Published in final edited form as:

Methods Mol Biol. 2020 ; 2133: 1–13. doi:10.1007/978-1-0716-0434-2_1.

Methods and Applications of Expressed Protein Ligation

Zhipeng A. Wang, Philip A. Cole

Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

Abstract

Expressed protein ligation is a method of protein semisynthesis and typically involves the reaction of recombinant protein C-terminal thioesters with N-cysteine containing synthetic peptides in a chemoselective ligation. The recombinant protein C-terminal thioesters are produced by exploiting the action of nature's inteins which are protein modules that catalyze protein splicing. This chapter discusses the basic principles of expressed protein ligation and recent advances and applications in this protein semisynthesis field. Comparative strengths and weaknesses of the method and future challenges are highlighted.

Keywords

Inteins; Protein splicing; Thioesters; Post-translational modifications; Protein labeling

1. Introduction: Protein semisynthesis and expressed protein ligation (EPL)

The study of proteins has taken center stage in biomedical investigation over the past several decades[1]. Powerful genetic tools including recombinant protein expression and site-directed mutagenesis[2] have allowed for the investigation of particular domains and individual amino acid contributions to the structure and function of complex proteins. Over the past twenty years, unnatural amino acid mutagenesis using nonsense suppression and related strategies has expanded the scope of protein study by providing a wide range of new chemical functionality into increasingly large and diverse protein types, addressing many biological challenges[3]. Nevertheless, in some cases it becomes desirable in protein biochemistry to have even greater chemical precision than that provided by genetic methods. Examples of such cases where the power of chemistry is especially beneficial in the study of proteins include the analysis of proteins with multiple diverse post-translational modifications, structural studies on segmentally isotopically labeled proteins, or the incorporation of two or more complex chemical probes at particular sites into proteins. One solution to these challenges is the use of chemical tools in the preparation of proteins[4]. Total synthesis of proteins typically involves generating peptides using solid phase synthesis and ligating such peptides together by chemoselective reactions including the powerful native chemical ligation reaction (NCL). The NCL reaction developed by Kent and

colleagues[5] is based on the early work of Wieland[6] and exploits the chemoselective reaction of an N-terminal cysteine (Cys) and a C-terminal thioester of two unprotected peptides. Such chemoselectivity derives from an initial trans-thioesterification between the thiol of the N-terminal Cys and C-terminal thioester followed by intramolecular aminolysis to afford a stable peptide bond. Total chemical synthesis of proteins has been successfully used to generate small proteins, usually less than 200 aa, but becomes difficult to execute on larger proteins because of the need for five or more peptide ligation reactions.

A compromise between genetic and chemical approaches to the study of proteins is embodied by protein semisynthesis. Protein semisynthesis refers to the technological merger of recombinant protein expression and chemical modifications, which are often introduced via chemoselective ligation reactions[7]. One of the major modalities of protein semisynthesis is expressed protein ligation (EPL)[7]. The EPL method exploits nature's inteins for generating recombinant proteins with C-terminal thioesters, which will further undergo NCL with an external peptide with an N-terminal Cys. Functionally related to introns in mRNA, inteins are protein domains that exist in a number of prokaryotic and yeast proteins that undergo protein splicing[8]. Protein splicing is analogous to RNA splicing in which two lateral protein domains, the N-extein (Ext^{N}) and C-extein (Ext^{C}), fuse to form a new polypeptide at the expense of the internal intein which is extruded.

2. Mechanisms of protein splicing, PTS and EPL

2.1 Intein based protein splicing

Intein-mediated protein splicing proceeds through multiple steps[8]. To begin with, the amide bond between the intein and Ext^{N} is cleaved by the first residue of the intein, most commonly a cysteine (Cys), forming a thioester with the Cys sidechain. Next, an acyl transfer reaction occurs in which the Ext^{N} is transferred to the first residue of the Ext^{C} , generally Cys, serine (Ser), or threonine (Thr). The intein appears to induce strain in the starting Ext^{N} -Cys peptide bond resulting in an equilibrium between the baseline amide and a thioester emanating from the intein Cys sidechain. This initial equilibrium provides an energetic source for intramolecular transacylation with the Ext^{C} . Subsequently, an asparagine (Asn) residue cleaves the main polypeptide chain to form a succinimidyl intermediate, resulting in the release of a branched thioester/ester from the intein, and the thioester/ester re-equilibrates to a normal backbone amide bond between the Ext^{N} and Ext^{C} (Fig. 1). Under some situations, the thioester/ester may undergo hydrolysis before the acyl transfer, leading to a side reaction commonly known as "N-terminal cleavage", which results in the release of free acid on the Ext^{N} . By judicious introduction of a mutations in the intein- Ext^{C} junction, the initial amide-thioester equilibrium in the Ext^{N} -Cys is arrested but can be exploited by large concentrations of exogenous thiols to produced C-terminal thioesters. This paves the way to obtain stable thioesters for use in semisynthesis.

2.2 Split intein, protein trans-splicing (PTS), and conditional protein splicing (CPS)

There are different families of inteins. Some natural inteins are split into separate segments on two different polypeptide chains, known as Int^{N} for the N-terminal fragment and Int^{C} for the C-terminal one. When the Int^{N} and Int^{C} come together non-covalently and form a

heterodimeric complex, the intact intein active fold is reconstituted and splicing activity is reestablished (Fig. 2). This mechanism has been termed “trans-splicing” because it links two separate peptide segments together[9]. Besides facilitating the purification of expressed proteins using split inteins as traceless affinity tags[10], like EPL, trans-splicing can also be used to generate site-specifically modified proteins.

With the discovery of new intein families and the development of intein techniques, the speed[11] and efficiency[12] of split inteins (natural and artificial) have been largely improved[13]. To gain tighter control of the trans-splicing process with environmental stimuli such as light, temperature, pH, or small molecules[14], conditional protein splicing (CPS) has been developed[15]. Such split inteins can be used to generate protein thioesters for use in EPL.

2.3 The EPL approach

In 1998, our group in collaboration with Tom Muir joined forces to develop EPL to generate site-specifically modified proteins. The fortuitous proximity of our two new groups at Rockefeller University helped to spark the development of this method. The Muir lab was expert in peptide synthesis and chemoselective ligations and our group brought recombinant protein production and enzymology skills. We both recognized that thioester peptides were difficult for most labs to access, especially at that time (although this situation has greatly improved[16]·[17]), and that inteins offered a potentially attractive solution to their production. Furthermore, it was our thought that EPL would be important for applying native chemical ligation to much larger proteins than those accessible by total synthesis (Fig. 3).

In the first examples of this method, we generated a post-translationally modified tyrosine kinase to explore the biochemistry of site-specific tyrosine phosphorylation at an unnatural location[7]. Shortly after our effort, Evans and colleagues[18] used EPL to generate a cytotoxic enzyme by producing it via two inert pieces that were ligated to form the full-length active enzyme. Over the years, EPL has been further refined and applied to address many biochemical challenges, particularly in the study of PTMs on proteins[19].

3. EPL for the preparation of non-natural and modified proteins

3.1 EPL vs. other protein chemical modification approaches in the study of modified proteins.

A number of methods have been developed for the installation of modifications in proteins and they are summarized in Table 1. Several efforts have exploited Cys modification strategies to introduce PTM mimics into proteins[20]·[21]·[22]·[23]. Overall, although Cys modification can be convenient, it generally requires only one Cys to be present and is limited in chemical scope[24]. Noncanonical amino acid incorporation (ncAA) via amber codon suppression is an increasingly popular method to generating modified proteins purely genetically. By simply installing a TAG amber stop codon into the target protein DNA sequence, an orthogonal tRNA synthetase-tRNA pair can recognize it to incorporate the modified amino acids[25]. However, a specific tRNA synthetase-tRNA mutant pair is

necessary for each single modification, which may not be suitable for all the non-natural structures, such as large modifications like ubiquitination[26]. Although a combination of ncAA with bioorthogonal reactions can extend the accessibility[27], proteins with multiple modifications are still hard to generate via this strategy due to reduced production efficiency.

Another means of artificial protein creation is to catalyze peptide ligations between synthetic peptides and expressed protein fragments with specialized enzymes including sortase[28], butelase[29], subtiligase[30] and transglutaminase[31] for peptide ligation, as well as lipoic acid ligase (LplA)[32]. Formylglycine generating enzyme (FGE)[33] for site-specific modification. These ligations can be rapid, but the yields are variable, and the enzymes can show particular sequence preferences that may be poorly compatible with particular applications[34].

3.2 Applications of EPL for the synthesis of non-natural proteins or modified proteins

As described above, EPL stands out as a robust and versatile strategy for the generation of proteins with site-specific PTMs. As an example, EPL has been often been used to investigate site-specific protein phosphorylation in cell signaling[37]. An early application[38] of this was in the analysis of SMAD2 and SMAD3 phosphorylation's role in oligomerization[39]. More recently, the influence of C-terminal phosphorylation of PTEN was analyzed[40]. It was found that such C-terminal phosphorylation induced a closed, inactive PTEN conformation with increased stability and resistance to HECT E3 ligase-mediated ubiquitination[41]. In this case, classical EPL which introduced an unnatural Cys residue in place of Tyr379 led to somewhat misleading results. That is, studies on Y379C phospho-PTEN underestimated the degree of conformational closure compared with the protein of native sequence (Fig. 4a)[42]. The use of subtiligase circumvented the Cys requirement and gave a more accurate assessment of phosphorylation impacts[43].

Recent EPL studies on the regulation of heavily investigated protein Ser/Thr kinase Akt helped clarify how it is modulated by C-terminal phosphorylation. Unexpectedly, two distinct mechanisms of activation were uncovered by Ser473 phosphorylation vs. Ser477/Thr479 dual phosphorylation (Fig. 4b)[44].

The tau protein is thought to play a central role in neurodegenerative diseases and is modified by Lys acetylation and other PTMs[45]. Elegant studies have recently been performed using EPL to elucidate the impact of a number of tau PTMs. As a protein with about 450 amino acid residues, a convergent ligation strategy involving five peptide segments was employed to assemble the full-length protein (Fig. 5). With these protein targets at hand, it was discovered that K280ac could strongly stimulate tau aggregation which might have implications in pathophysiology.

The functions of Ser *O*- β -N-acetylglucosamine (*O*-GlcNAc) modification on protein serine-threonine kinase casein kinase II (CK2) and on alpha-synuclein have been addressed using EPL[46]-[47]. It was found that such glycosylation can affect protein stability and aggregation.

Ypt1 is a small GTPase that is mono-prenylated and EPL was used to investigate this PTM[48]. The semisynthetic geranyl-Ypt1 formed a complex with RabGDI and a crystal structure was obtained to elucidate the structure of the complex.

Nucleosomes containing histone H2B with ubiquitination at the K120 site were prepared by EPL[36], and this helped reveal how H2B modification enhances methylation on histones H3 at Lys4 and Lys79[49]. Very recently, a high resolution cryoEM structure of the Dot1L methyltransferase in complex with the semisynthetic ubiquitin-modified nucleosome (and a norleucine replacement at Lys79 in H3) has revealed how Dot1L recognizes the ubiquitin and nucleosome to achieve specific methylation at Lys79[50].

3.3 Protein labelling

One of other important applications of EPL is to label proteins of interest for biophysical studies[51]. Fluorescence labeling is commonly used in FRET and binding studies of proteins. With the help of protein semisynthesis, a large variety of fluorophores have been added to protein targets. An early and impactful example involved incorporating into an SH3 domain the fluorescent probe 7-azatryptophan, which has distinct noninvasive fluorescence features for biophysical experiments[52]. A recent variant of the EPL approach has involved exploiting commercial NHS esters that are pretreated with MESNA to form thioesters that can be selectively linked to the N-termini of proteins. This approach avoids peptide synthesis and can be used by labs that wish to minimize chemical manipulation. In the inaugural example, the NHS ester strategy was employed to label WWP2 E3 ubiquitin ligase for studies on binding to its substrate PTEN and to clarify the stoichiometry of autoubiquitination[53].

In addition to incorporating fluorescence tags, isotopic labeling is a widely used strategy for the biophysical analysis of proteins of interest. The site-specific installation of NMR-sensitive isotopes, such as ^{13}C and ^{15}N , into target protein allows the direct assignment of resonances for structural determination[54]. EPL can be used for the ligation of isotopically labeled polypeptide fragment to another fragment with natural abundance isotopes. Known as “segmental isotopic labelling” or “block-labelling”, these methods can generate a sample for heteronuclear NMR analysis that possesses reduced spectral complexity relative to a fully labeled protein[55][56][57].

3.4 Other applications of EPL

Besides the labeling of EPL for biophysical studies, EPL and semisynthesis have also been applied to protein microarrays in a high-density manner for proteomics experiments. Protein-small molecule conjugates with an eye toward therapeutic applications[58], such as antibody-small molecule conjugates[12], have also been enabled by EPL[59]. EPL has been used to facilitate the expression of some otherwise-hard-to-achieve peptides or proteins[60]. EPL has also been applied to the production of cyclic peptides, toxic proteins, and proteins with uniform N-termini[61].

4. Summary and future challenges

Protein semisynthesis by expressed protein ligation and related methods have been slowly transforming the study of protein science. Continued refinements in the methods including the use of trans splicing, enzymes for ligation, and alternative chemoselective reactions[62] allow for enhanced applications. Nevertheless, continuing challenges in accessing the middle of large proteins by semisynthesis and generating semisynthetic proteins in vivo motivate current and future efforts in the field. We remain optimistic that investigators in the molecular biology and chemical biology communities will rise to these challenges and produce innovations that will increase the scope and influence of this technique in the coming years.

Acknowledgements-

We thank the NIH (GM62437 and CA74305) and the Leukemia and Lymphoma Society for financial support. We are grateful to Cole lab members past and present for their intellectual and experimental efforts that contributed to some of the work described here.

6. References

1. Kent SBH (2019) Novel protein science enabled by total chemical synthesis. *Protein Sci* 28:313–328. 10.1002/pro.3533 [PubMed: 30345579]
2. Fersht AR, Winter GP (1985) Redesigning enzymes by site-directed mutagenesis. *Ciba Found Symp* 111:204–218 [PubMed: 3860366]
3. Xiao H, Schultz PG (2016) At the Interface of Chemical and Biological Synthesis : An Expanded Genetic Code. *Cold Spring Harb Perspect Biol* 8:a023945 [PubMed: 27413101]
4. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Synthesis of proteins by native chemical ligation. *Sci York Then Washington-* 266:776
5. Dawson PE, Kent SBH (2000) Synthesis of Native Proteins by Chemical Ligation. *Annu Rev Biochem* 69:923–960 [PubMed: 10966479]
6. Wieland T, Bokelmann E, Bauer L, et al. (1953) Über Peptidsynthesen. 8. Mitteilung Bildung von S-haltigen Peptiden durch intramolekulare Wanderung von Aminoacylresten. *Justus Liebigs Ann Chem* 583:129–149
7. Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: A general method for protein engineering. *Proc Natl Acad Sci* 95:6705–6710. 10.1073/pnas.95.12.6705 [PubMed: 9618476]
8. Perler FB, Davis EO, Dean GE, et al. (1994) Protein splicing elements: inteins and exteins-a definition of terms and recommended nomenclature. *Nucleic Acids Res* 22:1125–1127 [PubMed: 8165123]
9. Mills KV, Lew BM, Jiang S -q., Paulus H (1998) Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein. *Proc Natl Acad Sci* 95:3543–3548. 10.1073/pnas.95.7.3543 [PubMed: 9520402]
10. Vila-Perelló, Perelló M, Liu Z, Shah NH, et al. (2013) Streamlined Expressed Protein Ligation Using Split Inteins. *J Am Chem Soc* 135:286–292. 10.1021/ja309126m [PubMed: 23265282]
11. David Y, Vila-Perelló M, Verma S, Muir TW (2015) Chemical tagging and customizing of cellular chromatin states using ultrafast trans-splicing inteins. *Nat Chem* 7:394–402. 10.1038/nchem.2224 [PubMed: 25901817]
12. Stevens AJ, Brown ZZ, Shah NH, et al. (2016) Design of a Split Intein with Exceptional Protein Splicing Activity. *J Am Chem Soc* 138:2162–2165. 10.1021/jacs.5b13528 [PubMed: 26854538]
13. Stevens AJ, Sekar G, Mostafavi AZ, et al. (2017) A promiscuous split intein with expanded protein engineering applications. *Proc Natl Acad Sci* 114:8538–8543. 10.1073/pnas.1701083114 [PubMed: 28739907]

14. Mootz HD, Muir TW (2002) Protein splicing triggered by a small molecule. *J Am Chem Soc* 124:9044–9045. 10.1021/ja026769o [PubMed: 12148996]
15. Di Ventura B, Mootz HD (2018) Switchable inteins for conditional protein splicing. *Biol Chem* 400:467–475. 10.1515/hsz-2018-0309
16. Blanco-Canosa JB, Dawson PE (2008) An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew Chemie - Int Ed* 47:6851–6855. 10.1002/anie.200705471
17. Zheng JS, Tang S, Huang YC, Liu L (2013) Development of new thioester equivalents for protein chemical synthesis. *Acc Chem Res* 46:2475–2484. 10.1021/ar400012w [PubMed: 23701458]
18. Evans TC, Benner J, Xu MQ (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* 7:2256–2264. 10.1002/pro.5560071103 [PubMed: 9827992]
19. Hackenberger CPR, Schwarzer D (2008) Chemoselective ligation and modification strategies for peptides and proteins. *Angew Chemie - Int Ed* 47:10030–10074. 10.1002/anie.200801313
20. Simon MD, Chu F, Racki LR, et al. (2007) The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell*. 10.1016/j.cell.2006.12.041
21. Huang R, Holbert MA, Tarrant MK, et al. (2010) Site-Specific Introduction of an Acetyl-Lysine Mimic into Peptides and Proteins by Cysteine Alkylation. *J Am Chem Soc* 132:9986–9987. 10.1021/ja103954u [PubMed: 20608637]
22. Spicer CD, Davis BG (2014) Selective chemical protein modification. *Nat Commun* 5:4740 10.1038/ncomms5740 [PubMed: 25190082]
23. Wang ZA, Liu WR (2017) Proteins with Site-Specific Lysine Methylation. *Chem - A Eur J*. 10.1002/chem.201701655
24. Bhat S, Hwang Y, Gibson MD, et al. (2018) Hydrazide Mimics for Protein Lysine Acylation to Assess Nucleosome Dynamics and Deubiquitinase Action. *J Am Chem Soc* 140:9478–9485. 10.1021/jacs.8b03572 [PubMed: 29991262]
25. Dai S, Tharp JM, Zeng Y, et al. (2016) A Genetically Encoded Allylsine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation. *Angew Chemie Int Ed* 56:212–216. 10.1002/anie.201609452
26. Wang ZA, Kurra Y, Wang X, et al. (2017) A Versatile Approach for Site-Specific Lysine Acylation in Proteins. *Angew Chemie - Int Ed* 56:1643–1647. 10.1002/anie.201611415
27. Chalker JM, Bernardes GJL, Davis BG (2011) A “tag-and-modify” approach to site-selective protein modification. *Acc Chem Res*. 10.1021/ar200056q
28. Wu M, Hayward D, Kalin JH, et al. (2018) Lysine-14 acetylation of histone H3 in chromatin confers resistance to the deacetylase and demethylase activities of an epigenetic silencing complex. *Elife* 7:. 10.7554/eLife.37231
29. Nguyen GKT, Qiu Y, Cao Y, et al. (2016) Butelase-mediated cyclization and ligation of peptides and proteins. *Nat Protoc* 11:1977–1988. 10.1038/nprot.2016.118 [PubMed: 27658013]
30. Henager SH, Chu N, Chen Z, et al. (2016) Enzyme-catalyzed expressed protein ligation. *Nat Methods* 13:925–927. 10.1038/nmeth.4004 [PubMed: 27669326]
31. Rachel NM, Toulouse JL, Pelletier JN (2017) Transglutaminase-Catalyzed Bioconjugation Using One-Pot Metal-Free Bioorthogonal Chemistry. *Bioconjug Chem* 28:2518–2523. 10.1021/acs.bioconjchem.7b00509 [PubMed: 28898047]
32. Uttamapinant C, White KA, Baruah H, et al. (2010) A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci* 107:10914–10919. 10.1073/pnas.0914067107 [PubMed: 20534555]
33. Rush JS, Bertozzi CR (2008) New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo - *Journal of the American Chemical Society (ACS Publications)*. 12240–12241
34. Fottner M, Brunner A-D, Bittl V, et al. (2019) Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nat Chem Biol* 15:276–284. 10.1038/s41589-019-0227-4 [PubMed: 30770915]
35. Siman P, Karthikeyan SV, Nikolov M, et al. (2013) Convergent Chemical Synthesis of Histone H2B Protein for the Site-Specific Ubiquitination at Lys34. *Angew Chemie Int Ed* 52:8059–8063. 10.1002/anie.201303844

36. McGinty RK, Kohn M, Chatterjee C, et al. (2009) Structure-activity analysis of semisynthetic nucleosomes: Mechanistic insights into the stimulation of Dot1L by ubiquitylated histone H2B. *ACS Chem Biol* 4:958–968. 10.1021/cb9002255 [PubMed: 19799466]
37. Chen Z, Cole PA (2015) Synthetic approaches to protein phosphorylation. *Curr Opin Chem Biol* 28:115–122. 10.1016/j.cbpa.2015.07.001 [PubMed: 26196731]
38. Wu J-W, Hu M, Chai J, et al. (2001) Crystal Structure of a Phosphorylated Smad2: Recognition of Phosphoserine by the MH2 Domain and Insights on Smad Function in TGF-Signaling involved in a ligand-specific signaling pathway; and the inhibitory Smads (I-Smads), including Smad6 and-7, which ne. *Mol Cell* 8:1277–1289 [PubMed: 11779503]
39. Qin BY, Lam SS, Correia JJ, Lin K (2002) Smad3 allosterically links TGF- β receptor kinase activation to transcriptional control. *Genes Dev* 16:1950–1963. 10.1101/gad.1002002 [PubMed: 12154125]
40. Bolduc D, Rahdar M, Tu-Sekine B, et al. (2013) Phosphorylation-mediated PTEN conformational closure and deactivation revealed with protein semisynthesis. *Elife* 2013:1–19. 10.7554/eLife.00691
41. Chen Z, Thomas SN, Bolduc DM, et al. (2016) Enzymatic Analysis of PTEN Ubiquitylation by WWP2 and NEDD4-1 E3 Ligases. *Biochemistry* 55:3658–3666. 10.1021/acs.biochem.6b00448 [PubMed: 27295432]
42. Chen Z, Dempsey DR, Thomas SN, et al. (2016) Molecular features of phosphatase and tensin homolog (PTEN) Regulation by C-terminal phosphorylation. *J Biol Chem* 291:14160–14169. 10.1074/jbc.M116.728980 [PubMed: 27226612]
43. Dempsey DR, Cole PA (2018) Protein Chemical Approaches to Understanding PTEN Lipid Phosphatase Regulation. *Methods Enzymol* 607:405–422. 10.1016/bs.mie.2018.05.007 [PubMed: 30149868]
44. Chu N, Salguero AL, Liu AZ, et al. (2018) Akt Kinase Activation Mechanisms Revealed Using Protein Semisynthesis. *Cell* 174:897–907.e14 10.1016/j.cell.2018.07.003 [PubMed: 30078705]
45. Haj-Yahya M, Lashuel HA (2018) Protein Semisynthesis Provides Access to Tau Disease-Associated Post-translational Modifications (PTMs) and Paves the Way to Deciphering the Tau PTM Code in Health and Diseased States. *J Am Chem Soc* 140:6611–6621. 10.1021/jacs.8b02668 [PubMed: 29684271]
46. Tarrant MK, Rho H-S, Xie Z, et al. (2012) Regulation of CK2 by phosphorylation and O-GlcNAcylation revealed by semisynthesis. *Nat Chem Biol* 8:262–269. 10.1038/nchembio.771 [PubMed: 22267120]
47. Levinea PM, Galesica A, Balanaa AT, et al. (2019) α -Synuclein O-GlcNAcylation alters aggregation and toxicity, revealing certain residues as potential inhibitors of Parkinson's disease. *Proc Natl Acad Sci* 116:1511–1519. 10.1073/pnas.1808845116 [PubMed: 30651314]
48. Rak A, Pylypenko O, Durek T, et al. (2016) Structure of Rab GDP-Dissociation Inhibitor in Complex with Prenylated YPT1 GTPase Published by : American Association for the Advancement of Science Stable URL : <http://www.jstor.org/stable/3835441> Linked references are available on JSTOR for this article. 302:646–650
49. Kim J, Guermah M, McGinty RK, et al. (2009) RAD6-Mediated Transcription-Coupled H2B Ubiquitylation Directly Stimulates H3K4 Methylation in Human Cells. *Cell* 137:459–471. 10.1016/j.cell.2009.02.027 [PubMed: 19410543]
50. Worden EJ, Hoffmann NA, Hicks CW, Wolberger C (2019) Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell* 176:1490–1501.e12 10.1016/j.cell.2019.02.002 [PubMed: 30765112]
51. Muralidharan V, Muir TW (2006) Protein ligation: An enabling technology for the biophysical analysis of proteins. *Nat Methods* 3:429–438. 10.1038/nmeth886 [PubMed: 16721376]
52. Muralidharan V, Cho J, Trester-Zedlitz M, et al. (2004) Domain-specific incorporation of noninvasive optical probes into recombinant proteins. *J Am Chem Soc* 126:14004–14012. 10.1021/ja0466199 [PubMed: 15506763]
53. Dempsey DR, Jiang H, Kalin JH, et al. (2018) Site-Specific Protein Labeling with N-Hydroxysuccinimide-Esters and the Analysis of Ubiquitin Ligase Mechanisms. *J Am Chem Soc* 140:9374–9378. 10.1021/jacs.8b05098 [PubMed: 30016585]

54. Romanelli A, Shekhtman A, Cowburn D, Muir TW (2004) Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction. *Proc Natl Acad Sci* 101:6397–6402. 10.1073/pnas.0306616101 [PubMed: 15087498]
55. Uegaki K, Oda N, Nakamura H, et al. (1998) Segmental Isotope Labeling for Protein NMR Using Peptide Splicing. *J Am Chem Soc* 120:5591–5592. 10.1021/ja980776o
56. Nabeshima Y, Mizuguchi M, Kajiyama A, Okazawa H (2014) Segmental isotope-labeling of the intrinsically disordered protein PQBP1. *FEBS Lett* 588:4583–4589. 10.1016/j.febslet.2014.10.028 [PubMed: 25447530]
57. Flynn JD, Jiang Z, Lee JC (2018) Segmental ^{13}C -Labeling and Raman Microspectroscopy of α -Synuclein Amyloid Formation. *Angew Chemie - Int Ed* 57:17069–17072. 10.1002/anie.201809865
58. Chiang MJ, Holbert MA, Kalin JH, et al. (2014) An Fc domain protein-small molecule conjugate as an enhanced immunomodulator. *J Am Chem Soc* 136:3370–3373. 10.1021/ja5006674 [PubMed: 24533830]
59. Flavell RR, Muir TW (2009) Expressed Protein Ligation (EPL) in the Study of Signal Transduction, Ion Conduction, and Chromatin Biology. *Acc Chem Res* 42:107–116. 10.1002/chin.200918258 [PubMed: 18939858]
60. Leibly DJ, Nguyen TN, Kao LT, et al. (2012) Stabilizing Additives Added during Cell Lysis Aid in the Solubilization of Recombinant Proteins. *PLoS One* 7:. 10.1371/journal.pone.0052482
61. Topilina NI, Mills KV. (2014) Recent advances in in vivo applications of intein-mediated protein splicing. *Mob DNA* 5:1–14. 10.1186/1759-8753-5-5 [PubMed: 24382139]
62. Bode JW (2017) Chemical Protein Synthesis with the α -Ketoacid-Hydroxylamine Ligation. *Acc Chem Res* 50:2104–2115. 10.1021/acs.accounts.7b00277 [PubMed: 28849903]

Figure 1. Classical mechanism of protein splicing. Following the removal of the succinimidyl intein intermediate, Ext^{N} joins with Ext^{C} through a native amide bond.

Figure 2.

Classical mechanism of protein trans-splicing (PTS). An intact intein in a functionally folded conformation is reconstituted after the formation of a heterodimeric complex of Int^N and Int^C , restoring normal splicing activity.

Figure 3.

The expressed protein ligation strategy. Intein cleavage generates the formation of a recombinant protein thioester fragment. This thioester fragment is then mixed with a synthetic N-terminal Cys containing peptide affording a semisynthetic protein via the native chemical ligation reaction.

Figure 4.

Scheme for the semisynthesis of phosphorylated PTEN (a) or Akt (b). (a) The PTEN aa1-378 recombinant protein thioester fragment is generated with MESNA and then ligated to N-Cys synthetic peptide (PTEN aa379-403) possessing site-specific phosphoserines and/or phosphothreonines. (b) The recombinant Akt PH domain-catalytic domain thioester fragment (aa1-459) is ligated to the N-Cys synthetic peptide C-tail peptide (aa460-480) possessing site-specific phosphoserines or phosphothreonines.

Figure 5.

The semisynthetic preparation of Tau protein with site-specific acetylations. A convergent strategy involving five-fragment stepwise ligations to produce full length Tau aa2-441 was achieved with site-specific acetylation. The longest N-terminal Tau aa2-245 recombinant protein thioester fragment was prepared using an intein.

Table 1:

Comparison of different chemical biology approaches for the preparation of posttranslationally modified proteins

Strategy	Example	Description	Advantage(s)	Disadvantage(s)	Reference
Amino acid (Cys) modification	H3K4me	PTM analogs are added to functional amino acid residue (mainly Cys)	chemically simple, convenient	Bioorthogonal reactions needed; Hard to separate side products with minor chemical differences	[20]
Enzyme-catalyzed ligation (example, Sortase)	H3K14ac	Enzyme catalyzed ligation of peptides	rapid reactions, can avoid Cys	Specialized enzymes needed, yields can be variable, restricted to certain amino acid sequences	[28]
Noncanonical amino acid mutagenesis (ncAA)	H3K4succ	An orthogonal tRNA-synthetase pair to suppress the amber stop codon to incorporate a non-natural amino acid	Similar to normal recombinant expression, accesses the middle of even larger proteins	Generally limited to one unnatural amino acid, tRNA synthetase mutant needed, can be difficult in eukaryotic expression systems	[26]
Protein total synthesis	H2B K34ub	Synthetic peptides are linked together via different ligation strategies	complete chemical control	Generally for smaller proteins, refolding often required	[35]
EPL	H2B K120ub	A synthetic peptide is linked to a larger piece of expressed protein	Small peptide easier to synthesize; no size limitation for target protein, refolding can be avoided	Mainly focused on C-terminal or N-terminal modifications, reactions can be slow	[36]