# Protein engineering

Protein engineering describes the intentional alteration of a protein's amino acid sequence, usually with the aim of achieving either:

- a better understanding of the relationship between a protein's primary and higher-level structure, or its structure and function; or
- the development of a protein variant which, relative to the wild-type protein, displays some enhanced property in the context of its commercial use.

***Site-directed mutagenesis*** ***Directed evolution***

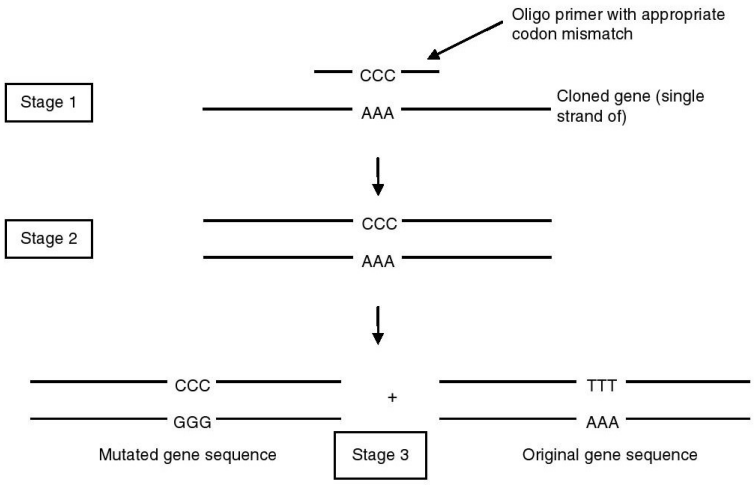
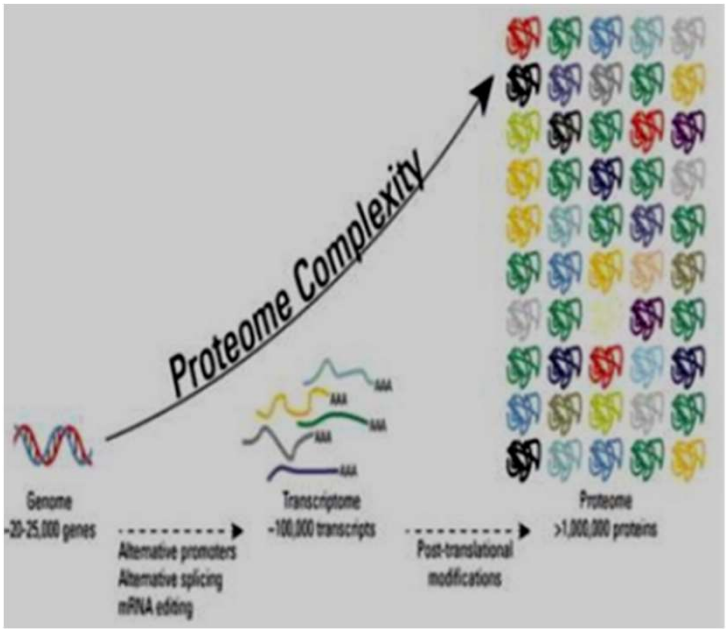
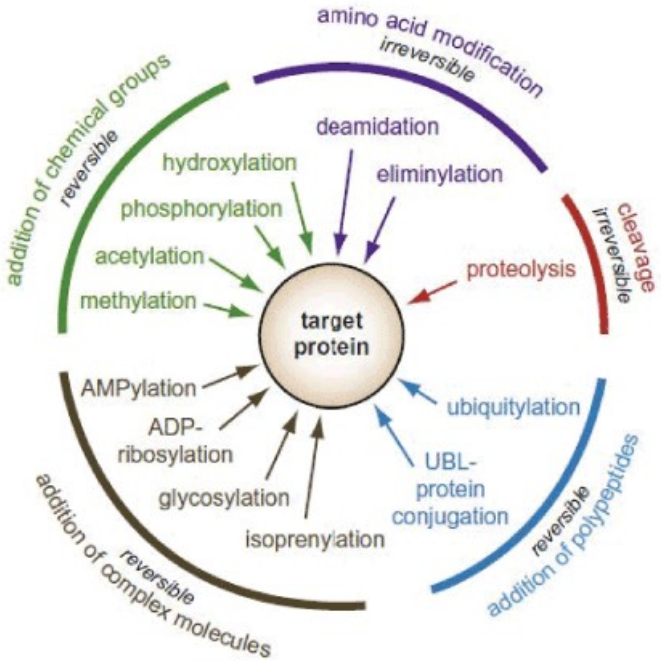


**Table 2.6** Representative engineered proteins which are now used commercially. These and other examples are considered in later chapters of this book.

| Protein | Use | Engineering detail | Chapter |
|---|---|---|---|
| Tissue plasminogen activator (tPA) | Thrombolytic agent | Various engineered products developed with altered amino acid sequences or with whole domains deleted in order to make clot degradation more efficient or lengthen serum half-life | 6 |
| Antibodies | Various, including cancer treatment | Various engineered products developed, including mouse-derived antibodies in which large segments have been replaced by human antibody domains (in order to reduce immunogenicity in humans), or the development of antigen-binding antibody fragments (which could for example penetrate tumours more effectively) | 7 |
| Fusion proteins | Various, including treating rheumatoid arthritis and cancer | Generation of novel hybrid proteins by combining one or more domains from two different proteins together. The fusion product 'Enbrel' for example consists of the extracellular domain of the tumour necrosis factor (TNF) receptor (allowing it bind TNF), fused to antibody constant (Fc) domains (which increases its serum half-life) | 7 |
| Engineered insulins | Diabetes | The replacement/alteration of amino acids in the insulin backbone in order to make the engineered product either faster-acting or slower-acting than un-engineered insulin | 8 |
| Various detergent proteases and amylases | Added to detergents to enhance cleaning | Removal/replacement of oxidation-sensitive amino acid residues, allowing the enzymes to retain activity in the presence of oxidants usually also present in detergents | 12 |
| DNA polymerase | PCR reactions | Enhances enzyme's affinity for DNA | 13 |



# Post-translational modifications (PTMs)



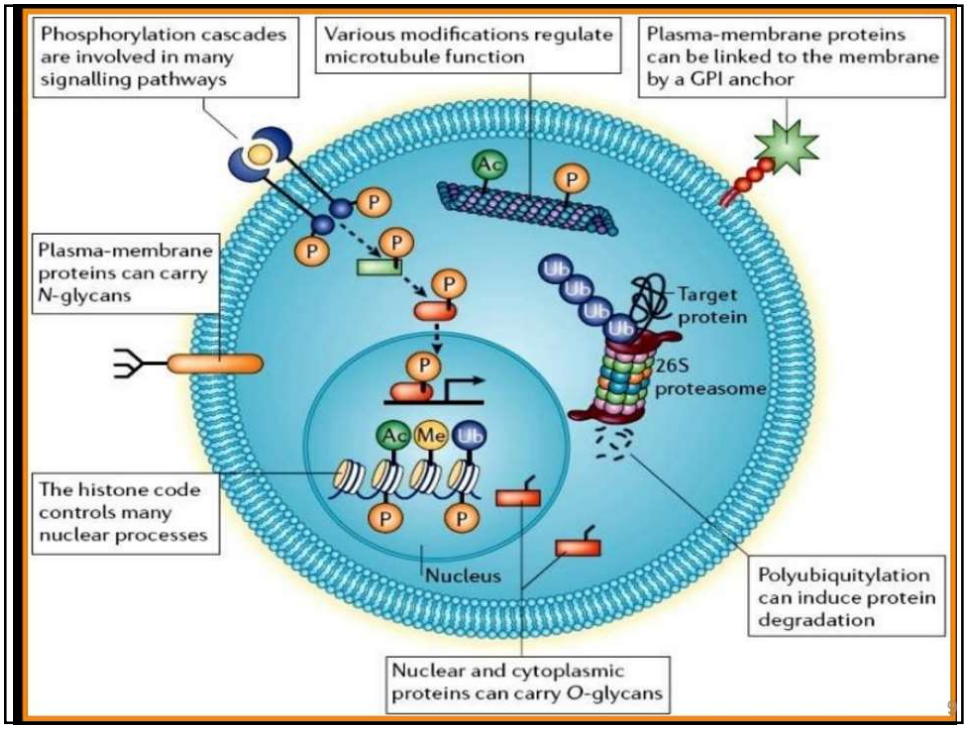
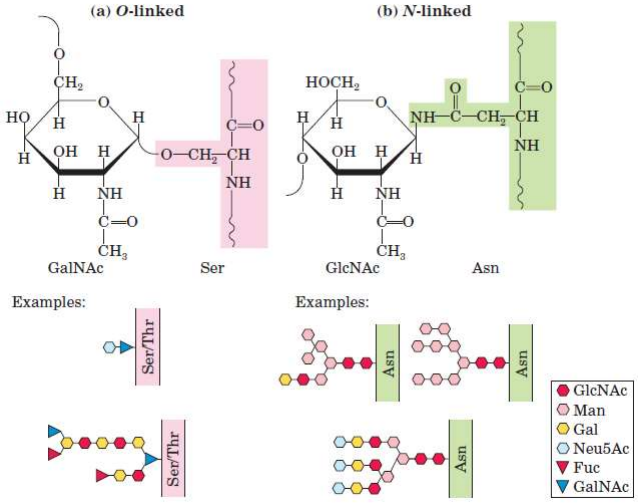
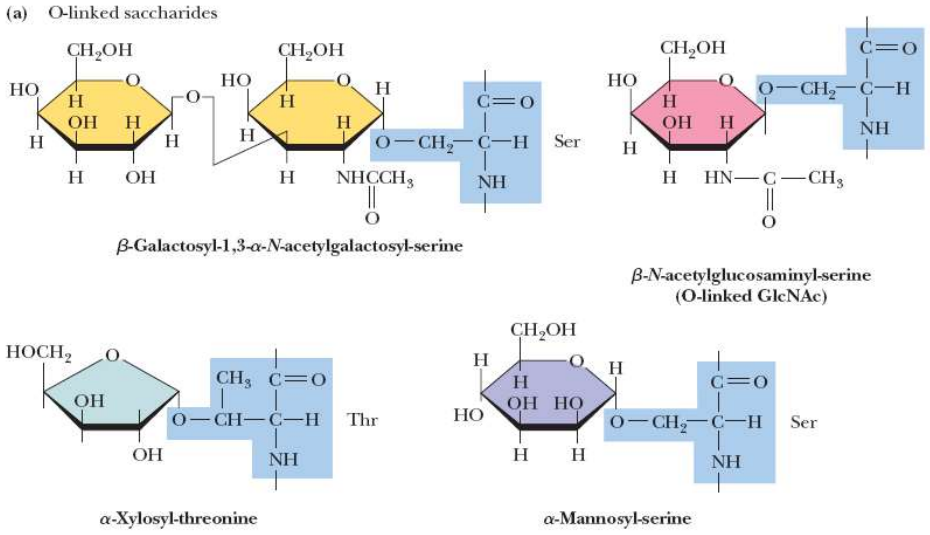



**Table 2.7** The more common forms of post-translational modifications that polypeptides may undergo. Refer to text for additional details.

| Modification | Comment |
|---|---|
| Glycosylation | For some proteins glycosylation can increase solubility, influence biological half-life and/or biological activity |
| Proteolytic processing | Various proteins become biologically active only on their proteolytic cleavage (e.g. some blood factors) |
| Phosphorylation | Influences/regulates biological activity of various regulatory proteins including polypeptide hormones |
| Acetylation | Modulation of target protein activity |
| Acylation | May help some polypeptides interact with/anchor in biological membranes |
| Amidation | Influences biological activity/stability of some polypeptides |
| Sulfation | Influences biological activity of some neuropeptides and the proteolytic processing of some polypeptides |
| Hydroxylation | Important to the structural assembly of certain proteins |
| γ-Carboxyglutamate formation | Important in allowing some blood proteins to bind calcium |
| ADP-ribosylation | Regulates biological activity of various proteins |
| Disulfide bond formation | Helps stabilize conformation of some proteins |

# Glycosylation




The carbohydrate residue linked to the protein in O-linked saccharides is usually an *N-acetylgalactosamine, but mannose, galactose, and xylose residues linked to protein* hydroxyls are also found.

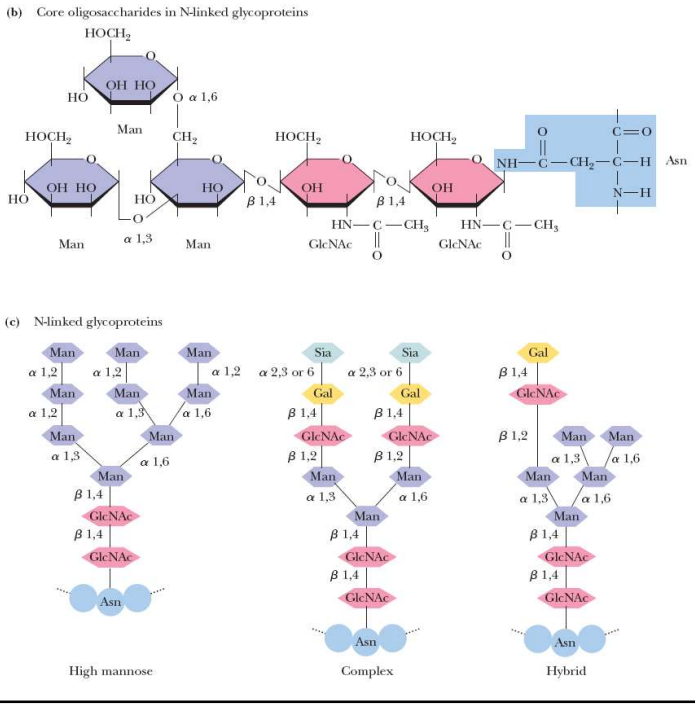
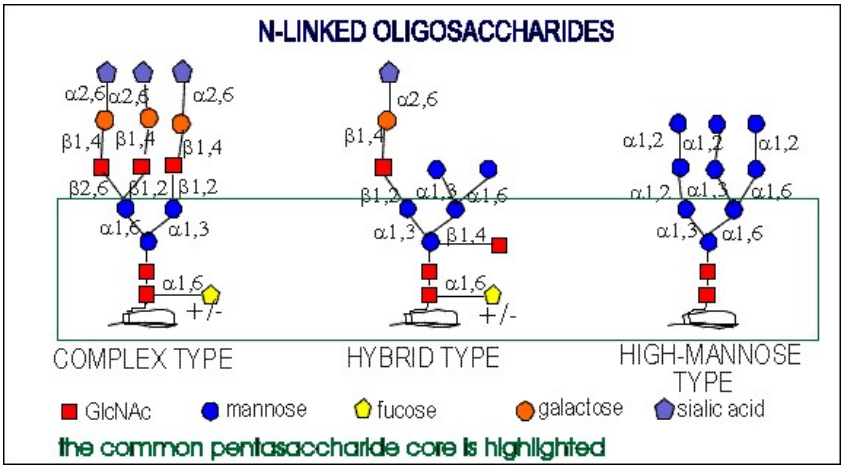
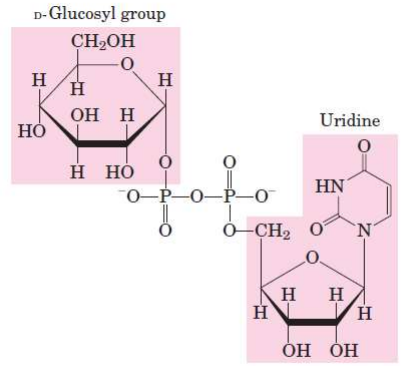
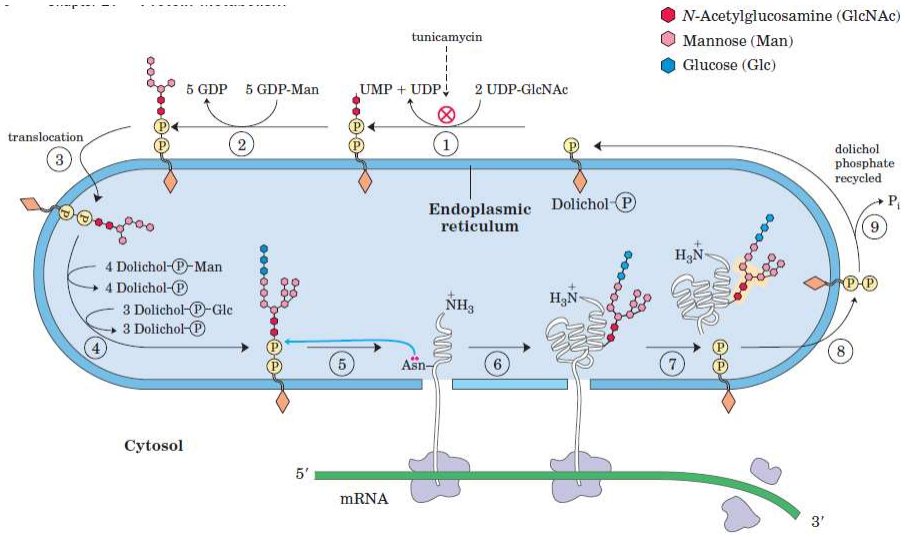



**FIGURE 7.32** The carbohydrate moieties of glycoproteins may be linked to the protein via (a) serine or threonine residues (in the O-linked saccharides) or (b) asparagine residues (in the N-linked saccharides). (c) N-linked glycoproteins are of three types: high mannose, complex, and hybrid, the latter of which combines structures found in the high mannose and complex saccharides.

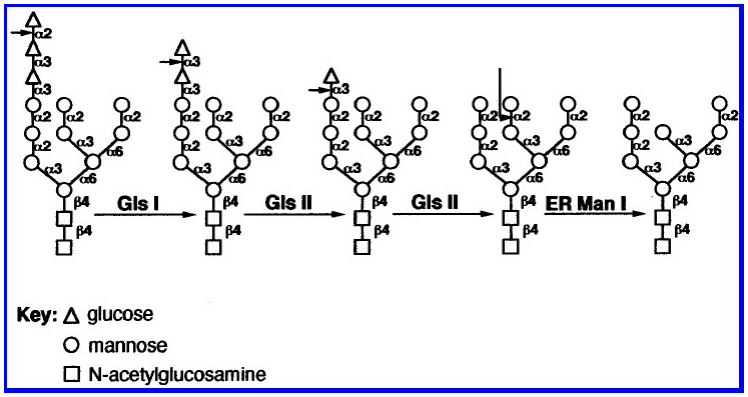




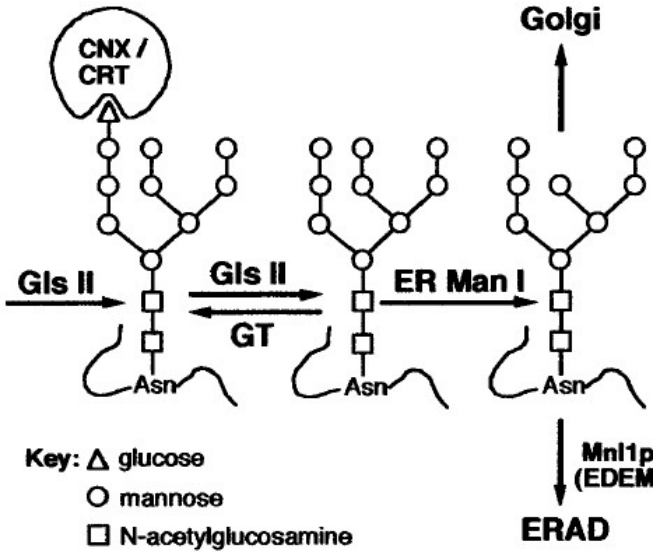

**UDP-glucose**
(a sugar nucleotide)



## Protein N-glycosylation in ER




## Protein quality control



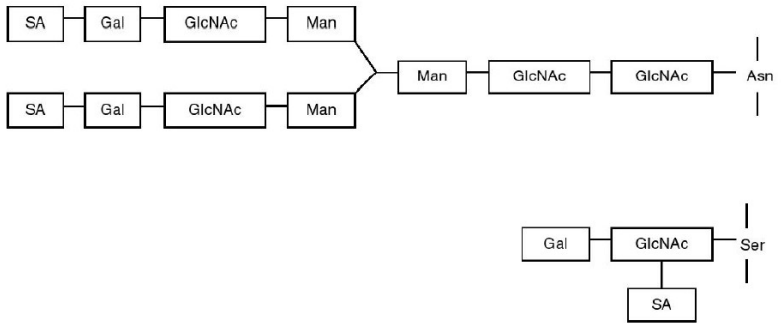



**Figure 2.18** Structure of two sample oligosaccharide side chains (one N-linked the other O-linked) found in glycoproteins. Man, mannose; Gal, galactose; SA, sialic acid; GlcNAc, *N*-acetylglucosamine.

## Sialic Acid Cleavage Can Serve as a Timing Device for Protein Degradation

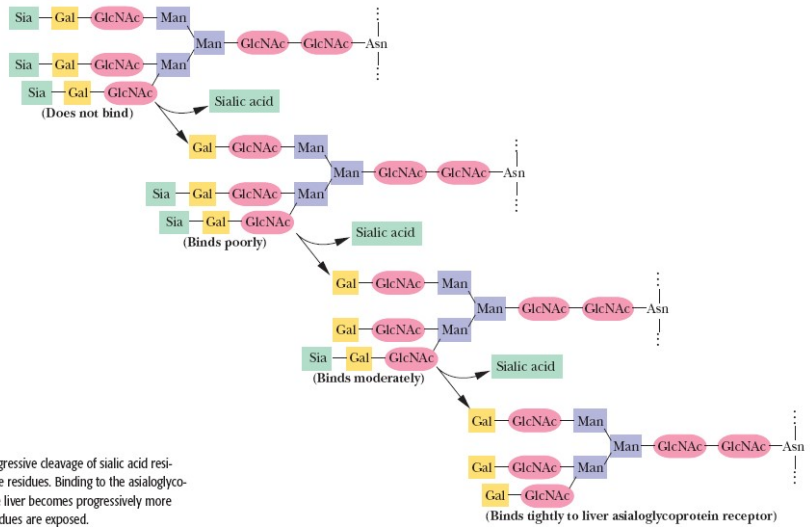
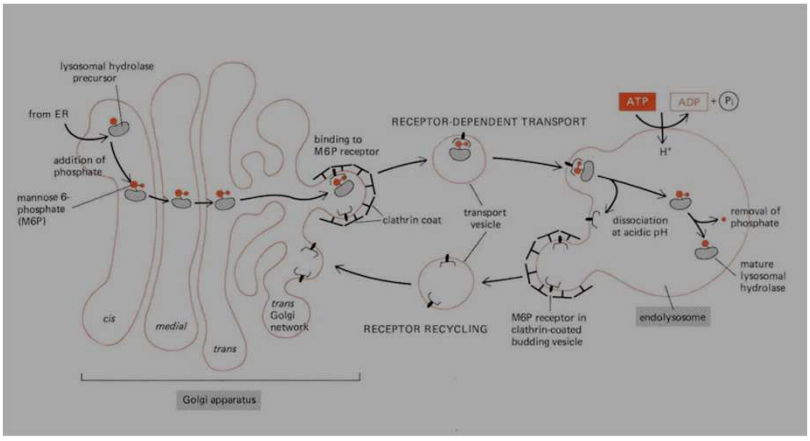
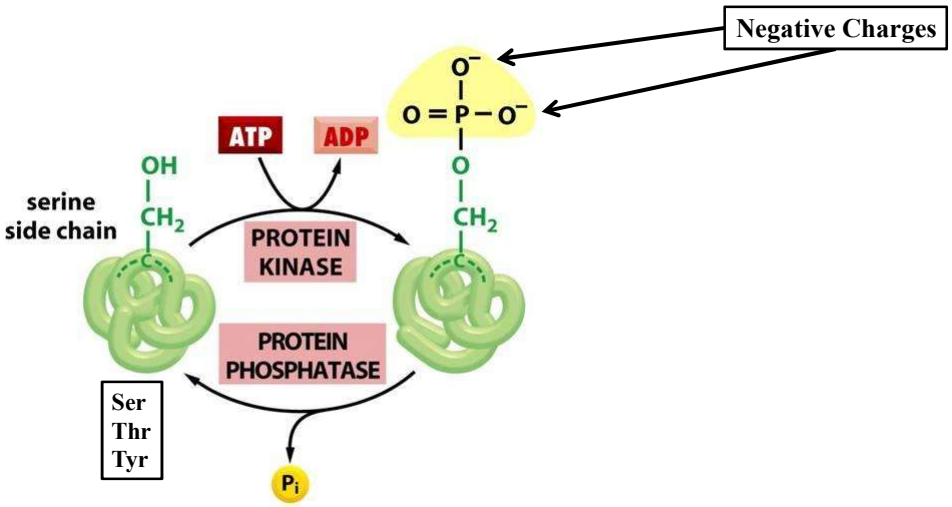


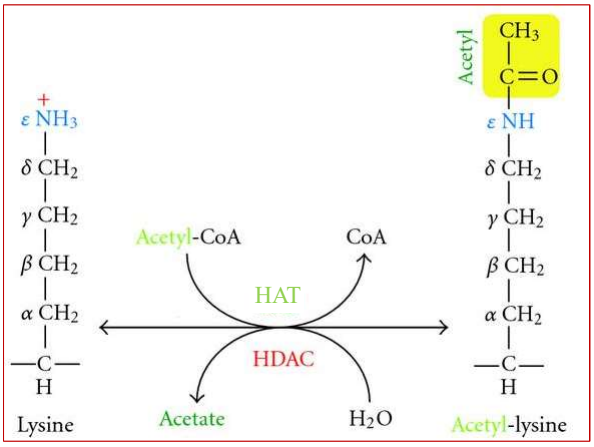

**FIGURE 7.35** Progressive cleavage of sialic acid residues exposes galactose residues. Binding to the asialoglycoprotein receptor in the liver becomes progressively more likely as more Gal residues are exposed.




# Phosphorylation

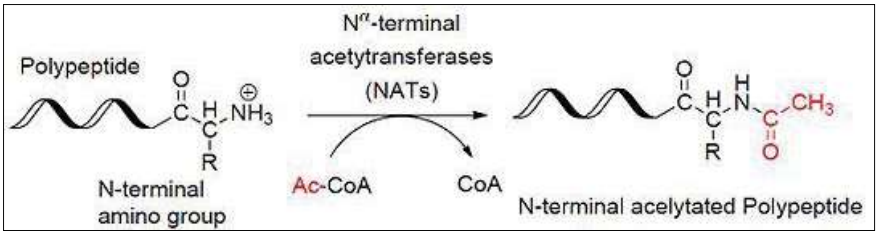


# Acetylation

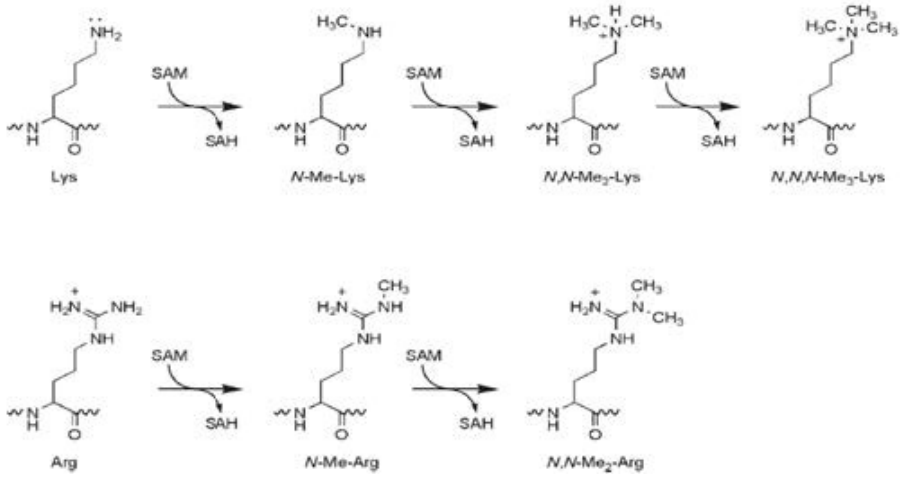



**1.Protein N-terminal acetylation**

**2.Protein lysine acetylation**

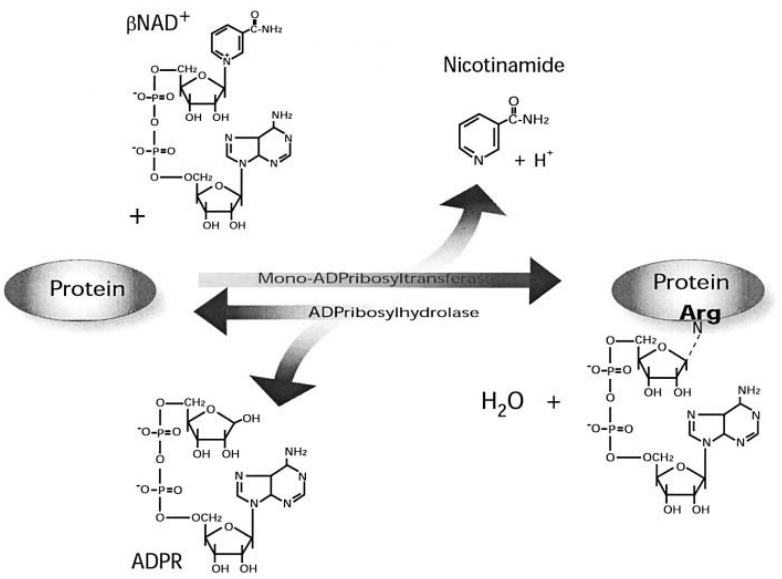


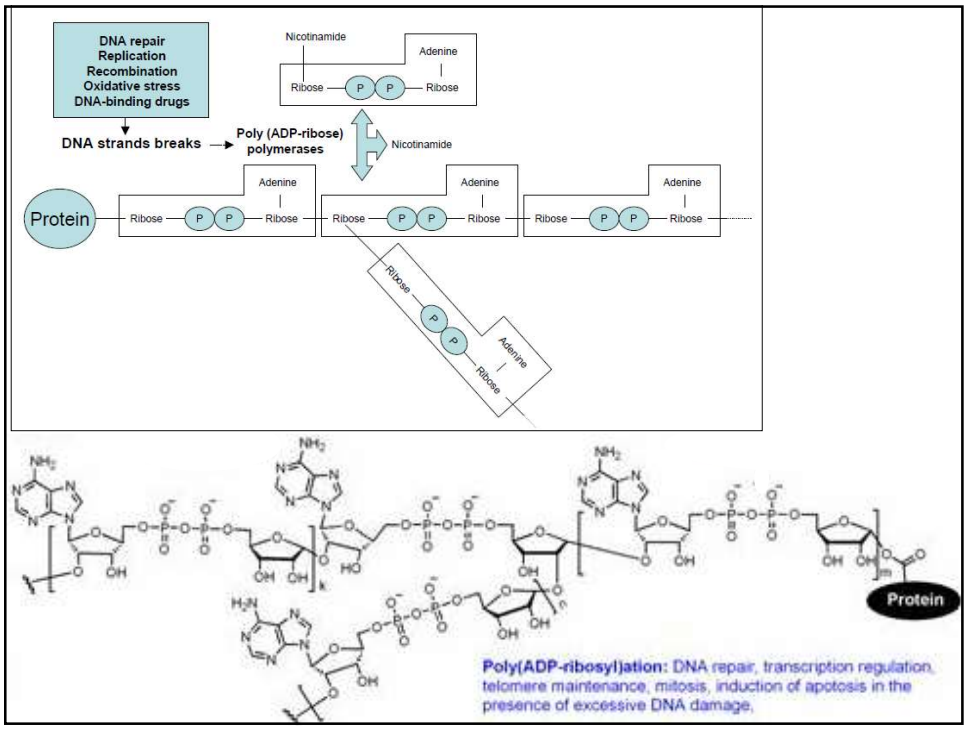

# Methylation

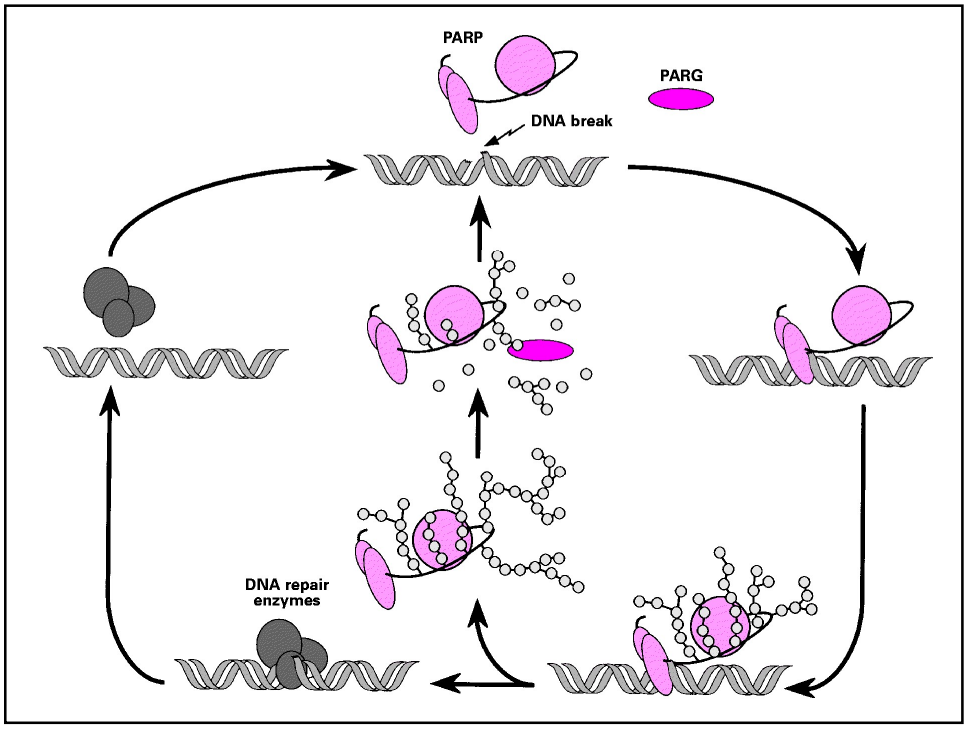



Methylation in *arginin* or *lysine* residues is an important PMT in proteins

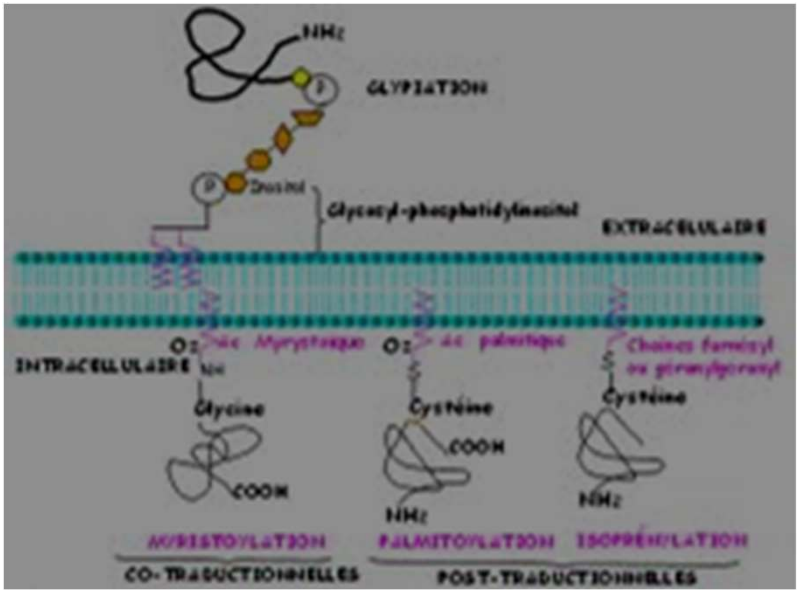
# ADP-ribosylation



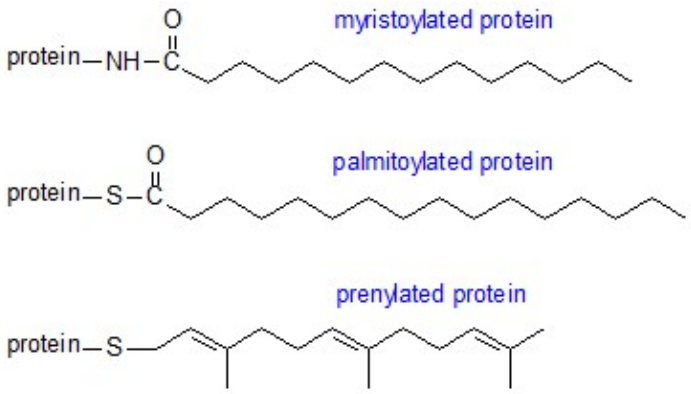




PARP
PARG
DNA break
DNA repair
enzymes

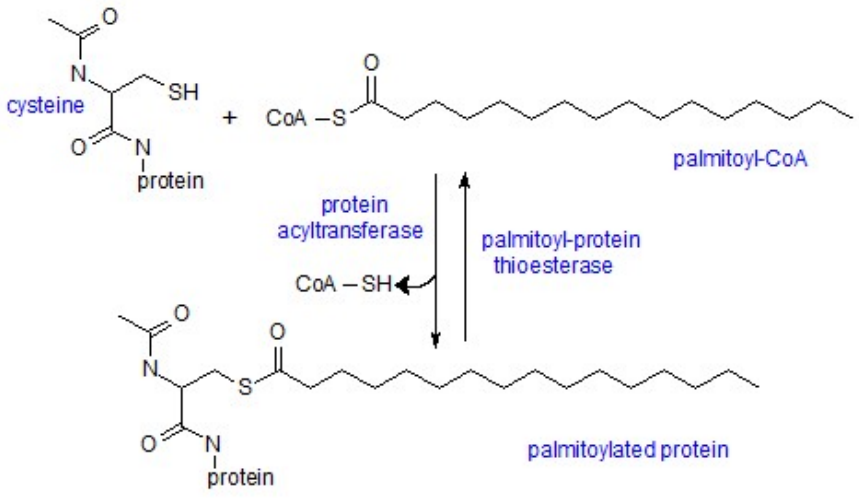
# Acylation


GLYPIATION
Glycosyl-phosphatidylinositol
EXTRACELLULAIRE
INTRACELLULAIRE
Glycine
Cystéine
Cystéine
CO-TRADUCTIONNELLES
POST-TRADUCTIONNELLES

# Acylation




# Palmitoylation



# Myristoylation

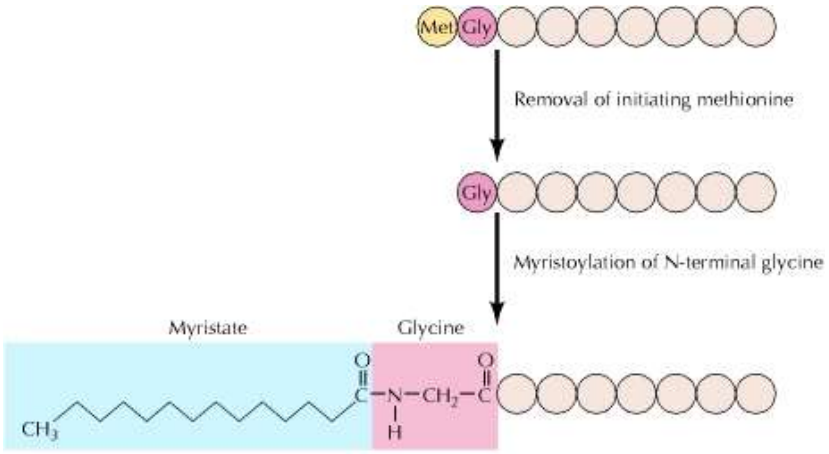
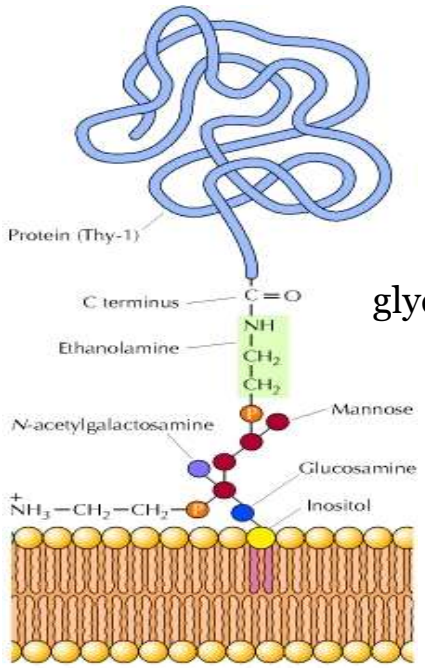



**Table 2.8** Representative examples of some acylated proteins. The fatty acid moiety attached and the location of eukaryotic-derived acylated proteins are also listed.

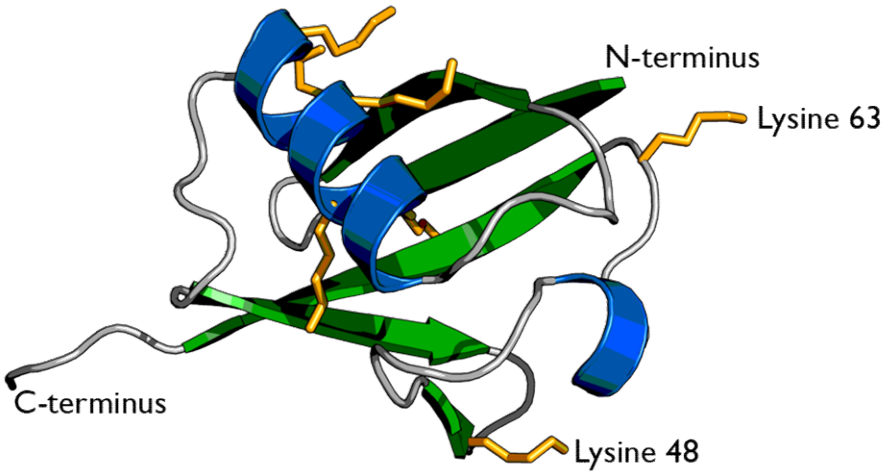
| Protein | Cellular location | Fatty acid |
|---|---|---|
| cAMP-dependent protein kinase | Cytoplasm | Myristic acid |
| Cytochrome $b_5$ reductase | ER and mitochondria | Myristic acid |
| $G_i$ and $G_o$ α-subunits | Plasma membrane | Myristic acid |
| Insulin receptor | Plasma membrane | Palmitic acid |
| Interleukin 1 receptor | Plasma membrane | Palmitic acid |
| Transferrin receptor | Plasma membrane | Palmitic acid |
| Rhodopsin | Disc membranes in retina | Palmitic acid |
| P55 and P28 | HIV | Myristic acid |
| VP4 | Picornaviruses | Myristic acid |
| P19 (gag) | HTLV I | Myristic acid |
| HA | Influenza virus | Palmitic acid |
| gE | Herpes simplex virus | Palmitic acid |

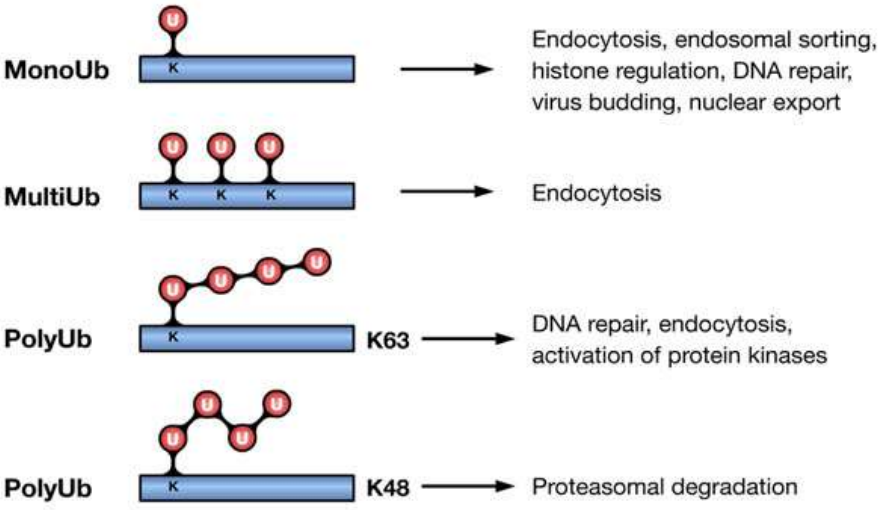


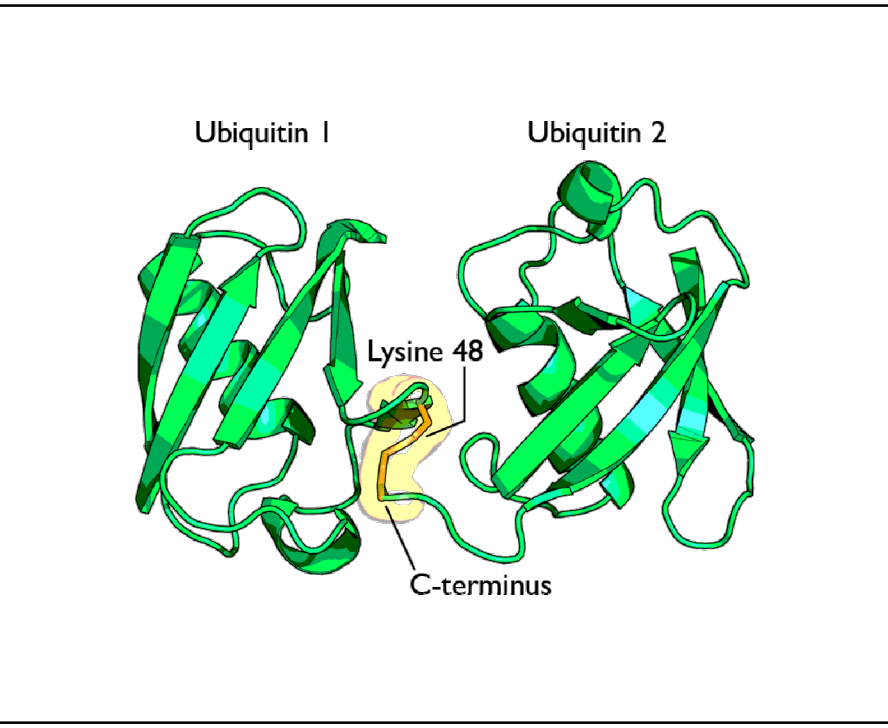

# GPI

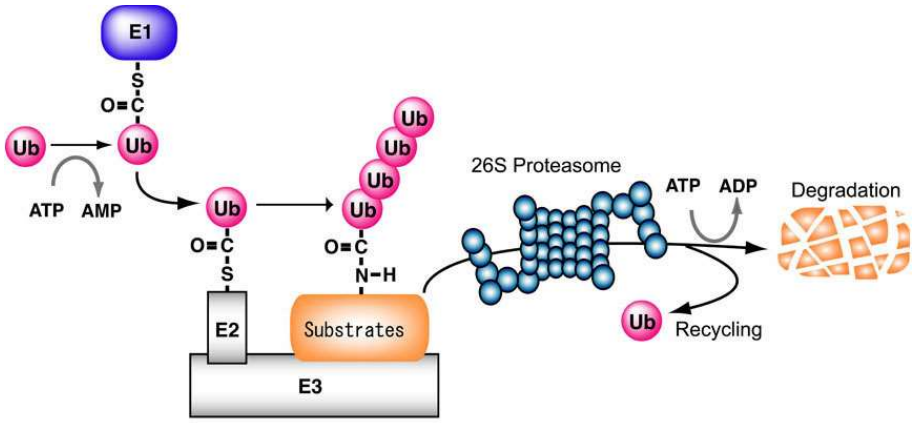
glycosylphosphatidylinositol

# Ubiquitination



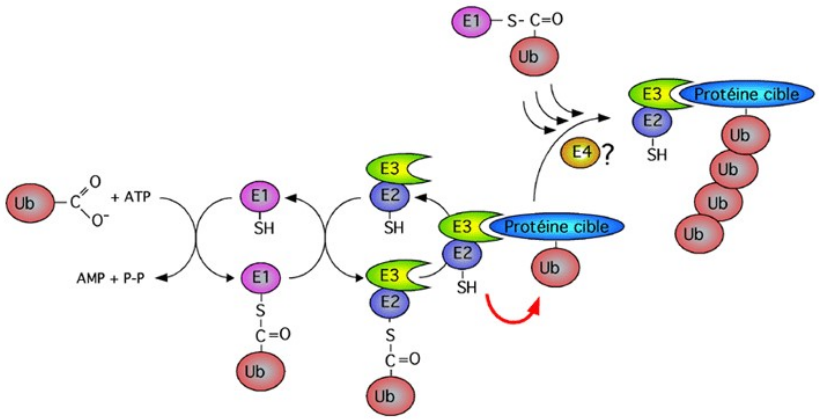

MonoUb
Endocytosis, endosomal sorting, histone regulation, DNA repair, virus budding, nuclear export
MultiUb
Endocytosis
PolyUb
K63
DNA repair, endocytosis, activation of protein kinases
PolyUb
K48
Proteasomal degradation


Ubiquitin 1
Ubiquitin 2
Lysine 48
C-terminus




1. Biochemical reactions of ubiquitination

# E4



# Proteasome

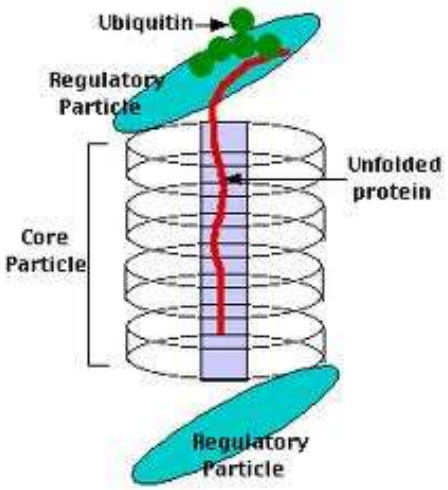



**Table 2.9** Some post-translational modifications (PTMs) associated with proteins for therapeutic use. The vast majority of such proteins are (recombinant forms of) native human extracellular proteins, and as such their PTM profile is biased towards PTMs characteristic of extracellular proteins derived from higher eukaryotes. The list is representative only, with specific examples being found throughout Chapters 6–9.

| Protein | Therapeutic application | PTM detail |
|---|---|---|
| Blood factor VIII | Haemophilia A (Chapter 6) | Glycosylation, disulfide bond formation, sulfation |
| Hirudin | Anticoagulant (Chapter 6) | Disulfide bond formation, sulfation |
| Tissue plasminogen activator (tPA) | Thrombolytic (Chapter 6) | Glycosylation, disulfide bond formation, proteolytic processing |
| α-Galactosidase | Fabry disease (Chapter 6) | Glycosylation, disulfide bond formation |
| Antibodies | Various (Chapter 7) | Glycosylation, disulfide bond formation |
| Insulin | Diabetes (Chapter 8) | Disulfide bond formation, proteolytic processing |
| Human growth hormone (hGH) | Dwarfism (Chapter 8) | Disulfide bond formation, phosphorylation |
| Erythropoietin (EPO) | Anaemia (Chapter 8) | Disulfide bond formation, glycosylation |
| Interferon β | Multiple sclerosis | Disulfide bond formation, glycosylation, phosphorylation |

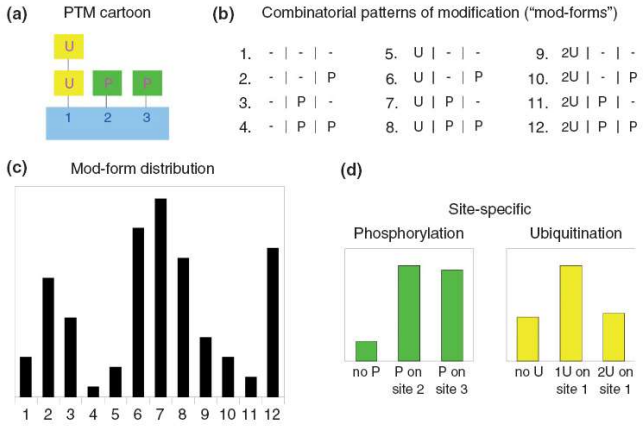



**FIGURE 6** | Mod-form distributions. (a) Cartoon depiction of a hypothetical substrate with 3 sites of modification; site 1 is ubiquitinated with a chain of up to two monomers; sites 2 and 3 are phosphorylated. (b) There are $12 = 3 \times 2 \times 2$ global patterns of modification, enumerated as shown. (c) A hypothetical mod-form distribution, showing the proportions in the population of each of the 12 mod-forms, following the numbering used in (b). The mod-form distribution can be viewed as a probability distribution, which gives, for each mod-form, the probability of finding a substrate molecule in that mod-form. The vertical scale has been omitted to focus on qualitative aspects. (d) In current practice, only limited information may be available. The separate phosphoryl- and ubiquityl-modifications calculated from (c) are shown, with the phosphoryl-modifications given as site-specific stoichiometries (the proportion of unphosphorylated substrate and of substrate phosphorylated on each site). Such summaries lose considerable information compared to the underlying mod-form distribution, making it harder to infer correlations between modification states and downstream responses.

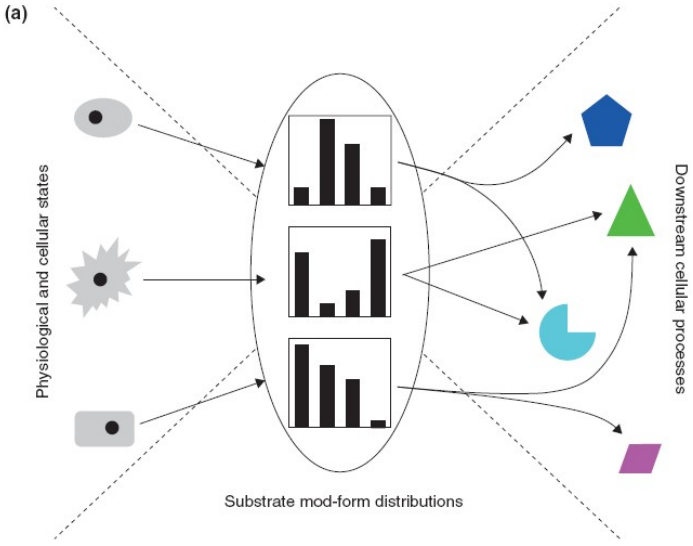



**FIGURE 7** | Encoding of information by mod-form distributions. (a) A bow-tie architecture describes the behavior of many of the examples discussed here. Distinct physiological and cellular states on the left can be represented ('encoded') by distinct mod-form distributions of a single substrate at the center of the bow-tie ('fan-in'). Each mod-form distribution can then orchestrate its own mix of downstream cellular processes, as on the right ('fan-out'). Figure 1 of Ref 45 reflects a similar architecture for the particular case of SRC-3. The mod-form distribution plays a central role here as the quantitative representation of the encoded information.