

Chemical Synthesis and Biological Function of Lipidated Proteins

Aimin Yang[§], Lei Zhao[§], and Yao-Wen Wu

Abstract Lipidated proteins play a key role in many essential biological processes in eukaryotic cells, including signal transduction, membrane trafficking, immune response and pathology. The investigation of the function of lipidated proteins requires access to a reasonable amount of homogenous lipid-modified proteins with defined structures and functional groups. Chemical approaches have provided useful tools to perform such studies. In this review we summarize synthetic methods of lipidated peptides and developments in the chemoselective ligation for the production of lipidated proteins. We introduce the biology of lipidated proteins and highlight the application of synthetic lipidated proteins to tackle important biological questions.

Keywords Click ligation · Diels–Alder ligation · Expressed protein ligation · GPI anchor · MIC ligation · Peptide synthesis · Protein lipidation · Rab · Ras · Rheb · Sortase-mediated protein ligation

Contents

1	Introduction	138
1.1	<i>N</i> -Myristoylation	141
1.2	Palmitoylation	142
1.3	Prenylation	143
1.4	GPI-Anchor Addition	144
1.5	Other Types of Lipidation	145
2	Synthesis of Lipidated Peptides	145
2.1	Preparation of Lipidated Cysteine Building Blocks	147

[§]Both authors contributed equally to this work.

2.2	Solution-Phase Synthesis of Lipidated Peptides	148
2.3	Solid-Phase Approach for the Synthesis of Lipidated Peptides	148
2.4	Synthesis of Lipidated Peptides by Combined Solution/Solid-Phase Approach	152
3	Synthesis of Lipidated Proteins	154
3.1	Assisted Solubilisation Strategy	155
3.2	Expressed Protein Ligation	158
3.3	MIC Ligation	161
3.4	Diels–Alder Ligation	163
3.5	Click Ligation	163
3.6	Sortase-Mediated Protein Ligation	164
4	Chemical Biology of Lipidated Protein	166
4.1	Cell Biological Studies of <i>S</i> -Palmitoylation Cycle of Ras GTPases	166
4.2	Biophysical Studies of Lipidated Ras GTPases	169
4.3	Structural Studies of Prenylated Rheb GTPases	170
4.4	Thermodynamic Basis of Rab GTPases Membrane Targeting	172
4.5	Biological Function of GPI-Anchors	173
4.6	Function of LC3-PE in Autophagosome Formation	176
5	Conclusions and Perspectives	178
	References	179




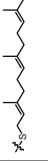
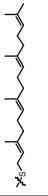

1 Introduction

A biological membranes system is typically formed by the combination of lipids and proteins. In eukaryotic cells, the plasma membrane, also referred to as the cell membrane, is a protective barrier which regulates what enters and leaves the cell. The endomembrane system is composed of different kinds of membranes which divide the cell into structural and functional compartments within a eukaryotic cell, such as the endoplasmic reticulum, Golgi apparatus, mitochondria, endosome and lysosome. Covalent modification of proteins with lipid anchors (protein lipidation) facilitates association of the lipidated proteins with particular membranes in eukaryotic cells. Protein lipidation is one of the most important protein post-translational modifications (PTMs). Studying lipidated protein function *in vitro* or *in vivo* is of vital importance in biological research.

A variety of lipids serve as lipid anchors attached to proteins, including fatty acids, isoprenoids, glycosylphosphatidylinositol (GPI) and cholesterol. Protein lipidation is not only essential for binding to membranes, but also for the protein–protein interactions and the regulation of the signalling process [1]. Therefore, lipid modification plays a critical role in the function and localization of proteins. So far, recombinant production of lipidated proteins has not been very successful and is particularly challenging in terms of homogeneity and output. In this review, we discuss the chemical synthesis of various lipidated proteins. We show a few examples of using synthetic lipidated proteins to elucidate their biological functions.

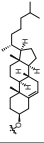
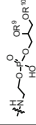
The four major types of protein lipidation are *N*-myristoylation, palmitoylation, prenylation and glycosylphosphatidylinositol-anchor (GPI-anchor) addition (Table 1).

Table 1 Types, properties and functions of different lipidations

Types of lipidations	Chemical structures of lipids modifications	Biological lipidated processes catalysed by	Modification sites	Reversibility	Cell localizations of lipidated proteins	Protein examples
<i>N</i> -Myristoylation		<i>N</i> -Myristoyl-transferase (NMT)	N-terminal of Gly	No	Mitochondrial membrane, Intracellular plasma membrane	Arf, Actin, PAK2
<i>S</i> -Palmitoylation		Palmitoyl Protein Thioesterases (PPT)	Side chain of Cys	Yes	Intracellular plasma membrane, Vesicle	H-Ras, Gs α
<i>O</i> -Palmitoylation		<i>O</i> -Acyltransferase	Side chain of Ser/Thr	n.d.	Extracellular plasma membrane	Wnt3a
Farnesylation		Farnesyl Transferase (FTase)	Side chain of Cys	No	Intracellular plasma membrane, vesicle, Golgi	Ras, Rho
Geranylgeranylation		Geranylgeranyl Transferase I (GGTase I)	Side chain of Cys	No	Intracellular plasma membrane, Golgi, endosome, lysosome, vesicle, other endomembranes	Rho, Rab, Ras
GPI-anchors addition		GPI Transamidase	C-terminal of Ala/Asn/ Asp/Cys/ Gly/Ser	Yes	Extracellular plasma membrane	Hamster brain PrP ^{Sc} , human CD52, human erythrocyte CD59

(continued)

Table 1 (continued)

Types of lipidations	Chemical structures of lipids modifications	Biological lipidated processes catalysed by	Modification sites	Reversibility	Cell localizations of lipidated proteins	Protein examples
Cholesterylation		Protein C-terminal domain undergoes an intertein like process, and acts as an intramolecular cholesterol transferase	C-terminal of Gly	n.d.	Extracellular plasma membrane	Hedgehog
Phosphatidylethanolaminylation		Atg7, Atg3, Atg12-Atg5 complex	C-terminal of Gly	Yes	Autophagosome, phagophore	Atg8/LC3

$R^{1-4} = \pm$ saccharide(s), $R^{5-7} = \pm$ phosphate(s), $R^8 = \pm$ fatty acid, Lipid = 1-alkyl-2-acylglycerol/1-alkyl-2-lysoglycerol/ceramide/diacylglycerol, $R^{9,10} =$ diacyl, alkylacyl, alkenylacyl
n.d.: not determined

1.1 *N*-Myristoylation

N-Myristoylation is an irreversible protein modification where a myristoyl group, a 14-carbon saturated fatty acid, is covalently attached via an amide bond to the N-terminal glycine residue. This type of protein modification was firstly identified as an “N-terminal blocking group” [2, 3]. In eukaryotic cells, *N*-myristoylation is mediated by the enzyme *N*-myristoyltransferase (NMT) which transfers the acyl group from myristoyl CoA to the N-terminal amine of proteins containing N-terminal GXXXS/T sequences [4]. This modification majorly occurs co-translationally and in some cases happens post-translationally. During co-translational modification, the N-terminal glycine is modified following the cleavage of N-terminal methionine residue by methionine aminopeptidases. Post-translational myristoylation typically occurs after a caspase cleavage, resulting in the exposure of an internal glycine residue. *N*-Myristoylation plays an essential role in protein–protein interactions and membrane targeting of proteins, which are involved in a wide range of signal transduction pathways. *N*-Myristoylation not only occurs on eukaryotic proteins, including Src family tyrosine kinases, Abl tyrosine kinase, cAMP-dependent protein kinase (PKA), α subunits of G proteins, ADP-ribosylation factors (Arfs), and Ca²⁺ sensor proteins (Recoverin, Hippocalcin, Neurocalcins, MARCKS), but also on bacterial and viral proteins, such as HIV-1 Gag [5].

N-Myristoylated proteins can switch between two distinct conformations, one conformation where the myristoyl group is exposed and available to promote membrane binding and the other conformation where the myristoyl moiety is sequestered in a hydrophobic binding pocket and not available for membrane binding. For instance, Arf, a member of the family of GTP-binding proteins of the Ras superfamily, is *N*-myristoylated in cells [6]. Arf functions as regulators of vesicular trafficking and actin remodelling. As the other members in the Ras superfamily, the switch between the inactive GDP-bound form and the active GTP-bound form of Arf GTPase is highly regulated by GTPase-activating proteins (GAPs) which accelerate the intrinsic GTP hydrolysis of GTPases and by guanine nucleotide exchange factors (GEFs) which facilitate exchange of GDP for GTP [7, 8]. In the GDP-bound Arf, the myristoylated N-terminal helix is in a shallow hydrophobic groove formed by loop λ 3. In the GTP-bound form, the extrusion of loop λ 3 from the GTPase core eliminates the binding site for the myristoylated N-terminus, which becomes available for membrane binding [9, 10]. The “myristoyl switch” can be used as a signal regulating cellular localization, membrane association and protein–protein interactions.

Moreover, *N*-myristoylation also plays a critical role in bacterial and viral entry. Although viruses and bacteria usually lack the enzyme NMT required for this modification, their proteins are consequently processed by NMTs of the hosts [11]. During the human immunodeficiency virus-1 (HIV-1) life cycle, an HIV protein, Gag, specifically assembles at the lipid raft region of the host cell membrane. The high concentration of Gags facilitates the viral particle budding. The

N-myristoylation plus basic motif is thought to target Gag to the plasma membrane. Without myristoylation, the host cells do not release any virus particles [12, 13].

1.2 Palmitoylation

Palmitoylation is the covalent attachment of preferred 16-carbon palmitic acid to a cysteine side chain (*S*-palmitoylation) and less frequently to a serine/threonine side chain of proteins (*O*-palmitoylation). Occasionally other long chain fatty acid moieties, including stearoyl (C18), oleoyl (C18:1) and arachidonyl (C20:4) chains, have also been found in acylated proteins. Acylation is therefore a more accurate description of this type of fatty acid modification [14]. The acyl transfer from palmitoyl CoA to the thiolate side chain of cysteine residues is an energy-neutral reaction, which is catalysed by palmitoyl acyl transferases (PATs). So far there is no consensus sequence in protein substrates which undergo *S*-palmitoylation. *S*-Palmitoylated proteins include transmembrane receptors (e.g. TGF α , adrenergic receptors, Rhodopsin), viral proteins (e.g. Influenza HA, HIV-1 gp160), Ras proteins, G α subunits, Src family tyrosine kinases (Src, Lck, Fyn, Hck, Lyn and Yes), etc. The G α subunits and Src family tyrosine kinases are both myristoylated and palmitoylated, and they contain consensus sequence MGC at their N-termini [15]. In contrast to *N*-myristoylation and prenylation, palmitoylation is usually reversible because of the thioester or ester connection [16, 17]. The reverse modification is catalysed by palmitoyl protein thioesterases (PPTs) [18, 19]. Because of the reversibility of palmitoyl modification, palmitoylation is a dynamic post-translational modification to regulate the subcellular localization and protein–protein interactions.

Cycles of palmitoylation and depalmitoylation regulate membrane binding of palmitoylated proteins. For example, the spatial cycle of H/N-Ras between the Golgi and the plasma membrane is dependent on the reversible *S*-palmitoylation at the C-termini [18, 20]. H-Ras and N-Ras are prenylated and palmitoylated. Farnesylation alone does not confer high membrane affinity. Farnesylated Ras molecules are solubilized by PDE δ [21] and rapidly diffuse throughout the cell until they become palmitoylated in the Golgi to acquire additional hydrophobicity and thereby higher affinity to membranes. Palmitoylated Ras are transported on vesicles via the secretory pathway, leading to an enrichment of Ras at the plasma membrane. The palmitoylated Ras at the plasma membrane slowly redistributes to all cellular membranes, where they are ubiquitously depalmitoylated by thioesterases, such as acyl protein thioesterase 1 (APT1). Rapid diffusion of depalmytoylated Ras molecules increases the probability of a Golgi encounter, in which they are repalmitoylated and transported to the plasma membrane. Thus, the acylation cycle maintains the spatial cycle for H- and N-Ras, which confers it with unique signal propagation characteristics [18, 20].

1.3 Prenylation

Prenylation is a posttranslational addition of C15 (farnesyl) or C20 (geranylgeranyl) isoprenyl groups via thioether linkages to the cysteine side chains at the C-termini of the protein substrates. Prenylation is of increasing interest since many prenylated proteins are involved in signal transduction pathways controlling cell growth and differentiation, cytoskeletal rearrangement, and vesicular transport [22]. Although a search of the human proteome revealed about 300 proteins which are potentially prenylated, only a fraction of these have been reported. So far, three protein prenyltransferases responsible for isoprenoid addition to proteins have been identified (for reviews see [23–25]). They can be classified into two categories according to their functions. One is the CaaX prenyltransferases: protein farnesyltransferase (FTase) and protein geranylgeranyltransferase type I (GGTase-I) recognize protein substrates containing a CaaX box (C is cysteine, a is usually an aliphatic amino acid, and X can be one of a variety of amino acids) at their C-termini. The other is Rab geranylgeranyltransferase (RabGGTase), also called protein geranylgeranyltransferase type II (GGTase-II), which mediates the addition of usually two geranylgeranyl groups to the C-terminal cysteines of Rab GTPases.

Substrates for FTase include Ras GTPases, which regulate signal transduction involved in cellular growth; nuclear lamin A and B, which form structural lamina on the inner nuclear membrane; the γ subunit of heterotrimeric G-protein transducin, which functions in visual signal transduction in the retina; the large-antigen component of the hepatitis δ virus; and yeast mating factors. Known targets of GGTase-I include γ subunits of heterotrimeric G-proteins and many small GTPases such as the Rho/Rac family. Prenylation of proteins enables them to associate with endoplasmic reticulum, where they are further modified in subsequent post-prenylation reactions, including proteolytic removal of the last three amino acids of the CaaX motif and subsequent carboxyl methylation [26].

RabGGTase has a very strict substrate preference and acts only on the members of the Rab GTPase family, which play a central role in membrane trafficking in eukaryotic cells [7, 8]. Unlike FTase and GGTase, RabGGTase does not recognize a short C-terminal sequence but requires an additional factor called Rab escort protein (REP) to recruit Rab protein. REP interacts with the unprenylated Rab protein preferentially in its GDP-bound form and mediates its recognition by RabGGTase. Since the specificity is already outsourced from the REP molecule, RabGGTase has essentially no sequence preference for the context of the prenylatable cysteines, and the C-terminal sequences occurring in Rab GTPases include CC, CXC, CCX, CCXX, CCXXX and CXXX [27, 28]. The conjugated prenyl group not only is a mediator of membrane association but also functions as a molecular handle for specific protein–protein interactions, such as interaction with GDP-dissociation inhibitor (GDI) which enables cycling the prenylated Rab proteins between different membranes [29, 30].

1.4 GPI-Anchor Addition

Glycosylphosphatidylinositol (GPI) anchors are found in many cell surface proteins in eukaryotes, which tether them to the extracellular side of the plasma membrane. The GPI anchor is attached to the C-terminus of a protein via a phosphoethanolamine linkage. GPIs and GPI-anchored proteins are built up in the ER and then the modified proteins transit to the cell surface. This posttranslational glycolipid modification is mediated by the GPI transamidase (GPI-T) in the ER lumen. The GPI anchor can be hydrolysed by phosphatidylinositol-specific phospholipase C or D (PLC or PLD) which releases the protein moiety into the extracellular milieu [31]. Almost all GPIs share a common core glycan structure, $\text{NH}_2(\text{CH}_2)_2\text{OPO}_3\text{H}-6\text{Man}\alpha 1 \rightarrow 2\text{Man}\alpha 1 \rightarrow 6\text{Man}\alpha 1 \rightarrow 4\text{GlcN}\alpha 1 \rightarrow 6\text{myo-Ino}1\text{-phospholipid}$ (Table 1). The glycan core can be decorated with several side-chain modifications, and the lipid moiety can vary between diacylglycerol, alkylacylglycerol, ceramide, etc. [32]. The variety of the different compositions results in high structural diversity among GPIs, which makes the studies complicated [33, 34]. There are diverse GPI-anchored proteins displayed at the cell surface, ranging from receptors (e.g. folate receptor, $\text{Fc}\gamma\text{RIII}$, CD14), cell surface antigens (e.g. Thy-1, CD antigens, Campath, LFA-3) to enzymes (e.g. alkaline phosphatase, carbonic anhydrase, dipeptidase). GPI-anchored proteins play vital roles in immune response, transmembrane signal transduction, cell contacts and migration, pathology of parasites, and oncogenesis [35].

For example, the glycopospholipid facilitates the protein lateral diffusion on the cell surface [36]. GPI-anchored proteins exhibit greater mobility than the transmembrane proteins. Moreover, GPIs can serve as an immunomodulator, which triggers the immune response by stimulating the ability of the immune system to produce antibody or sensitized cells. Nature killer T (NKT) cells could recognize GPIs, resulting in a rapid immune response to various parasitic pathogens [37]. The investigation of the structure-function relationship of GPIs could facilitate elucidating the principle of high mobility and immunomodulation of GPI-anchored proteins.

One of the famous examples of GPI-anchored proteins is prion protein (PrP), whose misfolded form is an infectious agent (PrP^{Sc}). PrP^{C} is a normal form of the protein, which can be digested by proteinase K and can be released from the cell surface by PLC which cleaves the GPI-anchor [38]. The mechanism of conversion from PrP^{C} to PrP^{Sc} is still unclear. GPI-anchor is implicated in the pathogenesis of prion disease. In the transgenic mice model, the engineered PrP lacking the GPI membrane anchor formed abnormal proteinase-resistant prion (PrP^{Sc}) amyloid deposits in their brains and hearts when infected with the murine scrapie, while infection of normal mice with a GPI-anchored PrP did not deposit amyloid with PrP^{Sc} in the brain or the heart [39]. Molecular dynamics simulation suggests that, unlike other lipid anchors, GPI-anchor is highly flexible and would maintain the protein at a certain distance from the membrane surface, with little influence on its structure or orientational freedom [40].

1.5 Other Types of Lipidation

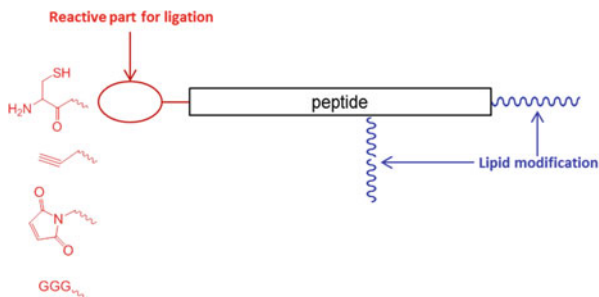
Besides four major types of lipidation, there are also many other types of lipidation, such as addition of cholesterol and phosphatidylethanolamine. Although some of them have been rarely identified so far, the importance in many biological processes is well established. To date, the only example with C-terminal modification of a cholesterol molecule in mammalian cells is the Hedgehog (Hh) family, which plays a critical role in regulating cellular differentiation and proliferation. The pro-Hh proteins (45 kDa) contain a C-terminal processing domain, which mediates the formation of a thioester intermediate and the subsequent addition of a cholesterol molecule in an intein-like process [41]. The resulting N-terminal 20-kDa fragment is further *S*-palmitoylated at the N-terminal cysteine by the Hedgehog acyl transferase. This *S*-acyl moiety migrates to the N-terminal amino group after an $S \rightarrow N$ acyl shift to form a stable amide bond [42]. Both lipidations are essential for the function of the Hh proteins. The cholesterol modification may play a role in regulating the Hh activity gradient to restrict the dilution and unregulated spread of Hh at the cell surface [43].

Another example of rarely found lipidation is phosphatidylethanolaminylation on the microtubule-associated protein light chain 3 (LC3, the mammalian homolog of yeast Atg8) family proteins, which play a key role in the formation of autophagosomes during the autophagy process. LC3 family proteins are the only phosphatidylethanolaminylated proteins identified so far. LC3 proteins require a phosphatidylethanolamine (PE) group attached to the C-terminal glycine for correct membrane localization and function. In mammalian cells, production of lipidated LC3 is controlled by two ubiquitin-like conjugation systems. Newly synthesized LC3 is processed by a protease, Atg4, to expose a C-terminal glycine. The resulting LC3 serves as a substrate for the addition of a PE molecule in a ubiquitin-like conjugation reaction catalysed by E1-like Atg7, E2-like Atg3, and the E3-like Atg12-Atg5:Atg16L complex (Atg16L complex). The Atg16L complex is generated by another ubiquitin-like conjugation system, in which Atg12 is conjugated to the lysine side chain of Atg5 in sequential reactions catalyzed by Atg7 and Atg10. There is no E3-like enzyme implicated in the Atg12-Atg5 conjugation. The Atg12-Atg5 conjugate further forms a complex with a multimeric protein, Atg16L. The Atg12-5 conjugate promotes LC3-PE formation, and Atg4 releases lipidated LC3 from the surface of closed autophagosomes [44, 45].

2 Synthesis of Lipidated Peptides

Preparation of lipidated proteins allows for in-depth study of protein function and the biological process in which the protein is involved. However, it is difficult to obtain lipidated proteins by using traditional biochemical approaches. Recent advances in protein ligation methods profoundly facilitate production of lipidated

Scheme 1 Overview of the lipidated peptide for ligation. In general, the lipidated peptide for ligation consists of three parts: peptide, lipid group and N-terminal reactive group (natural Cys and triglycine, and non-natural maleimide and alkyne moieties)



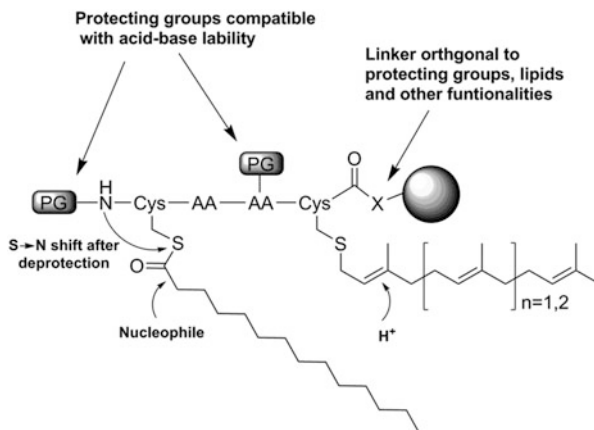
proteins by chemical synthesis. These ligation methods, including expressed protein ligation (EPL), maleimidocaproyl (MIC) ligation, Diels–Alder ligation, click ligation and sortase-mediated protein ligation, allow for ligation of lipidated peptides with expressed proteins [46]. Therefore, synthesis of lipidated peptides has been considered an important aspect in the preparation of lipidated proteins.

In general, the lipidated peptide for ligation usually consists of three parts: peptide, lipid moiety and N-terminal reactive group (natural Cys and triglycine, and non-natural maleimide and alkyne moieties) (Scheme 1). The strategy for synthesis of lipidated peptides depends upon the nature of the lipid group, peptide sequence, and the reactive group for ligation. Many different synthetic approaches have been reported, such as solution and/or solid-phase strategies, different protection strategies involving the *tert*-butoxycarbonyl (Boc) strategy and 9-fluorenylmethoxycarbonyl (Fmoc) strategy, and the methods for the incorporation of the lipid groups.

The solution phase approach is usually slow and laborious. Moreover, the increasing insolubility of the growing peptide chain in the reaction medium causes problems in both purification and the next coupling step. Solid-phase peptide synthesis (SPPS) has now become a widely used approach for peptide synthesis in the lab. The synthesis of lipidated C-terminal peptides of the Ras protein family typically involves preparation of lipidated amino acid building blocks, which are then incorporated into the elongating peptide chain, whereas the lipidated peptide containing phosphatidylethanolamine (PE) is prepared by coupling the lipid group to C-terminus of the peptide in solution. In this review, we briefly introduce the solution phase approach and the synthesis of lipidated building blocks and give an overview of the linker strategies for the SPPS of lipidated peptides. We focus on the synthesis of lipidated peptide specifically used for the ligation with proteins. More detailed information for the preparation of lipidated peptides has been reviewed by Waldmann and co-workers [47–49].

Synthesis of lipidated peptides, especially lipidated Rab and Ras peptides, is challenging because of several limitations (Scheme 2). First, prenyl groups, such as farnesyl or geranylgeranyl, cannot be combined with strong acid-labile protecting groups or linker systems because acids attack the double bonds and lead to isomerization of prenyl groups. Therefore, high concentrations of acid during the synthesis or for the release of the peptide from the solid support should be avoided.

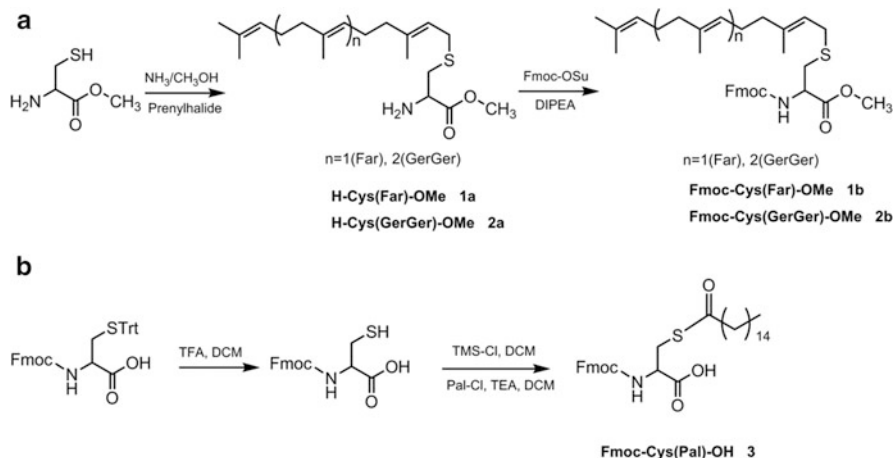
Scheme 2 General considerations for the synthesis of lipidated peptides of Ras family proteins. *PG* protecting group



Second, when a palmitoyl group is present, different conditions for the Fmoc deprotection and the coupling of amino acids should be chosen to minimize a nucleophilic attack on the thioester. Moreover, $S \rightarrow N$ acyl shift at the N-terminally unprotected Cys should be considered. Third, it should be considered that additional functional groups, which are often incorporated in the lipidated peptides for ligation or biological studies, such as maleimide, fluorophore and alkyne moieties, typically lead to additional restrictions for the synthetic strategy.

2.1 Preparation of Lipidated Cysteine Building Blocks

Incorporation of lipid groups can be performed in two ways. Either the lipidated cysteine building blocks are coupled into the peptide chain, or the lipid is introduced to the complete peptide backbone. The former approach is more suited for the synthesis of lipidated Ras and Rab peptides because of its flexibility. The prenylated cysteine building blocks (either with a farnesyl or a geranylgeranyl group) can be prepared by alkylation of the free thiolate of cysteine with prenyl chloride (Scheme 3a). The palmitoylated cysteine can be synthesized from Fmoc-Cys(Trt)-OH after removal of the trityl group and coupling with palmitoyl chloride (Scheme 3b) [50]. In the coupling of the Fmoc-Cys(Pal)-OH to the peptide in SPPS, to minimize the $S \rightarrow N$ acyl shift of the palmitoyl group, a fast removal of the Fmoc group is performed by 1,8-diazobicyclo[5.4.0]undec-7-ene (DBU), a non-nucleophilic hindered base. The coupling is then carried out immediately using preactivated amino acid and HATU as the coupling reagent [50].

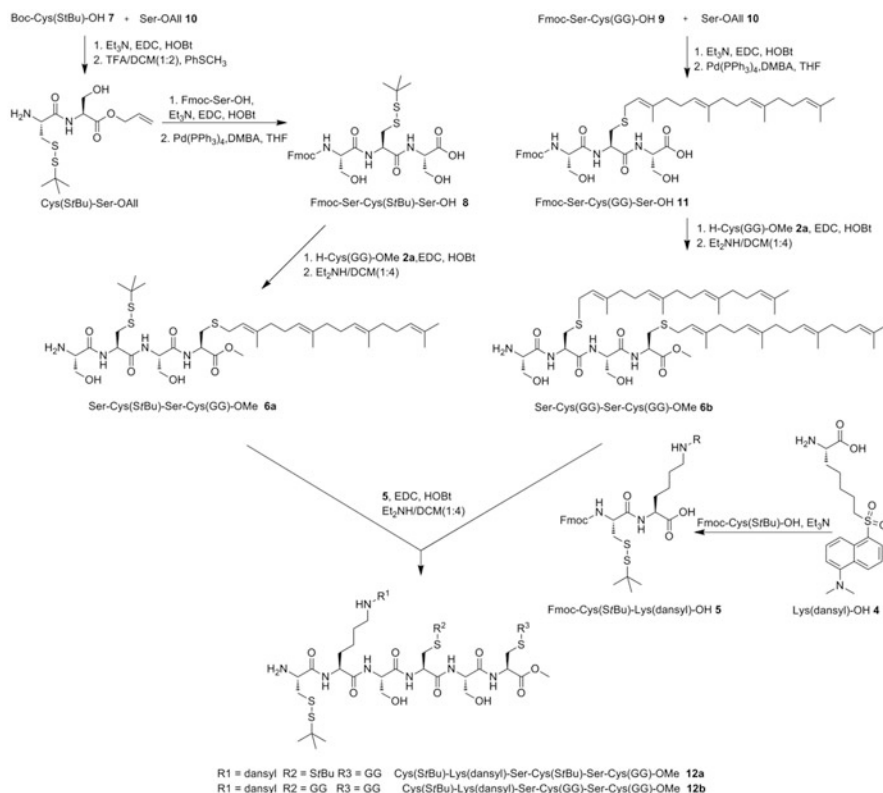


2.2 Solution-Phase Synthesis of Lipidated Peptides

The introduction of acid-sensitive prenyl groups and base-sensitive palmitoyl groups significantly limited the choice of orthogonal protecting groups for carboxy, amino, thiol and hydroxyl groups, which can be removed selectively under mild conditions. The synthesis of lipidated Ras peptides involves the combination of acid-labile *tert*-butyl ester function as carboxy protecting group, the Pd⁰-sensitive allyloxycarbonyl (Aloc) urethane function as amino-blocking group, mild acid-labile trityl-type protecting groups for masking lysine side chains, the reduction-labile *tert*-butyl disulfide function for protection of thiol groups and removal of Boc group with TMSOTf/lutidine [51, 52]. Because of the undesired cyclization in the linear elongation approach, prenylated peptides derived from Rab7 C-terminus (**12a**, **b**) were synthesized by applying a convergent approach using geranylgeranylated cysteine (**2a**) and ϵ -*N*-fluorescently labelled lysine (**5**) as building blocks (Scheme 4) [53, 54].

2.3 Solid-Phase Approach for the Synthesis of Lipidated Peptides

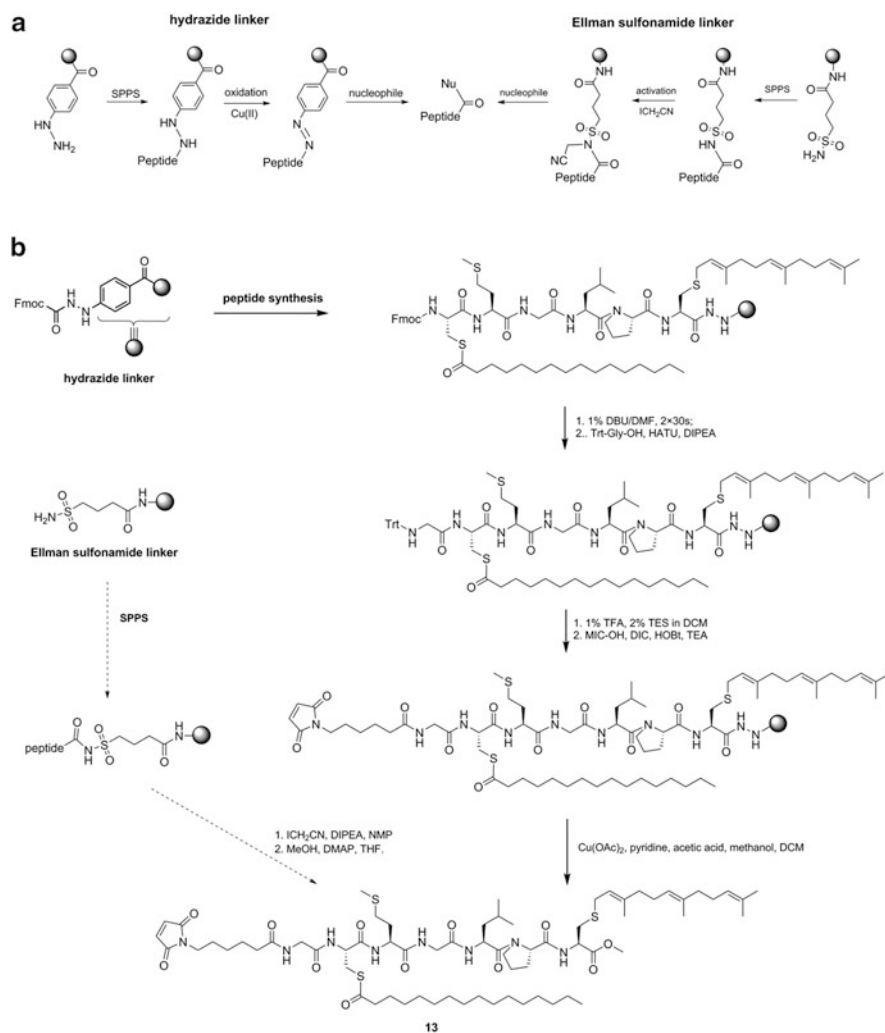
Solid-phase peptide synthesis (SPPS) is a fast and flexible approach and has been frequently used for the synthesis of lipidated peptides. SPPS allows for preparation of the desired peptides, including both natural and nonnatural modifications, with high purity and good yields in a short time. The linker chosen for the solid-phase synthesis of lipidated peptides is of utmost importance. High concentrations of acid during the synthesis or for the release of the peptide from the solid support should



Scheme 4 Synthesis of fluorescently labelled mono- and diprenylated Rab7 C-terminal hexapeptides using solution-phase approach

be avoided so as to keep the prenyl group intact. In the case where a palmitoyl group is present, different conditions for the Fmoc deprotection and the coupling of amino acids should be used to minimize a nucleophilic attack to the thioester and an S → N acyl shift. Finally, the linker should be able to afford the desired peptide as a C-terminal methyl ester in case this functionality is present in the native sequence.

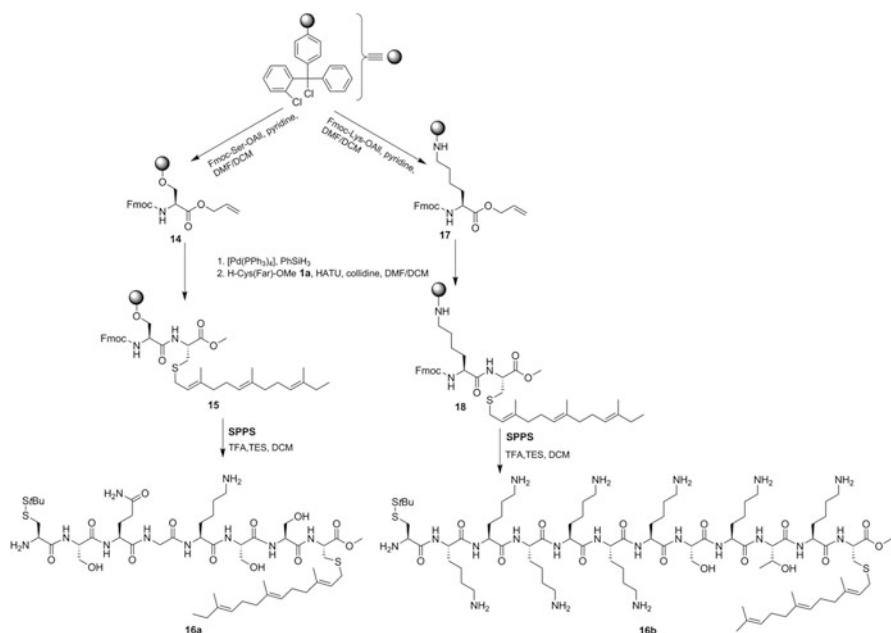
Not so many linkers meet all these requirements. Among them, the hydrazide linker and the Ellman sulfonamide linker are stable under acid and basic conditions, permitting the synthesis of the peptides and their orthogonal release from the solid support [47]. The hydrazide linker can be cleaved by oxidation to an acyldiazene followed by a nucleophilic attack by methanol or water to release the peptide with a C-terminal methyl ester or carboxylic acid, respectively (Scheme 5a) [55]. Oxidation can be performed with either Cu(AcO)₂/O₂ or NBS. Such cleavage conditions are orthogonal to prenyl and palmitoyl groups and classical protecting groups (Boc, Fmoc and Alloc). An example of using the hydrazide linker for the synthesis of N-Ras C-terminal peptide **13** is depicted in Scheme 5b (solid line) [53]. The hydrazide linker was also used to produce lipidated peptide by on-resin lipidation



Scheme 5 Synthesis of N-Ras protein C-terminus for MIC ligation using hydrazone linker and Ellman sulphonamide linker. (a) Cleavage of the hydrazone linker by oxidation and nucleophilic attack and Ellman sulfonamide linker by activation and nucleophilic attack of a nucleophile. (b) Synthesis of farnesylated and palmitoylated N-Ras C-terminus with maleimido group using hydrazone linker and Ellman sulfonamide linker

[56]. However, because free amines could attack the oxidized linker, deprotection of amines has to be performed after linker cleavage, in order to reduce the formation of undesired cyclic peptides [57].

The Ellman sulfonamide linker is stable under acid or basic conditions. It can be selectively alkylated with haloacetonitriles, and then becomes susceptible to nucleophilic attack, leading to release of the peptide from the solid support (Scheme 5a)



Scheme 6 Synthesis of the farnesylated and carboxymethylated C-terminal peptide of Rheb and K-Ras4B using the Trt linker

[58]. However, classical cleavage with a solution of methanol and DMAP leads to significant racemization of cysteine. An alternative approach was to release the peptide from the solid support using H-Cys(Far)-OMe as a nucleophile with microwave irradiation for 10 min. Peptide corresponding to N-Ras protein C-terminus **13** was synthesized using Ellman sulfonamide linker strategy as shown in Scheme 5b (dotted line) [59, 60].

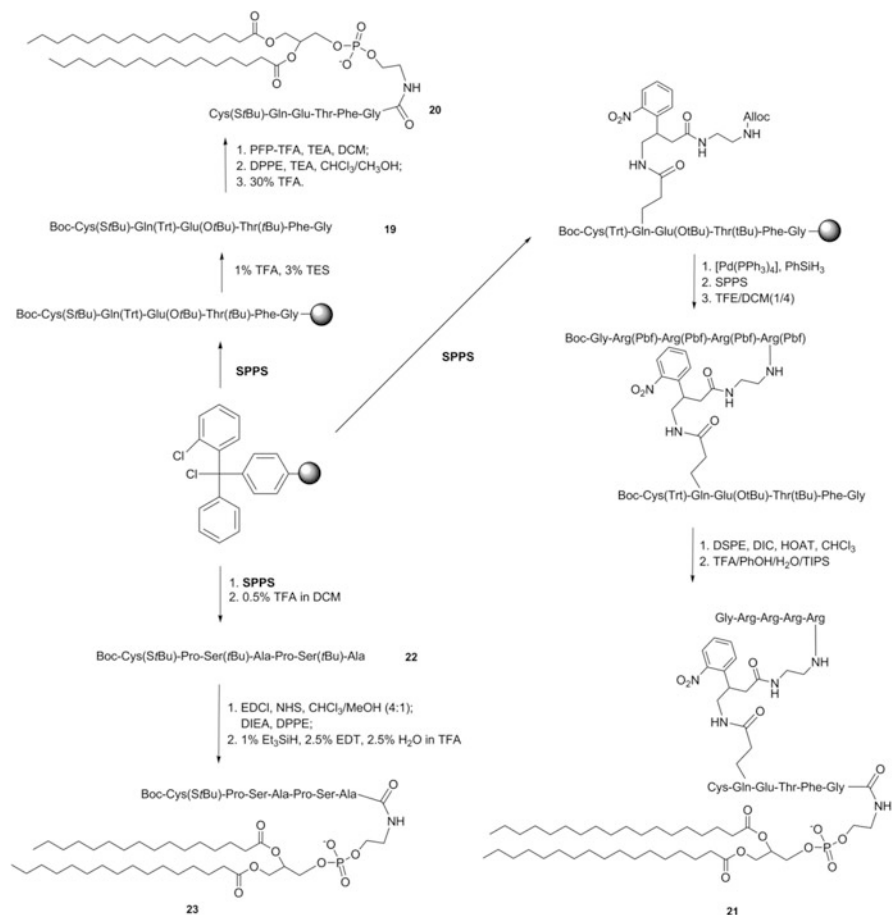
Another linker successfully applied to the synthesis of lipidated peptides is the trityl linker, which can be cleaved by treatment with low concentrations of acid (1% TFA) without affecting the integrity of the prenyl group. The main disadvantage of this linker is that the cleavage at the C-terminus generates free carboxylic acid. In order to obtain the C-terminal methyl ester, the last second C-terminal amino acid can be immobilized on the solid support via the side-chain. After incorporation of the designated prenylated cysteine methyl ester **1**, the peptide chain can be elongated and subsequently released from the solid support with 1% TFA, which simultaneously cleaves all acid-sensitive side-chain protecting groups without affecting the farnesyl group. Synthesis of the farnesylated and carboxymethylated C-terminal peptide of Rheb and K-Ras4B is shown as an example [61] (Scheme 6). For the synthesis of Rheb peptide, Fmoc-Ser-OAlI was loaded to the trityl resin through the side-chain hydroxyl group of Ser to form **14**. After selective removal of the allyl ester and coupling of the *S*-farnesylated cysteine methylester, the peptide chain **15** was elongated by Fmoc strategy with N-Fmoc-protected amino acids with

acid-labile side-chain protecting groups. Finally, treatment of the resin with 1% TFA and a scavenger released the desired peptide **16a**. The lipidated C-terminal peptide of K-Ras4B (**16b**) was synthesized by a similar strategy. In this case, Fmoc-Lys-OAll was loaded to the trityl resin through the side-chain amine (**17**). The side chain of other Fmoc-lysine building blocks was protected with the orthogonal allyloxycarbonyl (Aloc) group which can be liberated with palladium(0) and piperidine. After cleavage from the solid support, the peptide **16b** could be precipitated readily in diethyl ether (Scheme 6) [61]. It should be noted that the deprotection conditions of Aloc are incompatible with thioesters, such as palmitoylated cysteine, and the maleimide group, because of the use of piperidine and triphenylphosphine in the reaction.

2.4 Synthesis of Lipidated Peptides by Combined Solution/Solid-Phase Approach

2.4.1 Synthesis of Phosphatidylethanolaminylated Peptide

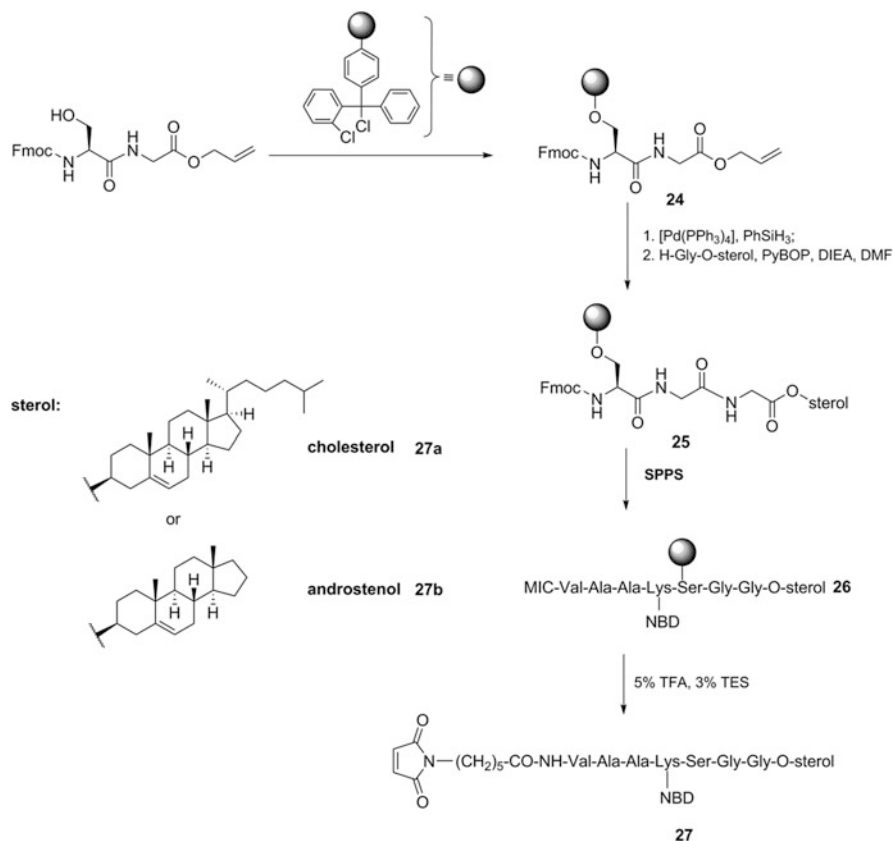
Synthesis of phosphatidylethanolaminylated (PE) peptide is performed by lipidation of peptide backbone in solution, which was synthesized using the trityl linker strategy via SPPS. PE-conjugated C-terminal peptide of LC3 (**20**) was synthesized on the chlorotriyl resin by means of the Fmoc strategy (Scheme 7) [62]. After release from the resin, the protected peptide **19** was subsequently activated by pentafluorophenyl trifluoroacetate as an activated ester and was coupled to 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine (DPPE) in solution to produce protected lipidated peptide. The desired lipidated peptide **20** can be obtained after removing all acid sensitive protecting groups by a high concentration of TFA. In another approach, the peptide **22** was preactivated with 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide HCl (EDCI) and *N*-hydroxysuccinimide (NHS), and was subsequently coupled to the DPPE in the presence of base *N,N*-diisopropylethylamine (DIEA). Removal of the peptide protecting groups with TFA afforded PE-conjugated peptide **23** [63]. In order to facilitate handling and solubilisation of LC3-PE protein, Liu and co-workers introduced a photolabile poly-Arg chain into the lipidated peptide [64]. To this end, the main peptide chain was prepared on the chlorotriyl resin, followed by an elongation of poly-Arg chain at the glutamine side chain, which is connected via a photosensitive nitrobenzyl linker. The branched peptide was cleaved off the resin and condensed with 1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine (DSPE). The final PE-peptide **21** was obtained after removal of the protecting groups.



Scheme 7 Synthesis of the PE conjugated peptides for EPL

2.4.2 Synthesis of Sterol-Modified Peptide

C-terminally sterol-modified heptapeptides derived from the Hedgehog protein were prepared by a combined solution and solid-phase approach with introduction of different functional and reporter groups, i.e. sterols, a fluorescent label for membrane binding assay, and a maleimidocaproyl (MIC) group for ligation to the protein [65]. Dipeptide Fmoc-Ser-Gly-OAll was prepared in solution and loaded to the trityl resin via the serine side chain. The C terminus of the immobilized dipeptide was coupled with glyceryl-sterol esters. The glyceryl-sterol esters were prepared by esterification of *tert*-butyloxycarbonyl (Boc)-protected glycine with the sterols using *N,N*-diisopropylcarbodiimide (DIC) and 4-(dimethylamino) pyridine (DMAP) followed by selective removal of the Boc group. N-terminal peptide chain elongation was achieved by means of SPPS to yield peptide **26** carrying an



Scheme 8 Synthesis of sterol-modified heptapeptides for MIC ligation

NBD group at a lysine side chain and a maleimide group at the N-terminus. The peptides were cleaved from the resin under very mild conditions, resulting in desired products **27** (cholesterol for **27a** and androstenol for **27b**) (Scheme 8) [65].

3 Synthesis of Lipidated Proteins

In general, there are two approaches which have been used to prepare lipidated proteins: (1) the incorporation of the lipid by lipid transfer enzymes and (2) the ligation of lipidated peptides with expressed proteins [66]. Recently, approaches using protein prenyltransferases have been used to obtain prenylated Ras family proteins. Protein prenyltransferases can tolerate diverse modifications of their lipid substrates [67, 68]. Therefore, bioorthogonal groups or probes can be incorporated into proteins containing a CaaX motif or Rab proteins via prenylation [69–72].

Such enzymatic approaches have limitations in the scope of manipulation of protein structure, and therefore are not suited for the preparation of proteins with different lipid moieties at multiple sites and/or with non-natural groups. Moreover, not all lipid transfer enzymes are readily recombinantly available. Chemical protein ligation methods have been developed in the past few years. These methods allow for site-specific lipid modifications of a protein and production in large quantities for cellular, biochemical and biophysical analyses (Table 2).



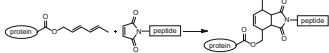


Each chemical ligation method, involving expressed protein ligation (EPL), MIC ligation, Diels–Alder ligation, click ligation and sortase-mediated protein ligation, has its own pros and cons. The choice of the approach depends on the nature of the target protein and design of protein synthesis. The EPL method has been applied to the synthesis of most lipidated proteins, affording the native peptide bond. However, the EPL reaction is relatively slow and sometimes leads to a low yield caused by hydrolysis of thioester. MIC ligation, Diels–Alder ligation and click ligation proceed much faster with a high yield. However, a non-natural linker is introduced in the protein-peptide conjugate, which could affect the function of lipidated proteins. Sortase-mediated protein ligation emerges as a fast ligation strategy with a good yield. We discuss some examples for the application of these ligation methods to the lipidated protein synthesis.

3.1 Assisted Solubilisation Strategy

The poor solubility of lipidated peptides in aqueous solution makes the lipidated protein ligation much more challenging. Moreover, lipidated proteins tend to aggregate and precipitate in solution, which renders them difficult to handle. Several assisted solubilisation techniques have been developed to overcome this problem, including detergent strategy, polyethylene glycol (PEG) tag, ploy-Arg tag and maltose binding protein (MBP) tag.

Detergent is the most popular strategy used for lipidated protein ligation. Detergent not only facilitates solubilisation of lipidated peptide but also drives the ligation reaction as a catalyst. In an early study on the synthesis of mono- and diprenylated Rab7 proteins, a wide range of detergents have been screened [54]. Although most detergents can solubilize the prenylated peptide, only 6 out of 76 detergents are able to support the ligation efficiently, including cetyltrimethylammoniumbromide (CTAB), lauryldimethylamine-*N*-oxide (LDAO), *N*-dodecyl-*N,N*-(dimethylammonio)butyrate (DDMAB), sodiumdodecyl sulfate (SDS), *n*-octyl-phosphocholine (FOS-Choline-8) and cyclohexyl-ethyl- β -D-maltoside (Cymal-2). The ligation efficiency is also dependent on the concentration of the detergent. A concentration above the critical micellar concentration (CMC) is necessary to drive the ligation reaction (unpublished results). It is conceivable that prenylated peptides which form higher order structures in aqueous solution can be made accessible to the protein via formation of mixed detergent micelles. However, it remains unclear why some detergents are dramatically more efficient

Table 2 Chemical ligation methods for the synthesis of lipidated proteins

Ligation methods	Requirements	Mechanism	Product	Examples
Expressed protein ligation	C-terminal thioester		Native peptide bond	Rab7, N-Ras, Rheb, K-Ras4B, D-Ral, YPT1, LC3-PE, lipidated GFP
	Lipidated peptide with an N-terminal cysteine			
Maleimide ligation	Protein with a C-terminal cysteine		Non-natural bond	H-Ras, N-Ras
	Lipidated peptide with an N-terminal maleimide			
Diels–Alder ligation	Protein with a C-terminal dienyl group		Non-natural bond	Rab7
	Lipidated peptide with an N-terminal maleimide			
Click ligation	Protein with a C-terminal azide		Non-natural bond	Rab1, Rab7
	Lipidated peptide with an N-terminal alkyne moiety			
Sortase-mediated protein ligation	Protein with a C-terminal LPXTG motif		Native peptide bond with additional amino acid sequence	K-Ras4B, lipidated GFP, GPI modified GFP
	Lipidated peptide with an N-terminal oligoglycine sequence			

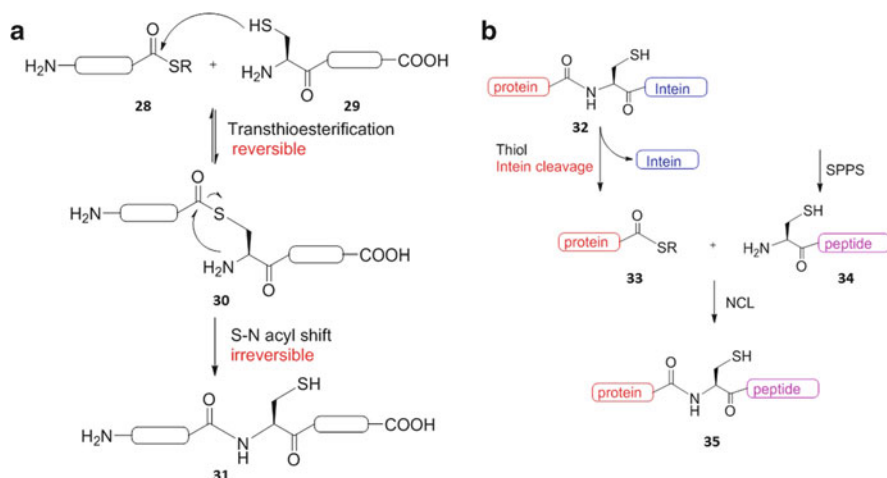
than others. Among these detergents, CTAB appears to be the most robust mediator of the ligation reaction and has been used to produce mono- and diprenylated Rab7 (Scheme 10a) [54].

Syntheses of farnesylated Rheb methyl ester by EPL and prenylated Rab1 and Rab7 proteins by click ligation were carried out efficiently in the presence of CTAB (Schemes 10b and 14b) [61, 73]. The native chemical ligation of PE-modified peptides with protein thioesters was performed in the presence of β -octylglucoside to afford GFP-PE (Scheme 10e) [63] and LC3-PE (Scheme 11) [62]. β -Octylglucoside was also used for the EPL of GPI-anchored GFP (Scheme 10f) [74]. Some other examples of application of detergent to the lipidated protein ligation include 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) for the EPL of geranylgeranylated Rab7 protein [75], dodecylmaltoside (DDM) and dodecyl-phosphocholine (DPC) for the EPL of lipidated rPrP^{Palm} (Scheme 10d) [76], DDM and deoxycholate (DOC) for the sortase-mediated protein ligation of lipidated GFP (Scheme 15b) [77]. Triton X-114 is widely used in the synthesis of lipidated Ras proteins because it can drive protein ligation as well as facilitate purification of ligation product (Scheme 12b, c) [78]. Triton X-114 has a low cloud point of 22°C. The reaction was carried out at 4°C, under which the reaction mixture is homogenous. A temperature shift to 37°C after reaction leads to a phase separation of the detergent phase from the aqueous phase. A further separation of the ligated from unligated protein can be performed by extraction with 11% Triton X-114, whereby lipidated proteins partition into surfactant droplets.

Because the presence of detergents could affect protein function, removal of detergents after ligation is usually required. However, dialysis is not always sufficient to eliminate detergents because of the strong interaction between the lipid group and the detergent. Extensive washing with organic solvents leads to denaturation of the protein. Thereby, another refolding step is required [54]. Recently, detergent-free strategies using traceless solubilisation tags have been developed.

PEG solubilisation tag has been used in the synthesis of lipidated murine prion protein (PrP) with two palmitoyl modifications as the GPI anchor mimic. PEG tag was introduced at the C terminus of the lipidated peptide, leading to a large increase in solubility (Scheme 10d) [76, 79]. Using this strategy, ligation reactions could be carried out in the absence of detergent and organic solvent with a fourfold increase in the yield. The PEG can be removed by proteolytic cleavage with TEV protease.

Poly-Arg tag and MBP tag have been employed in the synthesis of LC3-PE protein. Highly positive charged poly arginine chain makes the PE-peptide and protein soluble in aqueous solution [64]. This strategy allowed for the synthesis of lipidated proteins under detergent-free conditions without laborious screening of the solvents and additives. The poly-Arg tag is connected to the peptide via a photosensitive linker, which can be removed by UV irradiation (Scheme 11). In parallel, an MBP tag strategy was developed for the synthesis of LC3-PE. The MBP tag which is fused to the N-terminus of the LC3 protein thioester dramatically enhances the ligation efficiency, probably owing to the nonspecific association of the PE-peptide with the MBP tag. The EPL reaction was performed under folding conditions. The resulting MBP-LC3-PE protein is soluble in the buffer without detergents, making it facile to handle the lipidated LC3 protein. Before the analysis of the lipidated LC3, the MBP tag was removed by TEV protease (Scheme 11) [62].



Scheme 9 Mechanisms of (a) native chemical ligation and (b) expressed protein ligation

3.2 Expressed Protein Ligation

In the early 1990s, Kent and coworkers introduced the breakthrough approach of native chemical ligation (NCL), which is now a general method for chemical protein synthesis [80, 81]. In NCL, the thiol group of an N-terminal cysteine residue of an unprotected peptide **29** attacks the C-terminal thioester of another unprotected peptide **28** in an aqueous buffer to form a thioester intermediate **30**. The initial chemoselective transthioesterification in NCL is essentially reversible, whereas the subsequent S \rightarrow N acyl shift is spontaneous and irreversible. Thus, the reaction is driven to form a native amide bond specifically at the ligation site, even in the presence of unprotected internal cysteine residues (Scheme 9a). A number of refinements and extensions in ligation methodology and strategy have been developed (for a recent review see [46]).

The scope of application of NCL was significantly widened upon introduction of the approach referred to as expressed protein ligation (EPL) from the Muir laboratory [82, 83]. With EPL, both fragments containing C-terminal thioester and N-terminal cysteine, respectively, can be produced recombinantly (Scheme 9b). EPL emerged as a result of the advances in self-cleavable affinity tags for recombinant protein purification using intein chemistry. Inteins are protein insertion sequences flanked by host protein sequences (N- and C-exteins) and are eventually removed by a posttranslational process termed protein splicing. By means of a C-terminal Asp to Ala substitution on the intein to prevent the formation or breakdown of the branched intermediate, the protein can be trapped in an equilibrium between the thioester and the amide form. The engineered intein can then be cleaved by treatment with thiol reagents (such as 2-mercapoethanesulfonate, MESNA) via an intermolecular transthioesterification reaction, generating a recombinant protein thioester **33** which is ready to undergo NCL with a synthetic peptide

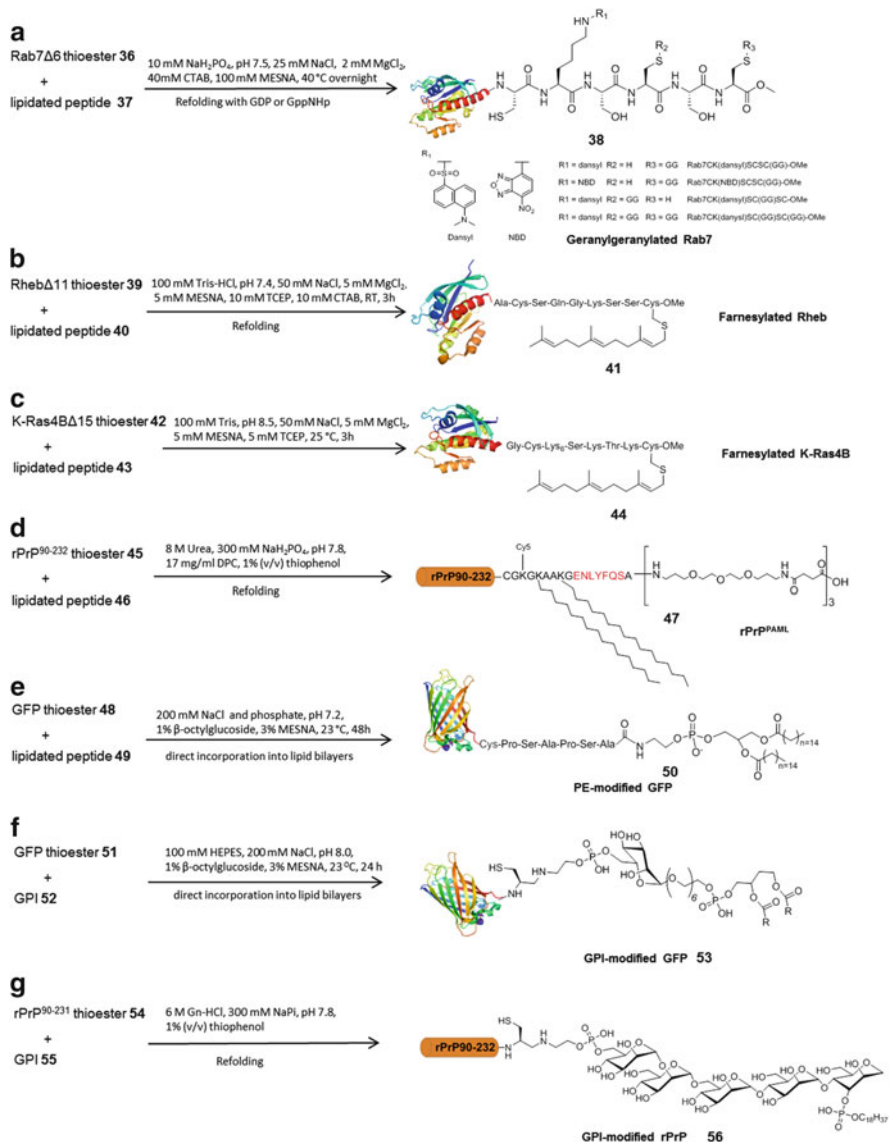
34 containing N-terminal cysteine. Until now, the EPL has been widely applied to produce proteins with post-translational modifications [84, 85].

The EPL approach requires a cysteine residue at the ligation site. Because cysteine is the second least common of the 20 amino acids in proteins, many proteins do not have a native cysteine residue. Even if the protein contains cysteine, it may not be a suitable ligation site. A simple solution is that a mutation to cysteine at the ligation site could be introduced in the ligated protein. Several issues concerning the choice of the ligation site should be considered. First, introduction of a mutation to Cys at the ligation site should minimally interfere with protein activity and function. Second, the synthetic C-terminal peptide length should be short to reduce the synthetic effort and the risk of protein folding. If a cysteine mutation is not tolerated, it is possible to perform the ligation reactions with an amino acid other than Cys or to convert Cys chemically to other amino acids or analogues [46].

Prenylated Rab proteins have been produced using the EPL approach (Scheme 10a) [29, 54, 75, 86, 87]. Rab proteins were expressed in *Escherichia coli* with a C-terminal fusion to an engineered intein, followed by a purification tag, chitin-binding domain (CBD). The Rab-intein-CBD fusion protein on the chitin beads was treated with MESNA to release Rab-thioester protein **36**, which is amenable for native chemical ligation with mono- or diprenylated peptides. After ligation, the protein either remained in solution or precipitated, depending on the ligation conditions, such as detergent and salt concentration. Washing of the ligation mixture with organic solvents led to extraction of the peptide and detergent to the organic phase and precipitation of protein. The protein pellet was dissolved in 6 M guanidinium chloride and was then refolded by stepwise dilution into the buffer containing CHAPS. The approach yielded correctly folded prenylated Rab proteins **38**.

Farnesylated Rheb (**41**) and K-Ras4B (**44**) methyl ester were obtained by EPL (Scheme 10b, c) [61]. Because of the presence of CTAB in the ligation reaction of Rheb, an extraction with organic solvent and subsequent refolding was required (Scheme 10b). In contrast, the C-terminal polybasic amino acid sequence of K-Ras4B mediates solubilisation of farnesylated peptide and protein. Thus, ligation of the peptide **43** with K-Ras4B thioester **42** was carried out in buffer without any detergent. Denaturation and refolding were not needed for the synthesis of farnesylated K-Ras4B.

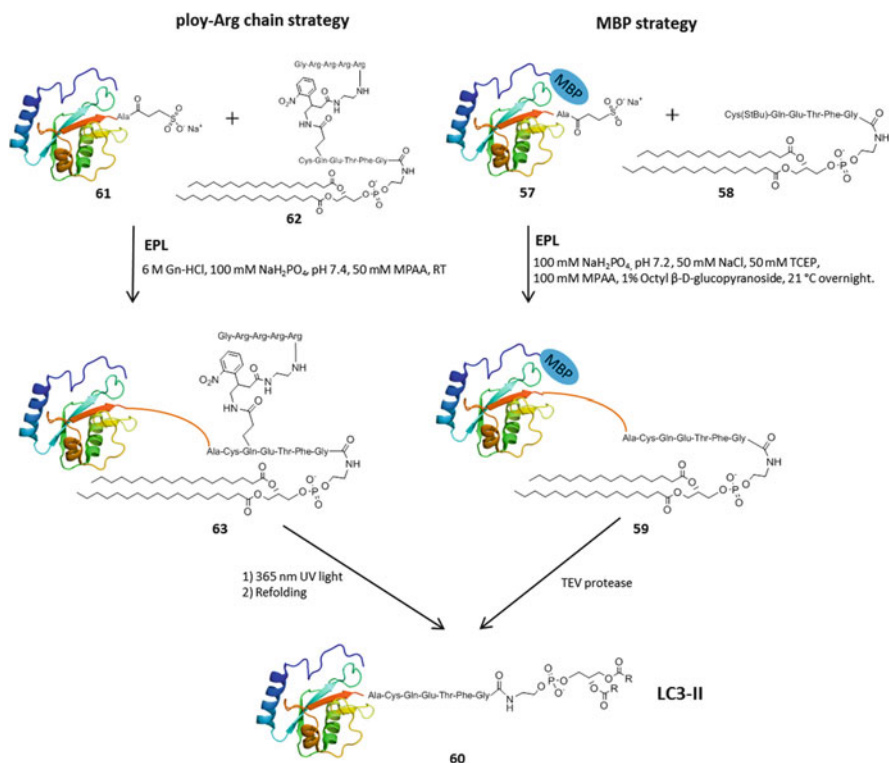
Many other examples of lipidated proteins generated via EPL include PrP^{Palm} **47** [76, 79], GFP-PE **50** [63], GPI-anchored proteins **53** [74, 88] and GPI-anchored PrP **56** [89] (Scheme 10). Bertozzi and co-workers prepared a series of GPI-protein analogues bearing different anchor structures to dissect the structure-function relationship of GPI-proteins (see the discussion in Sect. 4.5). After ligation of cysteine-bearing GPI analogues **52** with GFP-thioester **51**, the resulting GPI-anchored proteins **53** were extracted by 12% Triton X-114 at 37°C (Scheme 10f) [74, 88]. Seeberger and co-workers reported a synthetic strategy for the preparation of homogeneous GPI-anchored prion protein **56** by a similar strategy (Scheme 10g) [89]. Access to the GPI anchor **55** relies on the incorporation



Scheme 10 Semisynthesis of lipidated proteins by using EPL: (a) geranylgeranylated Rab7, (b) farnesylated Rheb, (c) farnesylated K-Ras4B, (d) rPrP^{ALM}, (e) PE-modified GFP protein, (f) GPI-modified GFP protein and (g) GPI-modified rPrP

of the cysteine residue into the GPI backbone before global deprotection and on the judicious selection of protecting groups.

A highly effective catalyst for native chemical ligation, (4-carboxymethyl) thiophenol (MPAA), was used in the EPL of LC3-PE protein (Scheme 11)

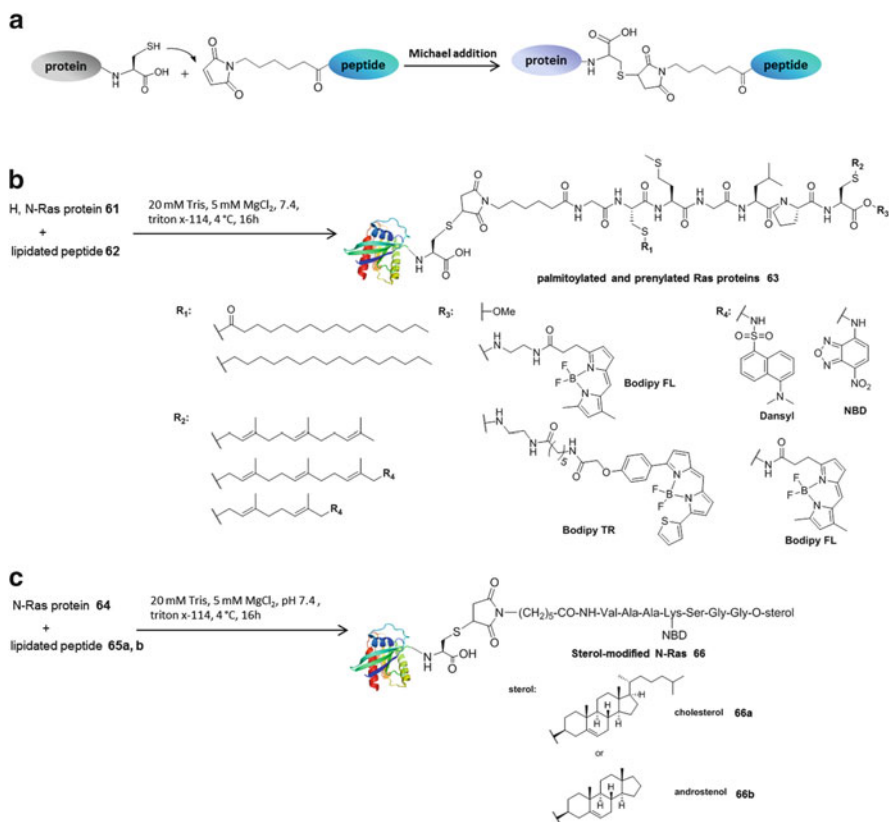


Scheme 11 Semisynthesis of LC3-PE by EPL

[90]. In the poly-Arg solubilisation strategy, the ligation of lipidated hexapeptide **62** with LC3¹⁻¹¹⁴-thioester **61** was performed under denaturing conditions. After cleavage of the poly-Arg tag by UV irradiation, the lipidated protein LC3-PE **60** was purified by HPLC and subsequently refolded by pulse dilution into refolding buffer [64]. In the MBP solubilisation strategy, the reaction of lipidated hexapeptide **58** with MBP-LC3¹⁻¹¹⁴-thioester **57** was performed under folding conditions [62]. The MBP tag was removed by proteolytic cleavage with TEV protease, followed by the amylose affinity chromatography.

3.3 MIC Ligation

The chemoselective Michael addition of sulfhydryl group to the maleimido group is a well-known conjugation reaction under neutral pH, which has been commonly used for the coupling of fluorophores to proteins with surface-exposed cysteine residues. The reaction was used to conjugate a maleimidocaproyl (MIC) peptide to a C-terminally truncated Ras protein bearing a C-terminal cysteine (Scheme 12).



Scheme 12 Synthesis of lipidated proteins using MIC ligation

Chemical synthesis allows for the incorporation of various types of lipids together with reporter groups required for biological studies, such as fluorophores, photoactivatable groups, different kinds of lipid groups, and nonhydrolysable palmitoyl thioester analogues. The modular nature of this approach also offers more opportunity for introducing additional non-natural building blocks. Although the site selectivity of the reaction is limited when more than one cysteine is present, structures of N-Ras and H-Ras suggest that the cysteine residues in the GTPase domain are buried in the fold and therefore are not easily accessible. C-terminally truncated N-Ras or H-Ras protein with a C-terminal cysteine introduced at position 181 was expressed in *E. coli*. The exposure of the C-terminal cysteine makes the ligation reaction fast and selective. The ratio of peptide to protein has to be limited and generally should not exceed 3:1 to prevent nonspecific reaction with internal cysteine residues. The MIC ligation was performed in the presence of Triton X-114 at 4°C. The ligated product was subject to extraction by 11% Triton X-114 at 37°C. Scheme 12 shows the preparation of a collection of semisynthetic H,

N-Ras proteins **63** with different modifications using the MIC ligation approach [78, 86, 91–93].

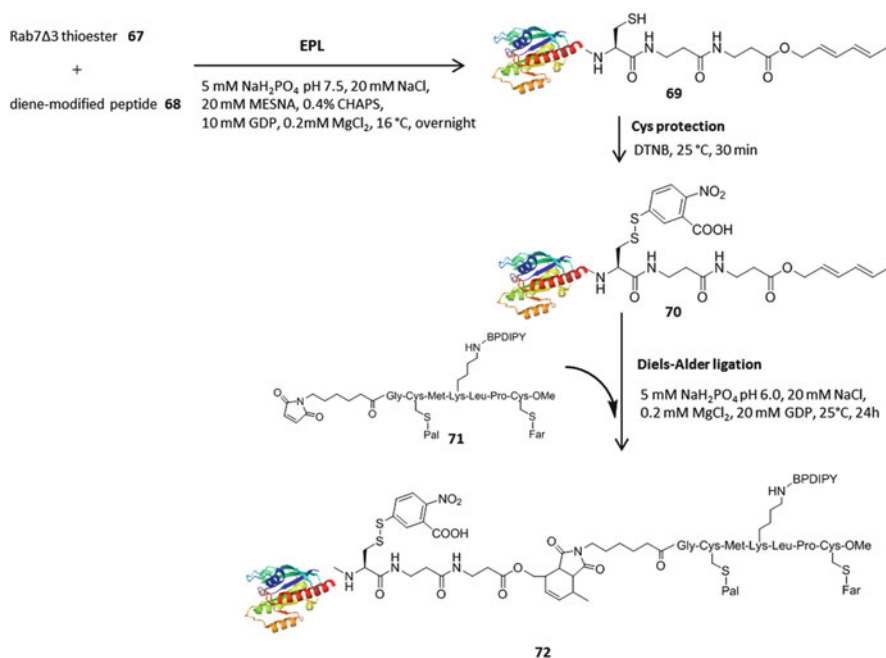
To understand the function of sterol anchors, fluorescently labelled heptapeptides **65a** and **65b** bearing sterol moieties were attached to the N-RasG12V(1–181) **64**, yielding the sterol-modified proteins **66a, b** (Scheme 12c). N-Ras was chosen as the protein moiety because this system offers the possibility to evaluate the membrane binding of different membrane anchors in cells [65].

3.4 *Diels–Alder Ligation*

The Diels–Alder reaction is a highly selective and fast transformation and can proceed in aqueous solution. Its compatibility with biomolecules has been explored elegantly in the bioconjugation and/or immobilization of oligonucleotides and other biomolecules. Wladmann and co-workers reported the development of the Diels–Alder cycloaddition as chemoselective ligation of peptides and proteins under mild conditions [94, 95]. This approach was successfully implemented by employing the Rab7 protein as a representative biologically relevant example. The peptide features a Cys residue at its N-terminus and a 2,4-hexadienyl ester at its C-terminus (Scheme 13). The Rab7-thioester **67** was ligated with the dienyl peptide **68** via EPL. To avoid undesired modification of the thiol group in the subsequent reaction with the maleimide, the accessible cysteine side chains were protected as disulfides by treatment with Ellmann's reagent immediately after the ligation reaction. The resulting protein dienyl ester **70** was ligated with the lipidated peptide **71** containing a maleimido group at the N-terminus to afford lipidated protein **72**. This strategy is suitable for the incorporation of BODIPY fluorophore which is unstable under the conditions of EPL [94, 95].

3.5 *Click Ligation*

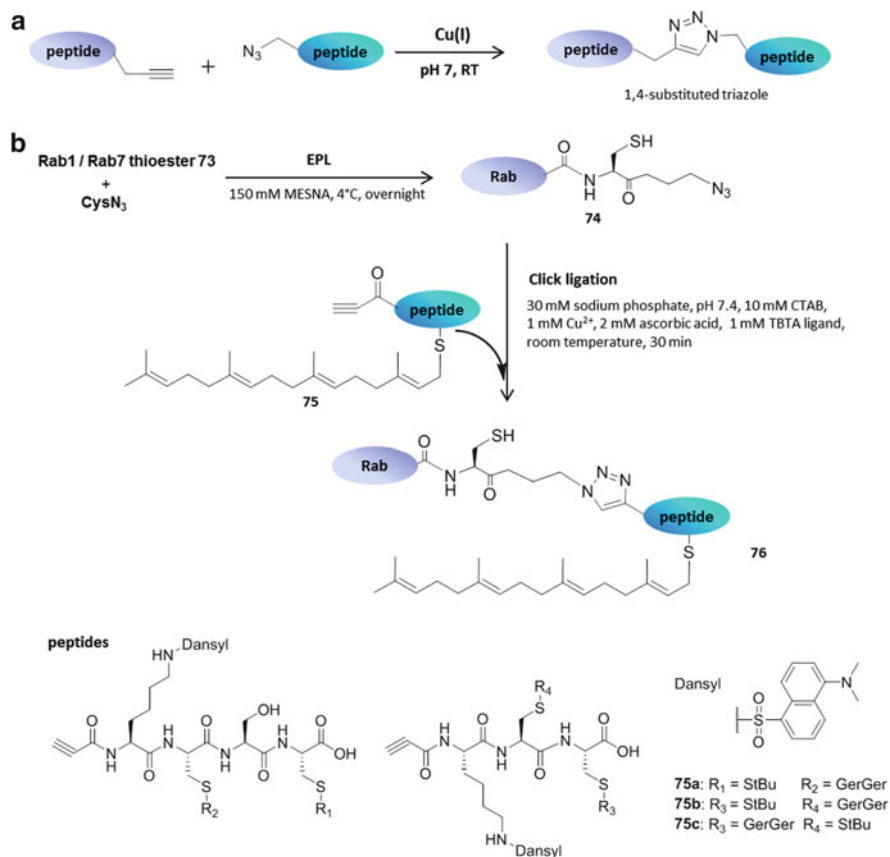
Cu(I)-catalyzed Huisgen 1,3-dipolar cycloaddition reaction, also referred to as the “click reaction”, is widely employed in protein/peptide modifications [96]. The click reaction was applied to the synthesis of geranylgeranylated Rab1 and Rab7 proteins [73]. There are several advantages to using click chemistry: first, the 1,2,3-triazole formed has only a low steric demand and is also regarded as a peptide-bond mimetic linker; second, the reaction proceeds quickly and selectively under neutral pH conditions at room temperature. The incorporation of the azide-modified cysteine, CysN₃, into the Rab protein by EPL is quantitative and efficient, in contrast to the EPL of prenylated peptides. The alkyne-containing peptides **75** were then coupled to the proteins with an azide **74** through the click reaction (Scheme 14). The ligation is fast and quantitative, which makes the purification of the ligated protein highly facile.



Scheme 13 Combination of EPL and Diels–Alder cycloaddition for the synthesis of a palmitoylated, and farnesylated Ras protein

3.6 Sortase-Mediated Protein Ligation

Sortase A (SrtA) is a transpeptidase from the Gram-positive bacterium *Staphylococcus aureus*. It catalyses attachment of proteins with an LPXTG motif to the cell wall. The motif is cleaved by SrtA at threonine residue, leading to formation of a thioester intermediate at the active centre cysteine of SrtA. A nucleophilic attack by the favoured α -amino group of the pentaglycine unit of peptidoglycan on the cell wall results in formation of a new peptide bond (Scheme 15a) [97, 98]. The LPXTG motif has been successfully transposed onto unstructured regions of other proteins to generate new sortase substrates. Protein substrates require only a five amino acid extension (LPETG), a modest insertion which is not expected to impede the function of most proteins and should also have minimal impact on the expression yield of these polypeptides (Scheme 15a). Recently, Ploegh and co-workers have developed a strategy using sortase-mediated transpeptidation as a means to install lipid modifications onto protein substrates in a site-specific fashion (Scheme 15b) [77]. The ligation of lipid-modified triglycine **78** and model protein eGFP **77** was successfully performed in the buffer with 150 μ M SrtA and 1% detergent (β -octylglucoside, DDM or deoxycholate). The His tag present on both the sortase and the C-terminus of eGFP substrate provided a convenient way for purification of the transpeptidation product **79** by Ni-NTA chromatography. The attachment

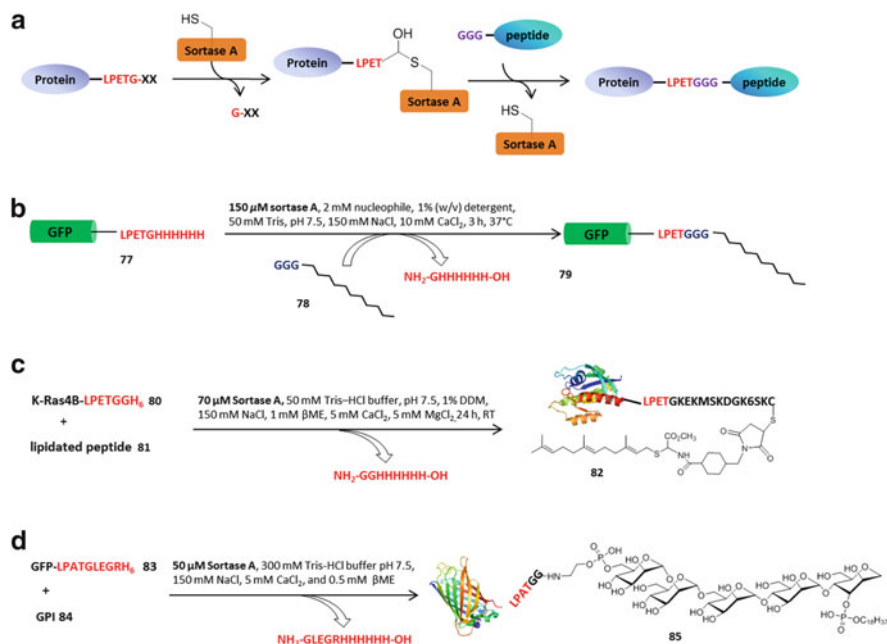


Scheme 14 Semisynthesis of geranylgeranylated Rab proteins by click ligation

of a range of hydrophobic modifications to eGFP was achieved in excellent yields (60–90%).

Another example of lipidated proteins successfully generated by sortase transpeptidation is lipidated K-Ras4B (Scheme 15c) [99]. The farnesyl group was attached to the cysteine of C-terminal K-Ras4B peptide via the sulfo-SMCC heterobifunctional crosslinker. The lipopeptide **81** bearing an N-terminal glycine was ligated to the K-Ras4B protein with an LPETG motif **80** in the presence of 70 μM SrtA and 1% (w/v) *n*-dodecylmaltoside (DDM).

Recently, Guo and co-workers reported sortase-mediated chemoenzymatic synthesis of a GPI-anchored protein [100]. The GPI anchor **84** featuring the common glycan core, a lipid and an additional double glycine unit was coupled to the model protein GFP **83** by SrtA (Scheme 15d). This work has demonstrated that SrtA could accept a complex GPI anchor, suggesting that SrtA-mediated protein ligation is a versatile approach for protein synthesis.



Scheme 15 Site-specific lipid attachment through sortase-mediated transpeptidation. (a) Mechanism of sortase-mediated ligation. (b) Semisynthesis of lipid modified GFP protein by sortase-mediated ligation. (c) Semisynthesis of lipidated K-Ras4B protein. (d) Semisynthesis of GPI modified GFP protein

4 Chemical Biology of Lipidated Protein

Protein crystallization, NMR, FTIR and AFM studies usually required large quantities of homogeneous proteins. Chemical approaches as shown above allow for production of reasonable amount of lipid modified proteins with well-defined structures as well as incorporation of reporter groups into proteins. These strategies have profoundly facilitated the structural, biophysical and cellular studies of the function of lipidated proteins. Some examples are discussed in this section.

4.1 Cell Biological Studies of S-Palmitoylation Cycle of Ras GTPases

Ras GTPases signalling is spatially organized by its specific intracellular localization on membranes or microdomains. Three isoforms of Ras protein (H-, N- and K-Ras) share a common C-terminal *S*-farnesylcysteine carboxymethyl ester, while N- and H-Ras have one and two adjacent *S*-palmitoylcysteine residues, respectively, and K-Ras has a polylysine cluster at the C-terminus. The three isoforms of

Ras protein take different intracellular trafficking modes because of different lipidated patterns. To elucidate how the S-palmitoylation cycle regulates the localization of N- or H-Ras, semi-synthetic lipidated N or H-Ras proteins with natural and unnatural lipidated patterns are required (Scheme 12b, Fig. 1a). Hexadecyl (HD) group is introduced as a non-cleavable palmitoyl (Pal) analogue, and a serine substitution of cysteine provides a non-palmitoylatable form. D-Cysteine and β -peptidomimetics are employed to study the specificity of palmitoylation enzymes. Bastiaens and co-workers investigated the retrograde trafficking of Ras from the PM to the Golgi apparatus in Madin–Darby canine kidney (MDCK) cells using the fluorescence recovery after photobleaching (FRAP) technique [18]. N-Ras(PalFar) and N-Ras(HDFar) were microinjected into the MDCK cells, respectively. The measurements clearly showed that PalFar protein localized normally to PM and the Golgi and displayed a similar fluorescence recovery kinetics at the Golgi to the wild type N-Ras. In contrast, HDFar protein localized unspecifically to the entire membrane system and did not display restricted Golgi or PM localization. Specific fluorescence recovery at the Golgi was not apparently observed in FRAP experiments (Fig. 1b). These findings suggest that retrograde PM-Golgi trafficking of H-Ras and N-Ras is mediated by de/repalmitoylation activities acting on Ras in different subcellular localizations.

To elucidate the site and kinetics of Ras palmitoylation, CysFar, a substrate for palmitoylation resembling the depalmitoylated N-Ras, was microinjected into the cell. A rapid accumulation of CysFar at the Golgi with a $t_{1/2}$ of 14 s was observed,

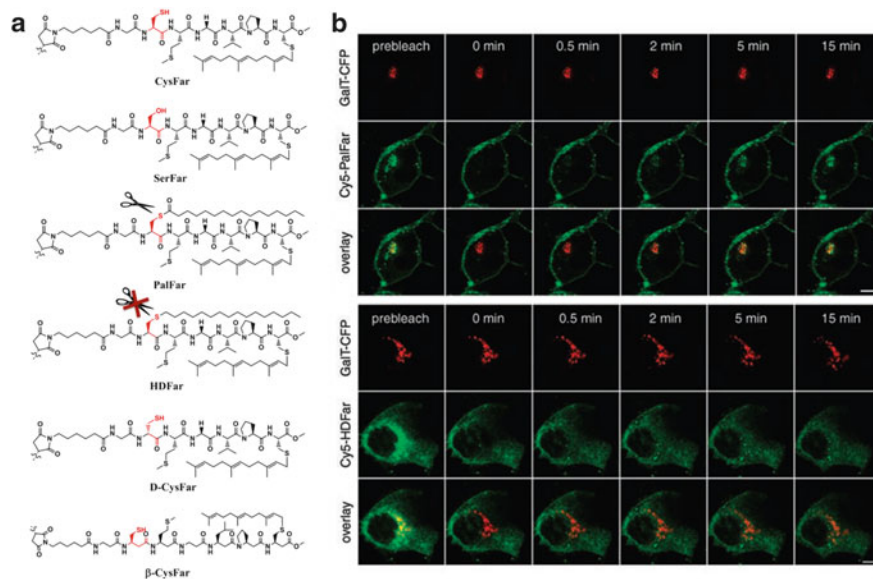
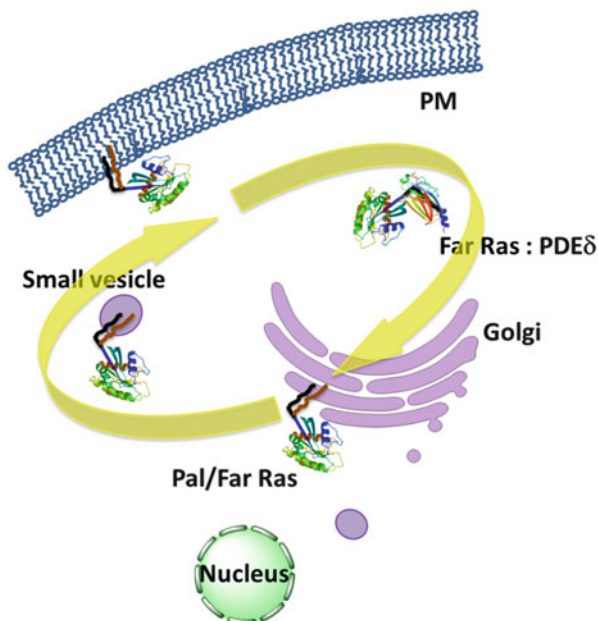


Fig. 1 (a) Semisynthetic N-Ras proteins with various C-terminal structures. (b) Cellular distribution and FRAP measurements of Cy5-N-Ras-PalFar and Cy5-N-Ras-HDFar at the Golgi. GalT is a Golgi marker

Scheme 16 Model for de/repalmitoylation cycle of H- and N-Ras in living cells



followed by PM localization at later time points, suggesting that the palmitoylated CysFar exits Golgi via the secretory pathway. In contrast, SerFar, a protein which cannot be palmitoylated, nonspecifically distributed over endomembranes. FRAP measurements showed that CysFar fluorescence recovery at the Golgi is 13-fold slower than SerFar, suggesting a stable membrane association because of palmitoylation. These results led to the conclusion that prenylated Ras is further palmitoylated at the Golgi apparatus within seconds [20]. Depalmitoylation was accessed by using PalFar, a substrate for depalmitoylation, and HDFar with a non-cleavable palmitoyl analogue. PalFar rapidly accumulated on the Golgi shortly after microinjection, whereas HDFar distributed all over the cell. PalFar is depalmitoylated before reaching Golgi, which is derived from the similar recovery kinetics of PalFar at the Golgi to CysFar. These experiments show that N-Ras is depalmitoylated everywhere in the cell on a time scale of seconds. Furthermore, to study the substrate specificity of the palmitoylation machinery, α -CysFar and β -CysFar (Fig. 1a) proteins were evaluated. Both proteins are rapidly trapped on the Golgi by palmitoylation with kinetics similar to CysFar. The results imply that no consensus sequence is involved in cellular palmitoylation and that there is no essential requirement for the de/repalmitoylation machinery to recognize any structure on the substrate other than the target cysteine side chain [20]. These studies reveal that the palmitoylation cycle plays a key role in Ras intracellular localization and translocation, thereby controlling Ras activity in different organelles (Scheme 16).

4.2 Biophysical Studies of Lipidated Ras GTPases

The lipidation of Ras plays a vital role in regulating the protein localization and function. The association with different membrane microenvironments, such as lipid rafts, is believed to regulate Ras signalling further. Lipid rafts can serve as “signalling platforms” involved in transducing extracellular stimuli into the cell. To investigate how the farnesylated and palmitoylated Ras proteins localize to different membrane microdomains, the fully lipidated Ras proteins are required. The 4,4-difluoro-4-bora-3a,4a-diaza-*s*-indacene (BODIPY) labelled and dual lipidated [hexadecylated (HD) as a nonhydrolysable palmitoyl group analogue and farnesylated (Far)] N-Ras protein was obtained by MIC ligation (Scheme 12b). The heterogeneous lipid bilayer systems were generated by 1-palmitoyl-2-oleoylphosphatidylcholine (POPC), bovine sphingomyelin (BSM) and cholesterol. The liquid ordered (l_o) domain, liquid disordered (l_d) phase, and gel or solid ordered (s_o) phase were controlled by different ratio of POPC/BSM/cholesterol mixture [101]. Winter and co-workers elucidated the interaction between lipidated Ras protein and membrane and investigated the distribution of Ras proteins in membrane microenvironments using two-photon fluorescence microscopy on giant unilamellar vesicles (GUVs) and tapping mode atomic force microscopy (AFM) [102]. The result of time-dependent partitioning of lipidated N-Ras in the different domains of GUVs indicates that the phase sequence of preferential binding of N-Ras to mixed-domain lipid vesicles is $l_d > l_o \gg s_o$.

Moreover, a series of N-Ras proteins with different lipid patterns (N-Ras Far/Far, N-Ras HD/Far, N-Ras HD/HD, N-Ras Far) and farnesylated K-Ras4B were prepared (Schemes 10c and 12b). By using time-lapse tapping-mode AFM, the partitioning of these N-Ras proteins to various membrane microenvironments are able to be detected. The results showed that GDP-bound N-Ras proteins bearing at least one farnesyl anchor (N-Ras Far/Far, N-Ras HD/Far, N-Ras Far) display comparable membrane partitioning behaviour and show diffusion of the protein into the l_o/l_d phase boundary region, suggesting that the bulky and rigid farnesyl anchor is responsible for the clustering of N-Ras proteins in the interfacial regions of membrane domains, thus leading to a decrease of the line energy (tension) between domains (Fig. 2) [103]. In contrast to N-Ras, farnesylated K-Ras4B induces formation of new protein-containing fluid domains within the bulk fluid phase (l_d) and is believed to recruit multivalent acidic lipids by an effective, electrostatic lipid sorting mechanism. Furthermore, the GDP-GTP exchange and thereby K-Ras4B activation leads to changes in G-domain orientation and a stronger enrichment of activated K-Ras4B in the signalling platform [104].

The lipidated Ras proteins were further studied under some extreme environmental conditions by monitoring the chemical or physical signal. For instance, pressure modulation has been applied in combination with FTIR spectroscopy to reveal equilibria between spectroscopically resolved conformations of the lipidated N-Ras. The measurements showed that increased pressure shifts the conformational equilibrium toward the more open and solvent exposed state 1, which is involved in

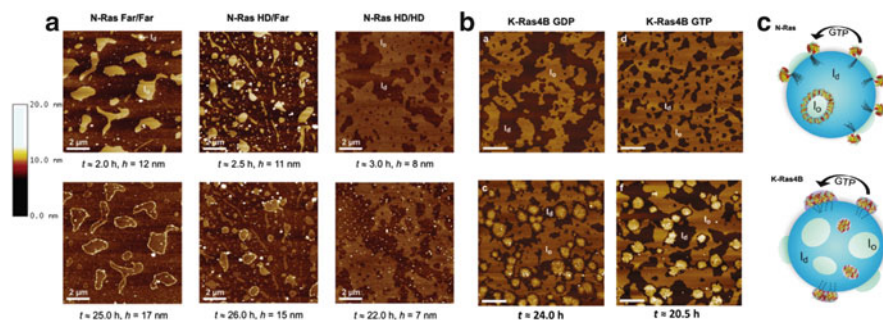


Fig. 2 (a) AFM images of the time-dependent partitioning of GDP-bound N-Ras Far/Far, N-Ras HD/Far, and N-Ras HD/HD into lipid bilayers consisting of DOPC/DPPC/Chol 1:2:1. (b) AFM images of the time-dependent partitioning of GDP- and GTP-bound K-Ras4B into lipid bilayers consisting of DOPC/DOPG/DPPC/DPPG/Chol 20:5:45:5:25. (c) Schematic model for N- and K-Ras localization in heterogeneous model biomembranes with liquid-disordered (l_d) and liquid-ordered (l_o) domains

more effective interaction with GEFs. Moreover, upon membrane interaction, high pressure induces the otherwise lowly populated state 3, which is accompanied by structural reorientations of the G domain at the lipid interface. These findings suggest that the membrane is involved in modulating Ras conformations, thereby regulating its effector and modulator interactions [105].

4.3 Structural Studies of Prenylated Rheb GTPases

Ras homologue enriched in brain (Rheb) protein is a key regulator of the mammalian target of rapamycin complex 1 (mTORC1) signalling pathway, involved in regulating cell growth, metabolism and proliferation. Similar to Ras protein, Rheb is *S*-farnesylated and methylated at its C terminal cysteine. *S*-Farnesylated Rheb (here referred to as F-Rheb) was generated by a combination of EPL and lipopeptide synthesis (Scheme 10b) [61], facilitating preparation of F-Rheb:PDE δ complex for crystallization. PDE δ was initially identified as a fourth subunit of rod-specific cGMP phosphodiesterase, PDE6. Wittinghofer and co-workers showed that PDE δ can bind and solubilize prenylated Ras, Rheb, Rho6 and G α_{i1} [106]. The structural studies of F-Rheb:PDE δ complex provide insights into the function of PDE δ as a GDI-like solubilizing factor involved in the transport of farnesylated small GTPases [107].

As shown in Fig. 3, PDE δ interacts with F-Rheb-GDP with a total buried surface area of 2,142 \AA^2 . Rheb C-terminal residues 177–181 contact PDE δ via main-chain atoms with a buried surface area of 1,007 \AA^2 which involves a PDE δ flexible loop (residues 111–117). This flexible loop is invisible in the crystal structure of PDE δ in complex with Arl2, suggesting it can adopt different conformations (Fig. 3a). The

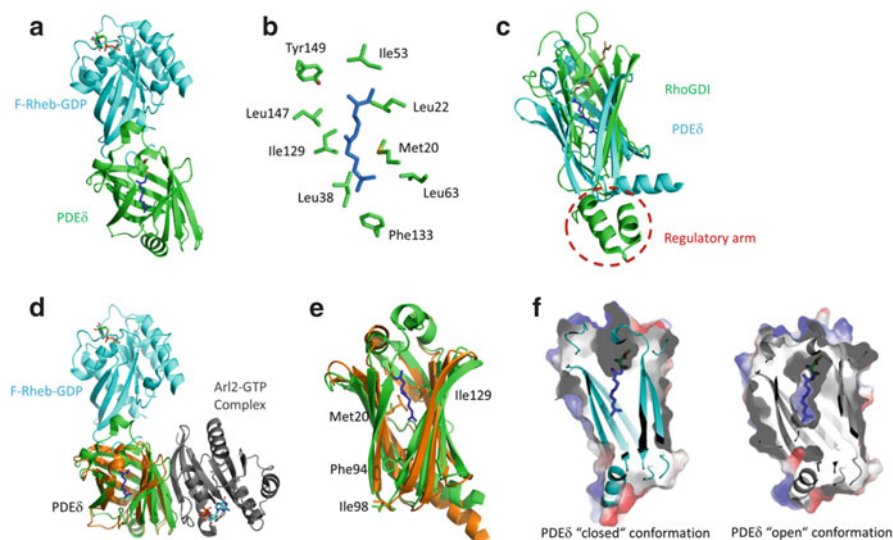


Fig. 3 Structure analysis of F-Rheb:PDE δ . (a) Ribbon representation of F-Rheb in cyan, with the farnesyl group in blue, in complex with PDE δ in green (PDB code 3T5G). GDP bound to Rheb is shown in ball-and-stick representation. (b) Residues forming the hydrophobic pocket of PDE δ are shown in green. The farnesyl group is shown in blue (PDB code 3T5I). (c) Superimposition of PDE δ in cyan on RhoGDI in green (PDB code 1DOA) with the RhoGDI regulatory arm marked by red dashed circle, the PDE δ -bound farnesyl in blue and the RhoGDI-bound geranylgeranyl group in gold. (d) Superimposition of F-Rheb:PDE δ as shown in a on the PDE δ :Arl2-GTP complex (PDB code 1KSJ) with PDE δ in orange and Arl2 in gray. (e, f) The lipid binding pockets of F-Rheb-bound PDE δ (open conformation) and Arl2-bound PGE δ (closed conformation)

main-chain interactions together with the flexibility of this loop support the notion of broad specificity of PDE δ . Several hydrophobic residues constitute the hydrophobic pocket for binding farnesyl moiety (Fig. 3b).

In contrast to RhoGDI, which features a “regulatory arm” by which it contacts Rho switch regions (Fig. 3c), there is no interaction between PDE δ and the switch regions of the Rheb. Moreover, the last three C-terminal residues of Rheb together with the farnesyl group penetrate much more deeply into the hydrophobic pocket of PDE δ , suggesting the interaction occurs mainly through the farnesylated C-terminus (Fig. 3c). These findings explain the nucleotide-independent binding of G proteins to PDE δ [106, 108]. PDE δ binds to Arl2 and Arl3 GTPases in a GTP-dependent manner [106, 109, 110] (Fig. 3d). Upon binding to Arl2-GTP, residues in the hydrophobic pocket, Met20, Ile129 and Phe94, are shifted toward the inside, leading to a clash with the farnesyl group (Fig. 3e). The conformation of PDE δ switches between the Arl2-bound closed conformation and the F-Rheb-bound open conformation (Fig. 3f). The fluorescence polarization measurements demonstrated that Arl2-GTP disrupts F-Rheb:PDE δ complex in a nucleotide-dependent manner by forming a low-affinity, rapidly dissociating ternary complex. Fluorescence lifetime imaging microscopy (FLIM) measurements also suggested

that Arl2 releases Rheb or N-Ras from PDE δ in cells. Therefore, Arl2 and Arl3 function as GDI-like displacement factors (GDFs), which allosterically regulate the release of farnesylated G proteins from PDE δ .

4.4 Thermodynamic Basis of Rab GTPases Membrane Targeting

Rab GTPases with more than 60 members in humans consist of the largest subgroup of Ras superfamily. Rab GTPases regulate vesicular transport through a spatiotemporally controlled GTPase cycle and their distinct membrane localization in cells. Cycling between the cytosol and membranes is an essential feature of the mode of action of Rabs, made possible by reversible interaction with GDP dissociation inhibitor (GDI), which can solubilize the geranylgeranylated Rab molecules in the cytosol. Membrane-bound GDI displacement factors (GDFs) were proposed to disrupt GDI:Rab complexes, leading to insertion of the prenylated Rab into the membrane in the GDP form and release of GDI into the cytosol (Scheme 17b). Since GDI is a generic regulator (only two isoforms in humans and one isoform in yeast known to date) for prenylated Rab proteins and only one GDF (Pra1 in humans and Yip3 in the yeast) with promiscuous activity on several different Rab proteins has been identified so far, it has been a perplexing question as to how individual Rabs are targeted to their cognitive membrane compartments specifically.

Elucidation of the thermodynamic basis of Rab membrane targeting requires analysis of interaction between prenylated Rab proteins (GDP/GTP-bound) and REP/GDI. Such analysis is made possible by generation of fluorescent labelled prenylated Rab proteins (Scheme 10a) [30, 54]. A series of Rab7-based protein probes with one or two isoprenyl moieties and fluorophores on the lipid moiety or the lysine side chain were prepared using the EPL technique. The semisynthetic method enables precise installation of GDP/GTP into Rab proteins to generate the “off” and “on” states, yielding for the first time homogeneous preparations of functionalized prenylated proteins in a well-defined nucleotide bound state [87].

Thermodynamic and kinetic analysis of the interaction of prenylated Rab proteins with regulatory factors provides insights into the mechanism of Rab membrane targeting. For example, Rab7 Δ 6CK(NBD)SCSC(G)-OMe (Rab7NBD-G) displays a four- to fivefold fluorescence enhancement upon binding to REP-1 or GDI-1. This signal change was used to perform fluorescence titration experiments to determine K_d values (Fig. 4a) [87]. These measurements indicated that replacement of GDP with GTP analog GppNHp leads to a reduction of the affinity of prenylated Rab proteins for their regulators REP-1 and GDI-1 by at least ca. three orders of magnitude. In the case of GTPase interaction with effector proteins, the affinity increases by several orders of magnitude on substitution of GDP by GTP. These reciprocal relationships are essential features of the Rab cycle, in which

nucleotide exchange coordinates membrane delivery, effector interactions and retrieval of Rabs from membranes.

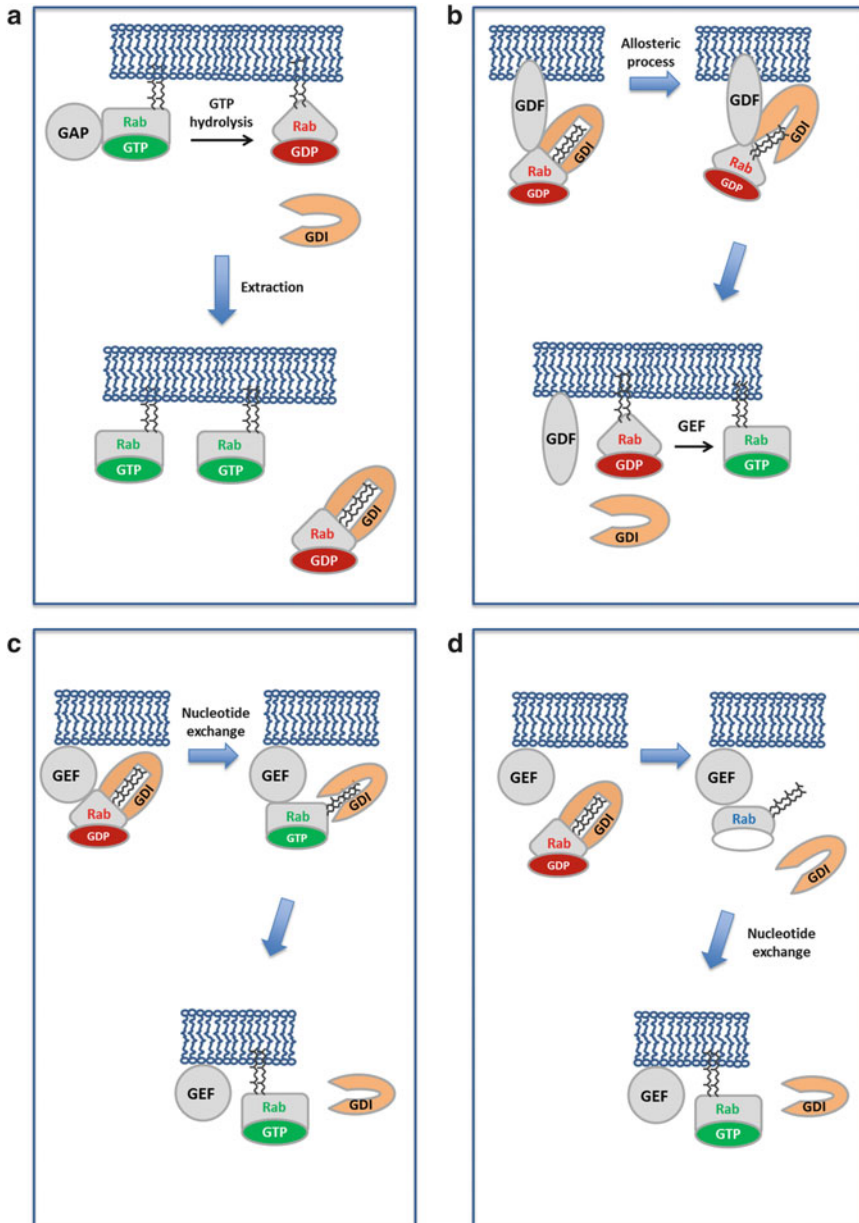
To study the relationship further between nucleotide exchange and Rab targeting to membranes, a RabGEF from *Legionella pneumophila* (DrrA) was used in investigating the effect of GEFs on the Rab:GDI complex. Kinetics of the complex interaction was monitored by a fluorescence change of Rab1-NBD-farnesyl (Rab1-NF) (Fig. 4b). DrrA-mediated exchange for GTP or GDP resulted in loss or recovery, respectively, of the Rab binding to GDI. These measurements suggest GEF activity is sufficient to disrupt Rab:GDI complex and could lead to membrane insertion.

As shown in this study, after the Rab extraction (Scheme 17a), GEF-mediated exchange of GDP for GTP dramatically reduces the affinity of Rabs to GDI and leads to an essentially irreversible dissociation of GDI. GEF-mediated nucleotide exchange plays a key role in providing the free energy to drive this process. The results obtained with DrrA suggest that GEF activity is necessary and sufficient to displace GDI (Scheme 17c), but the dissociation of the Rab:GDI complex is rate-limiting in this process (Scheme 17d). Therefore, GTP/GDP exchange catalyzed by a membrane-specific GEF is the thermodynamic determinant for the delivery to and stabilization of Rab on a particular membrane or membrane domain.

4.5 Biological Function of GPI-Anchors

Although many types of GPI-anchored proteins have been identified, the biological functions of the GPI anchor have yet to be elucidated at a molecular level. However, the structure-function relationship of GPI-anchor is difficult to study because of the heterogeneity and limited quantities of GPI-anchors from natural sources. Chemical synthesis of a series of GPI-protein analogues profoundly facilitates understanding the contribution of glycan components to the behaviour of GPI-proteins on the membrane. Bertozzi and co-workers generated fully modified GPI-anchored green fluorescent protein (GFP), which mimics the three domains of native GPI anchor (Fig. 5, Scheme 10f). The proteins were incorporated into supported lipid bilayers or loaded on the cell surface, and were analysed using fluorescence correlation spectroscopy (FCS) [74].

Native GPI-anchored proteins diffuse more rapidly in supported lipid bilayers than transmembrane proteins, presumably because the lipid tail of the GPI anchor does not extend completely through the lipid bilayer [111]. To investigate the relationship of GPI-anchor structure to the mobility on the membrane, the glycan core of GPI anchor was substituted with no (**87**), one (**88**) or two mannosyl units (**89**). These GPI anchored protein analogues were incorporated into supported lipid bilayers. The diffusion properties of GFP-2, GFP-3 and GFP-4 in supported lipid bilayers were investigated by FCS. From these FCS measurements, the characteristic correlation times (τ_D) and the diffusion coefficient (D), a physical measure of protein mobility, were obtained. GFP-4, which contains two monosaccharides in



Scheme 17 Models of modulation of Rab recycling and targeting of Rabs to membranes by the state of bound nucleotide. **(a)** The minimal model of Rab extraction. **(b)** GDF allosterically regulates GDI dissociation, followed by membrane attachment and GEF-mediated nucleotide exchange. **(c, d)** In the other models for GEF-mediated insertion, either there is direct interaction of GEF with the Rab:GDI complex, leading to **(c)** nucleotide exchange and Rab dissociation, or **(d)** spontaneous dissociation is rendered effectively irreversible by GEF activity and membrane attachment

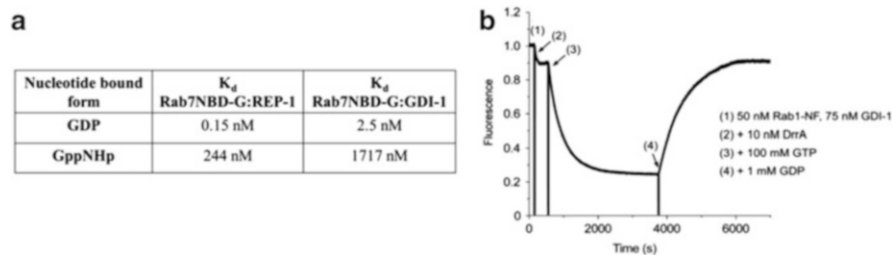


Fig. 4 Quantitative analysis of interaction of Rab7NBD-G with REP-1 and GDI-1. (a) K_d values of Rab7NBD-G interacting with REP-1 or GDI-1 in different nucleotide bound forms. (b) DrrA-mediated displacement of GDI-1. 50 nM Rab1-NF:GDI-1 complex was supplemented with 10 nM DrrA. Nucleotide exchange was triggered by adding 100 μ M GTP. Fluorescence was recovered by adding an excess of GDP (1 mM GDP)

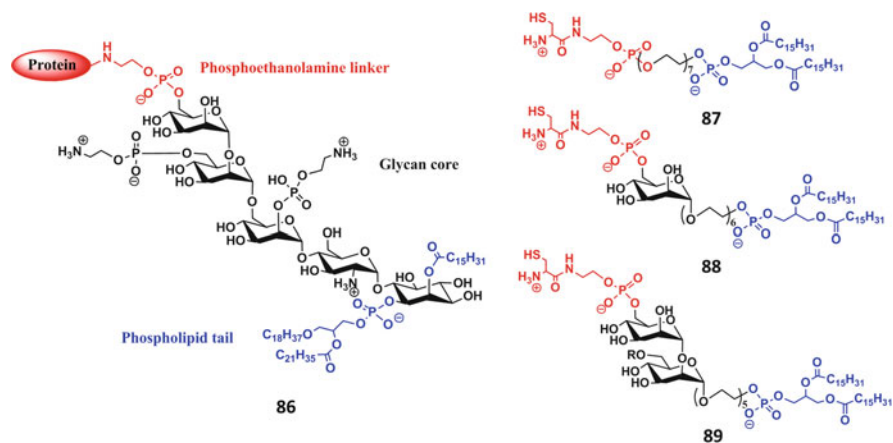


Fig. 5 Structures of native GPI-anchor (**86**) and GPI-anchor analogues (**87**), (**88**), and (**89**). These structures contain three domains of GPI-anchor: (1) a phosphoethanolamine linker (*red*), (2) the common glycan core (*black*) and (3) a phospholipid tail (*blue*). R is a GPI anchor side chain, such as galactose or phosphoethanolamine. The GPI-analogues were attached to GFP protein by EPL to produce GFP-2 (**87**), GFP-3 (**88**), GFP-4 (**89**)

the GPI anchor, diffused more rapidly than GFP-2 or GFP-3, which contains no or one monosaccharide residues, respectively (Fig. 6a, b). Moreover, the results also indicated that a protein attached to a native GPI anchor **86**, which contains four monosaccharide moieties, may move even more rapidly through the lipid bilayer.

For further elucidation of biological function of GPI anchor, the behaviour of those GPI anchor analogues together with the native GPI anchor was accessed in living cells. Transiently expressed native GPI-anchored protein, decay-accelerating factor (DAF) or the folate receptor (FR), GFP-2, GFP-3 or GFP-4 was tested on HeLa cell surface. FCS analysis revealed a correlation between the structure of the glycan core and lateral mobility in the cell membrane (Fig. 6c, d). GFP-2 displayed

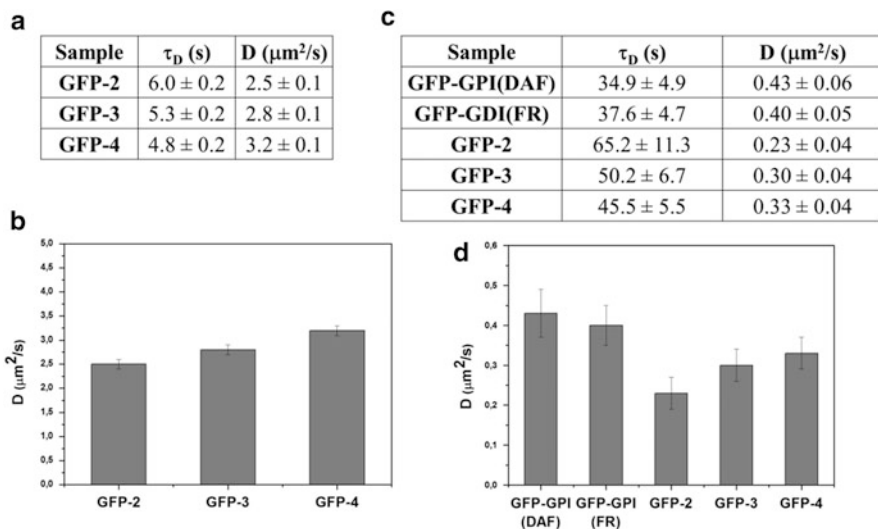


Fig. 6 Measurements of mobility by FCS. (a, b) The mobilities of GFP-2, GFP-3 and GFP-4 in a supported lipid bilayers. (c, d) The mobilities of native GPI-anchored proteins GFP-GDI (DAF), GFP-GDI (FR) and GPI analogues GFP-2, GFP-3 and GFP-4 on HeLa cell surface

significant lower diffusion kinetics than GFP-3, GFP-4 and the native GFP-GPIs. GFP-3 and GFP-4 also appeared to diffuse more slowly than the native GPI proteins. GFP-2 contains a highly flexible linker connecting to the lipid anchor. The flexible linker might permit a great movement of the protein attached. Thus, the protein may be allowed to engage in contacts with both lipid bilayer and other cell surface proteins, leading to the decrease of the mobility on the cell surface. The sugar units may rigidify the native GPI anchor so as to limit the interaction of the attached protein with the membrane, resulting in the increase of the mobility. Therefore, the GPI anchor is not only a membrane anchor, but also serves to prevent transient interactions of the attached protein with the lipid bilayer, thus permitting rapid diffusion in the membrane [74, 88].

4.6 Function of LC3-PE in Autophagosome Formation

Phosphatidylethanolaminylated LC3 family proteins (LC3-PE) are required for the elongation of autophagosomal precursors. However, the function of LC3-PE in promoting membrane tethering and hemifusion is controversial. Using in vitro reconstitution of Atg8 (LC3 in yeast) ubiquitin-like system, conjugation of yeast Atg8 to liposomes containing high concentrations (55%) of PE has been shown to promote the tethering and hemifusion of liposomes. Crosslinking of LC3 to liposomes through maleimide-coupling strategy induces membrane tethering and fusion. However, recent studies using both the reconstitution system and the

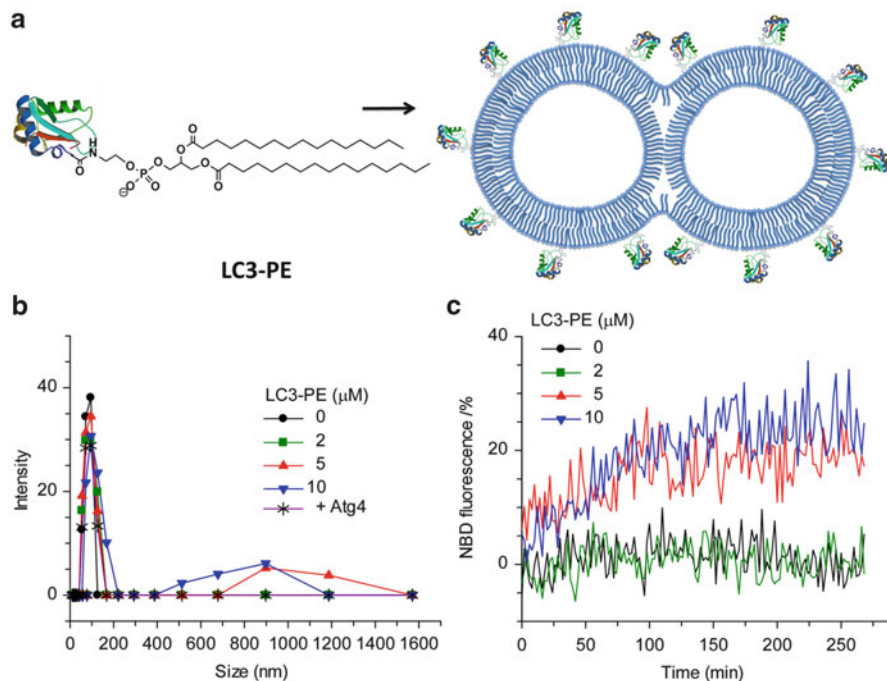


Fig. 7 Membrane tethering and fusion mediated by the semisynthetic MBP-LC3-PE protein *in vitro*. **(a)** A schematic view of LC3-PE-mediated liposomal hemifusion. **(b)** LC3-PE induces membrane tethering in a dose-dependent manner. **(c)** LC3-PE induces membrane fusion in a dose-dependent manner

maleimide-coupling strategy suggested that Atg8-PE/LC3-PE is unable to drive membrane fusion in the presence of physiological concentrations of PE (30%). Therefore, it is of great importance to be able to produce lipidated LC3 protein to study the role of LC3 in autophagosome formation. However, it is challenging to generate lipidated LC3 protein by reconstituting the LC3-PE conjugation reaction *in vitro* with purified protein components, because of the difficulties in recombinant production of mammalian proteins involved in the LC3-PE conjugation system.

Wu's lab and Liu's lab prepared LC3-PE using a semisynthetic approach (Scheme 11) [62, 64]. The semisynthetic LC3-PE allows for addressing the perplexing question on the membrane fusing activity of LC3-PE. MBP-LC3-PE was used in liposomal assays, since it is soluble in aqueous solution without detergents (Fig. 7a). The ability of MBP-LC3-PE to promote liposome tethering and fusion was determined by dynamic light scattering (DLS) and the lipid mixing assay, respectively [62]. Addition of MBP-LC3-PE to liposomes containing various concentrations of PE (30% and 55%) induced aggregate formation in a dose-dependent manner. In contrast, after treatment with catalytic amounts of Atg4B to cleavage PE, MBP-LC3-PE had no effect on liposome size distribution, in line with the fact that lipidation of LC3 is essential for membrane association and

function of LC3 (Fig. 7b). Membrane fusion activity was measured by the lipid mixing assay, in which fluorescence energy transfer from NBD-labelled lipid to rhodamine B (Rhod)-labelled lipid is reduced when a labelled liposome fuses with an unlabelled liposome. A dose-dependent induction of membrane fusion by MBP-LC3-PE was observed in the presence of 30% PE (Fig. 7c). These findings clearly demonstrate that LC3-PE mediates membrane tethering and fusion at physiological concentrations of PE.

5 Conclusions and Perspectives

Chemical approaches are invaluable means for the preparation of homogeneous lipidated proteins on a scale which permits X-ray crystal structure determination and many other biophysical studies. Moreover, chemical synthesis allows for manipulation of the protein structure and incorporation of difference functional groups into proteins. These strategies make it possible to investigate structure-function relationships, protein-protein interactions, protein-membrane interactions, intracellular localization and function of lipidated proteins *in vitro* and in cells. A combination of chemistry and biology has allowed the study of biological functions previously not possible through traditional biochemical approaches.

The toolbox of chemoselective methods for protein synthesis and modification has substantially expanded in the past few years [46, 112–115]. Many of these reactions proceed under physiological conditions, which are compatible with biological system. These methods allow for application of powerful synthetic chemistry to the modification of proteins. In particular, the recent development of bioorthogonal and rapid ligation reactions make it possible to label protein in cells and organisms [116]. In principle, many of these reactions are applicable to the synthesis of lipidated proteins to improve the yield and reduce the reaction time. In this sense, chemical ligation of lipidated proteins in cells would also be possible. In addition to the development of new ligation methods, another important issue for the synthesis of lipidated proteins is solubilization of lipidated peptides and proteins. Detergents have been shown to be a useful strategy. In some cases, they serve not only as a solubilizer but also as a catalysis for the ligation. However, detergents are usually not easily removed, and the presence of detergents could affect protein function. A high demand remains for assisted solubilisation techniques for the synthesis of lipidated proteins in a detergent-free manner. It is conceivable that in future many other lipidated proteins with diverse lipid modifications can be prepared and become essential tools for elucidation of various biological processes.

References

1. Resh MD (2006) *Nat Chem Biol* 2:584–590
2. Carr SA, Biemann K, Shoji S, Parmelee DC, Titani K (1982) *Proc Natl Acad Sci U S A* 79:6128–6131
3. Aitken A, Cohen P, Santikarn S, Williams DH, Calder AG, Smith A, Klee CB (1982) *FEBS Lett* 150:314–318
4. Farazi TA, Waksman G, Gordon JI (2001) *J Biol Chem* 276:39501–39504
5. Gheysen D, Jacobs E, de Foresta F, Thiriart C, Francotte M, Thines D, De Wilde M (1989) *Cell* 59:103–112
6. Liu Y, Kahn RA, Prestegard JH (2009) *Structure* 17:79–87
7. Hutagalung AH, Novick PJ (2011) *Physiol Rev* 91:119–149
8. Stenmark H (2009) *Nat Rev Mol Cell Biol* 10:513–525
9. Amor JC, Harrison DH, Kahn RA, Ringe D (1994) *Nature* 372:704–708
10. Goldberg J (1998) *Cell* 95:237–248
11. Maurer-Stroh S, Eisenhaber F (2004) *Trends Microbiol* 12:178–185
12. Pal R, Reitz MS Jr, Tschachler E, Gallo RC, Sarnagadharan MG, Veronese FD (1990) *AIDS Res Hum Retroviruses* 6:721–730
13. Provitera P, El-Maghrabi R, Scarlata S (2006) *Biophys Chem* 119:23–32
14. Casey PJ (1995) *Science* 268:221–225
15. Resh MD (1999) *Biochim Biophys Acta* 1451:1–16
16. Kang R, Wan J, Arstikaitis P, Takahashi H, Huang K, Bailey AO, Thompson JX, Roth AF, Drisdell RC, Mastro R, Green WN, Yates JR 3rd, Davis NG, El-Husseini A (2008) *Nature* 456:904–909
17. Wedegaertner PB, Bourne HR (1994) *Cell* 77:1063–1070
18. Rocks O, Peyker A, Kahms M, Verveer PJ, Koerner C, Lumbierres M, Kuhlmann J, Waldmann H, Wittinghofer A, Bastiaens PI (2005) *Science* 307:1746–1752
19. Tu Y, Wang J, Ross EM (1997) *Science* 278:1132–1135
20. Rocks O, Gerauer M, Vartak N, Koch S, Huang ZP, Pechlivanis M, Kuhlmann J, Brunsveld L, Chandra A, Ellinger B, Waldmann H, Bastiaens PI (2010) *Cell* 141:458–471
21. Chandra A, Grecco HE, Pisupati V, Perera D, Cassidy L, Skoulidis F, Ismail SA, Hedberg C, Hanzal-Bayer M, Venkitaraman AR, Wittinghofer A, Bastiaens PI (2012) *Nat Cell Biol* 14:148–158
22. Zhang FL, Casey PJ (1996) *Annu Rev Biochem* 65:241–269
23. Casey PJ, Seabra MC (1996) *J Biol Chem* 271:5289–5292
24. Maurer-Stroh S, Washietl S, Eisenhaber F (2003) *Genome Biol* 4:212
25. Nguyen UT, Goody RS, Alexandrov K (2010) *Chembiochem* 11:1194–1201
26. Clarke S (1992) *Annu Rev Biochem* 61:355–386
27. Wu YW, Goody RS, Abagyan R, Alexandrov K (2009) *J Biol Chem* 284:13185–13192
28. Li F, Yi L, Zhao L, Itzen A, Goody RS, Wu YW (2014) *Proc Natl Acad Sci U S A* 111:2572–2577
29. Rak A, Pylypenko O, Durek T, Watzke A, Kushnir S, Brunsveld L, Waldmann H, Goody RS, Alexandrov K (2003) *Science* 302:646–650
30. Wu YW, Tan KT, Waldmann H, Goody RS, Alexandrov K (2007) *Proc Natl Acad Sci U S A* 104:12294–12299
31. Yu S, Guo Z, Johnson C, Gu G, Wu Q (2013) *Curr Opin Chem Biol* 17:1006–1013
32. Tsai YH, Liu X, Seeberger PH (2012) *Angew Chem Int Ed Engl* 51:11438–11456
33. Paulick MG, Bertozzi CR (2008) *Biochemistry* 47:6991–7000
34. Levental I, Grzybek M, Simons K (2010) *Biochemistry* 49:6305–6316
35. Nosjean O, Briolay A, Roux B (1997) *Biochim Biophys Acta* 1331:153–186
36. Ishihara A, Hou Y, Jacobson K (1987) *Proc Natl Acad Sci U S A* 84:1290–1293
37. Schofield L, McConville MJ, Hansen D, Campbell AS, Fraser-Reid B, Grusby MJ, Tachado SD (1999) *Science* 283:225–229

38. Weissmann C (2004) *Nat Rev Microbiol* 2:861–871
39. Chesebro B, Trifilo M, Race R, Meade-White K, Teng C, LaCasse R, Raymond L, Favara C, Baron G, Priola S, Caughey B, Masliah E, Oldstone M (2005) *Science* 308:1435–1439
40. Zuegg J, Gready JE (2000) *Glycobiology* 10:959–974
41. Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ (1997) *Cell* 91:85–97
42. Buglino JA, Resh MD (2008) *J Biol Chem* 283:22076–22088
43. Guerrero I, Chiang C (2007) *Trends Cell Biol* 17:1–5
44. Mizushima N, Yoshimori T, Ohsumi Y (2011) *Annu Rev Cell Dev Biol* 27:107–132
45. Feng Y, He D, Yao Z, Klionsky DJ (2014) *Cell Res* 24:24–41
46. Hackenberger CPR, Schwarzer D (2008) *Angew Chem Int Ed Engl* 47:10030–10074
47. Brunsveld L, Kuhlmann J, Alexandrov K, Wittinghofer A, Goody RS, Waldmann H (2006) *Angew Chem Int Ed Engl* 45:6622–6646
48. Gerauer M, Koch S, Brunsveld L, Waldmann H (2009) Lipidated peptide synthesis. In: Begley TP (ed) *Wiley encyclopedia of chemical biology*, vol 2. Wiley, New Jersey, pp 520–530
49. Görmer K, Waldmann H, Brunsveld L (2010) Lipidation of peptides and proteins. In: Mander L, Lui H-W (eds) *Comprehensive natural products II: chemistry and biology*, vol 5. Elsevier, Oxford, pp 531–585
50. Lumbierres M, Palomo JM, Kragol G, Roehrs S, Muller O, Waldmann H (2005) *Chemistry* 11:7405–7415
51. Kadereit D, Deck P, Heinemann I, Waldmann H (2001) *Chemistry* 7:1184–1193
52. Kadereit D, Waldmann H (2000) *Chembiochem* 1:200–203
53. Brunsveld L, Watzke A, Durek T, Alexandrov K, Goody RS, Waldmann H (2005) *Chemistry* 11:2756–2772
54. Durek T, Alexandrov K, Goody RS, Hildebrand A, Heinemann I, Waldmann H (2004) *J Am Chem Soc* 126:16368–16378
55. Millington CR, Quarrell R, Lowe G (1998) *Tetrahedron Lett* 39:7201–7204
56. Ludolph B, Eisele F, Waldmann H (2002) *J Am Chem Soc* 124:5954–5955
57. Rosenbaum C, Waldmann H (2001) *Tetrahedron Lett* 42:5677–5680
58. Backes BJ, Ellman JA (1999) *J Org Chem* 64:2322–2330
59. Triola G, Gerauer M, Görmer K, Brunsveld L, Waldmann H (2010) *Chemistry* 16:9585–9591
60. Palomo JM, Lumbierres M, Waldmann H (2006) *Angew Chem Int Ed Engl* 45:477–481
61. Chen YX, Koch S, Uhlenbrock K, Weise K, Das D, Gremer L, Brunsveld L, Wittinghofer A, Winter R, Triola G, Waldmann H (2010) *Angew Chem Int Ed Engl* 49:6090–6095
62. Yang A, Li Y, Pantoom S, Triola G, Wu YW (2013) *Chembiochem* 14:1296–1300
63. Grogan MJ, Kaizuka Y, Conrad RM, Groves JT, Bertozzi CR (2005) *J Am Chem Soc* 127:14383–14387
64. Huang YC, Li YM, Chen Y, Pan M, Li YT, Yu L, Guo QX, Liu L (2013) *Angew Chem Int Ed Engl* 52:4858–4862
65. Peters C, Wolf A, Wagner M, Kuhlmann J, Waldmann H (2004) *Proc Natl Acad Sci U S A* 101:8531–8536
66. Goody RS, Alexandrov K, Engelhard M (2002) *Chembiochem* 3:399–403
67. Kho Y, Kim SC, Jiang C, Barma D, Kwon SW, Cheng J, Jaunbergs J, Weinbaum C, Tamanoi F, Falck J, Zhao Y (2004) *Proc Natl Acad Sci U S A* 101:12479–12484
68. Nguyen UT, Cramer J, Gomis J, Reents R, Gutierrez-Rodriguez M, Goody RS, Alexandrov K, Waldmann H (2007) *Chembiochem* 8:408–423
69. Duckworth BP, Zhang Z, Hosokawa A, Distefano MD (2007) *Chembiochem* 8:98–105
70. Dursina B, Reents R, Delon C, Wu Y, Kulharia M, Thutewohl M, Veligodsky A, Kalinin A, Evstifeev V, Ciobanu D, Szedlaczek SE, Waldmann H, Goody RS, Alexandrov K (2006) *J Am Chem Soc* 128:2822–2835
71. Nguyen UT, Guo Z, Delon C, Wu Y, Deraeve C, Franzel B, Bon RS, Blankenfeldt W, Goody RS, Waldmann H, Wolters D, Alexandrov K (2009) *Nat Chem Biol* 5:227–235

72. Berry AF, Heal WP, Tarafder AK, Tolmachova T, Baron RA, Seabra MC, Tate EW (2010) *ChemBiochem* 11:771–773
73. Yi L, Abootorabi M, Wu YW (2011) *ChemBiochem* 12:2413–2417
74. Paulick MG, Wise AR, Forstner MB, Groves JT, Bertozzi CR (2007) *J Am Chem Soc* 129:11543–11550
75. Alexandrov K, Heinemann I, Durek T, Sidorovitch V, Goody RS, Waldmann H (2002) *J Am Chem Soc* 124:5648–5649
76. Olschewski D, Seidel R, Miesbauer M, Rambold AS, Oesterhelt D, Winkelhofer KF, Tatzelt J, Engelhard M, Becker CF (2007) *Chem Biol* 14:994–1006
77. Antos JM, Miller GM, Grotenbreg GM, Ploegh HL (2008) *J Am Chem Soc* 130:16338–16343
78. Reents R, Wagner M, Kuhlmann J, Waldmann H (2004) *Angew Chem Int Ed Engl* 43:2711–2714
79. Chu NK, Becker CF (2009) *Methods Enzymol* 462:177–193
80. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) *Science* 266:776–779
81. Dawson PE, Kent SB (2000) *Annu Rev Biochem* 69:923–960
82. Muir TW (2003) *Annu Rev Biochem* 72:249–289
83. Muir TW, Sondhi D, Cole PA (1998) *Proc Natl Acad Sci U S A* 95:6705–6710
84. Siman P, Brik A (2012) *Org Biomol Chem* 10:5684–5697
85. Vila-Perello M, Muir TW (2010) *Cell* 143:191–200
86. Kuhn K, Owen DJ, Bader B, Wittinghofer A, Kuhlmann J, Waldmann H (2001) *J Am Chem Soc* 123:1023–1035
87. Wu YW, Oesterlin LK, Tan KT, Waldmann H, Alexandrov K, Goody RS (2010) *Nat Chem Biol* 6:534–540
88. Paulick MG, Forstner MB, Groves JT, Bertozzi CR (2007) *Proc Natl Acad Sci U S A* 104:20332–20337
89. Becker CF, Liu X, Olschewski D, Castelli R, Seidel R, Seeberger PH (2008) *Angew Chem Int Ed Engl* 47:8215–8219
90. Johnson EC, Kent SB (2006) *J Am Chem Soc* 128:6640–6646
91. Janosch S, Nicolini C, Ludolph B, Peters C, Völkert M, Hazlet TL, Gratton E, Waldmann H, Winter R (2004) *J Am Chem Soc* 126:7496–7503
92. Reents R, Wagner M, Schlummer S, Kuhlmann J, Waldmann H (2005) *ChemBiochem* 6:86–94
93. Bader B, Kuhn K, Owen DJ, Waldmann H, Wittinghofer A, Kuhlmann J (2000) *Nature* 403:223–226
94. de Araujo AD, Palomo JM, Cramer J, Kohn M, Schroder H, Wacker R, Niemeyer C, Alexandrov K, Waldmann H (2005) *Angew Chem Int Ed Engl* 45:296–301
95. de Araujo AD, Palomo JM, Cramer J, Seitz O, Alexandrov K, Waldmann H (2006) *Chemistry* 12:6095–6109
96. Kolb HC, Finn MG, Sharpless KB (2001) *Angew Chem Int Ed Engl* 40:2004–2021
97. Huang X, Aulabaugh A, Ding W, Kapoor B, Alksne L, Tabei K, Ellestad G (2003) *Biochemistry* 42:11307–11315
98. Frankel BA, Kruger RG, Robinson DE, Kelleher NL, McCafferty DG (2005) *Biochemistry* 44:11188–11200
99. Dementiev A (2012) *Protein Expr Purif* 84:86–93
100. Wu Z, Guo X, Gao J, Guo Z (2013) *Chem Commun (Camb)* 49:11689–11691
101. de Almeida RF, Fedorov A, Prieto M (2003) *Biophys J* 85:2406–2416
102. Nicolini C, Baranski J, Schlummer S, Palomo J, Lumbierres-Burgues M, Kahms M, Kuhlmann J, Sanchez S, Gratton E, Waldmann H, Winter R (2006) *J Am Chem Soc* 128:192–201
103. Weise K, Triola G, Brunsveld L, Waldmann H, Winter R (2009) *J Am Chem Soc* 131:1557–1564
104. Weise K, Kapoor S, Denter C, Nikolaus J, Opitz N, Koch S, Triola G, Herrmann A, Waldmann H, Winter R (2011) *J Am Chem Soc* 133:880–887

105. Kapoor S, Triola G, Vetter IR, Erlkamp M, Waldmann H, Winter R (2012) *Proc Natl Acad Sci U S A* 109:460–465
106. Hanzal-Bayer M, Renault L, Roversi P, Wittinghofer A, Hillig RC (2002) *EMBO J* 21:2095–2106
107. Ismail SA, Chen YX, Rusinova A, Chandra A, Bierbaum M, Gremer L, Triola G, Waldmann H, Bastiaens PI, Wittinghofer A (2011) *Nat Chem Biol* 7:942–949
108. Nancy V, Callebaut I, El Marjou A, de Gunzburg J (2002) *J Biol Chem* 277:15076–15084
109. Linari M, Hanzal-Bayer M, Becker J (1999) *FEBS Lett* 458:55–59
110. Veltel S, Kravchenko A, Ismail S, Wittinghofer A (2008) *FEBS Lett* 582:2501–2507
111. Groves JT, Dustin ML (2003) *J Immunol Methods* 278:19–32
112. Sletten EM, Bertozzi CR (2009) *Angew Chem Int Ed Engl* 48:6974–6998
113. Ramil CP, Lin Q (2013) *Chem Commun (Camb)* 49:11007–11022
114. Wu YW, Goody RS (2010) *J Pept Sci* 16:514–523
115. Takaoka Y, Ojida A, Hamachi I (2013) *Angew Chem Int Ed Engl* 52:4088–4106
116. Lang K, Chin JW (2014) *Chem Rev* 114:4764–4806