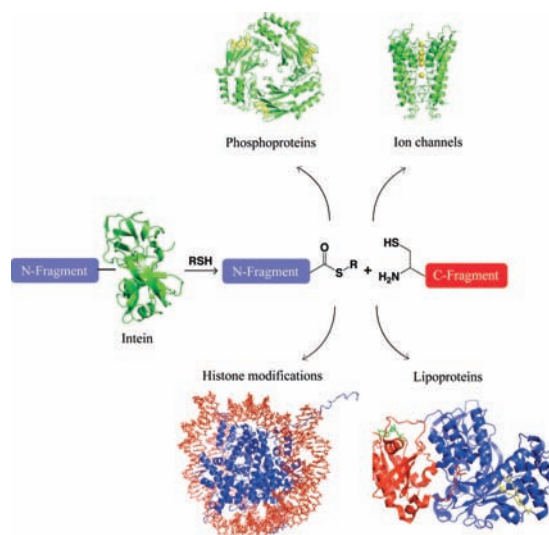# Expressed Protein Ligation (EPL) in the Study of Signal Transduction, Ion Conduction, And Chromatin Biology

ROBERT R. FLAVELL AND TOM W. MUIR*

*Laboratory of Synthetic Protein Chemistry, The Rockefeller University,
New York, New York 10065*

CONSPECTUS

Expressed protein ligation (EPL) is a semisynthetic technique in which a recombinant protein thioester, generated by thiolysis of an intein fusion protein, is reacted with a synthetic or recombinant peptide with an N-terminal cysteine to produce a native peptide bond. This method has been used extensively for the incorporation of biophysical probes, unnatural amino acids, and post-translational modifications in proteins. In the 10 years since this technique was developed, the applications of EPL to studying protein structure and function have grown ever more sophisticated.

In this Account, we review the use of EPL in selected systems in which substantial mechanistic insights have recently been gained through the use of the semisynthetic protein derivatives. EPL has been used in many studies to unravel the complexity of signaling networks and subcellular trafficking. Herein, we highlight this application to two different systems. First, we describe how phosphorylated or otherwise modified proteins in the TGF-β signaling network were prepared and how they were applied to understanding the complexities of this pathway, from receptor activation to nuclear import. Second, Rab-GTPases are multiply modified with lipid derivatives, and EPL-based techniques were used to incorporate these modifications, allowing for the elucidation of the biophysical basis of membrane association and dissociation.

We also review the use of EPL to understand the biology of two other systems, the potassium channel KcsA and histones. EPL was used to incorporate D-alanine and an amide-to-ester backbone modification in the selectivity filter of the KcsA potassium channel, providing insight into the mechanism of selectivity in ion conduction. In the case of histones, which are among the most heavily post-translationally modified proteins, the modifications play a key role in the regulation of gene transcription and chromatin structure. We describe how native chemical ligation and EPL were used to generate acetylated, phosphorylated, methylated, and ubiquitylated histones and how these modified histones were used to interrogate chromatin biology.

Collectively, these studies demonstrate the utility of EPL in protein science. These techniques and concepts are applicable to many other systems, and ongoing advances promise to extend this semisynthetic technique to increasingly complex biological problems.

This Account describes the application of expressed protein ligation (EPL) to the study of protein structure and function in selected systems. EPL is a semisynthetic technique in which a recombinantly expressed protein thioester is fused to any peptide containing an N-terminal cysteine. Because the C-terminal peptide may be prepared by synthetic methods, this allows for the incorporation of a variety of protein modifications or probes. In the decade since this technique was introduced, it has been applied to a range of targets, allowing for the site-specific incorporation of biophysical probes; attachment to surfaces, polymers, or nanoparticles; the introduction of post-translational modifications (PTMs) or unnatural amino acids; and the generation of proteins that are toxic or difficult to express by other means. Herein, we will provide a brief review of the development and methodological considerations of EPL (because this subject has been reviewed extensively elsewhere[1,2]) and describe, in detail, the application to selected systems in which the use of this technique has led to greater understanding of protein structure and function.

The development of recombinant DNA technology and techniques for the heterologous overexpression of polypeptides has allowed for the routine preparation of many proteins. However, these techniques are limited to the 20 naturally occurring amino acids. In many cases, it is desirable to obtain proteins with PTMs, unnatural amino acids, or biophysical probes. Thus, a variety of techniques have been developed for the preparation of proteins with unnatural amino acids, by amino acid replacement in auxotrophic expression strains or by re-engineering the cell's own biosynthetic machinery.[3,4] Additionally, a host of bioconjugation approaches can be used to attach biophysical probes or other modifications, and in many cases, these reactions can be applied site-specifically.[5] Finally, by the careful choice of expression systems, or the use of suitable modifying enzymes, a wide variety of post-translationally modified proteins can be isolated. However, all of these techniques can suffer from problems of low yield, heterogeneity in the isolated products, or limitations in the complexity and number of modifications incorporated.

In principle, a chemical approach could allow for the generation of a limitless array of modified proteins. The development, optimization, and automation of solid phase peptide synthesis (SPPS) allows for the routine synthesis of peptides up to 50 amino acids in length, the size of small protein domains.[6] Additionally, PTMs such as phosphates, lipids, and sugars can readily be incorporated into synthetic peptides. The fundamental limitation in the length of synthetic peptides can be overcome by the use of chemoselective ligation reactions, in which two or more fully unprotected peptides are linked together.

In 1994, the group of Kent extended the pioneering work of Wieland et al. on the reaction of amino acid thioesters with cysteine[7] with the development of native chemical ligation (NCL).[8] In NCL, an unprotected peptide with an N-terminal cysteine can react with a second peptide with a C-terminal thioester to form a peptide bond (Figure 1a). The simplicity and robustness of this approach has resulted in its widespread application.[9] In particular, we stress that the key feature of NCL, which helped to guide its development, was the need to avoid any protecting groups on either fragment. In general, this critical feature defines a peptide ligation reaction. We anticipate that continued technical advances will extend this approach to ever-larger and more complex molecules. Nevertheless, the current restriction of SPPS to peptides of 50−60 amino acids in length, combined with the difficulty in performing multiple sequential ligation steps, at present limits this technique to the synthesis of small to moderately sized proteins.

The fusion of the NCL technique with recombinant protein production was realized via two different approaches. First, a synthetic peptide thioester can be reacted by NCL with a recombinant protein with an N-terminal cysteine;[10] each fragment can be prepared via several different approaches.[2,10] Second, the fusion of a recombinant protein thioester with a synthetic N-terminal cysteine peptide was accomplished by re-engineering the process of protein splicing. In protein splicing, a protein segment termed an intein is excised out of two flanking sequences, named exteins, in a process analogous to the splicing of RNA introns (Figure 1b).[11] This fascinating autocatalytic process has been exploited for many different purposes by our group and others. However, one of the simplest uses of protein splicing has proven to be the most versatile: by mutating the intein to prevent the formation or breakdown of the branched intermediate, the protein can be trapped in equilibrium between the thioester and amide form. The mutated intein can then be cleaved by treatment with thiols, generating a recombinant protein thioester. The ligation of the recombinant protein thioester with a synthetic N-terminal cysteine peptide was first reported in 1998 (Figure 1c)[12−14] and has been termed expressed protein ligation (EPL) or, less frequently, intein-mediated protein ligation.

The scope of EPL was initially explored in a series of studies in which various synthetic modifications were incorporated into recombinant proteins.[1,2] Since this subject was last reviewed comprehensively, EPL has been used for a variety of
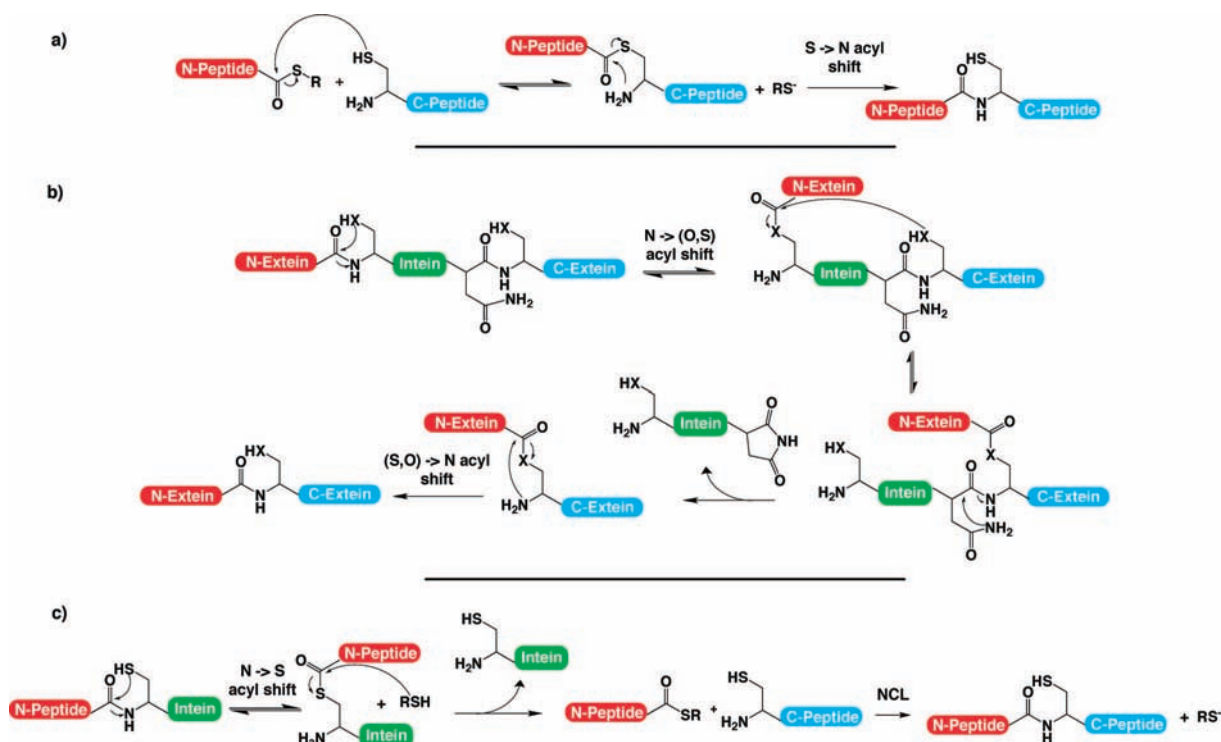
FIGURE 1. (a) Mechanism of NCL. In NCL, transthioesterification by the N-terminal cysteine is followed by an S → N acyl shift, generating a native peptide bond. (b) Mechanism of protein splicing. In protein splicing, a N → S or O acyl shift at the N-terminal residue of the intein is followed by *trans*(thio)-esterification to the residue C-terminal to the intein, generating a branched intermediate. The C-terminal asparagine of the intein then cleaves the C-terminal intein−extein bond, generating a succinimide product, and the fused exteins undergo an (S,O) → N acyl shift to create a native peptide bond. (c) Mechanism of EPL. In EPL, a thiol is used to generate a thioester from a mutated intein. The thioester can then react by NCL with an N-terminal cysteine peptide. R = alkyl, phenyl, benzyl, $CH_2CH_2SO_3Na$; X = O, S.

novel purposes, including the incorporation of orthogonal reactive groups for the purpose of evaluating novel protein conjugation reactions;[15,16] surface,[17] polymer,[18] biopolymer,[19] and nanoparticle conjugation;[20] the generation of protein oligomers;[21] the incorporation of positron emitting isotopes;[22] and the development of complex probes for following and manipulating protein structure in live cells.[23,24] In this Account, we will describe areas of research, primarily from our own group, in which EPL has been utilized beyond simple proof of principle studies to provide insights into the structure and function of complex protein systems.

Structural, Biochemical, and Biophysical Analysis of TGF-β Signaling

TGF-β family ligands are secreted polypeptides that regulate a variety of biological responses in growth and development, and dysregulation of TGF-β signaling is implicated in pathophysiogical conditions such as cancer. In the standard model of TGF-β signaling (Figure 2a), ligand binding induces the formation of a tetrameric receptor signaling complex consisting of two TβR-I and two TβR-II receptors.[25] The TβR-II receptor then phosphorylates TβR-I on a segment termed the GS-re-

gion, which activates the kinase activity on the TβR-I receptor, leading to autophosphorylation and also recruitment and phosphorylation of R-SMADs. The phosphorylated R-SMADs can then homotrimerize or heterotrimerize with co-SMADs and translocate to the nucleus and activate or repress gene transcription. In order to assess the role of the regulatory GS segment in TβR-I activation, a homogeneously tetraphosphorylated derivative of the receptor cytoplasmic domain was prepared by ligation of a tetra-phosphopeptide thioester with a recombinantly expressed C-terminal peptide.[26,27] The phosphorylated kinase had 40-fold increased kinase activity toward its substrate, SMAD2, compared with the unphosphorylated kinase. The phosphorylation activity was directed to the C-terminal tail of SMAD2, and the phosphorylated kinase could no longer be inhibited by FBKP12, a negative regulator of the pathway.[28] Furthermore, the phosphorylated protein bound to a positively charged basic patch on SMAD2. Thus, phosphorylation of the GS segment of TβR-I leads to a conformational switch from a state that favors binding to the inhibitor FKBP12 to a state favoring the binding and phosphorylation of its substrate, SMAD2.
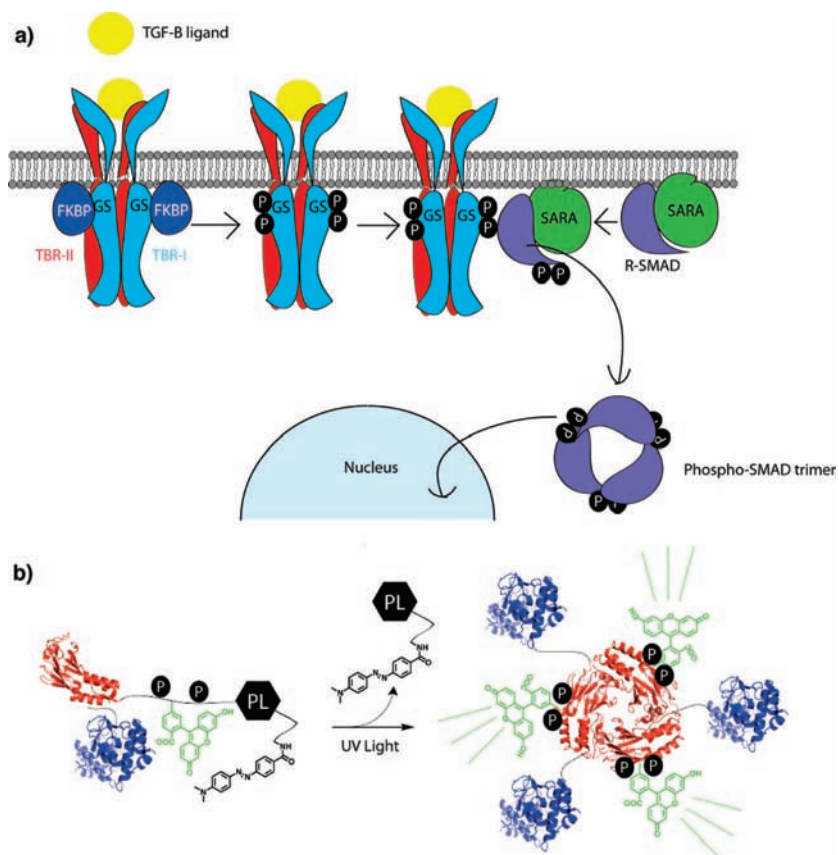
FIGURE 2. (a) TGF-β signaling and (b) simultaneous activation of SMAD2 trimerization and fluorescence using light. The protein consists of the MH1 and MH2 domains of SMAD2, with a modified C-terminus that is doubly phosphorylated (P), with fluorescein and a dabcyl quencher linked together via a photocleavable linker (PL). Treatment with UV light results in cleavage of the photolinker, relieving the quenching of the fluorescein and allowing trimerization and thus activation of the protein. PDB IDs: SMAD2 monomer 1dev; SMAD2 trimer 1khx; MH1 domain 1mhd.

Since the C-terminal tail of SMAD2 is the substrate of TβR-I, subsequent studies examined the role of phosphorylation in regulating the oligomerization, structure, and function of this transcription factor. A SMAD2 derivative, phosphorylated at serines 465 and 467, was prepared by EPL.[29] A crystal structure of the corresponding homotrimer revealed that the C-terminal phosphorylated region had extensive hydrogen bonding with the loop strand pocket of a neighboring SMAD2 within the so-called MH2 domain. Interestingly, the same region of the MH2 domain that mediates binding to the neighboring phosphorylated SMAD2 tail also mediates binding to the phosphorylated receptor, revealing the mechanism by which the phosphorylated SMAD2 switches from receptor binding to trimerization. A similar approach was applied to SMAD3, allowing the structural basis for heterotrimeric SMAD signaling to be elucidated.[30] The contribution of the phosphorylation of the individual C-terminal serines of SMAD2 to trimerization and activation was also studied using EPL.[31] It was shown that while serine 465 phosphorylation was sufficient to drive homotrimerization, two phosphates were required to stabilize the trimer when challenged with SARA, a factor that retains SMADs at the membrane. Furthermore, serine 465 phosphorylation activated the rate of phosphorylation of serine 467. Interestingly, this study represented the first use of a semisynthetic enzyme (TβR-I) on a semisynthetic substrate (SMAD2), both prepared by ligation techniques. Thus, bis-phosphorylation of the C-terminus of R-SMADs such as SMAD2 or SMAD3 induces stable trimerization of the protein.

Since SMAD trimerization is believed to drive nuclear localization and hence transcriptional activity, we developed approaches to control the oligomerization and function of SMADs based on photocaging. In one strategy, the C-terminal phosphates were caged using 2-nitrophenylethyl protecting groups, which could be removed using UV light.[32] An alternate approach was also developed, in which the C-terminus of the protein was blocked with a photocleavable linker, which was fused to a dabcyl quenching group (Figure 2b). Since the C-terminal carboxylate is also essential for intersubunit association, the photolinker blocks trimerization. Methion-

ine 466 of the protein was also replaced with a fluorescein−lysine derivative. Thus, the fluorescein remained quenched by the dabcyl group until the photolinker was cleaved by UV light, simultaneously activating protein activity (i.e., trimerization) and fluorescence.[23] An extension of this strategy demonstrated that both the phosphorylated and nonphosphorylated SMAD2 derivatives, labeled with two different fluorophores, could simultaneously be activated and imaged when microinjected into cells.[33] Collectively, these data support a model in which R-SMADs such as SMAD2 are retained in the cytoplasm by interaction with proteins such as SARA. Upon ligand binding and tetramerization of the receptor complex, TβR-I becomes multiply phosphorylated on the GS segment, leading to recruitment and phosphorylation of R-SMADs. Upon double phosphorylation of the R-SMAD on the two C-terminal serine residues, it can be released from SARA, homo- or heterotrimerize with co-SMADs, and subsequently translocate to the nucleus.

Structural Analysis of Rab GTPases

Rab GTPases are members of the class of monomeric GTPases that play critical roles in various membrane trafficking events inside cells. Like other G-proteins, Rabs are activated by binding to GTP, inactivated by hydrolysis of GTP to GDP and P_i, and anchored to membranes by post-translational modification with lipids at their C-termini.[34] Lipid modifications are particularly challenging to incorporate into proteins due to problems of solubility and stability of the resulting conjugates in aqueous media. Native Rab7 has two geranylgeranyl groups at its C-terminus, and this GTPase has been extensively studied using proteins produced by EPL. The development of effective ligation conditions and synthetic approaches to these lipid-modified proteins is a noteworthy achievement.[35,36]

The large-scale preparation of these lipidated proteins allowed for the crystal structure of the singly geranylgeranylated Rab/Ypt protein family member Ypt1 to be solved in complex with Rab GDP dissociation inhibitor (RabGDI, Figure 3a).[37] GDIs extract the lipid-modified GDP−Rabs from membranes and facilitate their movement to acceptor membranes, thereby regulating the subcellular localization of their GDP bound forms. This structure revealed that the geranylgeranyl group of Ypt1 was buried deep within a channel in RabGDI, explaining the molecular basis of membrane extraction (Figure 3a). Furthermore, the structural basis of the inhibition of GDP release by RabGDI was revealed, because several residues of RabGDI make close contacts with the region of Ypt1 that binds to the phosphates of GDP. However, this study did not suggest where the second geranylgeranyl moiety of Ypt1
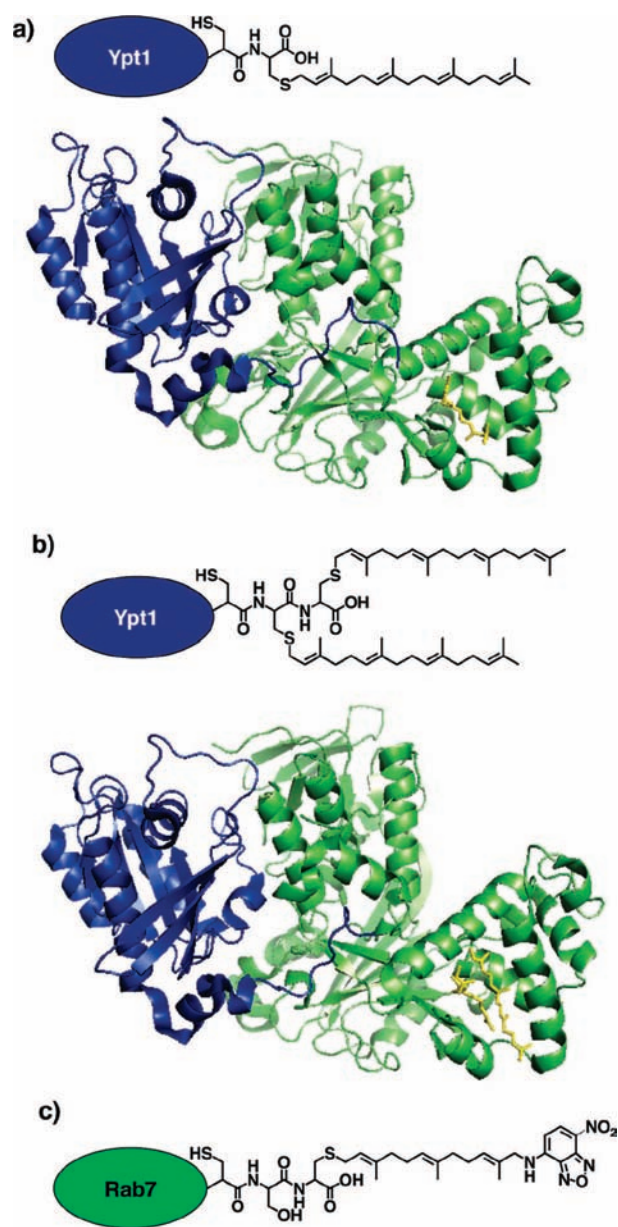

FIGURE 3. (a) Schematic of monoprenylated Ypt1 and crystal structure of monoprenylated Ypt1 (blue) in complex with RabGDI (green). The geranylgeranyl group is indicated in yellow. (b) Schematic of doubly prenylated Ypt1 and crystal structure of doubly prenylated Ypt1 (blue) in complex with RabGDI (green). The geranylgeranyl groups are indicated in yellow. (c) Schematic of Rab7-NBD-farnesyl.

would bind to RabGDI, and thus the structure of doubly geranylgeranylated Ytp1 was subsequently solved in complex with RabGDI (Figure 3b).[38] The structure revealed that one of the geranylgeranyl groups was buried in RabGDI, as in the monolipidated form, while the other was solvent exposed, lying along the protein surface. Importantly, the conclusions drawn from structural data were confirmed with biochemical assays, using a soluble, fluorescent, farnesylated mimic of Rab7, termed Rab7-NBD-farnesyl (Figure 3c).[39] By competitive titra-
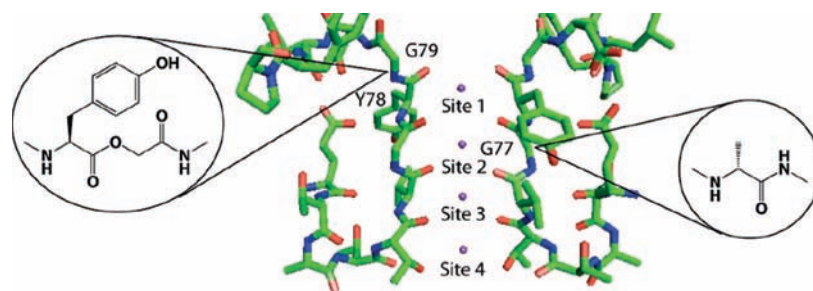
FIGURE 4. Structure of the selectivity filter of the potassium channel KcsA, with semisynthetic variants discussed in the text indicated (PDB ID 1K4C). For clarity, only two opposite monomers of the tetramer are shown.

tion of the Rab7-NBD-farnesyl with natively prenylated Rab variants, it was discovered that GDI binds both mono- and diprenylated forms equally well ($K_d = 1-5$ nM), while it has poor affinity for unprenylated Rab molecules. These experiments provide the structural and biochemical basis for GDI's role in membrane extraction of GDP bound geranylgeranylated Rabs.

Analysis of Potassium Channel Structure and Function

Potassium channels are integral membrane proteins, which permit the rapid and selective conduction of potassium ions across biological membranes. The crystal structure of the potassium channel KcsA, from *Streptomyces lividans*, revealed that the selectivity filter of the homotetrameric assembly was composed of a narrow, 12 Å channel lined with backbone carbonyl oxygen atoms (Figure 4).[40] Since ion selectivity is predominantly mediated by these carbonyl groups rather than amino acid side chains, standard site-directed mutagenesis cannot easily be used to systematically and directly perturb and study this aspect of channel function. In an attempt to further understand, the mechanisms underlying potassium channel function, we developed an EPL-based protocol to generate folded tetrameric potassium channels.[41,42] Since the entirety of the selectivity filter is contained within the synthetic portion of the channel, this method allows for the introduction of various unnatural amino acids into this segment.

In low potassium concentrations, the selectivity filter of KcsA collapses to a nonconductive or closed conformation.[43] The transition from this closed state to the conductive conformation occurs in response to high [K+] but not high [Na+].[44] However, it was unclear whether this structural adaptation was the sole basis of selectivity. By locking the filter in the conductive state, we thought it might be possible to address this important question. Inspection of the crystal structures revealed that Gly77 was in a left-handed helical conformation and that introduction of an additional steric bulk at this position might disfavor this closed conformation. Since the

left-handed helical conformation is favored for D-amino acids and glycine but not L-amino acids, a KcsA analog with Gly77 replaced by D-alanine was prepared.[45] Crystallographic analysis indicated that in the presence of high [K+], the D-alanine had no effect on the structure of the filter, which was anticipated. More importantly and consistent with the design, the filter of the D-alanine containing channel remained in the conductive conformation even in low [K+], which is in stark contrast to wild-type channels.[46] Remarkably, this analog was able to conduct Na+ in the absence of K+. However, the channel was still completely selective for K+ in the presence of Na+, which was unexpected. Thus, the ability to of the channel to adapt its structure in the presence of potassium or sodium ions is an important component of ion selectivity, as is the ability of multiple potassium ions to compete with sodium ions for binding in the conductive filter.

Structural data indicate that there are four contiguous K+ binding sites in the selectivity filter with each ion coordinated by eight oxygen atoms (mostly from main-chain carbonyls) in a square antiprismatic arrangement (Figure 4). However, only two potassium ions on average occupy the selectivity filter at any given time. Two discrete binding configurations, when sites 1 and 3 are occupied and when sites 2 and 4 are occupied, are believed to underlie ion conduction. Maximal ion conduction is predicted to occur when these two configurations are energetically similar. This model predicts that a perturbation at a site should be felt at its coupled site (e.g., sites 1 and 3) and increasing or decreasing the energy of potassium ion coordination at any site should reduce potassium current. To test this idea, a KcsA analog with an ester bond in place of the amide between Tyr78 and Gly79 was prepared.[47] This subtle replacement is nearly isosteric but was expected to result in the reduction in the electronegativity of the carbonyl oxygen by about one-half. Since the carbonyl of Tyr78 contributes to site 1 of the selectivity filter, this mutation would be expected to perturb ion binding in this part of the channel and thus coupled ion conduction. The crystal structure of this modified channel was solved under a variety of

conditions. The isosteric ester substitution had no effect on the structure of the backbone of the selectivity filter but as anticipated did result in a significant reduction of ion density at site 1. Importantly, K^+ current through this channel was reduced in a manner fully consistent with the conduction model. Thus, even slight perturbations in ion occupancy at site 1 result in the alteration of current to suboptimal levels. It is likely that further substitutions within the selectivity filter or in other positions in KcsA will allow for greater understanding of the basis of ion selectivity and conduction in this channel.

Semisynthetic Histones As Tools for Understanding Chromatin Biology

Eukaryotic DNA is organized into chromatin in the nucleus of cells. The fundamental unit of chromatin is the nucleosome, a protein−DNA particle composed of 147 base pairs of DNA wrapped around an octameric protein complex composed of two copies each of the four core histones, H3, H4, H2A, and H2B. The core histones have flexible N-terminal tails, which are heavily post-translationally modified. These modifications regulate both the structure and function of chromatin in transcription, replication, repair, and condensation. Both the position and chemical identity of the PTMs control its specific function, and many enzymes have been identified that are responsible for installing and removing modifications.[48] In order to assess the biochemical function of specific histone modifications, ideally one would want to use chemically defined nucleosomes harboring the modified histones of interest. However, as is the case with many PTMs, these are difficult to obtain by standard methods.

Peptide chemistry has long been used to study the biochemical effects of PTMs on histone tail peptides.[49] Recently, several groups have used ligation protocols to prepare homogeneously modified, full-length histones. In a pioneering study, Shogren-Knaak et al. prepared an H3 variant with a phosphoserine at position 10 and incorporated it into nucleosomal arrays.[50] The histone acetyltransferase Gcn5 had increased activity upon the modified nucleosomes, while the Gcn5-containing SAGA complex did not. The first result was consistent with peptide substrate studies, while the latter was not, demonstrating the advantage of studying the role of histone-modifying enzymes on intact substrates. By a similar method, a site specifically Lys16-acetylated H4 derivative was prepared. The authors incorporated this derivative in nucleosomal arrays and analyzed the resulting complexes by ultracentrifugation and in a nucleosome sliding assay.[51] The acetylated nucleosomes had a reduced propensity to form 30-nm-like fibers, a higher order chromatin structure. Further-
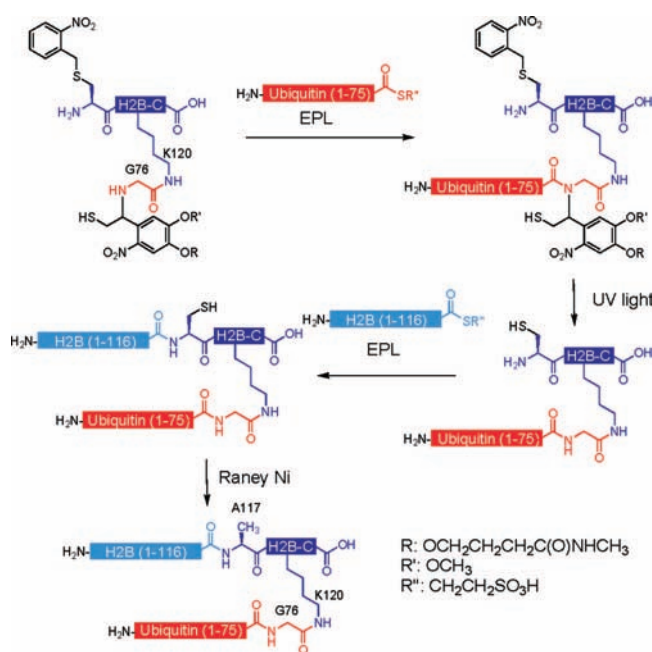

FIGURE 5. Synthetic strategy to generate ubiquitylated H2B. Ubiquitylated H2B peptide was prepared by ligation to a photocleavable auxiliary containing peptide. After simultaneous removal of the ligation auxiliary and deprotection of the N-terminal cysteine using UV light, the ubiquitylated peptide was reacted with the H2B(1−116) thioester. Finally, C117 was desulfurized with Raney nickel to produce intact ubiquitylated H2B.

more, the modified nucleosomes moved more slowly compared with unmodified nucleosomes in the sliding assay when treated with the chromatin remodeling complex, ACF. These data indicate that acetylation of H4 at position 16 promotes the formation of a decondensed chromatin structure, which may favor transcription. He et al. extended the semisynthetic histone approach by adding a desulfurization reaction to the synthetic procedure, thereby converting the cysteine to an alanine, allowing for a traceless ligation.[52] Using this approach, the authors generated acetylated and methylated H3 and acetylated H4 and demonstrated that they were substrates for histone-modifying enzymes.

In our own laboratory, a semisynthetic approach for generating ubiquitylated histones was developed based on the peptide ubiquitylation approach of Chatterjee et al.[53] This traceless peptide ubiquitylation approach makes use of a photocleavable ligation auxiliary, which is coupled to the side chain of a lysine residue. The ligation auxiliary can thus act as an N-terminal cysteine surrogate to react in an EPL reaction with a recombinant ubiquitin thioester. The ligation auxiliary could be removed with UV light, generating a natively ubiquitylated peptide. A similar approach was used to prepare sumoylated peptides. Recently, we extended this approach to full-length H2B by using a two-step EPL procedure (Figure 5).[54]

The ubiquitylated H2B was incorporated into nucleosomes, and it was demonstrated that this PTM activates the nucleosome for methylation on H3 Lys79 by the methyltransferase, Dot1. This work represents the first direct biochemical evidence of crosstalk between PTMs on different histones. It is likely that EPL and NCL approaches will continue to be used to interrogate the role of histone modifications in chromatin biology. For example, differentially modified histones could be incorporated into single nucleosomes to study cross-talk between modifications, and other modifications such as sumoylation, ADP ribosylation, or deimination could readily be studied using this approach.

Outlook and the Future of EPL

The studies reviewed herein hopefully provide a clear description of how EPL can be used to unravel the function of specific PTMs and how it can be used for detailed biophysical and structural studies of proteins. The fundamental strength of the technique is that it allows for the preparation of homogeneous semisynthetic proteins on a scale sufficient for most routine studies of protein structure and function. In principle, nearly any modification, or combination of modifications can be included due to the mild nature of the reaction conditions, as long as the final product is itself stable. Furthermore, careful optimization of the ligation conditions allows for the preparation of challenging targets, such as the potassium channel KcsA or lipidated Rab molecules (see above). Thus, EPL is a versatile technique for protein engineering and semisynthesis.

Several technical limitations in the application of EPL still remain. First, the modifications that are incorporated must be within 50 amino acids from the termini of the protein, or else multiple ligation steps are required. While this is possible, it increases the complexity of the synthesis substantially.[54,55] The second principal limitation is the requirement for a cysteine at the ligation junction. Progress has been made in this area by using auxiliaries or desulfurization reactions, allowing for ligation at phenylalanine, alanine, and glycine.[56−59] It is likely that with the development of additional ligation junctions, virtually any protein will be amenable to EPL modification. The final limitation is the difficulty of performing this reaction *in vivo*, in contrast to some other labeling technologies.[60] While EPL is unlikely to be re-engineered to proceed *in vivo*, due to side reactions of the thioester or N-terminal cysteine with endogenous metabolites, related intein-based technologies such as protein trans-splicing[61] are promising approaches for generating the types of complex modifications described in this Account in living systems.

EPL shows promise for extension into several novel areas. First, we predict that EPL will be increasingly used in the area of nanotechnology. EPL is a versatile technique for attaching proteins site-specifically to nanoparticles and surfaces.[17,20] In the next several years, it is likely that this technique will be used to generate more sophisticated organic/inorganic hybrids for various applications. A second area in which EPL will prove increasingly useful is in the generation of protein therapeutics. NCL has already been used to produce complex protein therapeutics.[62] EPL provides an important advantage over this approach, namely, that the cost of the generation of recombinant protein fragments is often lower than synthetic peptides. Third, we believe that this approach will be used for the study of increasingly complex PTMs, for example, GPI anchors,[63,64] or proteins with multiple modifications. Finally, EPL will continue to be applied to increasingly complex problems in the structural, biochemical, and biophysical analysis of proteins.

BIOGRAPHICAL INFORMATION

Robert R. Flavell received his bachelor's degree in mathematics and chemistry from Wesleyan University in Connecticut. Currently he is a student of the Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional M.D.−Ph.D. Program, pursuing his Ph.D. in the laboratory of Dr. Tom W. Muir.

Tom W. Muir received his B.Sc. (Hons, 1st class) in Chemistry from the University of Edinburgh in 1989 and his Ph.D. in Chemistry from the same institute in 1993 under the direction of Professor Robert Ramage. After postdoctoral studies with Stephen B. H. Kent at the Scripps Research Institute, he joined the faculty at the Rockefeller University in New York City in 1996, where he is now Richard E. Salomon Family Professor and Director of the Pels Center of Chemistry, Biochemistry and Structural Biology.

FOOTNOTES

*To whom correspondence should be addressed. E-mail: muirt@mail.rockefeller.edu.

REFERENCES

1   Muir, T. W. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **2003**, *72*, 249–289.

2   Muralidharan, V.; Muir, T. W. Protein ligation: An enabling technology for the biophysical analysis of proteins. *Nat. Methods* **2006**, *3*, 429–438.

3   Hortin, G.; Boime, I. Applications of amino acid analogs for studying co- and posttranslational modifications of proteins. *Methods Enzymol.* **1983**, *96*, 777–784.

4   Wang, L.; Xie, J.; Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.

5   Hermanson, G. *Bioconjugate Techniques*; Academic Press: San Diego, CA, 1996.

6   *Solid Phase Peptide Synthesis*; Fields, G. B., Ed.; Methods in Enzymology, Vol. 289; Academic Press: San Diego, CA, 1997.

7   Wieland, T.; Bokelmann, E.; Bauer, L.; Lang, H. U.; Lau, H. Bildung von S-haltigen Peptiden durch intramolekulare Wanderung von Aminoacylresten. *Liebigs. Ann. Chem.* **1953**, *583*, 129–149.

8   Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266*, 776–779.

9   Dawson, P. E.; Kent, S. B. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.

10 Erlanson, D. A.; Chytil, M.; Verdine, G. L. The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex. *Chem. Biol.* **1996**, *3*, 981–991.

11 Giriat, I.; Muir, T. W.; Perler, F. B. Protein splicing and its applications. *Genet. Eng. (N Y)* **2001**, *23*, 171–199.

12 Muir, T. W.; Sondhi, D.; Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705–6710.

13 Severinov, K.; Muir, T. W. Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. *J. Biol. Chem.* **1998**, *273*, 16205–16209.

14 Evans, T. C., Jr.; Benner, J.; Xu, M. Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci.* **1998**, *7*, 2256–2264.

15 Hooker, J. M.; Esser-Kahn, A. P.; Francis, M. B. Modification of aniline containing proteins using an oxidative coupling strategy. *J. Am. Chem. Soc.* **2006**, *128*, 15558–15559.

16 Lin, P. C.; Ueng, S. H.; Tseng, M. C.; Ko, J. L.; Huang, K. T.; Yu, S. C.; Adak, A. K.; Chen, Y. J.; Lin, C. C. Site-specific protein modification through Cu-I-catalyzed 1,2,3-triazole formation and its implementation in protein microarray fabrication. *Angew. Chem., Int. Ed.* **2006**, *45*, 4286–4290.

17 Camarero, J. A.; Kwon, Y.; Coleman, M. A. Chemoselective attachment of biologically active proteins to surfaces by expressed protein ligation and its application for "protein chip" fabrication. *J. Am. Chem. Soc.* **2004**, *126*, 14730–14731.

18 Marsac, Y.; Cramer, J.; Olschewski, D.; Alexandrov, K.; Becker, C. F. W. Site-specific attachment of polyethylene glycol-like oligomers to proteins and peptides. *Bioconjugate Chem.* **2006**, *17*, 1492–1498.

19 Burbulis, I.; Yamaguchi, K.; Gordon, A.; Carlson, R.; Brent, R. Using protein-DNA chimeras to detect and count small numbers of molecules. *Nat. Methods* **2005**, *2*, 31–37.

20 Becker, C. F. W.; Marsac, Y.; Hazarika, P.; Moser, J.; Goody, R. S.; Niemeyer, C. M. Functional immobilization of the small GTPase Rab6A on DNA-gold nanoparticles by using a site-specifically attached poly(ethylene glycol) linker and thiol place-exchange reaction. *ChemBioChem* **2007**, *8*, 32–36.

21 van Baal, I.; Malda, H.; Synowsky, S. A.; van Dongen, J. L.; Hackeng, T. M.; Merkx, M.; Meijer, E. W. Multivalent peptide and protein dendrimers using native chemical ligation. *Angew. Chem., Int. Ed.* **2005**, *44*, 5052–5057.

22 Flavell, R. R.; Kothari, P.; Bar-Dagan, M.; Synan, M.; Vallabhajosula, S.; Friedman, J. M.; Muir, T. W.; Ceccarini, G. Site-specific (18)F-labeling of the protein hormone leptin using a general two-step ligation procedure. *J. Am. Chem. Soc.* **2008**, *130*, 9106–9112.

23 Pellois, J. P.; Hahn, M. E.; Muir, T. W. Simultaneous triggering of protein activity, and fluorescence. *J. Am. Chem. Soc.* **2004**, *126*, 7170–7171.

24 Lee, Y. J.; Datta, S.; Pellois, J. P. Real-time fluorescence detection of protein transduction into live cells. *J. Am. Chem. Soc.* **2008**, *130*, 2398–2399.

25 Shi, Y.; Massague, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **2003**, *113*, 685–700.

26 Huse, M.; Holford, M. N.; Kuriyan, J.; Muir, T. W. Semisynthesis of hyperphosphorylated type I TGF beta receptor: Addressing the mechanism of kinase activation. *J. Am. Chem. Soc.* **2000**, *122*, 8337–8338.

27 Flavell, R. R.; Huse, M.; Goger, M.; Trester-Zedlitz, M.; Kuriyan, J.; Muir, T. W. Efficient semisynthesis of a tetraphosphorylated analogue of the Type I TGFbeta receptor. *Org. Lett.* **2002**, *4*, 165–168.

28 Huse, M.; Muir, T. W.; Xu, L.; Chen, Y. G.; Kuriyan, J.; Massague, J. The TGF beta receptor activation process: An inhibitor- to substrate-binding switch. *Mol. Cell* **2001**, *8*, 671–682.

29 Wu, J. W.; Hu, M.; Chai, J.; Seoane, J.; Huse, M.; Li, C.; Rigotti, D. J.; Kyin, S.; Muir, T. W.; Fairman, R.; Massague, J.; Shi, Y. Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling. *Mol. Cell* **2001**, *8*, 1277–1289.

30 Chacko, B. M.; Qin, B. Y.; Tiwari, A.; Shi, G.; Lam, S.; Hayward, L. J.; De Caestecker, M.; Lin, K. Structural basis of heteromeric smad protein assembly in TGF-beta signaling. *Mol. Cell* **2004**, *15*, 813–823.

31 Ottesen, J. J.; Huse, M.; Sekedat, M. D.; Muir, T. W. Semisynthesis of phosphovariants of Smad2 reveals a substrate preference of the activated T beta RI kinase. *Biochemistry* **2004**, *43*, 5698–5706.

32 Hahn, M. E.; Muir, T. W. Photocontrol of Smad2, a multiphosphorylated cell-signaling protein, through caging of activating phosphoserines. *Angew. Chem., Int. Ed.* **2004**, *43*, 5800–5803.

33 Hahn, M. E.; Pellois, J. P.; Vila-Perello, M.; Muir, T. W. Tunable photoactivation of a post-translationally modified signaling protein and its unmodified counterpart in live cells. *ChemBioChem* **2007**, *8*, 2100–2105.

34 Brunsveld, L.; Kuhlmann, J.; Alexandrov, K.; Wittinghofer, A.; Goody, R. S.; Waldmann, H. Lipidated Ras and Rab peptides and proteins - Synthesis, structure, and function. *Angew. Chem., Int. Ed.* **2006**, *45*, 6622–6646.

35 Durek, T.; Alexandrov, K.; Goody, R. S.; Hildebrand, A.; Heinemann, I.; Waldmann, H. Synthesis of fluorescently labeled mono- and diprenylated Rab7 GTPase. *J. Am. Chem. Soc.* **2004**, *126*, 16368–16378.

36 Brunsveld, L.; Watzke, A.; Durek, T.; Alexandrov, K.; Goody, R. S.; Waldmann, H. Synthesis of functionalized rab GTPases by a combination of solution- or solid-phase lipopeptide synthesis with expressed protein ligation. *Chemistry* **2005**, *11*, 2756–2772.

37 Rak, A.; Pylypenko, O.; Durek, T.; Watzke, A.; Kushnir, S.; Brunsveld, L.; Waldmann, H.; Goody, R. S.; Alexandrov, K. Structure of Rab GDP-dissociation inhibitor in complex with prenylated YPT1 GTPase. *Science* **2003**, *302*, 646–650.

38 Pylypenko, O.; Rak, A.; Durek, T.; Kushnir, S.; Dursina, B. E.; Thomae, N. H.; Constantinescu, A. T.; Brunsveld, L.; Watzke, A.; Waldmann, H.; Goody, R. S.; Alexandrov, K. Structure of doubly prenylated Ypt1: GDI complex and the mechanism of GDI-mediated Rab recycling. *EMBO J.* **2006**, *25*, 13–23.

39 Wu, Y. W.; Tan, K. T.; Waldmann, H.; Goody, R. S.; Alexandrov, K. Interaction analysis of prenylated Rab GTPase with Rab escort protein and GDP dissociation inhibitor explains the need for both regulators. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12294–12299.

40 Zhou, Y.; Morais-Cabral, J. H.; Kaufman, A.; MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 A resolution. *Nature* **2001**, *414*, 43–48.

41 Valiyaveetil, F. I.; MacKinnon, R.; Muir, T. W. Semisynthesis and folding of the potassium channel KcsA. *J. Am. Chem. Soc.* **2002**, *124*, 9113–9120.

42 Valiyaveetil, F. I.; Sekedat, M.; Muir, T. W.; MacKinnon, R. Semisynthesis of a functional K+ channel. *Angew. Chem., Int. Ed.* **2004**, *43*, 2504–2507.

43 Zhou, Y.; MacKinnon, R. The occupancy of ions in the K+ selectivity filter: charge balance and coupling of ion binding to a protein conformational change underlie high conduction rates. *J. Mol. Biol.* **2003**, *333*, 965–975.

44 Lockless, S. W.; Zhou, M.; MacKinnon, R. Structural and thermodynamic properties of selective ion binding in a K+ channel. *PLoS Biol.* **2007**, *5*, e121.

45 Valiyaveetil, F. I.; Sekedat, M.; MacKinnon, R.; Muir, T. W. Glycine as a D-amino acid surrogate in the K+-selectivity filter. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17045–17049.

46 Valiyaveetil, F. I.; Leonetti, M.; Muir, T. W.; Mackinnon, R. Ion selectivity in a semisynthetic K+ channel locked in the conductive conformation. *Science* **2006**, *314*, 1004–1007.

47 Valiyaveetil, F. I.; Sekedat, M.; MacKinnon, R.; Muir, T. W. Structural and functional consequences of an amide-to-ester substitution in the selectivity filter of a potassium channel. *J. Am. Chem. Soc.* **2006**, *128*, 11591–11599.

48 Kouzarides, T. Chromatin modifications and their function. *Cell* **2007**, *128*, 693–705.

49 Krieger, D. E.; Levine, R.; Merrifield, R. B.; Vidali, G.; Allfrey, V. G. Chemical studies of histone acetylation. Substrate specificity of a histone deacetylase from calf thymus nuclei. *J. Biol. Chem.* **1974**, *249*, 332–334.

50 Shogren-Knaak, M. A.; Fry, C. J.; Peterson, C. L. A native peptide ligation strategy for deciphering nucleosomal histone modifications. *J. Biol. Chem.* **2003**, *278*, 15744–15748.

51 Shogren-Knaak, M.; Ishii, H.; Sun, J. M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **2006**, *311*, 844–847.

52 He, S.; Bauman, D.; Davis, J. S.; Loyola, A.; Nishioka, K.; Gronlund, J. L.; Reinberg, D.; Meng, F.; Kelleher, N.; McCafferty, D. G. Facile synthesis of site-specifically acetylated and methylated histone proteins: reagents for evaluation of the histone code hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12033–12038.

53 Chatterjee, C.; McGinty, R. K.; Pellois, J. P.; Muir, T. W. Auxiliary-mediated site-specific peptide ubiquitylation. *Angew. Chem., Int. Ed.* **2007**, *46*, 2814–2818.

54 McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **2008**, *453*, 812–816.

55 Cotton, G. J.; Ayers, B.; Xu, R.; Muir, T. W. Insertion of a synthetic peptide into a recombinant protein framework: A protein biosensor. *J. Am. Chem. Soc.* **1999**, *121*, 1100–1101.

56 Yan, L. Z.; Dawson, P. E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **2001**, *123*, 526–533.

57 Marinzi, C.; Offer, J.; Longhi, R.; Dawson, P. E. An o-nitrobenzyl scaffold for peptide ligation: synthesis and applications. *Bioorg. Med. Chem.* **2004**, *12*, 2749–2757.

58 Crich, D.; Banerjee, A. Native chemical ligation at phenylalanine. *J. Am. Chem. Soc.* **2007**, *129*, 10064–10065.

59 Wan, Q.; Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew. Chem., Int. Ed.* **2007**, *46*, 9248–9252.

60 Chen, I.; Ting, A. Y. Site-specific labeling of proteins with small molecules in live cells. *Curr. Opin. Biotechnol.* **2005**, *16*, 35–40.

61 Giriat, I.; Muir, T. W. Protein semi-synthesis in living cells. *J. Am. Chem. Soc.* **2003**, *125*, 7180–7181.

62 Kochendoerfer, G. G.; Chen, S. Y.; Mao, F.; Cressman, S.; Traviglia, S.; Shao, H.; Hunter, C. L.; Low, D. W.; Cagle, E. N.; Carnevali, M.; Gueriguian, V.; Keogh, P. J.; Porter, H.; Stratton, S. M.; Wiedeke, M. C.; Wilken, J.; Tang, J.; Levy, J. J.; Miranda, L. P.; Crnogorac, M. M.; Kalbag, S.; Botti, P.; Schindler-Horvat, J.; Savatski, L.; Adamson, J. W.; Kung, A.; Kent, S. B.; Bradburne, J. A. Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* **2003**, *299*, 884–887.

63 Paulick, M. G.; Wise, A. R.; Forstner, M. B.; Groves, J. T.; Bertozzi, C. R. Synthetic analogues of glycosylphosphatidylinositol-anchored proteins and their behavior in supported lipid bilayers. *J. Am. Chem. Soc.* **2007**, *129*, 11543–11550.

64 Lschewski, D.; Seidel, R.; Miesbauer, M.; Rambold, A. S.; Oesterhelt, D.; Winklhofer, K. F.; Tatzelt, J.; Engelhard, M.; Becker, C. F. W. Semisynthetic murine prion protein equipped with a GPI anchor mimic incorporates into cellular membranes. *Chem. Biol.* **2007**, *14*, 994–1006.