Protein Modification

DOI: 10.1002/anie.200602979

# Linking Chemistry and Biology for the Study of Protein Function

*Daniel Rauh* and Herbert Waldmann*

**Keywords:**
native chemical ligation · phosphatase · protein modifications · proteins · semisynthesis

Within the last decade, a wealth of methods have been developed to manipulate protein structure and activity to unravel critical details of protein function at the molecular and cellular level. Notably, the use of genetic tools such as site-directed mutagenesis, knock-out and knock-in techniques, and the use of encoded reporters such as fluorescent proteins have facilitated the discovery of many protein functions within their molecular networks.

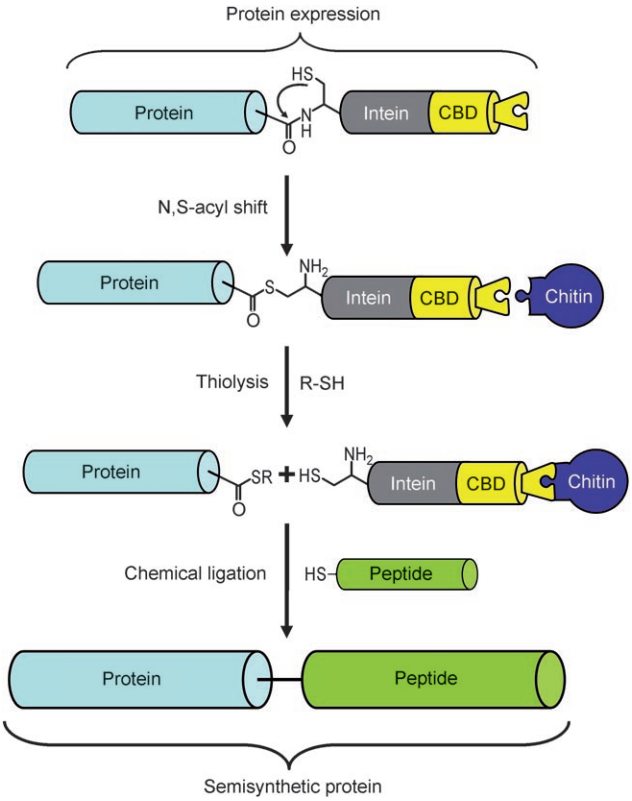
One of the next great challenges in biology is understanding the molecular details of how post-translational modifications affect protein function, stability, cellular localization, three-dimensional structure, and interaction with other molecules. Many of these biological investigations of protein function require access to protein molecules that are hard to obtain with purely biological methods such as site-directed mutagenesis and recombinant protein expression. The combination of chemical and biological techniques to chemoselectively modify proteins, however, has proved to be an excellent resource for studying protein function on a molecular level.

Based on early work by Wieland et al.,[1] a synthetic method known as native chemical ligation (NCL) has evolved to be a powerful approach in peptide chemistry. NCL allows the construction of large peptides through the condensation of two or more smaller peptides with C-terminal thioesters and N-terminal cysteines.[2] The semisynthetic version of NCL, known as expressed protein ligation (EPL), combines chemical synthesis with biological techniques and enables the fusion of synthetically produced peptides with recombinantly produced proteins.[3,4] EPL facilitates site-specific protein labeling with a broad range of physical probes, such as fluorophores, post-translational modifications, stable isotopes, spin labels, and unnatural amino acids, and has been applied to a wide range of protein-engineering problems[5] (Scheme 1).

Structural studies of post-translationally modified proteins are often hampered by standard protein-expression approaches simply owing to inhomogeneous protein preparations. The use of EPL for the detailed X-ray crystal structure analysis of proteins with modifications such as phosphorylation and prenylation proved to be extremely powerful in a variety of cases.[6] In a recent example, semisynthetic modification and protein engineering yielded preparative amounts of a mono- and diprenylated version of the GTPase Ypt1 and allowed, for the first time, crystallization and structure determination of Ypt1 in complex with its physiological modulator Rab GDP dissociation inhibitor (RabGDI).[7] Rab proteins are critical regulators of vesicular membrane transport and mediate various events such as docking, fusion, and the mobility of intracellular membranes. In this way, post-translational modifications are vital for protein function. The X-ray crystal structure of the Ypt1:RabGDI complex revealed that RabGDI undergoes a conformational change upon binding of Ypt1 to create a hydrophobic binding cavity for the prenyl group, which is otherwise anchored to the plasma membrane. The work by Goody and co-workers uncovered the location of the prenyl binding site, which was previously believed to be at a different location (Figure 1). Thus, further investigations led to the delineation of the mechanism by which Rab effectors, such as GDI and Rab escort proteins (REPs), deliver Rabs to and extract them from vesicular membranes. As these mutations in RabGDI are associated with mental retardation, biophysical and structural biological studies clarified the mode of action of this mutation: Rab proteins, like the prenylated Ypt1, are less-efficiently extracted from the membrane, and as such, membrane transport is disturbed, which is the molecular cause for this hereditary disease.[8,10]

Among other post-translational modifications, phosphorylation and dephosphorylation are the most important ways in which cellular proteins are modified to modulate their function and to transduce information between distinct cellular compartments within one cell through signaling pathways. To achieve the essential specificity to opti-

[*] Dr. D. Rauh
Chemical Genomics Centre of the Max-Planck-Society
Otto-Hahn-Strasse 15, 44227 Dortmund (Germany)
Fax: (+49) 231-9742-6479
E-mail: daniel.rauh@cgc.mpg.de
Prof. H. Waldmann
Department of Chemical Biology
Max-Planck-Institut für molekulare Physiologie
Otto-Hahn-Strasse 11, 44227 Dortmund (Germany)
And
Chemische Biologie
Fachbereich Chemie
Universität Dortmund
44227 Dortmund (Germany)
Fax: (+49) 231-133-2499
E-mail: herbert.waldmann@mpi-dortmund.mpg.de




***Scheme 1.*** Schematic diagram of EPL. A protein splicing element (intein), combined with an affinity tag such as a chitin binding domain (CBD) is fused to the C terminus of the target protein. The N-terminal cysteine of the intein tag initiates an N,S acyl shift by a nucleophilic attack on the carbonyl carbon atom. Thiolysis by exogenous thiols results in the recombinant protein thioester and the cleaved intein tag. The C-terminal thioester of the target protein reacts with an N-terminal cysteine of the synthetic peptide through native chemical ligation and forms the final semisynthetic protein. Beads with immobilized chitin facilitate efficient separation and purification of the constructs. Commercial systems have been developed for the intein-mediated purification with an affinity chitin-binding tag.[31]

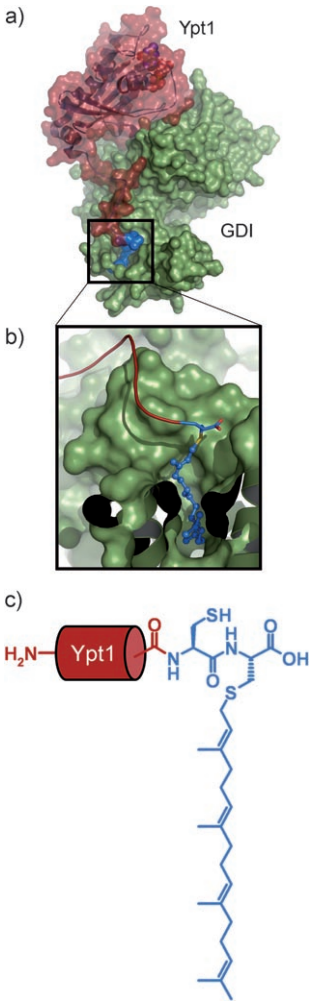



***Figure 1.*** Structure of the monoprenylated Ypt1:RabGDI complex (PDB code: 1ukv[7]). a) Surface and ribbon representation of RabGDI (green) bound to Ypt1 (red). The isoprenoid moiety is displayed in spheres (blue). b) Enlarged surface representation of the lipid binding site of RabGDI (green). The semisynthetic C terminus of Ypt1 is shown in blue and the isoprenoid chain is displayed in a ball and stick representation. The sulfur atom of the modified cysteine is colored yellow. c) Schematic representation of prenylated Ypt1. The synthetic dipiptide (blue) is ligated to the C terminus of Ypt1 (red).

mize modulation of enzyme activity in signaling events, nature often uses compartmentalization of protein kinases and phosphatases. The balanced interplay between protein kinases and phosphatases plays a crucial role in the control of cell signaling cascades.

As two of the largest families of enzymes in eukaryotes, kinases and (their physiological counterpart) phosphatases are involved in controlling both normal cellular processes as well as disregulated pathways. Aberrantly regulated protein kinases and phosphatases have been shown to play causative roles in diseases such as cancer, diabetes, and neurological and autoimmune disorders, making these classes of enzymes an important set of therapeutic targets across almost all disease areas.[11] However, a detailed understanding of protein phosphorylation and dephosphorylation is key to the future development of innovative medicine.[12]

With the microinjection of semisynthetic variants of the tyrosine phosphatase SHP-2 (bearing nonhydrolyzable phosphotyrosine analogues) into cells, Cole and co-workers investigated the role of phosphorylation of SHP-2 in the activation of the MAP kinase signaling pathway and paved the road of EPL applications to systems in vivo.[13] In a very recent article, the same group set out to investigate the role of phosphorylation in the function of the low-molecular-weight protein tyrosine phosphatase (LMW-PTP),[14] a widely expressed 18-kDa enzyme that is important in cell signaling events such as growth-factor-dependent mitogenesis, focal adhesion, and cell mobility.[15] The activity of LMW-PTP has initially been shown to be dependent on the response to growth-factor stimulation by phosphorylation of two tyrosine side chains (Tyr131 and Tyr132) by either platelet-derived growth factor (PDGF) kinase, Src kinase, or both.

Despite examples in literature, the detailed biochemical analysis of phosphorylation patterns on the function of tyrosine phosphatase is difficult to access owing to the intrinsic tendency of these enzymes to autodephosphorylate especially under physiological condi-

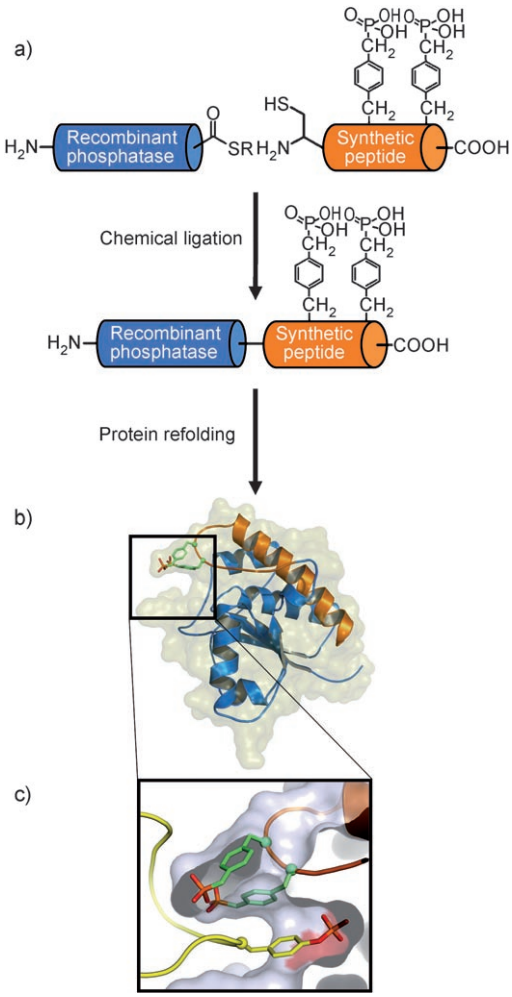
  

tions. In an elegant method, the authors circumvented the limitations of heterogeneous protein preparation by incorporating phosphonomethylene-L-phenylalanine (Pmp) as a known nonhydrolysable phosphotyrosine analogue. The phosphonate mimics phosphorylation at the previously predicted hotspots located on the C-terminal part of protein tyrosine phosphatase (PT-phosphatase). The N-terminal domain of human PT-phosphatase was produced as the thioester by fusion to an intein tag through recombinant techniques expressed in *Escherichia coli*. The C-terminal fragment of PT-phosphatase was synthesized on the solid phase as the *N*-Cys-containing peptide, equipped with the artificial residues (Pmp) at the two phosphorylation sites of interest. NCL was used for site-directed linking of the native N-terminal and the synthetic C-terminal domains of human PT-phosphatase. The new assembly was refolded by dialysis and purified to yield an enzymatically active PT-phosphatase (Scheme 2). The biochemical characterization of the semisynthetic LMW-PTP variants revealed a decreased, rather than the previously proposed increased,[16,17] enzyme activity when phosphotyrosine peptides derived from putative cellular substrates (PDGF and p190RhoGap) were used. These findings highlight the first example of a tyrosine phosphatase that is inhibited by phosphorylation and provide a novel model for the regulation of LMW-PTP and its functional role in cell signaling.

The work of Cole and co-workers[14] demonstrates, in a very convincing manner, the power of semisynthetic approaches to the study of protein function. Typically, the study of the biological function of enzymes such as phosphatases includes the application of knock-down and knock-out techniques (genetic) as well as the use of enzyme inhibitors (chemical genetics). However, both approaches would not have been applicable to the study of LMW-PTP described above as these methods do not give direct control of specific post-translational modifications to the protein of interest.[18–20] Thus, the synthesis and use of a semisynthetic phosphatase opened up an otherwise inaccessible opportunity to study protein function.

Aside from the examples of semisynthetically generated proteins for the investigation of lipidation and phosphorylation presented above, several groups have extended the use of ligation techniques to study the functional and mechanistic role of additional post-translational protein modifications such as the complex nature of glycosylation patterns.[21,22]

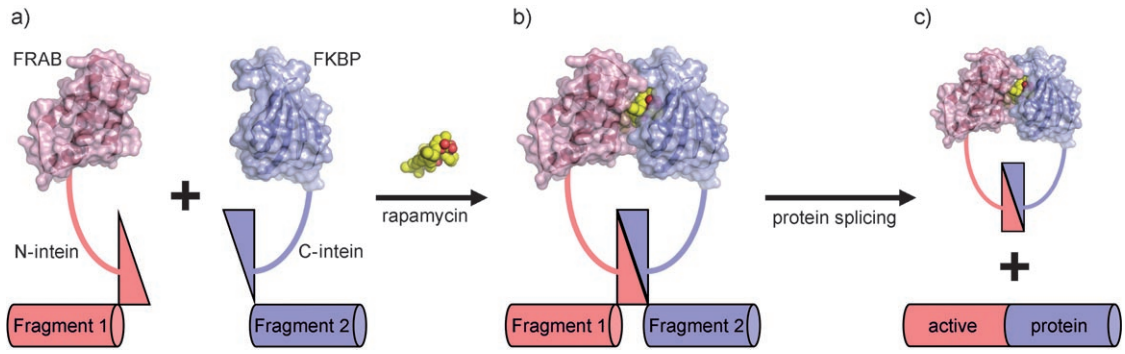
Although the methods presented herein have proven to be very powerful, certain limitations in the field of NCL and EPL still exist. One of the major limitations is the requirement for an N-terminal cysteine residue. Not every construct or design strategy allows for the presence of a cysteine at the site of ligation. For this reason, extension of the methodology to form native peptide bonds to an amino acid other than cysteine is desired. Approaches to overcome the requirement for cysteine residues might be found in the use of thiol auxiliaries that mimic cysteines and permit NCL-type reactions to occur when attached to the N terminus of the peptide[22,23] or by facilitating enzymes for the peptide-bond coupling.[24] As an additional limiting factor, under the chosen reaction conditions, the ligation steps tend to be compromised by low yields and the tendency of the protein constructs to denature . Although smaller proteins were successfully synthesized by NCL techniques on, for example, a water-compatible solid support,[25] the sequential ligation of peptides to form larger proteins remains a challenge for the future.




***Scheme 2.*** Generation of the semisynthetic LMW-PTP constructs through chemical ligation. a) The recombinant thioester of the phosphatase (blue) is ligated with the synthetic peptide (orange). Nonhydrolyzable phosphotyrosine mimics (Pmp, phenolic oxygen of the phosphotyrosine is substituted by a methylene) are incorporated into the synthetic peptides (orange). b) The semisynthetic construct is refolded to form biologically active LMW-PTP. c) Enlarged surface representation of the phosphatase binding site. Phosphorylation sites (green) of LMW-PTP are located at the substrate binding pocket. A putative phosphotyrosine substrate (yellow) is fitted into the binding cleft close to the catalytic center (red). The modeled structure is based on LMW-PTP in complex with a phosphate (PDB code: 1pnt[32]). Figure prepared with PyMol.[33]

However, the power of protein ligation to investigate protein function is evident and the technology is constantly improving,[26,27] aiming towards minimally invasive methods that allow analysis of target proteins within the ensemble of all other proteins in vivo. Along these lines, Mootz, Muir, and co-workers developed a technique in vivo for the functional reassembly of a target protein from inactive parts that are triggered in response to a small cell-permeable molecule[28–30] (Scheme 3). The method is called conditional protein splicing and is based on the fusion of inactive target protein domains to a split intein. The inteins are fused to either the rapamycin receptor or the rapamycin binding domain. The splicing of the protein frag-




***Scheme 3.*** Schematic diagram of conditional protein splicing. a) The protein of interest is split into two inactive fragments (1 and 2). The fragments are fused to one half of a split intein (N intein and C intein) and to either the FKBP-rapamycin-associated protein (FRAP) or the FK506-binding protein (FKBP). The fusion constructs are expressed in cells. b) On addition of the low-molecular-weight cell-permeable rapamycin (yellow = carbon atoms, red = oxygen atoms) to the cells, dimerization is induced and the intein halves are reconstituted. c) The reconstituted inteins initiates the splicing reaction of fragments 1 and 2 to form the active protein.

ments to the functionally active protein of interest in vivo occurs through dimerization of the split inteins only in the presence of rapamycin, thus providing a methodology to switch on rather than inhibit protein function by chemical entities. This underlines the extension of ligation techniques to the emerging area of chemical genetics. The technique provides innovative tools to probe protein function through perturbation of protein activity or protein–protein interactions in a rapid tunable manner that is necessary to track protein function within the dynamic and variable nature of biological processes.

Published online: December 29, 2006


[1] T. Wieland, E. Bockelmann, L. Bauer, H. U. Lang, H. Lau, *Justus Liebigs Ann. Chem.* **1958**, *583*, 129.

[2] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776.

[3] T. C. Evans, Jr., J. Benner, M. Q. Xu, *Protein Sci.* **1998**, *7*, 2256.

[4] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705.

[5] T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249.

[6] J. W. Wu, M. Hu, J. Chai, J. Seoane, M. Huse, C. Li, D. J. Rigotti, S. Kyin, T. W. Muir, R. Fairman, J. Massague, Y. Shi, *Mol. Cell* **2001**, *8*, 1277.

[7] L. Brunsveld, J. Kuhlmann, K. Alexandrov, A. Wittinghofer, R. S. Goody, H. Waldmann, *Angew. Chem.* **2006**, *118*, 6774; *Angew. Chem. Int. Ed.* **2006**, *45*, 6622; M. H. Gelb, L. Brunsveld, C. A. Hrycyna, S. Michaelis, F. Tamanoi, W. C. Van Voorhis, H. Waldmann, *Nat. Chem. Biol.* **2006**, *2*, 518.

[8] O. Pylypenko, A. Rak, T. Durek, S. Kushnir, B. E. Dursina, N. H. Thomae, A. T. Constantinescu, L. Brunsveld, A. Watzke, H. Waldmann, R. S. Goody, K. Alexandrov, *EMBO J.* **2006**, *25*, 13.

[9] A. Rak, O. Pylypenko, T. Durek, A. Watzke, S. Kushnir, L. Brunsveld, H. Waldmann, R. S. Goody, K. Alexandrov, *Science* **2003**, *302*, 646.

[10] R. S. Goody, A. Rak, K. Alexandrov, *Cell. Mol. Life Sci.* **2005**, *62*, 1657.

[11] L. Bialy, H. Waldmann, *Angew. Chem.* **2005**, *117*, 3880; *Angew. Chem. Int. Ed.* **2005**, *44*, 3814.

[12] A. Östman, C. Hellberg, F. D. Bohmer, *Nat. Rev. Cancer* **2006**, *6*, 307.

[13] W. Lu, D. Gong, D. Bar-Sagi, P. A. Cole, *Mol. Cell* **2001**, *8*, 759.

[14] D. Schwarzer, Z. Zhang, W. Zheng, P. A. Cole, *J. Am. Chem. Soc.* **2006**, *128*, 4192.

[15] G. Raugei, G. Ramponi, P. Chiarugi, *Cell. Mol. Life Sci.* **2002**, *59*, 941.

[16] M. Bucciantini, P. Chiarugi, P. Cirri, L. Taddei, M. Stefani, G. Raugei, P. Nordlund, G. Ramponi, *FEBS Lett.* **1999**, *456*, 73.

[17] P. Tailor, J. Gilman, S. Williams, C. Couture, T. Mustelin, *J. Biol. Chem.* **1997**, *272*, 5371.

[18] K. Hinterding, A. Knebel, P. Herrlich, H. Waldmann, *Bioorg. Med. Chem.* **1998**, *6*, 1153.

[19] M. A. Shogren-Knaak, P. J. Alaimo, K. M. Shokat, *Annu. Rev. Cell Dev. Biol.* **2001**, *17*, 405.

[20] D. P. Walsh, Y. T. Chang, *Chem. Rev.* **2006**, *106*, 2476.

[21] L. Liu, C. S. Bennett, C. H. Wong, *Chem. Commun.* **2006**, *21*.

[22] B. Wu, J. Chen, J. D. Warren, G. Chen, Z. Hua, S. J. Danishefsky, *Angew. Chem.* **2006**, *118*, 4222; *Angew. Chem. Int. Ed.* **2006**, *45*, 4116.

[23] J. Offer, C. N. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, *124*, 4642.

[24] Z. Machova, R. von Eggelkraut-Gottanka, N. Wehofsky, F. Bordusa, A. G. Beck-Sickinger, *Angew. Chem.* **2003**, *115*, 5065; *Angew. Chem. Int. Ed.* **2003**, *42*, 4916.

[25] E. C. Johnson, T. Durek, S. B. Kent, *Angew. Chem.* **2006**, *118*, 3361; *Angew. Chem. Int. Ed.* **2006**, *45*, 3283.

[26] C. Ludwig, M. Pfeiff, U. Linne, H. D. Mootz, *Angew. Chem.* **2006**, *118*, 5343; *Angew. Chem. Int. Ed.* **2006**, *45*, 5218.

[27] V. Muralidharan, T. W. Muir, *Nat. Methods* **2006**, *3*, 429.

[28] H. D. Mootz, E. S. Blum, T. W. Muir, *Angew. Chem.* **2004**, *116*, 5301; *Angew. Chem. Int. Ed.* **2004**, *43*, 5189.

[29] H. D. Mootz, E. S. Blum, A. B. Tyszkiewicz, T. W. Muir, *J. Am. Chem. Soc.* **2003**, *125*, 10561.

[30] H. D. Mootz, T. W. Muir, *J. Am. Chem. Soc.* **2002**, *124*, 9044.

[31] S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus, M. Q. Xu, *Gene* **1997**, *192*, 271.

[32] M. Zhang, M. Zhou, R. L. Van Etten, C. V. Stauffacher, *Biochemistry* **1997**, *36*, 15.

[33] W. L. DeLano, http://www.pymol.org **2002**.