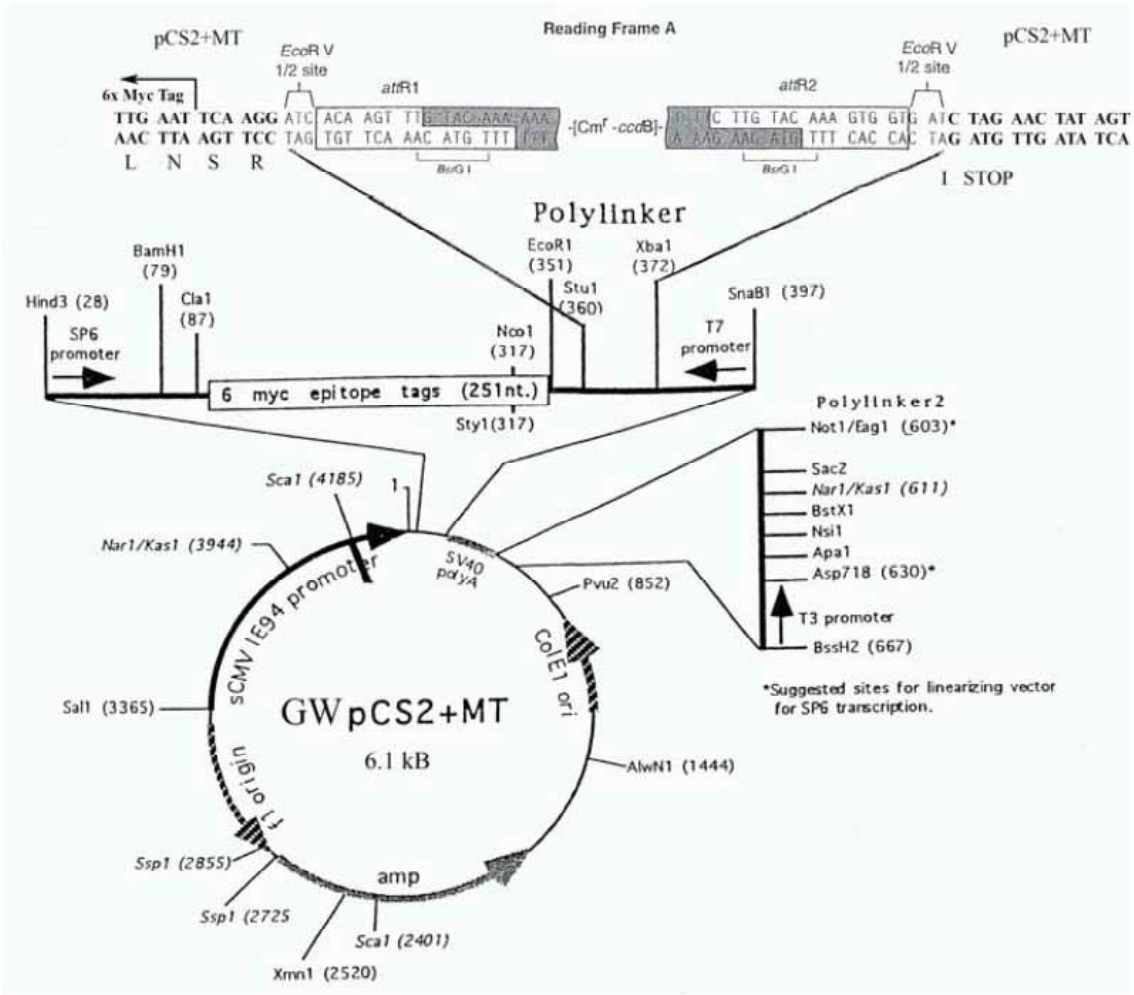



**Figure 2.2**

**Constructed Destination Vector GWpCS2+MT**

Vector map of GWpCS2+MT constructed by ligation of a gateway cassette (*reading frame A*) into blunted *Stu*1 and *Xba*1 sites.

NOTE: This image is included on page 33b in the print copy of the thesis held in the University of Adelaide Library.

**Figure 2.3**

**Commercially Available Destination Vectors**

**A.** Vector map of pDEST15 for the bacterial expression of N-terminal GST fusion proteins
**B.** Vector map of pEF-DEST51 for the mammalian expression of C-terminal V5 6xHis tagged fusion proteins
**C.** Vector map of pDEST20 for the transposition of the cloned gene of interest into a DH10Bac (Invitrogen) bacmid. Recombinant bacmid is expressed in insect cells and produces N-terminal GST fusion proteins.

(Figures adapted from the Invitrogen website (http://invitrogen.com))

## *Expression Clones*

Expression clones were constructed via an LR reaction between entry clone and destination vector (*table 2.3*). Correct insertion was verified by diagnostic restriction digests.

**Expression Clones**

| Name | Entry Clone | Destination vector |
|---|---|---|
| TDEE-MT | TDEE | GWpCS2$^{+}$MT |
| ERL-MT | ERL | GWpCS2$^{+}$MT |
| RKE-MT | RKE | GWpCS2$^{+}$MT |
| NPF-MT | NPF | GWpCS2$^{+}$MT |
| eGFP-MT | eGFP | GWpCS2$^{+}$MT |
| TDEE-GST | TDEE | pDEST15 |
| ERL-GST | ERL | pDEST15 |
| NPF-GST | NPF | pDEST15 |
| RKE-GST | RKE | pDEST15 |
| TDEE-V5 | TDEE | pEF-DEST51 |
| ERL-V5 | ERL | pEF-DEST51 |
| NPF-V5 | NPF | pEF-DEST51 |
| RKE-V5 | RKE | pEF-DEST51 |
| *Fam*-DUB-GST | *Fam*-DUB | pDEST15 |
| *HAUSP*-DUB-GST | *HAUSP*-DUB | pDEST15 |
| *Fam*-N-GST | *Fam*-N | pDEST20 |
| *Fam*-C-GST | *Fam*-C | pDEST20 |
| *Fam*$^{+}$-GST | *Fam* SSSRF+ | pDEST20 |
| FAM-Cys-GST | FAM-Cys (C1566S) | pDEST20 |

**Table 2.3:** Table of Expression Clones Generated through the Gateway transfer (LR reaction) of an entry clone into a destination vector.

## *Plasmid Preparation*

Plasmid DNA minipreps were performed using alkaline lysis as outlined in "Molecular Cloning: A Laboratory Manual" (Sambrook and Russell, 2001). Larger scale plasmid DNA preparations were made using the alkaline lysis based Quantum Prep Plasmid Midiprep and Quantum Prep Plasmid Maxiprep kits from Biorad (Regents Park, NSW).

### RECOMBINANT BACMID PREPARATION

Genes in the pDEST20 vector (for N-terminal GST fusion for baculovirus expression) (*figure 2.3 C*) were heat-shocked transformed into DH10Bac competent cells (Invitrogen) and transposition of the gene into the bacmid was selected for on plates containing: 50μg/ml kanamycin (Bacmid Selection), 7μg/ml gentamicin (expression clone selection), 10μg/ml tetracycline (Helper plasmid selection), 100μg/ml X-gal (blue/white colony chromogenic substrate), 40μg/ml IPTG (*lac*Zα induction). White colonies represent instances where the gene cassette has transposed (with the helper plasmid encoded transposase) from the expression clone to the bacmid, disrupting the *lac*Zα gene, interfering with X-gal processing. White colonies were re-streaked to ensure genetic homogeneity. DNA was prepared from cultures of the white colonies and analysed by electrophoresis and PCR. Electrophoretic gels of the colonies show slow migration of extremely large DNA bands that correspond to bacmid DNA (>135 kb). PCR analysis was performed to check for transposition of the insert. Flanking the transposition sites in the bacmid vector are M13/pUC sites. A PCR strategy exploiting these sites as well as gene-specific sequences was conducted on of the *Fam*-N/C-GST Bacmid with the Platinum Pfx DNA polymerase PCR system according to the manufactures instructions to verify correct insertion (*table 2.4*). An annealing temperature of 55°C and an extension time of 4 mins was used. PCR of the *Fam*-GST Bacmid was performed with the Expand High Fidelity PCR system according to the manufactures instructions. An annealing temperature of 49°C and an extension time of 6 mins was used.

| Fam-GST | | | Fam-N-GST | | | Fam-C-GST | | |
|---|---|---|---|---|---|---|---|---|
| Forward | Reverse | Predicted Size | Forward | Reverse | Predicted Size | Forward | Reverse | Predicted Size |
| M13 F | M13 R | × | M13 F | M13 R | × | M13 F | M13 R | × |
| M13 F | Famz1150-1133 | 3.1 kb | M13 F | Fam1192-1175 | 3.2 kb | M13 F | F7 R | 2.5 kb |
| NPF F | M13 R | 1.5 kb | Fam4663-4680 | M13 R | 1 kb | F9 F | M13 R | 1.2 kb |
| RKE F | RKE R | 1 kb | | | | | | |

M13 F/R sequencing:
Bacmid alone: 300bps
Transposition without insert: 3 kb
Transposition with insert: too big

**Table 2.4:** Sequencing primers and predicted PCR product sizes used to verify transposition of inserts from pDEST20 into the DH10Bac bacmid. × denotes fragments that should not give rise to fragments sizes listed below the table.

## *Cell culture and 293T transfections*

Human Embryonic Kidney 293T fibroblasts (HEK293T) cells were grown in DMEM containing 10% FCS, penicillin (50 units/ml) and streptomycin (50μg/ml). All cells were incubated in an air:10% $CO_2$ atmosphere at constant humidity. Cells for transfection were passaged the day prior to the transfection and plated at 5 x $10^5$ cells per 10cm dish, such that they would be at 20% confluence at the time of transfection. GWpCS2+MT vectors (6x Myc tag) were transiently transfected into 293T cells by electroporation of 20μg of linearised plasmid DNA. pEF-DEST51 vectors (V5-His tag) were transiently transfected by $CaPO_4$ as previously described in "Cells, a Laboratory Manual – Volume 2: Light Microscopy and Cell Structure" (Spector *et al.* 1998).

## Whole-Mount In Situ Hybridisation

A *usp9*-specific *in situ* probe was designed and cloned into the pGEMTeasy vector (Promega, Madison, WI, USA) as previously described. To produce linear, RNAase-free cDNA template for *in vitro* transcription, another PCR reaction was performed using M13 primers to amplify the *usp9* sequence as well as flanking vector sequence, which contains SP6 and T7 transcriptional start sites in opposing orientations. Due to the orientation of the insert, Antisense RNA was produced with SP6 primers while sense RNA from T7 (*table 2.1*).

For whole-mount in situ transcript hybridisation, zebrafish embryos were raised at 28ºC and staged as previously described (Kimmel *et al.* 1995). 24 hours post-fertilisation (hpf) up to 5 days post-fertilisation (dpf) embryos were treated with Phenylthiocarbamide (PTC) to inhibit their pigmentation. A 20mM stock solution in 70% DMSO (1:200 in embryo media) was applied to embryos whose chorions were punctured with hypodermic needles. Media and PTC was changed daily. *In situ* transcript hybridisation was performed as described by Jowett (1997) using single-stranded RNA probes labelled with digoxigenin-UTP (Roche, Basel, Switzerland). Photos were taken under Normaski optics under a light microscope.

## mRNA Injections Into Zebrafish Embryos

mRNA for injection into zebrafish embryos was *in vitro* transcribed using the mRNA mMessage mMachine kit (Ambion Inc., Austin, TX, USA) as per manufacturer's instructions. 1 μg of template linearised plasmid (HCRs were linearised with NotI) was used in each reaction. After precipitation, mRNA was resuspended in 20μl $H_2O$. In all instances, yields of mRNA were estimated at 500ng/μl. The RNA was then aliquoted into 5μl volumes (to minimise freeze thaw cycles) at dilutions of 1:10 (50ng/μl) and 1:5 (100ng/μl).

Pre-prepared injection needles were used to draw up aliquots of mRNA by capillary action and then loaded on to a gas pump injection apparatus positioned under a light microscope. With the use of fine forceps, individual 2-cell stage embryos were penetrated first through the chorion, then the yolk sac, then finally into one of the cells. A small amount of mRNA was then deposited within the cell and the needle removed. Embryos were then placed into a Petri dish of embryo media and allowed to develop at 28.5°C for the required amount of time. After observation, embryos were fixed in PBS 4% formaldehyde.

## *Protein Processing*

### BACTERIAL INDUCTION

GST fusions were expressed in the *E. coli* BL21 (DE3) strain. Pilot studies of induction and solubility of fusion proteins at 37°C and 25°C were conducted from starter cultures, diluted 1:100 into 50 ml cultures (luria broth, 50 μg/ml amp). Although aiming to induce at OD 0.6-0.8, due to the unusually slow growth of the HCR-GST cultures, induction was performed after 2-4hr of growth with [0.2mM] IPTG for 2.5-4hrs. Cells were lysed by sonication 3x 30 sec in 2mls TTBS with addition of PMSF before, between and after each interval to a final [1mM]. Lysed cells were then spun at 3000xg for 10 min at 4°C. The supernatant was removed and the pellet resuspended in 2mls TBS. Uninduced, induced, soluble and insoluble samples were analysed by SDS-PAGE. Large scale inductions were conducted at 37°C and 25°C using essentially the same protocol as for the pilot studies except; 400ml cultures were used, lysis was performed in 20ml TTBS by French Press and lysed cells were spun 3000xg for 20 min at 4°C.

Uninduced and Induced samples for gel loading were prepared by removing 2mls and 1ml respectively from the appropriate bacterial culture followed by spinning down a pellet and removing the supernatant. The pellet was then resuspended in 100μl of 2x protein load buffer, of which 10μl was loaded. The Soluble and Insoluble samples for gel loading were prepared from 20μl each of the prepared protein solutions, mixed with 100μl of 2x protein

load buffer, of which 10μl was loaded. All samples were boiled at 100°C for 5 minutes prior to loading.

The pDEST15 vector expresses GST fusions under the T7 promoter. The original bacterial strain designed for its induced expression is the BL21-SI (Gibco BRL, Invitrogen), which transcribes a chromosomal copy of T7 RNA polymerase (DE3 lysogen) in response to 0.3 M NaCl instead of IPTG. BL21-SI is an expression host designed for high-level, tightly regulated expression from the *pro*U, salt-inducible promoter. However, the BL21-SI strain was discontinued and not available for use at the time of experimentation. The vector can be expressed in a new strain called BL21-AI (Invitrogen) which is induced with L-arabinose and repressed by glucose (Invitrogen, 2004), however others have continued to use this vector with traditional IPTG induction (Evans *et al.* 2003).

### PREPARATION OF MAMMALIAN CELL LYSATES

Cells were harvested and lysed by resuspension in 1-2 pellet volumes of mammalian lysis buffer (0.5% triton X, 60mM KCl, 5mM $MgCl_2$, 2mM EDTA, 2mM EGTA, 30mM Hepes pH 7.4, 1mM $Na_3Vo_4$, 50mM NaF, 2mM PMSF, 10 μg/ml aprotinin), incubation with gentle rocking at 4°C for 30 mins, spin at 14,000 rpm in a microfuge, and removal of supernatant lysate. Protein concentration was assayed with Bradford reagent (Biorad).

### PREPARATION OF INSECT CELL LYSATES

Frozen pellets of cultures transfected with recombinant bacmid were obtained from Dr. Sassan Asgari. Cell pellets were resuspended in five volumes of lysis buffer per gram of cells with one "Complete Mini" protease inhibitor cocktail tablet per 10ml (Roche Applied Sciences). Insect cell lysis buffer (50mM Tris-HCL (pH 8.5), 5mM 2-mercaptoethaol, 100mM KCL, 1mM PMSF, 1% NP-40 at 4°C). Tubes were inverted for 1 minute to lyse the cells and then cellular debris removed by centrifugation at 10,000 x g for 10 minutes. The supernatant containing soluble protein was then transferred to a new tube. The pellet was resuspended in TTBS to constitute the insoluble fraction.

### PROTEIN PURIFICATION

Soluble fractions filtered with a 0.45 μm filter and bound to glutathione agarose. Bound proteins were eluted off the column with reduced glutathione (10 mM pH8.0) in TBS.

Samples were then concentrated to 2 mls and buffer exchanged in to binding buffer (50mM Tris/HCL at pH 8.0, 1 mM EDTA, 0.1% Triton X-100, 0.1% CHAPS). The protein sample was then concentrated with a Macrosep centrifugal device (Pall Life Sciences, Lane Cove, NSW) to around 6 mg/ml.

### PARTIAL PROTEOLYSIS

Partial proteolysis of FAM-Cys-GST was carried out with Chymotrypsin sequencing grade, Endoproteinase Glu-C (Protease V8) Trypsin, modified, sequencing grade (Roche Diagnostics, Mannheim, Germany) in their respective protease specific buffers and temperatures (See below) with the addition of DTT [1mM]. Reactions were stopped by the addition of PMSF [5mM] for 15 minutes on ice before addition of load buffer and boiling prior to gel loading.

Trypsin digestion buffer: TBS pH 8 @ 37°C
V8 protease digestion buffer: 100mM $NH_4HCO_3$ pH 7.8 @ 25°C
Chymotrypsin digestion buffer: 100mM Tris-HCl, 10mM $CaCl_2$ pH 7.8 @ 25°C

### PROTEIN PRECIPITATION

Scaled-up partial proteolysis reactions required the proteins to be precipitated in order to achieve a suitable volume for loading onto a gel. This was achieved by the addition of TCA (trichloroacetic acid) to a final concentration of 10%, followed by incubation of the sample on ice for 30 minutes to allow the protein to precipitate. Samples were then centrifuged at 14,000 rpm in a microfuge for 15 minutes, washed in an equal volume of 100% EtOH centrifuged for a further 15 minutes and air-dried. Samples then had load buffer added and were boiled at 100% for 5 min.

### PROTEIN ELECTROPHORESIS

Proteins samples and prestained RPN800 molecular weight size markers (Amersham Biosciences) or Benchmark molecular weight size markers (Invitrogen) were subjected to SDS-PAGE at 40mA on either 7.5% or 12.5% polyacrylamide resolving gels poured and

electrophoresed using the Miniprotean III system (Biorad) or large gels at 20mA on the SE 600 series electrophoresis unit (Hoefer).

## *Protein Detection*

### WESTERN BLOTTING

Proteins were electrophoretically transferred to nitrocellulose membrane (Biotrace NT; Pall Corporation, NY, USA) using a semi-dry transfer apparatus (Biorad) for 1 hour at 80mA. Membranes were immunoblotted with monoclonal or polyclonal antibodies as indicated in figure legends. Primary antibodies were detected with HRP conjugated secondary antibodies and visualised by ECL (Enhanced Chemi-Luminescence). Chemiluminescent bands were detected with SuperRX Fuji Medical X-ray Film (Fuji Photo Film, Tokyo, Japan). Chemiflorescent bands were detected using a Typhoon 8600 variable mode imager (Amersham Biosciences).

### SILVER STAIN

Polyacrylamide gels (0.75mm) were first fixed in fixative solution (10% Acetic Acid, 40% EtOH) for 30 minutes to overnight. Gels were incubated for 30 minutes to overnight in incubation solution (0.5M Sodium acetate (anhydrous), 8mM Sodium thiosulphate, 0.13% Glutaraldehyde, 30% EtOH (added after the other reagents were dissolved)). Gels were then washed three times for 5 minutes in MQ water. Gels then were incubated in silver solution (0.1% silver nitrate, 0.004% formaldehyde) for 40 minutes. Gels were then briefly rinsed in water and then developed in developing solution (2.5% sodium carbonate, 0.004% formaldehyde) for up to 15 minutes. Staining was stopped in stop solution (1.46% EDTA.$Na_2$.$2H_2O$) for 5-10 minutes. Incubation, silver and developing solutions were made just prior to use.

### COLLOIDAL COOMASSIE STAINING

After electrophoresis, gels were rinsed in water and then washed in water for a further 10 minutes. The gels were then fixed with 20% (w/v) Trichloroacetic Acid (TCA) overnight at room temperature to fix the proteins. After removal of the fixing solution, gels were covered with the colloidal Coomassie stain (0.1% (w/v) Coomassie blue in 20% (w/v) TCA) for 3 to 4 hours with gentle agitation. After aspiration of the stain, the acid was washed out of the gel with several changes of water over ~2 hours to enhance contrast. Gels stored in 5% acetic acid.

## *Mass Spectrometry*

The scaled-up digest gels containing the partial proteolysis products were submitted to the Hanson Institute's Protein Laboratory for mass spectrometry analysis at the Institute of Medical and Veterinary Science. The researchers there were responsible for the selection and excision of the correct bands from the gels guided by a provided scanned printout of the bands, as well as making decisions on whether particular bands were suitable for analysis. For instance, when excising bands from the chymotrypsin 1:500 lane (*figure 4.13 C*), care was taken to avoid cutting across a smear that runs vertically down the lane. Bands of interest were excised from the gel and the protein was treated with iodoacetamide (S-carbamidomethylation) prior to its full digestion with trypsin. The resultant peptides from bands were resolved through a short C18 reverse phase silica column into the Micromass quadrupole/time-of-flight (QTOF2) mass spectrometer for matrix-assisted laser desorption/ionisation (MALDI) peptide mapping. The eluted peaks were analysed and the monoisotopic ($C^{12}$, $H^{1}$, $N^{14}$, $O^{16}$, and $S^{32}$) forms of the component ions and their charge states were measured. Spectra were recorded as abundance versus mass/charge ratio (Thomson, Th), and were calibrated against a standard solution including glutathione, apomyoglobin and [Glu1]-Fibrinopeptide B. Further calibration was achieved through identifying a trypsin autolysis product, which was used as a lock mass. The raw data from each band was then transformed with MAXENT-3 which resolves the observed masses back to a single charge state corresponding to the mass of the peptide plus one proton for ease of analysis.

# Computer Analyses

## ASSEMBLY OF *USP9* SEQUENCE

Predicted zebrafish *Fam* sequence (*usp9*) was obtained by a BLAST search of the Zv4 assembly of zebrafish genomic sequence at the Sanger Institute (http://www.ensembl.org/Multi/blastview?species=Danio_rerio). Where amino acid residues differed from the pufferfish and other vertebrate orthologues, the sequence was verified by examining several electrophoretograms available from the Ensembl Trace Server (http://trace.ensembl.org/perl/ssahaview?server=danio_rerio).

To define the putative coding sequence of zebrafish *Fam* (*usp9,* Acc. no. DQ086492), a predicted mRNA sequence from the Zv4 assembly of zebrafish genomic sequence by the Sanger Institute (http://www.ensembl.org/Multi/blastview?species=Danio_rer-io) was aligned with the known *Fam* cDNA sequences from: Mouse *Famx* (DQ086491) and *Famy* (Acc. no. AJ307017) and Humans *Usp9x* (Acc. no. X98296) and *Usp9y* (Acc. no. Y13618), and with predicted mRNA sequences from *Fugu rubripes* (Sanger centre SINFRUG00000128324), *Gallus gallus* (Sanger centre ENSGALG00000016236), *Xenopus tropicalis* (Sanger centre ENSXETG000000154-89), *Canis familiaris* (Sanger centre ENSCAFG00000014207) and *Rattus norvegicus* (Sanger centre ENSRNOG00000003261). When positions of variation in the predicted zebrafish sequence relative to both the *Fugu rubripes* sequence and the other sequences were identified, or where, (in one instance), an intron-exon boundary was not conserved relative to mouse *Fam*, the variant sequence was checked by examination of electrophoretograms from the zebrafish genome project (Ensembl Trace Server; http://trace.ensembl.org/perl/ssahaview?server=danio_rerio). In this way, all uncertainties in the *usp9* open reading frame were resolved.

## PHYLOGENETIC ANALYSIS

Phylogenetic analysis was performed using the BioManager facility provided by the Australian National Genomic Information Service (ANGIS, http://www.angis.org.au/). The protein multiple sequence alignment shown in *figure 3.4* was constructed using Gendoc

Pairwise alignment (Nicholas, 1997) with default parameters, and was prepared for publication using the GenDoc program (Nicholas, 1997). The phylogenetic tree was constructed from known and predicted FAM orthologue amino acid sequences using Protdist. The Protdist distance matrix was generated under the Dayhoff PAM matrix method of amino acid substitution, and a phylogenetic tree subsequently constructed using the 'Neighbour Joining Method' (*figure 3.5*). Bootstrap analysis was conducted using Seqboot (Felsenstein, 1989) to produce 1,000 resampled datasets that were then analysed with Protdist (Felsenstein, 1989). Bootstrap values were finally generated using the Consense program (Felsenstein, 1989), and are shown in *figure 3.5*.

### STATISTICAL ANALYSIS OF INJECTION PHENOTYPES

Embryos injected with FAM HCR mRNA were scored according to the presence or absence of several observed phenotypes. The data was then processed by cross-tabulating embryo phenotype against the type of injected HCR mRNA. Included in the cross-tabulation were the negative controls of eGFP and $H_2O$. The data was then assessed for significant differences with a Pearson Chi-Square test to obtain a p-value. All statistical analyses were performed with SPSS version 10.

### FAM PEPTIDE MASS ANALYSIS

The raw data from the mass spectrometry analysis of each band was transformed with MAXENT-3 which resolves observed masses back to a single charge state (monoisotpic masses $[M^+H]^+$) corresponding to the mass of the peptide plus one proton for ease of analysis. This data was used to query the FAM-Cys-GST sequence in Findmod (http://ca.expasy.org/cgi-bin/findmod_form.pl) to find peptides that matched the FAM sequence. Search parameters included the likely carbamidomethyl modification to cysteine (from reaction of the band with iodoacetamide prior to full digestion with trypsin) and oxidation of methionine. Matching peptides were found with an error of ±50 ppm (parts per million) and allowed for up to 2 missed cleavages (Appendix A).

The occurrence of the matched peptides was manually plotted against a FAM-Cys-GST sequence that had been theoretically digested with trypsin (http://ca.expasy.org/cgi-bin/peptide-mass.pl). This table (Appendix B) shows the potential carbamidomethyl

modification to cysteine and oxidation of methionine modifications but does not include peptides < 500 Da.

### SECONDARY STRUCTURE PREDICTION

The primary sequence of both FAM (SSSRF+) and FAM-GST (SSSRF+) (including the additional residues in the linker between FAM and GST, and vector encoded C-terminal residues before the vector stop sequence) were analysed by the JUFO server (Meiler *et al.* 2002) found at www.jens-meiler.de. This server offers a protein secondary structure prediction from its primary sequence only. A neural network was trained with an amino acid property profile and the position based scoring matrix of a blast run (as used by Jones DT, 1999). It achieves a three state prediction accuracy of about 75%.

### INTRINSIC PROTEIN DISORDER, DOMAIN & GLOBULARITY PREDICTION

The FAM-GST sequence was analysed with GlobPlot version 2.2 (http://globplot.embl.de/) using default settings and the Russell/Linding definition (Linding *et al.* 2003).

### STATISTICAL ANALYSIS OF PROTEIN SEQUENCES

The primary sequence of FAM (SSSRF+) was analysed by the ISREC-Server (Version of April 11, 1996, http://www.isrec.isb-sib.ch/software/SAPS_form.html) to give a SAPS (Statistical Analysis of Protein Sequences) report. SAPS evaluates by various statistical criteria, a wide variety of protein sequence properties including composition, charge distribution, distribution of other amino acid types, repetitive structures, mulitplets, periodicity and spacing (Brendel *et al.* 1992).

### FOLD PREDICTION

Predicted domain sequences were analysed for potential structural similarities to other known folds by the LOOPP (Learning, Observing and Outputting Protein Patterns) server (version 3.0, http://cbsuapps.tc.cornell.edu/loopp.aspx). LOOPP is a fold recognition program that collates straightforward sequence alignments, sequence profiles, threading, secondary structure and exposed surface area predictions to merge them into a single score,

and generate atomic coordinates based on an alignment into a homologue template structure (Teodorescu *et al.* 2004).