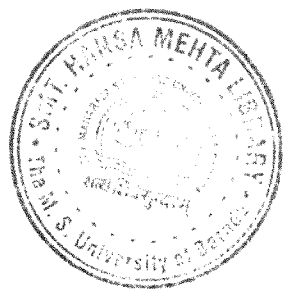

CHAPTER 1

GENERAL INTRODUCTION

1.1. PROTEINS: The sophisticated masters of cell

Proteins are very important molecules found in living organisms. They are the most abundant components within a cell making up more than half its dry weight. They have a range of indispensable roles as enzymes, regulatory proteins, transporters, storage proteins, contractile and motile proteins, defense proteins and structural proteins. In other words, proteins are responsible for carrying out almost every vital function in a cell. To carry out their tasks, proteins must fold into a complex and unique three-dimensional structure referred to as the native conformation. The information essential for a protein to attain its native conformation, that is its biologically active conformation, is determined by the stereo-chemical code present in the form of its amino acid sequence (Anfinsen, 1973 and 1975). But the question how a nascent synthesized polypeptide chain or protein is guided by the stereo-chemical code to its native state has been confounding. Further, there is an indispensable relationship between native structure of the protein and its function.

1.2. PROTEIN ARCHITECTURE

1.2.1. The amino acids as the building blocks of proteins

Amino acids are the chemical constituents of proteins. The structure of an amino acid has the following features: a central carbon atom, often referred to as C_α, is linked to an amino group, a carboxyl group, a hydrogen atom and a side chain that gives rise to the chemical variety of the amino acids.

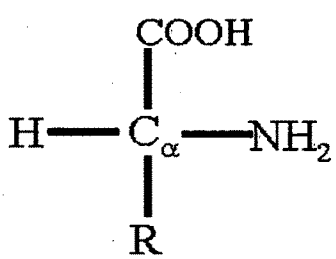

Figure 1.1. General structural formula of α-amino acid

The 'C_α' carbon is asymmetric in all amino acids except in glycine, which has only a proton as its side chain. All naturally occurring proteins have L-isomers of amino acids, exceptions being some of the insect poisons venoms and toxins which contain D-isomers at some places. The choice of L-isomer by early prebiotic systems as a building block during chemical evolution could be an accident which later on became fixed and self perpetuating.

As mentioned in section 1.2.1 side chains of amino acids have a wide chemical variety, which is important for the unique functions of proteins. These side chains can be

grouped into four categories: non-polar aliphatic, aromatic, polar uncharged and polar charged (positive or negative) amino acids.

1.2.1.1. Nonpolar aliphatic:

The simplest amino acids like glycine. alanine, valine, leucine, isoleucine, proline and methionine whose side chains are entirely aliphatic and hydrophobic belong to this class. In proteins alanine, valine, leucine and isoleucine tend to cluster together forming the core and help in stabilizing the structure by means of hydrophobic interactions. Proline contains an aliphatic side chain that is covalently bonded to α-amino group, forming an imide bond. Presence of proline introduces a kink in the polypeptide reducing its structural flexibility. Glycine, owing to its smallest side chain, makes no real contribution to hydrophobic interactions but is involved in the formation of turns because of least steric hindrance posed by the latter. Methionine, one of the two sulfur containing amino acids, has a nonpolar thioether group in its side chain. These amino acids have low solubility in water because they can form only van der Waals interactions with water molecules (Nelson and Cox, 2005).

1.2.1.2. Non-polar aromatic:

The amino acids phenylalanine, tyrosine and tryptophan have hydrophobic aromatic side chains. The hydroxyl group of tyrosine can form hydrogen bonds, and it is an important functional group of many enzymes. Tyrosine and tryptophan are significantly more polar than phenylalanine. Additionally hetero atoms like oxygen and nitrogen in their side chains, give rise to many bonding possibilities (Nelson and Cox, 2005).

1.2.1.3. Polar uncharged:

The uncharged members of this group include: serine, threonine, cysteine, asparagine and glutamine. The side chains of these amino acids are more soluble in water, or more hydrophilic, because they contain functional groups that form hydrogen bonds with water. The polarity of serine or and threonine is due to their hydroxyl groups and that of cysteine is because of its thiol group. In asparagine and glutamine the amide groups impart polarity (Nelson and Cox, 2005).

Cysteine is readily oxidized to form a covalently linked cystine, in which two cysteine molecules or residues are joined by a disulphide bond. The disulphide-linked residues play a special role in the structures of many proteins by forming covalent links within a polypeptide chain or between two different polypeptide chains (Nelson and Cox, 2005).

1.2.1.4. Polar charged amino acids:

The most hydrophilic side chains are those that are either positively or negatively charged. The amino acids in which the side chains having significant positive charge at pH 7.0 are lysine and arginine and those having significant nagative charge at pH 7.0 are aspartate and glutamate. Histidine, has an imidazole group having an ionizable side chain with a pKa near neutrality. In many enzyme catalyzed reactions histidine residue facilitates the reaction by serving as a proton donor or acceptor at physiological pH (Nelson and Cox, 2005).

1.2.2. .Primary, secondary, tertiary structures and quaternary organization

Depending on the nature of the side chain, an amino acid can be hydrophilic or hydrophobic, acidic or basic and it is this diversity in side chain properties that gives each protein its specific character. Thus, all structural and functional properties of proteins are derived from the chemical properties of the polypeptide chain, which in turn depends on the composition and sequence of amino acid it is made up of. There are four levels of structural organization that a protein possesses of: primary, secondary, tertiary and quaternary structures. Primary structure is defined as the linear sequence of amino acids in a polypeptide chain. The secondary structure refers to local hydrogen bonding interactions of the chain along the peptide backbone. Tertiary structure is a consequence of intra-chain and generally long range interactions, which include disulphide bonds, hydrogen bonds, hydrophobic interactions and van der Waals forces. The quaternary structure is present only in oligomeric proteins and it represents the organization of protein subunits.

1.2.3. Primary structure of a protein or polypeptide chain

In order to form a protein or polypeptide chain, amino acids are condensed forming a C-N bond or peptide bond between carboxyl group of one amino acid and the amino group of the next, with a loss of water molecule. The peptide bond angles and lengths are well known from x-ray crystallographic studies of protein and peptide structures. The peptide (C'-N) bond length is observed to be 1.33Å. This is considerably shorter than the adjacent (non-peptide) C_α-N bond length of 1.45Å, but the C'=O bond length of 1.23Å is longer than that of C=O bond of an aldehyde or ketone. These bond lengths and angles reflect the distribution of electrons between atoms due to differences in polarity of the atoms, and the hybridization of their bonding orbitals. The two more electronegative atoms, O and N, can bear partial negative charges, and the two less electronegative atoms, C and H, can bear partial positive charges. The peptide group consisting of these four atoms can be thought of as a resonance structure with partial double bond character, accounting for its intermediate bond length. Like any double bond, rotation about the peptide bond is restricted (Pauling, 1960). The successive Cα atoms occupy opposite corners along most peptide linkages, giving trans conformation to the peptide linkage in between them. The cis arrangement is unfavorable because of the steric hindrance experienced by side chains on adjacent Cα atoms. The angle of rotation between Cα and amino nitrogen denoted by Φ torsion angle, angle of rotation between Cα and C' carbon denoted by Ψ torsion angle provide structural flexibility to the polypeptide chain despite the restriction to rotation along the peptide bond (Ramachandran, Ramakrishnan and Sasisekharan, 1963).

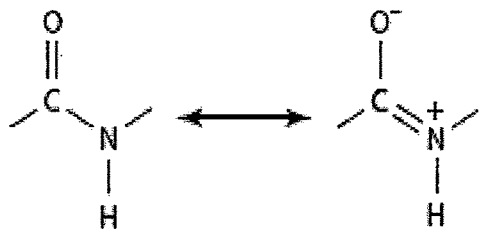

Figure 1.2. Resonance interaction of peptide bond

Peptide bond formation is not spontaneous nor is the reverse reaction, involving hydrolysis of the peptide bond. They can be accomplished chemically, under very vigorous conditions. Thus proteins are chemically and biologically stable. Proteins are heteropolymeric containing most or all of the different amino acids. Any region of a typical protein will therefore have a chemically heterogeneous environment. This heterogeneity is further amplified by the higher levels of protein structure (Nelson and Cox, 2005).

1.2.4. Secondary structures

1.2.4.1. α-helix:

The restrictions in the peptide bond rotations and steric hindrance between side chain and backbone are the origin of secondary structures. The steric restrictions arising within each residue and between residues are sequence dependent phenomena. However, a sequence of residues which have similar allowed space of Φ and Ψ can give rise to a chain segment that forms repetitive structures like α-helix or β-sheet. Thus, these secondary structures owe their formation to both backbone and side chain steric restrictions. The helix structure looks like a spring. The most common shape is a right handed α-helix defined by the repeat length of 3.6 amino acid residues and a rise of 5.4 Å per turn. The pitch and dimensions of the helix also bring the amide proton of $(i+4)^{th}$ residue into proximity to the carbonyl oxygen of i^{th} residue such that a hydrogen bond is formed.

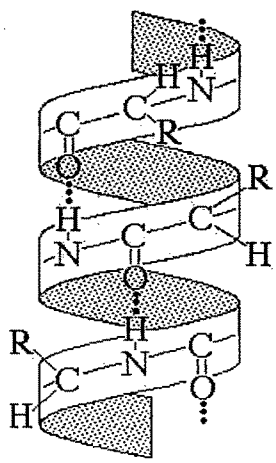

Figure 1.3. The right handed α-helix

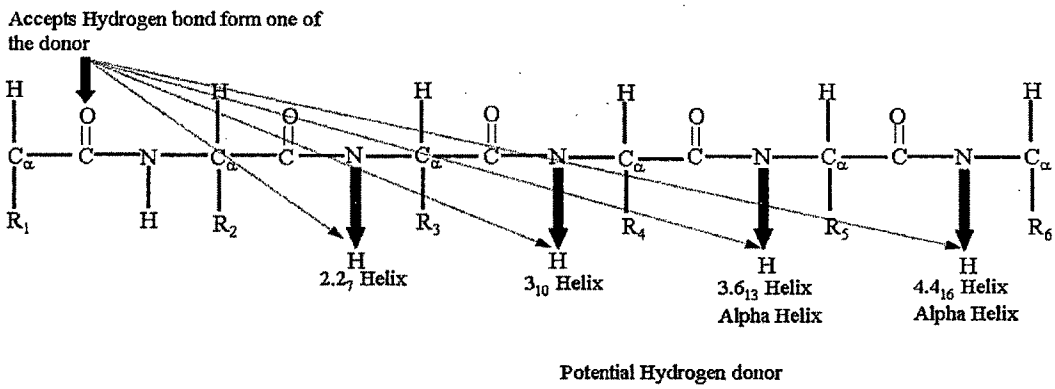

Figure 1.4. The hydrogen bonding pattern of several polypeptide helices

There are other types of helices and depending upon the hydrogen bond interaction these are classified as 2.2_7 helix, 3_{10} helix, the most abundant one 3.6_{13} (α helix) and 4.4_{16} (π helix). All peptide group hydrogen bond donors and acceptors form hydrogen bonds in the central part of the helical segment, but not at the ends. However, Φ and Ψ restrictions can

have the effect of pre-organizing the chain into a helical conformation, which may favor hydrogen bonding by enhancing the local concentration of donors and acceptors (Pauling, Corey and Branson, 1951; Pauling and Corey, 1951).

1.2.4.2. β-sheets:

β-strands are the other regular secondary structures that proteins form. These are extended structures in which successive peptide dipole moments alternate direction along the chain. Because it is an extended structure, Φ and Ψ steric hindrance is reduced in the β-strand (Richardson et al., 1978). Two or more strands can pair by hydrogen bonding and dipolar interactions to form a β-sheet. Unlike helical segments, all peptide group hydrogen bond donors and acceptors are satisfied not within but between β-strand segments. Individual β-strands do not have an independent existence contrasting to a helical segment. β-sheets can consist of either parallel or antiparallel strands, or a mixture of the two. In parallel strand the segments are connected by a loop or helix. In purely antiparallel sheets, segments that are sequentially next to each other in the primary structure often form adjacent strands. The numbers of hydrogen bonds which can exist for a polypeptide of a certain length are more in case of antiparallel than the parallel strands and thus are more stable (Pauling, Corey and Branson, 1951; Pauling and Corey, 1951.

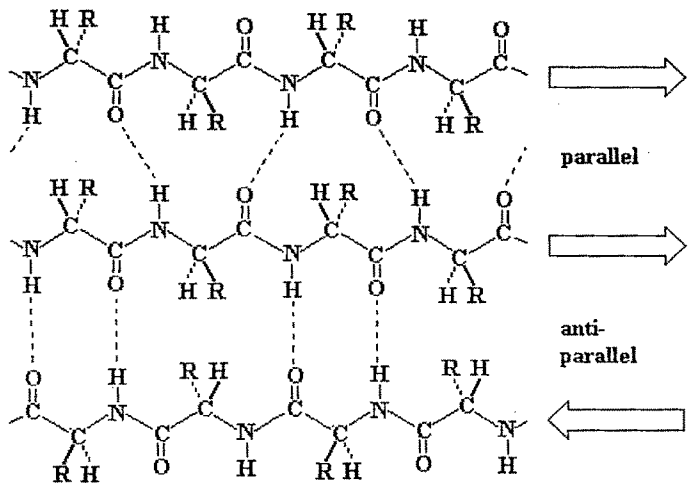

Figure 1.5. β-pleated sheets: parallel and anti-parallel

1.2.4.3. Turns:

Turn structures are also classified as secondary structural elements, but unlike helices and strands, they do not have a repeating geometry (Wilmot and Thornton, 1988). Rather, they can have well-defined spatial dispositions defined by certain values of Φ and Ψ

angles that often require specific residue types and/or sequences, as well as fixed hydrogen bonding patterns. Turns are essential for allowing the polypeptide chain to fold back upon itself to form tertiary interactions. Such interactions are generally long-range and result in compaction of the protein into a globular, often approximately spherical form. The turn regions are thus generally located on the outside of the globular structure, with helices and sheets forming its core. Turns on the surfaces of proteins display a wide range of dynamics,from quite mobile in cases where they form few interactions with the underlying protein surface to the ones quite fixed due to extensive tertiary contacts (White and Poet, 1987).

1.2.4.4. β-bulges

Sometimes an irregular region is found in β sheet involving two or more residues on one strand against one residue in another. As a consequence, two consecutive β type hydrogen bonds include two residues on one strand and a single residue on the other strand giving rise to a structure known as β bulge (White and Poet, 1987). The two residues on the bulged side are labeled **"1"** and **"2"**, whereas the residue on the opposite strand is labelled **"X"** (Richardson et al., 1978). The β-bulges are classified into five types, namely classic, G1, wide, bent and special types. Classic and wide include both parallel and antiparallel β bulges and G1 bulges are mostly antiparallel only. G1 bulges are further classified into: 1) G1G type with glycine at position first position of the bulge, 2) G1T type with glycine at first position and a Type I′ or Type II beta turn between position 2 and any other residue, with glycine at position i+2 of turn and 3) G1A type has any amino acid (excluding glycine) at first position. Bent type bulges have two residues in each of the β-strands and have no X position in this class. This arrangement of residues introduces a bend instead of one residue bulge as in other types of bulges. Special types of bulges include the bulges introduced by proline (Chan et al., 1993).

1.2.5. Tertiary structure

The side chain of amino acid residues project outwards from the secondary structural frame work of a protein, and are therefore available for interactions with other surfaces through hydrophobic contacts and various kinds of bonding interactions to form the tertiary

structure upon compaction. Proteins with highly organized tertiary structures generally have a well-developed core of hydrophobic residues contributed from most or all of the secondary structure elements in the chain. Thus, secondary and tertiary structures are in general closely interconnected. Protein secondary and tertiary structures are not independent of each other (Prabha and Rao, 2004) and are lost in an all or none manner upon changes in environment that disfavor the folded state, such as higher temperature or solvent additives.

1.2.6. Quaternary organization

The highest level of protein structural organization is the quaternary structure, found only in oligomeric proteins. The subunits that associate may or may not be identical and their organization may or may not be symmetric. The interdependence of tertiary and quaternary structures parallels the interdependence between secondary and tertiary structures, and suggest that the distinction among these levels of the protein structure organizational hierarchy are blurred and perhaps even misleading for our understanding of protein structural stability and folding.

1.3. PROTEIN FOLDING AND STABILITY

1.3.1. From polypeptide to a functional native conformation

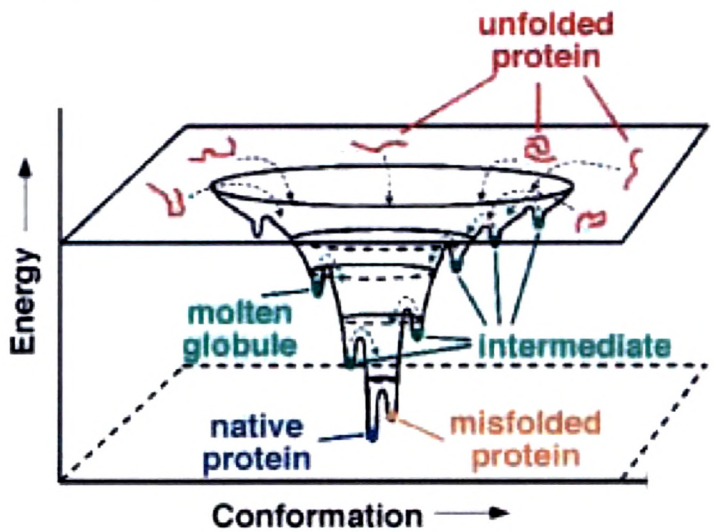

Figure 1.6. Folding funnel representing various state of energy levels

Protein folding obeys the laws of thermodynamics. A protein always folds so that it achieves the lowest possible energy (Anfinsen, 1973; Hagen, 2007). Theoretically a chain of 100 amino acids will take 10^{50} years to attain a thermodynamically stable native state if it has to try out all possible stable conformations one after another (Levinthal, 1968). However in vivo protein folding takes a few minutes. The fast and proper folding of protein in vivo is a shared effect produced by physiochemical environment of cell, assistance of chaperones, the post-translational modifications and various inter and intramolecular interactions. However, there is an ongoing debate over whether protein folding occurs co-translationally or post-translationally. There are evidences supporting both these models. There are proteins like cytochrome c, which require the C-terminal to interact with N-terminal to form the native state. On the other hand it was reported that certain peptides during synthesis are known to react with antibodies providing an evidence for formation of native like conformation during translation (Basharov, 2000). Posttranslational modifications of proteins in vivo also decide the path for proper folding. It is not yet completely understood how an amino acid chain folds into its tertiary structure in the short time scales in a cell. Thermodynamics of protein folding is described by folding funnels and energy landscapes (Dill, 1985; Bryngelson et al., 1987). Protein looks for its native state through a funnel of declining energy. It explores many folding routes till it reaches the completed tertiary structure the sate of lowest energy level. All non-native conformations of a protein possess higher energy than the native conformation which represents the thermodynamically most stable. The impediments present down the folding funnel represent the local energy barriers which trap the intermediate states or the molten globule states of folding.

1.4. MODELS PROPOSED TO EXPLAIN PROTEIN FOLDING

Several models have been proposed to explain the process of protein folding. Some of the prominent models include the Framework model (Karplus and Weaver, 1976), the diffusion collision model (Kim and Baldwin, 1982) and the hydrophobic collapse model (Chan and Dill, 1990).

The framework model proposes formation of secondary structures along the length of the polypeptide backbone in an unfolded protein. These regions of nascent secondary structure represent the initiation sites of protein folding. These sequences have high

propensity for the secondary structures they take up in native protein. Studies with several peptides showed they could attain native like structure in vitro independent of rest of the protein (Wright et al., 1988; Dyson et al., 1988a, 1988b; Waltho et al., 1989; Jaenicke, 1991; Montelione and Scheraga, 1989; McLeish et al., 1993) and some peptides could even display antigenic properties (Feiser., 1987), lending unequivocal support to the above model.

The diffusion collision model proposed that the secondary structures formed in unfolded protein initially are not stable. They are in dynamic equilibrium with unfolded state and hence are often referred to as secondary structural elements. These secondary structural elements diffuse in space and collide with each other. In the process they stabilize by establishing tertiary interactions. The major evidence for this model came from the folding studies of cytochrome c Roder et al. (1988) during their refolding experiments with cyt c observed that the N and C terminal regions form fluctuating helical structures. In this study the secondary structures were stabilized and the protein folded into native state in about 20ms. The time interval taken by the protein to fold was found to be in agreement with the time required for the two helices to diffuse in space and collide with each other theoretically (Bashford et al., 1990). The N and C terminal fragments formed marginally stable secondary structures in isolation. However when put together they formed stable secondary structures along with tertiary interactions (Wu et al., 1993).

The above models are based on formation of secondary structures and secondary structures are stabilized by hydrogen bonding. However, Chan and Dill opined that hydrogen bonding cannot act as a driving force for folding as there would not be any major gain or loss in terms of energy, whether the hydrogen bonds are formed with in the protein or with the surrounding water. Hence, according to hydrophobic collapse model, the non polar side chains of hydrophobic residues move away from water and interact among themselves to form the hydrophobic core of the folded protein. In the collapsed state the protein tries to attain minimum energy by searching for secondary and tertiary structures. This model of folding was successfully demonstrated with barnase (Serrano et al., 1992) and dihydrofolate reductase (Garvey et al., 1989).

1.4.1. Small proteins as model systems to understand the mechanism of protein folding

Small single domain proteins that undergo cooperative and reversible folding reactions have been used as model systems for protein folding studies (Went et al., 2004). Such models are manageable systems, as the folding reaction is not complicated permitting easy interpretation of results. The off-pathway side reactions, such as aggregation can be avoided (Judith, 2001; Baldwin and Rose, 1999). Such studies are usually performed under conditions that minimize off-pathway reactions, such as very high dilution and low temperature (Chothia et al., 1990; Kim et al., 1990). Ubiquitin is a 8.5kDa small protein which is extensively used as a favorite model system for protein folding studies (Recksteiner, 1998; Monia et al., 1990).

1.4.2. Ubiquitin: The lethal tag

In a series of elegant experiments discovered that a heat stable polypeptide was required for an energy-dependent proteolytic system in reticulocytes (Ciechanover et al., 1978; Hershko et al., 1980). This polypeptide subsequently was known as 'ubiquitin'. Subsequent studies helped identify Later on the enzymes that carry out ubiquitin–protein conjugation, termed E1, E2, and E3, were identified and characterized (Ciechanover et al., 1982; Hershko et al., 1983, 2000). The ATP-dependent protease that mediates the destruction of ubiquitin–protein conjugates (Hershko et al., 1984) was characterized by several laboratories in the 1990s, and is now called the 26S proteasome (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).

A large number of studies have confirmed the importance of this protein in the proteolytic system (Bachmair and Varshavsky, 1989) for the breakdown of intracellular proteins during various processes such as transcription, protein synthesis, cell cycle regulation, DNA repair, stress response and involvement in patho-physiological conditions (Ciechanoveret et al., 1984; Finley et al., 1984, 1987, 1989; Ozkaynak et al., 1984, 1987; Bachmair et al., 1986; Jentsch et al., 1987; Goebl et al., 1988; Bachmair and Varshavsky, 1989; Chau et al., 1989; Gonda et al., 1989; Bartel et al., 1990; Hochstrasser and Varshavsky, 1990; Johnson et al., 1990).

1.4.3. Ubiquitin: The conserved gene

Ubiquitin is a small protein found only in eukaryotic organisms and is not found in either eubacteria or archaebacteria and can exist either in free form or as part of a complex with other proteins. Among eukaryotes, ubiquitin protein sequence is highly conserved. (Gavilanes et al., 1982; Watson et al., 1978; Schlesinger et al., 1975; Schlesinger and Goldsteiner, 1975). In yeast (Wilkinson et al., 1986) and oat (Vierstra et al., 1986) replacement of amino acids is seen in only three places in the sequence of ubiquitin. Difference in only three positions from yeast (Wilkinson et al., 1986) to human implies strong sequence conservation and the vast majority of amino acids that make up ubiquitin are essential as apparently any mutations that have occurred over evolutionary history might have been removed by natural selection. The sequences of yeast, plant and human ubiquitins are shown figure 1.7, highlighting the positions of the residues where there are differences.

YEAST UBIQUITIN PROTEIN SEQUENCE

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

PLANT UBIQUITIN PROTEIN SEQUENCE

MQIFVKTLTG KTITLEVESS DTIDNVKAKI QDKEGIPPDQ QRLIFAGKQL
1 10 20 30 40 50
EDGRTLADYN 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 1.7. The amino acid sequences of yeast, plant and human ubiquitins. The residues which were changed have been highlighted.

1.4.4. The structure of ubiquitin

Ubiquitin is a compact, globular protein first isolated and crystallized in 1987 (Vijay kumar et al., 1987). The protein consists of a single polypeptide chain of 76 amino acids

bearing a molecular weight of 8565Da. It is without any cysteines, metal ions or cofactors. The x-ray crystallographic structure of ubiquitin reveals a globular α/β structure with hydrophobic core made up of five strands of β-sheet two parallel strands three antiparallel strands and 3.5 turns of α helix and 3_{10} helix characterized by a particular fold denoted as SSHSSS. The compact structure of ubiquitin has nine reverse turns (Vijay kumar et al., 1987).

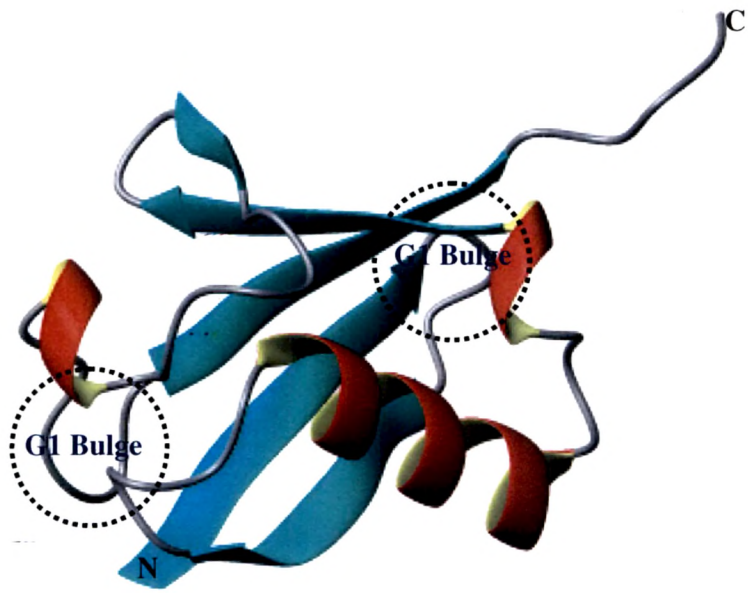

Figure 1.8. Ubiquitin structure

Ubiquitin contains two β-bulges. The first G1 β-bulge is present at the N-terminal end. Present on two antiparallel β-strands, M1 to T7 and G10 to V17 and involves Gly10 (1), Lys11 (2) and Thr7 (N). It is present along with Type I turn. The second G1 β- bulge is present at the C-Terminal end. Present on two parallel β-strands, E64 to R74 and M1 to T7 and involves Glu64 (1), Ser65 (2) and Gln2 (N). It is present along with a type II turn (Vijay kumar et al., 1987). There are nine reverse turns in the protein. Table 1.1. lists the type of turns, the residues involved and their Φ ,Ψ values.

Table 1.1. The details of nine reverse turns found in ubiquitin.

Residues				(Φ and Ψ Angles (Degree)		
I	**I+1**	**I+2**	**i+3**	**i+1**	**i+2**	**Type**

T 7	L8	T9	G10	73	-7	-101	15	I
E18	S19	S20	D21	-55	-25	-80	-8	I
P37	P38	D39	E40	-57	-32	-68	-16	III
F45	A46	G47	K48	48	46	62	22	III
G51	D52	G53	R54	-48	-42	-83	-9	I
T55	L56	S57	D58	-61	-36	-64	-30	III
L56	S57	D58	Y59	54	30	56	39	III
S57	D58	Y59	N60	56	39	91	5	I
E62	K63	E64	S65	55	143	67	19	II

1.4.5. Ubiquitin stability

The hydrophobic core of ubiquitin consists of the valine, leucine, isoleucine, and methionine residues buried within the interior of the molecule. This hydrophobic core may be a reason for the marked stability of ubiquitin largely contributed by hydrophobic interaction between Ile30 and Ile36 (Thomas et al., 2000). Any mutation in the core hydrophobic residue is marked by the loss of its stability. Strict conservation of sequence of ubiquitin leaves no scope for interspecies comparison to understand the role of individual residues. However, site directed mutagenesis has been successfully employed to elucidate specific functions of many residues contributing to the stability of ubiquitin. Val5, Val17, Leu67 buried nonpolar residues when substituted by Asn a polar residue lead to (>30%) decreases in the heat capacity upon unfolding and buried polar Gln41 when substituted with nonpolar residues Val or Leu lead to (>25%) increase in the heat capacity upon unfolding (Loladze et al., 2001). Similarly the contribution of solvent exposed charged residues was estimated in this protein by replacing the Arg residues in the position of Lys residues and carbamoylating the amino groups of Lys (Loladze et al., 2002). Analysis of the structure and stability of this modified protein showed that its stability is equal to the stability of the unmodified protein strengthens the fact that the stability of ubiquitin is mainly due to its hydrophobic core.

Electrostatic interactions in proteins influence the stability, solubility, catalysis, ligand binding, and redox potentials. Electrostatic interactions in ubiquitin were determined by the role of positive charges in modulating carboxyl p*K*a values, p*K*a values have been

determined by NMR for all 12 carboxyl groups in the wild-type ubiquitin and in variants where single lysines have been replaced by neutral residues. Aspartate p*K*a values in ubiquitin range from 3.1 to 3.8 and Glutamate p*K*a values range from 3.8 to 4.5 which is generally less for Asp and more or less similar to model compound values. The observed p*K*a values for Glu16, Glu18, Glu24, Asp32, Glu64 and Gly76 are within close proximity of the predicted values. These groups are probably not interacting significantly with other groups in ubiquitin. This is partially confirmed for Glu16, Glu18 and Asp32 by the mutagenesis experiments of the nearest lysine amino groups. Asp39 and Asp58 show moderately low p*K*a values (3.6) for the carboxyl groups are partially buried. This might be due to proximity effect of the guanidino groups of Arg72 and Arg74 for Asp39 and similarly, Arg54 for Asp58. Asp52 also shows moderately low p*K*a value (3.4) which makes a salt bridge to Lys27. Asp21 has the lowest p*K*a in ubiquitin, and both mutagenesis and electrostatics calculations indicate that Lys29 interaction is primarily responsible for this low p*K*a. The p*K*a for Glu34 in wild type ubiquitin is 4.5, very similar to model compound values, but the p*K*a in K11Q and K11T is unusually higher at about 5.3. NMR studies revealed significant interactions of Lys11 and Lys29 with Glu34 and Asp21, respectively (Sundd et al., 2002). Electrostatic interactions in ubiquitin due to the carboxyl p*K*a values that are different in the native and denatured states give rise to changes in stability with varying acid pH because these two states have different affinities for proton (Sundd et al., 2002).

Ubiquitin does not contain any Trp residues. Trp residues act as intrinsic fluorophores. In order to include a fluorescence probe in the sequence of ubiquitin Phe45 of ubiquitin was substituted with Trp and the folding kinetics was studied using fluorescence (Khorasanizadeh et al., 1993). Phe45 is located in a shallow hydrophobic pocket composed of Ala46, Ile61 and Leu67 (Vijay-Kumar et al., 1985).The tyrosine residue is involved in a reverse turn that is composed of residues 57 to 60. In addition, findings suggest that the tyrosine ring spans the large loop involving residues 51-59 and contributes to the stability of the loop by formation of a hydrogen bond between peptide nitrogen atom of Glu51 and side chain hydroxyl of the Tyr59 (Vijay-Kumar et al., 1985). Ubiquitin has been used as a model system to study the effect of loop insertions on the structure and stability of proteins. The effect of the loop is largely dependent upon the position of the insert and not on the

sequence or the length of the insert. Inserts into the 35–36 loop result in greater structural perturbation than inserts into the 9–10 loop, possibly due to the intrinsic flexibility/ stability of the two loops in the wild-type structure (Ferraro et al., 2005).

1.4.6. Folding of ubiquitin

Folding of ubiquitin using pulsed H-D exchange NMR indicated that the backbone amide protons of N-terminal β-sheet and α-helix are protected early. The C-terminal half of the protein exhibits relatively slow folding kinetics as the residues 59, 61 and 69 exhibit slow protection rates. Although, the amide protons located in some reverse turns like Leu56 (which is also located at the beginning of a 3_{10}-helix) also exhibit protection in the early phase (Briggs and Roder, 1992). Further evidence came from the studies on partially folded state of ubiquitin stabilized in 60/ 40% methanol/ water mixture, which revealed conservation of native secondary structural elements in the N-terminal half (Ub 1-21 and Ub 1-35). But the C-terminal half (Ub 36-76) which is predominantly β strand in character undergoes a transition to helical state (Harding et al., 1991; Stockman et al., 1993; Brutscher et al., 1997). Based on the above observations a model for folding of ubiquitin was suggested, where the N-terminal portion of the protein (spanning residues 1-35) serves as an autonomously folding chassis, which governs the folding of rest of the protein through tertiary interactions. A mutant thereof, (Thr→Asp9), was more stable and highly structured, possibly due to the introduction of a favorable interaction with Lys11 (Zerella et al., 1999).

Ubiquitin has two β-bulges. The first β-bulge and the N-terminal region of the protein have been the subjects of many studies. But the C-terminal region of the protein including the second β-bulge has been ignored for its slow folding kinetics.

1.4.6.1. N-terminal β-bulge

The first antiparallel β-bulge of ubiquitin is located in at the N-terminal region in the type I turn of the β-hairpin. This region displays native like structure in early folding intermediates. Moreover, occurrence of β-bulge in type I turn is unusual (Vijay-Kumar et al., 1987). Hence, it became a topic of extensive studies. Two peptides from ubiquitin 1-21 and 1-35 have been found to attain β-hairpin conformation autonomously in aqueous methanol (Cox et al., 1993). Later it was reported even in the absence of any organic solvent the

peptide1-17 can adopt native like structure, although to a much lower extent (Zerella et al., 1999). This observation confirms that the peptide has potential to act as an initiation site for protein folding. The β-bulge present in the hairpin is essential for the formation of latter structure in the N-terminal. Removal of Gly from this turn results in loss of structure (Chen et al., 2001). The question of occurrence of β-bulge as the cause or consequence of hairpin formation was answered by replacing the sequence TLTGK with type I turn forming NPDG. This replacement led to a non-native strand alignment demonstrating that the nature of turn dictates strand alignment in β-hairpin (Searle et al., 1995; Haque and Gellman, 1997). The importance of cross strand interactions in hairpin formation and stabilization were investigated by replacement of Thr9 by Asp in the peptide 1-71 and also by placing charged pairs at different positions along the length of the strands. In the former case increased stability of the hairpin due to charge interaction between the side chains of Asp9 and Leu11 was reported (Zerella et al., 2000). From the latest study it was concluded that cross strand interactions are less important for hairpin formation, although they contribute to stability of the structure (De Alba et al., 1997; Santiveri et al., 2000).

1.4.6.2. C-terminal β-bulge

The second parallel β-bulge also displays some unusual features. The parallel β-bulge bulge present at the C-terminal end is accompanied by a type II turn. Glu64 is the third residue in a type II turn. Glu64 is the first residue in parallel β-bulge (Vijay-Kumar et al., 1987). Generally the first residue in a parallel β bulge is a Gly (Chan et al., 1993). Presence of Glu64 is supported by unusual Φ and Ψ angles which are not under the allowed region of Glu in Ramchandran plot. Further these angles are observed with Gly (Vijay-Kumar et al., 1987). The reason for which nature selected Glu instead of Gly and conserved through evolution is interesting to be understood. Ile61 and Lys63 placed adjacent to Glu64. Ile61 has been reported to show somewhat slower protection in kinetic refolding experiments (Briggs and Roder, 1992) and Lys63 has a significant role in DNA repair mechanism (Spence et al., 1995). In view of this it needs to be understood whether this residue is important for folding and structural stability or for its function.

1.5. UBIQUITIN IN THE CELL

1.5.1. THE Ubiquitin gene family

In Sacchromyces cerevisiae, the ubiquitin gene family has four different loci from which ubiquitin is encoded, three of which *UBI1*, *UBI2* and *UBI3* are hybrid proteins in which ubiquitin (Ub) is fused to unrelated tail amino acid sequences where *UBI1* and *UBI2* are interrupted at identical positions by non-homologous introns. *UBI*1 and *UBI*2 encode identical 52-residue tails, known as UbL40 and *UBI3* encodes a different 76-residue tail UbS31 (*UBI*3). Hence, in the translation products ubiquitin is found fused to the ribosomal polypeptides L40 and S31, respectively. The tail residues are highly conserved between yeast and mammals. Specific endopeptidases cleave these precursor molecules to release ubiquitin moieties that are identical in sequence. In budding yeast, i.e. in vegetative cells of *S.cerevisiae* the ribosomal fusion proteins are responsible for the bulk of the free ubiquitin pool (Andre Catic and Hidde L. Ploegh, 2005; Ozkaynak et al., 1987). Before its proteolytic separation from L40 and S31, the ubiquitin moiety acts as a chaperone and thus facilitates ribosome assembly. *UBI4* the fourth gene, encodes a polyubiquitin precursor protein containing five ubiquitin repeats in head to tail spacer less arrangement. During exponential phase all genes are expressed. During stationary phase the expression of *UBI1* and *UBI2* is suppressed. *UBI4* is strongly induced by stress conditions, it contains 'heat shock box' upstream of the coding region. The essential function of UBI4 is to provide ubiquitin under various stress conditions namely heat stress (Finley et al., 1987; Ozkaynak et al., 1987), nutritional stress (Finley et al., 1987; Ozaknak et al., 1987), UV stress and antibiotic stress to the cell (Finley et al., 1994).

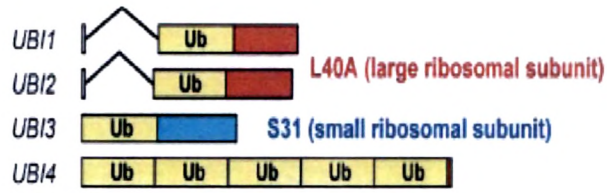

Figure 1.9. Ubiquitin gene family

After the fusion proteins are synthesized, another protein called ubiquitin C-terminal hydrolase cleaves the fusion proteins at the C-terminal end of ubiquitin. This either liberates an individual ubiquitin and ribosomal protein or liberates a set of ubiquitin monomers from the polyubiquitin. The number of ubiquitin repeats in the polyubiquitin locus varies among species, and some organisms express additional ubiquitin fusion proteins.

1.5.2. Protein degradation and the ubiquitin –proteasome pathway

Intracellular protein degradation contributes to many cellular regulatory mechanisms, including cell cycle control (Goebl et al., 1988), DNA repair (Jentsch et al., 1987), stress response (Ozkaynak et al., 1984; Finley et al., 1987), differentiation, signal transduction and metabolic controls. The conjugation of eukaryotic protein to ubiquitin is a prerequisite for its degradation by ATP dependent ubiquitin proteasome system (UPS) (Varshavsky, 1997; Weissman, 1997). Ubiquitin is joined reversibly to the target proteins via an isopeptide linkage through the C- terminal of ubiquitin to the ε-amino group of lysine (Lys) present on the acceptor protein (Ciechanover et al., 1980; Hershko et al., 1980). Ubiquitin conjugated to target protein undergoes further ubiquitination through any one of the ubiquitin's seven lysine residues present in the polypeptide. Lys29, Lys48, and Lys63 have been known to participate most frequently in polyubiquitination. Polyubiquitinated target proteins are recognized by the 26S proteasome for proteolytic degradation (Hershko et al., 1984, Baumeisteret et al., 1998; Lee and Goldberg, 1998; Rechsteiner, 1988; Groll and Huber, 2004; Pickart and Cohen, 2004; Wolfand Hilt, 2004; Rechsteiner and Hill, 2005). Efficient recognition of ubiquitinated substrates by the 26S proteasome requires a minimum targeting signal consisting of four ubiquitin moieties linked to each other through isopeptide bonds between Gly76 of ubiquitin and Lys48 (Chau et al., 1989; Pickart et al., 2000; Thrower et al., 2000). The diversity in the site of polyubiquitination is responsible for the different fates met by the target protein such as destabilization, rate of degradation, altered protein trafficking or functional modulation.

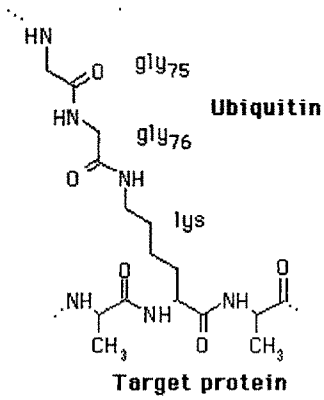

Figure 1.10. The isopeptide linkage

Polyubiquitin (poly-Ub) formed through Ub-Ub conjugation occurring with in the cells can have diverse structures due to the interaction between different lysine residues present in ubiquitin molecules. The existence of structurally distinct poly-Ub chains forms the basis for the diversity in Ub-dependent signaling. Recent studies also suggested that the deubiquitination of proteins (removal of ubiquitin from the substrate) also plays an

important role for the regulation of protein turn over. Mono-Ub and poly-Ub chains generally signal different fates for their target proteins. Monoubiquitination leads to altered trafficking in numerous pathways. Monoubiquitination of membrane proteins like receptors and transporters directs their internalization and degradation via lysosomes (Hicke, 1999; Hicke and Dunn, 2003), whereas poly-ubiquitination has multiple roles (Pickart et al., 2004).

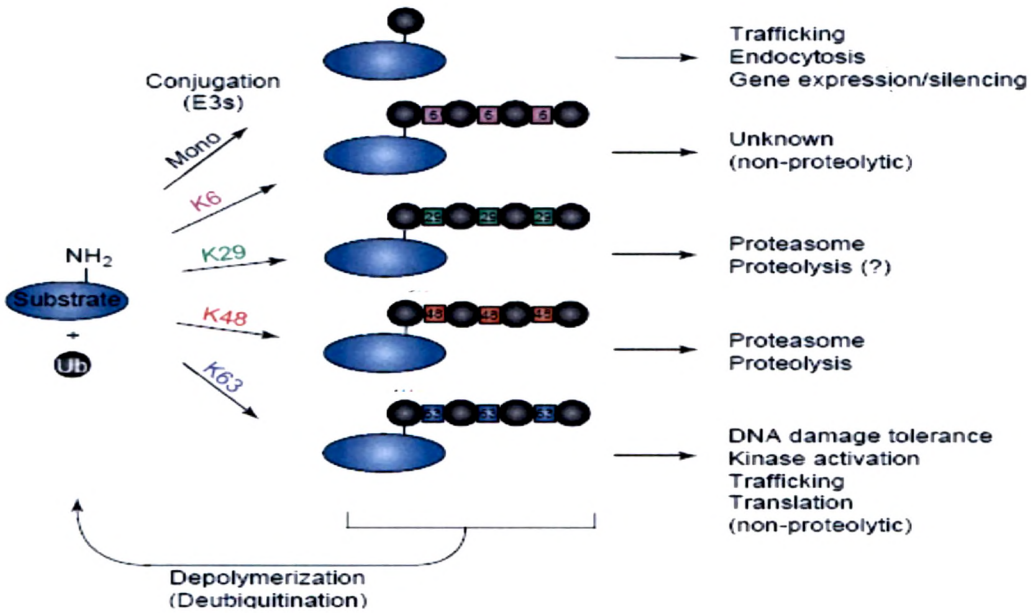

Figure 1.11. Polyubiquitin chains through different ubiquitin lysine residues. (Cecile M Pickart and David Fushman, 2004)

Lys48 linked chains are the principal proteasome delivery signals (Chan et al 1989, Finely et al., 1991, 1994; Hochstasser et al., 1991). Lys63 linked chains participates in DNA repair (Spence et al., 1995; Ulrich et al., 2002), the inflammatory response (Sun et al., 2004), protein trafficking (Hicke et al., 2003), and ribosomal protein synthesis (Spence et al., 2000). Lys11 linkage can signal proteasome degradation in vitro (Baboshina et al., 1996) but has not been known in any signaling pathway in vivo (Pickart et al., 2004). Tumor suppressor protein BRCA1 implicated in the pathogenesis of breast and ovarian cancer shows conjugation and poly-Ub chain formation via Lys6. These chains are recognized by 26S proteasomes but are processed differently from Lys48 linked chains (Nishikawa et al., 2004). Replacement mutation of Lys6 (Baboshina et al., 1996; Morris et al., 2004; Wu-

Baer., et al. 2003), Lys11 (Baboshina et al., 1996, Makhatadze et al., 2003), Lys27 (Spence et al., 1995), Lys29 (Spence et al., 1995; Arnason et al., 1994; Russel et al., 2004), Lys33 (Spence et al., 1995; Ott D.E et al., 2000), Lys 48 (Chau et al., 1989; Finley et al., 1994; Baboshina et al., 1996; Johnson et al., 1992, 1995) and Lys63 (Arnason et al., 1994; Spence et al., 1995) to Arg, double replacement mutation of Lys residues at 48 and 63 residues to Arg 63 and triple replacement of Lys 29, 48 and 63 by arginine (Arnason et al., 1994; Spence et al., 1995; Baboshina et al., 1996; Finley et al., 1994; Chau et al., 1989; Johnson et al., 1992, 1995, Russel et al., 2004) diable formation of poly-ubiquitin chains via these major known poly-ubiquitination sites resulting in reduction in poly-Ub chain length or conjugation rates Lys48Arg being lethal in yeast (Finley et al., 1994). Ubiquitin mutant containing no lysines with all lysines mutated to arginine also developed having no activity (Hershko, 1991 and 1985). Theses ubiquitin single, double and triple mutants can form an E1-catalyzed active thioester at the C-terminus allowing the molecule to be transferred to the lysines of substrate proteins.

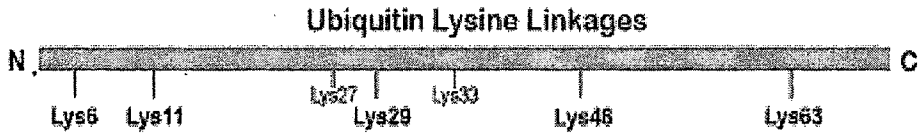

Figure 1.12. Positions of lysine residue in uniquitin.

Apart from the lysine resides present in the ubiquitin some of the other surface residues replacement like Pro19Ser and Pro19Ser Ala-28Ser Glu24Asp retains complete in vitro activity of protein. Tyr59Phe results in slightly decreased in vitro activity where as His68Lys and Tyr59Phe His68Lys show considerable decrease in the protein activity in vitro. Complete loss of in vitro protein activity results when core residue replacements like Leu67Asn and Leu69Asn were made signifying the importance of core. Moreover replacement mutation of Gly76 to Ala and deletion mutation like Leu73Δ and Leu73Δ Arg72Ser also results in complete loss of in vitro protein activity suggesting the active role of C-terminus in protein activity. These mutations resulted in partial or complete loss of ubiquitin function depending on the contributions they make towards the structure or activity (Ecker et al., 1987).

The first step in the ubiquitin system is activation of the carboxyl terminal glycine residue of ubiquitin by the **ubiquitin-activating enzyme (E1)** (Haas and Rose, 1982). This step is energy (ATP) dependent. In the second step, activated ubiquitin is transferred to a family of ubiquitin carrier proteins **ubiquitin-conjugating enzymes, (Ubc) (E2)**. Conjugation of ubiquitin to the target protein that is to be degraded can occur either by itself, or in cooperation with an **ubiquitin ligase (E3)**. E3s bind directly to substrate, and confer specificity and regulation to ubiquitination (Hochstrasser, 1996; Hershko et al., 1983).

Ubiquitin needs to be removed from tagged target protein prior to its entry in to the proteolytic core of proteasomes. The deubiquitinating enzymes belong to ubiquitin processing (UBP) and ubiquitin carboxy-terminal hydrolases (UBH) families (Hershko et al., 1980). In general, UBPs remove ubiquitin from poly-ubiquitinated proteins, whereas UBHs remove small adducts from ubquitin and regenerate free monomeric ubiquitin. Also, ubiquitin genes are transcribed and translated as a polyubiquitin chain, which then needs to be acted upon by UBHs to release single ubiquitin moieties (Weissman, 2001; Kim et al., 2003).

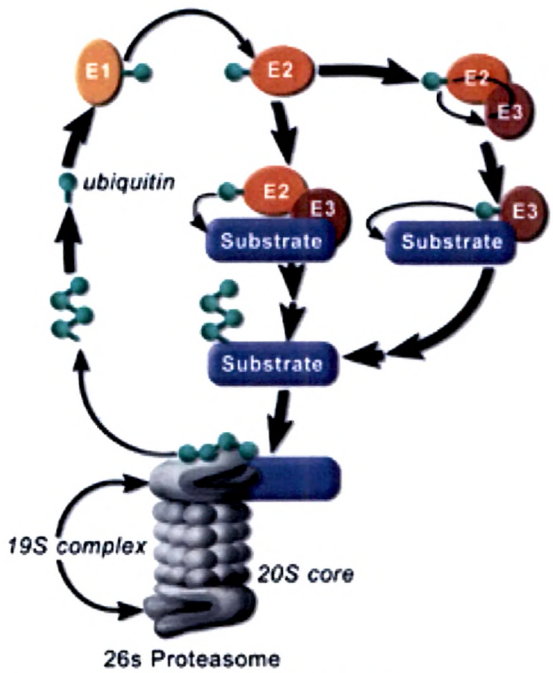

Figure 1.13.Ubiquitin proteasome system (UPS)

1.5.3. The proteolytic machinery: The ubiquitin proteasome system

The ubiquitin proteasome system (UPS) participates in ATP-dependent process of selective intracellular proteolysis. Ubiquitination of target proteins is essential for recognition and degradation through 26S UPS complex (Hershko and Ciechanover, 1998). Ubiquitin tagged proteins are subjected to proteolysis leading to the subsequent release of free and reusable ubiquitin. This process is also mediated by ubiquitin recycling enzymes such as isopeptidases also known as ubiquitin specific proteases (UBPS) and

deubiquitinating enzyme (Pickart and Cohen, 2004; Nandi et al., 2006). Assembly of the large 26S proteasome is a energy-dependent process. This 26S proteasome assembly is formed by two components, the 20S proteasome catalytic core and two 19S cap regulatory proteins, which form hood like structures on either side of 20S component. The 20S proteasome core complex contains four staked heptameric rings made up of 14 different polypeptides. The two α rings present at the ends of the cylindrical structure are formed by seven α-type proteins and seven β-type proteins form each of the two central β rings (Glickman et al., 1998; Nandi et al., 2006).

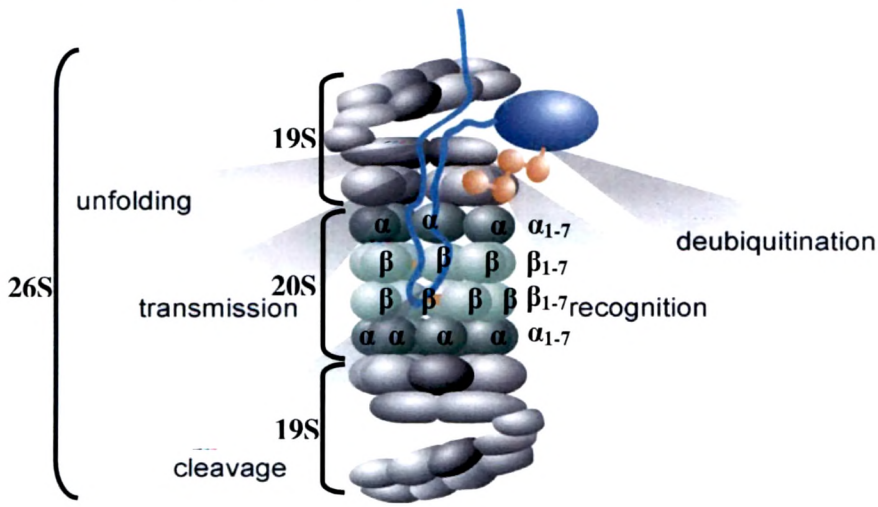

Figure 1.14. The 26S proteasome System. (Dieter H. Wolf and Wolfgang Hilt, 2004)

The catalytic sites are located on the inner surface of the β-subunits, whereas the α-subunits are responsible for controlling selective entry of substrate proteins. The 20S proteasome core complex encloses a cavity with three compartments joined by narrow passage. A polypeptide substrate unfolds in an ATP-dependent manner while passing through two narrow constrictions before it can be hydrolyzed. Protease activities are associated with three of the β-subunits (Baumeister et al., 1998), each having different substrate specificity.

- The **chymotrypsin-like** activity with preference for tyrosine or phenylalanine at the P1 (peptide carbonyl) position.
- The **trypsin-like** activity with preference for arginine or lysine at the P1 position.

- The **post-glutamyl** peptide hydrolyzing activity with preference for glutamate or the other acidic residue at the P1 position

Among the regulatory proteins, the 19S cap complex is important for regulation of the breakdown of ubiquitin-conjugated proteins by the 20S proteasome. The function of the 19S cap complex is to recognize (Thrower et al., 2000), unfold and transport (Verma and Deshaies, 2000), followed by and degradation of substrates by 20S proteasome (Baumeister et al., 1998). The 19S cap complexes and the 20S particleexist in dynamic equilibrium with the 26S proteasome complex (Pickart and Cohen, 2004).

Apart from selective recognition and removal of ubiquitin conjugated proteins by 26S UPS, ubiquitin-independent degradation is observed for certain proteins like ornithine decarboxylase. Ornithine decarboxylase is bound by another small protein called antizyme (Murakami et al., 1992).

1.5.4. Protein half life and substrate recognition

Metabolic stability of many proteins is a result of the varying half life of different proteins resulting from selective and regulated protein degradation by UPS. The presence of a Lys residue is necessary, but not sufficient for ubiquitination of target proteins for degradation. The cis-degradation signals in proteins are often referred to as degrons. Proteins must contain a 'degron' motif which is recognized by the components of the ubiquitination system. One such degron may be the N-terminal residue of a protein (Hershko et al., 1984; Bachmair et al., 1986). Other degron is the 'destruction box' or PEST sequences rich in Pro (P), Glu (E), Ser (S) and Thr (T). The proteins containing PEST sequences are rapidly degraded (Rogers, Wells and Rechsteiner 1986).

1.5.4.1. The N-end rule:

Certain N-terminal residues are found to be destabilizing a protein acting as a degradation signal. The N-end rule identifies these destabilizing residues and classifies them as N-degron, responsible for the in-vivo half-life of a protein. The N-end rule pathway is present in all organisms examined, from the bacterium *E. coli* to the eukaryotic mammalian cells. The N-degrons can bear different destabilizing N-terminal residues which are recognized by distinct targeting complexes (Varshavsky et al., 1996).

1.5.4.2. Primary Destabilizing Residues (N-d^p):

Destabilizing activity of these residues requires their physical binding by a protein called N-recognin, a type of E3 present in eukaryotes. N-recognin has at least two substrate binding sites designated as type I and type II sites. The type I binding site of N-recognin binds N-terminal Arg, Lys or His, the set of basic N-d^p residues, whereas the type II binding site binds Phe, Leu, Tyr or Ile the set of bulky hydrophobic N-d^p residues. In *S. cerevisiae* either of the sites of N-recognin can be mutationally inactivated without significantly disturbing the activity of the other (Varshavsky et al., 1996).

1.5.4.3. Secondary Destabilizing Residues (N-d^s):

These N-terminal residues are Asp and Glu in *S. cerevisiae* and Asp, Glu and Cys in mammalian cells. In eukaryotes their destabilizing activity requires their accessibility to Arg-tRNA-protein transferase (R-transferase). The R-transferase enzyme arginylates Asp and Glu present at the N-termini of a protein (Varshavsky et al., 1996).

1.5.4.4. Tertiary Destabilizing Residues (N-d^t):

These N-terminal residues are Asn and Gln. Destabilizing activity of these residues requires their accessibility to N-terminal amidohydrolase (Nt-amidase). Asn and Gln are deamidated by Nt-amidase if they are located at the N-terminus of a protein (Varshavsky et al., 1996).

1.5.4.5. Stabilizing Residues:

A stabilizing N-terminal residue is a "default" residue, and is stabilizing as components of an N-end rule pathway do not bind to it efficiently. Gly, Val and Met are the stabilizing residues (Varshavsky et al., 1996). In *E. coli* N-end rule does not include the N-d^t residues (Varshavsky et al., 1996)

The N-terminal residues in the non-compartmentalized proteins from both prokaryotes and eukaryotes are exclusively of the stabilizing class, while those in the compartmentalized proteins are largely of the destabilizing class. Posttranslational addition of destabilizing amino acids to the amino-terminal of target proteins in vivo is often carried out to accelerate their degradation under changed physiological state of the cell. These

additions are expected to occur during entry to and exit from the cell cycle, responses to chemical or physical stress and specific differentiation events like erythroid differentiation and spermatogenesis (Bachmair et al., 1986). A variety of factors apart from the N-end rule may combine to modulate a protein's half-life in vivo, like, the flexibility and accessibility of the protein's N-terminus, the presence of chemically blocking N-terminal groups such as the acetyl group, the distribution of ubiquitinatable Lys residues near the N-terminus, also the structure of the C-terminus. Additionally, the quaternary structure of proteins is also another parameter expected to modulate the impact of the N-end rule pathway on protein half-lives in vivo. Hence, the recognition of a N-terminal residue in a protein appears to regulates the metabolic stability of a protein in vivo (Bachmair et al., 1986).

1.5.5. The role of ubiquitin proteasome system in biological function and its pathogenesis.

Regulatory protein degradation by UPS plays crucial roles in a large variety of essential cellular pathways. These pathways are highly interconnected and any failures in proteasomal degradation can be a source of severe disturbance to the cell function, often becoming the cause for diseases. The proteins which are substrates of UPS play key roles in many basic cellular processes like regulation of cell cycle (Goebl et al., 1988) differentiation and development (Bowerman et al., 2006), DNA repair (Jentsch et al., 1987) and chromatin remodeling, transcriptional regulation (Hochstrasser and Varshavsky, 1990) and silencing, nutritional stress (Finley et al., 1987; Ozkaynak et al., 1987), modulation of cell surface receptors, ion channels and the secretory pathway (Hicke and Riezman, 1996), morphogenesis of neuronal networks, involvement in the cellular response to stress and extra cellular effectors (Ozkaynak et al., 1984; Finley et al., 1987), removal of misfolded and unwanted proteins and regulation of immune and inflammatory responses (Ciechanover and Iwai, 2004; Varshavsky, 2005).

Different phases of the cell cycle maintain the levels of cell cycle regulatory proteins cyclins, cyclin dependent kinase inhibitors. The levels of these proteins are modulated by the ubiquitin proteasome system (UPS). The UPS is essential for smooth cells cycle progression and thus differentiation and development (Murray, 2004). p53 a tumor suppressor is essential in regulation of cell cycle arrest, DNA repair and apoptosis. p53 levels are

maintained low under normal conditions of growth as they interact with Mdm2, an E3 ligase. DNA damaging signal phosphorylates p53, phosphorylation of p53 results in reduced interaction with Mdm2 causing cell death. Excessive degradation of p53 by UPS or mutations in p53 is often associated with different human cancers (Ciechanover and Iwai, 2004). NF-κB is an essential transcription factor involving inflammatory responses. Upon appropriate stimulation, inhibitor IκB is phosphorylated and degraded by the UPS releasing NF-κB. Free NF-κB enters into the nucleus, inducing expression of genes involved in the inflammatory response (Karin and Ben-Neriah, 2000). Cell surface receptors are targeted by proteasome system. In addition, self or microbial proteins are digested into peptides by the UPS and presented on MHC class I. Inhibition of UPS leads to impaired antigen presentation (Kloetzel, 2004).

Finally mutated, denatured and misfolded proteins are recognized specifically and are removed efficiently. Thus aberrations in the UPS have been implicated in the pathogenesis of several diseases, both inherited and acquired. The pathological states can be divided into two major groups (Burkhardt, 2007):

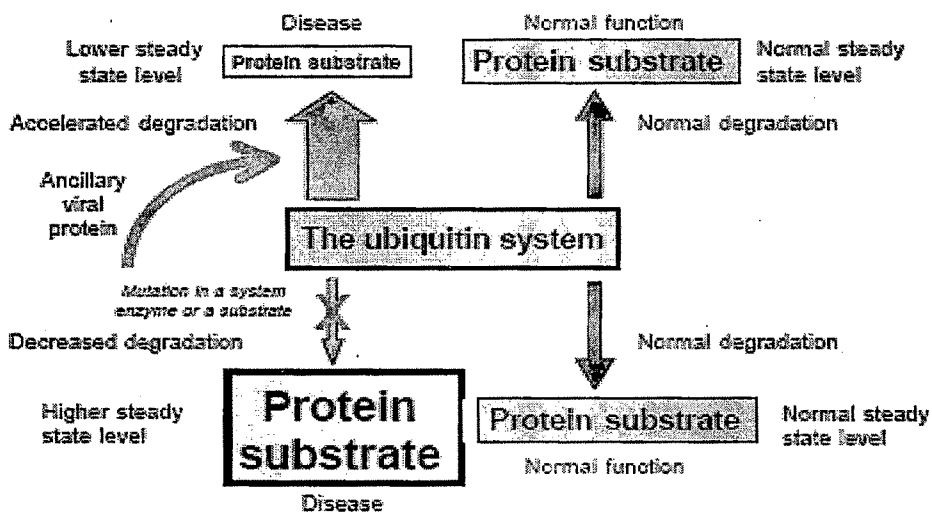

Figure 1.15. Pathogenesis of ubiquitin system related diseases

- **Loss of function:** Mutations defects in an enzymatic component or a target substrate that result in stabilization of certain proteins. Decreased degradation rates are a major

cause of neurodegenerative diseases e.g. Alzheimer's (Keck et al., 2003), Parkinson's (Mc Naught and Jenner, 2001) and Huntington (Zhou et al., 2003). This part explains the association of these conditions with aging. Decreased degradation of positive regulators of cell cycle lead to loss of control over cell cycle, a condition leading to cancer.

- **Gain of function**: It is associated with abnormal accelerated degradation of the protein target. Increased proteasome activity is characteristic of muscle wasting conditions and inflammation eg. Sepsis, cachexia and uraemia (Hobler et al., 1999). Proteasomes participate in degradation pro-apoptotic proteins, the tumor suppressor p53 and the negative regulators of cell cycle (Reed and clocks, 2003). Increased degradation of any of the protein is pathological.

Progressive misfolding of specific protein into aggregates which can injure and kill cells renders UPS to function improperly because of the burden on UPS.

1.6. RATIONALE FOR SELECTING YEAST UBIQUITIN AND YEAST AS A MODEL SYSTEM

Saccharomyces cerevisiae has been a popular and successful model system for understanding eukaryotic cellular and molecular biology (David Botstein and Gerald R. Fink, 1998). The reason for this success is experimental tractability, especially in applying classical and molecular genetic methods to understand the genes, proteins and their functions within the cell. Being a eukaryote, yeast shares many functional features at molecular and cellular levels with higher eukaryotes, which include the UPS system. *S.cerevisiae,* with its short doubling times, provides a convenient host system for eukaryotic protein expression and purification. In addition, it is amenable to genetic manipulations like site directed and random mutagenesis. Moreover ubiquitin of yeast differs from that of higher eukaryotes by only three residues. Thus yeast ubiquitin and yeast have been selected as a model system for the present studies.

1.7. THE ROLE OF CONSERVED RESIDUES IN THE PARALLEL G1 β-BULGE IN THE DETERMINATION OF STRUCTURE-FUNCTION RELATIONSHIPS OF UBIQUITIN

Today considerable amount of information is available on the structure of proteins. However, structure-function relationships in individual proteins evade our understanding. Consistent efforts with small, model proteins may supply answers to some of the puzzling questions and ultimately pave way for designer proteins with desirable properties. The main objective of the thesis is to study the structure function relationships in ubiquitin, a small, globular and highly conserved eukaryotic protein. An understanding on the structure-function relationships in ubiquitin may have the potential for biomedical implications considering the multiple and indispensable functions of the protein.

Chapter 1 presents a general introduction to protein structure, protein folding, the importance of structure in relation to the function of proteins, structure and functions of ubiquitin. The chapter also introduces the main objectives of the work and their chapter-wise presentation.

As mentioned earlier in section 1.4.6. of this chapter ubiquitin has two β-bulges. The first one is located in N-terminal region in the type I turn of the β-hairpin. This region displays native like structure in early folding intermediates. Occurrence of β-bulge in type I turn, however, is uncommon (Vijay-kumar at el., 1987). Hence, it became a topic for many investigations.

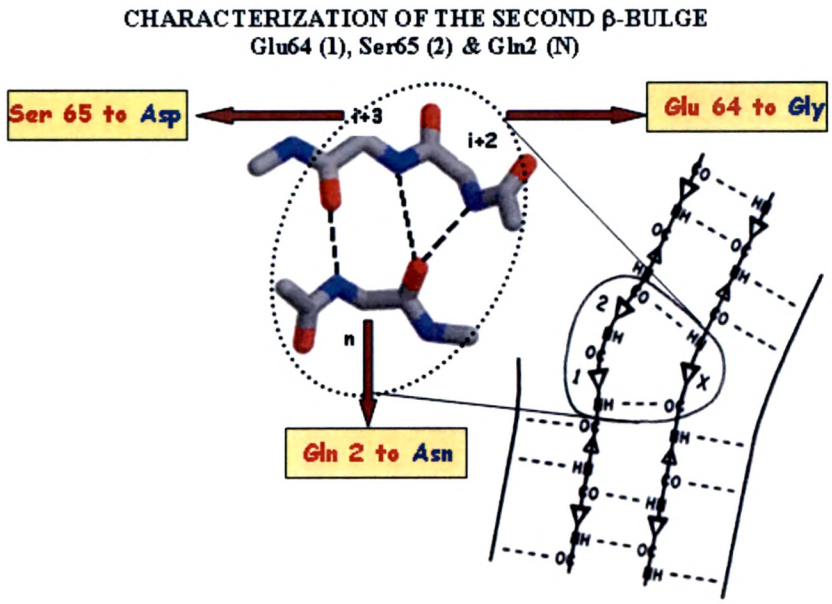

Figure 1.16. Position of Glu64, Ser65 and Gln2 in the β-bulge at C-terminal.

The second β-bulge also displays some unusual features. It is a parallel G1 β-bulge, which is very rare. Glu64 forms the third residue in a type II turn and first residue in the β-bulge (Vijay-kumar at el., 1987). Generally first residue in a β bulge is a Gly (Chan et al., 1993). Further, the homologs of ubiquitin of Rub1 and NEDD8 display Gly in this position (Jentsch et al., 2000). This unusual feature and its conservation through millions of years of evolution make it an interesting protein for structure function studies. Moreover, this acidic residue is adjacent to a basic residue Lys63, which has been found to be important for UV repair of DNA (Spence et al., 1995), resistance of cells to stress conditions (Arnason et al., 1994; Spence et al., 2000; Ozkaynak et al., 1987), and endosomal degradation of certain proteins (Galan et al., 1996). Significance of this structural feature in ubiquitin biology is the main focus of **Chapter 2.** In order to understand the importance of Glu64 in the structure, stability and function of ubiquitin, a variant of ubiquitin (UbE64G) has been engineered using site directed mutagenesis and characterized by circular dichroism (CD) and fluorescence spectroscopy. Its stability was evaluated by guanidine hydrochloride and thermal denaturation studies.

Similarly, the remaining two residues of the β-bulge Ser65 in the second position and Gln2 in the Nth position also have low propensity for the secondary structure. However, they are also totally conserved in all ubiquitins irrespective of the species. The % frequency of occurrence of these amino acids in a bulge in these positions is as low as 1.5 and 0.34 respectively (Chan et al., 1993). Further more structural homologs of ubiquitin or ubiquitin like proteins (UBL) such as SUMO-1, SUMO-2, SUMO-3 have Asp in place of Ser65 and RUB1, NEED8 has Asn replacing Gln2 (Jentsh et al., 2000). These

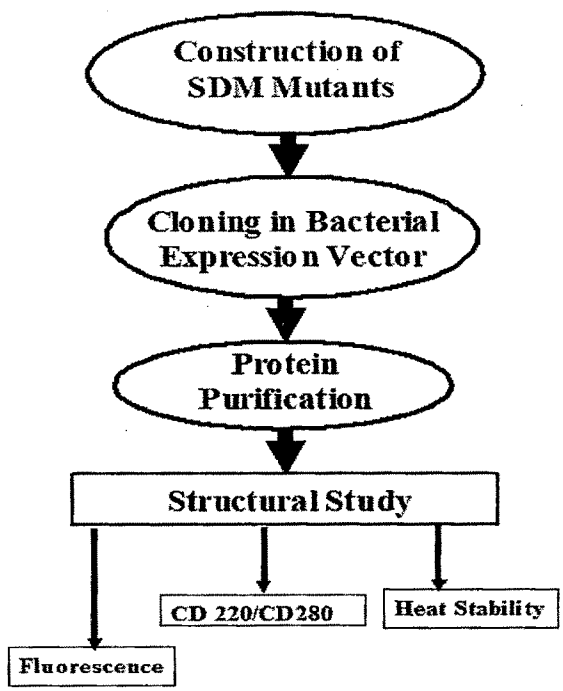

Figure 1.17. Schematic representation of proposed structural work.

residues may have their functional significance in ubiquitin. Hence, in **Chapters 3 and 4,** the structural importance of these residues was addressed by constructing S65D and Q2N substitutions and studying the effects on their secondary and tertiary structures and stability by using CD and fluorescence techniques.

Chapter 5 addresses the effect of mutations in ubiquitin on its function, by looking at their ability to complement the stress sensitive mutant of *Saccharomyces cerevisiae*. Ubiquitin is known to be participating in a variety of cellular functions as explained earlier and is important component of UPS. Ubiquitin rescues the cells under various stress conditions and plays an important role in maintaining the cellular physiology through regulating the turn over of various proteins (Ciechanover and Iwai, 2004; Varshavsky, 2005). Null mutant of ubiquitin is lethal. However, UBI4 mutant of *Saccharomyces cerevisiae*, lacking the *UBI4* gene cluster coding for polyubiquitin protein, expressed under stress conditions, has proper vegetative growth but lacks resistance to different stresses (Finley et al., 1987 and 1994; Ozkaynak et al., 1987). The ability of mutant ubiquitins UbE64G, UbS65D and UbQ2N to complement UBI4 cells under various stresses has been assessed.

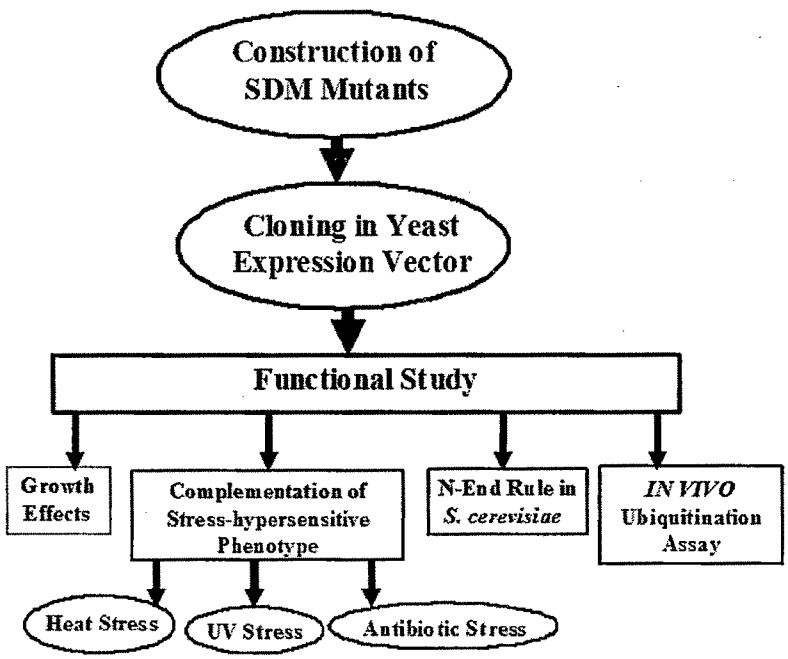

Figure 1.18. Schematic representation of proposed work for functional complementation.

As ubiquitin is absolutely essential for eukaryotes, mutations in ubiquitin can be employed strategically to remove certain unwanted cells. In order to obtain such a mutant error prone PCR technique (Arnheim, 1992) was used in **Chapter 6** to generate random mutants. Interestingly one of the mutants displayed a dosage dependent lethal phenotype killing the UBI4 mutant cells and leaving the wild-type cells unaffected. Primary structure of the protein was found and certain functional characterizations have been carried out.

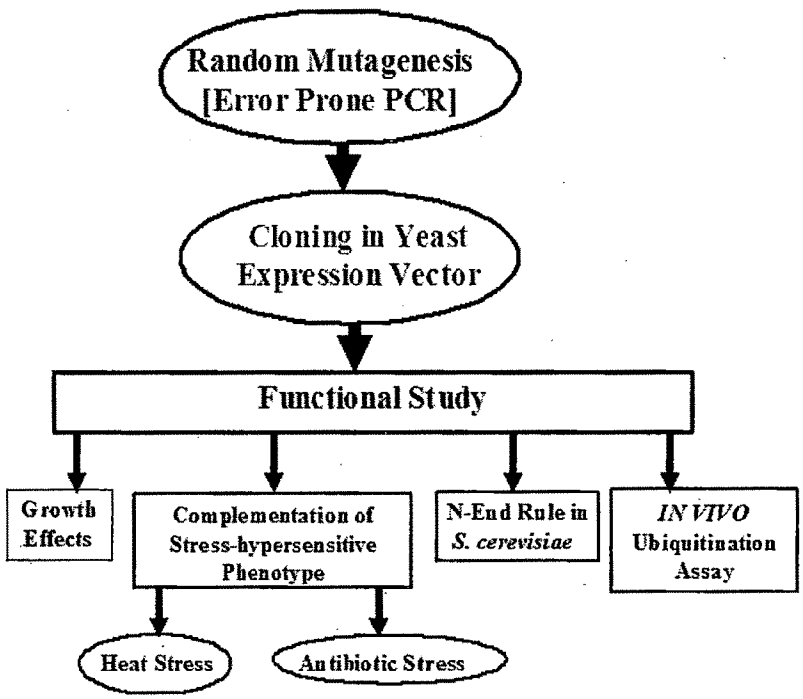

Figure 1.19. Schematic representation for screening of error prone PCR mutants and structure function relationships of ubiquitin.