

Protein Folding and Degradation

Scientific Foundations

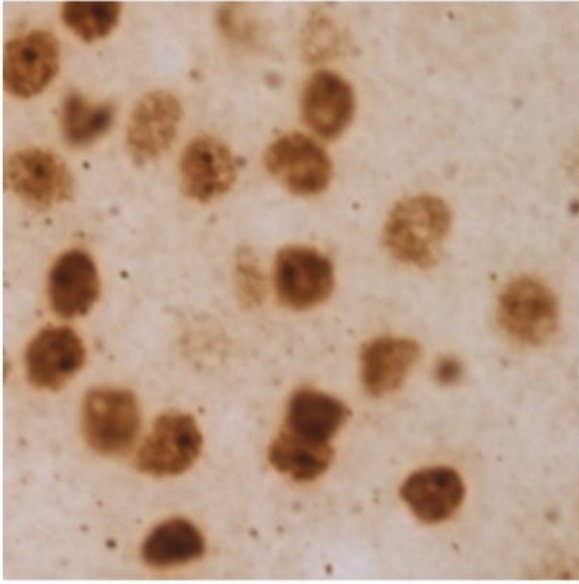
Learning Objectives

- Students should be able to explain how chaperones prevent aggregation of unfolded protein.
- Students should be able to differentiate the roles of HSP70 and HSP90 in protein folding.
- Students should be able to explain the roles ubiquitin plays in the life of a protein and diagram the pathway that adds ubiquitin to proteins.
- Students should be able to describe how the proteasome digests proteins.
- Students should be able to describe how glycosylation in the ER facilitates protein folding and identifies unfolded proteins.
- Students should be able to describe how proteins that sense unfolded protein in the ER and how their activation reduces the concentration of unfolded protein in the ER.
- Students should be able to define autophagy and its role in eliminating large macromolecules and organelles.

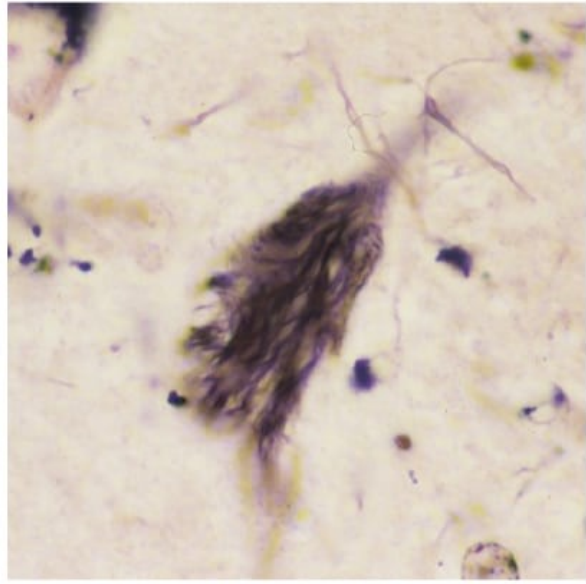
Biology textbooks can give the impression that molecular events proceed with high fidelity. However, many reactions are prone to errors that require molecular systems to detect and correct. For example, DNA replication introduces incorrect nucleotides and cells have machinery that scans the genome looking for mismatches in base pairing and repairs those mismatches. Nevertheless, some errors go undetected and can generate mutations in genes that change their amino acid sequence or regulation of expression.

Another source of error in cells is the folding of proteins into three-dimensional structures. Converting a linear chain of amino acids into a functional three-dimensional structure is a complex process and prone to errors that result in incorrectly or partially folded structures. One consequence of protein misfolding is that the cell has expended energy to generate a non-functional protein. A more dire consequence is that partially folded proteins can adhere to each other, creating aggregates of protein within cells. A large build up of protein aggregates can lead to apoptosis, which is the process of cell death. Some partially folded proteins will polymerize into fibrillar structures called amyloid. These amyloid fibers can damage tissues and trigger apoptosis.

Huntingtin



β -amyloid



Unfolded and partially folded protein polymerize into fibers and damage cells and tissues.

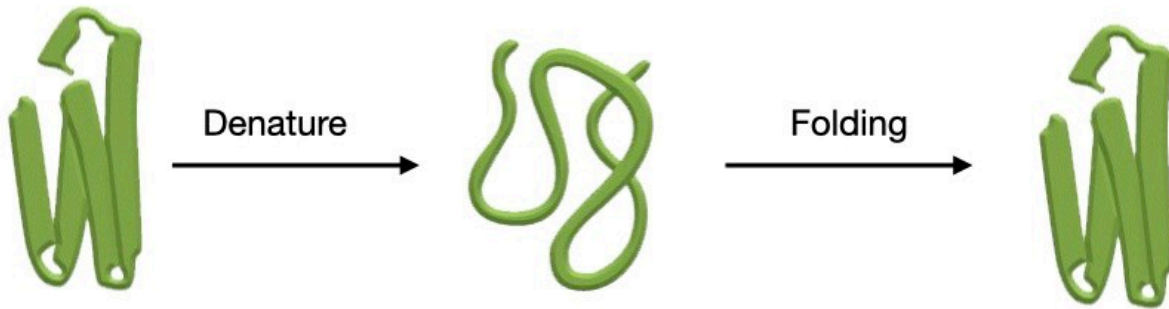
Because of the potential damage caused by unfolded protein, cells have extensive and elaborate mechanisms to facilitate proper folding of protein and to detect accumulation of unfolded protein.

How proteins fold

We have learned the molecular structure of thousands of proteins and from these structures we understand what maintains the three-dimensional shape of a protein. Most important are hydrophobic domains which bury themselves within the central region of the protein to avoid interacting with water. These hydrophobic domains must interact in a specific arrangement to maintain the structure of a protein. Certain polar amino acids within the hydrophobic domains must be positioned correctly so they can form hydrogen bonds. If the hydrophobic domains are misaligned, the polar amino acids can't form hydrogen bonds, and the charge within the hydrophobic domains weakens their interaction.

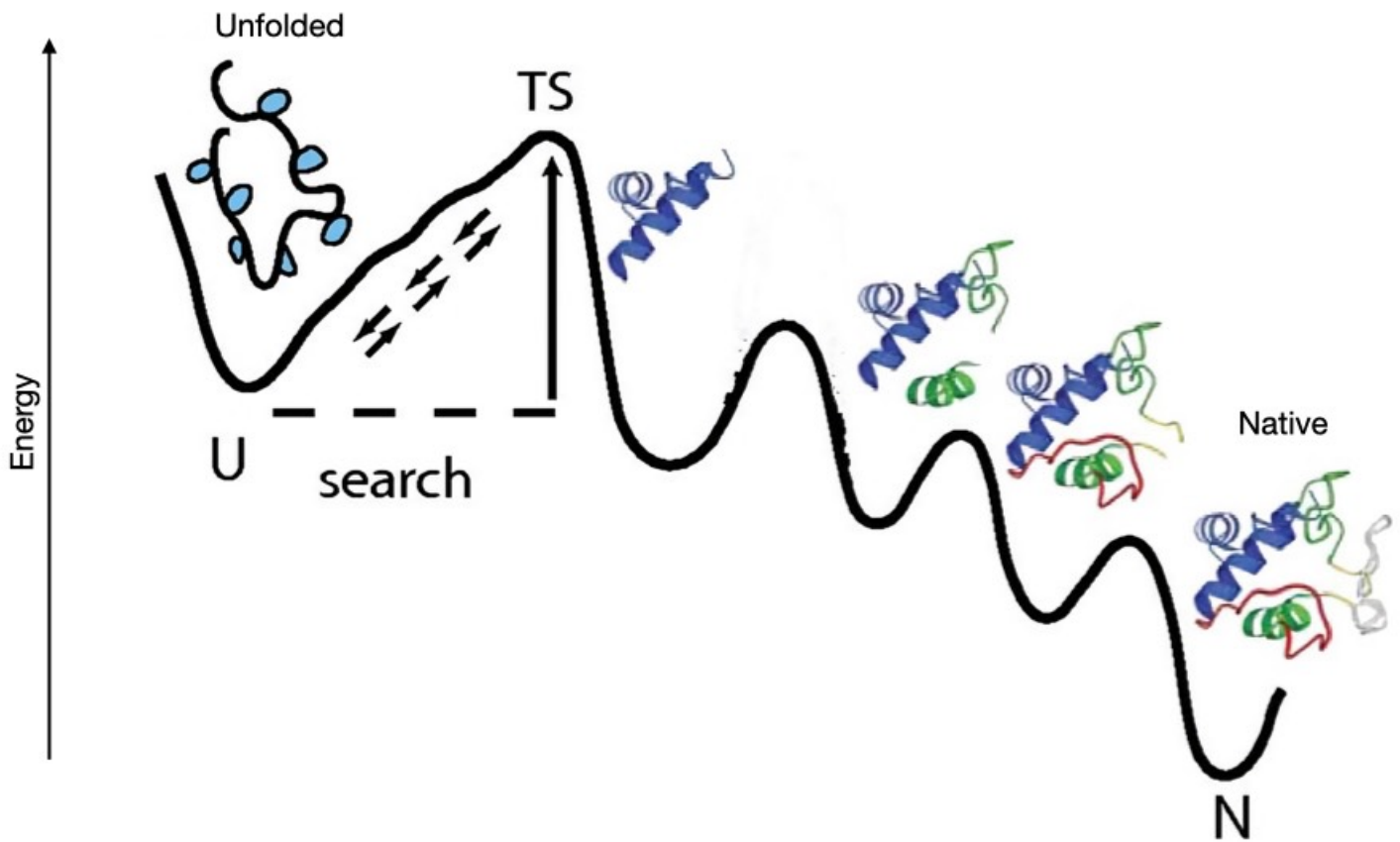
While we understand the final structure of many proteins, less clear is how polypeptide chain folds into its correct, native structure. Early studies on protein folding suggested that proteins fold rapidly and with high fidelity. Most of these early experiments, however, were performed with

small polypeptides (< 100 amino acids) in which a protein was denatured and then allowed to refold.



Proteins shorter than 100 amino acids can spontaneously fold into 3-D structure.

These early studies gave the impression that every molecule of a protein follows the same path towards its final three-dimensional structure. The unfolded protein has a high energy state and as it folds into its three-dimensional structure, its energy decreases at each step in the folding pathway.

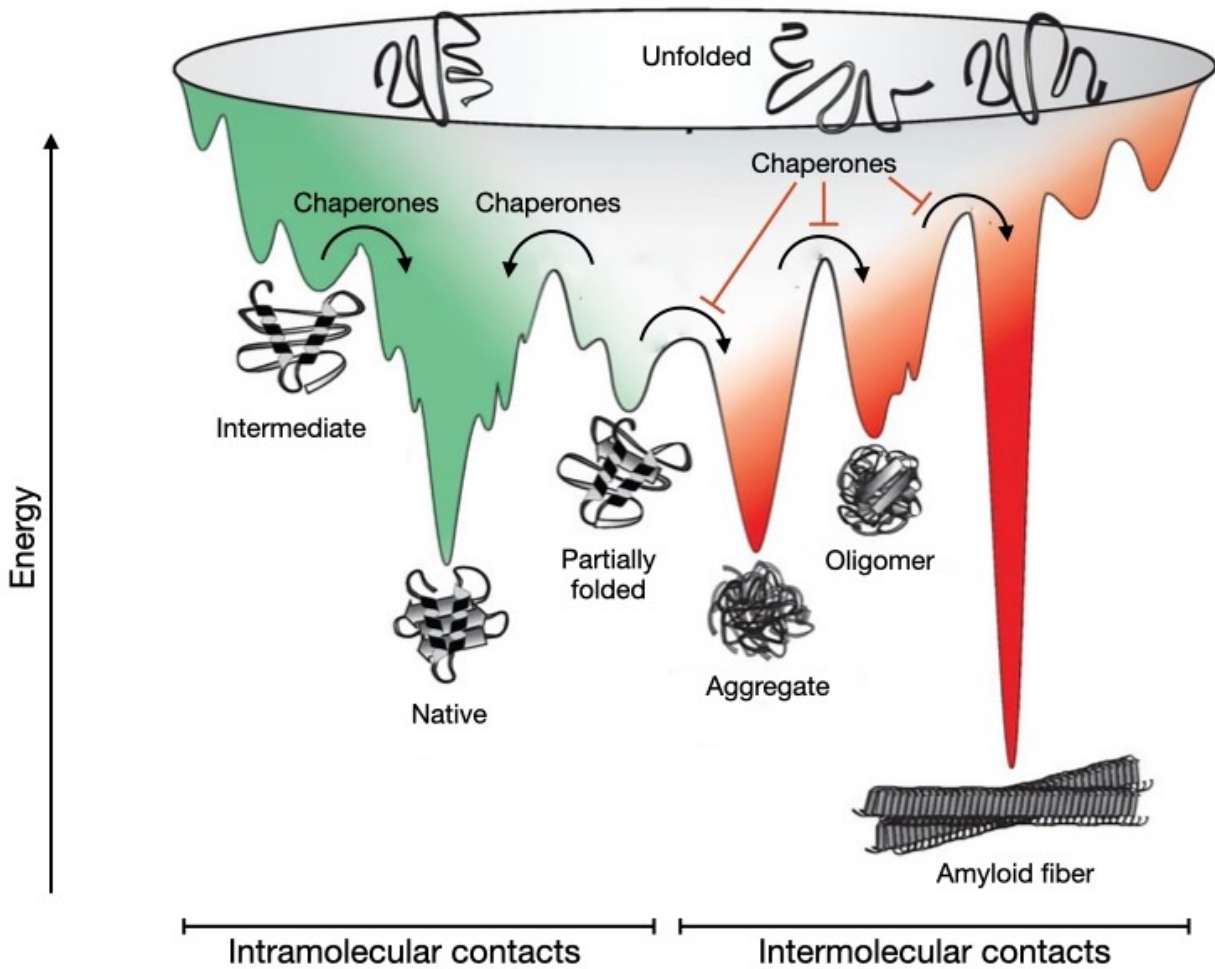


The energy state of proteins decrease as they fold into their three-dimensional structure.

In contrast to small polypeptides, longer polypeptides did not fold rapidly and in many cases failed to fold into a stable three-dimensional structure. Instead, the proteins aggregated into large insoluble complexes. Because 95% of the proteins encoded by the human genome are longer than 100 amino acids, cells need mechanisms to facilitate folding of proteins.

Our current model of how most proteins fold is that each polypeptide takes a different path toward the stable, three-dimensional structure. This model can be visually represented through an energy landscape of a protein. Higher regions of the landscape are the high energy states of the protein where it has less structure. As it folds, its energy decreases and it migrates down the landscape. The protein's stable, three-dimensional structure is the lowest energy state, but along the way there are other low energy states that function as valleys. If a protein's folding pathway takes it into one of these valleys, the protein becomes trapped because an input of energy is required to get it out of the valley. The valleys at higher states represent folded states that are far

from the native structure whereas those closer to the native energy state are folded states that are very similar to the native structure.



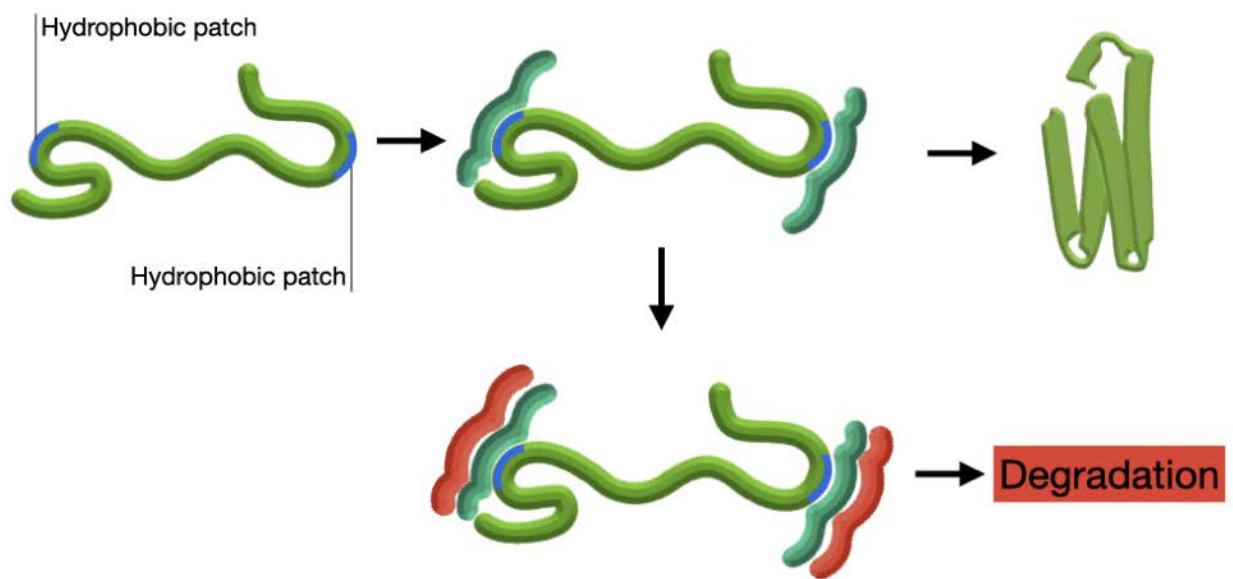
Large proteins follow several folding pathways that can lead to multiple states.

Some folding pathways lead not just to a partially misfolded protein but one that forms higher order complexes: aggregates, polymers and amyloid fibers. These complexes arise when misfolded proteins associate with each other, primarily through exposed hydrophobic domains. These larger complexes can decrease cell viability or trigger cell death.

Chaperones

How do cells reduce the probability that a protein falls into one of the incorrect valleys? Cells express a large class of proteins called chaperones that help nascent proteins find their correct three-dimensional structure. They do this by preventing interactions that lead to incorrectly folded

states (alternative low-energy states) or by providing the energy to move a protein out of an alternative low-energy state. The most common mechanism by which chaperones work is by forming temporary interactions with hydrophobic domains in a protein. This prevents interactions between mismatched hydrophobic domains from forming. Chaperones repeatedly bind and release from hydrophobic domains to prevent spurious interactions while allowing the correct domains to find each other. Chaperones can also dissociate mismatched hydrophobic domains and allow a misfolded protein another opportunity to fold correctly.

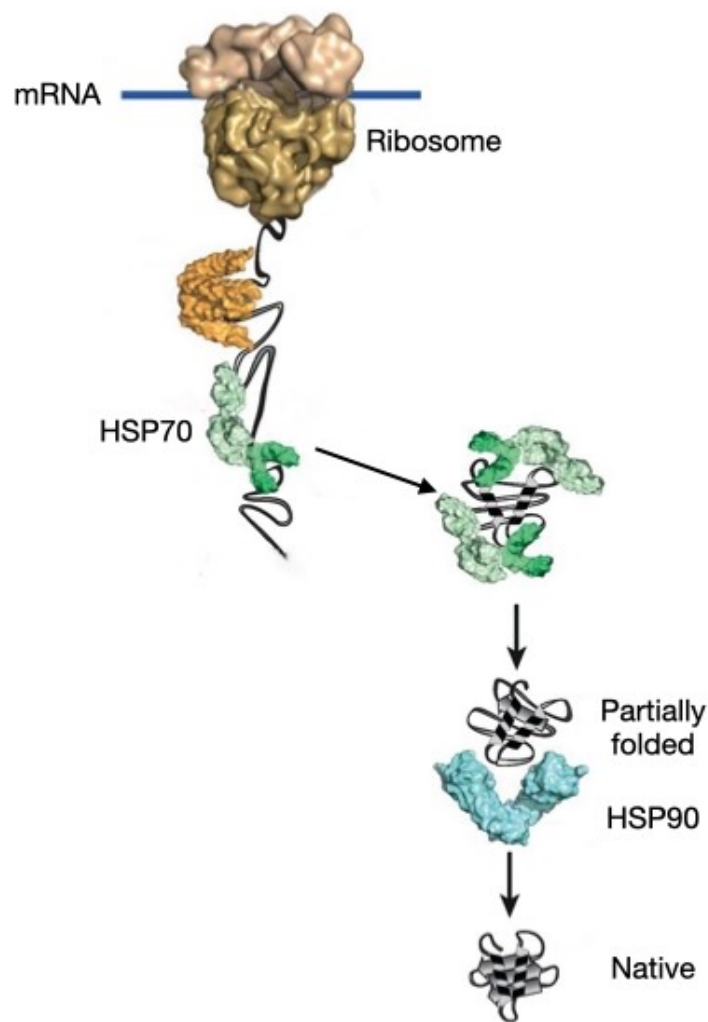


Chaperones bind hydrophobic domains to mediate folding or recruit degradation machinery.

Despite the help from chaperones, some proteins fail to fold correctly. In these instances, the chaperones recruit proteins which will target the nascent protein for degradation (see below).

The chaperone account for roughly 2% of all protein in a cell, making them one of the most abundant. The abundance of chaperones allows them to protect most new synthesized and unfolded protein during normal levels of protein expression. Two critical chaperones are HSP70 and HSP90. HSP stands for heat shock protein. HSP70 and HSP90 were discovered when cells are exposed to high temperatures. High temperature causes some proteins to unfold, and also increased the levels of HSP70 and HSP90 in cells. Both chaperones were found to help unfolded protein avoid aggregation by binding to their hydrophobic domains.

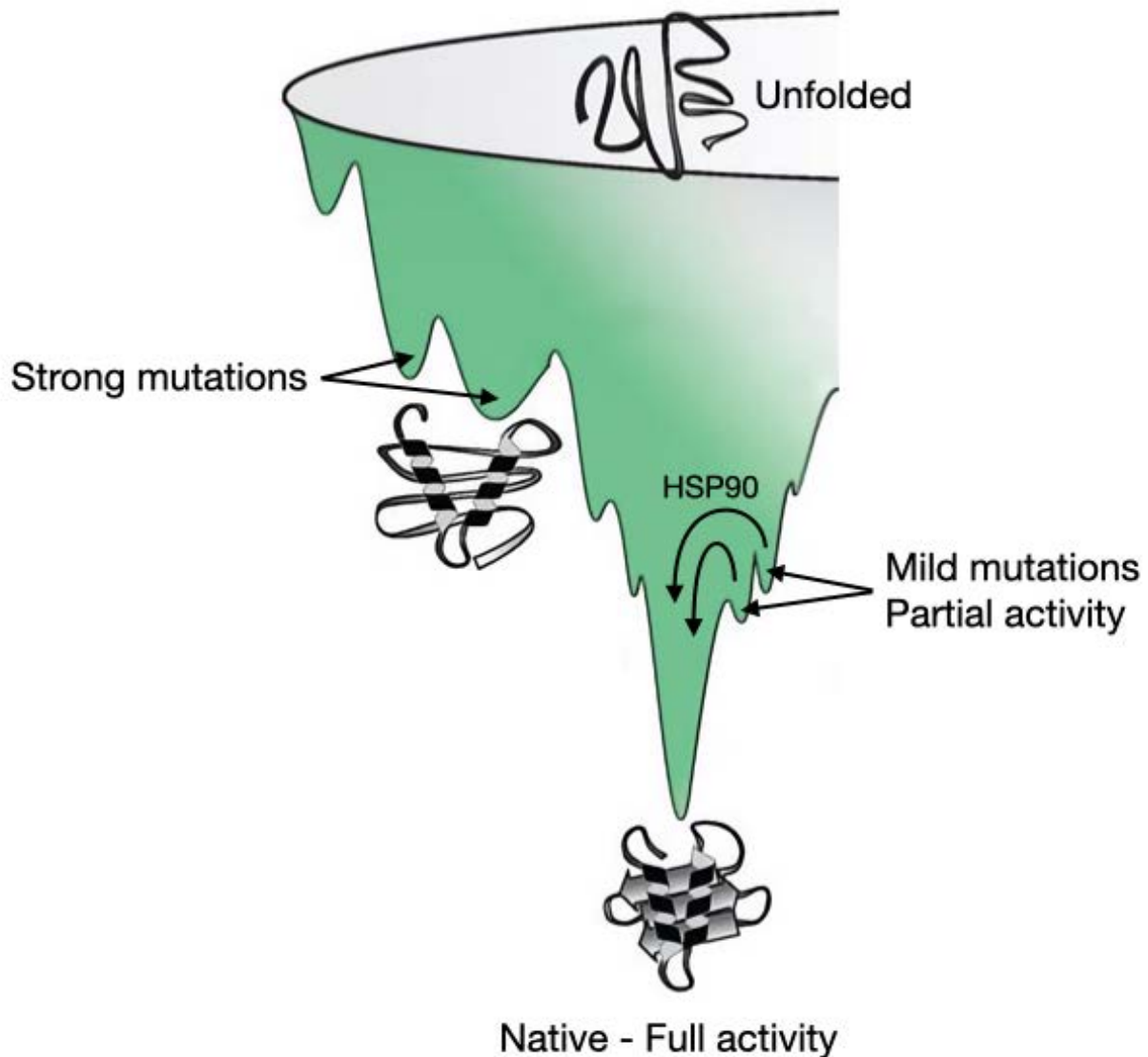
HSP70 and HSP90 function at different stages of protein folding. HSP70 associates with proteins soon after they are translated and probably helps a protein transition from secondary structure to its tertiary structure. In contrast, HSP90 associates with proteins much later in the folding pathway, often when they have achieved or are close to achieving their final structure. In fact, HSP90 seems to maintain an association with some proteins even after they are correctly folded. Thus, HSP90 helps proteins make minor changes in their structure to transition from a partially folded to native state.



Heat-shock proteins mediate different steps in protein folding pathways.

Another interesting feature of HSP90 is that it can mask mutations that cause a slight change in protein structure. Without HSP90, these mutations lead to partially folded states that are close to

the native state, and the small difference in structure changes the activity of the protein. HSP90 can bind to these partially folded structure and facilitate the conversion to the native state or very close to the native state to restore proper activity of the protein. Thus, individuals could have different genetic backgrounds but similar physiologies because HSP90 corrects the effect of the mutations on protein structure.



HSP90 converts proteins with mild mutations to their native structures to restore full activity.

One consequence of using HSP90 to mask genetic mutations is that to maintain normal cell physiology cells must have adequate supplies of HSP90. If HSP90 concentrations fall, then some of the mutated protein will not fold into its native structure reducing the activity of the protein and

potentially compromising cell physiology. For example, under normal conditions, cells express enough HSP90 to facilitate folding of most other proteins. If conditions change and those new conditions generate more unfolded protein in cells, then the amount of HSP90 available to facilitate folding of mutated protein will be reduced, which could alter the behavior of the cells. Many conditions increase unfolded protein, including increases in temperature, oxygen and glucose deprivation, infection and exposure to toxins.

Even with the help of chaperones, some proteins fail to fold into a proper three-dimensional structure. Cells need to eliminate these unfolded proteins to reduce the risk of proteins forming aggregates. Chaperones, in addition to helping proteins fold, will also target unfolded proteins for degradation. A protein that remains bound to chaperones for an extended period is unlikely to properly fold, so chaperones will recruit the machinery which leads to degradation of the protein. Thus, chaperones operate in two phases. In the first phase, they bind an unfolded protein and help it fold. In the second phase, when the protein fails to fold, they initiate a pathway that leads to degradation of the protein.

Protein Degradation

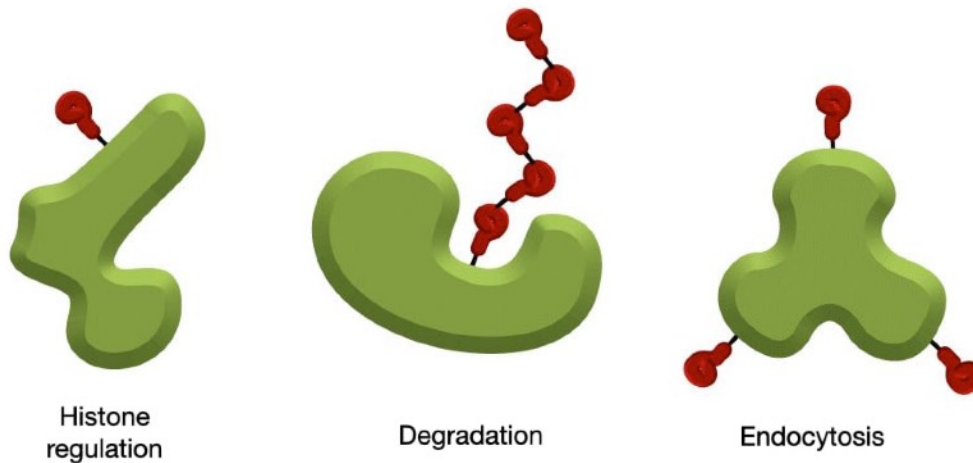
Ubiquitylation

Ubiquitylation mediates degradation of protein in the cytosol. Ubiquitin is a small polypeptide that is covalently attached to specific proteins. The presence of ubiquitin on a protein can trigger different events and depends on the amount and arrangement of ubiquitin. For example, a single ubiquitin on histones changes their arrangement on chromosomes. Multiple, single ubiquitins on receptor proteins triggers their inclusion in endocytic vesicles. A chain of ubiquitin on a protein triggers digestion of that protein.

Ubiquitin

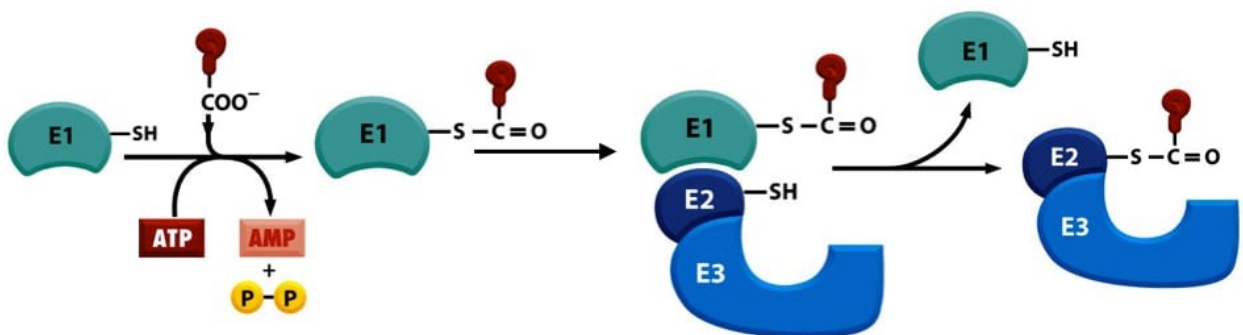


Ubiquitin is a small polypeptide that is covalently linked to proteins.



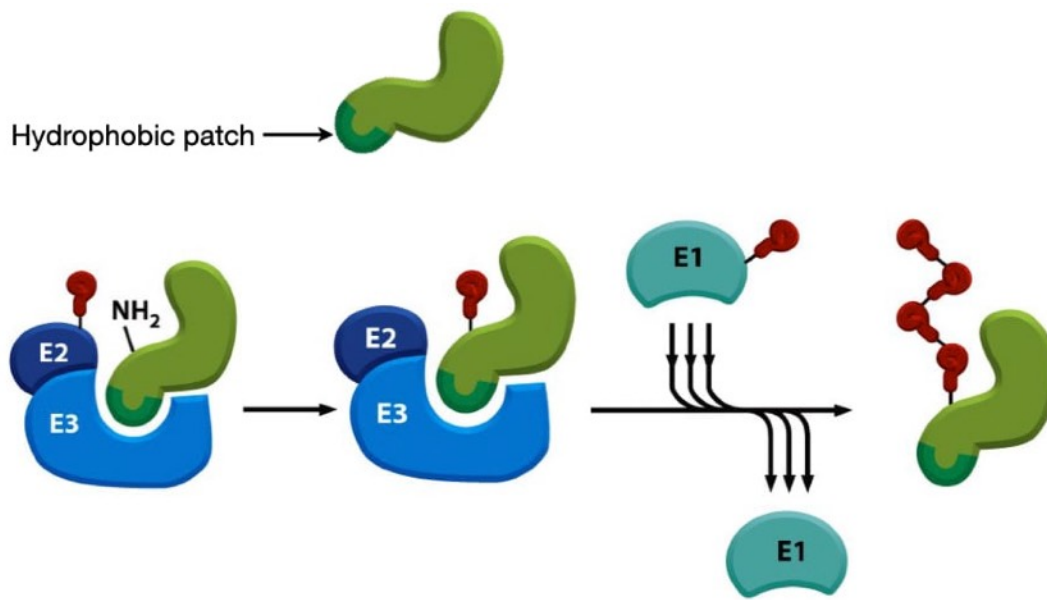
The pattern of ubiquitins on proteins have different biological meanings.

Adding ubiquitin to a protein is a multistep process that requires several enzymes that are divided into three classes: E1, E2 and E3. The E1 enzymes uses ATP to transfer ubiquitin onto itself. The E1 then associates with E2/E3 and transfers ubiquitin to E2. The E3 targets E2 to correct protein. Our genomes encode for more than 30 different E2s and hundreds of E3s allowing cells to target a variety of different proteins.



Three enzymes mediate ubiquitylation of proteins.

E3 recognizes signal on target protein, for example an exposed hydrophobic patch. Once bound to the target and E2 transfers ubiquitin onto the target protein. How a chain of ubiquitin is built is less clear. E1s could transfer more ubiquitin to E2 and E2 could add to the existing ubiquitin on the target protein.

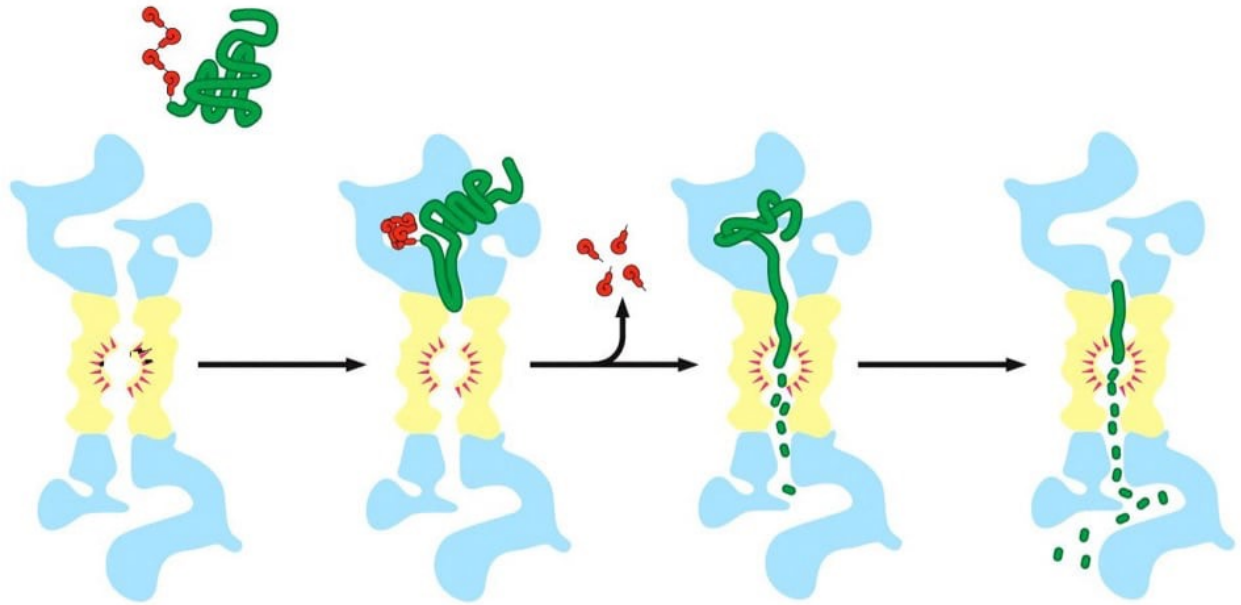


E2 and E3 target specific proteins for ubiquitylation.

When a protein fails to fold during the protein folding process, chaperones will recruit a complex of E2 and E3 ligase to initiate the ubiquitylation of the protein.

Proteasome

The proteasome is responsible for digesting most proteins in the cell and is highly abundant at 1% of total cell protein. The caps on the proteasome recognize a signal on protein that marks it for degradation, often ubiquitin chains. The proteins in the cap unfold the marked protein and then feed it into core region. The core region contains enzymes that digests protein.



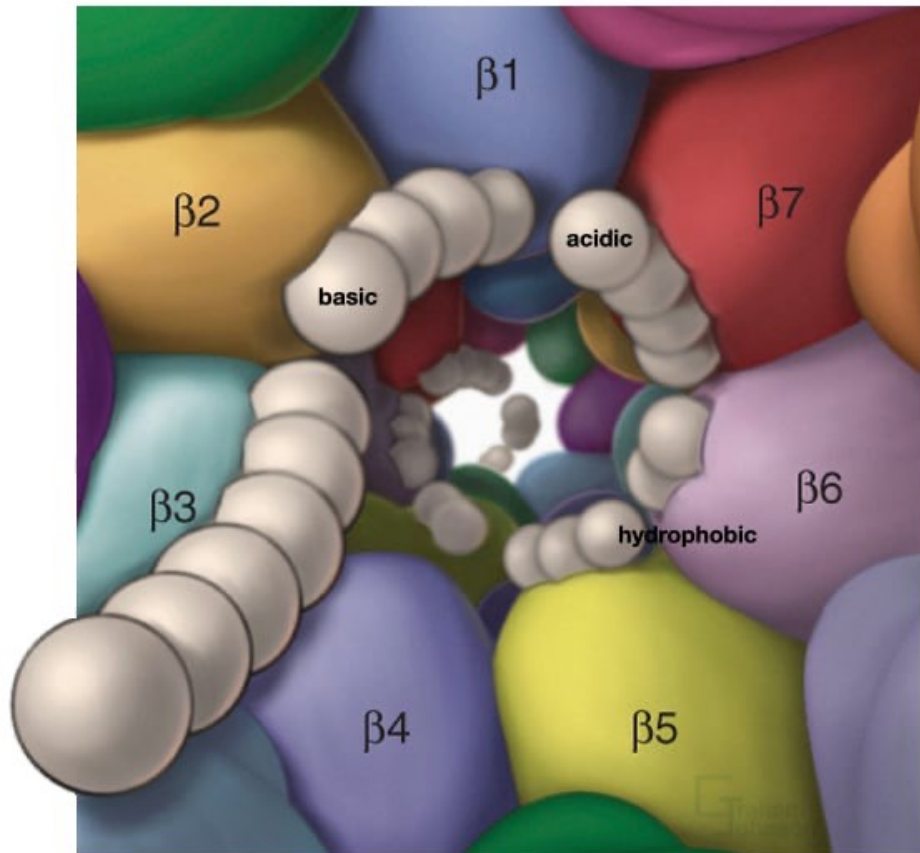
Proteasome

Proteasome is a large complex of proteins that digests ubiquitylated proteins.

The core contains several different types of proteases that cleave proteins after certain type of amino acids:

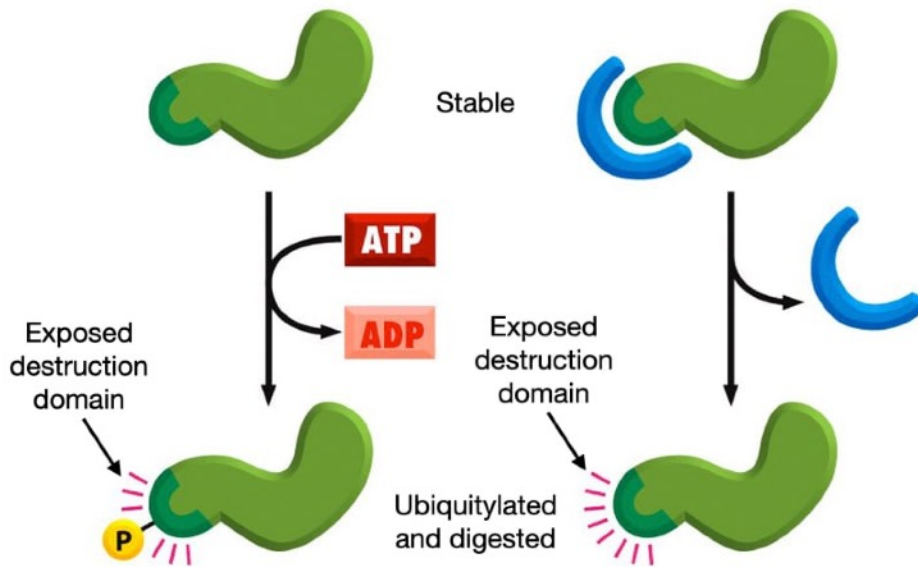
- Beta2 cleaves after basic residues.
- Beta1 cleaves after acidic residues.
- Beta5 cleaves after hydrophobic residues.

These proteases allow the proteasome to digest almost any protein into small peptides.



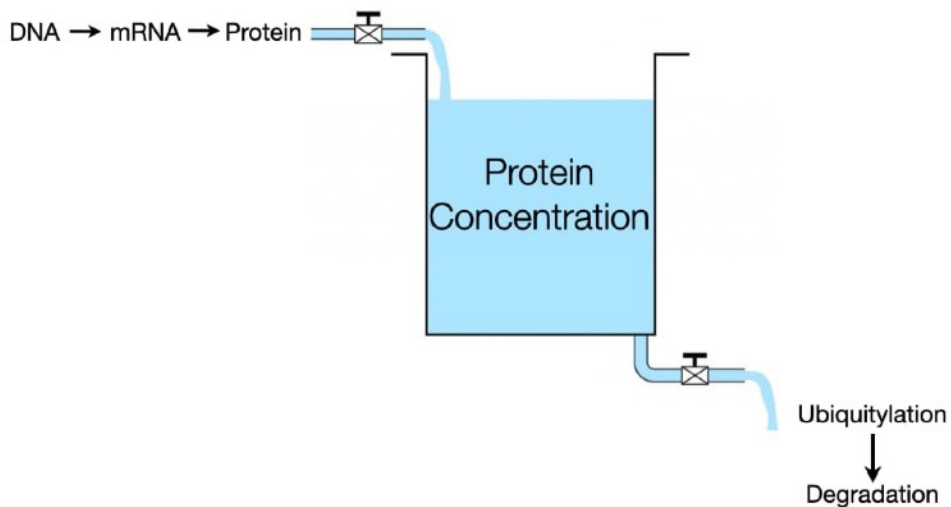
Proteasome contains several proteases that cleave at specific sites in proteins.

Cells use ubiquitin and the proteasome to maintain the concentration of specific proteins within defined ranges. Cells also turn off activity of certain proteins by degraded them. In both of these cases, correctly folded proteins are targeted by the ubiquitin machinery and then degraded by the proteasome. Many proteins contain a destruction domain that is recognized by E2/E3. The domain is hidden while the protein is stable or needed by the cell. Certain signals will expose the destruction domain leading to ubiquitylation and digestion. These signals include adding a phosphate to the domain which makes it recognizable by E2 and E3 or removing another protein that hides the destruction domain.



Ubiquitylation and proteasome also mark and digest folded proteins.

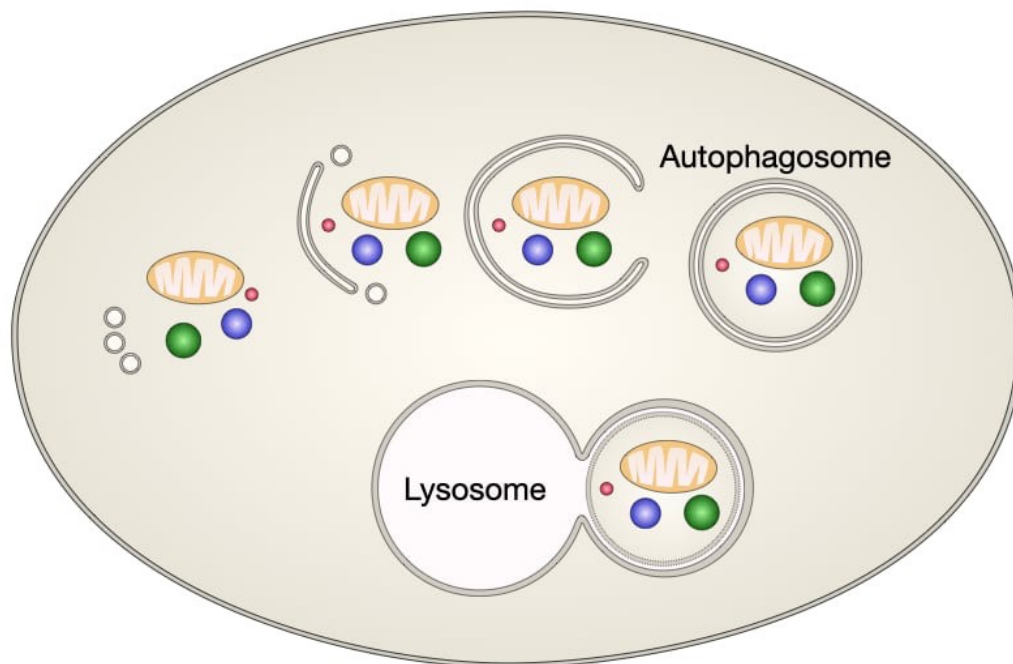
Cells maintain protein concentrations at specific levels by balancing production and destruction of proteins. Most proteins are synthesized at a certain rate and the rate of destruction via ubiquitylation must match the production rate to hold concentration steady. Increases in protein concentration could arise via new synthesis or by inhibiting destruction.



Cells maintain specific concentrations of proteins by balancing synthesis and protein degradation.

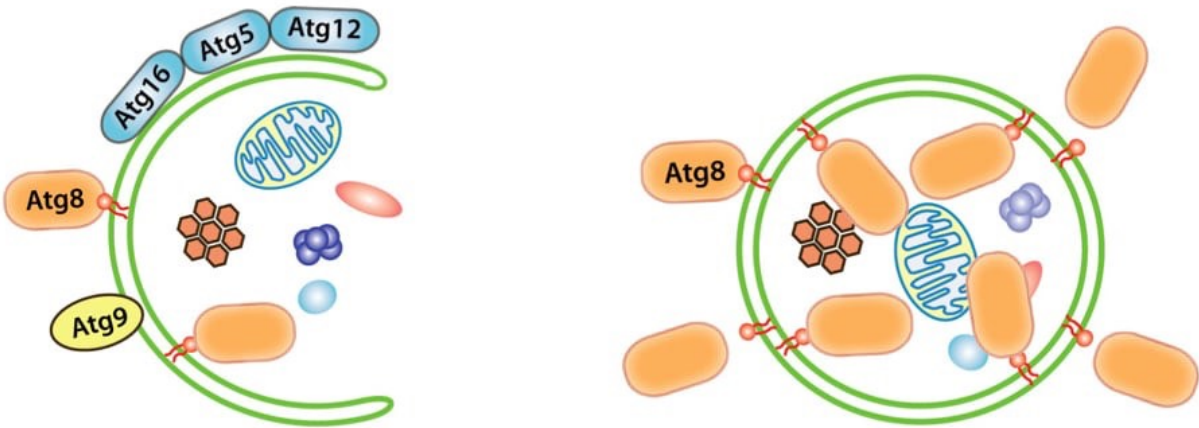
Autophagy

Autophagy or self-eating is the process in which the cell encloses a portion of its cytoplasm in a membrane-bound structure or autophagosome. Autophagosomes fuse with the lysosome where enzymes in the lysosome degrade the cytoplasmic macromolecules and organelles into smaller subunits (i.e. amino acids, nucleotides, sugars). These can be reused to make new macromolecules or catabolyzed to generate energy. Autophagy was originally thought to be used only to generate energy when cells can generate enough ATP by normal mechanisms (i.e. during starvation). During starvation, cells would randomly encompass a portion of their cytoplasm in an autophagosome to be digested. More recently, autophagy has been found to be a way for cells to get rid of old or damaged organelles and protein aggregates in the cytoplasm



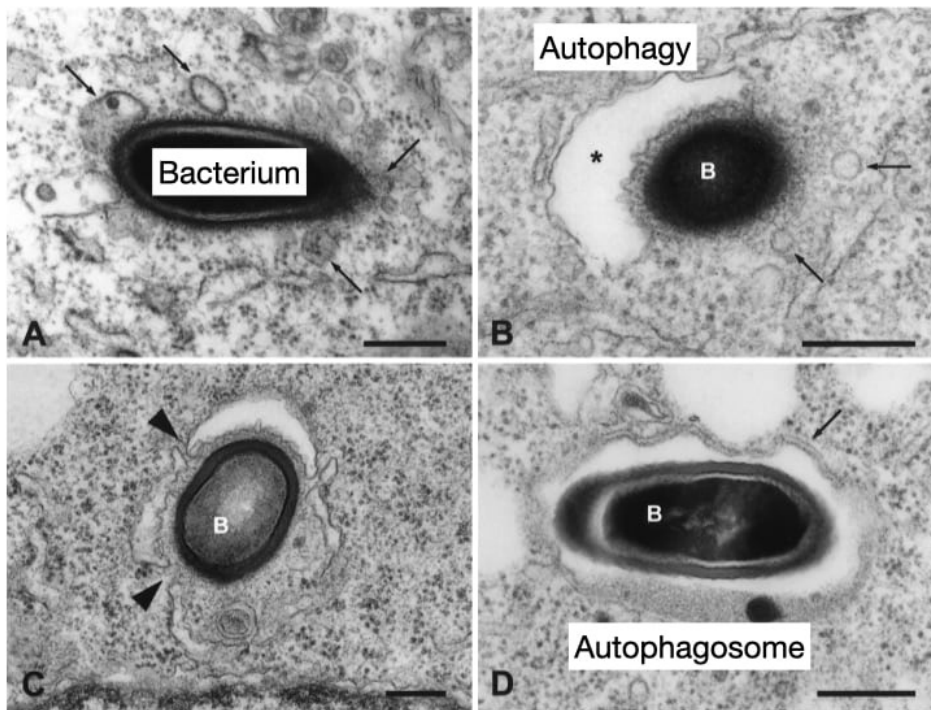
Autophagy mediates digestion of proteins and organelles.

Recent work on the mechanisms of autophagy have found that the process can target specific organelles and other material. A set of proteins called Atg facilitate the process of autophagy and the selection of certain marked organelles, such as mitochondria. Autophagy is also a method by which cells can degrade large aggregates of unfolded protein.



Atg proteins mediate specific and non-specific autophagy of cellular material.

In addition, to organelles and protein, autophagy can also rid cells of intracellular pathogens such as bacteria. For example, listeria can be encompassed in an autophagosome and then targeted to the lysosome for degradation.



Autophagy is part of the innate immune response that rids cells of intracellular pathogens.

Unfolded Protein in the ER

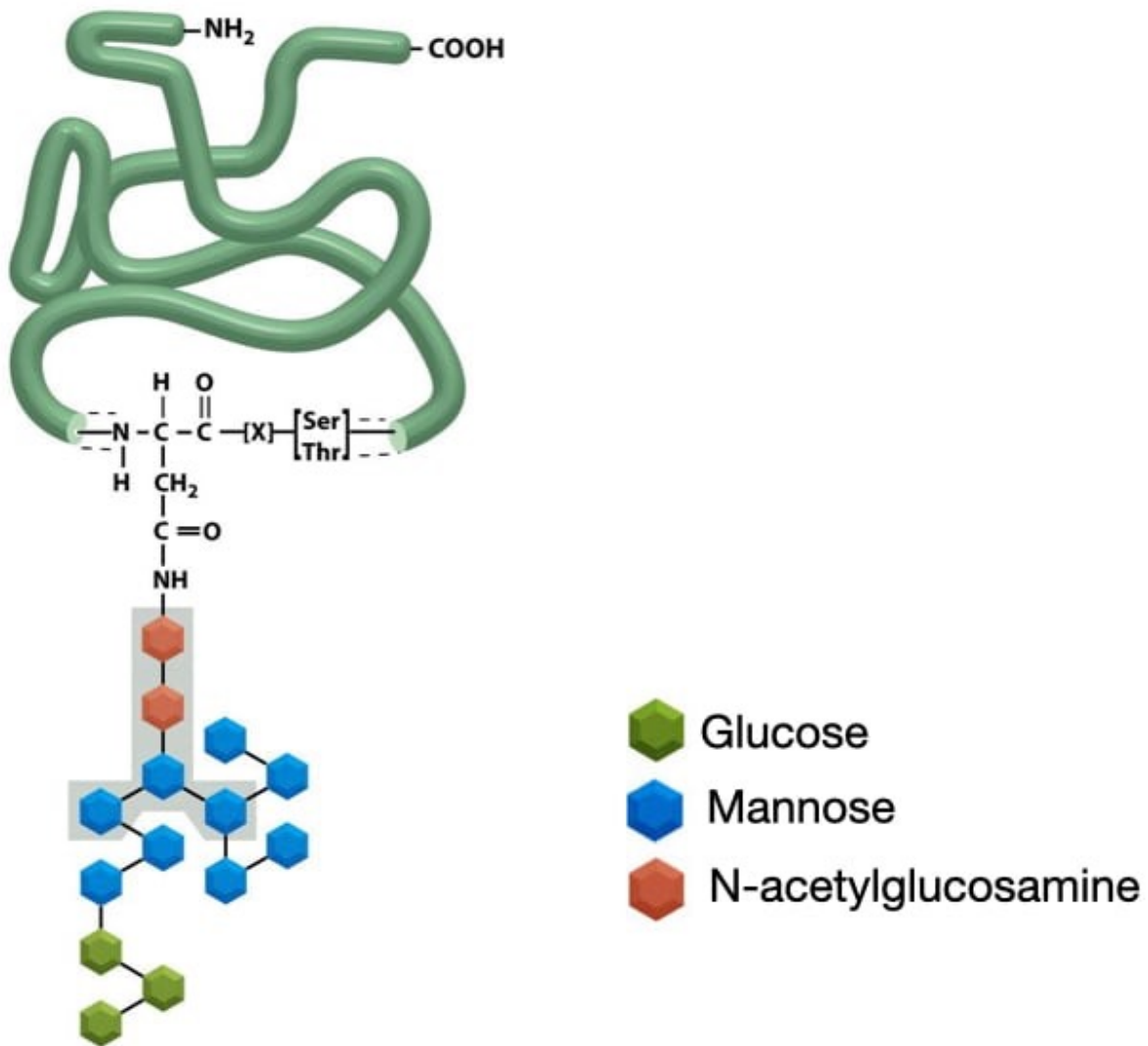
The ER can contain a high concentration of unfolded protein and the accumulation of unfolded protein puts stress on ER. The ER has mechanisms to find unfolded protein and help those proteins fold. The ER also has mechanism to detect the amount of unfolded protein and increase the ability of ER to handle unfolded protein.

Under normal conditions, secreted protein synthesized at basal rate. Often this rate is low enough so that proteins fold rather than aggregate. This allows the protein to proceed through the secretory pathway to its final destination. When stimulated, some cells greatly increase protein production in the ER. This increases the concentration of unfolded protein in ER, making it more likely that unfolded protein will aggregate rather than fold.

One cell type particularly susceptible to build up of unfolded protein in the ER is the beta-cell in the pancreas that produces insulin. When stimulated, 20% of the total mRNA and 30-50% of the total protein synthesis in the β -cell is insulin. The rate of insulin approaches 1 million molecules per minute per cell.

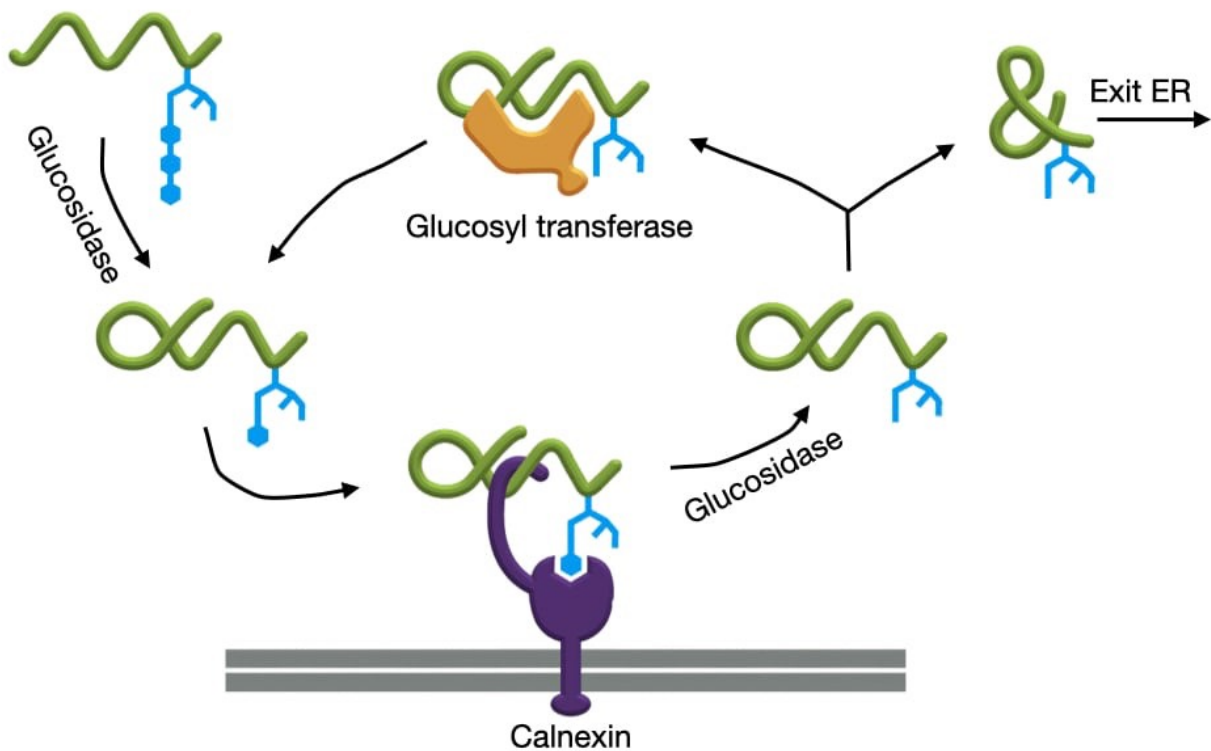
Identifying Unfolded Protein in the ER

Cells need a way to identify unfolded protein in the ER. Surprisingly, the sugar side chains attached to proteins play an important role. Most proteins in ER receive sugars via N-linked glycosylation which adds a tree of 14 sugars. The glucose in the tree will serve as a marker to detect unfolded proteins and target them to chaperones. The mannoses serve as a timer, targeting old, unfolded proteins for destruction.



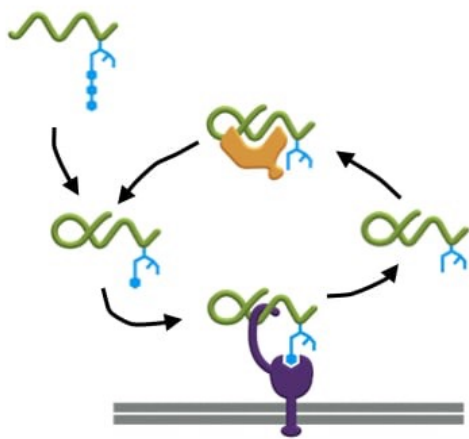
The pattern of sugars on glycosylated proteins marks them as unfolded proteins in the ER.

After synthesis and glycosylation, protein tries to fold. At the same time, glucosidase removes 2 glucoses from the sugar side chain. A side chain with a single glucose is recognized by an ER proteins called calnexin. Calnexin keeps unfolded protein in the ER and recruits chaperones that help protein fold. When glucosidase removes the final glucose, calnexin releases the protein. If the protein folds correctly, it leaves the ER. If not folded, glucosyl transferase binds the exposed hydrophobic domains and adds glucose to sugar side chain. This allows the unfolded protein to rebind calnexin and try to fold again.

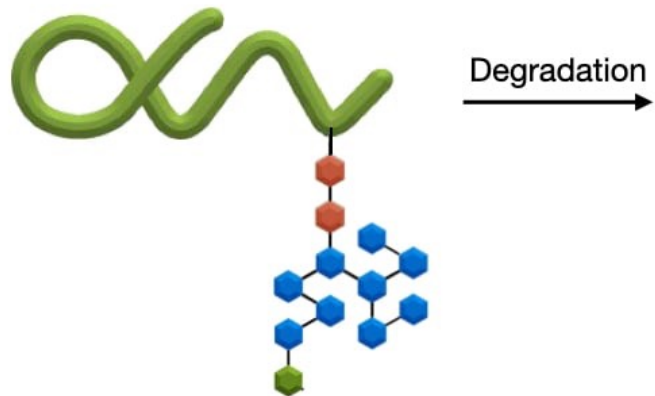


Calnexin and glucosyl transferase prevent unfolded proteins from leaving the ER.

On top of the folding cycle, is the machinery that target proteins for degradation. The ER contains a low concentration of mannosidase. While protein is going through folding cycle, there is a low probability that mannosidase will remove mannose residues. If mannose residues are trimmed, the protein is marked for degradation. The mannoses function as timer for proteins: the longer a protein remains in ER, the more likely it will have mannose removed and then be degraded.



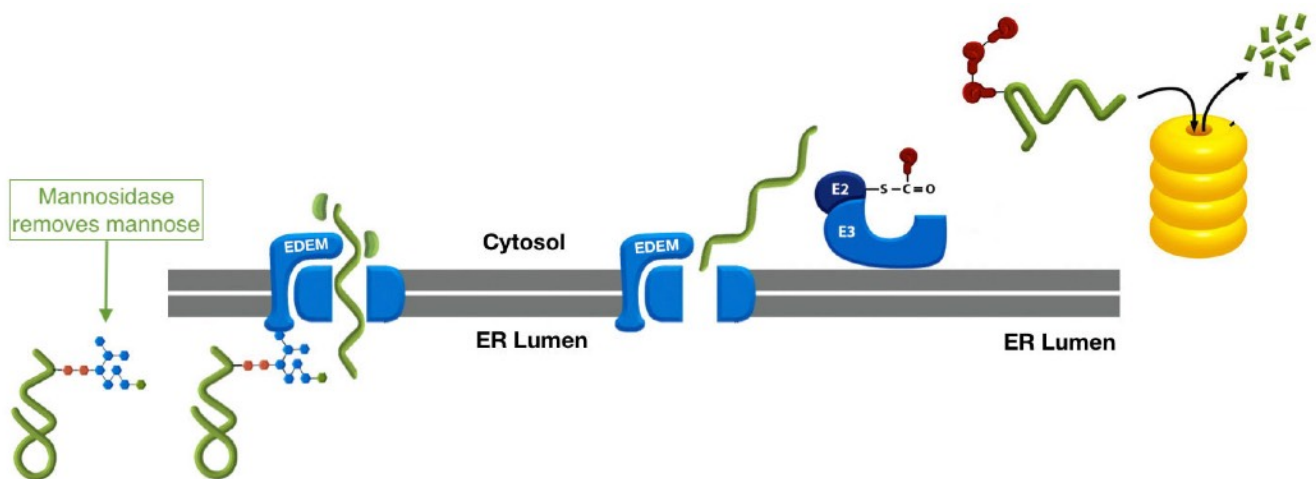
Folding Pathway



Mannosidase

Mannosidase triggers degradation pathway for ER proteins.

To degrade proteins that reside in the ER, cells need a mechanism to get the protein from the ER to the cytosol. EDEM (ER degradation-enhancing α -mannosidase-like protein) is a protein in the ER membrane that binds trimmed mannose. EDEM targets proteins to the retrotranslocator, which is a protein channel that spans the ER membrane. In an energy-dependent process, the unfolded protein is threaded through the retrotranslocator into the cytosol. Chaperones in the cytosol pull the protein through the channel.



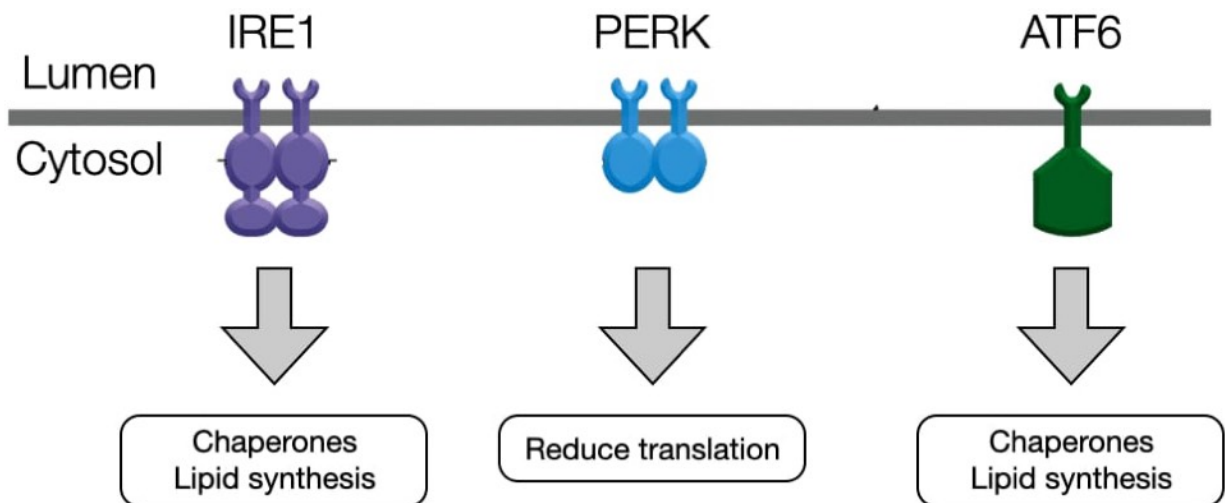
Unfolded proteins are exported from the ER and degraded in the cytosol.

After exiting the ER, ubiquitin ligases on cytoplasmic side of ER membrane catalyze poly-ubiquitylation of protein. Similar to unfolded cytosolic proteins, poly-ubiquitinated ER proteins are digested by the proteasome.

ER Unfolded Protein Response

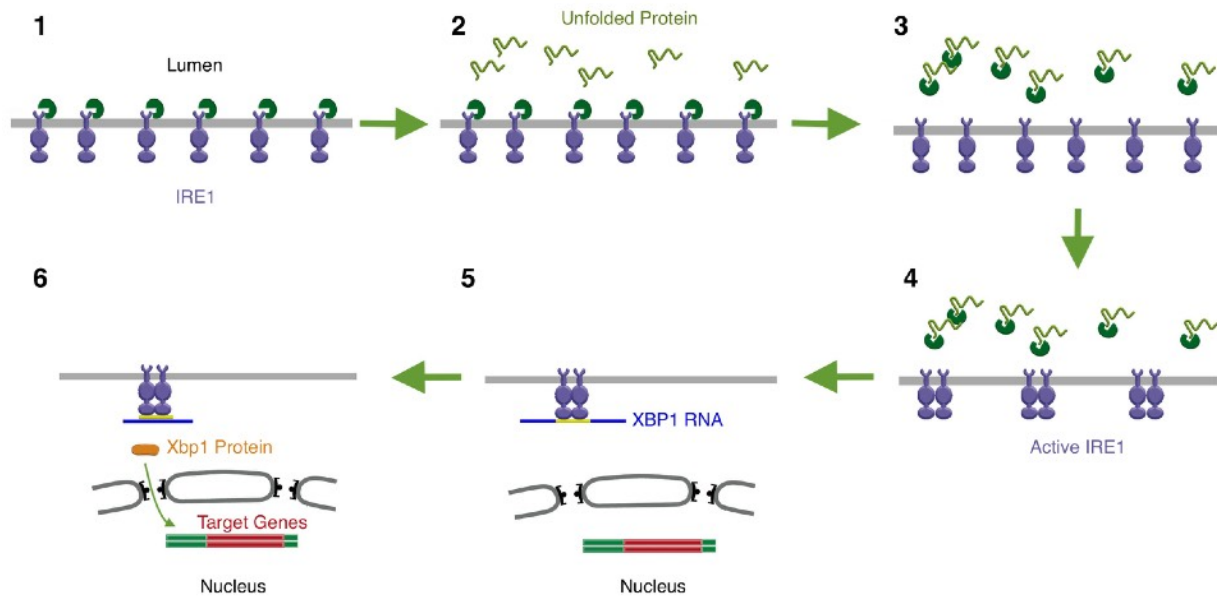
If the ER accumulates too many unfolded proteins, it creates ER stress. ER stress has recently been associated with several diseases, including diabetes, Alzheimer's, kidney disease and inflammatory bowel disease.

ER uses three different proteins to detect unfolded protein in ER: IRE1, PERK, ATF6. When activated by unfolded protein, the receptor elicit a cellular response to alleviate the amount of unfolded protein. PERK reduces global translation by inactivating a key translation initiation factor. IRE1 and ATF6 turn on genes that encode chaperones and lipid biosynthesis enzymes both of which expand the capacity of the ER to fold protein.



Three sensors detect amount of unfolded protein in ER and generate response.

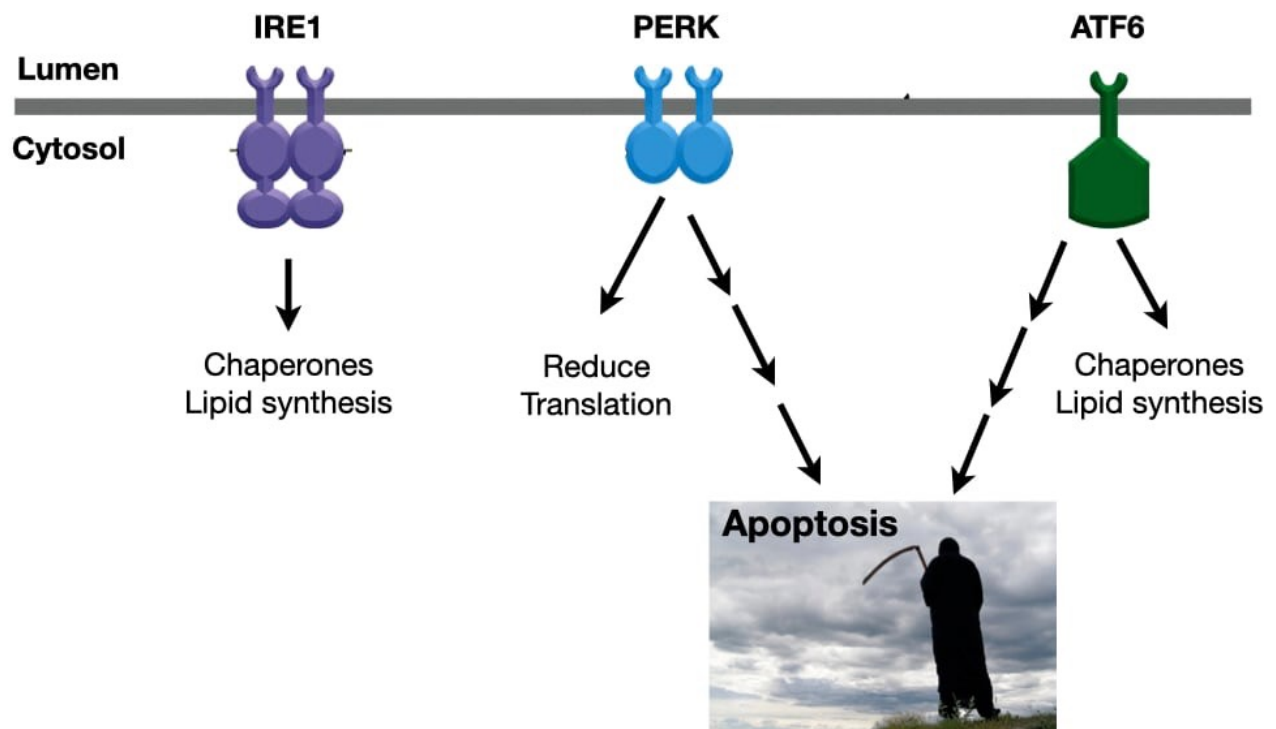
One potential mechanisms by which IRE1 detects unfolded protein is through dimerization. IRE1 is only active as a dimer and chaperones in the ER lumen bind IRE1 preventing dimerization. As the amount of unfolded protein rises in ER, the chaperone disassociates from IRE1 to bind the unfolded protein. IRE dimerizes, leading to its activation.



Unfolded protein removes chaperones from IRE1, leading to dimerization and activation.

How does ER, such as IRE1 protein activate genes in nucleus? IRE1 contains nuclease activity that removes an intron from an RNA called XBP1 in the cytosol, a tRNA ligase joins two exons to produce XBP1 RNA that can be translated to produce a transcription factor. The transcription factor is imported into nucleus and binds to upstream regulatory elements in genes that encode chaperones and lipid biosynthesis enzymes.

When unfolded proteins start to accumulate in the ER, cells respond through the activation of the different receptors and their downstream effectors. However, if these responses are insufficient and the amount of unfolded protein remains high, the receptors will trigger an alternative pathway that leads to cell death via apoptosis. Thus, prolonged ER stress can lead to death of cells.



Prolonged activation of PERK and ATF6 trigger the apoptosis pathway.