# Chapter 1
# Introduction

## 1.1. Introduction

Proteins are essential for biomolecular signalling and maintenance of cellular processes. Many proteins undergo modifications before serving as functional entities in different cells. These modifications are known as, "post-translational modifications." Protein post-translational modifications are important to regulate the functions and levels of various proteins of the proteome. These modifications influence almost all aspects of normal cell biology and pathogenesis.

Ubiquitination is one such post-translational modification that regulates many cellular processes and creates versatility in cell signalling. It functions in both cytoplasm and nucleus (Shabek and Ciechanover, 2010). The system involves addition of a small protein ubiquitin as a tag to target proteins for degradation. Ubiquitin can be attached to a substrate protein or to an ubiquitin molecule that is already attached to a substrate, with the latter resulting in an inter-ubiquitin linkage. Generally ubiquitin linkage is an isopeptide bond formed between the ε-amino group of one of lysines of the substrate protein and the carboxyl group of G76 of ubiquitin. While building polyubiquitin chains, similar isopeptide linkages are formed between the ubiquitin already bound to the substrate and incoming second ubiquitin molecule (Hershko et al., 1983). There are seven lysine residues in ubiquitin as K6, K11, K27, K29, K33, K48 and K63 and all of them participate in polyubiquitin linkages, with different linkages signifying different functional meanings. Polyubiquitin chains with K48 and K11 linkages serve as signals for degradation by proteasome (Ye and Rape, 2009); while linkages with other lysines regulate processes such as inflammation, translation, DNA repair, endosome trafficking as summarised in **Fig.1.1**. Ubiquitination exhibits variation in the length and position of chains, it could be monoubiquitination (transfer of single ubiquitin molecule to substrate), multi-monoubiquitination or polyubiquitination. Further, these chains could be linear or branched.

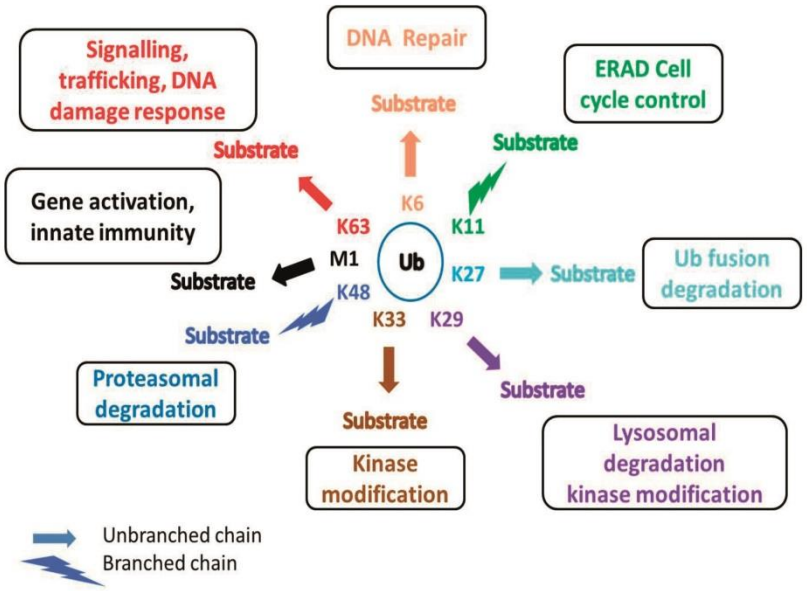


**Fig.1.1. Different ubiquitin linkage types and their role in cellular signalling.** Substrate ubiquitination with 8 possible linkage types (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, Lys63 and M1) of ubiquitin chains. Major functions are indicated in boxes.

### 1.1. The ubiquitin proteasome pathway

Ubiquitination is catalyzed by three enzymes E1 (ubiquitin activating enzyme), E2 (ubiquitin conjugating enzyme) and E3 (ubiquitin ligase) which act in a relay system (**Fig. 1.2**). In the cells, even though there is only a single kind of E1 (or two forms in case of human ubiquitin system) there are hundreds of E2s and nearly thousand E3s present, making the process of ubiquitination highly substrate specific (Schulman and Wade Harper, 2009). Ubiquitin must be activated before it can be attached to a substrate. This is done by ubiquitin activating enzyme or E1 forming a covalent thioester bond between itself and ubiquitin in a reaction driven by the energy released during ATP hydrolysis. The energy of the ATP hydrolysis is conserved in the thioester bond, creating an activated form of ubiquitin. Ubiquitin bound to E1 is then transferred to active site cysteine side chain of ubiquitin conjugating enzyme or E2, forming a second thioester linkage. E2-associated ubiquitin serves as a pool of activated ubiquitin, for transfer to proteins. The E2~Ub

thioester intermediate product functions with several families of E3 ubiquitin ligases to conjugate the Ub molecule to the lysine side chain of a substrate, which is accomplished by two different mechanisms depending on the class to which E3 belongs (Buetow and Huang, 2016). Four to five cycles of this process is repeated and a polyubiquitin chain is formed. Depending on the nature of substrate protein and the kind of linkages present in polyUb chain, either the protein is driven to 26S proteasome for degradation or for the ubiquitinated protein's specified role. Polyubiquitin chain is separated from the substrate and is cleaved into free Ub molecules, which are recycled.

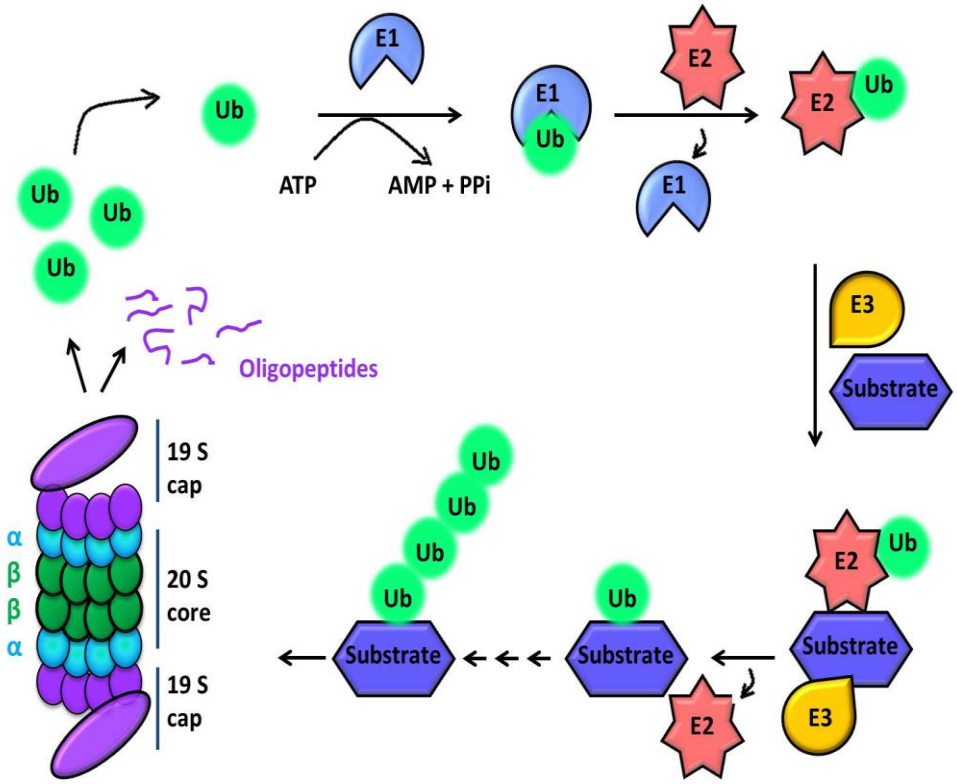


**Fig.1. 2. Schematic overview of the ubiquitin proteasome pathway.**

### 1.2. Ubiquitin activating enzyme (E1)

E1 is a multi-domain enzyme encoded by *UBA1* gene **(Fig. 1.3)**. The base of the enzyme is made of inactive and active adenylation domains (IAD and AAD, respectively). The domains arrange themselves into a pseudodimer structure and the latter constitutes the rigid body of the enzyme. Initially the AAD binds Ub, ATP, and $Mg^{2+}$ and this is followed by formation of ubiquitin adenylate (Lee and Schindelin, 2008). The enzyme constitutes two more domains known as the first and second catalytic cysteine half domains (FCCH and SCCH, respectively). The catalytic cysteine resides in the SCCH domain (Lee and Schindelin, 2008). Ubiquitin from ubiquitin adenylate, which is bound non-covalently to E1, is transferred to active site cysteine in SCCH domain. Lastly through its ubiquitin fold domain (UFD) located at the C-terminus, E1 binds with E2. Thus, E1 undergoes large conformational changes to perform its dual catalytic activities.


**Fig. 1.3. Domain organization of *UBA1*.** Representational arrangement of domains in *UBA1* of *S. cerevisiae* with residue numbers lining the domain boundaries.

#### 1.2.1. Transfer of ubiquitin to E2: interaction of E1 with E2

E2s interact with their related E1s with substantial affinity only if the E1 is carrying their modifier i.e. ubiquitin (Haas et al., 1988; Schulman and Wade Harper, 2009). Initial structural studies revealed that E1 undergoes a series of structural changes in presence of ATP and binds Ub to adenylate the Ub C-terminal glycine (Lee and Schindelin, 2008; Lois and Lima, 2005; Walden et al., 2003). Once the adenylation has occurred, the E1 releases pyrophosphate (Olsen et al., 2010). The enzyme goes through a whole network of conformational changes, which are essential for thioester bond formation and are referred as, "active site remodelling" (Olsen and Lima, 2013). Once

the thioester bond is established, the E1 Cys domain rotates back to allow open configuration and enable a second round of adenylation (Huang et al., 2007). Thus, the E1 is loaded with two Ub molecules and now can recruit an E2 (Haas et al., 1988).

When E1 is charged by ubiquitin it exposes cryptic sites on the UFD domain of E1 enabling E2s to interact with them and establish the E1-E2 complex (Haas et al., 1988; Huang et al., 2007; Lois and Lima, 2005). What was referred to as cryptic sites is now being identified as negatively charged groove **(Fig. 1.4)** within the ubiquitin fold domain (UFD) in the E1 that becomes available for recognition by two highly conserved Lys residues present in α-helix 1 of all ubiquitin E2s (Lee and Schindelin, 2008).

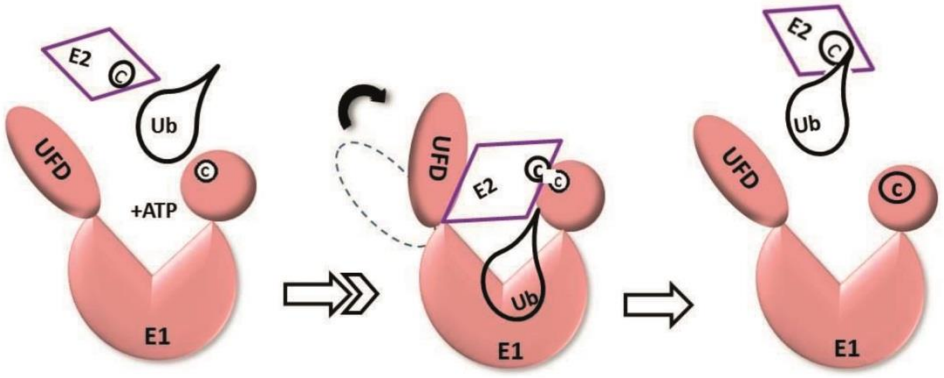


**Fig. 1.4**. **Transfer of ubiquitin to E2 by E1.** The ubiquitin fold domain, UFD, recruits E2 enzymes in a 'distal' conformation and subsequently rotates to a more 'proximal' conformation that places the E1 and E2 catalytic cysteines into proximity to facilitate E1-E2 thioester transfer.

Structural studies with UBC4 by the UFD of E1 from *Schizosaccharomyces pombe* establish basis for molecular recognition as contacts between the E1 Cys domain and E2, which are important for thioester transfer. Various mutational analyses reveal there is a high degree of conservation at the E1 UFD/E2 interface (Olsen and Lima, 2013).

### 1.2.2. Transferring ubiquitin: interaction of E2 with E3

Charged with ubiquitin, E2s interact with E3s to catalyse substrate ubiquitination. An E2 can interact with several different E3s (Ye and Rape, 2009). Differentiated structurally and functionally how they facilitate the transfer of ubiquitin from E2 to the predestined substrate, E3s are divided into three major classes. The first group of E3s contain RING (Really Interesting New Gene) domain or structurally related U-box. The RING and U-box groups act as matchmakers, bringing a substrate and a charged E2 together bringing them in direct contact **(**Scheffner et al**.,** 1995**).** The other two groups correspond to enzymes of the HECT (Homologous to E6-AP C-Terminus) domain family and RING-between-RINGs (RBRs). These two groups of E3s help in shuttling of ubiquitin from their cognate E2s to the targeted substrates involving an intermediate step where the ubiquitin is first transferred from the E2 to an active-site cysteine residue on the E3 ligase before it is conjugated to the target protein (Sluimer and Distel, 2018) (**Fig.1.5**).

The final outcome of substrate ubiquitination depends on E2-E3 cognate pair that is responsible for a specific response. All E2s characterized so far recognize E3s through the L1 and L2 loops and the N-terminal α-helix 1 on the E2 surface (Lee and Schindelin, 2008). It is not crucial that the residues involved in E3 recognition must be same, since a single E2 interacts with multiple E3s (Lee and Schindelin, 2008; Yin et al., 2009; Zhang et al., 2005).

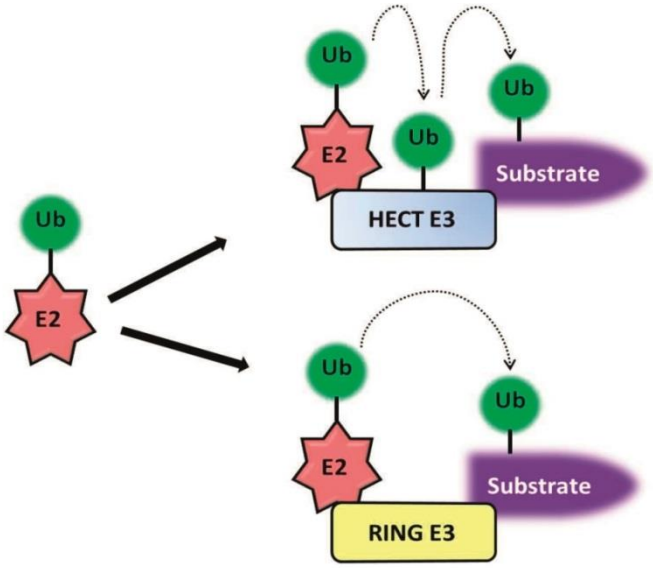


**Fig. 1.5. Transfer of activated ubiquitin from E2 to substrate with the help of E3.**

### 1.3. Ubiquitin conjugating enzymes (E2s)

(Previously also known as the carrier protein). Initially when the E2s were found they were named as E2-nk (where n denotes the molecular weight of the E2) and UBCn in yeast or UBCHn in humans (where n corresponds to the order of discovery). Later on, other E2s were labelled according to their discovery in genetic or proteomic screens, without a reference to their E2 function, for example Huntingtin‐interacting protein 2 (HIP2/E2‐25K/UBCH1/UBE2K). This leads to confusion as E2s from different organisms bearing similar names were not functionally related and most E2s have multiple names. To bring a solution to this and unify the nomenclature of mammalian E2s, bioinformatics‐driven system approach was incorporated in the identification of all predicted human E2s. This system used the form UBE2Xn, where the combination of letter 'X' and number 'n' specify different E2s (Ye and Rape, 2009). E2s have several structural features which influence their function. Hence, any deletion or mutation of the structural features leads to collapse of proteasomal degradation system. It was recently discovered that E2s recognize substrates that are to be degraded as per the N-

end rule (Sung et al., 1991). However, the role of E2s in polyubiquitin chain formation is still not completely understood. So far, 13 E2s in *Saccharomyces cerevisiae* and ~40 E2s in human have been identified (Stewart et al., 2016; Ye and Rape, 2009).

E2s are crucial in the UPS cascade, as they regulate both the topology of the poly-Ub chains and the processivity of the polyubiquitination reaction (van Wijk and Timmers, 2010; Ye and Rape, 2009). They catalyze attachment of Ub to different lysines on protein substrate or Ub, during mono- or poly-ubiquitination. This versatility in ubiquitin chain topology permits generation of diverse protein-Ub structures, which serve distinct functional purposes (Komander and Rape, 2012; van Wijk et al., 2009). All the members belonging to ubiquitin-conjugating enzymes (E2s) group are characterized by the presence of a highly conserved 150–200 amino acid ubiquitin-conjugating catalytic domain (UBC) (Hofmann and Pickart, 2001). Within this domain, a catalytic cysteine is embedded that accepts the activated Ub from E1 through a thioester bond, and subsequently participates in substrate conjugation by binding to E3 (van Wijk and Timmers, 2010; Ye and Rape, 2009). This domain constitutes 14–16 kDa stretch of amino acid residues, which are ~35% conserved among different family members and provide a binding platform for E1s, E3s, and the activated Ub (Burroughs et al., 2008a).

### 1.4. Classification of E2s

Ubiquitin conjugating enzymes are variously referred to as the ubiquitin carrier proteins, the mediators of ubiquitin conjugation. Structurally they are grouped into four classes, all of which have a conserved core catalytic domain (containing the active site cysteine), and some of which additionally have extensions at N and C terminals of the conserved core catalytic domain. The enzymes have been classified as: Class I - with only conserved core catalytic domain also known as the UBC domain, Class II – UBC domain with variable C terminal extension, Class III - UBC domain with variable N-terminal and Class IV- UBC domain with extensions on both the terminals

(**Fig. 1.6**) (Cook et al., 1993). The extensions contribute to the differences such as in the sub-cellular localization of E2s, stabilization of their interaction with E1 enzyme and modulation of the activity of the interacting E3s (van Wijk and Timmers, 2010).

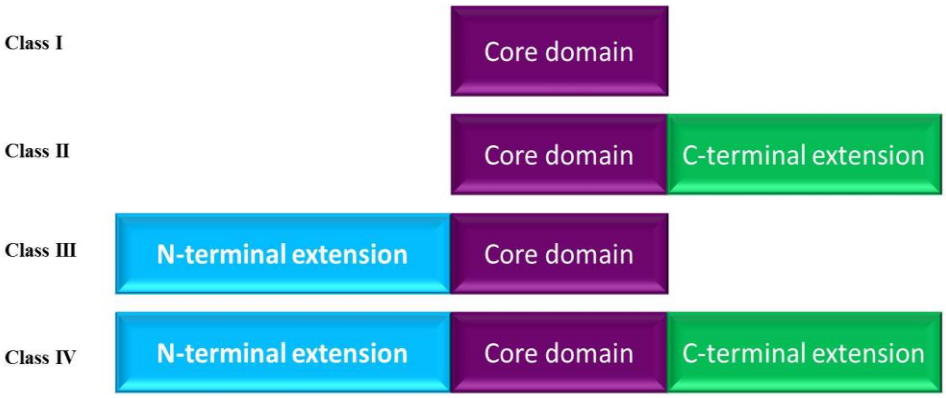


**Fig. 1.6. Structural classification of E2s.**

### 1.4.1. Breakdown of E2s

#### 1.4.1.1. UBC domain

The catalytic site or UBC domain or fold is the site of transthioesterification reaction and is the seat through which ubiquitin is transferred from E1 to E3 or targeted substrate. This domain is essentially present in all of the various classes of E2 enzymes. The secondary structure of UBC fold comprises of 4α helices, short $3_{10}$-helix and 4 antiparallel β sheets (van Wijk and Timmers, 2010). The β strands are centrally located, which are surrounded by the helices 1 and 2 (H1 and H2) while helix 3 and helix 4 (H3 and H4) border the other end (Burroughs et al., 2008b; Michelle et al., 2009; Winn et al., 2004a).Two distinct loops loop 1 (L1) and loop 2 (L2) originate within the meandering β sheet from the C-terminal end of UBC fold. Although UBC domain is highly conserved with respect to the helices and the sheets, the presence of loops displaying high levels of variability in sequence and length make the E2s conformationally distinct from each other. These loop regions are relatively flexible and are engaged in selection and binding of specific E3s (Burroughs et al., 2008b). E1 and E3 use partly overlapping surfaces, as depicted in **Fig. 1.7**, as they interact with the N-terminal helix or H1 as well as

region of L1 and L2 of E2. E2 must be free from E3 while being loaded with ubiquitin, since E1 and E3 share the same surface on the interaction site of E2 (Eletr et al., 2005).

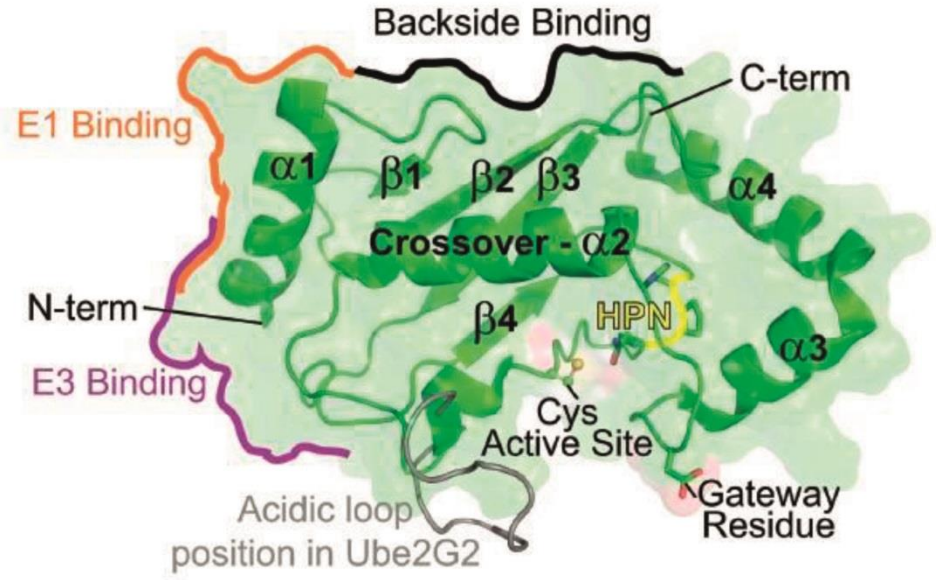


**Fig. 1.7**. **Important structural features of UBC domain (UBE2D3; PDB 2FUH).** The grey loop represents acidic loop of UBE2G2 (PDB 2CYX) on alignment with UBE2D3 (Stewart et al., 2016).

The catalytic cysteine present in a highly conserved loop connecting α helix 2 with β strand 4 and lies a shallow groove formed by the residues from the same loop and α helix 2 and 3 (Burroughs et al., 2008b). Certain residues in the vicinity of catalytic cysteine residue can participate in the transthioesterification reaction. A group of three residues make the HPN motif in which histidine plays a crucial role in stabilizing the structure of E2 active site (Cook and Shaw, 2012). During E2-E3 interaction aspargine interacts with the catalytic cysteine active site of E2 and stabilizes the oxyanion transition state (Wu et al., 2003) (**Fig. 1.8**). Residues surrounding the active site of E2 have been shown to influence the donor and acceptor in the ubiquitination reaction and play deterministic role in the chain formation. For example, UBE2D1 mostly leads to formation of K-11 linked chains, but mutation of a single residue from serine to alanine (Ser83Ala) favours the formation of K63

linked chains (Bosanac et al., 2011). Therefore the UBC domain not only provides the conformation best suited for covalent interaction with ubiquitin but also plays a role in selecting the lysine residue for building ubiquitin chain on the monoubiquitinated substrate.

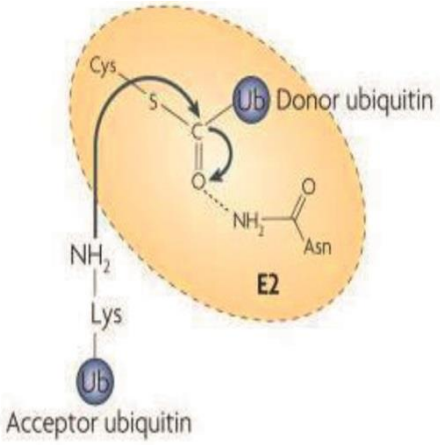


**Fig. 1.8**. **Model showcasing isopeptide bond formation during ubiquitin chain formation.** The side chain of a conserved Asn residue in the E2 interacts with the active-site Cys (carrying the donor ubiquitin), which stabilizes the oxyanion transition state of the nucleophilic attack by the Lys residue of the acceptor ubiquitin (Ye and Rape, 2009).

#### 1.4.1.2. UBC domain of all E2 enzymes: Class I enzymes with only UBC domain and domains flanking UBC domain in other E2s

**Class 1** E2 enzymes are unique from the rest as they do not possess additional domains shouldering UBC to assist in processes apart from ubiquitin conjugation. Class 1 enzymes of E2 essentially comprises of the core domain only as described in section 1.4.1.1. This core domain is the seat of transthioesterification reaction, wherein ubiquitin reacts with E2 through covalent interaction. However, it was found that the backside of the catalytic site has residues that play a key role in the formation of non-covalent interaction with another molecule of ubiquitin. This region is also called as ubiquitin interacting domain and has been hypothesized to play a role in building polyubiquitin molecules as a chain on the target substrate

(Schumacher et al., 2013) A non-covalent interaction between ubiquitin and specific residues on β sheet of E2, which is distinct from the catalytic site residues, assists in chain formation.

UBC4 and UBC5 of *Saccharomyces cerevisiae* belong to this class and are individually dispensable enzymes (**Fig.1.9**). These are responsible for selective degradation of abnormal proteins. Studies reveal that deletion of this domain leads to impaired cell growth, inviability at elevated temperatures or in the presence of amino acid analogs (Seufert and Jentsch, 1990).

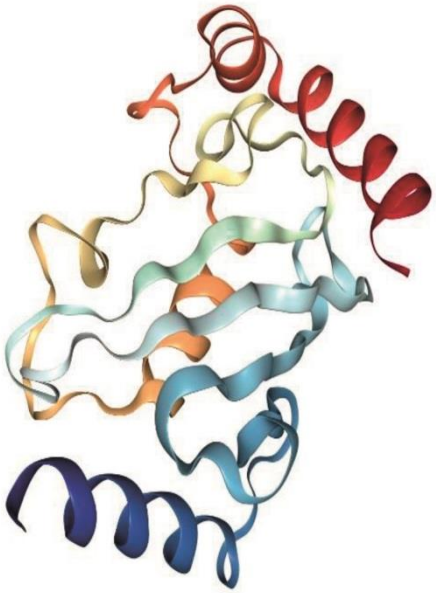

**Fig. 1.9. Cartoon view of UBC4 (PDB 1QCQ).**

### 1.4.1.3. C-Terminal Domain of Class II E2 enzymes

Free ubiquitin chains can be synthesized without the assistance from E3 ligases. The best example studied belongs to class II type E2s namely, E2-25k from mammalian source that can assemble polyubiquitin chain on Lys 48 of ubiquitin (Chen and Pickart, 1990). It consists of C-terminal tail, which forms the ubiquitin associated domain or UBA domain. The UBA domain is composed of a 3 helix bundle (Wilson et al., 2009). It is well conserved across mammalian species, but not in non-mammalian homologs. The highly conserved MGF loop of UBA domain forms the ubiquitin-binding surface. Majority of residues in the core domain are hydrophobic in nature (Wilson et al., 2009). The divergence in rest of the residues is responsible for specificity of ubiquitin binding (Wilson et al., 2011).

In *Saccharomyces cerevisiae*, UBC1 (**Fig. 1.10**) is the homolog of E2-25K. UBA is involved in protein-protein interactions via ubiquitin binding molecules (Dikic et al., 2009). UBA can bind to a second ubiquitin molecule in a non-covalent manner (Merkley and Shaw, 2004). Initial studies reported that UBA domain plays a critical role in polyUb chain synthesis (Haldeman et al., 1997). Deletion of the UBA domain affected the length of polyUb chain. Full length UBC1 is able to form polyUb which is of 10-12 molecules; while in presence of UBA a shorter polyUb chain is synthesised consisiting of only 4 ubiquitin molecules (Merkley et al., 2005). The direct involvement of UBA domain in regulating functions of the protein still remains unclear.

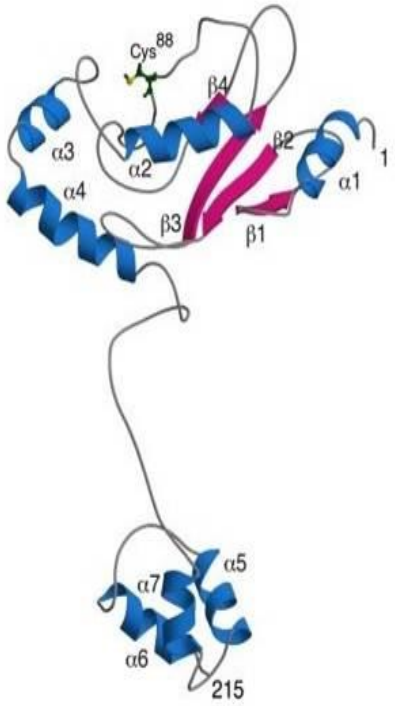


**Fig. 1.10. Ribbon diagram of UBC1 (Merkley et al., 2005).**

#### 1.4.1.4. N-terminal domain of Class III E2 enzymes

One of the intriguing questions about polyubiquitin chain building on substrate is how the entire process is regulated. Studies with UBE2E gave an interesting detail of the process. UBE2E (UBE2E1, UBE2E2 and UBE2E3) is an important member of class III family of E2s. Presence of N-terminal domain flanking UBC domain represents class III set of E2s. Like all other E2s, UBE2E plays the central role of a mediator in ubiquitination of the substrate by transferring ubiquitin to its cognate pair E3. Through the use of deletion studies it was shown that the core domain of UBE2E can build polyubiquitin chains but when the the N-terminal domain is intact it prevents chain building process, driving the enzyme preferentially to monoubiquitinylates substrate (Schumacher et al., 2013). Another study was conducted in which the N-terminal of UBE2E1 was fused to UBE2D2 (which contains only the core domain and is involved extensively in chain building process). It was observed that the resultant chimeric protein limited the

function of chain building. It has been suggested that the presence of intrinsically disordered residues at the N-terminal extension prevent donor ubiquitin at the active site of E2 from accessing the acceptor ubiquitin on the substrate (Schumacher et al., 2013). The UBC domain shares ~92% homology amongst the UBE2E family and mostly resembles other E2s as well (Sheng et al., 2012). However the difference in UBE2E members lies in the extension of the N-terminal which is between 37 and 52 amino acid in length. Besides, the N-terminal of UBE2E members have intrinsically disordered residues in certain stretches (Dunker et al., 2001). The N-terminal extension of UBE2E family has a number of serine residues in the intrinsically disordered stretches. These serine residues are often subjected to post-translational modifications to regulate the activity of the enzyme. Therefore, the main roles of N-terminus in the E2s are mainly incorporating the first ubiquitin, regulating the chain length of ubiquitination activity by limiting the polyubiquitin chain formation and recognizing the right substrate with highest fidelity.

#### 1.4.1.5. Class IV E2s with both N- and C-terminal domains

These are larger E2s when compared to other classes as they carry extensions on either side of the core domain. Reticulocyte maturation and haematopoiesis are governed by a large 1292 aa enzyme E2-230K (UBE2O) (Klemperer et al., 1989). This E2 can ubiquitinate endogenous substrates without the assistance from E3. Another gigantic E2 is apollon (BIRC6) having a size of 528kDa comprises of an N-terminal baculoviral IAP (inhibitor of apoptosis) repeat (BIR) domain and a C-terminal UBC fold. Utilizing its BIR and UBC domains, apollon can promote ubiquitination and proteosomal degradation of proapoptotic protein SMAC and inhibition of caspase-9 (Hao et al., 2004). It has been shown that apollon is involved in the midbody clearance during cytokinesis (Pohl and Jentsch, 2008). However, it has not yet been established how apollon as an enzyme with E2 activity is involved with such a function. One of the hypotheses is that it can recognize and ubiquitinate substrates independently and it does not require any E3 (Bartke et al., 2004).

### 1.4.2. Ubiquitin enzyme variants (UEVs)

These is another class of E2s which possesses UBC fold but lacks the active site cysteine and hence cannot directly participate in ubiquitin coupling (Hofmann and Pickart, 1999). Two of the members of this class are UEV1A (UBE2V1) and MMS2 (UBE2V2) on which extensive study has been done. They are known to form dimers with UBC13 (UBE2N) and are involved in forming lysine 63 ubiquitin chain on substrates. Interaction of UBC13 (UBE2N) with either UEV1A (UBE2V1) or MMS2 (UBE2V2) leads to variations in K63-linked ubiquitin chains (VanDemark et al., 2001). UBC13 (UBE2N)-UEV1A (UBE2V1) dimer is involved in NFκB activation while UBC13-MMS2 (UBE2V2) dimer catalyze K63 chain during DNA damage response (Andersen et al., 2005).

## 1.5. Models for ubiquitin chain assembly

With the discovery of ubiquitin pathway, only an overview of the roles played by the key enzymes has been established. However, some of the functional details of the enzymes involved in delivering ubiquitin to the substrate remain to be explored. One such fundamental role was to understand the chain building process. However with many ingenious experiments, scientists could now understand the importance of E2s in determining linkage specificity and length of ubiquitin chains that can strongly influence processivity of chain formation. The ubiquitin chain assemblage is commenced by transfer of first ubiquitin to Lys residue on substrate. Successively the E2-E3 pair switches to chain elongation, during which additional ubiquitin molecules are attached to the substrate bound ubiquitin. The choice is often made by the E2, whether a Lys residue in the substrate or ubiquitin will receive the next ubiquitin or not. Also it seems between pairs of E2s a work trend (i.e. division of labour) wherein a member of this couple is involved in chain initiation and the other takes up the elongation step (Rodrigo-Brenni and Morgan, 2007).One of the E2 pairs that represent a combinatorial function is UBC1 and UBC4 in Lys48 pathway of modification

of a substrate. UBC4 works as a chain initiator while UBC1 extends the ubiquitin chain. The Lys linkage specificity comes from E2s involved in chain elongation as they interact with substrate bound to ubiquitin. Example, UBE2S forms Lys-11 linked chains, UBE2K prefers Lys48 linked chains and UBE2N-UBE2V1 for Lys63 linked chains (Haldeman et al., 1997; Hofmann and Pickart, 1999; VanDemark et al., 2001). By contrast, other E2s are more selective in promoting ubiquitin chain initiation required for monoubiquitination. Example,UBE2T ubiquitinates specific Lys residue on its substrate FANCD2 but lacks any ubiquitin chain extension activity and does not co-operate with chain elongating E2s (Alpi et al., 2008).

On contrary to forming separate steps of chain initiation and elongation, certain E2s can form short chains of ubiquitin on the substrate followed by the chain elongation step by another E2. This helps in increasing the rate of ubiquitin chain formation. For example, UBE2C initiates formation of Lys11-linked ubiquitin chains on the substrate with the assistance from E3 APC/C promoting the degradation of a large number of substrates under a short stipulated time in mitosis (Rape et al., 2006).

There are few E2s that catalyze both the initiation and elongation of specific ubiquitin chains. The yeast E2 Cdc34 cooperates with the E3 SCF to add Lys48-linked ubiquitin chains to cell cycle inhibitor subunit of cyclin-dependent kinase 1 (Sic1), triggering Sic1 degradation and entry of cells into S phase (Verma et al., 1997)

The mechanism of chain building by three enzymes is yet to be unraveled but some of the models have been proposed to understand the cryptic polyubiquitation process. Each of the models requires experimental proof to state its plausibility which would then enable us to study the various forces imposed by the enzymes to enable ubiquitin transit from a relay of E1-E2-E3 to the targeted substrate. One of the significant points to consider in ubiquitin chain assembly mechanism is that E1 and E3 binding sites on E2 overlap, and their binding to E2 is mutually exclusive (Eletr et al., 2005). Thus, if sequential mechanism leads to building of ubiquitin chains on substrates, multiple cycles of E2-E3 binding and release are probably

necessary. Hypothetically, in E2-E3 complex if a stable E2 dimer is present and if E1 and E3 bind to different monomers, full release of E3 from E2 may be avoided, hence increasing processivity of chain formation. Other mechanisms may also require several rounds of cyclic release of E2 from E3 and are briefly discussed below (Hochstrasser, 2006).

### 1.5.1. Sequential model

Polyubiquitin chain formation can be described as an ensuing cycle of monoubiquitination steps. However, one of the main issues in not complying with this model is that the distal ubiquitin on the growing chain of the tagged substrate cannot act as an attacking nucleophile for ubiquitin to dissociate from a second molecule of activated E2~Ub because it is structurally remote. This model may hold true unless there is some mechanism that could explain looping out action by this growing ubiquitin chain between the substrate and E2 binding sites on the E3. E1 and E3 binding sites on E2 overlap therefore such looping out mechanism would require E2 to re-establish contact in each cycle for the transfer of ubiquitin to the chain. There are several variations to the standard model that have been put forward. One of the hypothesis is the "hit and run" theory which states that the E2 Ub conjugate dissociates from E3 and diffuses to the growing chain on the substrate (Deffenbaugh et al., 2003). This idea comes from the observation of E2 (Cdc34) which is released from E3 (SCFCdc34) for efficient polyubiquitination to occur. Another hypothesis to the sequential model proposes the presence of additional scaffolding like proteins that help in the addition of ubiquitin to the substrate called as E4s (Deffenbaugh et al., 2003). The exact mechanism of E4s contributing in the enhancement of chain assembly is still unclear but nevertheless these proteins are believed to function like E3s. The E4 proteins contain the U-box which shares the tertiary structure of the RING domain in E3s hence it can be concluded that E4s are nothing but E3 that assist in transferring ubiquitin from E2 to the desired substrate (Ohi et al., 2003). Another alternative theory to the sequential model was the formation of unanchored ubiquitin chains de novo before being plugged by the trio enzymes in the final transfer to the substrate

and this was supported by the fact that free ubiquitin chains were observed in the cell (Chen et al., 1991; van Nocker and Vierstra, 1993).

### 1.5.2. Indexation model

This model addresses the limitation of the chain length achieved in the polyubiquitnation process and not necessarily explains how the actual process occurs. This theory was designed on observing the structure of E3 which comprises of two lobes connected by a flexible linker. The building of ubiquitin chain occurs in the HECT domain of E3 and a flexible hinge between the two lobes allows repositioning of the lysine residue present in the distal ubiquitin on the chain that can attack ubiquitin conjugated to E2 (Verdecia et al., 2003). The chain is indexed to a limited length because of the physical constraints imposed by E3 structure.

### 1.5.3. Seesaw model

Seesaw model of polyubiquitin chain formation proposes that a pair of E2s, partner either as homodimers or heterodimers, pass the growing chain between their two active site cysteines, and build it to required size, before finally transferring to the substrate. Chain transfer is caused by the nucleophilic attack on the thioester-linked carbonyl of one ubiquitin by a lysine side chain of the other thioester linked ubiquitin. Another possible step in the cycle would be in the transfer of extended ubiquitin chain back to the other E2 cysteine by transthioesterification. This would mean that only one E2 of the dimer would be charged with E1 in all the cycles of the transfer and at some point a substrate lysine will attack the thioester at the base of the chain, resulting in substrate polyubiquitination. The latest ubiquitin added is always at the base of the chain rather than the distal end which can be tested experimentally (Hochstrasser, 2006).

### 1.5.4. Hybrid model

In this scheme, there is requirement for a second, non-covalent ubiquitin binding site, which is readily supplied by a protein such as E2 or E3. According to the hybrid model ubiquitin chain synthesis occurs by possible release and rebinding cycles and not by the sequential addition mechanism. The noncovalent ubiquitin binding site is acting much like a RING E3 to positioning ubiquitin to facilitate nucleophilic attack on the ubiquitin~E2 thioester. Besides, unlike in the indexation and seesaw models, the polyubiquitin chain remains linked to enzyme active site till it is transferred to substrate (**Fig. 1.11**). Several results are in support of the hybrid model. For instance, ubiquitin-chain assembly by UBC13-UEV heterodimers involves the positioning of a non-covalently bound ubiquitin for attack on a thioester-linked ubiquitin. The UBA domains of E2s such as those of E2-25K and UBC1, which are known to be unessential for ubiquitin-chain elongation, might help to transfer back the ubiquitin chain onto the E1 active-site cysteine, from which the chain can be once again transferred to E2 and finally onto the substrate. Ubiquitin binding domains are also present in some E3s but their functional significance for polyubiquitin-chain synthesis is unclear (Hochstrasser, 2006).

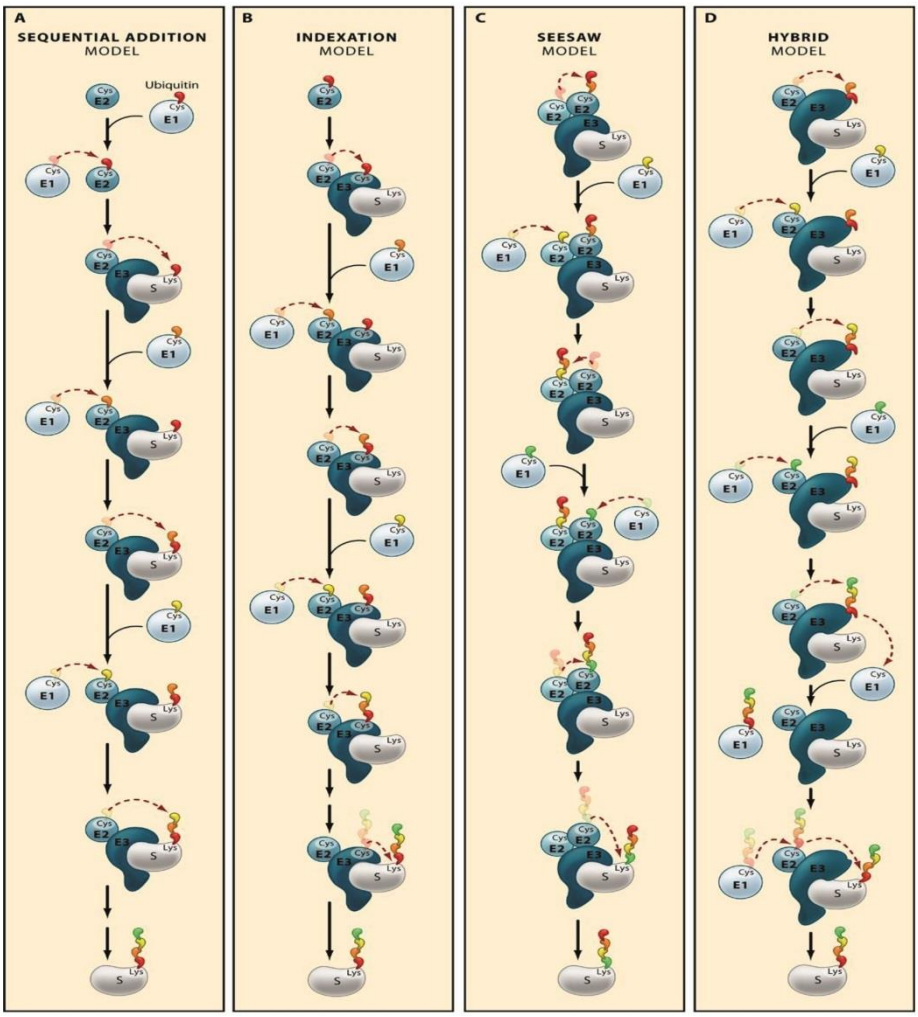


**Fig. 1.11. Models of polyubiquitin chain synthesis (Hochstrasser, 2006).**

### 1.6. Significance of ubiquitination

Understanding ubiquitin chain building processes from these models clarify the processivity that is engineered by the enzymes for chain formation**.** As defined by Hochstrasser "The processivity of ubiquitination is defined as the number of ubiquitin molecules transferred to the growing chain during a single round of substrate association with an E3. This can be determined by the affinity of a substrate for its E3 (that is how long the substrate remains bound to the E3 and therefore able to receive ubiquitin) and by the rate at which ubiquitin transfer is catalyzed by E2 (that is how fast ubiquitin is transferred during the time a substrate is bound to an E3)" (Hochstrasser, 2006). Higher processivity of chain assembly indicates the greater likelihood

that a substrate will receive ubiquitin chain that is long enough to be recognized by the subsequent substrate receptor. The processivity of ubiquitination reaction have a bearing on the biological magnitudes of the downstream effects. For example, timing of substrate degradation during mitosis required for cell cycle headway can be determined by the differences in the processivity of chain formation of APC/C substrate (Jin et al., 2008; Rape et al., 2006). By analyzing various models of chain formation, E2s can regulate chain processivity through several strategies. E2s can recognize specific motifs on substrate for rapid ubiquitin chain initiation, oligomerization of charged E2s, or preassembly of ubiquitin chains on E2 followed by *en bloc* transfer to increase processivity. Therefore, E2s are regulated in order to fine tune ubiquitin chain formation in response to upstream signals in cell physiology (Ye and Rape, 2009).

### 1.7. Physiological significance of E2s

E2 enzymes boost proteolysis by coupling multi-Ub chain assembly to proteolysis. Many E2s remain associated with proteasomes (Prabha et al., 2012; Tongaonkar et al., 2000). This offers a novel insight into the poorly understood mechanism of substrate translocation to the proteasome in vivo (Tongaonkar et al., 2000).

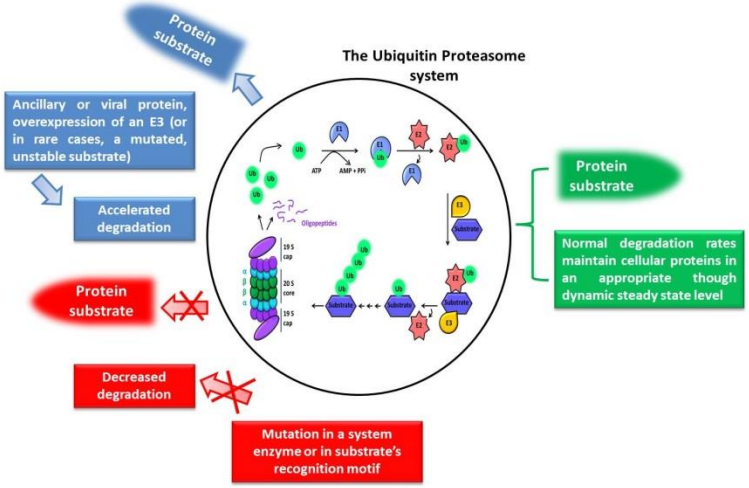


**Fig. 1.12. Regulation of protein levels by ubiquitination.**

Maintenance of protein level or proteostasis is very important aspect of survival of each and every cell (**Fig. 1.12**). Proteins of regulatory nature and those acting cell switches after completing their function undergo degradation. If there is any defect in ubiquitination machinery, it will lead to the accumulation of such proteins in the cell, leading to apoptosis, cancer and other diseased conditions. One of them is Alzheimer's disease (AD) which is a common cause of dementia and loss of cognitive function. The pathology behind AD is believed to be an outcome of accumulation of amyloid β peptide (Aβ) resulting in neurotic plaques (Small et al., 2001).

E2-25K functions like other E2s in ubiquitin conjugation but it also shows unusual ubiquitin ligase function, forming diubiquitin and polyubiquitin chains independently without any assistance from E3s (Chen and Pickart, 1990). Biochemical analysis suggested that E2-25K or HIP-2 is an unusual E2 with increased expression levels in AD patients. This ubiquitin ligase function is dependent on the core domain but the site of function was found in the 151-200 residues of C-terminal extension (Haldeman et al., 1997). On mutational analysis, it was found that the domain responsible for ligase activity in E2-25K played a role in AD pathogenesis. mutants E2-25K/Hip 2 lacking ubiquitin ligase activity due to deletion of C-terminal tail and a point mutation Ser86Tyr (Haldeman et al., 1997; Mastrandrea et al., 1998) responsible for the formation of unanchored polyubiquitin chains, reduced Aβ neurotoxicity. These results suggested that E2-25K indeed played a role in neurotoxicity.

A frameshift mutant of ubiquitin, ubiquitin-B ($UBB^{+1}$) is found in patients suffering from AD (Hilbich et al., 1993). It is composed of a 75 residues Ub moiety with a 19-residue C-terminal extension (van Leeuwen et al., 1998). The genes from which $UBB^{+1}$ mRNAs are transcribed contain numerous GAGAG motifs. Abnormal C-terminal sequence is the product of dinucleotide deletions (ΔGA) from within the GAGAG motif (van Leeuwen et al., 1998). Normally, β-amyloid precursor protein (β-APP) and $UBB^{+1}$ molecules are degraded by the 26S proteasome (van Leeuwen et al., 1998; van Leeuwen et al., 2006). However, in patients suffering from AD, $UBB^{+1}$ and Ub are found in the aggregation plaques of β-APP, which is indicative of UPS

dysfunction (Fischer et al., 2003; Oddo, 2008). Though, at basal level of expression $UBB^{+1}$ can be removed by the UPS (van Tijn et al., 2007), its increased levels of expression inhibits the 26S proteasome, resulting in the accumulation of aberrant proteins. Due to the absence of a C-terminal tail terminating in residue G76 as in normal Ub, the C-terminal of $UBB^{+1}$ cannot undergo activation and subsequent ligation to substrates. As an alternative, $UBB^{+1}$ serves as a scaffold for ligation of Ub molecules to produce polyUb that is anchored to $UBB^{+1}$ (on the unaffected K48 site). Therefore, when $UBB^{+1}$-anchored polyUb is targeted to the 26S proteasome, it acts as a functional antagonist, inhibiting the activity of the proteasome, leading to Aβ neurotoxicity (Song et al., 2003).

Several neurological diseases are caused by an expansion of polyglutamine repeats such as spinocerebellar ataxias and Huntington's disease. Expansion of the repeat above a critical length results in severe neurodegradation associated with neuronal intranuclear inclusions (NII) in the affected areas of the brain. Several studies point towards impairment of UPS as cause of pathogenesis of the disease, NIIs incorporate aggregated proteins that are not degraded in spite of ubiquitination. E2-25K or Hip-2 has been reported to be expressed in high levels in the affected areas of the brain in HD (de Pril et al., 2007). E2-25K via its UBA domain directly interacts with huntingtin and mediates the aggregation and toxicity of expanded huntingtin, resulting in cell death. It was suggested that E2-25K might be involved in aggregate formation and induced cell death.

Recent study has shown that aging increases levels of UBE2N which is linked to the aggregation of huntingtin protein in the synaposome and decrease in UPS functioning. Overexpression of UBE2N increases mutant huntingtin aggregation while reducing UBE2N decreases the aggregation of the protein as seen in mouse and culture models (Li and Li, 2011). As the functioning of the UPS decreases with age and there is an increase in UBE2N as mentioned above, both the factors together contribute to accumulation of mutant protein in the synaptosomes.

Thus, overall reduction in the functioning of the Ubiquitin proteasome pathway with ageing is a factor to be considered and while studying the above diseases it was observed in some cases that even though the UPS enzymes work properly, the proteasome does not, leading to the aggregation.

The edible mushroom *Volvariella volvacea* undergoes cryogenic autolysis. This was attributed to the E2 enzymes as the cold induced gene expression profiles in the mushroom showed significant increase in the UBE2 enzyme (Gong et al., 2016). Further studies showed that UBE2V, a type of UBE2 is up regulated, which was confirmed by qPCR analysis. Significant up regulation of UBE2V at low temperatures induces ubiquitination counterbalancing the regulated processes of the cell.

Ever since the E2s have been discovered to be the crucial players of ubiquitination cascade, researchers are excavating the process for better understanding on the various aspects of E2s like structure and folding, catalytic interactions and their biological significance. We are now entering a phase of understanding on how an E2 enzyme is selected for a given physiological context. This knowledge helps us in bridging the gap between understanding the biological significance of E2s and their association with the diseases. The spectrum of diseases associated with E2s is broad, ranging from immunological disorders, neurological diseases to cancer in mammals to disorders like tomato bushy stunt in plants.

With the modern advancements in scientific tools, it is now possible to study the various E2s and their interaction with cognate E3 ligases in the true biological context and help in developing better understanding on their mechanisms of action. The results from such studies could pave way for discovery of new therapeutic targets for diseases and eventually their cure.

## 1.8. Rationale for selecting yeast ubiquitin conjugating enzymes as the research problem and yeast as model system

*Saccharomyces cerevisiae* is the simplest eukaryotic organism which shares many mechanistic similarities with humans in many vital cellular processes (Karathia et al., 2011). It is therefore an important model organism to study and understand basic molecular processes in humans. Baker's or budding yeast (*Saccharomyces cerevisiae*) has long been a popular model organism for basic biological research. In the laboratory it is easy to manipulate, can cope with a wide range of environmental conditions and controls cell division in a similar way to our cells. Yeast chromosomes share a number of important features with human chromosomes. Being a eukaryote, yeast shares many functional features at molecular and cellular levels with higher eukaryotes, which include the UPS system. Yeast cells divide in a manner similar to human cells, but at a much faster rate with their doubling times in the range of 2-3 hrs. Hence, they provide a convenient host system for eukaryotic protein expression and purification. In addition, they are amenable to genetic manipulations like site directed and random mutagenesis. Thus, for understanding the intricacies of structure and function of ubiquitin conjugating enzyme, yeast had been selected as a model system for the present studies.

## 1.9. Brief introduction to chapters

Ubiquitination is carried out by a set of three enzymes, ubiquitin activating enzyme (E1), ubiquitin conjugating enzymes (E2s) and ubiquitin ligases (E3s). Free ubiquitin is activated by ubiquitin activating enzyme in the presence of ATP and transferred to ubiquitin conjugating enzyme. Ubiquitin conjugating enzyme further transfers it to either ubiquitin ligase carrying the substrate or directly to the substrate which is in association with ubiquitin ligase. The pairing of E2 and E3 is specific, which makes E2s as key mediators in building the polyubiquitin chains on a substrate. Thirteen E2s are present in *S. cerevisiae,* and nearly forty in humans.

**Chapter 1** presents a general introduction on ubiquitination and detailed review on structural and functional aspects of ubiquitin conjugating enzymes. It brings out the intricacies of structure of E2 enzymes and the details of their structure based classification. Interestingly, in certain cases it was shown that more than one E2 enzyme is involved in building polyubiquitin chains, as they carry out specific steps. Classification of E2s has been discussed on the basis of their structure and how the domains are used by different E2 enzymes differently to achieve ubiquitin conjugation. Various models proposed to explain the mechanisms adopted by E2 enzymes to build polyubiquitin chains on a substrate have been discussed in the wake of why each class of E2 takes a different strategy in doing the same function. Further, the direct and indirect roles played by E2 enzymes in the various diseases is also discussed. The chapter also introduces the main objectives of the work and their chapter-wise presentation.

As mentioned in section 1.4.1.3 and 1.7, the class II E2 enzyme UBC1 from *S. cerevisiae* is a flexible two-domain protein comprising of an N-terminal catalytic domain and a C-terminal UBA domain. E2-25k is a human homolog of yeast UBC1. E2-25k interacts with $UBB^{+1}$ which is a frame shift mutant of ubiquitin and is presumed to be one of the contributing factors to neurotoxic disorders. UBC1 can synthesize Lys48 linked free polyubiquitin chains in the absence of any E3 ligase. Sequence alignment between UBC1 and E2-25K protein sequences by Clustal Omega tool available online at https://www.ebi.ac.uk/Tools/msa/clustalo/, revealed that the catalytic N and C-terminal domains were similar in sequence with a major exception lying only in the linker region. The UBC1 linker is composed of 22 amino acids, while that of E2-25k has 6 amino acids. The reasons for the striking structural difference between these two proteins were investigated and the results were presented in **Chapter 2.** Chimeric protein (c-UBC1) was constructed by swapping UBC1 linker with that of E2-25k, to understand the importance of the length of linker, and the mutant proteins effect over growth, viability, polyubiquitin chain formation, thermotolerance and survival in the face of antibiotic and heat stresses were studied. Structure of the mutant protein c-UBC1 was characterized by CD and fluorescence spectroscopic techniques.

(Varsha Raimalani; Brinda Pamchamia and C. Ratna Prabha, Manuscript under preparation).

**Chapter 3** addresses the structural differences in two of the class I E2 enzymes of S. *cerevisiae* namely, UBC4 and UBC5 in relation to their functional importance. The genes of UBC4 and UBC5 encode almost identical proteins (92% identical residues) of 16 kDa molecular weight. Interestingly, UBC4 shares some of its functions with UBC5. However, most of UBC5 functions are identical to those of UBC4, making UBC5 functionally redundant. Furthermore, deletion of either gene can be complemented by the other. However, deletion of both the genes results in slow growth and loss of viability at elevated temperatures. In addition to the polyubiquitin gene *UBI4*, UBC4 and UBC5 are essential modules of a dynamic pathway of eukaryotic stress response. Among E2s, the enzymes UBC4, UBC5 and UBC1 constitute a sub-family, which is required for cell growth and viability. Single genes are dispensable, but mutant cells where all three genes are deleted are not viable (Seufert et al., 1990). This proves the important role played by these enzymes in ubiquitin-dependent protein degradation in eukaryotic cells. Although these enzymes are partially complementing in function, each one has its specific role. Hence, with the objective of understanding the structural and functional reasons for sequence differences it was decided that the 11 residues of UBC4 will be replaced one by one with those of UBC5. Subsequently, the resultant single mutants of UBC4 were characterized for the effects of mutations on growth, tolerance to heat and antibiotic stresses. Stability of mutant proteins was evaluated both theoretically and experimentally (Raimalani et al., 2019).