Lecture 5. Post-translational protein modifications and folding.

Learning outcomes:
1. Draw a functional connection between primary structure and higher-order spatial organization of polypeptides.
2. Explain the auxiliary role of chaperones in protein folding.
3. Give detailed examples of human disorders linked with protein misfolding.

Within the last few decades, scientists have discovered that the human proteome is vastly more complex than the human genome. While it is estimated that the human genome comprises between 20,000 and 25,000 genes, the total number of proteins in the human proteome is estimated at over 1 million. These estimations demonstrate that single genes encode multiple proteins. Genomic recombination, transcription initiation at alternative promoters, differential transcription termination, and alternative splicing of the transcript are mechanisms that generate different mRNA transcripts from a single gene.
The increase in complexity from the level of the genome to the proteome is further facilitated by protein post-translational modifications (PTMs). PTMs are chemical modifications that play a key role in functional proteomic because they regulate activity, localization, and interaction with other cellular molecules such as proteins, nucleic acids, lipids and cofactors.

Types of post-translational modification

There are many typs of protein modification, which are mostly catalyzed by enzymes that recognize specific target sequences in proteins. These modifications regulate protein folding by targeting specific subcellular compartments, interacting with ligands or other proteins, or by bringing about a change in their functional state including catalytic activity or signaling. The most common PTMs are:

Based on the addition of chemical groups

- Phosphorylation
- Acetylation
- Hydroxylation
- Methylation

Based on the addition of complex groups

- Glycosylation
- AMPylation
- Lipidation

Based on the addition of polypeptides

- Ubiquitination

Based on the cleavage of proteins

- Proteolysis

Based on the amino acid modification

- Deamidation

Chemical groups

Phosphorylation

Reversible phosphorylation of proteins involves addition of a phosphate group on serine, threonine, or tyrosine residues and is one of the important and extensively studied PTM in both prokaryotes and eukaryotes.
Several enzymes or signaling proteins are switched ‘on’ or ‘off’ by phosphorylation or dephosphorylation. Phosphorylation is performed by enzymes called ‘kinases’, while dephosphorylation is performed by ‘phosphatases’.
Addition of a phosphate group can convert a previously uncharged pocket of protein into a negatively charged and hydrophilic protein thereby inducing conformational changes in the protein.

Phosphorylation has implications in several cellular processes, including cell cycle, growth, apoptosis and signal transduction pathways. One example is the activation of p53, a tumor suppressor protein. p53 is used in cancer therapeutics and is activated by phosphorylation of its N-terminal by several kinases.

Acetylation

Acetylation refers to addition of acetyl group in a protein. It is involved in several biological functions, including protein stability, location, synthesis; apoptosis; cancer; DNA stability. Acetylation and deacetylation of histone form a critical part of gene regulation.

Acetylation of histones reduces the positive charge on histone, reducing its interaction with the negatively charged phosphate groups of DNA, making it less tightly wound to DNA and accessible to gene transcription. Acetylation of p53, a tumor suppressor gene, is crucial for its growth suppressing properties.

Hydroxylation

This process adds a hydroxyl group (-OH) to the proteins. It is catalyzed by enzymes termed as 'hydroxylases' and aids in converting hydrophobic or lipophilic compounds into hydrophilic compounds.

Methylation

Methylation refers to addition of a methyl group to lysine or arginine residue of a protein. Arginine can be methylated once or twice, while lysine can be methylated once, twice, or thrice. Methylation is achieved by enzymes called methyltransferases. Methylation has been widely studied in histones wherein histone methylation can lead to gene activation or repression based on the residue that is methylated.

Complex groups

Glycosylation

Glycosylation involves addition of an oligosaccharide termed 'glycan' to either a nitrogen atom (*N*-linked glycosylation) or an oxygen atom (*O*-linked glycosylation). N-linked glycosylation occurs in the amide nitrogen of asparagine, while the *O*-linked glycosylation occurs on the oxygen atom of serine or threonine.

Carbohydrates present in the form of N-linked or O-linked oligosaccharides are present on the surface of cells and secrete proteins. They have critical roles in protein sorting, immune recognition, receptor binding, inflammation, and pathogenicity. For example, N-linked glycans on an immune cell can dictate how it migrates to specific sites. Similarly, it can also determine how a cell recognizes 'self' and 'non-self'.

AMPylation

AMPylation refers to reversible addition of AMP to a protein. It involves formation of a phosphodiester bond between the hydroxyl group of the protein and the phosphate group of AMP.

Lipidation

The covalent binding of a lipid group to a protein is called lipidation. Lipidation can be further subdivided into prenylation, N-myristoylation, palmitoylation, and glycosylphosphatidylinositol (GPI)-anchor addition.

Prenylation involves the addition of isoprenoid moiety to a cysteine residue of a substrate protein. It is critical in controlling the localization and activity of several proteins that have crucial functions in biological regulation.

Myristoylation involves the addition of myristoyl group to a glycine residue by an amide bond. It has functions in membrane association and apoptosis. In palmitoylation, a palmitoyl group is added to a cysteine residue of a protein.

In GPI-anchor addition, the carboxyl-terminal signal peptide of the protein is split and replaced by a GPI anchor. Recent research in human genetics has revealed that GPI anchors are important for human health. Any defects in the assembling, attachment or remodeling of GPI anchors lead to genetic diseases known as inherited GPI deficiency.

Polypeptides

Ubiquitination

Ubiquitination involves addition of a protein found ubiquitously, termed ‘ubiquitin’, to the lysine residue of a substrate. Either a single ubiquitin molecule (monoubiquitination) or a chain of several ubiquitin molecules may be attached (polyubiquitination).

Polyubiquitinated proteins are recognized by the 26S proteasome and are subsequently targeted for proteolysis or degradation. Monoubiquitinated proteins may influence cell tracking and endocytosis.

Protein cleavage

Proteolysis

Proteolysis refers to breakdown of proteins into smaller polypeptides or amino acids. For example, removal of N-terminal methionine, a signal peptide, after translation leads to conversion of an inactive or non-functional protein to an active one.

Amino acid modification

Deamidation

Deamidation is the removal or conversion of asparagine or glutamine residue to another functional group. Asparagine is converted to aspartic acid or isoaspartic acid, while glutamine is converted to glutamic acid or pyroglutamic acid. This modification can change the protein structure, stability, and function.

Molecular chaperones are present in all organisms and are essential for cell survival. One of the major functions of molecular chaperones is to facilitate protein folding. Although the amino acid sequence of a protein contains the information required to adopt the native conformation, not all proteins can fold spontaneously. Unfolded polypeptides are generated during normal growth as the product of protein synthesis, but misfolded proteins arise as a consequence of cellular stresses, such as heat shock, oxidative stress, as well as pathological conditions. Molecular chaperones, including Hsp60s, Hsp70s, Hsp90s and sHsps, assist in the folding of unfolded and misfolded polypeptides by stabilization of folding intermediates and prevention of protein misfolding and aggregation. Several chaperones also function to reactivate aggregated proteins. For example both Clps and Hsp90s work with Hsp70s to salvage aggregated proteins. Additionally, some chaperones, such as Clps, interact with specific proteases and deliver unfolded and non-native proteins to compartmentalized proteases for degradation. There are also dedicated chaperones that act on one or a small number of proteins. The combined action of molecular chaperones therefore increases the cellular pool of native proteins while minimizing inactive proteins and potentially harmful protein aggregates. Since molecular chaperones have important roles in protein folding and remodeling, modulation of chaperone activity is linked to numerous diseases including cancer and amyloid disorders.

The questions for self - control:

1. What are the primary, secondary, tertiary and quaternary structures of proteins?
2. What is the protein folding and which enzymes help to provide this process?
3. Can you give the detailed examples of human diseases connected with protein misfolding?

Recommended readings:

1. Alberts et al., pp. 109-134;
2. Alberts et al., pp. 333-362;
3. Lodish et al., pp. 67-87.