Posttranslational Modification

11-10-05
Alteration of the Chain Termini

a. N-terminal Met
b. Addition of Terminal residues
c. Acetylation of N-terminus
d. Myristoylation of N-terminus
e. Glycosyl-phosphatidylinositol and Farnesyl Membrane Anchors at the C-terminal
f. Amidation of The C-terminus
g. Reversible Removal of the C-terminal Tyr of α-tubulin
a. N-terminal Met

The formyl group on the initiating Met residue of poly-peptides that are synthesized in prokaryotes is almost always removed by a deformylase enzyme. Only rarely is N-formyl Met found at the N-terminus of a mature protein.

In about half the proteins of both prokaryotic and eukaryotic cells, the initiating Met residue is removed from the nascent chain by a ribosome-associated Met-aminopeptidase. Whether it is removed depends primarily on the second amino acid residue. Small residues (Gly, Ala, Ser, Cys, and Thr) favor removal of the Met residue in prokaryotes; large, hydrophobic, and charged residues seem to prevent removal.

The enzyme responsible for removal of the Met residue may be saturated in prokaryotes if a protein is being synthesized in large quantities, or the terminus may be inaccessible if the protein is aggregated into inclusion bodies. Such proteins are often made as a mixture, with and without the N-terminal Met. Usually, only the Met residue is removed from the N-terminus, but in some cases an additional residue is also removed. The mechanism for this is not known.
b. Addition of Terminal Residues

- The only known instances of posttranslational addition of residues to the ends of polypeptide chains have been shown to occur \textit{in vitro} by transfer of residues from charged transfer RNA to the \(\alpha\)-amino groups of some peptides and proteins.

- No other components of the protein biosynthetic reaction are required, and the reactions are presumed to occur primarily in the cytoplasm.

- For example, a single enzyme from \textit{E. coli} transfers Leu and Phe residues from their respective tRNA molecules to the \(\alpha\)-amino groups of proteins with Arg or Lys residues at the N-terminus. A similar mammalian enzyme, arginyl tRNA-protein transferase, catalyzes the transfer of Arg to peptides with N-terminal Glu or Asp residues.

- No functions for these reactions, which have been demonstrated only \textit{in vitro}, were known for a long time, but they have recently been implicated in tagging proteins for degradation.
c. Acetylation of N-terminus

- Some 59-90% of eukaryotic proteins synthesized in the cytoplasm are isolated with their N-termini acetylated:

  \[
  \begin{array}{c}
  0 \\
  \hline
  \text{CH}_3-C-NH-
  \end{array}
  \]

- A variety of N\(^\alpha\)-acetyl transferase enzymes are thought to catalyze this reaction, using acetyl-CoA as the acetyl donor. The principal enzyme is loosely associated with ribosomes, consistent with the observation that the nascent chain of only 20-50 residues is usually acted on.

- Acetylation can occur whether or not the initiating Met residue is still present. Whether acetylation occurs depends to some extent on the nature of the N-terminal residue. In a survey using mutagenesis of the N-terminus of one particular protein, those forms acetylated had N-terminal Gly, Ala, Ser, and Thr residues.

- The initiating Met was retained and acetylated if the following residue was Asp, Glu, or Asn. Nevertheless, many exceptions to these rules are found in various proteins, and so it must be other properties of the protein that determine whether or not it is acetylated.
One important aspect is whether the acetylating enzyme is present. Acetylation of nascent chains is apparent only in cytoplasmic proteins because the N-terminal targeting peptides of other proteins are subsequently removed.

In other cases, acetylation can occur posttranslationally. For example, the melanocyte-stimulating hormone (α-MSH) is acetylated only in some parts of the pituitary. In contrast, another hormone, endorphin, shows just the opposite acetylation properties. In these cases, acetylation occurs only after cleavage of the hormone from a larger precursor, and the responsible acetyl transferase is present only in secretory granules. Another complication is the presence of enzymes that can remove acetylated residues from the amino end of polypeptides. Acetylation is common only for proteins made in eukaryotic cells, but not in their mitochondria or chloroplasts.
The initiating Met residue is removed from some proteins only after it is acetylated, and then the newly exposed N-terminus is acetylated.

This process may take place posttranslationally rather than during synthesis of the polypeptide chain. Not all the factors that determine whether or not the N-terminus of a protein is acetylated are known, nor is the function of acetylation.

For the protein chemist, the primary consequence of acetylation is to make the protein refractory to sequencing by the Edman degradation or by aminopeptidase.

Although acetylation is the main covalent modification made to the amino ends of proteins, a great variety of other modifications have been observed in particular cases, which include addition of formyl, pyruvoyl, fatty acyl, α-keto acyl, glucuronyl, and methyl groups.
Another common modification of the N-terminus is observed when the first residue is Gln.

In this case, the side chain reacts with the amino group to generate the pyroglutamyl residue, which has no amino group and is also refractory to sequence analysis. This modification reaction occurs spontaneously, but there are also enzymes that catalyze it *in vitro*.

The enzymes appear to be involved in the posttranslational processing of peptide hormones, which are usually synthesized as large precursors.

It seems likely, though it is not yet proven, that the state of the amino terminus of a protein has a large effect on the rate of degradation of the protein.
A number of cytoplasmic proteins are found with the myristoyl fatty acid linked to their terminal α-amino group:

\[
H_3C-(CH_2)_{12}-C(=O)-NH-C(=O)-
\]

The myristoyl group is added to the α-amino group during or shortly after biosynthesis of the chain.

The myristoyl group is transferred from myristoyl-CoA by the enzyme myristoyl-CoA: protein N-myristoyl transferase.

The information determining whether a protein is myristoylated is thought to reside in the first 7-10 residues of the polypeptide chain, but the rules are not precisely defined.
Proteins to which the myristoyl group is added always have Gly as the N-terminal amino acid, and this Gly is the one following the initiating Met residue. The next residue is almost always a small, uncharged residue (Ala, Ser, Asn, Gin, or Val).

A broader spectrum of residues occurs at positions 3 and 4, but residue 5 must be small and uncharged (Ala, Ser, Thr, Cys, or Asn).
The long hydrophobic myristoyl group is believed to cause at least some of the attached proteins to be loosely associated with membranes, but these proteins are basically soluble and are not tethered tightly to the membrane.

Other myristoylated proteins are not associated with membranes to any significant extent. The consequences of myristoylation are different from those of other lipid attachments, which cause the proteins to be firmly anchored in the membrane.

Some myristoylated proteins are protein kinases or phosphatases that have important roles in modulating cellular metabolism.

The myristoyl group may tend to keep its proteins in juxtaposition to particular membranes or other components of cellular regulatory circuits.
Some proteins localized on the cell surface have been found to be firmly anchored to the cell membrane through complex glycosyl-phosphatidylinositols attached to their terminal $\alpha$-carboxyl groups.

The structures of these anchor groups are very complex involving an ethanolamine group that is attached to the protein $\alpha$-carboxyl in an amide bond, through a complex glycan and an inositol phosphate to a diacylglycerol lipid that is embedded in the membrane.

The mode of assembly of this complex structure is not known, but it is thought to be preassembled and to be transferred to the polypeptide chain soon after completion of its translation in the endoplasmic reticulum.
Proteins that are modified in this way are synthesized with signal peptides and are directed into the ER.

They are intrinsically soluble, hydrophilic proteins, except for a hydrophobic tail at the carboxyl end of the chain.

This sequence, plus the 10 -12 preceding residues, is cleaved from the protein at about the same time that the glycosyl phosphatidylinositol group is added to the new terminal carboxyl group.

The sequence that is removed appears to be the signal for the modification, analogous to signal peptides for translocation to the ER.

Adding such a sequence to a protein that is normally secreted causes it to be modified and anchored to the membrane.
The primary purpose of this modification appears to be to anchor otherwise water-soluble proteins to the cell membrane.

The hydrophobic C-terminal peptide that is replaced does not seem to be a suitable transmembrane anchor.

One advantage of the glycosylphosphoinositide type of anchor over that provided by transmembrane segments in the polypeptide chain is that it permits the protein to diffuse much more rapidly within the plane of the membrane. Also, the glycosylphosphoinositide linkage can be hydrolyzed by specific enzymes, which suggests that anchoring might be regulated.

Another intriguing possibility arises from the involvement of phosphoinositides; their breakdown products after cleavage by phospholipase C have been identified in recent years as important second messengers that are produced in the cell in response to triggers at the cell surface. Cleavage of the glycosylphosphoinositol-anchored proteins produces the same second messengers, but outside the cell. This led to the suggestion that a similar phenomenon might be involved in the interaction of the anchored molecules with ligands or receptors at the cell surface.
Farnesyl and geranylgeranyl groups are attached to Cys residues at the C-termini of proteins synthesized with C-terminal sequences of the general type -Cys-Axx-Axx-Axx-Xxx, where Axx is a residue of an aliphatic amino acid (most often Leu, Ile, or Val) and Xxx is any C-terminal amino acid residue.

The three terminal residues are removed, the farnesyl or geranylgeranyl group is attached to the sulfur atom of the Cys side chain, and the $\alpha$ -carboxyl group is methylated.

Farnesyl groups are 15-carbon lipids that are derived from farnesol, which is a branched-chain, polyunsaturated hydrocarbon alcohol intermediate of sterol biosynthesis; geranylgeranyl groups are similar but longer, with 20 carbon atoms.

This type of modification is crucial for the biological activities of at least some especially important proteins, such as the products of the ras oncogenes, but the chemical basis is not known.
Structure of the glycosyl-phosphatidylinositol structures that anchor the C-termini of proteins to membranes. The C-terminus of the protein is shown (top), linked by an amide bond to an ethanolamine moiety. The structure of the glycan is indicated schematically (Man, mannose; Gal, galactose; GlcNH₂, glucosamine); the numbers refer to the atom numbers through which the covalent bonds occur. The myristyl fatty acids shown can be replaced by others; they are embedded in one layer of the lipid bilayer of the membrane. These structures can vary somewhat, particularly that of the glycan.
f. Amidation of C-Terminus

- A C-terminal amide group in place of the usual carboxyl group is a characteristic feature of many peptide hormones. It is often important for biological activity and contributes to biological stability of the hormone.

- The amide group is derived from a Gly residue originally present at the C-terminus: The two reactions are catalyzed by two sequential enzymes and involve ascorbic acid (vitamin C), copper, and oxygen. The hydroxylated intermediate is highly unstable.

- In many cases, the Gly residue modified in this way occurs at the C-terminus only after proteolytic cleavage of a larger precursor. Proteolytic processing occurs in the Golgi apparatus and maturing secretory granules. The enzymes that catalyze the amidation reaction are in the secretion granules.
g. Reversible Removal of the C-Terminal Tyr of \( \alpha \)-Tubulin

- Removal of the C-terminal residue in \( \alpha \)-tubulin is special because the removal is reversible. Virtually all \( \alpha \)-tubulins are synthesized in the cytoplasm with a Tyr residue at the C-terminus.

- This residue is removed in vivo by the enzyme tubulinyl tyrosine carboxypeptidase but is reinstated by another enzyme, tubulinyl tyrosine ligase. Addition of the Tyr residue to \( \alpha \)-tubulin requires only the free amino acid and ATP.
The details are not certain, but it appears that this reversible modification of α-tubulin is related to the role of this protein in forming microtubules.

Tubulin exists as a soluble heterodimer of α- and β-tubulins and reversibly aggregates into filamentous microtubules.

The tubulin carboxypeptidase acts preferentially on the polymerized protein, so microtubules gradually become depleted of the C-terminal Tyr residue on α-subunits.

The ligase acts primarily on soluble, monomeric tubulin, so monomers with Tyr predominate *in vivo* and are used primarily in assembling microtubules. Consequently, there is a cycle in which Tyr residues that are added to the soluble tubulin subunits are gradually removed after assembly of the monomers into microtubules.

Removal of the Tyr residues affects the dynamic properties of the microtubules, but the precise relationship is not certain.
Microtubules are involved in many cellular functions such as mitosis, morphogenesis, motility, and intracellular organelle transport.

To perform so many functions simultaneously, the microtubules may need to be differentiated.

The presence or absence of the C-terminal Tyr residue may be one such marker, but it is clear that additional modifications of tubulin may also be significant, such as acetylation of the N-terminus. In addition, a variety of closely related α- and β-tubulins are synthesized by most organisms.
2.4.3 Glycosylation

- Attachment of carbohydrates is one of the most prevalent posttranslational modifications of eukaryotic proteins, especially of secreted and membrane proteins, yet the process has no well-defined universal purpose.
- Indeed, the biological activities of many glycoproteins are not detectably different if the carbohydrates are removed, and glycosylation of proteins does not occur at all in prokaryotes.
- Those functions of the carbohydrates that have been detected thus far appear to be specific to each protein.
- In a few cases the carbohydrates are involved in biological activity of the protein, but they are more often important for its physical properties, such as solubility.
- Carbohydrates often lengthen the biological life of a protein by decreasing its rate of clearance from the serum. Being on the surfaces of proteins, the carbohydrates are often involved in interactions with other cells or molecules, such as immunoglobulins, cell-surface receptors, and proteases.
- The most relevant properties of glycosyl groups attached to proteins are (1) their variable structures, which permit specificity in their interactions with other molecules; (2) their hydrophilic natures, which keep them in aqueous solution; and (3) their bulk, which markedly affects the surface properties of the protein to which they are attached.
There are two types of glycosylation, called N-type or O-type depending on the atom of the protein to which the carbohydrate is attached.

N-type glycosylation occurs exclusively on the nitrogen atom of Asn side chains, whereas O-glycosylation occurs on the oxygen atoms of hydroxyls, particularly those of Ser and Thr residues.

N-glycosylation occurs cotranslationally soon after the Asn residue emerges into the ER.

Whereas O-glycosylation occurs primarily in the Golgi as a posttranslational modification.

Description of glycosylation is made difficult by the complexity of the carbohydrate structures attached to the proteins.

Not only are at least eight different sugar monomers used — galactose, glucose, fucose, mannose, N-acetyl galactosamine, N-acetyl glucosamine, sialic acids, and xylose—but they are joined by a variety of glycoside linkages, between their various functional groups.

The details of carbohydrate chemistry are omitted in the following discussion, and emphasis is placed on the protein, though this provides only half of the story.
FIGURE 2.20
Structures of carbohydrate–protein linkages commonly found in proteins. GlcNAc-Asn: N-acetylglucosaminyl-Asn (N-glycosylation); Xyl-Ser: xylosyl-Ser (linkage to glycosaminoglycan in proteoglycans); GalNAc-Ser: N-acetylgalactosaminyl-Ser; Gal-Hyl: galactosyl hydroxylysine (O-glycosylation).
a. N-Glycosylation of Asn Residues

- Carbohydrate to be attached to Asn residues is pre-assembled as a core structure attached to a membrane lipid, dolichyl phosphate.

- The assembly of this core structure in the ER by membrane-bound enzymes is the step that is blocked by the commonly used inhibitor of glycosylation, tunicamycin.

- This method of attachment of a preformed core structure is probably used because the target Asn residue is encountered only transiently as it emerges through the ER membrane.
FIGURE 2.21
Structures of typical oligosaccharide chains attached to proteins. I is the polymannose core oligosaccharide transferred to Asn residues in the sequence —Asn-Xaa-Ser/Thr/Cys— from the lipid pyrophosphoryldolichol; the original lipid-linked oligosaccharide contained an additional three glucose residues that are removed after transfer to the protein. II is a typical product of further processing of core structure I, in this case to produce a "complex" type of carbohydrate. In high-mannose types, additional mannose residues are added to the tips of the core structure I. Hybrid types, in which one branch is complex and another is high man- nose, are also found. III is a typical O-glycosidically linked structure, added to a Ser residue.

The abbreviations of the sugar residues are Man, mannose; Gal, galactose; GlcNAc, N-acetylglucosamine; GalNAc, N-acetylgalactosamine; and SA, sialic acid. The positions and anomeric configurations of the carbohydrate linkages are omitted for simplicity. (From J. A. Hanover and W. J. Lennarz, Arch. Biochem. Biophys. 211:1–19, 1981.)
The Asn residue that is glycosylated always occurs in a characteristic sequence: -Asn-Xaa-Ser-, -Asn-Xaa-Thr-, or -Asn-Xaa-Cys-. Xaa can be any residue except Pro, which also cannot immediately follow the tripeptide sequence.

This characteristic sequence, however, is not the only determinant for glycosylation because not all such Asn residues of proteins that enter the ER are modified. For example, proteins such as ovalbumin and deoxyribonuclease are fully glycosylated at a single site but contain at least one additional sequence that is not glycosylated.

Some proteins, such as bovine pancreatic ribonuclease, are glycosylated at an appropriate Asn residue in only a fraction of the molecules. Others, such as pancreatic elastase and carboxypeptidase, contain one or more potential glycosylation sites but are not glycosylated at all, even though they are synthesized and secreted by the same pancreatic cells that glycosylate other proteins.

Because glycosylation occurs to the nascent chain in the ER, the primary structure would be expected to be the primary determinant, but what distinguishes between the different Asn residues in the same tripeptide sequence is a major unsolved problem.
After attachment of a core glycan to the Asn residue of a protein in the ER, the glycan is extensively modified during passage of the protein through the ER and Golgi.

In some cases this modification primarily involves attachment of more mannose groups; in other cases a more complex type of structure is attached.

In the case of lysosomal proteins, the core mannose groups are phosphorylated.

The type of processing depends on the identity of the protein, the cell type, and the physiological state of the cell.

The processing of the glycosyl groups is typically variable, so an N-glycosylated protein is usually heterogeneous in its carbohydrate groups.

The various modifications to the core glycan take place in various parts of the ER and the Golgi, so the state of its carbohydrate is an excellent marker of where in the cell a protein has traveled.
b. O-Glycosylation

- Attachment of carbohydrates to the oxygen atoms of amino acid side chains occurs primarily in the Golgi apparatus.

- N-Acetyl galactosamine (GalNAc) groups are attached to Ser and Thr groups of certain proteins.

In the case of collagen, hydroxy-Lys (Hyl) and hydroxy-Pro residues are also modified in this way.
The signals that determine which Ser and Thr residues are glycosylated have not been apparent from just the amino acid residues surrounding them, though the Hyl residues in collagen that are glycosylated occur in a characteristic sequence, -Gly-Xaa-Hyl-Xaa-Arg-, where Xaa is any residue.

O-glycosylation occurs in proteins that are already folded and where the three-dimensional structure is probably important.

In some proteins, the amino acid residues that are glycosylated are clustered in the primary structure. The carbohydrate content of these proteins can be as high as 65 - 85 by weight, so the carbohydrate dominates the structure.

These regions of the protein tend not to have fixed conformations but to be mostly extended.

O-glycosylation was thought to be confined to proteins that pass through the ER and Golgi apparatus. but recently it has been found to occur in a surprising number of cytoplasmic and nuclear proteins. In this case, N-acetyl glucosamine (GlcNAc) groups are attached to the side chains of Ser and Thr residues, but little is known about the process.
c. Proteoglycans

- Proteoglycans are composed of a variety of protein backbones, to which one or many glycosaminoglycan chains are covalently attached. The glycosaminoglycans are repeating disaccharide chains of three types: chondroitin sulfate/dermatan sulfate, heparan sulfate/heparin, or keratan sulfate.
- They are sulfated to various degrees and are usually attached to the protein backbone through a xylose moiety linked to a Ser residue.
- They can also be attached to Ser residues through N-acetyl glucosamine residues and through the complex type of carbohydrates N-linked to Asn residues.
- The signal for attachment of chondroitin sulfate chains appears to be the sequence -Ser-Gly-Xaa-Gly- preceded by two or three acidic residues.
- Other glycosaminoglycan chains are attached to Ser residues that are followed by Gly. No other signals have yet been detected.
- What chains are attached to each core protein depends on other unknown aspects of the protein structure and on the cell in which it is made.
Proteoglycans are extreme examples of glycosylated proteins. The bulk of their structure is usually the large amount of carbohydrate that is attached to the polypeptide chain at very many sites.

Proteoglycans are secreted and in some cases also have N-linked oligosaccharides attached to other side chains.

Their physical and biological properties consequently are dominated by the carbohydrates, and the protein components have been very difficult to characterize chemically.

Information about the protein parts is finally becoming available with the ability to determine their primary structures using recombinant DNA techniques.

The greatly varied core proteins in proteoglycans have complex primary structures and are not simply unfolded polypeptide backbones. The core proteins vary in molecular weight from 11,000 to 220,000, and the number of glycosaminoglycan chains can vary from 1 to 100.

Besides having segments of the polypeptide chain involved in attachment of glycosaminoglycan chains, the core proteins have other domains that are involved in interactions with membranes, extracellular matrix components, and other molecules.
Proteoglycans are important constituents of the extra-cellular matrix of multicellular organisms, and they are also associated with most cells, on their surface and inside intracellular storage granules.

Most extra-cellular matrix proteins and many growth factors have binding sites for glycosaminoglycans. The detailed biological and physical characterization of proteoglycans is only just beginning.
Lipids frequently tether intrinsically soluble proteins to membranes. The polar group of the lipid is attached covalently to the protein, while the hydrophobic portion of the lipid is embedded in the membrane. Examples are the myristoyl groups attached to the N-terminus and the glycosyl-phosphatidylinositol and farnesyl groups attached to the C-terminus that have been described. In addition, palmitoyl groups can be attached in thioester linkages to the side chains of Cys residues:
No similarities among the sequences of proteins modified in this way have yet been noted.

Proteins modified with palmitoyl groups are usually firmly anchored in a membrane by the palmitoyl group.

In most cases, however, the proteins are intrinsic membrane proteins without the palmitoyl group and are synthesized on the rough endoplasmic reticulum.

The palmitoyl groups are added to the completed chain either in the ER or in the cis or medial parts of the Golgi apparatus. The Cys residues modified are usually 3-6 residues from the start of transmembrane segments, on the cytoplasmic side. In these cases, the protein would be integrated into the membrane whether or not it were acylated, so the reaction for palmitoylation is not known. The palmitoyl group is often labile, being removed and replaced. In other instances, palmitoylation occurs only if the C-terminal Cys residue is farnesylated.

Other fatty acids can also be esterified to proteins, and other types of linkages are thought to occur.

Much remains to be learned about this covalent modification, especially because many proteins involved in cell regulation are modified in this way.
2.4.5 Sulfation

- Sulfation of Tyr residues is another posttranslational modification that is limited to proteins that pass through the Golgi apparatus.
- The enzyme responsible, tyrosyl protein sulfo-transferase, is an integral membrane protein, with its active site in the lumen of the trans Golgi network.
- The donor of the sulfate groups is 3'-phosphoadenosine-S'-phosphosulfate.
- Tyr residues that are sulfated do not occur in recognizably similar sequences, though they are usually surrounded by acidic residues, with a paucity of basic residues.
- Four acidic residues are generally within five residues on either side of the Tyr that is sulfated, and one is usually the preceding residue. It is also important that the Tyr residue be on the surface of the protein conformation and that it be accessible; nearby glycosylation blocks sulfation.
- The functional roles of the recently recognized phenomenon of sulfation are just being discovered.
- There are indications that sulfation affects the biological activities of some neuropeptides, the proteolytic processing of some protein precursors, and intracellular transport of some secretory proteins.
2.4.6 $\gamma$-Carboxy-Glu Residues

- Certain Glu residues, particularly in proteins involved in blood clotting and bone structure, are carboxylated to yield the unusual residue $\gamma$-carboxyglutamic acid, generally abbreviated as Gla:
The enzyme responsible, vitamin K-dependent carboxylase, is an integral membrane protein with its active site in the lumen of the endoplasmic reticulum.

The carboxyl donor is \( \text{HCO}_3^- \), and the enzyme also requires the reduced form of vitamin K and Oz.

This is the first biochemical function found for vitamin K; its role appears to be to labilize or remove the \( \gamma \)-hydrogen atom of the Glu side chain that is to be replaced by the carboxyl group.
The Glu residues that are converted to Gla do not occur in any particular amino acid sequence, but they are generally in the first 40 residues of the mature protein.

In one case, Factor X, the propeptide has been shown to direct the carboxylation of the 12 Glu residues closest to the amino terminus.

There are also some sequence similarities among segments that are modified in various proteins.

The function of Gla residues is almost invariably linked to binding of Ca$^{2+}$ ions.

The second adjacent carboxyl group considerably increases the intrinsic ability of these residues to bind Ca$^{2+}$ ions, and Ca$^{2+}$ binding is invariably involved in the functions of the proteins modified.

The modification is essential for the functional properties of the proteins, as can be shown by synthesizing the proteins in the presence of vitamin K antagonists such as warfarin and dicumarol, which inhibit carboxylation and cause the proteins to be biologically inactive.
2.4.7 Hydroxylation

- Hydroxylation of certain Pro and Lys residues is an important step in the maturation and secretion of collagen.

- These modifications occur in the endoplasmic reticulum but only to procollagen. Pro residues that occur in the sequence -Xaa-Pro-Gly- are hydroxylated on the γ-carbon, and Lys residues in the sequence -Xaa-Lys-Gly- are hydroxylated on the δ-carbon:
Other Pro residues of certain types of collagen, occurring in the sequence -Gly-Pro-, are hydroxylated at the $\beta$-carbon:

Each of these modifications requires $O_2$, $\alpha$-ketoglutarate, and ascorbate (vitamin C) and is catalyzed by an enzyme: prolyl 4-hydroxylase, lysyl 5-hydroxylase, and prolyl 3-hydroxylase, respectively. Each of these enzymes contains ferrous iron, and ascorbate is needed to keep the iron atom reduced. One of the oxygen atoms of $O_2$ hydroxylates the side chain, and the other oxidizes the $\alpha$-ketoglutarate to succinate and $CO_2$.

These modifications are vital for the folding and assembly of mature collagen. $\gamma$-OH-Pro residues stabilize the collagen triple helix by introducing additional hydrogen bonding, and $\delta$-OH-Lys residues are necessary for the formation of certain cross-links and for the attachment of glycosyl groups.
A similar hydroxylation has recently been found to occur on certain Asn and Asp residues in a few proteins. In each case, the hydroxyl groups are added posttranslationally in a specific orientation to the $\beta$-carbon of the side chains; the hydroxyl group introduces a new center of chirality and is always the erythro isomer:

Which Asp and Asn residues are modified in this way seems to depend primarily on the three-dimensional conformation of the protein.

These modifications are thought to participate in $\text{Ca}^{2+}$ binding, although little is known.
2.4.8 Phosphorylation

- An increasing number of proteins are known to be phosphorylated at specific sites, usually reversibly and with important functional consequences.
- The phosphoryl groups are added by specific protein kinases, using ATP as the phosphoryl donor:
  \[
  \text{protein} + \text{ATP} \rightarrow \text{protein} - \text{P032-} + \text{ADP}
  \]
  The phosphoryl groups are removed by specific phosphatases:
  \[
  \text{protein} - \text{PO}_3^{2-} + \text{H}_2\text{O} \rightarrow \text{protein} + \text{HPO}_3^{2-}
  \]

The sum of these two reactions is simply hydrolysis of ATP, so the two reactions are catalyzed by different enzymes, and their activities are strictly controlled. In fact, it is through control of the kinases and phosphatases that the activities of the phosphorylated proteins are regulated.

Many hormones act by increasing the Intracellular concentration of second messengers—cyclic AMP, diacyl glycerol, or \( \text{Ca}^{2+} \) which in turn activate protein kinases that phosphorylate Ser and Thr residues of various proteins. The protein products of oncogenes and many growth-factor receptors have protein kinase activities that phosphorylate Tyr residues.
The sites of phosphorylation are usually the hydroxyl groups of specific Ser, Thr, or Tyr residues. But Asp; His, and Lys residues may also be phosphorylated.

From the point of view of protein structure and function, the most important aspect of the phosphoryl group appears to be its negative charge.
The various kinases have different specificities for different proteins. In any one protein, which residues are phosphorylated depends on the primary structure around them, on their accessibility to the kinase and on the specificity of the kinase enzyme.

The important cyclic AMP-dependent kinase has a strong preference for Ser residues that occur in the sequence Arg-Arg-(Xaa)$_n$-Ser-, where $n$ is usually 1 but can be 0 or 2.

Other Ser/Thr kinases similarly recognize Ser residues following one or two basic residues. In contrast, Tyr phosphorylation usually involves Tyr residues that occur in the sequence -Lys/Arg-(Xaa)$_3$-Asp/Glu-(Xaa)$_3$-Tyr-.

The folded conformation of the protein is also important for determining which residues are phosphorylated by any kinase, because short peptides with these sequences are usually relatively poor substrates of the kinases.

Phosphorylation almost invariably occurs to folded proteins well after their synthesis has been completed, and in contrast to many other posttranslational modifications, it occurs primarily in the cytoplasm.

Not all phosphorylation is functionally important—that of the milk protein casein is probably primarily of nutritional importance. In this case, the Ser residues phosphorylated are usually to the amino side of a number of acidic residues.
Another common modification, ADP-ribosylation, is similar to phosphorylation in that it acts reversibly in the cytoplasm and nuclei of cells to regulate various proteins. All eukaryotic cells seem to have enzymes called ADP-ribosyl transferases, which cleave the cofactor NAD$^+$ and transfer the ADP-ribosyl moiety to various side chains in a number of proteins:
The modification can occur at the nitrogen atoms of His, Arg, Asn, and Lys residues, at the carboxyl group of Glu, and at the α-carboxyl group of terminal Lys residues. In modifications to carboxyl groups, addition of a first ADP-ribosyl group is followed by addition of others to the 2' hydroxyls of either of the ribose groups. In this way, linear and branched poly(ADP-ribose) structures containing up to 65 ADP-ribose groups can be generated. This modification occurs primarily in the nucleus.

Multiple ADP-ribosyl groups can be removed by the enzyme poly(ADP-ribose) glycohydrolase, and the group attached directly to the protein can be removed by another enzyme, ADP-ribosyl protein lyase. A multitude of proteins are modified in this way, with a variety of effects on their activities, and no simple/coherent description of the normal physiological effects of ADP-ribosylation can be given. The importance of this modification is illustrated, however, by the toxic effects caused by the ADP-ribosylation of certain proteins by diphtheria, cholera, and pertussis toxins. These toxic modifications mimic and subvert the regulated physiological modifications that normally occur.
2.4.10 Bisulfide Bond Formation

- Disulfide bond formation between Cys residues is a common occurrence in proteins synthesized in the ER.
- The Cys residues linked by disulfide bonds are usually far apart in the primary structure, so disulfide formation between them is intimately linked with three-dimensional folding of the polypeptide chain.
- The mechanism of disulfide bond formation in vivo is uncertain, but it probably involves thiol-disulfide exchange between the protein synthesized with free SH groups on its Cys residues and small-molecule disulfide compounds.
- The predominant thiol compound in most cells is glutathione, which exists in both the thiol (GSH) and disulfide (GSSG) forms.
- Formation of one disulfide bond in a protein requires two sequential thiol-disulfide exchanges involving the mixed-disulfide intermediate:
The protein becomes oxidized and the glutathione is reduced, so it is convenient to define a disulfide oxidation-reduction potential, which in this case would be given by the ratio of the concentrations of GSSG and GSH: \([\text{GSSG}/\text{GSH}]^2\).

The chemical reaction can occur rapidly under physiological conditions and is reversible. In a protein with more than two Cys residues, formation of disulfide bonds is often followed by intra-molecular rearrangements ("shuffling") of the disulfides among the various Cys residues in the protein.

This is frequently the rate-limiting step in the entire process of forming multiple disulfides in a protein. Not surprisingly, an enzyme, protein-disulfide isomerase, is present in the ER to catalyze disulfide rearrangements in proteins and consequently to assist in their folding.
Which disulfides, if any, are formed in a protein depends on both the conformational properties of the protein, which determines whether and which Cys residue come into appropriate oxidation-reduction potential, which determines the intrinsic stability of protein disulfide bonds. The observed stability of individual protein disulfides are given by the equilibrium constant.

\[ K_{ss} = \frac{[\text{protein}_S^S] [GSH]^2}{[\text{protein}_S^H][GSSG]} \]  \hspace{1cm} (2.19)

The stabilities of protein disulfide bonds vary enormously. For unfolded proteins in which Cys residues are not separated by more than 50 residues the values of \( K_{ss} \) for different pairs of Cys residues are in the region of only 10\(^{-2}\) M. For folded proteins in which the folded conformation keeps the Cys residues in proximity for forming disulfides, the values of \( K_{ss} \) may be as high as 10\(^{5}\) M whereas Cys residues kept apart by the conformation have values of zero.

In the cytoplasm of most cells, glutathione is present at a total concentration of 1-10 mM, and only about 1% of it is present as GSSG; consequently the ratio \( [\text{GSSG}]/[\text{GSH}]^2 \) has a value between 1 and 10 M\(^{-1}\).
Protein disulfides are present under these conditions only if their value of $K_{ss}$ is greater than 1 M. Such large values of $K_{ss}$ occur only if the protein conformation brings pairs of Cys residues into suitable proximity. In this case, disulfides can be stable in folded proteins under such disulfide oxidation-reduction potential conditions, even though the cytoplasm is frequently said to be too reducing because the majority of the glutathione is in the thiol form. Intramolecular protein disulfides can be considerably more stable than the intermolecular disulfide of GSSG. In contrast Cys residues kept apart by the protein conformation largely remain in the thiol form under intracellular redox conditions.

The value of the disulfide oxidation-reduction potential of the lumen of the ER is not known, but it is unlikely to be much more oxidizing than that of the cytoplasm. Otherwise, disulfide bonds would be intrinsically too stable, and polypeptide chains synthesized in the ER would tend to form disulfide bonds between most of their Cys residues, not just those favored by the protein conformation.
2.4.11 Common Nonenzymatic, Chemical Modifications

- The covalent modifications described in the preceding sections occur to specific residues in certain proteins a specificity that is possible only because the modifications are catalyzed by enzymes and depend on the detailed structure of the protein modified. A great many chemical modifications occur chemically and spontaneously in the absence of enzymes. These modifications tend to occur to all appropriate residues in all proteins, depending only on the immediate chemical environment of the residues. Although many such modifications inactivate the modified proteins, most modifications occur at insignificant rates in vivo. Others occur more frequently; but at least in some cases, repair systems are available to minimize their consequences.

- One inevitable modification is oxidation by the $\text{O}_2$ that is necessary for most life, and by other oxidants such as peroxides, superoxide and hydroxyl radicals, and hypochlorite ions that are generated during metabolism. Enzymes such as superoxide dismutase, peroxidase, and catalase scavenge many such oxidants, but oxidation of proteins still occurs. Most susceptible are the sulfur atoms of Cys and Met residues.
The Cys thiol groups of intracellular proteins are generally protected by the high concentrations of glutathione (or other similar thiol compounds in some organisms) that are present in cells and maintained in the reduced thiol form by specific enzyme systems.

Met residues are readily oxidized chemically to the sulfoxide, which can have drastic functional consequences in some proteins. For example, $\alpha_1$-proteinase inhibitor ($\alpha_1$-antitrypsin) has a crucial Met residue in its active site; oxidation of this residue inactivates the inhibitor. The inhibitor normally inhibits serum elastase, a protease that when not controlled destroys the cell walls of the lung by digesting connective tissue proteins. Absence of $\alpha_1$-proteinase inhibitor from the serum causes pulmonary emphysema. Oxidation of the inhibitor and the severity of emphysema are increased by smoking. Oxidation of Met residues in cells, however, is reversed by the enzyme Met sulfoxide peptide reductase.

Another common chemical modification of proteins is deamidation of Asn and Gln residues. Deamidation of Asn residues usually converts a fraction of them to iso-Asp residues, in which the peptide bond occurs through the side chain. This can have severe effects on protein structure. Many chemical modifications of proteins lead to the degradation of the protein.
The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators
Rong-Gui Hu1*, Jun Sheng1*, Xin Qi2, Zhenming Xu1†, Terry T. Takahashi2 & Alexander Varshavsky1

The conjugation of arginine to proteins is a part of the N-end rule pathway of protein degradation. Three N-terminal residues—aspartate, glutamate and cysteine—are arginylated by ATE1-encoded arginyl-transferases. The oxidation of N-terminal cysteine is essential for its arginylation. The in vivo oxidation of N-terminal cysteine, before its arginylation, is shown to require nitric oxide. We reconstituted this process in vitro as well. The levels of regulatory proteins bearing N-terminal cysteine, such as RGS4, RGS5 and RGS16, are greatly increased in mouse ATE1 2/2 embryos, which lack arginylation. Stabilization of these proteins, the first physiological substrates of mammalian N-end rule pathway, may underlie cardiovascular defects in ATE1 2/2 embryos. This findings identify the N-end rule pathway as a new nitric oxide sensor that functions through its ability to destroy specific regulatory proteins bearing N-terminal cysteine, at rates controlled by nitric oxide and apparently by oxygen as well.
The N-end rule: Introduction
The N-end rule relates the in vivo half-life of a protein to the identity of its N-terminal residue. A ubiquitin-dependent pathway, called the N-end rule pathway, recognizes degradation signals (degrons) that include the signals called N-degrons (Fig. 1a). An N-degron consists of a protein’s destabilizing N-terminal residue and an internal Lys residue. The latter is the site of formation of a polyubiquitin chain. The N-end rule has a hierarchic structure. N-terminal Asn and Gln are tertiary destabilizing residues in that they function through their deamidation, by N-terminal amidohydrolases, to yield the secondary destabilizing residues Asp and Glu. The activity of N-terminal Asp and Glu requires their conjugation, by ATE1-encoded isoforms of Arg-tRNA-protein transferase (R-transferase), to Arg, one of the primary destabilizing residues. The latter are recognized by E3 ubiquitin ligases of the N-end rule pathway. In mammals, destabilizing N-terminal residues that function through their arginylation are not only Asp and Glu but also Cys (Fig. 1a), which is a stabilizing (unarginylated) residue in the yeast Saccharomyces cerevisiae. Known functions of the N-end rule pathway include the control of peptide import (through conditional degradation of the import’s repressor), the regulation of apoptosis (through degradation of a caspase-processed inhibitor of apoptosis) and the fidelity of chromosome segregation (through degradation of a conditionally produced cohesin’s fragment), as well as the regulation of meiosis, cardiovascular development in animals and leaf senescence in plants.
The N-terminal cysteine must be oxidized before its arginylation.

a, The mammalian N-end rule pathway. N-terminal residues are indicated by single-letter abbreviations for amino acids. Yellow ovals denote the rest of a protein substrate. The ‘cysteine’ (Cys) sector, in the upper left corner, describes the main discovery of this work: the NO-mediated oxidation of N-terminal Cys, with subsequent arginylation of oxidized Cys by ATE1-encoded isoforms of R-transferase. C* denotes oxidized Cys, either Cys-sulphinic acid (CysO2H) or Cys-sulphonic acid (CysO3H). Type 1 and type 2 primary destabilizing N-terminal residues are recognized by E3 ubiquitin (Ub) ligases of the N-end rule pathway, including UBR1 and UBR2. Through their other substrate-binding sites these E3 ubiquitin ligases also recognize internal (non-N-terminal) degrons in other substrates of the N-end rule pathway, denoted by a larger yellow oval.

b, MetAPs remove Met from the N terminus of a polypeptide if the residue at position 2 belongs to the set of residues shown.

c–j, N-terminal Cys must be oxidized before its arginylation. Three eight-residue peptides are denoted as X-p; their N-terminal residues (X) were either Asp, Cys or CysO3H. X-p was incubated with mouse ATE1-1 R-transferase at pH 7.5 in the presence of ATP, S. cerevisiae Arg-tRNA synthetase and tRNAs, followed by analyses of peptide products, either by capillary electrophoresis (CE) (c–h) or by MALDI–TOF MS (i, j).
Inhibiting farnesylation of progerin prevents the characteristic nuclear blebbing of Hutchinson–Gilford progeria syndrome

Introduction Hutchinson–Gilford progeria syndrome (HGPS) is an extremely rare and uniformly fatal “premature aging” disease in which all children die as a consequence of myocardial infarction or cerebrovascular accident at an average age of 12 years (8–21 years). HGPS is a sporadic autosomal dominant disease caused in nearly all cases by a de novo single-base substitution in codon 608 of exon 11 of the LMNA gene on chromosome 1. The LMNA gene encodes three proteins, lamin A (LA), LC, and LA10, all of which are components of the nuclear lamina, a dynamic molecular interface located inside the inner nuclear membrane. The lamina has now been shown to have significant roles in DNA replication, transcription, chromatin organization, nuclear shape, and cell division.

In the LMNA gene, 180 mutations have been reported, and currently, there are eight diseases in addition to HGPS (referred to as the “laminopathies”) that are associated with various mutations in this gene.
HGPS is almost always caused by a *de novo* point mutation in the lamin A gene (*LMNA*) that activates a cryptic splice donor site, producing a truncated mutant protein termed “progerin.”

- WT prelamin A is anchored to the nuclear envelope by a farnesyl isoprenoid lipid. Cleavage of the terminal 15 aa and the farnesyl group releases mature lamin A from this tether.

- In contrast, this cleavage site is deleted in progerin. The retention of the farnesyl group causes progerin to become permanently anchored in the nuclear membrane, disrupting proper nuclear scaffolding and causing the characteristic nuclear blebbing in HGPS cells.

- When the terminal CSIM sequence in progerin was mutated to SSIM, a sequence that cannot be farnesylated. SSIM progerin relocalized from the nuclear periphery into nucleoplasmic aggregates and produced no nuclear blebbing.

- Also, blocking farnesylation of authentic progerin in transiently transfected HeLa, HEK 293, and NIH 3T3 cells with farnesyltransferase inhibitors (FTIs) restored normal nuclear architecture. Last, treatment of both early and late-passage human HGPS fibroblasts with FTIs resulted in significant reductions in nuclear blebbing.

- Treatment with FTIs represents a potential therapy for patients with HGPS.
Introduction: Peptide toxins ("conopeptides") present in the venoms of marine cone snails (family Conidae, genus *Conus*) may be divided into two main structural groups: those with zero or one disulfide bridge and highly disulfide cross-linked peptides with two to five disulfide bridges, conventionally called "conotoxins". The mature toxins found in the venoms are processed from prepropeptide precursors produced by normal ribosomal translation by proteases and other posttranslational modification enzymes. Most of the >50000 different conotoxins belong to a relatively few major gene superfamilies (A, T, O, M, P, I, and S); the peptides that belong to a particular superfamily share considerable sequence identity in their signal peptides, and have a specific arrangement of Cys residues within the mature toxin (the "Cys" pattern).
FIGURE 1: Purification of de13a by HPLC. (A) Fractionation of the extract from four venom ducts by an analytical Vydas C\textsubscript{18} column eluted at room temperature with a linear gradient from 5 to 95\% solution B over the course of 90 min at 1 mL/min. HPLC solutions were 0.1\% (v/v) TFA in water (solution A) and 0.085\% (v/v) TFA in 90\% (v/v) aqueous ACN (solution B). The absorbance was monitored at 220 nm. (B) The peak highlighted by the arrow in panel A was further purified on the same column using a linear gradient starting with 10\% solution B and increasing solution B by 0.2\%/min at a rate of 1 mL/min. The peak highlighted by the arrow in panel B was further characterized.
FIGURE 4: Identification of 6-bromo-L-tryptophan and 5-hydroxy-L-lysine in peptide de13a. (A) Chromatographic profile of a standard mixture of PTH derivatives (20 pmol each) of the 20 genetically encoded amino acids; the one-letter code is employed. (B) Standards of PTH-6-bromo-L-tryptophan (W<sup>6</sup>), PTH-3-bromo-DL-tryptophan (W<sup>3</sup>), and PTH-(5R)-5-hydroxy-L-lysine (K) generated and chromatographed simultaneously under the conditions of automatic sequencing (Procise 491 protein sequencing system). (C) PTH derivative (arrow) obtained at position 14 of native peptide de13a. (D) PTH derivative (arrow) obtained at position 18. (E) PTH derivative (arrow) obtained at position 25.
A major peptide, de13a from the crude venom of *Conus delessertii* collected in the Yucatan Channel, Mexico, was purified. The peptide had a high content of posttranslationally modified amino acids, including 6-bromotryptophan and a nonstandard amino acid that proved to be 5-hydroxylysine.

The mature toxin has the sequence:

```
DCOTSCOTTCANGWECCKGYOCVNAKGCH*,
```

where O is 4-hydroxyproline, W 6-bromotryptophan, and K 5-hydroxylysine, the asterisk represents the amidated C-terminus.

The eight Cys residues are arranged in a pattern (C-C-C-CC-C-C) not described previously in conotoxins. This arrangement, for which we propose the designation of framework #13 or XIII, differs from the ones (C-C-CC-CC-C-C and C-C-C-C-CC-C-C) present in other conotoxins which also contain eight Cys residues. This peptide thus defines a novel class of conotoxins, with a new posttranslational modification not previously found in other *Conus* peptide families.
Introduction: Protein prenylation is a critical posttranslational modification that is found in 1% of mammalian proteins.

Prenylation, either farnesylation or geranylgeranylation, is required for the proper membrane association and activity of many signal transduction proteins.

Ras proteins are essential signaling proteins that are farnesylated on the cysteine sulfur of their C-terminal CaaX box, where a is often an aliphatic residue and X is typically either serine or methionine (alanine, glutamine, threonine, and, in certain cases, leucine can also serve as the X residue).

Oncogenic Ras has been implicated in up to 30% of all human cancers.

The prenylation of Ras is catalyzed by protein farnesyltransferase (FTase).

This enzyme transfers a 15-carbon isoprenoid from FPP, farnesyl diphosphate, to Ras thus allowing for proper membrane association.
General kinetic mechanism for protein farnesyltransferase.
Protein farnesyltransferase (FTase) catalyzes the post translational modification of many important cellular proteins, and is a potential anticancer drug target.

Crystal structures of the FTase ternary complex illustrate an unusual feature of this enzyme, the fact that the isoprenoid substrate farnesyl diphtosphate (FPP) forms part of the binding site for the peptide substrate.

This implies that changing the structure of FPP could alter the specificity of the FPP-FTase complex for peptide substrates.

A newly synthesized FPP analogue, 3-MeBFPP, is a substrate with three peptide cosubstrates, but is not an effective substrate with a fourth peptide (dansyl-GCKVL).

Addition of this analogue also inhibits farnesylation of dansyl-GCKVL by FPP. The differential substrate abilities of these four peptides with FPP-FTase and 3-MeBFPP-FTase complexes do not correlate with their binding affinities for these isoprenoid-enzyme complexes.
The possible mechanistic rationales for this observation, along with its potential utility for the study of protein prenylation.

- isoprenoid moiety of FPP forms part of the binding site for the peptide substrate. The modification of the farnesyl moiety could thus lead to looser or tighter binding of the peptide or protein substrates to the enzyme. Modifying the isoprenoid ligand can alter the specificity of the enzyme for its peptide substrate.

- Certain isoprenoid analogues may be able to alter FTase selectivity for its protein substrates in cells.

- The anticancer effects of FTIs are due to the inhibition of the farnesylation of several other proteins, including RhoB (8) and CENP-E (I). The determination of the identity of these “protein X” targets (3) is crucial for the rational use of FTIs in clinical settings.