INTRODUCTION

1.1. Background

The cell is a masterpiece of nature. It carries a bewildering variety of functions inside its tiny space, using a vast variety of biomolecules, and remains as an independent entity representing the single cellular organisms, while serving as the building block of the multicellular organisms. One of the functions that cell carries out is the degradation of proteins it no longer needs, so that they would not interfere with other activities and the amino acids released from degradation can be recycled to synthesize other proteins (Varshavsky, 1998). This also leads to the efficient use of limited resources like amino acids under starvation. This process of selective and controlled degradation is just as important for the proper functioning of the cell as constructive activities like protein synthesis and DNA replication are. This is amply demonstrated in the form of the diseases that result from improper or inefficient degradation of proteins, which include Alzheimer's disease and Parkinson's disease.

The molecular system that is chiefly responsible for selectively degrading proteins in the cell is known as the ubiquitin proteasome system, that plays the central role in the process of protein degradation. Ubiquitin is a small, 76 amino acid long, globular, single domain protein molecule that gets covalently attached to proteins which are to be degraded (Ciechanover et al., 1978; Hershko et al., 1980; Yadav et al., 2017). In doing so, it marksthose proteins as substrates for degradation. Proteasome is a large complex of multiple proteins. It recognises and degrades the proteins to which ubiquitin is attached. Proteasome was first discovered in 1984 and has been studied by several laboratories since then (Hershko et al., 1984 Baumeister et al., 1998; Lee and Goldberg, 1998; Recksteiner, 1998; Groll and Huber, 2004; Pickart and Cohen, 2004; Wolf and Hilt, 2004; Rechsteiner and Hill, 2005). An introduction to the structures and functioning of ubiquitin, proteasome and other associated proteins is presented here.

Ubiquitin is used to tag proteins for degradation, by a process called ubiquitination. Usually, proteins are ubiquitinated when they are unfolded or misfolded, either due to errors during protein synthesis or due to high temperature or some other form of stress. Proteins that might be intact and functional, but are no longer needed by the cell, are also ubiquitinated. Ubiquitin has structural features that make it an ideal molecule for use as a tag for degradation of proteins. These include its small size and single domain nature. Ubiquitin's core is very hydrophobic (Haririnia et al., 2008) and its surface residues form salt bridges with each other

(Makhatadze et al., 2003; Bhattacharya and Ainavarapu, 2008). Both these features make ubiquitin highly thermostable. This is significant, because one of the stress conditions in which proteins get unfolded and have to be degraded is high temperature. And ubiquitin has to be intact at high temperature for tagging unfolded proteins for degradation.

1.2. Use of Small Proteins to Understand Protein Structure and Folding

Small, single domain proteins are useful for protein structure and folding studies as they maintain stable structures and undergo cooperative and reversible folding (Went et al., 2004). The simple nature of their folding reactions makes interpretation of the results easy, making these small proteins ideal models to study protein structure and folding. For example, the simplicity of folding helps to prevent off-pathway reactions like aggregation (Judith, 2001; Baldwin and Rose, 1999). Nonetheless, protein folding studies are typically carried out under conditions that minimize the off-pathway reactions. These conditions include high dilution and low temperature. Ubiquitin is one such small single domain protein, which has a molecular weight of just 8.5 kDa and consists of just 76 residues, which is why it has also been used as an ideal model to study protein structure and folding (Recksteiner, 1998; Monia et al., 1990).

1.3. Structure of Ubiquitin

Ubiquitin was first discovered in 1975, and the efforts to unravel its structure go back to the 1980s. In 1985, structure of ubiquitin was solved at 2.8 Å resolution using X ray crystallography (Vijay-Kumar et al., 1985). This was followed in 1987 by a 1.8 Å resolution structure, also by X ray crystallography (Vijay-Kumar et al., 1987). The studies showed that ubiquitin has a pronounced hydrophobic core, consisting of valine, leucine, isoleucine, and methionine residues. The secondary structural features of ubiquitin include three and a half turns of α helix, a short 3_{10} helix, and a mixed β sheet composed of five strands and seven reverse turns (Fig. 1.1). Histidine, tyrosine and two phenylalanine residues were also found on

the surface of the molecule. Overall, the structure of ubiquitin was found to be very compact (Fig. 1.3), with large number of hydrogen bonds imparting structural stability to it. Around 87% of the polypeptide chain exists in the form of hydrogen bonded secondary structure. There are also some unusual structural features in ubiquitin, like the two reverse Asx turns, a parallel G1 β bulge, and a symmetrical hydrogen bonding region formed by two helices and two reverse turns.

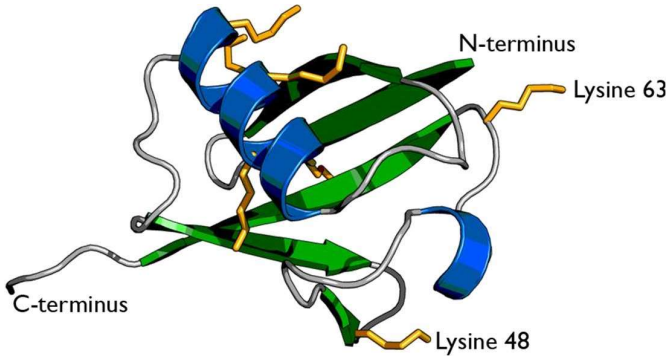

Figure 1.1: Tertiary structure of ubiquitin, showing α helices (green) and β sheets (blue). Image courtesy: Rogerdodd - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=4771409

Ubiquitin is a highly conserved protein, with just three residues differing between yeast and human ubiquitin polypeptides (Fig. 1.2) (Gavialnes et al., 1975; Watson et al., 1978; Schlesinger et al., 1975; Schlesinger and Goldsteiner, 1975; Wilkinson et al., 1986; Vierstra et al., 1986). This means that most residues are essential to either ubiquitin's structure or function. While the residues in the core of ubiquitin might play structural role only, many of the residues on the surface of the molecule play important roles in functional aspects too, like providing binding sites for other proteins. To study the role of surface residues in ubiquitin's structure, they were replaced with alanine one by one and the resultant proteins were characterized (Sloper-Mould et al., 2001). The study found that the surface residues essential for vegetative growth in yeast are clustered in three regions. One of these regions includes L8, I44, V70 and

their surrounding residues, and is essential for proteasomal degradation and endocytosis. The second region consists of F4 and its surrounding residues, and is also essential for endocytosis but not for proteasomal degradation. The third region consists of the C terminal tail. This region is essential for most of the functions of ubiquitin, owing to its role in polyubiquitin chain formation.

YEAST UBIQUITIN PROTEIN SEQUENCE

```
MQIFVKTLTG  KTITLEVESS  DTIDNVKSKI  QDKEGIPPDQ  QRLIFAGKQL
1          10          20          30          40          50
EDGRTLSDYN  IQKESTLHLV  LRLRGG
51         60          70          76
```

HUMAN UBIQUITIN PROTEIN SEQUENCE

```
MQIFVKTLTG  KTITLEVEPS  DTIENVKAKI  QDKEGIPPDQ  QRLIFAGKQL
1          10          20          30          40          50
EDGRTLSDYN  IQKESTLHLV  LRLRGG
51         60          70          76
```

Figure 1.2: Comparison of amino acid sequence of ubiquitin between yeast and humans. Residues that vary are shown in red.

Studies were done on the role of surface salt bridges in the structural stability of ubiquitin at pH 2.0. Salt bridges are found on ubiquitin surface between positively charged lysine residues and negatively charged glutamate and aspartate residues. When the lysines were replaced with neutral residues like glutamine and threonine, the salt bridges broke down, affecting structural stability as observed by NMR studies (Sundd et. al., 2002). It was also observed that various salts increase thermostability of ubiquitin due to anion binding, as measured by differential scanning calorimetry (Ibarra-Molero et al., 1999).

Based on these studies, a variant of ubiquitin has been engineered that is more structurally stable than the naturally occurring ubiquitin (Loladze and Makhatadze, 2002). This was done by first replacing all the arginine residues on ubiquitin's surface by lysine residues, and then carbomylation of the amino groups of the lysine residues. This study indicated that surface charge-charge interactions might not always be essential for structural stability of a protein. These observations helped frame the guidelines for engineering of surface charges to increase structural stability of proteins (Makhatadze et al., 2003).

There have also been studies on the importance of hydrophobic residues in the core of ubiquitin (Loladze et al., 2001; Loladze et. al., 2002). These studies involved replacing

different residues with site directed mutagenesis, and measuring the consequent changes in heat capacity using differential scanning calorimetry. Substitution of nonpolar residues in the core with polar residues decreased its structural stability. When naturally occurring polar residues in the core were replaced with nonpolar residues, it increased structural stability of ubiquitin. And when nonpolar residues in core were replaced with other nonpolar residues, it did not affect structural stability in any significant way. Another study investigated the importance of I30 and I36 hydrophobic residues found at the C terminus of the α helix of ubiquitin (Thomas et al., 2000). None of the 16 variants produced in the study were found to be as stable as wild type ubiquitin.

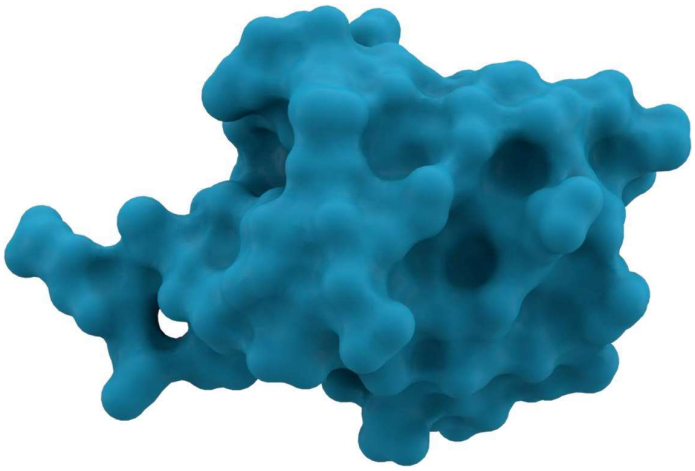

Figure 1.3: Surface representation of ubiquitin. Image courtesy: Thomas Splettstoesser (www.scistyle.com) - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=33687983

1.4. C Terminal G1 β Bulge of Ubiquitin

β-bulge is a structural feature found in some proteins. It is formed when in a β sheet, there are two residues in one strand opposite to the one residue in the adjacent strand (Fig. 1.4) (White and Poet, 1987). The space required by the extra residue causes this region of the β sheet to bulge out, hence the name. There are a few types of commonly occurring β-bulges (Chan et al., 1993; Craveur et al., 2013). One is called the classic β-bulge, which is found either within or at the edge of an antiparallel β-sheet. The first residue of this β-bulge is in α_R

conformation. The other type of β-bulge is the G1 β-bulge. G1 β-bulge is mostly found in antiparallel β sheets but can be found in parallel β sheets as well.

Ubiquitin contains two G1 β-bulges in its β sheet, one located closer to the N-terminal region and the other towards the C-terminal region of the protein. The C-terminal G1 β-bulge is made up of E64 and S65 on one strand and Q2 on the adjacent strand. The fact that these three residues occurring in G1 β bulge of ubiquitin are found very rarely in G1 β bulges of other proteins, makes this region of ubiquitin interesting. For example, in place of E64, S65 and Q2, there is a greater preference for D, G and N residues in the G1 β-bulges of other proteins. The E64, S65 and Q2 residues of G1 β-bulge in ubiquitin are highly conserved across species, suggesting that these residues have a role to play in the structure or functions of ubiquitin. Another unusual feature of this β bulge is that it is accompanied by a type II turn, such that while E64 is the first residue in the β bulge, it is the third residue in the type II turn (Vijay-Kumar et al., 1987).

As stated above, there is a greater preference for G than E in this position of G1 β bulge in other proteins (Chan et al., 1993). Moreover, the Φ and Ψ angles formed by E64 are unusual, as they do not fall in the allowed region for E residue in Ramchandran plot. Instead, these angle values are a characteristic of G residue (Vijay-Kumar et al., 1987). This further suggests that E64 has functional relevance for ubiquitin, which is why it is conserved despite these unusual features. I61 and K63 are two important residues adjacent to E64 and S65 of the β bulge. K63 is involved in K63 linked polyubiquitination (Spence et al., 1995), while I61 shows slow protection during kinetic refolding experiments (Briggs and Roder, 1992). I61 is also one of the residues involved in the UbEP42 substitution generated in our laboratory through random mutagenesis (Prabha et al., 2010), which is described in detail later in this chapter. Hence, the role of E64, S65 and Q2 residues of the C terminal G1 β bulge in the functioning of ubiquitin is worth studying.

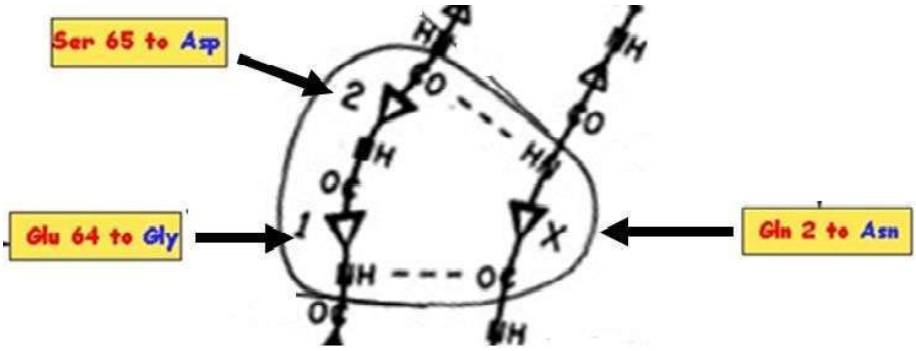

Figure 1.4: Structural of the C terminal G1 β bulge of ubiquitin. The portion within the circle shows the bulging out of one of the strands of the beta sheet due to the presence of the extra amino acid residue. The amino acid substitutions carried out at positions 2, 64 and 65 are shown.

1.5. N Terminal β Bulge of Ubiquitin

Ubiquitin has an antiparallel β bulge located in its N terminal region, in the type I turn of the β hairpin. As described in more detail below, the N terminal region of ubiquitin displays faster folding kinetics than the rest of the protein, and shows native structure early in the folding process (Briggs and Roder, 1992). Also, it is unusual for a β bulge to be found in a type I turn (Vijay-Kumar et al., 1987). This means the N terminal region of ubiquitin is worth studying. In one study, it was found that two peptides from ubiquitin, Ub1-21 and Ub1-35, can autonomously fold to β hairpin conformation in aqueous methanol (Cox et al., 1993). In fact, Ub1-21 can attain native like conformation in solution devoid of any organic solvent, though the extent to which it does so is much lower (Zerella et al., 1999). These reports suggest that the N terminal region of ubiquitin can be the site for initiation for the folding of ubiquitin, and the N terminal β bulge might play a central role in this process. Removing the G residue from the type I turn in the N terminal region led to loss of structure (Chen et al., 2001).

Whether the N terminal β bulge formation is a cause or result of the formation of the hairpin in this region was a long-standing question. To answer this question, the naturally occurring TLTGK sequence in the region was replaced by NPDG sequence, which led to formation of type I turn. As a result of this modification, non-native strand alignment was observed, suggesting that strand alignment in β hairpin is governed by the nature of the turn (Searle et al., 1995; Haque and Gellman, 1997). In another study, T9D substitution was carried

out in Ub1-71 peptide, and charged pairs were placed at different positions along the strand. The T9D substitution resulted in charge interaction between D9 and L11, making the hairpin more stable than the wild type sequence (Zerella et al., 2000). Although cross strand interactions do impart structural stability, they are not that vital for the formation of hairpin (De Alba et al., 1997; Santiveri et al., 2000).

1.6. Folding of Ubiquitin

Folding of ubiquitin has been studied using pulsed H-D exchange NMR. The study showed that the backbone amide protons in the N terminal region of β sheet and α helix are protected early in the folding process. In contrast, the folding kinetics of the C terminal half of ubiquitin is relatively slow. Specifically, residues Y59, I61 and L69 are characterised by slow protection rates. L56 residue, which is located at the beginning of 3_{10}-helix, also shows protection in early phase (Briggs and Roder, 1992). Studies on partially folded ubiquitin stabilized in 60/ 40% methanol/ water mixture showed that while the native secondary structural elements are conserved in the N terminal half of Ubiquitin (residues 1-35), in the C terminal half (residues 36-76) there is a transition from β sheet to a helical state (Harding et al., 1991; Stockman et al., 1993; Brutscher et al., 1997). These results suggest that the N terminal region of ubiquitin acts as an autonomously folding chassis, and determines the folding of the rest of the protein through tertiary interactions. Interestingly, T9D substitution in ubiquitin makes the protein more favourable than wild type, possibly because of electrostatic bonding between D9 and K11 residues.

1.7. Ubiquitination and its Types

As described before, ubiquitin is used by the cell as a tag to mark proteins for degradation. This tagging happens through the formation of an isopeptide bond between the ε amino group of a lysine on the substrate protein and the carboxyl group of the C terminal glycine of ubiquitin (Fig. 1.5) (Ciechanover et al., 1980; Hershko et al., 1980). If only one ubiquitin is attached to the substrate protein this way, the protein is said to be

monoubiquitinated. Alternatively, the ubiquitin molecule attached to the substrate protein can itself be ubiquitinated, and an entire chain of ubiquitin molecules covalently linked can be formed, called as polyubiquitin. In this case, the substrate protein is said to be polyubiquitinated. There are also instances in which single ubiquitin molecules are attached to different parts of the same substrate protein. In this case, the substrate protein is said to be multiubiquitinated.

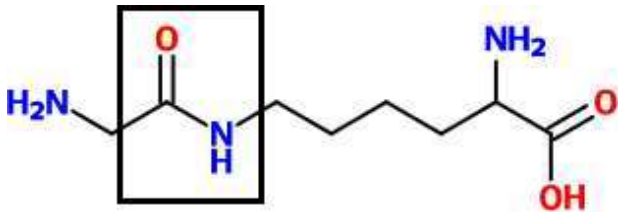

Figure 1.5: The isopeptide bond formed between the ε amino group of a lysine on the substrate protein and the carboxyl group of the C terminal glycine of ubiquitin. Image courtesy: Simon Caulton - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=27310148

Monoubiquitination is involved in many processes like DNA repair, transcriptional regulation, receptor transport, endosomal sorting, nuclear export, viral budding and histone regulation. Multiubiquitination is known to be involved in endocytosis. Polyubiquitintion is of several types, based on which lysine residue on ubiquitin is ubiquitinated. There are seven lysine residues in ubiquitin, namely K6, K11, K27, K29, K33, K48 and K63, and all of them can be ubiquitinated (Fig. 1.6, 1.7). If every ubiquitin molecule in a polyubiquitin chain is ubiquitinated on the same lysine residue, it is called a homotypic chain. And if different ubiquitin molecules in a polyubiquitin chain are ubiquitinated on different lysine residues, it is called a heterotypic chain. Sometimes, other proteins that are structurally related to ubiquitin, like SUMO and Nedd8, might also be found incorporated in a polyubiquitin chain. Such chains are called heterologous chains.

Figure 1.6: Location of the different lysine residues on ubiquitin.

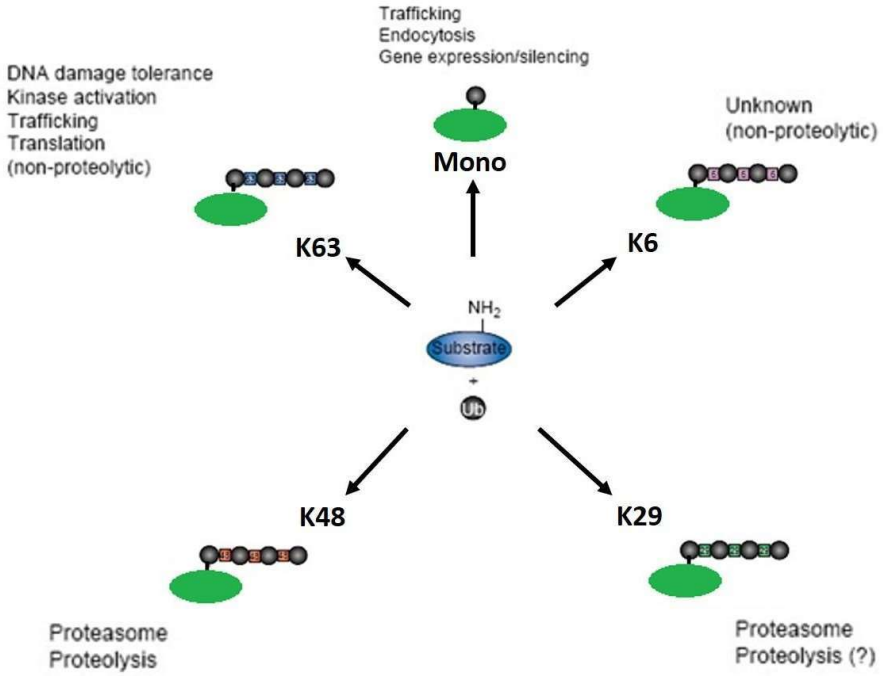

Figure 1.7: Types of ubiquitination and their functions.

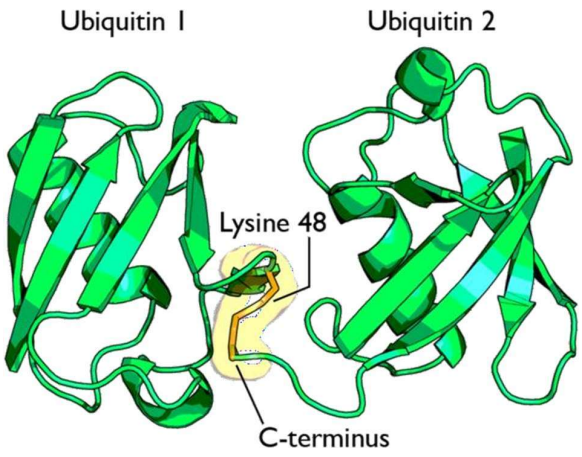

Figure 1.8: Topology of the K48 linked polyubiquitin chain. Image courtesy: Rogerdodd - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=4798852

The first polyubiquitin chain to be discovered was the K48 linked polyubiquitin chain (Fi. 1.8). It was found to be involved in proteasomal degradation of substrate proteins, one of the central functions of ubiquitin (Hershko et al., 1984, Baumeisteret et al., 1998; Lee and Goldberg, 1998; Rechsteiner, 1988; Chau et al., 1989; Pickart et al., 2000; Thrower et al., 2000; Groll and Huber, 2004; Pickart and Cohen, 2004; Wolfand Hilt, 2004; Rechsteiner and Hill, 2005). Another polyubiquitin chain that plays a prominent role in cell is the K63 linked polyubiquitin chain (Fig. 1.9). It is involved in processes like lysosomal degradation and autophagy (Kirkin et al., 2009), activation of protein kinases (Chen et al., 2009), protein trafficking (Hicke et al., 2003), DNA repair (Spence et al., 1995; Ulrich et al., 2002), inflammatory response (Sun et al., 2004) and regulation of ribosome activity (Spence et al., 2000). In addition, linear polyubiquitin chains have also been documented. In these polyubiquitin chains, the covalent linkage is not between the C terminal glycine of one ubiquitin and one of the lysine residues of other ubiquitin, but between C terminal glycine of one ubiquitin and the N terminal methionine of the other ubiquitin. These polyubiquitin chains are assembled by linear ubiquitin chain assembly complex (LUBAC) and are involved in NF-κB signalling (Iwai et al., 2009).

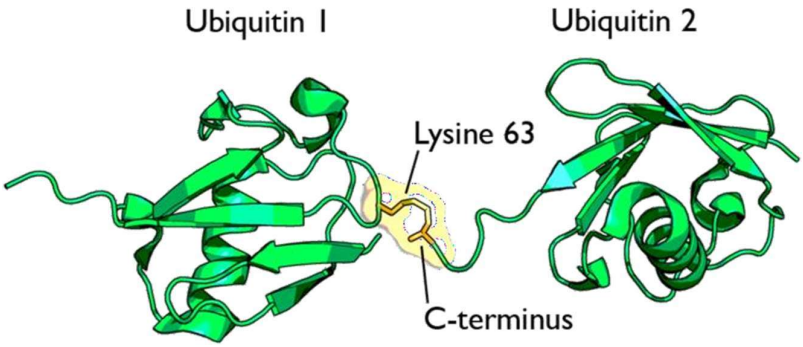

Figure 1.9: Topology of K63 linked polyubiquitin chain. Image courtesy: Rogerdodd - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=4798877

1.8. Ubiquitin Genes

In yeast, ubiquitin is encoded by four different genes, namely *UBI1*, *UBI2*, *UBI3* and *UBI4* (Fig. 1.10) (Ozkaynak et al., 1987). Out of these, *UBI1*, *UBI2* and *UBI3* produce a fusion of a ubiquitin and a ribosomal protein, while UBI4 produces a fusion of five ubiquitin molecules, arranged in head to tail fashion, without any spacers. These fusion proteins are separated from each other by C terminal hydrolase. *UBI1* and *UBI2* also have introns in their sequence, while *UBI3* and *UBI4* do not have any introns. Also, *UBI1* and *UBI2* are fused to the gene encoding the same ribosomal protein, namely L40, with a length of 52 amino acids, while *UBI3* is fused to a gene encoding a different ribosomal protein namely S31, with a length of 76 amino acids. It is believed that before their proteolytic separation from these ribosomal proteins, the ubiquitin moiety plays the role of a chaperone protein, helping in the assembly of ribosome. The proteolytic separation of the ubiquitin moieties from the ribosomal proteins is carried out by C terminal hydrolase.

UBI1, *UBI2* and *UBI3* are constitutively expressed genes that synthesise ubiquitin for the daily needs of the cell. *UBI1*, *UBI2* and *UBI3* genes account for the bulk of the free ubiquitin

pool in yeast (Catic and Ploegh, 2005; Ozkaynak et al., 1987). *UBI4*, on the other hand, is under the control of Heat Shock Box promotor, and hence is expressed only during stress conditions like high temperature (Finley et al., 1987; Ozkaynak et al., 1987), nutritional starvation (Finley et al., 1987; Ozaknak et al., 1987), UV exposure and antibiotic exposure (Finley et al., 1994). The function of *UBI4* gene, therefore, is to synthesise the additional ubiquitin needed by cell to degrade the misfolded and denatured proteins that are typically produced in large amounts as a result of stress condition.

Figure 1.10: The four genes that express ubiquitin in S. cerevisiae, showing introns (red), ubiquitin sequence (green), large ribosomal subunit sequence (light blue), small ribosomal subunit sequence (dark blue), extra C terminal/ N terminal amino acids (black).

1.9. Enzymes Involved in Ubiquitination

The process of ubiquitination is accomplished by a set of three enzymes, namely E1, E2 and E3 (Fig. 1.11) (Ciechanover et al., 1982; Hershko et al., 1983; 2000). The first step of ubiquitination is catalysed by E1. It binds to Mg-ATP complex and hydrolyses it into AMP and PPi (Haas and Rose, 1982). The energy released by this hydrolysis is used to form a high energy mixed anhydride bond between AMP and C terminal glycine residue ofubiquitin. This step primes ubiquitin for covalent attachment to substrate protein later. After this, a high energy thioester bond is formed between the C terminal glycine residue of the adenylated ubiquitin and a cysteine residue in E1. This step releases the AMP. The energy required to form the high energy thioester bond comes from the breaking of the high energy bond between AMP and C terminal glycine. Since the thioester bond is at a lower energy level than the anhydride bond between AMP and C terminal glycine, this whole process is spontaneous.

Ubiquitin is transferred from E1 to E2, the next enzyme in the process. As the thioester bond between C terminal glycine of ubiquitin and the active site cysteine of E1 breaks, a new thioester bond is formed between the C terminal glycine of ubiquitin and the active cysteine on E2 (Hershko et al., 1983). The energy needed to form this new thioester bond comes from the breakage of the older thioester bond between ubiquitin and E1. Since both these thioester bonds have the same energy level, there is no change in the energy level of ubiquitin during this process. There are many more types of E2s than there are E1s. For example, in *S. cerevisiae*, there are 14 E2s and higher organisms have even more. This diversity of E2s contributes to substrate specificity of ubiquitin system to some extent.

The final step in the process of ubiquitination is catalysed by the E3 enzyme. This is the step of forming an isopeptide bond between the C terminal glycine of a ubiquitin molecule accepted from an E2, and a lysine residue on the substrate protein (Hochstrasser, 1996; Hershko et al., 1983). The diversity of E3s in organisms is even more than that of E2s. This further increases substrate specificity of ubiquitin system. This high substrate specificity enables selective ubiquitination of proteins and through it, regulation of cellular activity. There are two main types of E3s, namely RING E3s and HECT E3s.

HECT E3 breaks the thioester bond between cysteine of E2 and C terminal glycine of ubiquitin, and uses the energy released to form a thioester bond between C terminal glycine of ubiquitin and its own cysteine residue. Later the HECT E3 breaks this thioester bond, and uses the energy released to form an isopeptide bond between C terminal glycine of ubiquitin and the ε-amino group of a lysine on the substrate protein. On the other hand, RING E3 does not form a covalent bond with ubiquitin. Instead, it simultaneously binds to the substrate protein and the E2 carrying ubiquitin molecule, and catalyses the transfer of the ubiquitin molecule from the E2 to the substrate protein.

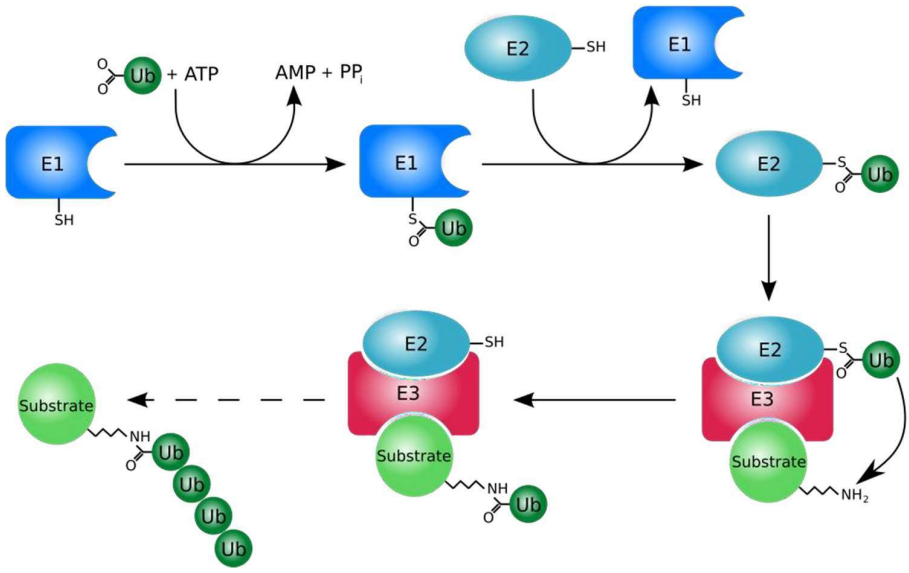

Figure 1.11: Enzymes involved in the process of ubiquitination. Image courtesy: Rogerdodd, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=7677277

1.10. Deubiquitinating Enzymes

Every process in the cell requires regulation so that it does not go out of control. In case of ubiquitination, this role is played by deubiquitinating enzymes, or DUBs (Wilkinson, 1997). DUBs remove ubiquitin attached to a substrate protein, a process called deubiquitination. Besides removing ubiquitin from substrate proteins, DUBs also break down polyubiquitin chains into individual ubiquitin molecules, thus replenishing the pool of free ubiquitin that can be used again, reducing the need to synthesise new ubiquitin. Thirdly, as mentioned before, the UBI1, UBI2 and UBI3 genes of ubiquitin produce ubiquitin fused to ribosomal proteins, and the UBI4 gene produces a fusion of four to five ubiquitin molecules. DUBs break down these fusions into individual ubiquitin molecules and ribosomal proteins. Finally, some unwanted reactions in cell produce useless chemical species like free ubiquitin adenylate, thiol esters between thiols like glutathione and the C terminus of ubiquitin, and amide derivatives of ubiquitin. DUBs break down these unwanted complexes to release free ubiquitin.

1.11. The Proteasome

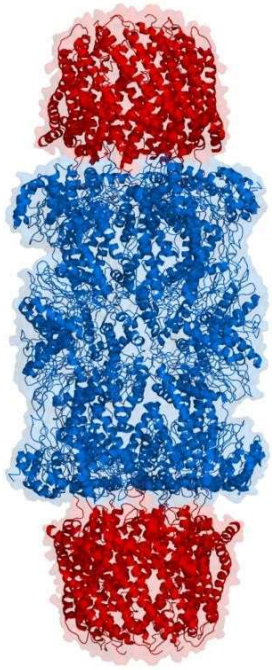

Figure 1.12: Tertiary and quarternary structure of the proteasome. The 20S subunit is shown in blue and the 19S subunit is shown in red. Image courtesy: Thomas Splettstoesser (www.scistyle.com) - Based on atomic coordinates of PDB 1FNT, rendered with open source molecular visualization tool PyMol (www.pymol.org), CC BY-SA 3.0, https://commons.wik

Proteasome is the protein complex that carries out one of the central functions of ubiquitin proteasome system, namely degradation of unwanted proteins (Hershko et al., 1984; Baumeister et al., 1998; Lee and Goldberg, 1998; Recksteiner, 1998; Groll and Huber, 2004; Pickart and Cohen, 2004; Wolf and Hilt, 2004; Rechsteiner and Hill, 2005). It belongs to a family of proteases called chambered proteases. Proteasome is a complex of many proteins and has two parts. One is the 20S subunit, which has the shape of a hollow cylinder, and the other is the 19S subunit (Fig. 1.12). Each of the two openings of the cylindrical 20S subunit is gated by one 19S subunit, which acts as a lid. The 20S subunit is 15 nm in length and 11 nm in diameter. It is made of two α rings and 2 β rings, stacked over each other in the order α β β α (Fig. 1.13). An α ring is made of seven α subunits and a β ring is made of seven β subunits (G Pühler, 1992). In the eukaryotic proteasome, three of the seven β subunits in a β ring have

proteolytic sites. And since there are two β rings in the proteasome, there are six proteolytic sites per proteasome.

The two β rings make up the central, catalytic chamber of the proteasome, containing six proteolytic sites. These proteolytic sites are not identical. The site at β1 has caspase like activity, the site at β2 has trypsin like activity and the site at β5 has chymotrypsin like activity (Heinemeyer, 1997; Baumeister et al., 1998). This diversity of catalytic activity increases the efficiency of protein degradation by proteasome. Attached to either end of the catalytic chamber formed by β rings is an α ring. These two α rings, together with the β rings adjacent to them, form two additional chambers on either side of the catalytic chamber. These two outer chambers store unfolded proteins before they are inserted into the catalytic chamber for degradation. The volume of the catalytic chamber is 84 nm^3 and the volume of the outer chambers is 59 nm^3 (Seemüller, 1995).

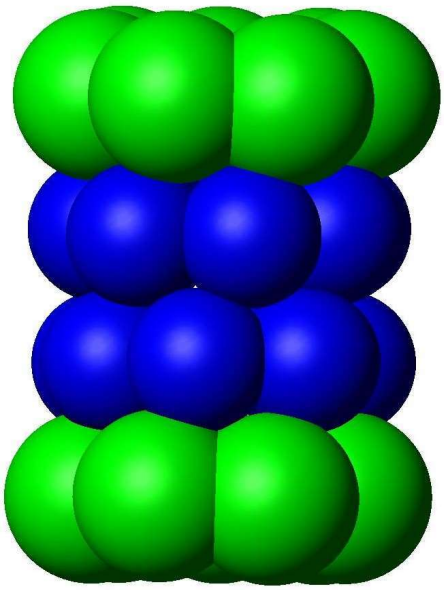

Figure 1.13: The four rings of the proteasome, that make up the inner, catalytic chamber and the two outer chambers. The two β rings are shown in blue and the two α rings are shown in green. Image courtesy: No machine-readable author provided. WillowW assumed (based on copyright claims). - No machine-readable source provided. Own work assumed (based on copyright claims)., CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=1455768

Access to the chambers is controlled by pores that are only 2 nm in diameter, which means only unfolded proteins can pass through them. The 19S complex, which is associated

with both ends of 20S complex, brings about a conformational change in the α subunits to allow entry of proteins into the proteolytic chamber. The 19S complex consists of 19 subunits and is made of two parts, the base and the lid. The base is composed of 8 subunits and unfolds and inserts substrate proteins into the 20S complex. This process requires energy, which the base complex gets from the ATPase activity of 6 of its 8 subunits. The lid is also composed of 8 subunits and allows only proteins tagged with polyubiquitin chains to be unfolded and inserted into the proteasome. This is where the proteasome gets its substrate specificity from.

1.12. Ubiquitin Fusion Degradation (UFD) Pathway

Studies have shown that a protein N terminally fused to a ubiquitin moiety has a short life *in vivo*, as the ubiquitin moiety fused to the protein serves as a degradation signal (Johnson, 1995). The proteolytic system involved in degrading such fusion proteins is called Ubiquitin Fusion Degradation (UFD) pathway. By analysing mutations that disturb this pathway, five genes were identified to be involved in the pathway. Two of the five genes, namely *UFD1* and *UFD5*, operate in the post ubiquitination steps. Out of the other three genes, *UFD3* controls the cellular concentration of ubiquitin, as deleting it reduces free ubiquitin concentration in the cell. *UFD2* and *UFD4* genes control the synthesis and topology of a multiubiquitin chain, which is connected to the ubiquitin moiety fused to the C-terminal of the protein.

1.13. N-End Rule and the Role of Ubiquitination

N-end rule is a rule that decides the half-life of a protein based on which amino acid residue is found at its N terminal (Varshavsky, 1996). In other words, the half-life of a protein is determined by its N terminal residue, which can act as a degron (degradation signal recognised by the proteolytic machinery). The rule applies to both prokaryotic and eukaryotic organisms. In eukaryotes, these N terminal residues form a part of N degrons, which are

recognised by ubiquitin E3 ligases. Hence, the identity of the N terminal residue determines the probability of an E3 ligase ubiquitinating the protein for degradation. However, the impact of any particular amino acid on the rate of degradation of the protein can vary from organism to organism. Based on their ability to act as degrons, the N terminal residues are divided into three categories, namely primary destabilizing residues (N-d^P), secondary destabilizing residues (N-d^s) and tertiary destabilizing residues (N-d^t).

N-d^P residues are recognised by an E3 called N recognin. N recognin has two substrate binding sites, namely type I and type II sites. Type I site binds to basic N-d^P residues, namely arginine, histidine and lysine. Type II site, on the other hand, binds to bulky hydrophobic N-d^P residues, namely phenylalanine, leucine, tyrosine, tryptophan and isoleucine. In *S. cerevisiae*, it is possible to disrupt one of these sites through mutagenesis, without disrupting the functioning of the other site (Varshavsky, 1996). The N-d^S residues in *S. cerevisiae* are aspartic acid and glutamic acid, both of which have acidic side chains. In mammalian cells, cysteine is a third N-d^S residue. These N-d^S residues are acted upon by R transferase, which arginylates aspartic acid and glutamic acid residues if they are located at the N terminal (Varshavsky, 1996). The N-d^T residues are asparagine and glutamine, both of which are polar residues with amide groups in their side chains. They are acted upon by Nt-amidase, which deamidates their side chains if they are located at the N terminal (Varshavsky, 1996).

Residues that are not recognized by the components of the N end rule pathway, even when they are located at the N terminal, are called stabilizing residues. These residues include glycine, valine and methionine. The non-compartmentalized proteins in both prokaryotes and eukaryotes almost always have stabilizing N terminal residues, while the compartmentalized proteins usually have destabilizing N terminal residues. *In vivo*, destabilizing residues are often added to proteins post translationally to reduce their half-life. These modifications are believed to occur during entry or exit from cell cycle, as a response to chemical or physical stress, and during cell differentiation, like spermatogenesis and erythropoiesis.

There are other factors too, apart from the identity of the N terminal residue, which determine the half-life of a protein *in vivo*. For example, a protein with a destabilizing N terminal residue might still have a long half-life, if its N terminal is buried in the core and hence is inaccessible to the components of the N end rule pathway. Acetylation or other modifications of N terminal residues can also prevent them from triggering the N end rule pathway. The distribution of lysine residues amenable to ubiquitination around the N terminus also

determines the extent to which the destabilizing N terminal residues can exert their effect. Other factors that affect the efficiency of destabilizing N terminal residues are the structure of C terminal and the overall quaternary structure of the protein complex. For example, the N terminal of a protein in a complex might be hidden from components of the N end rule pathway by another subunit in the complex. However, N end rule pathway plays a vital role in the degradation of proteins in the cell (Bachmair et al., 1986).

1.14. Role of Ubiquitin in Cell Function and Disease

Ubiquitin proteasome regulates a large variety of cellular pathways by selectively degrading proteins regulating those pathways. The many different processes in which ubiquitin proteasome system plays a key role include DNA repair (Jentsch et al., 1987), regulation of transcription (Hochstrasser and Varshavsky, 1990), cell cycle (Goebl et al., 1988), differentiation and development (Bowerman et al., 2006), nutritional stress (Finley et al., 1987; Ozkaynak et al., 1987), regulation of cell surface receptors, ion channels and secretory pathway (Hicke and Riezman, 1996), neural network morphogenesis, response to stress conditions and extracellular effectors (Ozkaynak et al., 1984; Finley et al., 1987), degradation of misfolded and denatured proteins and regulation of immune response (Ciechanover and Iwai, 2004; Varshavsky, 2005). Due to the vital role these pathways play in the cell's functioning, disruption of the proteasomal degradation of their regulatory proteins can have severe consequences, which can lead to disease.

One of the proteins involved in cell cycle arrest is the p53. In normal growth conditions, p53 levels are low as it is acted upon by Mdm2, an E3 ligase. However, at the end of cell division, or if the DNA is damaged, p53 undergoes phosphorylation, which prevents its proteasomal degradation, leading to cell cycle arrest and cell death. Due to this role of p53 in cell cycle arrest, excessive degradation of p53 can cause cell division to go out of control, leading to cancer (Ciechanover and Iwai, 2004). NF-κB is a transcription factor that is usually inhibited by IκB. Phosphorylation of IκB leads to its proteasomal degradation, which allows NF-κB to express genes involved in inflammatory response (Karin and Ben-Neriah, 2000).

Another role of ubiquitin proteasome system in immune response is degradation of antigen proteins so that the peptides produced can be presented by the MHC I molecules. Consequently, a defective ubiquitin proteasome system can lead to impaired antigen presentation, compromising the immune response. Moreover, impaired proteasomal degradation of denatured or unwanted proteins has been implicated in neural diseases like Alzheimer's and Parkinson's.

Pathological states caused by mutations can be grouped into two categories, namely loss of function and gain of function. In loss of function mutations, the defects in proteasomal degradation of substrate proteins leads to increase in half life of the proteins. Examples of such loss of function due to reduced proteasomal degradation of target proteins are neurodegenerative diseases like Alzheimer's disease (Keck et al., 2003), Parkinson's disease (Mc Naught and Jenner, 2001) and Huntington's disease (Zhou et al., 2003). Similarly, decreased proteasomal degradation of positive regulators of cell cycle can lead to uncontrolled cell division, leading to cancer. In contrast, gain of function mutations cause excessive proteasomal degradation of substrate proteins. The conditions caused by gain of function of ubiquitin proteasome system include inflammation and muscle wasting (Hobler et al., 1999), and cancer caused by excessive degradation of proapoptotic proteins or negative regulators of cell cycle. Other than the defects caused by mutations, ubiquitin proteasome system can behave improperly if it is excessively burdened by large number of misfolded or denatured proteins during stress conditions (Reed and clocks, 2003).

1.15. *Saccharomyces cerevisiae* as a Model System

Saccharomyces cerevisiae is one of the most popular model systems in biology (Botstein and Fink, 1998). This is because while being a unicellular microorganism, yeast has a very short doubling time, which enables fast experimental studies, being a eukaryote, yeast also shares many of the biomolecular features of the higher eukaryotes. This includes the ubiquitin proteasome system. Hence, yeast is placed at the sweet spot between the fast dividing prokaryotes and slow growing human like higher eukaryotes. As mentioned before, ubiquitin differs between yeast and humans by only three residues. Thus, findings of studies on yeast ubiquitin can be easily extrapolated on to human ubiquitin. Moreover, yeast is much more

robust than the mammalian cell lines, which makes it far easier to maintain. Due to these reasons, our laboratory has selected *Saccharomyces cerevisiae* as the model system to study structure function relationships in ubiquitin.

1.16. Use of β-bulge Substitutions to Study Structure Function Relationships

As mentioned before, the E64, S65 and Q2 residues of the G1 β-bulge of ubiquitin are unusual residues for the structure, as they are not commonly found in G1 β-bulges in other proteins. In G1 β-bulges of other proteins, residues G, D and N are found with much higher frequency in place of E, S and Q. And yet, residues E64, S65 and Q2 are highly conserved across species in ubiquitin, which suggests that they play a vital role in ubiquitin's structure and/ or functions. To understand their role, substitutions E64G, S65D and Q2N were carried out by our lab using site directed mutagenesis (Mishra et al., 2009; 2011; Sharma et al., 2012; 2015). The secondary structure of the single mutant proteins UbE64G, UbS65D and UbQ2N was only marginally different from UbF45W. UbF45W is structurally identical wild-type ubiquitin UbWt. However, the F45W substitution makes the protein fluorescent homologue. In the guanidinium chloride induced denaturation experiments it was found that UbE64G and UbQ2N were found to be more stable compared with UbF45W. Expression of the β-bulge mutants of ubiquitin had no effect on growth, N-end rule associated protein degradation. Expression of UbE64G in vivo shows due to complementation substrate proteins with Pro as the N-terminal residue, which undergo ubiquitination have extended half-lives. This reduced preference for Pro might be related to natural preference of glutamate at 64th position in ubiquitin. Functional significance of these β-bulge residues in ubiquitin might be due to the fact that the β-bulge lies right next to K63, the lysine residue involved in K63 linked polyubiquitination.

1.17. Isolation of UbEP42 Mutation through *in vitro* Evolution

Many efforts today are focused on developing novel proteins with certain desirable qualities, which makes them useful for certain applications, including in healthcare. These novel proteins are not naturally found in living organisms, but artificially engineered. One of the strategies for developing such novel proteins is *in vitro* evolution. *In vitro* evolution is a process in which a naturally occurring protein is subjected to random mutagenesis so that by sheer chance, a mutant version of the protein is produced that has some desirable features that were not found in the original protein. The rate of random mutagenesis can be increased either by exposing the cells to mutagens like mutagenic chemicals or UV rays, or through error prone PCR.

A naturally occurring variant of wild type ubiquitin has been documented in *Autografa californica*. It shows only 75% sequence similarity with wild type ubiquitin, and is capable of competitively inhibiting the formation of polyubiquitin chains. This shows that artificially produced ubiquitin variants could be used for inhibiting ubiquitin activity in case of diseases that are caused by an overexpression of ubiquitin. There have been many efforts to produce such artificial variants of ubiquitin using site directed mutagenesis, including some from our lab. But they have met little success so far in acting as inhibitors of ubiquitin activity. Random mutagenesis for *in vitro* evolution, however, can be a much more powerful tool for developing engineered proteins, as it scans through a huge number of different mutations in a very short time, increasing the chances of finding a useful variant.

While natural evolution happens over millions of years, especially in the slow growing higher organisms, *in vitro* evolution happens very fast due to the much shorter doubling times of microbes and the higher rate of artificially induced mutagenesis. Our lab chose *Saccharomyces cerevisiae* as the model organism for *in vitro* evolution of ubiquitin and error prone PCR as the method to induce mutagenesis. Error prone PCR is a variation of the PCR technique, in which the fidelity of DNA Polymerase is reduced to increase the rate of error. This can be done by various means like playing with the concentration of Mg ions, introducing Mn ions in the solution, or biasing the concentration of dNTPs. DNA polymerase needs Mg ion as a cofactor to carry out its catalytic activity. Hence, changing Mg ion concentration affects

the fidelity of DNA polymerase. Biasing the concentration of dNTPs also affects the fidelity of DNA polymerase, because it increases the chance of incorporating dNTP with higher-than-normal concentration and decreases the chance of incorporating the dNTP with lower-than-normal concentration.

Using error prone PCR, random mutants of ubiquitin were generated and screened for absence of complementation in *UBI4* mutants. Denoted as *ubi4* mutants, these strains lack the *UBI4* gene, and hence are sensitive to temperature stress. Among many mutants produced by successive rounds of error prone PCR, one mutant, which was named UbEP42, was particularly detrimental (Prabha et al., 2010). It caused lysis of *ubi4* mutant cells even when they were not under any kind of stress. This showed that UbEP42 interfered with the normal functioning of ubiquitin produced by *UBI1*, *UBI2* and *UBI3* genes (Prabha et al., 2010; Doshi et al., 2012). Sequence of this mutant revealed that there were seven base substitutions in its gene. Out of these seven base substitutions, three result in codons that code for the same amino acid that the original codon coded, resulting in no amino acid substitution. The remaining four base substitutions result in four amino acid substitutions, namely S20F, A46S, L50P and I61T.

In order to find out how each of these four substitutions contributed to the detrimental effect of the UbEP42 mutation, each of them was separately introduced in the ubiquitin gene through site directed mutagenesis and their effects were checked (Doshi et al., 2017). It was found that some of these substitutions are more detrimental than the others. The most lethal substitutions turned out to be L50P and I61T, while S20F and A46S were found to have no negative effect. L50P and I61T cause ubiquitin to be so structurally unstable and functionally defective.

1.18. Understanding the Importance of Conserved Residues in Ubiquitin through the Functional Characterisation of Mutants

The main objective of structural biology research is to understand structure-function relationships in macromolecules and especially proteins. Small proteins have served as models

in many studies aiming at unravelling structure-function relationships in proteins. Among small proteins ubiquitin stands out in many ways. **Chapter 1** gives an introduction to ubiquitin, its detailed structure, the two β-bulges of ubiquitin, types of ubiquitination, enzymes involved in ubiquitination, deubiquitinating enzymes, the N-end rule, UFD pathway and role of ubiquitin in cell function and disease. The chapter also describes the importance of yeast as a model system and presents the objectives of the work described in this thesis and the results obtained.

Ubiquitin is a small globular protein of 76 residues. The protein is highly stable due to extensive hydrogen bonding reinforcing its structure. It is not associated with any enzyme activity. Ubiquitin is conjugated to other proteins by forming isopeptide linkage through its C-terminal to the lysine sidechains of the substrate proteins. There are seven lysines in its structure. The number of molecules of ubiquitin attached to protein, the number of ubiquitin molecules attached to each other making a polyubiquitin chain, and the branching patterns in polyubiquitin chain convey different signals to the cell. Further, ubiquitin shows more than 96% sequence conservation through evolution. Therefore, it is not possible to draw conclusions by comparing the sequences of ubiquitin protein from different species. In order to understand the importance of residues in the structure and functions of ubiquitin mutant forms of ubiquitin were generated using site directed mutagenesis and error-prone PCR. The resultant β-bulge and UbEP42 derived double mutants were characterized by studying various functions.

Studies in our laboratory showed that while the structure of ubiquitin is not affected by these substitutions (as observed through UV CD spectra and guanidinium chloride denaturation curves), certain functional aspects of ubiquitin are affected (Mishra et al., 2009; Mishra et al., 2011). The substitutions caused an increased sensitivity towards cycloheximide, which is a translational inhibitor that acts on the elongation phase of protein synthesis by binding to the large subunit of ribosome. L28, a ribosomal protein in the large subunit, is known to undergo polyubiquitination, and the G1 β-bulge subjected to these substitutions lies adjacent to the K63 residue. There was no effect of the β-bulge substitutions on adherence to N-end rule, growth under normal conditions or survival under heat stress. This showed that the functional relevance of the β-bulge residues should be further probed using these substitutions.

Chapter 2 deals with further characterisation of the G1 β bulge, using the E64G, S65D and Q2N substitutions. The β bulge mutations did not affect the functions of protein sorting into MVBs and protein degradation by lysosomes. As it was observed earlier observation of no change in growth due to β bulge mutations There is no change produced in the Cdc28 protein

kinase levels due to the expression of β bulge mutations, supporting the. The studies on complementation efficiency of β bulge mutations on growth showed the antibiotic G418 affected growth of yeast, while gentamycin showed no negative effect. Hygromycin used at lower concentrations also produced no effect. K48 linked polyubiquitination is reduced with Q2N and E64G mutations of ubiquitin.

Eukaryotic cells cannot exist without ubiquitin. Hence, lethal mutations of ubiquitin may be employed for the removal of unwanted cells like cancer cells. Earlier, error-prone PCR was used to generate random mutations. Expression of one of the mutants UbEP42 showed it confers dosage dependent lethal phenotype on host cells and was found to be lethal to cells even under normal conditions. The UbEP42 mutation carries four substitutions in its sequence. The substitutions are S20F, A46S, L50P and I61D (Prabha et al., 2010; Doshi et al., 2012). In the present study we investigated the functional reasons for the observed lethality. Expression of UbEP42 in a UBI4-deleted stress-sensitive strain resulted in an increased generation time due to a delayed S phase caused by decreased levels of Cdc28 protein kinase. Cells expressing UbEP42 displayed heightened sensitivity towards heat stress and exposure to cycloheximide. Furthermore, its expression had a negative effect on the degradation of substrates of the ubiquitin fusion degradation pathway. However, UbEP42 is incorporated into polyubiquitin chains. Single mutations isolated from UbEP42 were studied to understand their functional contribution to the observed lethality (Doshi et al., 2017). Two mutations L50P and I61T were found to be the lethal mutations affecting almost all the functions tested.

Double mutations produce varied effects in proteins, depending on the structural relationship of the mutated residues, their role in the overall structure and functions of a protein. Here, in **Chapter 3** the double mutations derived from UbEP42 were studied in various combinations (Sharma et al., 2021). Six double mutants derived from the ubiquitin mutant UbEP42, namely S20F-A46S, S20F-L50P, S20F-I61T, A46S-L50P, A46S-I61T, and L50P-I61T, have been studied here to understand how they influence the ubiquitination related functions, by analysing their growth and viability, Cdc28 levels, K-48 linked polyubiquitination, UFD pathway, lysosomal degradation, endosomal sorting, survival under heat, and antibiotic stresses. The double mutation L50P-I61T is the most detrimental, followed by S20F-I61T and A46S-I61T. The double mutations studied here, in general, make cells more sensitive than the wild type to one or the other stress. However, the excessive negative effects of L50P and I61T are compensated under certain conditions by S20F and A46S mutations. The

competitive inhibition produced by these substitutions could be used to manage certain ubiquitination associated diseases.