CHAPTER 1: INTRODUCTION

THE UBIQUITIN PROTEASOME PATHWAY

The Importance of Protein Half-life Regulation

There are many developmental and physiological processes that occur in eukaryotic tissues that require rapid transitions in cellular state and protein composition. These transitions are generally accompanied by a restructuring or dismantling of the existing regulatory network and replacement with a new one (Hochstrasser, 1995). The rapid switches in protein content observed in cell cycle progression, cell lineage specification, embryogenesis, and the events following reception of an intercellular signal, are but a few examples of this (Callis and Vierstra, 2000; Ciechanover et al., 2000; Hochstrasser, 1995). In many cases, changes in protein levels need to occur at a rate that exceeds what is possible for regulation at the transcriptional or translational levels (Kornitzer and Ciechanover, 2000). One mechanism that cells use to achieve dramatic and fast flowing changes in cellular protein content involves regulation of protein half-life via degradation.

Protein half-life is a determinant of the time it takes for a protein to reach a new steady-state level following a change in the rate of its synthesis (Hochstrasser, 1995). By swiftly degrading existing protein, a cell can reach a new steady-state level for that protein at a much faster rate. Elimination of protein in this fashion also avoids inappropriate reactivation of the protein. Ubiquitylation and the proteasomal pathway play an essential role in the control of protein half-life regulation (Hochstrasser, 1995).

Ubiquitin and the Proteasome

Ubiquitin is a small 76 amino acid residue polypeptide that can be covalently attached to specific lysine residues on a protein. This modification is carried out in a step-wise process involving three classes of enzymes (Figure 1.1a). The first step involves energy dependent activation of the ubiquitin moiety by the ubiquitin-activating enzyme E1 (Haas and Rose, 1982). This enzyme links to ubiquitin through a high-energy thiol-ester bond between the active site cysteine (Cys) residue of E1 and the carboxyl-terminal glycine (Gly) residue of Ubiquitin (Haas and Rose, 1982). In most eukaryotic organisms there is a single gene encoding two or more isoforms of the E1 enzyme, which serve to activate all of the ubiquitin required by the cell (Handley-Gearhart et al., 1994; Pickart, 2001). Activated ubiquitin is transferred to a second family of enzymes, the ubiquitin conjugating/carrier proteins or E2s, through a thiol-ester intermediate (Hershko and Ciechanover, 1998). The last and most

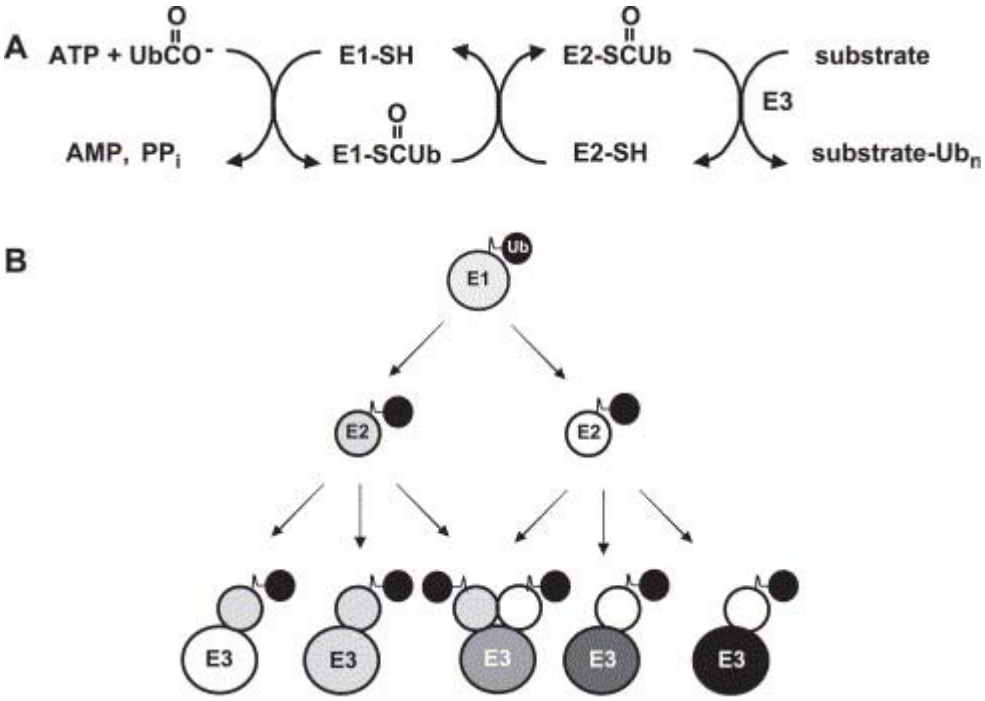


Figure 1.1: ***The Ubiquitin-Proteasomal Degradation Pathway.*** **A)** Ubiquitin is first activated in an ATP dependent fashion, by an ubiquitin activating enzyme E1. The activated ubiquitin is transferred to a second group of enzymes, the ubiquitin-conjugating enzyme or E2s, via a thiol ester intermediate. A third group of enzymes, the E3 ubiquitin ligases, present a specifically bound substrate to the E2 enzyme catalysing the attachment of ubiquitin to the substrate protein via an internal lysine residue of the substrate. The substrate protein becomes polyubiquitylated through the repeated action of the E2/E3 enzymes and this process may also utilise a fourth enzyme, E4. The polyubiquitylated substrate is subsequently degraded by the 26S proteasome. **B)** The specificity of the ubiquitin proteasome system is achieved by the expansion in number of enzymes involved at each step of the cascade. There are at least 18 known E2 enzymes (Pickart, 2001) encoded by the human genome and a much larger number of E3 enzymes (adapted from (Pickart and Eddins, 2004).

diverse series of enzymes in the process, the ubiquitin-protein ligase family of E3 enzymes, now present a specifically bound protein substrate to the ubiquitin bearing E2 enzyme (Laney and Hochstrasser, 1999). The ubiquitin moiety is subsequently transferred to the substrate protein in one of two ways. This occurs either via an E2 ubiquitin thiol ester intermediate (Hochstrasser, 1995) or via an additional step in which the ubiquitin moiety is transferred to a catalytic cysteine residue in the E3 enzyme, which finally transfers the ubiquitin to the substrate via an E3 ubiquitin thiol ester intermediate (Scheffner et al., 1995). Both processes result in the creation of an isopeptide bond between an internal Lys residue of the substrate and the activated carboxyl-terminal Gly of ubiquitin. Recognition of most proteins targeted for destruction by the proteasome requires attachment of a polyubiquitin chain containing at least four ubiquitin groups (Deveraux et al., 1994). In some cases, the process of polyubiquitylation may require an additional factor known as an E4, in keeping with established terminology (Koegl et al., 1999). Polyubiquitylation occurs through the attachment of additional ubiquitin moieties to an internal Lys residue located on the last ubiquitin molecule to be attached to the substrate protein (Ciechanover et al., 2000). Ubiquitin itself has seven internal lysine residues (amino acids 6, 11, 27, 29, 33, 48, and 63) with each potentially capable of acting as a substrate for polyubiquitin chain formation (Pickart and Fushman, 2004). It has been hypothesised however, that ubiquitin chains formed through linkage of ubiquitin moieties via lysine 48 of ubiquitin are preferentially recognised by the proteasomal machinery (Pickart, 1997).

Regulation of Protein Substrates by the 26S Proteasome and the Need for Substrate Specificity

Many of the proteins that are targeted to the proteasome for degradation are involved in cellular processes that need to be tightly regulated. For example, the ubiquitin proteasomal pathway is essential for controlled passage through all stages of the eukaryotic cell cycle (Hershko, 1997). In eukaryotes, members of the family of proteins known as the cyclin dependent kinases or cdks, are the central regulators of all phases of the cell cycle (Obaya and Sedivy, 2002). The ability of this small group of proteins to perform so many different activities is a result of their complex interactions with different members of a second family of proteins called the cyclins (Smits and Medema, 2001). Members of a third family of proteins known as the cyclin-dependent kinase inhibitors are also intimately involved in regulating the activity of the cyclin/cdk complexes (Obaya and Sedivy, 2002). Members of both the cyclin family and the cdk inhibitor family are temporally regulated by the ubiquitin-proteasome pathway during the cell cycle.

In mammals for example, cyclin E associates with cdk2 and is required for the transition from G1 into S-phase (Koff et al., 1992; Resnitzky et al., 1994). Cyclin E is only present in the cell at this stage and, in its complex with cdk2, it mediates its own destruction via phosphorylation and degradation by the ubiquitin-proteasome pathway (Clurman et al., 1996). Other cyclins regulated by the ubiquitin-proteasome pathway include the mitotic cyclins A and B. These two proteins are present through G2 and M-phases, and exit from mitosis requires their ubiquitylation by the anaphase promoting complex (APC) and subsequent degradation at the proteasome (Geley et al., 2001; Holloway et al., 1993). In yeast, it has been found that ubiquitin-mediated degradation of the cdk inhibitor Sic1 is required for transition from G1 to S-phase (Patton et al., 1998; Schwob et al., 1994). Similarly, the transition from a quiescent state to a proliferative state in mammals requires the proteasomal degradation of the mammalian cdk inhibitor p27 (Carrano et al., 1999; Sutterluty et al., 1999; Tsvetkov et al., 1999). Expression of cyclins and cdk inhibitors at inappropriate stages of the cell cycle can have disastrous results. Overexpression of cyclin E for example, can lead to accelerated entry into S-phase (Resnitzky et al., 1994). It is vital that the level of each of these factors is tightly controlled at each stage of the cell cycle. It is easy to see then, the need for a high degree of substrate specificity on the part of the degradation machinery of the cell. This specificity is required to ensure that only the appropriate combinations of cyclins/cdk inhibitors, are present or degraded at each phase of the cell cycle.

As discussed above, the remarkable specificity of the ubiquitin proteasome pathway is achieved, in part by the assorted E2 ubiquitin conjugating enzymes, but mainly by the variety of E3 ubiquitin ligases. There are at least six subclasses of E3 ligases characterised by their structure and/or the class of signal that they receive (Ciechanover et al., 2000) (Figure 1.b). These are: -

1) The HECT-domain (homologous to E6-AP carboxyl terminus) proteins, which are characterised by a 350 amino acid residue domain that can be found in the first described member of this class, E6-associate protein (Huibregtse et al., 1995). A conserved cysteine residue found in this region is responsible for catalysing the transfer of activated ubiquitin from an E2 enzyme, to the substrate (Scheffner et al., 1995). HECT-domain proteins mediate the degradation of, among other things, p53, the kidney epithelial Na^+ channel, and yeast uracil and amino acid permeases (Kornitzer and Ciechanover, 2000).

2) A complex of proteins known as the anaphase-promoting complex (APC) is the second class of E3 enzyme. This complex contains at least eight subunits and targets for destruction mitotic substrates, including cyclins A and B (Geley et al., 2001).

3) The E3α protein. This ligase recognises and ubiquitylates proteins containing bulky hydrophobic and basic amino-terminal residues (Kornitzer and Ciechanover, 2000).
4) A tumor suppressor known as the von Hippel-Lindau protein (pVHL). This protein targets hypoxia inducible factors (HIFs) for destruction (Maxwell et al., 1999). It is part of a complex that contains Elongins B and C, Cullin-2 and Rbx1/Roc1 and has been shown to be capable of polyubiquitylation (Iwai et al., 1999; Lisztwan et al., 1999).
5) Really interesting new gene (Ring) finger containing proteins such as c-cbl. These have also demonstrated the properties of E3 ligases and it is believed that the ring finger itself may be essential in this role (Lill et al., 2000; Longva et al., 2002; Thien et al., 2001). Substrates of these proteins include the epidermal growth factor (EGF) (Longva et al., 2002) and platelet derived growth factor (PDGF) (Miyake et al., 1999) receptors and CSF-1 (Lee et al., 1999).
6) Large complexes containing as many as six components known as the SCF (Skp1, yeast Cdc53, or mammalian Cullin, and F-box protein) complexes. These complexes, also containing Rbx1, Roc1 and Sgt1, act in conjunction with the E2 enzyme, Cdc34/Ubc3, and possibly with members of the UbcH5 class of ubiquitin-conjugating enzyme (Ciechanover et al., 2000). A common component containing Rbx1/Roc1, Skp1, and Cdc53/Cullin-1 is proposed, with specificity being provided by the variety of F-box proteins after which each complex is named (Ciechanover et al., 2000). For example, the $SCF^{\beta\text{-TrCP}}$ complex contains the β-TrCP F-box protein. Substrates of these complexes include the transcription factor E2F-1, and β-catenin (Koepp et al., 1999; Laney and Hochstrasser, 1999). This large variety of different E3 ubiquitin ligases, are thought to provide the ubiquitin proteasomal pathway with a high degree of specificity.

REGULATION OF PROTEIN TRAFFICKING BY UBIQUITYLATION

The Endocytic/Endosomal Pathway

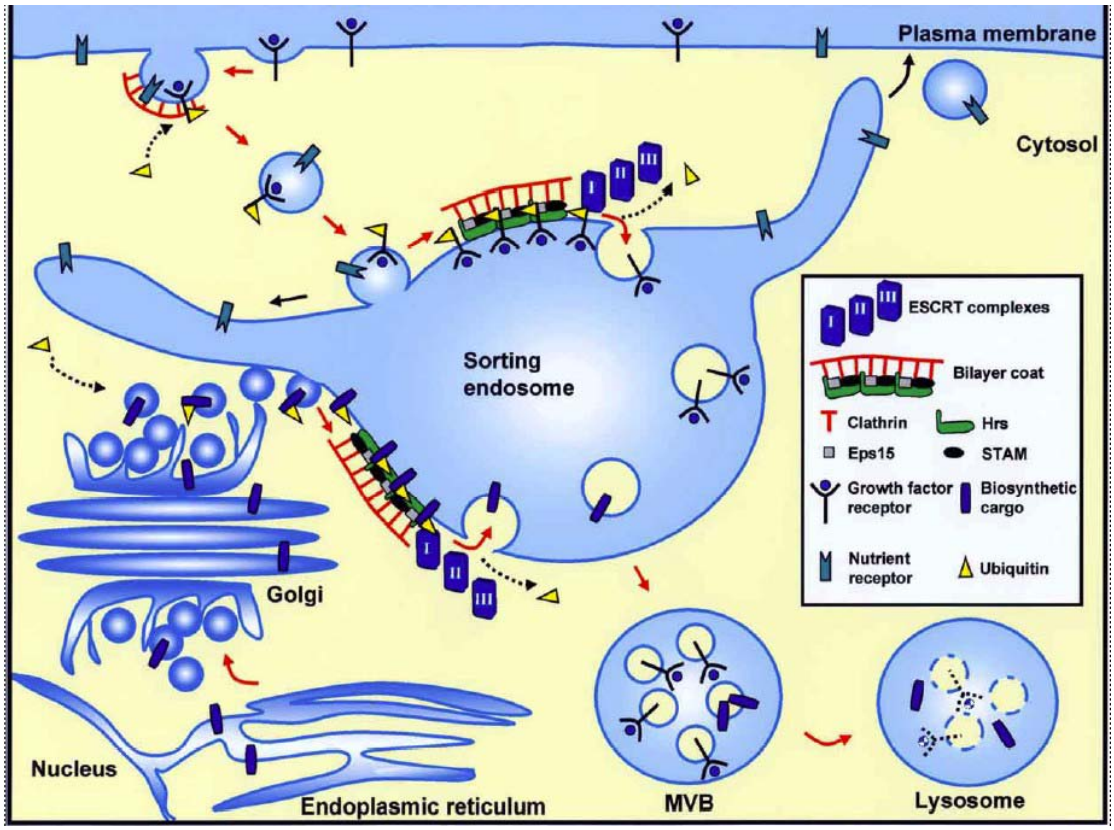
Whilst protein modification by ubiquitin is best known for its role in proteasomal degradation, it has more recently been recognised as an important player in the endocytosis of plasma membrane proteins and cargo sorting in the endosomal system. This association has been studied most extensively in the budding yeast *Saccharomyces cerevisiae*. The first evidence of an association between protein ubiquitylation and endocytosis came from studies involving the yeast plasma membrane protein Ste6p. The Ste6p protein undergoes constitutive endocytosis followed by degradation, occurring primarily in the vacuole (Berkower et al., 1994; Kolling and Hollenberg, 1994). The link between ubiquitylation and endocytosis came

from two observations. Firstly, in mutant yeast cells that are defective in endocytosis, ubiquitylated Ste6p protein accumulates at the plasma membrane. Secondly, the rate of vacuolar degradation of Ste6p was significantly reduced in mutant yeast cells deficient in members of the ubiquitin-conjugating enzyme family (Kolling and Hollenberg, 1994).

Subsequent studies into the endocytosis of plasma membrane proteins in yeast have strengthened the tie between ubiquitylation and endocytosis based on three types of observation. (i) Ubiquitylated forms of plasma membrane proteins accumulate at the cell surface in yeast strains lacking components of the endocytic machinery leading to the hypothesis that ubiquitylation precedes internalisation. (ii) Yeast cells that are defective in various subsets of the ubiquitylation machinery often display a reduction in the rates of degradation of plasma membrane proteins and a reduced ability to correctly internalise them (Rotin et al., 2000). (iii) Loss of the proteolytic components of the vacuole leads to strong reductions in the levels of degradation of many plasma membrane proteins, whereas their destruction is unaffected when regulatory or catalytic subunits of the 26S proteasome are lost (Hicke and Riezman, 1996).

Internalisation and degradation of plasma membrane proteins can be broken down into three main stages (Figure 1.2). (i) A cell receives an external signal such as a change in nutrient availability, or the binding of a ligand to its receptor, which results in the internalisation of the plasma membrane protein into either clathrin coated vesicles (Mousavi et al., 2004) or caveolae (Minshall et al., 2003; Mukherjee et al., 2006) before passage to an early endosome (EE). (ii) The EE is the site of the first sorting decision a cell makes in the internalisation process. Here the cell sorts proteins that are destined for destruction from those that are to be recycled back to the plasma membrane. Proteins that are to be recycled enter Rab11 positive recycling endosomes (Lock and Stow, 2005) and make their way back to the plasma membrane whereas proteins targeted for destruction pass through to the late endosome/pre-vacuolar compartment (LE/PVC). (iii) In the LE/PVC vacuolar/lysosomal resident proteins on the biosynthetic pathway from the trans-Golgi network (TGN) join those proteins that are en route to the vacuole/lysosome from the plasma membrane. The cell now needs to make a second sorting decision to separate proteins destined for destruction in the vacuolar/lysosomal lumen from those that are required in the limiting membrane of the vacuole/lysosome. The former are sorted into internal vesicles of the LE/PVC leaving the latter on the outer membrane and leading to the maturation of the LE/PVC into a multivesicular body (MVB). The MVB then fuses with the vacuole/lysosome delivering plasma membrane proteins to the vacuolar/lysosomal lumen and incorporating

Figure 1.2: ***The Endosomal Pathway and the Role of Ubiquitin*** Schematic diagram representing the different stages along the endocytic and biosynthetic routes, and the protein complexes that can be found at each stage. On the endocytic route, cell surface proteins such as growth factor receptors and nutrient receptors are targeted for internalisation by signals in their cytoplasmic tails (endocytic motifs or ubiquitylation). This serves to recruit adaptor proteins which in turn recruit coat components like clathrin. Recruitment of coat complexes causes budding of the plasma membrane and the formation of an endocytic vesicle. A decision that appears to be reliant on the ubiquitylated status of the cargo is then made as to whether the plasma membrane protein is recycled to the cell surface or continues towards degradation at the lysosome. Once the decision is made to degrade the protein at the lysosome, the cargo enters the late endosome or sorting endosome, where it joins proteins on the biosynthetic route. Here proteins are sorted by the ESCRT I-III complexes and biosynthetic proteins destined for the limiting membrane of the lysosome are separated from plasma membrane proteins destined for destruction by the formation of internal vesicles. Biosynthetic proteins remain associated with the limiting membrane of the forming multivesicular body (MVB) whereas proteins destined for destruction enter the intra-endosomal vesicles. Again, this sorting process seems to be reliant on the presence or absence of an ubiquitin tag. Those proteins tagged with ubiquitin enter the internal vesicles and those without an ubiquitin tag do not. From here plasma membrane proteins are delivered to the lysosomal lumen and biosynthetic proteins integrate into the limiting membrane of the lysosome when the MVB eventually fuses with the lysosome (adapted from (Raiborg and Stenmark, 2002).


Plasma membrane
Cytosol
Sorting endosome
ESCRT complexes
Bilayer coat
Clathrin
Hrs
Eps15
STAM
Growth factor receptor
Biosynthetic cargo
Nutrient receptor
Ubiquitin
Golgi
Nucleus
Endoplasmic reticulum
MVB
Lysosome

vacuolar/lysosomal resident proteins into the limiting membrane. Ubiquitylation plays a role in each of these steps.

Signals for Internalisation

Current research in eukaryotes has led to the view that there are two mechanisms that the cell uses to signal the internalisation of plasma membrane proteins, ubiquitin-independent and ubiquitin-dependent. In both cases, the process relies upon sequences found in the cytoplasmic portions of the protein. In brief, ubiquitin-independent signals include amino acid sequences consisting of a dileucine pair or those based on a tyrosine consensus motif of either NPXY or YXXZ (where X is any amino acid and Z is a hydrophobic amino acid such as leucine) (Kirchhausen et al., 1997). Less well understood is the mechanism of ubiquitylation dependent internalisation. In yeast mono-ubiquitylation, or di/tri-ubiquitylation through K63 links is sufficient to trigger internalisation of almost all plasma membrane proteins (Horak, 2003). Lysine residues found in one of two consensus cytoplasmic amino acid sequences, the [(D/E)xK(S/T)] motif or PEST-like sequences, have been identified as sites of ubiquitylation in most internalised yeast plasma membrane proteins. Furthermore, mono ubiquitylation or K63-linked polyubiquitylation of surface proteins containing [(D/E)xK(S/T)] motifs (Ste2p, Fur4p and Tat2p) (Beck et al., 1999; Hicke and Riezman, 1996; Marchal et al., 2000) and/or PEST-like sequences (Ste3p, Ste6p, Fur4p, and Mal61p) (Kolling and Losko, 1997; Marchal et al., 2000; Medintz et al., 2000; Roth et al., 1998) is sufficient to induce internalisation. The situation in higher eukaryotes seems to be much more complicated.

Probably the best-characterised examples of endocytosed plasma membrane proteins in higher eukaryotes are the receptor tyrosine kinases (RTKs). These include the platelet derived growth factor receptor (PDGFR), growth hormone receptor (GHR), the *c-kit* receptor (*c-kit*R), colony stimulating factor 1 receptor (CSF-1R), and the epidermal growth factor receptor (EGFR). Studies have clearly demonstrated that for each of these receptors, ligand binding stimulates activation of the receptor, either by dimerisation or conformational change, and subsequent phosphorylation on internal tyrosine residues (Levkowitz et al., 1999; Mori et al., 1995; Mori et al., 1992; Sengupta et al., 1988; Strous et al., 1996; Yee et al., 1994). Moreover, in all cases, this phosphorylation requires that the kinase domain of each receptor is active (Levkowitz et al., 1999; Mori et al., 1995; Mori et al., 1992; Sengupta et al., 1988; Strous et al., 1996; Yee et al., 1994). This ligand dependent phosphorylation of RTKs is often accompanied by multiubiquitylation (Mori et al., 1995). Following ligand stimulation of the receptor, a smear or ladder of anti ubiquitin immunoreative bands appears upon western

blotting. This smear or ladder begins at the predicted molecular weight of the mature receptor and continues upwards with species migrating at progressively higher molecular weights often separated by approximately 8 kDa (the predicted molecular weight of ubiquitin) (Mori et al., 1995). Whether or not this ubiquitylation serves as a signal for internalisation is a question of great debate.

Early studies identified the cbl family of proteins as major players in receptor tyrosine kinase down-regulation. Genetic work in *Drosophila* and *C. elegans* highlighted the role that the cbl protein plays in modulating signalling through the EGFR (Hime et al., 1997; Jongeward et al., 1995; Meisner et al., 1997; Yoon et al., 1995). There are now three known isoforms of the gene in mammals, c-*cbl* (also called cbl or cbl-A), cbl-b, and cbl-c (or cbl-3) all of which encode a RING finger type, E3 ubiquitin ligase (Thien and Langdon, 2001). The cbl family is characterised by multiple conserved domains that include an amino-terminal tyrosine kinase binding domain (TKB), a RING finger domain, and in the case of c-*cbl* and cbl-b, a proline rich region and a ubiquitin associated domain (UBA) (Thien and Langdon, 2001). At least two members of the family, c-*cbl* and cbl-b, have the ability to ubiquitylate EGFR (Ettenberg et al., 2001); (Waterman et al., 1999), and c-*cbl* has been shown to be capable of ubiquitylating TCR (Wang et al., 2001), PDGFR (Miyake et al., 1999), GHR (Goh et al., 2002), and CSF-1R (Lee et al., 1999).

Mechanistically, it has been determined that, following ligand stimulated activation and tyrosine phosphorylation of RTKs, cbl proteins themselves become tyrosine phosphorylated and are recruited to the signalling complex (Levkowitz et al., 1999; Levkowitz et al., 1998; Miyake et al., 1999; Thien and Langdon, 2001; Zhu et al., 1998) Using mutations of the receptors that have either lost kinase activity, or TKB mutations of the cbl proteins that no longer bind the receptors, it has been found that cbl recruitment promotes ubiquitylation and down regulation of the receptors (Ettenberg et al., 2001; Lill et al., 2000; Miyake et al., 1999; Mori et al., 1992). Taken together with the fact that multiple mono-ubiquitylation has been shown to be sufficient to induce internalisation of many of these receptors, it has been hypothesised that ubiquitylation mediated by cbl and other ubiquitin ligases, is indeed a signal for internalisation (Haglund et al., 2003).

On the other hand, significant evidence suggests that ubiquitylation may not be required, or that it is just one of a number of redundant signals. When a mutated form of the cbl protein that contained most of the N-terminus but lacked the RING finger domain (cbl-N), was overexpressed along with exogenous EGFR in the human embryonic kidney cell line,

HEK 293, two effects were noticed. 1) The cells showed a much-reduced capacity to down-regulate the receptor upon ligand stimulation corresponding to a loss of ubiquitylation of the receptor. 2) Immunofluorescence studies of the cells however, indicated that some receptor was still being internalised (Lill et al., 2000). Levels of receptor degradation were much reduced with cbl-N expression and the pattern of vesicular staining that was observed was nonetheless divergent from that seen for overexpression of wild-type protein or of a larger mutation that now carries the ring finger domain (Lill et al., 2000). This suggested that ubiquitylation is dispensable for internalisation but is required at later stages of the endosomal pathway. This research was strengthened by the observation that overexpression of c-*cbl,* or catalytically inactive mutations of the protein, in an NIH 3T3 fibroblast cell line that also expresses EGFR (NhEGFR cells), had no effect on receptor internalisation compared to control cells (Thien et al., 2001). Similar results were obtained in COS-1 cells where exogenous expression of cbl-N had no effect on ligand stimulated endocytosis of EGFR in immunofluorescence studies and pulse chase experiments with ^{125}I-EGF ligand (Longva et al., 2002). Probably the most convincing evidence comes from investigations of the characteristics of EGFR internalisation and degradation in cbl$^{-/-}$ mouse embryonic fibroblasts. It was found that uptake of fluorescently labelled EGF was unchanged in cells lacking cbl protein when compared to wild-type cells, even though ubiquitylation of EGFR was lost. Furthermore, examination of the internalisation of EGFR in the CHO-ts20 cell line, which harbours a temperature sensitive E1 activating enzyme that is non-functional at the restrictive temperature of 42°C, again found that ubiquitylation was dispensable for the internalisation of the receptor (Duan et al., 2003).

Most recently it was shown that ubiquitylation of the EGFR by cbl actually directs its internalisation via clathrin-independent routes and that cbl's function in clathrin-mediated endocytosis of the receptor is more likely to be ubiquitylation of downstream components of the endocytic machinery. In this set of elegant experiments a chimeric EGFR/ubiquitin fusion in combination with a suite of pharmacological and dominant negative inhibitors of endocytosis, were used first to show how the ubiquitin signal directs internalisation, and secondly, to dissect out which internalisation mechanism is dependent on this ubiquitylation. It was found that when HeLa cells were treated with two different physiologically relevant levels of EGF ligand, one "high" and one "low" level, internalisation of the EGFR proceeded differentially (Sigismund et al., 2005). Low levels led to the localisation of EGFR in clathrin coated pits. With high levels of ligand, it was seen that the receptor localised in approximately equal proportions to both caveolae and clathrin coated pits. Significant amounts of ubiquitylated receptor were only seen however, at high levels of EGF treatment, suggesting a

correlation between ubiquitylation status and internalisation via caveolae (Sigismund et al., 2005). If this were the case, one would expect the chimeric receptor to be internalised via this route. Indeed, when EGFR negative NR6 or Chinese hamster ovary cells were transiently transfected with the chimera, the internalised EGFR/ub fusion localised predominantly to caveolae. In contrast, the same cells transfected with wild-type EGFR internalised the receptor by equal proportions of caveolae and clathrin coated pits (Sigismund et al., 2005).

The proposed model that ubiquitylation of EGFR directs its internalisation via caveolae was strengthened by investigating each pathway of endocytosis in isolation. Amphiphysin is an adapter protein which recruits dynamin, a protein required for clathrin coated vesicles to pinch off from the plasma membrane, to sites of newly forming clathrin coated pits (Mousavi et al., 2004). Eps15 is another adapter protein that has been shown to be essential in clathrin mediated endocytosis and, like epsin, has been implicated in recruitment of AP complexes and cargo recognition (Mousavi et al., 2004). Co-transfection of NR6 or CHO cells with mutated proteins, eps15 (Δ95/295) and an amphiphysin fragment (A1/SSR), both of which interfere with formation of clathrin pits, perturbs internalisation of wild-type EGFR but not the EGFR/ub chimera. In contrast, use of the selective pharmacological inhibitors of non-clathrin mediated internalisation filipin, nystatin and low-dose genistein only mildly affected internalisation of wild-type EGFR while almost completely abolishing internalisation of the EGFR/ubiquitin fusion (Sigismund et al., 2005).

It seems likely that evolution has provided the cell with a number of redundant systems to ensure that the signals it receives through its plasma membrane receptors can be attenuated and controlled in the event that one of them fails. Indeed, it would not be surprising to find that both ubiquitin-dependent and ubiquitin-independent machinery controls the down-regulation of each of these receptors. What is clear is that ubiquitylation plays an essential role in the later stages of the endocytic process.

Ubiquitylation at the Early Endosome

Regardless of how it happens, once a plasma membrane protein is internalised, the next port of call is the early endosome (EE). The early endosome is often differentiated from other endosomal compartments by its tubular morphology, localisation at the cell periphery, and by kinetic observations that many endocytosed plasma membrane proteins enter this compartment rapidly following internalisation. Additionally, the EE is often defined by markers. One such marker is the small Rab-GTPase Rab5. Rab5 is a protein that is intimately

involved in both sequestration of cargo into endocytic particles from the plasma membrane (McLauchlan et al., 1998) and fusion of those particles to the EE in concert with the Rab5 effecter, early endosomal antigen 1 (EEA1), another common marker of EEs (Christoforidis et al., 1999; Gorvel et al., 1991). It appears that in this compartment, ubiquitin acts as a sorting signal to separate plasma membrane proteins destined for destruction, from those that are to be recycled back to the plasma membrane.

In yeast, there appear to be two modes of internalisation of the α-factor mating pheromone receptor, constitutive and **a**-factor dependent. In the constitutive pathway, a PEST-like sequence found in the protein leads to ubiquitylation of the protein, ultimately resulting in degradation of the protein at the vacuole (Davis et al., 1993; Roth and Davis, 1996; Roth et al., 1998). On the other hand, internalisation of the receptor via the **a**-factor dependent pathway proceeds through the NPFxD motif, which does not require ubiquitylation and results in the recycling of the protein back to the plasma membrane (Chen and Davis, 2000; Chen and Davis, 2002; Roth et al., 1998).

Similarly in higher eukaryotes, the EGFR has dual fates that are reliant upon the ligand that has stimulated it. The more acidic EGF stays bound to the receptor through the entire endocytic process, from internalisation and acidification of the endosome, to proteolysis of the receptor in the acidic conditions of the lysosome (Ebner and Derynck, 1991). TGFα on the other hand, dissociates from the receptor in the early stages of internalisation resulting in the recycling of the receptor back to the plasma membrane (Sorkin and Waters, 1993). It has been found that even slight decreases in the pH of the cellular environment (pH 7.4 to 7.0) prevent TGFα mediated phosphorylation of EGFR whereas EGF stimulated phosphorylation is stable with pH reductions from 7.4 to 6.0 (Longva et al., 2002). Furthermore, while stimulation with either ligand can lead to cbl recruitment and trigger EGFR ubiquitylation, EGF stimulation produces more stable ubiquitylation and a greater efficiency of receptor degradation (Longva et al., 2002). In addition cbl only remains associated with EGFR bound to EGF and, in contrast to TGFα bound EGFR, the EGF/EGFR/cbl complex is directed to the MVB (Longva et al., 2002). It is hypothesised that EGFR stimulated with TGFα recruits cbl and is ubiquitylated. The receptor is then internalised into an endocytic vesicle which becomes acidified as it enters the endosomal pathway causing TGFα to dissociate. The receptor is then dephosphorylated causing dissociation of cbl and allowing its deubiquitylation and sorting into recycling vesicles. EGF however, remains associated with the receptor in the early endosome, maintaining the phosphorylation and ubiquitylation of the receptor and leading to its degradation at the lysosome (Longva et al., 2002). It remains to be

seen whether this mechanism is specific to the EGFR, or whether similar mechanisms apply to other plasma membrane proteins of higher eukaryotes.

Ubiquitylation at the Late Endosome/Multivesicular Body

The final point in the endocytic pathway that utilises ubiquitin as a sorting signal is at the late endosome (LE) to multivesicular body (MVB) transition. This is also the point at which proteins on the biosynthetic pathway destined for the limiting membrane of the vacuole/lysosome join those en route to the lysosome for degradation. In order to ensure that both types of cargo reach their ultimate destination, the cell has evolved a system whereby the latter are sorted into inward budding vesicles while the former remain in the limiting membrane of the LE resulting in maturation of the compartment into an MVB. Although this is the case for the sorting of biosynthetic proteins like the Vph1p subunit of the vacuolar H^+-ATPase and the iron transporter Fth1p in yeast, a similar system has yet to be determined in mammalian cells (Horak, 2003; Raiborg and Stenmark, 2002). As with the early endosome, the LE/MVB compartment has been defined by its morphology, localisation within the cell, and by marker proteins. The LE/MVB compartment has a more or less spherical shape that develops smaller internal vesicles with the maturation of the compartment into an MVB. LE/MVBs are localised to an area in the cell that is in general much closer to the nucleus and markers such as Rab7 and Rab9 are used to visualise the compartment by immunofluorescence microscopy (Feng et al., 1995; Lombardi et al., 1993; Mukhopadhyay et al., 1997; Riederer et al., 1994).

Ubiquitylation as a sorting signal in the LE/MVB has been demonstrated in yeast, by investigating the sorting and trafficking of two types of protein cargo. The first cargo type, consisted of Vph1p, the integral membrane subunit of the vacuolar ATPase, and the iron transporter Fth1p, and normally traffic from the Golgi to the limiting membrane of the vacuole. The second class consisted of proteins that normally traffic back and forth between the Golgi-apparatus and the LE/MVB but never actually enter the vacuole, and was represented by the sorting receptor for carboxypeptidase Y Vsp10p. Both sets of proteins are not known to be ubiquitylated. Chimeric fusions between ubiquitin and these proteins were incorrectly sorted into invaginating vesicles of the MVB and finally into the intralumenal compartment of the vacuole where they underwent proteolytic degradation (Urbanowski and Piper, 2001). This suggested that the presence of an ubiquitin tag on proteins that reach the LE/MVB serves as a signal to the sorting machinery present here and redirects proteins from

the limiting membrane into the invaginating vesicles that form the MVB. This machinery has been identified as the ESCRT complexes.

The Role of the ESCRT Complexes in Endosomal Sorting

Yeast studies have identified a 17-member group of proteins that is required for endosomal sorting. These are the class E vacuolar protein-sorting proteins (VPS) so called because when they are mutated, transport of biosynthetic and endocytic proteins to the vacuole is disrupted. The proteins instead accumulate in large perivacuolar and multilamellar compartments known as class E compartments (Conibear and Stevens, 1998).

There are three complexes of proteins in yeast that are thought to be the central mediators of ubiquitin-dependent sorting at the LE/MVB stage. These are the ESCRT-I to ESCRT-III (endosomal sorting complex required for transport) protein complexes that act sequentially to sort protein cargo into internal vesicles of MVBs (Babst et al., 2002; Babst et al., 2002; Katzmann et al., 2001). ESCRT-I is a heterotrimeric protein complex that is cytosolic in nature and is made up of the UBC-like domain containing protein Vps23p and two other proteins, Vps28p and Vps37p. ESCRT-I is recruited to the endosomal membrane via interactions between Vps28p and phosphatidyl inositol-3-phosphate. Once there, it binds (directly or indirectly) to ubiquitylated cargo through the ubiquitin-binding domain of Vps23p and leads to sorting of the cargo into vesicles (Katzmann et al., 2001). The second complex in the sequence, ESCRT-II, is also cytosolic and is recruited transiently to the endosomal membrane. It consists of the three proteins, Vps22p, Vps25p and Vps36p (Babst et al., 2002). Extensive yeast-two hybrid analyses of both the mammalian (Martin-Serrano et al., 2003; von Schwedler et al., 2003) and yeast (Bowers et al., 2004) ESCRT complexes have now identified direct interactions between proteins of the ESCRT-I and ESCRT-II supporting the proposed model of sequential recruitment. ESCRT-II in turn promotes ESCRT-III complex formation (Babst et al., 2002).

ESCRT-III consists of four proteins that are found in a complex only when they are recruited to the membrane but are otherwise monomeric in the cytosol. Of the four proteins, Vsp2p/Vsp24p and Vsp20p/Snf7p form two functionally distinct sub-complexes. The Vsp20p/Snf7p sub-complex interacts with ESCRT-II to promote assembly of multiple copies of ESCRT-III at the membrane (Babst et al., 2002). It is believed that ESCRT-III functions to cluster ubiquitylated cargo bound for the lysosome into localised concentrations on the limiting membrane of the LE. The exact mechanism of membrane budding leading to the

formation of the MVB has yet to be determined though it is thought to use a mechanism different from the one used in clathrin coat formation. ESCRT-III also serves to recruit the deubiquitylating enzyme Doa4 which removes the now redundant ubiquitin tag from the sorted cargo, helping to maintain the cellular pool of ubiquitin (Amerik et al., 2000). The final step in the ESCRT-mediated sorting process leads to the recruitment of the AAA-ATPase Vps4 to the ESCRT-III complex. Here Vps4 promotes an ATP-dependent disassembly of ESCRT-III allowing the process to start again (Babst et al., 1997; Babst et al., 1998).

Sorting on the Biosynthetic Pathway

Understanding of the mechanisms of protein sorting at the TGN is an area that is still poorly understood and greatly debated. There are at least two cellular destinations to which proteins are known to be trafficked to from the TGN; i) the plasma membrane and ii) the endosomal/lysosomal system. Proteins that are to be secreted from the cell are also targeted to the plasma membrane through association with carrier proteins like the mannose-6-phosphate receptors. Targeting of proteins to the plasma membrane in polarized cells is directed towards the basolateral compartment or the apical compartment. Evidence suggests that this basolateral/apical segregation also occurs in non polarized cell types (Musch et al., 1996). Transport to the basolateral membrane compartment was originally postulated to be the default pathway with discrete signals required for sorting to the apical membrane. It is now clear that discrete sorting signals are required for both pathways. Basolateral sorting signals are well characterised and fall into two categories, tyrosine-based motifs and di- or mono-leucine motifs (Bonifacino and Traub, 2003; Casanova et al., 1991; Hunziker et al., 1991; Matter et al., 1992).

Consensus motifs for the tyrosine-based sorting signals are either of the form NPXY or YXXΦ, where N is an asparagine residue, P is a proline residue, X is any amino acid and Φ is any bulky hydrophobic residue. Of the two, only YXXΦ motifs are believed to be involved in the biosynthetic route. NPXY motifs are more commonly associated with endocytic trafficking of transmembrane proteins (Bonifacino and Traub, 2003). The two forms of leucine based sorting signals are of the form DXXLL, or (D/E)XXXL(I/L), where D is an aspartic acid residue, E is a glutamic acid residue, I is an isoleucine residue, and L is a leucine residue. Both forms of short peptide motif have the common properties of loose consensus motifs and localisation to the cytoplasmic domain of the transmembrane proteins that they are found in (Bonifacino, 2004). Recognition of both tyrosine-based and leucine-based sorting

signals is mediated by a group of multimeric, organelle-specific protein complexes known as the adapter protein complexes (AP) (Bonifacino and Traub, 2003).

Apical sorting signals are less well characterised. At least three types of protein modification have been reported as apical sorting signals including attachment of lipids (Brown et al., 1989; Lisanti et al., 1989; Lisanti et al., 1988) sugars in the form of N-glycans (Scheiffele et al., 1995) or O-glycans (Alfalah et al., 1999; Yeaman et al., 1997) and the presence of peptide motifs (Altschuler et al., 2003; Tai et al., 1999; Takeda et al., 2003). Apical signals can be found on the luminal/extracytoplasmic, transmembrane, or cytoplasmic domains of proteins (Rodriguez-Boulan and Musch, 2005).

The precise mechanism by which sorting occurs on the biosynthetic pathway is a topic which is still unclear and at least three models have been proposed (Rodriguez-Boulan and Musch, 2005) (Figure 1.3). Model A in Figure 1.3 predicts that all sorting of proteins on the biosynthetic pathway occurs at the TGN. Here, apical and basolateral membrane proteins are sorted based on the sorting signals found in their cytoplasmic tails. The sorting machinery described above packages proteins destined for the basolateral membrane compartment into clathrin coated vesicles (CCVs). The mechanisms by which apically targeted proteins are packaged and sent to the apical plasma membrane is very poorly understood although it has been suggested to involve the association of these proteins with lipid rafts (Simons and Ikonen, 1997; van Meer and Simons, 1988). Lysosomal targeting also occurs at the level of the TGN in this model. Lysosomal hydrolases and other lysosomal proteins, along with their associated mannose-6-phosphate receptors, are packaged into clathrin coated vesicles which bud off and head to the endosomal/lysosomal system.

A growing body of evidence suggests that much of the sorting on the biosynthetic pathway occurs in a post-Golgi compartment. Live cell imaging and electron microscopy of proteins trafficking out of the TGN have provided the basis for model B (Figure 1.3). The first indications of post-Golgi sorting stem from early pioneering work combining tracking of a Golgi transport intermediate using GFP-tagged cargo proteins with high resolution 3D immunoelectron microscopy (Polishchuk et al., 2000). A basolaterally targeted GFP-tagged marker protein, G-protein of vesicular stomatitis virus (VSVG) was used to identify carrier structures exiting from the Golgi. These carriers were tracked in their passage from the Golgi to the plasma membrane and the cell was then fixed at a point in the trafficking of the carrier that allowed it to be isolated in a relatively empty area of the cell and a 3D structural image obtained. What was observed was that VSVG appeared to predominantly exit the Golgi

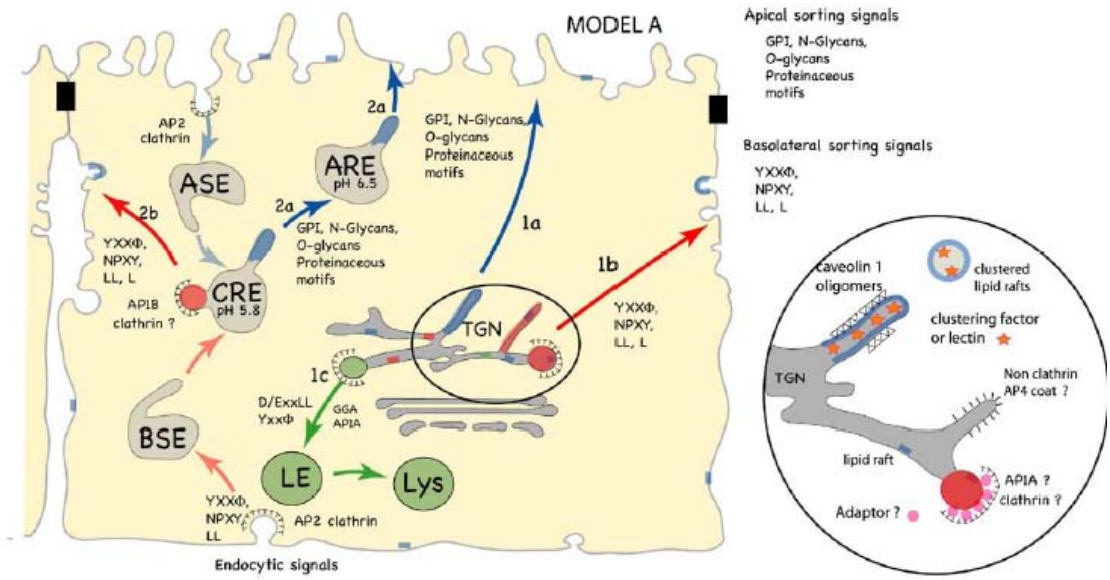
apparatus via relatively large (0.5-1.0μm) carriers with a tubular-saccular structure (Polishchuk et al., 2000). This work challenged the long held view that all carriers emerging from the Golgi apparatus took the form of small (60-100nm) round vesicular structures.

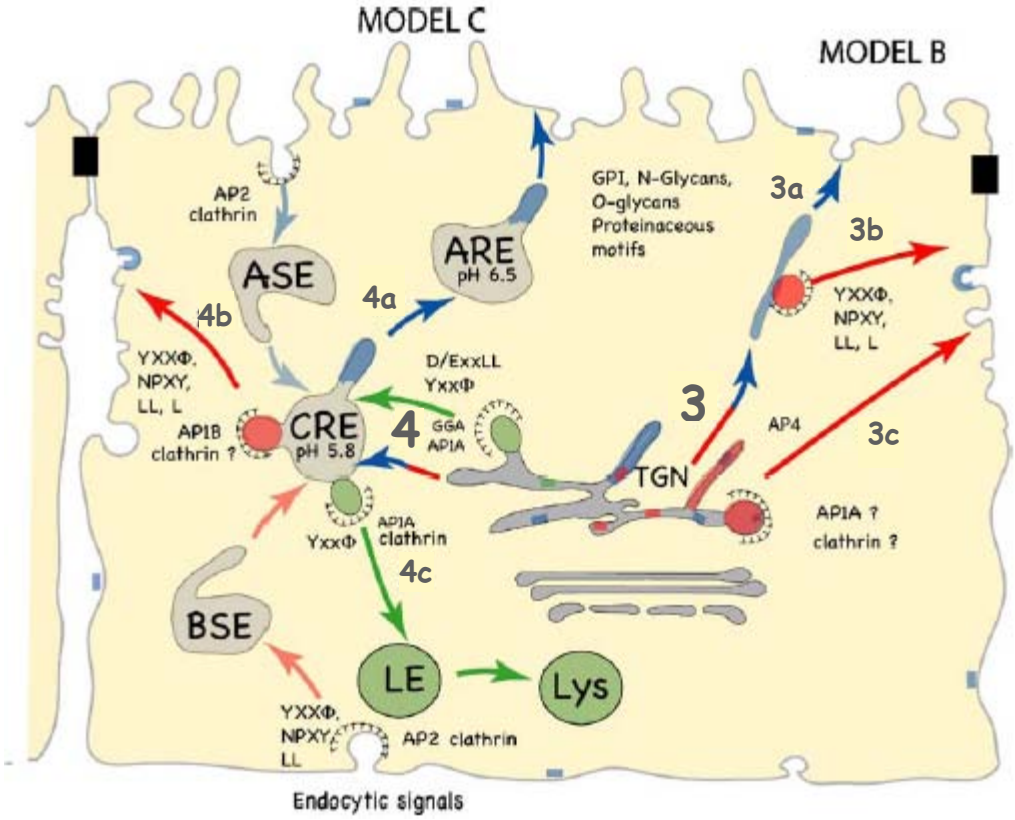
Formation of these large Golgi to plasma membrane carriers (GPCs) seems to occur through a novel mechanism. Fusion of smaller vesicles to form a large tubular-saccular structure appears not to be involved as demonstrated through the use of neutralising antibodies, or dominant negative mutations of proteins required for vesicle fusion (Polishchuk et al., 2003). A further argument against classical formation of these GPCs is that they display features only found on the TGN, a cisternal morphology with fenestrations, which suggests a more direct derivation from the TGN than coat mediated budding. Instead, it seems that these large GPCs are actually drawn out of the TGN by the action of microtubules (Polishchuk et al., 2003). Some cargo molecules transported by these structures, rhodopsin and amyloid precursor protein (APP), have the ability to interact with dynein and kinesin which supports the hypothesis (Kamal et al., 2000; Tai et al., 1999; Tai et al., 2001).

Further investigations revealed the surprising fact that coat proteins traditionally associated with the budding of small vesicular carriers from the TGN itself, were also found on the surface of these detached tubular-vesicular structures. This includes the clathrin adapter protein AP-1, the family of GGA (Golgi-localised, γ-adaptin ear containing, ADP-ribosylation factor (ARF) binding proteins) proteins and clathrin itself (Huang et al., 2001; Puertollano et al., 2003; Waguri et al., 2003). The investigations into Golgi ultra structure, taken together with cell biology studies looking at biosynthetic markers, have led to the development of model B in Figure 1.3. This model proposes that a large proportion of the sorting that takes place on the biosynthetic route occurs at post-TGN sites. The large tubular-saccular structures that appear to be drawn out of the TGN take with them a population of cargo proteins that are only partly segregated. These carriers are then directed to the plasma membrane (the apical surface in polarized cells) by transport on microtubules. Basolaterally or lysosomal/endosomal targeted cargos are subsequently removed from the carriers in transit by coat mediated membrane budding, or directly from the TGN itself.

A third model for sorting of proteins on the biosynthetic route (Figure 1.3) arises from studies investigating the biosynthetic pathways taken by membrane proteins using pulse chase experiments. In these investigations, metabolic ^{35}S labelling of membrane targeted cargo proteins and sorting signal mutations, 20°C trafficking blocks, and labelled endocytic trackers were used to determine the exocytic routes taken by those cargoes. It was found that

Figure 1.3: ***Models of Biosynthetic Trafficking from the Golgi-Apparatus*** Model A – This model proposes that all sorting occurs in the trans-Golgi-network (TGN). Hence basolateral and apical sorting signals are recruited by different machinery in the Golgi (circle) and are subsequently sorted into separate vesicular pathways (routes 1a and 1b). In this model, lysosomal sorting also occurs at the TGN. Clathrin coated pits form and bud off to the endolysosomal compartment (route 1c). Two endosomal systems mediate traffic back from the basolateral and apical surfaces. Membrane proteins first enter either apical or basolateral sorting endosomes (ASE, BSE) following internalisation and subsequently meet at the common recycling endosome (CRE). Once at the CRE, differential sorting mediated by similar signals to those at the TGN sort apical and basolateral proteins into different recycling pathways to the apical (2a) and basolateral (2b) compartments. A third compartment in the apical recycling route (2a) consists of a post-CRE compartment called the apical recycling endosome (ARE). Model B – This is the first of two models that propose that sorting occurs post-TGN. Here, tubular structures are seen to bud from the TGN that contain both apically and basolaterally destined membrane proteins. From these structures, clathrin coated pits form, into which basolateral proteins are sorted and targeted to the basolateral membrane (route 3b). Some basolateral protein sorting also occurs directly from the TGN (route 3c). The remaining (apical) proteins in the tubular structures are targeted towards the apical membrane (route 3a). Model C – In the second post-TGN sorting model, all biosynthetic proteins (apical, basolateral, and lysosomal) traffic from the TGN to the common recycling endosome (CRE) and from there are sorted into their respective routes, apical (4a), basolateral (4b), and lysosomal (4C) (adapted from Rodriguez-Boulan and Musch, 2005).


MODEL A
Apical sorting signals
GPI, N-Glycans,
O-glycans
Proteinaceous
motifs
Basolateral sorting signals
YXXΦ,
NPXY,
LL, L
AP2
clathrin
ASE
ARE
pH 6.5
CRE
pH 5.8
BSE
LE
Lys
TGN
2a
2b
1a
1b
1c
YXXΦ,
NPXY,
LL, L
AP1B
clathrin ?
D/ExxLL
YxxΦ
GGA
AP1A
YXXΦ,
NPXY,
LL
AP2 clathrin
Endocytic signals
caveolin 1
oligomers
clustered
lipid rafts
clustering factor
or lectin
Non clathrin
AP4 coat ?
lipid raft
AP1A ?
clathrin ?
Adaptor ?


MODEL C
MODEL B
AP2
clathrin
ASE
ARE
pH 6.5
CRE
pH 5.8
BSE
LE
Lys
TGN
GPI, N-Glycans,
O-glycans
Proteinaceous
motifs
4a
4b
4
4c
3
3a
3b
3c
YXXΦ,
NPXY,
LL, L
AP1B
clathrin ?
D/ExxLL
YxxΦ
GGA
AP1A
AP4
AP1A
clathrin
YxxΦ
AP1A ?
clathrin ?
YXXΦ,
NPXY,
LL
AP2 clathrin
Endocytic signals

the first port of call for newly synthesized transferrin and asialoglycoprotein receptors after exiting the TGN was an endosomal sorting compartment (Futter et al., 1995; Leitinger et al., 1995). This also occurred for other membrane targeted proteins including E-cadherin, polymeric IgA receptor, and the commonly used basolaterally targeted VSVG protein (Ang et al., 2004; Folsch et al., 2003; Lock and Stow, 2005; Orzech et al., 2000). These studies provided firm evidence that endosomes are intimately involved in the targeting and sorting of many membrane bound proteins on the biosynthetic route.

Targeting Proteins to the Lysosome

Transport of soluble lysosomal proteins and enzymes

Owing to their involvement in numerous lysosomal storage diseases, the biosynthesis and trafficking of lysosomal hydrolases and other soluble lysosomal proteins is well documented in the literature. Most soluble lysosomal proteins enter the lysosome on what is sometimes referred to as the direct route. On this pathway, newly synthesized lysosomal hydrolases are *N*-glycosylated on asparagine residues in the rough endoplasmic reticulum (ER). This is followed soon after by addition of *N*-acetylglucosamine phosphate (GlcNAc-P) to C-6-Hydroxyl groups of selected mannose residues of the oligosaccharide chain by UDP-*N*-acetylglucosamine: lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase (phosphotransferase). This processing occurs at the ER-Golgi intermediate compartment and continues on into the *cis*-Golgi (Dittmer and von Figura, 1999; Lazzarino and Gabel, 1988). The next step in the biosynthetic pathway occurs at the TGN and involves the uncovering of mannose-6-phosphate residues by the so-called uncovering enzyme (UCE) *N*-acetylglucosamine-1-phosphodiester α-*N*-acetylglucosaminidase by removal of the covering GlcNAc residue (Hasilik et al., 1980; Rohrer and Kornfeld, 2001; Tabas and Kornfeld, 1980; Varki and Kornfeld, 1980; Varki and Kornfeld, 1981). This process leads to the production of exposed mannose-6-phosphate (M6P) residues which are a recognition substrate for either of the two known mannose-6-phosphate receptors (M6PRs). These are the 46 kDa cation-dependent M6PR (CD-MPR) and the 300 kDa cation-independent M6PR (CI-MPR), also known as the insulin-like growth factor II receptor (IGF-IIR) for its additional ability to bind IGF-II.

Recognition of the M6P signal is mediated by the lumenal domain of both receptors. In the CI-MPR, the lumenal domain consists of 15 repeating segments of around 147 amino

acid residues each showing a 14 - 38% sequence identity (Lobel et al., 1988). There are two M6P binding sites in CI-MPR that consist of repeat segments 3 and 9, with repeat segment 11 providing the binding site for IGF-II (Garmroudi et al., 1996; Hancock et al., 2002; Marron-Terada et al., 2000; Schmidt et al., 1995). The CD-MPR lumenal domain is composed of one 159 residue segment that is similar to the repeating units of CI-MPR and forms just one binding site for M6P (Olson et al., 1999). Both receptors appear to function as homodimers thereby increasing their affinity for their cargo proteins (Byrd and MacDonald, 2000; Byrd et al., 2000; Olson et al., 1999).

Signals for the correct sorting and targeting of the M6PRs are contained in their cytoplasmic tails. It has been known for quite some time that the M6PRs exit the TGN in CCVs (Campbell and Rome, 1983; Schulze-Lohoff et al., 1985), and that these CCVs are associated with one of the four multimeric adapter protein complexes, AP-1 (Klumperman et al., 1993). At the time of these publications it was accepted that an important factor in the interaction of proteins with AP-1 was the presence of an acidic-cluster dileucine motif (AC-LL). Since both receptors were found to have such a motif located at the carboxyl terminus of their cytoplasmic tails, and these sequences were seen to be critical for the correct sorting of the receptors to the endolysosomal system (Chen et al., 1993; Chen et al., 1997; Johnson and Kornfeld, 1992; Johnson and Kornfeld, 1992), it was assumed that sorting of the M6PRs was reliant on an interaction between this motif and AP-1. Mutation of this motif in CD-M6PR however did not impair binding to AP-1 (Honing et al., 1997; Mauxion et al., 1996). Additionally, it was found that that there were in fact multiple binding sites for AP-1 in both the CD-M6PR (Honing et al., 1997) and the CI-M6PR (Dittie et al., 1997; Le Borgne et al., 1993), none of which included the AC-LL motif at the carboxyl terminus.

Recently the requirement for the carboxyl terminal AC-LL motif for correct endolysosomal delivery was explained with the discovery of the Golgi-localised, γ-adaptin ear containing, ADP-ribosylation factor (ARF) binding proteins (GGAs). The GGA proteins are a multidomain containing family of proteins with three members, GGA1, GGA2, and GGA3. While they are still relatively new players on the scene, their discovery has sparked the emergence of a large body of research into their function. This is due primarily to the presence in their multidomain structure of a VHS (Vps27, Hrs, and Stam) domain that is capable of binding the critical AC-LL of the M6PRs (Puertollano et al., 2001; Takatsu et al., 2001; Zhu et al., 2001), the final link in the story of M6PR mediated sorting of lysosomal acid hydrolases.

The three GGA proteins are arranged in a modular structure consisting of three well conserved domains and a less conserved flexible hinge region (Figure 1.4). Much is already known about the functions of each of these regions and this has allowed models of the action of the GGA proteins to be postulated. The currently proposed mechanism of action of the GGA proteins in sorting of cargo proteins at the TGN begins with their recruitment from the cytoplasm. The GGAs are recruited to the TGN following the localised activation of ARF (Shiba et al., 2003). GTP bound ARF is recognised by the GAT (GGA and Tom1) domain of the GGA proteins (Collins et al., 2003). This recruitment of the GGAs through their interaction with ARF•GTP brings their VHS domain into close proximity with the Golgi membrane where it is subsequently able to interact with the AC-LL motif of the M6PRs. Lastly, the GGA proteins are also believed to be involved in the packaging of the M6PRs and their cargo into the CCVs that transport them to their endosomal/lysosomal destinations although the precise mechanisms that facilitate this are at present unclear.

A flexible, unstructured hinge region that lies between the GAT domain and the carboxyl terminal GAE (γ-adaptin ear-like domain) interacts with clathrin, as does an as yet unspecified region of the GAE domain itself (Costaguta et al., 2001; Mullins and Bonifacino, 2001; Puertollano et al., 2001; Zhu et al., 2001). Additionally, the GAE domain is able to interact with and recruit accessory proteins that mediate membrane curvature and vesicle budding (Lui et al., 2003; Miller et al., 2003). There are three possibilities for the involvement of the GGA protein in formation of M6PR loaded CCVs. Firstly; it is possible that the GGA proteins are able to form these CCVs alone. Secondly, they could act in concert with the AP-1 complexes that are able to bind directly with the M6PRs. Lastly; it may be that they are involved in the initial stages of the formation of these carriers but then pass on the final stages of vesicle formation to AP-1. Some support for this final mechanism has been seen in the observation that a kinase associated with AP-1, CK-2, phosphorylates GGA1 and GGA3, an event that leads to their auto-inhibition through binding of the internal VHS domain to an AC-LL motif in the hinge region (Doray et al., 2002). This auto-inhibition would free the bound M6PRs and their cargo and allow AP-1 and other recruited accessory proteins to finish the job of packaging them into vesicles. There are strong arguments against the GGA proteins acting on their own as there have been no reports of free CCVs associated with the GGA proteins in the cytoplasm, only CCVs in close proximity to the TGN (Bonifacino, 2004).

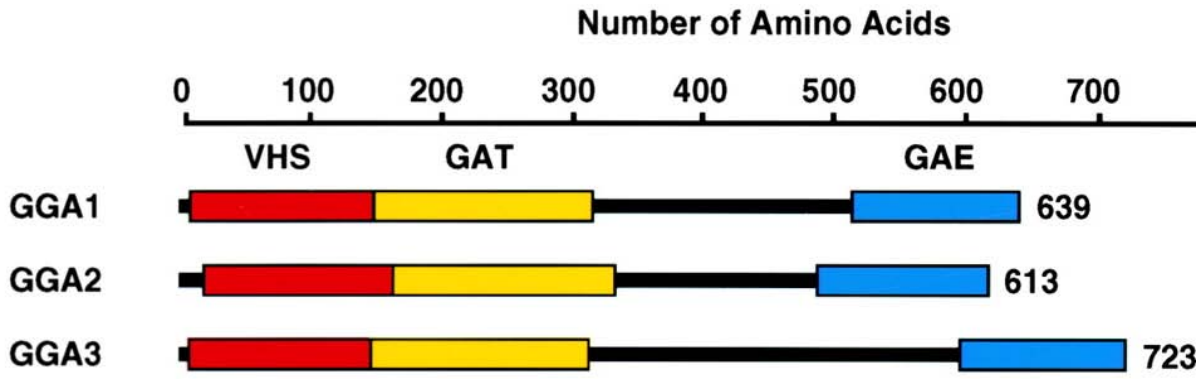


Figure 1.4: ***Domain organization of the three GGA (Golgi-associated, gamma-adaptin ear-like) proteins***. All three proteins have a common four domain structure. Situated at the N-terminus is a VHS (Vps27, Hrs, and Stam) domain. The VHS domain is responsible for binding to the DXXLL sorting motif of the Golgi cargo binding mannose-6-phosphate receptors. The GAT domain (GGA and Tom1) recruits the GGA proteins to the Golgi membrane via its interaction with activated GTP-bound Arf proteins. The GAT domain also associates with ubiquitin and may confer an additional level of specificity for cargo substrates. At the C-terminus resides a GAE (gamma-adaptin ear-like) domain. This domain, in combination with a flexible hinge region located between the GAT and GAE domains, interacts with and is responsible for recruiting clathrin and other adapter proteins that are required for formation of clathrin coated vesicles. Adapted from (Dell'Angelica et al., 2000).

Soon after budding from the TGN, packaged M6PR/lysosomal cargo vesicles lose their clathrin coat through the action of HSC70 in conjunction with auxilin (Greener et al., 2000; Umeda et al., 2000). Most investigations into the events proceeding this, point to the predominant destination of these vesicles as the early endosomal compartment (Ludwig et al., 1991; Press et al., 1998; Waguri et al., 2003). In the final stages of delivery to the lysosome, as the early endosome matures into the late endosome, the shift to low pH causes dissociation of lysosomal cargo from the M6PRs. This allows recycling of the M6PRs that would otherwise be degraded in the lysosome, and delivery of lysosomal proteins to their final destination with fusion of the late endosome and lysosomal compartments (Storrie and Desjardins, 1996).

Some soluble lysosomal proteins also enter the lysosomal compartment via an extended indirect route utilizing the M6PRs in an alternative pathway. Here, the two M6PRs may have independent roles to play. Evidence suggests that the CD-MPR is the major player in mediating secretion of soluble lysosomal proteins on the indirect route. CD-MPR overexpression results in increased secretion of soluble lysosomal proteins (Chao et al., 1990). CI-MPR, unlike CD-MPR, is known to be able to bind to and endocytose secreted lysosomal proteins (Distler et al., 1991; Hoflack et al., 1987; Tong et al., 1989; Watanabe et al., 1990). When CI-MPR is co-overexpressed with CD-MPR, the overactive secretion observed in the CD-MPR overexpressing cells is ablated. Additionally, untransfected cells treated with antibodies raised against the M6P binding site of CD-MPR show a decrease in secreted lysosomal proteins (Chao et al., 1990). Two possible hypotheses have been suggested to explain these results. The first possibility is that CI-MPR may also be active in secretion of soluble lysosomal proteins but its endocytic activity is far more active and hence there is an overall decrease in secreted M6P tagged proteins seen in the media. Alternatively, CI-MPR may not participate in secretion of soluble lysosomal proteins at all and its overexpression simply restores the balance between secretion and uptake that was perturbed with overexpression of CD-MPR. It seems unlikely that CD-MPR is the only factor involved in soluble lysosomal protein secretion however, as there have been reports of secretion of lysosomal hydrolases occurring in a CD-MPR independent fashion (Ludwig et al., 1993).

Some soluble lysosomal proteins also traffic to the lysosome without the need for interaction with the CD-MPRs. This certainly seems to be the case for the sphingolipid activator proteins (SAPs) and acid sphingomyelinase (ASM). The first indications of a MPR-independent route to the lysosome came from analyses of fibroblasts from patients with I-cell disease (ICD). This is a condition in which the phosphotransferase that is required for the

generation of the M6P recognition signal on soluble lysosomal proteins (described above) is non-functional (Reitman et al., 1981; Rijnboutt et al., 1991). Lack of the M6P signal caused missorting and inappropriate secretion of many of the soluble lysosomal proteins (Reitman et al., 1981). However some lysosomal proteins, like the SAPs, appeared to escape this fate and were correctly directed to the lysosome (Rijnboutt et al., 1991). Insight into how this was able to occur came from an investigation of the sorting properties of a relatively uncharacterised type-I membrane receptor, sortilin (Nielsen et al., 2001). When a chimeric sortilin protein, a fusion of its cytoplasmic domain with the transmembrane and extracellular ligand binding domains of CI-MPR, was expressed in cells it was able to functionally replace both M6PRs in the correct sorting of soluble lysosomal proteins (Nielsen et al., 2001). In addition, sortilin is able to physically interact with and direct the trafficking of prosaposin and $G_{M2}AP$ to the lysosome, in cooperation with the GGA proteins (Lefrancois et al., 2003). Similarly, ASM is also able to interact with sortilin and dominant negative mutations of sortilin disrupt the targeting of ASM to the lysosome (Ni and Morales, 2006).

Transport of membrane associated lysosomal proteins

Membrane-associated lysosomal proteins appear to take a different route to the lysosome from soluble lysosomal proteins. There are at least 18 known lysosomal membrane proteins all with diverse functions (Eskelinen et al., 2003 and references therein). By far the major proteinaceous constituents of the lysosomal membrane are the lysosomal membrane glycoproteins (lgps) which fall into two categories, the lysosomal associated membrane proteins of which there are two (LAMP-1 and LAMP-2), and the lysosomal integral membrane proteins also having two members (LIMP-1 and LIMP-2). Together these two groups of proteins make up approximately 50% of the total protein content of the lysosomal membrane (Marsh et al., 1987).

Both groups of proteins are highly glycosylated on their lumenal domains with both *N*- and *O*-linked glycans. This glycosylation however, is not required for the transport of these proteins to the lysosomal compartment, but rather for their stability once they have reached their destination (Barriocanal et al., 1986). Instead it has been found that correct targeting of these proteins requires either a tyrosine-based sorting signal (LAMP-1, LAMP-2, and LIMP-1) located on the carboxyl-terminal cytoplasmic tails of the LAMPs and LIMP-1, or an AC-LL based motif in the carboxyl-terminal cytoplasmic tail of LIMP-2 (Hunziker et al., 1996; Le Borgne et al., 1998; Peters and von Figura, 1994). All four proteins are predominantly

sorted to the endolysosomal system on an intracellular route in AP-3 derived vesicles formed independently of the involvement of AP-1 and the M6PRs (Fukuda, 1991; Honing and Hunziker, 1995; Karlsson and Carlsson, 1998; Le Borgne et al., 1998; Rous et al., 2002). Despite this, there are reports of LAMP-1 being sorted into AP-1 derived CCVs (Honing et al., 1996) and also of its delivery to the plasma membrane (Akasaki et al., 1995). While there is not much data in the literature describing the trafficking of the minor lysosomal membrane proteins, many of them have tyrosine or AC-LL based sorting motifs in their sequence and are presumably transported to the lysosome in a similar fashion to the LAMPs and LIMPs (Eskelinen et al., 2003 and references therein).

Ubiquitin on the Biosynthetic Pathway

The role of ubiquitin in sorting on the biosynthetic pathway in mammalian cells is not well understood. Most current information comes from studies in yeast. These have demonstrated the requirement of ubiquitin for correct sorting of carboxypeptidase S (Cps1p) into the internal vesicles of maturing MVB/late endosomes leading to its delivery into the lumen of the yeast vacuole (Katzmann et al., 2001; Reggiori and Pelham, 2001).

Cps1p is synthesised in a precursor form as a type-II integral membrane protein and is transported directly from the TGN to the late endosome. From here, maturation of the late endosome leads to the formation of an MVB in which Cps1p is sorted into internal vesicles and delivered to the vacuolar lumen upon fusion of the MVB with the vacuole (Odorizzi et al., 1998). The precursor form is then processed by resident vacuolar hydrolases to produce its soluble mature form (Spormann et al., 1992). The role of ubiquitylation in the sorting of Cps1p and two other proteins, Phm5p and Hmx1p, into invaginating vesicles of the MVB and delivery to the vacuolar lumen rather than the limiting membrane was investigated using a *Doa* deletion strain of yeast (Reggiori and Pelham, 2001). Doa4 is a deubiquitylating enzyme that has previously been implicated in sorting at the MVB where it's proposed main function is to recycle free ubiquitin to the cytosol for reuse (Amerik et al., 2000; Losko et al., 2001). It was found that GFP tagged Phm5p, Hmx1p, and Cps1p were mislocalised to the vacuolar membrane in *Δdoa4* mutant yeast cells (Reggiori and Pelham, 2001).

As deletion of Doa4 has a dramatic effect on the entire ubiquitin system in yeast due to a depletion of free ubiquitin, the observed missorting of these proteins may have been an indirect consequence of perturbations at other points in the vacuolar sorting process. To address this issue, the ubiquitylation status of two of the three proteins, Phm5p and Cps1p,

was investigated. Both proteins were found to be ubiquitylated *in vivo* (Reggiori and Pelham, 2001). Mutation of Lys8 in Cps1p resulted in loss of ubiquitylation and the same missorting of the protein to the limiting membrane of the vacuole that occurs in a *Δdoa4* mutant yeast strain (Katzmann et al., 2001). By using various yeast strains with mutations in genes that block the exit of Cps1p from the Golgi (*Δgga1Δgga2*, or the *sec7-1* temperature sensitive mutation), or prevent the fusion of Golgi derived Cps1p carrier vesicles with the LE/MVB (*pep12Δ*), it was determined that ubiquitylation of Cps1p occurs somewhere in transit between the Golgi and the LE/MVB (Katzmann et al., 2001).

The fate of the general amino-acid permease, Gap1p, is also reliant on its ubiquitylation status. Newly synthesised Gap1p transits through the TGN and is directed to the endosomal system where a sorting decision directing its final destination is made. This decision is regulated by the nutritional environment of the yeast cell. When yeast are grown in a nitrogen poor environment, Gap1p is directed to the plasma membrane and becomes active. In a nitrogen rich environment, Gap1p is redirected through the LE/MVB and is instead degraded in the vacuole (Roberg et al., 1997). The redirection of Gap1p to the vacuole occurs when the protein is ubiquitylated by a complex consisting of the E3 ligase Rsp5p, and two cooperating E4 proteins, Bul1p and Bul2p (Helliwell et al., 2001). The ubiquitin signal on Gap1p is recognised by the GAT domain of the GGA proteins and this interaction is necessary to target Gap1p to the degradative pathway (Scott et al., 2004). For this reason, it appears that ubiquitylation of Gap1p, as with Cps1p, is an event that occurs prior to the protein reaching the LE/MVB. It seems clear that ubiquitylation on the biosynthetic route is certainly involved in sorting of Cps1p, Phm5p, and Gap1p to the lysosomal/vacuolar system. However, it is yet to be determined how widely ubiquitylation is used to regulate both the activity of the sorting machinery, and as a sorting signal for recognition of cargo proteins at the TGN in both yeast and higher eukaryotes.

The Dual Roles of Ubiquitin in Protein Trafficking

It is now well established that ubiquitin can regulate protein trafficking in both a cis and a trans-manner (Hicke and Dunn, 2003). In its cis capacity, ubiquitin modification of transmembrane proteins can act as a signal for endocytosis and/or correct sorting of those proteins at all stages of the protein trafficking pathways, both endocytic and biosynthetic. The precise nature of what constitutes an ubiquitin-mediated signal in the endosomal pathways has not been firmly established and may depend on the nature of the protein in question or the organism being investigated. In yeast it appears that attachment of mono-ubiquitin or Lys63

linked di/tri-ubiquitin chains is sufficient to trigger endocytosis of all plasma membrane proteins (Horak, 2003). Similarly, as discussed above, covalent attachment of a single ubiquitin moiety is sufficient to sort proteins into the internal vesicles of MVBs (Urbanowski and Piper, 2001). However, this is not always the case, and may not be true in higher eukaryotes. In mammalian cells, the presence of a smear of anti-ubiquitin reactive bands of higher molecular weights than mature receptors suggests the attachment of multiple ubiquitin moieties (Cenciarelli et al., 1992; Miyazawa et al., 1994; Mori et al., 1995; Mori et al., 1992; Strous et al., 1996). It is yet to be determined whether this is multi-ubiquitylation (the attachment of multiple copies of a single ubiquitin to different lysine residues), polyubiquitylation (a chain of ubiquitin moieties attached to each other through internal lysine residues) or indeed Lys63-linked polyubiquitylation.

Whatever the nature of the ubiquitin signal may be, it is now known that there are many components of the protein trafficking machinery that have the ability to act as ubiquitin receptors. These are proteins that contain one or multiple copies of ubiquitin binding motifs of which there are now sixteen known types (Hurley et al., 2006). Ubiquitin binding motifs commonly found in protein trafficking machinery proteins include the ubiquitin interacting motif (UIM), ubiquitin associated domain (UBA), coupling of ubiquitin to endoplasmic reticulum associated degradation (CUE), ubiquitin E2 variant, Npl4 zinc finger domain, and the GGA and TOM1 domain (GAT) (Hicke and Dunn, 2003; Hurley et al., 2006). Typically the affinity of the interaction of these domains with ubiquitin is moderate to low (individual domains have K_Ds that range from 1 to 1000μM) and in most cases seems to centre around a hydrophobic patch on the surface of ubiquitin consisting primarily of Ile44, Leu8 and Val70 (Shih et al., 2003).

It is clear that ubiquitin is able to act to both initiate endocytosis of cell surface proteins as well as to direct their sorting. It also appears that ubiquitin can act to modulate the activity of the machinery directing these processes which forms the basis of its ability to act in trans. Studies of the lysosomally degraded growth hormone receptor (GHR) and LDL receptor related protein found that these proteins were stabilised by disruption of the proteasomal destruction machinery, suggesting that components of the endo-lysosomal destruction machinery are regulated by the proteasomal pathway (Melman et al., 2002; Strous et al., 1996). This is supported by the observation that many of these components are themselves mono-ubiquitylated, including epsin, Eps15, Hrs, CIN85 and the GGA proteins (Haglund et al., 2002; Oldham et al., 2002; Polo et al., 2002; Shiba et al., 2004; van Delft et al., 1997). It is often the case that this ubiquitylation requires the presence of an internal ubiquitin binding

domain or motif (Polo et al., 2002; Shih et al., 2003). Indeed UIMs can even promote mono-, multi-, and poly-ubiquitylation of other proteins (Miller et al., 2004).

There is much discussion in the literature about the role of ubiquitin modification of protein trafficking machinery. One possibility is that the combination of mono-ubiquitylation and ubiquitin binding domains/motifs serves as a scaffolding mechanism to build the transitory complexes required in the sequential processing and sorting of cargo proteins in the protein transport system (Di Fiore et al., 2003). In a similar manner ubiquitylation of an ubiquitin binding protein may serve to amplify the ubiquitin signal to more efficiently promote events like cargo sorting into vesicles and coat protein recruitment. It is also possible that mono-ubiquitylation of these proteins could have a protective or regulatory effect. For example, an intramolecular association of the ubiquitin binding domain/motif of these proteins with the internal attached mono-ubiquitin could prevent the inappropriate association of the ubiquitin binding domain/motif with other ubiquitylated cellular proteins or free ubiquitin (Polo et al., 2002). Another possibility is that the protein becomes activated as an ubiquitin receptor when a self inhibitory conformation, formed through an intramolecular association, is disrupted by a deubiquitylation event or conformational change brought on by protein-protein interactions (Haglund and Dikic, 2005). As with the ubiquitin signal, the precise nature of ubiquitin mediated regulation is likely to be context specific and vary from protein to protein.

The β-Grasp Fold – Ubiquitin-Like Modifiers, Ubiquitin-Domain Proteins, and Their Functions

A β-grasp fold structure is the three-dimensional configuration that ubiquitin and the ubiquitin-like proteins adopt. As ubiquitin was the first molecule identified having the β-grasp fold structure, the more recently discovered β-grasp proteins have been termed ubiquitin-like proteins. Some of these have high linear amino acid sequence homology to ubiquitin (NEDD8), while others have a lower degree of sequence homology (SUMO1). Ubiquitin-like proteins can be divided into two classes based on functional differences. Type 1 ubiquitin-like proteins are termed ubiquitin-like modifiers (UBLs) while type 2 ubiquitin-like proteins are termed ubiquitin domain proteins (UDPs).

The UBLs include Nedd8 (RUB1), small ubiquitin-like modifier 1 (SUMO-1), interferon stimulated gene 15 (ISG15), and APG12. All four proteins, like ubiquitin, become

covalently conjugated to target proteins on internal lysine residues through the action of E1, and E2 like proteins which, at least in the case of NEDD8, SUMO-1, and APG12, are specific for each (Jentsch and Pyrowolakis, 2000). Some UBLs have links with the ubiquitin pathway, for example SUMO-1 may compete with ubiquitin for conjugation to lysine residues in target proteins such as IκB and thereby have an inhibitory effect (Desterro et al., 1998). Substrates of Nedd8 appear to be limited to members of the cullin family, subunits of the oligomeric SCF family of ubiquitin ligases, providing a direct link between Nedd8 and the ubiquitin pathway (Gray and Estelle, 2000; Hori et al., 1999; Lammer et al., 1998; Liakopoulos et al., 1999; Liakopoulos et al., 1998; Osaka et al., 1998). The remaining UBLs, as well as SUMO-1, seem to have functions that have no relation to the ubiquitin pathway (Jentsch and Pyrowolakis, 2000).

UDPs are distinct from UBLs in that they are proteins with a diverse array of cellular functions. UDPs contain an ubiquitin-like domain (UBLD) that is translated in frame with the remainder of the protein and this is where their structural similarity to ubiquitin ends. Whereas the UBLs can be conjugated to other proteins and are generally synthesised as precursor proteins that are subsequently processed, the UBLDs of UDPs undergo neither of these processes. Instead, it appears that UBLDs are modular components like most other protein domains, and function as protein-protein interaction domains. Some UDPs are associated with the 26S proteasome. The yeast proteins Rad23/Rhp23 , Dsk2/Dph1, and their mammalian homologue and orthologue, hR23 and PLIC-2 respectively, interact with the yeast Rpn1 and mammalian S5a subunits of the 26S proteasome, respectively (Elsasser et al., 2002; Hiyama et al., 1999; Seeger et al., 2003; Walters et al., 2002). Rad23/Rhp23 and Dsk2/Dph1could both be further classified into a subpopulation of the UDPs consisting of those that not only have a UBLD, but also containing the ubiquitin binding UBA (ubiquitin associated) domain. It is believed that these proteins act as a shuttling system directing ubiquitylated proteins ,which they can bind to via their UBA, to the proteasome, via the interaction of their UBLD with Rpn1/S5a (Hartmann-Petersen and Gordon, 2004; Hartmann-Petersen et al., 2003). Other UDPs that have been found to interact with the proteasome include Bag1, Nub1, Nub1L, and parkin (Alberti et al., 2002; Kamitani et al., 2001; Luders et al., 2000; Sakata et al., 2003; Tanaka et al., 2003). The interaction of many of these proteins with the proteasome is mediated by a ubiquitin interacting motif on the S5a subunit (Hiyama et al., 1999; Sakata et al., 2003; Walters et al., 2002).

Other UDPs appear to have roles independent of the ubiquitin-proteasome pathway. IκB kinase is a complex consisting of three core components, IκBKα, IκBKβ, and NF-κB

essential modulator (NEMO) (Mercurio et al., 1999; Mercurio et al., 1997; Rothwarf et al., 1998). IκB kinase is indirectly responsible for the regulation of the NF-κB family of transcription factors through the phosphorylation of IκB inhibitors of NF-κB (Ghosh and Karin, 2002; Karin et al., 2004). Phosphorylation of IκBs leads to their ubiquitylation and subsequent destruction at the proteasome, freeing NF-κB to enter the nucleus and promote transcription of its target genes (Karin and Ben-Neriah, 2000). IκBKα is almost identical to IκBKβ in its domain structure organization and yet the two proteins have divergent properties (Yamamoto and Gaynor, 2004). One structural feature that might allow the two proteins to carry out different functions is a UBLD at the N-terminus of IκBKβ identified *in silico*. When the UBLD of IκBKβ was deleted the kinase lost its activity against IκB and failed to induce activity of NF-κB (May et al., 2004). Overexpression of the isolated UBLD of IκBKβ also resulted in reduced activity of NF-κB appearing to act as a dominant negative molecule (May et al., 2004). The UBLD was however redundant for complex formation (May et al., 2004). It was further determined that a leucine residue that appears to be conserved among many UDPs and ubiquitin, was critical for IκBKβ activation of NF-κB but not for the kinase activity of IκBKβ (May et al., 2004).

Given that there are many non-proteasomal associated proteins that contain ubiquitin interacting motifs and other ubiquitin binding domains, it would seem likely that some of these proteins may also be able to interact with UDPs through a UBLD-UIM interaction. Indeed this appears to be the case. PLIC-1 (ubiquilin) is a protein that is capable of interacting with not only eps15 and eps15R, but Hrs, ataxin-3 (AT3), HSJ1, and Hbp (Heir et al., 2006; Regan-Klapisz et al., 2005). PLIC-1 co-localised with eps15, Hrs, AT3, and HSJ1 at cytoplasmic puncta but only when both proteins were overexpressed (Heir et al., 2006; Regan-Klapisz et al., 2005). These puncta were found to be distinct from endocytic/endolysosomal structures and were determined to be aggresomes (Heir et al., 2006; Regan-Klapisz et al., 2005). PLIC-1 did not co-localise with markers for clathrin coated pits (CALM, clathrin, or AP2), early endosomes (transferrin), or late endosomes (CD63), but instead was found in structures that formed after incubation of cells with the proteasomal inhibitor MG132 or overexpression of CTFR (a protein that is known to promote formation of aggresomes) (Regan-Klapisz et al., 2005). Furthermore, when an aggresome inducing, epitope tagged expanded poly glutamine tract protein was overexpressed in cells, endogenous PLIC-1 and eps15 co-segregated with it in cytoplasmic aggregates (Heir et al., 2006). The UIM/UBL interaction was found to be crucial for both the formation and transport of aggresomes as

deletion of either the UIM from eps15, or the UBL from PLIC-1 disrupted both processes (Heir et al., 2006; Regan-Klapisz et al., 2005).

Parkin was identified as the gene responsible for the pathogenesis of autosomal recessive juvenile Parkinsonism (AR-JP) (Kitada et al., 1998; Matsumine et al., 1998). There are several disease related mutations in *parkin* including exonic deletions, insertions, and missense mutations (Lucking et al., 2000). *Parkin* encodes a 465 amino acid protein that contains an amino-terminal UBLD and an amino-terminal RING-finger E3 ubiquitin ligase domain (Kitada et al., 1998). Most disease causing point mutations that have been characterised are found in the RING-finger domain region of the protein and hence much work has focused on identifying the substrates of *parkin* (Dawson and Dawson, 2003; Lucking et al., 2000). Despite this, perhaps the most enlightening work that has been carried out on parkin has come from work focusing on the function of the UBLD of the protein. It had previously been speculated that the pathology of mutations in *parkin* arose from an inability of ubiquitylated substrates of parkin to be degraded in the proteasomal pathway (Dawson and Dawson, 2003). It might have been expected therefore, that the UBLD of parkin, like that of Rad23, would enable the protein to act not only as an E3 ligase, but to also act as a carrier, bringing its ubiquitylated substrates to the proteasome via a UBLD-S5a UIM interaction. Surprisingly it was found that the UBLD of parkin was not able to interact with the UIM of S5a (Fallon et al., 2006). While this does not discount the possibility that parkin's UBLD can bind to other regions of the proteasome, or that the interaction with the S5a UIM is too weak to be detected by GST-affinity binding experiments, it prompted the hypothesis that the parkin UBLD might bind to the UIMs of other proteins.

Indeed, it was found that the UBLD of parkin was able to interact with the UIM of eps15 but not with the UIM of epsin (Fallon et al., 2006). Eps15 was originally identified as a phosphorylation target of EGF activated EGFR (van Delft et al., 1997), therefore, the effect of EGF treatment on the interaction of parkin and eps15 was investigated. When Her14 cells were treated with EGF, a rapid and transient increase in the amount of parkin that could be co-precipitated with eps15 was observed and furthermore, EGFR itself was also found in the precipitates (Fallon et al., 2006). Eps15 is also known to be mono-ubiquitylated and its ubiquitylation was found to be dependent on its UIM domains (Klapisz et al., 2002; Polo et al., 2002; van Delft et al., 1997). Bringing together these facts, it was found that parkin could ubiquitylate eps15 both *in vitro* and upon EGF stimulation *in vivo*, and that this ubiquitylation was dependent on the UBLD/UIM interaction (Fallon et al., 2006). The transient nature of the interaction between parkin and eps15 could be explained by the previously suggested self

inactivation of eps15. It has been proposed that following ubiquitylation, an intramolecular self inhibition occurs whereby the UIM of eps15 wraps around and masks the mono-ubiquitin moiety preventing further ubiquitylation (Polo et al., 2002). It seems likely that this intramolecular interaction would also prevent access of the eps15 UIM to the UBLD of parkin (Fallon et al., 2006). Investigations of a cell line lacking parkin (HeLa) in the presence or absence of exogenously expressed parkin protein suggested that the main action of parkin in EGFR signaling is to regulate the rate of its endocytosis by binding of its UBLD to eps15 (Fallon et al., 2006). This leads to ubiquitylation of eps15, occluding access to its UIM and preventing interaction and endocytosis of EGFR (Fallon et al., 2006). Hence, it appears that UBLDs of UBD proteins may have many diverse functions within the cell and may act in a similar manner to ubiquitin in facilitating protein-protein interactions.

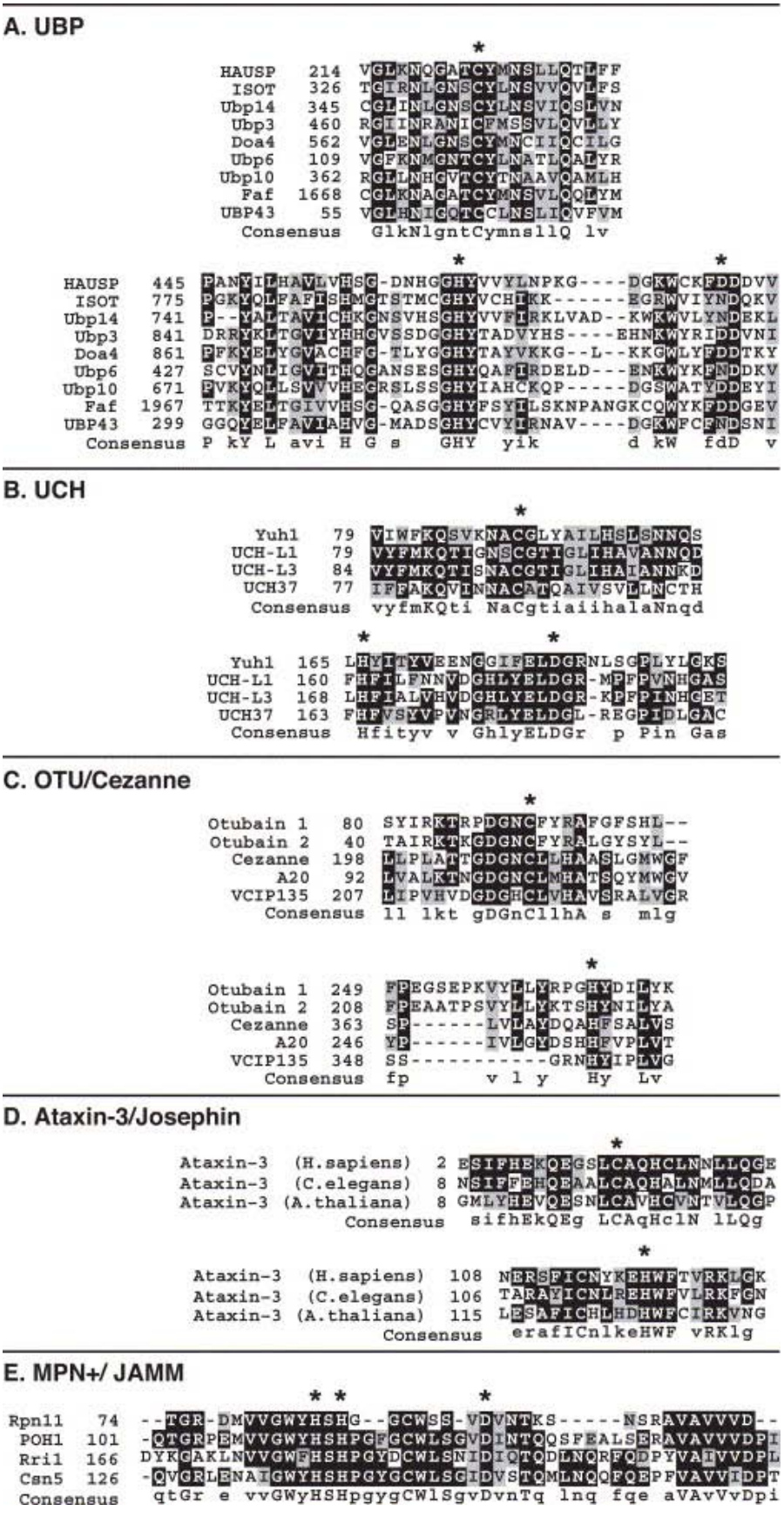
DEUBIQUITYLATION

Classes of Deubiquitylating Enzyme

As with many other forms of protein modification, ubiquitylation is a reversible process. A large and expanding family of deubiquitylating enzymes (DUBs), has members that are able to catalytically cleave ubiquitin-ubiquitin and ubiquitin-protein bonds (Baker et al., 1992; Tobias and Varshavsky, 1991; Wilkinson et al., 1989). Currently there are at least five recognised families of DUBs classified on the basis of sequence similarity and the proposed mechanism of action of the enzyme (Figure 1.5) (Nijman et al., 2005). Four of these families, the ubiquitin Carboxyl-terminal hydrolases (UCHs), ubiquitin specific peptidases (USPs), Machado-Joseph disease protein domain proteases (MJDs), and ovarian tumour proteases (OTUs) belong to the larger family of cysteine proteases. The remaining family of DUB, the JAMM (JAB1/MPN/MOV34 metalloenzyme) motif proteases are metalloproteases.

The founding member of the MJD family, ataxin-3, was identified as a DUB, along with a number of ataxin-3-like proteins, in a bioinformatic search for other forms of ubiquitin and ubiquitin-like modifier activating proteases (Scheel et al., 2003). This bioinformatics approach predicted that ataxin-3, and its related proteins, contain an amino-terminal protein fold domain with a 3-dimensional fold structure similar to that of the UCH and USP cysteine proteases (Scheel et al., 2003). A polyglutamine expansion in the gene encoding ataxin-3 leads to an inheritable neurodegenerative disease known as spinocerebeller ataxia type-3 or Machado-Joseph disease, giving rise to the name of this family of DUB (OMIM 607047). The cellular functions of the remaining MJD members are at present unclear.

Figure 1.5: ***Sequence alignments showing the conserved catalytic regions which define the five classes of deubiquitylating enzyme.*** Amino acid alignments showing consensus sequences found flanking the conserved catalytic residues (indicated by asterisks) of the five classes of deubiquitylating enzyme. **A)** The three catalytic residues of USP enzymes in the context of the consensus Cys (top) and His boxes (bottom). **B)** The catalytic Cys, His, and Asp residues of the UCHs. **C) & D)** The conserved (or predicted in the case of the ataxin-3 family) catalytic Cys and His residues of the OTU and ataxin-3 DUB families. **E)** Sequence alignment showing the core region of the MPN+/JAMM motif. SwissProt Database entries shown are: Yuh1 (*S. cerevisiae*; SW: P35127); UCH-L1 (human; SW: P09936); UCH-L3 (human; SW: P15374); UCH37 (bovine; SW: Q9XSJ0); HAUSP (human; SW: Q93009); ISOT (human; SW: P45974); Ubp14 (*S. cerevisiae*; SW: P38237); Ubp3 (*S. cerevisiae*; SW: Q01477); Doa4 (*S. cerevisiae*; SW: P32571); Ubp6 (*S. cerevisiae*; SW: P43593); Ubp10 (*S. cerevisiae*; SW: P53874); Faf (*D. melanogaster*; SW: P55824); UBP43 (mouse; SW: Q9WTV6); otubain 1 (human; SW: Q96FW1); otubain 2 (human; SW: Q96DC9); Cezanne (human; SW: Q9NQ53); A20 (human; SW: P21580); VCIP135 (rat; SW: Q8CF97); ataxin-3 (human; SW: P54252); ataxin- 3 (*C. elegans*; SW: O17850); ataxin- 3 (*A. thaliana*; SW: Q9M391); Rpn11 (*S. cerevisiae*; SW: P43588); POH1 (human; SW: O00487); Rri1 (*S. cerevisiae*; SW: Q12468); and Csn5 (human; EP: AAH01187). Adapted from (Amerik and Hochstrasser, 2004).


A. UBP
HAUSP 214 VGLKNQGATCYMNSLLQTLFF
ISOT 326 TGIRNLGNSCYLNSVVQVLFS
Ubp14 345 CGLINLGNSCYLNSVIQSLVN
Ubp3 460 RGIINRANICFMSSVLQVLLY
Doa4 562 VGLENLGNSCYMNCIIQCILG
Ubp6 109 VGFKNMGNTCYLNATLQALYR
Ubp10 362 RGLLNHGVTCYTNAAVQAMLH
Faf 1668 CGLKNACATCYMNSVLQQLYM
UBP43 55 VGLHNIGQTCCLNSLIQVFVM
Consensus GlkNlgntCymnsllQ lv
HAUSP 445 PANYILHAVLVHSG-DNHGGHYVVYLNPKG----DGKWCKFDDDVV
ISOT 775 PGKYQLFAFISHMGTSTMCGHYVCHIKK------EGRWVIYNDQKV
Ubp14 741 P--YALTAVICHKGNSVHSGHYVVFIRKLVAD--KWKWVLYNDEKL
Ubp3 841 DRRYKLTGVIYHHGVSSDGGHYTADVYHS----EHNKWYRIDDVNI
Doa4 861 PFKYELYGVACHFG-TLYGGHYTAYVKKG--L--KKGWLYFDDTKY
Ubp6 427 SCVYNLIGVITHQGANSESGHYQAFIRDELD---ENKWYKFNDDKV
Ubp10 671 PVKYQLLSVVVHEGRSLSSGHYIAHCKQP-----DGSWATYDDEYI
Faf 1967 TTKYELTGIVVHSG-QASGGHYFSYILSKNPANGKCQWYKFDDGEV
UBP43 299 GGQYELFAVIAHVG-MADSGHYCVYIRNAV----DGKWFCFNDSNI
Consensus P kY L avi H G s GHY yik d kW fdD v
B. UCH
Yuh1 79 VIWFKQSVKNACGLYAILHSLSNNQS
UCH-L1 79 VYFMKQTIGNSCGTIGLIHAVANNQD
UCH-L3 84 VYFMKQTISNACGTIGLIHAIANNKD
UCH37 77 IFFAKQVINNACATQAIVSVLLNCTH
Consensus vyfmKQti NaCgtiaiihalaNnqd
Yuh1 165 LHYITYVEENGGIFELDGRNLSGPLYLCKS
UCH-L1 160 FHFILFNNVDGHLYELDGR-MPFPVNHGAS
UCH-L3 168 LHFIALVHVDGHLYELDGR-KPFPINHCET
UCH37 163 FHFVSYVPVNGRLYELDGL-REGPIDLGAC
Consensus Hfityv v GhlyELDGr p Pin Gas
C. OTU/Cezanne
Otubain 1 80 SYIRKTRPDGNCFYRAFGFSHL--
Otubain 2 40 TAIRKTKGDGNCFYRALGYSYL--
Cezanne 198 LLPLATTGDGNCLLHAASLGMWGF
A20 92 LVALKTNGDGNCLMHATSQYMWGV
VCIP135 207 LIPVHVDGDGHCLVHAVSRALVGR
Consensus ll lkt gDGnCllhA s mlg
Otubain 1 249 FPEGSEPKVYLLYRPGHYDILYK
Otubain 2 208 FPEAATPSVYLLYKTSHYNILYA
Cezanne 363 SP------LVLAYDQAHFSALVS
A20 246 YP------IVLGYDSHHFVPLVT
VCIP135 348 SS----------GRNHYIPLVG
Consensus fp v l y Hy Lv
D. Ataxin-3/Josephin
Ataxin-3 (H.sapiens) 2 ESIFHEKQEGSLCAQHCLNNLLQGE
Ataxin-3 (C.elegans) 8 NSIFFEHQEAALCAQHALNMLLQDA
Ataxin-3 (A.thaliana) 8 GMLYHEVQESNLCAVHCVNTVLQGP
Consensus sifhEkQEg LCAqHclN lLQg
Ataxin-3 (H.sapiens) 108 NERSFICNYKEHWFTVRKLGK
Ataxin-3 (C.elegans) 106 TARAYICNLREHWFVLRKFGN
Ataxin-3 (A.thaliana) 115 LESAFICHLHDHWFCIRKVNG
Consensus erafICnlkeHWF vRKlg
E. MPN+/ JAMM
Rpn11 74 --TGR-DMVVGWYHSHG--GCWSS-VDVNTKS-----NSRAVAVVVD--
POH1 101 -QTGRPEMVVGWYHSHPGFGCWLSGVDINTQQSFEALSERAVAVVVDPI
Rri1 166 DYKCAKLNVVGWFHSHPGYDCWLSNIDIQTQDLNQRFQDPYVAIVVDPL
Csn5 126 -QVGRLENAIGWYHSHPGYGCWLSGIDVSTQMLNQQFQEPFVAVVIDPT
Consensus qtGr e vvGWyHSHpgygCWlSgvDvnTq lnq fqe aVAvVvDpi

A bioinformatics search also identified the OTU protein family (Makarova et al., 2000). The sequence of *Drosophila otu,* a gene involved in the development of the ovary (Steinhauer et al., 1989), was used as a starting point to search sequence databases for *otu* homologues (Makarova et al., 2000). Some sequence similarity was found between *otu* and its homologues and proteins of the arterioviruses which have cysteine protease activity (Makarova et al., 2000). It has since been found that at least three members of this family have DUB activity including Otubain-1, Otubain-2, and Cezanne (Balakirev et al., 2003; Evans et al., 2003). As with the MJD proteases, many of the remaining members of the OTU family have only recently been defined and as such have no prescribed cellular function or demonstrated DUB activity as yet.

In contrast to the UCH, USP, MJD and OTU deubiquitylating enzymes which are all cysteine proteases, the JAMM family of DUB has a catalytic core that is characteristic of metalloproteases. There are at least fourteen members of the JAMM domain proteases however seven of these may not be functional as proteases due to non-conservative amino acid substitutions in the zinc ion stabilising region of these proteins that is important for metalloprotease catalytic activity (Nijman et al., 2005). Three members of this family have demonstrated physiological roles with Rpn11/POH1 having ubiquitin chain editing activity resulting in recovery of ubiquitin at the proteasome (Lam et al., 1997; Verma et al., 2002), CSN5 being responsible for the deubiquitylating activity of the COP9 signalosome (Cope et al., 2002), and AMSH being involved in endocytosis (McCullough et al., 2004).

The catalytic core of the UCH class of DUBs shares a very close homology between members and has an average size of ~230 amino acids. This catalytic core is characterised by a catalytic triad of conserved Cys, His, and Asp residues that display a spatial geometry described as a papain fold and places them into the super family of cysteine proteases. The UCH class of DUBs are mainly involved in recycling monomeric ubiquitin (Ub) by breaking down isopeptide-linked chains of ubiquitin or by processing peptide-linked precursors (Ub polymers or Ub-protein translational fusions) (Liu et al., 1989; Pickart and Rose, 1985). It is thought that this specificity for smaller ubiquitin adducts is due to a steric constraint intrinsic in the fold structure of the catalytic core of the UCHs consisting of a loop structure that crosses over the catalytic cleft (Johnston et al., 1999). Although exceptions to this steric restraint have been observed, it is thought that in general the predominant substrates of these enzymes are small ubiquitin adducts. They are therefore thought to act as positive regulators of the proteasomal pathway by ensuring that cellular levels of free ubiquitin monomers are kept constant (Johnston et al., 1999).

The Ubp class of DUBs is characterised by a slightly larger catalytic core than that of the UCH family, which is on average 450 amino acid residues in length but can vary greatly among members (Wilkinson, 1997). The catalytic core contains two domains that are highly conserved between members of the family, the Cys and His boxes (Wilkinson 1997). These two regions contain the "catalytic triad" which consists of conserved cysteine, aspartic acid and histidine residues. Members of this family frequently have polypeptide insertions between the Cys and His boxes, amino-terminal extensions up to 1000 amino acids in length, and in some cases carboxyl-terminal extensions of a similar size (Wilkinson 1997). Some members of the family have functions that overlap those of the UCH family members. One example of this is the Ubp enzyme isopeptidase T. This enzyme is able to act specifically upon free-formed chains of ubiquitin that have an unanchored ubiquitin α-carboxyl group (Wilkinson et al., 1995). These chains can be generated by the action of the 26S proteasome itself or through synthesis by certain members of the E2 ubiquitin-conjugating enzyme family (Hochstrasser 1995). The action of isopeptidase T is essential in preventing inhibition of the proteasome, by removing these chains from the proteasomal subunits that bind them (Hochstrasser 1995).

The majority of Ubp enzymes have an as yet undefined function. The increasingly large number of members in this family indicates however, that Ubps are likely to act as so-called 'substrate specific' DUBs. It is proposed that the amino acid insertions found within the catalytic regions of these enzymes and/or the amino- and carboxyl- terminal extensions that are often associated with them confer this specificity. While this idea has been around for some time, only recently have specific targets for DUBs been identified. One such substrate is the aptly named 'guardian of the genome' p53. Loss of this short-lived tumour suppressor contributes to the development of around half of all human cancers (Ryan et al., 2001). In normal cells, the action of Mdm2, an E3 ubiquitin ligase, maintains low levels of p53 by mediating its degradation via the proteasome (Honda et al., 1997). In response to many forms of stress including DNA damage, oncogene activation and loss of cell growth and survival signals, p53 protein levels are stabilised mainly via post-translational events (Schuler and Green, 2001). In response to DNA damage for example, p53 becomes hyperphosphorylated, preventing its binding to Mdm2. Subsequently, this E3 ligase cannot ubiquitylate p53 resulting in its escape from degradation (Ryan et al., 2001). Recently it was discovered that p53 can also be stabilised by the Herpes simplex virus associated ubiquitin specific protease (HAUSP), a member of the Ubp family of deubiquitylating enzymes (Li et al., 2002).

HAUSP was identified as an interacting partner of p53 following GST-p53 affinity chromatography (Li et al., 2002). It was further shown that HAUSP was able to interact with p53 *in vivo* by co-immunoprecipitation with antibodies to both p53 and HAUSP (Li et al., 2002). Furthermore, HAUSP was able to specifically stabilise p53 in H1299 human lung carcinoma cells, a p53 negative cell line. Cells transfected with HAUSP and p53 showed increased levels of p53 over cells transfected with p53 alone (Li et al., 2002). This effect was specific to p53 as western analysis demonstrated that there was no change in p27, another protein regulated by the ubiquitin proteasome pathway, levels (Li et al., 2002). In addition, the specificity of HAUSP for p53 was further demonstrated by the inability of an unrelated Ubp, USP11, to bind to or stabilise p53 levels (Li et al., 2002).

Direct evidence of HAUSPs deubiquitylating ability was seen *in vitro*, when purified and ubiquitylated p53 was incubated with either wild-type or catalytically inactive HAUSP. The presence of a ubiquitylated ladder of p53 bands in untreated control samples was abolished by wild-type HAUSP but not the catalytically inactive cysteine mutation of HAUSP (Li et al., 2002). Additionally, the amino- and carboxyl- terminal extensions of HAUSP were shown to directly bind p53 (Li et al., 2002). Like most enzymes in the Ubp family, these extensions show no homology to any other protein. Hence this work has provided evidence that the Ubp group of enzymes can regulate the levels of specific, and indeed, vital protein substrates within the cell. Furthermore, they demonstrated that, at least in this case, recognition of a substrate by an Ubp can occur via the non-conserved regions of these proteins.

FAT FACETS, A DEUBIQUITYLATING ENZYME OF THE USP CLASS

Fat Facets and Fat Facets in Mouse

Genetic evidence has identified a substrate-specific deubiquitylating enzyme of the Ubp family found in *Drosophila*, Fat facets (Faf). Faf can antagonise both ubiquitylation and proteolysis (Wu et al., 1999). *Faf* mutant flies show a rough eye phenotype whereby loss or partial loss of *faf* function leads to aberrant cell fate decisions. This results in the production of extra photoreceptors in the facets of the eye (Fischer-Vize et al., 1992). In crossed strains also carrying a mutation in a gene encoding a component of the proteasome, the *UbcD1*gene, the FAF rough eye phenotype is suppressed indicating that a 50% decrease in one component of the proteasome can compensate for decreased Faf function (Wu et al., 1999). This provided evidence that Faf may have the ability to cleave the degradation tag from specific protein

substrates destined for destruction by the proteasome and thereby antagonise the degradation process (Chen et al., 2002). Indeed this was confirmed when a genetic screen for dominant enhancers of the *faf* rough eye phenotype identified the *liquid facets* (*lqf*) gene. *Lqf*, the *Drosophila* homologue of vertebrate epsin, has since been identified as the major substrate of Faf involved in eye development (Chen et al., 2002). In addition, genetic analyses suggest that *faf* may also be involved in the latter stages of eye development acting through the RTK/Ras/MAPK pathway (Fischer et al., 1997; Isaksson et al., 1997; Li et al., 1997). A second role for *faf*, also identified through genetic studies, is in embryonic development. A null mutation in *faf* causes female flies to become sterile (Fischer-Vize et al., 1992). Whereas embryonic development in normal flies involves development of the embryo as a syncytium for 14 nuclear divisions at which point the nuclei migrate to the cellular periphery and become cellularised, migration and cellularisation of the nuclei was defective in *faf* null embryos (Fischer-Vize et al., 1992). Stabilisation of the posterior patterning protein Vasa may be the role of Faf in this developmental process (Liu et al., 2003).

The murine homologue of the *faf* gene product, Fat Facets in Mouse (FAM), is co-linear with, and shows 50% identity and 70% similarity to, fat facets, and was first identified in a gene trap assay aiming to isolate genes up regulated in early mouse gastrulation and neuralation (Wood et al., 1997). FAM is also referred to as USP9X and DFFRX (*Drosophila* fat facets related X) in the literature, but for the purposes of this thesis the vertebrate X-chromosome homologues will be referred to simply as FAM. FAM is able to functionally replace *faf* in both of its critical roles in *Drosophila* however its complex temporal and tissue specific expression patterns and variable sub cellular localisation in mammals defy a simple explanation of its normal cellular function. FAM transcripts are found in several tissues of the developing mouse embryo including the mitotic cell populations of the gastrulating and neurulating embryo, post mitotic cells in the central nervous system, and the apoptotic regions between the digits (Wood et al., 1997). This complex and varied pattern of expression suggests FAM may have roles in several cellular and developmental processes.

The intracellular localisation of FAM protein shows a similar degree of complexity and is very likely to be dependent on the cellular environment. Dramatic relocalisation of the FAM protein has been observed in response to both developmental events and simple changes in cellular environment. In the unfertilised ovulated mouse egg FAM protein is localised to the spindle and chromosomes with weak punctate staining in the cytoplasm. Following fertilisation of the egg, there is an increase in the punctate cytoplasmic staining with a corresponding loss of staining in both of the pronuclei. Furthermore, this localisation changes

again in subsequent cleavage events up to the point of blastocyst formation, with staining becoming concentrated in a perinuclear ring (Pantaleon et al., 2001). FAM localisation can also shift in response to cellular signalling events. When cells of the non-transformed human keratinocyte cell line HaCat were exposed to soluble EGF, FAM localisation was altered from a primarily punctate cytoplasmic staining, to also include defined staining on the cell periphery (M. Scherer, PhD candidate, University of Adelaide, SA, personal communication). In most epithelial and fibroblastic cell types studied to date (T84, Caco2, L-cells and HEK293T) FAM displays primarily a punctate cytoplasmic staining with a strong perinuclear concentration at the Golgi apparatus (Mouchantaf et al., 2006; Murray et al., 2004). However, staining at the cell surface, and more specifically at points of cell-cell contact, has also been observed (Friocourt et al., 2005; Taya et al., 1998).

One context in which FAM function has been observed to be critical is in the pre-implantation mouse embryo. Antisense oligonucleotide knockdown of FAM delayed the progression of pre-implantation embryos from the 2-cell to the blastocyst stage, a defect most likely resulting from observed decreases in blastomere cleavage rates and a loss of cell-cell adhesion (Pantaleon et al., 2001). Another developmental stage in which FAM function may be critical is in development of the mouse eye, which may be consistent with Faf's role in *Drosophila* eye development. High levels of FAM expression have been observed in the retinal pigmented epithelium, and some neural cell types of the developing mouse eye although there have been no reports of defects in mammalian eye development linked to mutations in the FAM gene (Kanai-Azuma et al., 2000)

Fam is located on the sex chromosomes in mammals and has both X and Y homologues. In mice, the Y homologue, *Dffry* or *usp9y*, is expressed exclusively in the testes and maps to the Sxr^b deletion (Brown et al., 1998). This deletion is associated with an early post-natal blockage of spermatogonial proliferation and differentiation leading to absence of germ cells (Bishop et al., 1988; Mardon et al., 1989). The human Y homologue of *Fam* is closely associated with oligozoospermia (Sargent et al., 1999; Sun et al., 1999) and the human X homologue has been linked to the failure of oocytes to pass through the first meiotic prophase in Turner syndrome (Cockwell et al., 1991; Speed, 1986). Despite these associations, the substrates and precise role of *Fam* and its homologues in these processes have not yet been defined. Due to the complex nature of *Fam* expression and the lack of data tying FAM to specific cellular functions, much attention has been paid in identifying interacting partners and cellular targets of FAM activity to aid in the definition of its role in the cell and development.

FAM is a large protein of approximately 290 kDa (Wood et al., 1997). Its catalytic region, containing the characteristic Cys and His boxes conserved between all members of the Ubp family, spans amino acids 1476 - 1918 (Taya *et al.,* 1998). As is commonly found with Ubps, FAM has large amino- and carboxyl-terminal extensions that flank the catalytic region and are believed to be involved in substrate recognition (Friocourt et al., 2005; Wood et al., 1997). Until relatively recently, no other structural features were easily identifiable outside of the catalytic region of the protein. Bioinformatic searches have now identified a number of putative structural features of the protein. These include a predicted calmodulin binding motif, two predicted clathrin binding motifs or clathrin boxes, an AP-2 binding motif, an armadillo repeat region, and predicted annexin, and ubiquitin-like domains (Birney et al., 2006; Ginalski et al., 2004; Puntervoll et al., 2003) (Figure 1.6 a and b). Despite the early lack of information on the structure of the FAM protein, several substrates and interacting proteins have been identified.

FAM and its Substrates

FAM and Cell-Cell Adhesion

The first FAM interacting protein was discovered during a screen to identify proteins interacting with the tight junction associated protein AF-6 (acute lymphoblastic luekemia-1 fusion partner from chromosome 6) (Taya et al., 1998). Subsequent experiments found that FAM protein co-localised with a second cell-cell adhesion protein, β-catenin, and led to the hypothesis that FAM is involved in the regulation of cell-cell adhesion complexes (Taya et al., 1999; Taya et al., 1998).

AF-6/Afadin

The best characterised role of AF-6/Afadin is as a structural element of both adherens junctions (AJs) and of tight junctions (TJs) (Mandai et al., 1997; Yamamoto et al., 1997). There are three known splice variants of the protein varying in length at the C-terminus (Saito et al., 1998). Structural features that are common to all splice variants of the AF-6/Afadin protein include a ras binding domain, a PDZ domain, an U104 domain, and a myosin V-like domain, (Ponting, 1995). The function of the latter two domains in the context of the AF-6 protein is as yet uncharacterised (Ponting, 1995). The ras binding domain mediates the

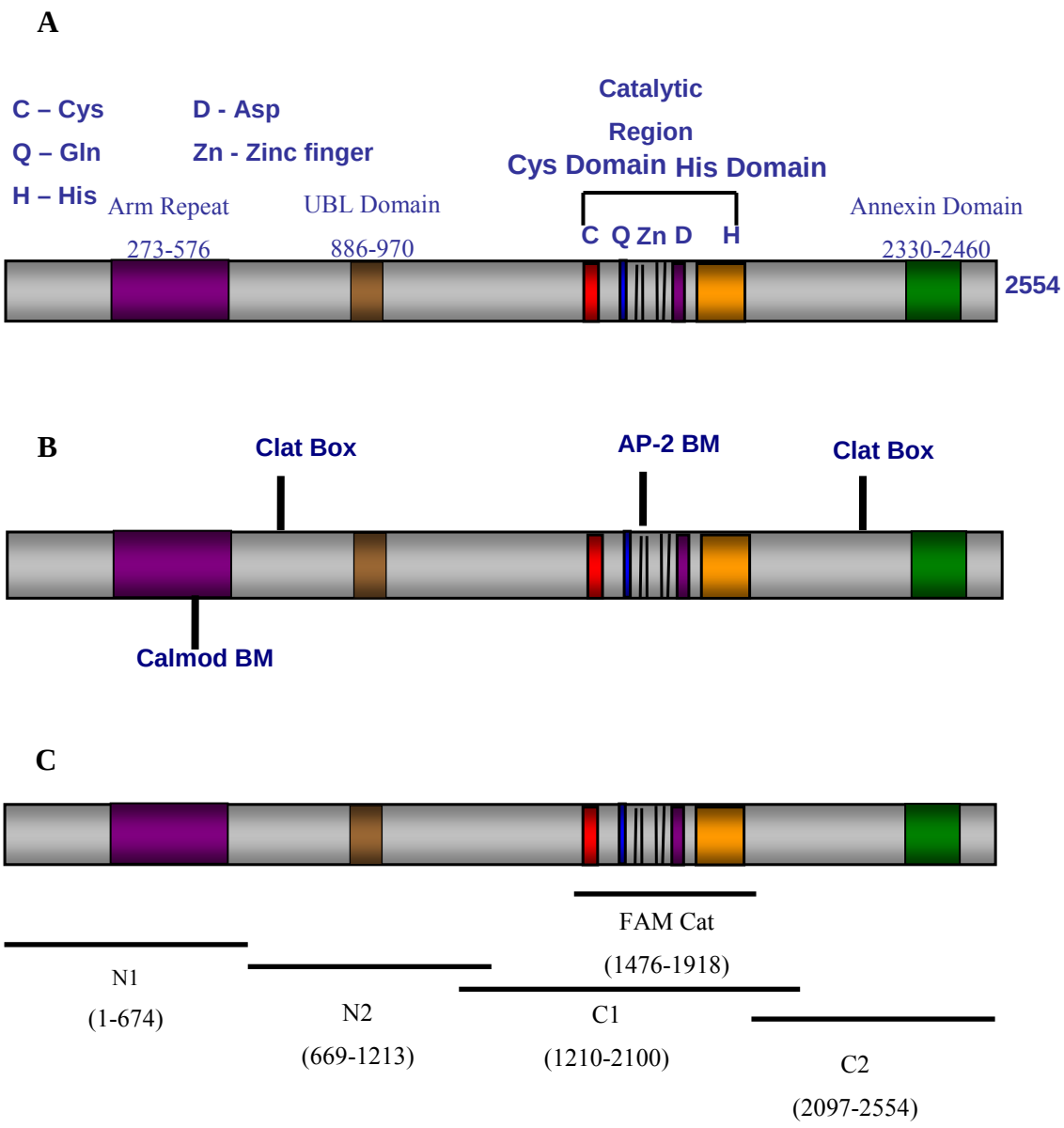


Figure 1.6: ***Putative Domains and Motifs of the FAM Deubiquitylating Enzyme*** **A)** Box diagram of the FAM deubiquitylating enzyme showing the major known and predicted structural features. These include a predicted armadillo repeat region at the N-terminus, followed by a predicted ubiquitin-like domain. The catalytic core of the enzyme is bounded by the cys and his boxes which contain the conserved catalytic cys and his residues. Also within this region is the conserved asp residue rounding out the catalytic triad, and a conserved gly residue and zinc finger motifs. **B)** A number other features of the FAM protein were discovered in a search for expressed linear motifs on the ELM server (Puntervoll et al., 2003). These include predicted binding site motifs for calmodulin (a calcium signalling responsive protein), AP-2, and two clathrin box motifs for interaction with clathrin. **C)** The regions of the FAM spanned by the four fragments and the catalytic core designed by Taya and colleagues (Taya *et al* 1998)

interaction of AF-6 with Ras GTPases and has implicated AF-6 in signal transduction pathways (Boettner et al., 2000; Kuriyama et al., 1996; Yamamoto et al., 1999). The ras binding domain has also been linked to the interaction of AF-6 with the tight junction protein ZO-1 (Yamamoto et al., 1997). PDZ domains have been characterised as protein-protein interaction domains and in the case of AF-6, have been shown to mediate it's interaction with the tight junction associated protein ZO-1 (Ebnet et al., 2000) and with Eph related receptor tyrosine kinases (Hock et al., 1998).

AF-6/Afadin and FAM

The interaction of AF-6 and FAM was demonstrated by their reciprocal co-immunoprecipitation from Mardine Darby canine kidney II (MDCK II) epithelial cells and mouse eyes (Taya et al., 1998; Kanai-Azuma and Wood, Unpublished data). In immunohistochemistry experiments, FAM was found to co-localise with AF-6 in the developing mouse eye (Kanai-Azuma et al., 2000) as well as at sites of cell-cell contacts in MDCK II cells (Taya et al., 1998). Ectopic expression of a fragment of FAM (amino acids 1210-2410), which contains both of the catalytic cys and his boxes, resulted in stabilisation of total AF-6 levels in the African green monkey kidney fibroblast cell line COS7, suggesting that AF-6 is a substrate for FAM's deubiquitylating activity (Taya et al., 1998).

In vitro GST-pulldown experiments demonstrated a direct interaction between FAM and AF-6. Using GST-tagged AF-6 and four *in vitro* transcribed and translated fragments of FAM (Figure 1.6c) (amino acids 1-669, 669-1213, 1210-2100, and 2097-2554), it was found that an 890 amino acid stretch of the FAM protein (amino acids 1210-2100) which contained both the catalytic Cys and His boxes, was the only fragment that could bind to AF-6 (Taya et al., 1998). This interaction was subsequently refined to a smaller 450 amino acid region, amino acids 1497-1918, designated FAM-Cat (Taya et al., 1998). The identification of AF-6 as a substrate of FAM and the co-localisation of AF-6 and FAM to points of cell-cell contact in MDCK II cells raised the possibility that FAM regulated cell adhesion. It was proposed that FAM might be recruited to sites of adherence complexes by a direct interaction with AF-6 (Taya et al., 1998). A second putative substrate of FAM was identified while investigating its co-localisation with AF-6 in MDCK II cells. β-catenin, a protein involved in the formation of adherens junctions, co-stained with FAM at the basolateral plasma membrane (Taya et al., 1998).

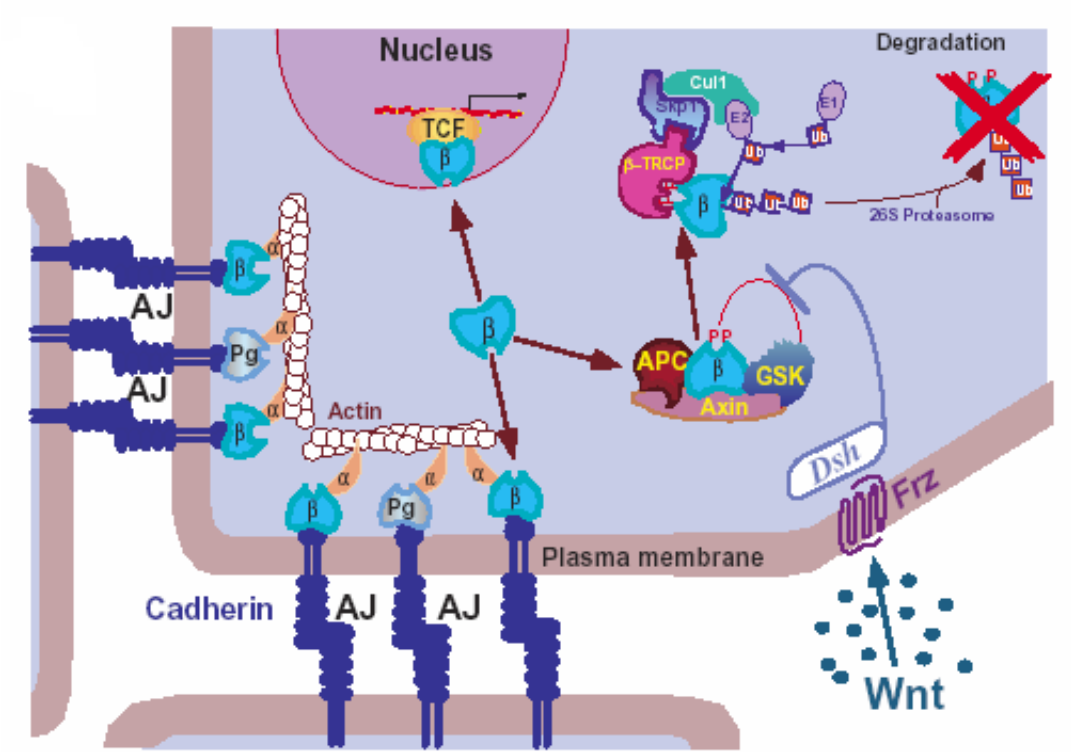
β-catenin and the Importance of its Regulation in the Cell

Adherens junctions are found at the basolateral sites of cell-cell contact. They are involved in cell-cell recognition and are often the first adhesion complexes to form between epithelial cells. β-catenin is essential in the formation of these junctions, and was first recognised for this role. In its cell adhesion capacity, β-catenin provides a structural link between cell adhesion molecules of the calcium-dependent classical cadherin family and the actin cytoskeleton. It does so by forming a bridge between classical cadherins such as E-cadherin and the actin binding molecule α-catenin (Hulsken et al., 1994).

A second role for β-catenin is as a vital player in the canonical wnt/wingless (wg) signal transduction pathway. This signalling pathway plays an important role in body axis formation, development of the central nervous system, axial specification in limb development and mouse mammary gland development (Akiyama, 2000). The essential role of β-catenin in mammalian embryonic development was demonstrated in mouse knock out studies. Loss of β-catenin was embryonic lethal at 6.5-7 days post coitum (Haegel et al., 1995). Specifically, β-catenin$^{-/-}$ embryos exhibited defects in anterior-posterior axis formation, demonstrated by the absence of the anterior differentiation markers Hex, Hesx1, Otx2 and engrailed. Also, β-catenin$^{-/-}$ embryos lack expression of *brachyury* and *goosecoid*, markers of posterior differentiation (Huelsken et al., 2000). Additionally, conditional knockout and transgenic mouse experiments have shown a crucial role for β-catenin in many other developmental processes (Chenn and Walsh, 2002; Chenn and Walsh, 2003; Huelsken et al., 2001; Jamora et al., 2003; Lickert et al., 2002 and reviewed in Van Amerongen, 2006).

The wnt/wg signalling pathway utilises the precise nature of the ubiquitin/proteasome system to ensure its correct function. In cells that do not receive a wnt/wg signal, cytoplasmic β-catenin that is not involved in cell adhesion is constantly produced and degraded in a futile cycle. In these cells, excess cytoplasmic β-catenin is found in multi-protein complexes containing the scaffold protein axin/conductin, the adenomatous polyposis coli (APC) protein, and glycogen synthase kinase-3β (GSK-3β). It is these complexes which target β-catenin for degradation by the proteasome (Figure 1.7) (von Kries et al., 2000). Following phosphorylation of β-catenin's amino-terminus by casein kinase 1 (CK-1) and GSK-3β, the protein is ubiquitylated by the E3 ligase β-TrCP/Slimb and subsequently, β-catenin is degraded in the proteasome (Clevers, 2006; von Kries et al., 2000).

Figure 1.7: ***The Dual Roles of β-catenin*****.** β-catenin plays an essential role in the formation of the adherens junctions (AJ) that are required for cell-cell adhesion. It does so by forming a bridge between members of the classical cadherin family and the actin cytoskeleton binding protein α-catenin. β-catenin also has an essential role in mediating wnt/wingless (wg) signalling. In the absence of a wnt/wg signal, β-catenin is constantly produced and degraded in a futile cycle. Cytoplasmic β-catenin is bound in a complex consisting of the scaffold protein axin/conductin, the adenomatous polyposis coli (APC) protein, and the serine/threonine kinase glycogen synthase kinase-3β (GSK-3β). In this complex, GSK-3β phosphorylates serine residues on the N-terminus of β-catenin, which acts as a signal for its ubiquitylation by the E3 ligase β-TrCP/Slimb. Following polyubiquitylation of β-catenin it is degraded by the 26S proteasome. When a member of the wnt family binds to its receptor, one of the membrane spanning frizzled (frz) receptors, a signalling cascade is activated. Binding of wnt to frz leads to the activation of disheveled (dsh). Dsh acts upon the APC/axin/conductin/GSK-3β complex and prevents phosphorylation of β-catenin. This leads to the accumulation of a cytoplasmic pool of β-catenin that is able to interact with members of the Tcf/LEF class of DNA binding proteins. These transcriptionally active heterodimers move into the nucleus where they induce the transcription of wnt/wg target genes, which include c-Myc and cyclin -D1. β = β-catenin, PG = plakoglobin, α = α-catenin. Adapted from (Zhurinsky et al., 2000).


Nucleus
Degradation
TCF
β
Cul1
Skp1
E2
E1
Ub
β-TRCP
26S Proteasome
AJ
Pg
AJ
α
β
Actin
APC
GSK
Axin
PP
Dsh
Frz
Plasma membrane
Cadherin
AJ
AJ
Wnt

In the presence of a wnt/wg signal, a signal cascade is initiated. This cascade begins with the binding of one of the secreted wnt family proteins to its receptor, a member of the Frizzled (Fz) family of transmembrane proteins, and co-receptor, Arrow/Lrp5/6. The Fz-Arrow/Lrp5/6 receptor complex then activates a cytoplasmic protein, dishevelled (Dsh or Dvl), either by a phosphorylation event possibly mediated by the serine/threonine kinase casein kinase 2 (CK-2), or another as yet unidentified mechanism (Boutros and Mlodzik, 1999; Clevers, 2006). In addition, wnt signalling leads to phosphorylation of the cytoplasmic tail of Lrp by GSK-3β and CK-1, which can act as a docking site for Axin (Davidson et al., 2005; Zeng et al., 2005). Activated Dvl, perhaps in combination with sequestration of Axin to the plasma membrane, inhibits the activity of GSK-3β thereby preventing phosphorylation of β-catenin (Wodarz and Nusse, 1998). The deactivation of GSK-3β via wnt signalling, leads to the accumulation of a cytoplasmic pool of β-catenin. This free β-catenin is able to move into the nucleus, where it binds to members of the TCF/LEF family of transcription factors (Behrens et al., 1996). Upon binding of β-catenin to one of these factors, a transcriptionally active heterodimer is formed. This transcription factor complex then proceeds to initiate transcription of wnt/wg target genes through the combined action of the DNA bindingdomain of TCF/LEF family members, and the transactivation domain found at the amino- and carboxyl- terminal regions of β-catenin (Figure 1.7) (Zhurinsky et al., 2000).

Target genes of the wnt/wg signalling pathway include the cell cycle regulator cyclin D1 whose promoter contains consensus Tcf/LEF binding sites and can be activated by β-catenin, and the proto-oncogene c-Myc (He et al., 1998; Tetsu and McCormick, 1999). It is not surprising therefore, that β-catenin in its capacity as a mediator of the wnt/wg signalling pathway, has the potential to lead to deregulated cell growth and cancer. This can occur either by mutations that decrease or abolish the function of factors involved in the down regulation of β-catenin, or by mutation of β-catenin itself (Polakis, 2007). The involvement of β-catenin in such a variety of pathogenic processes, and the inability of mouse embryos to develop in β-catenin knockout experiments, demonstrate the need for tightly controlled regulation of the levels of free cytoplasmic β-catenin in the cell.

For this reason, the β-catenin destruction complex is not the only control mechanism for regulating the wnt signalling pathway and β-catenin activity in the cell. Other nuclear proteins have been reported to interact with and modulate β-catenin transcriptional activity in the nucleus including Chibby and the Sox proteins (Sharpe et al., 2001; Takemaru et al., 2003). Regulation of the wnt pathway mediated by secreted proteins including antagonists of

the Arrow/Lrp receptors, the Dickkopf (Dkk) proteins (Glinka et al., 1998; Mao et al., 2002; Mao et al., 2001), and the Soluble Frizzled-Related Proteins (SFRPs) that may act to either titrate out wnt proteins or stabilise them has been reported (Logan and Nusse, 2004). Tyrosine phosphorylation of β-catenin has also been shown to modulate its ability to interact with the cadherin adhesion proteins and other nuclear factors (Piedra et al., 2001; Piedra et al., 2003; Roura et al., 1999). Indeed, it seems that even the pool of β-catenin in adhesion complexes at the plasma membrane does not escape regulation with the observation that the mature cadherin-catenin complex can be ubiquitylated by the E3 ubiquitin ligase Hakai (Fujita et al., 2002). Thus there are many levels at which β-catenin is regulated in the cell.

Interactions of FAM and β-catenin

Evidence suggests that β-catenin may also be regulated by FAM. As discussed above, FAM co-localises with β-catenin at adherens junctions in MDCK II cells (Taya et al., 1998). It also extensively co-localises with β-catenin in sub-confluent cultures of the human epithelial colorectal cancer cell line T84 (Murray et al., 2004). Reciprocal co-immunoprecipitation of FAM and β-catenin has also been demonstrated in a number of cell lines and tissues including MDCK II epithelial cells, mouse L fibroblasts (Taya et al., 1999), mouse embryonic stem (ES) cells, T84 cells, the human colonic adenocarcinoma cell line Caco-2, and from extracts taken from mouse brain (Murray et al., 2004 and L. Jolly, PhD candidate, The University of Adelaide, SA, personal communication).

Evidence suggests that β-catenin is stabilised by Fam. Mouse L cells stably transfected with an extended HA-tagged version of FAM-Cat (amino acids 1210-2410) showed a stabilisation of β-catenin *in vivo* (Taya *et al* 1999). In addition, in pulse chase experiments with FAM-cat (1210-2410) transfected cell lines, the half-life of β-catenin was increased from 5 minutes in control transfected cells to 10 minutes (Taya et al., 1999). Lastly, double immunofluorescence using antibodies to β-catenin and HA, was performed on HA-FAM cat (1210-2100) and HA clones. This revealed higher intensity staining with anti-β-catenin antibody in the HA-FAM-Cat clones, over clones expressing the HA tag alone (Taya et al., 1999).

A direct interaction between Fam and β-catenin has also been reported and a small amount of work has been done in identifying the interacting regions of both FAM and β-catenin (Taya et al., 1999). Deletion analysis of β-catenin revealed that its armadillo repeats

were critical for an interaction with FAM (Taya et al., 1999). However, while the minimal AF-6 binding region of FAM (amino acids 1497-1918) was also found to bind to GST tagged β-catenin (Taya et al., 1999), the ability of other parts of the FAM protein to bind to β-catenin was not investigated (Taya et al., 1999).

Functional Significance of the FAM/β-Catenin Interaction

The knockdown of FAM using antisense oligonucleotides in pre-implantation embryos indicated that a reduction in FAM levels was accompanied by a reduction in β-catenin levels (Pantaleon et al., 2001). Interestingly the β-catenin remaining in the blastomeres was still tightly localised at sites of cell-cell contact (Pantaleon et al., 2001). It was suggested that FAM might be responsible for stabilising the large cytoplasmic pool of β-catenin normally found in pre-implantation embryos (Pantaleon et al., 2001). In other cell types, FAM is instead associated with the pool of β-catenin that traffics with vesicular E-cadherin transiting to or from the plasma membrane (Murray et al., 2004).

In sub confluent populations of the intestinal epithelial cell line T84, β-catenin is localised to points of cell-cell contact and in cytoplasmic puncta. Upon reaching confluence, β-catenin becomes increasingly restricted to sites of cell-cell contact presumably as junctional complexes between cells mature and increase in number. In sub confluent cells, FAM co-localises with the β-catenin containing cytoplasmic puncta, but this co-localisation is lost upon these cells reaching confluence. Thus it seems that FAM protein does not associate with the mature E-cadherin/β-catenin adhesion complex. This is supported by immunoprecipitation data. Fractionation of intact protein complexes under non-denaturing conditions from both confluent and sub-confluent T84 cell extracts revealed that FAM protein is present in two pools, one of ~2 MDa and the other of 0.2-0.7 MDa. In sub confluent cells, β-catenin and E-cadherin are present in both pools whereas in confluent cells, E-cadherin is restricted to the larger pool, which also contains the AJ protein, p120ctn. The larger fraction containing p120ctn is likely to represent protein interactions from mature junctional complexes at the plasma membrane. Anti-FAM antibodies were used to co-immunoprecipitate proteins from both of these fractions but it was found that E-cadherin and β-catenin only co-immunoprecipitated from the smaller pool, suggesting that FAM did not associate with mature E-cadherin/β-catenin adhesion complexes (Murray et al., 2004).

Further support for the role of FAM in the trafficking of the E-cadherin/β-catenin complex was given by experiments performed in the E-cadherin negative mouse L fibroblast cell line. In these cells, FAM was found to have a staining pattern that was predominantly associated with the Golgi apparatus. Upon the introduction of E-cadherin however, FAM localisation was shifted to a predominantly punctate cytoplasmic staining pattern (Murray et al., 2004). Moreover, when FAM protein levels were increased in the MCF-7 cell line, which expresses both β-catenin and E-cadherin, a reciprocal increase in membrane localised β-catenin was observed while cytoplasmic and nuclear levels were unchanged (Murray et al., 2004). In these experiments, FAM protein was only found in the cytoplasmic fraction, suggesting that the increase in membrane associated β-catenin is a result of increased trafficking of the E-cadherin/β-catenin complex to the cell surface.

It is possible that FAM has a role in both the biosynthetic and endocytic trafficking of the E-cadherin/β-catenin complex. As discussed above, ubiquitylation of proteins has been demonstrated to be involved in all stages of protein transport studied to date. At the cell surface, the E-cadherin/β-catenin complex is known to be ubiquitylated by the ubiquitin ligase Hakai, in response to Src kinase (Fujita et al., 2002). It is possible that this ubiquitylation may be involved in the internalisation and intracellular transport of the E-cadherin/β-catenin complex and that FAM may modulate this process at points in the endosomal system or, in those cells in which FAM is seen at the cell periphery, at the plasma membrane. Indeed, ubiquitylation of the E-cadherin/β-catenin complex by Hakai is enhanced in the presence of DUB inhibitors (Fujita et al., 2002).

FAM and Proteins of the Trafficking Machinery

As might be expected from the complex spatial and tissue specific expression patterns, and the size of the FAM protein, it might interact with a number of proteins. Increasingly some of these interacting proteins and potential substrates have either demonstrated or potential roles in protein trafficking. The list includes the endocytic adaptor protein, epsin 1, the HECT domain E3 ligase itch/AIP4, and the microtubule stabilising protein, doublecortin (DCX). The roles of these interacting partners in protein trafficking pathways, and the evidence for their interaction with FAM are discussed below.

Epsin1

Vertebrate epsins are a family of adaptor proteins that are intimately involved in the endocytic process. At least four genes encoding epsin proteins have been identified in humans. Epsin 1 is the founding member and has been studied in the greatest detail (Chen et al., 1998). Epsin 2 has two isoforms and, like epsin 1, is highly expressed in the brain, but also has a broader expression pattern (Rosenthal et al., 1999). Epsin 3 was identified in experiments that reproduced extracellular matrix activation of keratinocytes (Spradling et al., 2001). Epsin 4 is the newest homologue to be identified and has a genetic association with schizophrenia (Pimm et al., 2005). All four members of the family are characterised by a multidomain structure that allows them to act in their capacity as adapter proteins in the events surrounding endocytosis. The structural features of the epsins include clathrin binding regions, DPW/DPF repeat regions, NPF repeats, and the ENTH (epsin NH2-terminal homology) domain. The clathrin binding domain allows the epsins to recruit clathrin to newly forming endocytic invaginations (Drake et al., 2000). DPW/DPF motifs allow the epsins to recruit AP-2, while the NPF motif has been demonstrated to interact with Eps15 and intersectin, both essential components of clathrin mediated endocytosis (Chen et al., 1998; Drake et al., 2000; Yamabhai et al., 1998). The ENTH domain is responsible for the interaction of the epsins with the plasma membrane as it is able to bind to phosphatydlinositol (4,5)-bisphosphate (PIP_2) (Itoh et al., 2001). *In vitro* reconstitution studies have demonstrated the ability of the ENTH domain of epsin 1 to cause curvature of lipid membranes and for this reason, the ENTH domain is believed to be involved in membrane invagination in endocytosis (Ford et al., 2002). The ENTH domain has also been found in a number of other proteins involved in protein trafficking (Itoh et al., 2001; Kay et al., 1999). In addition, Epsin 1 has three UIM domains that can bind to ubiquitin, although whether mono-, multi-, or indeed poly-ubiquitylation is recognised is still under debate (Barriere et al., 2006; Hawryluk et al., 2006; Hicke et al., 2005; Madshus, 2006).

The regulation of epsin 1 activity has been reported to occur through a number of mechanisms. Epsin 1 can be phosphorylated in a mitotically induced manner and is constitutively phosphorylated in nerve terminals. Phosphorylation of Epsin 1 decreases its ability to interact with the endocytic adapter protein AP-2 (Chen et al., 1999). Epsin 1 has also been found to be associated with the Ral1 GTPase, however the significance of this interaction is yet to be determined (Morinaka et al., 1999). The modification of epsin 1 that has probably sparked the most interest, and the most debate, is its ubiquitylation. Ubiquitylation of epsin 1 has been reported to require the presence of its UIMs and probably

occurs within the ENTH domain although the exact site(s) of ubiquitylation have not yet been identified (Miller et al., 2004; Oldham et al., 2002; Polo et al., 2002). Ubiquitiylation of epsin can occur in reponse to signalling by the EGFR (Oldham et al., 2002; Polo et al., 2002). In addition, the ubiquitylation status of epsin at nerve cell termini can be regulated by depolarisation. Treatment of synaptosomes (a model of nerve cell synapses) with potassium in the presence of extracellular calcium (Ca^{++}), results in depolarisation and a global decrease in ubiquitylated proteins. The low level of ubiquitylated epsin 1 found in resting nerve cell termini was also found to be decreased upon (Ca^{++} dependent) depolarisation (Chen et al., 2003). Ubiquitylation of epsin 1 has been proposed to reduce its ability to interact with AP-2, clathrin and lipids (Chen et al., 2003).

FAM and Epsin

The *Drosophila* homologue of FAM, Faf, has been shown both genetically and biochemically to interact with the *Drosophila* homologue of Epsin1, liquid facets (Lqf). The genetic evidence is based on four observations; (i) loss of function mutations of the *Lqf* gene dominantly enhanced loss of function mutations of *faf*. (ii) Loss of function of either protein produced an almost identical mutant eye phenotype. (iii) both genes were required in the same subset of cells in the eye and (iv) overproduction of Lqf through the presence of an extra copy of the gene compensated for the lack of Faf in the eye (Cadavid et al., 2000; Fischer et al., 1997). Biochemical data has also been published on the interaction of Faf and Lqf. Loss of Faf protein in the developing eye of $faf^{-/-}$ mutants, led to a corresponding two- to three-fold reduction in the levels of Lqf and this reduction could be reversed by addition of one copy of the *faf* gene. Faf and Lqf were co-immunoprecipitated from *Drosophila* embryos demonstrating that the two proteins could be found in a complex together (Chen et al., 2002). However, bacterially expressed or *in vitro* transcribed and translated partial Faf and full length Lqf protein were unable to interact in *in vitro* GST-affinity chromatography assays raising some doubt as to whether this is a direct interaction, or that only full length Faf protein can interact (Cadavid et al., 2000). As FAM is able to functionally replace Faf in the mutant eye phenotype (Chen et al., 2000) it suggests that FAM may be able to interact with the closest mammalian homologue of Lqf, epsin 1.

Indeed this does appear to be the case. FAM has been shown to co-localise with epsin 1 in sections of the rat brain, and anti-epsin 1 antibodies were able to co-immunoprecipitate FAM (Chen et al., 2003). Further evidence comes from the ability of a bacterially expressed GST fusion with the catalytic region of FAM (amino acids 1554-1953) to specifically bind to

epsin 1 (Chen et al., 2003). The enzyme/substrate relationship of FAM and epsin was supported by experiments in which FAM levels were decreased by siRNA knockdown. Cells subjected to calcium signalling demonstrated a global deubiquitylation of proteins including epsin 1. When cells were tranfected with FAM siRNA knockdown constructs however, ubiquitylated epsin 1 levels were stablised while global protein deubiquitylation was not significantly affected (Chen et al., 2003).

Itch/AIP4

The HECT domain containing E3 ligase Itch, also known as AIP4 (Atrophin-1-interacting protein 4), has a large and growing number of substrates that link its function to many cellular processes including immune responses and inflammation, and the regulation of epithelial and haematopoetic cell growth (Bai et al., 2004; Gao et al., 2004; Heissmeyer et al., 2004; Liu, 2004; Perry et al., 1998; Qiu et al., 2000). Itch is characterised by the presence of an amino-terminal C2 domain which is responsible for its intracellular localisation on endosomes, four WW domains, commonly used for mediating protein-protein interactions, and its HECT domain which confers the ubiquitin ligase activity on the protein(Angers et al., 2004; Liu, 2004). Some of the substrates that have been identified for Itch include the p53 related transcription factors p63 and p73 (Rossi et al., 2006; Rossi et al., 2005; Rossi et al., 2006), Notch and HEF1(Feng et al., 2004; Qiu et al., 2000), the G-protein coupled receptor CXCR4 (Marchese et al., 2003), the tight junction protein occludin (Traweger et al., 2002), and key components of the cellular protein trafficking machinery including Cbl, Hrs, and endophilin (Angers et al., 2004; Courbard et al., 2002; Magnifico et al., 2003; Marchese et al., 2003). This list is far from a complete but a common theme suggests that itch regulates intracellular trafficking, and in some instances, the lysosomal degradation of these proteins. Indeed, regulation of the ring finger E3 ligase deltex (DTX) by itch was the first demonstrated case of the involvement of K29-linked ubiquitylation in the lysosomal degradation of a protein (Chastagner et al., 2006).

FAM and Itch

FAM was initially identified as a binding partner for itch in an affinity chromatography assay using rat brain lysate (Mouchantaf et al., 2006). The interaction was confirmed *in vivo* in immunoprecipitation experiments where anti-itch antibodies were able to co-immunoprecipitate FAM. Interestingly, FAM was also co-immunoprecipitated with antibodies raised against endophilin, a previously characterised substrate of itch (Mouchantaf

et al., 2006). Endophilin is another protein that is intimately involved in clathrin dependent endocytosis (Mousavi et al., 2004). It interacts with amphiphysin and dynamin via an SH3 domain (Gad et al., 2000; Micheva et al., 1997; Ringstad et al., 1997), and its amino-terminal region has acetyltransferase activity that may be involved in the pinching off of newly formed clathrin coated vesicles (Mousavi et al., 2004). This suggests that FAM and itch are in a complex together at points in the cell where proteins are being endocytosed, further strengthening the involvement of FAM in protein trafficking. A direct interaction between FAM and itch was determined using far western blotting with purified GST-Itch and cell lysates from HEK (human embryonic kidney) 293T cells (Mouchantaf et al., 2006). No direct interaction was seen between purified GST-endophilin or GST alone and FAM (Mouchantaf et al., 2006). Co-immunoprecipitation of FAM with anti-endophilin antibodies is therefore likely to have occurred indirectly through the interaction of itch with endophilin. Finally, the binding site for FAM was identified as lying within the WW domains of itch through affinity chromatography experiments (Mouchantaf et al., 2006).

In addition to interacting with itch, FAM is able to modulate the stability of itch. Overexpression of V5-tagged FAM in HEK 293T cells also expressing epitope tagged flag-itch and myc-ubiquitin led to a loss of the ubiquitylated species of itch identified by both anti-Flag and anti-Myc antibodies (Mouchantaf et al., 2006)..Overexpression of FAM also resulted in an increase in the total cellular levels of itch (Mouchantaf et al., 2006). In addition, when FAM protein levels were reduced by transfection of two independent siRNA constructs directed against FAM transcripts, there was a corresponding drop in the level of endogenous itch (Mouchantaf et al., 2006). Furthermore, an analysis of the relative levels of both FAM and itch in a number cell lines and a range of different rat tissues, found a strong correlation. Tissues that had an abundance of FAM protein exhibited a correspondingly high amount of itch, whereas tissues with low levels of FAM showed a decreased amount of itch (Mouchantaf et al., 2006).

Doublecortin

The formation of the mature cerebral cortex is a complex process which involves neurogenesis, formation of synaptic connections, neuronal differentiation, and neuronal migration. The microtubule associated protein doublecortin (DCX) is essential for neuronal migration and differentiation in the mammalian cerebral cortex (Bai et al., 2003). Mutations in the *doublecortin* gene have been found to be the genetic basis behind the brain cortical dysgenesis disorders, subcortical laminar heterotrophia (SCLH) and type-I lissencephaly (des

Portes et al., 1998; Gleeson et al., 1998). Little is known about the precise function of DCX though it is known to bind to and stabilise microtubules (Horesh et al., 1999; Moores et al., 2004), a process that is regulated by phosphorylation, and can stimulate neuronal migration (Schaar et al., 2004; Tanaka et al., 2004). DCX has also been shown to associate with the μ subunits of the TGN and endocytic associated clathrin adaptor protein complexes AP-1 and AP-2 respectively (Friocourt et al., 2001).

FAM and Doublecortin

Yeast two-hybrid studies aimed at understanding the biochemical and mechanistic functions of DCX in neuronal migration and neural development, identified FAM as a novel binding partner (Friocourt et al., 2005). Two interacting FAM peptides were identified, the shortest of which corresponded to the carboxyl-terminal most 255 amino acids of FAM, a region that is highly conserved between vertebrate FAM and its *Drosophila* homologue, Faf (Friocourt et al., 2005). The binding region of FAM on DCX was further mapped to the second microtubule binding repeat region which corresponds to the region bounded by amino acids 110 to 214 of DCX (Friocourt et al., 2005). Interestingly, a known mutation in this region (R192W) from a patient with lissencephaly disrupted the interaction of DCX with the FAM C-terminus, suggesting that loss of the FAM/DCX interaction may be a causative factor in the disorder (Friocourt et al., 2005). To determine if the DCX/FAM interaction also occurred in an *in vivo* setting, COS-7 cells were cotransfected with expression constructs for DCX and either a GFP-tagged carboxyl-terminal fragment of FAM (corresponding to the region identified in the yeast two-hybrid), full length FAM protein, or a fragment containing the catalytic core of FAM (amino acids 1210-2410). Cell lysates were then subjected to immunoprecipitation with antibodies directed against Dcx, FAM, or the FAM fragment. Both the carboxyl-terminal fragment and full length FAM proteins were co-immunoprecipitated with DCX, but not the catalytic core containing fragment (Friocourt et al., 2005). It was further determined that DCX was not ubiquitylated in the cell types and tissues examined suggesting that DXC is not a substrate of FAM (Friocourt et al., 2005). Endogenous FAM and DCX proteins were able to be reciprocally co-immunoprecipitated from newborn mouse brain extracts, and a proportion of FAM was found to co-sediment with DCX containing microtubule fractions (Friocourt et al., 2005). Taken together, this information suggests that while DCX may not be a substrate of FAM, the two proteins work together in areas of the mouse brain and at points within the cell where dynamic changes occur. It is possible that DCX is involved in correctly localising FAM to areas where its deubiquitylating activity is

required, or that FAM has other roles that require DCX to act as a cofactor (Friocourt et al., 2005).

AIMS AND HYPOTHESES

Hypotheses

1. That the amino- and carboxyl-terminal regions outside of the catalytic core of FAM contain domains or motifs that are involved in the interaction of FAM with other proteins, or modulate FAM function
2. That FAM is involved in regulating protein trafficking from the Golgi apparatus to the lysosome.

Specific Aims and Objectives

The first hypothesis is explored in Chapter 3 of this thesis. A variety of commonly used *in vitro* binding assays were utilised to attempt to identify regions of the FAM protein outside of the catalytic core that might be involved in protein-protein interactions with known substrates and novel interacting partners. Yeast and mammalian two-hybrid screens were used to identify novel binding sites on FAM for the known substrates β-catenin, itch, and epsin 1. This was done in the hope that regions outside of the essential catalytic core of the protein could be mutated in such a way that individual interactions of FAM with other proteins could be disrupted without a total disruption of FAM function. This would provide the tools to begin to dissect, on a molecular basis, the specific role of FAM's regulation or interaction with each binding partner in isolation. Co-immunoprecipitation, GST, and Nickel affinity chromatography methods were also used in Chapter 3, with the aim of identifying novel binding partners for FAM using the putative ubiquitin-like domain of FAM as bait.

Based on the results from Chapter 3, and given that FAM has been found to co-localise with a number of markers for both the lysosome and the TGN (Murray et al., 2004), it was postulated that FAM may have a functional role in protein trafficking from the Golgi apparatus. In Chapter 4 of this thesis, the role of FAM in the trafficking of proteins from the Golgi to the lysosome was explored. An ELISA based assay that measures both the cellular and secreted protein levels of eleven known lysosomal proteins simultaneously was used here. The aim of this experiment was to determine the effect of both a gain and a loss of FAM

function in HEK 293T cells, on the correct trafficking of these proteins to the lysosome. Western analysis was used, in conjunction with this assay, to determine the amount of processing that occurs on any of the 11 proteins that were found to have increased or reduced protein levels compared to untreated cells. As all eleven of the proteins that were analysed in the multiplex ELISA assay have been established as main players in a variety of lysosomal storage diseases, the trafficking pathways that they take and the sites within these pathways at which processing occurs, are well established. This allowed us to speculate as to the sites at which any blockade of the trafficking of these proteins may have occurred. In addition, immunofluorescence techniques were used to attempt to identify changes in the localisation of these proteins in cells where FAM levels were altered.

CHAPTER 2: MATERIALS AND METHODS

MATERIALS AND ANTIBODIES

Cell Culture Materials:

Foetal Calf Serum (FCS), Dulbecco's Modified Eagle Medium (DMEM), and Phosphate Buffered Saline (PBS) for tissue culture usage were obtained from CSL Biosciences (Parkville, NSW) and Gibco BRL (Invitrogen, Mount Waverley, VIC).

Antibodies:

Monoclonal anti-V5 antibody was obtained from Invitrogen. Monoclonal anti-His antibody was obtained from Sigma-Aldrich (Sydney, NSW). Monoclonal anti-Gal4 DBD and anti-Gal4 AD antibodies were obtained from Becton-Dickinson (BD) Biosciences (North Ryde, NSW). Polyclonal anti-β-catenin antibody was obtained from Sigma-Aldrich. Rabbit polyclonal anti-FAM UBLD and anti-GST antibodies were raised against the putative UBLD domain of FAM (amino acids 886-970) expressed as a GST fusion protein. Purification of these antibodies is discussed in Chapter 4 of this thesis. Sheep polyclonal anti-GGA1, anti-GGA2, anti-GGA3, anti-Lamp1, anti-ASA, anti-I2S, and anti-4S, and goat anti-epsin1and anti-Stam were all kind gifts from Associate Professor Douglas Brooks and Dr Peter Meikle (Lysosomal Diseases Research Unit, Chemical Pathology, Women's and Children's Hospital, SA). Monoclonal anti-Hrs antibody was also a kind gift from Associate Professor Douglas Brooks originally sourced from Alexis Biochemicals (Sapphire Bioscience Pty. Ltd, Redfern, NSW). Monoclonal anti-GGA3 was obtained from BD Biosciences. Polyclonal anti-N1 FAM antibodies were raised against a synthetic peptide corresponding to the first 20 amino acids of murine FAM (TATTRGSPVGGNDNQGQAPC), denoted N1. N1 peptide coupled to the carrier protein keyhole limpet haemocyanin (KLH) was synthesised by and obtained from Mimotopes Pty. Ltd. (Clayton, VIC). Two rabbits, Sw1 and Sw2, were immunised with the N1-KLH peptide and antibodies were purified from the serum on an N1 affinity column obtained from Mimotopes Pty. Ltd. according to the manufacturer's instructions. Goat anti-rabbit and rabbit anti-mouse secondary antibodies conjugated to horseradish peroxidase (HRP) were obtained from DAKO corporation (Carpinteria, USA). Donkey anti-sheep/goat secondary antibody conjugated to HRP was obtained from serotec (North Carolina, USA). Sheep anti-rabbit secondary antibody conjugated to Cy3 was obtained from Sigma-Aldrich. Donkey anti-sheep and goat anti-mouse secondary antibodies conjugated to Alexa-555 or Alexa-488 fluorochromes respectively, were obtained from Molecular Probes (Invitrogen).

PLASMIDS

Vector Construction:

Yeast Two-Hybrid:

Bait fusion vectors

Mammalian expression vectors producing Gal4 DNA binding domain (DBD) fusions for use as bait in yeast-two hybrid binding assays were constructed by Gateway cloning technology (Invitrogen). pDONR201 plasmids containing the FAM fragments designated N1, N2, C1, C2, and FAM Cat (USP9X fragments spanning amino acids 1-674, 669-1213, 1210-2100, 2097-2554, and 1476-1918 respectively) were previously cloned by P. Khut (Masters Graduate, University of Adelaide, SA). Briefly, PCR reactions that introduced attB sites onto the 5' and 3' ends of the open reading frame, were performed on pRSET (N1, N2, C2), pBluescript SK(-) (C1), or pGEX4T1 (FAM Cat) plasmids containing these fragments of USP9X (a kind gift from Prof. K. Kaibuchi, Nara Institute of Science and Technology, Japan) The gene fragments were then cloned into pDONR201 (Invitrogen) as per the manufacturers instructions. LR reactions were performed between the donor vectors pDONR201 – N1, N2, C1, C2, and FAM Cat, and the destination vector pDEST32.

Prey fusion vectors

Gal4 activation domain (AD) fusions for use as prey in yeast-two hybrid binding assays were constructed by Gateway LR reactions. The Gateway donor vectors pDONR201 – N1, N2, C1, C2, β-catenin, the FAM binding region (FBR) of β-catenin (amino acids 184-535) (Taya et al., 1999), rat epsin 1, the cytoplasmic domain of E-cadherin designated ECT (amino acids 737-884) ((Lickert et al., 2000), and itch were recombined with the Gateway destination vector pDEST22 using LR reactions as per the manufacturers instructions (Invitrogen). pDONR201 vector for murine β-catenin was constructed by performing PCR using the Gateway primers listed in Table 2.1, on total cDNA from embryonic day 11.5 mouse ES cells. cDNA was generated by reverse transcription of total RNA isolated from embryonic day 11.5 mouse ES cells using Omniscript reverse transcriptase (Quiagen). The pDONR201 vector for the FBR of β-catenin was previously constructed by P. Khut (Masters

Graduate, The University of Adelaide, SA). The pDONR201 vector for rat epsin 1 was constructed by performing PCR using the primers listed in Table 2.1, on pcDNA3.1-Xpress Epsin 1 (a kind gift from Dr. Pietro De Camilli MD, Howard Hughes Medical Institute and Department of Cell Biology, Yale University School of Medicine, New Haven, USA). The pDONR201 vector for the cytoplasmic domain of E-cadherin was constructed by performing PCR using the primers listed in Table 2.1, on pGEX4T2-ECT737-884 (a kind gift from Prof. Rolf Kemler, Max Planck Institute of Immunology, Frieburg, Germany). The pDONR201 vector for human itch was constructed by performing PCR using the primers listed in Table 2.1, on pFLAG-CMV-2-Itch (a kind gift from Dr. Peter McPherson, Montreal Neurological Institute and Hospital, Montreal, Canada). In all instances, primers were designed such that the cDNA contained the first ATG start codon but had no 3' stop codon to allow for the generation of expression plasmids with both N- and C- terminal tags.

Table 2.1: Primers used for cloning of pDONR201 vectors for prey cDNAs

cDNA	Forward Primer	Reverse Primer	Product Size (bp)
Murine β-catenin	5'_ggggacaagtttgtacaaaaaag caggctcaATGGCTACTCA AGCTGACC_3'	5'_ggggaccactttgtacaagaaag ctgggtACAGGTCAGTATC AAACCAGG_3'	2346
Rat Epsin 1	5'_ggggacaagtttgtacaaaaaag caggctctATGTCGACATC ATCGCTGC_3'	5'_ggggaccactttgtacaagaaag ctgggtTTAGGAGGAAGG GGTTAGTG_3'	1700
Human Itch	5'_ggggacaagtttgtacaaaaaag caggctctATGGGTAGCCT CACCATGAAATCACAG C_3'	5'_ggggaccactttgtacaagaaag ctgggtCAGAAGGATTTGG ACAAGAG_3'	2700
E-Cadherin Cytoplasmic Domain	5'_ggggacaagtttgtacaaaaaag caggctctCCCAGAACGGT GGTCAAAG_3'	5'_ggggaccactttgtacaagaaag ctgggtAGTCGTCCTCGCC ACCGC_3'	451

Lowercase letters correspond to Gateway attB sequences and uppercase letters correspond to gene-specific homology sequences.

Bacterial Expression Vectors

GST Fusion Expression Vectors

Gateway Conversion of pGEX4T2

pGEX4T2 GW (Gateway converted N-terminal GST fusion expression vector with thrombin cleavage site [Promega Corp., Madison, WI]) was generated by blunt cloning Gateway conversion cassette A (Invitrogen) into the Sma I site of the multiple cloning site of pGEX4T2.

FAM Domain Vectors

GST fusion bacterial expression vectors for seven putative FAM domains (discussed in Chapter 3) were created by Gateway LR reactions. pDONR201 vectors containing the putative USP9X domains designated FD1-7 were constructed by BP reactions between PCR products of the primers listed in Table 2.2 and pDONR201. LR reactions were performed between these donor vectors and pGEX4T2 GW.

FAM UBLD expression vector

A GST fusion expression vector for the putative UBL domain of FAM (discussed in Chapter 3) was created by a Gateway LR reaction. A pDONR201 vector containing the cDNA for the putative FAM UBL domain was generated by performing PCR on pDONR201 FAM (containing the full length cDNA for FAM) with forward (5'_ggggacaagtttgtacaaaaaagcaggctcctctcttttatagttcga_3') and reverse (5'_ggggaccactttgtacaagaaagctgggtggagcttatttgtgtaag) Gateway primers. A Gateway BP reaction was then performed between the PCR product and pDONR201 to generate pDONR201 FAM-UBLD. A Gateway LR reaction between pDONR201 FAM-UBLD and pDEST15 produced the FAM-UBLD/GST fusion expression vector.

GGA3 Domain Vectors

pGEX4T2 GST fusion expression vectors for GGA3 -VHS, -VHSGAT, -GAT, and -HingeGAE were kind gifts from Professor Kazuhisa Nakayama (Kyoto, Japan).

Table 2.2: Primers used for cloning of pDONR201 vectors for FAM domain cDNAs

cDNA	Forward Primer	Reverse Primer	Product Size
FAM Domain 1	5’_ggggacaagtttgtacaaaaaagcaggctctCGAGTATGACAGCCACGAC_3’	5’_ggggaccactttgtacaagaaagctgggtTAAAACGATCGTGCAGTATCTG_3’	723
FAM Domain 2	5’_ggggacaagtttgtacaaaaaagcaggctctATGTCGACATCATCGCTGC_3’	5’_ggggaccactttgtacaagaaagctgggCTTTCAAACGATCAAAGCAAGAC_3’	2487
FAM Domain 3	5’_ggggacaagtttgtacaaaaaagcaggctctACACATTGTGTGTTTTGGATG_3’	5’_ggggaccactttgtacaagaaagctgggtCTTCTGCCACAGCACGAAC_3’	1035
FAM Domain 4	5’_ggggacaagtttgtacaaaaaagcaggctctCTCCTATGACATCGGTCAACC_3'	5’_ggggaccactttgtacaagaaagctgggtTGGATGCGGGAAAGATGAAG_3'	867
FAM Domain 5	5’_ggggacaagtttgtacaaaaaagcaggctctCTGAAGCACTTACTGAGTGGG_3'	5’_ggggaccactttgtacaagaaagctgggtAAACAACATTACATGTCAGT_3'	1461
FAM Domain 6	5’_ggggacaagtttgtacaaaaaagcaggctctGGCAAGATCACCTGTCTC_3'	5’_ggggaccactttgtacaagaaagctgggtAAAAGCGAAGCAATTTGACGG_3'	831
FAM Domain 7	5’_ggggacaagtttgtacaaaaaagcaggctctACTCTTGGCAGACTCACAG_3'	5’_ggggaccactttgtacaagaaagctgggtACTGATCCTTGGTCTGAGGA_3'	621

Lowercase letters correspond to Gateway attB sequences and uppercase letters correspond to gene-specific homology sequences.

6XHis Fusion Expression Vector

A bacterial expression vector for 6 x His epitope tagged FAM UBLD was generated by a Gateway LR reaction between pDONR201 FAM-UBLD and pDEST17 (N-terminal 6 x His fusion expression vector – Invitrogen).

Mammalian Two-Hybrid Vectors

Gateway Conversion of Mammalian Two-Hybrid Vectors

Mammalian expression vectors for Gal4 DBD (pEFBOS GAL4 DBD) or VP16 AD (pEFBOS VP16 AD) with expression driven from an elongation factor 1 α promoter (EF-1α) were a kind gift from Dr Dan Peet (University of Adelaide, SA). pEFBOS GAL4 DBD and pEFBOS VP16 AD were converted to Gateway vectors by blunt cloning of the Gateway conversion cassette B (Invitrogen) into the EcoRV site located in the multiple cloning site of both vectors.

Bait Fusion Vectors

Mammalian two-hybrid bait vectors were generated by Gateway LR reactions between pDONR201 -β-catenin, -β-catenin FAM binding region, or AF-6 FAM binding region and pEFBOS GAL4 DBD GW.

Prey Fusion Vectors

Mammalian two-hybrid prey vectors were generated by Gateway LR reactions between pDONR201 –FAM Cat (1553-1954) or EGFP and pEFBOS VP16 GW.

Mammalian Expression Vectors

pDEST40 base vector, pDEST40 FAM, and pDONR221 FAM were kind gifts from Tsunaeki Asai (Tokai University, Japan). pDONR221 and pDEST40 were originally sourced from Invitrogen. pDEST40 FAMΔUBLD was generated in a seven step process. Firstly two PCR reactions were performed. The first PCR reaction used the 5' Gateway cloning primer for the FAM N2 fragment in combination with a 3' primer (5'_GCGGATCC_BamHI_gtgtttaccacggaaagctcttg_3') that begins with a BamHI restriction site followed by a short region of FAM homology starting at the codon (base pairs 2662-4 coding for amino acid 885) immediately preceding the first codon for FAM UBLD (base pairs 2665-7 coding for amino acid 886). The second PCR utilised the 3' Gateway cloning primer for the FAM N2 fragment in combination with a 5' primer (5'_CGGGATCC_BamHI_aatatgccttcaaagtcctgatagc_3') that also begins with a BamHI

restriction site followed by a short region of FAM homology starting at the codon (base pairs2911-3 coding for amino acid 971) immediately proceeding the final codon for FAM UBLD (base pairs 2908-10 coding for amino acid 970). PCR reactions were performed using pDONR201-FAM N2 as a template. Both PCRs were then digested with BamHI and a sticky ligation performed to produce a 1.1 Kb fragment consisting of the FAM N2 fragment minus the coding region for FAM UBLD (which is replaced by the reformed BamHI restriction site). A Gateway BP reaction was then performed between the ligation product and pDONR201 to generate pDONR201 FAM N2ΔUBLD. pDONR201 FAM N2ΔUBLD was then digested with SpeI and BanI and the UBLD deleted fragment "sticky, sticky" ligated into the corresponding SpeI and BanI sites in pDONR221 FAM. A Gateway LR reaction was then performed between pDONR221 FAMΔUBLD and pDEST40 to generate pDEST40 FAMΔUBLD. pDEST51 C1566S FAM was created by M. Scherer (PhD candidate, University of Adelaide, SA). pcDNA3.1 6xHis GGA1, GGA2, and GGA3 were kind gifts from Professor Kazuhisa Nakayama (Kyoto, Japan).

Other Vectors:

FAM shRNA Knockdown Vectors

pRetro Super, pRetro Super USP9X 1800, pRetro Super USP9X 3601, and pRetro Super USP9X were kind gifts from René Bernards (The Netherlands Cancer Institute, Division of Molecular Carcinogenesis, The Netherlands).

Target Sequences on FAM mRNA for shRNAs

pRetro Super USP9X 1800:

bp 1800 – bp 1818: GGAAAUGCUUAGCUGAGAA

pRetro Super USP9X 3601:

bp 3601 – bp 3619: CGAACAGGTTTGTGAA

pRetro Super USP9X:

bp 4247 – bp 4265: GAACAGGAGAAACGGGTAT

Other Vectors

pG51EBLuc (mammalian two-hybrid reporter vector) was a kind gift from Dr Dan Peet (University of Adelaide, SA). pRL TK (renilla luciferase reporter vector for transfection efficiency control) was a kind gift from Dr Murray Whitelaw (University of Adelaide, SA)

Plasmid Preparation:

Plasmid DNA minipreps were performed using the alkaline lysis protocol described in "Molecular Cloning: A Laboratory Manual" (Sambrook and Russell, 2001). Plasmid DNA midi- and maxi-preps were performed using Quantum prep plasmid midi- and maxi-prep kits from Biorad (Regents Park, NSW).

YEAST STRAINS:

The yeast strain that was used for all transformations for the yeast two hybrid was the MAV203 strain (Invitrogen).

MAV203 Genotype:

MAV203: MAT α; leu2-3,112; trp1-901; his3 Δ200; ade2-101; cyh2R; can1R; gal4Δ; gal80Δ; GAL1::lacZ; HIS3UASGAL1::HIS3@LYS2; SPAL10 UASGAL1::URA3

Control Strains:

Five control strains were supplied with the ProQuest Two-Hybrid System with Gateway Technology (Invitrogen). The strains have the same genotype as MAV203 except that they are MAT **a**, and contain plasmids for the expression of cDNA inserts that can be used to demonstrate a range of interaction strengths from weak to strong (Chevray and Nathans, 1992; Du et al., 1996; Fields and Song, 1989; Vidal, 1997; Vidal et al., 1996; Vidal et al., 1996; Walhout et al., 2000).

BACTERIAL STRAINS:

The following baterial strains were used:

Cloning:

DH5α: supE44Δlac; U169 (ϕ80 lacZ ΔM15); hsdR17; recA 1; endA 1; gyrA96; thi 1; relA1
DB3.1: F-; gyrA462; endA1 Δ(sr1-recA); mcrB; mrr; hsdS20(rB-, mB-); supE44; ara-14;

Protein Expression:

BL21 (DE3): F-; ompT; hsdSB (rB-mB-); gal; dem; lon-; galK2; lacY1; proA2; rpsL20(SmR); xyl-5; λ-; leu mtl1
Rosetta (DE3): F– *ompT hsdS*$_B$(r_B^- m_B^-) *galdcm lacY1* (DE3) pRARE (*argU, argW, ileX, glyT, leuW, proL*) (Cm^r)

YEAST TRANSFORMATION AND GROWTH:

Preparation and transformation of Competent Yeast:

Competent MAV203 strain yeast were prepared and transformed as described in the protocol found on the Invitrogen website: https://catalog.invitrogen.com/index.cfm?fuseaction=iProtocol.unitSectionTree&objectId=8CA5CFA900950165C60D4B3199153CE8&treeNodeId=5EA00A1CE5E6CAD2B3074A1CB3B93A45

Growth and selection of transformants:

Transformed yeast strains were selected on SC –Leu –Trp plates. Cells were grown in an incubator at 25°C for the periods described in the Proquest Two-Hybrid System manual (Invitrogen). 50 mL liquid cultures for western analysis were grown in SC –Leu –Trp liquid media at 25°C with shaking.

BACTERIAL PROTEIN INDUCTION:

The *E. coli* BL21 (DE3) cell line was used for expression of most GST-fusion constructs unless otherwise stated. Small-scale studies were performed initially to determine optimum conditions for induction and solubility of proteins. 5mL cultures (LB, 100 μg/ml Ampicillin) were grown to an OD_{600} of 0.5 - 0.8 at 37°C prior to induction with [0.1 mM] iso-propylthio-β-D-galactopyranoside for 2.5 – 3 hrs. Cells were lysed by sonication (4 x 15 secs) in 200 μL of tris-tween-20 buffered saline (TTBS 0.05% Tween 20, 20 mM Tris pH 7.6, 150 mM NaCl). phenylmethylsulphonyl fluoride PMSF was added before each sonication and at the end of the last sonication resulting in a final PMSF concentration of 1 mM. Lysate was pelleted at 10,000 rpm for 10 minutes at 4°C and the supernatant collected (soluble fraction). The pellet was resuspended in 200 μL of TTBS. Uninduced, induced, insoluble and soluble samples were analysed by SDS PAGE. Test inductions were also carried out at 30°C and room temperature following the procedure outlined above excepting that incubation times following induction were increased (3.5 – 5 hrs at 30°C and overnight at room temperature ~25°C). Large scale inductions involved growth of 400ml cultures to an OD_{600} of 0.7-0.8 at 25°C followed by collection of a 2mL uninduced sample and induction with IPTG to a final concentration of 0.1mM. Cultures were incubated for a further 1-3hrs, a 2mL induced sample taken and the remaining culture pelleted at 4200 rpm for 20 mins. Pellets were resuspended in 40mLs of TTBS and lysed by sonication (3 x 30 second bursts) following addition of PMSF (final concentration of 1 mM). Soluble and insoluble fractions were separated and equal volumes uninduced, induced, soluble and insoluble fractions were analysed by SDS PAGE and Coomassie brilliant blue staining.

TISSUE CULTURE TECHNIQUES:

Cell Culture and Transfection of HEK 293T Cells

Human embryonic kidney (HEK) 293T cells were maintained in a 37°C, 5% CO_2, humidified incubator in Dulbecco's modified Eagle's medium (DMEM) with 10% (vol/vol) heat-inactivated foetal calf serum (FCS), penicillin (100U/mL), and streptomycin (100μg/mL). HEK 293T cells were fed every three days and passaged at 1:10 every four days or once cells had reached ~ 80% confluency. Transient transfection of these cells was achieved either with Lipofectamine2000 according to the manufacturer's instructions (Invitrogen), or by $CaPO_4$ transfection. $CaPO_4$ transfection was carried out according to the

protocol outlined in Cells: A laboratory manual (Spector et al., 1998). Briefly, 5 x 10^5 cells were plated on 10 cm tissue culture dishes to achieve a confluence of ~20% prior to transfection on the following day. 10-30 μg of DNA was diluted in MilliQ water to a volume of 450 μL. 50 μL of $CaPO_4$ solution I (2.5 M $CaCl_2$) was added to this followed by 500 μL of $CaPO_4$ solution II (50 mM BES (Sigma-Aldrich), 280 mM NaCl, 1.5 mM Na_2HPO_4, pH 6.95). The solutions were gently mixed and then incubated for 10-20 minutes before being added drop-wise into the culture with gentle swirling. The cells were then returned to the incubator and cultured overnight. 16-24 hours post transfection the media was aspirated and the cells washed very gently with PBS to remove precipitates. The cells were then re-fed and returned to the incubator for the periods indicated in figure legends. In experiments where fewer cells were required (eg. 6 well trays for immunofluorescence and 24 well trays for luciferase experiments), the number of cells, quantity of DNA, and the amounts of each solution were scaled accordingly. For the co-transfections described in Chapter 4, sorted cell multi-plex ELISA, ratios of each vector for each transfection (pDEST40 + pDEST51-V5-GFP or pDEST40-V5-FAM + pDEST51-V5-GFP) were normalized such that equimolar amounts of pDEST40 and pDEST40-V5-FAM were transfected.

Cell Culture and Transfection of COS-1 Cells

African green monkey kidney fibroblast cells (COS-1) were maintained in a 37°C, 5% CO_2, humidified incubator in DMEM with 10% (vol/vol) heat inactivated FCS, penicillin (100u/mL), and streptomycin (100μg/mL). Cells were refed every 2-3 days and maintained by passaging them 1:10 when they reached ~80% confluency. Transient transfection of these cells was achieved with Lipofectamine2000 according to the manufacturer's instructions.

DUAL LUCIFERASE ASSAYS:

Luciferase assays were performed using Promega's dual luciferase assay kit (Annandale, NSW, Australia). HEK 293T cells were grown in 24 well plates, transfected with appropriate plasmids and cultured for 72 hours. Note: cells were cultured for 72 hours post transfection as the $CaPO_4$ method of transfection required that the media be removed and the cells be washed 12-24 hours after addition of the transfection reagents. As HEK 293T cells are only loosely adherent, the cells needed to be grown for 72 hours in order to obtain enough material to use in the luciferase assays. Media was aspirated and cells were washed once with cold PBS. 100μl of passive lysis buffer was added to each well and incubated at room

temperature with shaking for 15 mins. 25µl of luciferase assay reagent II (LARII) was aliquoted into appropriately labelled 1.5ml microcentrofuge tubes and 1-2µl of each sample added. A reading of the luminescence produced by the firefly luciferase was immediately taken using a Turner Designs TD20/20 luminometer. Following this, 25µl of stop and glo reagent was added and a second luminescence reading, produced by the *Renilla* luciferase enzyme was taken. Calculation of the ratio of Firefly:*Renilla* luciferase activity normalises for transfection efficiency and returns a fold induction of luciferase activity.

PROTEIN ELECTROPHORESIS AND WESTERN BLOTTING:

Cells were harvested using a rubber policeman and were resuspended in an appropriate amount of mammalian lysis buffer (50 mM Tris pH7.5, 1mM EDTA, 50 mM NaCl, 0.5% [vol/vol] Triton X-100) containing a "Complete Mini" protease inhibitor cocktail tablet (1 per 10 mL – Roche Applied Sciences, Castle Hill, NSW, Australia). Cells were washed once in cold PBS and cell suspension pelleted in a refrigerated centrifuge at 4˚C for 1 minute (1200 rpm). Cells were lysed by sonication (3 x 10 second bursts) and Triton X-100 insoluble material removed by centrifugation at 4˚C for 30 minutes (16,000 rcf). The concentration of protein content in lysates was determined using a bicinchonic acid (BCA) protein assay kit (Pierce, Rockford, USA).

Alternatively, protein samples for use in the multiplex ELISA were obtained by harvesting and washing cells as described above. Cells were resuspended in 20 mM Tris, 250 mM NaCl, pH 7.0 buffer. The cells were then subjected to 6 x freeze (in liquid nitrogen) thaw (in room temperature water) cycles before undergoing sonication (3 x 10 second bursts). BCA protein assays were then performed to determine total protein content.

Protein samples and RPN800 (GE Healthcare) or Precision Plus Protein Kaleidoscope Standards (Biorad) molecular weight size markers were subjected to SDS PAGE on 7.5%, 10%, 12.5%, or 15% polyacrylamide resolving gels. Polyacrylamide gels were poured and electrophoresed using the Miniprotean III system (Biorad). Electrophoretic transfer of proteins to nitrocellulose membrane (Biotrace NT, Pall Corporation, N.Y., USA) was carried out using a Mini Trans-Blot Cell (Biorad) according to the manufacturer's instructions. Membranes were immunoblotted with monoclonal or polyclonal antibodies using antibody concentrations as directed by the manufacturer (for commercial antibodies) or at 1-5 µg/mL for in house purified polyclonal antibodies. Primary antibodies were detected using

appropriate HRP conjugated secondary antibodies and ECL reagent (amersham pharmacia, UK). Chemiluminescence was detected with SuperRX Fuji Medical X-ray Film (Fuji Photo Film, Tokyo, Japan).

IMMUNOPRECIPITATION:

HEK 293T cells were seeded into four to ten 10 cm tissue culture dishes per immunoprecipitation and grown to 70-80% confluence. Cells were harvested and lysed as described for western blotting. Lysate from each group of 4-10 flasks was pooled, cleared collectively and incubated with 30-50 μL of protein A agarose at 4ºC for 1 hour with shaking in 1.5ml microcentrifuge tubes. Antibodies used (10 μg of α-FAM or 10 μg of rabbbit pre-immune serum) were simultaneously incubated with 30-50 μL of protein A agarose at 4ºC for 1 hour with shaking in 1.5ml microcentrifuge tubes. Tubes containing antibodies were spun down at 4ºC for ~30s and supernatant was removed. The tube containing the pooled lysate was spun down and equal amounts of cleared lysate were transferred to each antibody containing tube. These were incubated for 1 hour with shaking at 4ºC. Beads were pelleted at 4ºC (~30s at 16,000 rcf) and the supernatant was removed. Pelleted beads were resuspended in 1ml of lysis buffer A and incubated at 4ºC for 5 mins. This was repeated five times for a total of six repeats followed by one wash with 1ml of ice cold PBS. The beads were pelleted once more and the supernatant was completely removed. Beads were resuspended in 25 μL of 1x protein sample buffer (0.2% w/v SDS, 10% v/v glycerol, 0.05% bromophenol blue w/v, 25 mM Tris pH 6.8, and 2% v/v β-mercaptoethanol) and boiled for 5 mins at 95ºC. Tubes were spun down and the supernatant was analysed by SDS PAGE and western blot analysis with appropriate antibodies.

ANTIBODY PURIFICATION PROCEDURES:

Protein Immobilisation

GST and GST-FAM-UBLD antigen columns were generated by covalently linking the purified expressed proteins to cyanogen bromide (CNBr) activated sepharose (GE healthcare, Buckinghamshire, England). Crosslinking was achieved as described by the manufacturer (GE healthcare) with all protein amounts and reaction volumes scaled for a bed volume of 1 mL of activated sepharose.

Purification

GST Antibody

A GST antigen column was prepared by washing with ten column volumes of MilliQ water at a flow rate of 1 mL per minute. The antigen column was pre-equilibrated with 10 column volumes of 1 x PBS at a flow rate of 1 mL per minute before the anti-GST antibody cleared serum was loaded at 0.5 mL per minute. The column was then washed with 10 column volumes of 1 x PBS at 1 mL per minute. Antibody was eluted with 0.1 M Phosphate buffer pH2.5 (orthophosphoric acid) in 1 mL fractions into 100 µL of 1.0 M Phosphate buffer, pH 9.0 at 1 mL per minute. Fractions were analysed by UV absorbance at 280 nm (A_{280}) to determine which fractions contained protein and the relevant fractions were pooled. To determine the quantity of antibody in the fractions the following equation was used: $A_{280}/1.35$ = IgG mg/ml. Pooled antibody fractions were dialysed against 1 x PBS overnight at 4°C with stirring and the buffer exchanged antibody sterile filtered (0.2 µm). The antibody was stored at 4°C with 0.1% sodium azide. The antigen column was washed with 10 column volumes of MilliQ water and stored in 20% ethanol at 4°C.

FAM-UBLD Antibody

Anti-FAM-UBLD antibody serum was filtered through a 0.2 µM filter. Anti-GST antibodies were removed from the serum by passing the serum through a column containing covalently immobilised GST five to six times (washing the column each time to remove the bound anti-GST antibodies). The GST-FAM-UBLD antigen column was then prepared by washing with ten column volumes of MilliQ water at a flow rate of 1 mL per minute. Anti-FAM-UBLD antibodies were then purified as described for the anti-GST antibodies.

GST-AFFINITY CHROMATOGRAPHY:

Purification of GST-Tagged Proteins on Glutathione Sepharose

GST fusion proteins were purified on Glutathione Sepharose 4B (GE Healthcare). Clarified lysate from bacterial protein inductions was incubated with 200 µL bed volume of glutathione sepharose in Poly-Prep columns (Biorad) for one hour at room temperature or

overnight at 4°C with shaking. Columns were allowed to settle by gravity, the supernatant drained, and a 100 µL aliquot saved for analysis. Columns were washed with 3 x 10 mL of binding buffer (10 mM Tris pH 6.8, 0.5% vol/vol Triton X-100, and 100 mM NaCl) before being eluted with 10 mM glutathione in 50 mM Tris pH 8.0. Glutathione was removed from the purified proteins by extensive buffer exchange with 50 mM Tris using Microsep Centrifugal Concentration Devices Pall Corporation (Lane Cove, NSW). Purified proteins were analysed by SDS PAGE with Coomassie Brilliant Blue staining and western blotting, and protein concentrations were estimated by BCA method combined with visual inspection of band intensity on Coomassie stained gels compared to BSA standards.

GST-Affinity Chromatography:

Column Method:

Purified fusion proteins (amounts are specified in the relevant figures) were incubated with 200 µL bed volume of glutathione sepharose for one hour at room temperature with rotation in 2 mL of binding buffer. Matrix beds were settled by gravity and supernatant was drained off. Total cellular or tissue lysate or purified 6 x His tagged FAM UBLD or Ubiquitin (amounts are specified in the relevant figures) was incubated with immobilised GST fusion protein for two hours at 4°C with rotation in 2 mL of binding buffer. Matrix beds were settled and supernatant was drained by gravity. Matrix was washed 2 x with 10 bed volumes of binding buffer, 1 x with 10 bed volumes of medium stringency wash buffer (10 mM Tris, 0.5% Triton X-100, 200mM NaCl), followed by one more wash with 10 bed volumes of binding buffer. Bound proteins were eluted with 10 mM Glutathione in 50 mM Tris pH 8.0, buffer exchanged, and analysed by SDS PAGE with Coomassie staining or western blotting.

PURIFICATION OF HIS-TAGGED PROTEINS ON A NICKEL COLUMN:

Nickel affinity purification of 6 x His FAM UBLD or Ubiquitin was achieved using Ni-NTA agarose (Invitrogen) according to the native purification protocol described in the product manual.

MICROSCOPY:

Indirect Immunofluorescence

HEK 293T or COS-1 cells were seeded onto 10 cm tissue culture dishes at cell densities that gave a confluence of ~20% on the following day prior to transfection (~5 x 10^5 cells). Cells were transfected by $CaPO_4$ (HEK 293T) or Lipofectamine2000 (COS-1) transfection reagents with the plasmid vectors described in each figure and incubated for 24 hours. Cells were then passaged onto glass coverslips in 6 well plates for indicated periods of time. Cells were fixed in 4% paraformaldehyde in PBS for 15 minutes at room temperature and permeabalised using 0.1% Triton-X in PBS for 5 minutes at room temperature. Blocking was performed using 1% BSA in PBS for 15 minutes at room temperature. Primary antibodies were diluted in 1% BSA/PBS as per the manufacturer's instruction and incubated with coverslips (one at a time for dual labelling) for 1 hour. Coverslips were washed 6 times with PBS before being incubated for 1 hour with secondary antibodies diluted in 1% BSA/PBS as per the manufacturer's instructions. Coverslips were washed 6 times with PBS. Nuclear staining was visualised using DAPI which was diluted to 10 μg/mL and incubated with coverslips for 5 minutes. Coverslips were mounted onto slides using fluorescent mounting medium (DAKO corporation) and visualised with a fluorescence microscope (Leitz model DMRB, Leica, North Ryde, NSW, Australia). Alexa Fluor488 and EGFP fluorescence were visualised using filtered light with a wavelength of 490 nm. Alexa Fluor555 and Cy3 fluorescence were visualised using filtered light with a wavelength of ~520 nm. DAPI nuclear staining was visualised using filtered light with a wavelength of ~360 nm.

MULTIPLEX ELISA ANALYSIS:

Multiplex ELISA analysis was performed by Mrs. Debbie Lang according to the protocol described in "Newborn screening for lysosomal storage disorders" (Meikle et al., 2006). Briefly, affinity purified sheep polyclonal antibodies raised against α-glucosidase, β-glucosidase, α-galactosidase, acid sphingomyelinase, Lamp-1, sulfamidase, iduronate-2-sulfatase, α-iduronidase, *N*-acetylgalactosamine-4-sulfatase, arylsulfatase A, and a monoclonal antibody raised against saposin-C were conjugated to carboxyl beads (Bio-Rad, Hercules, CA, USA) according to the manufacturer's instructions (Meikle et al., 2006). Conjugation of antibodies to create capture beads was performed with antibody concentrations of 9 μg/2.5 x 10^6 beads for all antibodies with the exception of anti-sulfamidase (5 μg/2.5 x 10^6 beads) and anti-acid sphingomyelinase and anti-α-iduronidase

(36 μg/2.5 x 10^6 beads) (Meikle et al., 2006). Each antibody was conjugated to a bead with a unique fluorescence intensity readout.

Protein quantitation of cell lysates from experimental and control samples was performed using the BCA method and values used to normalise the output from the Bio-Plex Protein Array System (Bio-Rad, Hercules, CA, USA). 12 μL (or 3 μL in cases where small amounts of lysate were available) of cell lysate was incubated with a mixture of antibody conjugated beads (5000 beads of each antibody for a total of 55000 beads per well) in 96 well multiscreen filtration plates for 16 hours at 4°C. Biotinylated reporter antibodies at 32 μg/L (except for anti-saposin C and anti-sulfamidase which were at 8 μg/L and anti-Lamp-1, anti-α-glucosidase, anti-α-iduronidase, and anti-iduronate-2-sulphatase which were at 16 μg/L) were simultaneously incubated with the lysate and beads. Beads were then washed by filtration, resuspended, and incubated (10 min, 20°C) in assay buffer (125 μL/well) containing streptavidin phycoerythrin conjugate (1.2 mg/L) (Molecular Probes, Eugene, OR, USA) (Meikle et al., 2006). Fluorescence intensity of the beads was then determined using a Bio-Plex Protein Array System with Bio-Plex Manager 3.1 software (Bio-Rad, Hercules, CA, USA) (Meikle et al., 2006). A minimum of 100 beads/well from each set were read and the fluorescence signal intensity translated into an estimate of the concentration of each analyte by reference to a corresponding calibration curve (Meikle et al., 2006).

SOLUTIONS:

Yeast Media:

YPAD

Yeast extract/Peptone/Adenine/Dextrose (YPAD) medium was prepared by adding 0.003% adenine hemisulfate to BD Biosciences Clontech YPD medium powder. Agar was added for solid media. Just prior to use, glucose was added (from a 40% w/v stock) to 2% v/v (glucose is not autoclavable).

SC and SC Dropout

SC – Leu (Synthetic Complete)

Yeast nitrogen base (Difco BRL, Detroit, MI) was mixed with SC – Leu – Trp dropout supplement (BD Biosciences Clontech) according to the manufacturers instructions. L-Tryptophan (Sigma) was added to a final concentration of 640μM (from a 40mM stock). Agar was added for solid media and glucose was added just prior to use to a final concentration of 2% v/v.

SC – Leu – Trp

SC – Leu – Trp was prepared as described for SC – Leu except that tryptophan was not added.

SC – Leu – Trp – His + 3AT

Yeast nitrogen base (Difco BRL, Detroit, MI) was mixed with SC –Leu – Trp – His dropout supplement and agar. Agar media was melted and 3-Amino-1,2,4-Triazole (3AT) was added from a 1M stock to final concentrations of 10, 25, 50, 75,or 100mM. 2% glucose was added just prior to use.

SC – Leu – Trp – Ura

SC – Leu – Trp – Ura was prepared as described for SC – Leu – Trp except SC – Leu – Trp – Ura dropout supplement was added instead of SC – Leu – Trp dropout supplement.

Other yeast solutions:

10 X TE buffer: 0.1M Tris, 10mM EDTA, pH 7.5

10 X LiAc: 1M Lithium Acetate (Sigma), pH 7.5 (with dilute acetic acid)

50% PEG 3350: 50% w/v PEG 3350 (Sigma)PEG/LiAc Solution: 1mL 10 X TE, 1mL 10 X LiAc, 40% PEG 3350

Z Buffer: 116.7mM Na2HPO4•7H2O, 40mM NaH2PO4•H2O, 10mM KCL, 1mM MgSO4•7H2O

X-Gal Stock: 20mg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal Sigma) in DMF

Z Buffer/X-Gal: 100mL Z buffer, 0.27mL β-mercaptoethanol (Sigma), 1.67mL X-Gal