Topics in Current Chemistry 362

Lei Liu *Editor*

Protein Ligation and Total Synthesis I

Springer

362
Topics in Current Chemistry

Aims and Scope

The series Topics in Current Chemistry presents critical reviews of the present and future trends in modern chemical research. The scope of coverage includes all areas of chemical science including the interfaces with related disciplines such as biology, medicine and materials science.

The goal of each thematic volume is to give the non-specialist reader, whether at the university or in industry, a comprehensive overview of an area where new insights are emerging that are of interest to larger scientific audience.

Thus each review within the volume critically surveys one aspect of that topic and places it within the context of the volume as a whole. The most significant developments of the last 5 to 10 years should be presented. A description of the laboratory procedures involved is often useful to the reader. The coverage should not be exhaustive in data, but should rather be conceptual, concentrating on the methodological thinking that will allow the non-specialist reader to understand the information presented.

Discussion of possible future research directions in the area is welcome.

Review articles for the individual volumes are invited by the volume editors.

Readership: research chemists at universities or in industry, graduate students.

More information about this series at
http://www.springer.com/series/128

Lei Liu
Editor

Protein Ligation and Total Synthesis I

With contributions by

J. Brailsford · S.J. Danishefsky · A. Fernández-Tejada · C.D. Hall · R.A. Jones · A.R. Katritzky · T. Kawakami · C.-F. Liu · F. Liu · L.R. Malins · J.P. Mayer · M.A.S. Moore · S.S. Panda · R.J. Payne · J.-H. Shieh · Y.-W. Wu · A. Yang · R. Yang · Q. Zhang · L. Zhao

Springer

Editor
Lei Liu
Department of Chemistry
Tsinghua University
Beijing
China

ISSN 0340-1022 ISSN 1436-5049 (electronic)
Topics in Current Chemistry
ISBN 978-3-319-19185-0 ISBN 978-3-319-19186-7 (eBook)
DOI 10.1007/978-3-319-19186-7

Library of Congress Control Number: 2015941105

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

As essential organic molecules of life, proteins have received the attentions of chemists since the beginning of synthetic organic chemistry. Today, chemical protein synthesis is not only a pursuit of pure science but also provides useful molecules with applications to biochemistry research and drug development. Synthetic chemistry enables a level of control of protein composition beyond that attainable by protein expression. Chemistry also holds promise for tuning the properties of a protein molecule at atomic resolution and thus can provide otherwise elusive insights into protein structure and function. For these reasons, chemical protein synthesis has been intensively explored in the field of chemical biology and its application has demonstrated the importance of modern synthetic chemistry to cutting-edge research in biomedicine.

The present and next issues of *Topics in Current Chemistry* collect a representative number of review chapters surveying some of the current research trends and technology levels in this important field. The chapters presented in the following pages are authored by some of the pioneers and active researchers in the field from different countries. These chapters reflect many of the important issues in the area, namely, development of novel chemical methods for the ligation of peptide segments, total and semi-synthesis of important protein targets, and application of state-of-the-art methods to solve problems in biochemistry research or drug development. I hope that the readers find the two issues to be an interesting read. I would like to thank all the authors for their excellent contributions. I would also like to thank Arun Manoj Jayaraman for the help given to me in handling the manuscripts.

Beijing, China — Lei Liu

Contents

Top Curr Chem (2015) 362: 1–26
DOI: 10.1007/128_2014_622

Published online: 25 March 2015

Total Synthesis of Glycosylated Proteins

Alberto Fernández-Tejada, John Brailsford, Qiang Zhang, Jae-Hung Shieh, Malcolm A.S. Moore, and Samuel J. Danishefsky

Abstract Glycoproteins are an important class of naturally occurring biomolecules which play a pivotal role in many biological processes. They are biosynthesized as complex mixtures of glycoforms through post-translational protein glycosylation. This fact, together with the challenges associated with producing them in homogeneous form, has hampered detailed structure-function studies of glycoproteins as well as their full exploitation as potential therapeutic agents. By contrast, chemical synthesis offers the unique opportunity to gain access to homogeneous glycoprotein samples for rigorous biological evaluation. Herein, we review recent methods for the assembly of complex glycopeptides and glycoproteins and present several examples from our laboratory towards the total chemical synthesis of clinically relevant glycosylated proteins that have enabled synthetic access to full-length homogeneous glycoproteins.

Keywords Chemoselective ligation • Glycopeptides • Glycoproteins • Glycosylation • Total synthesis

A. Fernández-Tejada (✉)
Laboratory for Bioorganic Chemistry, Molecular Pharmacology and Chemistry Program, 1275 York Avenue, New York, NY 10065, USA

Chemical and Physical Biology, CIB-CSIC, Ramiro de Maeztu 9, Madrid 28040, Spain
e-mail: aftejada@cib.csic.es

J. Brailsford, Q. Zhang, and S.J. Danishefsky (✉)
Laboratory for Bioorganic Chemistry, Molecular Pharmacology and Chemistry Program, 1275 York Avenue, New York, NY 10065, USA
e-mail: s-danishefsky@ski.mskcc.org

J.-H. Shieh and M.A.S. Moore
Cell Biology Program, Sloan Kettering Institute for Cancer Research, 1275 York Avenue, New York, NY 10065, USA

Contents

1 Introduction

The study of protein post-translational modification (PTM) [1] is an active area in protein research. Protein glycosylation is a major form of PTM which expands the functional diversity and structural heterogeneity of proteins. More than 50% of human proteins are reported to be glycosylated [2], a modification known to play an important role in a wide range of biological events, at both cellular (cell–cell interaction, signaling, immune response) and protein levels (molecular recognition, protein folding) [3–5]. Aberrant glycosylation of proteins is also implicated in a number of illnesses, including autoimmune diseases, infectious diseases, and cancer [6, 7]. As a result, glycans and glycoproteins are increasingly being investigated in clinical settings as potential targets for the development of therapeutics, diagnostics, and vaccines [8, 9]. For all these reasons, a great deal of effort has been directed toward elucidating the effects of protein glycosylation on the function, conformation, stability, and bioactivity of glycoproteins [10–13].

Structurally, all native glycoproteins can be classified into two major types: *N*-linked and *O*-linked glycoproteins. In *N*-linked glycoproteins, the glycan (*N*-acetylglucosamine, GlcNAc) is attached through a β-glycosidic linkage to the amide side chain of an asparagine within an Asn-Xaa-Ser/Thr consensus sequence. In common *O*-glycoproteins, the sugar (mainly *N*-acetyl-galactosamine, GalNAc and GlcNAc) is α- or β-linked to the hydroxyl group of serine or threonine residues, respectively (Fig. 1). Other forms of *O*-glycosylation are also observed with several different glycans such as mannose, xylose, glucose, or fucose [14].

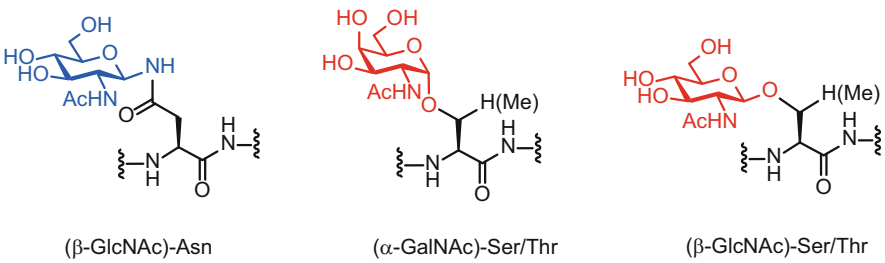


Fig. 1 Common (β-GlcNAc)-*N*-linked asparagine and (α-GalNAc)/(β-GlcNAc)-*O*-linked serine/threonine in glycosylated proteins

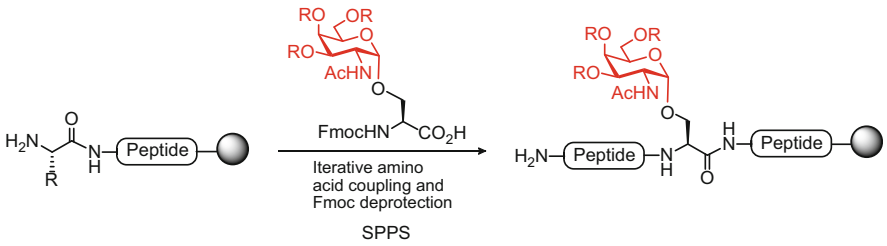
Unlike protein synthesis, protein glycosylation is not regulated genetically, and the biosynthesis of typical *N*- and *O*-glycoproteins often comprises multiple steps dictated by the relative activities of a large number of enzymes. As a result, glycoproteins are naturally expressed as complex mixtures of glycoforms possessing the same peptide backbone but differing in both the nature and site of glycosylation [15]. This structural diversity lays the molecular basis for a wide range of biological functions. However, the structural heterogeneity of glycosylation and the difficulty associated with obtaining homogeneous glycoproteins from biological sources has limited progress towards correlating glycoprotein structure with function [16]. The urgent need for chemically pure, synthetic glycoproteins for detailed structure–activity relationship studies and therapeutic applications has been the main driving force toward the development of several strategies for the preparation of these structures [17]. In particular, chemical synthesis has emerged as a practical way to gain access to well-defined glycoproteins, as is demonstrated through several relevant examples described herein.

2 Chemical Synthesis of Glycopeptides

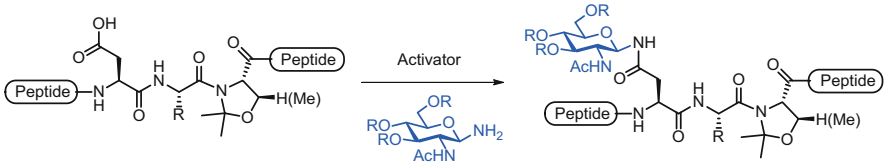
The chemical synthesis of glycoproteins requires the ability to prepare glycosylated peptide fragments efficiently [18, 19]. Peptide synthesis efforts typically rely on solid-phase peptide synthesis (SPPS) [20], a process which allows for the efficient assembly of peptide fragments comprising up to 50 amino acid residues. Of the two conventional SPPS methodologies (Boc method and Fmoc method), Fmoc-based SPPS is usually favored in the synthesis of glycopeptides because the glycosidic linkages are relatively labile in the presence of strong acid.

For the synthesis of a glycopeptide, bearing oligosaccharide side chains, there exist two general strategies: the cassette-based approach and the convergent method. In the more linear cassette approach, the glycan is attached to an amino acid residue and the resulting glycosyl amino acid is then used in SPPS to assemble the complete glycopeptide fragment (Scheme 1). In principle, the pre-formed glycosyl amino acid "cassette" can consist of either the fully elaborated oligosaccharide, or a monosaccharide residue, which, following SPPS, may be elongated further by enzymatic or chemical glycosylation to yield the glycopeptide containing the full glycan. Several research groups have used this stepwise, "cassette-based" approach to generate successfully *N*-linked [21, 22] and especially *O*-linked [23, 24] glycopeptide fragments related to natural glycoproteins. A particularly attractive application of this strategy is being implemented in a number of laboratories for the synthesis of mucin-type *O*-glycopeptides containing tumor carbohydrate antigens (Tn, TF, STn) as promising targets for anticancer vaccines [25–27].

In contrast, the convergent approach to glycopeptide synthesis relies on the direct conjugation of an oligosaccharide domain with the full-length peptide to generate the corresponding glycopeptide. Whereas the cassette strategy has been widely used for the synthesis of *O*-glycopeptides, the convergent method has been mainly applied for the preparation of *N*-glycopeptide fragments. In this approach,



Scheme 1 Linear "cassette"-based approach for the synthesis of *O*-linked glycopeptides



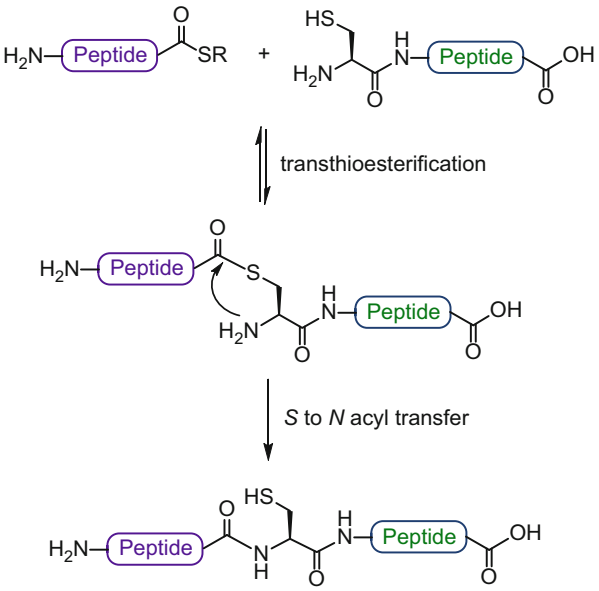
Scheme 2 Lansbury aspartylation reaction facilitated by a pseudoproline motif for the convergent synthesis of *N*-linked glycopeptides

the partially protected peptide is first synthesized by SPPS. Next, the *N*-glycan, bearing an anomeric amine, is directly introduced through an amide linkage to the free aspartic acid (Asp) side chain of the peptide under Lansbury aspartylation conditions [28]. The Asp residue to be coupled requires an additional orthogonal protecting group which can be selectively removed prior to Lansbury aspartylation. The carboxylic acid is activated using coupling agents (i.e., HATU, HBTU, or PyAOP), and the anomeric amine of the sugar reacts with the activated ester to form the amide bond. When the peptidyl Asp side chain is activated, care should be taken to minimize competitive intramolecular aspartimide formation [29]. To address this problem, we and Unverzagt independently developed a simple solution, involving emplacement of a temporary pseudoproline motif, derived from Ser/Thr, at the $n+2$ position (relative to Asp) of the Asn-Xaa-Ser/Thr consensus sequence (Scheme 2) [30, 31]. Adoption of this pseudoproline strategy has enabled the efficient and convergent syntheses of extended glycopeptide fragments, thereby providing enabling progress in the chemical synthesis of homogeneous complex *N*-glycopeptides [32].

3 Chemical Synthesis of Glycoproteins

3.1 Advances in Chemical Ligation

Extension of the methodologies described above to the assembly of full-length glycoproteins is hindered by a number of factors, including low coupling efficiency, poor solubility of protected polypeptides, acid-lability of the oligosaccharide

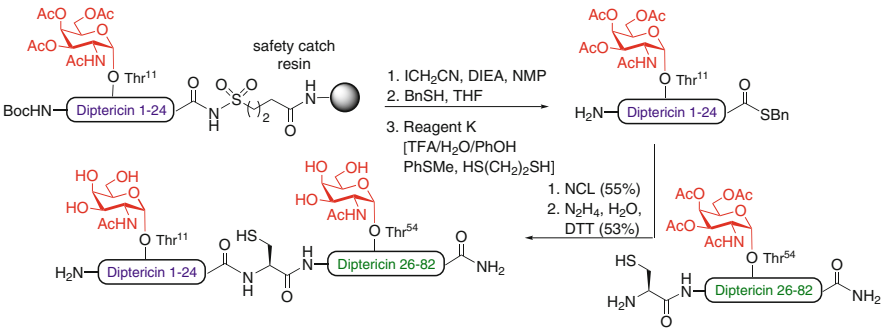


Scheme 3 Native chemical ligation (NCL)

chains, increased epimerization, and formation of side-products resulting in low yields and purities of the final target molecules. For these reasons, several chemoselective ligation methods have been developed which permit the assembly of complex glycoproteins from shorter glycosylated peptides [33].

Most notable among these is the powerful native chemical ligation (NCL) method, first described by Kent et al. in 1994 [34]. Over the past 20 years, NCL has proved to be a remarkably robust methodology for the efficient, epimerization-free assembly of proteins. As shown in Scheme 3, the process accomplishes chemoselective ligation, in aqueous solution, of two fully deprotected peptides, one of which presents a C-terminal thioester and the other an N-terminal cysteine residue. The reaction involves transthioesterification of the C-terminal thioester with the N-terminal cysteine to generate a transient thioester intermediate, which then undergoes an irreversible $S \rightarrow N$ acyl transfer to form the native peptide bond. Interestingly, large peptide fragments may be obtained by recombinant expression and further used as ligation partners through a method termed expressed protein ligation (EPL) [35, 36].

To date, NCL-based strategies have been successfully applied to the chemical synthesis of over 200 proteins [37]. This technology has also been adopted for the total synthesis of a number of biologically relevant glycoproteins [38]. The first significant examples were reported by Bertozzi and coworkers, who synthesized several mucin-type glycoproteins bearing GalNAc residues at the relevant *O*-glycosylation sites [39, 40]. One such target was diptericin, an 82-residue antimicrobial glycoprotein prepared via NCL between two fragments, a 58-mer Cys-glycopeptide and a 24-mer glycopeptide thioester (Scheme 4). Notably, an



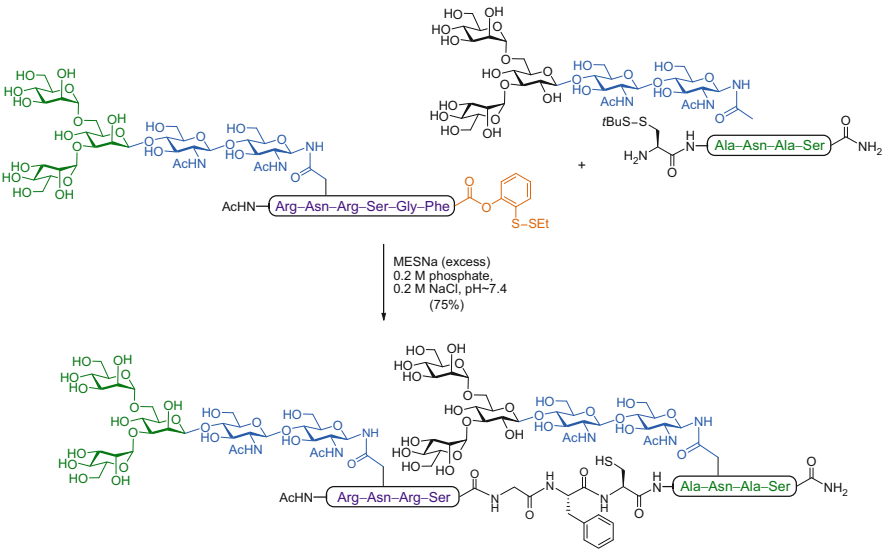
Scheme 4 Total synthesis of antimicrobial *O*-glycoprotein diptericin by NCL

alkane sulfonamide "safety-catch" linker was used for the Fmoc-based SPPS synthesis of the glycopeptide thioester [39]. Unverzagt et al. also used the "safety-catch" linker strategy followed by NCL to synthesize the first *N*-linked glycopeptide fragment carrying a complex-type oligosaccharide [41]. The first total chemical synthesis of a full-length glycoprotein bearing a complex type *N*-glycan was reported by Kajihara and coworkers, who prepared a single glycoform of the 76-amino acid chemokine monocyte chemotactic protein-3 (MCP-3) via two consecutive NCLs of three (glyco)peptide fragments, followed by protein folding (not shown) [42].

Despite these impressive examples, efforts to extend directly the NCL method toward the synthesis of complex glycopeptides have been complicated by the difficulties associated with synthesizing the requisite pre-formed glycopeptide thioester fragments, particularly by Fmoc-based SPPS [43]. To address this issue, our group has been engaged in the development of a number of synthetic strategies that may have broad utility for the preparation of homogeneously glycosylated peptides and proteins [44, 45]. In pursuing modified NCL strategies, we and others have sought to replace the labile C-terminal thioester with more stable latent thioester or alternative acyl donor groups.

In a key advance, we synthesized a C-terminal glycopeptide phenolic ester equipped with an *o*-disulfide moiety which, upon reductive treatment with subsequent unmasking of the free thiol, can rearrange to generate the corresponding aryl thioester via a reversible $O \rightarrow S$ acyl transfer [46]. The utility of this latent acyl donor in glycopeptide ligation has been widely demonstrated. In the example shown in Scheme 5, two glycopeptide fragments – one bearing the newly crafted *o*-disulfide phenolic ester at the C-terminus and the other a *t*Bu-thio-protected cysteine at the N-terminus – were subjected to reductive cleavage by the addition of sodium 2-mercaptoethanesulfonate (MESNa). The resulting glycopeptide intermediates underwent ligation in situ to furnish the doubly glycosylated peptide.

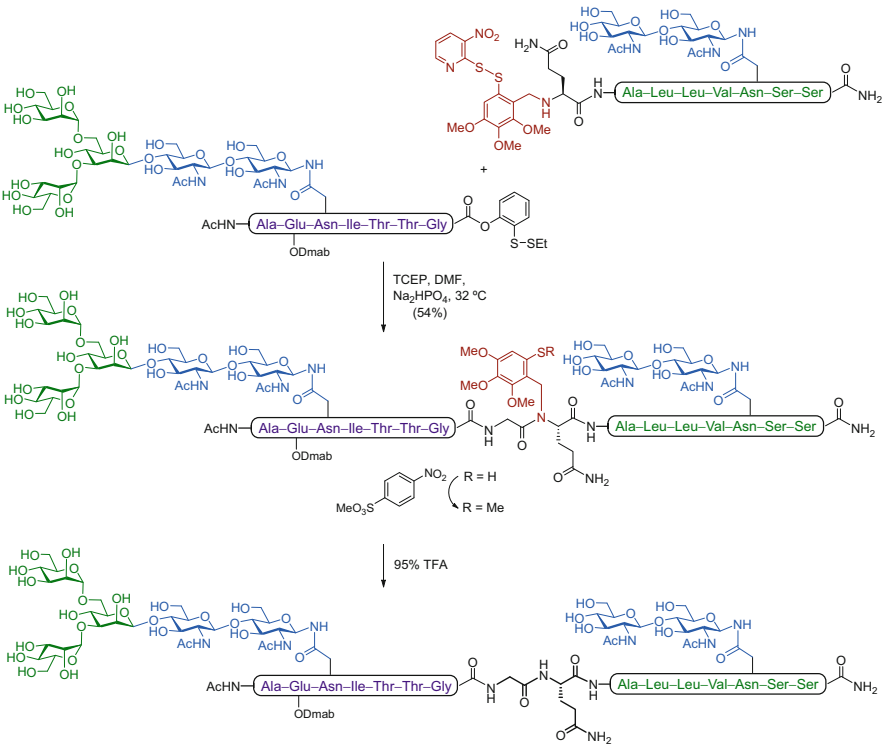
Although this method proved highly useful for a wide range of substrates, we next sought to increase the acyl transfer capability of the *o*-thiophenolic ester in order to improve ligation efficiency with peptides bearing hindered C-terminal amino acids. Along these lines, we found that, when activated *p*-nitrophenyl or



Scheme 5 NCL-based assembly of a complex model *N*-glycopeptide using an *o*-disulfide phenolic ester as latent thioester

p-cyanophenyl oxo-ester variants were employed as thioester surrogates [47], efficient ligation was attainable at sterically demanding C-terminal sites. Indeed, this strategy was ultimately adopted for the synthesis of a challenging glycopeptide fragment en route to erythropoietin (EPO) [48].

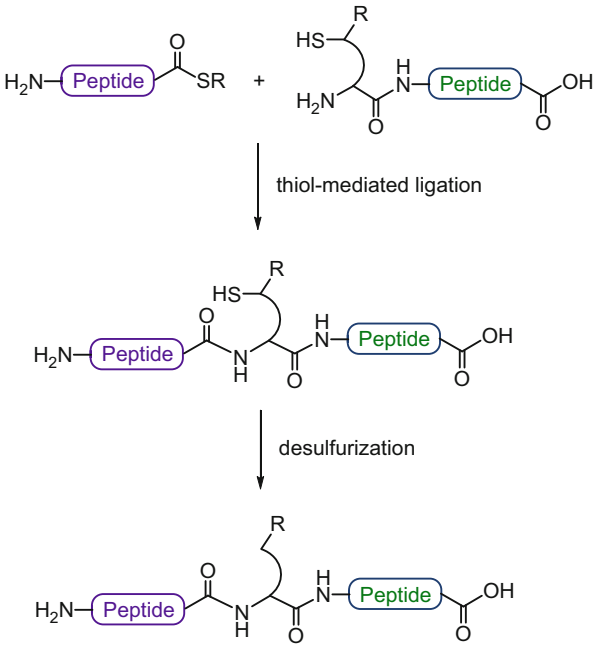
Another important limitation for the general applicability of the NCL is the requirement that an N-terminal cysteine be present on the target peptide or glycopeptide fragment. Given the relative scarcity of cysteine residues in naturally occurring proteins and glycoproteins, it is often difficult to locate a feasible ligation site containing this amino acid, which makes direct ligation at cysteine itself of somewhat limited practical value. As a consequence, a number of alternative strategies have emerged for the cysteine-free, ligation-based assembly of glycopeptides and glycoproteins [49]. Early studies led to the development of thiol-based auxiliaries, wherein a mercapto group present in a provisional N-terminal moiety serves to mimic the function of the cysteine thiol. A pioneering example by Macmillan and Anderson demonstrated the utility of this ligation approach at an accessible glycine–glycine junction in the synthesis of an *O*-glycopeptide fragment bearing a single GalNAc residue [50]. We have reported several examples of the use of such auxiliary-based ligations for the construction of complex glycopeptides bearing multiple *O*- and *N*-glycans [51]. In a representative case, a glycopeptide acceptor containing a cleavable thiol-containing trimethoxybenzyl auxiliary on the N-terminus was ligated with a masked glycopeptide thioester at a glycine – glutamine junction to afford the desired glycopeptide bearing two *N*-linked glycans (Scheme 6).



Scheme 6 Synthesis of a complex model *N*-glycopeptide using an auxiliary-based, cysteine-free ligation strategy

As shown, under the reducing ligation conditions, $O \rightarrow S$ acyl transfer occurred at the *o*-disulfide phenolic ester as previously described (see above), to provide the thioester, which underwent transthioesterification with the thiol-containing glycopeptide (upon in situ reduction of the auxiliary disulfide bond); the transient intermediate underwent an $S \rightarrow N$ acyl transfer to generate the thermodynamically favored amide bond of the doubly glycosylated peptide adduct. The auxiliary was subsequently removed through sequential methylation of the free thiol to prevent the reverse acid-mediated $N \rightarrow S$ acyl shift, followed by TFA treatment.

While useful as a tool for peptide or glycopeptide ligation, the general application of these cysteine-free ligation strategies presents some drawbacks. First, steric hindrance at the ligation site renders the key $S \rightarrow N$ acyl transfer step difficult and, as a result, this method can only be effectively applied to junctions where at least one of the amino acids is glycine or alanine. Another potential problem lies in the harsh acidic conditions required for cleavage of the thiol-auxiliary, which could compromise the labile glycosidic linkages within the glycopeptide. An alternative approach to the development of non-cysteine-based ligation strategies is to utilize a thiol-containing amino acid surrogate located at the N-terminus of the peptide fragment. In this strategy, the thiol moiety acts as a temporary "handle" to promote



Scheme 7 Non-cysteine ligation followed by desulfurization

ligation with the corresponding peptide coupling partner. After the ligation reaction, the thiol functionality is removed through desulfurization to afford the natural amino acid at the site of ligation (Scheme 7).

This idea was first introduced by Yan and Dawson, who took advantage of cysteine-based NCL and then converted the N-terminal cysteine residue into alanine by reduction of the side chain with Raney nickel or Pd/Al_2O_3 [52]. In essence, this method provided a means to use cysteine as an alanine surrogate, which is a much more abundant amino acid in natural protein sequences. Subsequent work from Pentelute and Kent showed that unprotected cysteine residues could be desulfurized in the presence of methionine and Acm-protected cysteine residues with Raney nickel, although a large excess of the metal was required [53]. In 1998, González and Valencia had introduced a photochemical method for the efficient removal of the thiol group of cysteine without metal reagents, using triethyl phosphite as the nucleophile and triethylborane in acetonitrile as a radical initiator [54]. Driven by our interest in the total synthesis of complex glycoproteins, which often contain extensive and sensitive functional groups, we developed a mild, metal-free protocol for the selective reduction of cysteine to alanine compatible with a wide range of functionalities, including carbohydrate, thiazolidine, biotin, and thioester moieties [55]. This novel desulfurization strategy is aqueous-compatible and employs TCEP as a reducing agent, in conjunction with the water-soluble radical initiator VA-044 and *tert*-butylthiol as a thiol additive. This mild,

radical-based protocol represented a significant improvement over previously reported metal-based NCL desulfurization methods and has been successfully applied in the synthesis of a number of complex proteins and glycoproteins [56–59]. Having identified a highly versatile cysteine reduction method, this ligation/desulfurization approach opened the door to a new set of thiol-based ligation strategies with a range of different amino acid residues. Thus, through recourse to desulfurization methods, the reach of the NCL logic was extended, in our laboratory and others, to include alanine [52, 55], valine [60, 61], threonine [62], leucine [63, 64], lysine [65, 66], proline [67, 68], phenylalanine [69], methionine [70, 71], serine [72], arginine [73], aspartate [74], tryptophan [75], and glutamate [76] as viable ligation sites. Overall, these non-cysteine-based thiol ligation strategies, which are referred to as extended NCL reactions, followed by post-ligation desulfurization, have greatly facilitated the total chemical synthesis of complex, biologically relevant proteins and glycoproteins.

3.2 *Total Synthesis of Homogeneous* N-*Glycosylated Proteins*

Our laboratory has a long-standing interest in the total synthesis of naturally occurring, complex *N*-linked glycoproteins of important therapeutic value. This pursuit arises from the motivation to understand the structural and functional effects of protein glycosylation at the molecular level and, hopefully, to bring these complex biologics into the realm of structures that can be accessed by chemical means for structure-activity relationship studies. We have primarily focused our synthetic efforts on the human glycoprotein hormones (hGPH), granulocyte–macrophage colony-stimulating factor (GM-CSF) and EPO.

3.2.1 Human Glycoprotein Hormones

We are interested in the synthesis of a family of glycoproteins involved in a variety of important biological processes – the human glycoprotein hormones (hGPH). The members of this family are follicle stimulating hormone (FSH), luteinizing hormone (LH), chorionic gonadotropin (CG), and thyroid stimulating hormone (TSH). These glycoproteins are composed of two non-covalently associated subunits, a common α-subunit (α-hGPH) and a unique hormone-specific β-subunit (β-FSH, β-LH, β-CG, and β-TSH), each of which contains diverse carbohydrates at defined glycosylation sites. As the α-subunit is shared amongst these glycoproteins, we initially focused on the synthesis of the α-subunit (α-GPH) as the first step to access the human glycoprotein hormones.

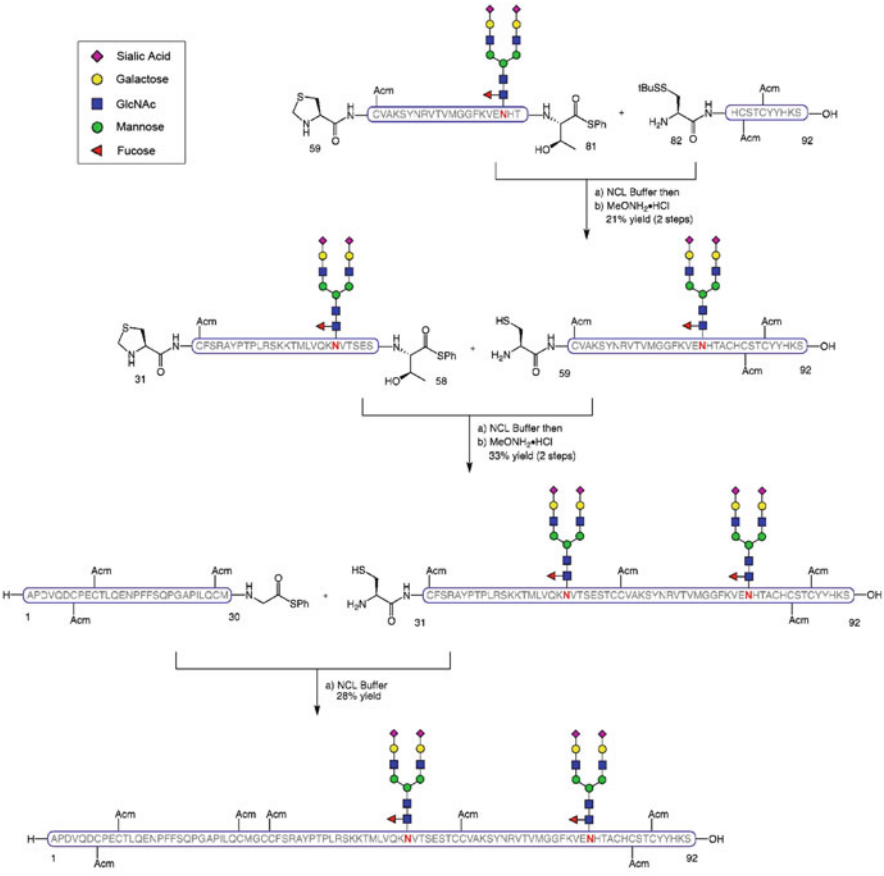
Human α-Glycoprotein Hormone (α-hGPH)

In our early studies, a simplified version of this domain incorporating chitobiose as a model glycan was synthesized as a proof of principle. The validated route was then applied to prepare the α-hGPH bearing the dibranched, sialylated and fucosylated *N*-linked dodecasaccharide. This complex-type *N*-glycan is a consensus sequence which has also been found to exist on the β-subunits of the human glycoproteins and incorporates the key features associated with these hormones. The challenging synthesis of this sugar, which involves approximately 70 chemical steps [77], utilizes glycal chemistry to build the core trisaccharide and highlights the use of the Sinaÿ radical glycosylation [78] for the simultaneous coupling of both biantennary side-chains of the dodecasaccharide. A seven-step deprotection procedure followed by Kochetkov amination [79] afforded the fully deprotected *N*-glycan bearing an anomeric amine (not shown).

With the synthetic dodecasaccharide amine in hand, we then built up the corresponding glycopeptide fragments using the Lansbury aspartylation reaction [28] with peptide thioesters α-hGPH[31–58] and α-hGPH[59–81], themselves prepared by Fmoc-based SPPS followed by single amino acid attachment under Sakakibara's epimerization-free conditions [80] (Scheme 8). After removal of all acid-labile protecting groups with a TFA-based cocktail, the fully deprotected (except Acm-containing cysteines) α-hGPH[82–92] and glycosylated α-hGPH [59–81] and α-hGPH[31–58] were then combined by means of a sequential two-step protocol involving ligation and subsequent cysteine unmasking. Lastly, the peptide thioester α-hGPH[1-30] (prepared by SPPS and direct thioesterification of the C-terminal glycine residue) was coupled with α-hGPH[31–92] via NCL to provide the Acm-protected primary sequence of α-hGPH featuring two defined, complex *N*-linked glycans [81].

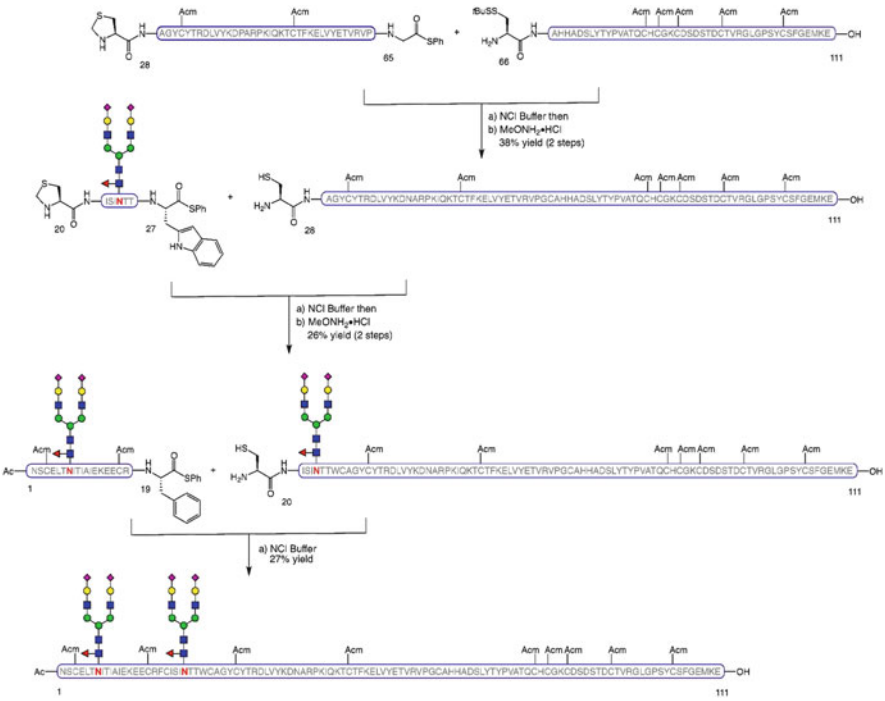
Human β-Follicle Stimulating Hormone (β-hFSH)

As our first complete target (with both α and β subunits synthesized), we selected hFSH, a member of the family of the gonadotropic hormones which play a key role in the regulation and maintenance of important reproductive processes. Naturally synthesized and secreted in the anterior pituitary gland, it has been used in clinical settings to treat anovulatory disorders associated with infertility [82]. Currently, hFSH is obtained from recombinant techniques as a complex mixture of glycoforms. This variability may well be the cause of the formidable liabilities and side effects associated with its use [83]. Faced with the unavailability of homogeneous forms of hFSH, chemical synthesis seems to be the most straightforward approach to overcome this limitation.



Scheme 8 Synthesis of the α-subunit of human glycoprotein hormones, α-hGPH[1–92]

The β-subunit of hFSH (β-FSH) was synthesized as described in Scheme 9 [84]. The individual peptide and glycopeptide segments were prepared by SPPS (with subsequent Sakakibara elongation or C-terminal glycine thioesterification to obtain the corresponding thioesters) and Lansbury aspartylation. Thus, fragment β-hFSH[66–111], peptide thioester β-hFSH[28–65], and glycosylated β-hFSH[20–27] were combined by NCL followed by cleavage of the N-terminal thiazolidine (Thz) blocking group in a stepwise manner. The last glycopeptide segment β-hFSH [1–19] was then joined under NCLconditions to form the glycosylated β-subunit of hFSH with two *N*-linked dodecasaccharides at the native glycosylation sites and Acm-protecting groups at the cysteine residues.



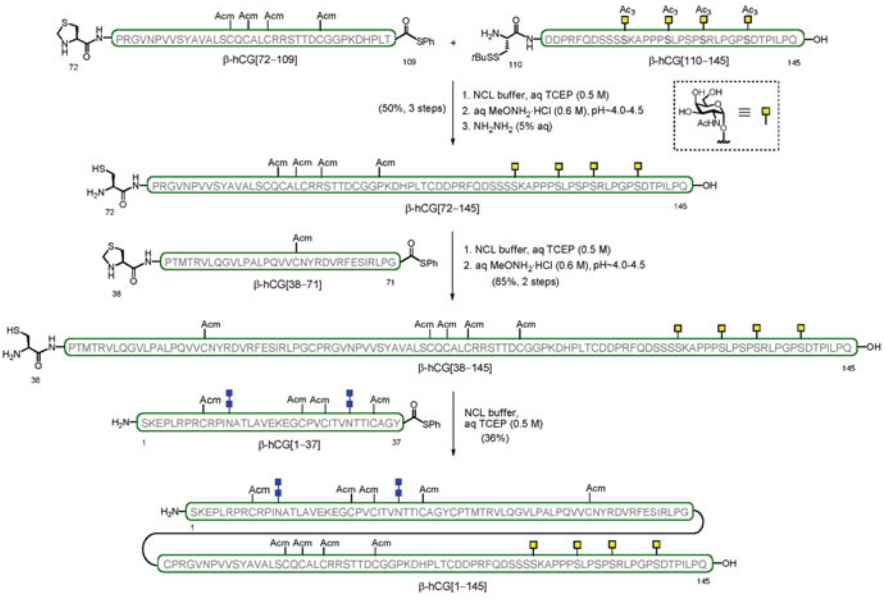
Scheme 9 Synthesis of the β-subunit of human follicle stimulating hormone, β-hFSH[1–111]

Human β-Luteinizing Hormone (β-hLH) and β-ChorionicGonadotropin Hormone (β-hCG)

Human luteinizing hormone (hLH) and chorionic gonadotropin hormone (hCG) belong to the family of the gonadotropins and are implicated in the stimulation of the gonadal and endocrine functions, as well as in pregnancy. hLH is released by the pituitary gland and hCG is produced in the human placenta. They both have similar physiological functions, and are used clinically as fertility drugs in reproductive medicine [85]. The β-subunit of hCG has been found to be overexpressed in several types of cancers and encompasses an epitope for hCG-based monoclonal antibodies. These features have made hCG a promising target for important biomedical and therapeutic applications [86, 87]. hLH and hCG share a high degree of similarity in their peptide sequence although hCG incorporates an extended C-terminal domain with four *O*-glycosylation sites. Thus, β-hLH contains 121 residues and only one *N*-glycosylation site (Asn30), whereas β-hCG comprises 145 amino acids and incorporates both *N*-linked (Asn13 and Asn30) and *O*-linked sugars (Ser121, Ser127, Ser132, and Ser138), making it the longest and most complex human glycoprotein hormone. Several studies investigating the importance of the oligosaccharide chains on the biological activity of these hormones suggest a more relevant role of the underlying carbohydrate residues vs the

peripheral sugar units [88]. We sought to gain access to homogeneous glycoforms of β-hLH and β-hCG by chemical means for further exploration of the influence of the glycans on hormone action and function. Initially, we accomplished the synthesis of the simpler β-subunit of hLH bearing chitobiose as a model glycan to explore the feasibility of the synthetic route. Upon successful completion of the target molecule by sequential NCL, we then applied this strategy to assemble the more complex hCG β-subunit bearing not only two *N*-glycans but also four *O*-linked sugars [89].

The synthesis included two key features to access both terminal domains of the molecule. A successful double Lansbury aspartylation enabled the simultaneous, convergent installation of two chitobiose units onto the N-terminus of the peptide backbone to provide β-hCG[1–37]. The C-terminal fragment bearing closely spaced *O*-glycosylation sites, β-hCG[110–145], was prepared following a more linear "cassette" approach, whereby the four *O*-linked glycosylamino acids were sequentially incorporated in a single, solid-supported synthesis. The assembly of the Acm-protected, full-length β-hCG containing two *N*-linked (chitobiose) and four *O*-linked glycans (GalNAc) was performed in a modular and convergent fashion by ligation of the individual (glyco)peptide fragments (i.e., *O*-glycosylated β-hCG[110–145], β-hCG[72–109], β-hCG[38–71], and *N*-glycosylated β-hCG[1–37]) from the C- to the N-terminus, using a multiple NCL-based strategy (Scheme 10). Importantly, this synthesis of the β-subunit of hCG represents the longest human glycoprotein hormone (hGPH) accessed as a single glycoform by strictly chemical means.



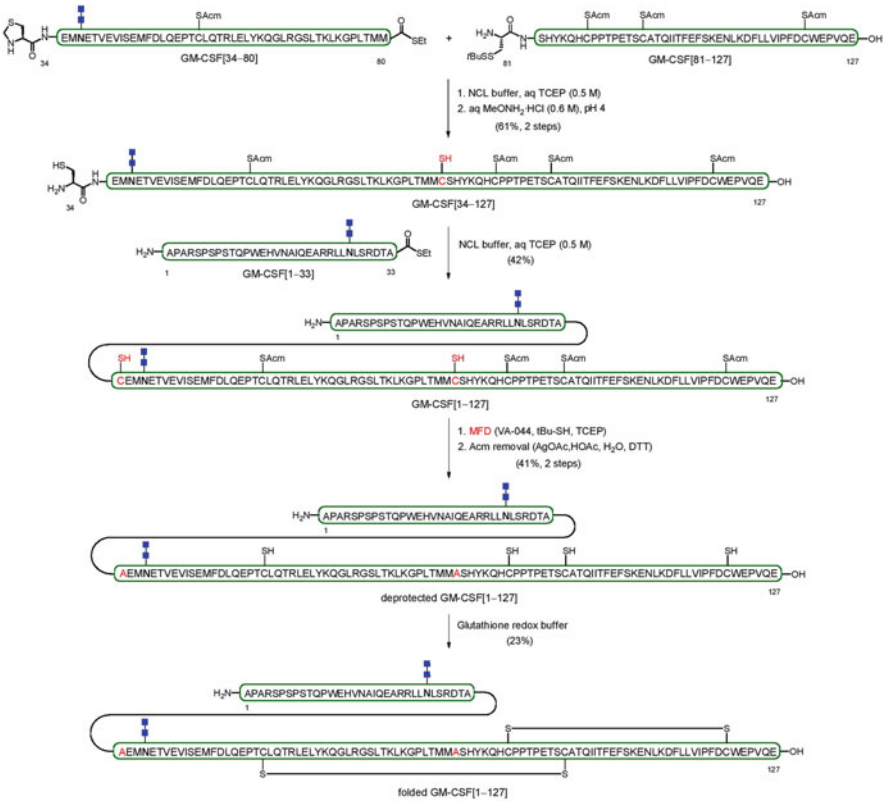
Scheme 10 Synthesis of β-subunit of human chorionic gonadotropin hormone, β-hCG[1–145]

Having accomplished the total synthesis of both α- and β-subunits of these glycosylated proteins, we are now investigating the late-stage challenges towards preparing homogeneous, folded heterodimeric hFSH, hLH, and hCG in our laboratory.

3.2.2 Granulocyte–Macrophage Colony-Stimulating Factor (GM-CSF)

Our continuing interest in complex, therapeutic-level glycoproteins led us to pursue the synthesis of homogeneous glycoforms of GM-CSF with the aim of studying the effect of glycosylation on the biological activity of this target. GM-CSF is a glycoprotein cytokine which promotes cell growth and proliferation and functions as an immune modulator. It is used to stimulate the immune system after bone-marrow transplant and chemotherapy and is being investigated as a vaccine immunoadjuvant [90]. The GM-CSF structure comprises 127 amino acids with both *N*- and *O*-glycosylation sites. Our aim was to establish a synthetic strategy to enable access to a number of homogeneous glycosylated variants of GM-CSF. We designed a route with maximal convergence whereby the full-length molecule would be assembled by merging three (glyco)peptide fragments using exclusively alanine ligations.

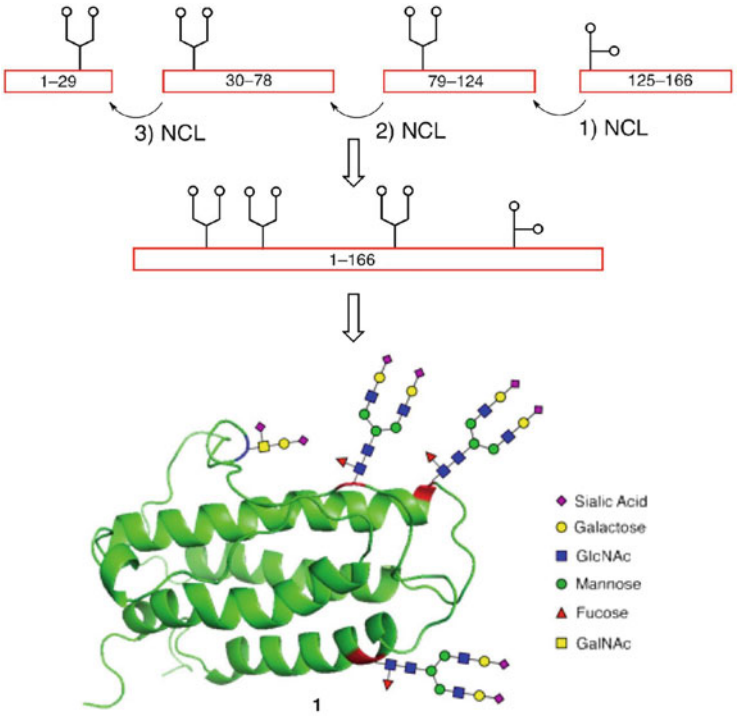
In the synthesis of a bis-*N*-glycosylated version containing the chitobiose disaccharide, glycopeptide GM-CSF[34–80] and peptide GM-CSF[81–127] were first connected via NCL to give, following Thz cleavage, GM-CSF[34–127] (Scheme 11) [91]. This fragment was then combined with GM-CSF[1–33], containing a chitobiose moiety, to provide complex glycopeptide intermediate GM-CSF[1–127]. The late-stage transformations towards the doubly-glycosylated GM-CSF congener involved reduction of the two free thiol groups via metal-free desulfurization (MFD) to reveal the native alanine residues, followed by deprotection of the four Acm-bearing cysteine residues. A monoglycosylated variant bearing a single chitobiose (at Asn37) was prepared via a similar synthetic route. Final folding of both fully synthetic analogues under glutathione redox buffering conditions provided the structurally defined GM-CSF glycoforms bearing either one (Asn37) or two (Asn27 and Asn37) chitobiose units at asparagine wild-type sites. Evaluation of the in vitro and in vivo biological activity of these synthetic GM-CSF samples, including a non-glycosylated congener, showed that they all function in a similar manner to commercially available GM-CSF, which is obtained recombinantly as a mixture of glycoforms.



Scheme 11 Synthesis of the bis-glycosylated GM-CSF glycoform

3.2.3 Erythropoietin (EPO)

Erythropoietin (EPO) is a glycoprotein hormone primarily implicated in the regulation of red blood cell production in the bone marrow, a process known as erythropoiesis [92–94]. Recombinantly derived EPO has been used for the treatment of anemia associated with renal failure and cancer chemotherapy [95]. However, production of recombinant EPO results in a complex mixture of glycoforms, where the primary protein sequence and the glycosylation sites are highly conserved but the carbohydrate domains present significant heterogeneity. Despite studies suggesting that specific sugar modifications affect the stability and erythropoietic activity of EPO [96, 97], the unavailability of single glycoforms has complicated elucidation of the relationship between EPO glycosylation and its biological activity. Chemical synthesis uniquely holds the potential to provide access to homogeneous EPO for systematic evaluation of the effect of the glycan structure on biological activity.

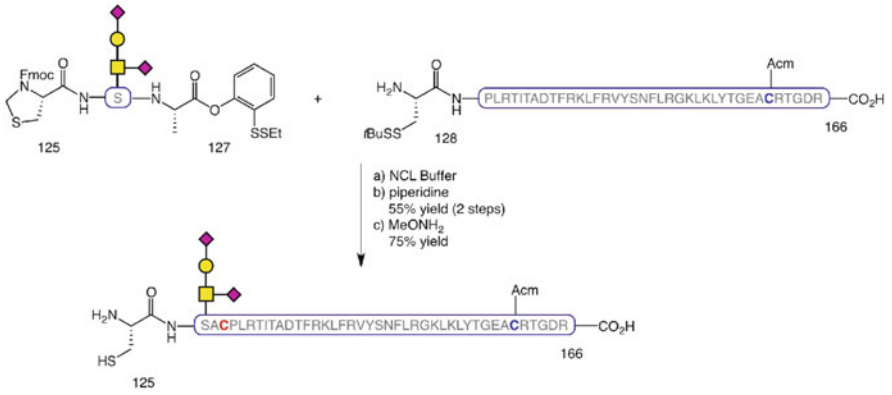


Scheme 12 Proposed synthetic strategy for convergent synthesis of erythropoietin, EPO[1–166] and ribbon diagram of the fully glycosylated protein

EPO is composed of a 166-amino acid backbone incorporating three *N*-linked glycans (Asn24, Asn38, and Asn83) and an *O*-linked glycan (Ser126). In our first total synthesis of homogeneously glycosylated EPO, we chose to include three simpler chitobiose units and the naturally-abundant glycophorin tetrasaccharide at the four native glycosylation sites [98]. From a retrosynthetic perspective, we envisioned gaining access to the full-length glycoprotein, EPO[1–166], by assembling four glycopeptide domains through iterative NCL-based ligations (Scheme 12).

The *O*-linked glycophorin was introduced into the peptide chain as a glycosylated serine cassette [99] via addition of a single amino acid at each end of the glycophorin-linked Ser residue followed by NCL of the resulting glycosylated tripeptide with the adjacent peptide segment to provide EPO[125–166] (Scheme 13).

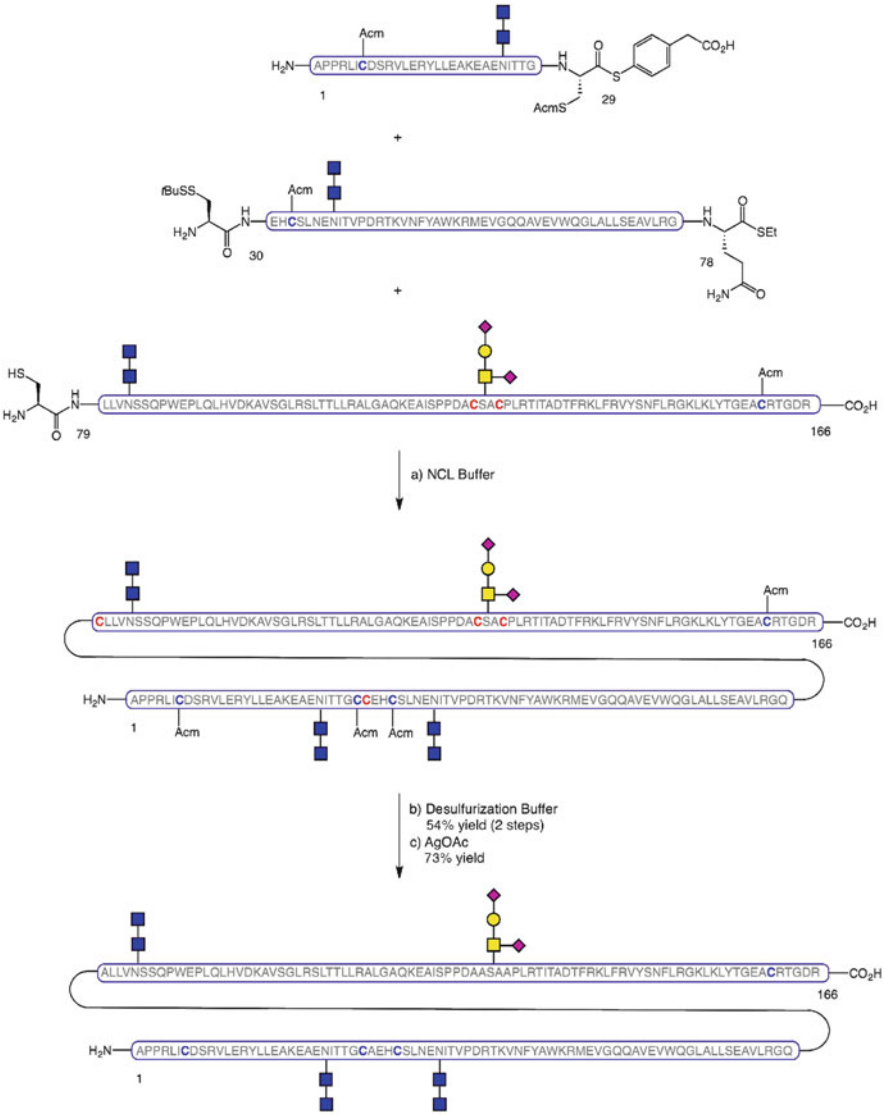
The chitobiose-containing *N*-glycopeptide constructs were prepared by our one-flask Lansbury aspartylation/deprotection protocol [30] from the corresponding peptide thioester fragments, themselves obtained by Fmoc-based SPPS followed by single amino acid attachment. On the basis of our previous experience toward a non-glycosylated version of EPO [59], we envisioned a synthetic strategy for



Scheme 13 Synthesis of EPO[125–166]

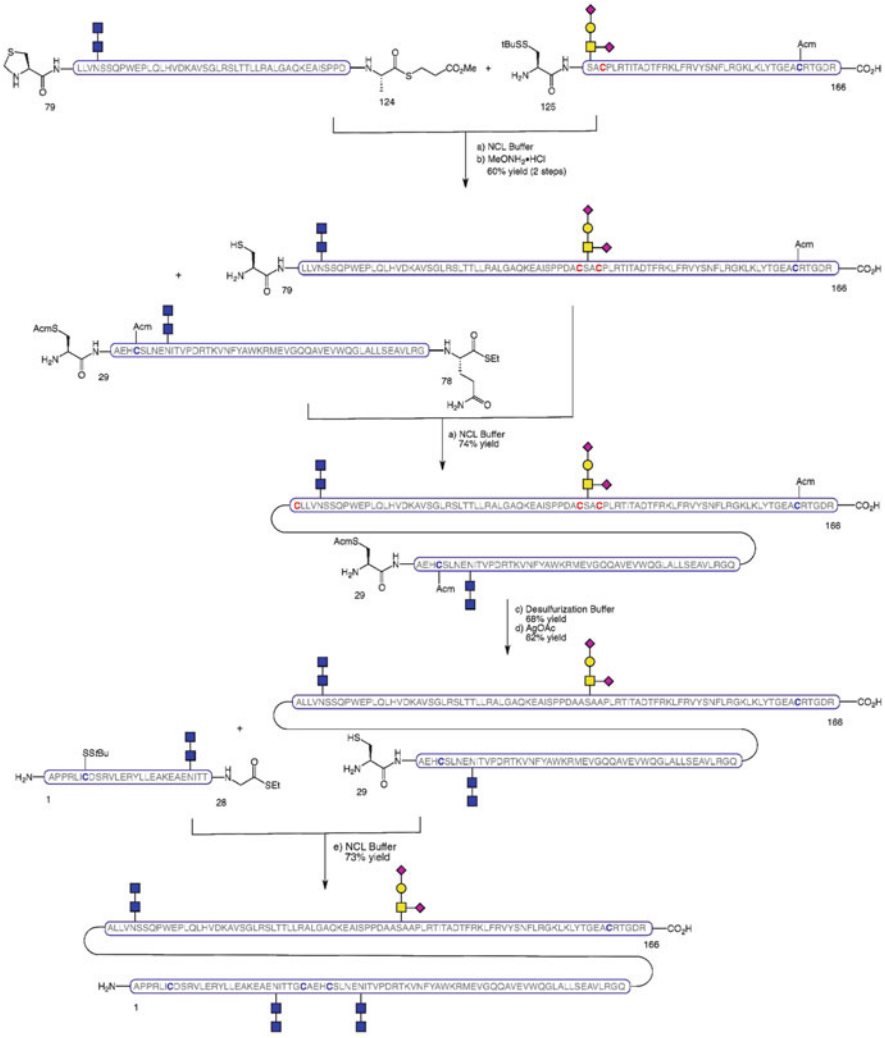
merging the four component glycopeptides via a series of iterative alanine ligations, followed by late-stage desulfurization to convert the non-native cysteines into the requisite alanine residues. This first-generation synthesis featured a kinetically controlled NCL (KCL)-based approach [100] to enable the one-flask coupling of three fully elaborated EPO fragments in a highly convergent fashion (Scheme 14). In this KCL route, the aryl thioester of EPO[1–29] undergoes transthioesterification with the N-terminal cysteine of EPO[30–78] at a faster rate than the alkyl thioester, resulting in selective ligation between Cys-29 and Cys-30. Thus, EPO[1–29] and EPO[30–78] were first merged by NCL; upon completion of this ligation, EPO[79–166], itself prepared by NCL between EPO[79–124] and EPO[125–166], was added to assemble the protein sequence. The synthesis was completed by MFD, followed by removal of the Acm groups to afford the primary structure of homogeneously glycosylated EPO.

An alternative, more linear strategy was also developed, whereby the full protein chain was assembled via NCL as the final step of the synthetic sequence. The main advantage of this route is that it avoids the need for post-ligation transformations on the full-length sequence prior to protein folding. The approach employs peptides which slightly differ from the fragments used in the convergent route in that the ligation site at the N-terminus is shifted by one residue. Thus, glycopeptide fragments EPO[125–166], EPO[79–124], and EPO[29–78] were combined by successive NCLs to afford the EPO[29–166] domain (Scheme 15). At this point, desulfurization of the non-native cysteines that participated in the ligations, and subsequent Acm deprotection on the remaining cysteine residues set the stage for the merger of the final peptide fragment, EPO[1–28], by NCL. The fully synthetic, homogeneously glycosylated EPO[1–166] thus obtained was then subjected to disulfide oxidation and protein folding to provide a simplified variant of folded EPO which, despite having truncated *N*-linked glycans, demonstrated erythropoietic activity.



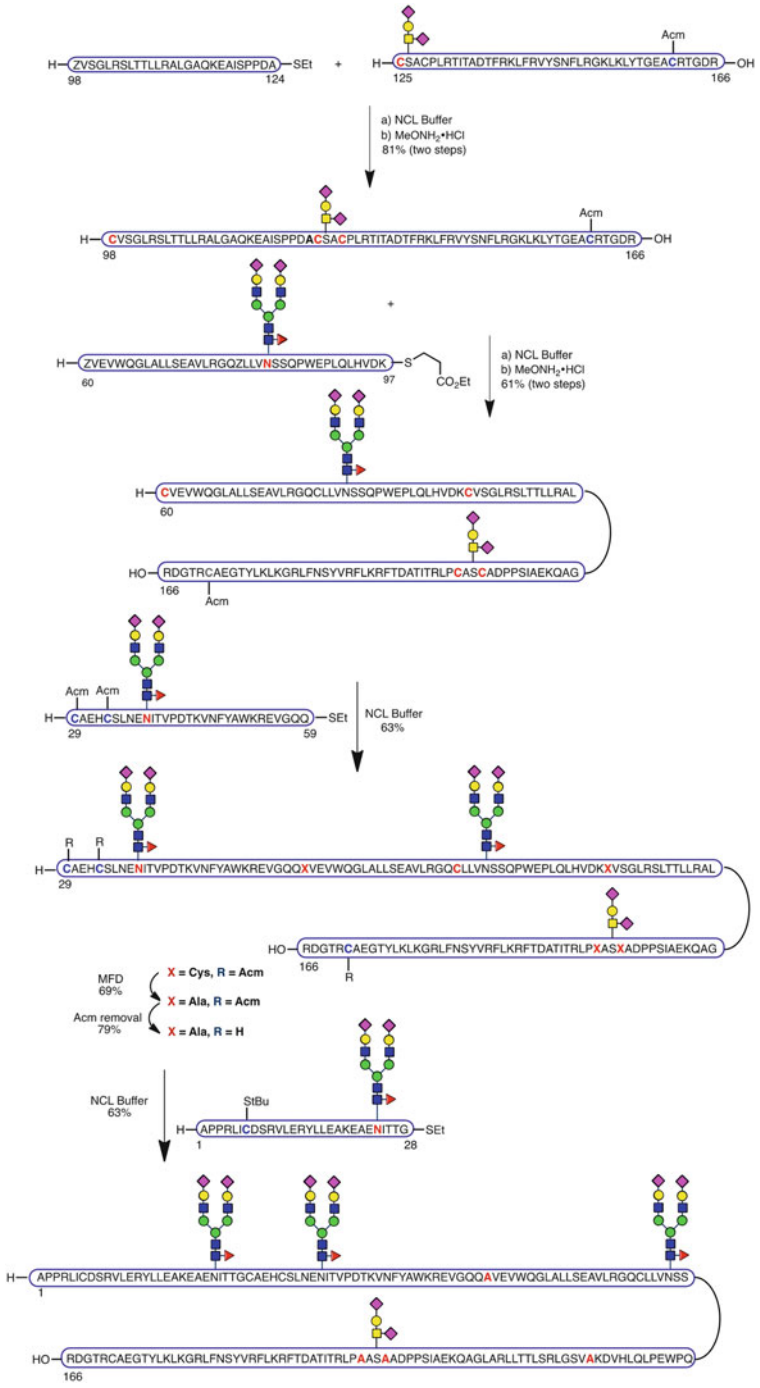
Scheme 14 Initial convergent synthesis of EPO

Notwithstanding the major breakthrough of generating a fully synthetic version of EPO as a single glycoform, our ultimate goal was to achieve the total synthesis of "wild-type" EPO with consensus carbohydrate structures of realistic, biological-level complexity at all four native glycosylation sites. Therefore, we set out on this effort with the objective of incorporating the elaborate biantennary dodecasaccharide at the three *N*-glycosylation sites and glycophorin at Ser126 [101]. We initially envisioned adapting our previously described linear strategy,



Scheme 15 Linear synthesis of erythropoietin, EPO[1–166]

developed in the context of the chitobiose glycoform, to this more complex target. However, the greater steric bulk of the full *N*-glycans rendered ligation of EPO[29–78] and EPO[79–166] fruitless. Accordingly, the synthetic route was modified and a different ligation site, more distant from the problematic, large dodecasaccharide, was selected. In the newly designed approach, the domain related to pre-EPO[29–124] was reconfigured into three novel fragments, *N*-glycosylated EPO[29–59] and EPO[60–97], and peptide segment EPO[98–124], which were sequentially assembled, starting from the *O*-glycosylated, C-terminal domain EPO[125–166] via a series of smooth alanine ligations followed by thiazolidine unmasking (Scheme 16).



Scheme 16 Synthesis of fully glycosylated EPO[1–166] bearing wild-type glycans at each of the four glycosylation sites

The resulting fragment EPO[29–166] was then subjected to fourfold MFD to reveal the required alanines. After cleavage of the Acm protecting groups, cysteine-based NCL between EPO[29–166] and EPO[1–28] afforded the complete primary sequence of EPO containing wild-type complex glycan structures at each of the four glycosylation sites. This unfolded material gave rise to a mass spectrum of marginal quality (see Supporting Information in Wang et al. [101]). Not surprisingly, the material assigned as the EPO primary sequence did not even exhibit in vitro activity.

Upon exposure of the EPO primary sequence to folding conditions, a product was obtained which was reported [101] to exhibit strong erythropoietic activity and, thus, presumed to correspond to a biologically active version of compound 1 (Scheme 12) (although our synthetic EPO showed suitable activity in an in vivo assay, literature precedent suggests that heterogeneous mixtures of tetra-antennary structures have approximately 7× the in vivo activity of analogous bi-antennary EPO glycoforms [102–104]). More recently, an approximately 1-year-old sample of this material, bearing identical mobility properties to those reported from the original sample (see Wang et al. [101]), was found to retain approximately 15% of the in vitro activity of Procrit. It has been found that thawing and re-freezing of the sample tend to erode its erythropoietic activity. In retrospect, however, the inability to obtain a supportive mass spectrum of this material raises concerns with respect to the nature of the folded product. Related sialic acid-containing recombinantly derived EPO mixtures of unspecified structure gave rise to useful mass spectra in the context of glycoform profiling studies ([105, 106] and references therein). The corresponding author of reference [101] assumes full responsibility for not taking cognizance of these precedents. Further clarification of the matter of folding of the presumed precursor would require access to freshly prepared material, which can only be accomplished by repetition of the total synthesis. This has recently been undertaken in order to settle these questions.

4 Conclusion

The total synthesis of glycosylated proteins has long posed a daunting challenge to practitioners of organic chemistry. Recent key developments in this field, including NCL, non-cysteine-based ligations, MFD, and novel glycosylation methodologies, have led to the realization of the long-standing goal of gaining synthetic access to homogeneous peptides, proteins, and glycoproteins. Several truncated-carbohydrate versions of important biologically active glycoproteins have been synthesized in homogeneous form by a combination of these methods. These include GM-CSF, TSH, and EPO. An application to EPO bearing sialic acid-containing biantennary 12-mer *N*-glycosides has also been reported. The nature of the previously described biologically active version of this compound warrants further investigation because we did not obtain a useful mass spectrum to support its structure.

In parallel to efforts to achieve the total synthesis of glycoproteins by strictly chemical means, a number of powerful chemical and biological technologies have emerged for making tailored, full-size glycoproteins bearing glycan-defined domains for the study of specific biological questions. These highly enabling strategies, which have been extensively reviewed elsewhere, include the chemoselective, site-specific glycosylation of recombinant proteins via the so-called "tag and modify" approach [107], in vitro chemoenzymatic glycosylation remodeling of heterogeneous glycoproteins [108], direct enzymatic glycosylation of proteins [109], specific glyco engineering of host biosynthetic pathways [108], and in vivo suppressor tRNA technology [110, 111].

All these approaches complement each other and it is envisaged that, either alone or in combination, they will continue to enable the generation of glycosylated proteins with predesigned modifications and increasing complexity for the elucidation of the molecular basis of protein glycosylation. This realization will certainly contribute to the development of novel and improved glycoprotein-derived therapeutics in the next few years.

Acknowledgements We thank the National Institutes of Health, William and Alice Goodwin and the Commonwealth Foundation for Cancer Research for generous financial support. AF-T gratefully acknowledges the European Commission (Marie Curie International Outgoing Fellowship) for funding. We acknowledge our colleagues of the Danishefsky laboratory whose work is presented in this manuscript. We thank Professors Yasuhiro Kajihara and Carlo Unverzagt for provocative exchanges and, particularly, for citing precedence on the feasibility of obtaining informative mass spectra of related EPO structures. We thank Rebecca Wilson for assistance in the preparation of the manuscript.

References

1. Walsh G, Jefferis R (2006) Nat Biotechnol 24:1241
2. Apweiler R, Hermjakob J, Sharon N (1999) BBA-Gen Subj 1473:4
3. Dwek RA (1996) Chem Rev 96:683
4. Varki A (1993) Glycobiology 3:97
5. Ohtsubo K, Marth JD (2006) Cell 126:855
6. Miller LH, Good MF, Milon G (1994) Science 264:1878
7. Dube DH, Bertozzi CR (2005) Nat Rev Drug Discov 4:477
8. Hudak JE, Bertozzi CR (2014) Chem Biol 21:16
9. Dalziel M, Crispin M, Scanlan CN, Zitzmann N, Dwek RA (2014) Science 343:1235681
10. Bosques CJ, Tschampel SM, Woods RJ, Imperiali B (2004) J Am Chem Soc 126:8421
11. Imperiali B, O′Connor SE (1999) Curr Opin Chem Biol 3:643
12. Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA (2002) Chem Rev 102:371
13. Li H, d'Anjou M (2009) Curr Opin Biotechnol 20:678
14. Spiro RG (2002) Glycobiology 12:43R
15. Rudd PM, Dwek RA (1997) Crit Rev Biochem Mol Biol 32:1
16. Rich JR, Wither SG (2009) Nat Chem Biol 5:206
17. Bertozzi CR, Kiessling LL (2001) Science 291:2357
18. Hojo H, Nakahara Y (2007) Biopolymers 88:308

19. Buskas T, Ingale S, Boons GJ (2006) Glycobiology 16:113R
20. Merrifield RB (1963) J Am Chem Soc 85:2149
21. Yamamoto N, Ohmori Y, Sakakibara T, Sasaki K, Juneja LR, Kajihara Y (2003) Angew Chem Int Ed 42:2537
22. Yamamoto N, Takayanagi A, Yoshino A, Sakakibara T, Kajihara Y (2007) Chem Eur J 13:613
23. Mathieux N, Paulsen H, Meldal M, Bock K (1997) J Chem Soc Perkin Trans 1:2359
24. Herzner H, Reipen T, Schultz M, Kunz H (2000) Chem Rev 100:4495
25. Gaidzik N, Westerlind U, Kunz H (2013) Chem Soc Rev 42:4431
26. Wilson RM, Danishefsky SJ (2013) J Am Chem Soc 135:14462
27. Lakshminarayanan V, Thompson P, Wolfert MA, Buskas T, Bradley JM, Pathangey LB, Madsen CS, Cohen PA, Gendler SJ, Boons GJ (2012) Proc Natl Acad Sci U S A 109:261
28. Cohen-Anisfeld ST, Lansbury PT (1993) J Am Chem Soc 115:10531
29. Bodanszky M, Natarajan S (1975) J Org Chem 40:2495
30. Wang P, Aussedat B, Vohra Y, Danishefsky SJ (2012) Angew Chem Int Ed 51:11571
31. Ullmann V, Rädisch M, Boos I, Freund J, Pöhner C, Schwarzinger S, Unverzagt C (2012) Angew Chem Int Ed 51:11566
32. Aussedat B, Vohra Y, Park PK, Fernández-Tejada A, Alam SM, Dennison SM, Jaeger FH, Anasti K, Stewart S, Blinn JH, Liao HX, Sodroski JG, Haynes BF, Danishefsky SJ (2013) J Am Chem Soc 135:13113
33. Payne RJ, Wong CH (2010) Chem Comm 46:21
34. Dawson P, Muir T, Clark-Lewis I, Kent SBH (1994) Science 266:776
35. Muir TW (2003) Annu Rev Biochem 72:249
36. Macmillan D, Bertozzi CR (2004) Angew Chem Int Ed 43:1355
37. Kent SBH (2009) Chem Soc Rev 38:338
38. Unverzagt C, Kajihara Y (2013) Chem Soc Rev 42:4408
39. Shin Y, Winans KA, Backes BJ, Kent SBH, Ellman JA, Bertozzi CR (1999) J Am Chem Soc 121:11684
40. Marcaurelle LA, Mizoue LS, Wilken J, Oldham L, Kent SB, Handel TM, Bertozzi CR (2001) Chem Eur J 7:1129
41. Mezzato S, Schaffrath M, Unverzagt C (2005) Angew Chem Int Ed 44:1650
42. Yamamoto N, Tanabe Y, Okamoto R, Dawson PE, Kajihara Y (2008) J Am Chem Soc 130:501
43. Mende F, Seitz O (2011) Angew Chem Int Ed 50:1232
44. Kan C, Danishefsky SJ (2009) Tetrahedron 65:9047
45. Yuan Y, Chen J, Wan Q, Wilson RM, Danishefsky SJ (2010) Biopolymers 94:373
46. Warren JD, Miller JS, Keding SJ, Danishefsky SJ (2004) J Am Chem Soc 126:6576
47. Wan Q, Chen J, Yuan Y, Danishefsky SJ (2008) J Am Chem Soc 130:15814
48. Yuan Y, Chen J, Wan Q, Tan Z, Chen G, Kan C, Danishefsky SJ (2009) J Am Chem Soc 131:5432
49. Haase C, Seitz O (2008) Angew Chem Int Ed 47:1553
50. Macmillan D, Anderson DW (2004) Org Lett 6:4659
51. Wu B, Chen J, Warren JD, Chen G, Hua Z, Danishefsky SJ (2006) Angew Chem Int Ed 45:4116
52. Yan LZ, Dawson PE (2001) J Am Chem Soc 123:526
53. Pentelute BL, Kent SBH (2007) Org Lett 9:687
54. González A, Valencia G (1998) Tetrahedron Asymmetry 9:2761
55. Wan Q, Danishefsky SJ (2007) Angew Chem Int Ed 46:9248
56. Shang S, Tan Z, Danishefsky SJ (2011) Proc Natl Acad Sci U S A 108:5986
57. Dong S, Shang S, Li J, Tan Z, Dean T, Maeda A, Gardella TJ, Danishefsky SJ (2012) J Am Chem Soc 134:15122
58. Li J, Dong S, Townsend SD, Dean T, Gardella TJ, Danishefsky SJ (2012) Angew Chem Int Ed 51:12263

59. Brailsford JA, Danishefsky SJ (2012) Proc Natl Acad Sci U S A 109:7196
60. Chen J, Wan Q, Yuan Y, Zhu J, Danishefsky SJ (2008) Angew Chem Int Ed 47:8521
61. Haase C, Rohde H, Seitz O (2008) Angew Chem Int Ed 47:6807
62. Chen J, Wang P, Zhu J, Wan Q, Danishefsky SJ (2010) Tetrahedron 66:2277
63. Harpaz Z, Siman P, Kumar KSA, Brik A (2010) ChemBioChem 11:1232
64. Tan Z, Shang S, Danishefsky SJ (2010) Angew Chem Int Ed 49:9500
65. Yang R, Pasunooti KK, Li F, Liu X-W, Liu C-F (2009) J Am Chem Soc 131:13592
66. Ajish Kumar KS, Haj-Yahya M, Olschewski D, Lashuel HA, Brik A (2009) Angew Chem Int Ed 48:8090
67. Shang S, Tan Z, Dong S, Danishefsky SJ (2011) J Am Chem Soc 133:10784
68. Townsend SD, Tan Z, Dong S, Shang S, Brailsford JA, Danishefsky SJ (2012) J Am Chem Soc 134:3912
69. Crich D, Banerjee A (2007) J Am Chem Soc 129:10064
70. Tam JP, Yu Q (1998) Biopolymers 46:319
71. Saporito A, Marasco D, Chambery A, Botti P, Monti SM, Pedone C, Ruvo M (2006) Biopolymers 83:508
72. Okamoto R, Kajihara Y (2008) Angew Chem Int Ed 47:5402
73. Malins LR, Cergol KM, Payne RJ (2013) ChemBioChem 14:559
74. Thompson RE, Chan B, Radom L, Jolliffe KA, Payne RJ (2013) Angew Chem Int Ed 52:9723
75. Malins LR, Cergol KM, Payne RJ (2014) Chem Sci 5:260
76. Cergol KM, Thompson RE, Malins LR, Turner P, Payne RJ (2014) Org Lett 16:290
77. Nagorny P, Fasching B, Li X, Chen G, Aussedat B, Danishefsky SJ (2009) J Am Chem Soc 131:5792
78. Zhang YM, Mallet JM, Sinaÿ P (1992) Carbohydr Res 236:73
79. Likhosherstov LM, Novikova OS, Derevitskaja VA, Kochetkov NK (1986) Carbohydr Res 146:C1
80. Sakakibara S (1995) Biopolymers 37:17
81. Aussedat B, Fasching B, Johnston E, Sane N, Nagorny P, Danishefsky SJ (2012) J Am Chem Soc 134:3532
82. Howles CM (1996) Hum Reprod Update 2:172
83. Pang SC (2005) Womens Health 1:87
84. Nagorny P, Sane N, Fasching B, Aussedat B, Danishefsky SJ (2012) Angew Chem Int Ed 51:975
85. Stenman UH, Tiitinen A, Alfthan H, Valmu L (2006) Hum Reprod Update 12:769
86. Talwar GP, Vyas HK, Purswani S, Gupta JC (2009) J Reprod Immunol 83:158
87. Iversen PL, Mourich DV, Moulton HM (2003) Curr Opin Mol Ther 5:156
88. Valmu L, Alfthan H, Hotakainen K, Birken S, Stenman UH (2006) Glycobiology 16:1207
89. Fernández-Tejada A, Vadola PA, Danishefsky SJ (2014) J Am Chem Soc 136:8450
90. Shi Y, Liu CH, Roberts AI, Das J, Xu G, Ren G, Zhang Y, Zhang L, Yuan ZR, Tan HS, Das G, Devadas S (2006) Cell Res 16:126
91. Zhang Q, Johnston EV, Shieh J-H, Moore MAS, Danishefsky SJ (2014) Proc Natl Acad Sci U S A 111:2885
92. Krantz SB (1991) Blood 77:419
93. Lacombe C, Mayeux P (1998) Haematologica 83:724
94. Koury MJ (2005) Exp Hematol 33:1263
95. Tonelli M, Hemmelgarn B, Reiman T, Manns B, Reaume MN, Lloyd A, Wiebe N, Klarenbach S (2009) CMAJ 180:E62
96. Mattio M, Ceaglio N, Oggero M, Perotti N, Amadeo I, Orozco G, Forno G, Kratje R, Etcheverrigaray M (2011) Biotechnol Prog 27:1018
97. Higuchi M, Oh-eda M, Kuboniwa H, Tomonoh K, Shimonaka Y, Ochi N (1992) J Biol Chem 267:7703
98. Wang P, Dong S, Brailsford JA, Iyer K, Townsend S, Zhang Q, Hendrickson RC, Shieh JH, Moore MAS, Danishefsky SJ (2012) Angew Chem Int Ed 51:11576

99. Schwarz JB, Kuduk SD, Chen X-T, Sames D, Glunz PW, Danishefsky SJ (1999) J Am Chem Soc 121:2662
100. Bang D, Pentelute BL, Kent SBH (2006) Angew Chem Int Ed 45:3985
101. Wang P, Dong S, Shieh J-H, Peguero E, Hendrickson R, Moore MAS, Danishefsky SJ (2013) Science 342:1357
102. Takeuchi M, Inoue N, Strickland TW, Kubota M, Wada M, Shimizu R, Hoshi S, Kozutsumi H, Takasaki S, Kobata A (1989) Proc Natl Acad Sci U S A 86:7819
103. Kochendoerfer GG, Chen S-Y, Mao F, Cressman S, Traviglia S, Shao H, Hunter CL, Low DW, Cagle EN, Carnevali M, Gueriguian V, Keogh PJ, Porter H, Stratton SM, Wiedeke MC, Wilken J, Tang J, Levy JJ, Miranda LP, Crnogoran MM, Kalbag S, Potti P, Schindler-Horvat J, Savatski L, Adamson JW, Kung A, Kent SBH, Bradburne JA (2003) Science 299:884
104. Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi B-K, Hopkins D, Wischnewski H, Roser J, Mitchell T, Strawbridge RR, Hoopes J, Wildt S, Gerngross TU (2006) Science 313:1441
105. Sanz-Nebot V, Benavente F, Vallverdú A, Guzman NA, Barbosa J (2003) Anal Chem 75:5220
106. Harazono A, Hashii N, Kuribayashi R, Nakazawa S, Kawasaki N (2013) J Pharm Biomed Anal 87:65
107. Chalker JM, Bernardes GJ, Davis BG (2011) Acc Chem Res 44:730
108. Wang LX, Lomino JV (2012) ACS Chem Biol 7:110
109. Wang LX, Amin MN (2014) Chem Biol 21:51
110. Wang L, Schultz PG (2005) Angew Chem Int Ed 44:34
111. Liu L, Bennett CS, Wong CH (2006) Chem Comm 42:21

Top Curr Chem (2015) 362: 27–88
DOI: 10.1007/128_2014_584

Published online: 18 December 2014

Modern Extensions of Native Chemical Ligation for Chemical Protein Synthesis

Lara R. Malins and Richard J. Payne

Abstract Over the past 20 years, native chemical ligation has facilitated the synthesis of numerous complex peptide and protein targets, with and without post-translational modifications, as well as the design and construction of a variety of engineered protein variants. This powerful methodology has also served as a platform for the development of related chemoselective ligation technologies which have greatly expanded the scope and flexibility of ligation chemistry. This chapter details a number of important extensions of the original native chemical ligation manifold, with particular focus on the application of new methods in the total chemical synthesis of proteins. Topics covered include the development of auxiliary-based ligation methods, the post-ligation manipulation of Cys residues, and the synthesis and utility of unnatural amino acid building blocks (bearing reactive thiol or selenol functionalities) in chemoselective ligation chemistry. Contemporary applications of these techniques to the total chemical synthesis of peptides and proteins are described.

Keywords Deselenization · Desulfurization · Native chemical ligation · Post-translational modifications · Protein synthesis

Contents

L.R. Malins (✉) and R.J. Payne (✉)
School of Chemistry, The University of Sydney, Camperdown, NSW 2006, Australia
e-mail: lara.malins@sydney.edu.au; richard.payne@sydney.edu.au

1 Introduction

Proteins are an important class of biomolecules which play a crucial role in a number of diverse physiological processes, including structure, transport, storage and catalysis. The immense structural and functional complexity of proteins is a consequence not only of the unique, genetically-encoded primary sequence of constituent amino acids, but also of the potential introduction of covalent post-translational modifications (PTMs), which occurs enzymatically after protein translation on the ribosome. Each of these factors has a profound impact on the three-dimensional structure, and ultimately the function and/or activity of a protein target.

The study of proteins and their use as novel therapeutics has been aided by the ability to access significant quantities of target molecules using biological expression of programmed genetic sequences. However, such technologies alone are insufficient to enable the flexible design of protein therapeutics, or even the detailed study of all native proteins comprising an organism's proteome, which is considerably more diverse than would be predicted simply from the finite size of the genome. The difference is due, in part, to the enzyme-mediated post-translational modification of proteins. The diverse array of PTMs, including acetylation, phosphorylation, glycosylation, methylation, hydroxylation, and ubiquitylation, among others, occur on the majority of the 20 common proteinogenic amino acid side-chains and, in some cases, on the amide backbone [1]. These modifications serve to greatly enhance proteome diversity beyond the intrinsic size of the genome, and also have significant effects on protein conformation, localization, and function. It is predicted that approximately 5% of the human genome is dedicated to encoding enzymes responsible for PTMs [2].

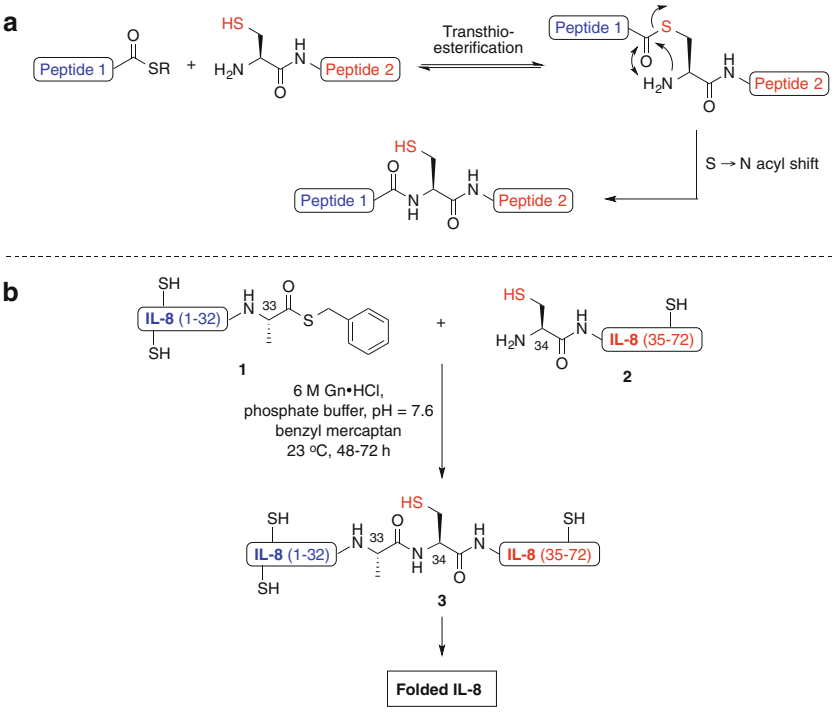
As a consequence of their incorporation through enzyme-mediated processes, PTMs, unlike the primary structure of proteins, are not under direct template control and are instead dictated by the relative intracellular concentration of processing enzymes. This has important implications in cell-signaling pathways, in which reversible modifications, such as phosphorylation and acetylation, dictate both the strength and duration of a signaling process [2]. As a result, the composition of the proteome is inherently dynamic, and the variability of PTMs results in the concomitant heterogeneity of native proteins derived or isolated from living systems. Given these considerations, the role of synthetic chemistry, in the context of the

total synthesis of proteins, becomes paramount to obtaining homogenous, post-translationally modified proteins for the purpose of understanding the biological role of such modifications. The therapeutic potential of modified peptide and protein targets may also benefit from the design and synthesis of engineered variants accessed in homogeneous form via chemical synthesis. Indeed, our understanding of the link between protein structure and function, and our ability to exploit this knowledge for the development of novel therapeutics, is closely connected to our aptitude in the chemical synthesis of complex targets. This chapter therefore focuses on the development of synthetic methods, specifically peptide ligation methods, for the convergent assembly of large polypeptides and proteins from smaller peptide fragments. Particular emphasis is placed on extensions to the native chemical ligation reaction, which has found enormous application in the chemical synthesis of proteins, with and without modifications, since the first report of the method two decades ago [3].

2 Native Chemical Ligation

The development of native chemical ligation by Kent and coworkers in 1994 marked a substantial step in the area of chemical protein synthesis by allowing, for the first time, the chemoselective construction of a native amide linkage between two fully unprotected peptides under mild, aqueous conditions [3]. Building upon a 1953 report by Wieland et al. for the linking of amino acids [4], the method involves the condensation of a peptide bearing an N-terminal cysteine (Cys) residue with a peptide bearing a C-terminal thioester (Scheme 1a). Mechanistically, the reaction occurs through a reversible transthioesterification reaction between the Cys thiol functionality and the C-terminal peptide thioester, generating an intermediate thioester. Spontaneous rearrangement of the bridged thioester moiety in a proximity-induced intramolecular $S \rightarrow N$ acyl shift then occurs through a five-membered ring intermediate to afford a new amide bond [3]. Importantly, the reaction proceeds in the presence of all native amino acid side-chains (including Cys residues distant from the ligation site) and enables a modular approach to the construction of proteins from peptide fragments prepared via iterative solid-phase peptide synthesis (SPPS) [5], a highly robust method for the construction of defined polypeptide sequences.

Kent and coworkers first demonstrated the power of the native chemical ligation methodology toward protein synthesis in the construction of human interleukin-8 (IL-8), a 72-amino acid polypeptide chain bearing 18 of the 20 common proteinogenic amino acids and 4 Cys residues, which form disulfide linkages in the native protein. This was accomplished via efficient ligation of a 33-amino acid peptide bearing a C-terminal benzyl thioester **1** with a 39-residue peptide **2** containing an N-terminal Cys residue. The reaction was performed in aqueous buffer (pH 7.6) containing a denaturing chaotropic salt (6 M guanidine hydrochloride) and afforded a single peptide product **3**, despite the presence of additional



Scheme 1 (**a**) Native chemical ligation. (**b**) Synthesis of human IL-8 by native chemical ligation

unprotected Cys residues in both peptide fragments (Scheme 1b). Selective reaction at the N-terminal Cys residue was attributed to the reversible nature of the initial transthioesterification reaction. Although internal Cys residues can react with the C-terminal benzyl thioester, the resultant thioester intermediates were unproductive in the ligation reaction because of the lack of a proximal amine group to promote the amide bond-forming $S \rightarrow N$ acyl shift, an irreversible process under the conditions employed for the ligation reaction. In the synthesis of IL-8, excess benzyl mercaptan was added to the reaction media to promote thiol exchange with the unproductive thioester intermediates, thereby regenerating thioester **1** and enabling productive transthioesterification with the desired N-terminal Cys residue of peptide **2**. Conveniently, the presence of excess thiol also served as a reducing agent to prevent the formation of Cys disulfides [3].

While the concept of peptide ligation chemistry had previously been explored by others, including Kemp and coworkers [6–8], Liu and Tam [9, 10], and Schnölzer and Kent for the construction of backbone-modified proteins [11], the total chemical synthesis of IL-8 by native chemical ligation marked the first preparation of a protein target, bearing a fully native polypeptide backbone, through the chemoselective condensation of unprotected peptide fragments. It was predicted that this modular synthetic approach would enable nearly unlimited variation in the covalent structure of proteins, thereby facilitating the systematic study of the

structure and function of these important macromolecules [3]. This has indeed proved to be the case with hundreds of proteins prepared using this robust methodology.

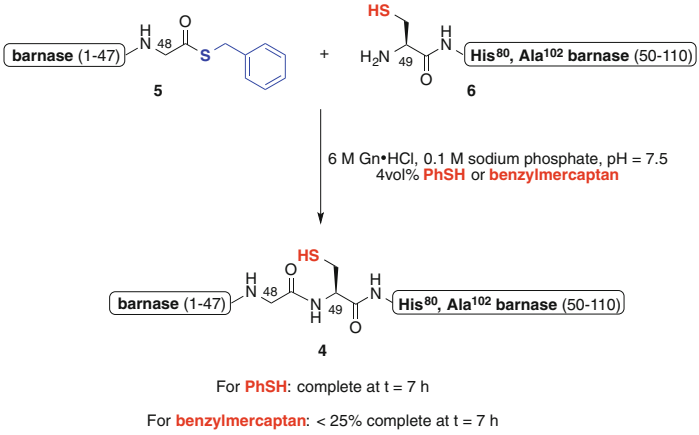
2.1 Scope and Mechanism

Shortly after the initial discovery of native chemical ligation, a number of studies were subsequently conducted with the aim of establishing the scope of the ligation methodology and deciphering the subtleties of the reaction mechanism. In particular, the dependence of the rate of ligation on the nature of the C-terminal peptide thioester moiety has been extensively studied.

Historically, C-terminal peptide thioesters for use in native chemical ligation have been prepared as the alkyl thioester derivatives using optimized in situ neutralization Boc-strategy SPPS [12]. The C-terminal benzyl thioesters initially employed by Kent and coworkers, for example, were prepared via Boc-SPPS using an HF-labile thioacid linker followed by a subsequent alkylation with benzyl bromide [3]. Advances in thioester linkers developed shortly thereafter facilitated the direct synthesis of peptide alkyl thioesters upon cleavage from the resin [13, 14]. The facile handling and preparation and the general stability of alkyl thioesters render these functionalities a convenient choice for use in native chemical ligation reactions. Nonetheless, in terms of acylating power, alkyl thioesters are relatively unreactive acyl donors, a result of the modest leaving group ability of the alkyl thiol component. The use of a more reactive peptide thioester, containing a better thiol leaving group, was first investigated in the seminal report of native chemical ligation, whereby a thioester bearing a 5-thio-2-nitrobenzoic acid (the reduced form of Ellman's reagent) leaving group was shown to facilitate rapid ligation [3].

As a corollary to this observation, one of the earliest general advances to the native chemical ligation method was the development and exploration of thiol ligation catalysts with a view to modulating thioester reactivity [15]. In 1997, Kent and coworkers found that the inclusion of thiophenol as an exogenous ligation additive served to facilitate thiol-thioester exchange with preformed alkyl thioesters, thereby promoting the in situ formation of the considerably more reactive peptide aryl thioester (Scheme 2) to facilitate more rapid native chemical ligation reactions. To enable a direct comparison between benzyl thiol and thiophenol as effective ligation additives, the authors explored the ligation-based assembly of a Cys-containing analogue of barnase **4**, a 110-amino acid microbial ribonuclease. Construction of the barnase analogue from barnase(1–48)-COS-benzyl **5**, a preformed peptide C-terminal benzyl thioester, and a modified N-terminal fragment, [Cys^{49}, His^{80}, Ala^{102}]barnase(49–110) **6**, in the presence of thiophenol proceeded to completion in approximately 7 h (Scheme 3). In contrast, in the presence of benzyl mercaptan, the ligation reaction was less than 25% complete after the same period of time. The observed rate enhancement for the ligation of a preformed benzyl thioester in the presence of exogenous thiophenol established in

Scheme 2 Thiophenol as an exogenous thiol additive for native chemical ligation



Scheme 3 The effect of thiophenol as an exogenous thiol additive for the construction of barnase analogue **4**

situ transthioesterification with thiol additives as a practical and general means of modulating the reactivity of C-terminal peptide alkyl thioesters for use in native chemical ligation [15].

More recently, a detailed study of the ability of various thiol additives to promote native chemical ligation identified the water-soluble aryl thiol mercaptophenylacetic acid (MPAA) as the optimal thiol catalyst [16]. This study also established an important relationship between ligation rate and the pK_a of the thiol additive. The most effective additives, such as thiophenol and MPAA, had pK_a values between approximately 6 and 8, thus maintaining sufficient nucleophilicity to promote the initial thiol-thioester exchange with the less reactive peptide alkyl thioester component, and adequate leaving group ability for effective transthioesterification with the Cys-containing peptide fragment [16]. Thiol additives with higher pK_a values (e.g., the water-soluble alkyl thiol sodium 2-mercaptoethanesulfonate (MESNa) [17], commonly employed in expressed protein ligation (EPL) [18]) rapidly exchanged with preformed alkyl thioester derivatives but were poor leaving groups in the reaction with a peptide bearing an N-terminal Cys residue. Conversely, thiol additives with a pK_a value less than six were inefficient ligation additives as a result of their inability to effectively activate preformed alkyl thioesters through thiol-thioester exchange [16]. It should also be noted that highly electrophilic acyl donors may increase the likelihood of thioester hydrolysis or epimerization at the α-center of the C-terminal amino acid residue,

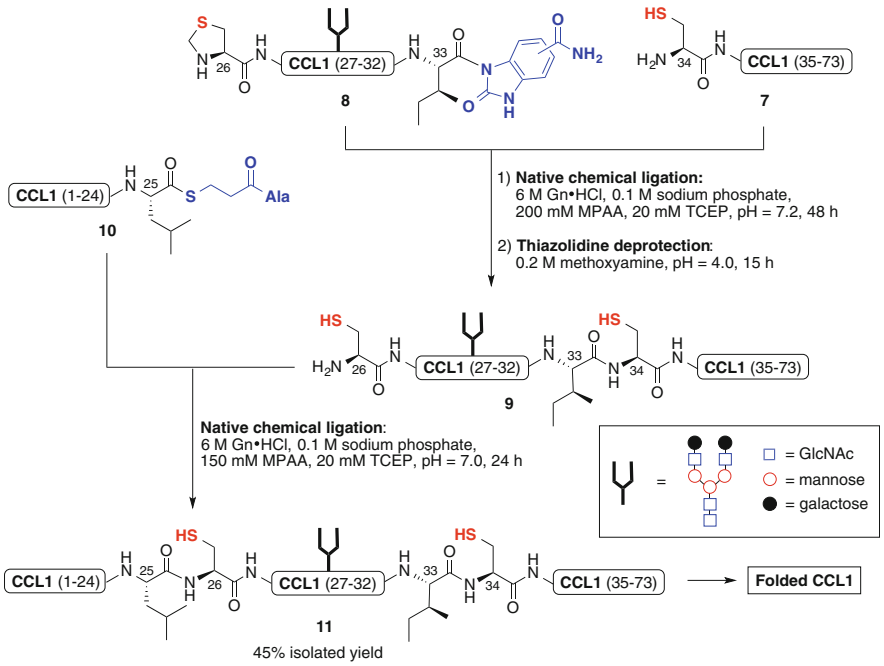
particularly with challenging ligation junctions. In addition to the pK_a of the thiol used in the reaction, these factors must be carefully considered in the design of new ligation additives. The development of other thiol catalysts for application in ligation chemistry is further discussed later in this chapter.

An equally important determinant of thioester reactivity in native chemical ligation reactions is the nature of the C-terminal residue bearing the reactive thioester functionality. In 1999, Dawson and coworkers carried out a comprehensive study of the rate of ligation as a function of the identity of the C-terminal amino acid residue [13]. In a series of model ligation reactions, whereby C-terminal thioesters derived from each of the 20 proteinogenic amino acids were examined, the reaction rate was shown to be closely correlated with the steric and electronic nature of the C-terminal thioester moiety. For example, reactions at sterically unencumbered C-terminal amino acid thioesters proceeded much faster than reactions employing bulky, β-branched amino acid thioesters, such as those derived from Ile, Val, and Thr, which, under the ligation conditions employed, required 48 h or more to reach completion. Electronic effects were likewise apparent, with Cys and His thioesters reacting at remarkably similar rates to the model Gly thioester, despite the increased steric bulk, and much faster than the corresponding Ala thioester [13]. Poor reactivity observed with the model Pro thioester was subsequently explained by the decreased electrophilicity of the Pro thioester carbonyl carbon resulting from an orbital overlap ($n \rightarrow \pi^*$) with the adjacent N-terminal amide oxygen atom [19].

The power of native chemical ligation together with an extended knowledge of the specific reactivity of C-terminal peptide thioesters have been exploited in the total chemical synthesis of numerous proteins to date [20–28], greatly expanding the scope and size of native protein targets within the grasp of synthetic chemists. In combination with SPPS as a robust approach to access peptide fragments bearing native amino acids as well as unnatural amino acid building blocks (derived from modern organic synthesis), the method has advanced the opportunity for protein engineering and structural remodeling. As a testament to the power of native chemical ligation, 20 years after its seminal report the technique is often employed, in near original form, for the total synthesis of complex and post-translationally modified proteins.

2.2 Modern Application

A recent, illustrative example from Okamoto et al. of the extensive applicability of native chemical ligation toward the study of protein structure and function is the total chemical synthesis of the glycosylated and non-glycosylated forms of the 73-amino acid chemokine protein CCL1 and a 74-amino acid, N-terminally extended variant, Ser-CCL1 [29]. The construction of these chemokines was accomplished using a convergent, three-component iterative native chemical ligation approach (see Scheme 4 for a representative example), with the glycosylated



Scheme 4 Total chemical synthesis of glycosylated CCL1 using an iterative native chemical ligation strategy

variants each containing a complex, *N*-linked asialo-nonasaccharide at the native glycosylation site, Asn[29].

Notably, preparation of the requisite peptide fragments for the construction of the target proteins strategically utilized a number of recent and powerful advances to the native chemical ligation motif. Using the synthesis of glycosylated CCL1 as a representative model (Scheme 4), the first step in the iterative ligation strategy involved the reaction of C-terminal Cys-containing peptide CCL1(34–73) **7**, prepared by Boc-SPPS, with bifunctional glycopeptide **8**. To enable chemoselective iterative ligation, the N-terminal Cys residue of glycopeptide **8** was masked as the corresponding thiazolidine (Thz) residue [30], which is orthogonal to the general conditions employed in solid-phase peptide synthesis, but is mildly removed in the presence of methoxyamine at pH 4.0. Furthermore, due to the incompatibility of glycans with the strongly acidic deprotection conditions (e.g., HF) employed in Boc-SPPS, glycopeptide fragment **8** was synthesized via 9-fluorenylmethoxycarbonyl (Fmoc)-SPPS, in which cleavage from the resin is accomplished using milder conditions. Nonetheless, thioester linkers are incompatible with the basic conditions employed for unmasking the N-terminal Fmoc group in standard Fmoc-SPPS. As such, the authors employed a diaminobenzoic acid (Dbz) linker for the Fmoc-based assembly of bifunctional peptide **8**, which bears a C-terminal peptide *N*-acylbenzimidazolinone (Nbz), a novel peptide thioester

precursor accessible from resin-bound *o*-aminoanilides, originally described by Dawson and coworkers in 2008 [31]. In the presence of thiol additives under standard native chemical ligation conditions, the peptide-Nbz group undergoes facile thiolysis to afford the corresponding peptide thioester.

Following the preparation of CCL1 peptide fragments **7** and **8**, the fragments were subsequently joined using native chemical ligation in aqueous denaturing buffer in the presence of 200 mM MPAA as a thiol catalyst and 20 mM of the water soluble phosphine reductant, tris(2-carboxyethyl)phosphine (TCEP), to prevent the formation of Cys disulfides (Scheme 4). The reaction required 48 h to reach completion because of the use of an acyl donor fragment bearing a bulky, β-branched isoleucine (Ile) residue as the C-terminal amino acid. Completion of the ligation, as monitored by HPLC-MS, was directly followed by thiazolidine deprotection in the presence of methoxyamine. Crude fragment **9**, bearing an unmasked N-terminal Cys residue, was then ligated to peptide thioester **10**, corresponding to CCL 1(1–24), affording the full-length glycosylated CCL1 **11** in 45% isolated yield for the two ligation steps following HPLC purification. Similar ligation protocols were employed for the synthesis of the non-glycosylated CCL1 variant and for the glycosylated and non-glycosylated Ser-CCL1 derivatives. Each of the CCL1 variants were folded and evaluated in a chemotaxis assay, allowing for a systematic study of the effect of glycosylation on the chemotactic activity of the proteins. Notably, Kent and coworkers were also able to prepare a non-glycosylated mirror image Ser-CCL1 protein composed entirely of D-amino acids using an analogous synthetic strategy [32]. This D-protein analogue was used to obtain the X-ray crystal structures of both Ser-CCL1 and glycosyl-Ser-CCL1 using racemic and quasi-racemic crystallization, respectively [32].

As the preparation of large quantities of complex, homogeneous post-translationally modified proteins and the construction of unnatural D-proteins represent synthetic feats currently unachievable using biological expression or the majority of other chemical methods, the work described by Okamoto et al. on CCL1 [29, 32] highlights the critical importance of native chemical ligation as an enabling tool for understanding protein structure and function. Importantly, native chemical ligation has also served as a platform for the development of a myriad of related synthetic technologies for application in chemoselective peptide ligations. The remainder of this chapter discusses a number of powerful extensions to the original native chemical ligation manifold, with a particular focus on the development of new technologies for the total chemical synthesis of proteins. Topics covered include the development of auxiliary-based methods for peptide ligation, the post-ligation manipulation of Cys residues, and the synthesis and utility of unnatural amino acid building blocks in native chemical ligation-like reactions. In the process, contemporary applications of these techniques to the total chemical synthesis of peptides and proteins are described.

3 Development of New Cysteine Ligation Surrogates

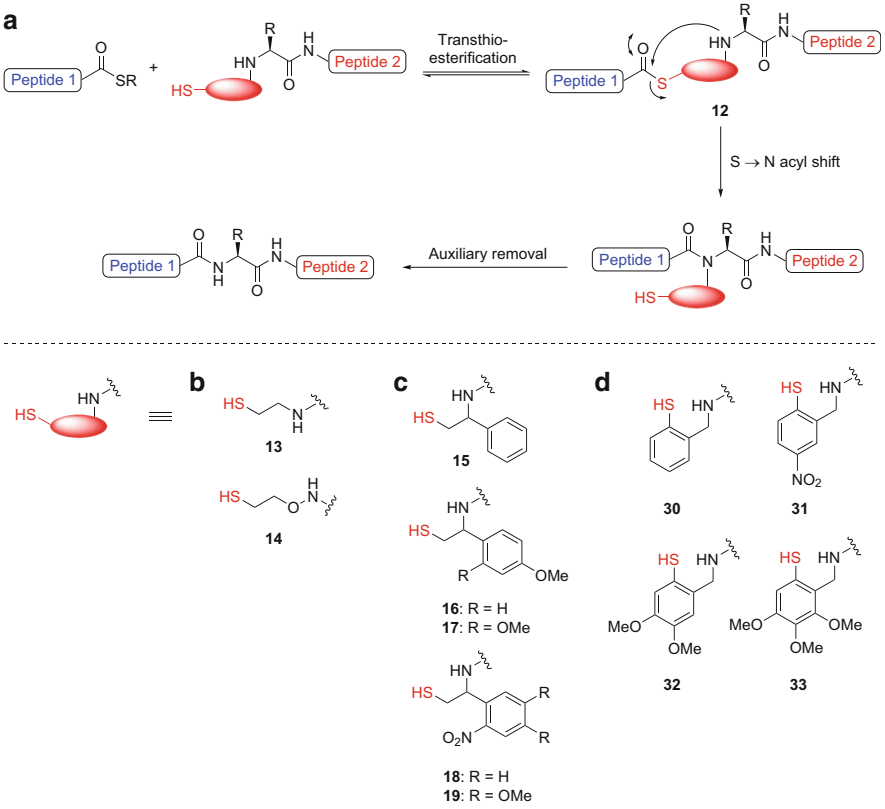
The successful disconnection of proteins using native chemical ligation is predicated on the presence of appropriately placed Cys residues within a target sequence. While there are countless targets that fulfill this unique requirement, the relative scarcity of Cys in naturally occurring proteins (1.1%) [33] precludes the ligation-based assembly of a number of desirable targets and represents a limitation to native chemical ligation in its original form. A number of innovative strategies have therefore been developed to address this intrinsic limitation. The following discussion provides an overview of some of the most significant discoveries, based on the overall logic of native chemical ligation, toward circumventing the reliance on N-terminal Cys residues.

3.1 Auxiliary-Based Methods

3.1.1 N-Terminal Ligation Auxiliaries

One of the earliest and most heavily explored means of circumventing the need for an N-terminal Cys residue has been the use of removable N-terminal thiol ligation auxiliaries to mimic the role of the Cys thiol group in the ligation reaction [27]. A summary of effort in this area is shown in Scheme 5. The general mechanism of N-terminal auxiliary-based ligation is analogous to that of native chemical ligation and is conceptually similar to the original "prior thiol capture" technique employed by Kemp and coworkers [6, 7], whereby a proximity-induced acyl shift event is templated by a traceless reaction auxiliary. In the case of N-terminal auxiliaries, the N-linked thiol reaction auxiliary participates in an initial transthioesterification step, generating the intermediate thioester **12**. A subsequent acylation of the auxiliary-bound secondary amine occurs via an $S \rightarrow N$ shift to generate an amide bond at the ligation site (Scheme 5a). Cleavage of the tethered reaction auxiliary following the ligation event then generates native peptide products. A variety of N-terminal reaction auxiliaries have been explored for this purpose (Scheme 5b–d), allowing access to select non-Cys ligation junctions.

The first N-terminal thiol auxiliary approach was developed by Kent and coworkers in 1996, with the installation of ethanethiol **13** and oxyethanethiol **14** auxiliaries (Scheme 5b) on the terminal amine functionality of peptides bearing N-terminal Gly and Ala residues [34]. These auxiliaries were shown to promote ligation with various C-terminal peptide thioesters in a number of model ligation reactions. Interestingly, when the oxyethanethiol auxiliary **14** was used to ligate peptides bearing steric bulk on either the N-terminal or C-terminal residue of the ligation junction, the authors detected substantial amounts of the unrearranged thioester intermediate (see **12**, Scheme 5a) in the ligation mixture. This observation suggested a decrease in the rate of acyl transfer of the thioester intermediate to the auxiliary-bound, secondary amine. It was postulated that the added steric bulk at the



Scheme 5 (**a**) Native chemical ligation facilitated by N-terminal ligation auxiliaries. (**b**) Oxyethanethiol and ethanethiol auxiliaries. (**c**) 1-Phenylethanethiol auxiliaries. (**d**) 2-Mercaptobenzyl auxiliaries

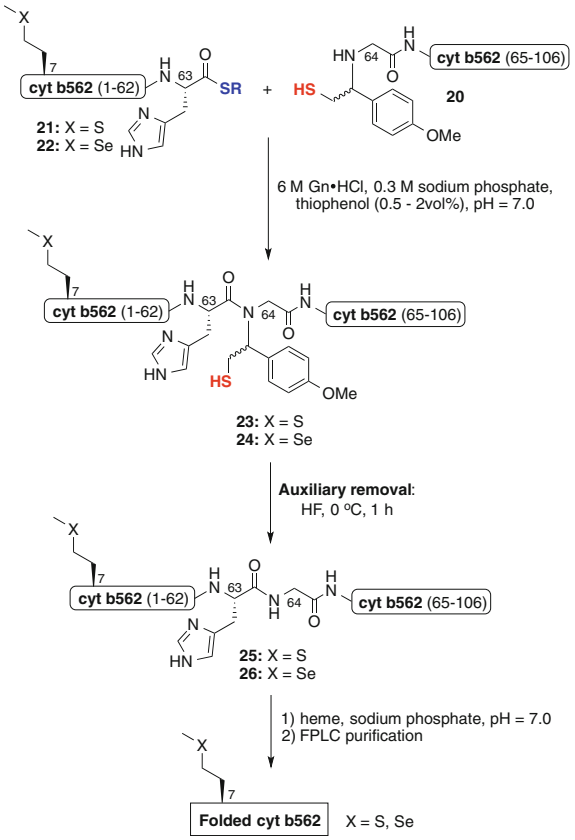
ligation junction in combination with the requirement for a six-membered ring intermediate in the intramolecular rearrangement of the oxyethanethiol-derived thioester (rather than the five-membered ring intermediate generated at unsubstituted Cys residues) served to slow substantially, or in some cases halt, the rate of the $S \rightarrow N$ acyl shift. Nonetheless, in ligation reactions in which amide bond formation was successfully mediated by the oxyethanethiol linker, removal of the ligation auxiliary was readily achieved through cleavage of the N–O bond upon treatment with Zn dust in acidic media. Despite some limitations, this technique served as a proof of concept for the generation of native peptide products following an auxiliary-based ligation strategy [34].

With the aim of further generalizing auxiliary ligations and increasing reaction rates, a number of 1-phenylethane thiol-based auxiliaries (**15**–**19**) (Scheme 5c) [35–39] were also explored. Importantly, these scaffolds were designed to facilitate intramolecular acyl shift through a five-membered ring intermediate and were

proposed to increase the rate of $S \rightarrow N$ acyl transfer relative to oxyethanethiol linker **14** [35]. An additional consideration in the strategic development of the methoxy benzyl auxiliaries **16** and **17** was the ease of removal following the ligation reaction [36, 40]. When bound to the N-terminal α-amino moiety of a peptide fragment, auxiliaries **16** and **17** were not labile to treatment with HF, and were thus compatible with standard Boc-SPPS conditions. The observed stability was attributed to the positive charge ensuing from protonation of the terminal amine under acidic cleavage conditions, which possibly disfavored the formation of a proximal benzylic cation during cleavage of the methoxybenzyl moiety [37]. However, amide bond formation as a consequence of the ligation reaction effectively increased the acid lability of the thiol auxiliary, which was readily cleaved in a post-ligation HF treatment. It is also important to note that, despite the potential for chirality in the 1-phenylethane thiol scaffold, auxiliaries **16** and **17** were synthesized as a mixture of diastereomers [36]. In a contemporary study by Dawson and coworkers on 1-phenylethane thiol derivative **15**, configuration at the benzylic position was determined to have no effect on the rate of ligation [35].

The first application of an N-terminal ligation auxiliary to the total synthesis of proteins was demonstrated through the use of *p*-methoxy auxiliary **16** in the ligation-based assembly of the 106-amino acid metalloprotein cytochrome b562 and an engineered analogue, [SeMet^{7}]cyt b562, containing a strategically placed selenomethionine residue to serve as an axial ligand for the cytochrome heme iron atom [37]. The synthesis of these two proteins was accomplished via ligation of N-terminal auxiliary bound peptide **20** to peptide thioesters **21** and **22**, bearing a native methionine residue or a selenomethionine residue, respectively, at position 7 (Scheme 6). The reaction was performed under aqueous, denaturing conditions in the presence of thiophenol as an exogenous thiol additive. Following the successful ligation, treatment of **23** and **24** with anhydrous HF facilitated clean removal of the N-terminal auxiliary to afford the full-length polypeptide chains **25** and **26**. The target protein molecules were then obtained after folding in the presence of heme, thus enabling a detailed analysis of the spectroscopic and electrochemical properties of wild-type cyt b562 relative to the engineered, selenomethionine-containing protein [37].

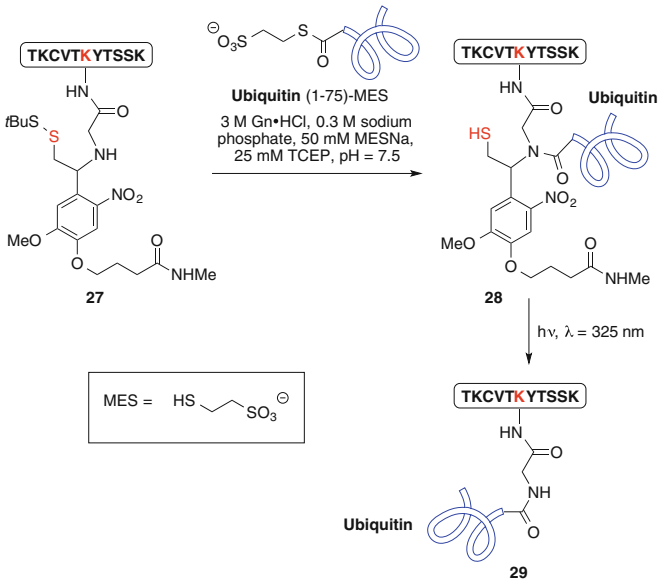
The successes of the 1-phenylethane thiol auxiliary scaffold in protein synthesis prompted the subsequent exploration of photolabile derivatives [35] (such as **18** and **19**) [39, 38] bearing the *ortho*-nitrobenzyl motif (see Scheme 5c), to allow for auxiliary removal without the use of harsh acidic conditions. These reaction auxiliaries were shown to facilitate model ligation reactions at Gly–Gly [38] and Ala–Gly [39] junctions, and could be removed by facile photolysis upon irradiation with UV light. Muir and coworkers have also reported the innovative application of a photolabile linker in auxiliary-mediated expressed protein ligation (EPL) [41] for the synthesis of a ubiquitylated peptide target [42]. In this study, the authors ligate recombinantly produced ubiquitin and SUMO (small ubiquitin-related modifier) thioesters to a small peptide fragment **27**, from the mammalian histone nucleoprotein H2B, bearing an auxiliary-modified Lys residue (see Scheme 7 for ubiquitin example). Following ligation, irradiation of the ubiquitylated product **28** at 325 nm



Scheme 6 Synthesis of cyt b562 and [SeMet^{7}]cyt b562 via native chemical ligation facilitated by an N-terminal ligation auxiliary

effected rapid removal of the photolabile auxiliary to afford chemically modified peptide **29**. Importantly, the final ubiquitin-modified protein was also shown to be a suitable substrate for the ubiquitin-dependent hydrolase UCH-L3, thus confirming the structural and functional integrity of the semi-synthetic protein [42]. A subsequent report from the Muir laboratory utilized the same photolytic ligation auxiliary, in combination with a second EPL reaction, to achieve the semi-synthesis of the full-length, chemically ubiquitylated H2B histone [43].

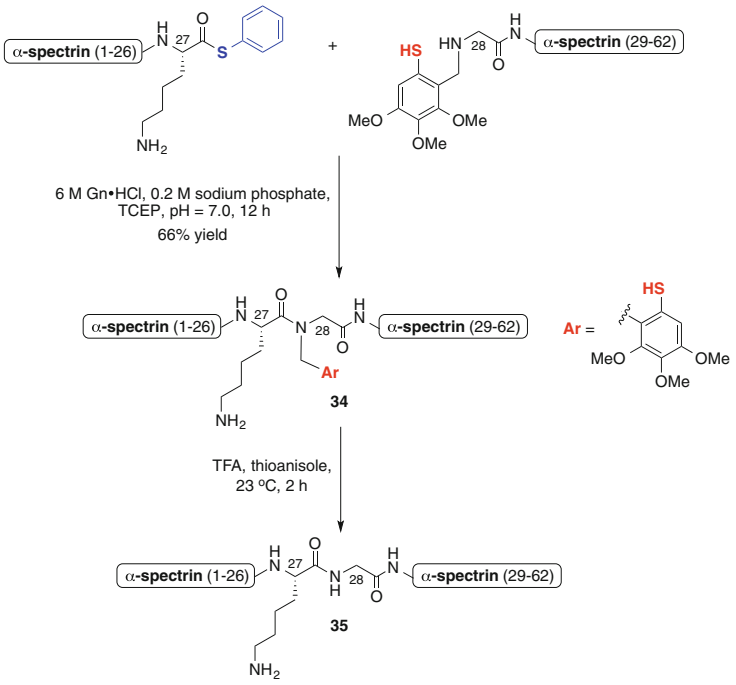
A third class of auxiliaries probed for use in ligation reactions employed the 2-mercaptobenzyl motif (e.g., **30**–**33**, Scheme 5d) [44–47]. The rationale for this tethered thiophenol-based scaffold was to exploit the rate enhancement observed in the presence of aryl thiol additives [15]. It was envisaged that transthioesterification of C-terminal peptide thioesters with the 2-mercaptobenzyl thiol auxiliaries would generate a highly acylating aryl thioester intermediate capable of facilitating rapid $S \rightarrow N$ acyl shift. Furthermore, substitution of the aromatic ring could be used to modulate the nucleophilicity of the thiol or enhance the lability of the auxiliary [44]. For example, increasing substitution of the aromatic ring with electron-donating functionalities (e.g., methoxy groups) was shown to enhance greatly the



Scheme 7 Peptide ubiquitylation using a photolabile N-terminal ligation auxiliary

acid lability of the auxiliaries, with the 4,5-dimethoxy-2-mercaptobenzyl auxiliary **32** [45, 46] cleaved upon treatment with strongly acidic trifluoromethane sulfonic acid (TFMSA) or bromotrimethylsilane (TMSBr) and the more electron-rich 4,5,6-trimethoxy-2-mercaptobenzyl derivative **33** effectively removed in the presence of TFA [47]. Notably, the comparatively mild conditions for removal of trimethoxybenzyl (Tmb) auxiliary **33** enabled its application in the synthesis of glycopeptides [48], including fragments derived from human erythropoietin (EPO) bearing complex *O*- and *N*-linked glycans [49, 50]. Tmb derivative **33** has likewise been applied to the synthesis of a 62-amino acid SH3 domain of the actin cross-linking protein α-spectrin using a Lys-Gly disconnection site (Scheme 8) [47]. This ligation reaction was complete in 12 h, affording auxiliary-bound peptide **34** in 66% isolated yield. Cleavage of the Tmb auxiliary in the presence of TFA subsequently provided the target SH3 domain **35** [47].

Although *N*-linked thiol ligation auxiliaries have expanded the scope of ligation chemistry to include non-Cys ligation sites, such methods are often plagued by additional limitations. Harsh conditions for the removal of some auxiliaries (particularly strongly acidic conditions such as HF and TFMSA) limit the application of these methodologies to peptides and proteins bearing acid labile functionalities, including post-translational modifications such as glycosidic linkages. Furthermore, the techniques generally exhibit poor sequence tolerance at both the C-terminus of the thioester component and the N-terminus of the auxiliary-bound peptide, with the majority of successful auxiliary-mediated ligations exclusively utilizing junctions where one or both termini are Gly residues [23, 27]. Additional steric bulk at the ligation junction resulting from the tethered auxiliary, together



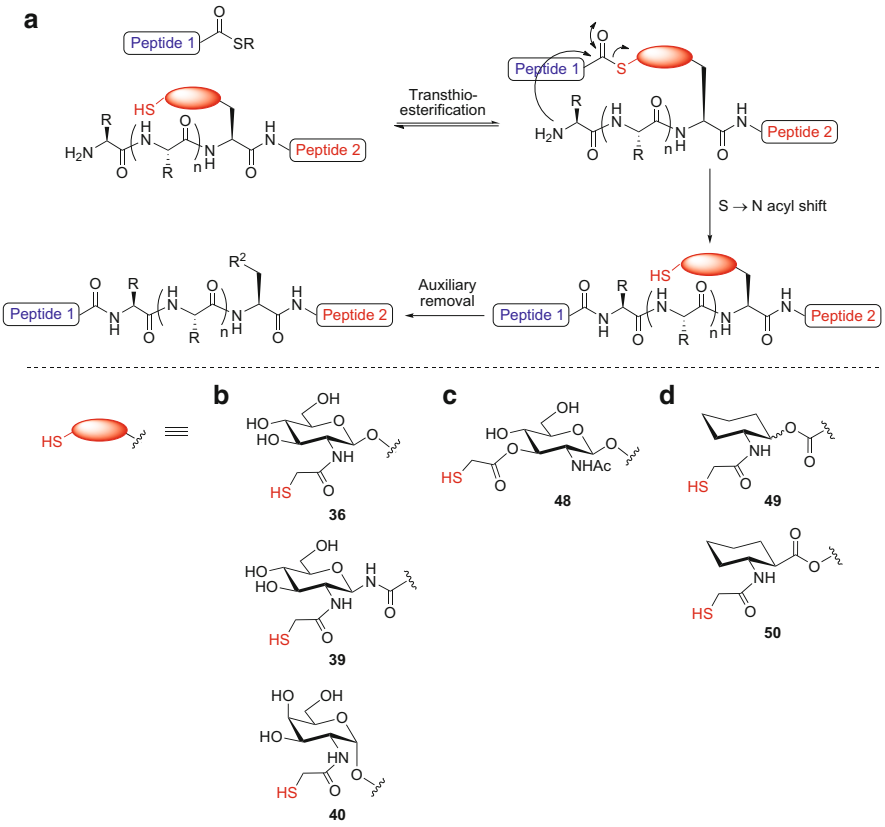
Scheme 8 Synthesis of the SH3 domain of α-spectrin using 4,5,6-trimethoxy-2-mercaptobenzyl ligation auxiliary **33**

with the requirement, in some cases, for larger ring-sized intermediates in the $S \rightarrow N$ acyl migration, greatly decrease the overall rate of ligation and limit general application of these auxiliaries in protein synthesis.

3.1.2 Side-Chain Ligation Auxiliaries

The development of side-chain ligation auxiliaries (Scheme 9) bearing reactive thiol functionalities tethered to the side-chain of an N-terminally located amino acid circumvented many of the issues associated with *N*-linked auxiliaries, including additional steric bulk at the terminal amine moiety [27, 51]. Mechanistically, the side-chain mediated process is very similar to native chemical ligation and *N*-linked auxiliary ligation, consisting of a thiol exchange reaction between a C-terminal peptide thioester and the side-chain thiol auxiliary, followed by an $S \rightarrow N$ acyl shift of the thioester intermediate to generate the new amide bond. A final step is then required to effect auxiliary removal and generate a native peptide product (Scheme 9a).

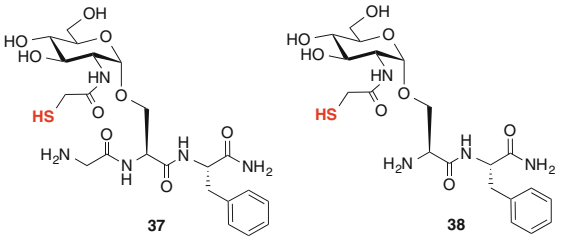
The first example of ligation via side-chain auxiliary, termed sugar-assisted ligation (SAL), was developed by Wong and coworkers in 2006 and involved the use of a glycopeptide whereby the reactive thiol auxiliary was appended to the C-2



Scheme 9 (**a**) Generalized ligation using side-chain auxiliaries. (**b**) First generation sugar-assisted ligation. (**c**) Second generation sugar-assisted ligation. (**d**) Side-chain assisted ligation

position of a β-*O*-linked carbohydrate moiety **36** (Scheme 9b) [52]. Ligation reactions were performed in aqueous buffer and, in contrast to ligation mediated by N-terminal auxiliaries, demonstrated a relatively broad sequence tolerance at the ligation junction [53]. Interestingly, reaction rates increased when the auxiliary-appended glycosylamino acid unit was incorporated as the penultimate residue as in model peptide **37**, rather than the terminal residue of the peptide fragment (compound **38**), despite the reliance on a larger ring-sized intermediate in the $S \rightarrow N$ acyl shift (Scheme 10) [52]. Following amide bond construction, the side-chain thiol functionality could be cleaved using a reductive desulfurization protocol (see below) [54] to yield glycopeptide products bearing a native acetamide at the C-2 position of the carbohydrate. The method was subsequently expanded to include *N*-linked [55] and α-*O*-linked [56] glycans (e.g., **39** and **40**, respectively, Scheme 9b). Enzymatic manipulation of the glycosylated peptide products (with or without the thiol handle present), including removal of the glycan [55] or elaboration of the appended monosaccharide unit through the action of glycosyltransferases [55, 57],

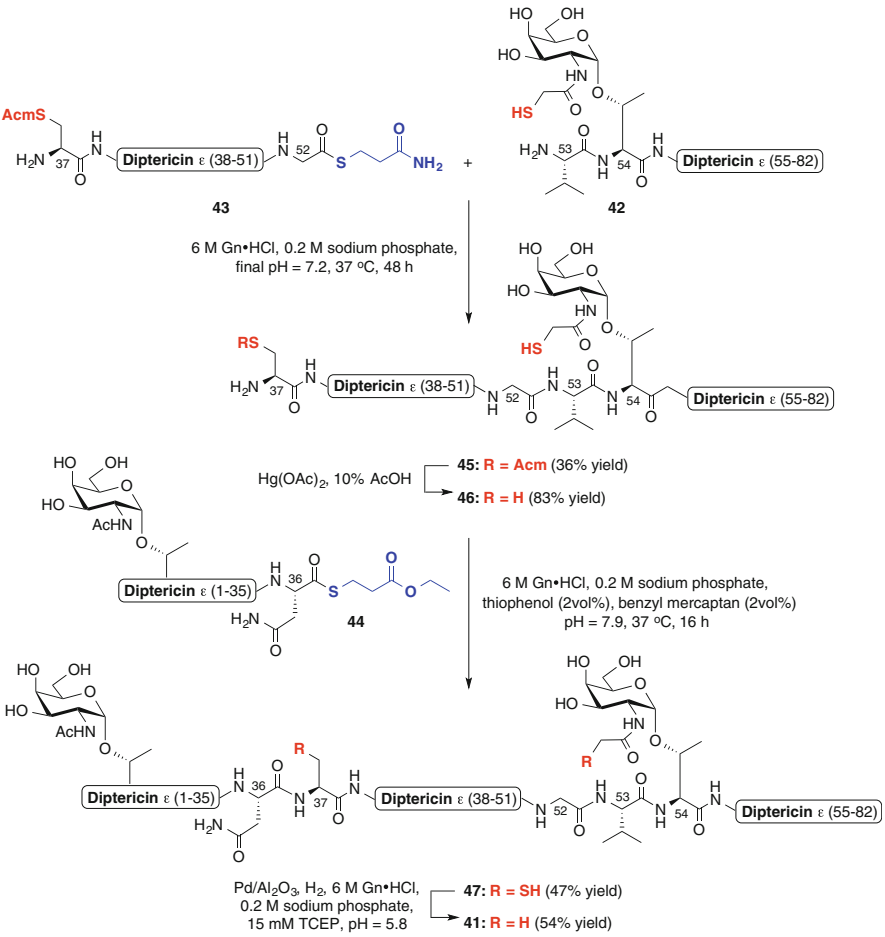
Scheme 10 Glycopeptides originally employed in sugar-assisted ligation (SAL)



further increased the scope and complexity of peptide and glycopeptide targets accessible using SAL.

Notably, the utility of SAL for the construction of glycoprotein targets was confirmed through the total chemical synthesis of diptericin ε **41**, an 82-amino acid, Cys-free antibacterial glycoprotein containing two galactosamine moieties α-*O*-linked to Thr^{10} and Thr^{54} (Scheme 11) [56]. It was envisaged that the glycoprotein could be synthesized from three segments, **42**, **43**, and **44**, in the *C*-to-*N* direction by employing a sequential SAL-native chemical ligation sequence, whereby the former is facilitated by a side-chain glycan auxiliary, and the latter a Cys residue installed as a temporary mutation. To this end, glycopeptide **42**, corresponding to the C-terminal region of diptericin ε and bearing a side-chain α-*O*-linked carbohydrate auxiliary at Thr^{54}, was first ligated to thioester **43** bearing an N-terminal Cys protected as the corresponding Cys acetamidomethyl (Acm) residue [58]. This sugar-assisted ligation was conducted in aqueous denaturing buffer at 37 °C and was complete in 48 h to afford peptide product **45**. Removal of the N-terminal Cys (Acm) group using mercury salts then provided glycopeptide **46**, bearing a free N-terminal thiol at position 37, poised for further functionalization. Accordingly, reaction between **46** and glycopeptide thioester **44** proceeded in 47% isolated yield to provide the 82-residue polypeptide chain **47**. A final global desulfurization reaction using conditions first described by Yan and Dawson [54] (explored in detail in Sect. 4.1) facilitated removal of the glycan-tethered thiol auxiliary as well as cleavage of the Cys side-chain thiol to generate diptericin ε **41**, with the native Ala residue at position 37 [56].

Although SAL in its original form greatly increased the number and flexibility of accessible ligation junctions, a major limitation of the technique was the reliance on a post-ligation reductive desulfurization protocol to facilitate cleavage of the glycan-appended auxiliary. These methods are incompatible with unprotected Cys thiols, and, as such, preclude the use of SAL for synthesis of peptides and proteins bearing native Cys residues. To address this shortfall, a second-generation SAL protocol was developed by Wong and coworkers in 2007, employing a modified auxiliary which could be removed in a mild and selective manner (Scheme 9c) [59]. In particular, this technique employed a thiol auxiliary with an ester linkage to the C-3 hydroxyl group of an *O*-linked glucosamine moiety **48**, which was easily removed via hydrazinolysis following ligation and without affecting unprotected Cys residues in the peptide sequence. Another important modification to the original SAL technique was the use of an organic cosolvent



Scheme 11 Synthesis of diptericin ε using sugar-assisted ligation (SAL)

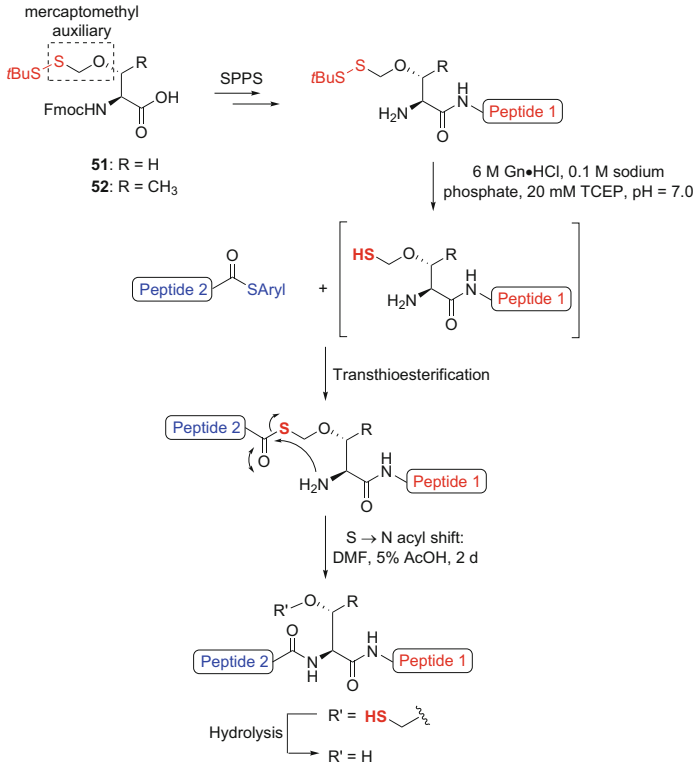
(e.g., *N*-methyl-2-pyrrolidone, NMP) in ligation reactions, serving to reduce hydrolysis of both the thioester component and the ester-linked auxiliary and therefore enhance ligation yields [59].

An additional SAL-inspired ligation strategy, developed in 2008 by Brik and coworkers, utilized a cyclohexyl or cyclopentyl ring auxiliary bound by an ester linkage to the side-chain of aspartic acid (Asp), glutamic acid (Glu), or serine (Ser) (e.g., **49** and **50**, Scheme 9d) [60]. In this method, the simplified side-chain carbocycle served an analogous role to the carbohydrate moiety in SAL. Following ligation, the auxiliary was rapidly cleaved in situ by the addition of NaOH [60]. Application of a Ser-linked cyclohexyl auxiliary using a sequential side-chain assisted ligation-native chemical ligation approach enabled the successful construction of the 86-residue polypeptide backbone of the regulatory protein

HIV-1 Tat [61]. Unfortunately, complications in the removal of the side-chain auxiliary hampered the total synthesis of the native protein in this study.

A remarkable feature of side-chain carbohydrate and cyclohexyl reaction auxiliaries is the ability to promote ligation effectively, despite the reliance of such systems on considerably larger ring size intermediates (14–15-membered) in the $S \rightarrow N$ acyl shift than those required for native chemical ligation or for N-terminal auxiliary-mediated ligation. As previously noted, SAL reactions employing the original carbohydrate bearing a C-2-linked thiol auxiliary proceeded faster with a single amino acid extension N-terminal to the glycosylamino acid moiety than when the glycosylamino acid was N-terminally located (see Scheme 10). These results suggest that the side-chain auxiliary may play an important role in appropriately positioning the intermediate bridged thioester for attack by the N-terminal amine [52]. Detailed studies probing the effect of multiple amino acid extensions N-terminal to the glycosylamino acid auxiliary in SAL [59, 62] demonstrated that ligation is feasible (though substantially slower) with as many as six amino acids appended to the N-terminus of the auxiliary-bound residue. However, it should be noted that ligation reactions bearing such large N-terminal extensions are likely to proceed, at least in part, via a direct aminolysis pathway [63, 64].

Another novel side-chain thiol auxiliary approach was developed in 2010 by Hojo et al. for ligation reactions at N-terminal Ser and Thr residues [65]. This method optimized the use of a mercaptomethyl group attached as a thiohemiacetal to the side-chain of Ser or threonine (Thr) (Scheme 12). As with native chemical ligation, the auxiliary was found to promote ligation through initial transthioesterification with a C-terminal peptide thioester followed by an $S \rightarrow N$ acyl shift, in this instance through a seven-membered ring intermediate, to generate a new amide bond. The inherent instability of the thiohemiacetal functional group was overcome, in part, through incorporation of the auxiliary-bound residue into peptides as the corresponding asymmetric disulfide **51** or **52**, allowing an in situ generation of the free auxiliary upon treatment with TCEP in the ligation reaction (Scheme 12). Furthermore, ligations were performed with preformed aryl thioesters to enhance the rate of the initial transthioesterification reaction. Under these conditions, rearrangement of the thioester intermediate through $S \rightarrow N$ acyl migration proved to be the rate-limiting step. To prevent hydrolysis of the unrearranged intermediate, the aqueous reaction buffer was diluted with DMF containing 5% acetic acid after the initial transthioesterification and left for 2 days to promote rearrangement. Following ligation, the susceptibility of the auxiliary to hydrolysis became a strategic advantage, whereby spontaneous cleavage of the thiohemiacetal afforded native products in a one-pot fashion. The method was utilized in the synthesis of the glycopeptide toxin contulakin-G, derived from the venom of *Conus geographus*, and for the preparation of human calcitonin [65].

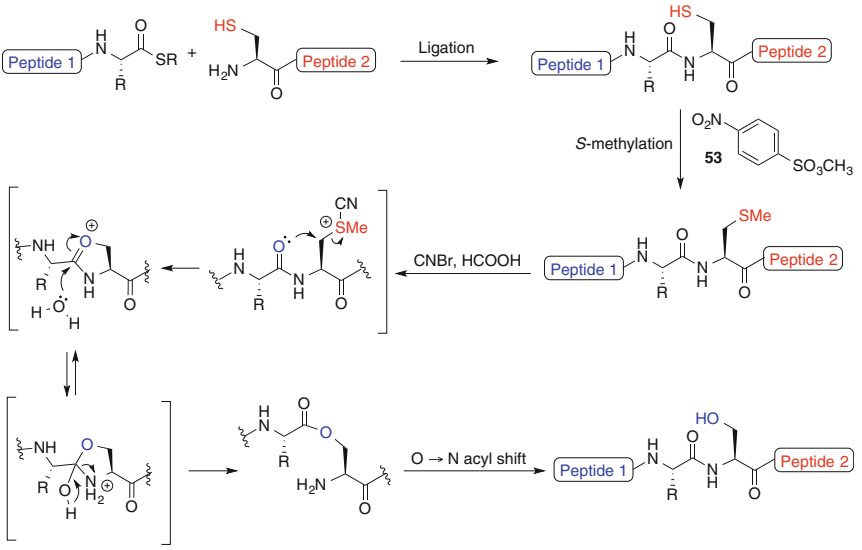


Scheme 12 Ligation at Ser and Thr using a mercaptomethyl side-chain auxiliary

4 Post-Ligation Manipulations

The contributions of N-terminal and side-chain auxiliary-mediated ligations have served to enhance greatly the scope of ligation chemistry beyond the original reliance on N-terminal Cys residues. Despite these successes, auxiliary methods generally require the multi-step preparation of specialized thiol auxiliaries and auxiliary-bound peptides, which reduces the overall efficiency of the techniques. As previously discussed, auxiliary-mediated ligations are also slower than native chemical ligation at Cys, requiring lengthy reaction times whereby hydrolysis and epimerization become significant competing pathways. As such, a conceptually appealing approach to increasing the scope of native chemical ligation without sacrificing the simplicity and efficiency of the technique is to explore the post-ligation modification of Cys residues [54, 66] for the generation of target peptides and proteins.

An innovative demonstration of the manipulation of Cys residues to expand the scope of accessible ligation junctions was reported in 2008 by Okamoto and Kajihara [67]. With the aim of preparing complex glycopeptides and proteins, the

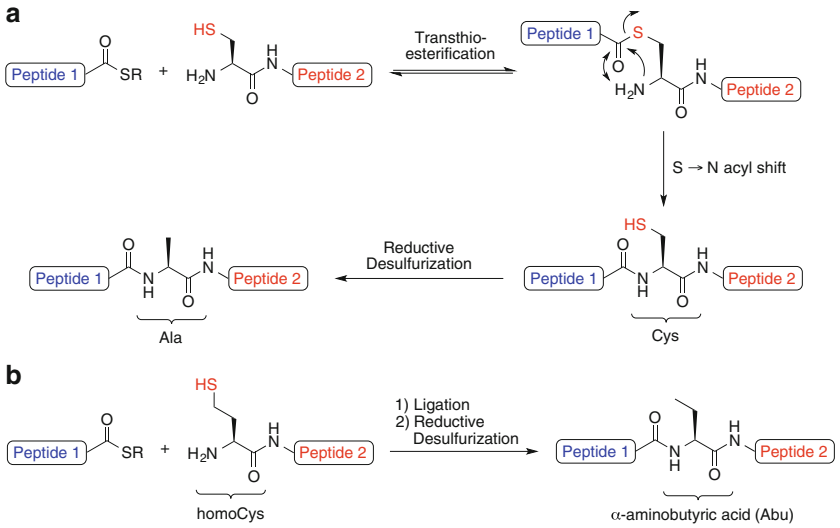


Scheme 13 Conversion of Cys to Ser following native chemical ligation

authors demonstrated a post-ligation conversion of Cys to the corresponding Ser residue using a three-step protocol (Scheme 13). The method first involved *S*-methylation of the Cys residue using methyl 4-nitrobenzenesulfonate **53** [68]. Subsequent reaction with CNBr in the presence of formic acid leads to the activation of the *S*-methyl group [69], facilitating attack of the *S*-methyl Cys β-carbon by the carbonyl oxygen of the neighboring amide bond. The resultant five-membered ring is converted to the *O*-ester peptide intermediate, which undergoes a subsequent O → N acyl shift at slightly basic pH (7–8) to generate the new amide bond. Initial application of this protocol on model systems demonstrated the overall feasibility of the approach. The potential for reaction of CNBr with Met residues, however, required the incorporation of Met as the corresponding sulfoxide so as to allow selectivity for the methyl Cys residues. Reduction of the sulfoxide to yield a native Met residue was accomplished following the conversion of Cys to Ser. The utility of the Cys to Ser transformation in the synthesis of complex glycopeptide fragments was also demonstrated through the construction of a glycopeptide fragment of erythropoietin, bearing a complex, *N*-linked glycan, along with the synthesis of a MUC1 40-mer peptide containing two copies of the T_N antigen [67].

4.1 Ligation-Desulfurization

In 2001, Yan and Dawson reported an elegant approach for the disconnection of peptide and protein targets at alanine (Ala) residues [54]. The concept combined a typical native chemical ligation reaction between a peptide bearing an N-terminal



Scheme 14 (**a**) Ligation at Cys and post-ligation reductive desulfurization to Ala. (**b**) Ligation at homoCys followed by reductive desulfurization to α-aminobutyric acid (Abu)

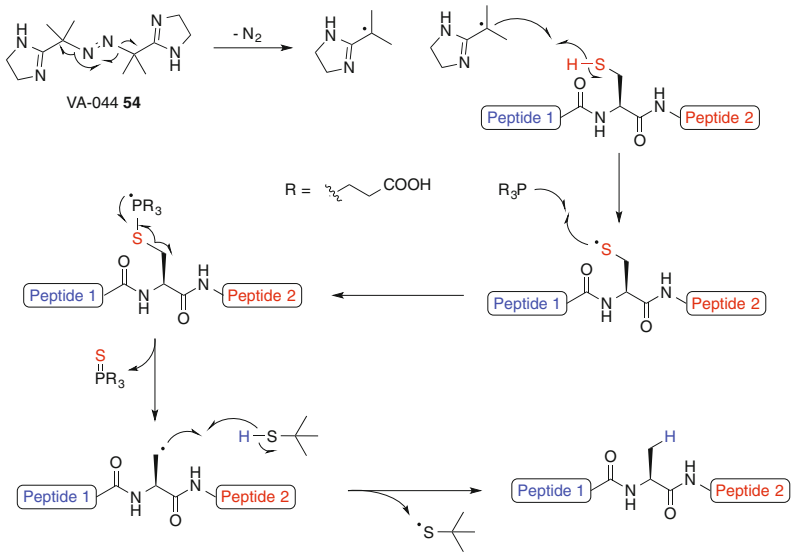
Cys residue and a C-terminal peptide thioester with a post-ligation, reductive desulfurization protocol which selectively removed the sulfhydryl moiety in the Cys side-chain to generate the corresponding Ala residue at the ligation junction (Scheme 14a). The technique combined the efficiency of Cys-promoted ligations while enabling access to ligation junctions containing the significantly more abundant Ala residue (8.9%) [33], thereby enabling the synthesis of a variety of previously inaccessible peptide and protein targets. To optimize the desulfurization protocol, the authors treated Cys-containing ligation products with a variety of metal reagents, including Pd on Al_2O_3, Pd on carbon, PdO and Raney Ni, in the presence of hydrogen gas. Excellent results were obtained with Pd on Al_2O_3, which effected high-yielding, global desulfurization of Cys residues while minimizing the over-reduction of the aromatic proteinogenic amino acids tyrosine (Tyr), phenylalanine (Phe), and tryptophan (Trp). Rapid and efficient desulfurization with Raney Ni was also demonstrated, although demethylthiolation of Met residues was evident with prolonged reaction times [54].

In the original publication, Yan and Dawson also demonstrated the post-ligation reduction of a homocysteine (homoCys) residue to the corresponding α-aminobutyric acid (Abu) (Scheme 14b) [54]. This application built upon earlier work by Tam and Yu [70], which demonstrated that ligation at homoCys followed by *S*-methylation provided a feasible approach to ligation at Met residues. The extension of the desulfurization protocol to include non-Cys thiols established the generality of the technique and led to the prescient notion of utilizing unnatural, thiol-derived amino acid derivatives to effect ligation (see Sect. 4.3 for more details) [54]. Importantly, the authors also demonstrated the utility of the Cys

ligation-desulfurization strategy in diverse peptide and protein systems through the syntheses of a cyclic microcin J25-like peptide, the 56-amino acid streptococcal protein G B1 (PGB1) domain, and [Ala^{49}]barnase, prepared by desulfurization of the previously reported 110-amino acid [Cys^{49}]barnase analogue [15].

Following this seminal report, research efforts have shifted to improving the selectivity of the desulfurization protocol to preclude the desulfurization of Cys residues (and other sulfur-containing functionalities) crucial to the native peptide sequence. Pentelute and Kent reported the desulfurization of unprotected Cys residues with Raney Ni in the presence of acetamidomethyl (Acm)-protected Cys residues, which could be unmasked following desulfurization to generate the free sulfhydryl side-chain of Cys [58]. In their work on sugar-assisted ligation (SAL), Wong and coworkers independently reported that the use of hydrogen gas and Pd/Al_2O_3 could effect selective removal of a glycan-bound thiol auxiliary in the presence of a Cys(Acm) residue [56].

However, broader issues with the use of metal-based reagents, including the adsorption of specific peptide sequences to the metal surface [71] and the undesired reduction of Met and thiazolidine-protected Cys residues [72], prompted the search for a milder, metal-free desulfurization protocol. To this end, Danishefsky and coworkers turned to a radical-promoted desulfurization method [72] based on an initial report by Hoffman et al. 50 years earlier on the desulfurization of thiols, under thermal or photochemical conditions, in the presence of trialkylphosphites [73]. The subsequent use of trialkylphosphines to promote desulfurization was also reported [74]. With the goal of mild and selective desulfurization of peptide systems in mind, Danishefsky and coworkers specifically employed the water-soluble phosphine TCEP, owing to its stability, ease of handling, and proven compatibility with proteinogenic amino acid side-chains and glycopeptide functionalities. Indeed, in the presence of TCEP, the water-soluble radical initiator 2,2′-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) **54**, and *t*BuSH in aqueous media, the selective desulfurization of Cys residues was readily achieved. The mechanism for this transformation presumably mirrors the one proposed by Walling et al. [75, 74] and involves the initial formation of a Cys thiyl radical in the presence of radical initiator **54** (Scheme 15). Addition of the sulfur-centered radical to TCEP generates a TCEP-adduct, which undergoes β-scission to produce an alanyl radical and a phosphine sulfide byproduct. Hydrogen abstraction from an exogenous thiol by the alanyl radical then generates the native Ala residue. Importantly, these conditions were shown to tolerate Cys(Acm) groups, Met residues, thiazolidine groups, and C-terminal thioester moieties. In the initial report, Danishefsky and coworkers employ the metal-free desulfurization protocol in a ligation-desulfurization approach to an *N*-linked glycopeptide fragment and the cyclic nonapeptide crotogossamide [72].

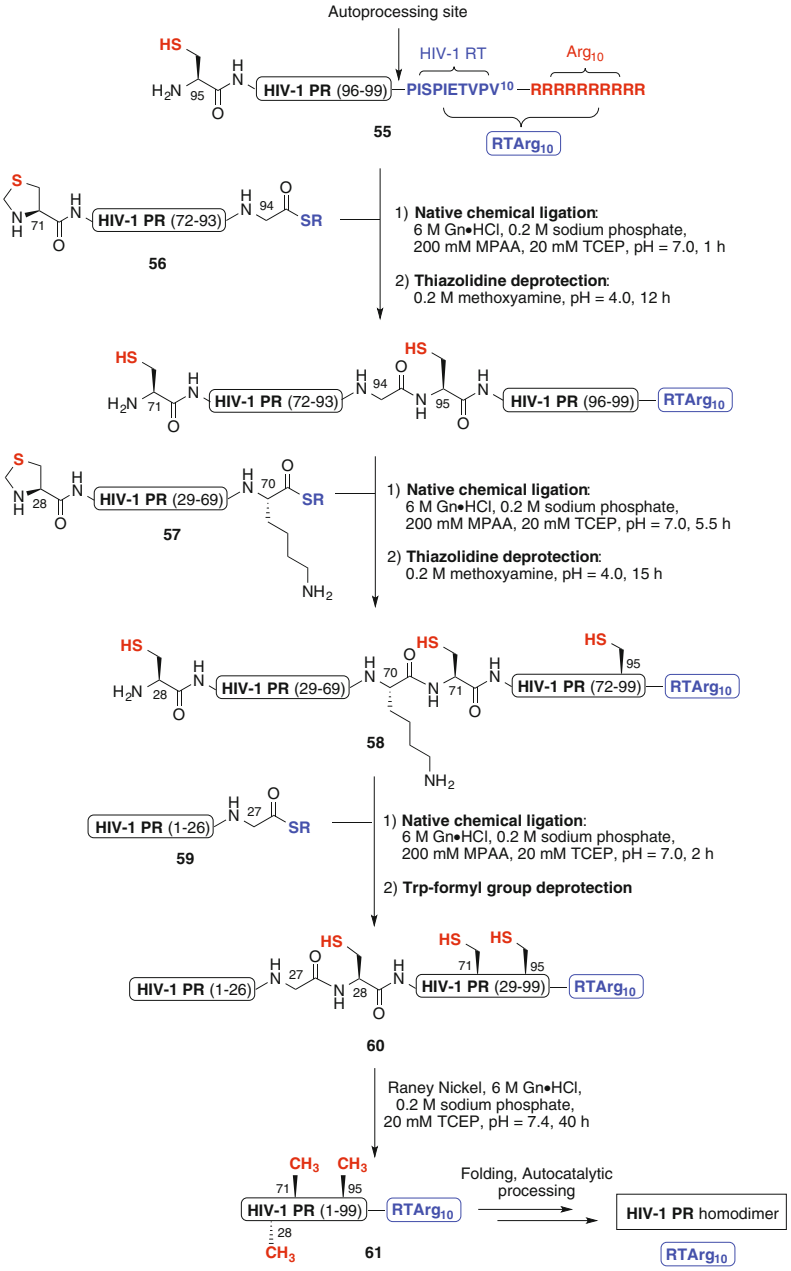


Scheme 15 Radical desulfurization of Cys residues initiated by VA-044 **54**

4.2 Ligation-Desulfurization in Protein Synthesis

Since its inception, the concept of ligation-desulfurization chemistry [76, 77] at Cys residues has been widely adopted for the synthesis of an enormous variety of peptides and proteins, including targets bearing complex PTMs [20, 21, 28, 78]. The following discussion outlines a number of recent, illustrative examples of the technique for the ligation-based assembly of protein targets.

In 2007, Kent and coworkers reported an elegant synthesis of a fully functional, homodimeric HIV-1 protease (PR) by combining ligation-desulfurization chemistry with a biomimetic autoprocessing strategy [79], taking advantage of the ability of the HIV-1 protease to catalyze its own removal from the Gag-Pol polyprotein precursor during HIV-1 maturation in vivo. In their initial studies toward the total chemical synthesis of the 99-amino acid HIV-1 PR monomer using ligation chemistry, the authors encountered considerable difficulties in solubilizing peptide intermediates, particularly the C-terminal fragment. A revised strategy therefore incorporated a C-terminal poly-Arg tag to aid in the solubility of the terminal fragment and subsequent ligation intermediates (Scheme 16). In order to facilitate removal of the solubility tag following construction of the monomeric polypeptide backbone, the poly-Arg sequence was attached to the C-terminus of the protease using a ten-residue linker sequence derived from the HIV-1 reverse transcriptase (RT) protein (which is proximally located in the Gag-Pol polyprotein precursor). The C-terminal construct **55** would therefore contain an autoprocessing site, and it was envisaged that the folded protease would readily cleave the modification.

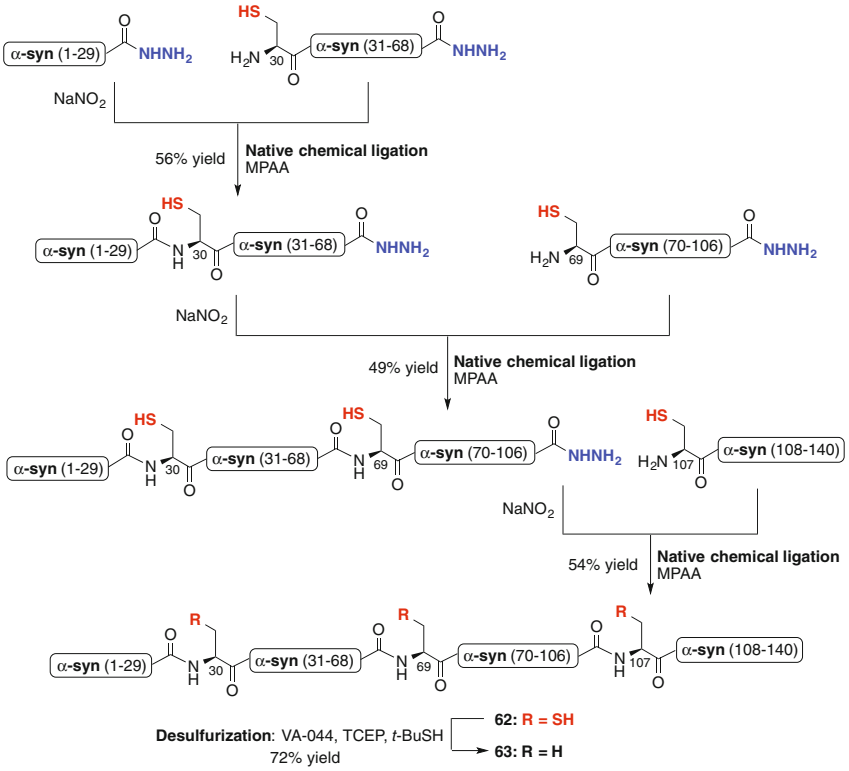


Scheme 16 Total synthesis of HIV-1 protease using an iterative ligation-autoprocessing strategy

Accordingly, construction of the full-length peptide was accomplished using a four-component iterative ligation-deprotection strategy in the *C*-to-*N* direction, whereby bifunctional thioesters **56** and **57** were prepared as N-terminal thiazolidines which could be easily unmasked following a ligation reaction to facilitate extension in the N-terminal direction (Scheme 16). As such, iterative ligation-thiazolidine deprotections employing peptide fragments **55**, **56**, and **57** rapidly afforded 28-99RTArg$_{10}$ **58**. After ligation of N-terminal thioester **59** to 28-99RTArg$_{10}$ **58** and removal of Trp formyl protecting groups, the 119-residue polypeptide **60** was globally desulfurized in the presence of Raney Ni to afford 1-99RTArg$_{10}$ **61** in 26% overall yield based on starting peptide **55**. As anticipated, autoprocessing of the C-terminal modification occurred concomitantly with folding of the protease to generate the target homodimeric protein (Scheme 16). The biological activity of the synthetic enzyme was further confirmed in a kinetic assay and the structure validated using X-ray crystallography [79].

Liu and coworkers recently reported the total chemical synthesis of α-synuclein, a Cys-free protein implicated in the development of Parkinson's disease, in another impressive application of ligation-desulfurization chemistry to the total chemical synthesis of proteins [80]. Herein, the authors utilize a four-component *N*-to-*C* ligation approach which takes advantage of C-terminal peptide acyl hydrazides, a technology pioneered in the Liu laboratory, as thioester surrogates for ligation chemistry [81, 82]. Briefly, the method first involves the activation of fully deprotected C-terminal peptide hydrazides with the oxidant $NaNO_2$, which chemoselectively affords an intermediate acyl azide. Thiolysis of the acyl azide using an aryl thiol (e.g., MPAA) then promotes the in situ formation of the peptide thioester, which is poised for use in ligation chemistry. Importantly, peptide hydrazides are easily prepared using Fmoc-SPPS and are able to serve as masked thioesters in iterative ligation strategies because of the requirement for an initial activation step (oxidation to the acyl azide). As such, the reactivity of bifunctional peptides containing a C-terminal hydrazide and an N-terminal Cys residue are carefully controlled to promote protein synthesis using iterative ligation chemistry. The ability of hydrazides to serve as latent thioesters was strategically exploited in the *N*-to-*C* construction of α-synuclein using an iterative oxidation-ligation approach (Scheme 17). To facilitate ligation, Cys mutants were temporarily introduced at residues 30, 69, and 107 in place of the native Ala residues. Following the iterative ligation sequence and construction of the full-length α-synuclein (1–140, A^{30,69,107}C) **62**, high-yielding conversion to the native protein **63** was accomplished using a global radical desulfurization protocol employing the conditions developed by Wan and Danishefsky [72].

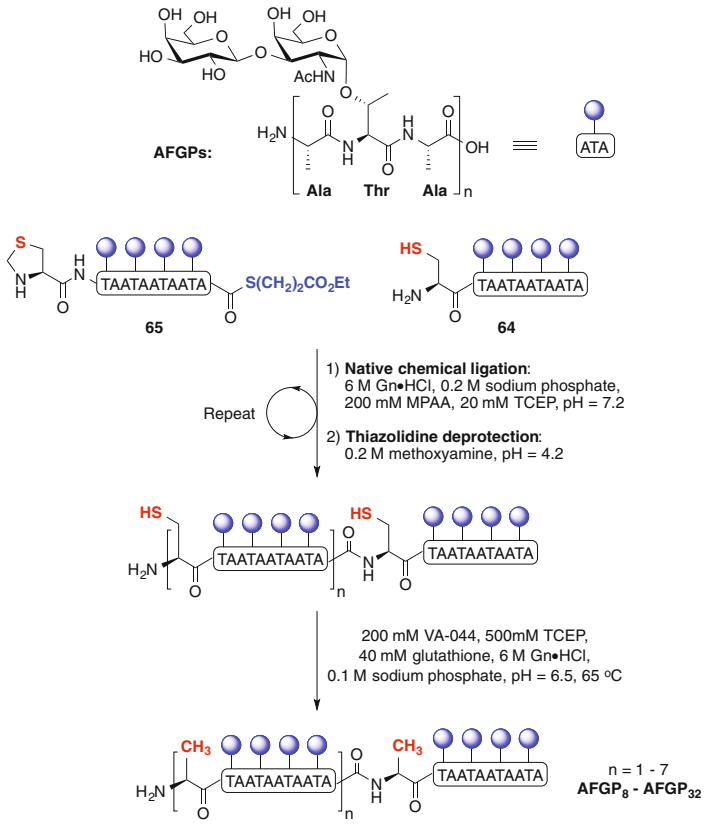
Post-ligation desulfurization has also been extensively applied to the synthesis of proteins bearing post-translational modifications. In 2012, Wilkinson et al. reported the construction of a library of homogeneous antifreeze glycoproteins (AFGPs) using ligation-desulfurization chemistry [83]. AFGPs are mucin-type glycoprotein natural products isolated from select Arctic and Antarctic fish, where they play a critical role in preventing the growth of ice crystals. Structurally, AFGPs are composed of multiple copies of the repeating tripeptide Ala-Thr-Ala/Pro, in



Scheme 17 Total synthesis of α-synuclein through an iterative ligation-desulfurization strategy employing peptide acyl hydrazides

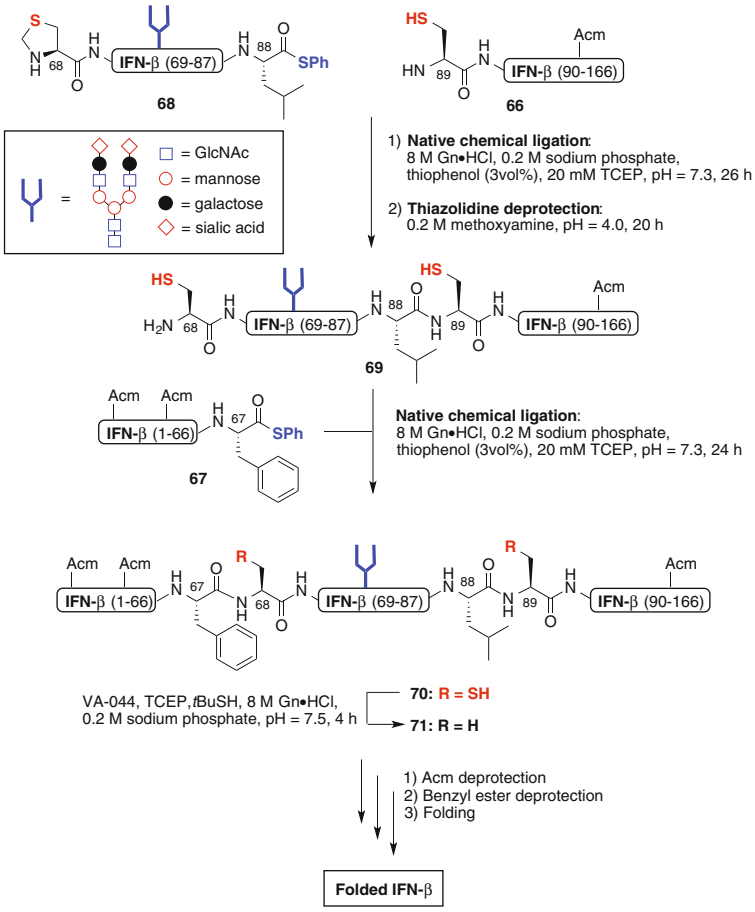
which each Thr residue is α-*O*-linked to the disaccharide β-D-galactosyl-(1→3)-α-*N*-acetyl-D-galactosamine (Scheme 18) and range in size from 4 to 50 repeat units [84, 85]. In an effort to access large quantities of homogeneous AFGPs for biological studies and applications in materials science, the authors designed a convergent ligation approach to homogeneous AFGPs bearing between 4 and 32 repeat units. Specifically, peptide fragment **64**, bearing an N-terminal Cys residue, and bifunctional peptide **65**, bearing an N-terminal thiazolidine and a C-terminal thioester, were used in iterative ligation chemistry to assemble increasingly large AFGP repeat units (Scheme 18). Upon reaching the desired chain length, the Cys ligation handles were readily converted to the native Ala residues via radical desulfurization [72] in the presence of VA-044, TCEP and glutathione [71] as a source of hydrogen atoms. The resulting library of homogeneous AFGPs (ranging in size from 1.2 to 19.5 kDa) enabled a comprehensive study of the effect of chain length on thermal hysteresis and ice recrystallization inhibition activities [83].

Kajihara and coworkers recently reported the total chemical synthesis of two glycoforms of the 166-amino acid human glycosyl-interferon-β, bearing a complex



Scheme 18 Construction of homogeneous antifreeze glycoproteins (AFGPs) using ligation-desulfurization chemistry

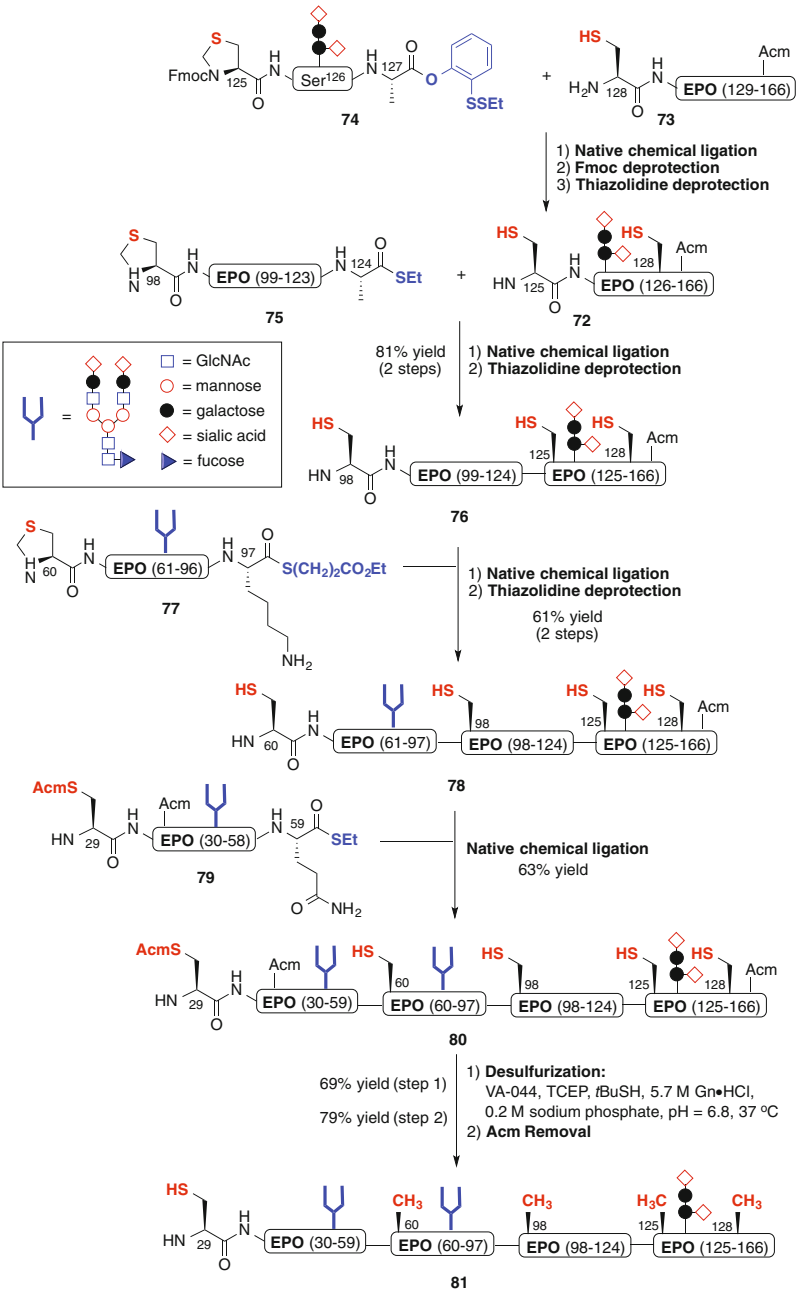
N-linked sialyl or asialo biantennary oligosaccharide [86]. Using a three-component synthetic strategy, the target was disconnected at two Ala ligation junctions, accessible via ligation-desulfurization chemistry (Scheme 19). Notably, interferon-β also contains three native Cys residues in the full-length sequence. However, the location of these residues was deemed unsuitable for the facile construction of the target protein. As such, the native Cys residues were incorporated as the corresponding Cys(Acm) residues in the target peptide fragments **66** and **67** to facilitate construction of the protein using ligation-desulfurization chemistry (Scheme 19). Initial ligation of peptide **66** and glycopeptide thioester **68**, bearing an N-terminal thiazolidine, was accomplished under standard aqueous conditions in the presence of thiophenol as a ligation catalyst. Removal of the thiazolidine afforded **69**, which was subsequently ligated with fragment **67** to afford the 166-amino acid polypeptide **70**. At this stage, radical desulfurization enabled conversion of the two ligation site Cys residues to the native Ala residues, affording glycopeptide **71**. Removal of the Cys(Acm) groups upon treatment with AgOAc in 90% AcOH and saponification of the benzyl ester protecting group (in the synthesis



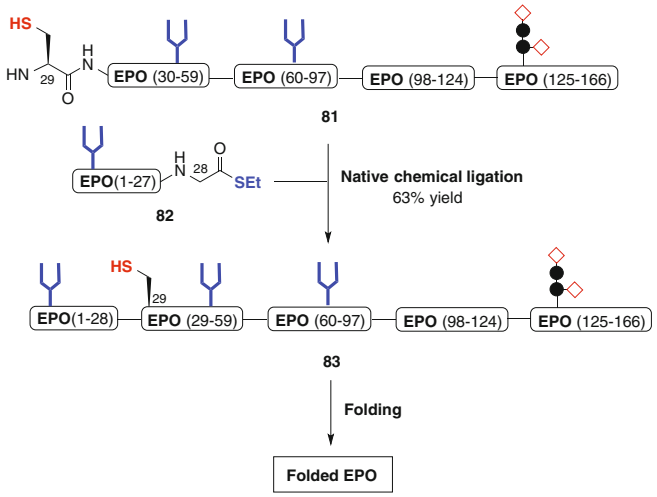
Scheme 19 Construction of homogeneous human glycosyl interferon-β (IFN-β) using ligation-desulfurization chemistry

of the sialylated glycoprotein variant) afforded the native protein glycoforms. Upon folding, both glycosylated variants of interferon-β were shown to suppress tumor growth in vivo [86].

Perhaps one of the most demonstrative applications to date of the power of ligation-desulfurization chemistry has been in the total chemical syntheses of the human glycoprotein hormone, erythropoietin (EPO) [87–90]. The most recently reported example described the first synthesis of the 166-amino acid glycoprotein as single glycoforms bearing natively linked glycans [87, 88]. In this groundbreaking work [91], the Danishefsky laboratory accessed multiple Ala ligation sites by employing a post-ligation metal-free radical desulfurization protocol [72] together with judicious protection of the native Cys residues, including those inappropriately positioned for ligation chemistry. A summary of the ligation strategy is shown in Schemes 20 and 21 [88].

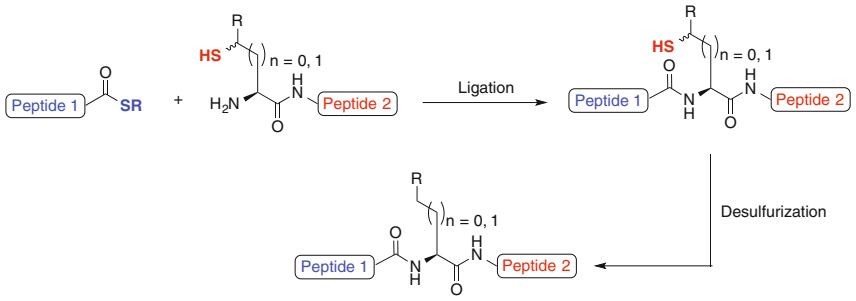


Scheme 20 Synthesis of EPO(29–166) **81** using ligation-desulfurization chemistry



Scheme 21 Total synthesis of homogenous EPO(1–166) as a single glycoform

To begin, C-terminal fragment EPO(125–166) **72**, bearing a Ser-linked glycophorin moiety, was first prepared via ligation of fragment **73** with the short glycopeptide **74** bearing a latent thioester moiety [92, 93] followed by an Fmoc deprotection and unmasking of the N-terminal thiazolidine (Scheme 20). Ligation of **72** with thioester EPO(98–124) **75** and subsequent thiazolidine deprotection then provided peptide **76** in 81% yield over the two steps. Glycopeptide thioester **77**, corresponding to EPO(60–97) and bearing a complex *N*-linked sialyl biantennary glycan, was next ligated to peptide fragment **76**. Another thiazolidine deprotection, followed by ligation with glycopeptide thioester EPO(29–59) **79**, afforded peptide **80** corresponding to residues 29–166 of the protein sequence. At this point, radical desulfurization of the four ligation-site Cys residues was accomplished in 69% yield to afford the corresponding Ala residues (Scheme 20). Removal of the Acm protecting groups then unmasked the native Cys residues, including N-terminal Cys^{29} to afford **81**, which was appropriately positioned for a subsequent ligation with glycopeptide thioester EPO(1–28) **82** to afford the full-length, native polypeptide sequence **83** (Scheme 21). Importantly, upon folding, chemically-derived EPO displayed potent erythropoietic activity in both in vitro and in vivo assays [88]. The size and complexity of the target EPO glycoform pushed the limits of modern protein synthesis and thus served as a potent validation of the utility of ligation-desulfurization chemistry for the construction of post-translationally modified proteins.



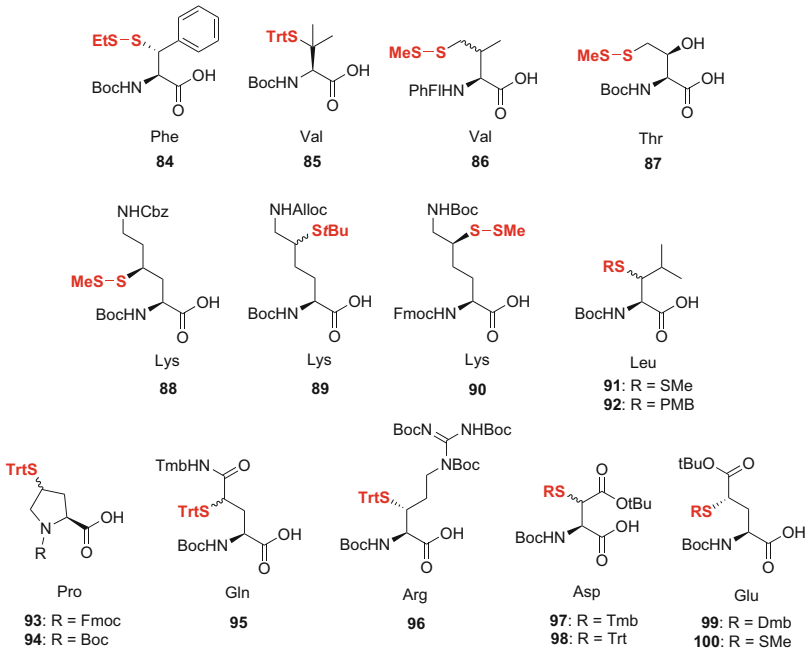
Scheme 22 Ligation-desulfurization chemistry at β- ($n = 0$) or γ-thiol ($n = 1$) amino acids

4.3 Ligation-Desulfurization at Thiol-Derived Amino Acids

The impact of Yan and Dawson's seminal work [54] on post-ligation desulfurization has extended far beyond access to Ala ligation junctions. As previously discussed in their seminal report, the authors also established the intellectual framework for ligation at a variety of additional non-Cys sites by demonstrating ligation-desulfurization at homoCys residues. These results prompted the proposal that unnatural, β- or γ-thiol amino acid derivatives could be utilized in a similar manner to enable disconnections at other proteinogenic amino acids (Scheme 22) [54]. This idea has fuelled an intense focus on the development of unnatural, thiol-derived amino acids for use in ligation reactions (Schemes 22 and 23) [76, 77]. A concise overview of the synthetic strategies employed in the construction of these valuable building blocks has recently been reported [94]. As such, the following discussion serves to outline the application of these novel building blocks (e.g., **84**–**100**, Scheme 23) in ligation-desulfurization chemistry for the synthesis of peptides and proteins.

4.3.1 Phenylalanine

The first application of post-ligation desulfurization beyond Cys to Ala conversions was in the demonstration of ligation disconnections at Phe residues [95, 96]. In 2007, Crich and Banerjee reported the synthesis of β-thiol derivative **84** (Scheme 23), beginning with L-Phe methyl ester. The synthetic pathway utilized chemistry originally developed for the bromination of the benzylic position of aromatic amino acid residues and subsequent conversion to the β-hydroxy analogues [97–99]. Synthetic Phe building block **84** was first shown to facilitate ligation reactions with amino acid thioesters. Following incorporation into a model peptide **101** using Fmoc-SPPS, the building block was successfully used to mediate ligation with peptide thioesters **102** and **103**, bearing C-terminal Gly and Met residues, respectively, in good yields (Scheme 24). Removal of the β-thiol



Scheme 23 β- and γ-thiol amino acid derivatives for ligation-desulfurization chemistry

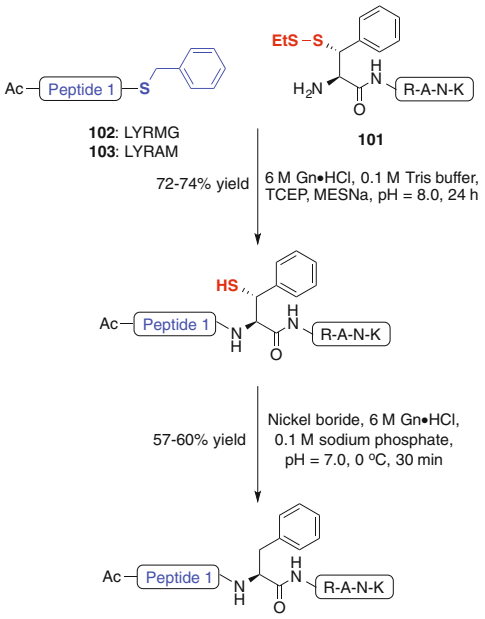
moiety following ligation was achieved via hydrogenolytic desulfurization with nickel boride, thereby generating the native Phe residue at the ligation junction [96]. The results of this initial study served as a critical proof-of-concept for the development of subsequent thiol-derived proteinogenic amino acids.

4.3.2 Valine

Shortly after the report of ligation at Phe building block **84**, Seitz and coworkers reported the first strategy for ligation at Val by employing a commercially available penicillamine building block (Boc-Pen(Trt)-OH, **85**, Scheme 23) [71]. Following incorporation into peptides, penicillamine-mediated ligations were successfully demonstrated for peptide thioesters bearing C-terminal Gly, His, Met, and Leu residues, proceeding in 12–48 h. The relatively lengthy reaction times were attributed to the additional steric bulk associated with the use of a tertiary thiol. Following ligation, application of a slight variation on Wan and Danishefsky's metal-free radical desulfurization protocol [72], using glutathione (rather than *t*BuSH) as the hydrogen atom donor, afforded native peptide products in excellent yields [71].

Danishefsky and coworkers independently reported access to Val ligation junctions via a synthetic γ-thiol Val building block **86** [100]. Using a ten-step

Scheme 24 Ligation-desulfurization at β-thiol phenylalanine



synthesis beginning with Fmoc-Asp-O*t*Bu, the authors were able to access both diastereomers of **86**. Following incorporation into model peptides, comparative rate studies revealed that ligations mediated by both diastereomers of γ-thiol Val proceeded significantly faster than the corresponding reactions mediated by penicillamine (β-thiol Val), despite the requirement for a six-membered ring intermediate in the $S \rightarrow N$ acyl shift for the γ-thiol variants. The enhanced rate of ligation relative to penicillamine is likely owing to the greater reactivity and decreased steric bulk of the primary thiol of building block **86** relative to the tertiary thiol of **85** (Scheme 23). Ligation at γ-thiol Val was also utilized in the high-yielding synthesis of a glycopeptide bearing an *N*-linked disaccharide through ligation of a peptide bearing an N-terminal γ-thiol Val residue with a glycopeptide *ortho*-thiophenolic ester [100].

4.3.3 Threonine

The Danishefsky laboratory also reported a ten-step synthetic approach to γ-thiol Thr **87** from H-Met-OMe (Scheme 23), expanding the repertoire of ligation-desulfurization chemistry to include Thr ligation sites [101]. Following incorporation of the building block into peptides, its utility in ligation-desulfurization chemistry was demonstrated in a number of diverse model systems. Notably, the γ-thiol Thr residue was capable of facilitating ligation with a variety of C-terminal acyl donors, including at sterically encumbered C-terminal Val, Ile, and Pro residues, through activation as the corresponding *p*-nitrophenyl ester acyl donors.

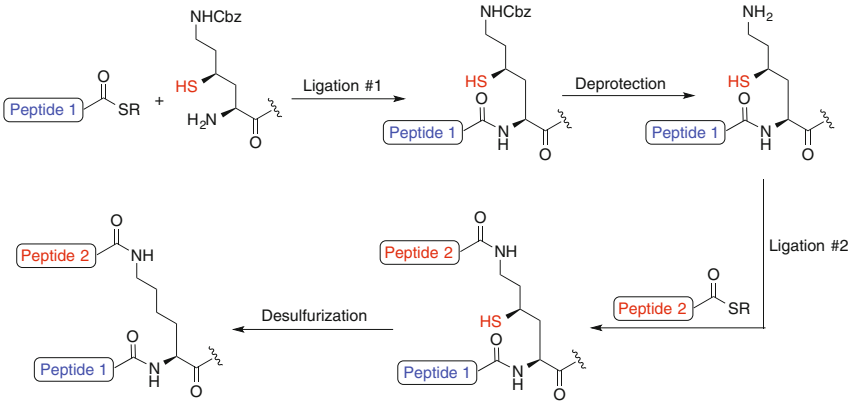
Application of the Thr ligation strategy followed by radical desulfurization also enabled the synthesis of a glycopeptide bearing a complex *N*-linked hexasaccharide moiety [101].

4.3.4 Lysine

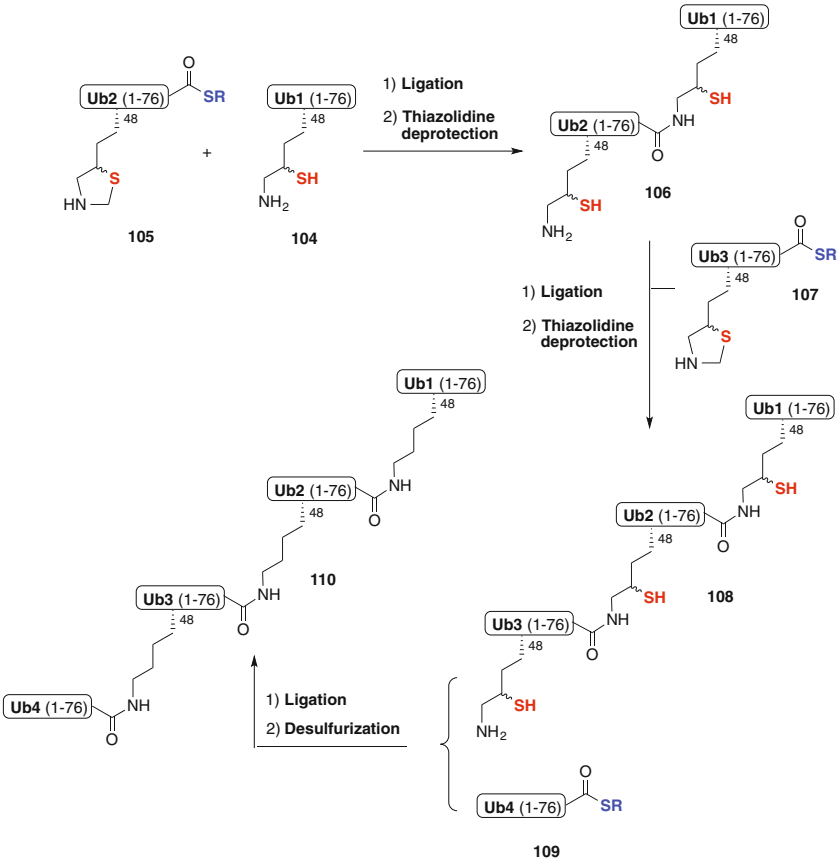
In 2009, Liu and coworkers published the synthesis of a γ-thiol Lys derivative **88** which was capable of mediating ligation at both the α- and ε-amino groups of Lys (Scheme 25), each via a six-membered ring intermediate in the $S \rightarrow N$ acyl shift [102]. Specifically, a side-chain Cbz-protected γ-thiol Lys derivative was first ligated at the α-amino group. Deprotection of the Cbz group unmasked the ε-amino group for a subsequent ligation, and a final desulfurization protocol rendered a native Lys residue, derivatized at both the α- and ε-positions (Scheme 25). This dual ligation protocol is particularly attractive for the synthesis of post-translationally modified peptides and proteins given the variety of functionalization occurring naturally at the ε-amino moiety of Lys, including acetylation, ubiquitylation, and methylation. Access to the key γ-thiol Lys building block **88** (Scheme 23) was accomplished in 16 steps from Fmoc-Asp-O*t*Bu by first employing the method of Guichard and coworkers [103] to access a 4-hydroxy-Lys derivative bearing a side-chain Cbz protecting group. Lys building block **88** was shown to effectively mediate a number of ligations at both the α- and ε-amino groups, and was used in the preparation of side-chain ubiquitylated and biotinylated peptide products [102].

Two δ-thiol Lys derivatives, **89** [104] and **90** [105], have been independently reported for use in side-chain ligation approaches to ubiquitylation. In particular, δ-thiol Lys building block **89** has been extensively employed by Brik and coworkers to study the ubiquitylation of α-synuclein [104, 106], a 140-amino acid presynaptic protein implicated in a number of neurodegenerative diseases, through protein semi-synthesis. Ovaa and coworkers have also reported the incorporation of Lys building block **90** in place of native Lys residues in the 76-amino acid ubiquitin protein sequence to facilitate the synthesis of a library of diubiquitin conjugates using ligation-desulfurization chemistry [105]. The same group recently reported a concise synthetic route to an additional γ-thiol Lys derivative and determined that both γ- and δ-thiol Lys are equally efficient in facilitating the synthesis of diubiquitin conjugates [107].

Another impressive display of the applicability of thiol-derived amino acid building blocks in the synthesis of large proteins through ligation-desulfurization chemistry was the total chemical synthesis of a 304-amino acid tetraubiquitin protein by Brik and coworkers in 2011 [108]. The polyubiquitin chain was constructed using an iterative ligation approach facilitated by a δ-thiol Lys residue positioned at Lys^{48} in the ubiquitin chain (Scheme 26). Ligation of Ub1 fragment **104**, bearing a δ-thiol Lys residue, with Ub2 thioester **105**, containing a thiazolidine protected δ-thiol Lys residue, was accomplished using standard ligation conditions in the presence of exogenous benzylmercaptan and thiophenol. Removal of the



Scheme 25 Dual native chemical ligation at Lys mediated by a γ-thiol Lys derivative



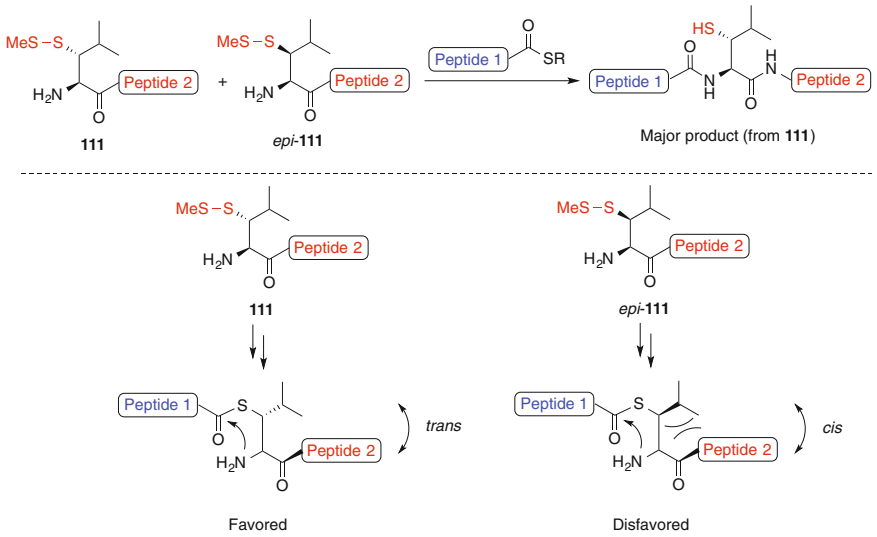
Scheme 26 Synthesis of a 304-amino acid tetraubiquitin protein using ligation-desulfurization at δ-thiol Lys

thiazolidine positioned diubiquitin **106** for subsequent reaction with another Ub thioester, Ub3 **107**. Thiazolidine deprotection and a final ligation of triubiquitin **108** with ubiquitin thioester Ub4 **109**, bearing a native Lys residue at position 48, then afforded a tetraubiquitin adduct containing three unnatural δ-thiol Lys residues. A final conversion of the unnatural amino acid derivatives to the native Lys residues using a global radical desulfurization protocol furnished the 304-amino acid tetraubiquitin **110** [108], currently the second largest protein prepared to date by total chemical synthesis.

4.3.5 Leucine

Ligation at Leu has been independently demonstrated by Danishefsky and coworkers [109] and Brik and coworkers [110], each developing a seven-step synthetic approach to suitably protected Leu building blocks, **91** and **92** (Scheme 23), respectively, beginning with commercially available β-hydroxy-L-Leu. Danishefsky and coworkers prepared both β-epimers of the target Leu building block **91** by beginning the synthesis with both *erythro*- and *threo*-β-hydroxy-L-Leu precursors [109]. In contrast, Brik and coworkers reported the synthesis of a single diastereomer of β-thiol Leu **92**, beginning exclusively with *threo*-β-hydroxy-L-Leu [110]. Following the preparation of the target building blocks, both groups demonstrated the utility of the Leu derivatives in ligation-desulfurization chemistry. Interestingly, in a competition experiment, Danishefsky and coworkers were able to show that diastereomeric β-thiol Leu derivatives reacted at substantially different rates, with peptide **111** reacting approximately 20 times faster than the β-epimer *epi*-**111** (Scheme 27). This selectivity was thought to be a result of a *trans* relationship between the β-isopropyl group and the peptide chain imposed by the putative five-membered ring transition state in the $S \rightarrow N$ acyl shift in peptide **111**, while a corresponding *cis* relationship would predominate for *epi*-**111** (Scheme 27). The authors also postulate that the rate of initial transthioesterification for the two epimeric thiols with the C-terminal peptide thioester is altered by the ability of the proximal α-amino group to participate in a base catalysis step to generate the reactive thiolate at the β-position. This proposed intramolecular proton transfer step would be similarly affected by the *cis* and *trans* orientation of the β-isopropyl group and the peptide chain [109].

Brik and coworkers further demonstrate the synthetic capability of ligation-desulfurization at β-thiol Leu through the total synthesis of the HIV-1 Tat protein, which the authors had previously attempted using a side-chain auxiliary approach [61] (see Sect. 3.1.2). In this instance, HIV-1 Tat was disconnected at a β-thiol Leu residue and a native Cys residue for an iterative ligation strategy employing fragments **112**, **113**, and **114** prepared via Fmoc-SPPS (Scheme 28) [110]. Ligation between peptide **112**, containing an N-terminal β-thiol Leu residue, and C-terminal peptide thioester **113**, bearing an N-terminal Cys thiazolidine, was first carried out to generate an intermediate peptide ligation product, which was subsequently desulfurized to generate the native Leu residue in peptide **115**. It should be noted



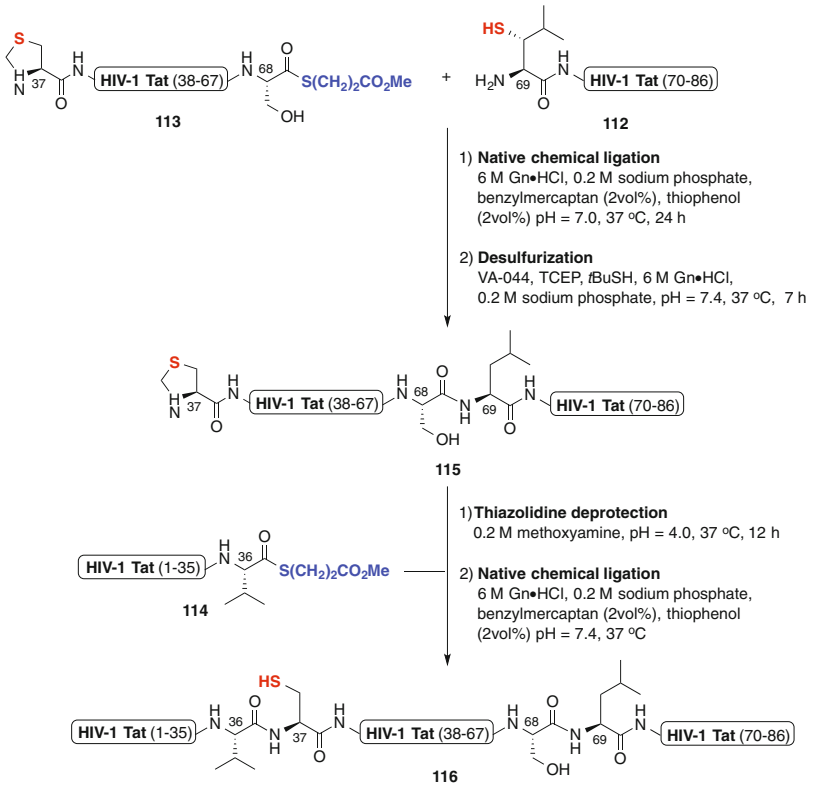
Scheme 27 Competition experiment between β-thiol Leu epimers and proposed origin of the observed selectivity

that a final, global desulfurization protocol was not employed to avoid the use of protecting groups on the N-terminal HIV-1 Tat fragment **114**, which is rich in Cys residues. Deprotection of intermediate thiazolidine **115** followed by ligation with thioester **114** afforded the full-length HIV-1 Tat protein **116** [110].

4.3.6 Proline

An approach to Pro ligation junctions employing the commercially available, protected γ-thiol Pro derivatives **93** and *epi*-**93** (Scheme 23) was offered by Danishefsky and coworkers in 2011 [111]. As with the diastereomeric Leu derivatives, a substantial rate difference was observed for the two γ-epimers, with only the *trans* derivative of **93** capable of facilitating ligation with C-terminal peptide thioesters. Once again, steric hindrance in the cyclic transition state of the $S \rightarrow N$ acyl shift was implicated as an explanation for the rate differential. Notably, in subsequent reports, Danishefsky and coworkers also demonstrated the utility of *trans* Pro derivative **93** in the ligation-based assembly of a large glycopeptide fragment of EPO [112, 113].

A synthetic approach to both diastereomers of the Boc-protected γ-thiol Pro derivative **94** was also reported by Otaka and coworkers [114]. The authors of this study confirmed the finding that use of the *trans* Pro derivative **94** was essential to facilitate ligation. The relative rate differential between rapid Cys-mediated ligations and ligation reactions mediated by *trans* γ-thiol Pro residues was also exploited to effect a one-pot, dual kinetically controlled ligation reaction.



Scheme 28 Total synthesis of the HIV-1 Tat protein using ligation-desulfurization at β-thiol Leu

This methodology strategically utilized an *N*-sulfanylethylanilide (SEAlide) peptide as a latent thioester moiety (see [115] for more details on SEA peptide technologies) [114].

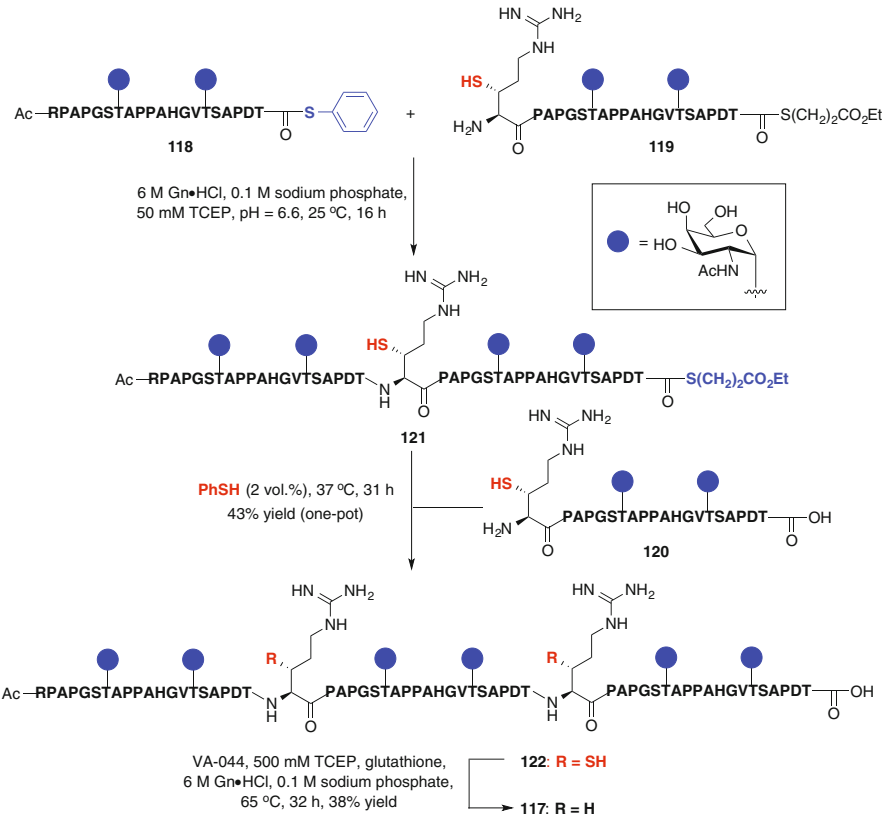
4.3.7 Glutamine

A ligation-desulfurization approach to Gln ligation junctions using a γ-thiol Gln derivative was reported by Brik and coworkers in 2012 [116]. The preparation of a diastereomeric mixture of the requisite γ-thiol Gln building block **95** was accomplished in ten steps from L-Asp. Ligation chemistry with a variety of model C-terminal peptide thioesters was unhindered by the diastereomeric γ-thiols. Interestingly, attempted radical desulfurization of γ-thiol Gln-containing peptides produced a complex mixture of products. As such, removal of the ligation auxiliary was accomplished via reductive desulfurization with nickel boride, affording enantiomerically pure peptide products [116].

4.3.8 Arginine

Payne and coworkers recently reported the synthesis of a β-thiol arginine (Arg) building block from commercially available Garner's aldehyde [117], a configurationally stable α-amino aldehyde, for use in ligation-desulfurization chemistry [118]. The Arg derivative **96** (Scheme 23) was shown to facilitate ligation with peptide thioesters bearing a range of functionality at the C-terminal position, and detailed kinetic studies indicated that ligation rates were inversely proportional to steric bulk at the C-terminal thioester residue, largely mirroring studies performed by Dawson and coworkers on native chemical ligation at Cys residues [13]. Interestingly, removal of the β-thiol auxiliary using radical desulfurization was substantially slower than the corresponding desulfurization of Cys to Ala. It was proposed that the Arg guanidine side-chain moiety was interfering with the standard radical desulfurization mechanism. Nonetheless, the Arg building block was successfully utilized in the ligation-based assembly of a 60-amino acid homogeneous MUC1 glycopeptide **117**, corresponding to three copies of the 20-residue MUC1 variable number tandem repeat (VNTR) sequence and bearing six *O*-linked glycans (Scheme 29).

Construction of the 60-amino acid MUC1 glycopeptide was achieved in a one-pot fashion from peptide fragments **118**, **119**, and **120** using a kinetically-controlled ligation strategy, first reported by Kent and coworkers for the construction of crambin [119]. The technique capitalizes on the innate reactivity difference between alkyl and aryl thioesters to perform iterative ligations in a rapid and efficient manner, while minimizing protecting group manipulations and intermediary purifications. This strategy has been successfully applied to the synthesis of a number of complex protein targets, including human lysozyme [120], HIV-1 protease [121], and full-length glycosylated EPO [87]. For the synthesis of MUC1 60-mer **117**, it was envisaged that kinetically-controlled, β-thiol Arg-mediated ligations between three functionalized MUC1 VNTR fragments would facilitate rapid construction of the target glycopeptides (Scheme 29). To this end, peptide thiophenyl thioester **118** was first ligated with bifunctional peptide **119** bearing an N-terminal β-thiol Arg and a C-terminal peptide alkyl thioester. The large increase in reactivity associated with aryl thioester **118** relative to the corresponding alkyl thioester **119** effectively promoted the intermolecular condensation of the two fragments to yield the 40-residue intermediate **121** rather than the competing cyclization or oligomerization of peptide **119** by reaction at the alkyl thioester. Importantly, this reaction took place in the absence of a thiol additive to avoid in situ activation of the alkyl thioester through transthioesterification. Without isolation, ligation of intermediate product **121** with peptide **120** containing an N-terminal β-thiol Arg residue was accomplished via addition of the final peptide fragment, along with 2 vol.% thiophenol, to generate the 60-amino acid polypeptide **122** in 43% isolated yield. Removal of the Arg β-thiol auxiliaries using a double desulfurization reaction then afforded the native glycopeptide **117** [118].



Scheme 29 Synthesis of a MUC1 glycopeptide oligomer using a kinetically-controlled ligation-desulfurization strategy at β-thiol Arg

4.3.9 Aspartic Acid

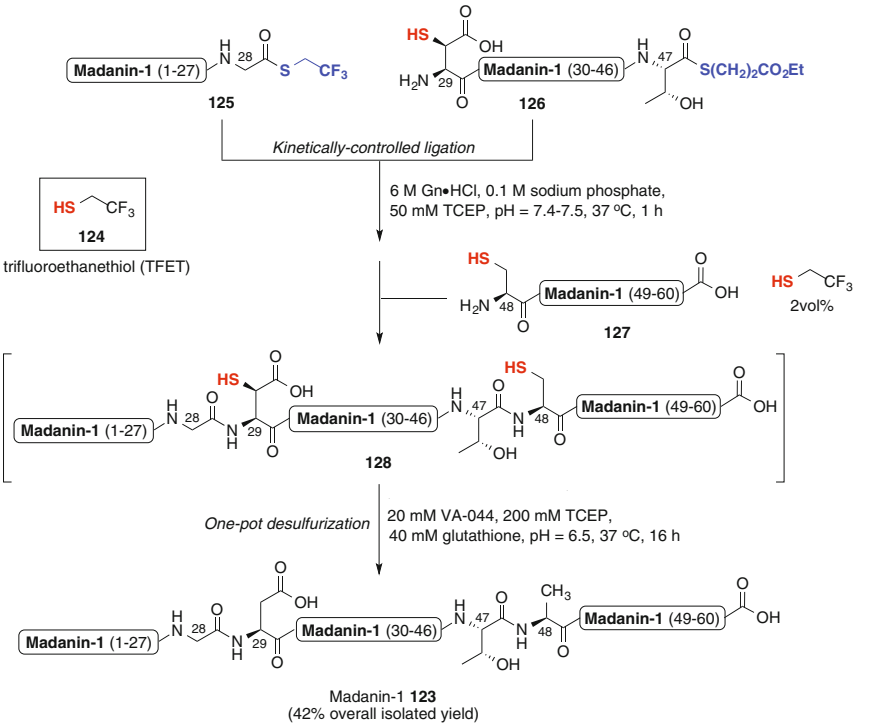
A concise, three-step synthesis of β-thiol Asp derivative **97** (Scheme 23) was developed in 2013 by Thompson et al. from commercially available Boc-Asp (O*t*Bu)-OH by employing a key sulfenylation reaction to install the requisite protected-thiol moiety [122]. Building block **97** was shown to facilitate the high-yielding synthesis of native peptides through ligation-desulfurization chemistry, and ligation reactions were found to proceed with equal efficiency, regardless of the configuration at the β-position. Interestingly, the authors of this report also demonstrated that β-thiol Asp residues could be selectively desulfurized in the presence of unprotected Cys residues upon treatment with TCEP and dithiothreitol (DTT) at 65 °C and pH 3, in the absence of a radical initiator. As standard reductive and radical-based desulfurization methods are unselective, application of these techniques requires the protection of all native Cys residues in the target sequence. In contrast, the ability to chemoselectively remove the β-thiol Asp ligation auxiliary

abrogates the need for protecting group manipulation in protein targets with functionally important Cys residues. The utility of the ligation-chemoselective desulfurization protocol was demonstrated through an efficient, one-pot synthesis of the N-terminal, extracellular domain of the chemokine receptor CXCR4. The target CXCR4(1–38) fragment contained a native Cys residue (although at an intractable Pro-Cys junction) and two post-translational modifications – an *N*-linked glycan and Tyr *O*-sulfation [122].

Tan and coworkers subsequently reported the 7-step synthesis of a modified β-thiol Asp derivative **98** and the application of this building block in the total synthesis of the 60-amino acid neuropeptide human galanin-like peptide (hGALP) using ligation-desulfurization chemistry [123]. Notably, the authors of this report utilize the β-epimer of the thiol Asp derivative previously employed for the ligation-based assembly of CXCR4(1–38) [122]. In contrast to the rate differential observed with epimeric β-thiol Leu derivatives **91** [109], these results confirm that configuration at the β-position does not have a large impact on the efficiency of ligation reactions at β-thiol Asp residues [122, 123].

Very recently, the application of the β-thiol Asp derivative **97** [122] has been reported in the one-pot, three-component total synthesis of madanin-1 **123** (Scheme 30), a small, Cys-free thrombin inhibitory protein derived from the hard tick *H. longicornis* [124]. In this study, the target 60-amino acid protein **123** was disconnected at Asp and Ala ligation junctions using ligation-desulfurization chemistry mediated by β-thiol Asp and Cys, respectively. Interestingly, for the construction of the target protein, the authors also reported the use of a novel thiol additive, trifluoroethanethiol (TFET) **124**, which facilitates the facile application of a one-pot ligation-desulfurization protocol [124]. Previous attempts to streamline the ligation-desulfurization strategy into a one-pot process have been hampered by the use of aryl thiol ligation additives which, because of their radical quenching ability [125], complicate the radical desulfurization process [126–128]. Given the importance of aryl thiol additives in modulating thioester reactivity and promoting rapid ligation reactions [15, 16], a number of approaches have aimed to facilitate the post-ligation removal of aryl thiols. The liquid–liquid extraction of thiophenol [128] and the development of bifunctional aryl thiol catalysts which can be captured using a resin-bound aldehyde following the ligation reaction [126] have recently been employed to facilitate one-pot ligation desulfurization reactions.

The alkyl thiol TFET was designed as an alternative additive in ligation reactions to circumvent the issues posed by aryl thiol additives in radical desulfurization reactions while maintaining efficiency as a ligation catalyst [124]. Because the pK_a of TFET (=7.30) is comparable to the pK_a of common aryl thiol ligation additives, it was postulated that this alkyl thiol would be sufficiently nucleophilic to promote initial transthioesterification with the unactivated C-terminal peptide thioester and would also maintain sufficient leaving group ability to promote the acylation of the Cys thiol moiety. The efficacy of the additive was showcased in a kinetically-controlled one-pot total synthesis of madanin-1 **123** (Scheme 30). To this end, preformed TFET thioester **125**, corresponding to residues 1–28 of madanin-1 was ligated with bifunctional peptide madanin-1 (29–47) **126**, bearing an N-terminal



Scheme 30 One-pot total synthesis of madanin-1 **123** using kinetically-controlled ligation-desulfurization chemistry with trifluoroethanethiol (TFET) **124**

β-thiol Asp and an unactivated C-terminal alkyl thioester. Without isolation, madanin-1 (48–60) **127**, containing an N-terminal Cys residue, was added to the ligation mixture, along with 2 vol.% TFET to activate the latent C-terminal alkyl thioester and promote a second ligation reaction. The 60-amino acid ligation product **128** was then subjected directly to radical desulfurization, cleanly affording the target protein **123** in 42% isolated yield over the three steps. The one-pot total synthesis of madanin-1 **123** therefore served as a proof-of-concept for the efficiency of β-thiol Asp residues in ligation-desulfurization chemistry, as well as application of the new thiol additive, TFET, in the one-pot construction of Cys-free proteins [124].

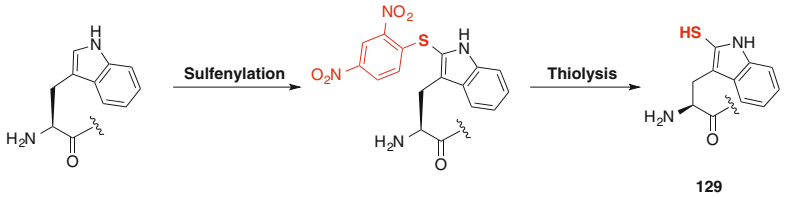
4.3.10 Glutamic Acid

Using a similar synthetic approach to the one employed for the synthesis of β-thiol Asp building block **97**, Cergol et al. reported the preparation of γ-thiol Glu derivatives **99** and **100** (Scheme 23) and their application in ligation-desulfurization chemistry [128]. Initial attempts to incorporate building block **99** into model

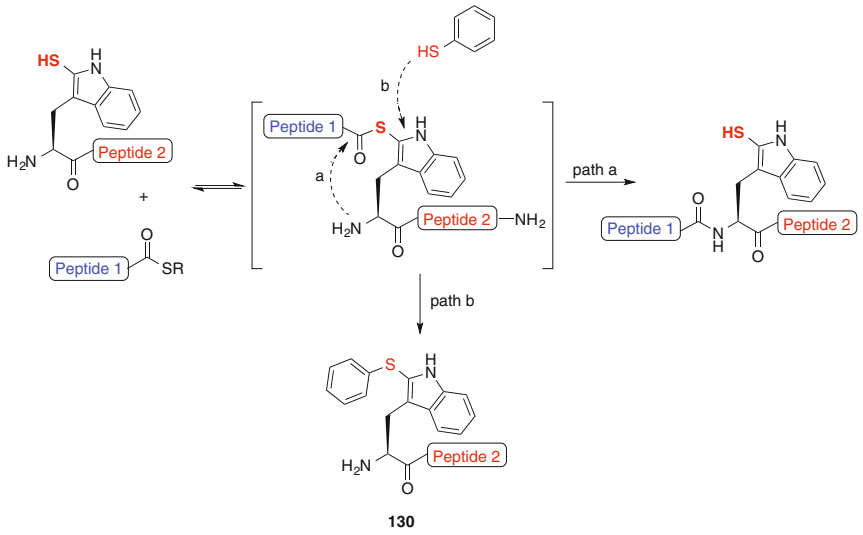
peptides using Fmoc-SPPS were complicated by the instability of the γ-thiol Glu derivative in the acidic cleavage cocktail. It was determined that, at acidic pH, the deprotected γ-thiol could facilitate nucleophilic cleavage of the amide backbone, resulting in loss of the terminal thiol-derived amino acid. Subsequent incorporation of the asymmetric disulfide building block **100** circumvented this issue, allowing for the efficient construction of model peptides. Ligation reactions proceeded in high isolated yields and could be followed by desulfurization using either a one- or two-pot protocol. To facilitate one-pot desulfurization, ligation reactions were initially performed using thiophenol as an additive, which could be removed almost entirely from the reaction mixture by extraction with diethyl ether [128]. Subsequently, the application of γ-thiol Glu derivative **100** in a TFET-mediated iterative ligation-desulfurization strategy enabled the efficient one-pot total chemical synthesis of chimadanin-1, another small thrombin inhibitory protein isolated from the hard tick *H. longicornis* [124].

4.3.11 Tryptophan

An innovative approach to ligation-desulfurization at N-terminal Trp residues has recently been reported [129]. Based on results from Scoffone et al. in the late 1960s [130, 131], the method utilizes a chemoselective sulfenylation protocol, followed by a mild and selective thiolysis reaction [132], to install a thiol ligation auxiliary at the 2-position of the Trp indole ring (e.g., **129**, Scheme 31) in unprotected peptides or resin-bound peptide intermediates. This methodology therefore eliminated the need to synthesize a protected thiol-derived Trp amino acid building block. Ligation reactions at model peptides bearing the 2-thiol Trp moiety **129** were found to proceed in good yields with preformed thiophenyl thioesters in the absence of an exogenous thiol additive. Interestingly, in the presence of exogenous aryl thiol, significant quantities of the 2-thioether byproduct **130** were observed (Scheme 32). Mechanistically, following the initial transthioesterification step, the positioning of the thiol auxiliary at the 2-position of the Trp indole ring required the $S \rightarrow N$ acyl shift to proceed through a seven-membered ring intermediate (path a, Scheme 32). It was postulated that, in the presence of excess thiophenol, a slower $S \rightarrow N$ acyl migration step allowed the intermediate bridged thioester to be intercepted at the C-2 position of the indole ring with exogenous aryl thiol (path b) before the intramolecular acyl shift could occur (Scheme 32). Following optimization of the ligation conditions, reductive desulfurization using Pd on Al_2O_3 in the presence of H_2 gas cleanly afforded native peptide products. The technique was also applied to the synthesis of a glycopeptide fragment of the N-terminal extracellular domain of the chemokine receptor CXCR1. In this example, installation of the key thiol auxiliary was performed using an efficient solid-phase sulfenylation protocol [129].



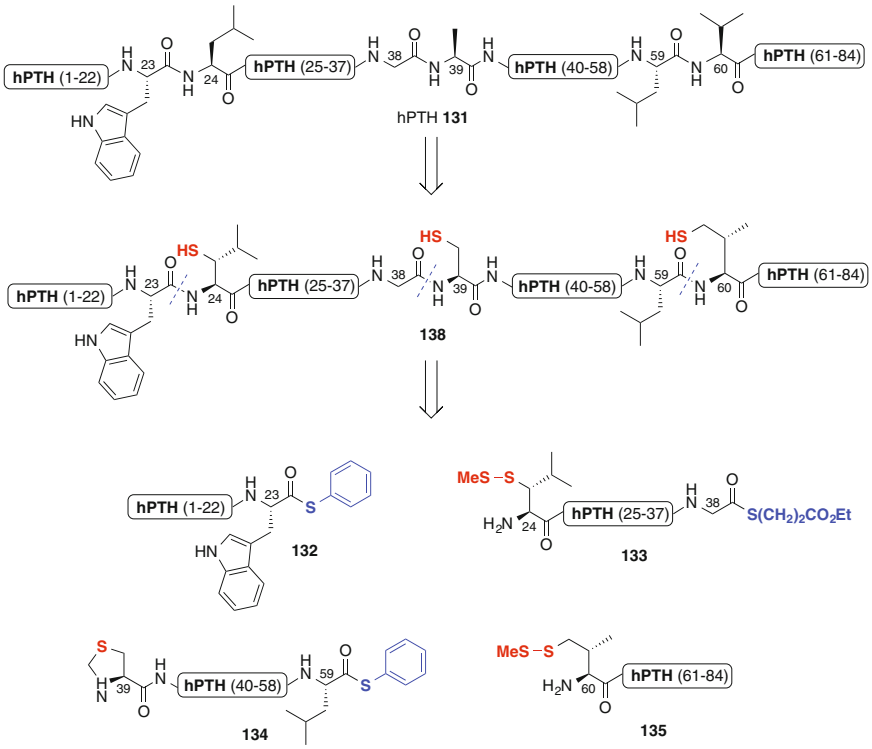
Scheme 31 Synthesis of 2-thiol Trp derivatives (e.g., **129**) through the sulfenylation of Trp



Scheme 32 Transthioesterification of 2-thiol Trp and subsequent $S \rightarrow N$ acyl shift (path a) or interception with exogenous thiophenol (path b)

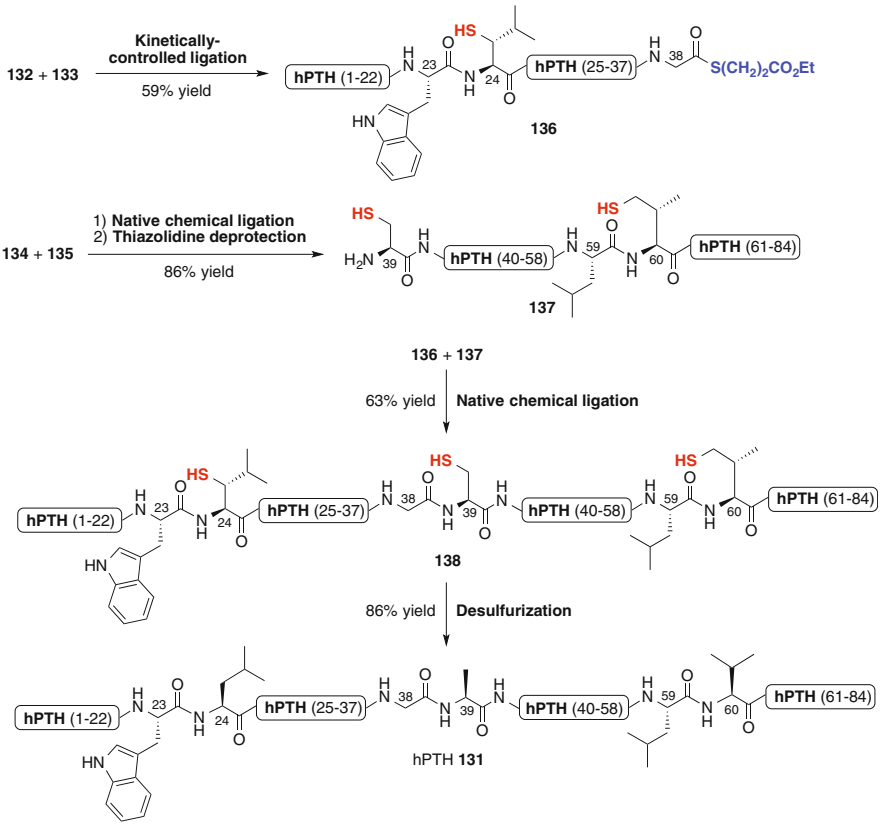
4.3.12 Protein Synthesis via Ligation at Non-Cys Sites

The flourish of activity in the development of building blocks for ligation-desulfurization chemistry has provided enormous flexibility for the disconnection of target proteins. As discussed, these building blocks are used in combination with ligation at Cys or ligation-desulfurization at Ala for the construction of complex proteins. A powerful example of the scope of ligation using a combination of these building blocks was provided by Danishefsky and coworkers in 2011 with the construction of human parathyroid hormone (hPTH) **131** using a convergent ligation-global desulfurization strategy employing Cys, β-thiol Leu, and γ-thiol Val ligation disconnections (Scheme 33) [133]. It was envisaged that the target hPTH protein could be synthesized from peptide fragments **132–135** by combining iterative ligation technologies for the construction of proteins in both the *N*-to-*C* and *C*-to-*N* directions (Scheme 34).



Scheme 33 Retrosynthetic analysis for the total chemical synthesis of hPTH **131**

First, thiophenyl thioester **132** and bifunctional peptide **133**, bearing an N-terminal β-thiol Leu residue and a C-terminal alkyl thioester were ligated in a kinetically-controlled ligation reaction, affording hPTH(1–38) **136** as the alkyl thioester derivative in 59% yield (Scheme 34). Peptide fragments **134** and **135** were then joined using a γ-thiol Val-mediated ligation reaction. Subsequent thiazolidine deprotection afforded hPTH(39–84) **137** bearing a free N-terminal Cys residue for further functionalization. Accordingly, standard native chemical ligation of peptide **137** with thioester **136** facilitated construction of the full-length polypeptide sequence **138** in 63% yield. Radical desulfurization of the three thiol ligation auxiliaries finally afforded the native protein hPTH **131** in 86% yield following HPLC purification. Notably, the hPTH prepared in this study using total chemical synthesis was shown to be significantly more pure than the comparative recombinant hPTH reference sample used to confirm the identity of the protein. The superior quality of the product obtained using chemical methods attests to the importance of ligation chemistry, including non-Cys ligation strategies, for the production of homogeneous proteins for use in biological studies. It is therefore envisaged that the ligation-desulfurization motif will continue to be widely utilized in the total chemical synthesis of target proteins, particularly as thiol-derived amino

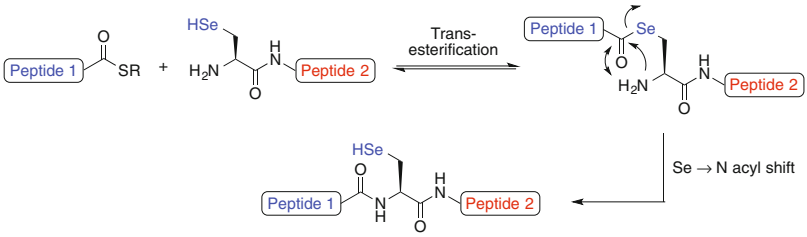


Scheme 34 The synthesis of hPTH **131** using a convergent, iterative ligation strategy followed by global desulfurization

acid building blocks become increasingly accessible, e.g., through commercial sources.

4.4 Ligation at Selenocysteine

Selenocysteine (Sec) is often considered to be the 21st amino acid [134], and there are a number of naturally occurring proteins that contain structurally and functionally crucial Sec residues [135]. These considerations, together with the structural similarities between Cys and Sec, render native chemical ligation at N-terminal Sec residues a logical extension of chemoselective ligation technologies. Indeed, in 2001, the laboratories of van der Donk, Hilvert, and Raines independently reported native chemical ligation at Sec [136–138]. Mechanistically, Sec-mediated ligations were proposed to proceed in an analogous fashion to ligation at Cys (Scheme 35),



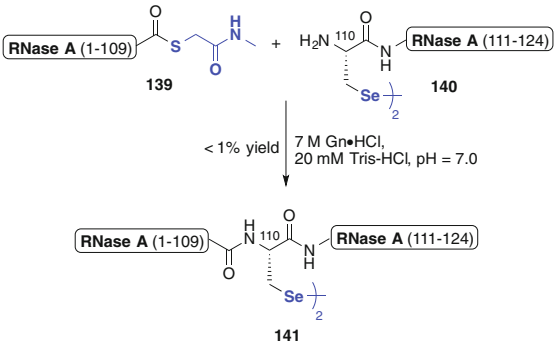
Scheme 35 Native chemical ligation at selenocysteine (Sec)

whereby an initial, reversible transesterification reaction between a C-terminal peptide thioester and an N-terminal Sec residue first proceeds to form a bridged selenoester intermediate. Intramolecular rearrangement of the selenoester in a *Se* → *N* acyl shift through a five-membered ring intermediate subsequently generates a new amide bond.

Despite similar mechanistic considerations, Sec displays a number of distinct physicochemical properties from its sulfur counterpart which have important implications in ligation chemistry. In particular, the pK_a of the Sec selenol functionality is approximately 5.24–5.63 [139, 140], implying that at physiological pH, Sec exists primarily in the anionic selenolate form. The corresponding Cys thiol, however, has a pK_a of 8.25 [139], and remains largely protonated at neutral pH. Selenolates also display enhanced nucleophilicity [139] and leaving group abilities relative to their thiolate analogues [141]. Collectively, these unique chemical properties suggest that ligation at Sec residues might proceed faster than ligation at the corresponding Cys residues, particularly at slightly acidic pH. Raines and coworkers indeed demonstrated that ligation reactions with Sec in the presence of ligation buffer and TCEP proceeded 10^3-fold faster than the analogous Cys ligation at pH 5.0, leading to the proposition that Sec ligations may be performed chemoselectively in the presence of Cys residues [138].

However, in the absence of TCEP, van der Donk and coworkers observed sluggish reactivity in ligation reactions facilitated by N-terminal Sec residues. The authors speculated that the decreased rate of reaction was a consequence of a low steady-state concentration of reactive selenol in the ligation mixture, with the starting Sec peptide existing primarily as the symmetrical diselenide or the selenylsulfide variant under the ligation conditions employed [136]. The unique redox properties of diselenides, particularly their relative stability and large negative reduction potential [142, 143], support this hypothesis. Furthermore, the observation by Hilvert and coworkers that Sec-mediated ligation reactions did not proceed in the absence of TCEP and exogenous thiophenol also suggests that the overall rate of reaction at Sec is heavily dependent on the ability to reduce diselenides in the ligation mixture [137]. The observation that the inclusion of TCEP in Sec-mediated ligations resulted in the formation of deselenization byproducts [136], however, has fuelled the adoption of thiol additives, such as MPAA, as mild reductants in Sec ligation chemistry [144].

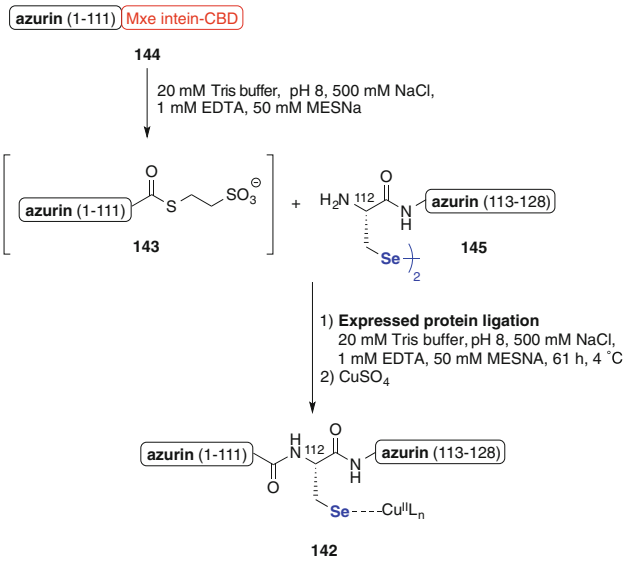
Scheme 36 Synthesis of RNase A using a Sec-mediated expressed protein ligation reaction



There have been a number of applications of Sec-mediated ligation chemistry for the synthesis of diverse peptide and protein targets, including a Sec analog of the C-terminal fragment of ribonucleotide reductase [136] and a fully folded Cys38Sec mutant of the 58-amino acid bovine pancreatic trypsin inhibitor (BPTI) [137]. The ligation technology has also been used in conjunction with expressed protein ligation to construct a Sec analog of ribonuclease A (RNase A) [138]. To this end, a 109-amino acid recombinant thioester **139** was produced in *E. coli* and subsequently ligated to the short, synthetic selenopeptide **140**, prepared using SPPS, to provide the target semi-synthetic protein **141** (Scheme 36). The Cys110Sec RNase A analogue **141** was isolated in very low yields because of the poor recovery of recombinant thioester, but nonetheless displayed ribonucleolytic activity consistent with the wild-type enzyme, suggesting that the Sec analog was properly folded [138].

A novel application of Sec ligation chemistry in protein engineering has been the preparation of a semi-synthetic analogue of the type 1 blue copper protein, azurin, containing a Cys112Sec mutation [145]. Cys^{112} is involved in the active site of the wild-type redox metalloprotein through coordination to the copper ligand. It was therefore envisaged that synthesis of a Sec^{112} variant might provide important insight into the structure and function of the protein. Synthesis of the target protein **142** was accomplished using expressed protein ligation of an N-terminal recombinant thioester **143** corresponding to residues 1–111 of azurin (Scheme 37). The recombinant thioester was generated in situ from the corresponding fusion protein **144** upon treatment with MESNa. Ligation with selenopeptide **145**, bearing an N-terminal Sec residue afforded the full-length Cys112Sec azurin. Addition of copper sulfate then produced the ligand-bound protein **142** in a yield of ~0.4 mg/L of culture. Synthesis of the engineered protein enabled a detailed comparison of the electronic absorption spectra and reduction potential of the engineered variant with wild-type azurin [145].

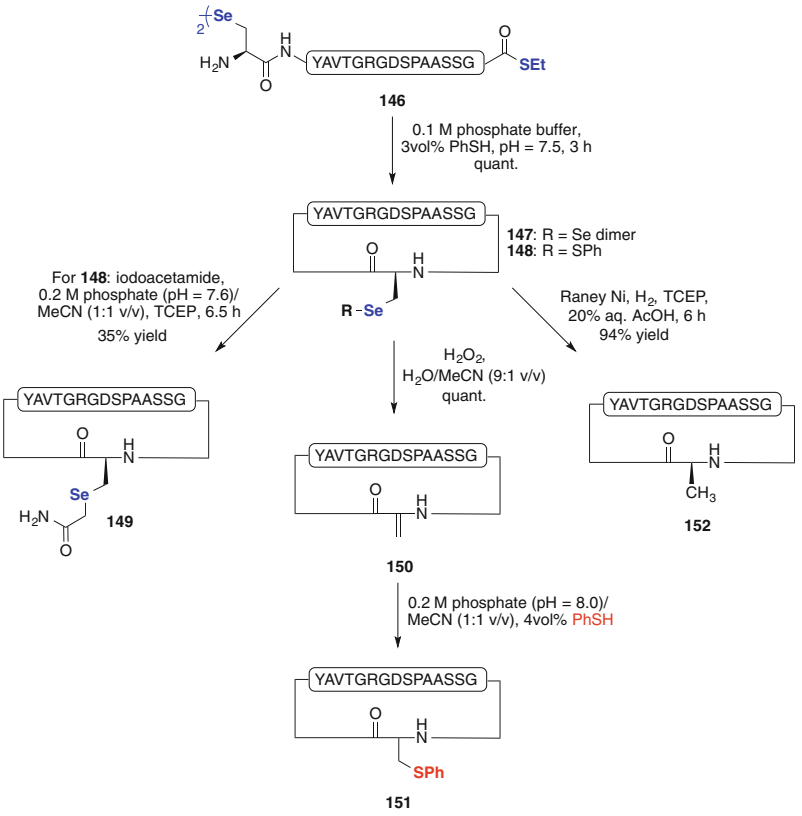
As with ligation at Cys, there has been considerable interest in the post-ligation manipulation of Sec residues to facilitate access to diverse ligation junctions. In 2002, Hilvert and coworkers reported the synthesis of a cyclic peptide from



Scheme 37 Synthesis of Cys112Sec azurin using expressed protein ligation

bifunctional precursor **146** using a Sec ligation strategy (Scheme 38) [146]. Following construction of the cyclic selenopeptide (isolated as the symmetrical diselenide **147** and the seleno-thiophenyl sulfide **148**), the Sec residue was used as a handle for chemical manipulation. Alkylation of selenylsulfide **148** with iodoacetamide in the presence of TCEP generated the selenoether derivative **149** in 35% yield. Oxidative elimination with hydrogen peroxide afforded the dehydroalanine (Dha) derivative **150**, which could be further functionalized in a thiol Michael reaction to generate thioethers such as **151**. Finally, reductive deselenization of **146** in the presence of Raney Ni, akin to the post-ligation desulfurization of Cys residues [54], afforded peptide **152**, bearing the corresponding Ala residue at the ligation junction [146]. Interestingly, in their original report of metal-free radical desulfurization, Danishefsky and coworkers also reported the extension of the radical protocol to the deselenization of Sec to Ala [72].

Despite the potential for diverse post-ligation modifications at Sec residues, Sec-mediated ligation chemistry has not been widely adopted for the routine construction of target peptides and proteins which do not contain Sec residues in the final product [147]. The lack of commercially available Sec building blocks for direct incorporation into peptides using standard SPPS and the ability to modify readily the side-chain of Cys residues [54, 66, 148] have generally favored the use of standard native chemical ligation at Cys.

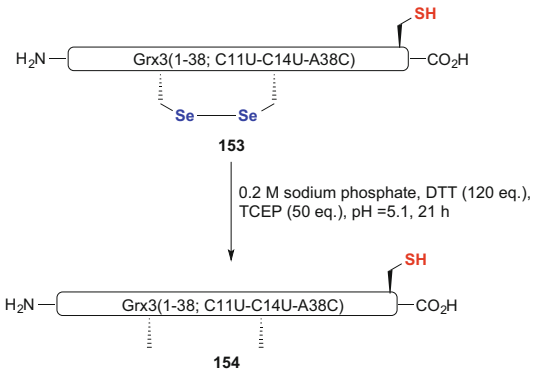


Scheme 38 Sec-mediated backbone cyclization followed by side-chain functionalizations

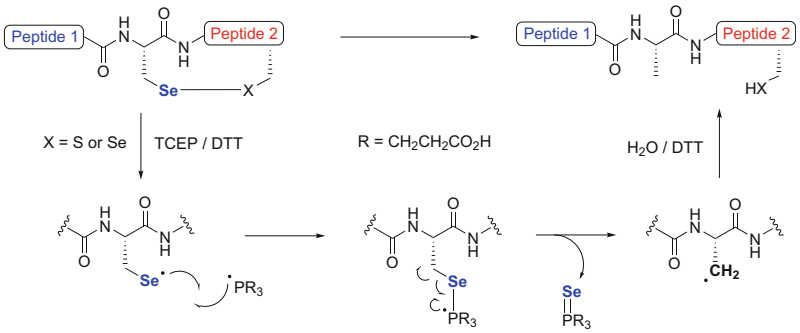
4.5 Ligation-Deselenization Chemistry

A 2010 report by Dawson and coworkers outlining the mild and selective deselenization of Sec in the presence of unprotected Cys residues [144] has contributed to a resurgence of interest in Sec-mediated ligation chemistry as a general strategy for the synthesis of peptides and proteins [147]. At the time of this seminal report, all known protocols for the reductive or radical-based desulfurization of Cys (see above) effected global cleavage of unprotected thiols within the target sequence. As such, the synthesis of peptides and proteins bearing native, non-ligation site Cys residues using ligation-desulfurization chemistry demanded the use of side-chain Cysprotecting groups. By providing the first chemoselective approach to Ala ligation junctions, Dawson and coworkers provided an important tool for the construction of proteins from fully deprotected peptide precursors [144].

The deselenization protocol involved the treatment of Sec-containing peptides, at room temperature and in aqueous media, with excess TCEP in the presence of dithiothreitol (DTT). The utility of the reaction was demonstrated on a model

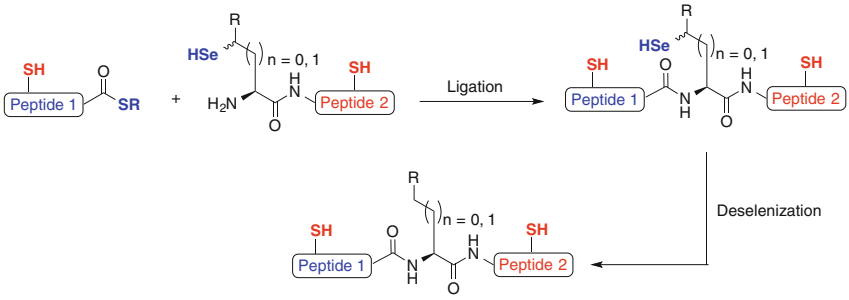


Scheme 39 Chemoselective deselenization of Sec to Ala in the presence of an unprotected Cys residue

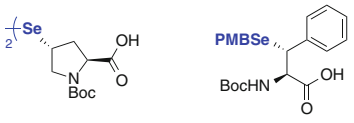


Scheme 40 Proposed mechanism for the phosphine-mediated deselenization of Sec

system and in a larger polypeptide corresponding to Grx(1–38) **153** containing two Cys to Sec mutations at positions 11 and 14, and a single Ala to Cys mutation at residue 38 (Scheme 39). Under the optimized deselenization conditions, the two Sec to Ala conversions were accomplished to afford compound **154** with only minor amounts of the globally reduced Grx (1–38) product observed (Scheme 39). Notably, a Grx(1–38) Cys mutant, containing no Sec residues, was completely stable to the deselenization conditions [144]. Mechanistically, the phosphine-mediated deselenization reaction was thought to proceed via a radical pathway (Scheme 40), similar to the one proposed for the radical desulfurization of thiols by trialkylphosphites and trialkylphosphines [74, 75] and implicated by Danishefsky and coworkers in their development of the metal-free radical desulfurization protocol [72]. Although proceeding through similar pathways, the observed selectivity of the deselenization reaction for Sec over Cys might be attributed to the preferential formation of selenium-centered radicals over the corresponding sulfur-centered radicals, particularly under the mild conditions employed for deselenization.



Scheme 41 Chemoselective ligation-deselenization at β- ($n = 0$) and γ- ($n = 1$) selenol amino acids



155 156

Scheme 42 γ-Selenoproline **155** and β-selenophenylalanine **156** building blocks

By allowing for the chemoselective conversion of Sec to Ala in the presence of unprotected Cys residues, the ligation-deselenization strategy effectively provided a means of accomplishing, without protecting group manipulations, ligation at Ala and Cys junctions in the same protein target. In a similar manner to the extension of ligation-desulfurization chemistry to include thiol-derived proteinogenic amino acid building blocks (see Scheme 23) [54], Dawson and coworkers also proposed that the logic of chemoselective ligation-deselenization chemistry could be extended to include synthetic selenol-derived amino acids [144]. Specifically, non-proteinogenic building blocks bearing a suitably positioned β- or γ-selenol auxiliary could facilitate a Sec-mediated ligation reaction to afford an unnatural selenopeptide product. The mild and chemoselective removal of the selenol auxiliary in the presence of TCEP could then be effected in the presence of unprotected Cys residues elsewhere in the target sequence (Scheme 41), thereby expanding the scope and flexibility of chemoselective ligation chemistry.

Chemoselective ligation-deselenization chemistry has recently been extended through the preparation of a γ-selenoproline building block **155** [113] and a β-selenophenylalanine derivative **156** [149] (Scheme 42). The *trans*-γ-selenoproline derivative **155** was prepared by Danishefsky and coworkers in three steps from a protected hydroxyproline precursor. The *trans*-derivative was chosen over the corresponding γ-epimer on the basis of favorable results obtained from earlier studies with the analogous *trans*-γ-thiol derivatives **93** and **94** [111, 113, 114]. Following incorporation into model peptides, building block **155** was shown to mediate ligation effectively with a variety of C-terminal peptide thioesters, including Gly, Ala, Phe, and Val. Notably, ligation reactions were performed in

the absence of TCEP to avoid premature deselenization of the selenol ligation auxiliary. The authors instead used MPAA [16] as both an exogenous thiol exchange catalyst and a mild reductant for the generation of free selenol from the starting diselenide [144]. Deselenization reactions were performed in a one-pot fashion through the sequential addition of DTT and TCEP. Importantly, the chemoselectivity of the deselenization protocol was also confirmed in the presence of an unprotected γ-thiol Pro residue [113].

The preparation of β-selenophenylalanine derivative **156** and its application in ligation-deselenization chemistry was subsequently reported [149]. The synthesis of the key protected amino acid building block was accomplished in seven steps from Garner's aldehyde [117]. Notably, Garner's aldehyde has also served as a synthetic precursor for the preparation of β-thiol Arg derivative **96** by Payne and coworkers [118] and has been proposed as a common starting point for the divergent synthesis of both β-thiol and β-selenol amino acid derivatives [149]. Following the synthesis of **156** and its incorporation into model peptides using standard Fmoc-SPPS, a number of ligation reactions were performed to evaluate the utility of the building block in ligation-deselenization chemistry. Reactions proceeded in moderate to good yields for the majority of C-terminal peptide thioesters examined, requiring 24–48 h to reach completion. The slow rates of reaction were attributed to the relative stability of the starting diselenide and the reliance on MPAA, rather than the more powerful reductant TCEP, to liberate free selenol to promote the ligation reaction. Interestingly, deselenization of the purified ligation materials with TCEP and DTT led to substantial formation of peptide derivatives bearing diastereomeric β-hydroxy Phe at the ligation junction. The prevalence of this byproduct was dramatically reduced when ligation-deselenization reactions were performed as a one-pot protocol in the presence of exogenous MPAA. Under these conditions, the rate of deselenization also dramatically decreased, perhaps because of the ability of aryl thiols such as MPAA to act as competitive radical scavengers [126, 127, 144]. Nonetheless, native peptide products, including those bearing unprotected, non-ligation site Cys residues, were isolated in good yields following the one-pot protocol for ligation-deselenization at β-selenophenylalanine [149].

Although the application of ligation-deselenization chemistry has thus far been limited to model systems and small polypeptides, it is envisaged that this technology will also be amenable to the construction of proteins. By minimizing the need for late-stage protecting group manipulations, chemoselective ligation-deselenization chemistry will enhance the availability of ligation junctions and serve to expedite the construction of complex targets [147].

5 Conclusion

Over the last 20 years, native chemical ligation has ushered in a new era in the total chemical synthesis of proteins by enabling the efficient, programmed construction of native structures, including those bearing post-translational modifications, as

well as the modular synthesis of strategically engineered protein variants. This chapter has summarized the importance of native chemical ligation and highlighted a number of modern extensions to the original technology which have facilitated these synthetic feats. In particular, novel methods for Cys-free ligation, including auxiliary-mediated ligation and ligation-desulfurization chemistry, employing both Cys residues and synthetic thiol-derived amino acid variants, have been extensively explored. These techniques have served to increase dramatically the availability of synthetically viable ligation junctions and have expanded the flexibility of modern ligation chemistry for the construction of diverse targets. In addition, the advent of chemoselective ligation-deselenization chemistry has provided a promising new strategy for the manipulation of proteins in the absence of protecting groups. It is predicted that these powerful new tools, and additional developments in chemoselective ligation technologies, will continue to fuel the construction of increasingly more complex protein targets in the years to come.

References

1. Walsh CT, Garneau-Tsodikova S, Gatto GJ Jr (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed 44:7342–7372
2. Walsh C (2006) Posttranslational modification of proteins: expanding nature's inventory. Roberts and Co Publishers, Englewood, Colo
3. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Synthesis of proteins by native chemical ligation. Science 266:776–779
4. Wieland T, Bokelmann E, Bauer L, Lang HU, Lau H (1953) Über peptidsynthesen. 8. Mitteilung bildung von S-haltigen peptiden durch intramolekulare wanderung von aminoacylresten. Justus Liebigs Ann Chem 583:129–149
5. Merrifield RB (1963) Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. J Am Chem Soc 85:2149–2154
6. Kemp DS, Leung S-L, Kerkman DJ (1981) Models that demonstrate peptide bond formation by prior thiol capture. I. Capture by disulfide formation. Tetrahedron Lett 22:181–184
7. Kemp DS, Kerkman DJ (1981) Models that demonstrate peptide bond formation by prior thiol capture. II. Capture by organomercury derivatives. Tetrahedron Lett 22:185–186
8. Fotouhi N, Galakatos NG, Kemp DS (1989) Peptide synthesis by prior thiol capture. 6. Rates of the disulfide-bond-forming capture reaction and demonstration of the overall strategy by synthesis of the C-terminal 29-peptide sequence of BPTI. J Org Chem 54:2803–2817
9. Liu C-F, Tam JP (1994) Chemical ligation approach to form a peptide bond between unprotected peptide segments. Concept and model study. J Am Chem Soc 116:4149–4153
10. Liu C-F, Tam JP (1994) Peptide segment ligation strategy without use of protecting groups. Proc Natl Acad Sci U S A 91:6584–6588
11. Schnoelzer M, Kent SBH (1992) Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. Science 256:221–225
12. Schnoelzer M, Alewood P, Jones A, Alewood D, Kent SBH (1992) In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. Int J Pept Protein Res 40:180–193
13. Hackeng TM, Griffin JH, Dawson PE (1999) Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. Proc Natl Acad Sci U S A 96:10068–10073

14. Camarero J, Adeva A, Muir T (2000) 3-Thiopropionic acid as a highly versatile multidetachable thioester resin linker. Lett Pept Sci 7:17–21
15. Dawson PE, Churchill MJ, Ghadiri MR, Kent SBH (1997) Modulation of reactivity in native chemical ligation through the use of thiol additives. J Am Chem Soc 119:4325–4329
16. Johnson ECB, Kent SBH (2006) Insights into the mechanism and catalysis of the native chemical ligation reaction. J Am Chem Soc 128:6640–6646
17. Evans TC Jr, Benner J, Xu MQ (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. Protein Sci 7:2256–2264
18. Muir TW (2003) Semisynthesis of proteins by expressed protein ligation. Annu Rev Biochem 72:249–289
19. Pollock SB, Kent SB (2011) An investigation into the origin of the dramatically reduced reactivity of peptide-prolyl-thioesters in native chemical ligation. Chem Commun 47:2342–2344
20. Kent SB (2009) Total chemical synthesis of proteins. Chem Soc Rev 38:338–351
21. Payne RJ, Wong CH (2010) Advances in chemical ligation strategies for the synthesis of glycopeptides and glycoproteins. Chem Commun 46:21–43
22. Raibaut L, Ollivier N, Melnyk O (2012) Sequential native peptide ligation strategies for total chemical protein synthesis. Chem Soc Rev 41:7001–7015
23. Macmillan D (2006) Evolving strategies for protein synthesis converge on native chemical ligation. Angew Chem Int Ed 45:7668–7672
24. Hackenberger CPR, Schwarzer D (2008) Chemoselective ligation and modification strategies for peptides and proteins. Angew Chem Int Ed 47:10030–10074
25. Unverzagt C, Kajihara Y (2013) Chemical assembly of N-glycoproteins: a refined toolbox to address a ubiquitous posttranslational modification. Chem Soc Rev 42:4408–4420
26. Haase C, Seitz O (2008) Extending the scope of native chemical peptide coupling. Angew Chem Int Ed 47:1553–1556
27. Dirksen A, Dawson PE (2008) Expanding the scope of chemoselective peptide ligations in chemical biology. Curr Opin Chem Biol 12:760–766
28. Gamblin DP, Scanlan EM, Davis BG (2009) Glycoprotein synthesis: an update. Chem Rev 109:131–163
29. Okamoto R, Mandal K, Ling M, Luster AD, Kajihara Y, Kent SBH (2014) Total chemical synthesis and biological activities of glycosylated and non-glycosylated forms of the chemokines CCL1 and Ser-CCL1. Angew Chem Int Ed 53:5188–5193
30. Bang D, Kent SB (2004) A one-pot total synthesis of crambin. Angew Chem Int Ed 43:2534–2538
31. 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 Chem Int Ed 47:6851–6855
32. Okamoto R, Mandal K, Sawaya MR, Kajihara Y, Yeates TO, Kent SB (2014) (Quasi-) racemic X-ray structures of glycosylated and non-glycosylated forms of the chemokine Ser-CCL1 prepared by total chemical synthesis. Angew Chem Int Ed 53:5194–5198
33. UniprotKB/TrEMBL Protein Database Release 2014_07 Statistics (2014) http://www.ebi.ac.uk/uniprot/TrEMBLstats. Accessed 1 Sept 2014
34. Canne LE, Bark SJ, Kent SBH (1996) Extending the applicability of native chemical ligation. J Am Chem Soc 118:5891–5896
35. Marinzi C, Bark SJ, Offer J, Dawson PE (2001) A new scaffold for amide ligation. Bioorg Med Chem 9:2323–2328
36. Botti P, Carrasco MR, Kent SBH (2001) Native chemical ligation using removable Nα-(1-phenyl-2-mercaptoethyl) auxiliaries. Tetrahedron Lett 42:1831–1833
37. Low DW, Hill MG, Carrasco MR, Kent SB, Botti P (2001) Total synthesis of cytochrome b562 by native chemical ligation using a removable auxiliary. Proc Natl Acad Sci U S A 98:6554–6559

38. Kawakami T, Aimoto S (2003) A photoremovable ligation auxiliary for use in polypeptide synthesis. Tetrahedron Lett 44:6059–6061
39. Marinzi C, Offer J, Longhi R, Dawson PE (2004) An o-nitrobenzyl scaffold for peptide ligation: synthesis and applications. Bioorg Med Chem 12:2749–2757
40. Clive DL, Hisaindee S, Coltart DM (2003) Derivatized amino acids relevant to native peptide synthesis by chemical ligation and acyl transfer. J Org Chem 68:9247–9254
41. Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. Proc Natl Acad Sci U S A 95:6705–6710
42. Chatterjee C, McGinty RK, Pellois JP, Muir TW (2007) Auxiliary-mediated site-specific peptide ubiquitylation. Angew Chem Int Ed 46:2814–2818
43. McGinty RK, Kim J, Chatterjee C, Roeder RG, Muir TW (2008) Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. Nature 453:812–816
44. Offer J, Dawson PE (2000) N-alpha-2-Mercaptobenzylamine-assisted chemical ligation. Org Lett 2:23–26
45. Kawakami T, Akaji K, Aimoto S (2001) Peptide bond formation mediated by 4,5-dimethoxy-2-mercaptobenzylamine after periodate oxidation of the N-terminal serine residue. Org Lett 3:1403–1405
46. Vizzavona J, Dick F, Vorherr T (2002) Synthesis and application of an auxiliary group for chemical ligation at the X-gly site. Bioorg Med Chem Lett 12:1963–1965
47. Offer J, Boddy CNC, Dawson PE (2002) Extending synthetic access to proteins with a removable acyl transfer auxiliary. J Am Chem Soc 124:4642–4646
48. Macmillan D, Anderson DW (2004) Rapid synthesis of acyl transfer auxiliaries for cysteine-free native glycopeptide ligation. Org Lett 6:4659–4662
49. Wu B, Chen J, Warren JD, Chen G, Hua Z, Danishefsky SJ (2006) Building complex glycopeptides: development of a cysteine-free native chemical ligation protocol. Angew Chem Int Ed 45:4116–4125
50. Chen J, Chen G, Wu B, Wan Q, Tan Z, Hua Z, Danishefsky SJ (2006) Mature homogeneous erythropoietin-level building blocks by chemical synthesis: the EPO 114–166 glycopeptide domain, presenting the O-linked glycophorin. Tetrahedron Lett 47:8013–8016
51. Kumar KS, Brik A (2010) Accessing posttranslationally modified proteins through thiol positioning. J Pept Sci 16:524–529
52. Brik A, Yang YY, Ficht S, Wong CH (2006) Sugar-assisted glycopeptide ligation. J Am Chem Soc 128:5626–5627
53. Brik A, Wong CH (2007) Sugar-assisted ligation for the synthesis of glycopeptides. Chemistry 13:5670–5675
54. Yan LZ, Dawson PE (2001) Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. J Am Chem Soc 123:526–533
55. Brik A, Ficht S, Yang YY, Bennett CS, Wong CH (2006) Sugar-assisted ligation of N-linked glycopeptides with broad sequence tolerance at the ligation junction. J Am Chem Soc 128:15026–15033
56. Yang YY, Ficht S, Brik A, Wong CH (2007) Sugar-assisted glycoprotein synthesis. J Am Chem Soc 129:7690–7701
57. Bennett CS, Dean SM, Payne RJ, Ficht S, Brik A, Wong CH (2008) Sugar-assisted glycopeptide ligation with complex oligosaccharides: scope and limitations. J Am Chem Soc 130:11945–11952
58. Pentelute BL, Kent SBH (2007) Selective desulfurization of cysteine in the presence of Cys (Acm) in polypeptides obtained by native chemical ligation. Org Lett 9:687–690
59. Ficht S, Payne RJ, Brik A, Wong CH (2007) Second-generation sugar-assisted ligation: a method for the synthesis of cysteine-containing glycopeptides. Angew Chem Int Ed 46:5975–5979
60. Lutsky MY, Nepomniaschiy N, Brik A (2008) Peptide ligation via side-chain auxiliary. Chem Commun 10:1229–1231

61. Ajish Kumar KS, Harpaz Z, Haj-Yahya M, Brik A (2009) Side-chain assisted ligation in protein synthesis. Bioorg Med Chem Lett 19:3870–3874
62. Payne RJ, Ficht S, Tang S, Brik A, Yang YY, Case DA, Wong CH (2007) Extended sugar-assisted glycopeptide ligations: development, scope and applications. J Am Chem Soc 129 (44):13527–13536
63. Payne RJ, Ficht S, Greenberg WA, Wong CH (2008) Cysteine-free peptide and glycopeptide ligation by direct aminolysis. Angew Chem Int Ed 47:4411–4415
64. Thomas GL, Hsieh YSY, Chun CKY, Cai ZL, Reimers JR, Payne RJ (2011) Peptide ligations accelerated by N-terminal aspartate and glutamate residues. Org Lett 13:4770–4773
65. Hojo H, Ozawa C, Katayama H, Ueki A, Nakahara Y, Nakahara Y (2010) The mercaptomethyl group facilitates an efficient one-pot ligation at Xaa-Ser/Thr for (glyco) peptide synthesis. Angew Chem Int Ed 49:5318–5321
66. Chalker JM, Bernardes GJL, Lin YA, Davis BG (2009) Chemical modification of proteins at cysteine: opportunities in chemistry and biology. Chem Asian J 4:630–640
67. Okamoto R, Kajihara Y (2008) Uncovering a latent ligation site for glycopeptide synthesis. Angew Chem Int Ed 47:5402–5406
68. Heinrikson RL (1970) Selective S-methylation of cysteine in proteins and peptides. Biochem Biophys Res Commun 41:967–972
69. Gross E, Morell JL (1974) The reaction of cyanogen bromide with S-methylcysteine: fragmentation of the peptide 14–29 of bovine pancreatic ribonuclease A. Biochem Biophys Res Commun 59:1145–1150
70. Tam JP, Yu Q (1998) Methionine ligation strategy in the biomimetic synthesis of parathyroid hormones. Biopolymers 46:319–327
71. Haase C, Rohde H, Seitz O (2008) Native chemical ligation at valine. Angew Chem Int Ed 47:6807–6810
72. Wan Q, Danishefsky SJ (2007) Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. Angew Chem Int Ed 46:9248–9252
73. Hoffmann FW, Ess RJ, Simmons TC, Hanzel RS (1956) The desulfurization of mercaptans with trialkyl phosphites. J Am Chem Soc 78:6414
74. Walling C, Basedow OH, Savas ES (1960) Some extensions of the reaction of trivalent phosphorus derivatives with alkoxy and thiyl radicals; a new synthesis of thioesters. J Am Chem Soc 82:2181–2184
75. Walling C, Rabinowitz R (1957) The reaction of thiyl radicals with trialkyl phosphites. J Am Chem Soc 79:5326
76. Rohde H, Seitz O (2010) Ligation-desulfurization: a powerful combination in the synthesis of peptides and glycopeptides. Biopolymers 94:551–559
77. Dawson PE (2011) Native chemical ligation combined with desulfurization and deselenization: a general strategy for chemical protein synthesis. Isr J Chem 51:862–867
78. Chalker JM (2013) Prospects in the total synthesis of protein therapeutics. Chem Biol Drug Des 81:122–135
79. Johnson EC, Malito E, Shen Y, Rich D, Tang WJ, Kent SB (2007) Modular total chemical synthesis of a human immunodeficiency virus type 1 protease. J Am Chem Soc 129:11480–11490
80. Zheng J-S, Tang S, Qi Y-K, Wang Z-P, Liu L (2013) Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. Nat Protocols 8:2483–2495
81. Fang GM, Li YM, Shen F, Huang YC, Li JB, Lin Y, Cui HK, Liu L (2011) Protein chemical synthesis by ligation of peptide hydrazides. Angew Chem Int Ed 50:7645–7649
82. Fang GM, Wang JX, Liu L (2012) Convergent chemical synthesis of proteins by ligation of peptide hydrazides. Angew Chem Int Ed 51:10347–10350
83. Wilkinson BL, Stone RS, Capicciotti CJ, Thaysen-Andersen M, Matthews JM, Packer NH, Ben RN, Payne RJ (2012) Total synthesis of homogeneous antifreeze glycopeptides and glycoproteins. Angew Chem Int Ed 51:3606–3610

84. Garner J, Harding MM (2010) Design and synthesis of antifreeze glycoproteins and mimics. ChemBioChem 11:2489–2498
85. Peltier R, Brimble MA, Wojnar JM, Williams DE, Evans CW, DeVries AL (2010) Synthesis and antifreeze activity of fish antifreeze glycoproteins and their analogs. Chem Sci 1:538–551
86. Sakamoto I, Tezuka K, Fukae K, Ishii K, Taduru K, Maeda M, Ouchi M, Yoshida K, Nambu Y, Igarashi J, Hayashi N, Tsuji T, Kajihara Y (2012) Chemical synthesis of homogeneous human glycosyl-interferon-beta that exhibits potent antitumor activity in vivo. J Am Chem Soc 134:5428–5431
87. Wang P, Dong S, Brailsford JA, Iyer K, Townsend SD, Zhang Q, Hendrickson RC, Shieh J, Moore MAS, Danishefsky SJ (2012) At last: erythropoietin as a single glycoform. Angew Chem Int Ed 51:11576–11584
88. Wang P, Dong S, Shieh J-H, Peguero E, Hendrickson R, Moore MAS, Danishefsky SJ (2013) Erythropoietin derived by chemical synthesis. Science 342:1357–1360
89. Murakami M, Okamoto R, Izumi M, Kajihara Y (2012) Chemical synthesis of an erythropoietin glycoform containing a complex-type disialyloligosaccharide. Angew Chem Int Ed 51:3567–3572
90. Liu S, Pentelute BL, Kent SBH (2012) Convergent chemical synthesis of [lysine24, 38, 83] human erythropoietin. Angew Chem Int Ed 51:993–999
91. Payne RJ (2013) Total synthesis of erythropoietin through the development and exploitation of enabling synthetic technologies. Angew Chem Int Ed 52:505–507
92. Warren JD, Miller JS, Keding SJ, Danishefsky SJ (2004) Toward fully synthetic glycoproteins by ultimately convergent routes: a solution to a long-standing problem. J Am Chem Soc 126:6576–6578
93. Chen G, Warren JD, Chen J, Wu B, Wan Q, Danishefsky SJ (2006) Studies related to the relative thermodynamic stability of C-terminal peptidyl esters of O-hydroxy thiophenol: emergence of a doable strategy for non-cysteine ligation applicable to the chemical synthesis of glycopeptides. J Am Chem Soc 128:7460–7462
94. Wong CTT, Tung CL, Li X (2013) Synthetic cysteine surrogates used in native chemical ligation. Mol Biosyst 9:826–833
95. Botti P, Tchertchian S (2006) Side chain extended ligation. WO/2006/133962
96. Crich D, Banerjee A (2007) Native chemical ligation at phenylalanine. J Am Chem Soc 129:10064–10065
97. Easton CJ, Hutton CA, Tan EW, Tiekink ERT (1990) Synthesis of homochiral hydroxy-α-amino acid derivatives. Tetrahedron Lett 31:7059–7062
98. Easton CJ, Hutton CA, Roselt PD, Tiekink ERT (1994) Stereocontrolled synthesis of β-hydroxyphenylalanine and β-hydroxytyrosine derivatives. Tetrahedron 50:7327–7340
99. Crich D, Banerjee A (2006) Expedient synthesis of threo-beta-hydroxy-alpha-amino acid derivatives: phenylalanine, tyrosine, histidine, and tryptophan. J Org Chem 71:7106–7109
100. Chen J, Wan Q, Yuan Y, Zhu JL, Danishefsky SJ (2008) Native chemical ligation at valine: a contribution to peptide and glycopeptide synthesis. Angew Chem Int Ed 47:8521–8524
101. Chen J, Wang P, Zhu JL, Wan Q, Danishefsky SJ (2010) A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides. Tetrahedron 66:2277–2283
102. Yang RL, Pasunooti KK, Li FP, Liu XW, Liu CF (2009) Dual native chemical ligation at lysine. J Am Chem Soc 131:13592–13593
103. Marin J, Didierjean C, Aubry A, Casimir JR, Briand JP, Guichard G (2004) Synthesis of enantiopure 4-hydroxypipecolate and 4-hydroxylysine derivatives from a common 4,6-dioxopiperidinecarboxylate precursor. J Org Chem 69:130–141
104. Ajish Kumar KS, Haj-Yahya M, Olschewski D, Lashuel HA, Brik A (2009) Highly efficient and chemoselective peptide ubiquitylation. Angew Chem Int Ed 48:8090–8094
105. El Oualid F, Merkx R, Ekkebus R, Hameed DS, Smit JJ, de Jong A, Hilkmann H, Sixma TK, Ovaa H (2010) Chemical synthesis of ubiquitin, ubiquitin-based probes, and diubiquitin. Angew Chem Int Ed 49:10149–10153

106. Hejjaoui M, Haj-Yahya M, Kumar KS, Brik A, Lashuel HA (2011) Towards elucidation of the role of ubiquitination in the pathogenesis of Parkinson's disease with semisynthetic ubiquitinated alpha-synuclein. Angew Chem Int Ed 50:405–409
107. Merkx R, de Bruin G, Kruithof A, van den Bergh T, Snip E, Lutz M, El Oualid F, Ovaa H (2013) Scalable synthesis of γ-thiolysine starting from lysine and a side by side comparison with δ-thiolysine in non-enzymatic ubiquitination. Chem Sci 4:4494–4498
108. Kumar KS, Bavikar SN, Spasser L, Moyal T, Ohayon S, Brik A (2011) Total chemical synthesis of a 304 amino acid K48-linked tetraubiquitin protein. Angew Chem Int Ed 50:6137–6141
109. Tan ZP, Shang SY, Danishefsky SJ (2010) Insights into the finer issues of native chemical ligation: an approach to cascade ligations. Angew Chem Int Ed 49:9500–9503
110. Harpaz Z, Siman P, Kumar KSA, Brik A (2010) Protein synthesis assisted by native chemical ligation at leucine. ChemBioChem 11:1232–1235
111. Shang SY, Tan ZP, Dong SW, Danishefsky SJ (2011) An advance in proline ligation. J Am Chem Soc 133:10784–10786
112. Dong S, Shang S, Tan Z, Danishefsky SJ (2011) Toward homogeneous erythropoietin: application of metal free dethiylation in the chemical synthesis of the Ala79-Arg166 glycopeptide domain. Isr J Chem 51:968–976
113. Townsend SD, Tan Z, Dong S, Shang S, Brailsford JA, Danishefsky SJ (2012) Advances in proline ligation. J Am Chem Soc 134:3912–3916
114. Ding H, Shigenaga A, Sato K, Morishita K, Otaka A (2011) Dual kinetically controlled native chemical ligation using a combination of sulfanylproline and sulfanylethylanilide peptide. Org Lett 13:5588–5591
115. Raibaut L, El Mahdi O, Melnyk O (2014) Solid phase protein chemical synthesis. Topics Curr Chem. doi:10.1007/128_2014_609
116. Siman P, Karthikeyan SV, Brik A (2012) Native chemical ligation at glutamine. Org Lett 14:1520–1523
117. Garner P (1984) Stereocontrolled addition to a penaldic acid equivalent: an asymmetric synthesis of threo-β-hydroxy-L-glutamic acid. Tetrahedron Lett 25:5855–5858
118. Malins LR, Cergol KM, Payne RJ (2013) Peptide ligation-desulfurization chemistry at arginine. ChemBioChem 14:559–563
119. Bang D, Pentelute BL, Kent SB (2006) Kinetically controlled ligation for the convergent chemical synthesis of proteins. Angew Chem Int Ed 45:3985–3988
120. Durek T, Torbeev VY, Kent SB (2007) Convergent chemical synthesis and high-resolution X-ray structure of human lysozyme. Proc Natl Acad Sci U S A 104:4846–4851
121. Torbeev VY, Kent SB (2007) Convergent chemical synthesis and crystal structure of a 203 amino acid "covalent dimer" HIV-1 protease enzyme molecule. Angew Chem Int Ed 46:1667–1670
122. Thompson RE, Chan B, Radom L, Jolliffe KA, Payne RJ (2013) Chemoselective peptide ligation-desulfurization at aspartate. Angew Chem Int Ed 52:9723–9727
123. Guan X, Drake MR, Tan Z (2013) Total synthesis of human galanin-like peptide through an aspartic acid ligation. Org Lett 15:6128–6131
124. Thompson RE, Liu X, Alonso-García N, Pereira PJB, Jolliffe KA, Payne RJ (2014) Trifluoroethanethiol: an additive for efficient one-pot peptide ligation – desulfurization chemistry. J Am Chem Soc 136:8161–8164
125. Rohde H, Schmalisch J, Harpaz Z, Diezmann F, Seitz O (2011) Ascorbate as an alternative to thiol additives in native chemical ligation. ChemBioChem 12:1396–1400
126. Moyal T, Hemantha HP, Siman P, Refua M, Brik A (2013) Highly efficient one-pot ligation and desulfurization. Chem Sci 4:2496–2501
127. Siman P, Blatt O, Moyal T, Danieli T, Lebendiker M, Lashuel HA, Friedler A, Brik A (2011) Chemical synthesis and expression of the HIV-1 Rev protein. ChemBioChem 12:1097–1104
128. Cergol KM, Thompson RE, Malins LR, Turner P, Payne RJ (2014) One-pot peptide ligation-desulfurization at glutamate. Org Lett 16:290–293

129. Malins LR, Cergol KM, Payne RJ (2014) Chemoselective sulfenylation and peptide ligation at tryptophan. Chem Sci 5:260–266
130. Scoffone E, Fontana A, Rocchi R (1966) Selective modification of tryptophan residue in peptides and proteins using sulfenyl halides. Biochem Biophys Res Commun 25:170–174
131. Scoffone E, Fontana A, Rocchi R (1968) Sulfenyl halides as modifying reagents for polypeptides and proteins. I. Modification of tryptophan residues. Biochemistry 7:971–979
132. Wilchek M, Miron T (1972) Conversion of tryptophan to 2-thioltryptophan in peptides and proteins. Biochem Biophys Res Commun 47:1015–1020
133. Shang S, Tan Z, Danishefsky SJ (2011) Application of the logic of cysteine-free native chemical ligation to the synthesis of human parathyroid hormone (hPTH). Proc Natl Acad Sci U S A 108:5986–5989
134. Bock A, Forchhammer K, Heider J, Leinfelder W, Sawers G, Veprek B, Zinoni F (1991) Selenocysteine: the 21st amino acid. Mol Microbiol 5:515–520
135. Lobanov AV, Hatfield DL, Gladyshev VN (2009) Eukaryotic selenoproteins and selenoproteomes. Biochim Biophys Acta 1790:1424–1428
136. Gieselman MD, Xie L, van Der Donk WA (2001) Synthesis of a selenocysteine-containing peptide by native chemical ligation. Org Lett 3:1331–1334
137. Quaderer R, Sewing A, Hilvert D (2001) Selenocysteine-mediated native chemical ligation. Helv Chim Acta 84:1197–1206
138. Hondal RJ, Nilsson BL, Raines RT (2001) Selenocysteine in native chemical ligation and expressed protein ligation. J Am Chem Soc 123:5140–5141
139. Huber R, Criddle RS (1967) Comparison of the chemical properties of selenocysteine and selenocystine with their sulfur analogs. Arch Biochem Biophys 122:164–173
140. Arnold AP, Tan KS, Rabenstein DL (1986) Nuclear magnetic resonance studies of the solution chemistry of metal complexes. 23. Complexation of methylmercury by selenohydryl-containing amino acids and related molecules. Inorg Chem 25:2433–2437
141. Pleasants JC, Guo W, Rabenstein DL (1989) A comparative study of the kinetics of selenol/diselenide and thiol/disulfide exchange reactions. J Am Chem Soc 111:6553–6558
142. Besse D, Siedler F, Diercks T, Kessler H, Moroder L (1997) The redox potential of selenocystine in unconstrained cyclic peptides. Angew Chem Int Ed 36:883–885
143. Nauser T, Dockheer S, Kissner R, Koppenol WH (2006) Catalysis of electron transfer by selenocysteine. Biochemistry 45:6038–6043
144. Metanis N, Keinan E, Dawson PE (2010) Traceless ligation of cysteine peptides using selective deselenization. Angew Chem Int Ed 49:7049–7053
145. Berry SM, Gieselman MD, Nilges MJ, Van der Donk WA, Lu Y (2002) An engineered azurin variant containing a selenocysteine copper ligand. J Am Chem Soc 124:2084–2085
146. Quaderer R, Hilvert D (2002) Selenocysteine-mediated backbone cyclization of unprotected peptides followed by alkylation, oxidative elimination or reduction of the selenol. Chem Commun 2620–2621
147. Malins LR, Mitchell NJ, Payne RJ (2014) Peptide ligation chemistry at selenol amino acids. J Pept Sci 20:64–77
148. Chalker JM, Bernardes GJL, Davis BG (2011) A "tag-and-modify" approach to site-selective protein modification. Acc Chem Res 44:730–741
149. Malins LR, Payne RJ (2012) Synthesis and utility of beta-selenol-phenylalanine for native chemical ligation-deselenization chemistry. Org Lett 14:3142–3145

Top Curr Chem (2015) 362: 89–106
DOI: 10.1007/128_2014_613

Published online: 10 April 2015

Chemical Methods for Protein Ubiquitination

Renliang Yang and Chuan-Fa Liu

Abstract In eukaryotic cells, many proteins undergo extensive post-translational modifications (PTMs) such as methylation, acetylation, phosphorylation, glycosylation, and ubiquitination. Among these, ubiquitination is a particularly interesting PTM from both structural and functional viewpoints. In ubiquitination, the C-terminal carboxyl group of the small ubiquitin protein is attached to the ε-amine of a lysine residue of a substrate protein through an isopeptide bond. Ubiquitination has been shown to be involved in the regulation of many cellular processes including protein degradation and gene expression. And dysfunction of these processes is implicated in many human diseases. Despite many years of intensive research, a large number of protein ubquitination events remain poorly characterized. The challenge lies with the tremendous difficulties in isolating homogeneously modified proteins from biological samples for structural and functional studies. Enzymatic ubiquitination in vitro often has limited practical value due to the large number of substrate-specific E3 ligases and the difficulties in identifying or isolating these enzymes. Chemical approaches to the preparation of ubiquitinated proteins provide a powerful solution, and the development of such approaches has been the subject of intense research by many research laboratories. This review summarizes the methodological developments of protein chemical ubiquitination in recent years.

Keywords Post-translational modification · Ubiquitin · Ubiquitination · Site-specific modification · Peptide ligation

R. Yang and C.-F. Liu (✉)
School of Biological Sciences, Nanyang Technological University, Singapore, Singapore
e-mail: cfliu@ntu.edu.sg

Contents

1 Introduction

As a fundamental mechanism to modulate protein function in the cell, post-translational modification (PTM) regulates almost all biological processes [1, 2]. The side chain ε-amine of lysine provides a platform for several important PTMs such as methylation, acetylation, and ubiquitination [1, 2]. Among these, ubiquitination is a particularly interesting PTM from both structural and functional viewpoints [3–9]. In contrast to most other PTMs such as phosphorylation, methylation, and acetylation, in which the modifying groups are rather simple, ubiquitination involves the transfer of an entire, albeit small, protein onto a protein substrate [3–9]. Ubiquitin is a 76-amino acid protein ubiquitously distributed in all tissues of eukaryotic organisms. In the post-translational ubiquitination process, the C-terminal carboxyl group of ubiquitin becomes attached to the ε-amine of a lysine residue of the substrate protein through an isopeptide bond [3–9]. More interestingly, as any of the seven lysine residues or the amino terminus of ubiquitin can still accept another ubiquitin molecule, the modified proteins often carries polyubiquitin chains of the same or mixed linkages [9]. This gives a great number of possibilities to ubiquitinating cellular proteins. These different ubiquitin modifications have different structural effects which are interpreted by the various effector proteins in the cell and thereby impart many different outcomes to the targeted proteins [9]. For example, K48 polyubiquitination signals for their degradation via the proteasome system, whereas other forms of polyubiquitination or monoubiquitination can change their function, alter their cellular location or trafficking, or inhibit/enhance their interactions with other proteins in a signaling pathway [10, 11]. Biologically, ubiquitination is achieved through the consecutive action of three enzymes – ubiquitin-activating enzymes (E1), ubiquitin-conjugating enzymes (E2), and ubiquitin ligases (E3) [3–9]. As with other PTMs, ubiquitination is reversible. The biochemical processes of ubiquitination and deubiquitination are tightly controlled, and their dysfunction is implicated in many human diseases. However, despite many years of intense research, a large number of protein ubquitination events remain poorly characterized. A great challenge in the structural and functional analysis of this PTM lies with the tremendous difficulties in isolating

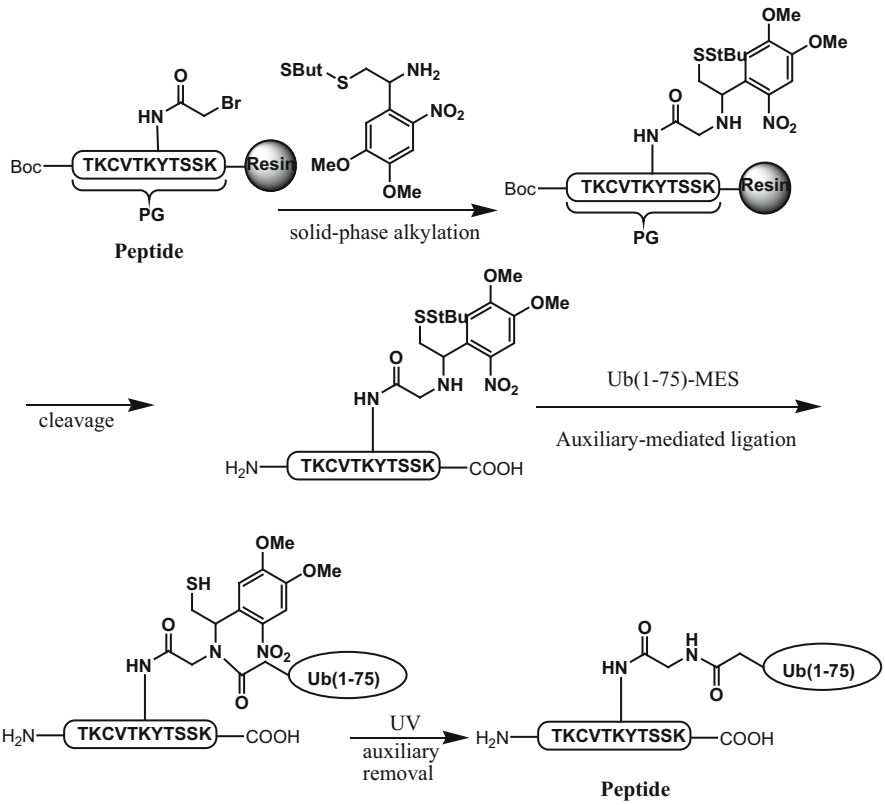
homogeneously modified proteins from biological samples. Because of the difficulties in identifying or isolating the substrate specific E3 ligases, enzymatic ubiquitination in vitro often has limited practical value. Chemical approaches to the preparation of ubiquitinated proteins provide a powerful solution, and the development of such approaches has been the subject of intense research by many research laboratories. This review summarizes the methodological developments of protein chemical ubiquitination in recent years.

2 Preparation of Ub-Proteins with Native Isopeptide Bonds

2.1 Nα-Auxiliary-Mediated Ubiquitination

In 2007, Muir et al. reported the first method of chemical ubiquitination of a synthetic peptide through Nα-auxiliary-mediated ligation [12]. In this method, site-specific ubiquitination was achieved through the ligation at Gly75–Gly76 junction between a ubiquitin(1–75)-thioester and the side chain of a lysine residue pre-acylated by Gly76 bearing an Nα ligation auxiliary in the substrate peptide. After ligation, the photolabile auxiliary was removed by UV irradiation to generate the native isopeptide linkage. With this method, they initially synthesized a ubiquitinated peptide NH_2-**TKCVTKYTSSK**-COOH corresponding to the residues 115–125 of human histone H2B (Scheme 1). Later, they also applied this strategy to the synthesis of full-length Lys120-ubiquitinated H2B via multi-step ligations (Scheme 2) [13]. This method represents the first chemical method for the synthesis of native ubiquitinated peptides/proteins. The only limitation of the method is that the auxiliary-mediated ligation is sluggish as it occurs at a sterically hindered secondary amine.

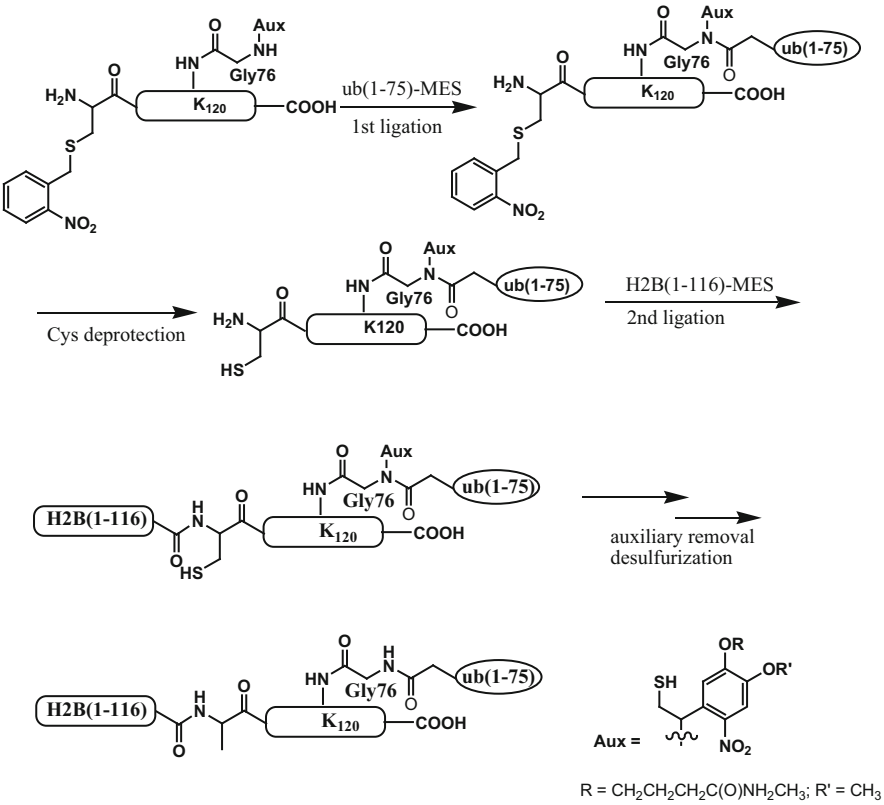
In Muir's auxiliary-mediated ubiquitination approach, the auxiliary has to be introduced onto peptides during solid-phase peptide synthesis (SPPS). When using this method for the synthesis of ubiquitinated proteins, one or more ligation steps may be required to construct the full-length substrate protein bearing the auxiliary. Recently, our group developed an auxiliary-mediated ubiquitination method in which the ligation auxiliary could be introduced onto recombinant proteins genetically (Scheme 3) [14]. In our method, the lysine involved in ubiquitination was first genetically incorporated as azidonorleucine (Anl) into the recombinant protein of interest as mediated by an evolved methionyl-tRNA synthetase. After protecting all the amines with *t*-Boc, Anl was reduced to lysine under mild conditions. A Gly derivative bearing the ligation auxiliary at its α-amine was installed on this lysine side chain through the active OSu ester. Site-specific ubiquitination was then achieved through auxiliary-mediated ligation between ub(1–75)-thioester and the Boc-deprotected acceptor protein. After ligation, the auxiliary group was removed by treating the ligation product with a cocktail of trifluoroacetic acid/triisopropylsilane/H_2O (95/2.5/2.5) (Scheme 3).



Scheme 1 General scheme of Nα-auxiliary-mediated ubiquitination

Besides the synthesis of ubiquitinated peptides and proteins, the Nα-auxiliary-mediated ubiquitination method has also been used for the synthesis of ubiquitin C-terminal derivatives. Recently, Liu et al. reported the semi-synthesis of ubiquitin C-terminal conjugate of 7-amino-4-methylcoumarin (ub-AMC) with such a method [15]. First, a ubiquitin hydrazide with residues 1–75 was generated through hydrazinolysis of a ubiquitin-intein fusion protein. The ub(1–75)-hydrazide was then ligated with Nα-(auxiliary)-Gly-AMC through Nα-auxiliary-mediated ligation. After ligation, the auxiliary was removed to generate the ub-AMC with a native peptide bond (Scheme 4).

The aforementioned Nα-auxiliary-mediated ubiquitination methods all employed substituted benzylamine-derived auxiliaries. Recently, Chatterjee et al. reported an auxiliary-mediated ubiquitination method employing 2-mercaptoethoxyl group as the auxiliary [16]. The 2-mercaptoethoxyl auxiliary was introduced to the α-amine of Gly76 which had been attached to the lysine side chain during SPPS. The ubiquitination was achieved through the reaction between

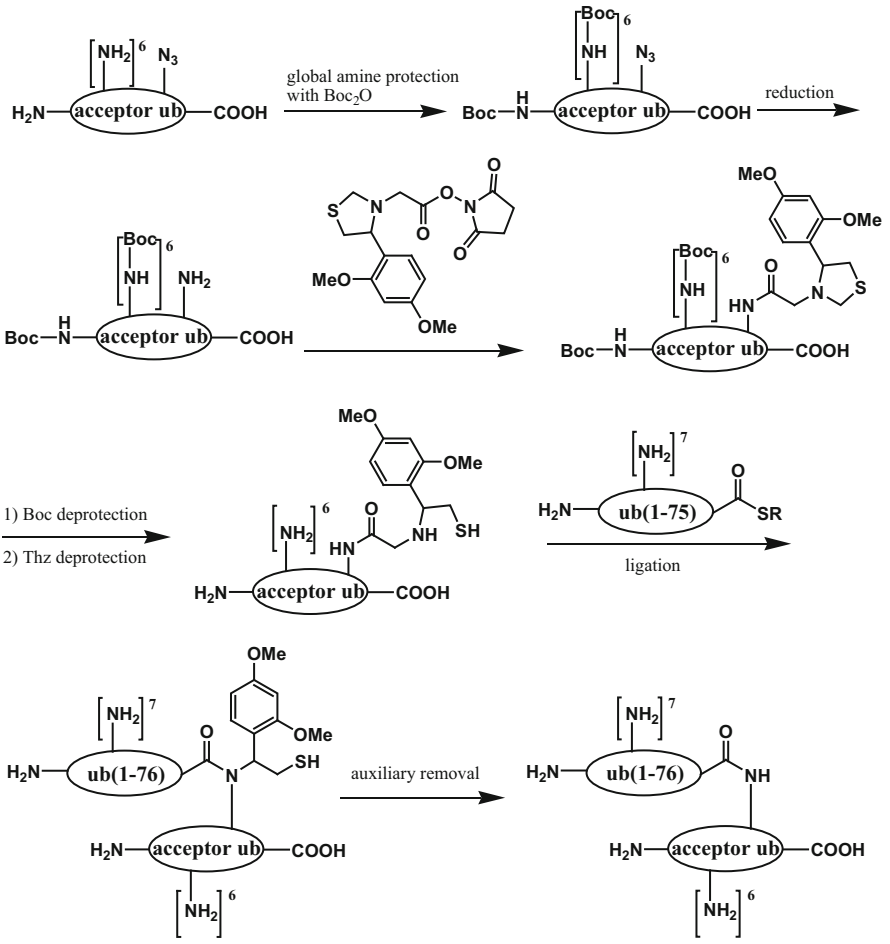


Scheme 2 General scheme for the synthesis of K120 ubiquitinated human H2B through N-α-auxiliary-mediated ubiquitination

ub(1–75)-thioester and the peptide with the auxiliary. After ligation, the auxiliary could be removed by activated Zn in acidic HPLC buffer containing 6 M guanidine hydrochloride (Scheme 5).

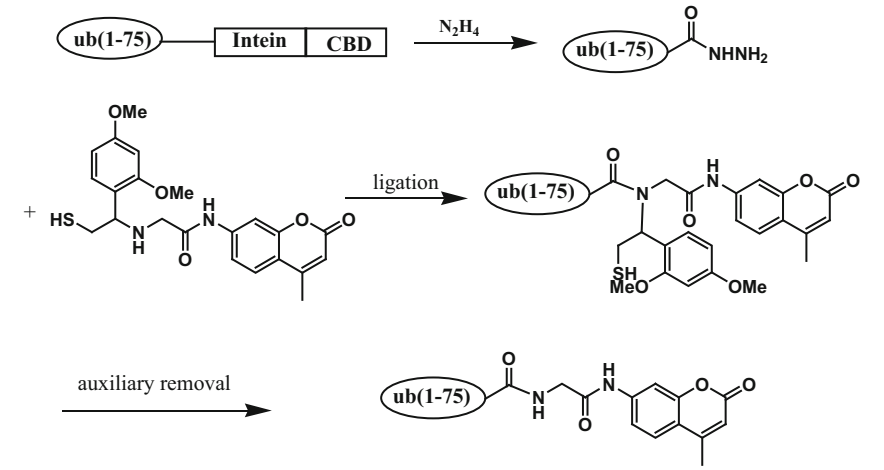
2.2 Thiolated-Lysine-Mediated Ubiquitination

Since the development of the Nα-auxiliary-mediated ubiquitination method, many researchers have been actively seeking the development of alternative and more efficient chemical ubiquitination methods. Our group [17, 18] and Brik's group [19–21] have independently reported the γ- and δ-thiolysine-mediated ubiquitination methods, respectively. In these two approaches, site-specific ubiquitination was achieved through chemical ligation involving full-length ubiquitin(1–76) thioester and the ε-amine of a lysine residue in the peptide

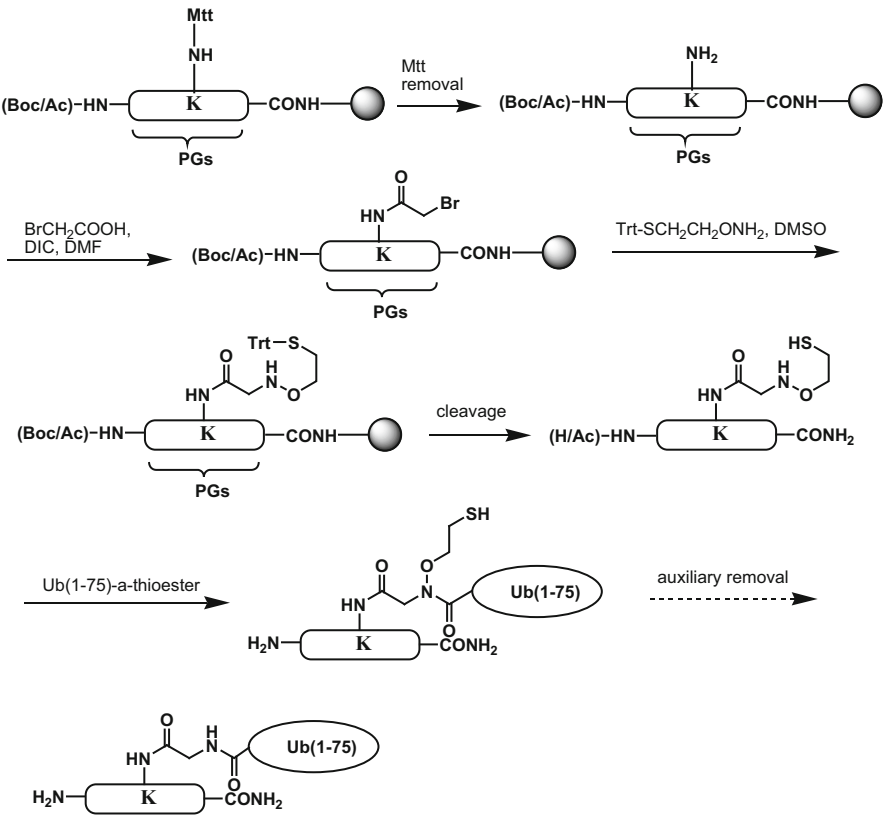


Scheme 3 General scheme for the synthesis of linkage-specific diubiquitins through genetic incorporation of azidonorleucine and Nα-auxiliary mediated ligation

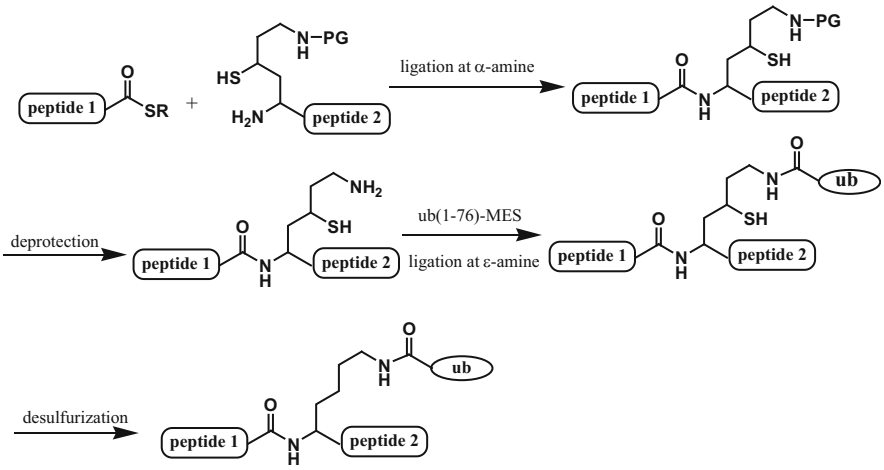
or protein substrate. In contrast to the Nα-auxiliary-mediated ubiquitination approach, this ligation process was mediated by a simple thiol group located at either the γ- (our approach) or δ- (Brik's approach) position of the lysine side chain. As the ligation occurred at the non-sterically hindered primary ε-amine, the reaction was very efficient (Schemes 6 and 7). With these methods, both ubiquitinated peptides and proteins were synthesized. Chin's group also reported a method for genetic incorporation of δ-thiolysine which therefore allows for ubiquitination of recombinant proteins [22]. The only limitation of the thiolated-lysine methods was that it took many steps to synthesize the unnatural thiolysine derivatives. Recently, Ovaa's group reported a simplified method for the



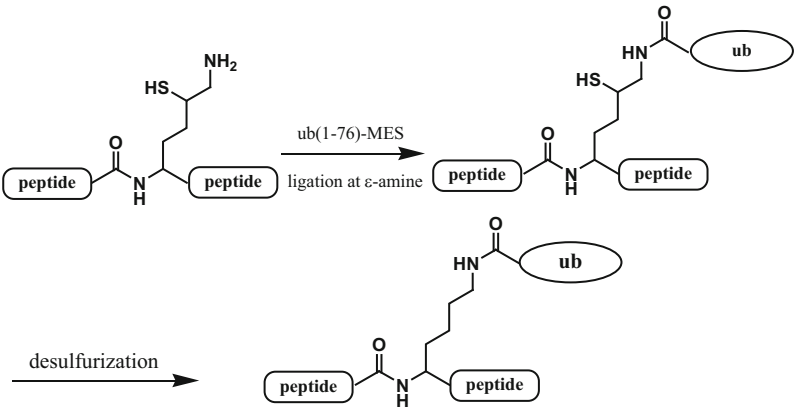
Scheme 4 Synthesis of ub-AMC through Nα-auxiliary-mediated ligation



Scheme 5 Nα-auxiliary-mediated ubiquitination employing 2-mercaptoethoxyl auxiliary



Scheme 6 General scheme of dual native chemical ligation at lysine for peptide ubiquitination



Scheme 7 General scheme of chemical ubiquitination through δ-thiolysine mediated ligation

synthesis of γ-thiolysines [23, 24]. This enabled the scale-up production of γ-thiolysine and can significantly enhance the application of our γ-thiolysine-mediated ubiquitination approach.

2.3 *Ubiquitination Through Ag^{+}-Mediated Activation of Ubiquitin C-Terminal Thioester*

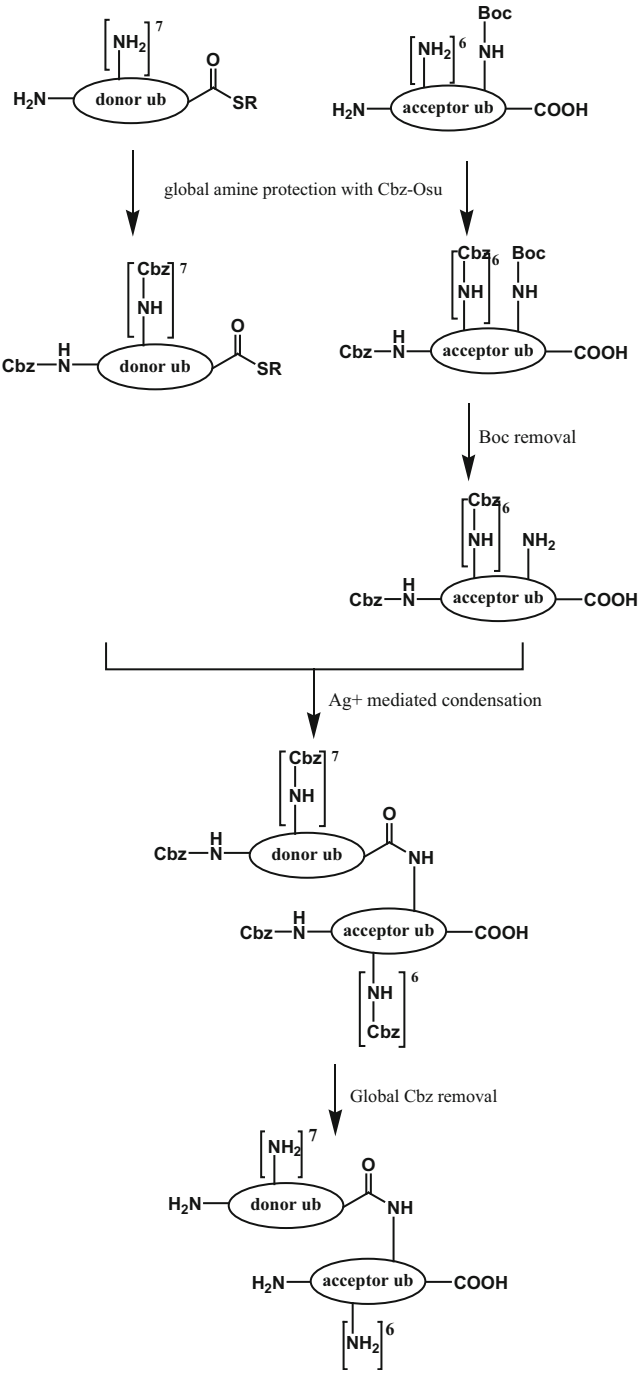
All these chemical ubiquitination methods mentioned above were through the chemoselective ligation between unprotected peptides or proteins. Recently, Chin

et al. reported the genetically encoded orthogonal protection and activated ligation (GOPAL) approach for the synthesis of K6- and K29-linked diubiquitins [25]. To synthesize diubiquitins using the GOPAL approach, the isopeptide bond was formed through Ag^{+}-mediated condensation between two partially protected (with Cbz) ubiquitin proteins (Scheme 8). One was a full-length ubiquitin thioester with all its amino groups protected. The other was the acceptor ubiquitin with all the amine protected except the ε-amine of the lysine where ubiquitination occurred. As the condensation occurred between two relatively large partially protected proteins, the reaction was less efficient than those chemical ligation approaches. For the acceptor ubiquitin, the lysine residue to be ubiquitinated was introduced genetically as Lys(Boc) which was deprotected after all other amines were protected with Cbz. More recently, Cropp and Fushman modified the approach and applied it to the synthesis of more complicated oligo-ubiquitins (Schemes 9 and 10) [26]. In their approach, Alloc, instead of Cbz, was used as the protecting group for the amines which were not involved in ubiquitination. At the end of the ubiquitin linkage construction, Alloc group could be globally removed by a ruthenium complex instead of strong acid treatment when Cbz was used as the protecting group. Another important modification was that the E1 ubiquitin activating enzyme was employed to thioesterify the C-terminal COOH of mono-, di-, or tri-ubiquitin to generate the reactive intermediates. Different from the protein splicing approach used by Chin, where a monoubiquitin thioester was generated through thiolysis of ubiquitin-intein fusion protein, their method could generate polyubiquitin thioesters, which was crucial for polyubiquitin construction.

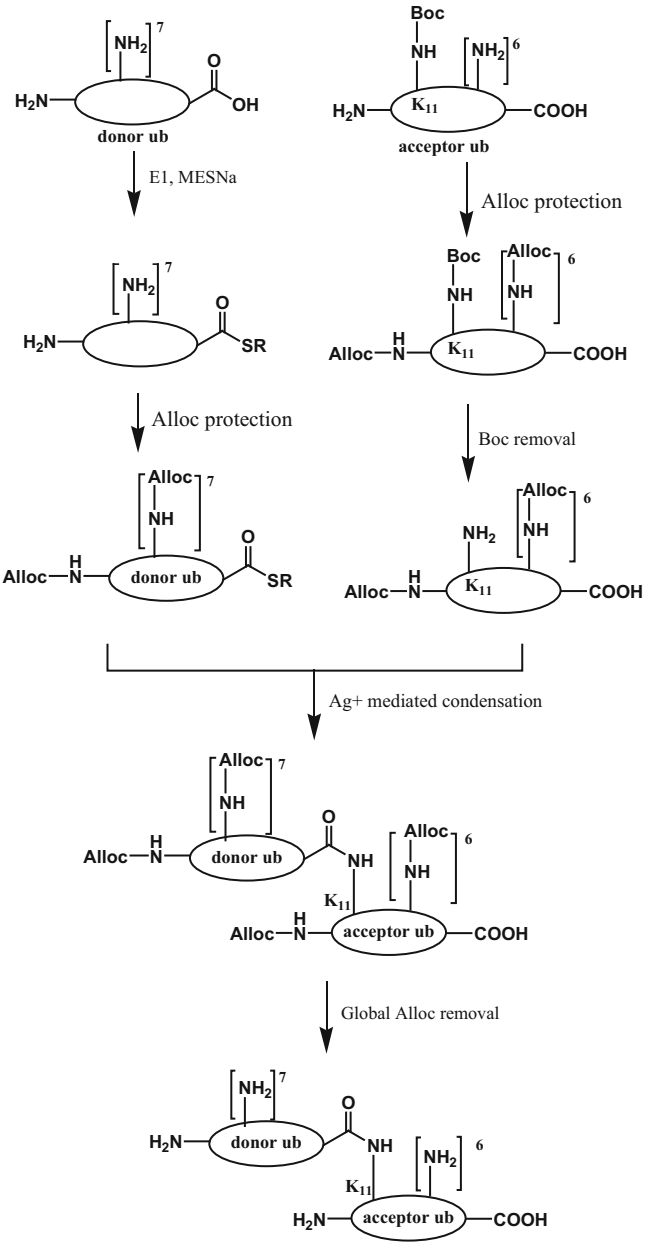
3 Preparation of Ub-Proteins with Non-Native Linkages

Synthesis of ubiquitinated peptides and proteins with the native isopeptide linkage is very challenging. During the development of chemical ubiquitination methods, researchers frequently need to balance the necessity for nativity of the ubiquitin linkage against the labor requirements of the synthetic task. Many relatively simple methods to generate non-native or mimetic ubiquitin conjugates have therefore been developed.

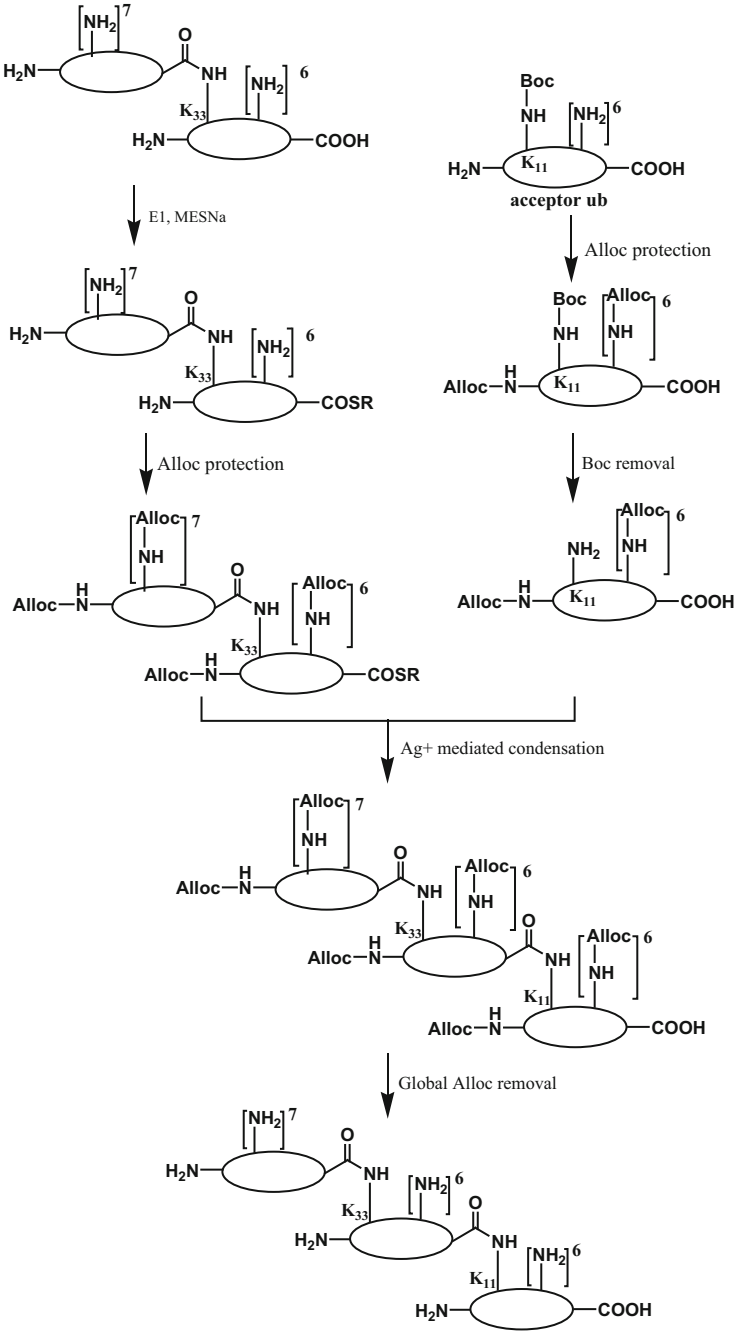
Ubiquitin conjugates with Gly76 to D-Cys or Ala mutation were synthesized by Chan's and Muir's groups, respectively (Scheme 11) [27–29]. Chan et al. reported the genetic incorporation of Nε-(D-cysteinyl)-L-lysine as an unnatural amino acid, which was mediated by the pyrrolysyl-tRNA synthetase and tRNA pair [27]. The Cys residue was employed to mediate native chemical ligation between ub(1–75)-thioester and the cysteinyl-lysine side chain on the acceptor ubiquitin. In Chan's method, the D-Cys was left unchanged after ligation. Muir's group attached the L-cysteine to the lysine side chain amine through solid phase peptide synthesis. After ligation, the thiol of cysteine residue could be converted to alanine through desulfurization [28, 29].



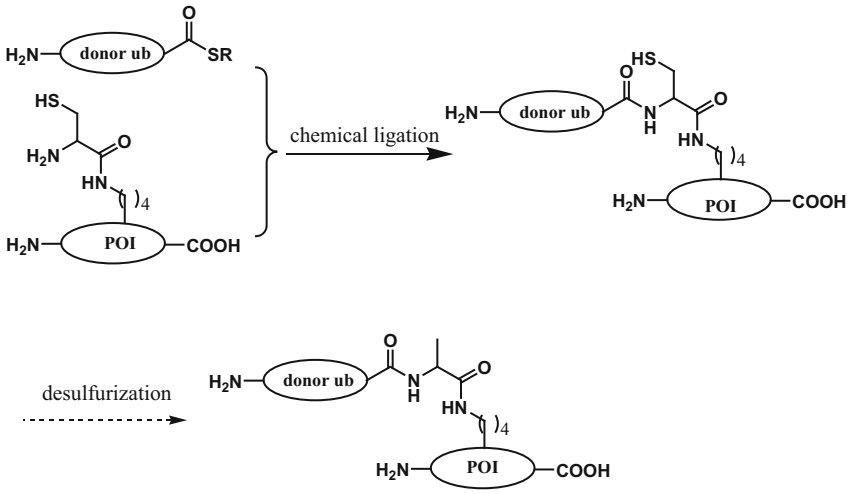
Scheme 8 General scheme of GOPAL approach for the synthesis of linkage-specific diubiquitins



Scheme 9 General scheme for the synthesis of K11-linked diubiquitin using the revised GOPAL approach



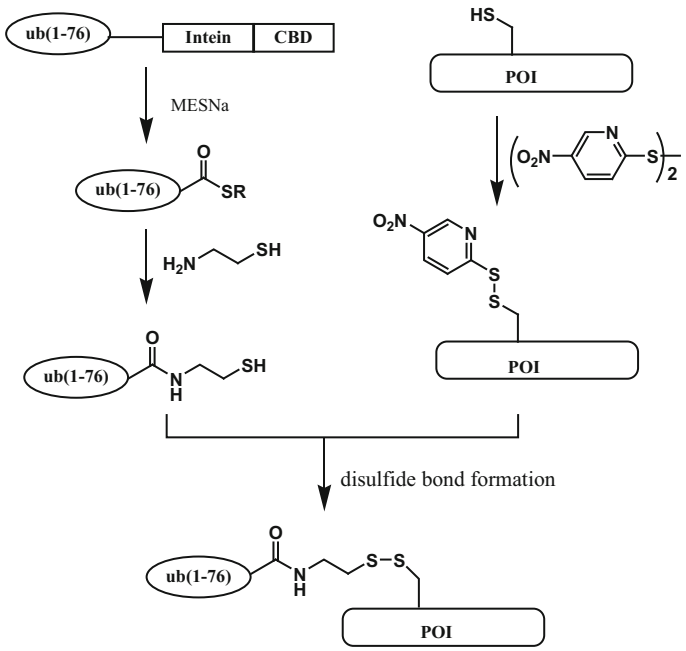
Scheme 10 General scheme for the synthesis of K33 and K11 mix-linked triubiquitin using the revised GOPAL approach



Scheme 11 General scheme for the synthesis of G76C or G76A ubiquitin conjugates. *POI* – protein of interest

Non-native ubiquitin conjugates with disulfide bond were also reported [30–33]. In this type of conjugates, a full-length ubiquitin with a C-terminal 2-mercaptoethyl amide moiety was first prepared by reacting cysteamine with ubiquitin thioester; the thiol at its C-terminal 2-mercaptoethyl amide was then linked to the cysteine residue of the substrate protein through a disulfide bond (Scheme 12). The method is very simple and suitable for the preparation of ubiquitinated proteins in large quantities as both reacting components can be easily obtained as recombinant proteins. The limitations of the method are that the linkage is much longer than the native lysine side chain and that the disulfide bond is also susceptible to reducing conditions.

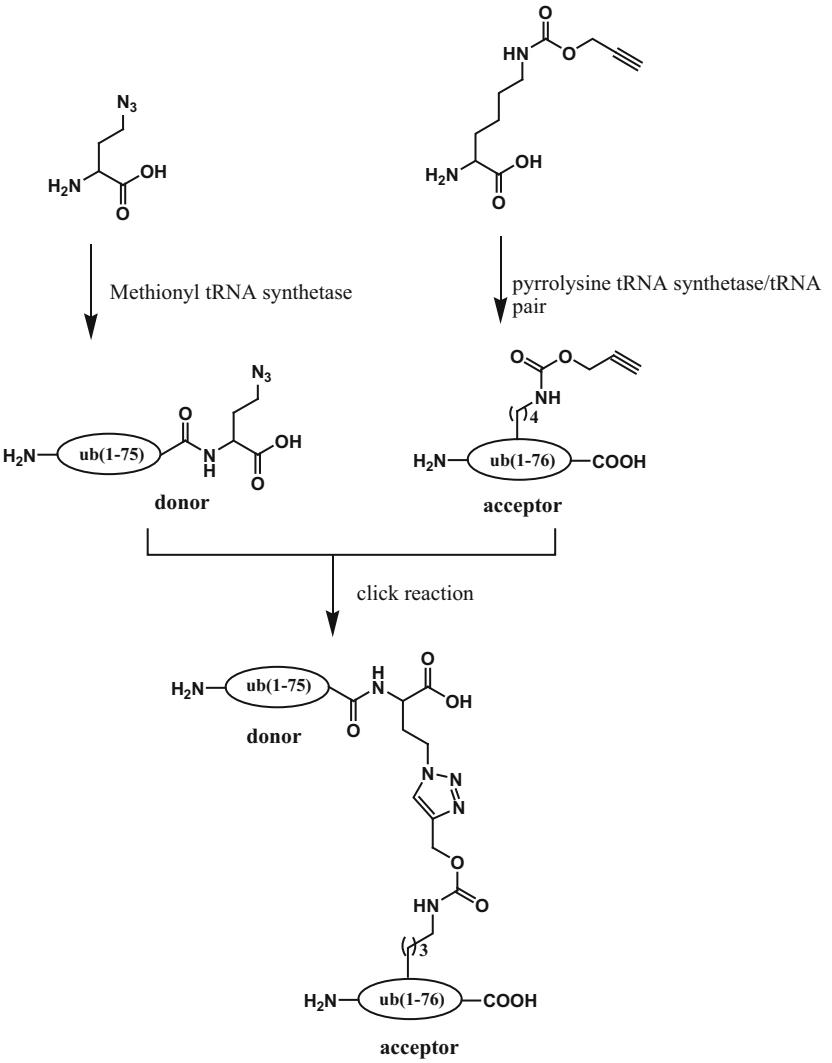
Triazole-linked ubiquitin conjugates were also demonstrated in the synthesis of proteins modified by ubiquitin or ubiquitin-like proteins (Schemes 13 and 14) [34–36]. The triazole linkage was formed through copper-catalyzed click reaction between an alkyne and an azide. Both the alkyne and azide are non-natural functional groups and have been introduced through genetic incorporation as part of the unnatural amino acids [34] or through post-expression chemical manipulation of recombinant proteins [35, 36]. In the genetic incorporation approach, an azide functional group was introduced when azidohomoalanine was incorporated into the donor ubiquitin at the C-terminal end. The incorporation was mediated by methionyl-tRNA synthetase in methionine auxotrophic bacterial strains grown in minimal media. The alkyne functional group was introduced when alkyne-bearing pyrrolysine analogs were incorporated into the receptor ubiquitin (Scheme 13). Mootz et al. reported another way to generate triazole-linked ubiquitinated/



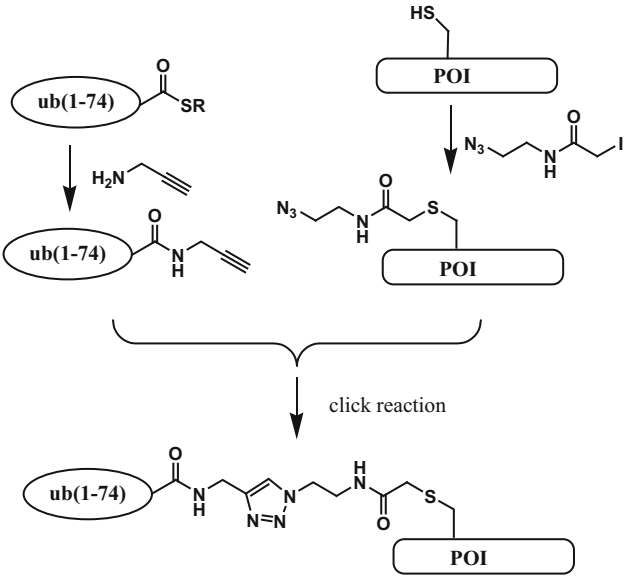
Scheme 12 General scheme for disulfide-directed site-specific ubiquitination. *POI* protein of interest

sumoylated proteins [36] (Scheme 14). In their approach, an alkyne group was introduced onto the C-terminus of ubiquitin(1–74) or ubiquitin-like proteins (such as SUMO) through reacting propargyl amine with the ubiquitin or SUMO thioester. An azide group was introduced to the side chain of a cysteine residue in the protein to be modified through alkylation with iodoacetamide ethyl azide. Click reaction was then performed to link ubiquitin (or SUMO) with the substrate protein. The triazole linkages formed were resistant to ubiquitin-deconjugating enzymes.

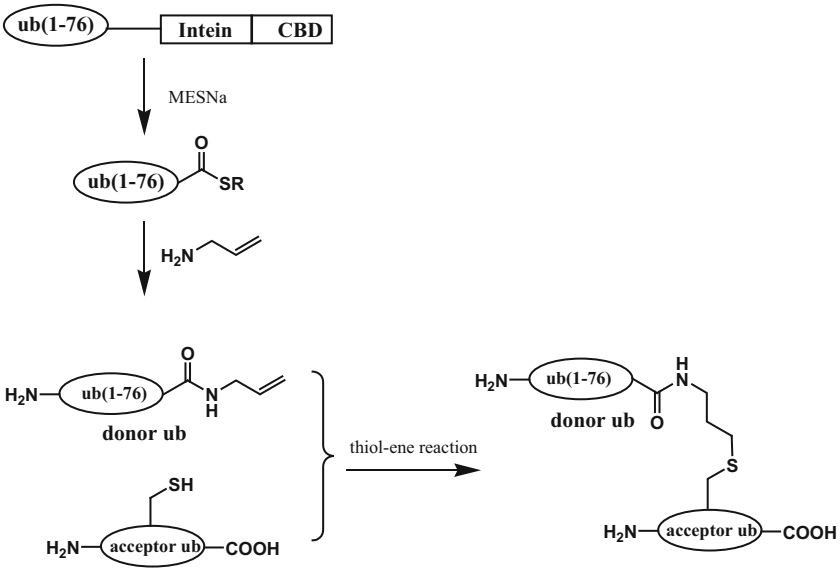
Ubiquitin conjugates with a thioether linkage have also been reported [37, 38]. The thioether linkage was formed through free-radical thiol-ene click reaction between an allyl group introduced at the C-terminus of donor ubiquitin and the thiol group of a cysteine residue of in the substrate protein (Scheme 15). The introduction of the ally group was achieved through reacting allylamine with the full-length ubiquitin thioester. The thioether linkage generated resembled very much the native isopeptide linkage of ubiquitination except that the former had a slightly longer "lysine" side chain. However, the thiol-ene reaction was inefficient in this setting and a yield of only about 30% was obtained. The probable reason was that the large size of the two reacting components imposed a molar concentration which was too low to sustain the radical chain reaction.



Scheme 13 Triazole-linked ubiquitin conjugate formed between genetically incorporated azide and alkyne



Scheme 14 Triazole-linked ubiquitin conjugate with post-translationally installed alkyne and azide. *POI* protein of interest



Scheme 15 General scheme for the synthesis of thioether-linked diubiquitins through thiol-ene click reaction

4 Conclusion

Chemical synthesis of site-specifically modified proteins has provided enabling techniques for protein structure-function studies. The development of the above methods has helped to make the difficult-to-obtain ubiquitinated proteins more available for biophysical and biochemical characterizations. Obviously, the ub-protein conjugates with non-native linkages are much easier synthetic targets and in many cases such conjugates are useful reagents for functional studies. Nevertheless, it is desirable to obtain ubiquitinated proteins with the native isopeptidic linkage to understand the complete and genuine structural/functional implications of this important post-translational modification. Without doubt, native chemical ligation [39] is still the most useful technique for the chemical installation of ubiquitin onto synthetic or recombinant proteins.

Acknowledgment We thank the A*Star Science and Engineering Research Council for financial support (SERC 112 120 2017).

References

1. Walsh CT, Garneau-Tsodikova S, Gatto GJ Jr (2005) Angew Chem Int Ed 44:7342–7372
2. Walsh CT (2006) Posttranslational modification of proteins. Roberts and Company, Englewood
3. Hochstrasser M (2000) Nat Cell Biol 2:E153–E157
4. Pickart CM (2001) Annu Rev Biochem 70:503–533
5. Kerscher O, Felberbaum R, Hochstrasser M (2006) Annu Rev Cell Dev Biol 22:159–180
6. Hochstrasser M (2009) Nature 458:422–429
7. Chen ZJ, Sun LJ (2009) Mol Cell 33:275–286
8. Ikeda F, Crosetto N, Dikic I (2010) Cell 143:677–681
9. Li W, Ye Y (2008) Cell Mol Life Sci 65:2397
10. Hochstrasser M (1995) Curr Opin Cell Biol 7:215
11. Wang C, Deng L, Hong M, Akkaraju GR, Inoue J-I, Chen ZJ (2001) Nature 412:346
12. Chatterjee C, McGinty RK, Pellois J-P, Muir TW (2007) Angew Chem Int Ed 46:2814–2818
13. McGinty RK, Kim J, Chatterjee C, Roeder RG, Muir TW (2008) Nature 453:812–816
14. Yang R, Bi X, Li F, Cao Y, Liu C-F (2014) Chem Commun 50:7971
15. Li Y-T, Liang J, Li J-B, Fang G-M, Huang Y, Liu L (2014) J Pept Sci 20:102
16. Weller CE, Huang W, Chatterjee C (2014) Chembiochem 15:1263
17. Yang R, Pasunooti KK, Li F, Liu X-W, Liu C-F (2009) J Am Chem Soc 131:13592–13593
18. Yang R, Pasunooti KK, Li F, Liu X-W, Liu C-F (2010) Chem Commun 46:7199–7201
19. Kumar KSA, Haj-Yahya M, Olschewski D, Lashuel HA, Brik A (2009) Angew Chem Int Ed 48:8090–8094
20. Kumar KSA, Spasser L, Erlich LA, Bavikar SN, Brik A (2010) Angew Chem Int Ed 49: 9126–9131
21. Kumar KSA, Bavikar SN, Spasser L, Moyal T, Ohayon S, Brik A (2011) Angew Chem Int Ed 50:6137–6141
22. Virdee S, Kapadnis PB, Elliott T, Lang K, Madzak J, Nguyen DP, Riechmann L, Chin JW (2011) J Am Chem Soc 133:10708–10711
23. El Oualid F, Merkx R, Ekkebus R, Hameed DS, Smit JJ, De Jong A, Hilkmann H, Sixma TK, Ovaa H (2010) Angew Chem Int Ed 49:10149–10153

24. Merkx R, De Bruin G, Kruithof A, Van Den Bergh T, Snip E, Lutz M, El Oualid F, Ovaa H (2013) Chem Sci 4:4494–4498
25. Virdee S, Ye Y, Nguyen DP, Komander D, Chin JW (2010) Nat Chem Biol 6:750–757
26. Castañeda C, Liu J, Chaturvedi A, Nowicka U, Cropp TA, Fushman D (2011) J Am Chem Soc 133:17855–17868
27. Li X, Fekner T, Ottesen JJ, Chan MK (2009) Angew Chem Int Ed 48:9184–9187
28. McGinty RK, Köhn M, Chatterjee C, Chiang KP, Pratt MR, Muir TW (2009) ACS Chem Biol 4:958–968
29. Fierz B, Kilic S, Hieb AR, Luger K, Muir TW (2012) J Am Chem Soc 134:19548–19551
30. Chatterjee C, McGinty RK, Fierz B, Muir TW (2010) Nat Chem Biol 6:267–269
31. Chen J, Ai Y, Wang J, Haracska L, Zhuang Z (2010) Nat Chem Biol 6:270–272
32. Meier F, Abeywardana T, Dhall A, Marotta NP, Varkey J, Langen R, Chatterjee C, Pratt MR (2012) J Am Chem Soc 134:5468–5471
33. Fierz B, Chatterjee C, McGinty RK, Bar-Dagan M, Raleigh DP, Muir TW (2011) Nat Chem Biol 7:113–119
34. Eger S, Scheffner M, Marx A, Rubini M (2010) J Am Chem Soc 132:16337–16339
35. Weikart ND, Mootz HD (2010) Chembiochem 11:774–777
36. Sommer S, Weikart ND, Brockmeyer A, Janning P, Mootz HD (2011) Angew Chem Int Ed 50:9888–9892
37. Valkevich EM, Guenette RG, Sanchez NA, Chen Y-C, Ge Y, Strieter ER (2012) J Am Chem Soc 134:6916–6919
38. Trang VH, Valkevich EM, Minami S, Chen Y-C, Ge Y, Strieter ER (2012) Angew Chem Int Ed 51:13085–13088
39. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Science 266:776–779

Top Curr Chem (2015) 362: 107–136
DOI: 10.1007/128_2014_575

Published online: 5 November 2014

Peptide Thioester Formation via an Intramolecular *N* to *S* Acyl Shift for Peptide Ligation

Toru Kawakami

Abstract In chemical protein synthesis, peptide building blocks are prepared by solid-phase peptide synthesis (SPPS), and then connected by chemical ligation methods. The peptide thioester is one of key building blocks used in chemical protein synthesis, and improvements in the Fmoc SPPS procedure for preparing such thioesters would be highly desirable. In this review we focus on a method for peptide thioester synthesis based on the use of an intramolecular *N* to *S* acyl shift reaction as a key reaction. Amide and thioester forms at the thiol-containing residue are in equilibrium as a result of a reversible intramolecular acyl shift, which is detectable by ^{13}C NMR. The amide form is favored under neutral conditions, while the thioester predominates under acidic conditions. Thiol auxiliaries can be employed to facilitate the formation of a thioester from an amide via an intramolecular *N–S* acyl shift, and the peptide thioester is formed after intermolecular transthioesterification in the presence of excess amounts of thiols. Even under neutral conditions, thiol auxiliary-containing peptides can be ligated with a cysteinyl peptide via an intramolecular *N–S* acyl shift, followed by native chemical ligation (NCL) in a one-pot reaction. These procedures can be applied to the chemical synthesis of proteins which are post-translationally modified.

Keywords Ligation · *N–S* acyl shift · Peptide thioester · Post-translational modification · Protein synthesis · Transthioesterification

T. Kawakami (✉)
Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
e-mail: kawa@protein.osaka-u.ac.jp

Contents

Abbreviations

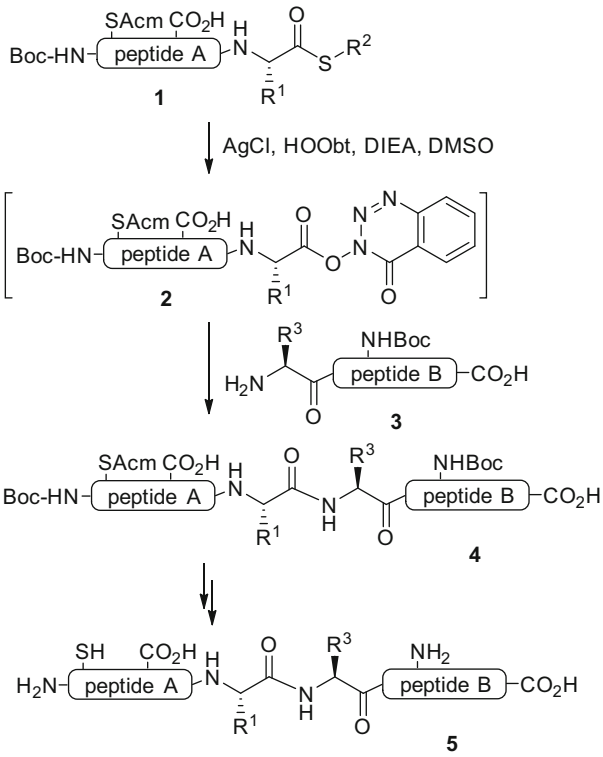
Acm	Acetamidomethyl
Boc	*tert*-Butoxycarbonyl
CPC	Cysteinylprolylcysteine
CPE	Cysteinylproline ester
DEAD	Diethyl azodicarboxylate
DIEA	*N*,*N*-Diisopropylethylamine
DIPCI	*N*,*N*′-Diisopropylcarbodiimide
DKP	Diketopiperazine
Dmmb	4,5-Dimethoxy-2-mercaptobenzyl
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
Fmoc	9-Fluorenylmethoxycarbonyl
Gdn	Guanidine
HBTU	2-(1*H*-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HFBA	Heptafluorobutyric acid
HOBt	1-Hydroxybenzotriazole
HOObt	3,4-Dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine
HOSu	*N*-Hydroxysuccinimide
MAP	Methionine aminopeptidase
MBHA	4-Methylbenzhydrylamine
MESNa	Sodium 2-mercaptoethanesulfonate
MPA	3-Mercaptopropionic acid
NAC	*N*-Alkyl cysteine
NCL	Native chemical ligation
NMR	Nuclear magnetic resonance
PDF	Peptide deformylase

PURE system	Protein synthesis using recombinant elements system
RP-HPLC	Reversed-phase high performance liquid chromatography
SPPS	Solid-phase peptide synthesis
t-Bu	*tert*-Butyl
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TFMSA	Trifluoromethanesulfonic acid
Tle	*tert*-Leucine
Trt	Trityl

1 Introduction

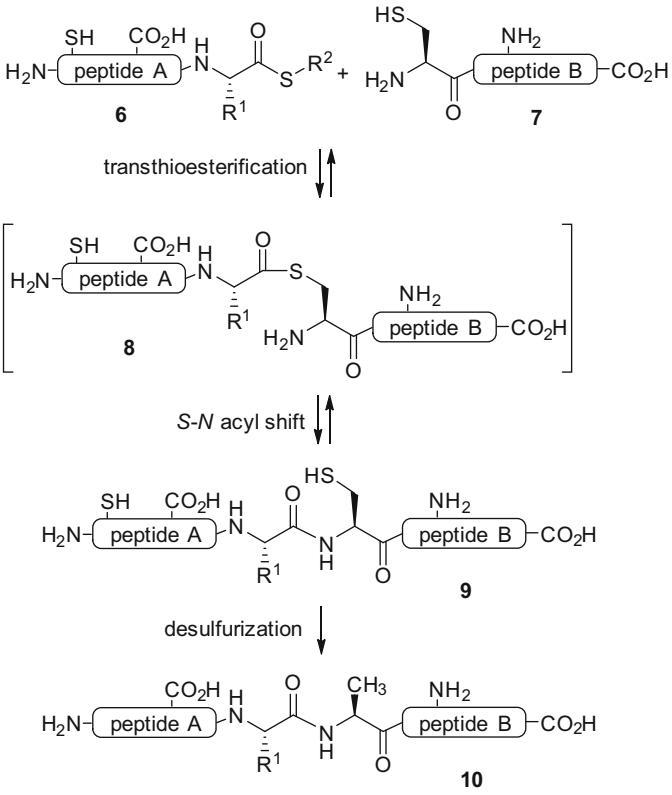
Peptides can now be conveniently and rapidly synthesized by solid-phase peptide synthesis (SPPS) [1, 2]. SPPS is an efficient method for the synthesis of a peptide, which contains less than 50 amino acid residues, but is generally difficult to apply to protein synthesis. In order to synthesize a longer peptide, peptide segments, which can be synthesized by SPPS, are condensed using ligation methodology. Producing native peptide bonds by chemical ligation has been dramatically advanced by the introduction of the peptide thioester as a building block [3–6]. Typical ligation methods using the thioester are shown in Schemes 1, 2, and 3. In the thioester method, a partially protected peptide thioester **1** is used as a building block and condensed with a partially protected peptide **3** via an active ester **2**, in the presence of silver ions as an activating reagent in an organic solvent such as DMSO (Scheme 1) [7–10]. This method can be applied to ligation at any given amino acid residue. Native chemical ligation (NCL) is a chemoselective reaction, in which an unprotected peptide thioester **6** is ligated with a cysteinyl peptide **7** in a neutral aqueous buffer solution (Scheme 2) [6, 11]. Although this method requires a cysteine residue at the ligation site, the cysteine residue can be transformed into an alanine residue after the reaction is complete [12, 13]. When amino acid residues, to which the thiol group is introduced on the side chain, are placed at the N-terminus, ligation at the corresponding amino acid residues can be realized [14–25]. Several types of thiol auxiliaries have been used as alternatives to the cysteine residue. In these cases, the thiol auxiliary is attached to the N-terminal amino group for extended chemical ligation, thus retaining the advantageous features of NCL [26–32]. In Scheme 3, an auxiliary group, 4,5-dimethoxy-2-mercaptobenzyl (Dmmb), is shown as an example [28].

The peptide thioester is a common building block used in ligation methodology, and improved methods for preparing such thioesters would be highly desirable. Peptide thioesters can be prepared in a straightforward manner via the use of *tert*-butoxycarbonyl (Boc) SPPS. For example, starting from a thioester resin, a Boc-Gly-SCH_2CH_2CO-β-Ala-MBHA resin, the peptide chain can be elongated by the standard Boc SPPS and the peptide thioester is then removed by treatment with anhydrous hydrogen fluoride [8, 10]. However, the preparation of peptides with



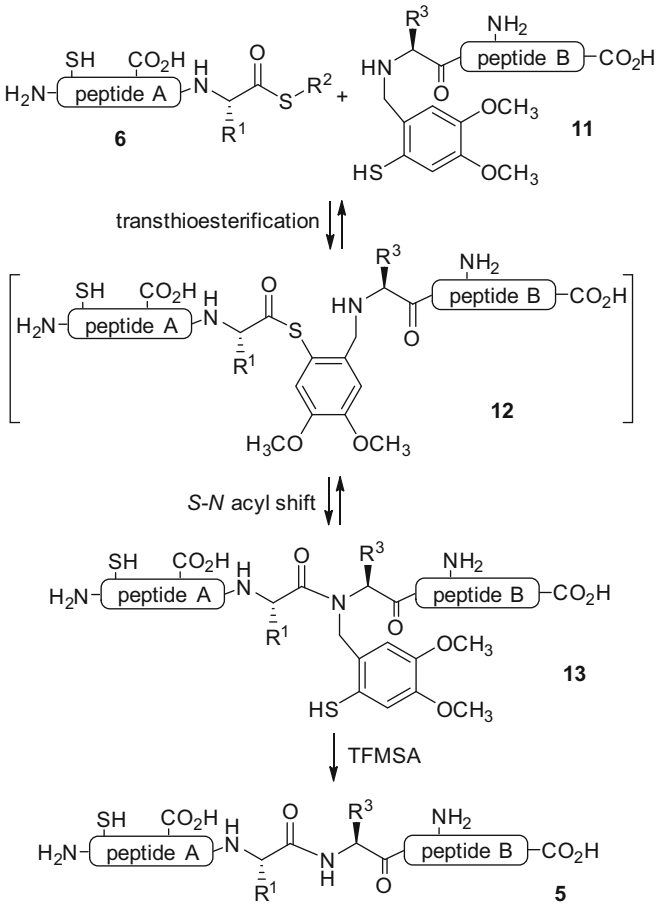
Scheme 1 Thioester method

post-translational modifications such as glycosylation [33–35] and phosphorylation [36] by Boc SPPS continues to be a formidable task. On the other hand, the 9-fluorenylmethoxycarbonyl (Fmoc) SPPS method is suitable for the preparation of peptides that are post-translationally modified, although the preparation of a peptide thioester using the Fmoc SPPS procedure continues to be a challenge, because the thioester decomposes in the presence of piperidine, which is commonly used to remove the Fmoc group. In the direct synthesis from a thioester resin, weak nucleophilic bases can be used. In this case, the Fmoc group is removed by treatment with a mixture of 1-methylpyrrolidine, hexamethylenimine, and 1-hydroxybenzotriazole (HOBt), permitting the thioester bond to remain intact [37]. Using this approach, peptide thioesters can be prepared, although chiral amino acid residues at the thioester position are partially racemized during the removal of the Fmoc group in the chain elongation cycle [38]. Indirect methods have also been examined [39]. For example, a peptide chain can be elongated on a sulfonamide linker, and the linker is then activated by alkylation of the sulfonamide, which can be replaced by an intermolecular reaction with an external thiol to produce a peptide thioester [40, 41]. This procedure is simple to use, but has some drawbacks. For example, the sulfonamide moiety can undergo acetylation during the capping cycle [42] and methionine residues are frequently alkylated during the



Scheme 2 Native chemical ligation (NCL) and desulfurization

activation step [43]. So far, a number of synthetic methods for preparing peptide thioesters based on the Fmoc SPPS procedure have been studied. These methods include the use of a Wang linker [44], an aryl hydrazine linker [45], a trithioortho ester [46], an in situ *O* to *S* acyl shift reaction [47, 48], the formation of *N*-acylurea [49] or pyroglutame [50] for activation, and the use of a peptide hydrazide [51] or a peptidyl *N*-acetylguanidine [52] intermediate. Each method may overcome some of the difficulties associated with peptide thioester synthesis by the Fmoc SPPS procedure. As an alternative approach, an intramolecular *N* to *S* acyl shift reaction has emerged as one of key reactions for the preparation of peptide thioesters. In this review we focus on methods for peptide thioester production via the intramolecular *N*–*S* acyl shift reaction in chemical procedures. The other recent reviews for preparation of peptide thioesters by Fmoc SPPS are also available [53–55].

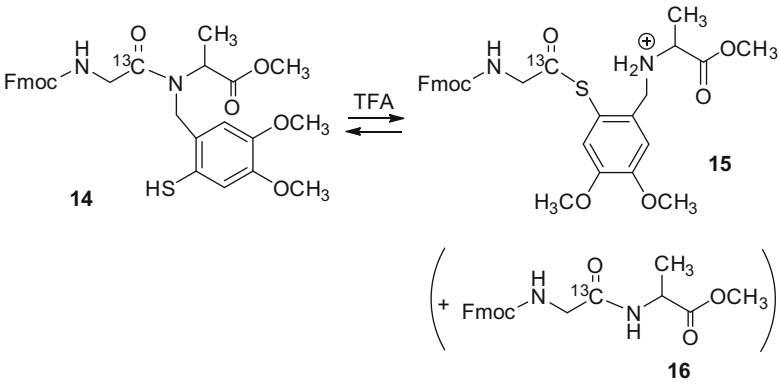


Scheme 3 Extended chemical ligation

2 Intramolecular *N* to *S* Acyl Shift Reaction

2.1 At the Thiol Auxiliary, Dmmb

The Dmmb group was originally developed as a thiol-mediated ligation auxiliary, and can be removed by treatment with 1 M trifluoromethanesulfonic acid (TFMSA) in trifluoroacetic acid (TFA) after completion of the ligation (Scheme 3) [28]. On the other hand, when the peptide **13** containing Dmmb group was treated with TFA without TFMSA, a new compound **X** was produced, which was eluted earlier than peptide **13** on reversed-phase high performance liquid chromatography (RP-HPLC), and had a mass identical to that of **13** [29]. Compound **X** disappeared when the pH of the solution was increased to 8. This suggests that compound **X** corresponded to the thioester intermediate **12**, which was formed by the reversal of the ligation reaction, thus involving an intramolecular *N*–*S* acyl shift reaction.



Scheme 4 TFA treatment of the Dmmb-containing ^{13}C labeled peptide

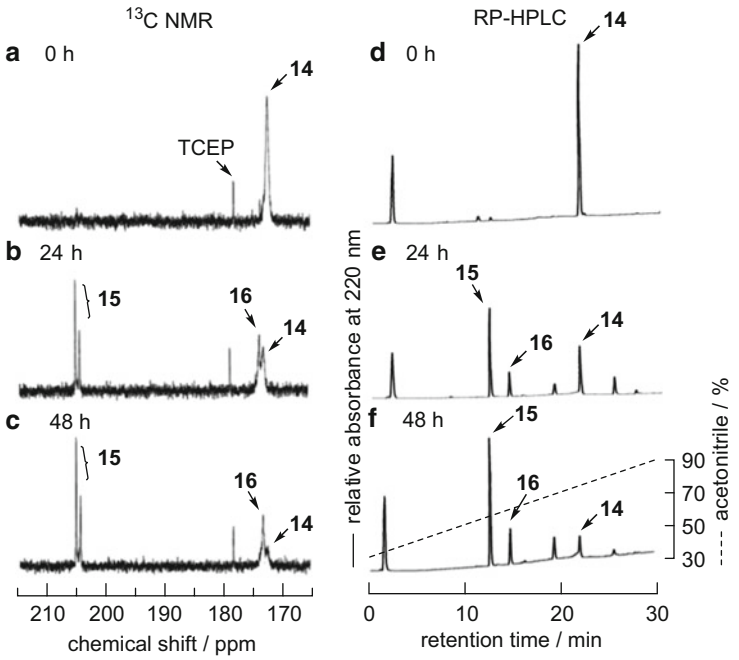


Fig. 1 Reaction of Dmmb-containing ^{13}C labeled peptide **14** in a TFA solution (14% $CDCl_3$, 0.5% TCEP•HCl (*v*/*v*/*w*)). (**a**–**c**) ^{13}C NMR spectra after indicated reaction time. (**d**–**f**) RP-HPLC. Column: YMC-Pack Pro C18 (4:6 × 150 mm); eluent: 0.1% TFA in aq CH_3CN, 1.0 mL/min

In order to confirm the formation of thioester from the Dmmb-containing peptide, the ^{13}C-labeled peptide, Fmoc-Gly(1-^{13}C)-D,L-(Dmmb)Ala-OCH_3 (**14**) was treated with TFA and analyzed by ^{13}C NMR and RP-HPLC (Scheme 4, Fig. 1) [56]. The ^{13}C NMR signal at 172.8 ppm, corresponding to amide **14**, was shifted, and signals at 204.6 and 205.3 ppm, corresponding to thioester **15**,

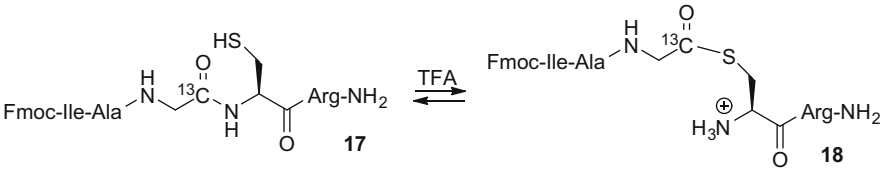
appeared. Two signals of the thioester are indicative of the presence of distinct conformers. A small amount of peptide **16** without Dmmb group was observed under the conditions used.

On the RP-HPLC, amide **14** and thioester **15** were eluted at 21 min and at 12 min, respectively. These two materials had identical masses. The ratio of the amide and the thioester (1:4) was in good agreement with the two different analytical methods, indicating that the thioester **15** is formed and is relatively stable under these analytical conditions.

2.2 At the Cysteine Residue

A thiol auxiliary, Dmmb mediated thioester formation, above, and a cysteine residue might also play a role in the formation of a thioester. It was suggested that a thioester is formed during the treatment of a protected cysteine-containing peptide with anhydrous hydrogen fluoride [57]. In order to confirm the formation of thioester from a cysteine-containing peptide, the ^{13}C-labeled peptide, Fmoc-Ile-Ala-Gly(1-^{13}C)-Cys-Arg-NH_2 (**17**) was treated with TFA and analyzed by ^{13}C NMR and RP-HPLC (Scheme 5, Fig. 2) [56]. On ^{13}C NMR the signals at 172.5 and 172.7 ppm, corresponding to amide **17**, were shifted to 201.8 ppm, corresponding to thioester **18**. This amide and thioester mixture reached equilibrium (1:4) after almost 1 month.

On the RP-HPLC, amide **17** was eluted at 23 min, and thioester **18** at 19 min. These two materials had identical masses. The ratio of amide and thioester (4:1) was, however, much different from the NMR results, showing a 1:4 ratio. These results indicate that cyateine-containing peptide is in equilibrium between amide and thioester, as a result of an intramolecular acyl shift, and the thioester is predominant in a concentrated TFA solution, whereas this thioester is transformed into an amide, even under the acidic conditions of 0.1% TFA in aq acetonitrile, which was used in the RP-HPLC analysis.



Scheme 5 TFA treatment of the Cys-containing ^{13}C labeled peptide

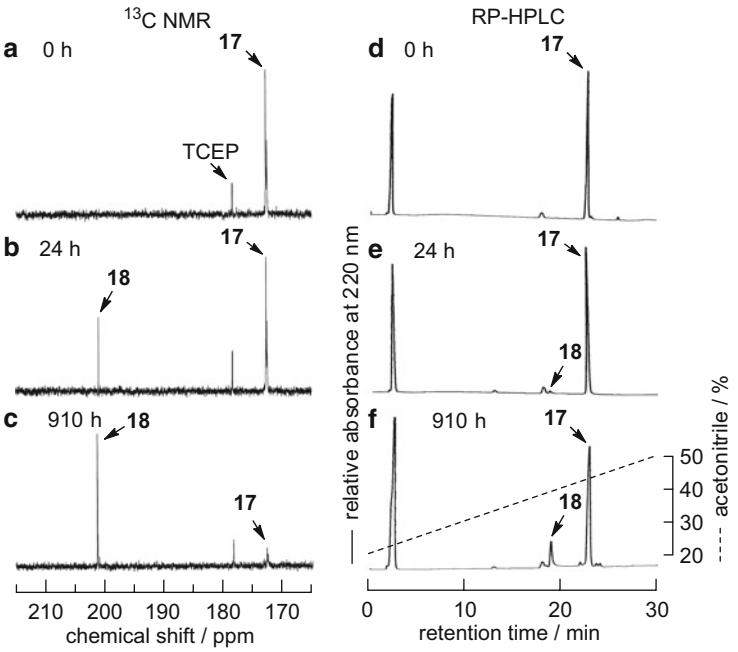
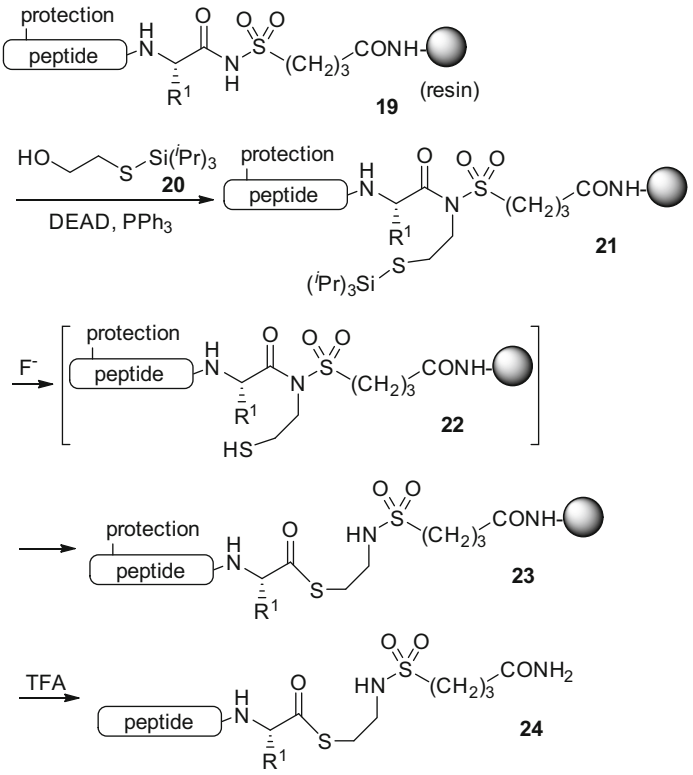


Fig. 2 Reaction of Cys-containing ^{13}C labeled peptide **17** in a TFA solution (29% $CDCl_3$, 0.5% TCEP (*v*/*v*/*w*)). (**a**–**c**) ^{13}C NMR spectra after indicated reaction time. (**d**–**f**) RP-HPLC. Column: YMC-Pack Pro C18 (4:6 × 150 mm); eluent: 0.1% TFA in aq CH_3CN; 1.0 mL/min

3 Peptide Thioester Synthesis via Intramolecular *N* to *S* Acyl Shift

3.1 Active Acyl Structure

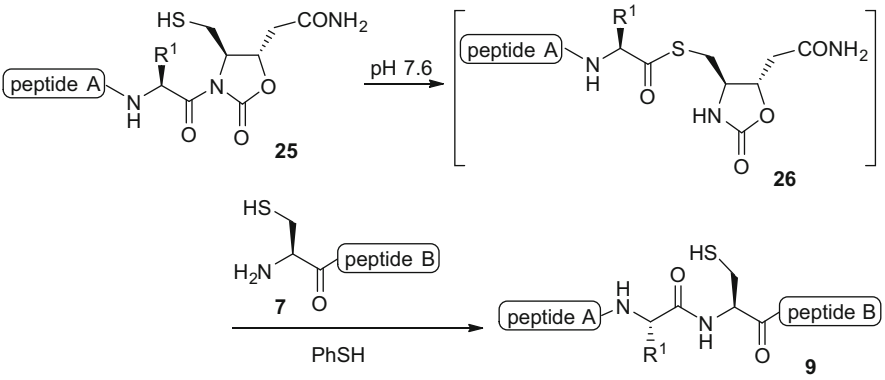
An active acyl moiety toward nucleophiles has advantages for use in intramolecular *N*–*S* acyl shift reactions, and the resulting thioester is basically stable, which means the reverse *S*–*N* acyl shift does not occur. The safety-catch strategy using the sulfonamide linker had been applied to peptide thioester synthesis by an intermolecular reaction with an external thiol after the peptide chain elongation was complete [40, 41]. The concept for peptide thioester synthesis via the intramolecular *N*–*S* acyl shift reaction taking advantage of the safety catch strategy has been reported (Scheme 6) [58]. The peptide sulfonamide resin **19** is alkylated for activation with *S*-protected mercaptoethanol **20** to give the alkylated sulfonamide **21**. The removal of the triisopropylsilyl protecting group results in the formation of a thioester via the thiol intermediate **22**, which undergoes an intramolecular *N*–*S* acyl shift, followed by TFA treatment to liberate the peptide thioester **24** from the protecting groups and the solid support.



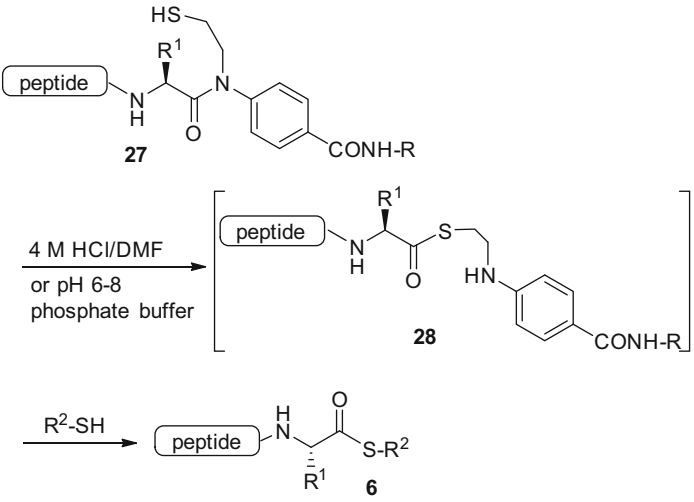
Scheme 6 Strategy for peptide thioester synthesis by intramolecular *N–S* acyl shift on sulfonamide linker

The acyl oxazolidinone structure **25** was reported as another active structure, in which an *N–S* acyl shift reaction can be performed under neutral conditions, and ligation with cysteinyl peptide **7** can then proceed [59]. Although the synthesis of hBNP32-NH_2 was successfully performed, the acyl oxazolidinone was partially decomposed by piperidine treatment and was potentially racemized during peptide chain elongation by Fmoc SPPS (Scheme 7).

Anilides are slightly more active than alkyl amides. *N*-Sulfanylethylanilide (SEAlide) **27** can be transformed into the thioester via intramolecular *N–S* acyl shift under both acidic and neutral conditions [60, 61]. Phosphate salts are needed for thioester formation under neutral conditions, in which ligation with the cysteinyl peptide is performed (Scheme 8).



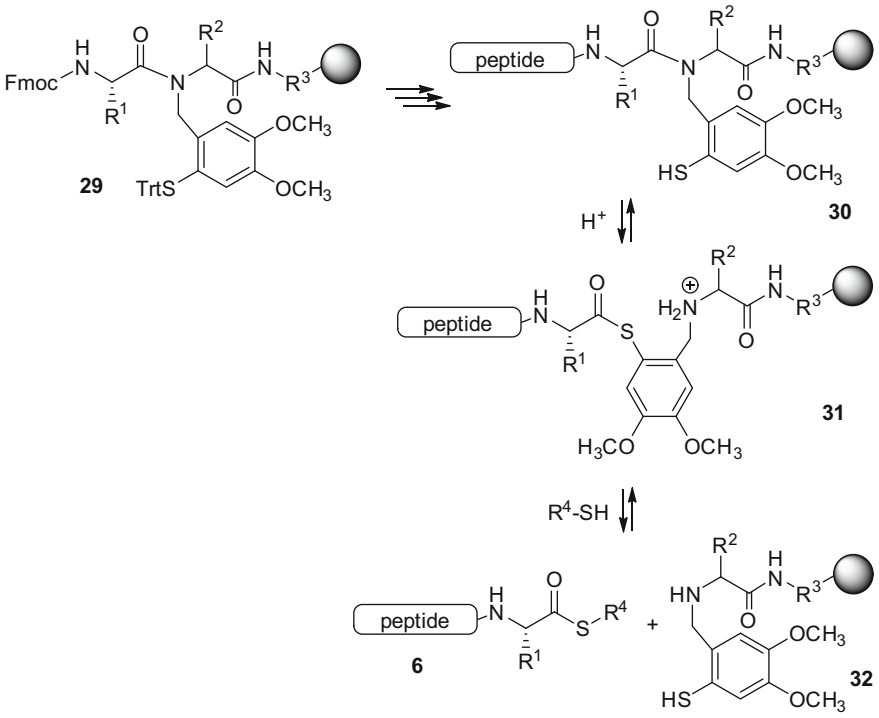
Scheme 7 Acyl oxazolidinone



Scheme 8 Anilide, SEAlide

3.2 Thiol Auxiliary, Dmmb

The Dmmb group can mediate thioester formation via an intramolecular *N*–*S* acyl shift reaction, and this thioester is relatively stable under the RP-HPLC conditions in which aqueous 0.1% TFA is used as the eluent, although under neutral conditions it is easily transformed into an amide by intramolecular *S*–*N* acyl shift. This



Scheme 9 Dmmb-mediated peptide thioester SPPS

thioester intermediate, however, can be transformed into a 'stable thioester,' which contains no amino group in the vicinity, by intermolecular transthioesterification, prior to the intramolecular *S–N* acyl shift, upon treatment with an excess of thiol under neutral conditions. This corresponds to the reverse scheme for ligation (Scheme 3). This process was applied to peptide thioester synthesis via Fmoc SPPS (Scheme 9) [56, 62–64]. A Dmmb-containing peptide does not contain a thioester bond itself, and the peptide chain is elongated on the Dmmb-containing resin **29** by using standard Fmoc SPPS methodology. After the peptide chain assembly, the resin **30** is treated with the acid such as TFA or hydrochloric acid to remove protecting groups and to promote an intramolecular *N–S* acyl shift, resulting in the formation of the thioester intermediate **31** on the resin. This intermediate can be transformed into the peptide thioester **6** by treatment with an excess amount of a thiol, such as 2-mercaptoethenesulfonic acid (MESNa), under neutral conditions. Under optimized conditions, peptide thioester with 29 amino acid residues, BPTI(1–29)-$SCH_2CH_2SO_3H$ (**33**), was obtained in 15% yield; thus, protected BPTI(1–29)-D,L-[Dmmb(Trt)]Ala-Phe-NH-ChemMatrix resin was prepared from the Fmoc-Leu-D,L-[Dmmb(Trt)]Ala-Phe-NH-ChemMatrix resin, and was treated with TFA containing scavengers for 2 h, followed by 0.25 M HCl in 50% aq acetonitrile at 37°C for 3 h. The resulting thioester intermediate was treated with 1.0 M MESNa in a mixture of sodium phosphate buffer (pH 7.0) and

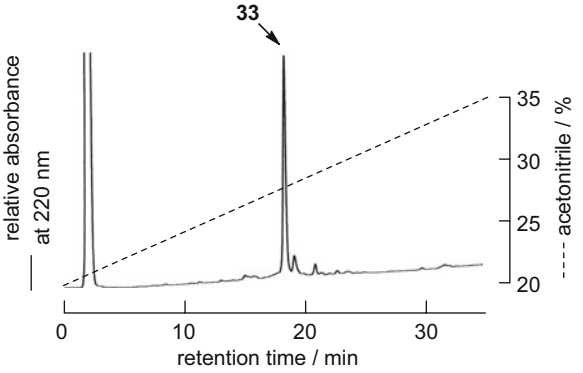


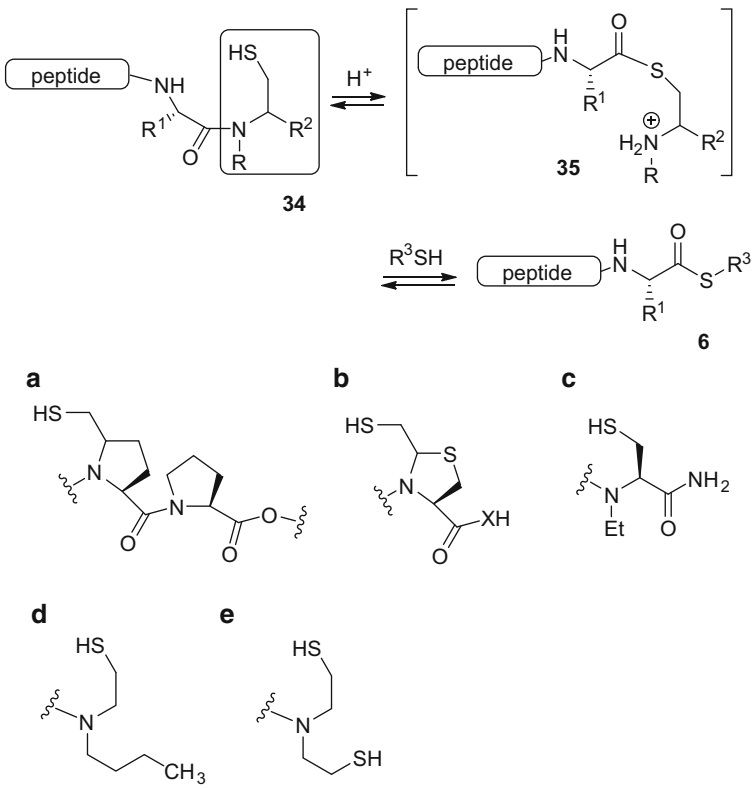
Fig. 3 RP-HPLC profile of the crude BPTI(1–29)-$SCH_2CH_2SO_3H$ (**33**) prepared via an on-resin *N*–*S* acyl shift reaction and transthioesterification. Column: YMC-Pack Pro C18 RS (4.6 × 150 mm); eluent: 0.1% TFA in aq CH_3CN, 1.0 mL/min.

acetonitrile (1:1, *v*/*v*) for 1 h to give peptide thioester **33**. The RP-HPLC of the crude material is shown in Fig. 3. 4-Mercaptophenylacetic acid (MPAA) can be used instead of MESNa to provide the corresponding *S*-aryl thioester, which is known to be a useful building block for native chemical ligation [65].

3.3 tert-*Amide with Vicinal Thiol Group*

In the reaction at the Dmmb group, an *N*–*S* acyl shift occurs at the *tert*-amide, resulting in the formation of a *sec*-amine, which is less reactive for an *S*–*N* acyl shift than the primary amine. Based on the *tert*-amide structure, several types of devices **34** for thioester formation by an intramolecular *N*–*S* acyl shift were reported, which contain a vicinal amino thiol moiety (Scheme 10). A basic concept is that the thioester intermediate **35** is formed under acidic conditions, in which the thioester **35** is in equilibrium with the starting amide **34**, and the intermediate **35** is converted into the 'stable thioester' **6** without a β-amino group, in the presence of an excess amount of an 'external thiol,' R^2SH, such as 3-mercaptopropionic acid (MPA) and MESNa.

The 5-mercaptomethylproline residue is one of such devices used for promoting intramolecular *N*–*S* acyl shift reactions (Scheme 10a) [66]. In an aqueous solution containing MPA, peptide **34a** is transformed into the thioester of MPA. This reaction is promoted by microwave irradiation. The thiazolidine structure can substitute for proline (Scheme 10b) [67]. This thiazolidine ring is constructed from a cysteine residue on the solid support, which simplifies the preparation of the thioester precursor.



Scheme 10 Peptide thioester formation at the *tert*-amide via an intramolecular *N*–*S* acyl shift

tert-Amide structures can be simplified by the use of *N*-alkyl cysteine (NAC) residues, such as *N*-ethyl cysteine (Scheme 10c), which can be easily prepared by the reductive amination of acetaldehyde with *S*-trityl (Trt) cysteine [68]. The introduction of an amino acid residue to an *N*-ethylcysteine residue on the solid support results in a low yield, and therefore dipeptides, *Xaa*-(Et)Cys(Trt), are prepared first and then introduced on the solid support [69]. The conversion rate of a peptide containing an *N*-alkyl cysteine residue at the C-terminus is much faster than that for a peptide with 5-mercaptomethylproline, even at low MPA concentrations and without microwave irradiation. The *tert*-amide structure can be further simplified as an *N*-(2mercaptoethyl)-*N*-butyl amide (Scheme 10d) [70]. This peptide amide is prepared on a Cl-Trt(2-Cl) resin, the efficiency of which is similar to that for thioester formation with the NAC group.

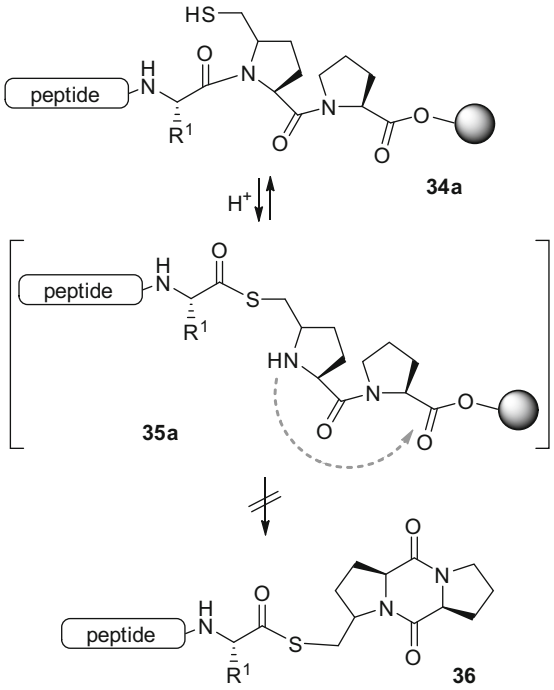
In order for the *N*–*S* acyl shift to take place, the planar amide bond must be in a conformation where the 2-mercaptoethyl group is *anti* to the carbonyl oxygen. *N*,*N*-Bis(2-sulfanylethyl)amide (SEA or BMEA) is introduced at the C-terminal of a peptide, in which the SEA group can facilitate the *N*–*S* acyl shift without any dependence on conformations because of the presence of two 2-sulfanylethyl groups (Scheme 10e) [71, 72]. While an MPA thioester is obtained from SEA peptide **34e** under acidic conditions, the ligation of SEA peptide can also proceed under neutral pH conditions with a cysteinyl peptide in a one-pot reaction. Microwave irradiation significantly accelerates the ligation reaction.

3.4 tert-*Amide with Irreversibility*

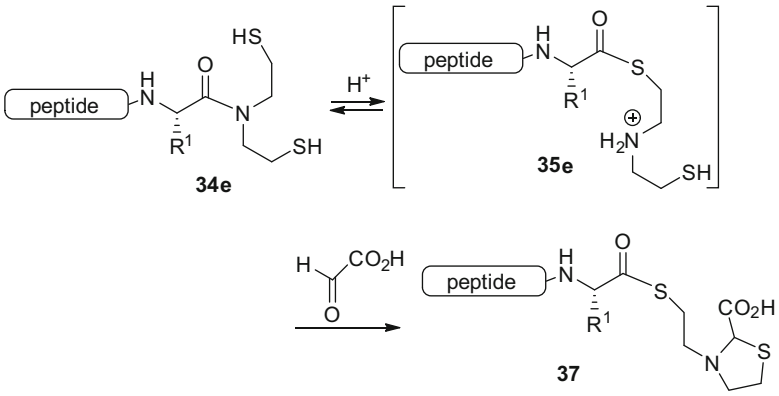
In the synthesis of peptide thioesters by the intramolecular *N*–*S* acyl shift reaction, a drawback is the reverse reaction of the *N*–*S* acyl shift, and thereby an *S*–*N* acyl shift reaction, which reproduces the starting amide. When the amino group is removed after the *N*–*S* acyl shift, the resulting thioester is stabilized, which shifts the equilibrium in the direction of the thioester. One of solutions to this dilemma is intramolecular amidation with different acyl groups from the thioester. The 5-mercaptomethylprolylproline structure (Scheme 10a) was designed to produce diketopiperazine (DKP) **36** after an *N*–*S* acyl shift (Scheme 11), but failed [66]. This strategy was realized later using the cysteinylproline ester (CPE) structure (see below).

A 1,2-aminoethanethiol moiety, which is produced after an *N*–*S* acyl shift reaction of the SEA peptide **34e** (Scheme 10e), can be reacted with glyoxylic acid to produce a thizolidine ring, resulting in *tert*-amine formation (Scheme 12) [73]. Furthermore, this thiozolidine thioester **37** is more reactive than the thioester of MPA in the ligation reaction with a cysteine peptide.

The other concept was reported, in which the enamide structure **39** is transformed into the enamine **40** by an *N*–*S* acyl shift, followed by hydrolysis to give the ketone **41**, thereby removing the amino group (Scheme 13) [74].

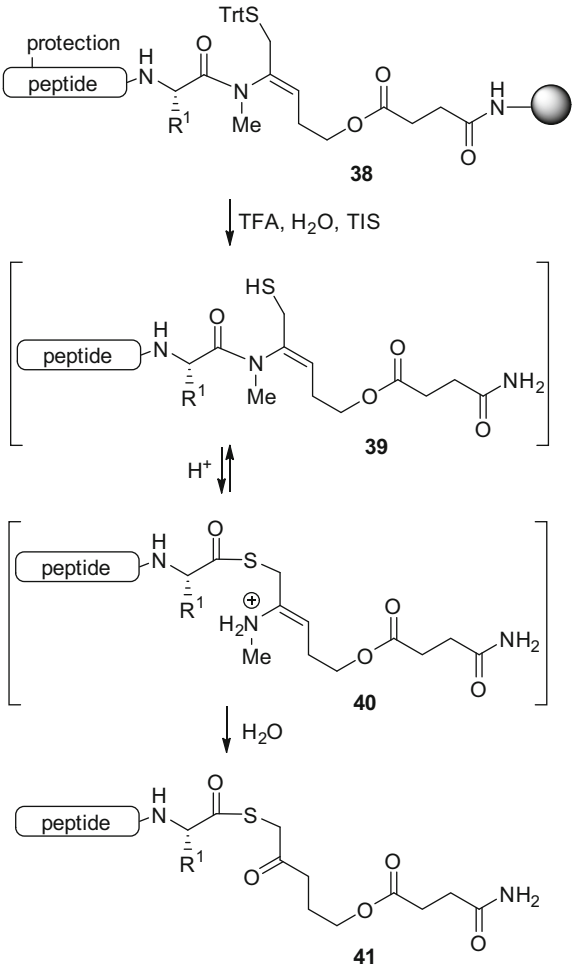


Scheme 11 Proposed peptide thioester stabilization by the DKP formation



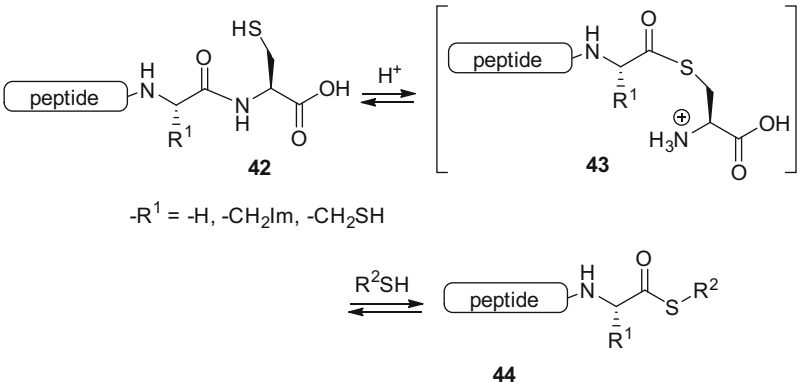
Scheme 12 Peptide thioester stabilization by converting *sec*-amine into *tert*-amine

Scheme 13 Peptide thioester stabilization by removal of the amino group

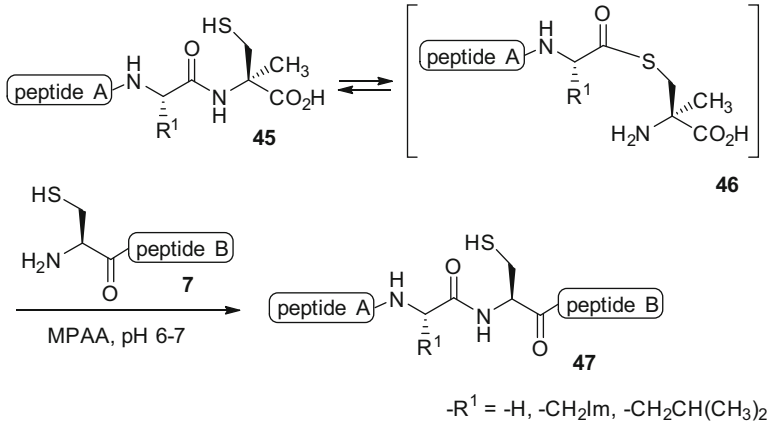


3.5 C-Terminal Cysteine

In the previous sections, thioester formation at the *tert*-amide moiety was discussed, and the thioester intermediate was predominant in the equilibrium mixture under acidic conditions. The intermediate was sufficiently stable, even under neutral conditions, to allow it to react with the external thiol by intermolecular transthioesterification. On the other hand, a thioester intermediate at the cysteine residue is easily transformed into the amide, even under aqueous acidic conditions such as a 0.1% aqueous TFA solution. Special sequences containing a cysteine residue were, however, reported. Peptide **42** containing a Gly-Cys, His-Cys, or Cys-Cys sequence can be transformed into the corresponding glycine, histidine, or cysteine thioester **44** in the presence of thiols under acidic conditions (Scheme 14) [75, 76]. The C-terminal cysteine carboxylic acid is more easily converted into the thioester than



Scheme 14 Peptide thioester formation at the cysteine residue



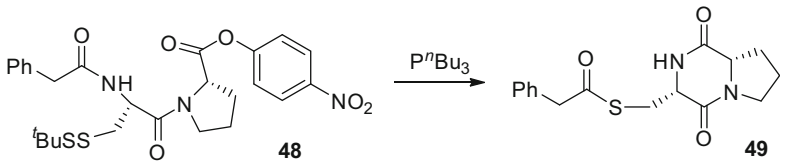
Scheme 15 Ligation via *N–S* acyl shift at the α-methylcysteine residue

the C-terminal cysteine amide. Peptide **42** is composed of only a proteinogenic structure, and a recombinant peptide thioester could be prepared [77].

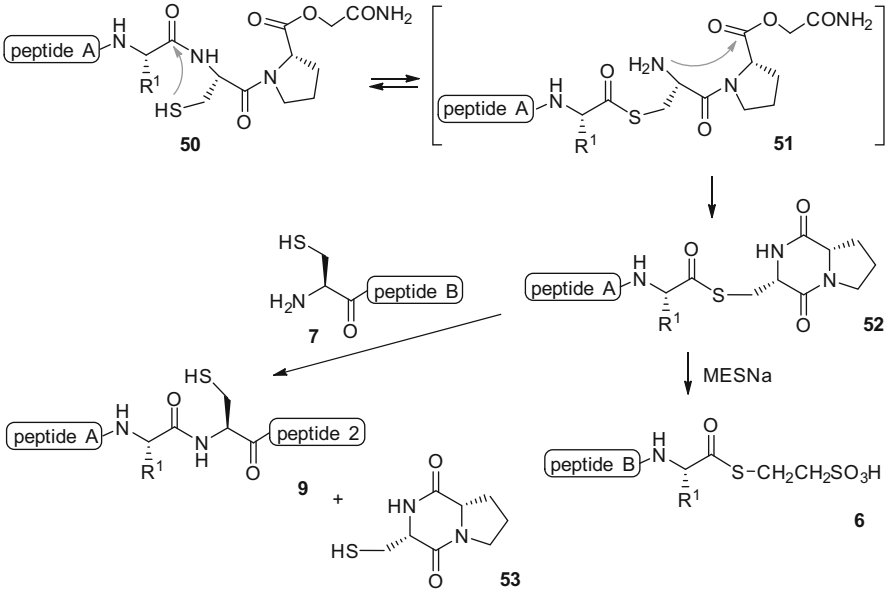
Recently, ligation via an *N–S* acyl shift at the α-methylcysteine site was reported [78]. By using an α-methylcysteine residue, the ligation with a cysteinyl peptide **7** was performed successfully, not only at the glycine or histidine residue but also at the leucine residue under neutral, aqueous conditions (Scheme 15).

3.6 Cysteinylproline Ester (CPE)

In order to stabilize the thioester intermediate at the cysteine residue, the amino group, produced as a result of an *N–S* acyl shift, needs to be protected. An interesting reaction was reported in 1985, in which the acyl cysteinylproline active



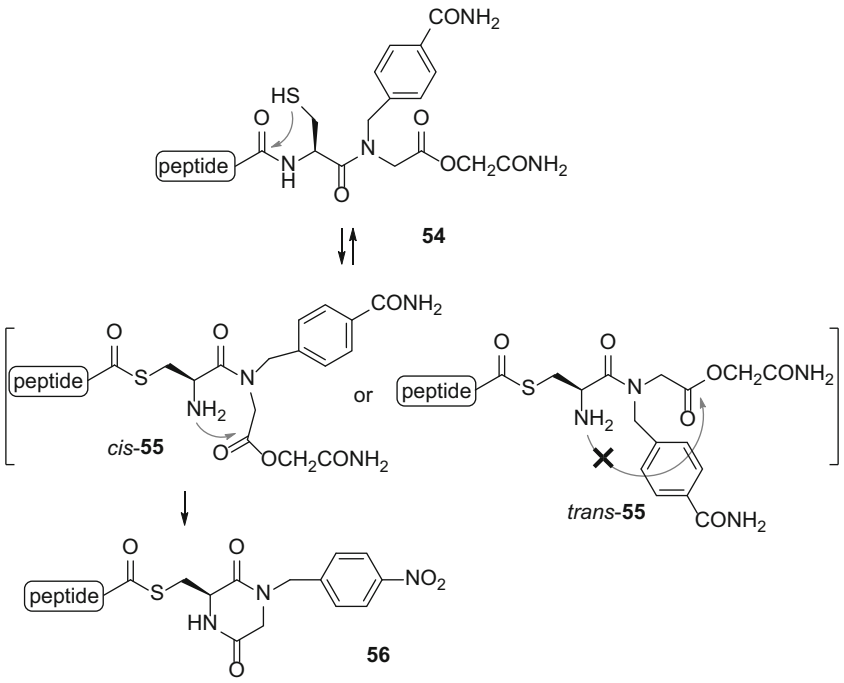
Scheme 16 Formation of DKP thioester



Scheme 17 CPE ligation

ester **48** was transformed into the DKP thioester **49** via an intramolecular reaction (Scheme 16) [79].

Based on this structure, a cysteinylproline ester (CPE) auto-activating unit was designed for use in thioester formation (Scheme 17) [80, 81]. When peptide **50**, containing a CPE structure at the C-terminus, is incubated under neutral, aqueous conditions, the DKP thioester is spontaneously formed via an intramolecular *N*–*S* acyl shift, followed by DKP formation (Scheme 17). In the presence of an additional thiol, such as MESNa, the corresponding thioester **6** is formed. On the other hand, in the presence of the cysteinyl peptide **7** in the reaction mixture, the ligation product **9** is obtained in a one-pot reaction. During the incubation, the chiral amino acid residue at thioester **6** is gradually racemized under the slightly basic conditions used, especially in the case of sensitive amino acid residues, such as serine. It is noteworthy that epimerization occurs after the formation of the thioester, and not



Scheme 18 DKP thioester formation from a CPE peptide derivative

during the thioesterification step. Epimerization is suppressed when the pH of the buffer is reduced and in the case of ligation with a cysteinyl peptide.

For DKP thioester formation, the presence of an ester group in the CPE moiety is required. When the ester is replaced by a carboxylic acid or amide, no ligation product is produced. On the other hand, the proline residue can be replaced by an N-substituted glycine residue, such as sarcocine (*N*-methylglycine). At the DKP formation step, the amide bond of the Cys-(R)Gly moiety of the thioester intermediate **55** must adopt a *cis* conformation for cyclization to occur. The *cis* conformation at the N-substituted glycine residue is preferred to that at the Pro residue [82]. As a result, substitution of an N-substituted glycine for proline in the CPE moiety accelerates the formation of the DKP thioester, and the bulky substituents, such as benzyl groups, are more effective [83] (Scheme 18).

An alternative possibility for enhancing the rate of peptide thioester formation was reported, in which an amino acid residue is added after the ester moiety of the CPE group [83]. A mixture of CPE peptides containing additional amino acid residues at the C-terminus, Ala-Lys-Leu-Arg-Phe-Gly-Cys-Pro-OCH_2CO-*Xaa*-NH_2 (**57**) (*Xaa*=none, **a**; Gly, **b**; Ala, **c**; Val, **d**; Tle (*tert*-leucine), **e**; Pro, **f**) was treated in the presence of MESNa under neutral conditions, and a plot was constructed for the rate of conversion of peptides **57** (Fig. 4), in which the thioester, Ala-Lys-Leu-Arg-Phe-Gly-$SCH_2CH_2SO_3H$ (**58**), was formed. The CPE peptides

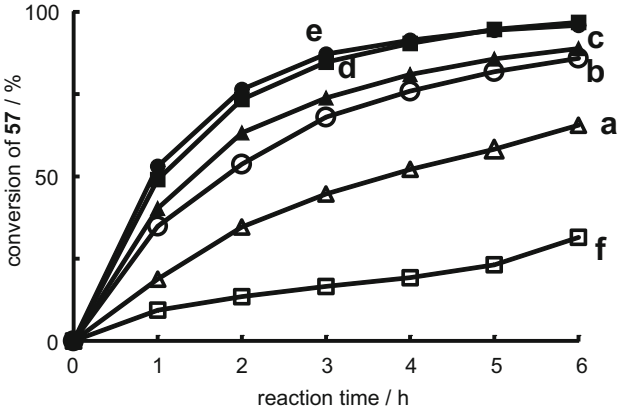
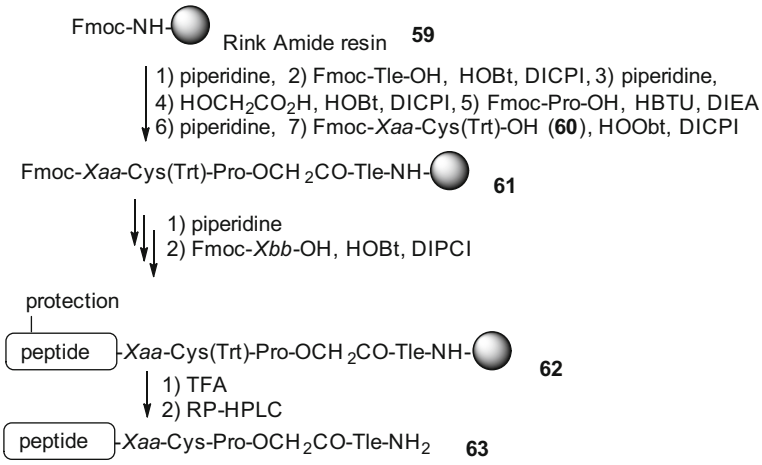


Fig. 4 Conversion rate of the CPE peptides containing additional amino acid residue at the C-terminus of the CPE moiety, Ala-Lys-Leu-Arg-Phe-Gly-Cys-Pro-OCH_2CO-*Xaa*-NH_2 (**57**) (*Xaa*=none, **a**; Gly, **b**; Ala, **c**; Val, **d**; Tle (*tert*-leucine), **e**; Pro, **f**) into thioester, Ala-Lys-Leu-Arg-Phe-Gly-$SCH_2CH_2SO_3H$ (**58**) in 0.1 M sodium phosphate buffer (pH 7.8) at 37°C in the presence of 50 mM MESNa, 20 mM TCEP, and 6 M Gdn

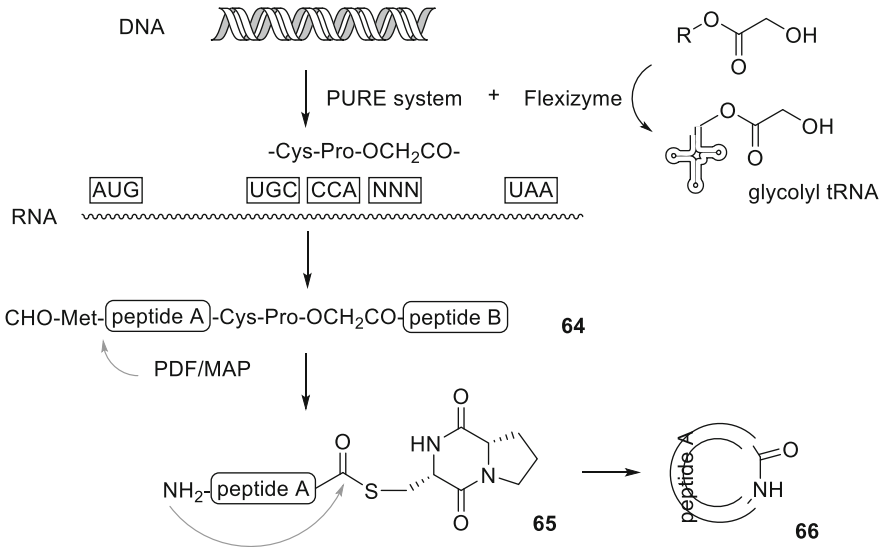
57d, **e** which contain an additional amino acid residue with a bulky side chain, such as Val and Tle, were converted into the thioester more rapidly, and the reaction rate was similar to that for peptide **54**, which contains an *N*-benzyl glycine derivative instead of the proline residue.

A CPE peptide can be easily prepared by the Fmoc SPPS procedure (Scheme 19). In order to prevent DKP formation during peptide chain elongation, the dipeptide, Fmoc-*Xaa*-Cys(Trt)-OH (**60**), must be introduced onto the Pro-OCH_2NH-resin. Standard Fmoc SPPS is then applied to peptide chain elongation, and the usual TFA treatment, followed by RP-HPLC purification, gives the CPE peptide.

Furthermore, a CPE peptide can be prepared using a cell-free peptide synthesis system by genetic code reprogramming [84–86] to introduce glycolic acid and non-proteinogenic amino acids into a peptide, in which protein synthesis using recombinant elements (PURE) system is used [87] (Scheme 20). These glycolic acid and non-proteinogenic amino acids were charged onto any desired tRNA by treatment with flexyzyme (highly flexible tRNA acylation ribozymes) [88]. In this system the CPE peptides **64** were transformed into the corresponding cyclic peptides **66** in the presence of peptide deformylase (PDF) and methionine aminopeptidase (MAP), which removed the N-terminal formyl group and the methionine residue to expose the N-terminal amino group, allowing intramolecular cyclization to proceed via the DKP thioester **65**.



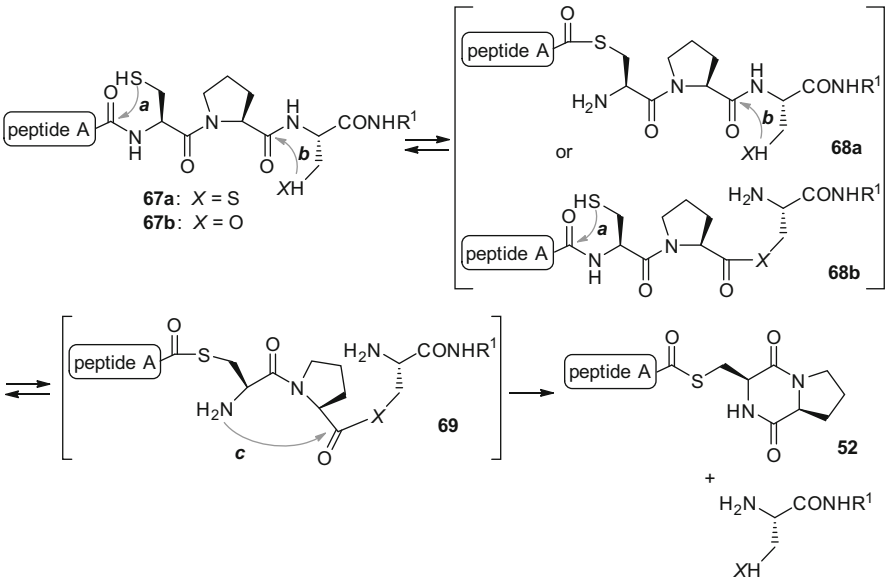
Scheme 19 Preparation of a CPE peptide



Scheme 20 Synthesis of CPE peptide in the translational system and formation of the cyclic peptide library

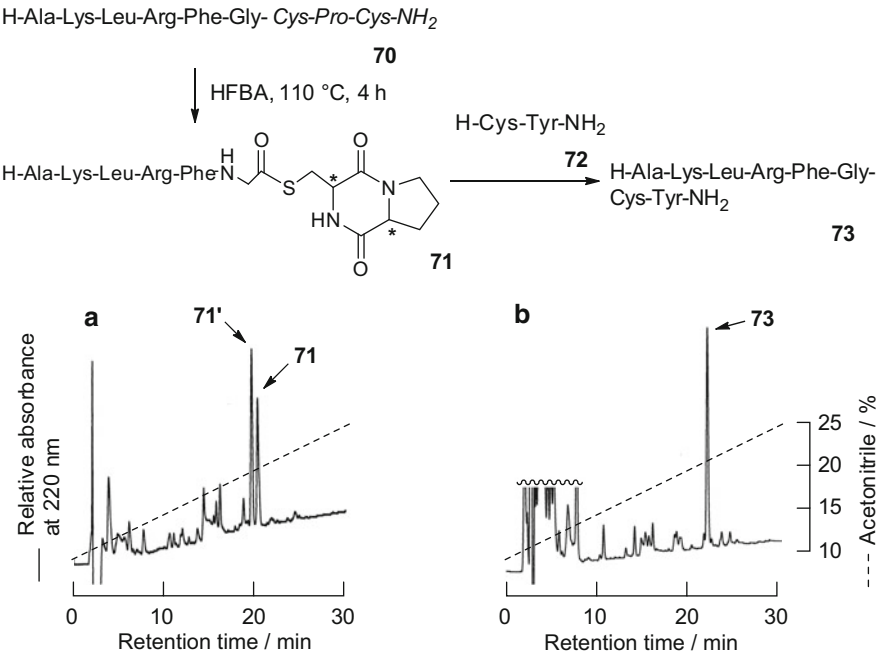
3.7 Cysteinylprolylcysteine (CPC)

The CPE structure contains an ester moiety, and is produced by chemical synthesis or by the special translational system, shown above. On the other hand, it was found that a peptide containing a Cys-Pro-Cys sequence (CPC peptide) without an ester group is also transformed into the peptide DKP thioester (Scheme 21) [89].



Scheme 21 Proposed reaction mechanism for the formation of DKP thioester from CPC or CPS peptide

An initial search for the thioester-forming sequence indicated that a peptide **67** containing a Cys-Pro-Cys/Ser sequence appeared to be converted into the DKP thioester. An *N*–*S*/*O* acyl shift at the second Cys or Ser residue (path ***b***) would produce a Cys-Pro (thio)ester structure **68b**. Once the CPE structure is formed, the DKP thioester **52** would be obtained via an *N*–*S* acyl shift at the first Cys residue (path ***a***) followed by DKP formation (path ***c***). The order of reactions '***a***' and '***b***' would not be critical for the overall reaction. As of this writing, DKP thioester formation from the CPC peptide was demonstrated only in model systems (Scheme 22). When the CPC peptide, H-Ala-Lys-Leu-Arg-Phe-Gly-Cys-Pro-Cys-NH_2 (**70**), was treated at 110°C under acidic conditions, dilute hydrochloric acid or heptafluorobutyric acid (HFBA), the corresponding DKP thioester **71** was obtained. Although epimerization in the DKP moiety was observed, the reaction mixture was reacted with cysteinyl peptide, H-Cys-Tyr-NH_2 (**72**), to produce **73** as a single isomer.

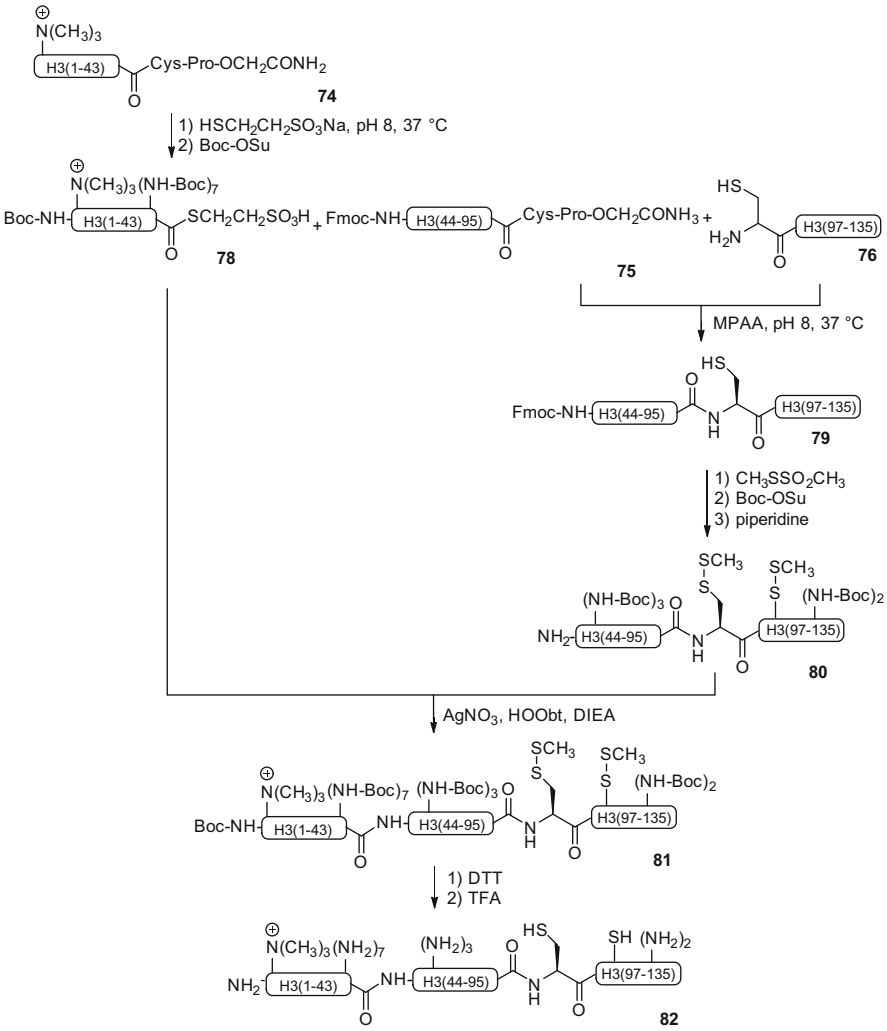


Scheme 22 DKP thioester formation from CPC peptide. RP-HPLC profiles of reaction mixtures: (**a**) after treatment of CPC peptide in HFBA at 110°C for 4 h; (**b**) after ligation of the crude DKP thioester with cysteinyl peptide. Column: YMC-Pack Pro C18 (4.6 × 150 mm); eluent: 0.1% TFA in aq CH_3CN, 1.0 mL/min

4 Protein Synthesis via *N* to *S* Acyl Shift Based on the CPE System

The Fmoc SPPS procedure can be used in peptide thioester synthesis based on intramolecular *N*–*S* acyl shift reactions. One of advantages of using Fmoc SPPS is that post-translationally modified peptides, i.e., phosphorylated, glycosylated peptides can be produced. Post-translational modifications play major roles in the regulation of numerous biological processes.

Histones are core proteins which form nucleosomes with DNA, in which an octamer of two copies of each of four different histones H2A, H2B, H3, and H4 are wrapped by DNA. This nucleosome is a basic unit of chromatin. Post-translational modifications of histones, such as acetylation, methylation, and phosphorylation, play an important role in gene regulation [90]. Although the synthesis of histones H2B [91] and H4 [92, 93], which contain no Cys residues, has been reported based on the combination of NCL and a desulfurization strategy by several groups, the synthesis of the Cys residue-containing histone H3 is difficult and has enjoyed limited success [72, 92, 94]. Recently, one-pot synthesis of H3 using peptide hydrazide as a peptide thioester precursor was reported [95].



Scheme 23 Synthesis of trimethyllysine-containing histone H3 by sequential ligation based on the use of CPE peptides

Here, synthesis of such a modified histone is presented as a representative example in the use of CPE peptide as a peptide thioester precursor [96]. H3, which contains a total of 135 amino acid residues with an N^{ϵ}-trimethyl Lys^{9} residue, was synthesized. H3 was divided for synthetic purposes into three segments, [Lys $(Me_3)^{9}$]H3(1–43)-Cys-Pro-OCH_2CONH_2 (**74**), Fmoc-H3(44–95)-Cys-Pro-OCH_2CONH_2 (**75**), and H3(96–135) (**76**), which were sequentially ligated by the native chemical ligation and thioester methods (Scheme 23). These segments were prepared by Fmoc SPPS. Peptides **75** and **76** were first ligated under NCL

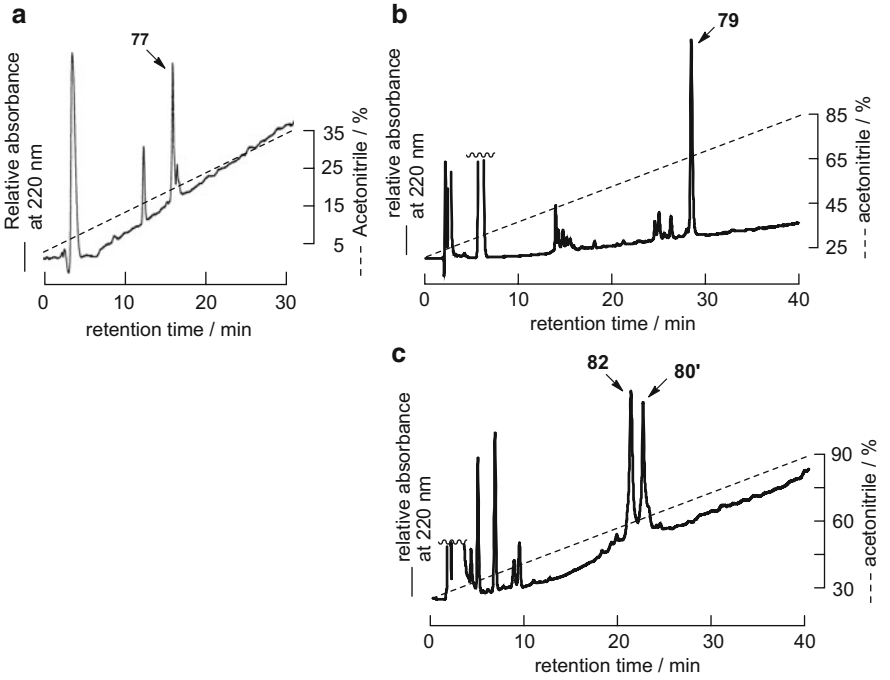


Fig. 5 RP-HPLC of (**a**) reaction mixture of the thioesterification of CPE peptide **74**, (**b**) reaction mixture of the ligation of CPE peptide **75** and cysteinyl peptide **76**, (**c**) final reaction mixture after removal of the protecting groups. Column: YMC-Pack ProC18 (4.6 × 150 mm) (**a**) or YMC-Pack C8 (4.6 × 150 mm) (**b**, **c**), eluent: aq CH_3CN containing 0.1% TFA, flow rate: 1.0 mL/min. **80′**: H3 (44–135).

conditions to give the Fmoc-H3(44–135) (**79**) (Fig. 5b). The thiol and amino groups of **79** were protected by methylthio and Boc groups, respectively, and the N-terminal Fmoc group was then removed to give [$Cys(SCH_3)^{96,110}$, $Lys(Boc)^{56,64,79,115,122}$]H3(44–135) (**80**). Prior to the second ligation, the CPE peptide **74** was transformed into the corresponding thioester, [$Lys(Me_3)^9$]H3(1–43)-$SCH_2CH_2SO_3H$ (**77**), by treatment with MESNa (Fig. 5a), and Boc groups were introduced to the amino group, resulting in the production of Boc-[$Lys(Boc)^{4,14,18,23,27,36,37}$, $Lys(Me_3)^9$]H3(1–43)-$SCH_2CH_2SO_3H$ (**78**). The thioester **78** was then ligated with peptide **80** in DMSO under the conditions of the thioester method. Although this ligation reaction proceeded very slowly, probably as the result of the properties of this protein, [$Lys(Me_3)^9$]H3 (**82**) was eventually obtained after removing the protecting groups (Fig. 5c).

Conclusions

Chemical protein synthesis has progressed rapidly in the period following the introduction of the peptide thioester. The thioester method and NCL are major methods for ligation using peptide thioesters as building blocks. The recent demand for preparing peptides and proteins that are post-translationally modified has prompted the development of synthetic methods for the peptide thioesters using the Fmoc-SPPS. One solution is to utilize intramolecular *N*–*S* acyl shift reactions to form thioesters, in which the peptide chain is elongated on a solid support by Fmoc-SPPS and the peptide thioester is then formed by an intramolecular *N*–*S* acyl shift as the key reaction, followed by intermolecular transthioesterification. This procedure corresponds to the reverse of the thiol-mediated ligation of peptides, such as NCL. In nature a biological *N*–*S* acyl shift is the initial step for protein splicing [97]. In this case, the enzymatic action of the intein domain promotes an *N*–*S* acyl shift and transthioesterification. Intein-mediated thioester formation is also used in the production of recombinant peptide thioesters for expressed protein ligation [98]. In chemical conditions, the amide and thioester forms are in equilibrium, and these species can be detected by ^{13}C NMR and RP-HPLC. Various types of skeletons containing the thiol group were developed for intramolecular *N*–*S* acyl shifts, including the use of cysteine residues in the synthesis of peptide thioesters. Using these methods, various types of peptide thioesters can be synthesized and applied to protein synthesis as building blocks. Chemical synthesis makes it possible to introduce a wide variety of modifications in proteins which are synthesized using this methodology. These novel methodologies have great potential for advancing protein science in general.

References

1. Merrifield RB (1963) J Am Chem Soc 85:2149
2. Giralt E et al (2004) Synthesis of peptides on solid supports. In: Goodman M, Felix A, Moroder L, Toniolo C (eds) Synthesis of peptides and peptidomimetics E22a. Georg Thieme Verlag, Stuttgart
3. Hojo H (2014) Curr Opin Struct Biol 26:16
4. Hackenberger CPR, Schwarzer D (2008) Angew Chem Int Ed 47:10030
5. Aimoto S (2001) Curr Org Chem 5:45
6. Dawson PE, Kent SBH (2000) Annu Rev Biochem 69:923
7. Hojo H, Aimoto S (1991) Bull Chem Soc Jpn 64:111
8. Hojo H, Kwon Y, Kakuta Y, Tsuda S, Tanaka I, Hikichi K, Aimoto S (1993) Bull Chem Soc Jpn 66:2700
9. Kawakami T, Kogure S, Aimoto S (1996) Bull Chem Soc Jpn 69:3331
10. Aimoto S (1999) Biopolymers (Pept Sci) 51:247
11. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Science 266:776
12. Yan LZ, Dawson PE (2001) J Am Chem Soc 123:526
13. Wan Q, Danishefsky SJ (2007) Angew Chem Int Ed 46:9248

14. Tam JP, Yu Q (1998) Biopolymers (Pept Sci) 46:319
15. Botti P, Tchertchian S (2006) WO 2006/133,962
16. Chen J, Wan Q, Yuan Y, Zhu J, Danishefsky SJ (2008) Angew Chem Int Ed 47:8521
17. Haase C, Rohde H, Seitz O (2008) Angew Chem Int Ed 47:6807
18. Yang R, Pasunooti KK, Li F, Liu XW, Liu CF (2009) J Am Chem Soc 131:13592
19. Tan Z, Shang S, Danishefsky SJ (2010) Angew Chem Int Ed 49:9500
20. Chen J, Wang P, Zhu J, Wan Q, Danishefsky SJ (2010) Tetrahedron 66:2277
21. Shang S, Tan Z, Dong S, Danishefsky SJ (2011) J Am Chem Soc 133:10784
22. Ding H, Shigenaga A, Sato K, Morishita K, Otaka A (2011) Org Lett 13:5588
23. Simon P, Karthikeyan SV, Brik A (2012) Org Lett 14:1520
24. Malins LR, Cergol KM, Payne RJ (2013) ChemBioChem 14:559
25. Rohde H, Seitz O (2010) Biopolymer (Pept Sci) 94:551
26. Canne LE, Bark SJ, Kent SBH (1996) J Am Chem Soc 118:5891
27. Botti P, Carrasco MR, Kent SBH (2001) Tetrahedron Lett 42:1831
28. Kawakami T, Akaji K, Aimoto S (2001) Org Lett 3:1403
29. Vizzavona J, Dick F, Vorherr T (2002) Bioorg Med Chem Lett 12:1963
30. Offer J, Boddy CNC, Dawson PE (2002) J Am Chem Soc 124:4642
31. Kawakami T, Aimoto S (2003) Tetrahedron Lett 44:6059
32. Offer J (2010) Biopolymer (Pept Sci) 94:530
33. Hojo H, Nakahara Y (2000) Curr Protein Pept Sci 1:23
34. Seitz O, Heinemann I, Mattes A, Waldmann H (2001) Tetrahedron 57:2247
35. Grogan MJ, Pratt MR, Marcaurelle LA, Bertozzi CR (2002) Annu Rev Biochem 71:593
36. Kawakami T, Hasegawa K, Aimoto S (2000) Bull Chem Soc Jpn 73:197
37. Li X, Kawakami T, Aimoto S (1998) Tetrahedron Lett 39:8669
38. Hasegawa K, Sha YL, Bang JK, Kawakami T, Akaji K, Aimoto S (2001) Lett Pept Sci 8:277
39. Futaki S, Sogawa K, Maruyama J, Asahara T, Niwa M, Hojo H (1997) Tetrahedron Lett 38:6237
40. Ingenito R, Bianchi E, Fattori D, Pessi A (1999) J Am Chem Soc 121:11369
41. Shin Y, Winans KA, Backes BJ, Kent SBH, Ellman JA, Bertozzi CR (1999) J Am Chem Soc 121:11684
42. Mezzato S, Schaffrath M, Unverzagt C (2005) Angew Chem Int Ed 44:1650
43. Flavell RR, Huse M, Goger M, Trester-Zedlitz M, Kuriyan J, Muir TW (2002) Org Lett 4:165
44. Sewing A, Hilvert D (2001) Angew Chem Int Ed 40:3395
45. Camarero JA, Hackel BJ, de Yoreo JJ, Mitchell AR (2004) J Org Chem 69:4145
46. Brask J, Albericio F, Jensen KJ (2003) Org Lett 5:2951
47. Botti P, Villain M, Manganiello S, Gaertner H (2004) Org Lett 6:4861
48. Warren JD, Miller JS, Keding SJ, Danishefsky SJ (2004) J Am Chem Soc 126:6576
49. Blanco-Canosa JB, Dawson PE (2008) Angew Chem Int Ed 47:6851
50. Tofteng AP, Sorensen KK, Conde-Frieboes KW, Hoeg-Jensen T, Jensen KJ (2009) Angew Chem Int Ed 48:7411
51. Fang GM, Li YM, Shen F, Huang YC, Li JB, Lin Y, Cui HK, Liu L (2011) Angew Chem Int Ed 50:7645
52. Okamoto R, Morooka K, Kajihara Y (2012) Angew Chem Int Ed 51:191
53. Zheng JS, Tang S, Huang YC, Liu L (2013) Acc Chem Res 46:2475
54. Mende F, Seitz O (2011) Angew Chem Int Ed 50:1232
55. Kang J, Macmillan D (2010) Org Biomol Chem 8:1993
56. Nakamura K, Sumida M, Kawakami T, Vorherr T, Aimoto S (2006) Bull Chem Soc Jpn 79:1773
57. Sakakibara S, Shimonishi Y, Okada M, Kishida Y (1967) Removal of protective groups by anhydrous hydrogen fluoride. In: Beyerman HC, van de Linde A, van den Brink WM (eds) Peptides: proceedings of the 8th European peptide symposium, North-Holland Publishing Company, Amsterdam
58. Ollivier N, Behr JB, El-Mahdi O, Blanpain A, Melnyk O (2005) Org Lett 7:2647

59. Ohta Y, Itoh S, Shigenaga A, Shintaku S, Fujii N, Otaka A (2006) Org Lett 8:467
60. Tsuda S, Shigenaga A, Bando K, Otaka A (2009) Org Lett 11:823
61. Sato K, Shigenaga A, Tsuji K, Tsuda S, Sumikawa Y, Sakamoto K, Otaka A (2011) ChemBioChem 12:1840
62. Kawakami T, Sumida M, Nakamura K, Vorherr T, Aimoto S (2005) Tetrahedron Lett 46:8805
63. Nakamura K, Mori H, Kawakami T, Hojo H, Nakahara Y, Aimoto S (2007) Int J Pept Res Ther 13:191
64. Nakamura K, Kanao T, Uesugi T, Hara T, Sato T, Kawakami T, Aimoto S (2009) J Pept Sci 15:731
65. Johnson ECB, Kent SBH (2006) J Am Chem Soc 128:6640
66. Nagaike F, Onuma Y, Kanazawa C, Hojo H, Ueki A, Nakahara Y, Nakahara Y (2006) Org Lett 8:4465
67. Sharma RK, Tam JP (2011) Org Lett 13:5176
68. Hojo H, Onuma Y, Akimoto Y, Nakahara Y, Nakahara Y (2007) Tetrahedron Lett 48:25
69. Hojo H, Kobayashi H, Ubagai R, Asahina Y, Nakahara Y, Katayama H, Ito Y, Nakahara Y (2011) Org Biomol Chem 9:6807
70. Taichi M, Hemu X, Qiu Y, Tam JP (2013) Org Lett 15:2620
71. Ollivier N, Dheur J, Mhidia R, Blanpain A, Melnyk O (2010) Org Lett 12:5238
72. Hou W, Zhang X, Li F, Liu CF (2011) Org Lett 13:386
73. Dheur J, Ollivier N, Melnyk O (2011) Org Lett 13:1560
74. Zheng JS, Chang HN, Wang FL, Liu L (2011) J Am Chem Soc 133:11080
75. Kang J, Richardson JP, Macmillan D (2009) Chem Commun 407
76. Kang J, Reynolds NL, Tyrrell C, Dorin JR, Macmillan D (2009) Org Biomol Chem 7:4918
77. Adams AL, Cowper B, Morgan RE, Premdjee B, Caddick S, Macmillan D (2013) Angew Chem Int Ed 52:13062
78. Burlina F, Papageorgiou G, Morris C, White PD, Offer J (2014) Chem Sci 5:766
79. Zanotti G, Pinnen F, Lucente G (1985) Tetrahedron Lett 26:5481
80. Kawakami T, Aimoto S (2007) Chem Lett 36:76
81. Kawakami T, Aimoto S (2009) Tetrahedron 65:3871
82. Sui Q, Borchardt D, Rabenstein DL (2007) J Am Chem Soc 129:12042
83. Kawakami T, Kamauchi A, Harada E, Aimoto S (2014) Tetrahedron Lett 55:79
84. Kawakami T, Ohta A, Ohuchi M, Ashigai H, Murakami H, Suga H (2009) Nat Chem Boil 5:888
85. Ohshiro Y, Nakajima E, Goto Y, Fuse S, Takahashi T, Doi T, Suga H (2011) ChemBioChem 12:1183
86. Kang TJ, Hayashi Y, Suga H (2011) Angew Chem Int Ed 50:2159
87. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T (2001) Nat Biotech 19:751
88. Murakami H, Ohta A, Ashigai H, Suga H (2006) Nat Methods 3:357
89. Kawakami T, Shimizu S, Aimoto S (2010) Bull Chem Soc Jpn 83:570
90. Jenuwein TC, Allis D (2001) Science 293:1074
91. Chiang KP, Jensen MS, McGinty RK, Muir TW (2009) ChemBioChem 10:2182
92. He S, Bauman D, Davis JS, Loyola A, Kishioka K, Gronlund JL, Reinberg D, Meng F, Kelleher N, McCafferty DG (2003) Proc Natl Acad Sci U S A 100:12033
93. Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL (2006) Science 311:844
94. Shimko JC, North JA, Bruns AN, Poirier MG, Ottesen JJ (2011) J Mol Biol 408:187
95. Li J, Li Y, He Q, Li Y, Lia H, Liu L (2014) Org Biomol Chem 12:5435
96. Kawakami T, Akai Y, Fujimoto H, Kita C, Aoki Y, Konishi T, Waseda M, Takemura L, Aimoto S (2013) Bull Chem Soc Jpn 86:690
97. Anraku Y, Satow Y (2009) Proc Jpn Acad Ser B 85:409
98. Muralidharan V, Muir TW (2006) Nat Methods 3:429

Top Curr Chem (2015) 362: 137–182
DOI: 10.1007/128_2014_582

Published online: 3 December 2014

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

[$]Both authors contributed equally to this work.

A. Yang, L. Zhao, and Y.-W. Wu (✉)
Chemical Genomics Centre of the Max Planck Society, Otto-Hahn-Str. 15, 44227 Dortmund, Germany
e-mail: yaowen.wu@mpi-dortmund.mpg.de

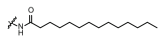
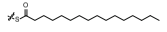
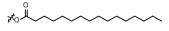
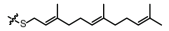
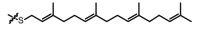
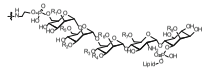
1 Introduction

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.

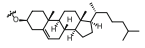
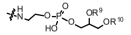
The four major types of protein lipidation are *N*-myristoylation, palmitoylation, prenylation and glycosylphosphatidylinositol-anchor (GPI-anchor) addition (Table 1).

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
N-Myristoylation		*N*-Myristoyl-transferase (NMT)	N-terminal of Gly	No	Mitochondrial membrane, Intracellular plasma membrane	Arf, Actin, PAK2
S-Palmitoylation		Palmitoyl Protein Thioesterases (PPT)	Side chain of Cys	Yes	Intracellular plasma membrane, Vesicle	H-Ras, Gsα
O-Palmitoylation		*O*-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 intein 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 Ca^{2+} 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 λ3. In the GTP-bound form, the extrusion of loop λ3 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. TGFα, adrenergic receptors, Rhodopsin), viral proteins (e.g. Influenza HA, HIV-1 gp160), Ras proteins, Gα subunits, Src family tyrosine kinases (Src, Lck, Fyn, Hck, Lyn and Yes), etc. The Gα 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 PDEδ [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, $NH_2(CH_2)_2OPO_3H$-6Manα1 → 2Manα1 → 6Manα1 → 4GlcNα1 → 6myo-Ino1-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, FcγRIII, 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 glycophospholipid 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].

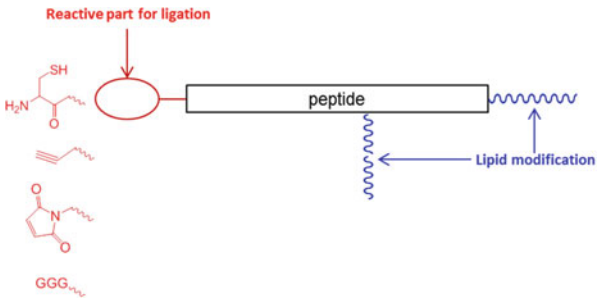
1.5 Other Types of Lipidation

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



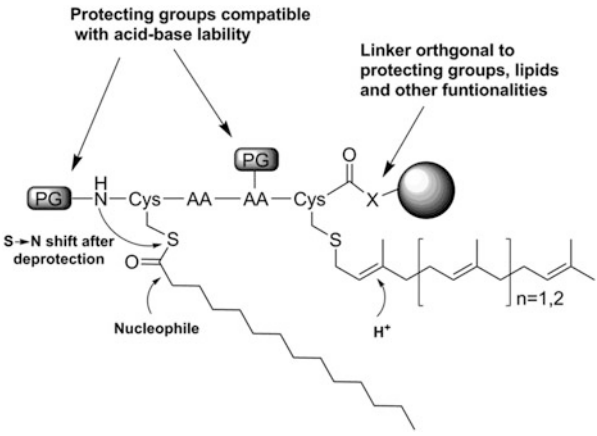
Scheme 1 Overview of the lipidated peptide for ligation. In general, the lipidated peptide for ligation consists of three parts: peptide, lipid group and N-terminal reactive group (natural Cys and triglycine, and non-natural maleimide and alkyne moieties)

proteins by chemical synthesis. These ligation methods, including expressed protein ligation (EPL), maleimidocaprolyl (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.

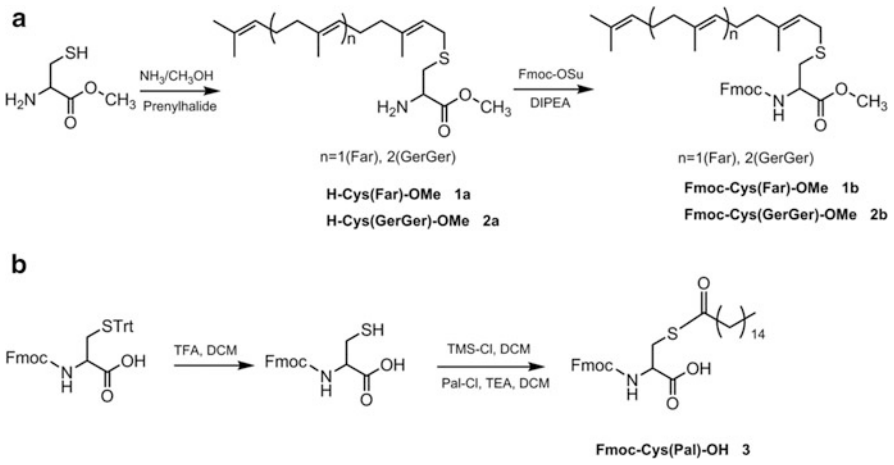


Scheme 2 General considerations for the synthesis of lipidated peptides of Ras family proteins. *PG* protecting group

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 → 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 → 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].



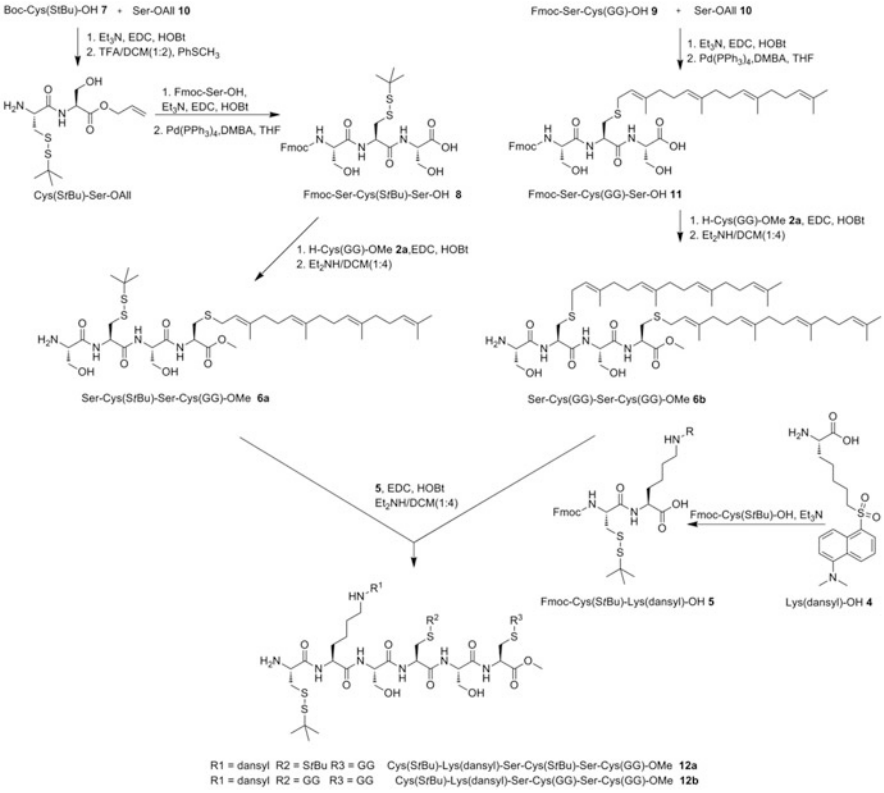
Scheme 3 Synthesis of lipidated cysteine building block

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 Pd^0-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

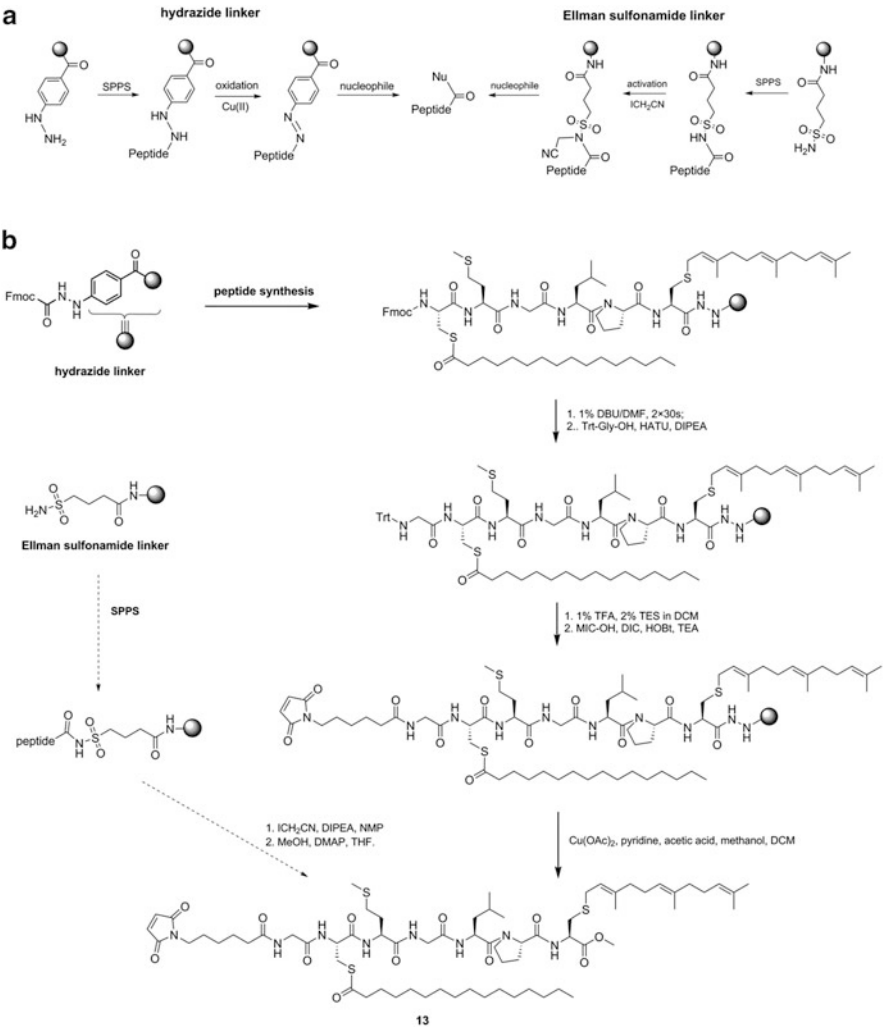
Solid-phase peptide synthesis (SPPS) is a fast and flexible approach and has been frequently used for the synthesis of lipidated peptides. SPPS allows for preparation of the desired peptides, including both natural and nonnatural modifications, with high purity and good yields in a short time. The linker 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.

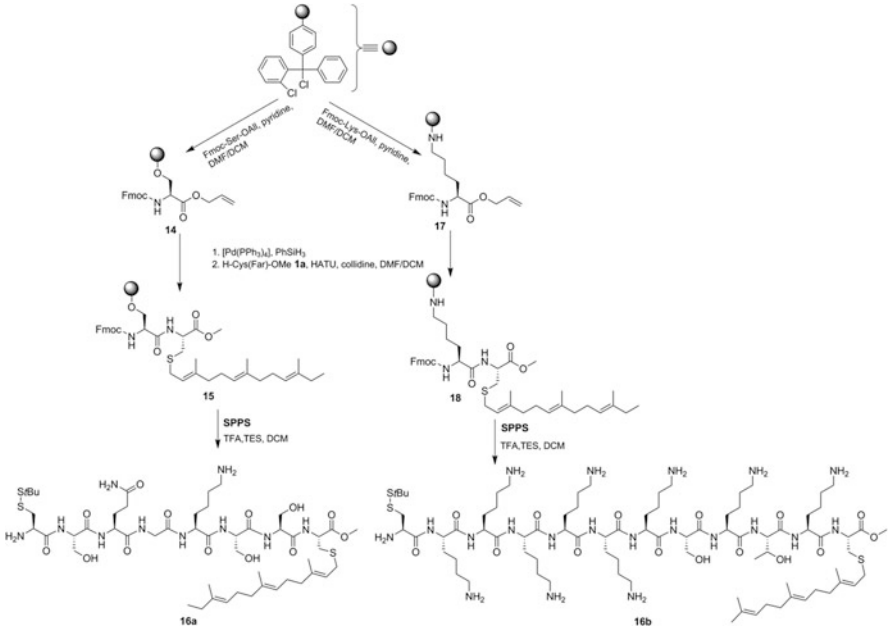
Not so many linkers meet all these requirements. Among them, the hydrazide linker and the Ellman sulfonamide linker are stable under acid and basic conditions, permitting the synthesis of the peptides and their orthogonal release from the solid support [47]. The hydrazide linker can be cleaved by oxidation 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 $Cu(AcO)_2/O_2$ or NBS. Such cleavage conditions are orthogonal to prenyl and palmitoyl groups and classical protecting groups (Boc, Fmoc and Aloc). 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 hydrazide linker and Ellman sulphonamide linker. (**a**) Cleavage of the hydrazide linker by oxidation and nucleophilic attack and Ellman sulphonamide linker by activation and nucleophilic attack of a nucleophile. (**b**) Synthesis of farnesylated and palmitoylated N-Ras C-terminus with maleimido group using hydrazide linker and Ellman sulphonamide 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 racemerization 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].

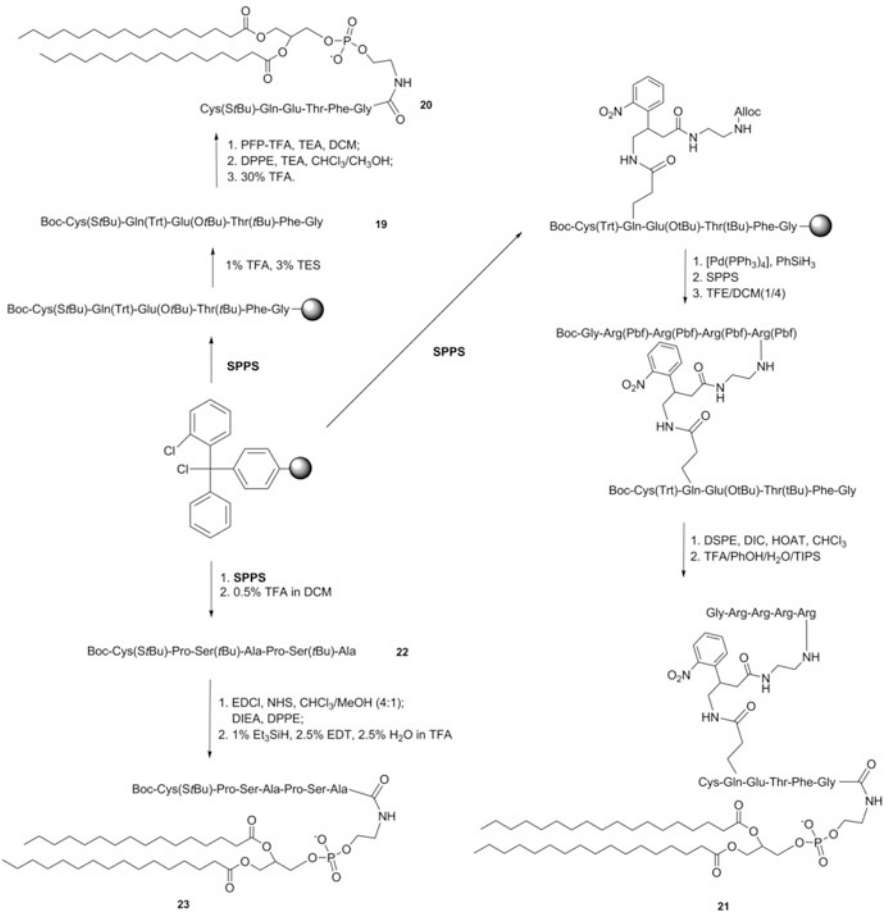
Another linker successfully applied to the synthesis of lipidated peptides is the trityl linker, which can be cleaved by treatment with low concentrations of acid (1% TFA) without affecting the integrity of the prenyl group. The main disadvantage of this linker is that the cleavage at the C-terminus 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-OAll 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

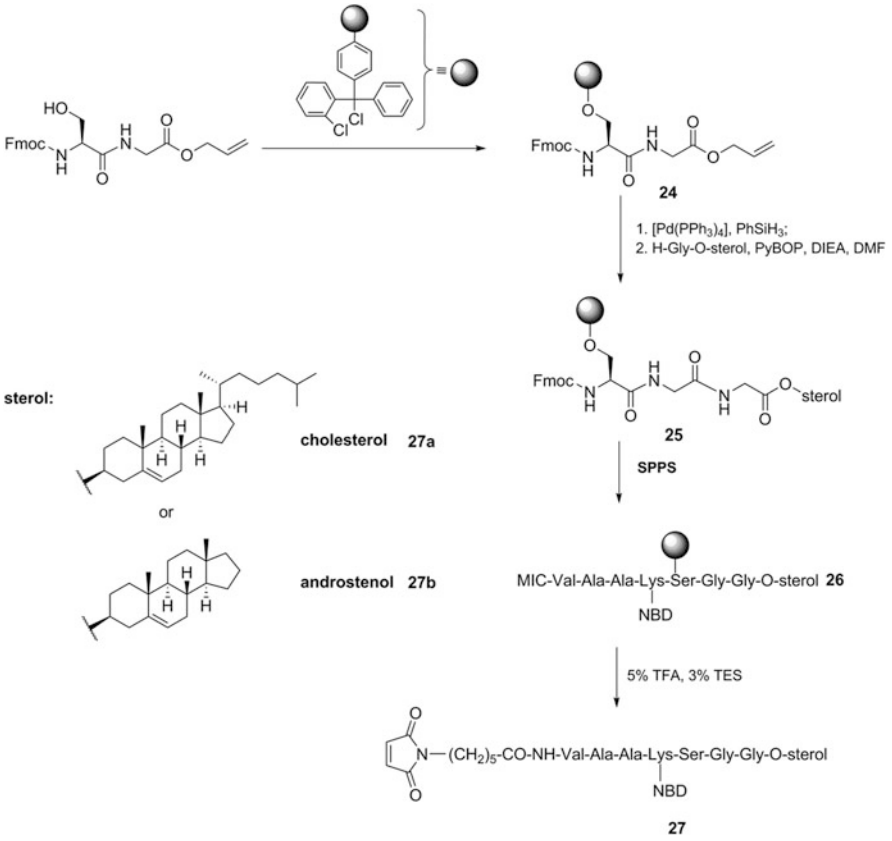
Synthesis of phosphatidylethanolaminylated (PE) peptide is performed by lipidation of peptide backbone in solution, which was synthesized using the trityl linker strategy via SPPS. PE-conjugated C-terminal peptide of LC3 (**20**) was synthesized on the chlorotrityl resin by means of the Fmoc strategy (Scheme 7) [62]. After release from the resin, the protected peptide **19** was subsequently activated by pentafluorphenyl trifluoracetate 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 (EDCl) 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 chlorotrityl 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 glycyl-sterol esters. The glycyl 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).

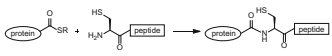
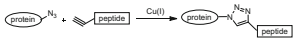
Each chemical ligation method, involving expressed protein ligation (EPL), MIC ligation, Diels–Alder ligation, click ligation and sortase-mediated protein ligation, has its own pros and cons. The choice of the approach depends on the nature of the target protein and design of protein synthesis. The EPL 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 Lipidated peptide with an N-terminal cysteine		Native peptide bond	Rab7, N-Ras, Rheb, K-Ras4B, D-Ral, YPT1, LC3-PE, lipidated GFP
Maleimide ligation	Protein with a C-terminal cysteine Lipidated peptide with an N-terminal maleimide		Non-natural bond	H-Ras, N-Ras
Diels–Alder ligation	Protein with a C-terminal dienyl group Lipidated peptide with an N-terminal maleimide		Non-natural bond	Rab7
Click ligation	Protein with a C-terminal azide Lipidated peptide with an N-terminal alkyne moiety		Non-natural bond	Rab1, Rab7
Sortase-mediated protein ligation	Protein with a C-terminal LPXTG motif Lipidated peptide with an N-terminal oligoglycine sequence		Native peptide bond with additional amino acid sequence	K-Ras4B, lipidated GFP, GPI modified GFP

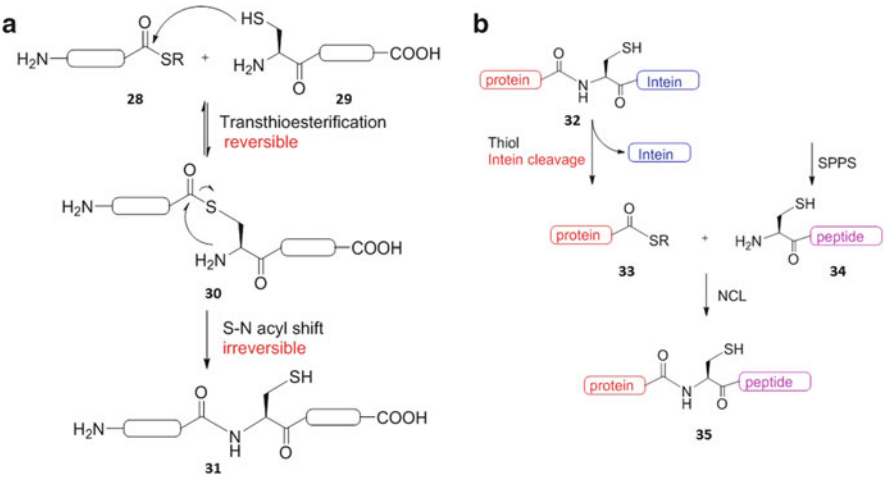
than others. Among these detergents, CTAB appears to be the most robust mediator of the ligation reaction and has been used to produce mono- and diprenylated Rab7 (Scheme 10a) [54].

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 → 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

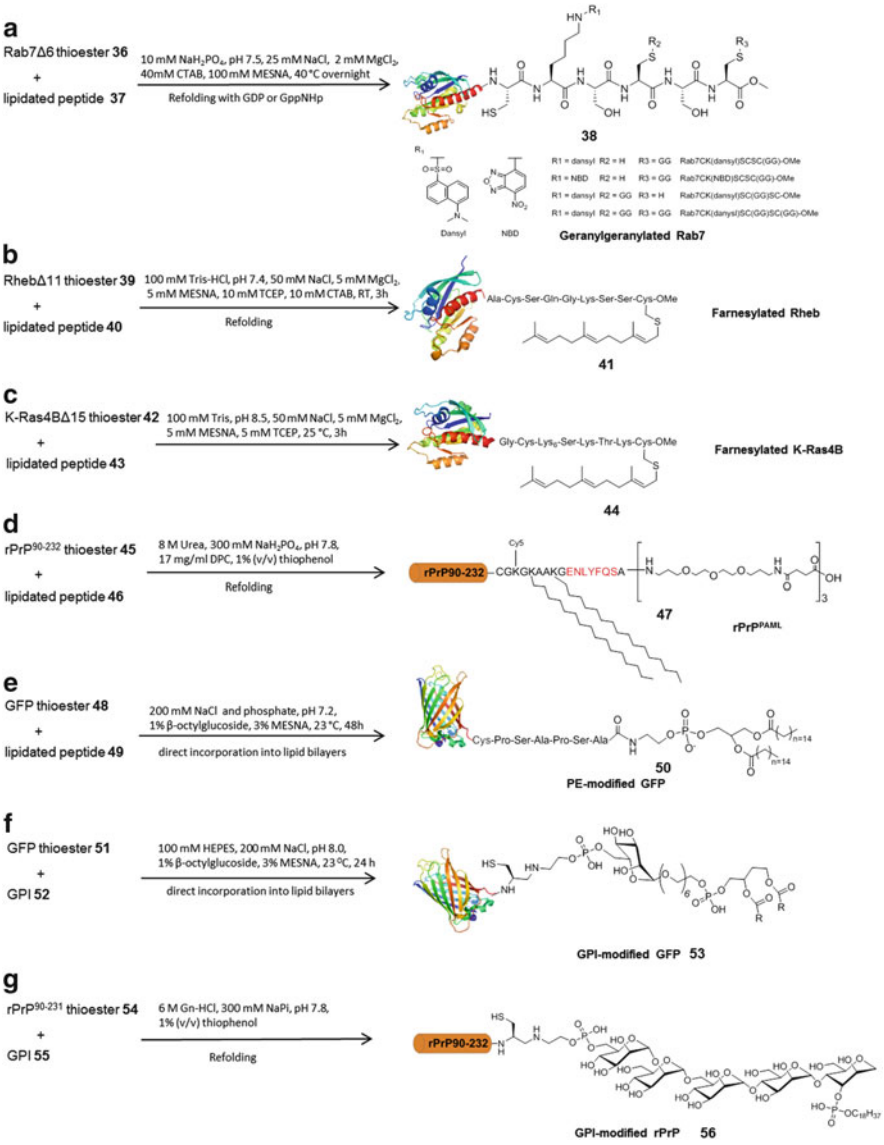
34 containing N-terminal cysteine. Until now, the EPL has been widely applied to produce proteins with post-translational modifications [84, 85].

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.

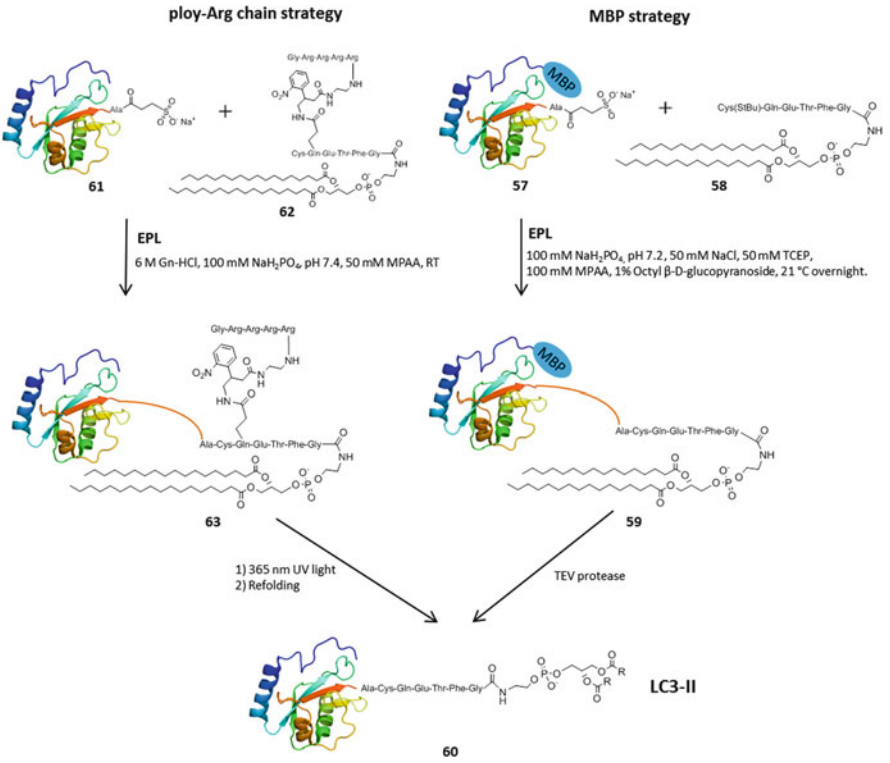
Many other examples of lipidated proteins generated via EPL include PrP^{Palm} **47** [76, 79], GFP-PE **50** [63], GPI-anchored proteins **53** [74, 88] and GPI-anchored PrP **56** [89] (Scheme 10). Bertozzi and co-workers prepared a series of GPI-protein analogues bearing different anchor structures 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**) $rPrP^{PALM}$, (**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)

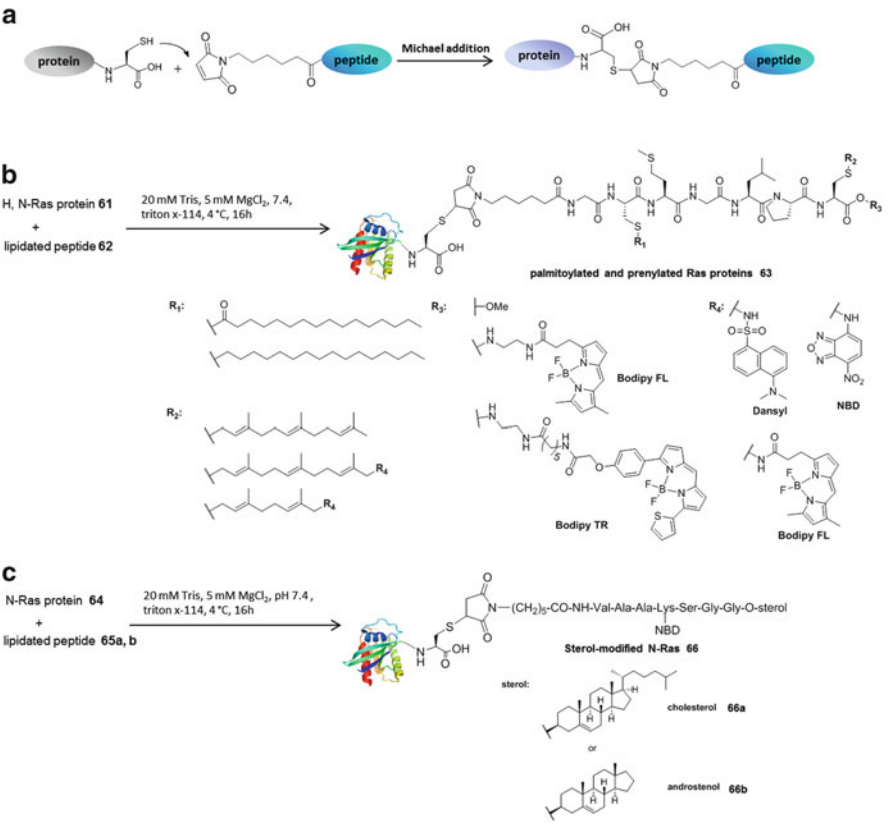


Scheme 11 Semisynthesis of LC3-PE by EPL

[90]. In the poly-Arg solubilisation strategy, the ligation of lipidated hexapeptide **62** with LC3$^{1–114}$-thioester **61** was performed under denaturing conditions. After cleavage of the ploy-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-LC3$^{1–114}$-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].

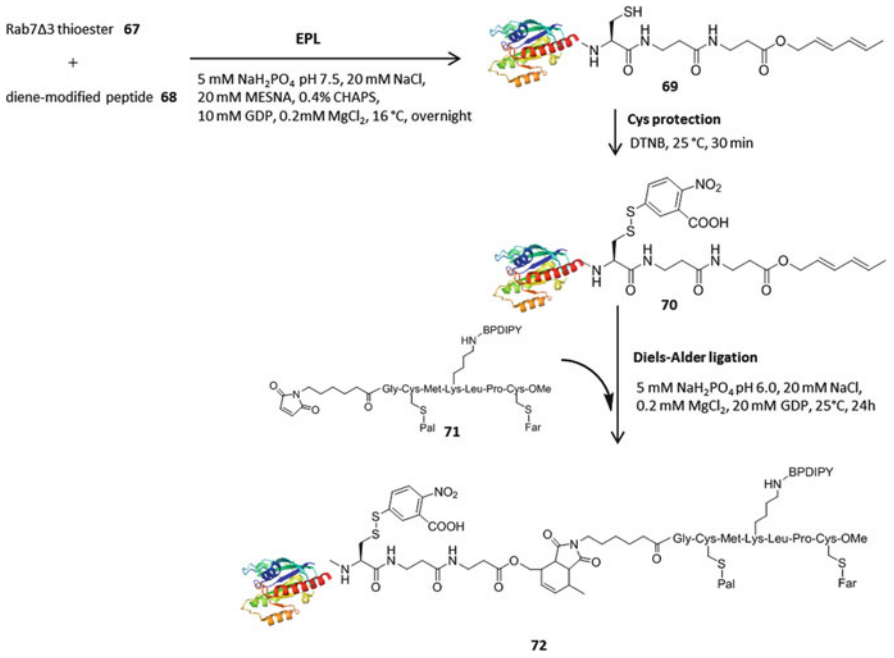
To understand the function of sterol anchors, fluorescently labelled heptapeptides **65a** and **65b** bearing sterol moieties were attached to the N-RasG12V(1–181) **64**, yielding the sterol-modified proteins **66a, b** (Scheme 12c). N-Ras was chosen as the protein moiety because this system offers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

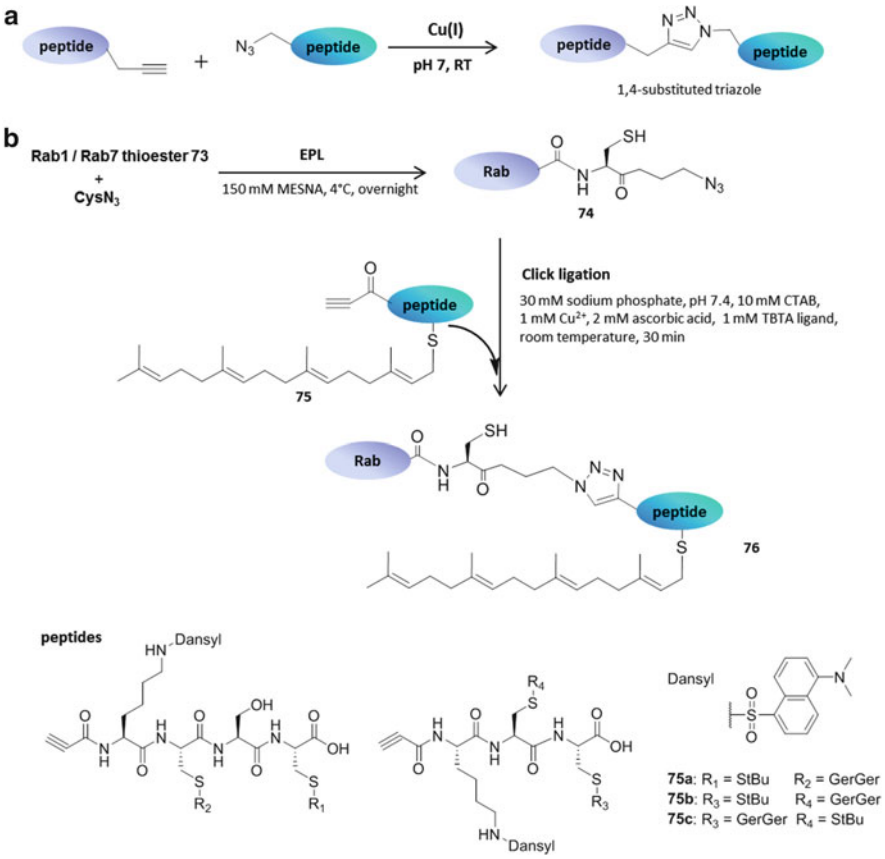
Cu(I)-catalyzed Huisgen 1,3-dipolar cycloaddition reaction, also referred to as the "click reaction", is widely employed in protein/peptide modifications [96]. The click reaction was applied to the synthesis of geranylgeranylated Rab1 and Rab7 proteins [73]. There are several advantages to using click 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, $CysN_3$, 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 μM 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

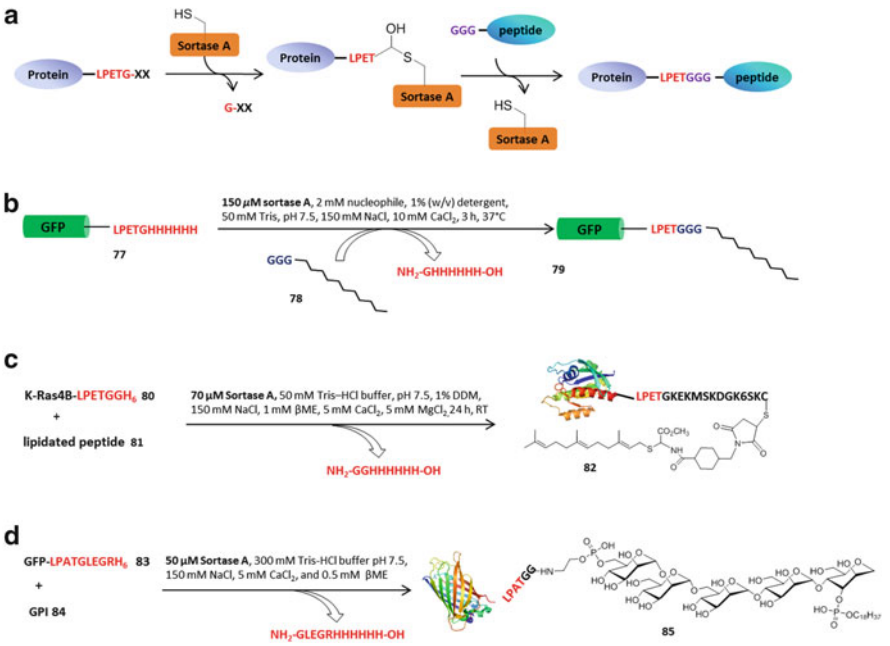


Scheme 14 Semisynthesis of geranylgeranylated Rab proteins by click ligation

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,

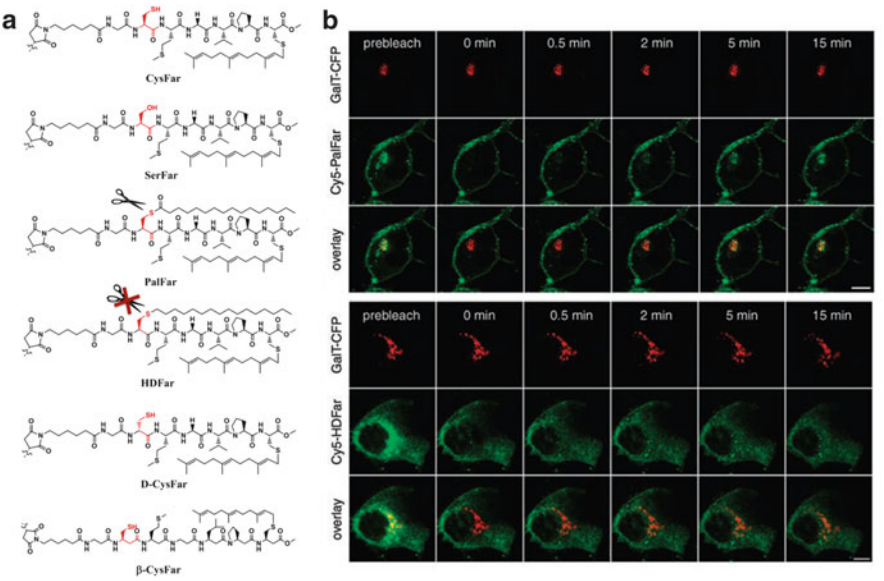
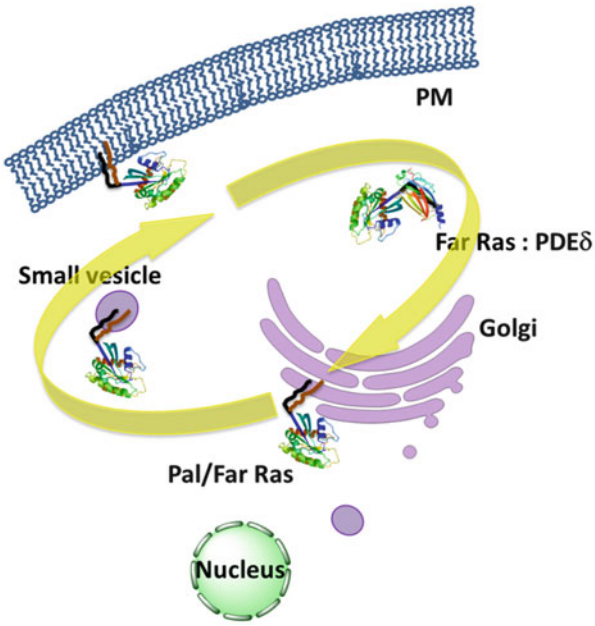


Fig. 1 (**a**) Semisynthetic N-Ras proteins with various C-terminal structures. (**b**) Cellular distribution and FRAP measurements of Cy5-N-Ras-PalFar and Cy5-N-Ras-HDFar at the Golgi. GalT is a Golgi marker



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, D-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

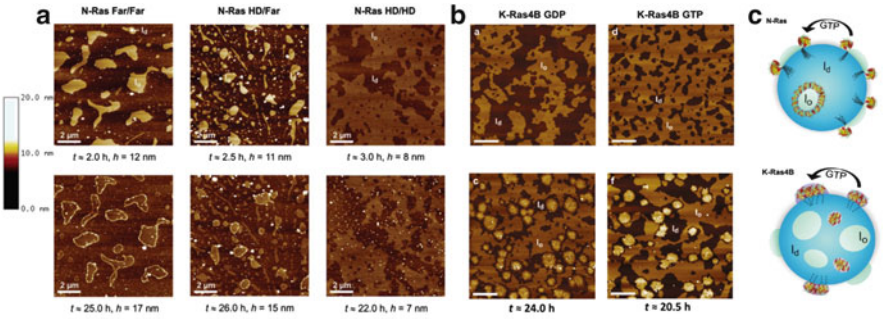


Fig. 2 (**a**) AFM images of the time-dependent partitioning of GDP-bound N-Ras Far/Far, N-Ras HD/Far, and N-Ras HD/HD into lipid bilayers consisting of DOPC/DPPC/Chol 1:2:1. (**b**) AFM images of the time-dependent partitioning of GDP- and GTP-bound K-Ras4B into lipid bilayers consisting of DOPC/DOPG/DPPC/DPPG/Chol 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:PDEδ complex for crystallization. PDEδ was initially identified as a fourth subunit of rod-specific cGMP phosphodiesterase, PDE6. Wittinghofer and co-workers showed that PDEδ can bind and solubilize prenylated Ras, Rheb, Rho6 and $G\alpha_{i1}$ [106]. The structural studies of F-Rheb:PDEδ complex provide insights into the function of PDEδ as a GDI-like solubilizing factor involved in the transport of farnesylated small GTPases [107].

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

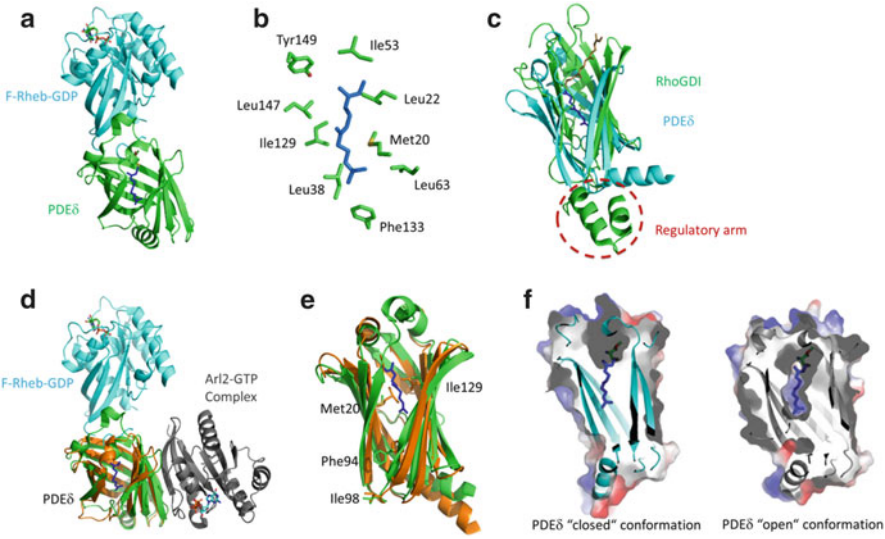


Fig. 3 Structure analysis of F-Rheb:PDEδ. (**a**) Ribbon representation of F-Rheb in *cyan*, with the farnesyl group in *blue*, in complex with PDEδ in *green* (PDB code 3T5G). GDP bound to Rheb is shown in *ball-and-stick* representation. (**b**) Residues forming the hydrophobic pocket of PDEδ are shown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

main-chain interactions together with the flexibility of this loop support the notion of broad specificity of PDEδ. 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 PDEδ 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 PDEδ, suggesting the interaction occurs mainly through the farnesylated C-terminus (Fig. 3c). These findings explain the nucleotide-independent binding of G proteins to PDEδ [106, 108]. PDEδ 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 PDEδ 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:PDEδ 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 PDEδ in cells. Therefore, Arl2 and Arl3 function as GDI-like displacement factors (GDFs), which allosterically regulate the release of farnesylated G proteins from PDEδ.

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, Rab7Δ6CK(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.

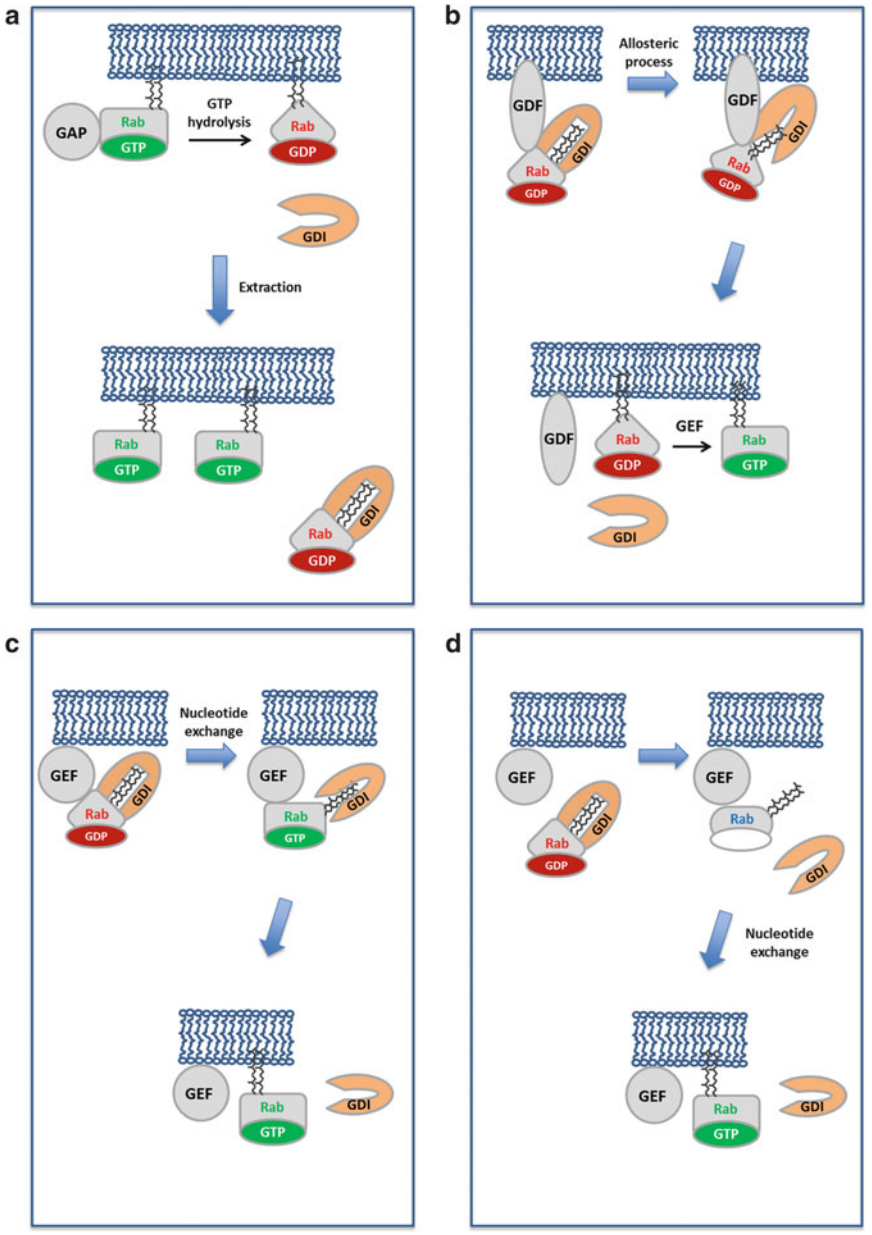
To study the relationship further between nucleotide exchange and Rab targeting to membranes, a RabGEF from *Legionella pneumophila* (DrrA) was used in investigating the effect of GEFs on the Rab:GDI complex. Kinetics of the complex interaction was monitored by a fluorescence change of Rab1-NBD-farnesyl (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

a

Nucleotide bound form	K_d Rab7NBD-G:REP-1	K_d Rab7NBD-G:GDI-1
GDP	0.15 nM	2.5 nM
GppNHp	244 nM	1717 nM

b

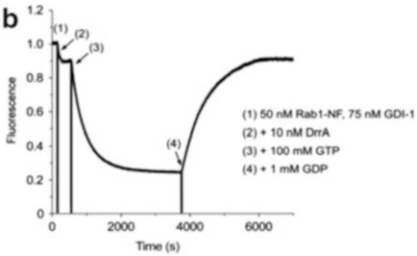


Fig. 4 Quantitative analysis of interaction of Rab7NBD-G with REP-1 and GDI-1. (**a**) K_d values of Rab7NBD-G interacting with REP-1 or GDI-1 in different nucleotide bound forms. (**b**) DrrA-mediated displacement of GDI-1. 50 nM Rab1-NF:GDI-1 complex was supplemented with 10 nM DrrA. Nucleotide exchange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

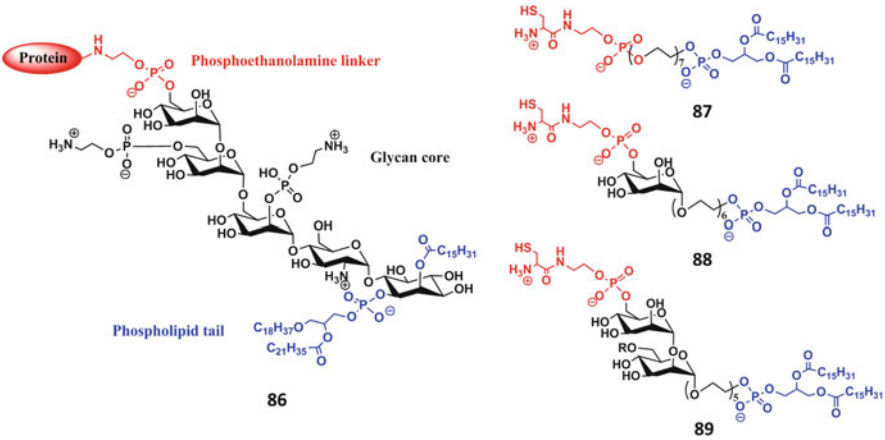


Fig. 5 Structures of native GPI-anchor (**86**) and GPI-anchor analogues (**87**), (**88**), and (**89**). These structures contain three domains of GDI-anchor: (1) a phosphoethanolamine linker (*red*), (2) the common glycan core (*black*) and (3) a phospholipid tail (*blue*). R is a GPI anchor side 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

a

Sample	τ_D (s)	D ($\mu m^2/s$)
GFP-2	6.0 ± 0.2	2.5 ± 0.1
GFP-3	5.3 ± 0.2	2.8 ± 0.1
GFP-4	4.8 ± 0.2	3.2 ± 0.1

c

Sample	τ_D (s)	D ($\mu m^2/s$)
GFP-GPI(DAF)	34.9 ± 4.9	0.43 ± 0.06
GFP-GDI(FR)	37.6 ± 4.7	0.40 ± 0.05
GFP-2	65.2 ± 11.3	0.23 ± 0.04
GFP-3	50.2 ± 6.7	0.30 ± 0.04
GFP-4	45.5 ± 5.5	0.33 ± 0.04

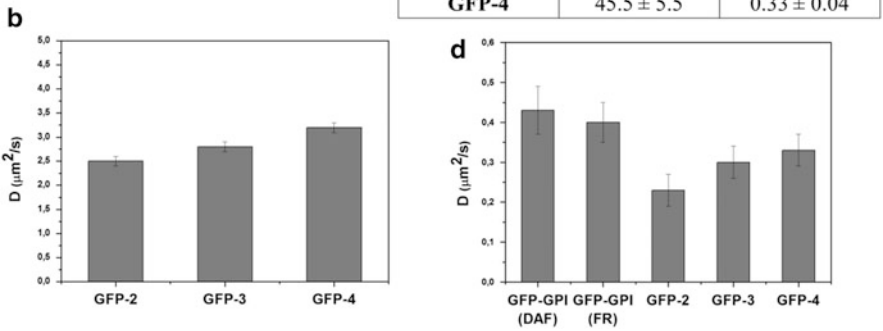


Fig. 6 Measurements of mobility by FCS. (**a**, **b**) The mobilities of GFP-2, GFP-3 and GFP-4 in a supported lipid bilayers. (**c** ,**d**) The mobilities of native GPI-anchored proteins GFP-GDI (DAF), GFP-GDI (FR) and GPI analogues GFP-2, GFP-3 and GFP-4 on HeLa cell surface

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

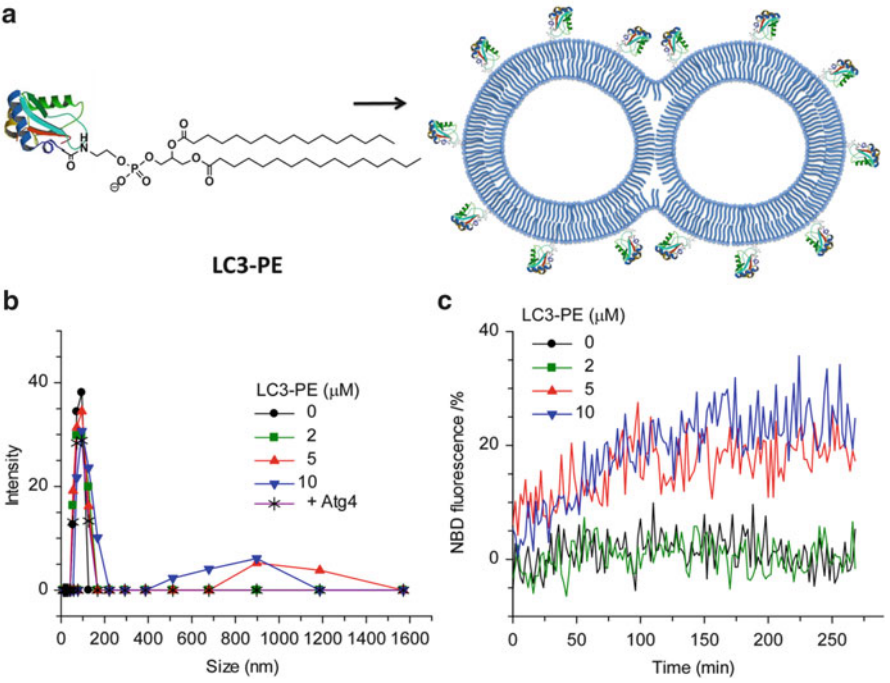


Fig. 7 Membrane tethering and fusion meditated by the semisynthetic MBP-LC3-PE protein in vitro. (**a**) A schematic view of LC3-PE-mediated liposomal hemifusion. (**b**) LC3-PE induces membrane tethering in a dose-dependent manner. (**c**) LC3-PE induces membrane fusion in a dose-dependent manner

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, Sarngadharan 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, Drisdel 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, Gormer 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, Szedlacsek 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, Winklhofer 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

Top Curr Chem (2015) 362: 183–228
DOI: 10.1007/128_2014_598

Published online: 24 February 2015

Protein Chemical Synthesis in Drug Discovery

Fa Liu and John P. Mayer

Abstract The discovery of novel therapeutics to combat human disease has traditionally been among the most important goals of research chemists. After a century of innovation, state-of-the-art chemical protein synthesis is now capable of efficiently assembling proteins of up to several hundred residues in length from individual amino acids. By virtue of its unique ability to incorporate non-native structural elements, chemical protein synthesis has been seminal in the recent development of several novel drug discovery technologies. In this chapter, we review the key advances in peptide and protein chemistry which have enabled our current synthetic capabilities. We also discuss the synthesis of D-proteins and their applications in mirror image phage-display and racemic protein crystallography, the synthesis of enzymes for structure-based drug discovery, and the direct synthesis of homogenous protein pharmaceuticals.

Keywords Drug discovery · Mirror image phage-display · Native chemical ligation · Peptide desulfurization · Peptide ligation · Peptide synthesis · Protein chemical synthesis · Protein pharmaceutical · Racemic protein crystallography

F. Liu (✉)
Calibrium LLC, 11711 N. Meridian Street, Carmel, IN 46032, USA
e-mail: fa.liu@calibrium.com

J.P. Mayer (✉)
Calibrium LLC, 11711 N. Meridian Street, Carmel, IN 46032, USA

Department of Chemistry, Indiana University, 800 E. Kirkwood Ave., Bloomington, IN 47405, USA
e-mail: johmayer@indiana.edu

Contents

Abbreviations

AAA	Amino acid analysis
Acm	Acetamidomethyl
BAL	Backbone amide linker
BHA	Benzhydrylamine
Boc	*tert*-Butyloxycarbonyl group
Bpoc	Biphenylisopropoxycarbonyl
Bzl	Benzyl
CBD	Chitin-binding domain
Cbz	Carboxybenzyl or benzyloxycarbonyl
CTC	2-Chlorotrityl chloride
DBU	1,8-Diazabicyclo[5. 4. 0]undec-7-ene
DCC	*N*,*N*′-Dicyclohexylcarbodiimide
DIEA	*N*,*N*-Diisopropylethylamine
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethyl sulfoxide
EPL	Expressed protein ligation
EPO	Erythropoietin
ESI	Electrospray ionization
FAB	Fast atom bombardment
Fmoc	9-Fluorenylmethoxycarbonyl group
GCSF	Granulocyte-colony stimulating factor
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HF	Hydrogen fluoride
HIV-PT	Human immunodeficiency virus-protease

HMPA	4-Hydroxymethylphenoxyacetic acid
HMPB	4-Hydroxymethyl-3-methoxyphenoxybutyric acid
HOBt	1-Hydroxylbenzotriazole
HOOBt	3,4-Dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine
HPLC	High-performance liquid chromatography
ivDde	1-(4,4-Dimethyl-2,6-dioxocyclohex-1-ylidene)-3-methylbutyl
KCL	Kinetically controlled ligation
MALDI	Matrix-assisted laser desorption/ionization
MBHA	4-Methylbenzhydrylamine
MESNa	Sodium 2-mercaptoethanesulfonate
MIPD	Mirror image phage-display
MPAA	4-Mercaptophenylacetic acid
MS	Mass spectrometry
NCL	Native chemical ligation
NHS	*N*-Hydroxysuccinimide
NMP	*N*-Methyl-2-pyrrolidone
NMR	Nuclear magnetic resonance
PAL	Peptide amide linker
PAM	4-Hydroxymethyl-phenylacetamidomethyl
PEG	Polyethylene glycol
rDNA	Recombinant DNA
RP	Reverse-phase
RSV	Raus-sarcoma virus
SAR	Structure-activity relationship
SPPS	Solid-phase peptide synthesis
*t*Bu	*tert*-Butyl
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TFET	Trifluoroethanethiol
Thz	1,3-Thiazolidine-4-carboxylic acid
TPO	Thrombopoetin
VEGF	Vascular endothelial growth factor

1 Introduction

Proteins constitute a universally essential class of macromolecules which perform a wide range of specialized functions in living systems. Examples of these functions include the enzymatic catalysis of metabolic pathways, hormonal signaling in the endocrine system, and antibody mediated defense in the immune system. Proteins also perform critical structural roles, for example as the muscle proteins actin and myosin. The study of protein structure and function is therefore essential to our understanding of life and the advancement of medicine [1].

Proteins have traditionally been obtained from their native sources through extraction and isolation, a method which is inherently limited by the availability of the source material and low concentration of the target protein. In addition, an extracted protein of a non-human sequence may be less than ideal for therapeutic application. Nevertheless, extraction of bovine and porcine pancreatic glands served as the sole source of insulin for over 60 years, while cadaver pituitary glands provided the initial source of human growth hormone. Semi-synthesis, typically accomplished by reverse proteolysis of a native precursor, offered an important means of primary sequence modification for early protein structure-function studies and has also been utilized commercially [2]. More recently, de novo methods of protein production have been developed which are not constrained by availability of the natural source and, importantly, are not restricted to the native sequence. The first of these is recombinant DNA (rDNA) technology which utilizes transfected exogenous genes to express the target protein in a bacterial, yeast, or mammalian cell. This process has become enormously significant from not only a discovery but also a commercial perspective as one of the key technologies of the biotechnology revolution. A number of highly important therapeutic proteins including insulin, growth hormone, interferon, granulocyte-colony stimulating factor (GCSF) were the earliest commercial products made possible by this technology.

An additional method, the production of proteins by chemical or total synthesis, as differentiated from semi-synthesis mentioned above, has emerged as a highly versatile tool for the study of protein structure and function. The ability to modify side-chains or portions of the backbone site-specifically, or to introduce isotopically labeled amino acids and then correlate these with biological activity, represents an enormous advance in the chemical biology of proteins. In this chapter we have focused on the impact of this technology on several areas of drug discovery: enzyme total synthesis, mirror image phage-display, racemic crystallography, and protein pharmaceutical synthesis.

2 A Brief History of Modern Chemical Protein Synthesis

2.1 Peptide Chemistry in the Early Years, 1880s–1950s

The first formal studies of peptide bond formation can be traced back to 1881, when Curtius obtained Bz-Gly-Gly through the treatment of glycine silver salt with benzoyl chloride [3]. A notable later milestone was Fisher's synthesis of an octadecapeptide LGGGLGGGLGGGGGGGGG in 1907, where the α-halogen was utilized as the precursor of α-amine as a means of protection during the peptide bond formation [4].

It had long been recognized that the efficiency of peptide synthesis could be dramatically increased with the development of reversible amine protections. This realization inspired Bergman to modify Fischer's ethoxycarbonyl function and

create the versatile carboxybenzyl group (Cbz) in 1932 which maintained α-carbon chirality during carboxyl activation and could be removed under mild cleavage conditions [5]. The Cbz group represented a milestone event in the history of peptide chemistry, and was subsequently widely adopted and utilized in the synthesis of numerous peptides, such as glutathione [6], carnosine [7], and oxytocin [8].

The acyl azide of Curtius [9, 10] and the acid chloride of Fisher [11] were the predominant methods of peptide bond formation until 1955 when Sheehan [12], Khorana [13], and their co-workers introduced the highly efficient *N*,*N*′-dicyclohexylcarbodiimide (DCC) coupling reagent. In contrast to the older reagents, DCC did not require carboxyl group pre-activation, and could be used in situ. The convenience and efficiency of DCC were quickly recognized and incorporated into many peptide synthesis protocols.

The introduction of the acid labile *tert*-butyloxycarbonyl group (Boc) in 1957 by Carpino [14], Anderson [15], and McKay and Albertson [16] marked the next breakthrough in peptide synthesis. The Boc group remains intact during hydrogenation and under conditions of the Birch reduction [17], and therefore represents complete orthogonality to Cbz and other benzyl-based groups. Its use in combination with Cbz greatly enhanced the flexibility of earlier protecting group schemes, illustrated by the synthesis of 39-mer β-corticotropin adrenocorticotrophic hormone (ACTH) by Schwyzer and Sieber in 1963 [18].

2.2 Solid-Phase Peptide Synthesis (SPPS), 1960s–1980s

Growing interest in peptide hormones together with the dramatic progress in protecting groups and coupling reagents in the 1950s set the stage for perhaps the most significant advance in peptide chemistry to date, the realization of Merrifield's ingenious concept of solid-phase peptide synthesis (SPPS) in 1963 [19]. The procedure was initiated with attachment of the α-carboxyl of the C-terminal residue to a polymer support via an ester bond, followed by elongation through a repetitive process of coupling and deprotection of carbamate-protected α-amino acids. The Boc/Bzl protecting group scheme, DCC activation, and HF resin cleavage/global deprotection were found to be highly compatible with the SPPS protocol and the technique was quickly adopted [20]. Merrifield's SPPS inarguably revolutionized the efficiency of peptide chemistry, shortening many syntheses from years to days. Automation, a unique and highly advantageous feature of SPPS, was first demonstrated in 1968 in Merrifield's lab with commercial peptide synthesizers becoming available by the 1980s. The impact of SPPS was recognized by the award of the 1984 Nobel Prize in Chemistry to R. B. Merrifield. It should be recalled that a nearly simultaneous report by Letsinger and Kornet proposed an alternative solid-phase approach which relied on anchoring of the α-amino group of the N-terminal residue and elongation in the N-to-C direction [21]. Activation of the C-terminal

carboxyl group of the resin-bound peptide, however, was accompanied by significant levels of epimerization, which discouraged the development of this method.

The 1970s brought numerous advances, many of which served to improve further the performance and convenience of SPPS. The use of DCC, despite the excellent reactivity of the *O*-acyl iso-urea intermediate, was associated with various side reactions, including the rearrangement to an unreactive *N*-acyl urea and the loss of α-chiral integrity. The introduction of 1-hydroxylbenzotriazole (HOBt) as a DCC coupling additive by König and Geiger in 1970 dramatically improved reaction efficiency [22]. Possessing a similar acidity as acetic acid, HOBt protonates the basic center in the DCC-generated *O*-acyl iso-urea intermediate and therefore prevents the racemization of the activated residue. It also prohibits the unreactive *N*-acyl urea formation by converting the over-activated *O*-acyl iso-urea to the less but still sufficiently reactive HOBt ester.

The need for successful base-labile α-amino group protection was met by the introduction of the 9-fluorenylmethoxycarbonyl group (Fmoc) by Carpino and Han in 1970 [23, 24]. The Fmoc group is completely orthogonal to the *tert*-butyl-based protecting groups and can easily be removed through β-elimination initiated by bases such as piperidine [25]. These features have led to the adoption of Fmoc/*t*Bu as an important alternative to the classic Boc/Bzl. More recently, the concerns about the use of HF and its incompatibility with phospho- and glycopeptides have made Fmoc/*t*Bu-SPPS generally more preferable to the Boc/Bzl-SPPS.

The 1970s and 1980s witnessed development of a number of solid-phase linkers, an area which had progressed dramatically since the days of chloromethylpolystyrene (Merrifield resin) (Fig. 1) [26]. These include the 4-hydroxymethylphenylacetamidomethyl resin (PAM resin) introduced by Merrifield and Mitchell to improve stability under prolonged TFA exposure [27], and the benzhydrylamine (BHA) linker of Pieta and Marshall [28], devised to simplify the synthesis of C-terminal carboxamides, a feature of many biologically active peptides. Stewart and Matsueda further modified the BHA linker for optimal balance between TFA stability and HF cleavage efficiency, resulting in the 4-methylbenzhydrylamine resin (MBHA resin) [29]. Since then, PAM and MBHA have been the standard linkers for Boc-SPPS for peptides with C-terminal acids and amides, respectively.

The introduction of the Fmoc group in 1970 prompted the design of more acid-sensitive linkers which can be cleaved in TFA. This was generally accomplished through the addition of electron-donating substituents to the benzyl or benzhydryl linkers employed in Boc chemistry. Examples include the Wang [30] and 4-hydroxymethylphenoxyacetic acid resins (HMPA resin) [31, 32], which can both be effectively cleaved in 95% TFA, as well as the more acid labile 4-hydroxymethyl-3-methoxyphenoxybutyric acid resin (HMPB resin) (Fig. 1) [33]. Fmoc/*t*Bu compatible linkers for generation of C-terminal amides began to emerge in 1987 and included the peptide amide linker (PAL) resin [34], the Rink amide resin [35], and the Sieber resin [36]. While the former two linkers require 95% TFA for cleavage, the Sieber requires only 1%, allowing the side-chain protecting groups to remain intact. A similar option is available in the case of the trityl-based resin, such as 2-chlorotrityl chloride resin (CTC resin), which also permits release of side-chain-protected peptide acid in the presence of 1% TFA [37, 38].

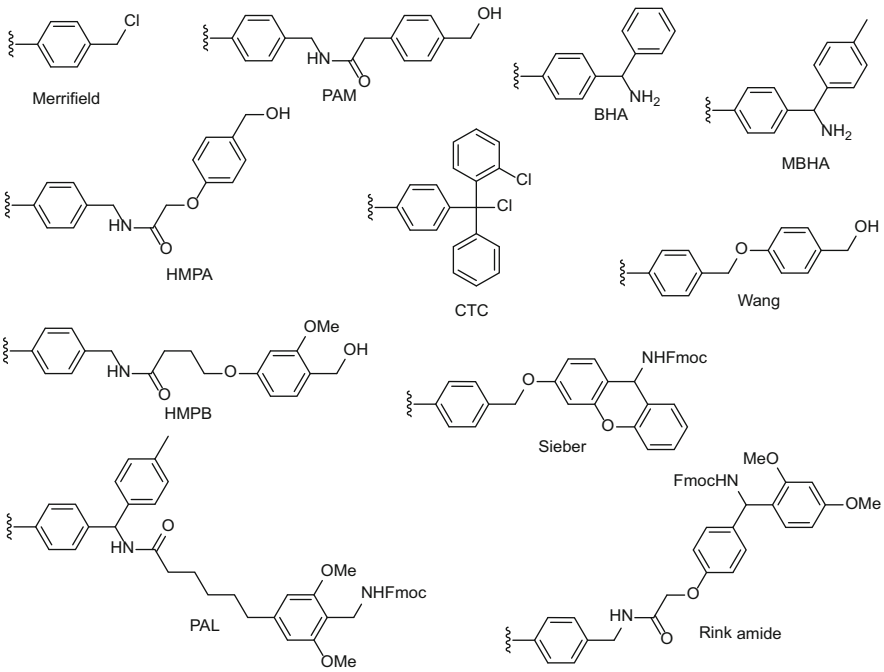


Fig. 1 Selected resin linkers

Since the invention of SPPS, divinylbenzene-cross-linked polystyrene has been the predominant solid support used. There have been several alternatives developed along the way, with the most important advancement being the introduction of PEG, including PEG-grafted polystyrene resin [39, 40] and all-PEG-based resin (ChemMatrix) [41, 42]. PEG-based resins can produce crude products with better quality than the polystyrene resins, particularly in the case of difficult sequences [43]. This superiority is explained by the H-bond acceptor function of PEG which interrupts peptide inter-chain H-bonding, effectively inhibiting peptide chain aggregation [44].

2.3 Protein Synthesis by Peptide Ligation, 1990s-Present

The capability of SPPS has made the high quality preparation of peptides of up to ~40 residues a routine exercise. However, for longer sequences, alternative methods have generally been required to achieve satisfactory results. A traditional approach dating back to the 1960s involved the convergent condensation of fully protected fragments [45, 46]. The practical application of this method has frequently been limited by the poor solubility of protected peptide fragments and by the epimerization of the C-terminal α-carbons during activation. Nevertheless, this

method has on occasion been successfully utilized on an industrial scale as in the manufacture of the anti-HIV agent enfuvirtide (Fuzeon), a 36-mer peptide, on a near-ton scale [47]. A notable advance first reported by Blake in 1981 and subsequently refined by Aimoto in 1999 involved silver ion-mediated peptide ligation which did not require side-chain carboxyl group protection [48, 49].

While it had been recognized that unprotected peptides benefit from better solubility and handling, it was also clear that their ligation would place higher demands on chemoselectivity. A number of such methods have been reported, including the formation of thioester via C-terminal thiocarboxyl and N-terminal bromoacetyl [50], oxime ligation via glyoxylyl functionality and an N-terminal aminooxy acetyl group [51], and the formation of thiazolidine via a C-terminal aldehyde and an N-terminal cysteine mediated by imine capture [52]. Despite the conceptual importance and utility of these methods, none are capable of producing a native peptide bond. A number of enzymatic ligation methods were developed to address this issue, such as the engineered sublitigase reaction [53], but, unfortunately, the restrictive N-terminal sequence requirements and the poor solubility of the peptides in the ligation buffer have limited this method's practical application.

A major breakthrough was achieved by Dawson et al. in 1994 with their discovery of native chemical ligation (NCL) [54]. This approach joins two peptides through a thiol/thioester exchange and a subsequent S-to-N acyl-shift. The use of NCL (discussed in detail below) offers exquisite chemoselectivity, generally excellent solubility of the peptides in the ligation buffer, and, above all, the ability to form a native peptide bond. It is appropriate to credit the work of Wieland et al. who first demonstrated the thiol-thioester exchange/S-to-N acyl shift reaction sequence during the synthesis of di and tri-peptides in 1953 [55], and the thiol capture strategy of Kemp et al. demonstrated in 1986 [56, 57]. The discovery of NCL represents without question a revolutionary advance in peptide and protein chemistry, as evident by its immediate extensive applications in the synthesis of various proteins [58–60].

Muir et al. in 1998 extended the capability of NCL by expressed protein ligation (EPL) where one or both ligation partners can be produced recombinantly [61], facilitating the use of larger native protein sequences in the ligation reaction.

It is important to cite several other recently developed methods for peptide ligation, which include the traceless Staudinger ligations of Bertozzi [62], Raines [63], and their co-workers, the decarboxylative amide ligation (KAHA ligation) developed by Bode et al. [64], and the Ser/Thr ligation of Li et al. [65, 66]. While these methods can be considered complementary to NCL, they have various constraints which limit their general utility.

2.4 Analytical Methods

While beyond the scope of this review, advances in analytical techniques were an essential counterpart to the evolution of synthetic methodology. Amino acid

analysis, N-terminal sequencing, high-performance liquid chromatography, and mass spectrometry have served an indispensable role in the structural characterization of peptides and proteins.

Amino acid analysis (AAA) developed by Stein and Moore [67] is the oldest method and is still in use today. Although it has historically been used to determine peptide composition and stoichiometry, it is primarily used today to assess peptide content [68]. For the synthetic peptide chemist this method offers a means of monitoring the progress of a synthesis through analysis of a crude peptide or resin hydrolysate.

Edman degradation or N-terminal sequence analysis [69–71] has not only been instrumental in the characterization of countless native peptides and proteins, but also offers significant utility in the context of synthetic peptide chemistry. The sequencing of a peptide can rapidly identify the presence and position of an amino acid deletion or side-chain modification. In a technique known as "preview analysis" [72] the sequencing of a peptide mixture containing a peptide of length "n" reveals the presence of a deletion peptide of length "n−1" through detection of a particular residue one cycle earlier than expected. The extent of such preview is also an excellent independent determinant of peptide homogeneity. Similarly to AAA, N-terminal sequence analysis can be carried out on crude or resin-bound peptides.

High-performance liquid chromatography (HPLC) has been enormously important both as a preparative and analytical tool [73]. It is particularly well suited for the resolution of heterogeneous products of SPPS containing, in addition to the desired material, deletion and termination products. The basic concept can be utilized through several modes of application including size-exclusion, ion exchange, and reversed-phase (RP-HPLC), the last being the most commonly used for peptide application.

Mass spectrometry (MS) offers a direct and efficient confirmation of the molecular weight of a synthetic peptide and has largely become the method of choice for this application. Fast atom bombardment (FAB), matrix-assisted laser desorption/ionization (MALDI), and electrospray ionization (ESI) modes have all been successfully used for peptide characterization. Of the three, ESI [74] stands out as particularly advantageous by virtue of its ready compatibility with LC/MS systems. This integration enables assignment of molecular weights to individual chromatographic peaks in the LC trace of a crude peptide and can also assess the homogeneity of individual fractions during preparative chromatography.

The greater size and complexity of synthetic protein targets also places increased demands on analytical methods required for their rigorous characterization. These criteria should include a reliable assessment of homogeneity by HPLC or isoelectric focusing (IEF), mass measurement by ESI, and confirmation of correct tertiary structure through multi-dimensional nuclear magnetic resonance (NMR) spectroscopy and/or high resolution X-ray crystallography. Interpretation of biological data is only meaningful if the above analytical criteria are satisfied.

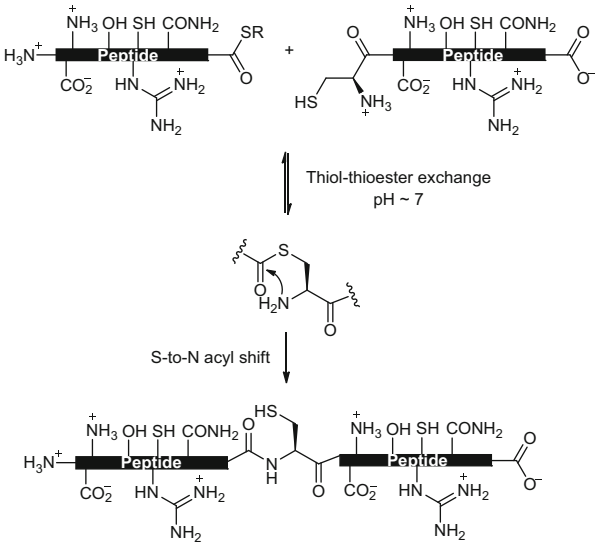
3 Key Chemistries for Modern Chemical Protein Synthesis

3.1 Native Chemical Ligation

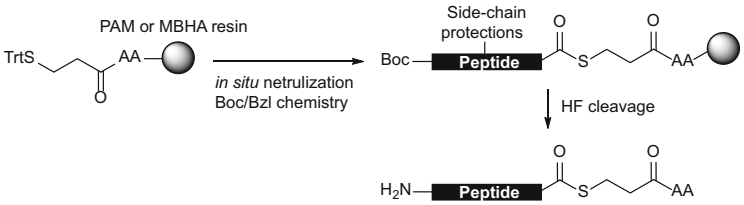
Native chemical ligation selectively conjugates two unprotected peptide segments, resulting in the formation of a native peptide bond (Scheme 1) [54]. The reaction between the C-terminal thioester of one segment and the N-terminal cysteine of another is initiated by a thiol-thioester exchange and generates a thioester-linked intermediate. The subsequent intra-molecular S-to-N acyl shift proceeds through a reversible five-membered transition state producing a stable native peptide bond, driving the equilibrium in the desired direction. While aryl thioesters are highly reactive and ligate readily, the alkyl thioesters, in contrast, are less reactive and require a thiol catalyst, such as 4-mercaptophenylacetic acid (MPAA) to facilitate their ligation [75]. The reaction is typically conducted in a phosphate buffer (pH ~7) and chaotropic agents such as guanidine or urea are included in the buffer to assist in the dissolution of the peptides. In addition, a water soluble disulfide bond reducing agent such as tris(2-carboxyethyl)phosphine (TCEP) is generally present to prevent cysteine oxidation.

3.2 Preparation of Peptide C-Terminal Thioesters

The efficient preparation of the C-terminal thioester peptide component is an essential requirement for conducting a successful NCL reaction. The chemical synthesis of peptide thioesters either by Boc-SPPS or Fmoc-SPPS has been the



Scheme 1 Native chemical ligation



Scheme 2 The preparation of peptide thioesters by Boc-SPPS (*AA* amino acid)

subject of intensive investigation; in addition, these intermediates can also be obtained by means of recombinant expression.

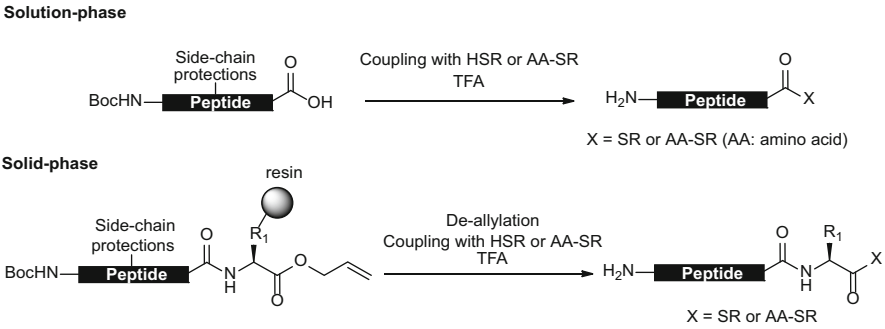
3.2.1 Preparation of Peptide C-Terminal Thioester by Boc-SPPS

C-terminal thioester peptides can be prepared in a straightforward manner by using Boc-SPPS [76, 77]. The thioester group is typically an integral component of the resin linker and stepwise chain elongation is conducted under in situ neutralization conditions developed by Kent et al. to ensure the efficiency of couplings (Scheme 2) [78].

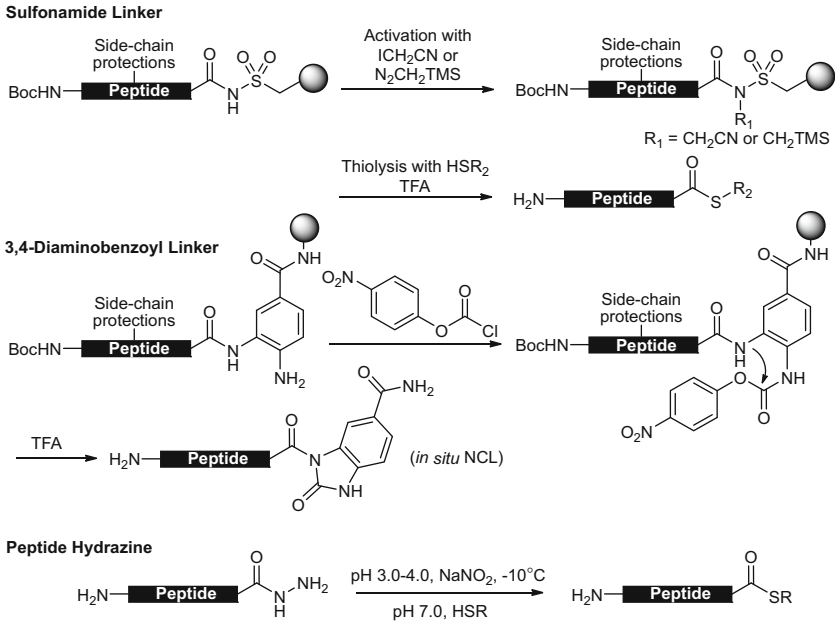
3.2.2 Preparation of Peptide C-Terminal Thioesters by Fmoc-SPPS

The inconvenience and risk of HF handling prompted the development of Fmoc-SPPS compatible protocols for thioester preparation. The principal obstacle is the instability of the thioester when exposed to 20% piperidine, the standard conditions for Fmoc-removal, and numerous alternative methods have been developed to circumvent this problem [79]. The most straightforward involve using less nucleophilic conditions, such as a mixture of 25 vol.% 1-methylpyrrolidine, 2 vol.% hexamethyleneimine, and 2% HOBt (w/v) in NMP and DMSO (1:1, v/v) [80, 81]; or a mixture of 1% HOBt (w/v) and 1 vol.% DBU in DMF [82]. These methods are effective when preparing short sequences.

A second method to avoid piperidine aminolysis of the thioester group involves performing the thioesterification of the C-terminal carboxyl as the last coupling step (Scheme 3). This conversion can be performed in solution or on resin, although both approaches share the risk of C-terminal epimerization. In addition, the poor solubility of the fully protected peptide can often complicate the solution-phase procedure [83, 84]. In contrast, the difficulty associated with the solid-phase procedure is the need to tether the C-terminus reliably. The earlier approach of anchoring from the backbone (backbone amide linker or BAL) is theoretically feasible but in practice limited by the low coupling yields at the *N*-alkylated C-terminal residues [85]. Subsequent approaches involving side-chain anchoring have proven to be more successful, although the requirement of anchoring at one of the two C-terminal residues has hindered their utility [85–89].



Scheme 3 The preparation of peptide thioester by C-terminal coupling

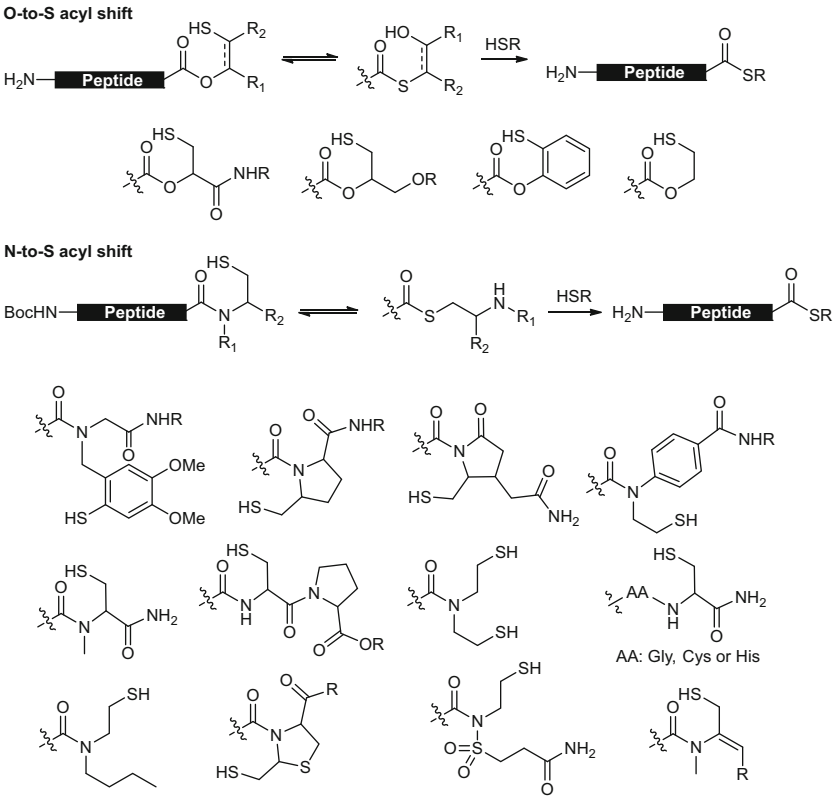


Scheme 4 Examples of peptide thioester preparation by C-terminal activation and thiolysis

The third method of preparing thioesters by Fmoc-SPPS involves C-terminal activation of the assembled peptide followed by thiolysis (Scheme 4). One widely used method utilizes Ellman's modification of Kenner's sulfonamide linker [90, 91]. The resin-bound peptide is treated with iodoacetonitrile or trimethylsilyldiazomethane to activate the acylsulfonamide bond, then exposed to free thiol to release the resulting thioester peptide from resin. A final global deprotection by TFA treatment provides the desired thioester. The reported drawbacks include the risk of epimerization during the slow loading step and side-chain modifications during the activation step [92, 93]. An alternative procedure recently introduced by

Dawson et al. involves peptide assembly on a 3,4-diaminobenzoyl linker, acylation of the free amine with 4-nitrophenyl chloroformate, followed by spontaneous ring closure. The resulting *N*-acylbenzimidazolone can either be readily converted to a thioester or directly used in an in situ ligation [94]. More recently, peptide C-terminal hydrazides have been reported by Liu et al. as convenient and efficient thioester precursors [95]. The reaction of a C-terminal hydrazide with $NaNO_2$ at pH 3.0–4.0 at −10°C forms an acyl azide, which is subsequently converted to the thioester prior to ligation. The C-terminal hydrazide can be easily obtained through Fmoc-SPPS except in the case of a C-terminal Asp, Asn, or Gln. Another advantage of the peptide hydrazide is that it can be produced by recombinant expression through hydrazinolysis of the protein thioester intermediate generated by the intein-mediated method [61]. Another approach from this category is based on the backbone pyroglutamylimide linker reported by Jensen et al., although both the formation of pyroglutamylimide and its conversion to the thioester require prolonged treatment at increased temperature [96].

A fourth method involves an intramolecular O-to-S or N-to-S acyl shift process following peptide chain assembly (Scheme 5). Typically, thiol-functionalized ester

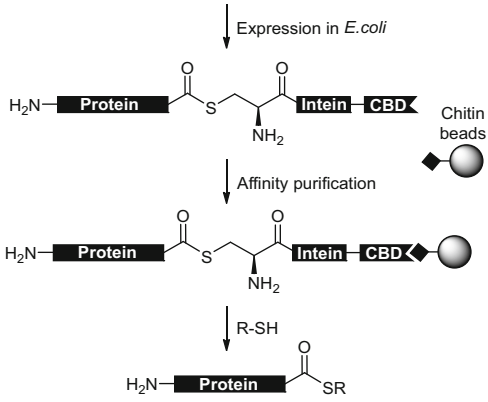


Scheme 5 Examples of peptide thioester preparation by O-to-S or N-to-S acyl shift

or amide-based linkages are incorporated during the solid-phase synthesis. TFA cleavage then liberates the thiol group which undergoes rearrangement via a 5-membered transition state, forming the corresponding thioester. The equilibrium is driven toward the thioester by excess thiols or irreversible native bond formation if used for in situ ligation. The advantage of the O-to-S acyl shift is that it generally proceeds rapidly under mild conditions [97]. Limitations involve the concurrent hydrolysis of the alkyl ester (10–20%), particularly in the case of beta-branched residues such as Val, Ile, and Thr (50–70%) [98–100] and the potential for C-terminal racemization during formation of the aryl ester [101]. Since the major challenges for the N-to-S acyl shift procedure involve slow conversion kinetics, most efforts have focused on accelerating this process under mild reaction conditions [102–121].

3.2.3 Preparation of Protein C-Terminal Thioesters by Expressed Protein Ligation

The preparation of protein C-terminal thioesters by expressed protein ligation (EPL) involves a manipulated protein splicing process [61]. Protein splicing is a post-translational process in which a thioester-mediated intramolecular rearrangement results in the extrusion of an internal sequence (intein). In EPL the intein is mutated so that, while it is still capable of facilitating the formation of the thioester intermediate, the splicing is not completed (Scheme 6). A fusion of the mutated intein and a chitin-binding domain (CBD) enables the affinity purification of the chimera by chitin beads, with a subsequent thiol treatment promoting the release of the desired protein thioester.



Scheme 6 Preparation of a protein thioester by EPL

3.3 Ligation at Non-Cys Junctions

The 1–2% composition percentage of cysteine in proteins [122] limits the practical application of conventional NCL. To circumvent this constraint and enhance the utility of NCL, a number of alternative approaches have been devised involving thiol handle-mediated as well as thiol handle-free approaches.

3.3.1 Thiol Handle-Mediated Ligation at Non-Cys Junction

A thiol handle can, in principle, be introduced through an auxiliary group [123] or by direct attachment to a side-chain. While auxiliaries can be anchored either to the backbone or to a side-chain, each approach involves inherent limitations (Fig. 2). When an auxiliary is attached to the backbone, the subsequent coupling to this residue is typically slow, which limits its utility to less hindered junctions. Nevertheless, successful examples have been reported by Kent [124, 125], Botti [126], Dawson [127], Danishefsky [128], and others [123, 129]. Use of a side-chain auxiliary circumvents the above difficulty but requires an appropriate side-chain anchoring site. This method has been successfully demonstrated by Wong et al. with carbohydrate-based auxiliaries at the side-chains of Ser or Thr [130, 131] and Brik et al. with cyclohexane or cyclopentane-based auxiliaries on the side-chains of Asp, Glu, Ser, or Thr [132, 133].

The direct attachment of a thiol to a side-chain requires a post-ligation desulfurization step for its removal. This approach was first reported by Yan and Dawson, who extended hydrogenolytic desulfurization of cysteine to the chemical synthesis of proteins by NCL, utilizing Ala as a new ligation junction (Scheme 7) [134]. In the same study, they also noted that not only Raney nickel but also palladium could be used as the catalyst. In addition, the application was expanded to other residues such as Leu, Val, and Ile. A subsequent refinement of this method

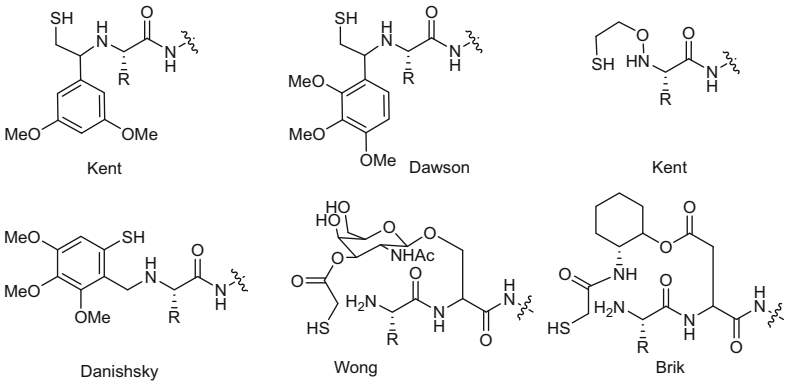
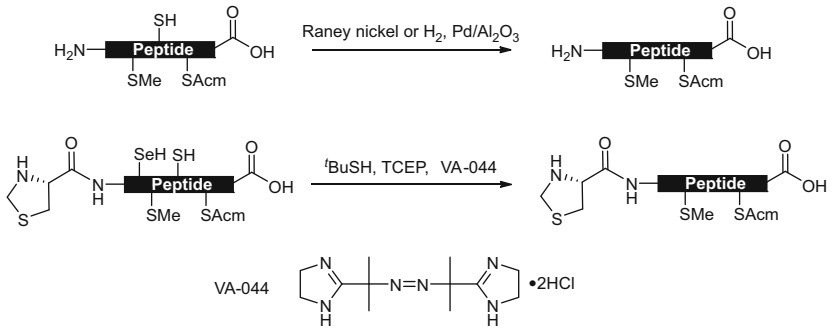


Fig. 2 Examples of auxiliary-assisted chemical ligation



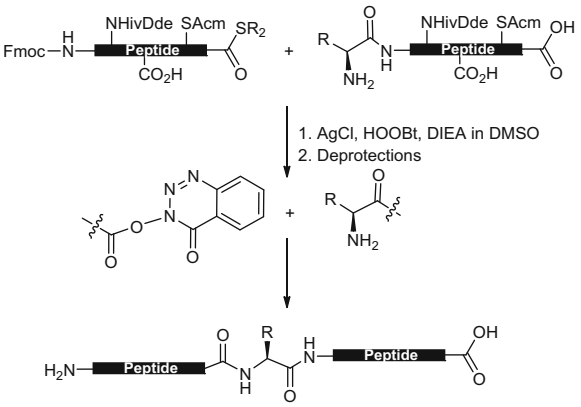
Scheme 7 Post-ligation desulfurization

by Kent and Wong and their co-workers [131, 135] achieved selective desulfurization of Cys in the presence of Cys(Acm) and Met. However, Danishefsky et al. observed that the metal-based desulfurization was not compatible with (4R)-1,3-thiazolidine-4-carboxylic acid (Thz) [136, 137] and subsequently introduced a new metal free, free radical-based method [138]. The key reagents of Danishefsky's homogeneous desulfurization include a water-soluble phosphine TCEP and a water-soluble free-radical initiator 2,2′-azobis[2-(2-imidazolin-2-yl) propane]dihydrochloride (VA-044), which are not only compatible with Thz, Cys (Acm), and Met, but also equally capable of reducing selenocysteine to Ala. Subsequent efforts in this area have focused on "one-pot" ligation-desulfurization protocols, which are complicated by the radical quenching properties of thiophenols, such as MPAA. Nevertheless, several successful one-pot protocols have been reported, including the use of sodium 2-mercaptoethanesulfonate (MESNa) in place of thiophenol [139], and the post-ligation removal of thiophenol by liquid- [140] or solid-phase extraction [141]. Most recently, Payne et al. described the use of a non-free radical quenching alkyl thiol trifluoroethanethiol (TFET) as an effective MPAA surrogate permitting one-pot post ligation desulfurization [142].

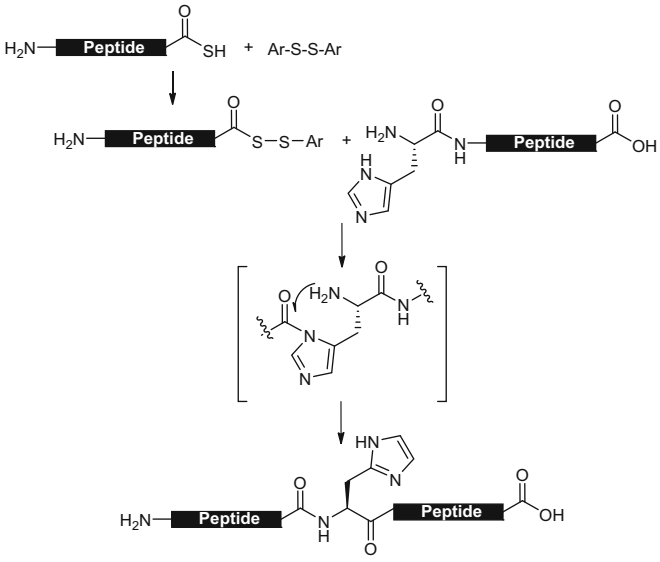
3.3.2 Thiol Handle-Free Ligation at Non-Cys Junction

In addition to the fragment condensation of side-chain fully protected segments, a number of improved thiol handle-free ligation methods have been developed, which complement NCL and enhance our capability of synthesizing proteins by chemical methods.

A strategy based on silver-ion mediated selective activation of a C-terminal thiolcarboxyl group initially reported by Blake [48] was modified by Aimoto et al., employing the more oxidation- and hydrolysis-resistant thioester in place of thioacid and HOBt or 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine (HOOBt) as an additive to facilitate the formation of the activated ester (Scheme 8) [49, 143]. This activated ester-based coupling requires the protection of side-chains of Lys



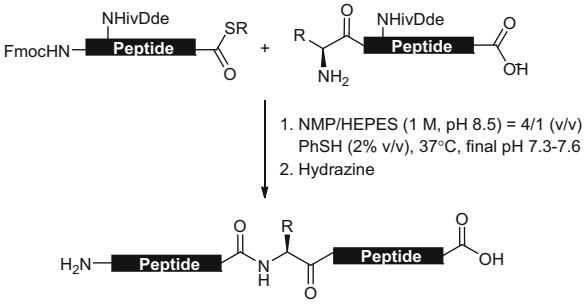
Scheme 8 Silver ion-mediated ligation



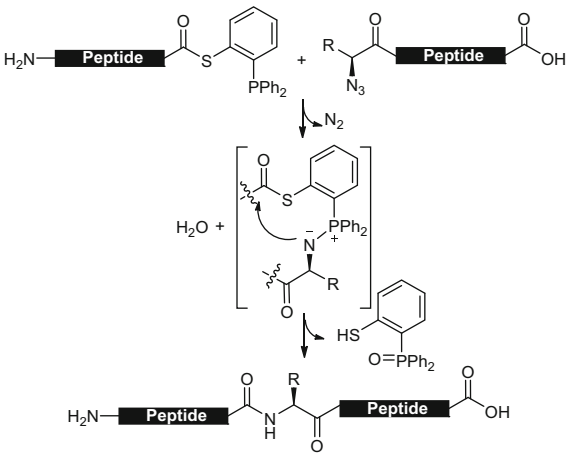
Scheme 9 Histidine ligation

(by ivDde) and Cys (by Acm) and the N-terminal amine (by Fmoc). Recently, Danishefsky et al. reported another modification of this method, where the thioester was masked as a protected *ortho*-thiol-containing phenolic ester [144]. Epimerization at the ligation site, and in some cases poor fragment solubility, are the main concerns when employing these approaches.

As a second approach, Tam et al. demonstrated a selective histidine ligation, in which the C-terminal thiolcarboxyl was activated by Ellman's reagent in situ, captured by the imidazole ring of the N-terminal histidine and subsequently transferred to its α-amine to furnish a native peptide bond (Scheme 9) [145]. The high



Scheme 10 Direct ligation of a peptide C-terminal thioester with another N-terminus



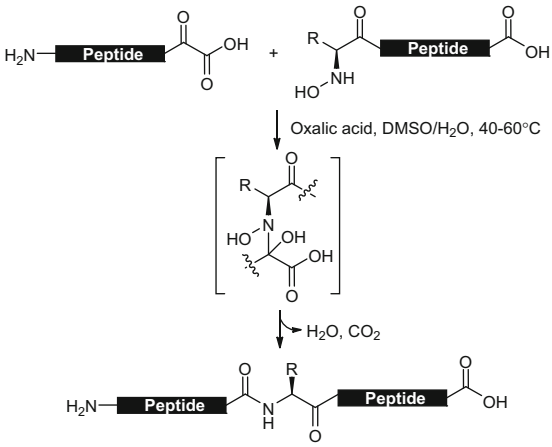
Scheme 11 Staudinger ligation

ligation selectivity between the α-amine of N-terminal histidine and the ε-amine of internal lysines was achieved by conducting the reaction in a low pH (5.7) buffer. The in situ activation of the thiolcarboxyl by Ellman's reagent was necessary since the unactivated thiolcarboxyl does not ligate.

In a third approach, the C-terminal thioester and the N-terminal amine are coupled directly, with the competing reaction, the hydrolysis of the thioester, effectively suppressed through the use of a mixture of organic solvent and aqueous buffer (Scheme 10). While the ligation is compatible with the free side-chain of Cys, protection is required for the amine groups [146]. Follow-up studies reported that the phosphate group of an N-terminal phosphorylated Ser or Thr could significantly improve the ligation yield, presumably because of acyl capture by phosphate and the subsequent intramolecular acyl transfer [147]. A similar yield enhancement effect was also observed for the N-terminal Asp or Glu [148].

In another example, Raines and Bertozzi and their co-workers modified the Staudinger reaction to an amide-bond-forming ligation, which utilizes a C-terminal phosphinobenzenethioester and an N-terminal α-azide (Scheme 11).

Scheme 12 KAHA ligation



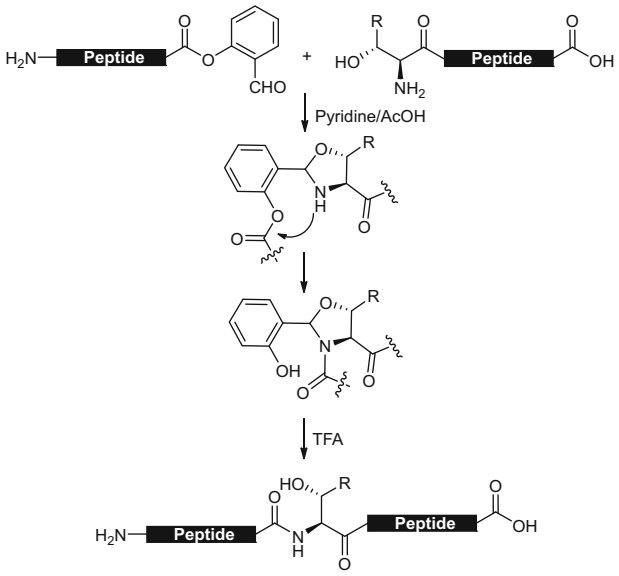
The intermediate iminophosphorane undergoes a six-membered transition state-mediated S-to-N acyl shift to furnish the native amide bond [62, 63, 149].

Bode et al. devised a unique chemoselective amide ligation by the decarboxylative condensations of α-ketoacids and *N*-alkylhydroxylamines (KAHA ligation) (Scheme 12) [64, 150–154]. This process requires neither coupling reagents nor catalysts, produces only water and carbon dioxide as by-products, and tolerates unprotected amino acid functional groups. It is also completely orthogonal to NCL and theoretically can be utilized at any junction. However, the limited access to *N*-alkylhydroxylamines and α-ketoacids restricts broader application of this chemistry.

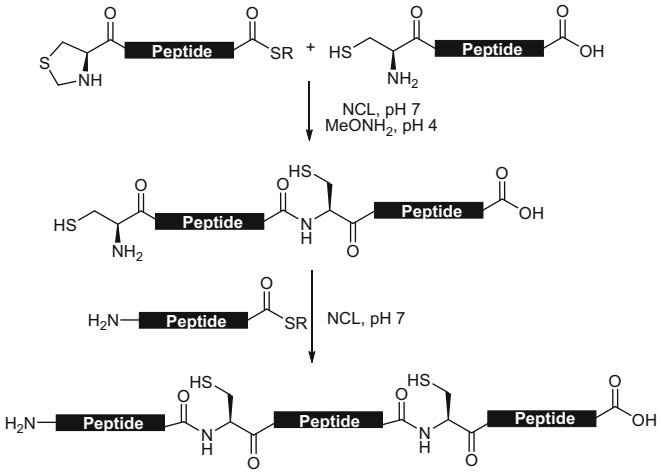
Li et al. reported a chemoselective ligation between a C-terminal ester of *O*-salicylaldehyde and an N-terminal Ser or Thr which after forming an *N*,*O*-benzylidene acetal intermediate, undergoes an intra-molecular six-membered transition state-mediated O-to-N acyl shift to generate an amide bond (Scheme 13) [65, 66]. The subsequent post-ligation TFA treatment removes salicylaldehyde and furnishes a native peptide bond.

3.4 Multiple Segment Ligation Strategies

The chemical synthesis of an average-sized protein of about 300 residues, requires multiple ligations. The minimization of the intermediate purification steps and use of a convergent route are necessary to ensure a high overall yield and a rapid throughput. Kent et al. have devised a number of strategies in pursuit of this goal, including the one-pot sequential ligation [137], His-tag-assisted ligation [155], solid-phase-assisted ligation [156, 157], and kinetically controlled ligation (KCL) [158–160].

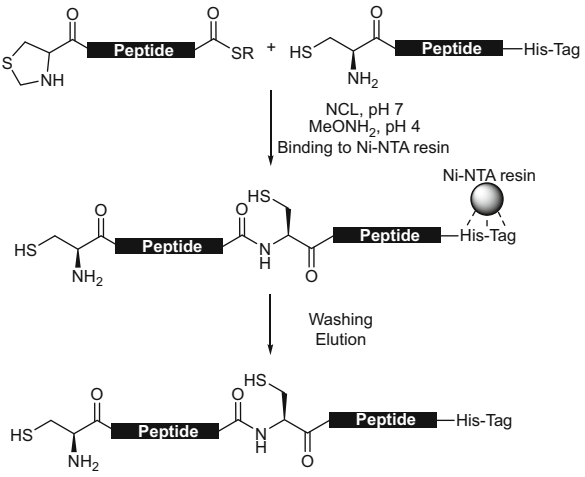


Scheme 13 Ligation at Ser or Thr junction

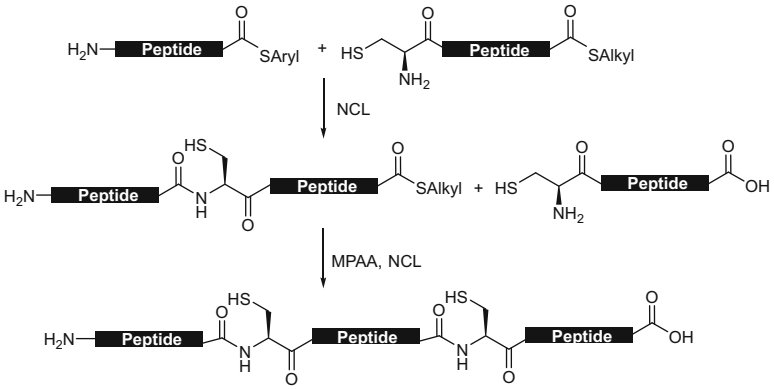


Scheme 14 One-pot sequential ligation

In the one-pot sequential ligation, Thz [136, 137] is utilized as the protected N-terminal cysteine with its liberation facilitated by the addition of methoxyamine at pH 4. The successful application of this strategy relies largely on a nearly quantitative yield of each intermediate ligation step (Scheme 14) [137].



Scheme 15 His-tag-assisted ligation



Scheme 16 Kinetically controlled ligation

Both solid-phase and His-tag-assisted ligation strategies take advantage of the principal benefits of solid-phase synthesis where the excess reagents are removed by simple filtration. An important distinction is that, while the His-tag ligation is conducted in solution-phase, the latter is conducted on solid support. Both approaches greatly reduce the number of intermediate purification steps (Scheme 15) [155–157].

KCL is made possible by the dramatic difference in reactivity between an aryl thioester and alkyl thioester. The significantly faster reaction kinetics of aryl thioesters enable selective ligation in the presence of an alkyl thioester, whose subsequent ligation is initiated by the addition of a catalyst such as MPAA. This approach provides the flexibility of conducting a ligation from the N-terminal to the C-terminal direction (Scheme 16) [158–160].

3.5 The Folding of a Synthetic Protein

A unique three-dimensional structure is an indispensable feature of a functional protein. To reproduce the essential physico-chemical and biological properties of a native protein, a chemically synthesized polypeptide chain must be folded into its correct tertiary structure. In most cases, the course of the folding reaction is directed thermodynamically by its primary sequence. In the native environment, the folding process occurs in the protein-rich cytosol and is aided by sophisticated "chaperone" machineries [161]. In contrast, the folding of a synthetic protein involves a single species and takes place under much simpler conditions. In general, a synthetic single domain linear polypeptide can be efficiently folded in an aqueous buffer into its fully active native form; in addition, a number of multi-domain proteins have been successfully produced by the same methods. For cysteine-containing proteins, optimized pH and oxidative or redox conditions are utilized to facilitate the disulfide bond formation. A modest concentration of a chaotropic agent such as 0.5–1.5 M urea or guanidine may be included to maintain solubility and prevent aggregation [58, 60].

4 Applications of Protein Chemical Synthesis to Drug Discovery

4.1 Chemical Synthesis of D-Proteins and Their Application to Drug Discovery

4.1.1 Mirror Image Phage-Display

A novel approach to drug discovery, enabled entirely by the recent advances in protein chemical synthesis, is "Mirror Image Phage-Display" (MIPD), a powerful technique designed to facilitate identification of non-native, all-D-amino acid peptide ligands. This approach has been successful in generating high affinity ligands to SH3 domains, viral entry inhibitors, and oncoprotein antagonists [162–165]. Remarkably, it has been particularly effective in identifying ligands which disrupt protein–protein interactions which have been considered an "undruggable" class of targets [166, 167]. To provide a meaningful perspective on the potential of this powerful new technology, it is necessary to view it in the context of existing recombinant and chemical methods of de novo peptide ligand discovery. Since their introduction almost three decades ago, peptide libraries have been utilized for a range of applications including receptor target screening, enzyme inhibitor identification, and vaccine development. The earliest reports of this method involved construction of chemically [168] or recombinantly [169, 170] derived libraries for the identification of immunogenic epitopes within a native protein sequence. Phage display technology, originally introduced by Smith [171] for this purpose, relied on

mutagenesis of filamentous pIII phage each displaying one peptide on its surface, to construct hexapeptide libraries of approximately 4×10^7 sequences. The resulting phage libraries were then surveyed for binding against a biotin labeled antibody, eluted, and amplified. Several rounds of affinity purification would then iteratively select high-affinity binders. The DNA sequences of the coding region of the high-affinity phage were then sequenced and the binding epitope sequences deduced. The deduced sequences not only correlated with the known sequence specificities but also revealed completely unrelated surrogate sequences of similar affinities. A simultaneous report by Devlin [172] employed a similar method to construct a library of pIII fusion sequences to identify nine streptavidin-selective peptides whose binding was inhibited in the presence of free biotin. The broad utility of this new technique was quickly recognized and validated through application to therapeutically relevant targets. The identification of peptide-based concanavalin A ligands [173], inhibitors of neutrophil elastase [174], and antagonists of the IL-1 receptor [175] demonstrated the unique and wide-ranging capabilities of this method. Two impressive examples emerged from the Affymax group which reported the discovery of small peptide erythropoietin (EPO) hormone agonists [176], with a subsequent report from the same group disclosing sequences which functioned as full agonists of the thrombopoetin (TPO) receptor [177].

While the large size of the phage-display libraries (10^{13}) as well amplification of binders enables efficient screening of enormous sequence diversity, the obligatory restriction to all L-amino acids imposes a limitation. In 1993 Kim et al. addressed this shortcoming and expanded the scope of phage-display technology beyond coded amino acids [162]. The work was conceptually inspired by an elegant demonstration of "reciprocal chiral specificity" by Kent et al. [178]. The L and D forms of the HIV protease enzyme were synthesized and shown to catalyze the cleavage of substrates of only their corresponding configuration: the L-enzyme cleaved only the L-substrate while the D-enzyme cleaved only the D-substrate. This specificity was also conserved with respect to corresponding enantiomeric inhibitors of the L- and D-enzymes. The use of mirror image phage display technology offers an important option of obtaining all-D amino acid sequences. The method entails chemical synthesis of an all-D amino acid target molecule (a mirror image of the native L-amino acid form) which is then used to screen a phage display library. The adhering phage hits are amplified, their sequences deduced and chemically synthesized using D-amino acids. By virtue of their mirror image relationship, these peptides now bind to the native target (Fig. 3). The original validation of this concept was provided by Kim and co-workers who synthesized a 60 residue, all-D amino acid protein corresponding to the homology 3 domain (SH3 domain) of c-Src and utilized this as a target to screen a phage library expressing 10-residue random peptides [162]. When the best native sequences binding the D-SH3 domain were resynthesized in their mirror image all-D form they were shown to bind to the native L-SH3 domain target. Heteronuclear magnetic resonance (NMR) studies of the L-SH3 and D-SH3 domains with their respective ligands revealed that the two binding interactions utilize distinct receptor contacts. In addition, while the all-D

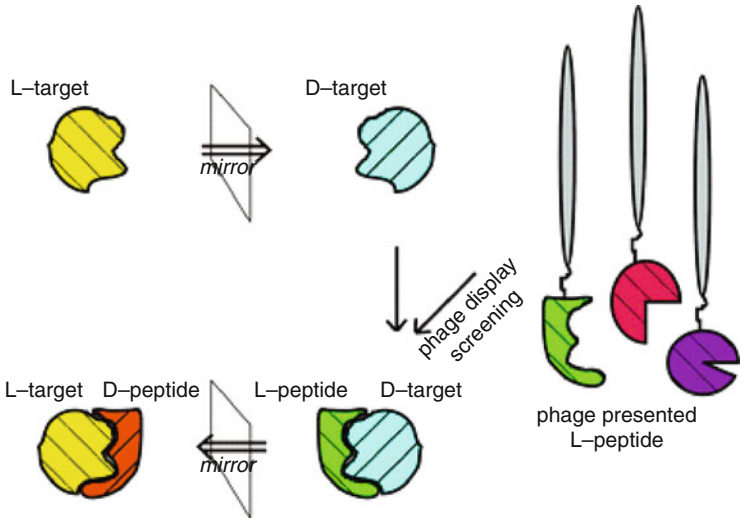


Fig. 3 The principle of mirror image phage-display. The D-enantiomer of any target is chemically synthesized and used for phage display which identifies an L-peptide ligand. The D-enantiomer of the identified L-peptide ligand is chemically synthesized and it binds to the L-enantiomeric target [179]. (Reproduced from [179] with permission from The Royal Society of Chemistry)

ligands bound tightly in their cyclic form, they lost binding affinity in their linear form. The same group provided further validation of this technology by targeting the gp41 pocket, a highly conserved region of the HIV envelope glycoprotein [163]. The gp41 pocket residues were fused to a soluble trimeric coiled coil forming a polypeptide sequence referred to as IQN17. This protein sequence was synthesized in its all-D form denoted D-IQN17 and used to screen a bacteriophage library identifying 12 peptides, 9 of which shared sequence homology and bound selectively to the pocket. Kay et al. subsequently utilized the consensus sequence ($CX_5EWXWLC$) from the above study as a basis for further optimization to develop dimeric and trimeric inhibitors of sub-nanomolar potency [180].

MIPD has also been utilized for potential diagnostic use in Alzheimer's disease. Willbold et al. used the all-D version of Aβ(1–42) amyloid peptide as a target to screen a commercial phage display library to generate a dominant consensus sequence whose all D-amino acid analog "D-pep" bound the native Aβ(1–42) sequence with submicromolar affinity [181]. In addition, a fluorescein-labeled derivative of D-pep was shown to stain native Aβ amyloid fibrils specifically, suggesting it may be useful as a diagnostic probe for amyloid plaques.

The Willbold group also explored the possibility of a therapeutic application to Alzheimer's disease using MIPD-derived peptides [182]. A series of peptides with a dominant consensus sequence referred to as "D3" was shown to reduce Aβ aggregation in a thioflavin fluorescence assay and also disrupt pre-existing Aβ aggregates in vitro. Remarkably, administration of a 9-μg dose of D3 into the hippocampus reduced Aβ plaque load in transgenic APP and PS1 mice after a 4-week treatment.

The utility of MIPD was further demonstrated in cancer immunotherapy with the disclosure of synthetic peptide mimics of human mucin 1 (MUC1), a protein which is expressed in several types of cancer [183]. The authors identified several peptides with no homology to MUC1 which were nevertheless capable of stimulating T-cell response to MUC1 antigen-expressing tumors. Presumably these would be less likely to be neutralized (recognized as "self") by the immune system and deliver a sustained therapeutic response.

The work of the Lu group in the discovery of native and all-D peptides which inhibit the MDM2/MDMX oncoprotein interaction with P53 is notable in its use of NCL to construct the required biotinylated targets and illustrates the impact of NCL in expanding the scope MIPD. An initial report described the NCL synthesis of $^{25–109}$MDM2 and $^{24–108}$MDMX in their unmodified and site-specifically biotinylated forms and the use of these as a target to screen a duodecimal phage library [164]. Two consensus sequences were obtained which both bound MDM2 and MDMX with low nanomolar Kd values. A complex with MDM2 and MDMX with the more potent of the two inhibitors was used to solve the crystal structures of the two respective complexes. These findings fundamentally furthered our understanding of the structural basis of p53 binding and provided insights for drugging this important target. In a subsequent publication from the same group [165], the authors used the MIPD strategy against an all-D MDM2 protein to identify peptide ligands termed ^{D}PMI-α and ^{D}PMI-β which bound to MDM-2 with Kd values of 219 and 34.5 nM, respectively. While unmodified ^{D}PMI-α and related peptides were unable to traverse cell membranes, they proved capable of inhibiting the growth of a mouse glioblastoma when administered in an RGD-liposome formulation [184].

Perhaps the best singular demonstration of the utility of MIPD as drug discovery tool is the identification of D-peptide ligands binding to vascular endothelial growth factor-A (VEGF-A) reported by Mandal [185]. VEGF-A, an angiogenic growth factor whose overexpression correlates with poor prognosis in breast cancer, has been successfully targeted by monoclonal antibodies. The B1 domain of streptococcal protein G (GB1) was chosen as scaffold for phage display affinity screening against the D-VEGF-A (8–109) target. The 56-residue lead inhibitor L-RFX001 bound the D-VEGF-A protein with a Kd of 95 ± 8 nM while D-RFX001 bound the L-VEGF-A isomer with a Kd of 85 ± 12 nM. The ligands also bound with reciprocal chiral specificities: L-RFX001 bound only the D-VEGF target and D-RFX001 only the L-VEGF target. The D and L forms of the VEGF proteins and the enantiomeric ligands were co-crystallized and used to solve their high resolution structure.

4.1.2 Racemic Protein Crystallization

The seminal contributions of Dorothy Hodgkin and Max Perutz involving the structural elucidations of insulin [186] and hemoglobin [187], respectively, stand out as early examples of the power of X-ray protein crystallography, while the HIV

protease work highlighted in the next section highlights its recent importance in drug discovery. The structural characterization of a macromolecule of interest, whether an enzyme, a hormone, or receptor is considered key to its validation as a drug discovery target. Along with NMR, X-ray crystallography continues to serve a primary method of 3D structure determination. One of the principal challenges in utilizing this powerful technique is the requirement of well-ordered protein crystals suitable for X-ray diffraction studies. Unlike small organic molecules, proteins are inherently difficult to crystallize, typically requiring considerable trial-and-error testing to identify optimal conditions. The fundamental reason for this difficulty was addressed by Wukovitz and Yeates [188] who proposed that, in contrast to small organic molecules, proteins tend to be highly solvated in their crystal lattice and utilize a narrow range of chiral space groups with nearly one-third falling into the $P2_12_12_1$ group. This is dictated by the universally homochiral nature of all native proteins which favors the formation of non-centrosymmetrical crystals. The suggestion of MacKay [189] that co-crystallization of equimolar quantities of a D and L protein would crystallize more readily by virtue of centrosymmetry was remarkably prescient. The synthesis of the L and D enantiomers of bacterial rubredoxin, a 45-amino acid metal binding protein containing 4 disulfide bonds by Zawadzke and Berg [190] validated MacKay's prediction. CD analysis of crystals formed by mixing the two enantiomers indicated no signal, confirming the presence of the racemate. Solution of rubredoxin structure using molecular replacement methods demonstrated the centrosymmetric structure of the racemic crystals within the P1 (bar) space group. The racemic method was further validated through the work of Toniolo [191] who determined the structure of trichogin A IV through crystallization of its D and L forms. Through the mid-1990s racemic crystallography was effectively restricted to relatively small proteins and peptides under 50 amino acids in length. The advent of NCL [54, 58–60] largely removed this size limitation and dramatically expanded the scope of this important technique. Over the past 15 years a number of protein structures have been solved through the use of racemic crystallography and have validated the early theoretical assumptions of MacKay [189], Yeates [188], and Berg [192]. Remarkably, the majority of the racemic protein crystals synthesized to date fall into the P1 (bar) space group, in excellent agreement with the predictions of Wukovitz and Yeates [188].

The structure of racemic sweet-tasting protein monellin was solved by molecular replacement and found to form the P1 (bar) crystal space group [193]. The synthetic L-enantiomer of the 81-mer snow flea antifreeze protein (L-sfAFP) crystallized only after a lengthy effort did not yield useful X-ray structure data [194]. In contrast, the racemic solution of L-sfAFP and D-sfAFP yielded crystals in ~50% of the conditions explored for L-sfAFP (Fig. 4). Excellent results were also obtained from "quasi-racemic" crystals composed of D-sfAFP and the L-enantiomer in which the Asn11 was replaced with "pseudo-Se-Gln" (a seleno-Cys residue alkylated with bromoacetamide). The resulting quasi-racemic mixture produced crystals under various conditions and was used to generate Synchrotron X-ray data to a resolution

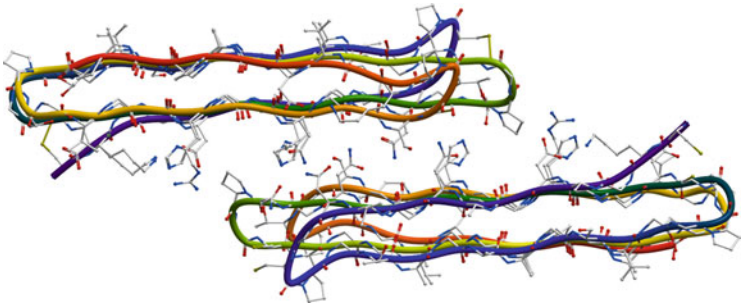


Fig. 4 X-Ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers (PDB code: 3BOG) [194]

of 1.2 Å. Within the span of 2–3 years a number of important proteins were synthesized in their D and L forms by NCL and their structures solved by direct methods. These include BmBKTx1, a 31-amino acid calcium-activated potassium channel blocker isolated from scorpion toxin [195], plactacin, a 40-amino acid antimicrobial protein of fungal origin [196], kaliotoxin, a 38-residue potassium channel blocker [197], and DKP ester insulin [198]. Studies of vascular endothelial growth factor (VEGF) [185] discussed above in reference to their use of MIPD also involved the use of racemic crystallography.

4.2 Chemical Synthesis of Enzymes

The possibility that an enzyme could be assembled entirely by chemical methods is not new and was first raised by Emil Fischer in a note to his contemporary Adolf Bayer in 1905: "My entire yearning is directed toward the first synthetic enzyme. If its preparation falls into my lap with the synthesis of a natural protein material, I will consider my mission fulfilled." [199]. While Fisher made important contributions to peptide chemistry culminating in the synthesis of an 18-mer peptide [4], his dream of enzyme total synthesis went unfulfilled in his lifetime. The synthetic and analytical technologies required to accomplish this ambitious goal would take decades to develop [58].

The first successful efforts in enzyme total synthesis involved ribonuclease in the late 1960s and were disclosed in simultaneous publications by Merrifield and Gutte at the Rockefeller [200] and Denkwalter and co-workers at Merck [45, 46, 201–203]. The Rockefeller group's bovine ribonuclease A (RNase A) synthesis utilized automated, stepwise Boc/Bzl chemistry to assemble the 124 amino acid sequence. Following conversion to the S-sulfonate derivative and purification by ion-exchange and gel chromatography, the material was reduced to the free sulfhydryl form and air oxidized to form the four disulfide bonds. Upon further purification the protein was shown to be identical to the native enzyme by paper

electrophoresis and exhibited appropriate selectivity with respect to an RNA substrate 2′, 3′-cyclic cytidine phosphate (active) vs DNA substrate 2′, 3′-cyclic guanosine phosphate (inactive).

The Merck approach relied on classic fragment condensation methods to assemble the ribonuclease S (RNase-S) fragment comprising residues 21–124 of RNase-A. The sequence had previously been shown to exhibit full enzymatic activity when assayed in the presence of the 1–20 ("S-peptide") fragment [204]. The use of *N*-carboxyanhydrides [205], *N*-thiazolidinediones [206], and *N*-hydroxysuccinimide esters [207] yielded 19 minimally protected fragments requiring only Cbz protection for lysine and the recently introduced S-Acm group for cysteine [208]. The resulting fragments were coupled through a convergent scheme using azide activation [10] to yield the protected 21–124 sequence. Deprotection and air oxidation yielded active enzyme when tested in the presence of S-peptide.

A later synthesis of RNase-A reported by Fujii and Yajima in 1981 [209–215] shared a number of features with the Merck strategy, such as a convergent fragment assembly and azide mediated fragment coupling. In contrast with the Merck method, however, the newer synthesis used substantially shorter fragments as well as a more elaborate side-chain protecting group scheme, including Bzl groups for Glu, Asp and 4-MeBzl for Cys. Global side-chain deprotection by treatment with trifluoromethanesulfonic acid–thioanisole followed by gel chromatography and air oxidation afforded the crude RNase enzyme. In addition, unlike the earlier Rockefeller and Merck syntheses, the Japanese group benefited from the use of affinity chromatography, which dramatically increased the specific activity of the final material when tested against substrate.

While incomplete, the ambitious efforts of Kenner and co-workers to prepare lysozyme are relevant to the discussion. Beyond demonstrating the capability of the synthetic method, one of the stated goals of the work was to study the effect of variations in lysozyme's structure on enzymatic activity [216]. The fragment-based synthetic strategy used *tert*-butyl protection for Ser, Thr, Asp, and Glu, adamantyloxycarbonyl for Orn and Lys, Acm for Cys, and variously Cbz, Boc, and biphenylisopropoxycarbonyl (Bpoc) for the N-termini of fragments. DCC/NHS activation was utilized for fragment coupling throughout. The strategy suffered from the well-known limitations encountered during the fragment-based RNase syntheses.

The AIDS crisis beginning in the late 1980s prompted the next significant series of advances in the chemical synthesis of enzymes. HIV protease (HIV-PT), a homodimeric aspartyl protease with homologies to Raus-sarcoma virus (RSV) [217–219], was shown to be essential for the assembly and maturation [220] of the HIV virus. Inhibition of this enzyme emerged as one of the early strategies for arresting the HIV replication cycle and became the central focus of structure-based or "rational" drug design efforts to develop clinically efficacious drugs [221]. Isolation of the enzyme from its native source, however, was complicated by two factors: the risk of exposure to the virus and the low concentration of the HIV-PT in mature viral particles. Fortunately, sufficient quantities of HIV-PT were obtained by both recombinant [222] and chemical [223] methods. Initial working models based on crystallographic studies of "unliganded" recombinant HIV-PT of Navia et al. [224]

and structural homology to Raus sarcoma protease [218, 225], while useful, differed in several important respects such as the precise structure of the dimer interface. Only later studies using HIV-PT chemically synthesized by Kent et al. [226] provided a refined crystallographic model and resolved important differences between the HIV-PT- and RSV-PT-based models. More importantly, synthetic enzyme was also used in the structural studies of enzyme-inhibitor complexes, a seminal contribution to the design of the first generation HIV-PT inhibitors [227]. In the first study [228], the structure of the HIV-PT complexed with the reduced scissile bond isostere-based inhibitor MVT-101 ($K_i = 780$ nM against synthetic protease) was determined at 2.3 Å resolution and facilitated mapping of the inhibitor's extended amino acid side-chains within the active site. Comparison of the inhibitor bound and unliganded structures noted that the enzyme undergoes a conformational rearrangement as it binds the inhibitor. A follow-up study [229] of an HIV-PT complex with the more potent hydroxyethylamine-based inhibitor JG-365 ($K_i = 0.66$ nM) provided a structural basis for the difference in potency between it and the less potent MVT-101. Analysis of the complex revealed a hydrogen-bonding interaction between the hydroxyl group of the more potent inhibitor with Asp25 and Asp125 carboxyl groups of the enzyme. Information gained through the above studies including a high-resolution three-dimensional structure of the enzyme, the position and conformation of the inhibitor in the enzyme's binding site, and specific contact points between the molecules provided key insights to the multidisciplinary teams involved in the discovery of the first generation HIV-PT inhibitors. The mid-1990s approvals of Saquinavir (Invirase), Ritonavir (Norvir), Indinavir (Crixivan), and Nelfinavir (Viracept) validated the concept of structure-based drug discovery and the role of chemical protein synthesis within this paradigm [227].

While a number of important HIV-PT studies had been conducted with recombinant material, subsequent work by Kent et al. dramatically demonstrates the unique capability of chemical protein synthesis when applied to structure-function problems in chemical biology. The nucleophilicity of Asp-25 and its role in HIV-PT catalysis was investigated through the chemical synthesis and NMR study of an analog incorporating a 13C-labeled Asp25 residue within each monomer [230]. Measurement of the 13C-NMR chemical shifts highlighted important differences in the ionization states of the Asp25 residues when comparing pepstatin-bound and the unliganded enzymes. In the presence of inhibitor, only one Asp25 residue was protonated, the other unprotonated while in the absence of inhibitor both Asp25 residues were chemically equivalent and deprotonated. The catalytic function of the characteristic hairpin loops of HIV-PT, also called "flaps," was explored through an approach referred to as "backbone engineering" [231, 232]. The Ile50 amide is thought to play a key role by providing hydrogen bond stabilization to the enzyme-substrate complex. Replacement of the native amide bond of Ile50 in each flap with an isosteric CO–S– bond lacking hydrogen bonding capability resulted in a dramatic 3,500-fold reduction in kcat relative to control enzyme. The question of whether the presence of one native Ile50 bond would

maintain catalytic activity was answered through the synthesis of a covalent "tethered" dimer where only one of the Ile50 amides was replaced by an ester isostere. The singly substituted enzyme maintained full intrinsic activity, confirming the single flap catalysis model. The Asp-25 13C label studies and the backbone modifications were instrumental in providing key clues to the mechanism of this important enzyme and demonstrated unequivocally the singular power of chemical synthesis to achieve "atom-by-atom" control. It should be noted that the seminal structural studies utilizing chemically synthesized HIV protease took place before the advent of NCL and relied upon highly optimized stepwise SPPS protocols and ligation via non-native amide bond surrogates. The enormous capability of NCL and subsequent advances such as convergent assembly, kinetically controlled ligation (KCL) [158], and desulfurization at cysteine residues [134, 138] greatly expanded the capabilities of chemical protein synthesis. An excellent example of this enhanced capability was exemplified through two more recent syntheses of HIV-PT published by Kent et al. in 2007. The first [233] was accomplished through a sequential four-component NCL assembly of the native 1–99 sequence extended at the C-terminus with an additional 10 residues of the HIV-PT open reading frame connected to an (Arg)10 solubility tag. Following Raney nickel desulfurization of the four cysteines, the folded protein underwent autoprocessing, removing the 20 amino acid extension to yield the mature HIV-PT(1-99) in excellent yield. The second approach involved application of KCL [158] to achieve a convergent ligation of the (1-40)-(α-thioarylester) and the Cys-(A42-A99)-(α-thioalkyl ester) sequences (Fig. 5) [234]. The resulting (A1-A99)-(α-thioarylester) obtained through transthioesterification was then joined to the Cys-Gly4-(B1-B99) fragment

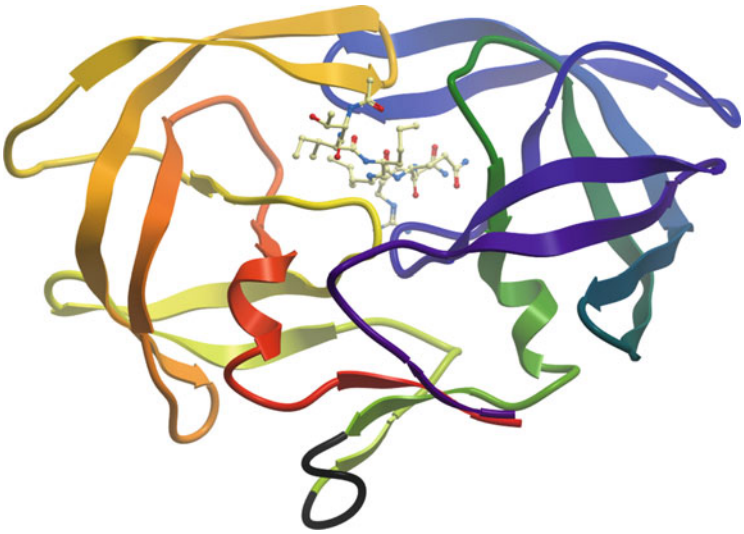


Fig. 5 X-Ray crystallographic structure of the "covalent dimer" HIV protease (shown as *ribbon*) complexed with inhibitor MVT-101 (shown as *stick*). The linker QGGGG is shown as *ribbon in black* (PDB code: 2O40) [234]

by NCL. Conversion of the cysteines at ligation sites with α-bromoacetamide to ψ-Gln followed by folding and dialysis furnished the catalytically active enzyme.

The extensive focus on HIV-PT reflected its role as a critically important drug discovery target. The methodological advances which enabled the tremendous advances in this area, however, have also increased our capability to synthesize other enzymes, some of which are also of therapeutic importance. The synthesis and characterization of the human type II secretory phospholipase A2 (sPLA2), an enzyme which catalyzes the hydrolysis of phospholipid esters, represent an example of a medicinally relevant enzyme prepared by total chemical synthesis [235]. The preparation of the 126-amino acid, 7-disulfide sequence was accomplished by NCL methodology involving a 2-component ligation at the Gly58-Gly59 site. The chemical synthesis of the D and L forms of 4-oxalocrotonate isomerase, an enzyme which catalyzes the isomerization 2-oxo-4(E)-hexanedioate to 2-oxo-3(E)-hexanedioate [236], further demonstrated the utility of this approach in the elucidation of catalytic mechanisms. While the two 4-oxalocrotonate isomerase enantiomers demonstrated equal catalytic activity when tested against an achiral substrate, the reaction was shown to proceed via a stereoselective mechanism at the C5 carbon atom by deuterium labeling. Another, more recent example is the synthesis of lysozyme [159] accomplished by a convergent KCL mediated assembly of the 130-amino acid polypeptide followed by oxidation of the 4 native disulfides to yield enzymatically active material. The landmark 1969 ribonuclease A synthesis of Gutte and Merrifield was revisited by Kent et al. [237] who employed a one-pot KCL strategy to assemble first the 65–124 fragment intermediate which they used in a subsequent one-pot ligation to form the full length ribonuclease protein. The completed linear protein was folded using a redox system and rigorously characterized by several methods including X-ray crystallography. Recently Wu et al. synthesized and characterized the human T-cell leukemia virus type I (HTLV-1) protease, a potentially important target for the treatment of adult T-cell leukemia [238]. The synthesis of this enzyme and its truncated version resolved an important controversy regarding the role of its C-terminal region. Studies with three substrates indicated that deletion of residues 116–125 reduced its catalytic efficiency by a factor of 10.

The above syntheses are not only significant by virtue of the importance of particular enzymes prepared, but by their illustration of the important innovations which have greatly enhanced the capabilities of NCL far beyond the original two-fragment ligation.

4.3 Protein Pharmaceuticals

The direct preparation of drug candidates represents another potentially important niche for chemical protein synthesis. Although most currently marketed FDA approved biologics are produced by fermentation, chemical synthesis can be

fundamentally important in the drug discovery phase, where homogeneity and structural variation can be advantageous.

4.3.1 Chemical Synthesis of Erythropoietin

Erythropoietin (EPO), a cytokine produced by the kidney which regulates the production of red blood cells, is a glycoprotein hormone of 166 amino acids. This protein undergoes extensive and variable glycosylation at four distinct sites (Asn24, 38, 83, and Ser126) when expressed in cell culture, yielding a mixture of complex and largely inseparable glycoforms. In fact, the commercial preparations of EPO, Procrit (Janssen), Epogen (Amgen), and Aranesp (Amgen), prescribed for anemia associated with chronic kidney disease or cancer chemotherapy, are each supplied as a mixture of glycoforms. However, studies have shown that the carbohydrate motifs at the four sites could significantly impact the in vitro stability and in vivo performance of EPO [239–241]. As a result, the chemical synthesis of EPO analogs, which enable homogenous modification by glycans or glycan-mimicking polymers, is essential to correlating carbohydrate modification with biological activity [242, 243].

In 2001, Flitch et al. reported the semi-synthesis of a homogeneously glycosylated EPO by producing the protein in *E. coli* with Asn to Cys mutations and subsequent modification with glycosyl-β-*N*-iodoacetamide at the mutated sites [244]. In 2003, Kochendoerfer et al. at Gryphon Therapeutics reported a full chemical synthesis of EPO with two mono-dispersed polymers attached at residues 24 and 126 [245]. A follow-up report in 2005 from the same group detailed a correlation between the in vivo half-life, overall charge, and molecular weight of EPO [246]. In 2008 Macmillan et al. reported a semi-synthetic approach utilizing cyanogen bromide cleavage of *E. coli*-derived material to generate the 29–166 fragment for ligation to the synthetic 1–28 fragment [247]. This work was extended to synthesize an EPO analog with an undecasaccharide attached at sites 24 and 30 [248]. In 2012, Kent et al. published another synthesis of a non-glycosylated EPO analog with Asn to Lys mutation at sites 24, 38, and 83 [249]. Working toward a full chemical synthesis of glycosylated EPO, Kajihara et al. reported their synthesis of an EPO analog with a complex undecasaccharide attached at the native glycosylation site Asn83 [250]. A 10-year effort by the Danishefsky group requiring the development of a number of specialized synthetic methods [101, 128, 138, 144, 243, 251–253] culminated with the report of fully synthetic EPO with "bio-level" complex polysaccharides attached at all four native glycosylated sites [254]. The synthetic routes reported by the Gryphon and Danishefsky groups are highlighted below.

With the aim of improving the pharmacokinetic properties and efficacy of EPO, the scientists at Gryphon designed a synthetic erythropoiesis protein (SEP), in which glycosylation of native EPO at residues Asn24, 38, 83, and Ser126 was replaced through the attachment of two negatively charged branched polymers of mono-dispersed molecular weight at residues 24 and 126 (Fig. 6a, b). Native

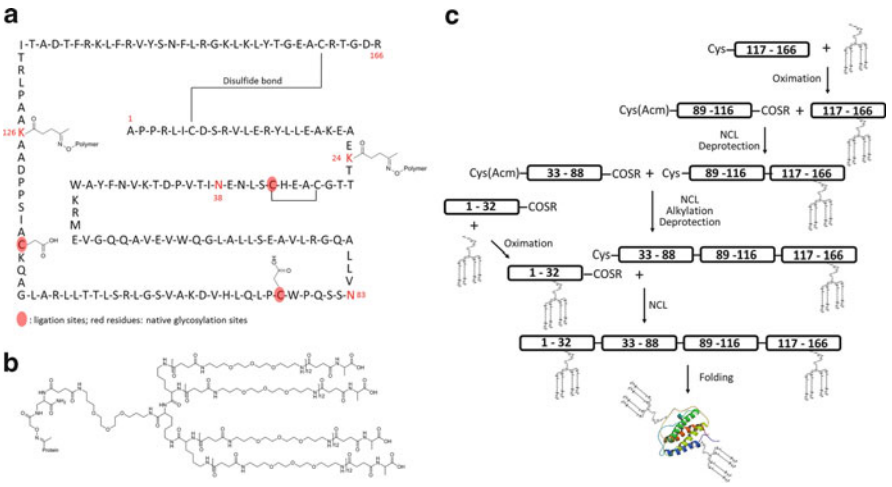


Fig. 6 Gryphon SEP route

residues Asn24 and Ser126 were substituted with Lys(*N*-levulinyl) to enable chemoselective attachments of the polymer through oxime ligation. The distribution of the native cysteine residues (Cys7, 29, 33, and 161) did not permit efficient use of NCL, and therefore two additional cysteines (at residues 89 and 117) were introduced in the Cys-free region of 34–160. After NCL, two non-native cysteines were alkylated with bromoacetic acid, thus mimicking the native glutamic acids. A sequential C-to-N ligation strategy was employed for the fragment assembly starting with the attachment of a mono-dispersed polymer chain to segment 117–166 through an oxime linkage, with the resulting product then ligated with segment 89–116 and followed by Acm deprotection by $Hg(OAc)_2$. Similar schemes were utilized to obtain the linear form of EPO 1–166 with polymers attached at residues 24 and 126, which was subsequently subjected to folding to furnish the SEP with two native disulfide bonds. The reaction yields were 15–30% for solid-phase peptide synthesis and purification, 30–50% for oxime-forming ligation, 40–70% for NCL and purification, and 25–40% for final folding and purification. This process was scalable by a factor of 100 with comparable yields, and more than 100 mg of material was generated. The comparison of the SEP analog with native microbial-derived EPO found similar in vitro potencies, with the former demonstrating a prolonged half-life and enhanced performance in animal models (Fig. 6c).

In 2002, the Danishefsky group set out to prepare a homogeneous "wild-type" EPO polypeptide glycosylated at the three conserved N-linked sites as well as at the single O-linked center. After an effort of ~10 years, this objective was completed in 2013. The synthetic plan called for the installation of a tetrasaccharide glycophorin motif at Ser126, and a consensus dodecasaccharide at all three Asn-linked sites (Fig. 7). In addition to NCL, a number of synthetic innovations proved instrumental to the success, including the metal-free desulfurization [138], the *o*-mercaptoaryl ester rearrangement-mediated NCL [252], and the one-flask aspartylation

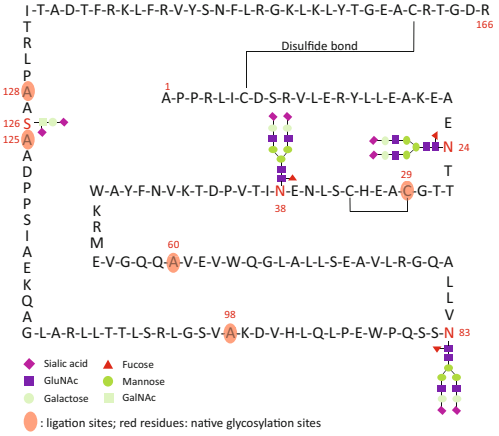


Fig. 7 Structure of EPO synthesized by the Danishefsky group

[253, 255]. The free radical-based desulfurization offers great selectivity between a free thiol and Cys(Acm) or Thz, facilitating the utilization of four Ala sites (Ala60, 98, 125, and 128) as NCL junctions in addition to the native Cys29. The modified aspartylation procedure, improved by the introduction of pseudoproline dipeptide segments, was also critical to the efficient preparation of the N-linked glycopeptide segments. The segment assembly was conducted by sequential ligation in the C-to-N direction utilizing Thz or Cys(Acm) as the protected forms of N-terminal Cys. All steps proceeded in good to excellent yields, including the critical one-step metal-free desulfurization of segment 29–166 which selectively converted four cysteines to alanines at sites 60, 98, 125, and 128 in the presence of three Cys (Acm) with a yield of 69%. The subsequent treatment with AgOAc removed all Acm groups in a 79% yield, liberating the free cysteines for the final ligation with segment 1–28. The resulting linear precursor was successfully folded to yield a biologically fully active EPO analog at sub-milligram scale, representing the first synthetic glycoform with glycosylation at bio-level complexity (Fig. 8).

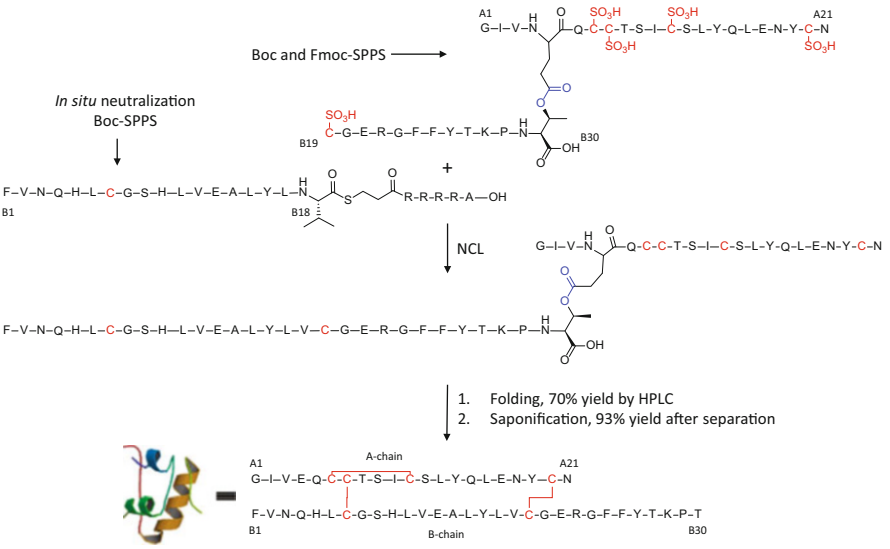
4.3.2 Chemical Synthesis of Insulin

Insulin is a mini-protein hormone produced by pancreatic beta cells, which promotes the uptake of glucose into skeletal muscle and fat tissue [256]. It possesses a unique structural topology, consisting of an A-chain of 21 residues and a B-chain of 30 residues linked by two disulfide bonds, and a third intra-A chain disulfide. Since its identification in the 1920s by Banting and Best [257], insulin and its analogues have served as life-saving medicines for diabetic patients, and will undoubtedly continue to do so until pancreas regeneration becomes clinically practical [256]. The industrial scale production of insulin relied for many years solely on the extraction of bovine and porcine pancreatic glands, with this supply now met entirely by microbial fermentation. Numerous early structure activity relationship studies (SAR) utilized the semi-synthetic preparation of analogs [258–264]. These

Fig. 8 Danishefsky route towards a fully glycosylated EPO

were, however, inherently limited in scope by providing access only to the N-termini of A- and B-chains and the C-terminus of the B-chain [265]. Recombinant DNA technology, which has been the mainstay of large scale insulin production [266, 267], provides a means to substitute each residue independently, but with the restriction that only coded amino acids can be introduced. These limitations pose significant constraints on the scope of insulin discovery research. The total chemical synthesis of insulin offers the ability to expand the structural diversity at each position, facilitating more thorough investigations into insulin's chemistry, biology, and therapeutic applications. In addition, the methods developed for insulin synthesis are also applicable to the synthesis of other members of the insulin-like-peptide family [268, 269].

The elucidation of the structure of insulin in the 1950s by Sanger et al. prompted attempts to synthesize the hormone chemically [270]. These remarkable achievements were accomplished over a 10-year span by three independent international teams: the consortium of Chinese institutes [271, 272], the group of Helmut Zahn at the University of Aachen [273], and the group of Panayotis Katsoyannis at the Biochemical Institute of Pittsburgh [274]. Solution-phase routes were utilized throughout the preparation of the 21-mer A-chain and the 30-mer B-chain in their S-sulfonate forms, which were subsequently combined under re-dox conditions to form the three native disulfide bonds and provided the biologically active protein. Shortly afterwards, Merrifield et al. achieved the same goal with application of the recently invented SPPS technique for the synthesis of A- and B-chains [275]. Because of the typically limited yield of the chain combination reaction, directed-disulfide bond formation approaches were investigated and successfully accomplished by Sieber et al. in 1974 [276] and later by Kiso et al. in 1993 [277]. In addition to the challenge of correctly forming the three native disulfide bonds, the insolubility of A-chain posed another major obstacle to achieving a highly efficient synthesis. A poly-lysine tag was introduced as a solubilizing motif for the synthesis of insulin A-chain by Wade et al. in 2009 and demonstrated by the synthesis of insulin glargine (Lantus) [278]. Recently, Kent [198, 279, 280], Hoeg-Jensen [281], and DiMarchi [282] and their co-workers reported the use of a bio-mimetic approach where the N-terminus of A-chain and the C-terminus of B-chain are linked together to form a mini-proinsulin enabling highly efficient folding. The various linkers explored so far include an oxime-containing PEG linker [279], amide bond [282], short peptide [281], and ester bond [198, 280]. More recently, Liu et al. at Eli Lilly and others have reported improved directed-disulfide bond formation methods, including a one-pot procedure [283, 284], solubility enhancement by insertion of isoacyl segments [285], and iodine-free procedures to eliminate side reactions induced by iodine oxidation [286, 287]. Overall, these methods can be classified into two categories based on the strategy used for construction of the native disulfide bond: (1) methods exploiting the insulin protein structure to help direct disulfide bond formation using both a direct two-chain combination method and an indirect method using a biomimetic single-chain insulin intermediate [198, 271, 273–275, 279–282] and, (2) those methods using differential thiol chemical reactivity to help direct disulfide bond formation [276–278, 283,

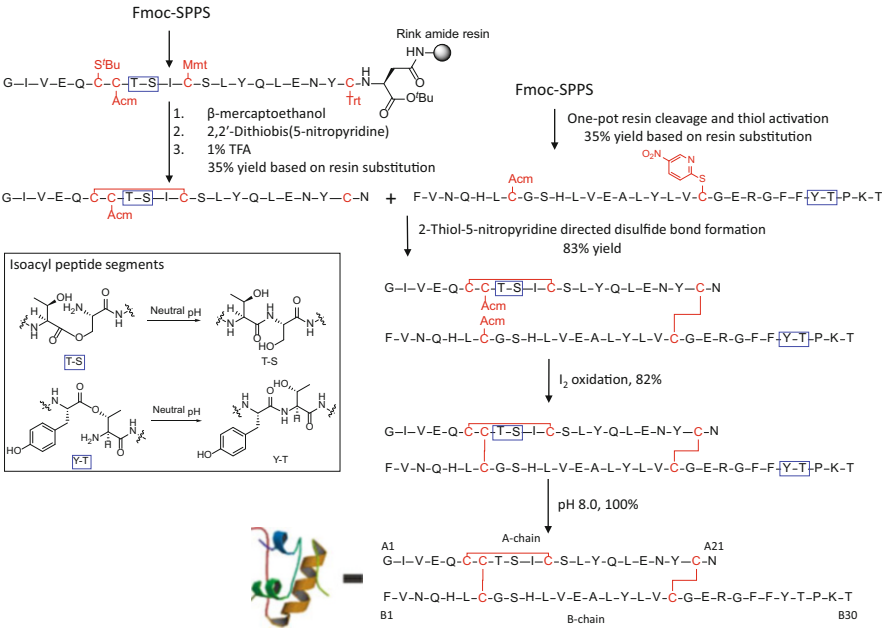


Scheme 17 Synthesis of insulin lispro via an "ester-insulin" intermediate

285–287]. The "ester-insulin" approach reported by Kent et al. [280] and the "isoacyl-insulin" approach reported by Liu et al. [285] are detailed below as illustrations of each strategy. Both methods have been demonstrated as effective routes to produce insulin analogs at milligram scale.

The elegant utilization of an ester bond as a means of linking the A-chain N-terminus to the B-chain C-terminus enabled an enzyme-free, traceless excision of the linker (Scheme 17) [280]. In this synthesis of insulin lispro (Humalog), the A- and B-chain were joined together via an ester bond between the side-chains of two native residues, GluA4 and ThrB30. To obtain the necessary segments for NCL, thioester segment B1–B18 was prepared by standard in situ neutralization Boc-SPPS [78], while the B19–A21 segment was assembled by a combination of Fmoc-SPPS (for residue A1-A21) and Boc-SPPS (for B19-B30). The 5 cysteines of segment B19-A21 were converted to their respective S-sulfonates to facilitate purification and improve handling. Separately, a tetra-arginine tag was installed at the C-terminus of thioester segment B1–B18 to improve its solubility. The ligation of these two segments was conducted under standard NCL conditions, followed by the folding of the resulting depsi-linear precursor to form three native disulfide bonds. In the last step, saponification furnished biologically active insulin lispro. Both the folding and saponification steps proceeded with high yields of 70% and 93% respectively.

The protein folding approach, which relies on the conformational characteristics of the insulin molecule, may have limited utility in cases where a specific mutation affects the analog's stability [288–290]. The chemically directed disulfide bond formation is theoretically able to generate analogs with any mutation, a key advantage of this approach. However, although the orthogonal chemistries of the



Scheme 18 Synthesis of human insulin by isoacyl A and B-chains

protected cysteine pairs is sufficient to enable the sequential formation of three disulfide bonds, the handling difficulties of the hydrophobic A-chain significantly decreases the overall yield [277, 278, 283]. A recent report by Liu et al. demonstrated the insertion of isoacyl segments into A- and B-chains as an effective way to circumvent this issue (Scheme 18) [285]. The isoacyl segments Thr-Ser and Tyr-Thr were placed at A8-A9 and B25-B26 of the A- and B-chains, respectively. The increased solubility imparted by isoacyl motifs enabled efficient purification of A-chain by standard C18 reverse-phase HPLC (RP-HPLC), and also increased the recovery of B-chain from RP-HPLC purification two- to threefold. The isoacyl motifs rearranged smoothly to their corresponding native peptide bond forms when exposed to neutral or basic pH conditions. The overall yield of human insulin based on substitution of the starting resin used for the A-chain synthesis was 24%, representing the highest yield reported to date.

5 Conclusion

The chemical synthesis of proteins has become a practical reality only within the last two decades. The advent of NCL chemistry in 1994 effectively overcame the limits of conventional stepwise SPPS and dramatically increased the range of protein targets accessible by chemical means. This relatively recent innovation

has played a significant role in the development of novel drug discovery technologies, impacting both small molecule and large molecule projects. As is characteristic of transformational technologies, its future uses are not easy to anticipate and the examples discussed above are undoubtedly only the first of many potential applications of this unique and powerful technology.

Acknowledgement The authors wish to thank Prof. Stephen B. H. Kent of the University of Chicago for his helpful comments and suggestions during the preparation of this manuscript.

References

1. Nelson DL, Cox MM (2004) Lehninger principles of biochemistry, 4th edn. W. H. Freeman, New York
2. Wallace CJA (ed) (2000) Protein engineering by semisynthesis. CRC, Boca Raton
3. Curtius T (1881) J Prakt Chem 24:239
4. Fischer E (1907) Ber Dtsch Chem Ges 40:1754
5. Bergmann M, Zervas L (1932) Ber Dtsch Chem Ges 65:1192
6. Harington CR, Mead TH (1935) Biochem J 29:1602
7. Sifferd RH, du Vigneaud V (1935) J Biol Chem 108:753
8. du Vigneaud V, Ressler C, Swan JM, Roberts CW, Katsoyannis PG (1954) J Am Chem Soc 76:3115
9. Curtius T (1904) J Prakt Chem 70:57
10. Curtius T (1902) Ber Dtsch Chem Ges 35:3226
11. Fischer E (1905) Ber Dtsch Chem Ges 38:605
12. Sheehan JC, Hess GP (1955) J Am Chem Soc 77:1067
13. Khorana HG (1955) Chem Ind 19:1087
14. Carpino LA (1957) J Am Chem Soc 79:98
15. Anderson GW, McGregor AC (1957) J Am Chem Soc 79:6180
16. McKay FC, Albertson NF (1957) J Am Chem Soc 79:4686
17. Birch AJ (1944) J Chem Soc 430
18. Schwyzer R, Sieber P (1963) Nature 199:172
19. Merrifield RB (1963) J Am Chem Soc 85:2149
20. Sakakibara S, Shimonishi Y, Okada M, Kishida Y (1967) Peptides. In: Proceedings of the 8th European peptide symposium, Noordwijk, The Netherlands, p 44
21. Letsinger RL, Kornet MJ (1963) J Am Chem Soc 85:3045
22. König W, Geiger R (1970) Chem Ber 103:788
23. Carpino LA, Han GY (1970) J Am Chem Soc 92:5748
24. Carpino LA, Han GY (1972) J Org Chem 37:3404
25. Fields GB, Noble RL (1990) Int J Pept Protein Res 35:161
26. Merrifield RB (1964) Biochemistry 3:1385
27. Mitchell AR, Erickson BW, Ryabtsev MN, Hodges RS, Merrifield RB (1976) J Am Chem Soc 98:7357
28. Pietta PG, Marshall GR (1970) J Chem Soc Chem Commun 650
29. Matsueda GR, Stewart JM (1981) Peptides 2:45
30. Wang SS (1973) J Am Chem Soc 95:1328
31. Sheppard RC, Williams BJ (1982) Int J Pept Protein Res 20:451
32. Sheppard RC, Williams BJ (1982) J Chem Soc Chem Commun 587
33. Riniker B, Flörsheimer A, Fretz H, Sieber P, Kamber B (1993) Tetrahedron 49:9307
34. Albericio F, Barany G (1987) Int J Pept Protein Res 30:206

35. Rink H (1987) Tetrahedron Lett 28:3787
36. Sieber P (1987) Tetrahedron Lett 28:2107
37. Barlos K, Gatos D, Schäfer W (1991) Angew Chem Int Ed 30:590
38. Barlos K, Chatzi O, Gatos D, Stavropoulos G (1991) Int J Pept Protein Res 37:513
39. Zalipsky S, Albericio F, Barany G (1985) Peptides: structure and function. In: Proceedings of the 9th American peptides symposium, p 257
40. Bayer E, Hemmasi B, Albert K, Rapp W, Dengler M (1983) Peptides. In: Hruby VJ, Rich DH (eds) The 8th American peptide symposium. Pierce Chemical Company, Rockford, p 87
41. Cote S (2005) New polyether based monomers, crosslinkers, and highly crosslinked amphiphile polyether resins. WO 2005012277 A1 20050210, Matrix Innovation Inc, Quebec
42. Rademann J, Grotli M, Meldal M, Bock K (1999) J Am Chem Soc 121:5459
43. Garcia-Martin F, Quintanar-Audelo M, Garcia-Ramos Y, Cruz LJ, Gravel C, Furic R, Cote S, Tulla-Puche J, Albericio F (2006) J Comb Chem 8:213
44. Li W, Xiao X, Czarnik AW (1999) J Comb Chem 1:127
45. Denkewalter RG, Veber DF, Holly FW, Hirschmann R (1969) J Am Chem Soc 91:503
46. Hirschmann R, Nutt RF, Veber DF, Vitali RA, Varga SL, Jacob TA, Holly FW, Denkewalter RG (1969) J Am Chem Soc 91:507
47. Bray BL (2003) Nat Rev Drug Discovery 2:587
48. Blake J (1981) Int J Pept Protein Res 17:273
49. Aimoto S (1999) Biopolymers 51:247
50. Schnoelzer M, Kent SBH (1992) Science 256:221
51. Rose K (1994) J Am Chem Soc 116:30
52. Liu C-F, Tam JP (1994) J Am Chem Soc 116:4149
53. Chang TK, Jackson DY, Burnier JP, Wells JA (1994) Proc Natl Acad Sci USA 91:12544
54. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Science 266:776
55. Wieland T, Bokelmann E, Bauer L, Lang HU, Lau H (1953) Liebigs Ann Chem 583:129
56. Kemp DS, Galakatos NG (1986) J Org Chem 51:1821
57. Fotouhi N, Galakatos NG, Kemp DS (1989) J Org Chem 54:2803
58. Dawson PE, Kent SBH (2000) Annu Rev Biochem 69:923
59. Nilsson BL, Soellner MB, Raines RT (2005) Annu Rev Biophys Biomol Struct 34:91
60. Kent SBH (2009) Chem Soc Rev 38:338
61. Muir TW, Sondhi D, Cole PA (1998) Proc Natl Acad Sci USA 95:6705
62. Saxon E, Bertozzi CR (2000) Science 287:2007
63. Nilsson BL, Kiessling LL, Raines RT (2000) Org Lett 2:1939
64. Bode JW, Fox RM, Baucom KD (2006) Angew Chem Int Ed 45:1248
65. Zhang Y, Xu C, Lam HY, Lee CL, Li X (2013) Proc Natl Acad Sci USA 110:6657
66. Li X, Lam HY, Zhang Y, Chan CK (2010) Org Lett 12:1724
67. Moore S, Spackman DH, Stein WH (1958) Anal Chem 30:1185
68. Smith AJ (1997) Methods Enzymol 289:419
69. Grant GA, Crankshaw MW, Gorka J (1997) Methods Enzymol 289:395
70. Edman P, Begg G (1967) Eur J Biochem 1:80
71. Edman P (1950) Acta Chem Scand 4:283
72. Matsueda GR, Haber E, Margolies MN (1981) Biochemistry 20:2571
73. Mant CT, Kondejewski LH, Cachia PJ, Monera OD, Hodges RS (1997) Methods Enzymol 289:426
74. Burdick DJ, Stults JT (1997) Methods Enzymol 289:499
75. Johnson ECB, Kent SBH (2006) J Am Chem Soc 128:6640
76. Hackeng TM, Griffin JH, Dawson PE (1999) Proc Natl Acad Sci USA 96:10068
77. Hojo H, Kwon Y, Kakuta Y, Tsuda S, Tanaka I, Hikichi K, Aimoto S (1993) Bull Chem Soc Jpn 66:2700
78. Schnoelzer M, Alewood P, Jones A, Alewood D, Kent SBH (1992) Int J Pept Protein Res 40:180
79. Mende F, Seitz O (2011) Angew Chem Int Ed 50:1232

80. Bu X, Xie G, Law CW, Guo Z (2002) Tetrahedron Lett 43:2419
81. Li X, Kawakami T, Aimoto S (1998) Tetrahedron Lett 39:8669
82. Clippingdale AB, Barrow CJ, Wade JD (2000) J Pept Sci 6:225
83. Futaki S, Sogawa K, Maruyama J, Asahara T, Niwa M (1997) Tetrahedron Lett 38:6237
84. Yamamoto N, Tanabe Y, Okamoto R, Dawson PE, Kajihara Y (2008) J Am Chem Soc 130:501
85. Alsina J, Yokum TS, Albericio F, Barany G (1999) J Org Chem 64:8761
86. Tulla-Puche J, Barany G (2004) J Org Chem 69:4101
87. Ficht S, Payne RJ, Guy RT, Wong C-H (2008) Chem Eur J 14:3620
88. Lelievre D, Barta P, Aucagne V, Delmas AF (2008) Tetrahedron Lett 49:4016
89. Li L, Wang P (2007) Tetrahedron Lett 48:29
90. Kenner GW, McDermott JR, Sheppard RC (1971) Chem Commun 636
91. Backes BJ, Virgilio AA, Ellman JA (1996) J Am Chem Soc 118:3055
92. Flavell RR, Huse M, Goger M, Trester-Zedlitz M, Kuriyan J, Muir TW (2002) Org Lett 4:165
93. Quaderer R, Hilvert D (2001) Org Lett 3:3181
94. Blanco-Canosa JB, Dawson PE (2008) Angew Chem Int Ed 47:6851
95. Fang G-M, Li Y-M, Shen F, Huang Y-C, Li J-B, Lin Y, Cui H-K, Liu L (2011) Angew Chem Int Ed 50:7645
96. Tofteng AP, Soerensen KK, Conde-Frieboes KW, Hoeg-Jensen T, Jensen KJ (2009) Angew Chem Int Ed 48:7411
97. Eom KD, Tam JP (2011) Org Lett 13:2610
98. Zheng JS, Cui HK, Fang GM, Xi WX, Liu L (2010) ChemBioChem 11:511
99. Botti P, Villain M, Manganiello S, Gaertner H (2004) Org Lett 6:4861
100. Liu F, Mayer JP (2013) J Org Chem 78:9848
101. Warren JD, Miller JS, Keding SJ, Danishefsky SJ (2004) J Am Chem Soc 126:6576
102. Kawakami T, Sumida M, Nakamura K, Vorherr T, Aimoto A (2005) Tetrahedron Lett 46:8805
103. Ollivier N, Behr JB, El-Mahdi O, Blanpain A, Melnyk O (2005) Org Lett 7:2647
104. Nagaike F, Onuma Y, Kanazawa C, Hojo H, Ueki A, Nakahara Y, Nakahara Y (2006) Org Lett 8:4465
105. Ohta Y, Itoh S, Shigenaga A, Shintaku S, Fujii N, Otaka A (2006) Org Lett 8:467
106. Kang J, Richardson JP, Macmillan D (2009) Chem Commun 407
107. Ollivier N, Dheur J, Mhidia R, Blanpain A, Melnyk O (2010) Org Lett 12:5238
108. Hou W, Zhang X, Li F, Liu CF (2011) Org Lett 13:386
109. Sharma R, Tam JP (2011) Org Lett 13:5176
110. Kawakami T, Aimoto S (2007) Chem Lett 36:76
111. Zheng JS, Chang HN, Wang FL, Liu L (2011) J Am Chem Soc 133:11080
112. Hojo H, Onuma Y, Akimoto Y, Nakahara Y, Nakahara Y (2007) Tetrahedron Lett 48:25
113. Nakamura K, Mori H, Kawakami T, Hojo H, Nakahara Y, Aimoto S (2007) Int J Pept Res Ther 13:191
114. Hojo H, Murasawa Y, Katayama H, Ohira T, Nakaharaa Y, Nakahara Y (2008) Org Biomol Chem 6:1808
115. Ozawa C, Katayama H, Hojo H, Nakahara Y, Nakahara Y (2008) Org Lett 10:3531
116. Kang J, Reynolds NL, Tyrrell C, Dorin JR, Macmillan D (2009) Org Biomol Chem 7:4918
117. Nakamura K, Kanao T, Uesugi T, Hara T, Sato T, Kawakami T, Aimoto S (2009) J Pept Sci 15:731
118. Tsuda S, Shigenaga A, Bando K, Otaka A (2009) Org Lett 11:823
119. Katayama H, Hojo H, Shimizu I, Nakahara Y, Nakahara Y (2010) Org Biomol Chem 8:1966
120. Richardson JP, Chan C-H, Blanc J, Saadi M, Macmillan D (2010) Org Biomol Chem 8:1351
121. Taichi M, Hemu X, Qiu Y, Tam JP (2013) Org Lett 15:2620
122. Pe'er I, Felder CE, Man O, Silman I, Sussman JL, Beckmann JS (2003) Proteins 54:20
123. Offer J (2010) Biopolymers 94:530
124. Canne LE, Bark SJ, Kent SBH (1996) J Am Chem Soc 118:5891

125. Botti P, Carrasco MR, Kent SBH (2001) Tetrahedron Lett 42:1831
126. Low DW, Hill MG, Carrasco MR, Kent SBH, Botti P (2001) Proc Natl Acad Sci USA 98:6554
127. Offer J, Boddy CNC, Dawson PE (2002) J Am Chem Soc 124:4642
128. Wu B, Chen J, Warren JD, Chen G, Hua Z, Danishefsky SJ (2006) Angew Chem Int Ed 45:4116
129. Li J, Cui H-K, Liu L (2010) Tetrahedron Lett 51:1793
130. Ficht S, Payne RJ, Brik A, Wong C-H (2007) Angew Chem Int Ed 46:5975
131. Yang Y-Y, Ficht S, Brik A, Wong C-H (2007) J Am Chem Soc 129:7690
132. Lutsky M-Y, Nepomniaschiy N, Brik A (2008) Chem Commun 1229
133. Spasser L, Ajish Kumar KS, Brik A (2011) J Pept Sci 17:252
134. Yan LZ, Dawson PE (2001) J Am Chem Soc 123:526
135. Pentelute BL, Kent SBH (2007) Org Lett 9:687
136. Villain M, Vizzavona J, Rose K (2001) Chem Biol 8:673
137. Bang D, Kent SBH (2004) Angew Chem Int Ed 43:2534
138. Wan Q, Danishefsky SJ (2007) Angew Chem Int Ed 46:9248
139. Siman P, Blatt O, Moyal T, Danieli T, Lebendiker M, Lashuel HA, Friedler A, Brik A (2011) ChemBioChem 12:1097
140. Cergol KM, Thompson RE, Malins LR, Turner P, Payne RJ (2014) Org Lett 16:290
141. Moyal T, Hemantha HP, Siman P, Refua M, Brik A (2013) Chem Sci 4:2496
142. Thompson RE, Liu X, Alonso-Garcia N, Pereira PJB, Jolliffe KA, Payne RJ (2014) J Am Chem Soc 136:8161
143. Hojo H, Matsumoto Y, Nakahara Y, Ito E, Suzuki Y, Suzuki M, Suzuki A, Nakahara Y (2005) J Am Chem Soc 127:13720
144. Chen G, Wan Q, Tan Z, Kan C, Hua Z, Ranganathan K, Danishefsky SJ (2007) Angew Chem Int Ed 46:7383
145. Zhang L, Tam JP (1997) Tetrahedron Lett 38:3
146. Payne RJ, Ficht S, Greenberg WA, Wong C-H (2008) Angew Chem Int Ed 47:4411
147. Thomas GL, Payne RJ (2009) Chem Commun 4260
148. Thomas GL, Hsieh YSY, Chun CKY, Cai ZL, Reimers JR, Payne RJ (2011) Org Lett 13:4770
149. Saxon E, Armstrong JI, Bertozzi CR (2000) Org Lett 2:2141
150. Chiang YL, Russak JA, Carrillo N, Bode JW (2012) Helv Chim Acta 95:2481
151. Fukuzumi T, Ju L, Bode JW (2012) Org Biomol Chem 10:5837
152. Ogunkoya AO, Pattabiraman VR, Bode JW (2012) Angew Chem Int Ed 51:9693
153. Pattabiraman VR, Ogunkoya AO, Bode JW (2012) Angew Chem Int Ed 51:5114
154. Pusterla I, Bode JW (2012) Angew Chem Int Ed 51:513
155. Bang D, Kent SBH (2005) Proc Natl Acad Sci USA 102:5014
156. Canne LE, Botti P, Simon RJ, Chen Y, Dennis EA, Kent SBH (1999) J Am Chem Soc 121:8720
157. Johnson ECB, Durek T, Kent SBH (2006) Angew Chem Int Ed 45:3283
158. Bang D, Pentelute BL, Kent SBH (2006) Angew Chem Int Ed 45:3985
159. Durek T, Torbeev VY, Kent SBH (2007) Proc Natl Acad Sci USA 104:4846
160. Torbeev Vladimir Y, Kent SBH (2007) Angew Chem Int Ed 46:1667
161. Sigler PB, Xu Z, Rye HS, Burston SG, Fenton WA, Horwich AL (1998) Annu Rev Biochem 67:581
162. Schumacher TNM, Mayr LM, Minor DL Jr, Milhollen MA, Burgess MW, Kim PS (1996) Science 271:1854
163. Eckert DM, Malashkevich VN, Hong LH, Carr PA, Kim PS (1999) Cell 99:103
164. Pazgier M, Liu M, Zou G, Yuan W, Li C, Li C, Li J, Monbo J, Zella D, Tarasov SG, Lu W (2009) Proc Natl Acad Sci USA 106:4665
165. Liu M, Pazgier M, Li C, Yuan W, Li C, Lu W (2010) Angew Chem Int Ed 49:3649
166. Whitty A, Kumaravel G (2006) Nat Chem Biol 2:112
167. Sackett DL, Sept D (2009) Nat Chem 1:596

168. Geysen HM, Meloen RH, Barteling SJ (1984) Proc Natl Acad Sci USA 81:3998
169. Cwirla SE, Peters EA, Barrett RW, Dower WJ (1990) Proc Natl Acad Sci USA 87:6378
170. Scott JK, Smith GP (1990) Science 249:386
171. Smith GP (1985) Science 228:1315
172. Devlin JJ, Panganiban LC, Devlin PE (1990) Science 249:404
173. Scott JK, Loganathan D, Easley RB, Gong X, Goldstein IJ (1992) Proc Natl Acad Sci USA 89:5398
174. Roberts BL, Markland W, Ley AC, Kent RB, White DW, Guterman SK, Ladner RC (1992) Proc Natl Acad Sci USA 89:2429
175. Yanofsky SD, Baldwin DN, Butler JH, Holden FR, Jacobs JW, Balasubramanian P, Chinn JP, Cwirla SE, Peters-Bhatt E et al (1996) Proc Natl Acad Sci USA 93:7381
176. Wrighton NC, Farrell FX, Chang R, Kashyap AK, Barbone FP, Mulcahy LS, Johnson DL, Barrett RW, Jolliffe LK, Dower WJ (1996) Science 273:458
177. Cwirla SE, Balasubramanian P, Duffin DJ, Wagstrom CR, Gates CM, Singer SC, Davis AM, Tansik RL, Mattheakis LC, Boytos CM, Schatz PJ, Baccanari DP, Wrighton NC, Barrett RW, Dower WJ (1997) Science 276:1696
178. Milton RCL, Milton SCF, Kent SBH (1992) Science 256:1445
179. Funke SA, Willbold D (2009) Mol BioSyst 5:783
180. Welch BD, VanDemark AP, Heroux A, Hill CP, Kay MS (2007) Proc Natl Acad Sci USA 104:16828
181. Wiesehan K, Buder K, Linke RP, Patt S, Stoldt M, Unger E, Schmitt B, Bucci E, Willbold D (2003) ChemBioChem 4:748
182. van Groen T, Wiesehan K, Funke SA, Kadish I, Nagel-Steger L, Willbold D (2008) ChemMedChem 3:1848
183. James TC, Bond U (2012) PLoS One 7:e49728
184. Liu M, Li C, Pazgier M, Li C, Mao Y, Lv Y, Gu B, Wei G, Yuan W, Zhan C, Lu W-Y, Lu W (2010) Proc Natl Acad Sci USA 107:14321
185. Mandal K, Uppalapati M, Ault-Riche D, Kenney J, Lowitz J, Sidhu SS, Kent SBH (2012) Proc Natl Acad Sci USA 109:14779
186. Blundell TL, Cutfield JF, Cutfield SM, Dodson EJ, Dodson GG, Hodgkin DC, Mercola DA, Vijayan M (1971) Nature 231:506
187. Green DW, Ingram VM, Perutz MF (1954) Proc R Soc London Ser A 225:287
188. Wukovitz SW, Yeates TO (1995) Nat Struct Biol 2:1062
189. Mackay AL (1989) Nature 342:133
190. Zawadzke LE, Berg JM (1992) J Am Chem Soc 114:4002
191. Toniolo C, Peggion C, Crisma M, Formaggio F, Shui X, Eggleston DS (1994) Nat Struct Biol 1:908
192. Zawadzke LE, Berg JM (1993) Proteins 16:301
193. Hung LW, Kohmura M, Ariyoshi Y, Kim SH (1999) J Mol Biol 285:311
194. Pentelute BL, Gates ZP, Tereshko V, Dashnau JL, Vanderkooi JM, Kossiakoff AA, Kent SBH (2008) J Am Chem Soc 130:9695
195. Mandal K, Pentelute BL, Tereshko V, Kossiakoff AA, Kent SBH (2009) J Am Chem Soc 131:1362
196. Mandal K, Pentelute BL, Tereshko V, Thammavongsa V, Schneewind O, Kossiakoff AA, Kent SBH (2009) Protein Sci 18:1146
197. Pentelute BL, Mandal K, Gates ZP, Sawaya MR, Yeates TO, Kent SBH (2010) Chem Commun 46:8174
198. Avital-Shmilovici M, Mandal K, Gates ZP, Phillips NB, Weiss MA, Kent SBH (2013) J Am Chem Soc 135:3173
199. Fruton JS (1992) A skeptical biochemist. Harvard University Press, Cambridge
200. Gutte B, Merrifield RB (1969) J Am Chem Soc 91:501

201. Jenkins SR, Nutt RF, Dewey RS, Veber DF, Holly FW, Paleveda WJ Jr, Lanza T Jr, Strachan RG, Schoenewaldt EF, Barkemeyer H, Dickinson MJ, Sondey J, Hirschmann R, Walton E (1969) J Am Chem Soc 91:505
202. Strachan RG, Paleveda WJ Jr, Nutt RF, Vitali RA, Veber DF, Dickinson MJ, Garsky V, Deak JE, Walton E, Jenkins SR, Holly FW, Hirschmann R (1969) J Am Chem Soc 91:503
203. Veber DF, Varga SL, Milkowski JD, Joshua H, Conn JB, Hirschmann R, Denkewalter RG (1969) J Am Chem Soc 91:506
204. Haber E, Anfinsen CB (1961) J Biol Chem 236:422
205. Denke-walter RG, Schwam H, Strachan RG, Beesley TE, Veber DF, Schoenwaldt EF, Barkemeyer H, Paleveda WJ Jr, Jacob TA, Hirschmann R (1966) J Am Chem Soc 88:3163
206. Dewey RS, Schoenewaldt EF, Joshua H, Paleveda WJ Jr, Schwam H, Barkemeyer H, Arison BH, Veber DF, Denkewalter RG, Hirschmann R (1968) J Am Chem Soc 90:3254
207. Anderson GW, Zimmerman JE, Callahan FM (1964) J Am Chem Soc 86:1839
208. Veber DF, Milkowski JD, Denkewalter RG, Hirschmann R (1968) Tetrahedron Lett 3057
209. Yajima H, Fujii N (1981) J Am Chem Soc 103:5867
210. Yajima H, Fujii N (1980) J Chem Soc Chem Commun 115
211. Fujii N, Yajima H (1981) J Chem Soc Perkin Trans 1:819
212. Fujii N, Yajima H (1981) J Chem Soc Perkin Trans 1:811
213. Fujii N, Yajima H (1981) J Chem Soc Perkin Trans 1:797
214. Fujii N, Yajima H (1981) J Chem Soc Perkin Trans 1:804
215. Fujii N, Yajima H (1981) J Chem Soc Perkin Trans 1:789
216. Kenner GW (1977) Proc R Soc Lond B Biol Sci 197:237
217. Pearl LH, Taylor WR (1987) Nature 328:482
218. Miller M, Jaskolski M, Rao JKM, Leis J, Wlodawer A (1989) Nature 337:576
219. Pearl LH, Taylor WR (1987) Nature 329:351
220. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RAF, Scolnick EM, Sigal IS (1988) Proc Natl Acad Sci USA 85:4686
221. Huff JR (1991) J Med Chem 34:2305
222. Debouck C, Gorniak JG, Strickler JE, Meek TD, Metcalf BW, Rosenberg M (1987) Proc Natl Acad Sci USA 84:8903
223. Schneider J, Kent SB (1988) Cell 54:363
224. Navia MA, Fitzgerald PMD, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP (1989) Nature 337:615
225. Weber IT, Miller M, Jaskolski M, Leis J, Skalka AM, Wlodawer A (1989) Science 243:928
226. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SBH (1989) Science 245:616
227. Wlodawer A, Vondrasek J (1998) Annu Rev Biophys Biomol Struct 27:249
228. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SBH, Wlodawer A (1989) Science 246:1149
229. Swain AL, Miller MM, Green J, Rich DH, Schneider J, Kent SBH, Wlodawer A (1990) Proc Natl Acad Sci USA 87:8805
230. Smith R, Brereton IM, Chai RY, Kent SBH (1996) Nat Struct Biol 3:946
231. Baca M, Kent SBH (1993) Proc Natl Acad Sci USA 90:11638
232. Baca M, Kent SBH (2000) Tetrahedron 56:9503
233. Johnson ECB, Malito E, Shen Y, Rich D, Tang WJ, Kent SBH (2007) J Am Chem Soc 129:11480
234. Torbeev VY, Kent SBH (2007) Angew Chem Int Ed 46:1667
235. Hackeng TM, Mounier CM, Bon C, Dawson PE, Griffin JH, Kent SB (1997) Proc Natl Acad Sci U S A 94:7845
236. Fitzgerald MC, Chernushevich I, Standing KG, Kent SBH, Whitman CP (1995) J Am Chem Soc 117:11075
237. Boerema DJ, Tereshko VA, Kent SBH (2008) Biopolymers 90:278
238. Li C, Li X, Lu W (2010) Biopolymers 94:487

239. Narhi LO, Arakawa T, Aoki KH, Elmore R, Rohde MF, Boone T, Strickland TW (1991) J Biol Chem 266:23022
240. Higuchi M, Oheda M, Kuboniwa H, Tomonoh K, Shimonaka Y, Ochi N (1992) J Biol Chem 267:7703
241. Mattio M, Ceaglio N, Oggero M, Perotti N, Amadeo I, Orozco G, Forno G, Kratje R, Etcheverrigaray M (2011) Biotechnol Prog 27:1018
242. Kent SBH (2013) Angew Chem Int Ed 52:11988
243. Wilson RM, Dong S, Wang P, Danishefsky SJ (2013) Angew Chem Int Ed 52:7646
244. Macmillan D, Bill RM, Sage KA, Fern D, Flitsch SL (2001) Chem Biol 8:133
245. Kochendoerfer GG, Chen SY, Mao F, Cressman S, Traviglia S, Shao H, Hunter CL, Low DW, Cagle EN, Carnevali M, Gueriguian V, Keogh PJ, Porter H, Stratton SM, Wiedeke MC, Wilken J, Tang J, Levy JJ, Miranda LP, Crnogorac MM, Kalbag S, Botti P, Schindler-Horvat J, Savatski L, Adamson JW, Kung A, Kent SBH, Bradburne JA (2003) Science 299:884
246. Chen S-Y, Cressman S, Mao F, Shao H, Low DW, Beilan HS, Cagle EN, Carnevali M, Gueriguian V, Keogh PJ, Porter H, Stratton SM, Wiedeke MC, Savatski L, Adamson JW, Bozzini CE, Kung A, Kent SBH, Bradburne JA, Kochendoerfer GG (2005) Chem Biol 12:371
247. Richardson JP, Macmillan D (2008) Org Biomol Chem 6:3977
248. Hirano K, MacMillan D, Tezuka K, Tsuji T, Kajihara Y (2009) Angew Chem Int Ed 48:9557
249. Liu S, Pentelute BL, Kent SBH (2012) Angew Chem Int Ed 51:993
250. Murakami M, Okamoto R, Izumi M, Kajihara Y (2012) Angew Chem Int Ed 51:3567
251. Tan Z, Shang S, Halkina T, Yuan Y, Danishefsky SJ (2009) J Am Chem Soc 131:5424
252. Kan C, Trzupek JD, Wu B, Wan Q, Chen G, Tan Z, Yuan Y, Danishefsky SJ (2009) J Am Chem Soc 131:5438
253. Wang P, Dong S, Brailsford JA, Iyer K, Townsend SD, Zhang Q, Hendrickson RC, Shieh J, Moore MAS, Danishefsky SJ (2012) Angew Chem Int Ed 51:11391
254. Wang P, Dong S, Shieh J-H, Peguero E, Hendrickson R, Moore MAS, Danishefsky SJ (2013) Science 342:1357
255. Ullmann V, Rädisch M, Boos I, Freund J, Pöhner C, Schwarzinger S, Unverzagt C (2012) Angew Chem Int Ed 51:11566
256. Litwack G (ed) (2009) Vitamins and hormones, vol 80, Insulin and IGFs. Academic, New York, pp 1–513
257. Banting FG, Best CH (1922) J Lab Clin Med 7:251
258. Morihara K, Ueno Y (1991) Biotechnol Bioeng 37:693
259. Zahn H, Gattner HG, Fischer W, Jonczyk A, Wiese A (1987) Adv Biosci 65:167
260. Inouye K, Watanabe K, Morihara K, Tochino Y, Kanaya T, Emura J, Sakakibara S (1979) J Am Chem Soc 101:751
261. Morihara K, Oka T, Tsuzuki H (1979) Nature 280:412
262. Ueno Y, Morihara K (1989) Biotechnol Bioeng 33:126
263. Sievert D, Gattner HG, Lenz V, Hoecker H (1991) Biomed Biochim Acta 50:S197
264. Cerovsky V (1991) Tetrahedron Lett 32:3421
265. Mayer JP, Zhang F, DiMarchi RD (2007) Biopolymers 88:687
266. Galloway JA, Chance RE, Su KSE (1991) Int Congr Ser Excerpta Med 944:23
267. Frank BH, Chance RE (1983) Muench Med Wochenschr 125:14
268. Wade JD, Lin F, Hossain MA, Shabanpoor F, Zhang S, Tregear GW (2009) The chemical synthesis of relaxin and related peptides a historical perspective. In: BryantGreenwood GD, Bagnell CA, Bathgate RAD (eds) Relaxin and related peptides: 5th international conference, vol 1160. Wiley-Blackwell, New York, p 11
269. Belgi A, Hossain MA, Tregear GW, Wade JD (2011) Immunol Endocr Metab Agents Med Chem 11:40
270. Ryle AP, Sanger F, Smith LF, Kitai R (1955) Biochem J 60:541
271. Kung YT, Du YC, Huang WT, Chen CC, Ke LT, Hu SC, Jiang RQ, Chu SQ, Niu CI, Hsu JZ, Chang WC, Cheng LL, Li HS, Wang Y, Loh TP, Chi AH, Li CH, Shi PT, Yieh YH, Tang KL, Hsing CY (1965) Sci Sinica 14:1710

272. Tu YT, Chang YS, Lu TH, Chou CL (1961) Sci Sin 10:84
273. Meienhofer J, Zahn H, Schnabel E, Bremer H, Zabel R, Sroka W, Okuda T, Brinkhof O, Brandenb D, Klosterm H (1963) Z Naturforsch Pt B 18:1120
274. Katsoyan P, Tometsko A, Zalut C (1966) J Am Chem Soc 88:166
275. Marglin A, Merrifield RB (1966) J Am Chem Soc 88:5051
276. Sieber P, Kamber B, Hartmann A, Johl A, Riniker B, Rittel W (1974) Helv Chim Acta 57:2617
277. Akaji K, Fujino K, Tatsumi T, Kiso Y (1993) J Am Chem Soc 115:11384
278. Hossain MA, Belgi A, Lin F, Zhang S, Shabanpoor F, Chan L, Belyea C, Truong H-T, Blair AR, Andrikopoulos S, Tregear GW, Wade JD (2009) Bioconjugate Chem 20:1390
279. Sohma Y, Kent SBH (2009) J Am Chem Soc 131:16313
280. Sohma Y, Hua Q-X, Whittaker J, Weiss MA, Kent SBH (2010) Angew Chem Int Ed 49:5489
281. Tofteng AP, Jensen KJ, Schaeffer L, Hoeg-Jensen T (2008) ChemBioChem 9:2989
282. Zaykov AN, Mayer JP, Gelfanov VM, DiMarchi RD (2014) ACS Chem Biol 9:683
283. Liu F, Luo EY, Flora DB, Mayer JP (2013) Org Lett 15:960
284. Han J, Cheng S, DiMarchi RD (2009) Pept Sci 371
285. Liu F, Luo EY, Flora DB, Mezo AR (2014) Angew Chem Int Ed 53:3983
286. Karas JA, Scanlon DB, Forbes BE, Vetter I, Lewis RJ, Gardiner J, Separovic F, Wade JD, Hossain MA (2014) Chem Eur J 20:9549
287. Liu F, Liu Q, Mezo AR (2014) Org Lett 16:3126
288. Guo ZY, Wang S, Tang YH, Feng YM (2004) BBA-Proteins Proteom 1699:103
289. Guo ZY, Zhang Z, Jia XY, Tang YH, Feng YM (2005) Acta Bioch Bioph Sin 37:673
290. Zhang ZJ, Wu L, Qiao ZS, Qiao MQ, Feng YM, Guo ZY (2008) Protein J 27:192

Top Curr Chem (2015) 362: 229–266
DOI: 10.1007/128_2014_608

Published online: 25 March 2015

Applications of Chemical Ligation in Peptide Synthesis via Acyl Transfer

Siva S. Panda, Rachel A. Jones, C. Dennis Hall, and Alan R. Katritzky✠

Abstract The utility of native chemical ligation (NCL) in the solution or solid phase synthesis of peptides, cyclic peptides, glycopeptides, and neoglycoconjugates is reviewed. In addition, the mechanistic details of inter- or intra-molecular NCLs are discussed from experimental and computational points of view.

Keywords Acyl migration • Acyl transfer • Chemical ligation • Computational studies • Cyclic peptides • Glycopeptides • Neoglycoconjugates • SPPS

Contents

✠ Katritzky was deceased at the time of publication.

S.S. Panda (✉), R.A. Jones, and C. Dennis Hall
Department of Chemistry, Center for Heterocyclic Compounds, University of Florida, Gainesville, FL 32611-7200, USA
e-mail: sspanda12@gmail.com

1 Introduction to Native Chemical Ligation

Native chemical ligation (NCL) has shown great utility and is a proven method for the preparation of peptides, cyclic peptides, proteins, and glycoproteins via synthetic or semi-synthetic pathways [1–5]. The concept of NCL dates back to the pioneering work of Wieland et al. [6], but it was not until 1994 that this method gained widespread attention when Kent et al. reported its application in the synthesis of interleukin-8, a cytokine responsible for the proliferation of B cells during immune response [7].

Approaches to ligation of two peptide segments include prior thiol capture [8], NCL [1], conformationally assisted ligation [9], and Staudinger ligation [10], of which NCL is the most widely used method. NCL involves the chemoselective coupling of two protein or peptide fragments, one containing a C-terminal thioester and the other, typically, an N-terminal Cys residue (Fig. 1). The two components combine via an intramolecular *S*- to *N*-acyl shift to produce, irreversibly, the ligated amide bond (a native peptide bond) at the point of ligation. The fact that this reaction occurs in aqueous solution and in the absence of protecting groups has placed this powerful technology at the forefront of protein synthesis. The driving force for NCL is the formation of the thermodynamically stable amide link, at around neutral pH [3].

The rate of the ligation reaction depends on the C-terminal amino acid. Dawson et al. reported that all 20 amino acids could be used, but amino acids such as Pro, Val, and Ile gave reduced reaction rates (Fig. 2) [10]. Additives such as urea and guanidinium chloride can be added to the buffer solution to prevent the aggregation of amino acids or peptides in the solution phase synthesis [11].

The requirement of a Cys residue (Xaa-Cys, where Xaa=any amino acid) at the ligation site can be problematic because not all proteins contain a Cys residue and

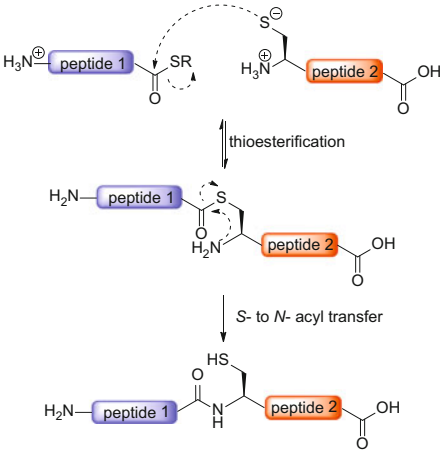


Fig. 1 Intermolecular chemical ligation (NCL)

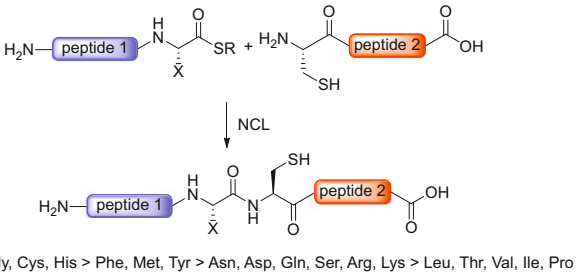


Fig. 2 NCL with all amino acids

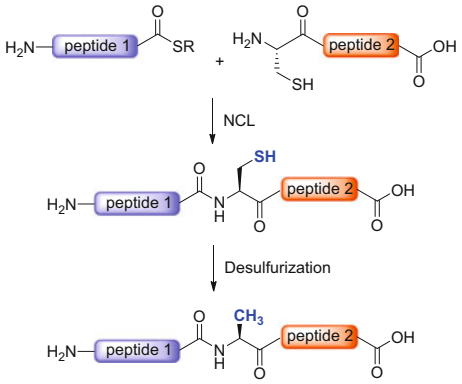


Fig. 3 NCL and subsequent desulfurization

the Cys residues may be present at locations in the protein which are not appropriate for NCL [12]. This has resulted in efforts to carry out ligations at Xaa-Xaa sites in which a thiol group is attached to a side chain of the N-terminal amino acid. Subsequent desulfurization then affords the desired peptide (Fig. 3) [13]. Desulfurization of thiol-modified amino acids can be achieved using Raney nickel or Pd/Al_2O_3, or under free-radical-based conditions [12, 14]. To date, the ligation-desulfurization technique has been achieved at Ala [12], Phe [15], Val [16], Lys [17, 18], Thr [19], Leu [20], Pro [21], Gln [22], and, most recently, Arg residues [23]. The design of thiol-containing amino acids was recently reviewed by He et al. [24]. However, where cysteine residues are present in the peptide sequence at positions other than the ligation junction, protection of the side chain is required to prevent unselective desulfurization. In this instance the acetamidomethyl group is typically used [25], and Pentelute and Kent were the first to show its application in NCL [26].

Dawson et al. demonstrated that the addition of thiophenol or benzyl mercaptan can increase the rate of ligation [27] and Johnson and Kent studied the effect that various thiols may have on NCL [28]. It was found that aryl thiols exchange with

peptide alkyl thioesters to form peptide aryl thioesters, which then act as efficient leaving groups, thus facilitating the ligation. Mercaptophenylacetic acid (MPPA) was found to be a more effective thiol additive than those previously used.

2 Solid Phase vs Solution Phase Chemical Ligation

The most successful method of fragment condensation for the synthesis of polypeptides and proteins in solution phase is NCL, reported by Dawson for the first time in 1994 [7]. This was a significant contribution because NCL overcomes one of the main limitations of solid phase peptide synthesis (SPPS), namely the production of long peptide sequences (>50 amino acid residues) [1, 3, 29]. NCL may be used in both solution and solid phases; solution phase NCL has been used for the synthesis of small peptides and cyclic peptides [30–32] whereas SPPS is more widely applied in polypeptide and protein synthesis.

SPPS has proved more useful than solution phase techniques in terms of ligation reactions for long peptides or polypeptides [33], because both N-terminal Cys-containing peptides and C-terminal thioester-containing peptides can be efficiently prepared on solid supports. Zhang et al. used 3-mercaptopropionyl MBHA

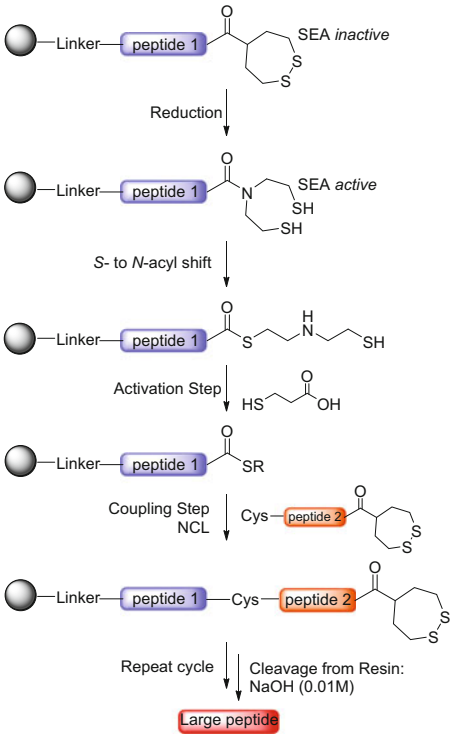


Fig. 4 *N*- to *C*-assembly of peptides

resin for the preparation of thioesters using Boc chemistry [34]. However, harsh acidic conditions (HF/anisole, 9:1) were required to cleave the peptide thioester from the solid support. Clippingdale et al. reported the synthesis of peptide thioesters via Fmoc SPPS in order to avoid the use of HF [35]. However, piperidine, widely used for Fmoc-deprotection, may cause hydrolysis of peptide-α-thioesters, and thus methods to overcome this difficulty have been the subject of a number of studies [36].

Raibaut et al. demonstrated the advantages of SPPS for preparing peptide sequences for use in NCL. An efficient solid-phase synthesis of large polypeptides was achieved by iterative ligations of bis(2-sulfanylethyl)amido (SEA) peptide segments [37]. Sequential NCL by N- to C-elongation cycles between the supported peptide thioester (blocked with SEA) and a free C-terminal thiol group-SEA activated N-terminus peptide (Fig. 4) allowed the synthesis of peptide thioesters containing 60 amino acids and the assembly of five peptide segments to give a 15-kDa polypeptide [38].

3 Intramolecular Chemical Ligation (Acyl Migration)

In general, NCL means intermolecular ligation but the phrase has also been applied to intramolecular ligation, often called *acyl migration*, which occurs when an acyl group migrates from X→N (X=S, O, N) within an isopeptide (Fig. 5). Recently, Panda et al. investigated chemical ligation from isopeptides in the solution phase via different cyclic transition states [39].

3.1 S- to N-Acyl Migration

Isopeptide ligation is an alternative method for the synthesis of cysteine peptides via an intramolecular chemical ligation by an entropically favored mechanism. *S*- to *N*-Acyl migration via various cyclic transition states was investigated by carrying out the ligation with mono-isopeptides under microwave irradiation (50°C, 50 W,

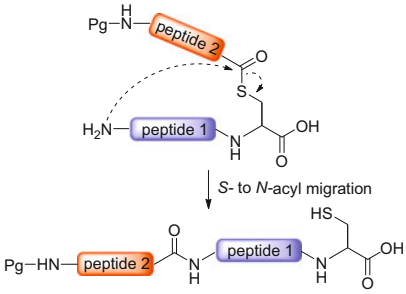
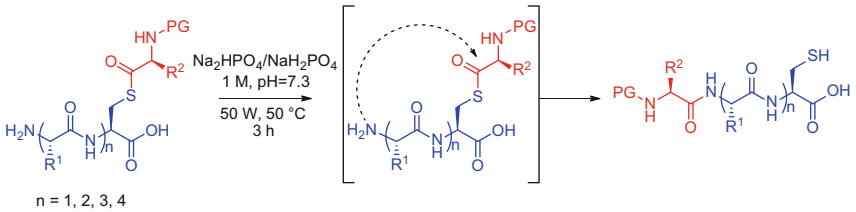
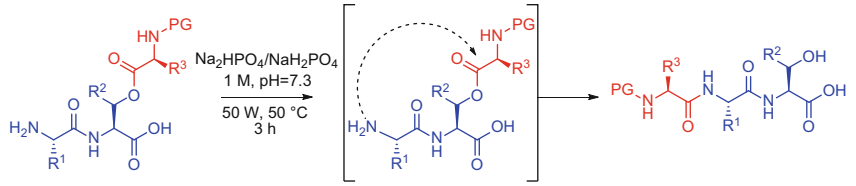


Fig. 5 Intramolecular chemical ligation



Scheme 1 Intramolecular ligation studies via *S*- to *N*-acyl migration



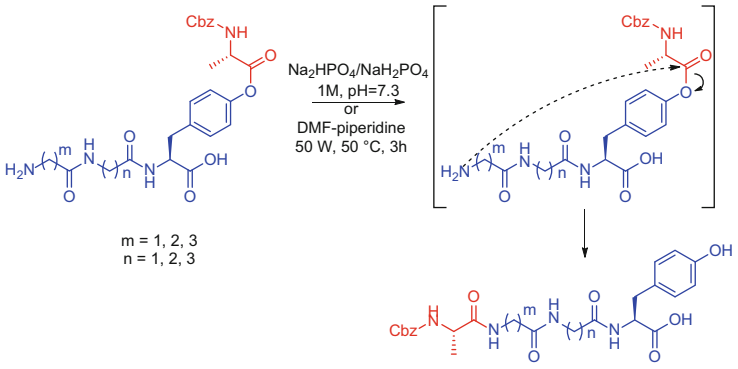
Scheme 2 Ligation studies via *O*- to *N*-acyl migration

1–3 h) using 1 M NaH_2PO_4/Na_2HPO_4 phosphate buffer to maintain pH 7.3 (Scheme 1). The feasibility of intramolecular acyl migrations via 5- to 19-membered cyclic transition states was demonstrated and the yields of long-range *S*- to *N*-acyl transfers were found to depend on the size of the macrocyclic transition state (TS). Thus the relative rates, based on yields, depended on the ring size of the TS in the order, 5 >10 >11 >14, 16, 17 >12 >13, 15, 19 >18 >>>9 >8 [40–44].

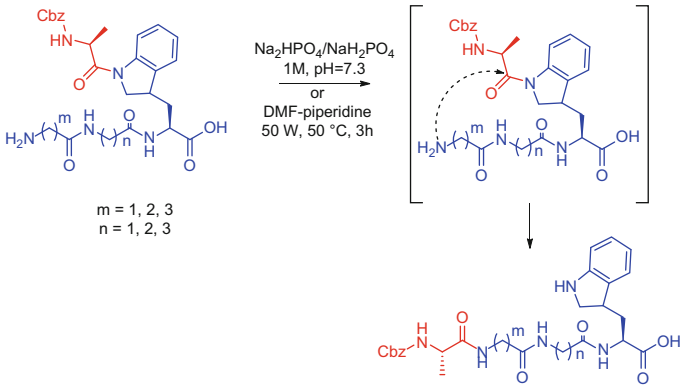
3.2 O- to N-Acyl Migrations

Chemical ligation of serine isopeptides via *O*- to *N*-acyl transfer with 8- and 11-membered TSs occurs without the use of an auxiliary group (Scheme 2) [45]. This is in contrast to cysteine isopeptides, in which the 8-membered TS was disfavored even under basic conditions. Intramolecular acyl transfer of Thr isopeptide through 5- and 9-membered TSs was favored over 8- and 11-membered TSs (Scheme 2) [46].

Chemical ligation studies of Tyr isopeptides under microwave irradiation (50°C, 50 W, 3 h) using 1 M phosphate buffer and a DMF-piperidine medium showed that intramolecular *O*- to *N*-acyl transfer occurs via 11- to 13-membered TSs under basic conditions and with 14- to 18-membered TSs in aqueous media (Scheme 3) [47].



Scheme 3 Ligation studies via *O*- to *N*-acyl migration



Scheme 4 Ligation studies via *N*- to *N*-acyl migration

3.3 N- to N-Acyl Migrations

The intramolecular chemical ligation of tryptophan isopeptides via *N*- to *N*-acyl migration occurs through 7- to 18-membered cyclic TSs forming the native peptides in basic, non-aqueous media rather than aqueous buffered conditions (Scheme 4) [48].

4 Applications of Native Chemical Ligation

4.1 Synthesis of Cyclic Peptides

In 1944 Gause and Brazhnikova discovered Gramicidin S, a cyclic peptide [49], and used it in the treatment of septic gunshot wounds during the Second World War.

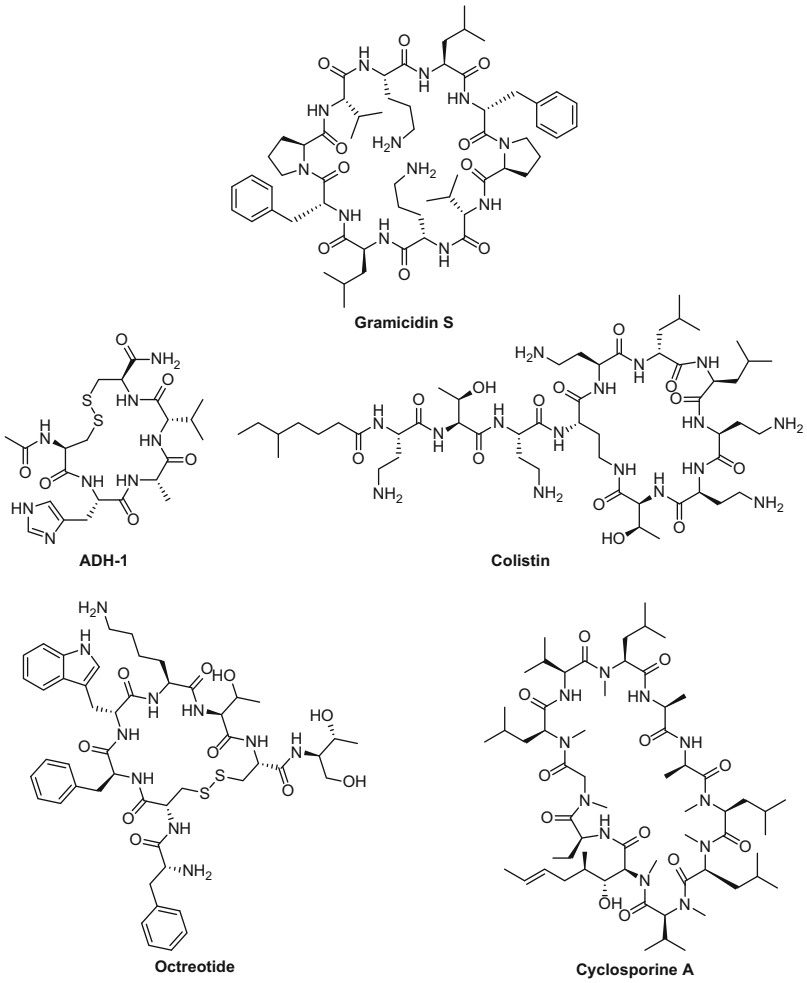


Fig. 6 Selected cyclic peptides

Since then, both natural and unnatural cyclic peptides have become important synthetic targets because of their potential applications as antibiotics and other therapeutic agents [50–54]. Selected examples include anticancer agents (ADH-1), antibiotics (colistin), growth hormone inhibitors (octreotide), and immunosuppressant agents (cyclosporine A) (Fig. 6) [55].

The constrained conformation of cyclic peptides often results in increased *exo*- and *endo*peptidase resistance, enhanced binding affinity, and in certain cases, increased cell penetration compared to their linear counterparts. Numerous strategies, both in solution and solid-phase, have been reported for the synthesis of cyclic peptides [56–58]; NCL, the reaction of a C-terminal peptide thioester with an

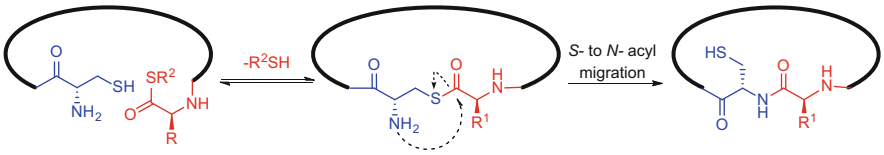


Fig. 7 Native chemical ligation applied to the head-to-tail cyclization of peptides

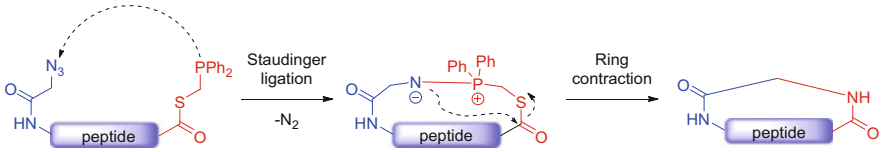


Fig. 8 The head-to-tail macrocyclization of peptides through a traceless Staudinger ligation strategy

N-terminal cysteine peptide, is now an established method for production of peptides with a cyclized backbone [59–61].

Dawson demonstrated ligation strategies on 1,2-aminothiols which depend on a capture/rearrangement mechanism to link two peptide fragments under mild conditions [7]. The cyclization process involves a reaction between a weakly activated C-terminal thioester and an unprotected N-terminal cysteine residue. The thermodynamic stability of an amide bond over a thioester is again the driving force behind this reaction, made possible through a proximity-driven *S*- to *N*-acyl migration. Zhang and Tam used the above methodology to synthesize cyclic peptides in a head-to-tail fashion (Fig. 7) [62].

Hackenberger and Kleineweischede reported a traceless Staudinger ligation for the head-to-tail macrocyclization of peptides without a deprotection step (Fig. 8) [63]. In this strategy, a phosphine tethered to a thioester at the C-terminus of a peptide reacts intramolecularly with an azide at the N-terminus to form the cyclic peptide.

Recently, cLac (cyclic peptide-mimicking lactadherin) was synthesized using NCL and studied for phosphatidylserine (PS) recognition [64, 65]. The linear precursors for the synthesis of cLac derivatives were prepared following standard protocols for automated Fmoc-peptide synthesis and the cyclic peptides (cLac variants) were obtained in the presence of 4-mercaptophenylacetic acid and isolated via HPLC in yields of 60–70% (Fig. 9). All the synthesized cLac variants were labeled with the thiol-reactive fluorescein-5-maleimide in DMF containing 2% *N*-methylmorpholine (NMM) by taking advantage of the free cysteine side chain [65].

The phosphatidylserine (PS) recognition study suggests that the cLac peptide effectively mimics the PS binding mechanism of lactadherin, in which multiple polar residues are conformationally preorganized by the protein or the cyclic peptide to balance various noncovalent forces (desolvation, hydrogen bonding, and salt bridges) for specific PS recognition [65].

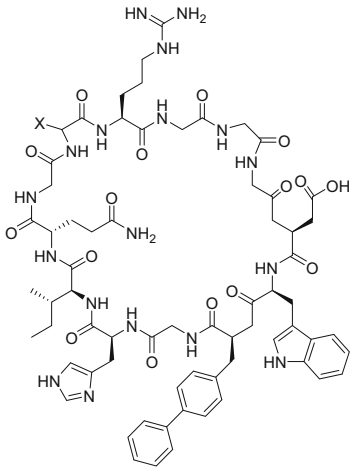
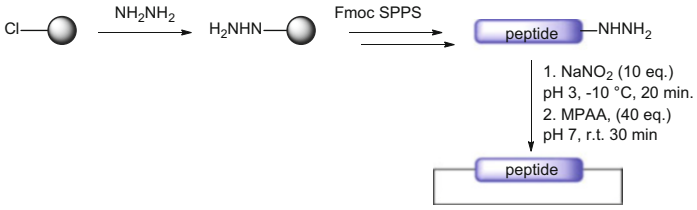


Fig. 9 cLac and a cLac variant



Scheme 5 Synthesis of cyclic peptides on a solid support

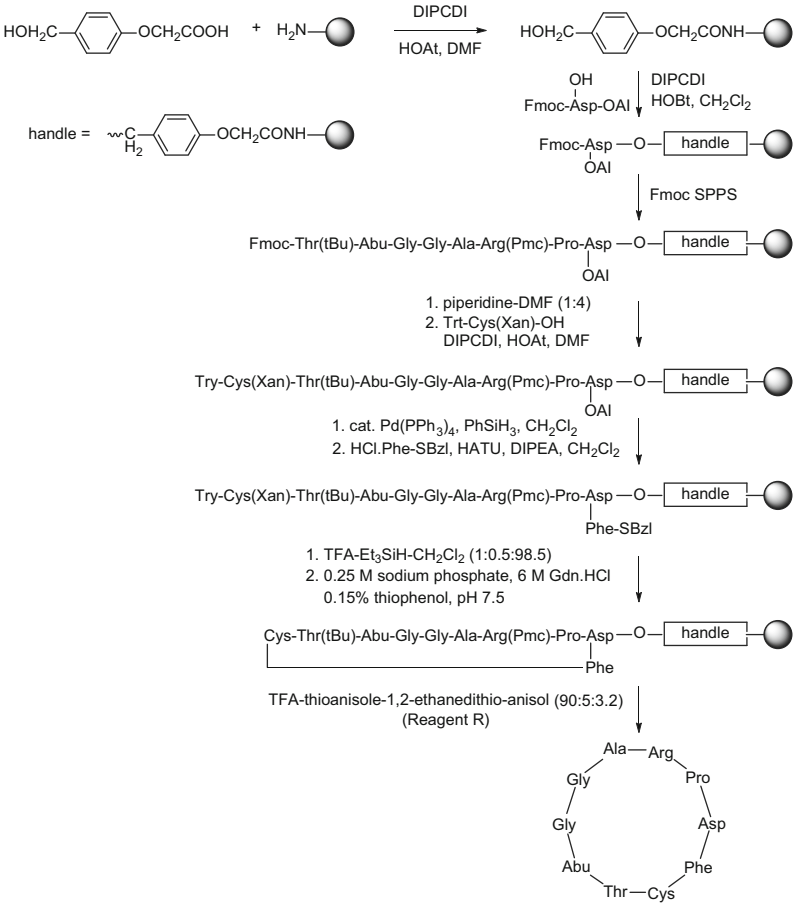
Liu et al. reported a modified version of NCL to synthesize the native peptide from a C-terminal peptide hydrazide and an N-terminal Cys under $NaNO_2$-mediated activation [66, 67]. This method (hydrazide ligation) was also applied in the synthesis of cyclic peptides. One of its important advantages is that peptide hydrazides can be easily prepared through routine Fmoc SPPS. The linear hydrazide peptides were synthesized from hydrazine-Trt(2-Cl) by following standard Fmoc SPPS. The peptide hydrazides cyclize in two steps, in a one-pot fashion, in the presence of $NaNO_2$ and the thio-additive MPAA. Aqueous phosphate buffer containing 6.0 M guanidinium chloride was used as the solvent system. A mixture of an organic solvent with aqueous phosphate buffer also works as medium for this transformation (Scheme 5). A number of cyclic peptides were prepared by following this methodology in 18–65% yields (Table 1) [68].

SPPS is often used for the preparation of linear precursors required for NCL. Barany and Tulla-Puche reported NCL on resin to avoid tedious purification steps, usually necessary after each step. The linear precursor was prepared using 1-hydroxy-7-aza-benzotriazole (HOAt) or 1-hydroxybenzotriazole (HOBt) and *N*,

Table 1 Cyclic peptides prepared by hydrazine ligation

Entry	Cyclic peptide	Yield of linear precursor (%)	Yield of cyclic peptide (%)
1	Cys-Asn-Pro-Ile-Trp-Gly-Ile-Val	76	59
2	Cys-Asn-Pro-Ile-Trp-Gly-Ile-Gly	65	55
3	Cys-Asn-Pro-Ile-Trp-Gly-Ile-Ala	72	63
4	Cys-Asn-Pro-Ile-Trp-Gly-Ile-Leu	70	60
5	Cys-Asn-Pro-Ile-Trp-Gly-Ile-Lys	72	57
6	Cys-Asn-Pro-Ile-Trp-Gly-Ile-Cys	55	42
7	Cys-Asn-Pro-Ile-Trp-Gly-Ile-His	75	65
8	Cys-Asn-Pro-Ile-Trp-Gly-Ile-Thr	60	45
9	Cys-Asn-Pro-Ile-Trp-Gly-Ile-Glu	59	39
10	Cys-Asn-Pro-Ile-Trp-Gly-Ile-Trp	46	35
11	Cys-Tyr-Leu-Ala-Gly	72	48
12	Cys-Val-Tyr-Gly-Gly-Ala	77	54
13	Cys-Trp-Leu-Val-Asp-Cys(Acm)-Pro	58	45
14	Cys-Thr-Arg-Gly-Phe-Arg-Cys-Val	70	52
15	Cys-Leu-Leu-Gly-Ile-Cys-Arg-Cys-Ile	68	48
16	Cys-Thr-Lys-Ser-Ile-Pro-Pro-Gln-Cys-Tyr-Ser	38	30
17	Cys-Thr-Lys-Ser-Ile-Pro-Pro-Ile-Lys-Phe-(D)Pro-Pro-Arg	53	45
18	CTKSIPPICTFSIPPI	45	34
19	CRCVGRRCLCRCFGRTCI	52	41
20	CLFKPFLSCGSGSDGGVCPKILKKCRRDSDCPGACICRGNGY	23	18

N′-diisopropylcarbodiimide (DIPCDI) as coupling agents. Key aspects of on-resin NCL include Fmoc/*t*Bu chemistry, side-chain anchoring, allyl protection of the penultimate residue to allow introduction of the C-terminal thioester later in the synthetic sequence, a new derivative, Trt-Cys(Xan)-OH, which facilitates selective



Scheme 6 Synthesis of cyclic peptides via on-resin NCL

and mild removal of both protecting groups. The synthesis of *cyclo*(Cys-Thr-Abu-Gly-Gly-Ala-Arg-Pro-Asp-Phe) using on-resin NCL is illustrated in Scheme 6 [69].

Fukuzumi et al. introduced a strategy in which linear peptides bearing side chains with unprotected functional groups could be cyclized under high dilution conditions [70]. This strategy, known as α-ketoacid-hydroxylamine amide-ligation, is achieved by the Fmoc-based SPPS of linear peptides bearing a C-terminal sulfur ylide linker, which acts as a 'masked' α-ketoacid (Fig. 10). The introduction of the *N*-hydroxylamine can also be carried out on a solid support. Subsequent global deprotection and cleavage from the solid support affords the linear peptide which, following sulfur ylide oxidation, undergoes the desired cyclization. The general applicability of this strategy was demonstrated by the preparation of five natural product cyclic peptides (Table 2).

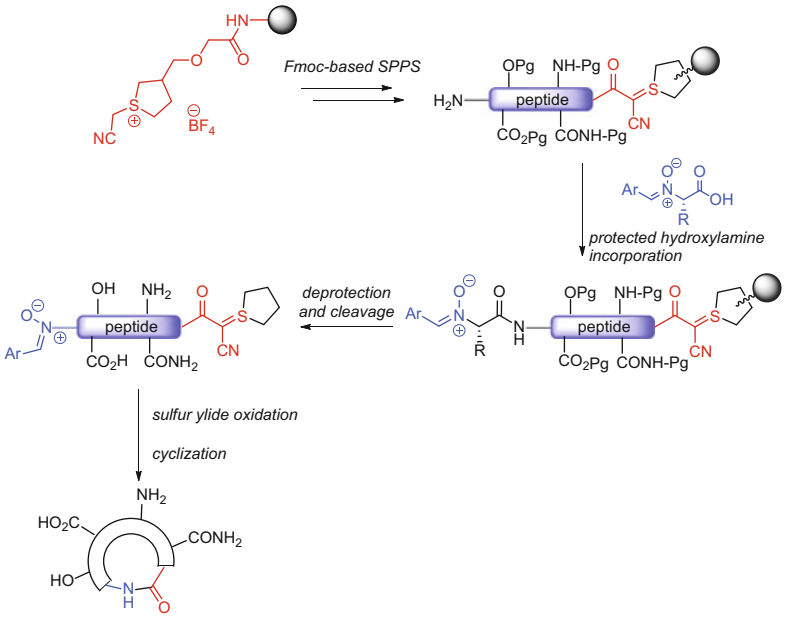


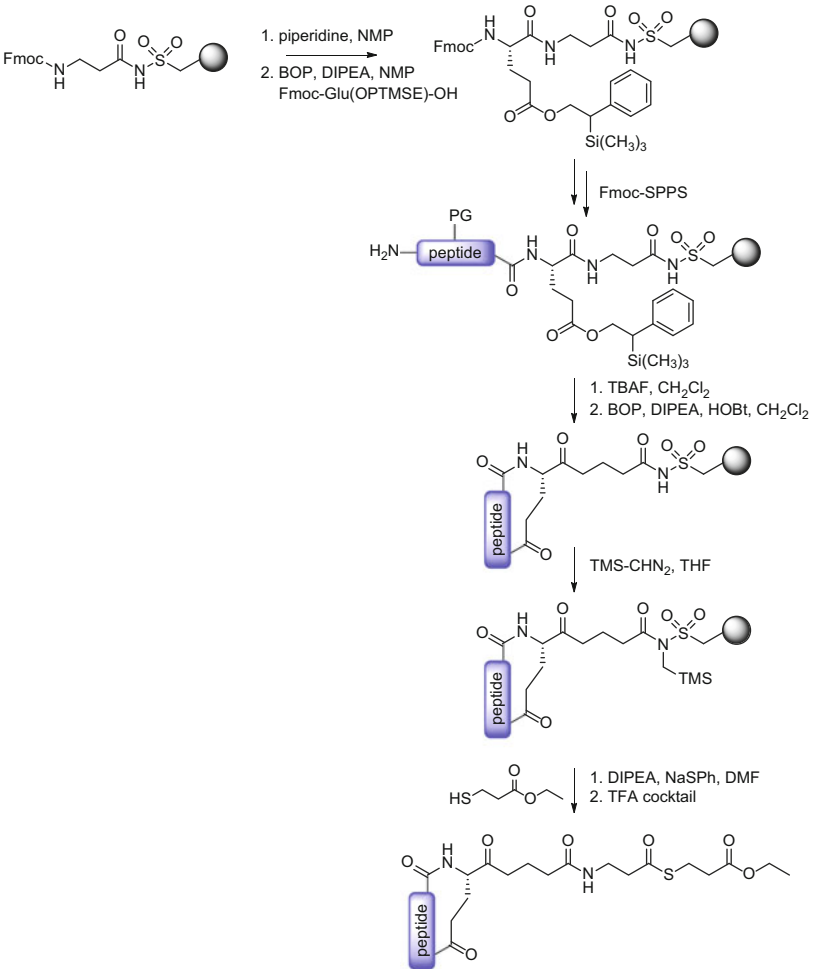
Fig. 10 General strategy for the preparation of unprotected linear peptides with *N*-terminal hydroxylamine and C-terminal α-ketoacids for direct, reagent-less cyclizations

Table 2 Cyclic peptides prepared by α-ketoacid-hydroxylamine amide-ligation

Entry	Cyclic peptide	HPLC yield (%)	Isolated yield (%)
1	**Gramicidin S:** *cyclo*(-^{D}FPVOL^{D}FPVOL-)	13	13
2	**Tyrocidine A:** *cyclo*(-^{D}FPF^{D}FNQYVOL-)	36	22
3	**Hymenamide B:** *cyclo*(-FPPNFVE)	15	8
4	**Semi gramicidin S:** *cyclo*(-^{D}FPVOL-)	17	15
5	**Stylostatin A:** *cyclo*(-AISN^{D}FPL-)	13	7

Cyclization conditions: 0.1 M $(COOH)_2$, 0.001 M DMF: H_2O (50:1), 40°C for 48 h

van de Langemheen et al. synthesized cyclic peptides containing a thioester handle using a 'sulfo-click' linker. In this approach the 'sulfo-click' linker used for the synthesis of the linear precursor was prepared by Fmoc SPPS (Scheme 7) [71]. Three different cyclic peptide sequences were synthesized, corresponding to the loops present in HIV protein gp120 interacting with CD4 as found in the X-ray structure of the gp120-CD4 complex. On the basis of this structure, the ^{365}SGGDPEIVT^{373}, ^{424}INMWQEVGKA^{433}, and ^{454}LTRDGGN^{460} peptide sequences were selected for the preparation of cyclic peptide thioesters (Fig. 11) [72]. HIV-gp120 plays a crucial role in the first steps of HIV-infection through its attachment to the CD4 receptor [73]. Preventing attachment of gp120 to cells and/or using gp120 as a starting point to develop a vaccine may offer alternative approaches to avoid the further spread of HIV [74, 75].



Scheme 7 Synthesis of linear precursors for NCL by Fmoc SPPS

Chen et al. developed a strategy for preparing cyclic peptides via in situ generation of a thioester resulting from disulfide reduction; subsequent NCL results in the desired peptide (Fig. 12) [76]. This strategy was used to synthesize linear glycopeptides, which, after thioester formation, resulted in cyclization to form a model glycopeptide in 73% yield (Scheme 8).

The synthesis of branched peptides using masked side-chain thioester derivatives of Asp and Glu which are compatible with Fmoc-SPPS is an important goal. Boll et al. synthesized cyclic and branched chain peptides using bis (2-sulfanylethyl)amido (SEA) side-chain derivatives of Asp and Glu via Fmoc SPPS [77]. The tail-to-side-chain cyclization via an in situ reduction of both acyclic and cyclic disulfides with tris(2-carboxyethyl)phosphine (TCEP) triggered the SEA intramolecular ligation. Glu derivatives cyclized more readily than the Asp analogues and without formation of side products (Scheme 9).

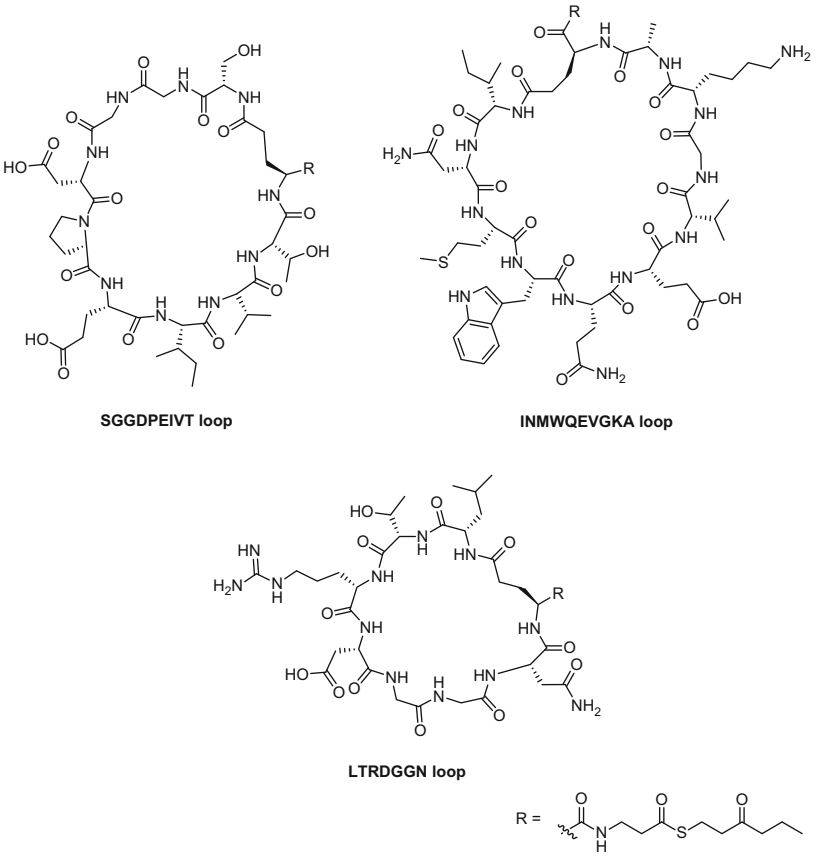


Fig. 11 ^{365}SGGDPEIVT^{373}, ^{424}INMWQEVGKA^{433}, and ^{454}LTRDGGN^{460} loops

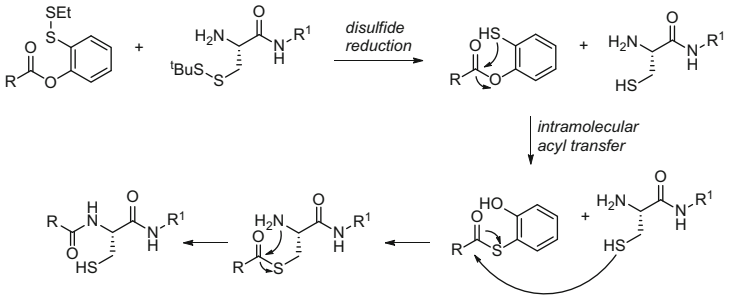
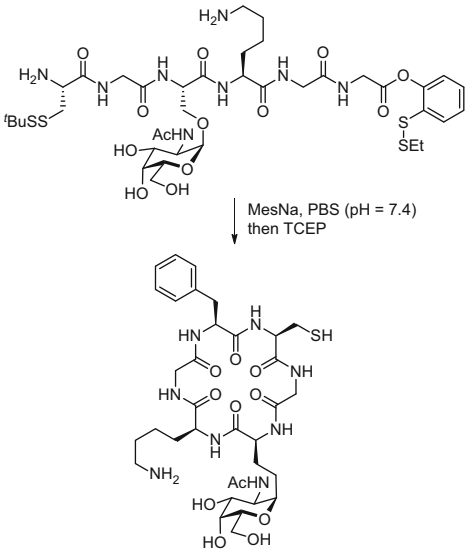
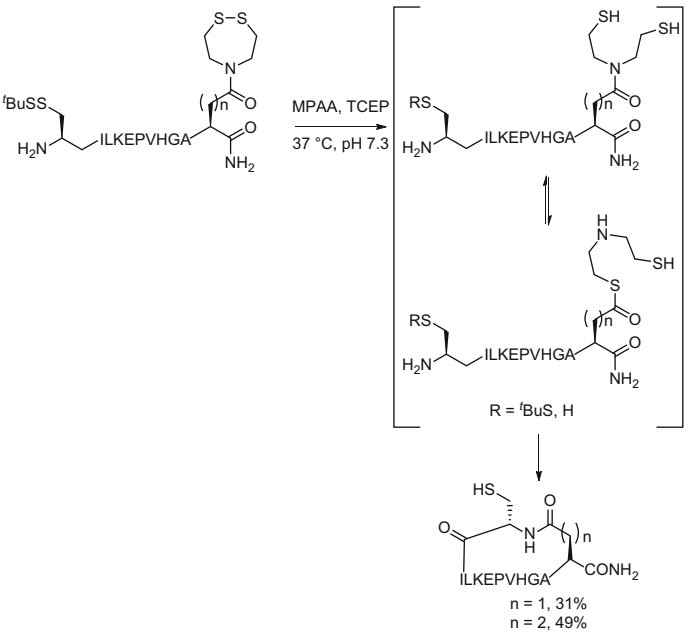


Fig. 12 Proposed mechanism for the in situ generation of thioesters and subsequent ligation



Scheme 8 Synthesis of cyclic glycopeptides following the in situ generation of thioesters



Scheme 9 Tail to side-chain cyclization using bis(2-sulfanylethyl)amido (SEA) ligation

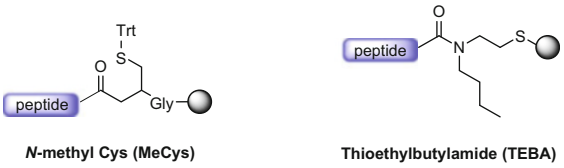
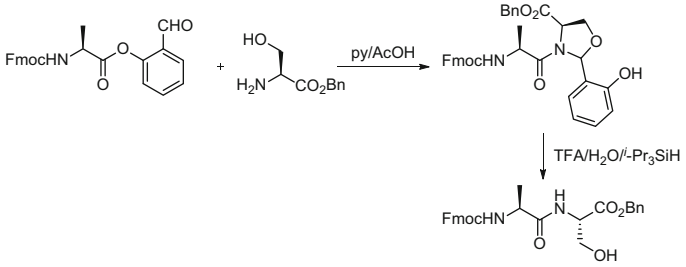


Fig. 13 Thioester surrogates



Scheme 10 NCL via an *O*- to *N*-acyl transfer

There are several drawbacks associated with the use of thioester surrogates such as SEA, in SPPS, especially when the thioester surrogate is attached to the resin via an MeCys linker [78]. These limitations include the need for multiple coupling reactions to attach the C-terminal amino acid to the resin and the possibility that the protected MeCys linker can, during peptide elongation, undergo β-elimination and then piperidine conjugate addition to form MeAla(Pip) as a side product [79]. In order to overcome these limitations, Taichi et al. developed the thioethylbutylamido (TEBA) group as an alternative to SEA-derived thioester surrogates (Fig. 13). The utility of the TEBA-thioester surrogate was demonstrated by the synthesis of cysteine-rich cyclic peptides [78, 80].

Li et al. developed an efficient method for the synthesis of peptides and proteins. In this method, an *O*-salicylaldehyde ester at the C-terminus reacts with N-terminal serine or threonine to realize peptide ligations via an *O*- to *N*-acyl transfer (Scheme 10) [81].

The utility of this ligation approach has been demonstrated through the convergent syntheses of therapeutic peptides (ovine-corticoliberin and Forteo) and the human erythrocyte acylphosphatase protein (∼11 kDa) [82]. The requisite peptide salicylaldehyde ester precursor is prepared in an epimerization-free manner via Fmoc–solid-phase peptide synthesis. This approach was also used for the synthesis of Daptomycin (Fig. 14), a lipodepsipeptide isolated from *Streptomyces roseoporus* which was obtained from a soil sample from Mount Ararat (Turkey) [83].

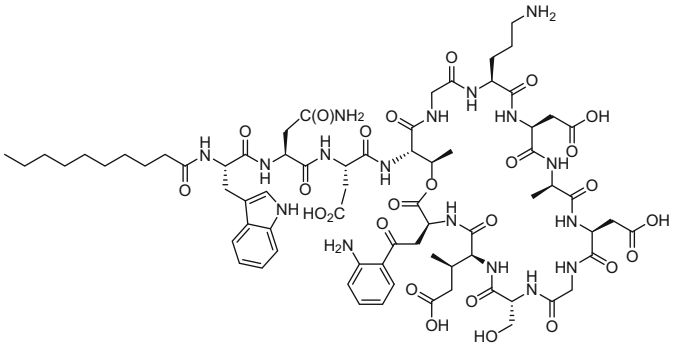


Fig. 14 Daptomycin

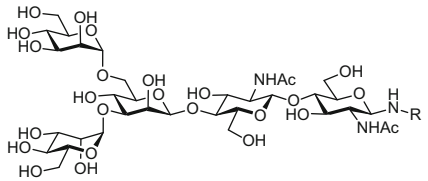


Fig. 15 Pentasaccharide core fragment present on N-linked glycopeptides where R=Asp

4.2 Synthesis of Glycopeptides

Post-translational modifications of proteins and peptides can have profound effects on the overall biological and chemical properties of these biopolymers. In particular, glycosylation of proteins and peptides results in highly complex structures known as glycoproteins [84, 85]. The presence of glycans can ensure the stability and activity of glycoproteins, with biological functions such as cell adhesion, differentiation, and growth [86–88]. Carbohydrates are linked to proteins via either a *N*-glycosidic bond at an asparagine residue or an *O*-glycosidic bond at a serine or threonine residue [89]. N-Linked glycoproteins are biosynthesized in a process commencing in the endoplasmic reticulum (ER) in which a 14-mer oligosaccharide is transferred to the amide nitrogen of an asparagine residue, in an Asn-Xxx-Ser/Thr sequon, where Xxx is any amino acid apart from proline, from a dolichol phosphate. Glycosidases then truncate the 14-mer to a pentasaccharide core fragment which is further modified to afford the final *N*-glycoprotein, which is in turn either transported to the cell surface or secreted (Fig. 15) [90].

The most prevalent class of O-linked glycoproteins are the mucins, in which a galactosamine (GalNAc) monosaccharide is linked to the peptide backbone via an α-glycosidic bond from GalNAc to a Ser or Thr residue [91]. The GalNAc monosaccharide, often known as the T_N antigen (Fig. 14), is then modified at the C-3 or C-6 positions with GalNAc, galactose, or glucosamine via α- or β-glycosidic bonds,

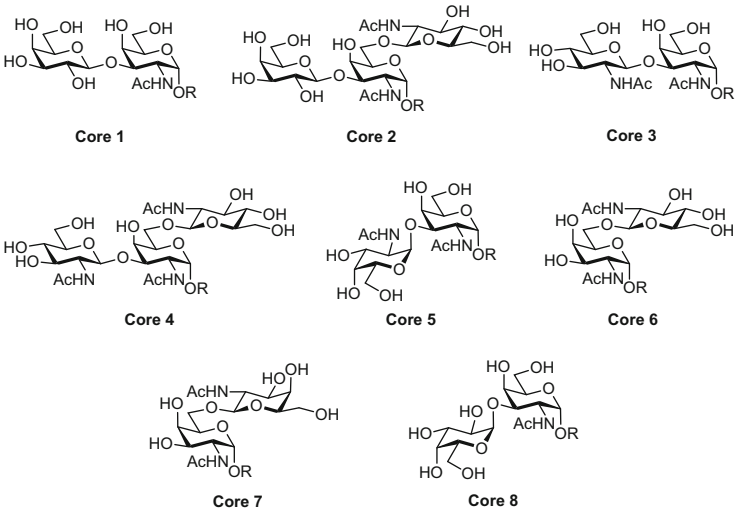


Fig. 16 Core structures of mucin O-linked glycans where R is a Ser or Thr residue [91]

resulting in a series of di- and trisaccharides which are known as the core structures (Fig. 16).

O-Linked glycoproteins are biosynthesized in a continuous process that occurs in the ER and the Golgi apparatus. This is not a template driven process, but rather is subject to numerous sequential and competitive enzymatic pathways, and the *O*-glycans vary according to the cell lineage, tissue location, and developmental stage of the cell [91, 92]. Furthermore, as the pattern of *O*-glycosylation alters in response to mucosal infection and inflammation, O-linked glycoproteins have important implications in a wide range of diseases, including cystic fibrosis, Crohn's disease, and cancers [93–99]. Aberrant glycosylation can result in truncated core structures, lacking backbone motifs, which leave the T_N and T antigens or their sialylated versions exposed, providing disease markers which might have important implications in the design and synthesis of anti-cancer vaccines (Fig. 17) [100, 101].

To determine the exact role of glycoproteins and glycopeptides in biological systems, it is necessary to access pure samples of these compounds. Furthermore, glycoproteins and glycopeptides are important compounds in drug discovery because of their potential therapeutic benefits [102, 103]. However, because of the dynamic and heterogeneous nature of biological systems, progress in glycobiology has been hindered, and it is not currently possible to isolate single glycoforms from natural sources [104]. Thus chemical synthesis of glycoproteins and glycopeptides is an important area of research [105, 106].

Chemical ligation is a very attractive method for the synthesis of glycoproteins [107], and NCL, traceless Staudinger ligation [108], and sugar-assisted ligation (SAL) (Fig. 18) [109–111] have all been used for this purpose. SAL has been used for the synthesis of a large number of glycopeptides [112–114], but difficulties were encountered in the presence of larger glycans [115].

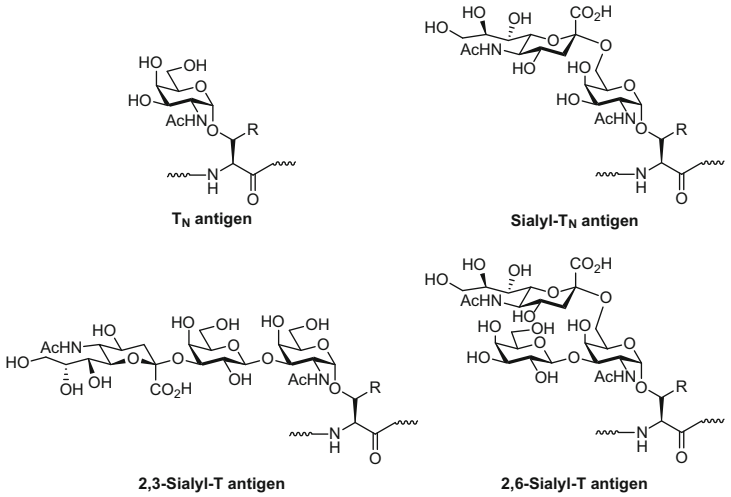


Fig. 17 Common tumor-associated carbohydrate antigens (TACAs) where R=H (Ser) or CH_3 (Thr)

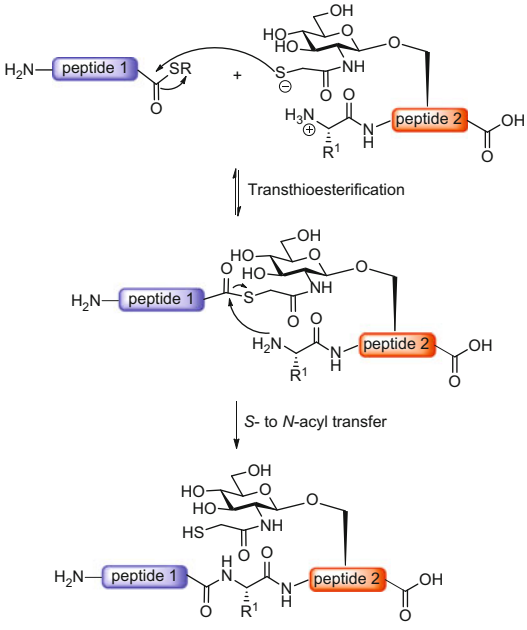


Fig. 18 Sugar-assisted ligation (SAL)

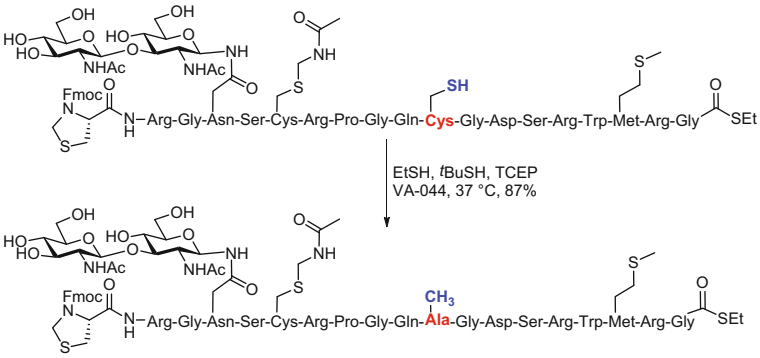
The synthesis of glycoproteins has been reviewed recently [84, 85, 106, 116, 117], and hence the focus here is on the use of NCL for the synthesis of glycopeptides and glycopeptide mimetics, the so-called 'neoglycopeptides' [118, 119]. Although NCL has been applied to the synthesis of glycoproteins and glycopeptides [120], the synthesis of glycopeptides is inherently challenging [121].

Boc-SPPS, traditionally used to prepare peptides required for NCL, is not used for the synthesis of glycopeptides because it is incompatible with acid labile glycosidic linkages such as sialyl- or fucosyl-glycosidic bonds. However, Murakami et al. were able to prepare a sialylglycopeptide by substituting the harsh acidic conditions (HF) typically used to remove Boc protecting groups for an acidic deprotection cocktail of TFA/TfOH/DMS/*m*-cresol (5:1:3:1) [122]. Alternatively, Fmoc-SPPS, used in glycopeptide synthesis, is not compatible with the synthesis of peptide-α-thioesters, as the thioester may be cleaved by the piperidine used to remove Fmoc protecting groups. Because the Fmoc strategy is more widely used than Boc strategies, especially in automated SPPS, methods to overcome difficulties associated with Fmoc-protected peptide-α-thioesters have received a great deal of attention [36].

The requirement of a Cys residue at the ligation junction for successful NCL can be problematic when attempting to synthesize glycopeptides because of the low natural abundance of Cys residues in nature (only 1.7%). This results in a lack of Cys residues in the target glycoprotein or the need to synthesize glycopeptide sequences of 30–50 amino acid residues. It is worth mentioning that the synthesis of N-linked glycopeptide chains is more challenging than the corresponding amino acid sequence without an attached glycan, which can lead to difficulties in glycopeptide synthesis [106, 123, 124]. Consequently, a large number of techniques have been developed, including Ser or Thr ligation strategies [125] and the addition of thiol groups to amino acids [24] which, subject to successful desulfurization, might afford the target glycopeptide. Desulfurization of peptides was originally achieved with either Raney nickel or Pd/Al_2O_3 [12], but these conditions are not always compatible with desulfurization of glycopeptides [16, 126].

The in situ generation of thioesters has been applied to the synthesis of cyclic glycopeptides bearing a single carbohydrate moiety (Fig. 12) [76]. Wan and Danishefsky applied this strategy to the synthesis of glycopeptides containing strategically placed Cys residues to facilitate NCL. Following NCL, a radical-based, metal-free desulfurization process using TCEP and *t*BuSH enabled the Cys residue to be converted into the desired Ala (Scheme 11) [14]. The NCL free-radical-based desulfurization was subsequently reported to be applicable to thiol-modified Val residues [127] and thiol-protected Thr residues [19].

Efforts have been made to improve free-radical desulfurization strategies and in this respect one-pot strategies are particularly attractive because they avoid time-consuming purification after each reaction. Moyal et al. [128] recently reported a one-pot ligation-desulfurization protocol and subsequently Thompson et al. developed a one-pot ligation-desulfurization protocol, using a novel thiol additive, 2,2,2-trifluoroethanethiol, for the synthesis of short peptide fragments [129]. Neither of these strategies has yet been applied to glycopeptide synthesis.



Scheme 11 Free-radical desulfurization of a glycopeptide

To achieve a more widely applicable method for glycopeptide synthesis, Okamoto and Kajihara developed a strategy for the conversion of Cys to a Ser residue after NCL. Thus, *S*-methylation of Cys, followed by an intramolecular rearrangement activated by CNBr, results in an *O*- to *N*-acyl shift, which affords an *O*-ester peptide intermediate. A second *O*- to *N*-acyl shift generates the desired glycopeptide (Fig. 19) [130]. This strategy was used to access a fragment (residues 79–98) of EPO, N-linked glycopeptide and a repeat sequence of MUC-1, an O-linked glycoprotein. Furthermore, this procedure can be carried out in the presence of methionine residues, when protected as the sulfoxide. Subsequently, Okamoto et al. reported that the substitution of Cys to Ser could be accomplished in the presence of acid-labile sialyl-glycosidic linkages such as those found on sialyl-T_N antigens [131].

NCL has been achieved at unprotected Ser and Thr residues via an *O*- to *N*-acyl transfer [81]. Hojo et al. employed a slightly different strategy in which a mercaptomethyl group was used to protect the Ser and Thr residues. In this instance the ligation step occurs through an *S*- to *N*-acyl transfer via a seven-membered ring and this approach was used to access O-linked glycoprotein contulakin-G (Fig. 20) [132].

N- to *S*-acyl shifts have been used to access peptide thioesters (Fig. 21) [133, 134]. Macmillan et al. used this strategy to prepared thioesters, which were subsequently used in NCL to access model glycopeptides based on the erythropoietin (EPO) amino acid sequence [135, 136].

Hsieh et al. prepared *S*-glycosylated peptides using NCL (Scheme 12) [137]. In this instance, the glycopeptides fragments were prepared using Fmoc SPPS and the subsequent NCL proceeded in yields of over 80% for the three examples. Disulfide-bridge formation at key cysteine residues achieved the natural product, bacteriocin glycopeptide Sublancin 168, and two derivatives bearing alternative sugars.

Glycoconjugate mimetics, known as neoglycoconjugates, which bear an 'unnatural' linkage between the carbohydrate and aglycon moieties, have been widely explored, as this allows access to novel glycopeptides which might have enhanced activity relative to the naturally occurring glycopeptide [119]. The triazole unit is

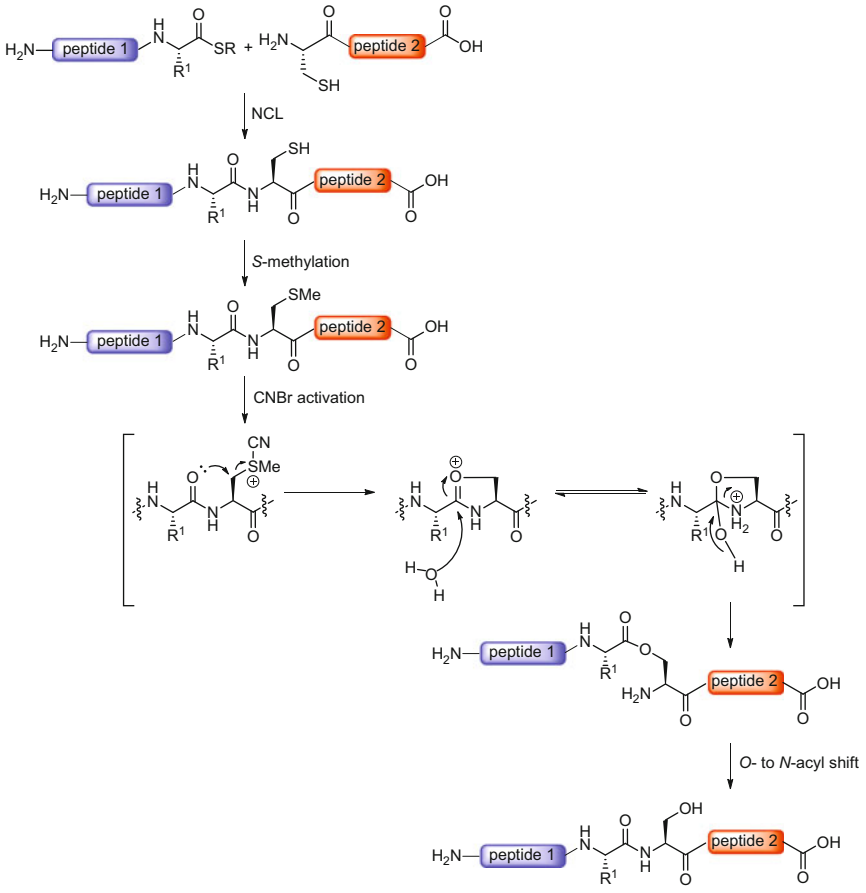


Fig. 19 Conversion of Cys to Ser following NCL

one of the most widely studied 'unnatural' glycosidic linkages and, in this context, Macmillan and Blanc synthesized a neoglycopeptide, using Boc-SPPS, in which the peptide sequence was based upon human erythropoietin. Cleavage of the neoglycopeptide from the solid support afforded an 11-mer peptide sequence which was able to undergo NCL with a suitable thioester to afford the target neoglycoconjugate (Scheme 13), neatly demonstrating that NCL is compatible with triazole moieties [138].

Lee et al. subsequently synthesized similar neoglycopeptides using a one-pot strategy in which two propargyl peptides were coupled via NCL; subsequent 1,3-dipolar cycloaddition afforded the glycosylated peptide (Fig. 22) [139].

Finally, it has been shown that NCL has applications in the context of synthesizing RNA mimics [140], and selective desulfurization again plays an important role in producing target molecules [141].

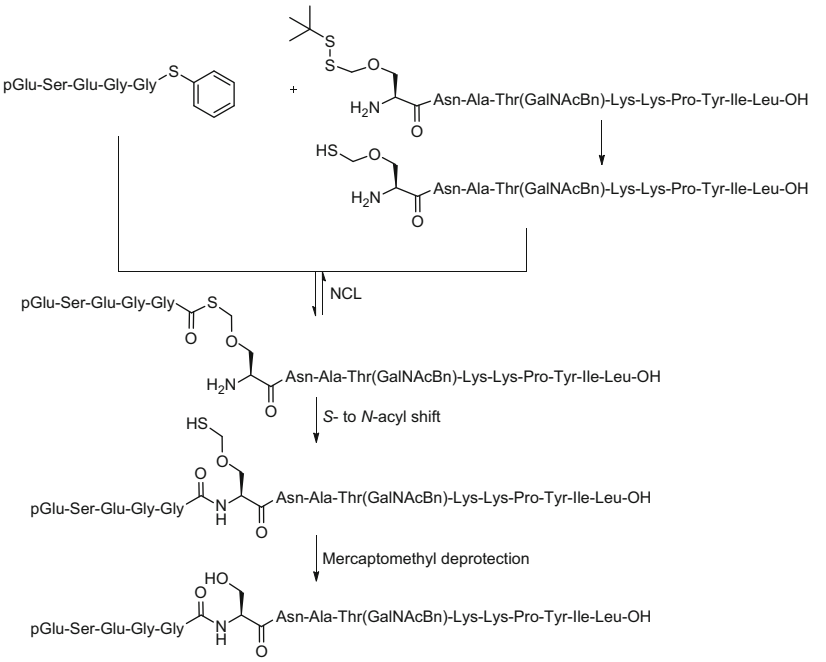


Fig. 20 Synthesis of contulakin-G via an *S*- to *N*-acyl transfer

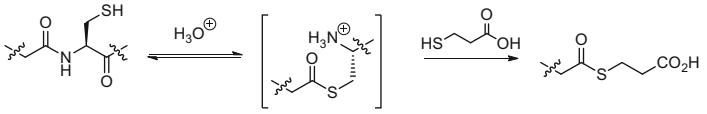
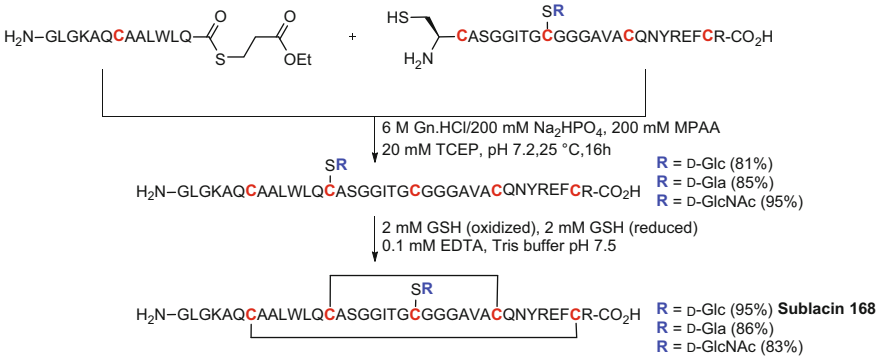
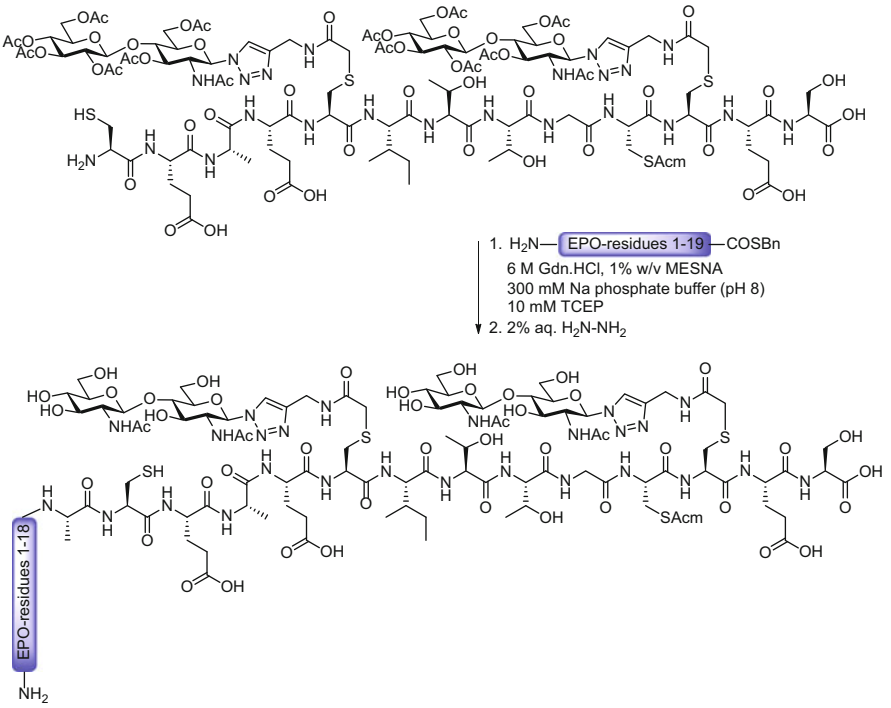


Fig. 21 Synthesis of peptide thioesters via an *N*- to *S*-acyl shift



Scheme 12 Synthesis of *S*-linked glycopeptides via NCL



Scheme 13 NCL of neoglycopeptides containing a triazole motif

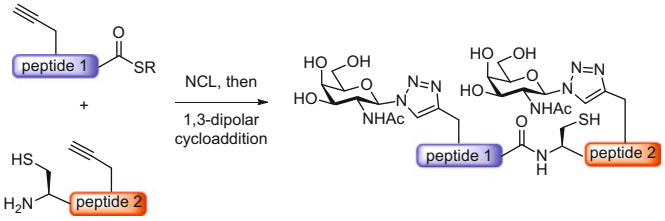
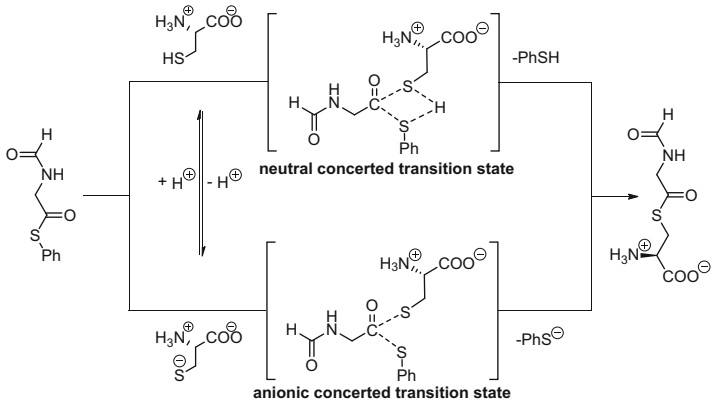


Fig. 22 One-pot NCL followed by 1,3-dipolar cycloaddition

5 Computational Rationalization of Chemical Ligation

Chemical ligation has been widely reported and discussed, and numerous attempts have been made to increase reactivity and improve yields to make the peptide ligation method more effective. In particular, mechanistic studies of the reaction have played an important role in rationalizing the rate of ligation.

Wang et al. investigated the mechanism of NCL and calculated energy barriers for various C-terminal amino acids in order to compare their reactivity [142]. In the

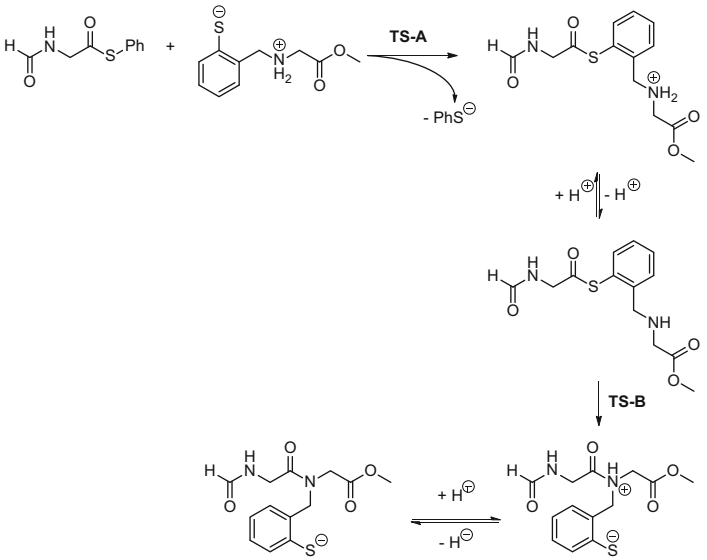


Scheme 14 Hypothetical routes for thioesterification

reaction between a peptide thioalkyl ester and a Cys-peptide in the presence of an aryl thiol catalyst, it is alleged that both the thiol–thioester exchange step and the *trans*-thioesterification step proceed by a concerted S_N2 displacement, whereas the intramolecular rearrangement occurs by an addition-elimination mechanism (Scheme 14).

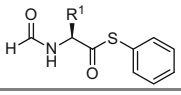
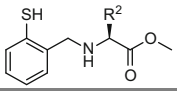
The energy barrier for the thiol–thioester exchange step depends on steric hindrance associated with the side-chain of C-terminal amino acids, whereas that of the acyl-transfer step depends on steric hindrance caused by the side-chain of the N-terminal amino acid. In auxiliary-mediated peptide ligation, between a peptide thiophenyl ester and an *N*-2-mercaptobenzyl peptide, the thiol–thioester exchange step and intramolecular acyl-transfer step proceed by a concerted S_N2-type displacement mechanism (Scheme 15). For N-terminal Gly, the thioester exchange is rate limiting, whereas the acyl transfer is the rate-limiting step for N-terminal non-Gly amino acids (Table 3) [142]. When the difference in $\Delta G^{\ddagger}$ between the two transition states is, for example, 3 kcal mol^{-1}, then the reaction with the lower activation energy is 150 times faster (Table 3).

When compared to intermolecular NCL, there have been far more computational studies into the intramolecular chemical ligation of isopeptides. Monbaliu et al. discussed the computational approach and developed the first systematic theoretical background for an *n-exo*-trig intramolecular *S*-to *N*-acyl transfer [143]. Cyclic TSs in the ring size range $n = 5$–10 were controlled by enthalpic factors and a classical range of $\Delta G^{\ddagger}$ values were found. The calculations also emphasized that the substituents at R^1 and R^2 (Fig. 23b) had little impact on the nature of the TS. The preorganization of the structures and, in particular, the emergence of stabilizing hydrogen bonds in intramolecular TSs appeared as major factors governing the variations of $\Delta G^{\ddagger}$ with ring size. In isomeric structures, the presence of an internal non-natural amino acid (β-Ala or GABA) directly after



Scheme 15 Mechanism of auxiliary-mediated peptide ligation

Table 3 Energy barriers of auxiliary-mediated peptide ligations for various conjugation sites

Entry	C-terminal thioester 	N-terminal amino acid 	$\Delta G^{\ddagger}$ (TS-A)	$\Delta G^{\ddagger}$ (TS-B)
1	Gly (R^1=H)	Gly (R^2=H)	29.2	29.4
2	Gly (R^1=H)	Ala (R^2=CH_3)	29.0	32.6
3	Gly (R^1=H)	Val (R^2=*i*Pr)	30.4	33.1
4	Ala (R^1=CH_3)	Gly (R^2=H)	31.8	29.5
5	Ala (R^1=CH_3)	Ala (R^2=CH_3)	≈32	32.3
6	Ala (R^1=CH_3)	Val (R^2=*i*Pr)	≈33	35.8
7	Val (R^1=*i*Pr)	Gly (R^2=H)	34.9	29.4
8	Val (R^1=*i*Pr)	Ala (R^2=CH_3)	≈35	36.3
9	Val (R^1=*i*Pr)	Val (R^2=*i*Pr)	≈36	38.4

the Cys residue favored hydrogen bonds (5.7–8.4 kcal mol^{-1}), which stabilize the TS. The competition between the intra- and intermolecular acyl transfers is driven by parameters which govern the approach of the reactive termini, for example, the ring strain (Fig. 23) [143].

The feasibility and rate of intramolecular ligation depends on various factors such as preorganization energy, hydrogen bonding, and the distance between the

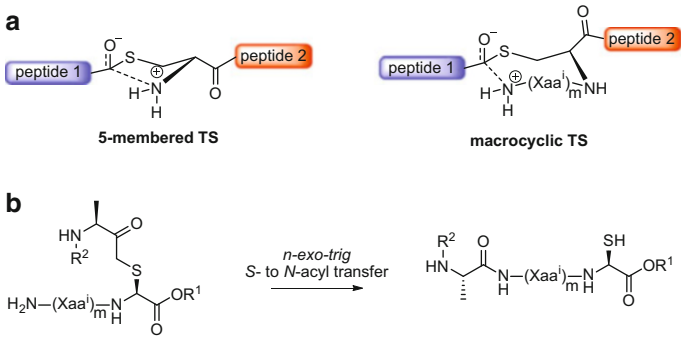
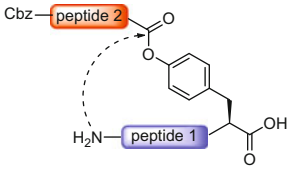


Fig. 23 (**a**) 5-*Exo*-trig *S*- to *N*-acyl transfer as commonly encountered in classical NCL and (*S*)-isopeptide rearrangements (*left*) and *n*-*exo*-trig *S*- to *N*-acyl transfer in internal Cys NCL and extended isopeptides rearrangements (*right*). (**b**) Isomerization of (*S*)-acyl isopeptides to native peptide analogues if $m = 0$, $n = 5$

reaction sites. Oliferenko and Katritzky rationalized the curious behavior of the intramolecular ligation of Cys-containing isopeptides by computational chemistry. Preorganization is an important factor in chemical ligation because it occurs through a cyclic TS [144]. The more easily the starting material achieves an appropriate cyclic conformation, the higher the probability of intramolecular reaction. Hydrogen bonding and NH–π interactions play a major role in the stabilization of a preorganized conformer [44]. Intramolecular ligation via eight-membered cyclic TSs of Cys-containing isopeptides was disfavored, whereas, surprisingly, Ser-containing isopeptides showed an efficient *O*- to *N*-acyl migration via eight-membered cyclic TSs. The *O*- to *N*-acyl migration of Tyr-containing isopeptides showed intramolecular chemical ligation in the presence of a base with 12- to 14-membered cyclic TSs and in a buffered medium with 15- to 19-membered cyclic TSs. The reactivity of Tyr isopeptides is described in terms of preorganization energy, hydrogen bonding and bond distance (Table 4) [145].

Biswas et al. reported, the design of a predictive model using statistical techniques to correlate the relative abundance (ligated product percentage) with quantitative structural activity/property relationship (QSAR/QSPR) [146]. The genetic algorithm linear regression method was performed using QSARINS software [147], which establishes a correlation between the dependent variable (property/response relative to the abundance of ligated peptide) and independent variables (molecular descriptors or factors) (Table 5). It was found that the percentage of ligated peptides increases with both a shorter spatial *b*(N-C) distance and a higher Balaban index (Fig. 24).

Table 4 Determination of parameters governing *O*- to *N*-acyl migrations for Tyr isopeptides

Cyclic TS size	Ligated product (%)[a]	*b*(N-C) (Å)	Hydrogen bonding		$\Delta E_{preorg} = E_{min} - E_{preorg}$ (kcal/mol)
			Distance (Å)	Angle (°)	
					
12	95.3	4.851	N(22)-H(50)----N(1)		−0.021
			1.944	163.0	
13	85.2	4.765	N(23)-H(53)----N(1)		−0.535
			2.010	162.8	
14	97.0	3.779	N(22)-H(52)----N(34)		0.00
			1.948	167.2	
15	94.6	3.398	N(1)-H(37)----O(28)		−3.301
			1.885	174.8	
16	87.0	3.246	N(24)-H(56)----O(29)		−1.757
			1.913	171.3	
17	96.8	3.656	N(3)-H(41)----O(30)		−2.723
			2.088	152.5	
18	99.9	3.289	N(21)-H(55)----N(39)		−2.360
			1.908	163.7	
19	98.1	2.990	N(21)-H(56)----N(39)		−1.500
			1.970	149.1	

[a]Relative abundance calculated from HPLC-MS

Table 5 Statistical model for the relative abundance of ligated peptides

ID	Factors[a]	Coeff[b]	s[c]	t[d]
$R^2 = 0.93$, $F = 106.1$, $s = 9.92$, $n = 2$, $N = 18$, R^2cvloo = 0.90, R^2adj = 0.88				
I	Intercept	0.1439		
g_1	Distance – *b*(N-C)	−36.143	−0.626	42.50
g_2	Balaban index	134.304	0.406	37.50

[a]Represents the molecular descriptor based on best multilinear regression (BMLR) stepwise model
[b]Coefficients of respective factors including the intercept
[c]Standard error
[d]Student's *t*-test (*t*-criterion)

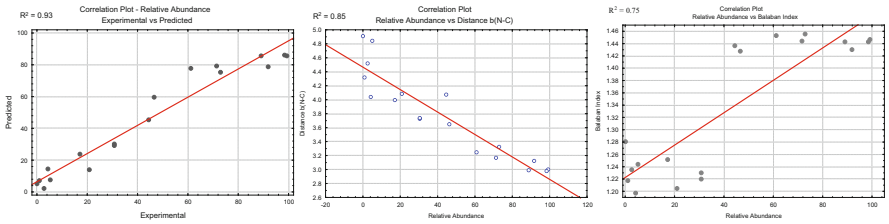


Fig. 24 (**a**) Correlation plot for relative abundance model of ligated product. (**b**) Correlation between relative abundance and distance *b*(N-C). (**c**) Correlation between relative abundance and Balaban index

6 Conclusions

The increasing demand for peptides and chemically modified peptides has produced a significant expansion in alternative synthetic routes for peptide and protein synthesis. Native chemical ligation has proved to be a very useful tool for the synthesis of proteins, peptides, cyclic-peptides, glycopeptides, and neoglycoconjugates. In spite of ligation-desulfurization techniques, the general requirement of an N-terminal cysteine and a C-terminal thioester remains a limitation. However, considerable efforts have been made to generalize NCL. Both solution and solid phase peptide synthesis have been used for NCL, and NCL is the more acceptable method for preparing peptides because of the high-yielding isolation of long-chain peptides and cyclic peptides with a lower likelihood of thioester racemization. Because thioesterification can occur at mildly acidic or neutral pH, NCL promises to have a major impact on the synthesis of peptides and peptide mimetics and is particularly suitable for the synthesis of natural products, cyclic peptides, and glycopeptides. The rate of acyl transfer in chemical ligation reactions depends on various factors and these have been analyzed by computational studies. This theoretical rationalization should assist in the preparation of new cyclic peptides and glycopeptides via chemical ligation.

Acknowledgements We thank the University of Florida and The Kenan Foundation for financial support.

References

1. Dawson PE, Kent SBH (2000) Synthesis of native proteins by chemical ligation. Annu Rev Biochem 69:923–960
2. Hackenberger CPR, Schwarzer D (2008) Chemoselective ligation and modification strategies for peptides and proteins. Angew Chem Int Ed 47:10030–10074
3. Kent SBH (2009) Total chemical synthesis of proteins. Chem Soc Rev 38:338–351
4. Boltje TJ, Buskas T, Boons GJ (2009) Opportunities and challenges in synthetic oligosaccharide and glycoconjugate research. Nat Chem 1:611–622

5. Wang P, Danishefsky SJ (2010) Promising general solution to the problem of ligating peptides and glycopeptides. J Am Chem Soc 132:17045–17051
6. Wieland T, Bokelmann E, Bauer L, Lang HU, Lau H (1953) Über Peptidsynthesen. 8. Bildung von S-haltigen Peptiden durch intramolekulare Wanderung von Aminoacylresten. Liebigs Ann Chem 583:129–149
7. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Synthesis of proteins by native chemical ligation. Science 266:776–779
8. Kemp DS, Buckler DR (1991) Highly enantioselective amide ligation by prior thiol capture. Tetrahedron Lett 32:3013–3016
9. Beligere GS, Dawson PE (1999) Conformationally assisted protein ligation using *C*-terminal thioester peptides. J Am Chem Soc 121:6332–6333
10. Kohn M, Breinbauer R (2004) The Staudinger ligation – a gift to chemical biology. Angew Chem Int Ed 43:3106–3116
11. Millevoi S, Chiaraluce R, Consalvi V, Giangiacomo L, Pasquo A, Linda Britton K, Stillman TJ, Rice DW, Engel PC (1998) A monomeric mutant of Clostridium symbiosum glutamate dehydrogenase: comparison with a structured monomeric intermediate obtained during refolding. Protein Sci 7:966–974
12. Yan LZ, Dawson PE (2001) Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. J Am Chem Soc 123:526–533
13. Dawson PE (2011) Native chemical ligation combined with desulfurization and deselenization: a general strategy for chemical protein synthesis. Isr J Chem 51:862–867
14. Wan Q, Danishefsky SJ (2007) Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. Angew Chem Int Ed 46:9248–9252
15. Crich D, Banerjee A (2007) Native chemical ligation at phenylalanine. J Am Chem Soc 129:10064–10065
16. Haase C, Rohde H, Seitz O (2008) Native chemical ligation at valine. Angew Chem Int Ed 47:6807–6810
17. Kumar KSA, Haj-Yahya M, Olschewski D, Lashuel HA, Brik A (2009) Highly efficient and chemoselective peptide ubiquitylation. Angew Chem Int Ed 48:8090–8094
18. Yang R, Pasunooti KK, Li F, Liu X-W, Liu C-F (2009) Dual native chemical ligation at lysine. J Am Chem Soc 131:13592–13593
19. Chen J, Wang P, Zhu JL, Wan Q, Danishefsky SJ (2010) A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides. Tetrahedron 66:2277–2283
20. Harpaz Z, Siman P, Kumar KSA, Brik A (2010) Protein synthesis assisted by native chemical ligation at leucine. ChemBioChem 11:1232–1235
21. Shang SY, Tan ZP, Dong SW, Danishefsky SJ (2011) An advance in proline ligation. J Am Chem Soc 133:10784–10786
22. Siman P, Karthikeyan SV, Brik A (2012) Native chemical ligation at glutamine. Org Lett 14:1520–1523
23. Malins LR, Cergol KM, Payne RJ (2013) Peptide ligation-desulfurization chemistry at arginine. ChemBioChem 14:559–563
24. He Q-Q, Fang G-M, Liu L (2013) Design of thiol-containing amino acids for native chemical ligation at non-Cys sites. Chin Chem Lett 24:265–269
25. Veber D, Milkowski J, Varga S, Denkewalter R, Hirschmann R (1972) Acetamidomethyl. A novel thiol protecting group for cysteine. J Am Chem Soc 94:5456–5461
26. Pentelute BL, Kent SBH (2007) Selective desulfurization of cysteine in the presence of Cys (Acm) in polypeptides obtained by native chemical ligation. Org Lett 9:687–690
27. Dawson PE, Churchill MJ, Ghadiri MR, Kent SBH (1997) Modulation of reactivity in native chemical ligation through the use of thiol additives. J Am Chem Soc 119:4325–4329
28. Johnson ECB, Kent SBH (2006) Insights into the mechanism and catalysis of the native chemical ligation reaction. J Am Chem Soc 128:6640–6646

29. Kent SBH (1988) Chemical synthesis of peptides and proteins. Annu Rev Biochem 57:957–989
30. Offer J (2010) Native chemical ligation with N^{α} acyl transfer auxiliaries. Biopolymers (Pept Sci) 94:530–541
31. Payne RJ, Wong C-H (2010) Advances in chemical ligation strategies for the synthesis of glycopeptides and glycoproteins. Chem Commun 46:21–43
32. Macmillan D, De Cecco M, Reynolds NL, Santos LFA, Barran PE, Dorin JR (2011) Synthesis of cyclic peptides through an intramolecular amide bond rearrangement. ChemBioChem 12:2133–2136
33. Palomo JM (2014) Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides. RSC Adv 4:32658–32672
34. Zhang LS, Tam JP (1999) Lactone and lactam library synthesis by silver ion-assisted orthogonal cyclization of unprotected peptides. J Am Chem Soc 121:3311–3320
35. Clippingdale AB, Barrow CJ, Wade JD (2000) Peptide thioester preparation by Fmoc solid phase peptide synthesis for use in native chemical ligation. J Pept Sci 6:225–234
36. Mende F, Seitz O (2011) 9-Fluorenylmethoxycarbonyl-based solid-phase synthesis of peptide α-thioesters. Angew Chem Int Ed 50:1232–1240
37. Raibaut L, Adihou H, Desmet R, Delmas AF, Aucagne V, Melnyk O (2013) Highly efficient solid phase synthesis of large polypeptides by iterative ligations of bis(2-sulfanylethyl)amido (SEA) peptide segments. Chem Sci 4:4061–4066
38. Boll E, Drobecq H, Ollivier N, Raibaut L, Desmet R, Vicogne J, Melnyk O (2014) A novel PEG-based solid support enables the synthesis of >50 amino-acid peptide thioesters and the total synthesis of a functional SUMO-1 peptide conjugate. Chem Sci 5:2017–2022
39. Panda SS, Hall CD, Oliferenko AA, Katritzky AR (2014) Traceless chemical ligation from *S*-, *O*-, and *N*-acyl isopeptides. Acc Chem Res 47:1076–1087
40. Katritzky AR, Abo-Dya NE, Tala SR, Abdel-Samii ZK (2010) The chemical ligation of selectively *S*-acylated cysteine peptides to form native peptides via 5-, 11- and 14-membered cyclic transition states. Org Biomol Chem 8:2316–2319
41. Katritzky AR, Tala SR, Abo-Dya NE, Ibrahim TS, El-Feky SA, Gyanda K, Pandya KM (2011) Chemical ligation of *S*-scylated cysteine peptides to form native peptides via 5-, 11-, and 14-membered cyclic transition states. J Org Chem 76:85–96
42. Panda SS, El-Nachef C, Bajaj K, Al-Youbi AO, Oliferenko A, Katritzky AR (2012) Study of chemical ligation via 17-, 18- and 19-membered cyclic transition states. Chem Biol Drug Des 80:821–827
43. Ha K, Chahar M, Monbaliu JCM, Todadze E, Hansen FK, Oliferenko AA, Ocampo CE, Leino D, Lillicotch A, Stevens CV, Katritzky AR (2012) Long-range intramolecular $S \rightarrow N$ acyl migration: a study of the formation of native peptide analogues via 13-, 15-, and 16-membered cyclic transition states. J Org Chem 77:2637–2648
44. Bol'shakov O, Kovacs J, Chahar M, Ha K, Khelashvili L, Katritzky AR (2012) *S*- to *N*-acyl transfer in *S*-acylcysteine isopeptides via 9-, 10-, 12-, and 13-membered cyclic transition states. J Pept Sci 18:704–709
45. El Khatib M, Elagawany M, Jabeen F, Todadze E, Bol'shakov O, Oliferenko A, Khelashvili L, El-Feky SA, Asiri A, Katritzky AR (2012) Traceless chemical ligations from *O*-acyl serine sites. Org Biomol Chem 10:4836–4838
46. Panda SS, Elagawany M, Marwani HM, Caliskan E, El Khatib M, Oliferenko A, Almary KA, Katritzky AR (2014) Chemical ligation from *O*-acyl isopeptides via 8- and 11-membered cyclic transition states. Arkivoc iv:91–106
47. Popov V, Panda SS, Katritzky AR (2013) Ligations from tyrosine isopeptides via 12-to 19-membered cyclic transition states. J Org Chem 78:7455–7461
48. Popov V, Panda SS, Katritzky AR (2013) Ligations of *N*-acyl tryptophan units to give native peptides via 7-, 10-, 11- and 12-membered cyclic transition states. Org Biomol Chem 11:1594–1597

49. Gause GF, Brazhnikova MG (1944) Gramicidin S and its use in the treatment of infected wounds. Nature 154:703–703
50. White CJ, Yudin AK (2011) Contemporary strategies for peptide macrocyclization. Nat Chem 3:509–524
51. Tyndall JDA, Nall T, Fairlie DP (2005) Proteases universally recognize beta strands in their active sites. Chem Rev 105:973–999
52. Craik DJ (2006) Seamless proteins tie up their loose ends. Science 311:1563–1564
53. Hamada Y, Shioiri T (2005) Recent progress of the synthetic studies of biologically active marine cyclic peptides and depsipeptides. Chem Rev 105:4441–4482
54. Daly NL, Craik DJ (2011) Bioactive cystine knot proteins. Curr Opin Chem Biol 15:362–368
55. Sewald N, Jakube H-D (2009) Peptides: chemistry and biology, 2nd edn. Wiley, Weinheim
56. Kates SA, Solé NA, Albericio F, Barany G (1994) In: Basava C, Anantharamaiah GM (eds) Peptides: design, synthesis and biological activity. Birkhäuser, Boston, pp 39–58
57. Davies JS (2003) The cyclization of peptides and depsipeptides. J Pept Sci 9:471–501
58. Gilon C, Mang C, Lohof E, Friedler A, Kessler H (2002) In: Goodman M, Felix AM, Moroder L, Toniolo C (eds) Synthesis of peptides and peptidomimetics (Houben-Weyl E22b: Methods of organic chemistry). Georg Thieme, Stuttgart/New York
59. Camarero JA, Muir TW (1997) Chemoselective backbone cyclization of unprotected peptides. Chem Commun 1369–1370
60. Shao Y, Lu WY, Kent SBH (1998) A novel method to synthesize cyclic peptides. Tetrahedron Lett 39:3911–3914
61. Clark RJ, Craik DJ (2012) In: Wittrup KD, Gregory LV (eds) Methods enzymol. Academic Press, Maryland Heights, pp 57–74
62. Zhang LS, Tam JP (1997) Synthesis and application of unprotected cyclic peptides as building blocks for peptide dendrimers. J Am Chem Soc 119:2363–2370
63. Kleineweischede R, Hackenberger CPR (2008) Chemoselective peptide cyclization by traceless Staudinger ligation. Angew Chem Int Ed 47:5984–5988
64. Zheng H, Wang F, Wang Q, Gao J (2011) Cofactor-free detection of phosphatidylserine with cyclic peptides mimicking lactadherin. J Am Chem Soc 133:15280–15283
65. Hosseini AS, Zheng H, Gao J (2014) Understanding lipid recognition by protein-mimicking cyclic peptides. Tetrahedron 70:7632–7638
66. Fang G-M, Li Y-M, Shen F, Huang Y-C, Li J-B, Lin Y, Cui H-K, Liu L (2011) Protein chemical synthesis by ligation of peptide hydrazides. Angew Chem Int Ed 50:7645–7649
67. Fang G-M, Wang J-X, Liu L (2012) Convergent chemical synthesis of proteins by ligation of peptide hydrazides. Angew Chem Int Ed 51:10347–10350
68. Zheng J-S, Tang S, Guo Y, Chang H-N, Liu L (2012) Synthesis of cyclic peptides and cyclic proteins via ligation of peptide hydrazides. ChemBioChem 13:542–546
69. Tulla-Puche J, Barany G (2004) On-resin native chemical ligation for cyclic peptide synthesis. J Org Chem 69:4101–4107
70. Fukuzumi T, Ju L, Bode JW (2012) Chemoselective cyclization of unprotected linear peptides by α-ketoacid-hydroxylamine amide-ligation. Org Biomol Chem 10:5837–5844
71. van de Langemheen H, Brouwer AJ, Kemmink J, Kruijtzer JAW, Liskamp RMJ (2012) Synthesis of cyclic peptides containing a thioester handle for native chemical ligation. J Org Chem 77:10058–10064
72. van de Langemheen H, van Hoeke M, Quarles van Ufford HC, Kruijtzer JAW, Liskamp RMJ (2014) Scaffolded multiple cyclic peptide libraries for protein mimics by native chemical ligation. Org Biomol Chem 12:4471–4478
73. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393:648–659
74. Franke R, Hirsch T, Overwin H, Eichler J (2007) Synthetic mimetics of the CD4 binding site of HIV-1 gp120 for the design of immunogens. Angew Chem Int Ed 46:1253–1255

75. Chamorro C, Kruijtzer JAW, Farsaraki M, Balzarini J, Liskamp RMJ (2009) A general approach for the non-stop solid phase synthesis of TAC-scaffolded loops towards protein mimics containing discontinuous epitopes. Chem Commun 821–823
76. Chen JH, Warren JD, Wu B, Chen G, Wan Q, Danishefsky SJ (2006) A route to cyclic peptides and glycopeptides by native chemical ligation using in situ derived thioesters. Tetrahedron Lett 47:1969–1972
77. Boll E, Dheur J, Drobecq H, Melnyk O (2012) Access to cyclic or branched peptides using bis (2-sulfanylethyl)amido side-chain derivatives of Asp and Glu. Org Lett 14:2222–2225
78. Taichi M, Hemu X, Qiu YB, Tam JP (2013) A thioethylalkylamido (TEA) thioester surrogate in the synthesis of a cyclic peptide via a tandem acyl shift. Org Lett 15:2620–2623
79. Lukszo J, Patterson D, Albericio F, Kates SA (1996) 3-(1-Piperidinyl)alanine formation during the preparation of *C*-terminal cysteine peptides with the Fmoc/t-Bu strategy. Lett Pept Sci 3:157–166
80. Hemu X, Taichi M, Qiu YB, Liu DX, Tam JP (2013) Biomimetic synthesis of cyclic peptides using novel thioester surrogates. Biopolymers 100:492–501
81. Li X, Lam HY, Zhang Y, Chan CK (2010) Salicylaldehyde ester-induced chemoselective peptide ligations: enabling generation of natural peptidic linkages at the serine/threonine sites. Org Lett 12:1724–1727
82. Zhang YF, Xu C, Lam HY, Lee CL, Li XC (2013) Protein chemical synthesis by serine and threonine ligation. Proc Natl Acad Sci U S A 110:6657–6662
83. Lam HY, Zhang YF, Liu H, Xu JC, Wong CTT, Xu C, Li XC (2013) Total synthesis of daptomycin by cyclization via a chemoselective serine ligation. J Am Chem Soc 135:6272–6279
84. Davis BG (2002) Synthesis of glycoproteins. Chem Rev 102:579–602
85. Gamblin DP, Scanlan EM, Davis BG (2009) Glycoprotein synthesis: an update. Chem Rev 109:131–163
86. Dwek RA (1996) Glycobiology: toward understanding the function of sugars. Chem Rev 96:683–720
87. Varki A (1993) Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 3:97–130
88. Dube DH, Bertozzi CR (2005) Glycans in cancer and inflammation. Potential for therapeutics and diagnostics. Nat Rev Drug Discov 4:477–488
89. Haase C, Seitz O (2007) In: Wittmann V (ed) Glycopeptides and glycoproteins. Springer, Berlin/Heidelberg, pp 1–36
90. Unverzagt C, Kajihara Y (2013) Chemical assembly of *N*-glycoproteins: a refined toolbox to address a ubiquitous posttranslational modification. Chem Soc Rev 42:4408–4420
91. Marcaurelle LA, Bertozzi CR (2002) Recent advances in the chemical synthesis of mucin-like glycoproteins. Glycobiology 12:69r–77r
92. Varki A (2006) Nothing in glycobiology makes sense, except in the light of evolution. Cell 126:841–845
93. Therkildsen MH, Mandel U, Christensen M, Dabelsteen E (1993) Simple mucin-type Tn and sialosyl-Tn carbohydrate antigens in salivary gland carcinomas. Cancer 72:1147–1154
94. Brockhausen I, Yang J-M, Burchell J, Whitehouse C, Taylor-Papadimitriou J (1995) Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells. Eur J Biochem 233:607–617
95. Hanisch FA (2001) *O*-Glycosylation of the mucin type. Biol Chem 382:143–149
96. Burchell J, Mungul A, Taylor-Papadimitriou J (2001) *O*-Linked glycosylation in the mammary gland: changes that occur during malignancy. J Mammary Gland Biol Neoplasia 6:355–364
97. Larsson JMH, Karlsson H, Crespo JG, Johansson MEV, Eklund L, Sjövall H, Hansson GC (2011) Altered *O*-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. Inflamm Bowel Dis 17:2299–2307

98. Stuchlová Horynová M, Raška M, Clausen H, Novak J (2013) Aberrant *O*-glycosylation and anti-glycan antibodies in an autoimmune disease IgA nephropathy and breast adenocarcinoma. Cell Mol Life Sci 70:829–839
99. Remmers N, Anderson JM, Linde EM, DiMaio DJ, Lazenby AJ, Wandall HH, Mandel U, Clausen H, Yu F, Hollingsworth MA (2013) Aberrant expression of mucin core proteins and *O*-linked glycans associated with progression of pancreatic cancer. Clin Cancer Res 19:1981–1993
100. Anish C, Schumann B, Pereira Claney L, Seeberger Peter H (2014) Chemical biology approaches to designing defined carbohydrate vaccines. Chem Biol 21:38–50
101. Gaidzik N, Westerlind U, Kunz H (2013) The development of synthetic antitumour vaccines from mucin glycopeptide antigens. Chem Soc Rev 42:4421–4442
102. Kajihara Y, Izumi M, Hirano K, Murase T, Macmillan D, Okamoto R (2011) Elucidating the function of complex-type oligosaccharides by use of chemical synthesis of homogeneous glycoproteins. Isr J Chem 51:917–929
103. Galan MC, Benito-Alifonso D, Watt GM (2011) Carbohydrate chemistry in drug discovery. Org Biomol Chem 9:3598–3610
104. Ficht S, Payne RJ, Guy RT, Wong C-H (2008) Solid-phase synthesis of peptide and glycopeptide thioesters through side-chain-anchoring strategies. Chem Eur J 14:3620–3629
105. Kiessling LL, Splain RA (2010) Chemical approaches to glycobiology. Annu Rev Biochem 79:619–653
106. Kajihara Y, Yamamoto N, Okamoto R, Hirano K, Murase T (2010) Chemical synthesis of homogeneous glycopeptides and glycoproteins. Chem Rec 10:80–100
107. Hojo H, Katayama H, Nakahara Y (2010) Progress in the ligation chemistry for glycoprotein synthesis. Trends Glycosci Glycotechnol 22:269–279
108. Liu L, Hong Z-Y, Wong C-H (2006) Convergent glycopeptide synthesis by traceless Staudinger ligation and enzymatic coupling. ChemBioChem 7:429–432
109. Brik A, Yang Y-Y, Ficht S, Wong C-H (2006) Sugar-assisted glycopeptide ligation. J Am Chem Soc 128:5626–5627
110. Brik A, Ficht S, Yang Y-Y, Wong C-H (2006) Sugar-assisted ligation of *N*-linked glycopeptides with broad sequence tolerance at the ligation junction. J Am Chem Soc 128:15026–15033
111. Brik A, Wong C-H (2007) Sugar-assisted ligation for the synthesis of glycopeptides. Chem Eur J 13:5670–5675
112. Ficht S, Payne RJ, Brik A, Wong C-H (2007) Second-generation sugar-assisted ligation: a method for the synthesis of cysteine-containing glycopeptides. Angew Chem Int Ed 46:5975–5979
113. Yang Y-Y, Ficht S, Brik A, Wong C-H (2007) Sugar-assisted ligation in glycoprotein synthesis. J Am Chem Soc 129:7690–7701
114. Payne RJ, Ficht S, Tang S, Brik A, Yang Y-Y, Case DA, Wong C-H (2007) Extended sugar-assisted glycopeptide ligations: development, scope, and applications. J Am Chem Soc 129:13527–13536
115. Bennett CS, Dean SM, Payne RJ, Ficht S, Brik A, Wong C-H (2008) Sugar-assisted glycopeptide ligation with complex oligosaccharides: scope and limitations. J Am Chem Soc 130:11945–11952
116. Wang L-X, Amin MN (2014) Chemical and chemoenzymatic synthesis of glycoproteins for deciphering functions. Chem Biol 21:51–66
117. Westerlind U (2012) Synthetic glycopeptides and glycoproteins with applications in biological research. Beilstein J Org Chem 8:804–818
118. Peri F, Nicotra F (2004) Chemoselective ligation in glycochemistry. Chem Commun 623–627
119. Specker D, Wittmann V (2007) In: Wittmann V (ed) Glycopeptides and glycoproteins. Springer, Berlin/Heidelberg, pp 65–107

120. Shin Y, Winans KA, Backes BJ, Kent SBH, Ellman JA, Bertozzi CR (1999) Fmoc-based synthesis of peptide-(alpha)thioesters: application to the total chemical synthesis of a glycoprotein by native chemical ligation. J Am Chem Soc 121:11684–11689
121. Buskas T, Ingale S, Boons G-J (2006) Glycopeptides as versatile tools for glycobiology. Glycobiology 16:113R–136R
122. Murakami M, Okamoto R, Izumi M, Kajihara Y (2012) Chemical synthesis of an erythropoietin glycoform containing a complex-type disialyloligosaccharide. Angew Chem Int Ed 51:3567–3572
123. Liu L, Bennett CS, Wong CH (2006) Advances in glycoprotein synthesis. Chem Commun 21–33
124. Hojo H, Nakahara Y (2007) Recent progress in the field of glycopeptide synthesis. Biopolymers 88:308–324
125. Monbaliu JCM, Katritzky AR (2012) Recent trends in Cys- and Ser/Thr-based synthetic strategies for the elaboration of peptide constructs. Chem Commun 48:11601–11622
126. Garner J, Jolliffe KA, Harding MM, Payne RJ (2009) Synthesis of homogeneous antifreeze glycopeptides via a ligation-desulfurisation strategy. Chem Commun 6925–6927
127. Chen J, Wan Q, Yuan Y, Zhu JL, Danishefsky SJ (2008) Native chemical ligation at valine: a contribution to peptide and glycopeptide synthesis. Angew Chem Int Ed 47:8521–8524
128. Moyal T, Hemantha HP, Siman P, Refua M, Brik A (2013) Highly efficient one-pot ligation and desulfurization. Chem Sci 4:2496–2501
129. Thompson RE, Liu XY, Alonso-Garcia N, Pereira PJB, Jolliffe KA, Payne RJ (2014) Trifluoroethanethiol: an additive for efficient one-pot peptide ligation-desulfurization chemistry. J Am Chem Soc 136:8161–8164
130. Okamoto R, Kajihara Y (2008) Uncovering a latent ligation site for glycopeptide synthesis. Angew Chem Int Ed 47:5402–5406
131. Okamoto R, Souma S, Kajihara Y (2009) Efficient substitution reaction from cysteine to the serine residue of glycosylated polypeptide: repetitive peptide segment ligation strategy and the synthesis of glycosylated tetracontapeptide having acid labile sialyl-T-N antigens. J Org Chem 74:2494–2501
132. Hojo H, Ozawa C, Katayama H, Ueki A, Nakahara Y, Nakahara Y (2010) The mercaptomethyl group facilitates an efficient one-pot ligation at Xaa-Ser/Thr for (glyco) peptide synthesis. Angew Chem Int Ed 49:5318–5321
133. Kang J, Macmillan D (2010) Peptide and protein thioester synthesis via *N*→*S* acyl transfer. Org Biomol Chem 8:1993–2002
134. Macmillan D, Adams A, Premdjee B (2011) Shifting native chemical ligation into reverse through *N*→*S* acyl transfer. Isr J Chem 51:885–899
135. Ackrill T, Anderson DW, Macmillan D (2010) Towards biomolecular assembly employing extended native chemical ligation in combination with thioester synthesis using an *N*→*S* acyl shift. Biopolymers 94:495–503
136. Premdjee B, Adams AL, Macmillan D (2011) Native *N*-glycopeptide thioester synthesis through *N*→*S* acyl transfer. Bioorg Med Chem Lett 21:4973–4975
137. Hsieh YSY, Wilkinson BL, O'Connell MR, Mackay JP, Matthews JM, Payne RJ (2012) Synthesis of the bacteriocin glycopeptide sublancin 168 and *S*-glycosylated variants. Org Lett 14:1910–1913
138. Macmillan D, Blanc J (2006) A novel neoglycopeptide linkage compatible with native chemical ligation. Org Biomol Chem 4:2847–2850
139. Lee DJ, Mandal K, Harris PWR, Brimble MA, Kent SBH (2009) A one-pot approach to neoglycopeptides using orthogonal native chemical ligation and click chemistry. Org Lett 11:5270–5273
140. Geiermann AS, Polacek N, Micura R (2011) Native chemical ligation of hydrolysis-resistant 3′-peptidyl-tRNA mimics. J Am Chem Soc 133:19068–19071

141. Geiermann AS, Micura R (2012) Selective desulfurization significantly expands sequence variety of 3′-peptidyl-tRNA mimics obtained by native chemical ligation. ChemBioChem 13:1742–1745
142. Wang C, Guo Q-X, Fu Y (2011) Theoretical analysis of the detailed mechanism of native chemical ligation reactions. Chem Asian J 6:1241–1251
143. Monbaliu JCM, Dive G, Stevens CV, Katritzky AR (2013) Governing parameters of long-range intramolecular *S*-to-*N* acyl transfers within (*S*)-acyl isopeptides. J Chem Theory Comput 9:927–934
144. Oliferenko AA, Katritzky AR (2011) Alternating chemical ligation reactivity of *S*-acyl peptides explained with theory and computations. Org Biomol Chem 9:4756–4759
145. Panda SS, Oliferenko AA, Marwani HM, Katritzky AR (2014) Effects of preorganization and hydrogen bonding on intramolecular chemical ligation of (*N*)- and (*O*)-acyl isopeptides. Mendeleev Commun 24:75–77
146. Biswas S, Kayaleh R, Pillai GG, Seon C, Roberts I, Popov V, Alamry KA, Katritzky AR (2014) Long-range chemical ligation from *N*→*N* acyl migrations in tryptophan peptides via cyclic transition states of 10-to 18-members. Chem Eur J 20:8189–8198
147. Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S (2013) QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. J Comput Chem 34:2121–2132

Index