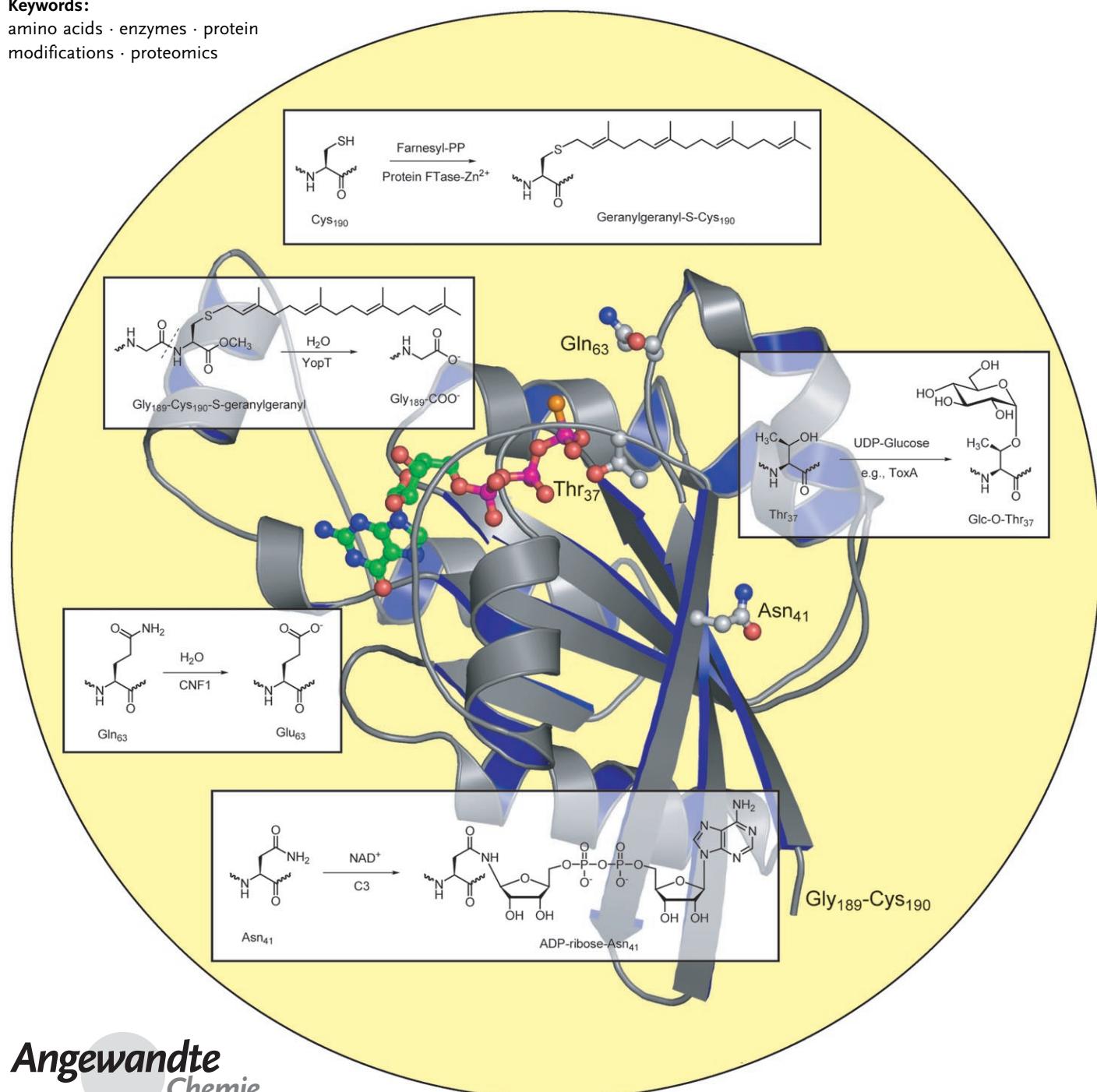


Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications

Christopher T. Walsh,* Sylvie Garneau-Tsodikova, and Gregory J. Gatto, Jr.

Keywords:

amino acids · enzymes · protein modifications · proteomics



**Angewandte
Chemie**

The diversity of distinct covalent forms of proteins (the proteome) greatly exceeds the number of proteins predicted by DNA coding capacities owing to directed posttranslational modifications. Enzymes dedicated to such protein modifications include 500 human protein kinases, 150 protein phosphatases, and 500 proteases. The major types of protein covalent modifications, such as phosphorylation, acetylation, glycosylation, methylation, and ubiquitylation, can be classified according to the type of amino acid side chain modified, the category of the modifying enzyme, and the extent of reversibility. Chemical events such as protein splicing, green fluorescent protein maturation, and proteasome autoactivations also represent posttranslational modifications. An understanding of the scope and pattern of the many posttranslational modifications in eukaryotic cells provides insight into the function and dynamics of proteome compositions.

1. Introduction

There are two major mechanisms for expanding the coding capacity of the 6000 (yeast) to 30,000 (human) genes in eukaryotic genomes to generate diversity in the corresponding proteomes, the inventory of all proteins in a cell or organism. Proteomes may be two to three orders of magnitude more complex (>1 000 000 molecular species of proteins) than the encoding genomes would predict. The first route of diversification of proteins is at the transcriptional level, by mRNA splicing, including tissue-specific alternate splicing.^[1,2] This is a central topic in RNA metabolism in eukaryotic biology.

The second route to proteome expansion is the focus of this Review: covalent posttranslational modification (PTM) of proteins at one or more sites.^[3] As the name implies, these are covalent modifications that occur after DNA has been transcribed into RNA and translated into proteins. The nascent or folded proteins, which are stable under physiological conditions, are then subjected to a battery of specific enzyme-catalyzed modifications on the side chains or backbones. Proteome diversification by covalent modification occurs in prokaryotes but is much more extensively encountered in nucleated cells, both in terms of types of modifications and frequency of occurrence. About 5% of the genomes of higher eukaryotes can be dedicated to enzymes that carry out posttranslational modifications of the proteomes.

Two broad categories of protein PTM occur (Scheme 1). The first subsumes all enzyme-catalyzed covalent additions of some chemical group, usually an electrophilic fragment of a cosubstrate, to a side chain residue in a protein. The side chain modified is usually electron rich, acting as a nucleophile in the transfer. The second category of PTM is covalent cleavage of peptide backbones in proteins either by action of proteases or, less commonly, by autocatalytic cleavage. Limited proteolysis to control location, activity, and lifetime of each protein in

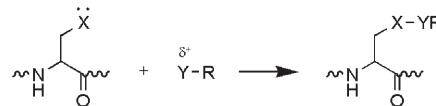
From the Contents

1. Introduction	7343
2. Covalent Addition: The Main Acts	7345
3. Covalent Addition: The Supporting Cast	7354
4. Cataloguing the Posttranslational Modification	7360
5. Multiple and Tandem Posttranslational Modification of Proteins	7361
6. Reversible versus Irreversible Posttranslational Modification	7362
7. Controlled Proteolysis	7364
8. Autocleavage and Peptide-Bond Rearrangement	7365
9. Peptide Bond Rearrangement without Autocleavage	7368
10. Conclusions	7369

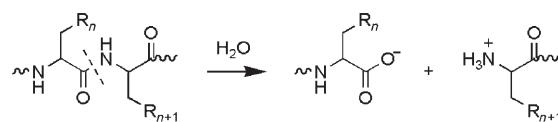
intracellular and extracellular milieus is a central strategy for the regulation of the composition and function of proteomes.

Protein covalent modifications can be sorted along several axes. One is by the identity of the protein side chain modified; 15 of the 20 common proteinogenic amino acid side chains

1. Covalent modification



2. Cleavage of protein backbone



Scheme 1. Two categories of posttranslational modifications of proteins: 1) covalent modification of a nucleophilic amino acid side chain by an electrophilic fragment of a cosubstrate; 2) cleavage of a protein backbone at a specific peptide bond.

[*] Prof. C. T. Walsh, Dr. S. Garneau-Tsodikova, Dr. G. J. Gatto, Jr.
Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
Boston, MA 02115 (USA)
Fax: (+1) 617-432-0348
E-mail: Christopher_walsh@hms.harvard.edu

undergo such diversification (Table 1). Another classification is that by the fragment of cosubstrate or coenzyme that is enzymatically coupled to the protein and the concomitant chemical nature of the protein modification. This catalogue includes *S*-adenosylmethionine (SAM)-dependent methylation, ATP-dependent phosphorylation, acetyl CoA dependent acetylation, NAD-dependent ADP ribosylation, CoASH-dependent phosphopantetheinylation, and phosphoadenosinephosphosulfate (PAPS)-dependent sulfurylation. A third axis of categorization of PTM is by the new function enabled by the covalent addition. These include gain in catalytic function of enzymes that have acquired tethered biotinyl, lipoyl, and phosphopantetheinyl groups, changes of subcellular address for proteins undergoing various lipid modifications (prenylation, palmitoylation, glycosyl phosphatidylinositol (GPI) anchor attachment), and targeting of the modified protein for proteolytic destruction by ubiquitylation to mark transport to lysosomes or proteasomes.



Christopher T. Walsh, born in 1944, majored in biology at Harvard and completed his PhD in biochemistry in the lab of Fritz Lipmann at the Rockefeller Institute of Medical Research. He was on the MIT faculty (1972–1987), and since 1987 has been at Harvard Medical School. He served as Chair of the Dept. of Chemistry at MIT (1982–1987) and of the Dept. of Biological Chemistry and Molecular Pharmacology at Harvard Medical School (1987–1995). His research interests lie in enzyme and inhibitor mechanisms and in the biosynthesis of non-ribosomal peptide antibiotics.



Sylvie Garneau, born in Québec, Canada, received her BSc (1995) and MSc (1997) in chemistry from the Université Laval, where she worked under the supervision of Robert Chênevert and Perséphone Canonne. She completed her PhD in chemistry in January 2003 at the University of Alberta with John C. Vederas on the studies of new antimicrobial agents acting on bacterial cell walls. She is currently a postdoctoral fellow with Christopher T. Walsh at Harvard Medical School, studying halogenation and pyrrole formation various of natural products.



Gregory J. Gatto, Jr., born in 1972, majored in chemistry at Princeton University, where he worked under the direction of Martin Semmelhack. In 2003, he received his MD and PhD degrees from the Johns Hopkins University School of Medicine. There, he worked in the lab of Jeremy Berg on the structural biology of peroxisomal targeting signal recognition. He is currently an NIH postdoctoral fellow in the lab of Christopher T. Walsh at Harvard Medical School, studying the biosynthesis of the macrolide immunosuppressants.

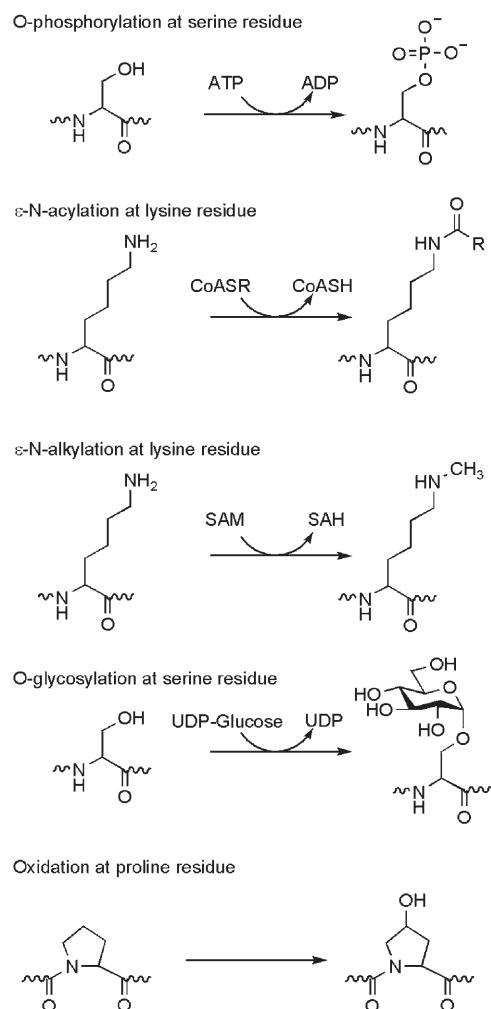
Table 1: Posttranslational protein modifications at the side chains.^[a]

Residue	Reaction	Example
Asp	phosphorylation isomerization to isoAsp	protein tyrosine phosphatases; response regulators in two-component systems
Glu	methylation carboxylation polyglycation polyglutamylation	chemotaxis receptor proteins Gla residues in blood coagulation tubulin tubulin
Ser	phosphorylation O-glycosylation phosphopantetheinylation autocleavages	protein serine kinases and phosphatases notch O-glycosylation fatty acid synthase pyruvamidyl enzyme formation
Thr	phosphorylation O-glycosylation	protein threonine kinases/phosphatases
Tyr	phosphorylation sulfation <i>ortho</i> -nitration TOPA quinone	tyrosine kinases/phosphatases CCR5 receptor maturation inflammatory responses amine oxidase maturation
His	phosphorylation aminocarboxypropylation N-methylation	sensor protein kinases in two-component regulatory systems diphthamide formation methyl CoM reductase
Lys	N-methylation N-acetylation by acetyl, biotinyl, lipoyl, ubiquityl groups C-hydroxylation	histone methylation histone acetylation; swinging-arm prosthetic groups; ubiquitin; SUMO (small ubiquitin-like modifier) tagging of proteins collagen maturation
Cys	S-hydroxylation (S-OH) disulfide bond formation phosphorylation S-acylation S-prenylation protein splicing	sulfenate intermediates protein in oxidizing environments PTPases Ras Ras intein excisions
Met	oxidation to sulfoxide	Met sulfoxide reductase
Arg	N-methylation N-ADP-ribosylation	histones G _{Sα}
Asn	N-glycosylation N-ADP-ribosylation protein splicing	N-glycoproteins eEF-2 intein excision step
Gln	transglutamination	protein cross-linking
Trp	C-mannosylation	plasma-membrane proteins
Pro	C-hydroxylation	collagen; HIF-1α
Gly	C-hydroxylation	C-terminal amide formation

[a] No modifications of Leu, Ile, Val, Ala, Phe side chains are known. A more extensive list can be found in reference [3].

2. Covalent Addition: The Main Acts

The five most common types of covalent additions to proteins are phosphorylation, acylation, alkylation, glycosylation, and oxidation, which are catalyzed by dedicated PTM enzymes (Scheme 2). The protein products obtained in this



Scheme 2. Five major types of covalent additions to protein side chains: phosphorylation, acylation, alkylation, glycosylation, oxidation.

manner make up subsets of the proteome of an organism: the phosphoproteome, the acyl proteome, the alkyl proteome, the glycoproteome, and the oxidized proteome. In turn, each of these subproteomes can contain substantial diversity.

2.1. Protein Phosphorylation

The mammalian phosphoproteomes have phosphoSer (pS), phosphoThr (pT) and phosphoTyr (pY) residues with a split of about 90:10 (pS, pT/pY; Figure 1).^[4] Bacterial and fungal phosphoproteomes will also have phosphoHis and phosphoAsp residues derived from proteins in two-component signal-transduction cascades.^[5] The pathogenic bacte-

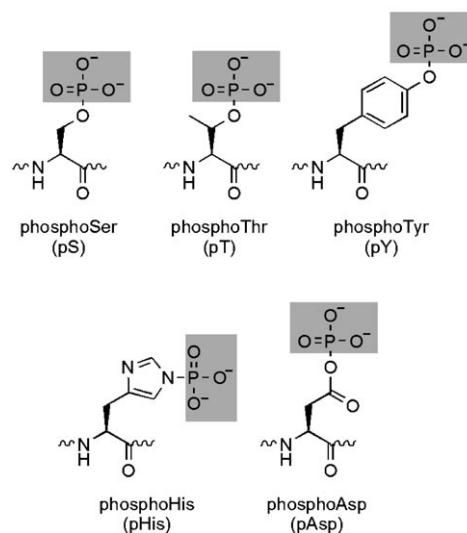


Figure 1. Phosphorylated forms of amino acid side chains in proteins: phosphoSer (pS); phosphoThr (pT); phosphoTyr (pY); phosphoHis (pHis); phosphoAsp (pAsp).

rium *Pseudomonas aeruginosa* exhibits more than 60 such pathways.^[6]

The enzymes dedicated to protein phosphorylation are among the largest class of PTM enzymes. This superfamily of protein kinases have been termed the kinase, with over 500 members in the human kinase.^[7] If these acted on average on only 20 different protein substrates, or different residues within a smaller subset of proteins, 10000 distinct molecular forms of phosphorylated proteins would be produced. This is most probably a substantial underestimate of the true size of phosphoproteomes of higher eukaryotes, in which phosphorylation sites can be predicted but, as yet, cannot be completely measured. For example, the enzymatic activity of Abl protein kinase is modulated by phosphorylation at up to 11 distinct residues (Tyr, Thr, Ser; Figure 2).

Introduction of the charged, dianionic tetrahedral phosphate group induces altered conformations in local protein microenvironments^[8] and is often paired with cationic arginine side chains (Figure 3). These local reorganizations of protein domains often create the architectural impetus for signal initiation, for example, in the four parallel MAP kinase pathways in eukaryotes and in the activation of many membrane receptor tyrosine kinases during autophosphorylation. The enzymatic conversion of a neutral OH side chain to dianionic phosphate has proven to be such a useful conformation switch in protein-domain restructuring that it has evolved into a major recurring theme in eukaryotic proteome diversity.

2.2. Protein Acylation

The most common acyl chains found in proteins that have undergone posttranslational modifications are C₂ (acetyl, e.g. histone tail acetylations),^[9] C₁₄ (myristoylation at glycine N termini),^[10] and C₁₆ (palmitoylated-S-Cys residues).^[11] The 8-kDa chain of the small protein ubiquitin^[12] (and congeners)

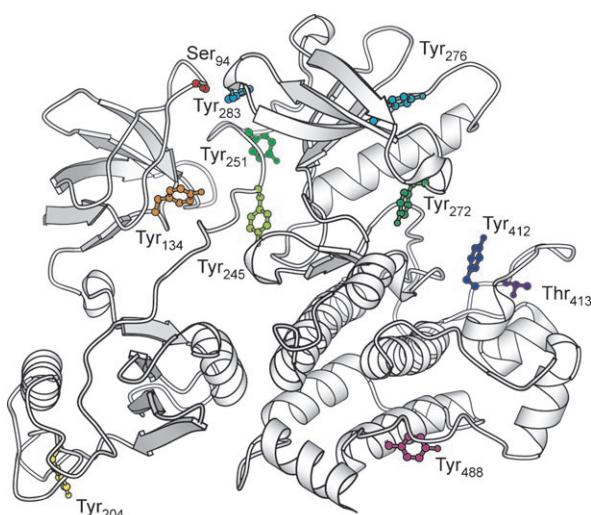


Figure 2. Eleven phosphorylation sites in the tyrosine kinase Abl, including Ser, Thr, and Tyr residues, color coded from red near the N terminus to purple near the C terminus. This figure and all other three-dimensional structural representations in this Review were generated with MolScript.^[122]

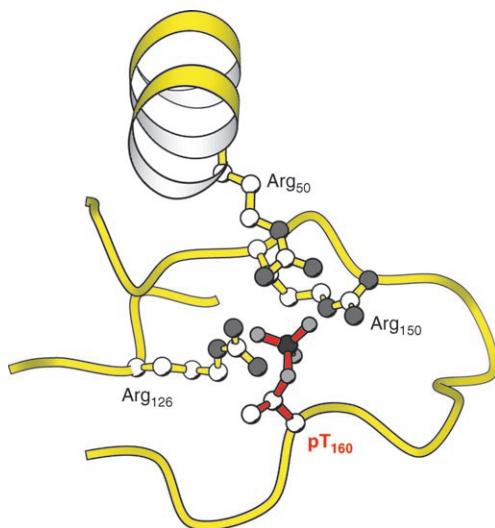


Figure 3. Conversion of a neutral OH side chain to a dianionic PO_3^{2-} side chain recruits cationic Arg side chains to make bifurcated charge-pairing interactions that can restructure the microenvironment of a protein region, induce a conformational change, and thereby initiate or propagate signaling information to partner proteins or small molecules. Shown is the interaction in Cdk2 of pThr₁₆₀ with the guanidinium side chains of Arg₅₀, Arg₁₂₆, and Arg₁₅₀.

is enzymatically transferred as an acyl moiety by ubiquityl ligases and chemically falls into this group of PTMs. The biological consequences of the acylation of a given protein with C₂, C₁₄, C₁₆, or 8-kDa chains are vastly different.

2.2.1. ϵ -N-Acetylation of Lysine

Acetylations of multiple lysine residues in histone N-terminal tails or at the C-terminus of the p53 transcription factor^[13] are viewed as an integral part of the epigenetic code

that controls selective gene transcription. The acetyl-group donor is the primary metabolite acetyl CoA, and dedicated histone acetyltransferase (HAT) isoenzymes select distinct combinations of the $\epsilon\text{-NH}_2$ groups of Lys side chains in the N-terminal tails of histones: two Lys residues on histone H2A, four on H2B, four on H3, and four on H4 (Figure 4). Given

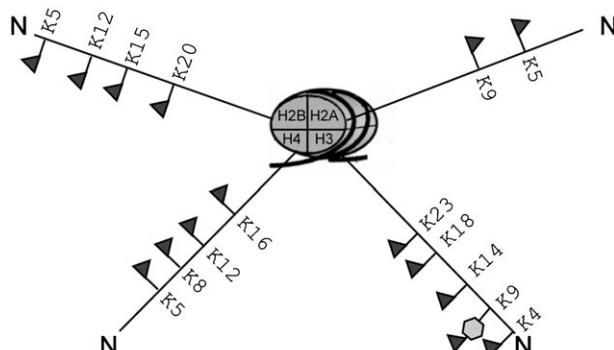


Figure 4. There are 15 Lys residues in the N-terminal tails of H2A, H2B, H3, and H4 that are sites for possible enzymatic acetylation. The dark gray flags represent sites of acetylation on the indicated lysine (K) residues.

two copies of each histone per octamer core in nucleosomes, there are $14 \times 2 = 28$ potential Lys side chains available for acetylation. Yeast histones with up to 13 acetylations have been reported,^[14] which reflects almost 50% posttranslational modification. The combinatorial possibilities for differentially acetylated histone tails in nucleosomes become astronomically large.

The acetyl groups on Lys side chains convert potentially cationic side chains into groups, thus altering the charge distribution. The N-acetyl Lys group is also specifically recognized by discrete protein domains, termed bromodomains (Figure 5), which are embedded in transcription factors and associated proteins. Thus, the acetylation state of histone tails can regulate the recruitment of transcription factor machinery that controls the initiation of transcription of the genes in the region of chromatin covered by those nucleosomes.^[14] In Section 5, we describe how five Lys residues at the C terminus of the transcription factor p53 can either be acetylated or ubiquitylated. Acetylation blocks the Lys side chains from ubiquitylation and prolongs the half-life of the p53 protein molecule.

2.2.2. N-Myristoylation and S-Palmitoylation

Myristoylation of eukaryotic proteins at an N-terminal glycine residue is catalyzed by the PTM enzyme protein N-myristoyltransferase, which utilizes the C₁₄ myristoyl CoA as the donor substrate.^[15] As protein synthesis is initiated with N-terminal methionine residues, cotranslational hydrolysis of the Met₁-Gly₂ bond by methionine aminopeptidase is a prerequisite to myristoylation (Scheme 3). The newly generated free amino group of the now N-terminal Gly is the nucleophile in the acylation reaction. The introduced hydrophobic C₁₄ fatty acyl group can be a membrane-directing

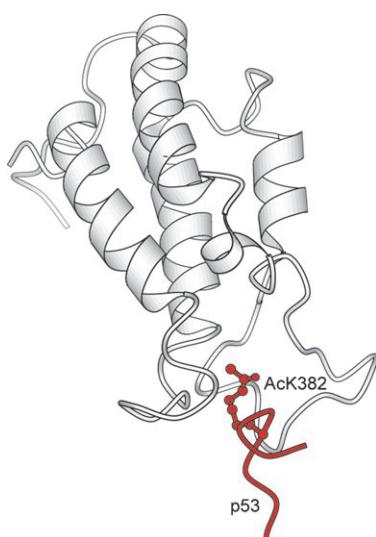
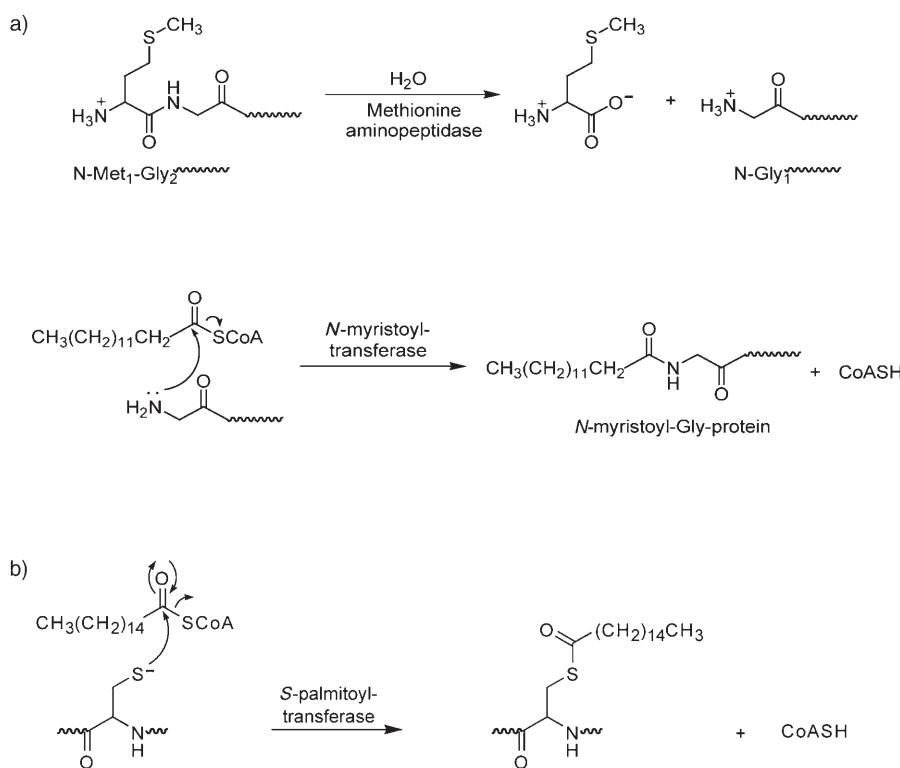


Figure 5. The acetylation status of the histone N-terminal tail can recruit partner proteins to control transcriptional activity on a nucleosome. The interaction of an acetyl-ε-NH-Lys side chain on a peptide fragment of histone H3 with a bromodomain of coactivator protein is redrawn from reference [32].



Scheme 3. a) Prior cleavage of the Met₁-Gly₂ peptide bond by methionine aminopeptidase liberates the NH₂ of the Gly residue for N-myristylation by N-myristoyltransferase with myristoyl CoA as donor. b) The thiolate side chain of a Cys residue as nucleophile towards palmitoyl CoA, catalyzed by palmitoyl-S-protein transferases.

group to move proteins to membrane interfaces.^[10] Additionally the myristoyl tail can switch from being buried in a cleft within the protein in one state to being available for membrane insertion in another conformational state.^[16]

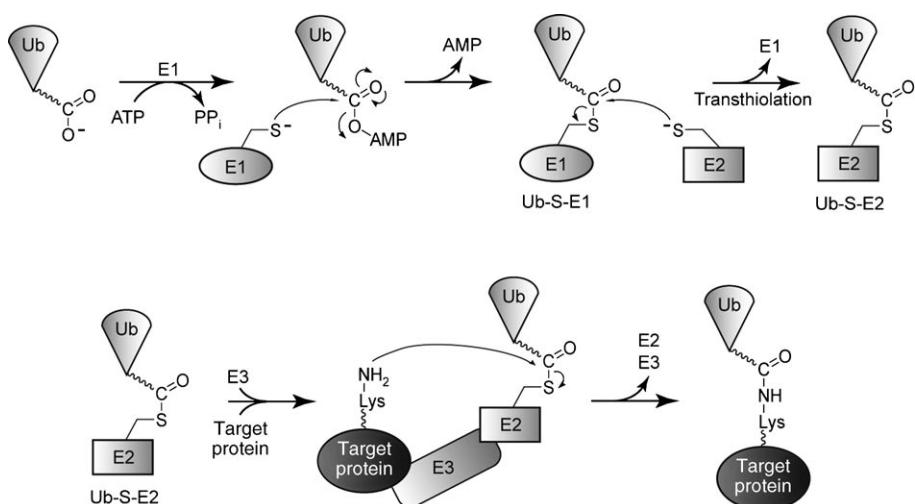
Examples of N-myristoylated proteins include HIV Gag protein and protein kinase A.^[17]

Palmitoylation of proteins requires the C₁₆ fatty acyl CoA as donor for the PTM acyltransferases, but the acyl group is typically transferred to the sulphydryl side chains of Cys residues rather than to N-termini of proteins (Scheme 3).^[11] Among the most-well-studied examples are S-palmitoylation of cysteines in the C-terminal regions of proteins such as the Ras GTPase. The lipidation of these nucleophilic thiolate side chains is consequential for partitioning Ras from the cytoplasm to membrane interfaces to meet up with its signaling partner proteins.^[10] As explained in Section 5, the S-palmitoylation of Ras isoforms is part of a cascade of reversible posttranslational modifications involved in maturation and membrane anchoring of modified forms of Ras.^[18] Protein substrates for S-palmitoylation can be transmembrane receptors such as the CD8α chain and the CCR chemokine receptor or cytoplasmic proteins such as the protein tyrosine kinase Lck.^[11]

2.2.3. Mono- and Polyubiquitylation of Proteins

An extension of the logic of posttranslational transfer of low-molecular-weight (C₂, C₁₄, C₁₆) acyl chains to proteins is the acylation on lysine ε-amino groups by the carboxy terminus of the 8-kDa protein ubiquitin. In analogy to acyl CoA donors of the electrophilic acetyl, myristoyl, and palmitoyl acyl groups, the C-terminal carboxy group of the 76-residue ubiquitin must be preactivated for acyl transfer. The activation principle is the same as that of the acyl thioester system, but eukaryotic cells make use of ubiquityl-S-protein as donors instead of ubiquityl CoA.^[19,20] The enzymatic activation machinery involves an enzyme 1 to make ubiquityl-AMP and a set of about a dozen thiol-containing enzymes 2 (Scheme 4) that provide the active-site nucleophile to capture the ubiquityl group as a set of ubiquityl-S-enzyme 2 family members. In general a third set of proteins, known collectively as enzyme 3 variants, are then required to catalyze the efficient transfer of the activated ubiquityl protein tag to Lys side chains of client proteins. Some of the enzyme 3 subclasses are multicomponent catalysts, with up to four subunits, which provide selectivity

for a given protein target.^[21] There are several hundred isoforms of such E3 ubiquityl ligases in higher eukaryotes, which allow subtle discrimination among many target proteins selected for ubiquitylation.^[22]



Scheme 4. Activation of the C-terminus of ubiquitin (at residue Gly₇₆) by enzyme 1 to make Ub-AMP, followed by transfer to a Cys thiolate in the active site of enzymes 2 to yield Ub-S-E2. Enzymes 3 can act as chaperones and recruiters of specific proteins for ubiquitylation at Lys side chains.

Unlike the small-molecule acyl groups described in Section 2.2.2., the ubiquityl acyl moiety provides an information-rich architectural scaffold (Figure 6) that can be read by particular partner proteins that control the downstream

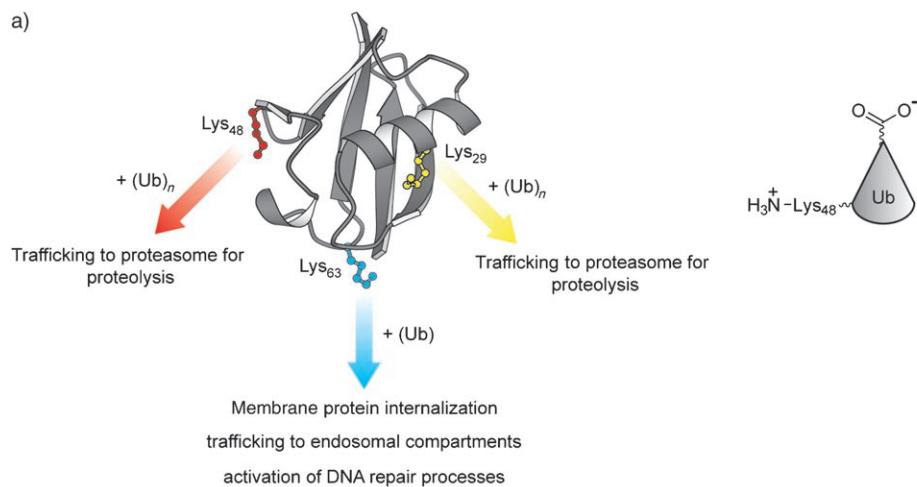
biological response. Two types of protein ubiquitylation can be distinguished according to the number of ubiquitins added: monoubiquitylation and polyubiquitylation. Polyubiquitylation specifically refers to the enzymatic construction of chains of ubiquitin molecules. As shown in Figure 6a, ubiquitin has Lys side chains on different surfaces and tandem attachment could go from any Lys unit on one ubiquitin monomer to the C-terminal Gly₇₆ on the preceding monomer. It appears that polyubiquitin chains are built up most often through Lys₄₈ side chains, although chains tethered through Lys₆₃ are also well-known.^[23] The X-ray structure of a Ub₄ chain is shown in Figure 6b. It is not clear how E3 ligases act processively to build up Ub_n chains (up to Ub₂₀) tethered to proteins in cells.

Monoubiquitylation and polyubiquitylation consign proteins to relocation in cells,

most often with the net consequence of proteolytic degradation, albeit by quite different mechanisms. Protein monoubiquitylation at a Lys-ε-NH₂ group by an E3 ubiquitin ligase can initiate relocation of transmembrane receptor proteins from the plasma membrane to the trans Golgi network sorting compartments.^[24–26] The covalent Ub tag is information-rich and recruits various partner proteins that contain one or more of several variants of Ub-binding domains. The partner proteins can act as chaperones for internalization of the ubiquitylated receptor, import into early endosomes, and transit to lysosomes in which lysosomal proteases can cause hydrolytic degradation (Scheme 5).^[23] These pathways are part of the homeostatic regulation of receptor density and lifetimes at plasma membranes.

In contrast to the sorting fate of monoubiquitylated proteins, polyubiquitylation sends modified proteins to the chambered proteases that constitute the proteasome. Tandem attachment of four or more ubiquitin molecules to such a lysine side chain constitutes an architectural signal that recruits protein chaperones with ubiquitin-binding domains. The chaperone complexes then escort the marked protein to the proteasomes,^[27] where the chaperones dissociate, and the polyubiquitin chain is removed hydrolytically, perhaps during ATP-driven unfolding of the targeted protein (Scheme 6). The unfolded protein is then threaded

a)



b)

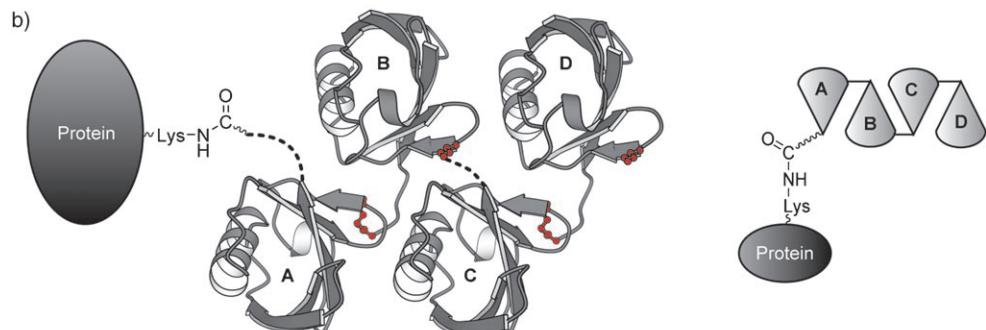
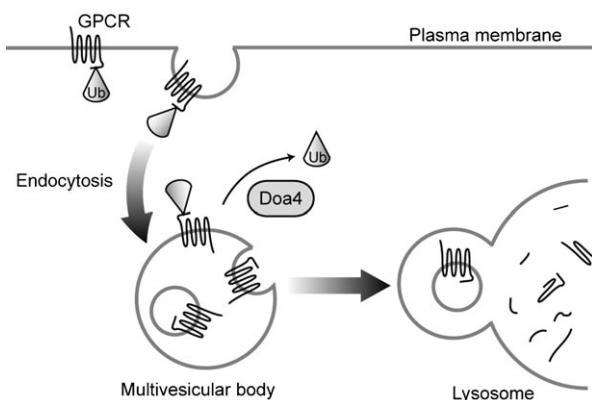
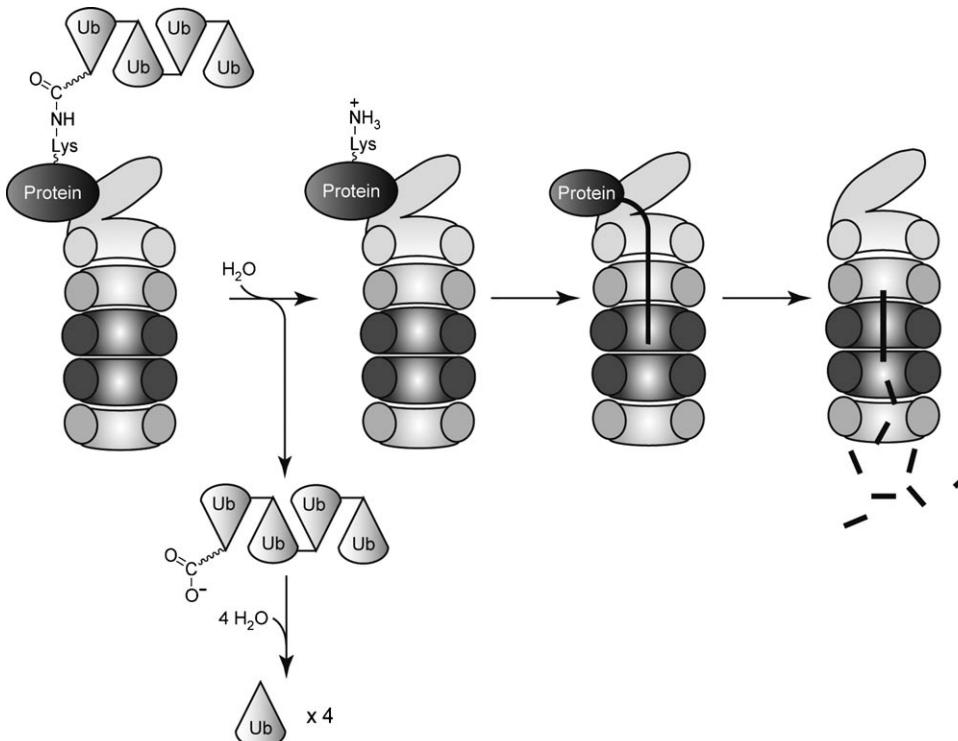


Figure 6. a) 3D trace of the 76-residue ubiquitin: Lys_{29,48,63} side chains on different faces of ubiquitin offer different surfaces for tandem conjugation of growing polyubiquityl chains; b) structure of a tetraubiquityl unit, the minimum chain length to direct polyubiquitylated proteins to the proteasome.



Scheme 5. Recognition of the Ub architecture in a monoubiquitin-tagged protein by partner proteins/chaperones that have one or more ubiquitin-recognition domains for transit to the secretory system.

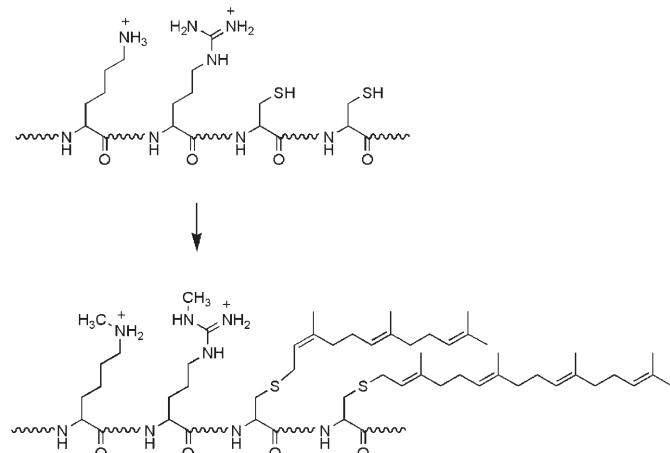
into the chamber of the proteasome where the active sites of the protease subunits degrade it to small peptides. The characteristic temporal control of proteins destroyed during the cell cycle, such as the cyclin subunits of cyclin-dependent protein kinases,^[28,29] is effected by the E1-E2-E3 ubiquitin ligase machinery. The activity of various multisubunit E3 ligases can be controlled by posttranslational states of the catalysts or of the target proteins, such as phosphorylation of particular Ser and Thr residues.



Scheme 6. Recognition of Ub_n-tagged protein for chaperoning to proteasomes where the Ub_n tag is retrieved by hydrolysis of the isopeptide link to the target protein; the target protein is unfolded and threaded into the chamber of the proteasome.

2.3. Protein Alkylation

Alkyl substituents are attached regiospecifically to proteins by posttranslational modification enzymes. The three common alkyl groups transferred are the methyl (C₁), for example, in histone methylations of Lys and Arg side chains,^[30] or the C₁₅ and C₂₀ isoprenyl (farnesyl and geranylgeranyl) groups (Scheme 7).^[31] The small C₁ and the large



Scheme 7. Alkyl groups transferred to protein side chains: the methyl group from Sadenosylmethionine (SAM) is transferred most often to Lys and Arg side chains (although O-, S-, and C-methylations of protein side chains are known); the two isoprenyl units transferred by protein prenyltransferases to Cys side chains are the C₁₅ (farnesyl) and the C₂₀ (geranylgeranyl) groups from the corresponding prenyl diphosphate substrates.

hydrophobic C₁₅ and C₂₀ groups each serve to introduce hydrophobicity but they do so to very different degrees.

2.3.1. N-Methylations

Whereas C-, O-, and S-methylations of protein side chains are known,^[3] the reactions of most contemporary interest are the N-methylations of Lys and Arg side chains, particularly on the same histone tails that are acetylated. Indeed, covalent posttranslational N-methylation of histone tails complements acetylation as the second main part of writing and reading the histone code (Figure 7). For example, 7 of the first 36 residues, Arg_{2,17,26} and Lys_{4,9,27,36} of histone H3 are known to be methylated

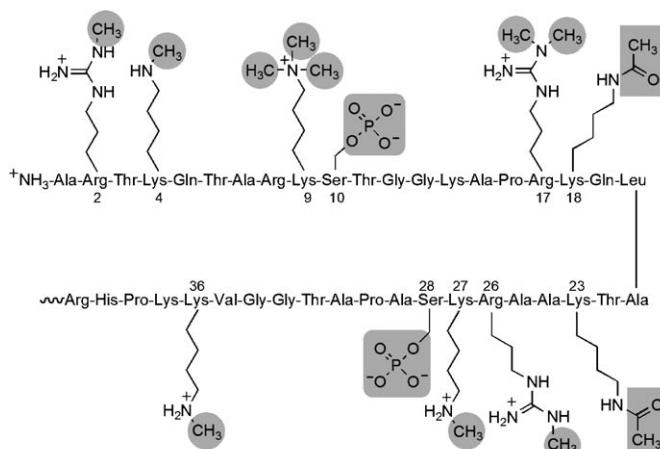
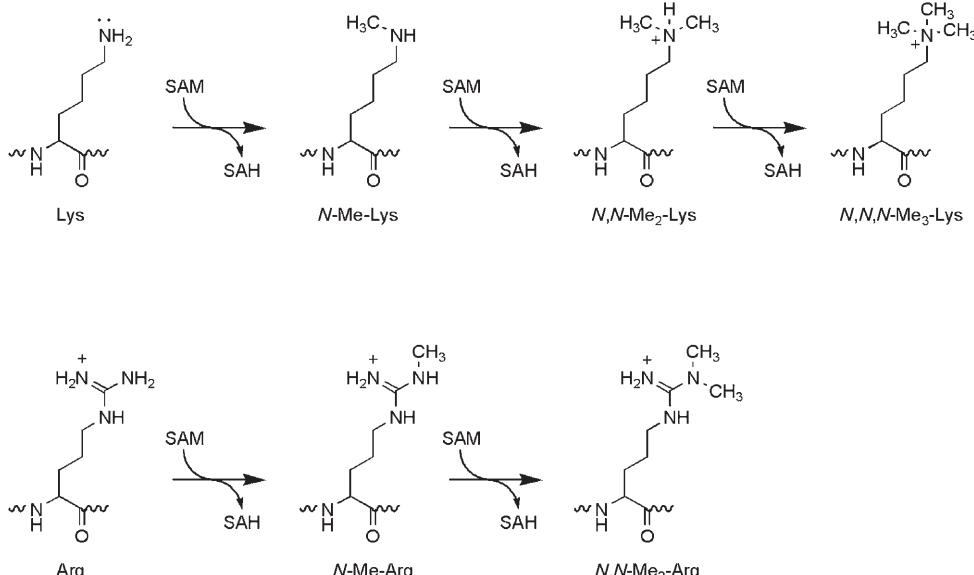


Figure 7. N-methylations can occur at Arg_{2,17,26} and Lys_{4,9,27,36} of histone H3. Lys₁₈ and Lys₂₇ can be acetylated and Ser₁₀ and Ser₂₈ phosphorylated.

by a family of histone methyltransferases,^[32] several of which are residue-specific. An additional layer of information content and combinatoric complexity is enabled by the fact that Lys-ε-NH₂ groups can be progressively mono-, di-, or trimethylated, again with distinct distributions shown by different methyltransferases. Analogously both monomethyl- and dimethylArg residues are observed (Scheme 8). The size and hydrophobicity differences between monomethyl- and trimethyl substituents on Lys side chains enable selective recruitment of proteins involved in transcriptional control. For example, trimethyl-Lys₉ in H3 recruits partner protein HP1 by binding to its chromodomain (Figure 8) as part of transcription factor and coactivator protein complex assemblies.^[14,32]



Scheme 8. Progressive mono-, di-, and trimethylation of Lys side chains and mono- and dimethylation of Arg side chains in histone N-terminal tail regions. SAH = S-adenosylhomocysteine.

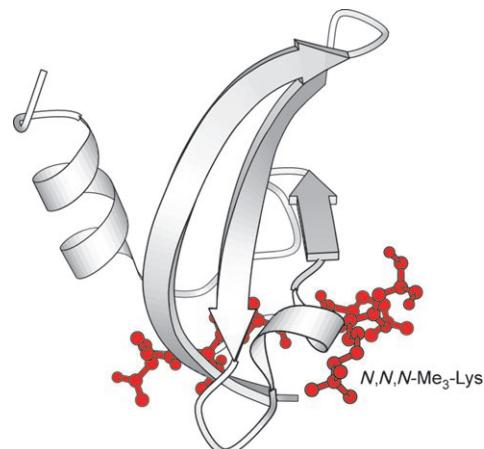


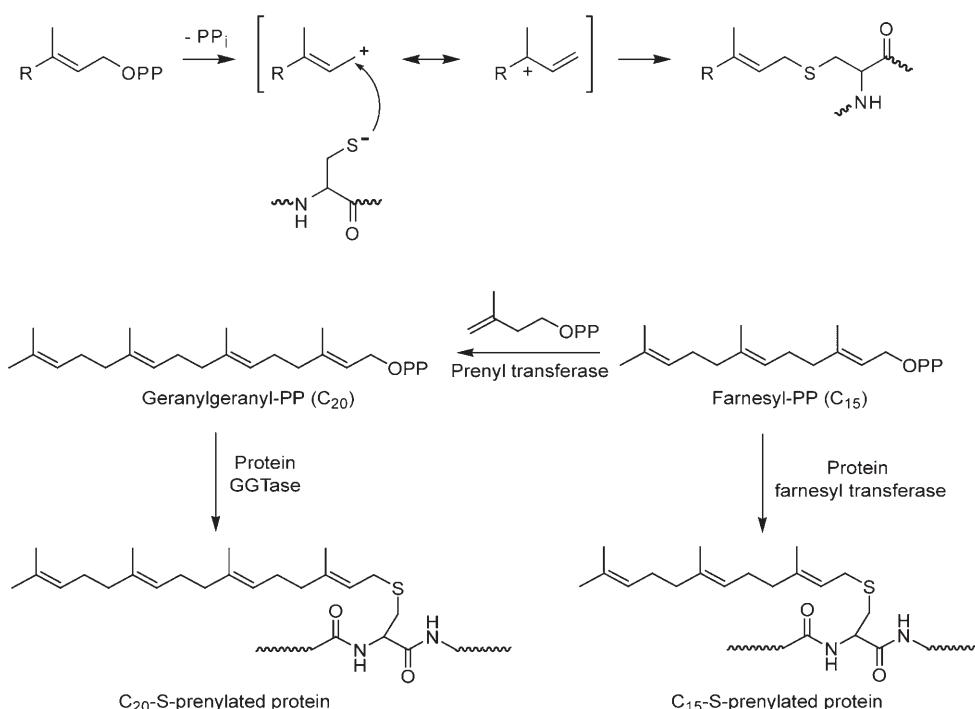
Figure 8. N,N,N-Tri-methyl-Lys₉ of histone H3 as a ligand to recruit protein HP1 through its chromo domain.

N-Acetylations and N-methylations of side chains in histone tails can have opposite effects on gene transcriptional silencing and activation. Histone H3 is just one of the four histones, each present as a dimer, in the nucleosome core. The balance of the four acetylations and seven methylations on the tail of histone H3 give, on this subunit alone, $11 \times 2 = 22$ sites for titration of transcriptional coactivator and corepressor complexes.

2.3.2. Protein S-Prenylation

The C₁₅ farnesyl and C₂₀ geranylgeranyl lipid groups are built up from C₅ isoprenyl diphosphate primary metabolites by iterative alkyl extension by action of C–C bond-forming enzymes.^[33] The farnesyl and geranylgeranyl-PP molecules can be used for further isoprene elongation (e.g. C₁₅ dimerization to squalene in the cholesterol biosynthetic pathway) or they can be utilized as electrophilic alkyl donors in posttranslational protein prenylation (Scheme 9).^[34]

There are protein farnesyltransferases and protein geranylgeranyltransferases, which are αβ heterodimers that share a common α subunit. The Ras GTPase superfamily have members that can be prenylated on Cys thiolate side chains, some with the C₁₅ and some with the C₂₀ prenyl chain. Typically, Ras family proteins that have a CaaX motif at the C terminus are farnesylated, when X is a small amino acid such as Ala or Ser. When X is Leu as in both Rac and RhoA GTPases, then the cysteine residue (C) is geranylgeranylated (Scheme 10 a). The Rab subfamily of GTPases, over 60 in number, have two cysteine residues at or near the C terminus, for example, in a CCXX arrangement. Both cysteine groups undergo posttranslational gera-



Scheme 9. Mechanism for Cys S-isoprenylation by protein prenyltransferases.

nylgeranylation (Scheme 10b), thus introducing two C₂₀ lipid anchors.^[35] Rab proteins cycle between membrane vesicles in the secretory system; escort proteins are used to control location and cycling.^[36,37]

Some proteins undergo both N- and S-acylation and some combination of S-prenylation. All of these lipid anchors drive

the modified proteins to partition more to membranes, thus controlling subcellular localizations.

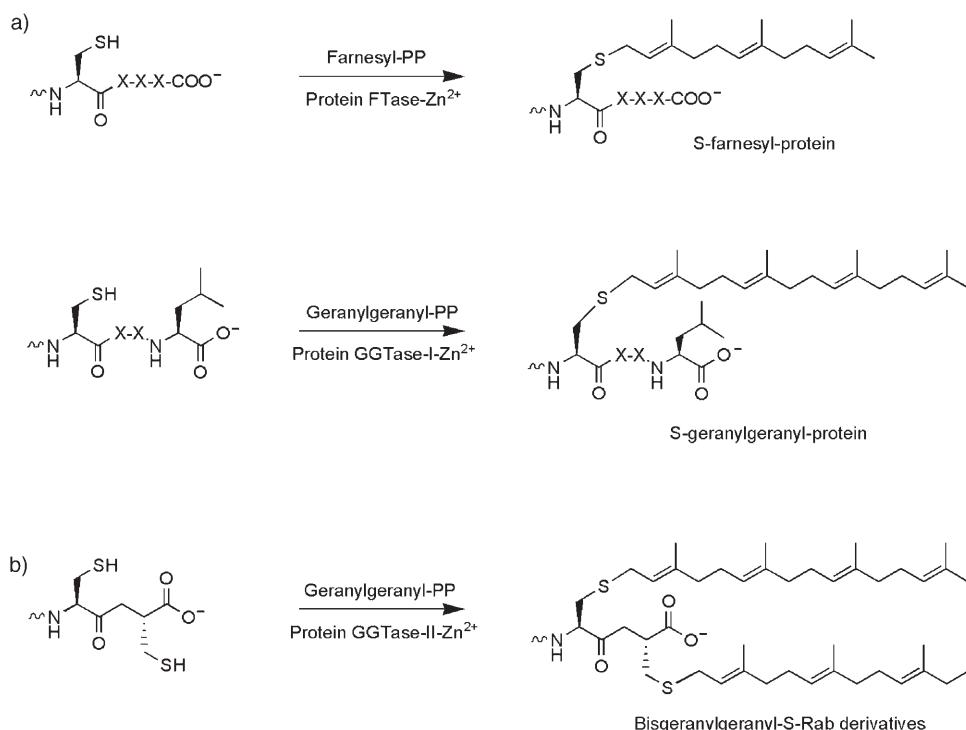
2.4. Protein Glycosylation

Covalent glycosylation of proteins is relatively rare in prokaryotes and quite common in eukaryotes. C-glycosylation, O-glycosylation, and N-glycosylation of proteins are known, but C-glycosylation, specifically mannosylation of C2 of the indole ring of tryptophan residues,^[38] is quite rare.

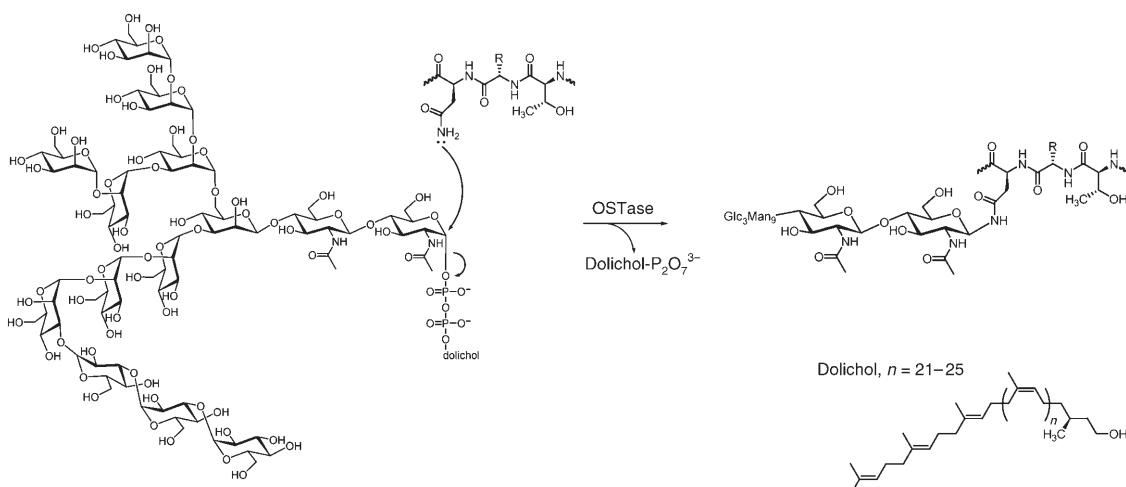
2.4.1. N-Glycosylation

N-glycoproteins are both more common and typically more complex in structure and architecture than O-gly-

coproteins in eukaryotes.^[39] The branching glycan unit in N-glycoproteins is preassembled on a lipid diphosphate scaffold by a series of membrane-associated glycosyltransferases in the endoplasmic reticulum.^[40] The assembled N-glycan unit that serves as a substrate for the multisubunit oligosaccharyltransferase is a tetradecasaccharyl-PP-dolichol substrate



Scheme 10. Prenylation reactions at the C termini of the Ras protein superfamily: a) farnesylation of the C terminus of Ras at CaaX (X = Ala, Ser). Geranylgeranylation of Rac at CaaX (X = Leu); b) double geranylgeranylation of the CC carboxy terminus of Rab proteins.

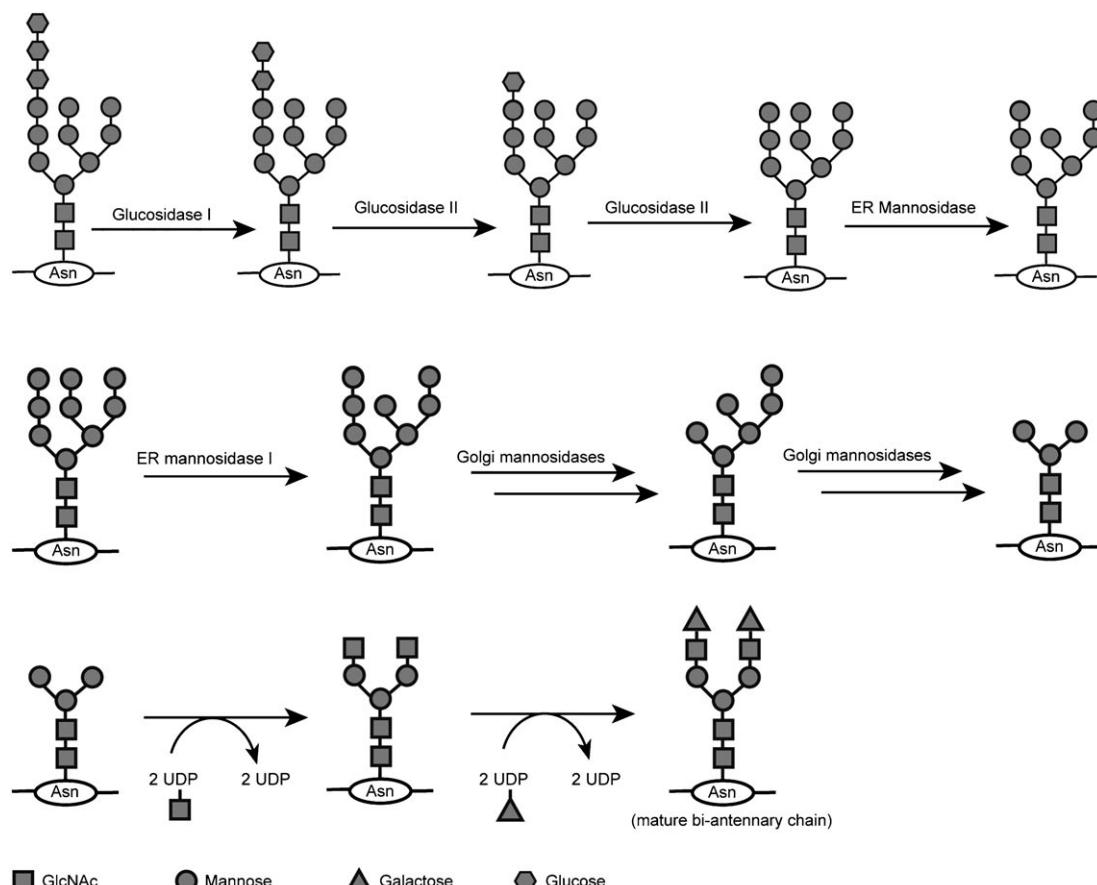


Scheme 11. The branched tetradecasaccharyl-PP substrate is the donor substrate in the Asn N-glycosylation reaction catalyzed by the oligosaccharyltransferase that initiates N-glycoprotein modifications.

(Scheme 11).^[41] The side-chain atoms that are modified in N-glycoprotein biogenesis are the carboxamide nitrogen atoms of asparagine groups, almost always in the sequence Ser/Thr-X-Asn. The low nucleophilicity of the Asn CONH₂ is thought to be enhanced by hydrogen bonding to the Ser/Thr-OH side

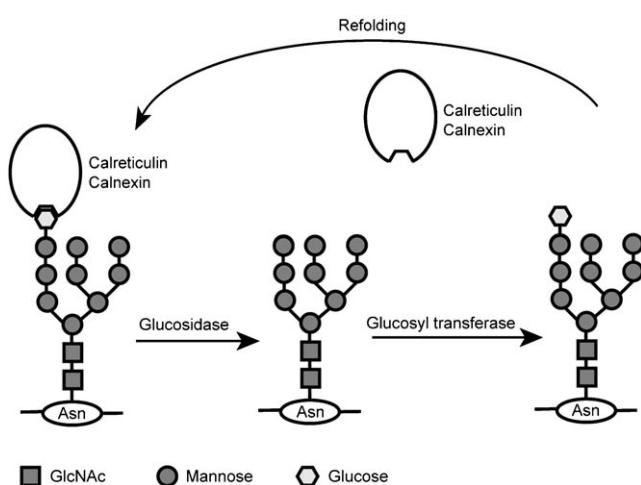
chain but the initial glycan transfer step is still poorly understood.

The initial tetradecasaccharyl chain Glc₃Man₉(GlcNAc)₂ then undergoes a remarkable enzymatic hydrolytic trimming and refashioning of the identity and linkages of the N-glycan chains (Scheme 12). The first hydrolytic tailoring enzymes are



Scheme 12. Progressive trimming of the initial N-linked Glc₃Man₉(GlcNAc)₂ glycan chain to Man₉(GlcNAc)₂ in the ER and then to the core pentasaccharide Man₉(GlcNAc)₂ in the Golgi complex before being built back up to mature N-glycan chains found on N-glycoproteins that have transited the secretory system.

two glucosidases in the ER lumen which generate the dodecasaccharyl $\text{GlcMan}_9(\text{GlcNAc})_2$ N-linked proteins. These are ligands for calnexin and calreticulin, protein chaperones that recognize the dodecasaccharyl chain and help refold the nascent glycoproteins as they are extruded through the ER membrane into the lumen. As the remaining Glc is hydrolyzed by the ER glucosidase, the chaperones lose affinity for the undecasaccharyl-N-protein products. A UDP-glucose glycoprotein glucosyltransferase^[42] puts the Glc residue back on the undecasaccharide unit to give the calreticulin and calnexin chaperones another round of assisted refolding. Thus, this is a refolding/protein-quality-control way station during glycoprotein secretion. If the glycoprotein cannot be refolded after several cycles of deglycosylation/reglycosylation, it is targeted for export back to the cytoplasm. There it undergoes polyubiquitylation, and proteasome-mediated degradation of the unfolded protein by a glycoprotein-targeting E3 ligase as part of the ER-accelerated degradation (ERAD) quality-control pathway (Scheme 13).^[42]



Scheme 13. Cycling of the oligosaccharyl chain between $\text{GlcMan}_9(\text{GlcNAc})_2\text{-Asn}$ and $\text{Man}_9(\text{GlcNAc})_2\text{-Asn}$ in the ER from opposing action of glucosidase and UDP-glucose glycoprotein glucosyltransferase. The Glc-containing oligosaccharyl chain is recognized by the chaperone proteins calreticulin and calnexin to help refold the nascent N-glycoproteins that have been glycosylated and extruded cotranslationally into the ER lumen. The glycosylation cycle is part of the protein quality-control system for proteins secreted into the endoplasmic reticulum.

For $\text{Man}_9(\text{GlcNAc})_2\text{-N-glycoproteins}$ that have refolded and passed into the Golgi compartment, six of the mannose residues are then trimmed hydrolytically by mannosidases to yield the $\text{Man}_3(\text{GlcNAc})_2$ pentasaccharyl core found in all mature N-glycoproteins. At this point the branched oligosaccharide core is rebuilt back up to biantennary and triantennary oligosaccharides characteristic of mature N-glycoproteins present on the cell surface.^[43] The multiplicity of glycosyltransferases in the Golgi can create enormous diversity in the mature N-glycan chains.

Multiple Asn side chains can be glycosylated in a given protein, and the identity of the glycan chains at each Asn

residue can vary depending on stochastic encounters with at least 10 trimming and rebuilding enzymes during passage through the ER and Golgi compartments.^[41,43] It has been estimated that about a third of all proteins that enter secretory pathways in eukaryotic cells may be N-glycosylated, and so tens of thousands of glycoprotein variants may coexist in eukaryotic cells. For example, 52 glyco forms of the prion protein have been reported.^[44,45]

2.4.2. O-Glycosylation

O-glycosyl chains in eukaryotic proteins are generally shorter and less complex than those in N-glycoproteins. Many proteins contain the monosaccharide GlcNAc^[46] that is put on by a specific O-GlcNAc transferase and removed by a corresponding hydrolase. Other proteins such as the signaling protein Notch contain tri- and tetrasaccharides in the EGF repeat domains (Figure 9).^[47,48] O-glycosylation is a crucial

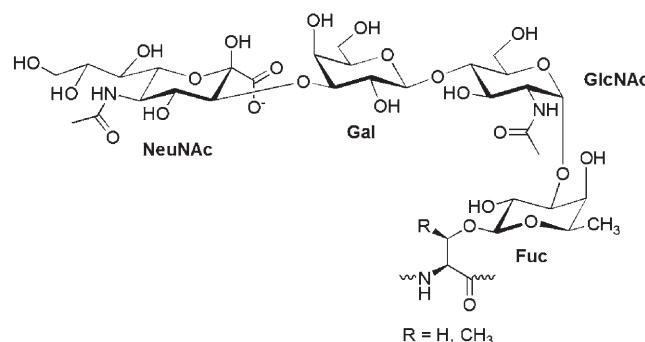
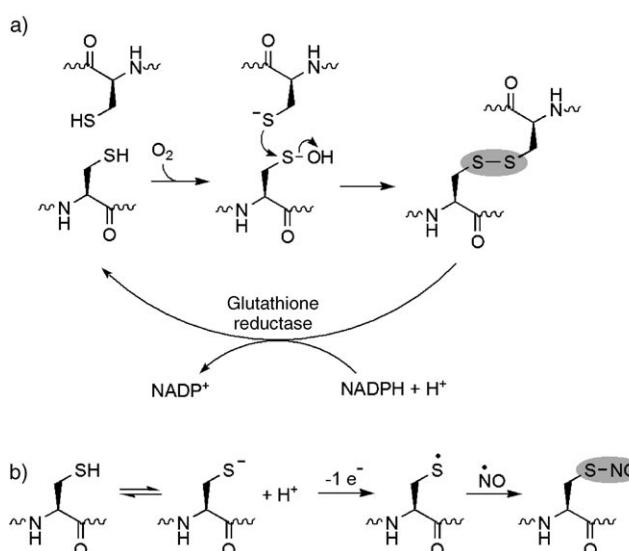


Figure 9. The O-linked tetrasaccharide sialyl- α -2,3-Gal-x-1,4-GlcNAc- β -1,3-fucosyl-Ser attached to the protein Notch during its transit through the secretory pathway on the way to the cell surface.

part of the maturation of Notch during its transit through the secretory pathway to the cell surface. The short O-linked sugar chains are important in a variety of functional contexts, from modulating transcription factor activity,^[49] to acting as essential recognition elements in signaling by Notch at cell surfaces.^[50,51]

2.5. S-S Bond Formation

Two main types of linkages serve to cross-link proteins, or portions of proteins, covalently. By far the more common are disulfide links from oxidation of cysteinyl residue thiolate side chains.^[52,53] The cytoplasmic and nuclear compartments in eukaryotic cells are reducing microenvironments, as reflected in the 100:1 ratio of the redox-active tripeptide glutathione in reduced (GSH) to oxidized (GSSG) state.^[54] The high reducing ratio is maintained by the high levels of NAD(P)H and enzymes, such as glutathione reductase and thioredoxin reductase^[5] that use the reduction potential of NAD(P)H to re-reduce disulfides in proteins that have become oxidized (Scheme 14a). As proteins pass through the secretory pathway in eukaryotic cells, the levels of total glutathione and reduced nicotinamide coenzyme fall, compartments become

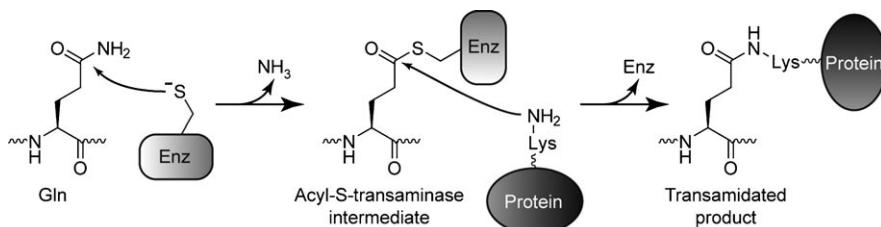


Scheme 14. Oxidation of thiolate side chains of cysteine residues: a) oxidation of dithiols to disulfides (for example via sulfenic acid intermediates) and reversible reduction back to dithiols by glutathione reductase action; b) oxidation of Cys-S⁻ side chain to S-nitrosyl-Cys by nitric oxide (·NO).

more oxidizing, and disulfide links predominate. The disulfide bonds may stabilize protein architectures as proteins reach cell outer surfaces or are excreted into extracellular spaces. The formation of disulfide bonds in zymogen forms of pancreatic protease as they are packaged into the oxidizing microenvironment of zymogen granules is a prototypical case.

The mechanism for oxidation of protein dithiols to disulfides typically involves oxidation of the electron-rich thiolate side chains of Cys residues. One-electron oxidation would yield thiyl radicals that could dimerize to the disulfides. Alternatively, thiolate side chains can be oxygenated by a variety of oxygen-derived oxidants (peroxide, hydroxyl radical) to generate sulfenic acid (-SOH) side chains. Capture of the sulfenate by a neighboring Cys-S⁻ generates disulfides. Regeneration of the dithiol forms is mediated by thiol-disulfide interchange using reduced glutathione or the low-molecular-weight dithiol protein thioredoxin (TSH). The oxidized GSSG or TSST are recycled at the expense of NADPH oxidation by thioredoxin reductase and glutathione reductases.

The electron-rich thiolates and the thiyl radicals can be captured by other oxidants and radicals, including ·NO. Such S-nitrosylation of thiolate side chains of cysteine residues is



Scheme 15. Nonoxidative cross-links introduced by transglutaminases; the amide bond in Glu side chains is replaced by the ε-NH of Lys to create Glu-ε-Lys isopeptide cross-links.

documented in many proteins by radical species derived from nitric oxide (Scheme 14b). These Cys-SNO moieties have been proposed to be widespread in oxidative signaling events.^[55]

The second PTM cross-link strategy is nonoxidative and involves transglutaminase catalysis in which glutamine side chains in protein substrates are deaminated via acyl-S-transaminase intermediates, which are captured by Lys-ε-NH₂ groups to effect net transamidations (Scheme 15).^[56,57]

3. Covalent Addition: The Supporting Cast

While the five chemical types of posttranslational modifications discussed in Section 2 are abundant and well-characterized in cells, there are many additional classes of purposeful enzymatic modification of proteins that expand the metabolic and signaling capacities of organisms.

3.1. Protein Hydroxylation

One additional category of protein posttranslational modification is enzyme-mediated hydroxylation. Hydroxylations occur at nonnucleophilic sites in aminoacyl side chains to generate 3-OH-Pro, 4-OH-Pro, and 5-OH-Lys (Figure 10) in

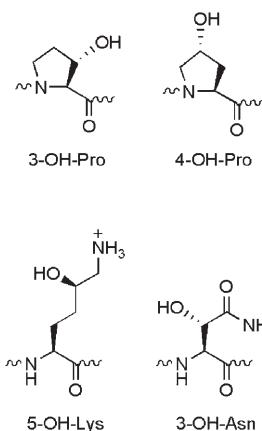


Figure 10. Hydroxylated amino acid residues generated by posttranslational Fe²⁺-dependent monooxygenases: 3-OH-Pro, 4-OH-Pro, 5-OH-Lys, 3-OH-Asn.

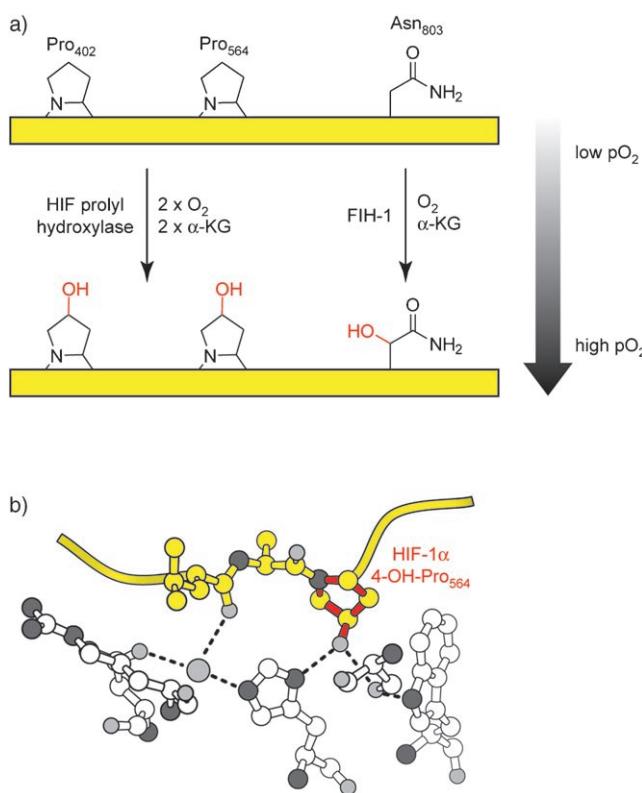
collagen at Pro-Gly and Lys-Gly sites. These hydroxylations are key modifications for proper maturation of collagen fibers.^[58] Some of the 5-OH-Lys residues are then subsequently tandemly glycosylated on the newly introduced OH group to create an O-disaccharide linkage.

4-Hydroxyproline modifications, about tenfold more abundant than the hydroxylations at C3 of Pro residues, are involved in the triple helical strands of collagen, with the 4-OH pointing away from the helix.

A third side chain in which a CH₂ group is converted into CH-OH is

Asn (to 3-OH-Asn, Figure 10) in a small number of proteins, including the transcription factor HIF (hypoxia inducible factor).^[59] HIF transcription is initiated at low partial pressures of O₂ and induces the transcription of hundreds of genes, including the gene that encodes erythropoietin to make more red cells that can carry more O₂ to hypoxic tissues.

The HIF-1 α subunit in the HIF $\alpha\beta$ heterodimer is post-translationally hydroxylated by hydroxylases, one specifically acting at two Pro residues to create 4-OH-Pro^[60] and the other acting at a particular Asn residue to generate 3-OH-Asn.^[61] These are determinative changes in the oxygen-sensing cascade in mammalian tissues (Scheme 16).^[62] The lifetime of the HIF-1 α subunit in cells is controlled by proteolysis in

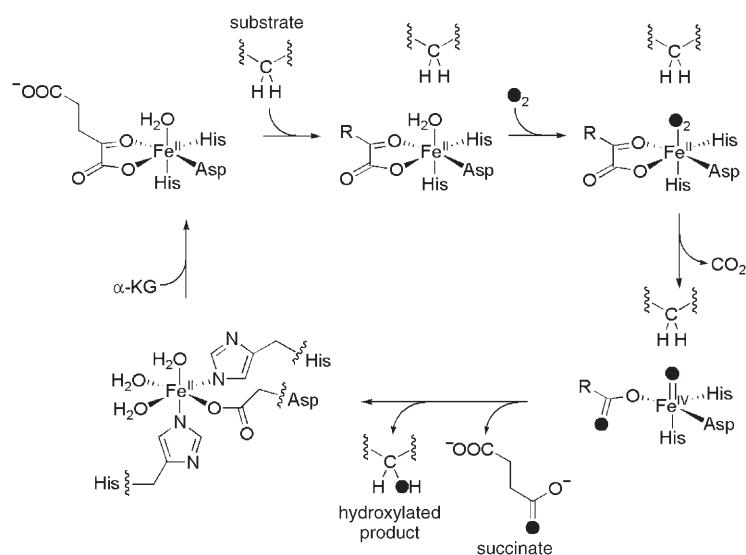


Scheme 16. a) Hydroxylation of Pro and Asn residues in the HIF-1 α subunit; b) interaction of the HO-Pro₅₆₄ residue of HIF with the E3 ligase that will catalyze polyubiquitylation of HIF.

the ubiquitylation pathway, which involves chaperoned escort to the proteasome, unfolding, and threading into the proteasome chamber for proteolysis to limit peptides. The polyubiquitylation of HIF-1 α is carried out by a particular E3 ligase, the von Hippel-Lindau (VHL) protein.^[63,64] The hydroxylation status of Pro₄₀₂ and Pro₅₆₄ in HIF-1 α controls the affinity for the VHL E3 ligase.^[60] At low pO₂ in cells the Pro hydroxylase is not saturated with its substrate O₂ and has low activity. At high pO₂ the hydroxylase is active and converts the two Pro residues into 4-OH-Pro residues. The hydroxy side chain in HO-Pro₅₆₄ provides about a 1000-fold

tighter binding of modified HIF-1 α over unmodified HIF-1 α for the VHL ubiquitin ligase.^[60] This is the mechanism for selective polyubiquitylation of the hydroxylated forms of HIF-1 α , which leads to its proteasome-mediated destruction at high pO₂ levels but its persistence at low pO₂. Persistence means a longer lifetime for the heterodimeric HIF-1 $\alpha\beta$ and the longer gene-transcriptional-activation response characteristic of hypoxia.

The protein hydroxylases that catalyze the side-chain hydroxylations noted in this section belong to the family of non-heme Fe^{II} monooxygenases that have two His and one Asp side chains to provide three of the six coordination sites to the Fe^{II} (Scheme 17).^[65] Two additional coordination sites



Scheme 17. Mechanism of protein hydroxylation by the nonheme Fe^{II} monooxygenases. The active site iron(II) is coordinated by three residues from within the protein (two His, one Asp). Subsequent coordination by $\alpha\text{-ketoglutarate}$ and O_2 results in the cleavage of dioxygen and formation of a high-valent $\text{Fe}^{\text{IV}}=\text{O}$ species. This iron complex cleaves the unactivated C-H bond of the substrate, yielding hydroxylated product and succinate.

are filled by cosubstrate $\alpha\text{-ketoglutarate}$ and the sixth by O_2 . When both O_2 and $\alpha\text{-KG}$ are bound, the organic diacid is oxidatively decarboxylated to succinate. O_2 is cleaved in such a way that one atom ends up in succinate and the other is coordinated to the iron as a high-valent $\text{Fe}^{\text{IV}}=\text{O}$. This high-valent oxoiron complex is a sufficiently powerful oxidant to cleave the unactivated C-H bonds at C3 and C4 of Pro residues, C5 of Lys residues, and C3 of Asn residues to generate transient 'CH radicals and $\text{Fe}^{\text{III}}\text{-OH}$. OH[·] transfer from the $\text{Fe}^{\text{III}}\text{-OH}$ to the carbon-centered radical yields the hydroxylated protein side chains. The polarity of these side-chain hydroxylations is quite distinct from the bulk of the other posttranslational modifications considered in this Review. The amino acid side chains undergoing modification are not electron rich or nucleophilic. Instead, the iron-based modifying enzyme generates a powerful oxidant and leads to homolytic cleavage of unactivated C-H bonds with regio- and stereospecificity.

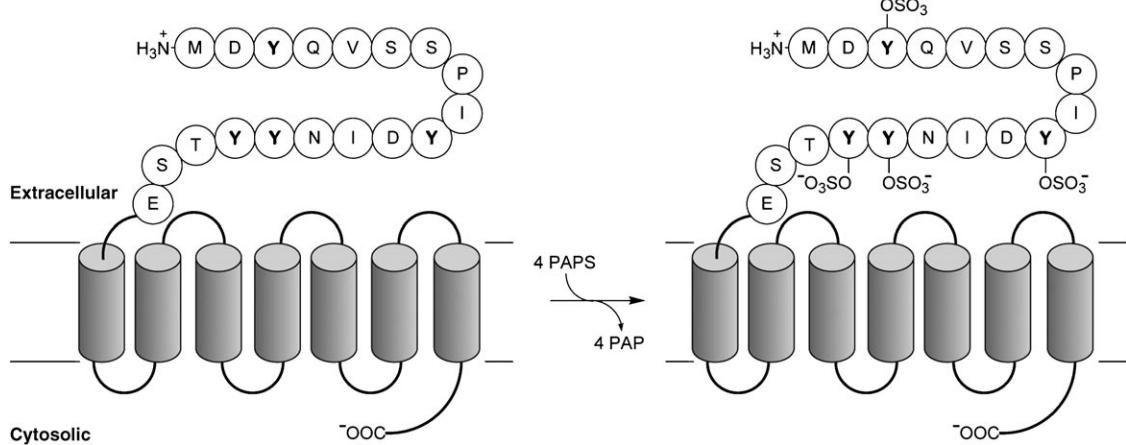
3.2. Protein Sulfur Transfer

Phosphoryl groups are not the only inorganic moieties transferred to protein side chains. The SO_3^- group is also transferred from phosphoadenosine phosphosulfate (PAPS) to tyrosine side chains in proteins.^[66] PAPS is the biological reagent for transfer of activated sulfuryl (SO_3^-) groups to nucleophiles in both small molecules (e.g. oligosaccharides such as heparin^[67]) as well as proteins and proteoglycans. Enzymatic sulfation of four tyrosine residues occur in the Golgi complex at the N-terminus of the CCR5 receptor (Scheme 18) during its transit to the plasma membrane^[68] where the N-terminus is displayed at the extracellular surface. This cluster of anionic Tyr- OSO_3^- modifications is important for recognition by the CCR5 chemokine ligand.

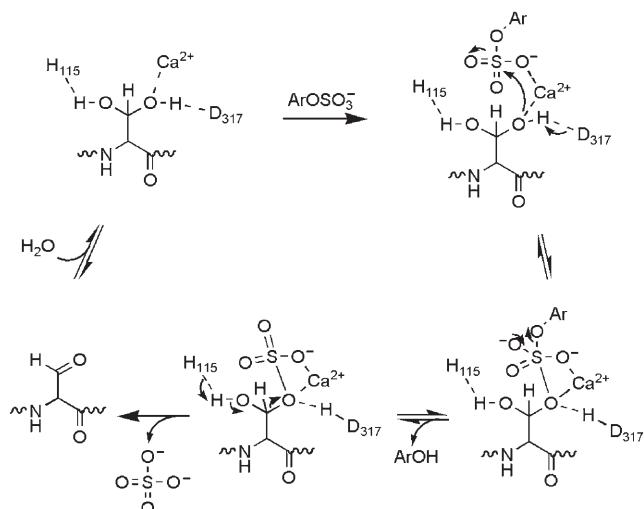
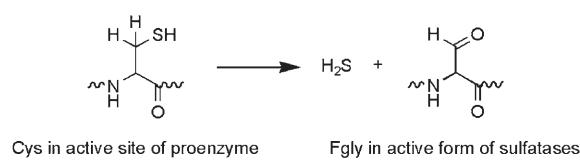
Once sulfated molecules are internalized, the sulfate ester bond is hydrolyzed enzymatically by sulfatases in the secretory compartment, predominantly in lysosomes.^[69–71] The set of sulfatases known to degrade aryl sulfate ester substrates are themselves in inactive proenzyme forms until they undergo posttranslational activation. This involves oxidative conversion of an active-site cysteine residue into an aldehyde in the form of formylglycine (Fgly).^[70] This conversion of a thiolate nucleophile into an electrophilic carbonyl is followed by a hydration equilibrium to distribute the Fgly between the aldehyde and the *gem* diol, the aldehyde hydrate (Scheme 19). It is the hydrated form of the aldehyde group in the formylglycine side chain that initiates covalent attack on sulfated protein substrates bound in the sulfatase active sites^[72] with resultant O- SO_3^- bond cleavage.

3.3. Protein Modification by Bacterial Toxins

Bacteria that invade eukaryotic cells secrete a complement of proteins into the host cell to neutralize its defense mechanisms. Among these virulence factors are three types of enzymes that act as posttranslational modification catalysts for ADP ribosylation, glucosylation, and deamidation of host target proteins.^[3]



Scheme 18. Transfer of four SO_3^- groups from PAPS to the phenolate oxygen atoms of four side chains of Tyr residues at the N-terminal region of the CCR5 receptor during its passage through the secretory compartments to the cell surface.



Scheme 19. Oxidative conversion of an active site Cys-S- to the aldehyde of formylGly (Fgly) converts inactive precursors of sulfatases to active catalysts. Hydration of the Fgly side chain generates a *gem* diol, which is the active nucleophile that attacks the $\text{Ar}-\text{OSO}_3^-$ substrate on sulfur to initiate the O- SO_3^- bond cleavage.

3.3.1. ADP Ribosylation

Several famous bacterial exotoxins, including the cholera toxin, the diphtheria toxin, the pertussis toxin, and the botulinum C3 toxin,^[73] are ADP ribosyltransferases. The donor substrate of the ADP ribosyl moiety is the readily available coenzyme NAD. The positively charged nicotinamide group departs in a transition state involving the ribaoxacarbenium ion of the transferring ADP ribosyl group (Scheme 20a). This ion can be captured by robust nucleophiles in protein cosubstrates in the active site of the toxin, such as thiolate side chains of Cys, for example when pertussis toxin modifies the α subunit of the inhibitory GTPase G_i that regulates cyclic AMP production. The transferring ribaoxacarbenium ion can also be captured by weak nucleophiles such as the guanidino group of Arg in the α subunits of the G_s GTPase by the cholera toxin. Another

example of the capture of the potent electrophilic form of the ADP ribosyl moiety is the modification of the weakly nucleophilic Asn₄₁ in the Rho subfamily of small GTPases by the C3 toxin from *Clostridium botulinum*, ultimately leading to a net depolymerization of the actin mesh work in the host cell.

A fourth example of bacterial toxin ADP ribosyltransferase activity on a specific host protein with deleterious consequence is the action of the diphtheria toxin on His₇₁₅ in the protein synthesis factor eukaryotic elongation factor 2 (eEF-2). That histidine in mature eEF-2 has already undergone a preparatory set of posttranslational modifications involving the aminocarboxypropyl transfer from cosubstrate S-adenosylmethionine (SAM), N,N,N-trimethylation by three additional SAM molecules, and glutamine-mediated amidation to convert His₇₁₅ into a diphthamide residue. This is the molecular form of eEF-2 that undergoes ADP ribosylation by the diphtheria toxin (Scheme 20b) and rendered inactive in its essential elongation functions, bringing host protein synthesis to a halt in the infected cell.^[73]

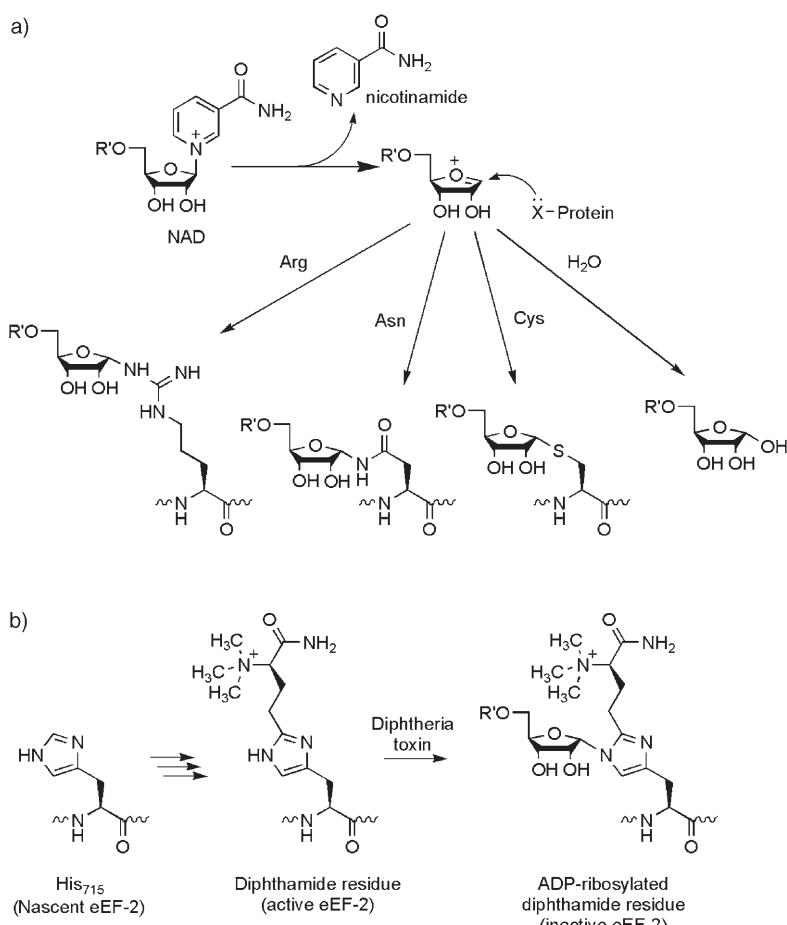
3.3.2. Other Modification Activities of Bacterial Toxins

The Ras and Rho GTPase families are also targets of additional bacterial toxin enzymes that inactivate the GTPases by chemical modifications distinct from ADP ribosylation. One is the lethal toxin protein from *Clostridium sordelli* which also is a protein glycosyltransferase. In this case it is not transfer of the ADP ribosyl moiety from NAD but instead the transfer of a glucosyl group from UDP-glucose that is catalyzed by the protein toxin. The target nucleophile is the β -OH of Thr₃₅ in Ras. This O-glycosylation blocks the catalytic activity of Ras.^[74,75]

A distinct type of posttranslational modification strategy is deamidation of Gln₆₁ in Rho by the cytotoxic necrotizing protein from pathogenic strains of *E. coli*.^[76] This Gln carboxamide side chain is in the GTPase active site and its hydrolysis to the γ -COO⁻ of Glu₆₁ disrupts the active-site machinery. Altogether, the small GTPase families, because they act as thermodynamic switches at so many cellular intersections of signaling and metabolism, are subjected to a diverse set of programmed chemical modifications. The covalent modifications alter localization and control function both in normal maturation and in pathogen interceptions.

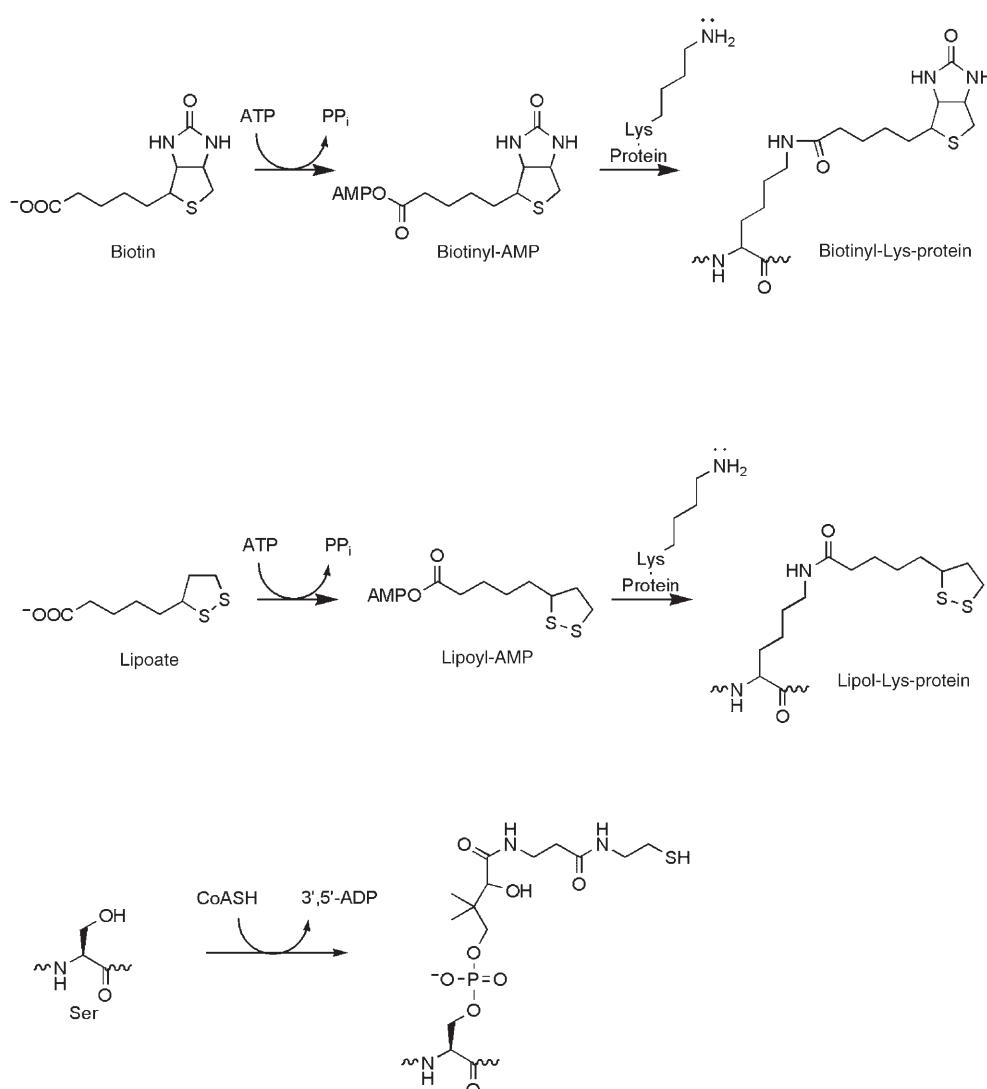
3.4. Installation of Swinging-Arm Cofactors

Several enzymes central to primary metabolism are nonfunctional until posttransla-



Scheme 20. ADP ribosylation of protein substrates by bacterial toxins: a) Cleavage of NAD releases nicotinamide and generates a stabilized ribaoxacarbenium ion in the active site. This can be captured by a range of cosubstrate nucleophiles, from the robust Cys-S⁻ in pertussis toxin catalysis to the weak Asn and Arg side chains in botulinum and cholera toxins. Water is a natural nucleophile and leads to NAD glycohydrolase side reactions; b) diphtheria toxin catalyzes ADP ribosylation of a modified histidine residue in the eukaryotic host cell protein synthesis elongation factor eEF-2. The residue, His₇₁₅ in the nascent eEF-2 undergoes five posttranslational steps of transfer of the methionyl moiety from SAM, N,N,N-trimethylation, and amidation to yield the diphthamide residue at position 715. This is the residue targeted for ADP ribosylation on N₃ of the imidazole ring by the diphtheria toxin.

ally primed with prosthetic groups that provide key functional groups to enable acyl- and carboxyl-transfer chemistry. The acyl-transfer coenzymes are lipoic acid and phosphopantetheine. The carboxyl-transfer prosthetic group is biotin (Scheme 21). All three cofactors are covalently attached to side chains of target apo proteins by dedicated activating/loading enzymes. Biotin and lipoate can be activated as the acyl AMP species and captured by the ϵ -NH₂ of lysine side chains presented in folded 100-residue domains of the target proteins to create the biotinylamide and lipoamide linkages. The full reach of the Lys biotin and Lys lipoate chains is about 20 Å, leading to the historical connotation as swinging-arm prosthetic groups that can visit different domains in multi-enzyme complexes.^[77] In analogous logic the phosphopantetheinyl moiety of CoASH as donor substrate can be captured by the β -OH side chain of a specific Ser side chain in a 80–100-residue acyl-carrier protein domain, again to create a tethered prosthetic group on a 20-Å pivot. In this case the chemical bond is not an amide linkage but a phosphodiester.



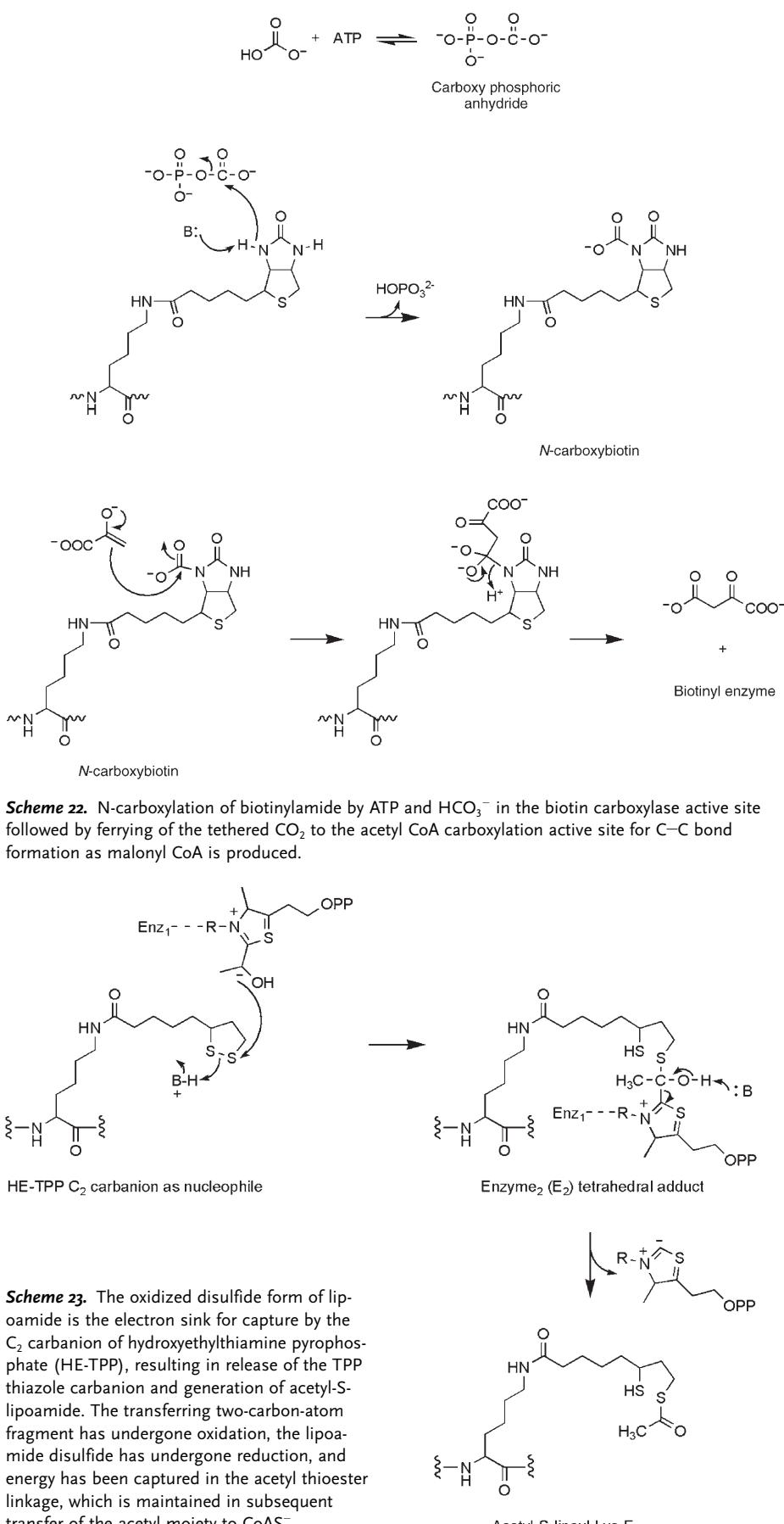
Scheme 21. Coenzymes that are tethered in the active sites of enzymes to act as swinging-arm prosthetic groups, carrying CO₂ or acyl groups between active sites: biotin and the resultant biotinylamide-Lys; lipoate and the resultant lipoamide-Lys; CoASH and the resultant pantetheinyl-OPO₃-Ser linkages.

The tethered biotinyl cofactor is used for the transfer of C₁ groups in the form of CO₂. In carboxylases that fix HCO₃[−] into acetyl CoA and propionyl CoA, to yield malonyl and methylmalonyl CoA products, there are multiple subsites or distinct subunits with different chemical functions. In the biotin carboxylase subunit, ATP is used as cosubstrate with HCO₃[−] to generate a transient carboxyphosphate mixed anhydride that is captured by the biotinyl-Lys on the N₁ ureido nitrogen atom to produce N₁-carboxybiotinyl-Lys-enzyme (Scheme 22). This form of fixed CO₂ is shuttled to the active site where the C₂ carbanion on the acetyl moiety of acetyl CoA is generated and used to attack the N-carboxybiotinyl tether. This action leads to C–C bond formation as acetyl CoA is carboxylated to malonyl CoA, one of the key building blocks for fatty acid biosynthesis in cells.

The other two tethered coenzymes, lipoamide and pantetheinyl-phosphate, are used to ferry substrate-derived acyl groups between active sites of multidomain enzymatic assembly lines.^[77,78]

The lipoamide prosthetic group is found in all α -keto acid dehydrogenase complexes that carry out oxidative decarboxylation, for example, of pyruvate at the end of glycolysis and of α -ketoglutarate in the citric acid cycle. For example, in pyruvate dehydrogenation the disulfide form of the lipoamide cofactor is the electron sink for ring-opening capture by the C₂ carbanion of hydroxethylthiamine-PP (Scheme 23). The transferring two-carbon fragment has been oxidized from the acetaldehyde to the acetate oxidation state and captured as an activated acetyl-S-lipoamide thioester as the disulfide link of oxidized lipoamide is reduced. This is the prototypic redox/energy-capture role for the lipoamide prosthetic group. In the second half-reaction, the acetyl-S-lipoamide arm moves to a separate active site and docks next to a CoASH molecule, thus allowing acetyl transfer to the thiolate of CoAS[−], an isoenergetic transfer that now releases the oxidized acetyl moiety as the diffusible cellular energy currency acetyl CoA.

A third example of cofactor tethering, the pantetheinyl moiety on acyl-carrier protein domains (ACPs), also offers a terminal nucleophilic thiolate for capture of acyl groups. Its central role in primary metabolism is in fatty acid biosynthesis in which acyl chains are built up by two carbon atoms at a time by cycles of Claisen condensation followed by redox tailoring to convert

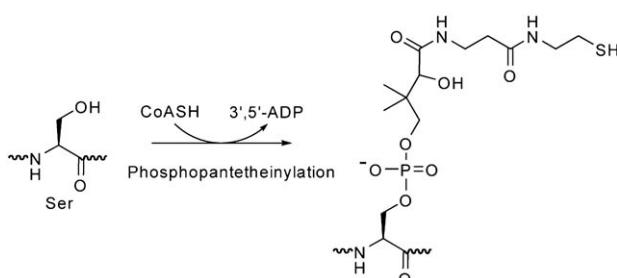


elongated beta keto acyl-S-ACP intermediate into the beta methylene acyl-S-ACPs for the next round of C_2 -unit elongation (Scheme 24).^[78] The acyl groups are proposed to visit ketosynthase, ketoreductase, dehydratase, and enoyl reductase active sites sequentially while tethered to the flexible phosphopantetheinyl arm.

It would appear that Nature has invented the coenzyme covalent tethering strategy in these three cases to fix small-molecule acyl fragments or CO_2 and ferry them to distinct sites on multi-enzyme complexes and assembly lines. There are other examples in which cofactors are tethered to lysine side chains of proteins, through hydrolyzable imine links between the Lys-NH₂ and an aldehyde carbon atom in the cofactor. This is true for the aldehyde form of vitamin B6, pyridoxal phosphate, in all pyridoxal phosphate dependent enzymes, and in the visual pigment proteins, the rhodopsins, in which vitamin A aldehyde (retinal) is the aldehydic chromophore.

3.5. Posttranslational Carboxylation of Glutamyl Residues for Bidentate Calcium Binding

A family of proteins involved in blood coagulation in mammals undergoes posttranslational modifications at sets of closely spaced glutamate residues during passage through secretory compartments on their way to the extracellular space. These modifications involve fixation of CO_2 to the γ -methylene carbon atoms of Glu residues, thus creating malonyl-type side chains, known as γ -carboxy Glu (Gla). The Gla side chains provide the possibility for bidentate chelation of divalent cations, of which interaction with Ca^{2+} ions is most significant. The proteins that undergo tandem γ -Glu carboxylation include the proenzyme forms of proteases such as prothrombin, proFac-



Scheme 24. The terminal thiolate of the phosphopantetheinyl prosthetic group is the nucleophile that captures the starting acyl fragment in fatty acid biosynthesis and serves as the platform on which the acyl chain is elongated by two carbon atoms in each subsequent cycle. The pantetheinyl arm carries the growing acyl chain between the active sites of ketosynthase, ketoreductase, dehydratase, and enoyl reductase in each cycle, converting the β -ketone into the four-electron-reduced β -methylene oxidation state.

tor IX, and proFactor X.^[79] The carboxylation of 10–12 glutamate side chains to γ -carboxy-Glu residues in a 40-residue stretch of the proenzyme forms of proteases creates local high density of bidentate chelators for Ca^{2+} ions (Figure 11). The conformations of the Gla domains are altered in the presence of Ca^{2+} and drive the association of the proteases on platelet surfaces, leading to the formation of protein complexes and the activation of neighboring proteases to initiate and propagate blood coagulation cascades.^[79]

Unlike the carboxylation reactions discussed in Section 3.4, which use tethered biotin as the CO_2 carrier between the active sites of the enzymes, the Gla-forming carboxylations of glutamyl residue side chains do not use biotinyl amide dependent enzymes. Instead, the naphthoquinone vitamin K, in its only well-defined role in mammalian metabolism, is the requisite cofactor for CO_2 fixation in the Gla modifications. In fact it is the dihydro-naphthoquinol form of vitamin K, that is the active form for the vitamin K dependent protein carboxylase. O_2 is also a required cosubstrate and the mechanism of the carboxylase (Scheme 25) is proposed to involve formation of the K-OOH peroxy adduct. Cyclization of this quinone hydroperoxide to the alkoxide anion of the 2,3-epoxide of vitamin K generates the strong base required to abstract a proton from the glutamyl- γ CH_2 side chains. The carbanion generated is required to attack CO_2 and form the new C–C bond of the malonyl side chains of Gla residues in the product.

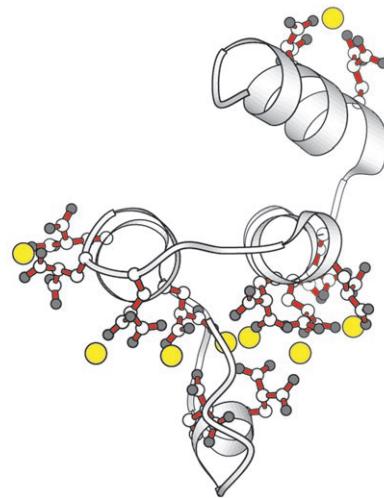
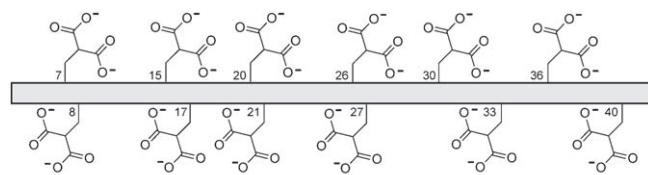
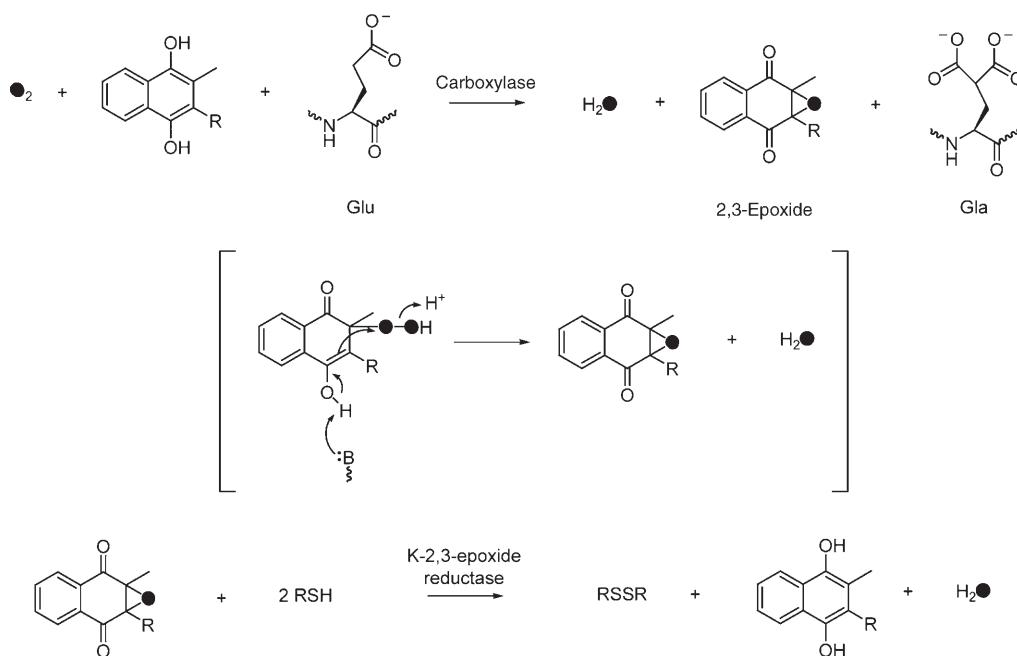


Figure 11. Twelve closely spaced Glu residues in the carboxylation domain of the pro form of the coagulation protease factor IX are modified to γ -carboxy-Glu (Gla) residues and provide a high local concentration of bidentate malonyl side chains for Ca^{2+} coordination. The 3D fold of dodecaGla region complexed to eight calcium ions (yellow spheres) emphasizes the divalent metal-ion-dependent structuring of this region of the protein.

4. Cataloguing the Posttranslational Modification

Consideration of major and minor categories of post-translational modifications above leads to at least hundreds of thousands, perhaps millions, of possible molecular variants of proteins in eukaryotic cells. The analytical problem is immense, and the challenge to integrate across the subproteomes and decipher connections for a systems biology perspective even greater. Much of the contemporary effort in PTM research involves the development of methodologies to evaluate such inventories. Mass-spectrometric approaches are dominant owing to issues such as femtomolar sensitivity of detection and simultaneous identification of many peptide fragments bearing a particular type of chemical modification.^[80] Notable recent studies include the detection of hundreds of pS and pT peptides from the phosphoproteome of yeast^[81] and the isolation of His-tagged ubiquitylated proteins of yeast. The latter method allows the identification of more than 100 ubiquitylated proteins that contain polyubiquityl chains connected through different Lys residues in the ubiquitin monomers.^[82]

At any moment in time, sampling of the proteome in a given organism or cell provides only a snapshot of a highly dynamic process, confounding the analytical problem and ultimately arguing for time-resolved inventories. Heterogeneity can arise in several ways. Because posttranslational modification enzymes do not work off templates, the mod-



Scheme 25. Proposed mechanism for the function of the dihydro form of vitamin K (KH_2) as cosubstrate in the Glu to Gla posttranslational protein carboxylations. KH_2 reacts with cosubstrate O_2 to produce a hydroperoxy vitamin K intermediate, which can proceed to the 2,3-epoxyvitamin K alkoxide. The alkoxide is argued to be a strong enough base to abstract one of the Glu $\gamma\text{-CH}_2$ hydrogen atoms as a proton, transiently generating the γ -carbanion required to attack CO_2 to produce the new C–C bond in the Gla product residue.

ifications are stochastic and likely to be incomplete across a target protein population. The fractional efficiency of modification of a site, for example, N-glycosylation of a particular Asn residue in an Asn-X-Ser sequence favorable for modification, may depend on the amount of time the protein region bearing that Asn is unfolded during passage across the ER membrane. The accessibility of the carboxamido nitrogen atom may be fleeting. The concentration of oligosaccharyl-N-transferase in the microenvironment may be limiting such that the dwell time of protein substrate and enzyme catalyst is too short to ensure complete modification at 100 % of that site in the protein population. In the multistep maturation of N-glycan $\text{Man}_3(\text{GlcNAc})_2$ core in glycoproteins there are up to a dozen such substrate–catalyst encounters. A 90 % yield at each step would produce the kind of complex and heterogeneous glycoprotein mixtures that are often detected. As noted earlier, the prion protein has been shown to exist in some 52 glycoforms.^[45]

5. Multiple and Tandem Posttranslational Modification of Proteins

Posttranslational modifications often occur at multiple sites or in tandem cascades that are crucial for function. Thus, the Abl tyrosine kinase is found phosphorylated at 11 different sites, (nine tyrosines, one serine, one threonine) spread over the different catalytic and regulatory domains of the protein^[83] (see Figure 2). In principle, $11! = 40\,420\,800$ distinct phosphorylated isoforms are possible just for this one protein alone. Fractional occupancy is likely at

the distinct sites, making for a large nested array of different phospho forms of this one protein.

Multiple lysine acetylations and methylations, one serine phosphorylation, and one N-terminal ubiquitylation are typical for the tails of the histone octamers in nucleosomes so that modifications at 28 sites are possible as described in Sections 2.2 and 2.3. (Surely not all of the $28!$ possibilities will have been explored in Nature.) Kelleher and co-workers^[84] recently devised bioinformatic methods coupled with high-resolution mass spectrometry to predict, search for, and identify particular histone variants. For example, they detected one particular hexamodified form of the N-terminus of histone H3 with an $[M+238]$ mass signature (Figure 12) to address such a functional covalent diversity of protein isoforms. (This is the equivalent of finding one needle in a haystack. The systems approach to PTMs would be to assess how many such needles altogether are in the “haystack” of proteins.) The multiple modifications of histone tails are

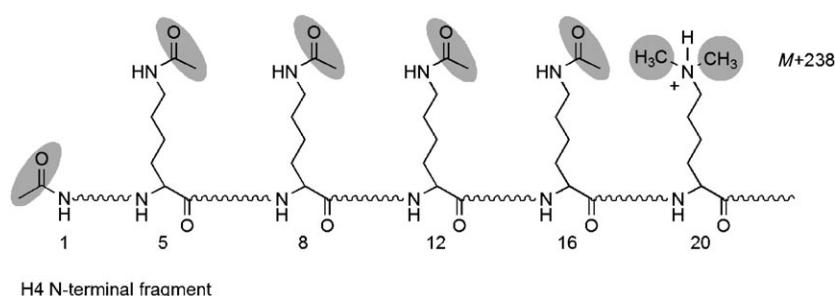
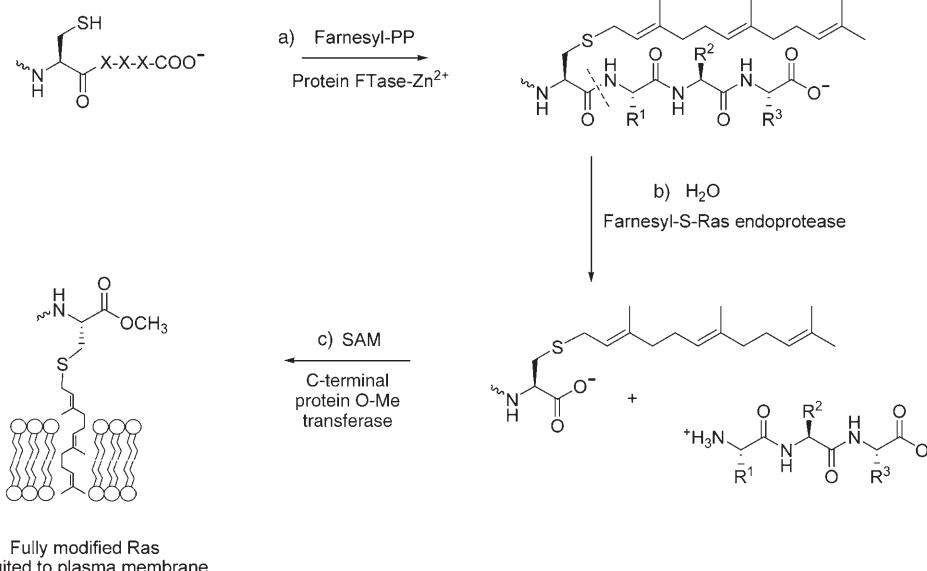


Figure 12. Tandem enzymatic posttranslational tailoring of the N-terminal tail of histone H4 generates a heptamodified protein in which the N-terminal residue is acetylated, lysines_{5,8,12,16} are acetylated, and lysine₂₀ is N,N-dimethylated.

presumed to be written and read by enzymatic machinery in specific temporal patterns for selective recruitment of transcriptional repressors and activator complexes.

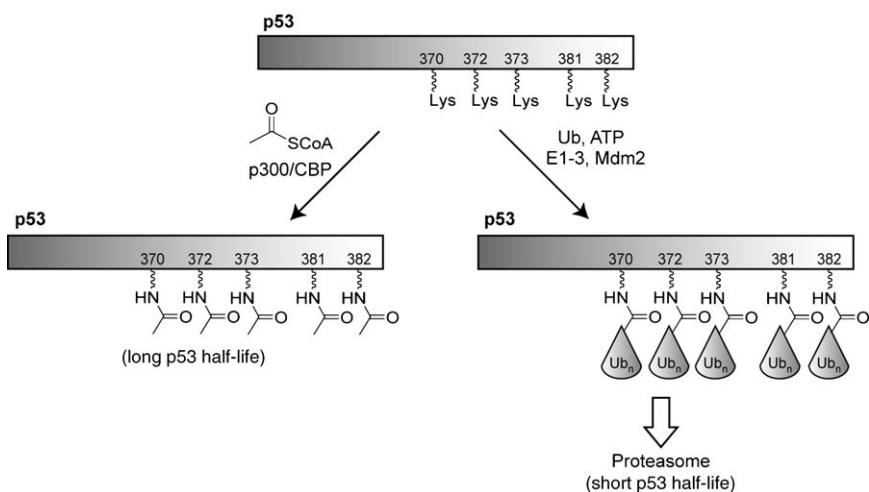
Tandem posttranslational modifications are known in many other proteins and comprise the multilayered informational content driving the molecular logic of posttranslational modifications. The consecutive four-step modification of the C-termini of Ras proteins—1) S-prenylation, 2) S-palmitoylation of neighboring cysteine residues, 3) specific endoproteolytic cleavage to reveal one of the cysteines as the new C-terminus, and 4) methylation of that new C-terminal carboxylate—comprise the integrated maturation process that moves modified Ras to membranes to dock with its upstream protein kinase partners (Scheme 26).^[35] We have noted above in Section 3.3 that the related RhoA family of GTPases can undergo ADP-ribosylation, O-glucosylation, and deamidation of the active-site Gln. Adding in this sequence of prenylation, proteolysis, and C-terminal carboxymethylation for Rho brings the total to seven specific steps of posttranslational modification. The mature RhoA can also be cleaved by the YopT protease from *Yersinia pestis* as that pathogen executes part of its virulence program.^[85]

The threshold nature of tandem posttranslational modifications to provoke a biological signal is illustrated clearly in the successive addition of a minimum of four ubiquitin 8-kDa tags to a target protein to set off the cascade of events leading to proteolysis. A polyubiquitin moiety of length at least four ubiquitin units appears to be the threshold for recognition by protein chaperones to send the tagged proteins to proteasomes.^[27]



Scheme 26. Multistep modification of the Ras GTPase involves a) S-prenylation, b) endoproteolysis, and c) C-terminal O-methylation.

Finally, there is good evidence for competition between posttranslational modifications, with opposing functional consequences for the target proteins. Two such examples involving competition between ubiquitylation and acetylation are 1) SMAD7 protein in TGF β signal transduction pathway^[86] and 2) five lysine side chains near the C-terminus of the transcription factor p53. The Lys ϵ -NH₂ residues can be acetylated or ubiquitylated (Scheme 27) and then extended to polyubiquitin chains, leading to proteolytic removal of p53 or SMADs. The acetylations block ubiquitylations and consequently lengthen the lifetime of the proteins.^[13]



Scheme 27. Competition between acetylation and ubiquitylation at the ϵ -NH of Lys_{370,372,373,381,382} near the C terminus of the transcription factor p53.

6. Reversible versus Irreversible Posttranslational Modification

Depending on the biological purpose of a particular covalent modification of a protein, reversibility may or may not be an important parameter to control. The prototype of reversible modification is protein phosphorylation, consistent with its evolution to the dominant role in protein-based signaling in eukaryotes. Of the five major categories of PTMs noted in Section 2 (phosphorylation, acylation, glycosylation, thiol-disulfide chemistry, and alkylation), all but alkylation have dedicated enzymes, often large enzyme families, that catalyze the removal of the covalent modifications. The enzymes that reverse phosphorylation, acylation, and glycosylation are, by and large, specific hydrolases, whereas disulfide bonds are cleaved by reductases. We note below that an oxidative enzymatic route has now been discovered for alkylation that involves removal of N-methyl substituents from the histone tails.

Protein kinases usually occur as low-activity “off” forms in basal states in the absence of a specific stimulus. When a signal is propagated and the kinases activated, often involving covalent autophosphorylation of Thr residues to pT or TXY sequons to doubly phosphorylated pTxpY loci in MAP kinases, the activated protein kinases modify their partner target proteins with much higher catalytic efficiency.^[8] The cyclic AMP activated protein kinase A, for example, phosphorylates over 100 proteins^[87] on Ser and Thr residues to propagate signals.

To control the duration and intensity of phosphoprotein signaling, the signals need to be turned off. Termination is accomplished by hydrolytic removal of the PO_3^{2-} group by phosphoprotein phosphatases (Scheme 28a). There are about 150 protein phosphatases encoded in the human genome,

including pS-, pT-selective phosphatases, pY phosphatases, and dual-specificity (pS and pY) phosphatases.^[88] Of these, 107 are pY protein phosphatases.^[89] The much smaller number of pS/pT phosphatases is balanced by the presence of many regulatory subunits that control subcellular location and substrate recognition for the pS/pT hydrolytic enzymes.

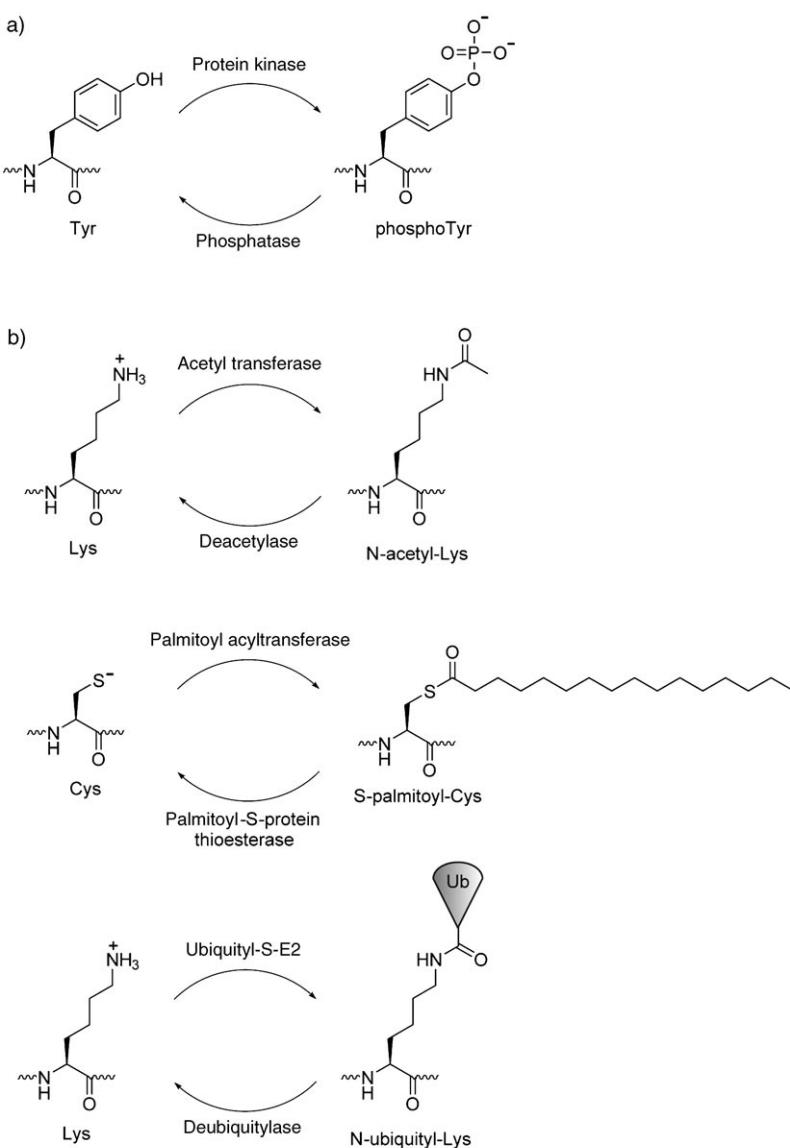
One practical consequence of PTM enzyme-mediated reversibility is that the phosphoproteome content in any cell at any given time represents the balance of activity of the 500 protein kinases/150 protein phosphatases towards their diverse protein substrates, allowing an almost infinite number of set points to a cell and true complexity for phosphoproteomicists.^[90,91]

Protein acetylation, protein ubiquitylation, and protein-S-palmitoylation are three other classes of covalent modifications that are readily reversed (Scheme 28b), and so qualify for regulation and signaling roles. Histone acetyltransferases (HATs) are opposed functionally by a family of histone deacetylases (HDACs).^[92,93] Some HDACs are catalytic domains embedded in multimodular proteins dynamically recruited to acetylated and methylated lysine tails of histones. A separate family of NAD-cleaving histone deacetylases, the sirtuins, are involved in gene silencing functions and couple energy metabolism with transcriptional regulation.^[94,95]

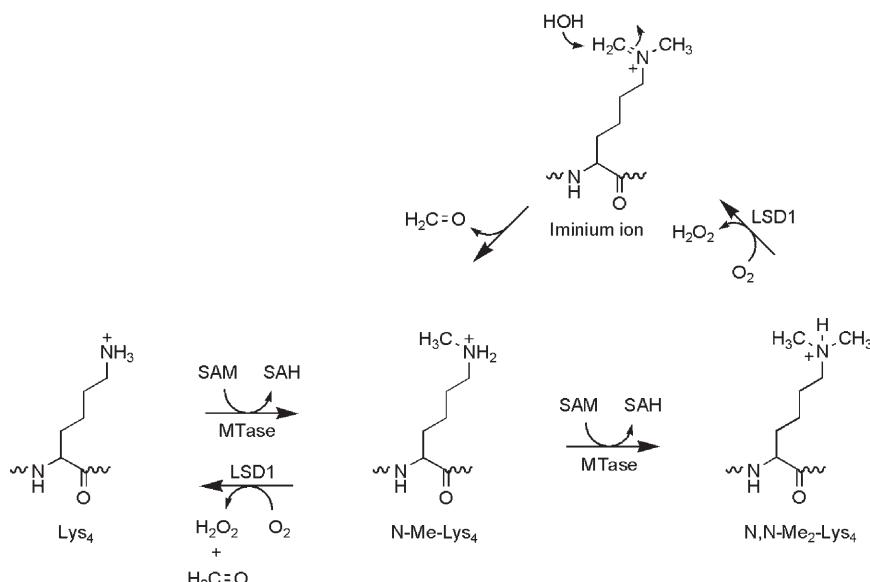
Histone Lys-N-methylation, unlike Lys-acetylation, is not susceptible to enzymatic or non-enzymatic hydrolytic reversal because there is no obvious path for hydrolytic cleavage of the N-alkyl bonds. Thus, in terms of writing and rewriting the histone code, it has been assumed that acetylation can be readily erased, but not methylation. However, the recent discovery of FAD-dependent methyl-Lys deaminases provides an oxidative route for counteracting the action of histone methyltransferases.^[96] Oxidative enzymatic conversion of an N-methyl-Lys into the $\text{CH}_2=\text{NH}$ -Lys product linkage (Scheme 29) now creates a product imine labile to hydrolysis to yield the unmodified $\text{H}_2\text{N-Lys}$ residue and 1 equivalent of formaldehyde.

The dozens/hundreds of protein ubiquitin ligases create ubiquityl-ε-NH-Lys-protein isopeptide bonds in tagged proteins. These linkages are resistant to normal proteases but the isopeptide bonds are cleaved by a family of several dozen deubiquitylases (DUBs), presumably selective for subsets of ubiquitylated proteins at different parts of the cell at different times and rescuing those from proteasome destruction.

Palmitoylated cysteine residues involve covalent thioester linkages. They can hydrolyze non-enzymatically, but this can be slow in membranes and there are specific palmitoyl



Scheme 28. Reversibility of covalent modifications in proteins; a) Phosphoprotein phosphatases oppose the effects of protein kinases; b) protein acetylations, S-palmitoylations, and ubiquitylations are reversed by deacetylases, palmitoyl-S-protein thioesterases, and deubiquitylases.



Scheme 29. Reversal of lysine ϵ -N-methylation occurs by oxidation (not hydrolysis) by action of a flavoprotein to generate the hydrolytically labile imine product.

protein thioesterases^[11] to control lifetime and consequent localization of the protein at membrane interfaces.

In contrast to the above reversible categories of post-translational modifications are sets that are functionally irreversible. Cys-S-prenylations are thioether linkages, and in contrast to the Cys-S-palmitoyl thioester linkages are not hydrolytically reversible. (Again, oxidative routes are known but probably occur during prenylated protein degradation in lysosomes.) Thus the enzymatic strategy for attaching the two common lipid anchors for proteins, palmitoylation or prenylation, provides reversible or irreversible modification of proteins targeted to membranes. The C-carboxylation of Glu residues to Gla residues is irreversible as is side-chain hydroxylation, for example, of Pro and Asn residues on HIF-1 α in oxygen sensing.

The quintessentially irreversible posttranslational modification is the second major category of covalent change to proteins: the proteolytic cleavage of peptide bonds.

7. Controlled Proteolysis

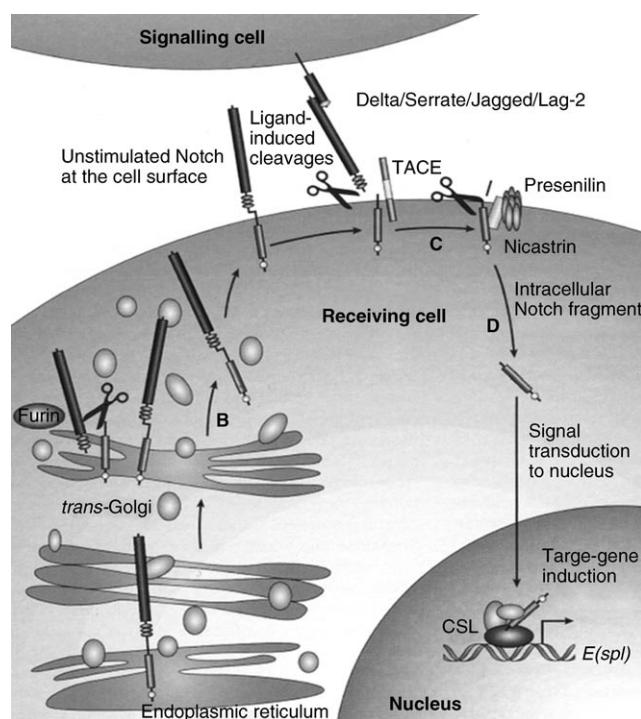
The life cycle of every protein in intracellular and extracellular compartments in an organism is controlled by homeostatic functioning of proteases, which cleave the covalent peptide backbones to release the constituent amino acids back into the monomer pool. Large subsets of proteins within eukaryotic cells may undergo consecutive limited proteolytic clipping as part of the normal temporal and spatial maturation process. These controlled proteolytic cuts at specific peptide sequences within a given set of protein substrates are effected by proteases that do not degrade but rather are involved in protein-substrate maturation and tailoring processes.

Essentially every protein that enters the endoplasmic reticulum in eukaryotic cells undergoes cleavage of the N-

terminal 25–30 amino acids, the signal sequence that specified transit into the ER, by signal peptidases.^[99] This first step in protein maturation can be followed by action of proprotein convertases later in the Golgi and *trans* Golgi network^[100,101] of the secretory compartment. A prototypical example is the cleavage of proinsulin to insulin, whose two chains are connected by disulfide bonds. The hormone cholecystokinin undergoes six–eight cleavages and trimming proteolytic maturations to convert the 115 residue initial translational product into the eight-residue sulfated mature hormone.^[102]

The maturation of the O-glycoprotein Notch by limited proteolysis occurs in four spatially and temporally distinct steps (Scheme 30): 1) signal peptide cleavage in the ER; 2) cleavage by proprotein convertases into two subunits that remain associated in the *trans* Golgi network on the way to the cell surface; 3) removal of the extracellular domain at the plasma membrane by action of sheddase-type proteolytic activity, triggered upon engagement of protein ligands;

4) regulated intramembrane proteolysis (RIP) of the truncated Notch as the fourth maturation event.^[99,103,104] The cytoplasmic stub of Notch, now with only a few remaining



Scheme 30. Maturation of the Notch protein by four sequential proteolytic cleavages at specific sites and specific places in a eukaryotic cell: a) Cleavage of the N-terminal signal peptide upon secretion into the endoplasmic reticulum (not shown in figure); b) hydrolytic clip by proprotein convertases during passage through the *trans* Golgi network; c) hydrolytic clip by a protein sheddase at the cell surface; d) regulated intermembrane proteolysis releases the cytoplasmic stub of Notch to go to the nucleus and act as transcription factor. Reproduced with permission from reference [121].

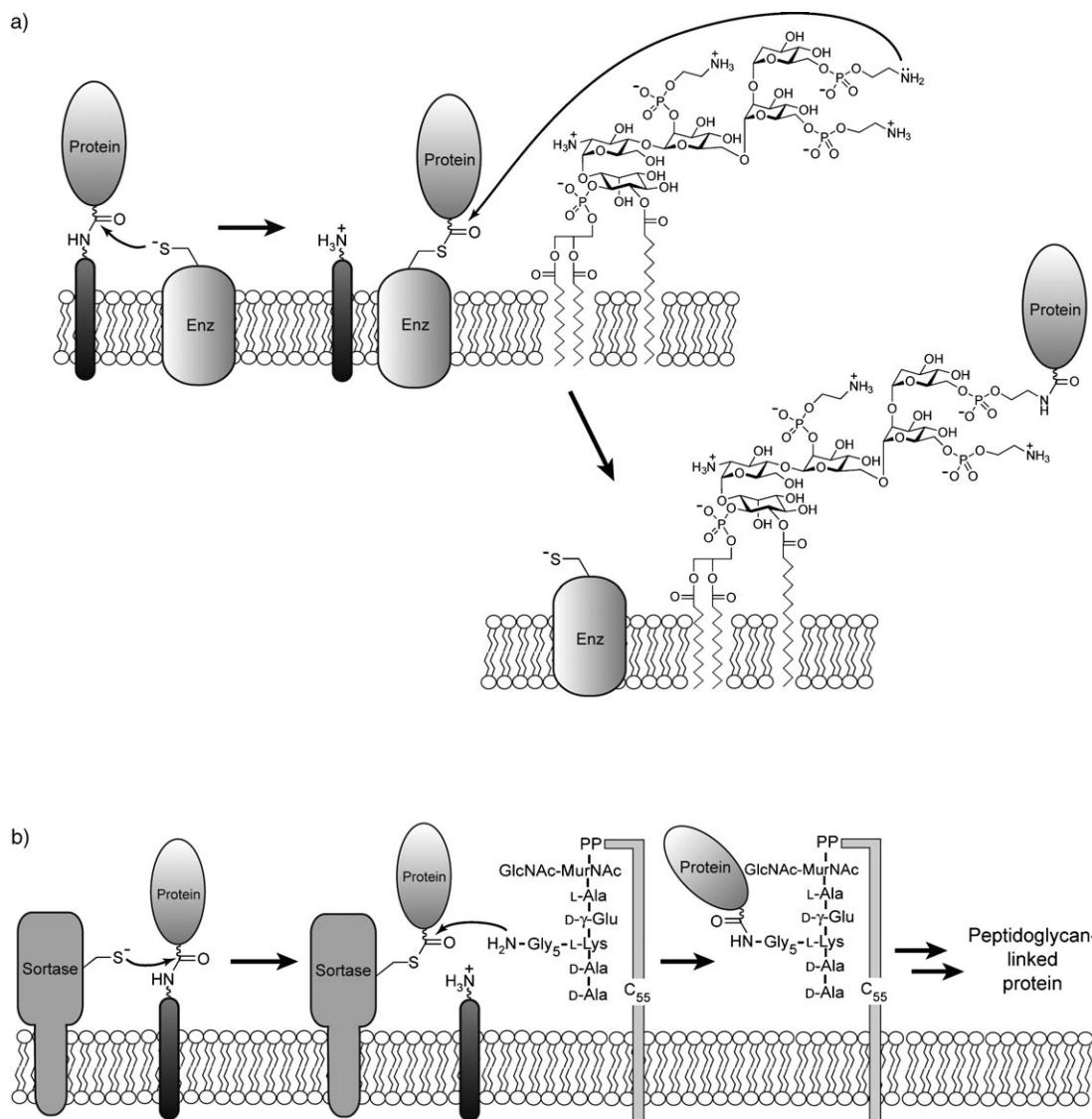
amino acids in the vesicle membranes, can then partition back to the cytoplasm, enter the nucleus, associate with partner proteins, and act as a selective transcriptional activator. The sequence of four limited proteolytic steps, separated in time and space, control the location and activity of Notch. It can be directed through the ER and the Golgi complex to the cell surface, function as a cell surface receptor, and then be liberated back into the cytoplasm as a transcriptionally activating fragment that can target the nucleus. Doubtless there are many other examples of multistep action of proteases to direct the location and function of client proteins and control their lifetimes in cells.

A variant of controlled proteolysis occurs when a short C-terminal peptide is excised from protein substrates by protease-like catalysts that generate transient covalent proteinyl-S-Cys-enzyme intermediates (Scheme 31). In the coupling of GPI lipid anchors to eukaryotic proteins, the transferring protein acyl group is specifically captured by an

ethanolamine group in the cosubstrate GPI anchor,^[105] a net switching of the C-terminal peptide of the initial translation product by the GPI anchor. In the action of bacterial sortases,^[106] the incoming amine nucleophile that captures the staphylococcal proteins to be displayed as antigens at the bacterial cell surface is a Gly or Ala-NH₂ terminus from cross-bridges in the peptidoglycan layer of the cell walls. Both of these C-terminal modifications are net transamidations, in which the incoming nucleophile is an amine to give an aminolytic product rather than water in hydrolysis.

8. Autocleavage and Peptide-Bond Rearrangement

There is a set of autocatalytic processes that leads folded proteins to catalyze rearrangements of the backbone connectivity at one or more specific peptide bonds in the folded proteins. Most often these rearrangements lead to cleavage of



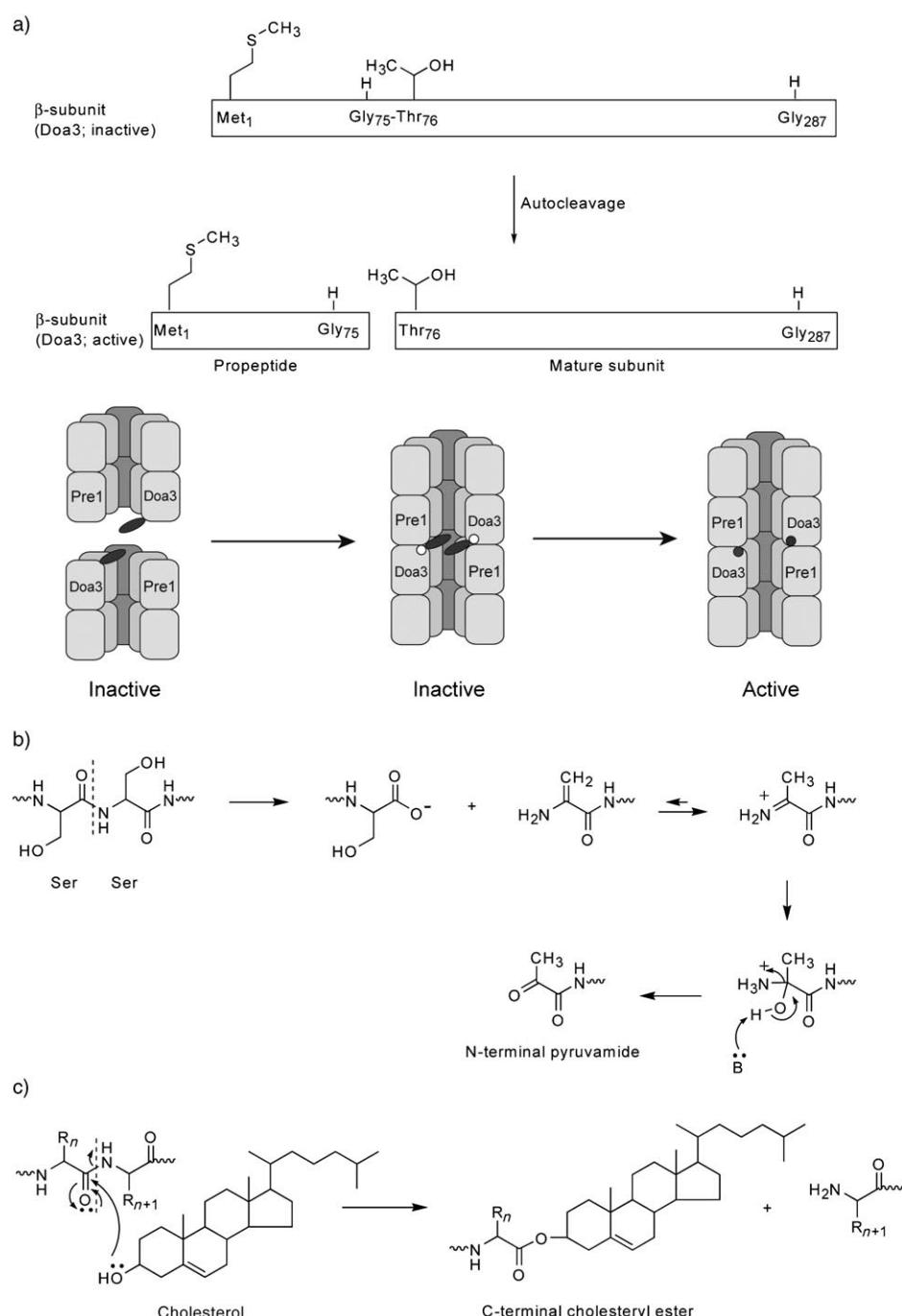
Scheme 31. Protease-like catalysts use active-site Cys nucleophiles to transfer protein-substrate-derived acyl fragments to nonprotein amine acceptors via acyl-S-Cys enzyme intermediates: a) GPI anchor attachment through enzymatic transamidation; b) cross-linking and display of proteins to the peptidoglycan cross-bridges in *Staphylococcus aureus* cell wall assembly by transamidation by the enzyme sortase.

those specific peptide bonds. The fate of the acyl fragment of the autocleaving peptide can vary depending on the capturing nucleophile. Water as cosubstrate results in hydrolysis. An alcohol as cosubstrate generates an ester product, as in the capture of the N-terminal fragment of Hedgehog protein by cholesterol. Capture by an amine, when the amine is downstream in the same protein, is the essence of protein splicing, although the amine capture is indirect as we shall note. A common type of intermediate is observed for all three such capture processes.

A second major variant of autocatalytic peptide backbone rearrangements occurs in two contexts (Section 9). One is the autoconversion of a tripeptide moiety into the highly conjugated aromatic fluorophore of the green fluorescent proteins and its relatives. The second is a related rearrangement of a tripeptide in the active site of the pro forms of phenylalanine and histidine deaminases to create the imidazolone cofactors that serve as essential electron sinks in deaminase catalysis.

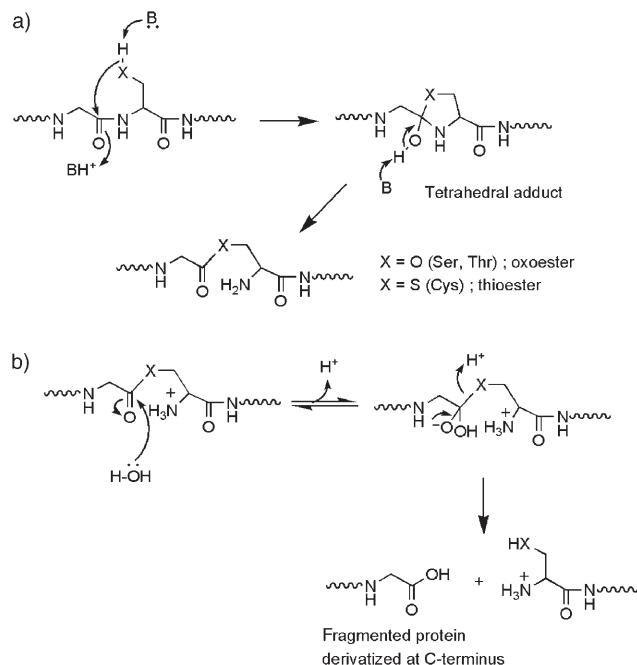
8.1. Autocleavage of Specific Peptide Bonds in Proteins During Precursor Activation

A relatively small subset of folded proteins have the capacity to convert themselves from a single-chain inactive precursor form into an active two-chain form by hydrolysis of a specific peptide bond in the precursor (Scheme 32). Three examples of variants of autoproteolytic activation are: 1) cleavage of the precursor single chain form of the β -subunits of proteasomes; 2) cleavage and uncovering of an N-terminal pyruvamide group in aspartate α -decarboxylase; 3) cleavage and activation of Hedgehog family proteins resulting in alcoholysis of the peptide bond by cholesterol. All three enzymes use the side chain of a Ser, Thr, or Cys residue to attack the immediately adjacent upstream peptide carbonyl to generate a five-membered tetrahedral adduct



Scheme 32. Autoproteolysis at specific peptide linkages to create the active forms of enzymes in the fragmented protein products; a) autocleavage of the precursor forms of the β subunits of proteasomes to liberate the catalytic Thr nucleophile; b) autocleavage of the precursor form of aspartate α -decarboxylase generates the N-terminal pyruvamide electron sink; c) autocleavage of the Hedgehog precursor and capture by the 3-OH of cholesterol generates the active N-terminal fragment as a membrane-tethered cholestryl ester.

(Scheme 33). If that adduct is protonated on nitrogen and reopens with cleavage of the C–N bond, then the peptide bond has been cleaved. The two parts of the protein chain are still held together, by an O-ester (Ser, Thr) or an S-thioester (Cys) linkage. This is the common thread of all three examples. The O-ester or S-thioester linkage is now labile to capture by even relatively weak nucleophiles such as water,



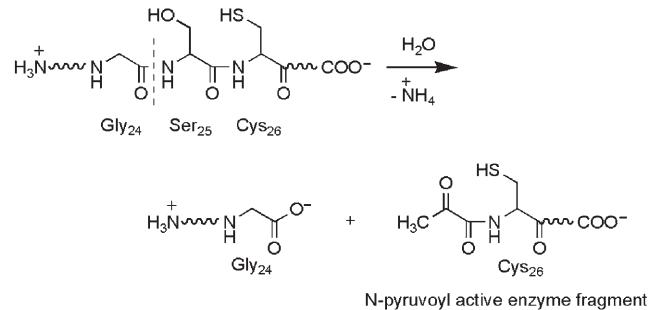
Scheme 33. Peptide-bond autocleavage mechanism proceeds via a) a tetrahedral cyclic adduct arising from attack of a nucleophilic side chain on the immediate upstream peptide bond; resolution of the adduct proceeds through C–N bond cleavage, breaking the original peptide bond and generating an oxo/thioester connectivity; b) the ester can be captured by a range of nucleophiles, including water.

the alcohol group of cholesterol, or downstream Ser and Cys side chains.

The proteasome is organized in an $\alpha_7\beta_7\beta_7\alpha_7$ four-layered cylinder.^[107] In yeast, three of the seven β subunits in each heptad ring are active, arising by autocleavage of the precursor β chains at Gly₇₅–Thr₇₆, releasing the Thr as the new N-terminus of the cleaved β subunit.^[108–110] Cleavage occurs when the side chain of Thr₇₆ attacks the adjacent peptide carbonyl to initiate cleavage. The liberated amino group of this new Thr₁ is the nucleophile in catalysis as protein substrates are cleaved, explaining why the precursor is inactive.^[111] The proteasomal β subunit is one of several proenzymes that autocleave and liberate the N-terminal nucleophile as the catalytic residue.^[109]

A related class of precursor forms of enzymes uses the side chain CH₂OH of a specific serine residue to initiate cleavage in the folded precursor on the adjacent upstream peptide carbonyl. As noted above, the tetrahedral adduct decomposes to an O-ester, cleaving the original C–N linkage of the peptide bond. Elimination of the RO group from this ester yields an N-terminal dehydroalanyl residue at the downstream chain of the two-chain product (Scheme 34).^[109] The N-terminal dehydroalanyl moiety hydrates and ketonizes to accumulate as a pyruvamide group at the N terminus of the active enzyme. This directed autoproteolytic process has uncovered a ketone moiety, an N-terminal electrophile rather than the N-terminal nucleophile above, which acts as an active-site electron sink for the substrate aspartate. Decarboxylation of the aspartyl imino enzyme followed by hydrolysis of the bound product imine yields β -alanine, a required intermediate in CoASH biosynthesis.^[112]

Aspartate decarboxylase from *E. coli*



Scheme 34. Peptide autocleavage at an X-Ser linkage can be followed by elimination of the Ser- β -OH and tautomerization to an N-terminal pyruvamide group. This is an electron sink that can engage in imine formation with amino acid substrates to facilitate kinetically accessible substrate carbanions, in this case for net decarboxylation of Asp.

ysis of the bound product imine yields β -alanine, a required intermediate in CoASH biosynthesis.^[112]

The third variant of precursor protein autoproteolysis is represented by the maturation of the protein Hedgehog that signals at plasma membranes of eukaryotic cells after two-dimensional diffusion in the plane of the membrane. The protein is tethered to the membrane by a C-terminal covalent lipid anchor, in this case by esterification of a cholesterol moiety with the C-terminal carboxylate.^[113] This arises from the autocleavage intermediate common to the above two cases in which rearrangement of a peptide bond in the precursor to an oxoester occurs by an identical mechanism. Now, rather than capture of the acyl fragment of the precursor protein by water for net hydrolysis, there is a specific binding site for cholesterol. Its 3'-OH is the kinetically competent nucleophile, yielding the peptidyl cholesterol ester as the cholesterolysis product that is biologically active (see Scheme 32).

These three variations of folded protein-precursor autocleavage in Scheme 32 yield either a) two normal peptide fragments, b) a downstream fragment with an N-terminal pyruvamide, or c) an upstream fragment with a C-terminal cholesterol ester. This shows the versatility of this autocatalyzed peptide bond route by selective control of the fate of the common rearranged ester intermediate. In all three cases the peptide bond fragmentations are irreversible.

8.2. Autocleavage and Religation: Protein Splicing

In the final variant of autocleavage of protein precursors the ester intermediate (or thioester intermediate when a Cys thiolate attacks the adjacent upstream peptide bond) is captured by an intramolecular nucleophile rather than an external one such as HOH or ROH. The internal nucleophile is a side chain Ser, Thr, or Cys in a folded downstream domain of the precursor protein. This is the essence of protein splicing, autocleavage, and peptide-bond religation, practiced by over a hundred bacterial and yeast proteins,^[109, 114] including DNA polymerases and ATPases. The first peptide bond is

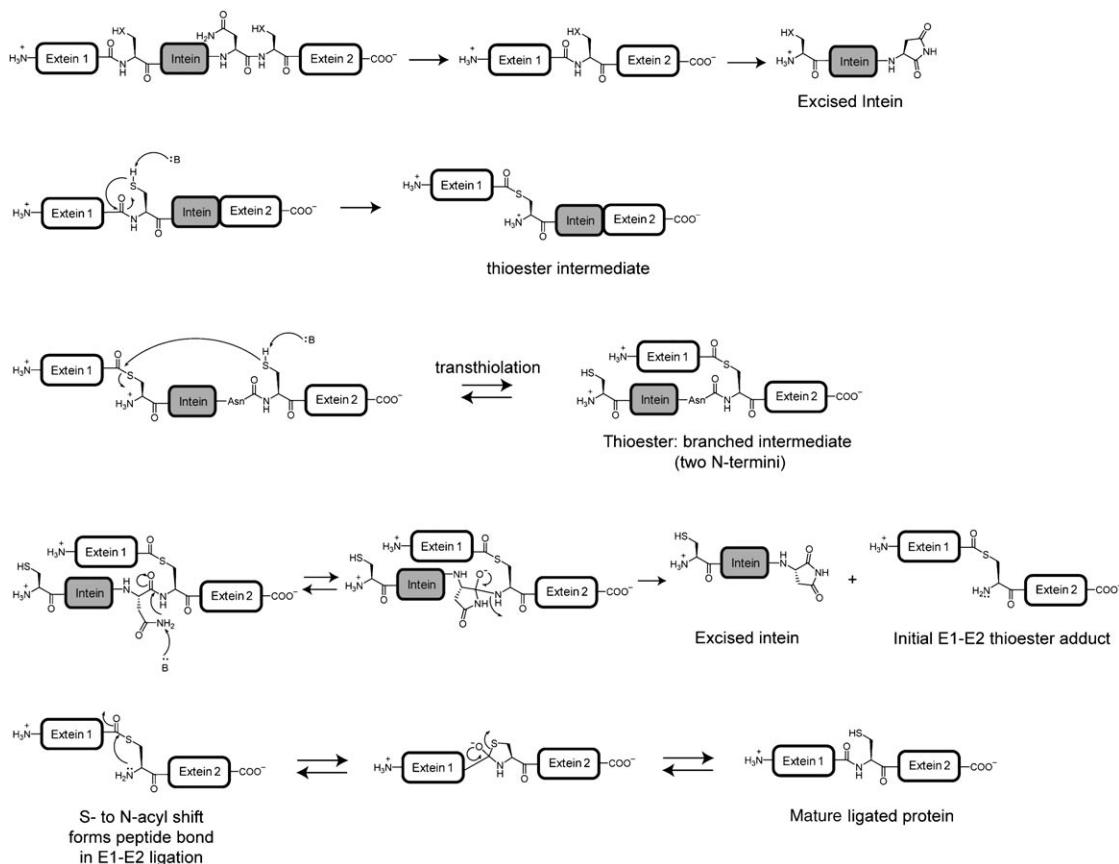
cleaved at the junction of upstream extein and intein. The internal nucleophile capturing the extein₁-intein-O-ester intermediate is at the boundary of the same intein and extein₂ (Scheme 35). The new intermediate has a lariat structure. Excision of the intein is completed with participation of an Asn side chain at the downstream boundary of the intein. This yields a second ester (thioester) intermediate as the lariat is resolved. Re-formation of the peptide bond joining extein₁-extein₂ is driven by thermodynamically favored acyl-O or acyl-S to N shifts and completes an in-frame ligation.

This is ancient protein biochemistry in an evolutionary sense (found in archaeal proteins) and may be the harbinger of the other autocleavage reactions noted in the above section. The net reversibility of such peptide-bond cleavage is the remarkable outcome and is controlled by structural features in the folded intermediates to keep water from being a competent, interfering nucleophile to the intramolecular acyl transfers within the self-splicing proteins. To the extent that this is ancient protein chemistry, it emphasizes that acyl oxoester and acyl thioester intermediates were primordial species in protein reactions; they are still intermediates in many posttranslational reactions in contemporary protein maturations. Furthermore, protein splicing has many practical applications in modifications of recombinant proteins with synthetic peptides by protein ligation.^[109,115]

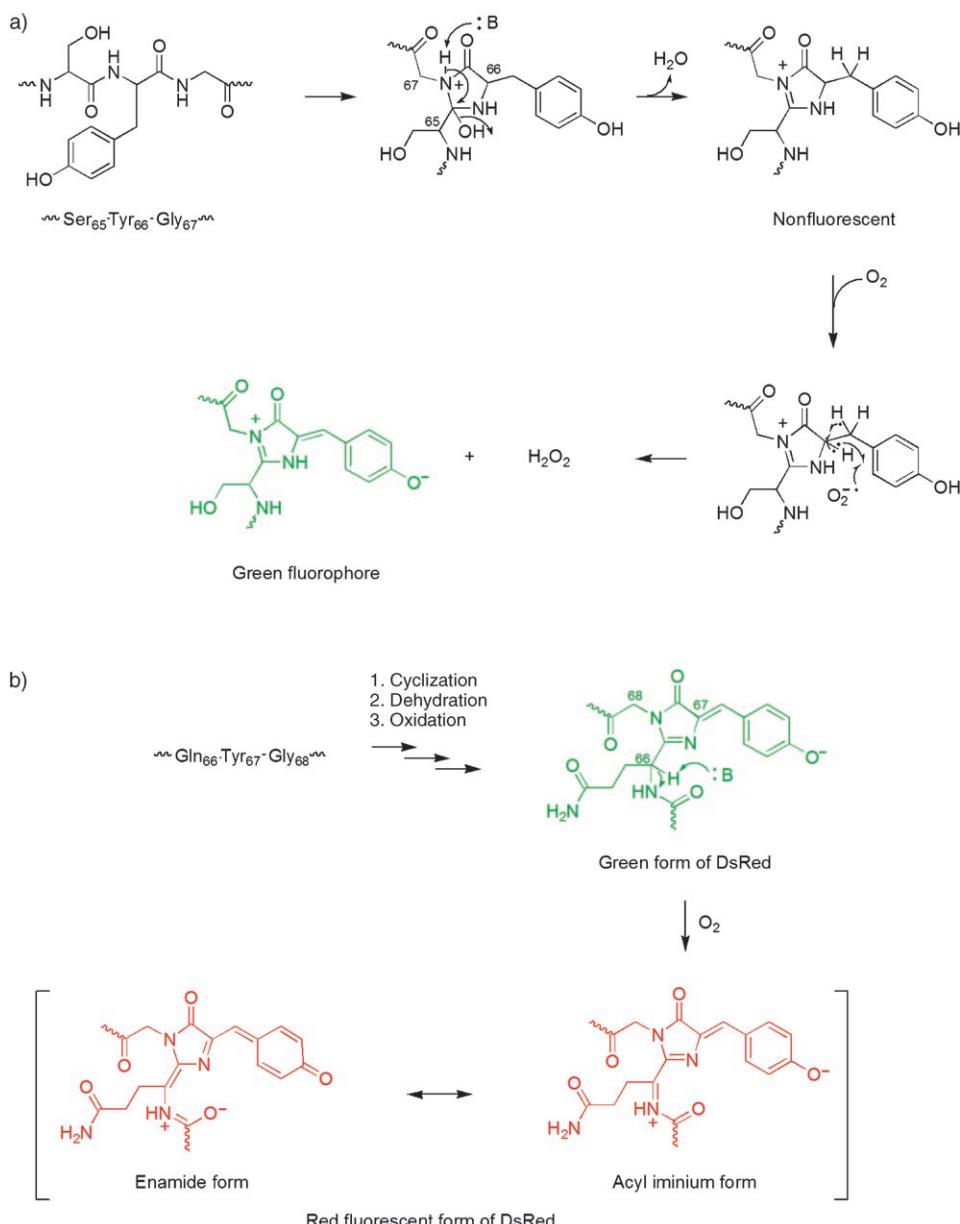
9. Peptide Bond Rearrangement without Autocleavage

9.1. Fluorophore Formation in Green Fluorescent Protein

An additional spectacular class of autocatalyzed PTM rearrangement of the peptide backbone in folded proteins is the maturation of chromoproteins of the green fluorescent protein (GFP) and red fluorescent protein families (Scheme 36a).^[116] The precursor protein folds into a β cylinder structure and the native conformation is required for the subsequent chromophore generation. A tripeptide loop of Ser₆₅Tyr₆₆Gly₆₇ in the folded colorless GFP precursor is sterically compressed, populating a conformer that allows attack of the Gly₆₇ amide N-H on the adjacent peptide carbonyl to generate a five-membered tetrahedral adduct reminiscent of the initial steps in the rearrangements described in Section 8. This adduct is dehydrated, and the resultant stable cyclic species is slowly autoxidized to create a double bond in conjugation with the phenol ring of Tyr₆₆. This oxidative last step generates the chromophore with absorption maximum at 506 nm and the green fluorescence useful to the producing coelenterate for energy-harvesting functions. A variant with QYG in the starting tripeptide is found in related coelenterates and yields the DsRed fluorophore after rearrangement and oxidative maturation (Scheme 36b).^[117] Fluo-



Scheme 35. Autocleavage, intein excision, and peptide religation during protein splicing: thioester formation during peptide bond autocleavage, followed by transthiolation to the cysteine at the intein-extein₂ boundary creates a lariat intermediate, that resolves by participation of the Asn at the C terminus of the intein. The resultant thioester reforms the extein₁-extein₂ peptide bond by acyl S to N shift.



Scheme 36. a) Autoconversion of $\text{Ser}_{65}\text{-Tyr}_{66}\text{-Gly}_{67}$ in the precursor of green fluorescent protein (GFP) to the green fluorescent form: formation of a cyclic tetrahedral adduct followed by dehydration generates the heterocycle that is not yet fluorescent. Extension of the conjugated system is a slow, oxygen-dependent oxidation that brings the Tyr chromophore into conjugation and completes fluorophore formation. b) A related protein with a $\text{Gln}_{66}\text{-Tyr}_{67}\text{-Gly}_{68}$ tripeptide instead of the SerTyrGly sequence is the fluorophore in the DsRed protein from coral. It proceeds through a green intermediate on the way to the final red form.

rophores with altered colors, including cyan and gold, have been generated by protein engineering and evolution approaches, allowing many fluorescence-energy-transfer studies between pairs of variant GFPs.

9.2. The Methyleneimidazolone Chromophore

Related rearrangements of a tripeptide sequon in a loop of the folded proenzyme forms of phenylalanine and histidine deaminases produce a related cyclic imidazolone group that

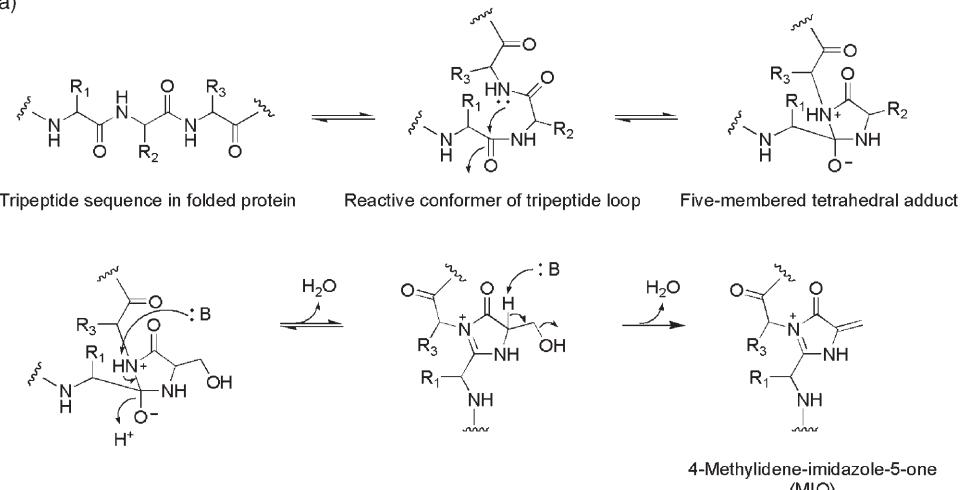
then functions as the active-site electrophile for substrate amine deamination (Scheme 37a).^[118,119] In the *Pseudomonas* histidine deaminase, $\text{Ala}_{142}\text{Ser}_{143}\text{Gly}_{144}$ in a loop region autoconverts to a compact heterocycle in which the Gly_{144} amide nitrogen is proposed to attack the carbonyl of Ala_{142} . The cyclic tetrahedral adduct can lose water and form the C=N of the imidazolone. Loss of the -OH from the Ser_{143} side chain forms the new double bond in conjugation to produce the 4-methylene-5-imidazolone (MIO) prosthetic group. This set of transformations fashions an electrophilic heterocyclic cofactor from the tripeptide loop to generate the active enzyme. The MIO heterocycle in the mature, active enzyme could be attacked by the substrate amino group. An elimination reaction initiated by β -H removal would then generate urocanate and the amino moiety still attached to the cofactor (Scheme 37b). Release of NH_3 regenerates the starting MIO cofactor for the next catalytic cycle.^[120]

10. Conclusions

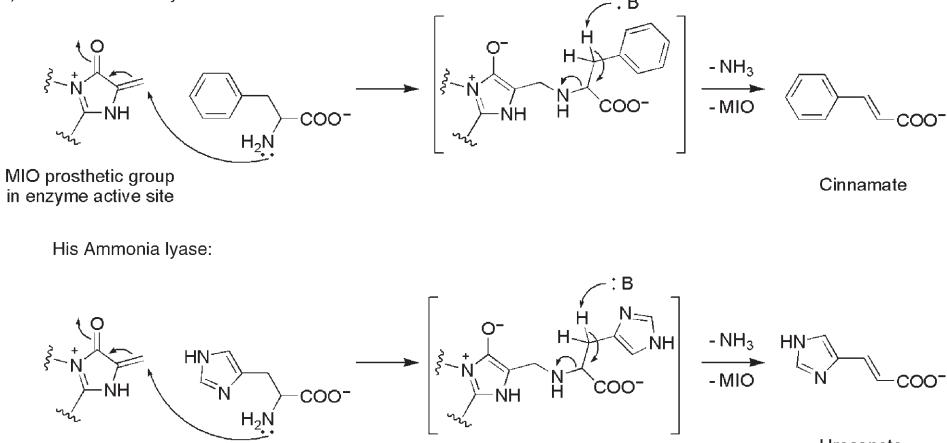
This summary of posttranslational modifications of proteins has not attempted exhaustive coverage of the more than 200 known covalent modifications (see reference [3] for more complete coverage). Rather, emphasis has been on how the addition of chemical groups from

common cofactors and coenzymes to side chains of 15 of the 20 amino acids found in proteins expands the proteome structurally and functionally. The dramatic enhancement of the capabilities of the limited scaffold of genetically encoded protein backbones creates new functional capacities. The modified proteins now have expanded opportunities for catalysis, initiation and termination of signal cascades, integration of information at many metabolic intersections, and alteration of cellular addresses. The posttranslational diversification of the proteome illuminates the underlying molecular logic for epigenetic acquisition of new protein functions.

a)



b) Phe Ammonia lyase:



Scheme 37. a) A rearrangement of a tripeptide loop, analogous to that in GFP maturation, occurs in autoactivation of histidine and phenylalanine ammonia lyases. Ala₁₄₂Ser₁₄₃Gly₁₄₄ is converted into a compact heterocycle termed MIO (4-methylene-5-imidazolone). b) The MIO is an electrophilic cofactor that can be attacked by the amino groups of Phe or His to set up net α,β -elimination of NH₂ and H to produce ammonia and the olefinic, deaminated acids cinnamate and urocanate.

The controlled cleavage of specific peptide bonds in particular protein substrates, especially those transiting eukaryotic secretory pathways, shows explicitly that proteases need not be blunt instruments in cellular control inventories. The life cycles of proteins secreted to the cell surface, engaged by ligands, and then retrieved to the nucleus by consecutive rounds of limited proteolysis show ingenious and sophisticated use of the molecular logic of proteolytic trimming of precursors.

Proteins can clearly help themselves for posttranslational creation of new functional groups. Folded protein precursor forms exhibit remarkable properties for autocatalytic conversion of peptide loops into cyclic derivatives for catalysis or light-energy harvesting. Controlled autocatalytic cleavages, presumably vestigial capacities of proteins during evolution, set up N-terminal nucleophiles, N-terminal electrophiles (pyruvamides), and C-terminal cholesterol esters by routing common intermediates down different decomposition pathways. The apotheosis of these autocleavages are the error-free

rearrangements in protein splicing reactions in which inteins are excised and exteins self-splice in frame to convert inactive precursors into mature proteins. Splicing may have been an important route to shuffling domains within proteins as a part of multidomain protein evolution pathways.

A given protein can be posttranslationally modified at many residues with the same group, for example, eleven phosphorylations of Abl, five acetylations at the C terminus of p53. Or a protein can be subjected to tandem modifications by several kinds of covalently introduced groups, as exemplified by the two lipidations of Ras followed by regiospecific proteolysis and C-terminal O-methylation, or the multiple modifications of Rho by bacterial protein toxins that disrupt eukaryotic cell cytoskeletal apparatus. The coordinated orchestration of acetylations, methylations, phosphorylations and ubiquitylations of histone tails on nucleosomes give insight into the finely tuned molecular logic of protein posttranslational modifications for gene-expression control.

With the advent of many variations of high-resolution mass spectrometry it is possible to detect and localize covalent changes in proteins beyond the genetically encoded sequence. Cataloguing the protein variants in various subproteomes (such as the phosphoproteome, the ubiquitylated proteome, the molecular variants of histone modifications in nucleosomes of differing transcriptional activity) will continue

to be important parameters to allow a full description of the protein composition of proteomes. This information will be a necessary preamble to understanding how the diverse molecular forms of proteins carry out their integrated functions.

We thank M. Fischbach for assistance with the production of the frontispiece.

Received: March 21, 2005

Published online: November 3, 2005

- [1] D. L. Black, *Annu. Rev. Biochem.* **2003**, *72*, 291–336.
- [2] T. Maniatis, B. Tasic, *Nature* **2002**, *418*, 236–243.
- [3] C. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, B. Roberts, Colorado, **2005**.
- [4] M. Mann, S. E. Ong, M. Gronborg, H. Steen, O. N. Jensen, A. Pandey, *Trends Biotechnol.* **2002**, *20*, 261–268.
- [5] J. A. Hoch, T. J. Silhavy, *Two Component Signal Transduction*, ASM, Washington, **1995**.

- [6] A. Rodrigue, Y. Quentin, A. Lazdunski, V. Mejean, M. Foglino, *Trends Microbiol.* **2000**, *8*, 498–504.
- [7] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, *Science* **2002**, *298*, 1912–1934.
- [8] L. N. Johnson, R. J. Lewis, *Chem. Rev.* **2001**, *101*, 2209–2242.
- [9] K. Zhang, K. E. Williams, L. Huang, P. Yau, J. S. Siino, E. M. Bradbury, P. R. Jones, M. J. Minch, A. L. Burlingame, *Mol. Cell. Proteomics* **2002**, *1*, 500–508.
- [10] M. D. Resh, *Biochim. Biophys. Acta* **1999**, *1451*, 1–16.
- [11] M. J. Bijlmakers, M. Marsh, *Trends Cell Biol.* **2003**, *13*, 32–42.
- [12] C. L. Brooks, W. Gu, *Curr. Opin. Cell Biol.* **2003**, *15*, 164–171.
- [13] M. Li, J. Luo, C. L. Brooks, W. Gu, *J. Biol. Chem.* **2002**, *277*, 50607–50611.
- [14] B. M. Turner, *Cell* **2002**, *111*, 285–291.
- [15] D. R. Johnson, R. S. Bhatnagar, L. J. Knoll, J. I. Gordon, *Annu. Rev. Biochem.* **1994**, *63*, 869–914.
- [16] S. McLaughlin, A. Aderem, *Trends Biochem. Sci.* **1995**, *20*, 272–276.
- [17] D. A. Johnson, P. Akamine, E. Radzio-Andzelm, M. Madhusudan, S. S. Taylor, *Chem. Rev.* **2001**, *101*, 2243–2270.
- [18] O. Rocks, A. Peyker, M. Kahms, P. J. Verveer, C. Koerner, M. Lumbierres, J. Kuhlmann, H. Waldmann, A. Wittinghofer, P. I. Bastiaens, *Science* **2005**, *307*, 1746–1752.
- [19] C. M. Pickart, *Annu. Rev. Biochem.* **2001**, *70*, 503–533.
- [20] C. M. Pickart, R. E. Cohen, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 177–187.
- [21] N. Zheng, B. A. Schulman, L. Song, J. J. Miller, P. D. Jeffrey, P. Wang, C. Chu, D. M. Koepf, S. J. Elledge, M. Pagano, R. C. Conaway, J. W. Conaway, J. W. Harper, N. P. Pavletich, *Nature* **2002**, *416*, 703–709.
- [22] T. Cardozo, M. Pagano, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 739–751.
- [23] J. D. Schnell, L. Hicke, *J. Biol. Chem.* **2003**, *278*, 35857–35860.
- [24] J. S. Bonifacino, L. M. Traub, *Annu. Rev. Biochem.* **2003**, *72*, 395–447.
- [25] L. Hicke, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 195–201.
- [26] D. J. Katzmann, G. Odorizzi, S. D. Emr, *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 893–905.
- [27] A. Hershko, A. Ciechanover, *Annu. Rev. Biochem.* **1998**, *67*, 425–479.
- [28] S. I. Reed, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 855–864.
- [29] A. W. Murray, *Cell* **2004**, *116*, 221–234.
- [30] R. Marmorstein, *Nat. Rev. Mol. Cell Biol.* **2003**, *303*, 1–7.
- [31] R. Roskoski, Jr., *Biochem. Biophys. Res. Commun.* **2003**, *303*, 1–7.
- [32] S. Khorasanizadeh, *Cell* **2004**, *116*, 259–272.
- [33] C. Walsh, *Enzymatic Reaction Mechanisms*, Freeman, San Francisco, **1979**.
- [34] J. A. Glomset, M. H. Gelb, C. C. Farnsworth, *Trends Biochem. Sci.* **1990**, *15*, 139–142.
- [35] F. L. Zhang, P. J. Casey, *Annu. Rev. Biochem.* **1996**, *65*, 241–269.
- [36] O. Pylypenko, A. Rak, R. Reents, A. Niculae, V. Sidorovitch, M. D. Cioaca, E. Bessolitsyna, N. H. Thoma, H. Waldmann, I. Schlichting, R. S. Goody, K. Alexandrov, *Mol. Cell* **2003**, *11*, 483–494.
- [37] A. Rak, O. Pylypenko, T. Durek, A. Watzke, S. Kushnir, L. Brunsved, H. Waldmann, R. S. Goody, K. Alexandrov, *Science* **2003**, *302*, 646–650.
- [38] A. Furmanek, J. Hofsteenge, *Acta Biochim. Pol.* **2000**, *47*, 781–789.
- [39] Y. Mechref, M. V. Novotny, *Chem. Rev.* **2002**, *102*, 321–369.
- [40] A. Helenius, M. Aeby, *Annu. Rev. Biochem.* **2004**, *73*, 1019–1049.
- [41] E. S. Trombetta, A. J. Parodi, *Adv. Protein Chem.* **2001**, *54*, 303–344.
- [42] J. Roth, *Chem. Rev.* **2002**, *102*, 285–303.
- [43] E. S. Trombetta, A. J. Parodi, *Annu. Rev. Cell Dev. Biol.* **2003**, *19*, 649–676.
- [44] P. M. Rudd, T. Endo, C. Colominas, D. Groth, S. F. Wheeler, D. J. Harvey, M. R. Wormald, H. Serban, S. B. Prusiner, A. Kobata, R. A. Dwek, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13044–13049.
- [45] P. M. Rudd, A. H. Merry, M. R. Wormald, R. A. Dwek, *Curr. Opin. Struct. Biol.* **2002**, *12*, 578–586.
- [46] K. Vosseller, K. Sakabe, L. Wells, G. W. Hart, *Curr. Opin. Chem. Biol.* **2002**, *6*, 851–857.
- [47] D. J. Moloney, L. H. Shair, F. M. Lu, J. Xia, R. Locke, K. L. Matta, R. S. Haltiwanger, *J. Biol. Chem.* **2000**, *275*, 9604–9611.
- [48] D. J. Moloney, V. M. Panin, S. H. Johnston, J. Chen, L. Shao, R. Wilson, Y. Wang, P. Stanley, K. D. Irvine, R. S. Haltiwanger, T. F. Vogt, *Nature* **2000**, *406*, 369–375.
- [49] L. Wells, S. A. Whalen, G. W. Hart, *Biochem. Biophys. Res. Commun.* **2003**, *302*, 435–441.
- [50] N. Haines, K. D. Irvine, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 786–797.
- [51] T. Okajima, A. Xu, K. D. Irvine, *J. Biol. Chem.* **2003**, *278*, 42340–42345.
- [52] N. M. Giles, A. B. Watts, G. I. Giles, F. H. Fry, J. A. Littlechild, C. Jacob, *Chem. Biol.* **2003**, *10*, 677–693.
- [53] C. Jacob, G. I. Giles, N. M. Giles, H. Sies, *Angew. Chem.* **2003**, *115*, 4890–4907; *Angew. Chem. Int. Ed.* **2003**, *42*, 4742–4758.
- [54] P. T. L. Chivers, M. R. Raines, *Protein Disulfide Isomerase: Cellular Enzymology of the CXXC Motif*, Marcel Dekker, New York, **1998**, pp. 487–505.
- [55] J. S. Stamler, S. Lamas, F. C. Fang, *Cell* **2001**, *106*, 675–683.
- [56] M. Griffin, R. Casadio, C. M. Bergamini, *Biochem. J.* **2002**, *368*, 377–396.
- [57] L. Lorand, R. M. Graham, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 140–156.
- [58] J. A. Vranka, L. Y. Sakai, H. P. Bachinger, *J. Biol. Chem.* **2004**, *279*, 23615–23621.
- [59] C. W. Pugh, P. J. Ratcliffe, *Nat. Med.* **2003**, *9*, 677–684.
- [60] W. C. Hon, M. I. Wilson, K. Harlos, T. D. Claridge, C. J. Schofield, C. W. Pugh, P. H. Maxwell, P. J. Ratcliffe, D. I. Stuart, E. Y. Jones, *Nature* **2002**, *417*, 975–978.
- [61] L. A. McNeill, K. S. Hewitson, T. D. Claridge, J. F. Seibel, L. E. Horsfall, C. J. Schofield, *Biochem. J.* **2002**, *367*, 571–575.
- [62] C. J. Schofield, P. J. Ratcliffe, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 343–354.
- [63] P. H. Maxwell, M. S. Wiesener, G. W. Chang, S. C. Clifford, E. C. Vaux, M. E. Cockman, C. C. Wykoff, C. W. Pugh, E. R. Maher, P. J. Ratcliffe, *Nature* **1999**, *399*, 271–275.
- [64] M. Ohh, C. W. Park, M. Ivan, M. A. Hoffman, T. Y. Kim, L. E. Huang, N. Pavletich, V. Chau, W. G. Kaelin, *Nat. Cell Biol.* **2000**, *2*, 423–427.
- [65] M. J. Ryle, R. P. Hausinger, *Curr. Opin. Chem. Biol.* **2002**, *6*, 193–201.
- [66] E. Chapman, M. D. Best, S. R. Hanson, C.-H. Wong, *Angew. Chem.* **2004**, *116*, 3610–3632; *Angew. Chem. Int. Ed.* **2004**, *43*, 3526–3548.
- [67] I. Capila, R. J. Linhardt, *Angew. Chem.* **2002**, *114*, 426–450; *Angew. Chem. Int. Ed.* **2002**, *41*, 391–412.
- [68] M. Farzan, T. Mirzabekov, P. Kolchinsky, R. Wyatt, M. Cayabyab, N. P. Gerard, C. Gerard, J. Sodroski, H. Choe, *Cell* **1999**, *96*, 667–676.
- [69] K. von Figura, A. Hasilik, *Annu. Rev. Biochem.* **1986**, *55*, 167–193.
- [70] K. von Figura, B. Schmidt, T. Selmer, T. Dierks, *Bioessays* **1998**, *20*, 505–510.
- [71] S. R. Hanson, M. D. Best, C.-H. Wong, *Angew. Chem.* **2004**, *116*, 5858–5886; *Angew. Chem. Int. Ed.* **2004**, *43*, 5736–5763.
- [72] R. von Bulow, B. Schmidt, T. Dierks, K. von Figura, I. Uson, *J. Mol. Biol.* **2001**, *305*, 269–277.

- [73] *Bacterial Protein Toxins* (Eds.: K. Aktories, I. Just), Springer, Berlin, **2000**.
- [74] I. R. Vetter, F. Hofmann, S. Wohlgemuth, C. Herrmann, I. Just, *J. Mol. Biol.* **2000**, *301*, 1091–1095.
- [75] M. Geyer, C. Wilde, J. Selzer, K. Aktories, H. R. Kalbitzer, *Biochemistry* **2003**, *42*, 11951–11959.
- [76] G. Schmidt, P. Sehr, M. Wilm, J. Selzer, M. Mann, K. Aktories, *Nature* **1997**, *387*, 725–729.
- [77] R. N. Perham, *Annu. Rev. Biochem.* **2000**, *69*, 961–1004.
- [78] C. T. Walsh, *Antibiotics: Actions, Origins, Resistance*, ASM, Washington, **2003**.
- [79] B. Furie, B. A. Bouchard, B. C. Furie, *Blood* **1999**, *93*, 1798–1808.
- [80] C. Shang, T. Shibahara, K. Hanada, Y. Iwafune, H. Hirano, *Biochemistry* **2004**, *43*, 6281–6292.
- [81] S. B. Ficarro, M. L. McCleland, P. T. Stukenberg, D. J. Burke, M. M. Ross, J. Shabanowitz, D. F. Hunt, F. M. White, *Nat. Biotechnol.* **2002**, *20*, 301–305.
- [82] J. Peng, D. Schwartz, J. E. Elias, C. C. Thoreen, D. Cheng, G. Marsischky, J. Roelofs, D. Finley, S. P. Gygi, *Nat. Biotechnol.* **2003**, *21*, 921–926.
- [83] H. Steen, M. Fernandez, S. Ghaffari, A. Pandey, M. Mann, *Mol. Cell. Proteomics* **2003**, *2*, 138–145.
- [84] J. J. Pesavento, Y.-B. Kim, G. K. Taylor, N. L. Kelleher, *J. Am. Chem. Soc.* **2004**, *126*, 3386–3387.
- [85] F. Shao, J. E. Dixon, *Adv. Exp. Med. Biol.* **2003**, *529*, 79–84.
- [86] L. Izzi, L. Attisano, *Oncogene* **2004**, *23*, 2071–2078.
- [87] J. B. Shabb, *Chem. Rev.* **2001**, *101*, 2381–2411.
- [88] M. D. Jackson, J. M. Denu, *Chem. Rev.* **2001**, *101*, 2313–2340.
- [89] A. Alonso, J. Sasin, N. Bottini, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, T. Mustelin, *Cell* **2004**, *117*, 699–711.
- [90] M. Mann, R. C. Hendrickson, A. Pandey, *Annu. Rev. Biochem.* **2001**, *70*, 437–473.
- [91] M. Mann, O. N. Jensen, *Nat. Biotechnol.* **2003**, *21*, 255–261.
- [92] H. H. Ng, A. Bird, *Trends Biochem. Sci.* **2000**, *25*, 121–126.
- [93] C. M. Grozinger, S. L. Schreiber, *Chem. Biol.* **2002**, *9*, 3–16.
- [94] A. A. Sauve, I. Celic, J. Avalos, H. Deng, J. D. Boeke, V. L. Schramm, *Biochemistry* **2001**, *40*, 15456–15463.
- [95] A. A. Sauve, V. L. Schramm, *Biochemistry* **2003**, *42*, 9249–9256.
- [96] Y. Shi, F. Lan, C. Matson, P. Mulligan, J. R. Whetstone, P. A. Cole, R. A. Casero, *Cell* **2004**, *119*, 941–953.
- [97] K. D. Wilkinson, *Semin. Cell Dev. Biol.* **2000**, *11*, 141–148.
- [98] T. Gan-Erdene, K. Nagamalleswari, L. Yin, K. Wu, Z. Q. Pan, K. D. Wilkinson, *J. Biol. Chem.* **2003**, *278*, 28892–28900.
- [99] M. Paetzl, A. Karla, N. C. J. Strynadka, R. E. Dalbey, *Chem. Rev.* **2002**, *102*, 4549–4579.
- [100] S. S. Molloy, E. D. Anderson, F. Jean, G. Thomas, *Trends Cell Biol.* **1999**, *9*, 28–35.
- [101] N. C. Rockwell, D. J. Krysan, T. Komiyama, R. S. Fuller, *Chem. Rev.* **2002**, *102*, 4525–4548.
- [102] G. A. Eberlein, V. E. Eysselein, M. T. Davis, T. D. Lee, J. E. Shively, D. Grandt, W. Niebel, R. Williams, J. Moessner, J. Zeeh, H. E. Meyer, H. Goebell, J. R. Reeve, Jr., *J. Biol. Chem.* **1992**, *267*, 1517–1521.
- [103] J. Arribas, A. Borroto, *Chem. Rev.* **2002**, *102*, 4627–4638.
- [104] Y. Ye, M. E. Fortini, *Semin. Cell Dev. Biol.* **2000**, *11*, 211–221.
- [105] K. Ohishi, N. Inoue, Y. Maeda, J. Takeda, H. Riezman, T. Kinoshita, *Mol. Biol. Cell* **2000**, *11*, 1523–1533.
- [106] S. K. Mazmanian, H. Ton-That, O. Schneewind, *Mol. Microbiol.* **2001**, *40*, 1049–1057.
- [107] D. Stock, P. M. Nederlof, E. Seemuller, W. Baumeister, R. Huber, J. Lowe, *Curr. Opin. Biotechnol.* **1996**, *7*, 376–385.
- [108] E. Seemuller, A. Lupas, W. Baumeister, *Nature* **1996**, *382*, 468–471.
- [109] H. Paulus, *Annu. Rev. Biochem.* **2000**, *69*, 447–496.
- [110] M. Groll, M. Bochtler, H. Brandstetter, T. Clausen, R. Huber, *ChemBioChem* **2005**, *6*, 222–256.
- [111] Y. D. Kwon, I. Nagy, P. D. Adams, W. Baumeister, B. K. Jap, *J. Mol. Biol.* **2004**, *335*, 233–245.
- [112] A. Albert, V. Dhanaraj, U. Genschel, G. Khan, M. K. Ramjee, R. Pulido, B. L. Sibanda, F. von Delft, M. Witty, T. L. Blundell, A. G. Smith, C. Abell, *Nat. Struct. Biol.* **1998**, *5*, 289–293.
- [113] J. A. Porter, S. C. Ekker, W. J. Park, D. P. von Kessler, K. E. Young, C. H. Chen, Y. Ma, A. S. Woods, R. J. Cotter, E. V. Koonin, P. A. Beachy, *Cell* **1996**, *86*, 21–34.
- [114] C. J. Noren, J. Wang, F. B. Perler, *Angew. Chem.* **2000**, *112*, 458–476; *Angew. Chem. Int. Ed.* **2000**, *39*, 450–466.
- [115] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705–6710.
- [116] R. Y. Tsien, *Annu. Rev. Biochem.* **1998**, *67*, 509–544.
- [117] D. Yarbrough, R. M. Wachter, K. Kallio, M. V. Matz, S. J. Remington, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 462–467.
- [118] M. Baedeker, G. E. Schulz, *Structure* **2002**, *10*, 61–67.
- [119] T. F. Schwede, J. Retey, G. E. Schulz, *Biochemistry* **1999**, *38*, 5355–5361.
- [120] J. C. Calabrese, D. B. Jordan, A. Boodhoo, S. Sariaslani, T. Vannelli, *Biochemistry* **2004**, *43*, 11403–11416.
- [121] M. E. Fortini, *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 673–684.
- [122] P. J. Kraulis, *J. Appl. Crystallogr.* **1991**, *24*, 946–950.