Topics in Current Chemistry 363

Lei Liu *Editor*

Protein Ligation and Total Synthesis II

Springer

363
Topics in Current Chemistry

Editorial Board:

H. Bayley, Oxford, UK
K.N. Houk, Los Angeles, CA, USA
G. Hughes, CA, USA
C.A. Hunter, Sheffield, UK
K. Ishihara, Chikusa, Japan
M.J. Krische, Austin, TX, USA
J.-M. Lehn, Strasbourg Cedex, France
R. Luque, Córdoba, Spain
M. Olivucci, Siena, Italy
J.S. Siegel, Tianjin, China
J. Thiem, Hamburg, Germany
M. Venturi, Bologna, Italy
C.-H. Wong, Taipei, Taiwan
H.N.C. Wong, Shatin, Hong Kong
V.W.-W. Yam, Hong Kong, China
S.-L. You, Shanghai, China

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 II

With contributions by

J.W. Bode · P.K. Chaffey · O. El Mahdi · M.L. Gardner · X. Guan · C.J. Howard · J. Ma · O. Melnyk · A. Otaka · J.J. Ottesen · L. Raibaut · F. Rohrbacher · K. Sato · A. Shigenaga · J.C. Shimko · Z. Tan · Q. Wan · T.G. Wucherpfennig · R.R. Yu · C. Zeng · J. Zeng

Springer

Editor
Lei Liu
Department of Chemistry
Tsinghua University
Beijing
China

ISSN 0340-1022 ISSN 1436-5049 (electronic)
Topics in Current Chemistry
ISBN 978-3-319-19188-1 ISBN 978-3-319-19189-8 (eBook)
DOI 10.1007/978-3-319-19189-8

Library of Congress Control Number: 2015941105

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

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) 363: 1–32
DOI: 10.1007/128_2014_597
© Springer International Publishing Switzerland 2014
Published online: 12 March 2015

Chemical Protein Synthesis with the KAHA Ligation

Florian Rohrbacher, Thomas G. Wucherpfennig, and Jeffrey W. Bode

Abstract Since the first report of the chemoselective amide bond forming reaction between α-ketoacids and hydroxylamines in 2006, the KAHA (α-ketoacid-hydroxylamine) ligation has advanced to a useful tool for the routine synthesis of small to medium sized proteins and cyclic peptides. In this chapter we introduce the concept of KAHA ligation starting with the synthesis and properties of hydroxylamines and α-ketoacids, methods for their incorporation into peptides, and give an insight into the mechanism of the KAHA ligation. We cover important improvements including sequential ligations with 5-oxaproline, traceless synthesis of peptide α-ketoacids and show their application in chemical protein synthesis and cyclic peptide synthesis. Recent developments of the KAT (potassium acyl trifluoroborate) ligation and its application as fast and chemoselective bioconjugation method are described and an outlook on ongoing work and possible future developments is given at the end of the chapter.

Keywords Chemical protein synthesis • Cyclic peptides • KAHA ligation • O • N-acyl shift • Peptide hydroxylamines • Peptide α-ketoacids

F. Rohrbacher and T.G. Wucherpfennig
Laboratorium für Organische Chemie, Department of Chemistry and Applied Biosciences, ETH Zürich, 8093 Zürich, Switzerland

J.W. Bode (✉)
Laboratorium für Organische Chemie, Department of Chemistry and Applied Biosciences, ETH Zürich, 8093 Zürich, Switzerland

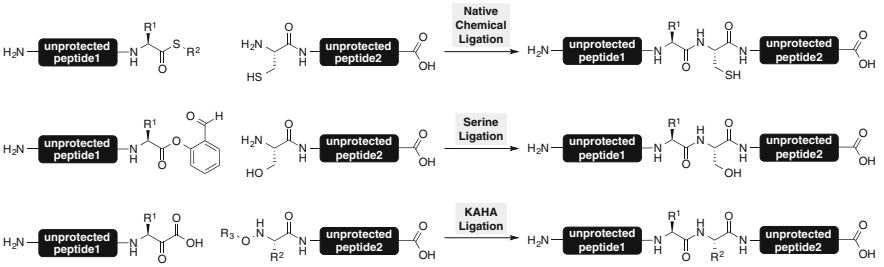
Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, 464-8602 Chikusa, Nagoya, Japan
e-mail: bode@org.chem.ethz.ch

Contents

1 Introduction

1.1 Overview of Chemical Ligations for Protein Synthesis

The ability to synthesize biologically active proteins chemically in a controlled way and with defined sequences is one of the greatest achievements of synthetic chemistry in the last 20 years [1]. Essential for further progress was the development of the solid-phase peptide synthesis technique developed by Merrifield in 1963, as this method established a very general and scalable synthesis of peptides [2]. Despite tremendous improvements over the last five decades, the size of peptides obtained by SPPS is commonly limited to 40–60 amino acid residues, considerably smaller than average eukaryotic proteins consisting of around 360 amino acid residues [3]. In addition to fragment couplings of fully protected peptide segments, the most widely used approach to overcome this restriction is to assemble the final protein from multiple, easy to access peptide building blocks by chemical ligation methods (Scheme 1) [4, 5]. The best implementations use completely unprotected peptide segments, work at suitably low substrate concentrations, and give natural amide bonds at the ligation site.



Scheme 1 Chemical ligation methods for the synthesis of peptides and proteins

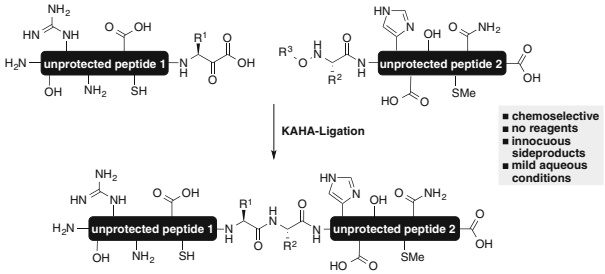
A milestone was the development of the native chemical ligation (NCL) by Kent and coworkers in 1994, as discussed in great detail in other sections of this book [6]. In brief, NCL exploits the chemoselective reaction of C-terminal peptide thioesters and peptides with N-terminal thiols such as cysteine to yield a natural amide bond [7]. Initial limitations included the requirement for relatively rare cysteine residues [8] at the ligation site and challenges to synthesize C-terminal peptide thioesters, but significant improvements by numerous research groups rendered NCL a robust and reliable method to synthesize proteins chemically [9–16].

Another relatively new chemical ligation method using different functional groups is the serine/threonine ligation (STL) introduced by Li in 2013 [17], and based on elegant prior art by Tam [18]. C-terminal peptide salicylaldehyde esters react with N-terminal serine or threonine residues, yielding a cyclic *N*,*O* benzylidene acetal intermediate which can be directly cleaved to yield a natural amide bond at the ligation site. Current limitation of this coupling include limited scope for the C-terminal residue bearing the salicylaldehyde ester, the requirement of relatively high substrate concentrations, and the use of organic solvents which may limit solubility and biocompatibility [19].

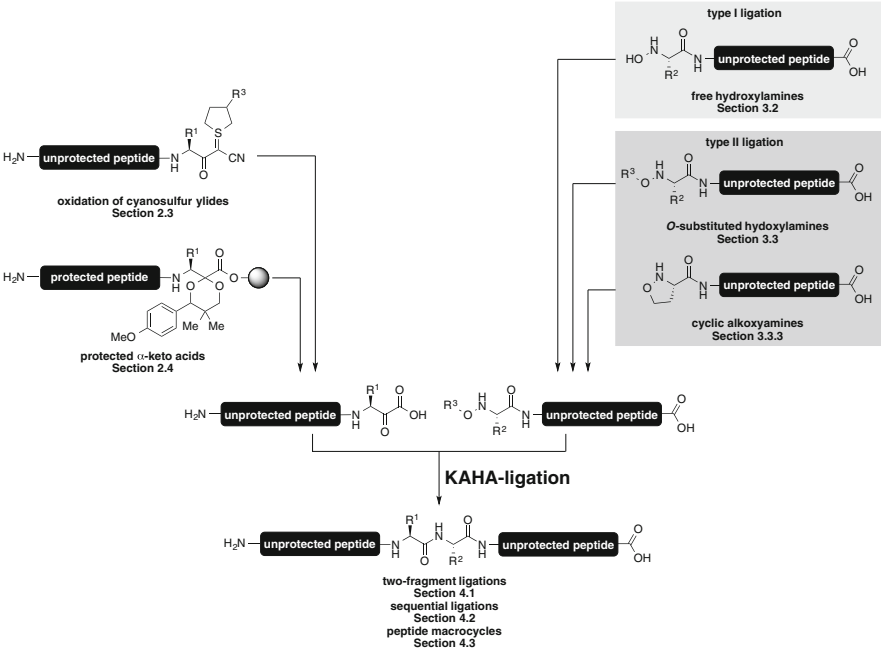
A third ligation method using distinct and orthogonal functional groups is the α-ketoacid-hydroxylamine (KAHA) ligation developed by Bode and coworkers in 2006 [20, 21], described in more detail in the following sections.

1.2 KAHA Ligation

As the name suggests, the α-ketoacid-hydroxylamine ligation employs C-terminal peptide α-ketoacids and N-terminal peptide hydroxylamines or derivatives which react chemoselectively to give an amide bond at the ligation site (Scheme 2). This reaction does not need any reagents or catalysts, and proceeds at slightly elevated temperatures in mixtures of water and organic solvent without the formation of problematic byproducts [22]. The mildly acidic reactions typically increase the



Scheme 2 α-Ketoacid-hydroxylamine (KAHA) ligation



Scheme 3 Overview of different functional groups used in KAHA-ligation

solubility of protein fragments and deter the hydrolysis of sensitive functional groups.

1.2.1 Types of KAHA Ligation

Over the years, different monomers and synthetic approaches for peptide hydroxylamines and peptide α-ketoacids have been developed (Scheme 3). The different synthetic routes to peptide α-ketoacids are discussed in Sect. 2. Regarding the different hydroxylamine derivatives evaluated for KAHA ligation, differences

were found not only in terms of stability and reactivity – they also seem to operate under different reaction mechanisms, leading to sometimes unexpected outcomes as discussed in Sects. 1.2.2 and 4.2.1. Based on the molecular structure and reaction mechanism, the ligation reactions are divided into two categories: type I ligation with free peptide hydroxylamines discussed in Sect. 4.1 and type II with O-substituted hydroxylamines discussed in Sect. 4.2.

1.2.2 Mechanism

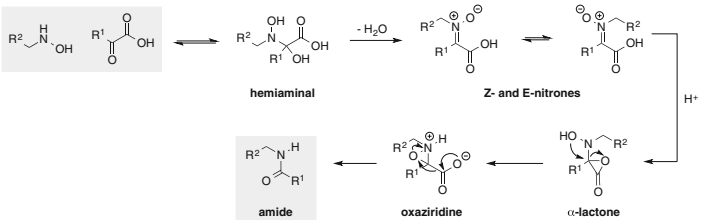
Type I

In order to identify the mechanism of type I ligations, we have extensively studied the reaction of ^{18}O labeled hydroxylamines and α-ketoacids [23, 24]. To our surprise, we found that the oxygen atom of the newly formed amide originates from the hydroxylamine. This can be explained by the initial formation of nitrones – which are observable intermediates of the reaction – and subsequent attack of the carboxylate to give an α-lactone. The α-lactone can rearrange to an α-oxaziridinyl acid which undergoes decarboxylation to afford the final amide product. This pathway is also supported by the observation that the potassium salt of the α-oxaziridinyl acid can be independently prepared and isolated, and cleanly carboxylates to give an amide upon addition of 1 equiv. of TFA (Scheme 4).

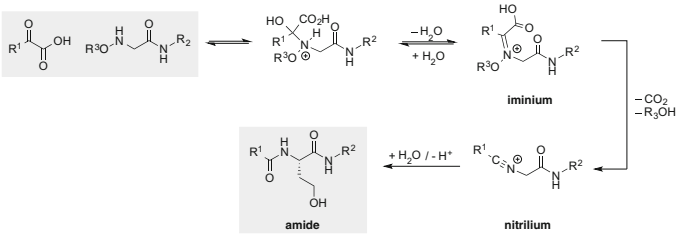
Type II

In contrast to type I ligations, KAHA ligation with O-substituted peptide hydroxylamines in ^{18}O labeled water leads to incorporation of ^{18}O into the product. Although we have not fully elucidated the mechanism of type II ligations, the most likely pathway involves the formation of an iminium which dehydrates to a nitrilium. As with the Ritter reaction, addition of water gives the amide product. In prior studies this pathway has been proposed by Sucheck for type I reactions [25] (Scheme 5).

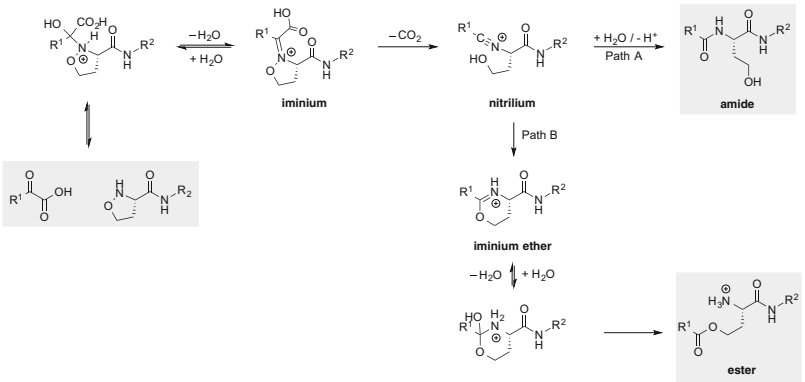
The proposed pathway of type II ligations is supported by the surprising outcome of reactions with the cyclic hydroxylamine 5-oxaproline. The major product of this



Scheme 4 Mechanism for type I ligations



Scheme 5 Possible mechanism for type II ligations



Scheme 6 Mechanism for KAHA ligations with 5-oxaproline

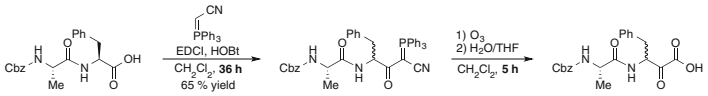
reaction is the ester and not the expected amide [26, 27]. Most likely, the alcohol – which is released upon decarboxylation – can intercept the intermediate nitrilium in an intramolecular addition. The resulting cyclic imino ether is hydrolyzed under acidic conditions to give the ester. As expected, ^{18}O is incorporated into both the ester and amide products if the KAHA ligation with 5-oxaproline is carried out in ^{18}O-labeled water (Scheme 6).

2 α-Ketoacids

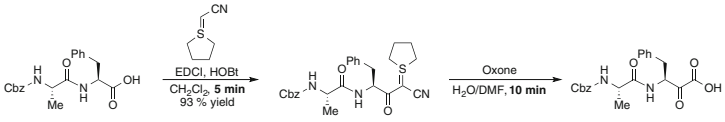
2.1 General Properties of α-Keto Acids

α-Ketoacids have been known for a long time; the first example, pyruvic acid, was prepared by Berzelius in 1835 [28, 29]. They play an important role in the metabolism of amino acids and are found as synthetic intermediates in prokaryotic and eukaryotic cells [30]. Because of their highly electrophilic carbonyl group, simple α-ketoacids are prone to nucleophilic attack, resulting in polymerization, cyclization, or other addition products [31]. Depending on the pH value, simple

Scheme 7 Stability of α-ketoacids



Scheme 8 Synthesis of peptide α-ketoacids based on a phosphorus ylide



Scheme 9 Synthesis of peptide α-ketoacids based on a sulfur ylide

α-ketoacids are also prone to form hydrates or enolates, resulting in epimerization of stereogenic centers in the β position [32].

In contrast, we have found that peptide α-ketoacids are remarkably stable and tolerate unprotected amino acid side chains. They do not epimerize in aqueous solutions under acidic conditions, rendering them stable towards standard peptide handling and purification methods, including reverse phase HPLC and lyophilization (Scheme 7).

2.2 Synthesis of Peptide α-Ketoacids

2.2.1 Phosphorus Ylides

The first method to synthesize peptide α-ketoacids was developed by Wasserman and coworkers and is based on stable phosphorus ylides, which were oxidized to form acyl cyanides which undergo in situ hydrolysis (Scheme 8). In our hands this method delivered the desired products but suffered from long reaction times for the coupling of the phosphonium salt, resulting in epimerization of the neighboring stereocenter [33]. Additionally, the oxidation required the use of toxic ozone gas, very long reaction times (4–6 h), low temperatures (−78°C), and the use of inert organic solvents in which the solubility of the peptides was found to be problematic.

2.2.2 Sulfur Ylides

In seeking suitable alternatives for the phosphonium ylides, we investigated the sulfur ylide analogues (Scheme 9) [34]. We were pleased to find that the coupling

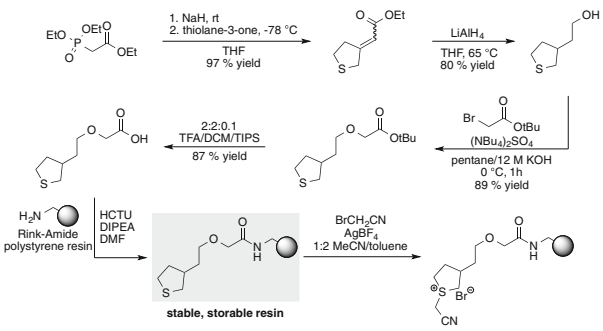
reaction of carboxylic acid and sulfonium salt readily proceeded quickly and in an epimerization-free manner.

During optimization studies, we found that solid Oxone, a cheap commercially available mixed salt containing potassium peroxymonosulfate, conveniently oxidizes sulfur ylides to the corresponding α-ketoacid with short reaction times (5–30 min) in mixtures of organic solvents and water, improving the solubility of the peptide substrates significantly. In further studies we confirmed that this relatively mild method tolerates all unprotected amino acid residues except for the oxidation-prone, sulfur-containing cysteine and methionine residues Tryptophan residues were also sometimes problematic. Despite this limitation, the sulfurylide approach offered powerful and convenient access to most classes of α-ketoacids.

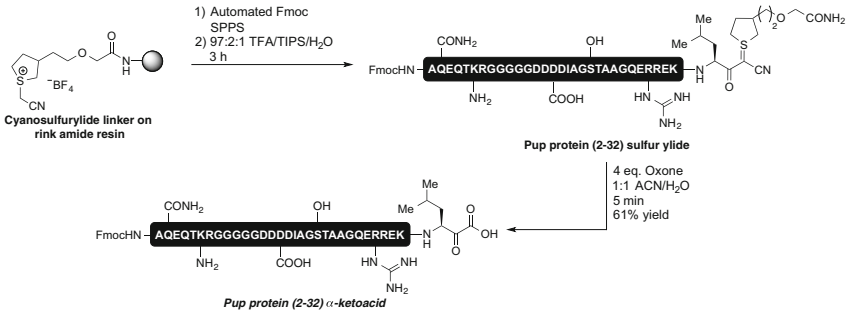
2.3 *Solid Supported Linker for Peptide Sulfur-Ylide Synthesis*

In order to simplify the preparation of C-terminal peptide α-ketoacids, we sought to develop a version of the sulfur ylide suitable for synthesis on a solid phase. A solid supported reagent would render the preparation of the peptide sulfurylides in a transparent process simply by using a suitable resin at the beginning of the synthesis. To achieve this, a suitable linker with a carboxylic acid was attached to the thiolane ring and immobilized on Rink amide polystyrene resin by regular amide coupling (Scheme 10) [35]. After alkylation on a solid phase a solid supported version of the sulfonium salt is obtained, ready for coupling with standard Fmoc-amino acids (Fmoc = 9-fluorenylmethyl carbamate) to give the corresponding solid supported sulfur ylides, which are generally stable and storable.

To illustrate the versatility of this protocol, the synthesis of a 31-mer peptide α-keto acid required for the synthesis of Pup protein is shown in Scheme 11 [36, 37]. The first amino acid residue was coupled under standard SPPS conditions



Scheme 10 Solid supported sulfur ylide for the synthesis of peptide α-ketoacids



Scheme 11 Synthesis of Pup (2–32) α-ketoacid using the sulfur ylide resin

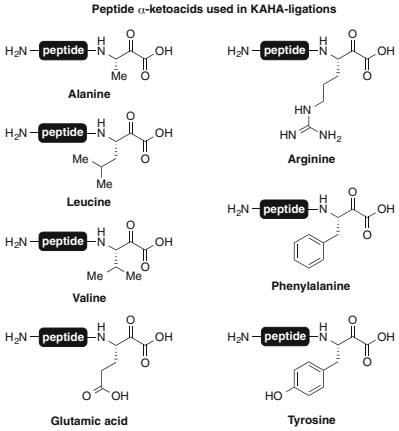
and the peptide chain extended with an automated peptide synthesizer. Cleavage under standard conditions using triisopropyl silane and water as scavenger and purification by HPLC gave the pure Pup (2–32) sulfur ylide, which could be oxidized with Oxone under mild conditions to give 61% yield of Pub (2–32) α-ketoacids after purification.

The described protocol is quite general and over the years we have evaluated numerous C-terminal residues suited well as C-terminal peptide α-keto acids (Scheme 12), in terms both of preparation and of performance in KAHA ligations. C-terminal peptide α-ketoacids up to 48 residues in length have been prepared by oxidation of the corresponding sulfurylide.

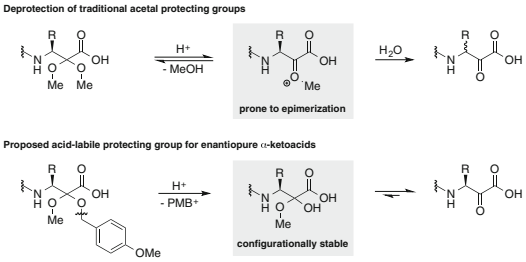
2.4 Protecting Groups for C-Terminal α-Ketoacids

Despite its great success, the sulfurylide protocol for synthesizing peptide α-ketoacids has two major limitations. First, although the oxidation conditions are fairly mild and are tolerated by most unprotected functionalities, they are incompatible with cysteine, methionine, and tryptophan residues present in the peptide segment, as these form oxidation products. Second, the oxidation protocol requires an additional manipulation and purification step at the late stage of the synthesis. This reduces the overall yield and decreases its attractiveness for certain applications, such as the preparation of cyclic peptides.

Unprotected α-ketoacids and simple derivatives such as α-ketoesters are unstable in the conditions used during SPPS, especially the basic conditions used for removal of N-terminal Fmoc-protection. We envisioned developing a protected α-ketoacid monomer which would fulfill several requirements: (1) it must be stable to all conditions used in SPPS; (2) it should give the peptide α-ketoacid directly without further manipulation under standard acidic cleavage conditions from the solid support; (3) the α-ketoacid must not undergo epimerization at any point during the synthesis, especially upon resin cleavage [38, 39].



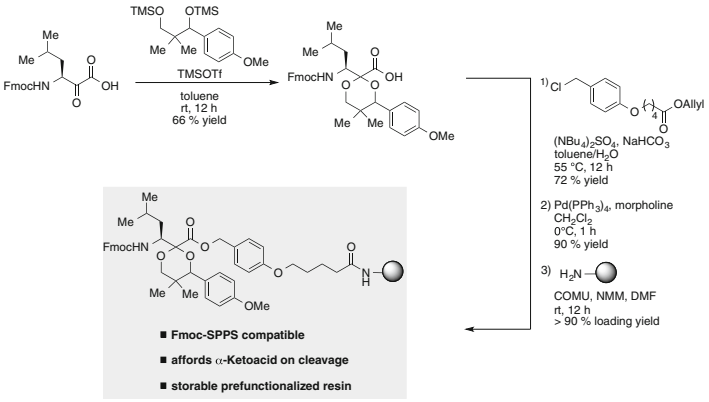
Scheme 12 Overview of peptide α-ketoacids utilized in KAHA-ligations



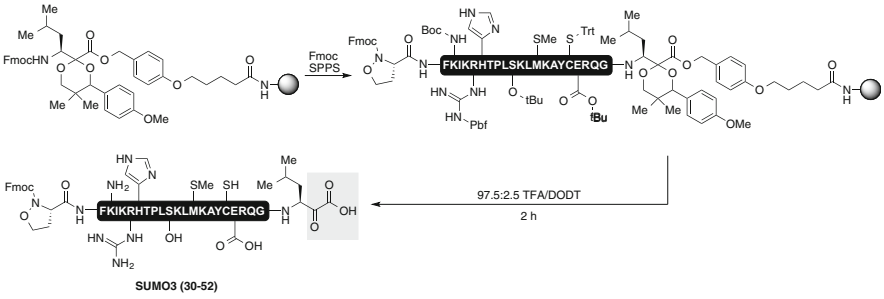
Scheme 13 Design of a protecting group for α-ketoacids that is removable without epimerization

The third requirement of epimerization-free cleavage was found to be the most challenging. Simple masking of the keto group of the α-ketoacid as acyclic dimethyl acetal was feasible, but partial epimerization upon cleavage was observed in some cases. To avoid this, we postulated that a change in deprotection mechanism might help. For example, an electron-donating aromatic substituent could facilitate the cleavage and therefore avoid epimerization (Scheme 13).

The synthesis of the simplest implementation of this design, a mono- or di-*para*-methoxybenzyl acetal, could not be achieved. After considerable experimentation, we found that a cyclic acetal derived from 1-(4-methoxyphenyl)-2,2-dimethylpropane-1,3-diol gave the protected α-ketoacid monomer in good yield, and deprotection occurred under standard acidic resin cleavage conditions in an epimerization-free manner to give peptide α-ketoacids directly (Scheme 14). Thiol scavengers such as 2,2′-(ethylenedioxy) diethanethiol (DODT) or 1,2-ethanedithiol have to be used in the cleavage cocktail, as the frequently used triisopropyl silane leads to reduction side products of the α-ketoacid. For simplification of the immobilization of the protected α-ketoacid monomer on a solid phase, a suitable Wang-type linker was attached to the protected α-ketoacid monomer. The protected



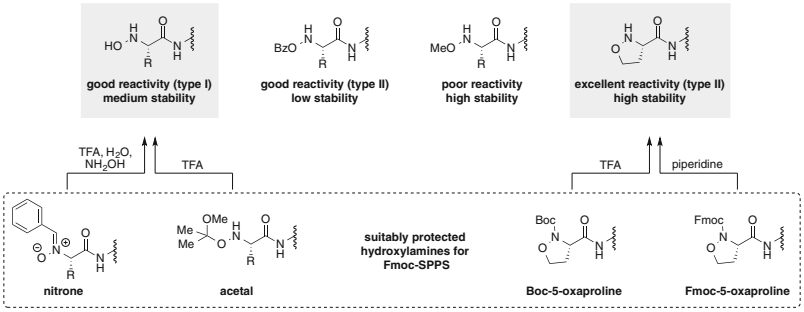
Scheme 14 Synthesis of a protected α-ketoacid resin



Scheme 15 Synthesis of a bifunctional peptide α-ketoacid using the prefunctionalized resin

α-ketoacid monomer with the linker can be loaded onto a variety of resins under standard coupling conditions to provide a stable and storable resin ready for use in SPPS synthesis.

To demonstrate its utility, the synthesis of a bifunctional peptide fragment required for the synthesis of SUMO3 protein bearing an N-terminal protected oxaproline residue and a C-terminal α-ketoacid is shown in Scheme 15. Starting from the protected α-ketoacid resin, the peptide chain was elongated by automated Fmoc-SPPS under standard conditions. After the synthesis was complete, the peptide was cleaved off the resin using 2.5% DODT in TFA and the desired peptide α-ketoacid directly obtained after purification by HPLC without additional manipulation steps. Notably, the sequence contains a cysteine and methionine, which would be incompatible with the previous oxidative strategy.



Scheme 16 Overview of hydroxylamine derivatives for KAHA ligation

3 Hydroxylamines

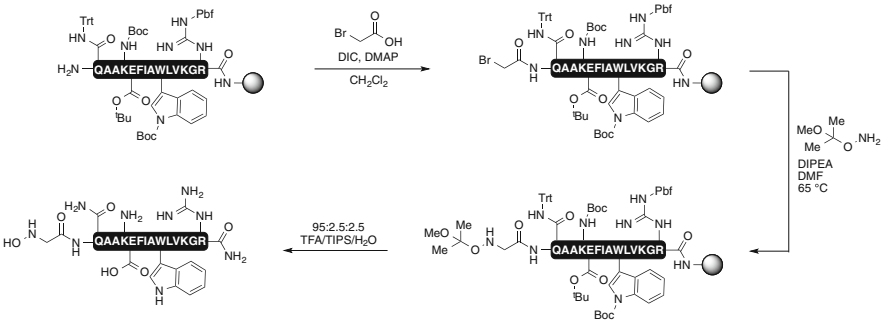
3.1 Overview

During our ongoing efforts to identify and synthesize peptide hydroxylamines, we have observed notable differences in both reactivity and stability of diversely substituted hydroxylamines (Scheme 16). Whereas, for example, OBz peptide-hydroxylamines show good reactivity, they are not sufficiently stable under ligation and SPPS conditions. On the other hand, *O*-alkyl peptide-hydroxylamines are very stable, but show poor reactivity in the KAHA ligation. So far we have identified two main classes of hydroxylamines, which are both stable and reactive enough for protein synthesis: O-unsubstituted hydroxylamines (type I ligations) and the cyclic hydroxylamine 5-oxaproline (type II ligations). In order to render these peptide-hydroxylamines suitable for Fmoc-SPPS, we have developed suitable protecting groups and synthesis strategies. In the following sections this is discussed in detail.

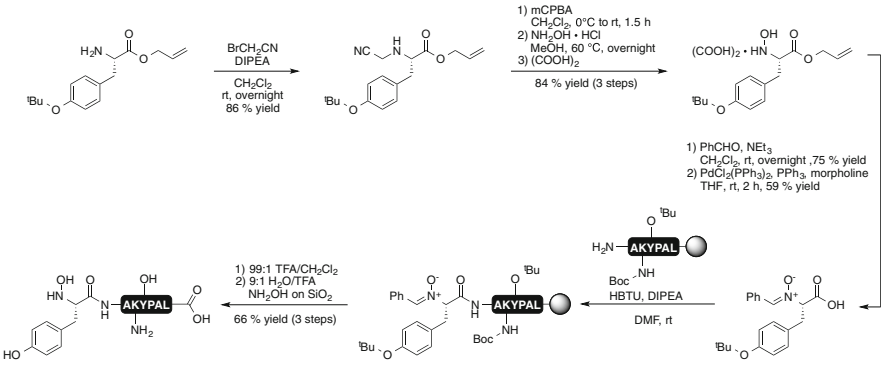
3.2 O-Unsubstituted Peptide Hydroxylamines

3.2.1 On-Resin Synthesis by Nucleophilic Substitution

As shown in Scheme 17, hydroxylamine-functionalized peptides can be prepared on-resin by coupling the free N-terminus with bromoacetic acid and subsequent nucleophilic displacement by an *O*-acetal protected hydroxylamine [40]. Upon acidic cleavage from the resin and simultaneous removal of the protecting groups, a peptide with an N-terminal hydroxylamine glycine was obtained.



Scheme 17 On-resin synthesis of a peptide hydroxylamine by nucleophilic substitution



Scheme 18 Synthesis of nitrone-protected *N*-hydroxy aminoacid building blocks for Fmoc SPPS

3.2.2 Nitrone-Protected *N*-Hydroxy Aminoacid Building Blocks

The method described in the previous section allows the synthesis of peptide hydroxylamines with an N-terminal glycine. As it is desirable to introduce other functionalized *N*-hydroxyl aminoacids, we have developed a method to prepare nitrone-protected *N*-hydroxy aminoacid building blocks amenable to direct incorporation into peptides by standard Fmoc SPPS without the loss of stereochemical integrity (Scheme 18) [41]. By following a modification of the procedure reported by Fukuyama [42], O-protected aminoacids were alkylated in good yields to afford the cyanomethyl amines. Following a one-pot oxidation to the corresponding nitrones and subsequent hydrolysis with hydroxylamine hydrochloride, the free *N*-hydroxylamino acids were obtained. In order to render these building blocks compatible with Fmoc SPPS, the hydroxylamine needs to be protected with a suitable protecting group. We found that the corresponding benzylidene nitrones were bench stable and could be incorporated by standard couplings. The hydroxylamines were converted to the benzylidene nitrones with benzaldehyde, followed by removal of the allyl ester. Using standard coupling reagents, the

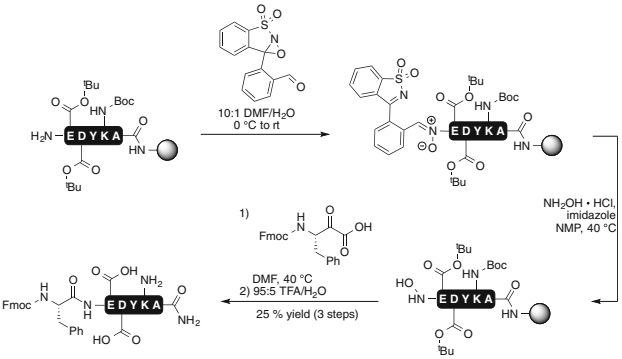
N-benzylidene amino acid nitrones were coupled onto the resin-bound protected peptide. The peptide can be cleaved by anhydrous TFA without removal of the nitrone protecting group. To remove the nitrone protecting group, the benzaldehyde must be scavenged from the reaction mixture to avoid reattachment. This can be realized by treatment of the peptide with 10% aqueous TFA and repeated, passing through a short column of C18 silica gel and a resin-bond hydroxylamine to afford the unprotected peptide with an N-terminal free hydroxyl amine. The main disadvantage of this method is the poor stability of most N-terminal peptide hydroxylamines. In many cases they undergo oxidation to the corresponding oxime, particularly the hydroxylamines of less hindered amino acid residues.

3.2.3 In Situ Preparation of Peptides with N-Terminal Nitrone-Protected *N*-Hydroxy Aminoacids

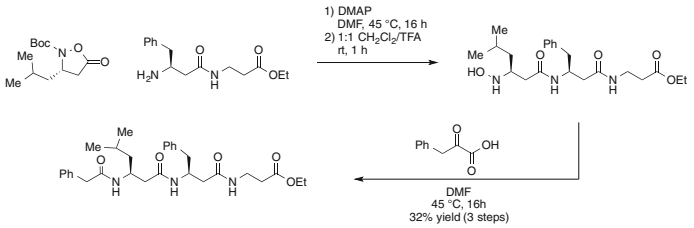
Many procedures for the preparation of hydroxylamines rely on the oxidation of imines to nitrones, which can be hydrolyzed to liberate a free hydroxylamine. To combine the imine formation and subsequent oxidation in a single reagent, we developed an *N*-sulfonyloxaziridine-based reagent featuring an aldehyde in proximity to the reactive oxaziridine moiety (Scheme 19) [43]. Upon reaction with the free N-terminus of a peptide, the intermediately formed imine is oxidized to the nitrone, which can be hydrolyzed on-resin by treatment with hydroxylamine hydrochloride and imidazole in NMP. The resulting *N*-hydroxy amine peptide can undergo on-resin ligations or can be cleaved from the resin [44].

3.2.4 Synthesis of β-Peptides with Isoxazolidin-5-One Building Blocks

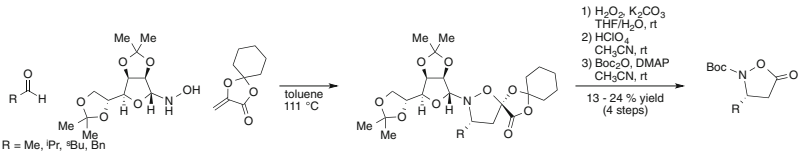
An alternative approach to the synthesis and incorporation of β^3-*N*-hydroxyl amino acids is shown in Scheme 20 and relies on the synthesis of substituted isoaxazolidin-5-one monomers, which can be regarded as activated β^3-aminoacids



Scheme 19 On-resin synthesis of hydroxylamines with an *N*-sulfonyloxaziridine-based reagent



Scheme 20 Incorporation of β^3-*N*-hydroxyl amino acids by reaction with functionalized isoxazolidin-5-one building blocks



Scheme 21 Synthesis of enantiopure isoxazolidin-5-one building blocks

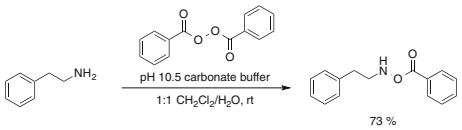
[45]. Simple mixing of a primary amine with the isoaxazolidin-5-one monomer in the presence of DMAP results in the formation of an amide bond to afford *N*-(Boc)-*N*-hydroxy β^3-peptides. Subsequent removal of the Boc protecting group with 50% TFA liberates the free *N*-hydroxy β^3-peptide.

The synthesis of these monomers is achieved by a (2+3) cycloaddition between protected acrylates and nitrones, which are formed in situ by a D-gulose-derived hydroxylamine and aldehydes (Scheme 21). The product of the cycloaddition can be purified to >99:1 dr by recrystallization. Subsequent oxidative decarboxylation with H_2O_2 under mildly basic conditions and acidic cleavage of the chiral auxiliary affords the enantiopure isoxazolidin-5-one monomers, which are *N*-Boc protected with Boc_2O and used for incorporation into peptides in solution or on-resin [46].

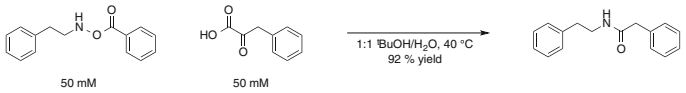
3.3 O-Substituted Hydroxylamines

3.3.1 OBz Hydroxylamines

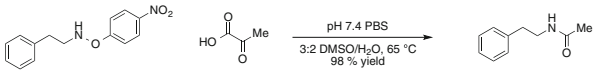
Based on initial observations by Gambarjan [47], and further improvement by Ganem [48], Phanstiel developed a direct method for the synthesis of OBz substituted hydroxylamines starting from the free amines (Scheme 22) [49]. The amines were allowed to react with benzoyl peroxide in a biphasic solvent system buffered at pH 10.5 to afford the corresponding OBz derivatives. While this method is convenient for simple amines, it is not possible to prepare OBz α-aminoacids with this method. Phanstiel demonstrated that OBz hydroxylamines can undergo



Scheme 22 Preparation of OBz hydroxylamines from free amines



Scheme 23 KAHA ligation with OBz hydroxylamines



Scheme 24 KAHA ligation with an *O*-4-nitrophenyl hydroxylamine

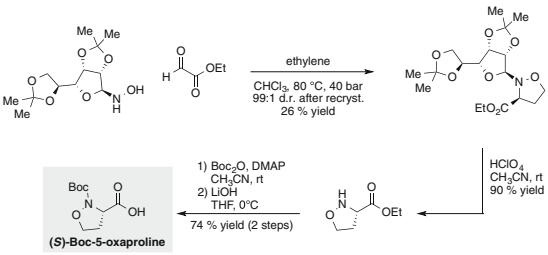
KAHA ligation with phenylpyruvic acid affording *N*-acetyl phenethylamine under neutral aqueous conditions at 40°C (Scheme 23) [50].

3.3.2 *O*-Aryl Hydroxylamines

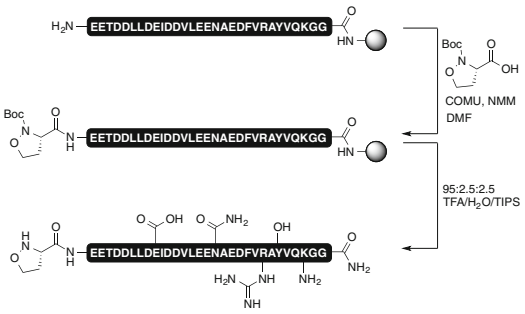
For simple substrates, Phanstiel showed that *O*-aryl substituted hydroxylamines undergo KAHA ligation (Scheme 24). Interestingly, this reaction proceeds in high yields, even in neutral buffered conditions [51]. Preliminary biological studies showed that *O*-4-nitrophenyl hydroxylamine had no toxic effect on Chinese hamster ovary cells after incubation at 100 μM concentration for 48 h at 37°C, demonstrating the potential of the KAHA ligation for biological applications.

3.3.3 Cyclic Hydroxylamines

As part of further efforts to identify chemically stable yet highly reactive hydroxylamines, we found that the cyclic hydroxylamine shown in Scheme 25 perfectly met both our requirements for handling and reactivity in the KAHA ligation [36, 37]. This building block, Boc (X)-5-oxaproline, can be synthesized in a similar fashion to that described in Sect. 3.2.4 for the synthesis of the isoxazolidin-5-one monomers, by a variation of the method reported by Vasella in 1981 [52]. The nitrone formed by the gulose-derived hydroxylamine and ethyl glyoxylate reacts in a (3+2) cycloaddition with ethylene to afford the desired cyclic product as a single diastereomer after recrystallization. After cleavage of the chiral auxiliary and *N*-Boc protection, the ethyl ester was hydrolyzed with LiOH to afford the enantiopure



Scheme 25 Synthesis of Boc-5-oxaproline using a chiral auxiliary



Scheme 26 Coupling of (X) Boc-5-oxaproline onto a resin-bound peptide with standard coupling reagents

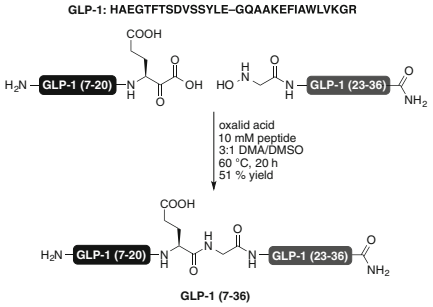
Boc-5-oxaproline monomer. This building block can be incorporated into a peptide chain by standard coupling reagents as shown in Scheme 26. Using an analogous route starting from *tert*-butyl glyoxylate, Fmoc- (S)-5-oxaproline can also be prepared.

4 Ligations/Protein Synthesis

4.1 Ligations with Type I Hydroxylamines

The original version of the KAHA ligation reported in 2006 mostly employed unsubstituted hydroxylamines as ligation partners. Advantageously, multiple synthetic routes towards unsubstituted hydroxylamine peptides existed (see Sect. 3) and an amide bond with a natural amino acid residue at the ligation site was obtained. However, in the context of peptide purification and ligation, the unsubstituted hydroxylamines are often sensitive towards oxidation, yielding unreactive oximes. Additionally, type I ligations are sometimes slowed by the presence of water and generally require organic solvents, which is disadvantageous for solubility for many peptide fragments.

Scheme 27 Synthesis of GLP-1 (7–36)



Despite these limitations, the KAHA ligation with unsubstituted hydroxylamine has been successfully employed to synthesize human GLP-1 (7–21) protein as depicted in Scheme 27 [40].

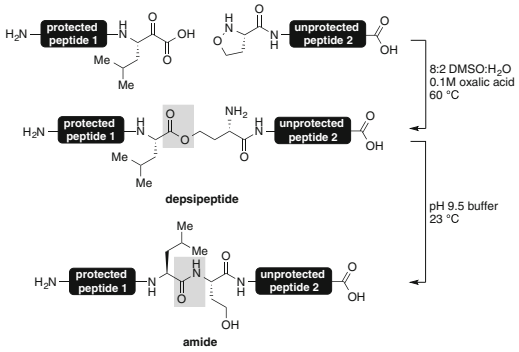
The synthesis involved a two-fragment, one-ligation strategy by combining Fragment 1 GLP-1 (7–21), obtained via oxidation of the corresponding peptide sulfur ylide, and the N-terminal glycine hydroxylamine Fragment 2 GLP-1 (22–36), prepared by the method described in Sect. 3.2.1. The ligation was performed in a mixture of 3:1 DMA/DMSO containing 2 equiv. of oxalic acid at 60°C with a minimal excess of the hydroxylamine fragment (1.05 equiv.) at a peptide fragment concentration of 10 mM. After purification by HPLC, the ligation product GLP-1 (7–36) was obtained in 51% yield.

4.2 Ligations with 5-Oxaproline

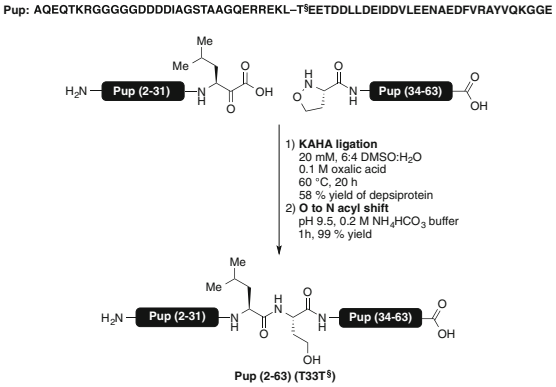
Ligations with 5-oxaproline as hydroxylamine component were found to react at comparable rates to the previously used unsubstituted hydroxylamines while offering much higher stability under the reaction conditions. This results in a higher overall performance and ease of implementation in comparison with previous cases. However, the amino acid residue at the ligation site is a non-canonical homoserine residue, but if the ligation site is chosen properly (e.g., mutation of serine, threonine, methionine, or asparagine residue to homoserine), the impact of this rather conservative mutation on structure or function is small [38, 39].

4.2.1 Formation and Rearrangement of Depsipeptides

During studies in dipeptide model systems, it was found that ligations with 5-oxaproline give unexpectedly the depsipeptide with an ester linkage rather than the anticipated amide (Scheme 28, see also mechanism in the section "Type II") [26, 27]. Conveniently, these depsipeptides are readily rearranged to the amides in basic buffers. They might also offer an advantage in terms of handling, as



Scheme 28 Formation of depsipeptides in the KAHA-ligation with 5-oxaproline and *O* to *N* acyl shift to the amide



Scheme 29 Synthesis of Pup (2–63) by a two-segment ligation

depsipeptides are in general more polar and more soluble than their amide counterparts [53].

4.2.2 Two-Segment Ligations with 5-Oxaproline

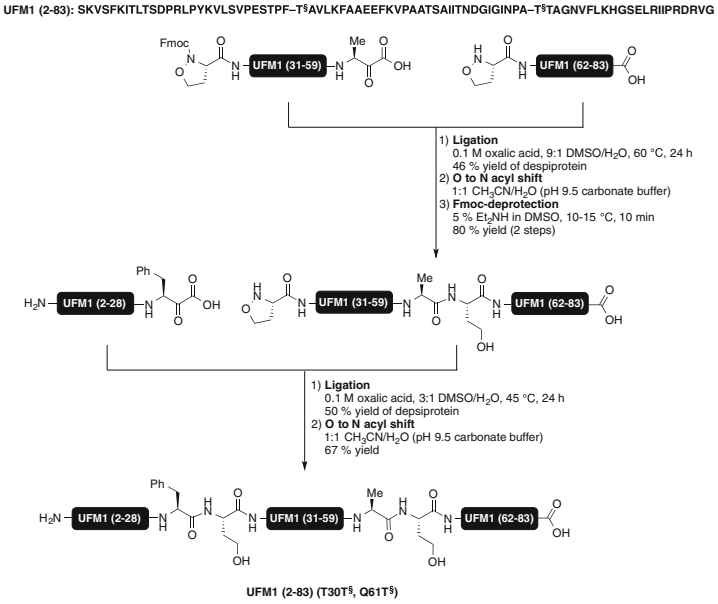
As a representative example for a two-segment, one-ligation strategy employing 5-oxaproline, the synthesis of Pup (2–63) protein is shown in Scheme 29. The two fragments were reacted at a concentration of 20 mM in 6:4 DMSO:H_2O at 60°C and the depsiprotein was obtained in 58% yield after purification. Subsequent *O* to *N* acyl shift in pH 9.5 NH_4HCO_3 buffer and removal of the volatile buffer by lyophilization gave the desired Pup (2–63) protein in 99% yield.

4.2.3 Multi-Segment Ligations

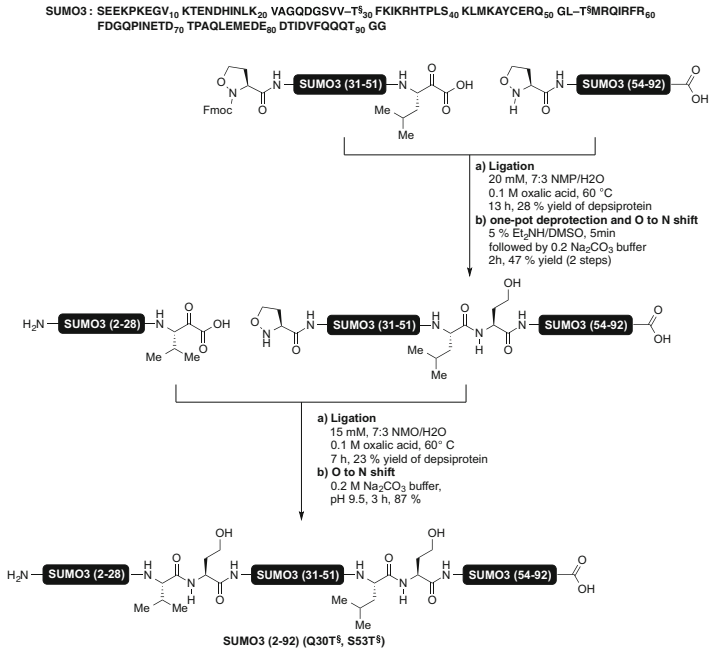
As the protein segment size is limited by SPPS, larger protein targets have to be assembled from more than two segments in multi-segment ligations. The bifunctional internal segments have both a N-terminal hydroxylamine and a C-terminal α-ketoacid. In order to prevent possible side reactions such as cyclization or oligomerization, one of these functionalities has to be temporarily blocked. Conveniently, this is realized by a base labile Fmoc-protection of the hydroxylamine.

UFM-1

One of the first examples of a multi-segment KAHA ligation is the synthesis of the important modifier protein UFM1 [54, 55]. In this three-segment, two-ligation strategy (Scheme 30), the N-terminal 5-oxaproline residue of the bifunctional internal segment 2 UFM1 (29–61) is protected by an Fmoc group to avoid side reactions. Segments 2 UFM1 (30–60) and segment 3 UFM1 (62–83) were reacted in 9:1 DMSO:H_2O and the depsiprotein was obtained in 46% yield after purification by HPLC. As expected, the carbamate-protected 5-oxaproline residue remained intact and unreactive during the ligation. Subsequent *O* to *N* acyl shift in a mixture of 1:1 CH_3CN and aqueous carbonate buffer at pH 9.5 followed by Fmoc-deprotection with 5% Et2NH in DMSO gave UFM1 (30–83) in 80% yield over two steps. The second ligation between segment 1 UFM1 (2–29) and segment 2 + 3



Scheme 30 Synthesis of UFM1 (2–83) by a three-segment sequential ligation

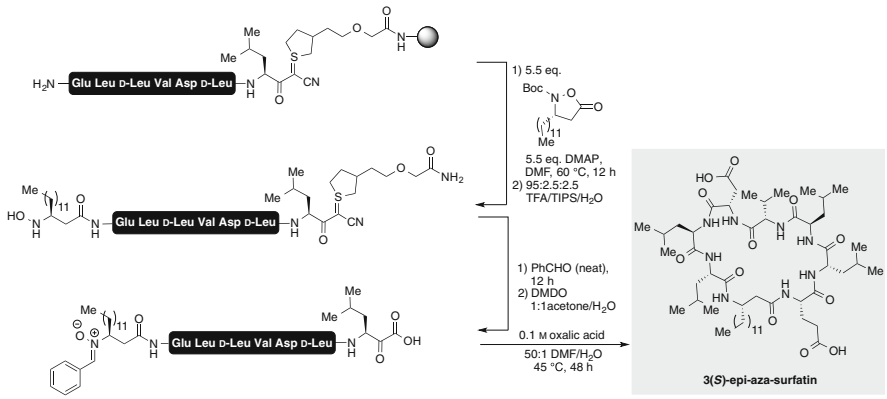


Scheme 31 Synthesis of SUMO3 by a three-segment sequential ligation

UFM1 (30–83) in 3:1 DMSO:H_2O gave the depsiprotein UFM1 (2–83) in 50% yield after purification by HPLC. The *O* to *N* acyl shift gave the final UFM1 (2–83) protein in 67% yield. This synthesis demonstrates the practicability of the temporary protection of 5-oxaproline residues required for sequential ligations.

SUMO3

The synthesis of the small ubiquitin-like protein SUMO3 is the first example of protein synthesis using the protected α-ketoacid monomer exemplified in Sect. 2.4 [38, 39]. This was necessary because of the presence of the oxidation sensitive residues Met43 and Cys47, which prevented the use of the oxidation strategy illustrated in Sect. 2.3. The assembly of SUMO3 (2–92) protein was performed following a three-segment, two-ligation strategy outlined in Scheme 31. The first ligation between N-terminal Fmoc-protected segment 2 SUMO3 (30–52) and segment 3 SUMO3 (53–92) in a mixture of 7:3 NMP:H_2O gave the depsiprotein SUMO3 (30–92) in 28% isolated yield. It should be noted that performing the ligation in the commonly used solvent DMSO resulted in significant amounts of byproducts arising from oxidation of cysteine and methionine residues. The Fmoc-deprotection and *O* to *N* acyl shift were performed in a one-pot fashion to give SUMO3 (30–92) in 47% overall yield. Notably, the Fmoc-deprotection had to be



Scheme 32 Synthesis of Epi-aza-surfactin by KAHA ligation

performed under dilute conditions (<1 mM) to avoid trapping of the formed dibenzofulvene by cysteine residues. Because of the basic conditions and the presence of free cysteine residues, the *O* to *N* acyl shift has to be conducted in the presence of a reducing agent (TCEP) to avoid the formation of disulfide adducts. The second ligation between segment 1 SUMO3 (2–29) and segment 2+3 SUMO3 (30–92) proceeded readily within 7 h, despite the increased steric demand of the valine α-ketoacid, and gave the depsiprotein SUMO3 (2–92) in 23% yield. The final *O* to *N* acyl shift delivered SUMO3 (2–92) in 87% yield. Remarkably, it was demonstrated in subsequent studies that the two introduced homoserine mutations (Q30T§ and S53T§) do not affect the bioactivity of the synthetic SUMO3 protein. Biochemical assays demonstrated that the synthetic SUMO3 protein is readily recognized by the SUMOylation enzymes and transferred on a substrate protein.

4.3 Peptide Macrocycles

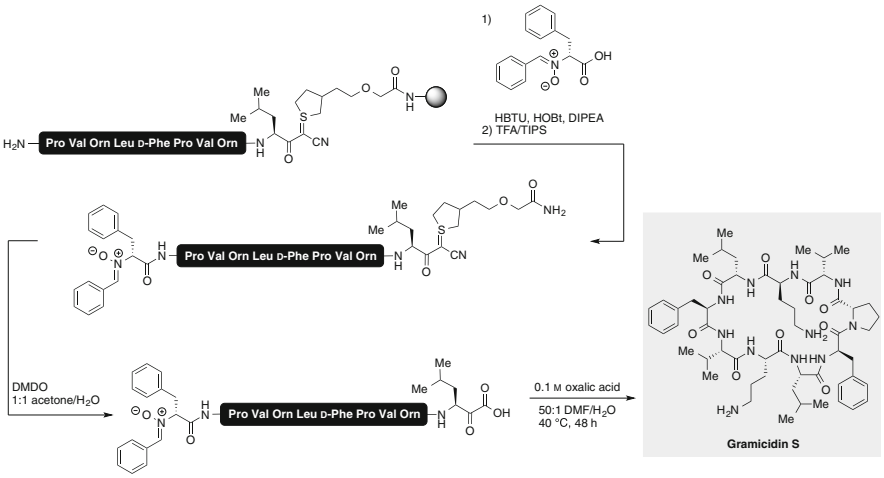
4.3.1 Macrocyclizations with Free Hydroxylamines: Synthesis of Epi-Aza-Surfactin

Surfactin is a principle member of the lipopeptide family and a powerful antibiotic. It features an ester linked β-hydroxy fatty acid in the macrocycle. We were interested in the epi-aza-isomer of surfactin and envisioned a cyclization of the peptide at the fatty acid residue using our isoxazolidin-5-one monomers (see Sect. 3.2.4) [46]. As depicted in Scheme 32, the linear peptide was assembled by Fmoc SPPS starting from the sulfur ylide linker (see Sect. 2.3). In the final step we coupled the fatty acid isoxazolidin-5-one monomer in the presence of DMAP onto the resin. After cleavage with TFA we obtained the C-terminal sulfur ylide with N-terminal hydroxylamine. Since the free hydroxylamine is not stable during the oxidation of the sulfur ylide to the α-ketoacid, it was converted into the benzylidene

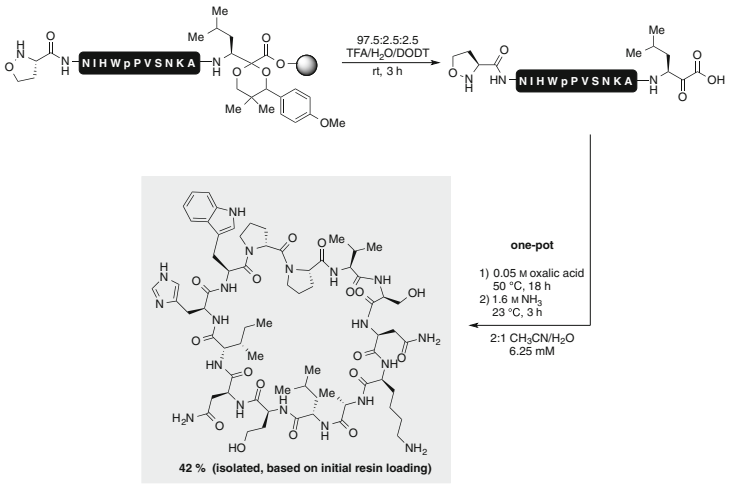
nitrone with benzaldehyde. Subsequent oxidation with DMDO afforded the peptide α-ketoacid with an N-terminal benzylidene nitrone. This linear precursor peptide was warmed to 45°C for 48 h in 50:1 DMF/H_2O in the presence of 0.1 M oxalic acid. Under these conditions, the nitrone protecting group was slowly hydrolyzed and the free hydroxylamine continuously liberated, resulting in low concentration of the active linear precursor and smooth cyclization.

4.3.2 Synthesis of Natural Products with Nitrone-Protected Monomers

For the synthesis of cyclic peptides consisting solely of α-amino acids, we have developed a method based on the incorporation of nitrone-protected hydroxylamine amino acid monomers (see Sect. 3.2.2) [44]. The linear peptides were assembled by Fmoc SPPS and the terminal nitrone-protected *N*-hydroxy aminoacid was coupled with standard reagents (Scheme 33). Using an anhydrous deprotection cocktail with TFA, the nitrone protecting group remains intact and the sulfur ylide could be oxidized with DMDO to the peptide α-ketoacid. Warming this linear precursor peptide to 40°C in 50:1 DMF/H_2O in the presence of 0.1 M oxalic acid led to hydrolysis of the nitrone protecting group followed by cyclization of the peptide. Using this method, five macrocyclic peptide natural products were prepared in isolated yields of 7–22%.



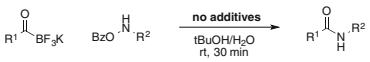
Scheme 33 Synthesis of cyclic peptides with nitrone-protected *N*-hydroxy aminoacid building blocks



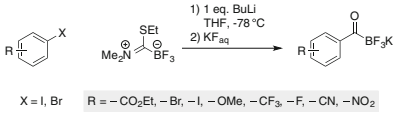
Scheme 34 One-pot preparation of homoserine-containing cyclic peptides and depsipeptides

4.3.3 General and Efficient Synthesis of Macrocyclic Peptides with 5-Oxaproline

The methods for peptide macrocyclization described above required oxidation of the sulfur ylide in solution to afford the C-terminal α-ketoacid. This necessitated an additional manual step and a suitably protected hydroxylamine. To reduce the number of manual steps and to facilitate the synthesis process, we utilized both protected α-ketoacid resins (see Sect. 2.4) and 5-oxaproline (see Sect. 3.3.3) into an efficient method for peptide cyclization. Using this approach, we prepared a library of 24 cyclic peptides (8–20 residues) by parallel peptide synthesis. Starting from protected α-ketoacid resins, the linear peptide, including the N-terminal 5-oxaproline residue, is assembled by automated Fmoc SPPS as shown in Scheme 34. After cleavage with TFA, the crude linear peptides are cyclized for 18 h at 50°C followed by a one-pot addition of ammonia to convert the cyclic depsipeptides into cyclic peptides. If the *O* to *N* acyl shift is not performed, the cyclic depsipeptides can be isolated. These conditions proved to be general for all 24 cyclic peptides synthesized. As a comparison, we prepared the same library by cyclization of sidechain-protected linear peptides in solution with HATU. We found that the KAHA cyclization leads to cyclic peptides of a considerable higher crude purity (54% for KAHA vs 33% for standard cyclization). The KAHA cyclization approach also needs significantly less time for the library synthesis and opens the possibility of preliminary biological screenings without purification.



Scheme 35 Rapid amid bond forming ligation between a potassium acyl trifluoroborate (KAT) and an OBz hydroxylamine



Scheme 36 One-step preparation of potassium acyltrifluoroborates

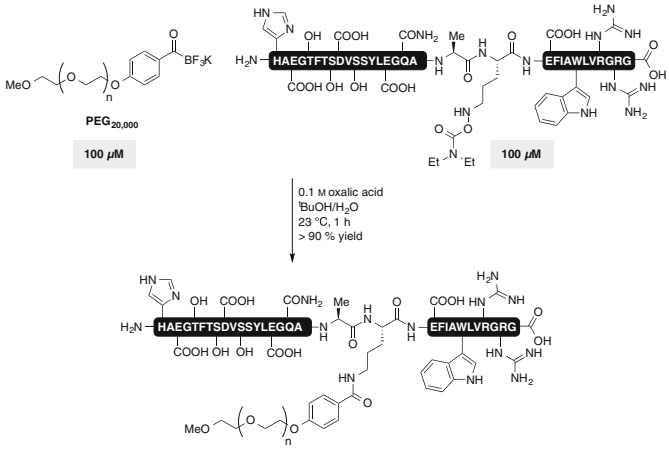
5 Potassium Acyltrifluoroborate (KAT) Ligation

In addition to α-ketoacids, we have recently identified potassium acyl trifluoroborates (KATs) as a new class of reaction partners which react chemoselectively with hydroxylamines to afford amides (Scheme 35) [56, 57]. The reaction is extremely fast under aqueous conditions, with a second-order rate constant of 20 $M^{-1}s^{-1}$. The high rate constant makes this reaction particularly interesting as a tool for bioconjugation with precious starting materials where a 1:1 stoichiometry is desirable.

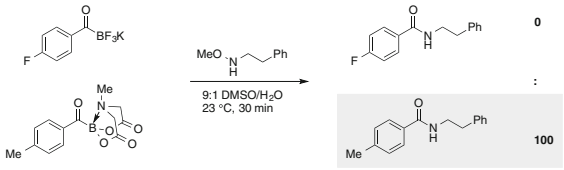
Unfortunately, poor synthetic access has limited the application and further exploration of acyl trifluoroborates. To overcome this problem, we have recently reported a reagent for the one-step conversion of aryl- and heteroarylhalides into acyltrifluoroborates as shown in Scheme 36 [58, 59]. Adding butyl lithium to a mixture of the thioformamide-derived reagent and an arylhalide followed by quenching the reaction with aqueous KF allows the preparation of a variety of substituted aromatic acyl trifluoroborates. Sensitive functional groups such as esters, nitriles, and nitro groups are also tolerated.

The KAT ligation proceeds fast at room temperature even under diluted conditions (100 μM) with a 1:1 stoichiometry of hydroxylamine and KAT (Scheme 37). This was, for example, demonstrated for the conjugation of an unprotected 31-residue peptide with a PEG 20,000 reagent as shown in Scheme 37 [60].

Replacing the fluorine of potassium acyl trifluoroborates with suitable ligands can modulate the reactivity of acyl boron compounds. MIDA (*N*-methyliminodiacetyl) boronates, which can be prepared in one step starting from acyl trifluoroborates, react chemoselectively and fast with the very stable *O*-Me substituted hydroxylamines, which have been shown to be unreactive towards α-ketoacids and acyl trifluoroborates (Scheme 38) [61].



Scheme 37 Chemoselective PEGylation of a 31-residue peptide by a $PEG_{20,000}$-functionalized KAT



Scheme 38 Chemoselective reaction of a MIDA boronate with an OMe substituted hydroxylamine in the presence of 1 equiv. of KAT

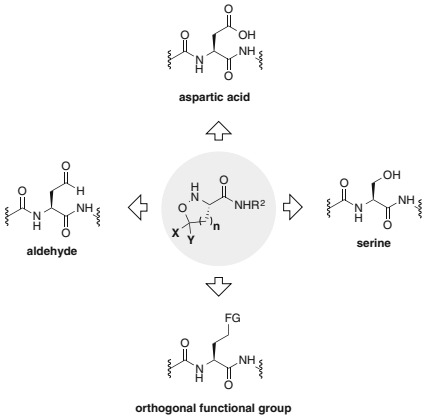
6 Outlook

Despite the success of the KAHA ligation in the total synthesis of proteins and the synthesis of cyclic peptides, there is still space for further improvement of the existing method. The following sections give a perspective on the possible future development of the KAHA ligation.

6.1 Development of New Cyclic Hydroxylamines

One of the current limitations of the KAHA ligation with 5-oxaproline lies in the formation of an unnatural homoserine residue. Although we have shown that the biological activity of proteins with several homoserine mutations remains unaltered, it would be desirable to develop cyclic monomers that would form natural amino acids such as serine or aspartic acid by changing the ring-size or

Scheme 39 Possible residues at the ligation site with new cyclic hydroxylamines



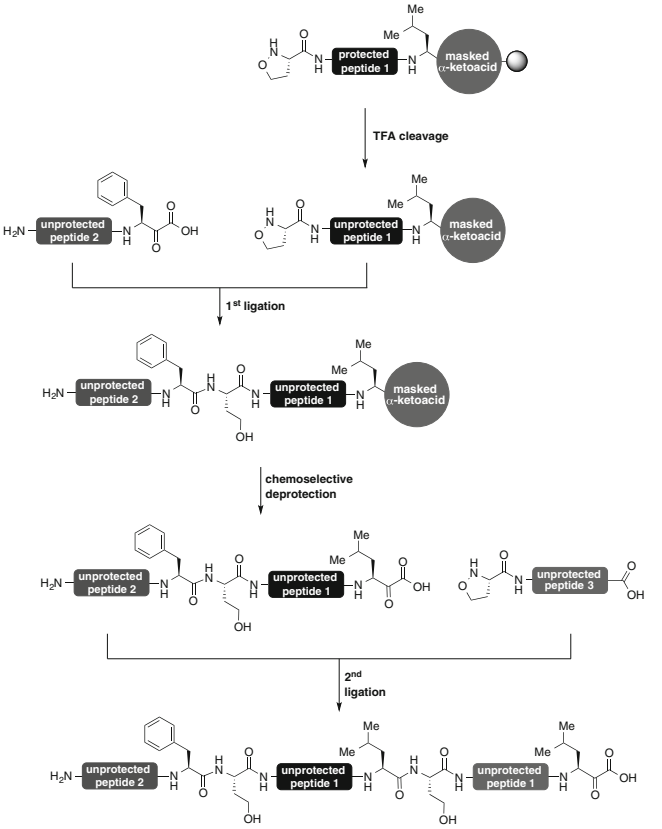
the substituents at the ring as shown in Scheme 39. Besides creating natural amino acids residues, it would be highly desirable to introduce selectively functional groups with orthogonal reactivity such as aldehydes at the ligation site. This would open an avenue for site-selective conjugation reactions.

6.2 *Orthogonally Protected α-Ketoacids*

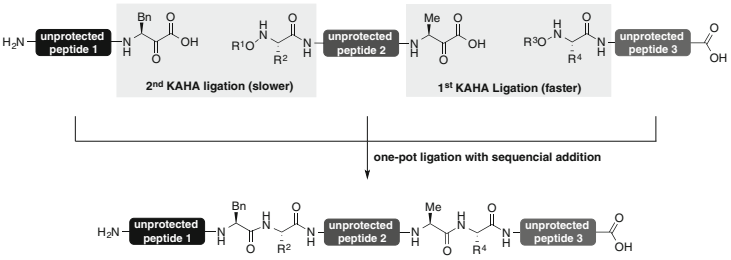
The synthesis of large proteins typically involves sequential segment ligations. To achieve this, either the N-terminus (hydroxylamine) or the C-terminus (α-ketoacid) has to be protected. As described in Sect. 4.2.3, Fmoc-5-oxaproline can be used as a temporarily protected hydroxylamine at the N-terminus, allowing sequential ligations in the C-to-N direction. For the C-terminus the sulfur ylide can be regarded as a temporary protecting group, which can be removed under oxidative conditions. This oxidation, however, is not compatible with cysteine, methionine, and tryptophan. An alternative strategy to protect orthogonally the α-ketoacid is therefore desired to increase further the flexibility in possible ligation strategies (Scheme 40).

6.3 *Kinetically Controlled Ligations*

A notorious problem of protein synthesis by ligation in general is the loss of material with each purification step by HPLC. To increase the overall yield it is therefore highly desirable to avoid any unnecessary purification steps. Kinetically controlled KAHA ligations (Scheme 41) would allow a second subsequent ligation in the same vial and without prior purification. This could be realized by using two distinctively-substituted hydroxylamines differing in their reactivity towards α-ketoacids. The first rapid KAHA ligation would occur between a peptide with a



Scheme 40 Multi-segment ligations with orthogonally protected α-ketoacids



Scheme 41 Kinetically controlled KAHA ligations

very reactive N-terminal hydroxylamine and a bifunctional peptide with C-terminal α-ketoacid and a less reactive N-terminal hydroxylamine. Subsequent addition of another peptide with C-terminal α-ketoacid would lead to a second, slower, KAHA ligation, affording a large unprotected peptide consisting of three peptide fragments without the need for intermediate purification steps.

6.4 Combining KAHA Ligation and Native Chemical Ligation

The native chemical ligation has proved to be a very powerful method in a great number of examples. Despite its undoubted success, the synthesis of peptide thioesters and the low natural abundance of cysteine remain the bottleneck. By developing conditions that allow the orthogonal use of the KAHA ligation and the NCL, the individual strengths of each ligation method could be combined and individual drawbacks could be reduced. This orthogonal combination of two powerful ligation methods would open an avenue for chemical protein synthesis and bioconjugation.

6.5 Summary

In summary, the KAHA ligation is maturing into a powerful and versatile companion to native chemical ligation for chemical protein synthesis. The initially esoteric reaction partners – C-terminal peptide α-ketoacids and N-terminal peptide hydroxylamines – can now be accessed by simple, traceless methods which are fully compatible with standard practices and reagents for Fmoc-SPPS. Further developments, including the design of new cyclic hydroxylamine monomers for the incorporation of natural and unnatural residues into the ligation site, improved methods for multi-segment ligation, and combination with NCL promise that the next few years will see further growth of this method into a user-friendly approach to the preparation of small and medium-sized proteins. The unique reaction mechanism has inspired new variants, including the KAT ligation, which promise to simplify further the coupling of large segments and eventually folded protein domains.

References

1. Kent SBH (2009) Chem Soc Rev 38:338–351
2. Merrifield RB (1963) J Am Chem Soc 85:2149–2154
3. Brocchieri L, Karlin S (2005) Nucleic Acids Res 33:3390–3400
4. Hackenberger CPR, Schwarzer D (2008) Angew Chem Int Ed 47:10030–10074
5. Hackenberger CPR, Schwarzer D (2008) Angew Chem 120:10182–10228
6. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Science 266:776–779
7. Wieland T, Bokelmann E, Bauer L, Lang HU, Lau H (1953) Justus Liebigs Ann Chem 583:129–149
8. Nilsson BL, Soellner MB, Raines RT (2005) Annu Rev Biophys Biomol Struct 34:91–118
9. Mende F, Seitz O (2011) Angew Chem Int Ed 50:1232–1240
10. Mende F, Seitz O (2011) Angew Chem 123:1266–1274
11. 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 50:7645–7649

12. Fang G-M, Li Y-M, Shen F, Huang Y-C, Li J-B, Lin Y, Cui H-K, Liu L (2012) Angew Chem 124:10493–10496
13. Blanco-Canosa JB, Dawson PE (2008) Angew Chem Int Ed 47:6851–6855
14. Blanco-Canosa JB, Dawson PE (2008) Angew Chem 120:6957–6961
15. Hojo H, Onuma Y, Akimoto Y, Nakahara Y, Nakahara Y (2007) Tetrahedron Lett 48:25–28
16. Warren JD, Miller JS, Keding SJ, Danishefsky SJ (2004) J Am Chem Soc 126:6576–6578
17. Zhang Y, Xu C, Lam HY, Lee CL, Li X (2013) Proc Natl Acad Sci U S A 110:6657–6662
18. Liu C-F, Tam JP (1994) J Am Chem Soc 116:4149–4153
19. Wong CTT, Li T, Lam HY, Zhang Y, Li X (2014) Front Chem 2:28
20. Bode JW, Fox RM, Baucom KD (2006) Angew Chem Int Ed 45:1248–1252
21. Bode JW, Fox RM, Baucom KD (2006) Angew Chem 118:1270–1274
22. Ju L, Bode JW, Toma T, Fukuyama T (2010) Org Synth 87:218–225
23. Pusterla I, Bode JW (2012) Angew Chem Int Ed 51:513–516
24. Pusterla I, Bode JW (2012) Angew Chem 124:528–531
25. Sanki AK, Talan RS, Sucheck SJ (2009) J Org Chem 74:1886–1896
26. Wucherpfennig TG, Rohrbacher F, Pattabiraman VR, Bode JW (2014) Angew Chem Int Ed. doi:10.1002/anie.201406097
27. Wucherpfennig TG, Rohrbacher F, Pattabiraman VR, Bode JW (2014) Angew Chem. doi:10.1002/ange.201406097
28. Berzelius JJ (1835) Ann Phys (Leipzig) 36:1
29. Waters KL (1947) Chem Rev 41:585–598
30. Miflin BJ, Lea PJ (1977) Annu Rev Plant Physiol 28:299–329
31. Cooper AJL, Ginos JZ, Meister A (1983) Chem Rev 83:321–358
32. Meister A (1951) J Biol Chem 190:269–276
33. Bode JW (2013) In: Gooßen LJ (ed) Inventing reactions. Springer, Berlin/Heidelberg, pp 13–33
34. Rich DH, Bernatowicz MS (1983) J Org Chem 48:1999–2001
35. Ju L, Bode JW (2009) Org Biomol Chem 7:2259–2264
36. Pattabiraman VR, Ogunkoya AO, Bode JW (2012) Angew Chem Int Ed 51:5114–5118
37. Pattabiraman VR, Ogunkoya AO, Bode JW (2012) Angew Chem 124:5204–5208
38. Wucherpfennig TG, Pattabiraman VR, Limberg FRP, Ruiz-Rodríguez J, Bode JW (2014) Angew Chem Int Ed. doi:10.1002/anie.201407014
39. Wucherpfennig TG, Pattabiraman VR, Limberg FRP, Ruiz-Rodríguez J, Bode JW (2014) Angew Chem. doi:10.1002/ange.201407014
40. Wu J, Ruiz-Rodríguez J, Comstock JM, Dong J, Bode JW (2011) Chem Sci 2:1976–1979
41. Medina SI, Wu J, Bode JW (2010) Org Biomol Chem 8:3405–3417
42. Tokuyama H, Kuboyama T, Amano A, Yamashita T, Fukuyama T (2000) Synthesis 9:1299–1304
43. Fukuzumi T, Bode JW (2009) J Am Chem Soc 131:2009–3865
44. Fukuzumi T, Ju L, Bode JW (2012) Org Biomol Chem 10:5837–5844
45. Juarez-Garcia ME, Yu S, Bode JW (2010) Tetrahedron 66:4841–4853
46. Huang Y-L, Frey R, Juarez-Garcia ME, Bode JW (2012) Heterocycles 84:1179–1191
47. Gambarjan S (1909) Ber Dtsch Chem Ges 42:4003–4013
48. Biloski AJ, Ganem B (1983) Synthesis 7:537–538
49. Wang QX, King J, Phanstiel O IV (1997) J Org Chem 62:8104–8108
50. Arora JS, Kaur N, Phanstiel O IV (2008) J Org Chem 73:6182–6186
51. Kumar S, Sharma R, Garcia M, Kamel J, McCarty C, Muth A, Phanstiel O IV (2012) J Org Chem 77:10835–10845
52. Vasella A, Voeffray R (1981) J Chem Soc Chem Commun 97–98
53. Carpino LA, Krause E, Sferdean CD, Schümann M, Fabian H, Bienert M, Beyermann M (2004) Tetrahedron Lett 45:7519–7523
54. Ogunkoya AO, Pattabiraman VR, Bode JW (2012) Angew Chem Int Ed 51:9693–9697
55. Ogunkoya AO, Pattabiraman VR, Bode JW (2012) Angew Chem 124:9831–9835

56. Dumas AM, Molander GA, Bode JW (2012) Angew Chem Int Ed 51:5683–5686
57. Dumas AM, Molander GA, Bode JW (2012) Angew Chem 124:5781–5784
58. Erős G, Kushida Y, Bode JW (2014) Angew Chem Int Ed 53:7604–7607
59. Erős G, Kushida Y, Bode JW (2014) Angew Chem 126:7734–7737
60. Noda H, Erős G, Bode JW (2014) J Am Chem Soc 136:5611–5614
61. Noda H, Bode JW (2014) Chem Sci 5:4328–4332

Top Curr Chem (2015) 363: 33–56
DOI: 10.1007/128_2014_586
© Springer International Publishing Switzerland 2014
Published online: 3 December 2014

Chemical Synthesis of Proteins Using *N*-Sulfanylethylanilide Peptides, Based on N–S Acyl Transfer Chemistry

Akira Otaka, Kohei Sato, and Akira Shigenaga

Abstract Native chemical ligation (NCL), which features the use of peptide thioesters, is among the most reliable ligation protocols in chemical protein synthesis. Thioesters have conventionally been synthesized using *tert*-butyloxycarbonyl (Boc)-based solid-phase peptide synthesis (SPPS); however, the increasing use of 9-fluorenylmethyloxycarbonyl (Fmoc) SPPS requires an efficient preparative protocol for thioesters which is fully compatible with Fmoc chemistry. We have addressed this issue by mimicking the naturally occurring thioester-forming step seen in intein-mediated protein splicing of the intein–extein system, using an appropriate chemical device to induce N–S acyl transfer reaction, avoiding the problems associated with Fmoc strategies. We have developed *N*-sulfanylethylanilide (SEAlide) peptides, which can be synthesized by standard Fmoc SPPS and converted to the corresponding thioesters through treatment under acidic conditions. Extensive examination of SEAlide peptides showed that the amide-type SEAlide peptides can be directly and efficiently involved in NCL via thioester species in the presence of phosphate salts, even under neutral conditions. The presence or absence of phosphate salts provided kinetically controllable chemoselectivity in NCL for SEAlide peptides. This allowed SEAlide peptides to be used in both one-pot/N–to–C-directed sequential NCL under kinetically controlled conditions, and the convergent coupling of large peptide fragments, which facilitated the chemical synthesis of proteins over about 100 residues. The use of SEAlide peptides, enabling sequential NCL operated under kinetically controlled conditions, and the convergent coupling, were used for the total chemical synthesis of a 162-residue monoglycosylated GM2-activator protein (GM2AP) analog.

A. Otaka (✉), K. Sato, and A. Shigenaga
Institute of Health Biosciences and Graduate School of Pharmaceutical Sciences, Tokushima University, Shomachi, Tokushima 770-8505, Japan
e-mail: aotaka@tokushima-u.ac.jp

Keywords Kinetically controlled NCL · N−S Acyl transfer · Native chemical ligation · *N*-Sulfanylethylanilide peptide · One-pot/sequential NCL · Protein chemical synthesis · Thioester

Contents

Abbreviations

Ac	Acetyl
Acm	Acetamidemethyl
Ar	Aryl
Boc	*tert*-Butoxycarbonyl
Bu	Butyl
DMF	Dimethylformamide
Fmoc	9-Fluorenylmethyloxycarbonyl
GM2	Ganglioside GM2
GM2AP	GM2-activator protein
GM3	Ganglioside GM3
Gn	Guanidine
hANP	Human atrial natriuretic peptide
HEPPS	3-[4-(2-Hydroxyethyl)piperazin-1-yl]propane-1-sulfonic acid
HexA	β-Hexosaminidase A
HSPro	Sulfanylproline
KCL	Kinetically controlled NCL
MBom	4 Methoxybenzyloxymethyl
Me	Methyl
MPAA	(4-Carboxymethyl)thiophenol
NCL	Native chemical ligation

OTf	Trifluoromethanesulfonate
Ph	Phenyl
rt	Room temperature
SEAlide	*N*-Acyl-*N*-sulfanylethylaniline
SPPS	Solid-phase peptide synthesis
t-Bu	*tert*-Butyl
TCEP	Tris(2-carboxyethyl)phosphine (TCEP)
TFA	Trifluoroacetic acid
Thz	1,3-Thiazolidine-4-carbonyl
Tr	Triphenylmethyl (trityl)

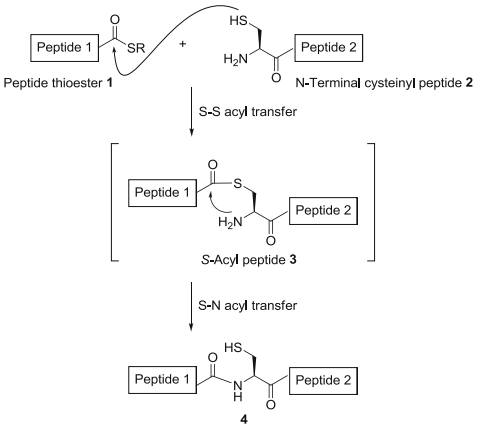
One letter or three letters abbreviations for amino acids are used as follows:

A (Ala)	Alanine
C (Cys)	Cysteine
D (Asp)	Aspartic acid
E (Glu)	Glutamic acid
F (Phe)	Phenylalanine
G (Gly)	Glycine
H (His)	Histidine
I (Ile)	Isoleucine
K (Lys)	Lysine
L (Leu)	Leucine
M (Met)	Methionine
N (Asn)	Asparagine
P (Pro)	Proline
Q (Gln)	Glutamine
R (Arg)	Arginine
S (Ser)	Serine
T (Thr)	Threonine
V (Val)	Valine
W (Trp)	Tryptophan
Y (Tyr)	Tyrosine

1 Introduction

The chemical synthesis of proteins up to about 150 residues is now achievable because of the development of fragment ligation protocols. Among such ligation protocols, native chemical ligation (NCL), developed by Kent and co-workers, has been receiving increasing attention as an indispensable synthetic platform applicable to a wide variety of proteins (Scheme 1) [1–5].

Scheme 1 NCL reaction mechanism of NCL



The most striking characteristic of NCL is that the chemoselective reaction of peptide thioesters **1** with *N*-terminal cysteinyl peptides **2** occurs between the thioester moiety and the sulfanyl group of the cysteine to afford native peptides/proteins **4** via intermediary *S*-acyl peptides **3**. The NCL protocol, which is widely used in protein chemistry, requires thioesters **1** as essential synthetic intermediates, and therefore a broad range of synthetic procedures for thioesters have been investigated. Originally, thioesters were prepared using Boc-based SPPS of thioester-linked C-terminal amino acids [6–10]; however, the increasing popularity of Fmoc SPPS for peptide synthesis has prompted the development of Fmoc SPPS-compatible synthetic methods for thioesters. A serious limitation of the use of Fmoc SPPS is that the thioester linkage for attachment of the C-terminal amino acids on the solid support is highly susceptible to the base treatment required for Fmoc removal, resulting in side reactions such as decomposition of the thioester linkage and epimerization of the C-terminal amino acids (Fig. 1).

In this context, considerable efforts have been made by many research groups, including ours, to develop methodologies for the preparation of peptide thioesters which are fully compatible with Fmoc protocols [11]. For literature about Fmoc-based synthesis of peptide thioesters using safety-catch type linker and related system, see [12–22]; for literature about Fmoc-based synthesis of peptide thioesters using O–S acyl transfer reaction, see [23–27]; for literature about Fmoc-based synthesis of peptide thioesters using miscellaneous protocols, see [28–33]. In this account, we present our work on this important issue. We have developed an artificial "chemical device" which facilitates the Fmoc-based synthesis of

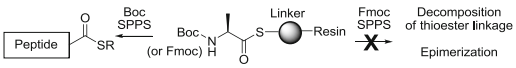


Fig. 1 Thioester synthesis by Boc SPPS, and problems encountered in Fmoc-based thioester synthesis

thioesters. The device has been proven to function as a useful platform for one-pot/N−to−C-directed sequential NCL using more than one thioester fragment, and has been used in chemical protein synthesis.

2 Development of N−S Acyl Transfer Device for Thioester Synthesis with Practical Applications in Peptide/Protein Synthesis

2.1 Naturally Occurring Thioester Formation: Intein−Extein System

One rational way of preparing thioesters using Fmoc protocols is to mimic the N−S acyl transfer step seen in intein-mediated protein splicing of the intein−extein system [34–39] using an appropriate chemical device (Fig. 2).

The use of a chemical device to induce N−S acyl transfer enables Fmoc-based elongation of peptide chains to be performed on C-terminal amino acids linked on resins via amide bonds, followed by N−S acyl transfer; this process should be free from reported side reactions (for literature about Fmoc-based synthesis of peptide thioesters using N–S acyl transfer reaction except for the use of *N*-acyl-*N*-sulfanylethyl type linker, see [40–62]). The intein−extein system functions as a self-editing machine for proteins in low organisms (Scheme 2). Extein fragments (N- and C-extein) split by intein are ligated to give a mature extein sequence **10** by the action of intein, and then the intein itself is cut from the precursor protein **5**. In intein-mediated protein splicing, sequentially occurring N−S, S−S, and S−N acyl transfers (**5** to **6**, **6** to **7**, and **9** to **10**, respectively) are key steps. Chemical bases involved in such acyl transfer steps have provided useful insights for developing a wide variety of chemical methodologies for peptide/protein chemistry. In particular, because the first N−S acyl transfer serves as a naturally occurring thioester formation step, mimicking this step using an artificial chemical device provides a method for solving the problems encountered in Fmoc-based thioester synthesis.

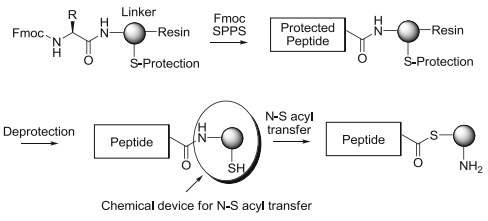
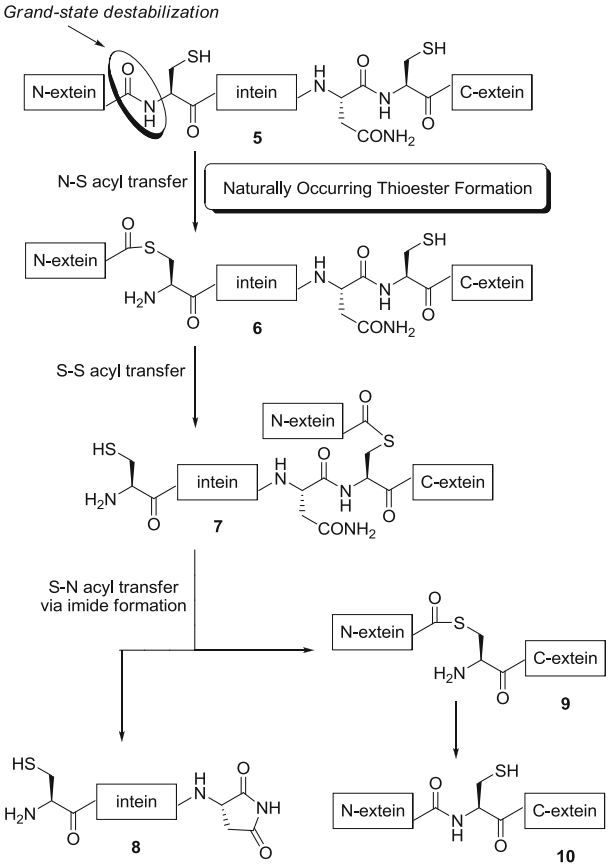


Fig. 2 Concept of peptide thioester synthesis mediated by N–S acyl transfer



Scheme 2 Mechanism of intein-mediated protein splicing

2.2 *Chemistry Seen in Naturally Occurring Thioester Formation*

Ground-state destabilization of the peptide bond has been reported to be responsible for the N−S acyl transfer involved in intein-mediating protein splicing, as shown in Fig. 3. An elegant ^{13}C and ^{15}N NMR spectroscopic study of the amide bond located at the N-extein−intein junction (Xaa-Cys) showed that the amide bond was twisted, its planarity disappeared, and it became susceptible to nucleophilic attack by the sulfanyl group of the cysteine residue located at the junction [39, 63]. Twisting of the amide bond followed by nucleophilic attack by the neighboring functional group leads to naturally occurring thioester formation. We attempted to develop a methodology for producing peptide thioesters, based on ground-state destabilization of the amide bond, by mimicking the N−S acyl transfer step in the intein−extein system.

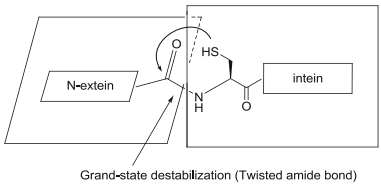


Fig. 3 Ground-state destabilization of peptide bond between *N*-extein and intein

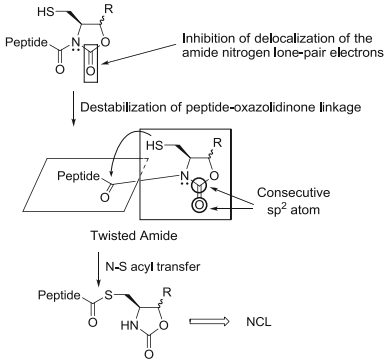


Fig. 4 Synthesis of peptide thioesters using peptidyloxazolidinone system as amide-activating device

2.3 Peptidyloxazolidinone as Thioester Precursor

Various synthetic protocols for the conversion of amide-type peptides synthesized by Fmoc protocols to the corresponding thioester-type peptides have been studied using the N−S acyl transfer reaction. Generally, the stability of the amide bond is attributable to the planar double-bond character resulting from the $n \rightarrow \pi^*$ interaction between the electron-filled non-bonding nitrogen orbital and the anti-bonding carbonyl orbital. As mentioned above, the N−S acyl transfer observed in the intein−extein system is attributable to the disappearance of the planar character of the amide bond involved. On the assumption that inhibition of delocalization of the nitrogen lone pair electrons and appropriate positioning of a sulfanyl group should result in the desired N−S acyl transfer through disappearance of amide bond planarity, we initially used an S-protected-cysteine-derived peptidyloxazolidinone system, in which delocalization of the nitrogen lone pair electrons toward the *exo*-amide linkage is inhibited by the electron-withdrawing ring carbonyl group; such an *exo*-peptide bond is twisted and susceptible to nucleophilic attack by the sulfanyl group, leading to the formation of peptide thioesters (Fig. 4) [42]. Based on this idea, we attempted to develop an Fmoc SPPS-compatible synthetic protocol for a thioester precursor. The use of an oxazolidinone system under Fmoc conditions afforded a peptidyloxazolidinone derivative as a thioester precursor, which was then converted to a peptide thioester under neutral conditions. However, some decomposition of the peptide−oxazolidinone linkage and epimerization of the C-terminal amino acids were seen during peptide chain elongation because of over activation of the *exo*-carbonyl group by the oxazolidinone ring.

2.4 N-*Sulfanylethylaniline Linker as Alternative to Oxazolidinone*

We assumed that the over activation in the oxazolidinone system was attributable to the electron-withdrawing ring carbonyl, as an adjacent two sp^2 atom system, so we thought that the use of a carbon double-bond as a weaker electron-withdrawing sp^2 atom system should result in the development of a chemical device fully compatible with Fmoc chemistry. On the basis of this assumption, we examined the feasibility of using an *N*-sulfanylethylaniline linker as a device for inducing N−S acyl transfer for thioester synthesis (Fig. 5) [64]. Although the less reactive nucleophilic nature of the *N*-alkylaniline moiety of the *N*-sulfanylethylaniline linker **11** caused concern about the difficulty of attaching the C-terminal Fmoc amino acids to the linker, Fmoc amino acids without acid-sensitive protecting groups such as *t*-Bu or Tr were coupled by the reaction of Fmoc amino acid chlorides with sodium anilide **12** to afford the desired *N*-Fmoc amino acyl-*N*-sulfanylethylaniline linker **13** (Scheme 3). The chlorides were prepared by standard treatment of amino acid derivatives with $SOCl_2$. Exploration of solutions to the incompatibility of this $SOCl_2$-mediated protocol with amino acid derivatives with acid-labile protection identified a $POCl_3$−Et_3N system as reagents for coupling with the sodium anilide **12** [65]. Except in the case of His(Tr), the use of $SOCl_2$ or the $POCl_3$−Et_3N system was free from epimerization. In terms of incorporation of a histidine derivative, 4-methoxybenzyloxymethyl-protection (Fmoc-His(MBom)-OH) proved to be suitable [66]. We denote the resulting *N*-acyl-*N*-sulfanylethylaniline derivative **13**, "i.e., the *N*-sulfanylethylanilide," by "SEAlide" in our research.

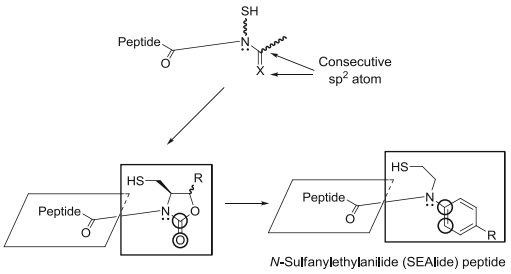
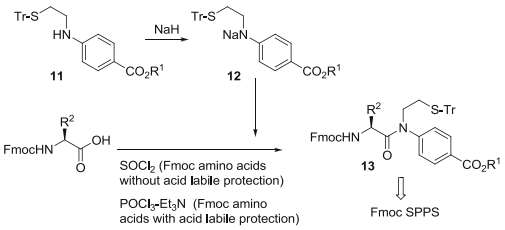


Fig. 5 Activation of peptide bond in peptidyloxazolidinone and SEAlide systems is achieved by presence of consecutive sp^2 atom adjacent to peptide bond



Scheme 3 General scheme for synthesis of *N*-Fmoc amino acyl-*N*-sulfanylethylaniline linker for preparation of SEAlide peptide

2.5 Initial Observation on of SEAlide Peptide

As shown in Fig. 6, standard Fmoc SPPS on the Fmoc amino acid-incorporated SEAlide-linked resin **14** followed by treatment with TFA-based reagent cocktail efficiently gave amide-type SEAlide peptides **15**. During the chain elongation step in Fmoc protocols, decomposition of the anilide linkage and epimerization of the C-terminal amino acids, partly seen in the case of the oxazolidinone system, were not observed. Furthermore, the reaction of the protected peptide on the sulfanylethylaniline linker with a deprotecting agent consisting of TFA and scavengers, under standard deprotecting conditions (room temperature and 1−2 h), scarcely gave thioester-type SEAlide peptides. Preliminary examination of the N−S acyltransfer-mediated conversion of the amide-type SEAlide peptide **15** obtained to thioester-type SEAlide peptide **16** showed that this conversion occurred under acidic conditions such as 4 M HCl in DMF, but apparently not under neutral to slightly basic conditions (Fig. 6) [64]. At this stage, it seemed reasonable to conclude that the amide-type SEAlide peptides, synthesized by Fmoc chemistry, could be converted to the corresponding thioester-type SEAlide peptides only under acidic conditions (4 M HCl in DMF); the resulting thioesters could then be subjected to NCL with cysteinyl peptides to yield ligated peptides, although the acidic conversion would be accompanied by partial epimerization of the chiral C-terminal amino acids.

2.6 Attempted Sequential NCL Using SEAlide Peptides

The chemical synthesis of proteins over 100 amino acid residues generally requires sequential NCL protocols featuring more than one NCL coupling [67–72]. The successful use of SEAlide peptides as thioester precursors would enable the use of SEAlide peptides in sequential NCL protocols involving two or more thioester fragments. Stepwise coupling of thioesters requires N-terminal cysteinyl thioester (or thioester precursor) fragments. The minimal requirement of such fragments for use in sequential NCL is that the cysteinyl residue in the N-terminal cysteinyl thioester reacts intermolecularly with the thioester moiety in the coupling partner,

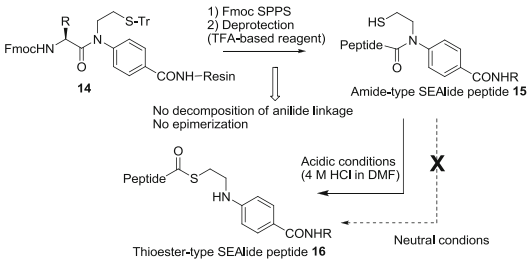
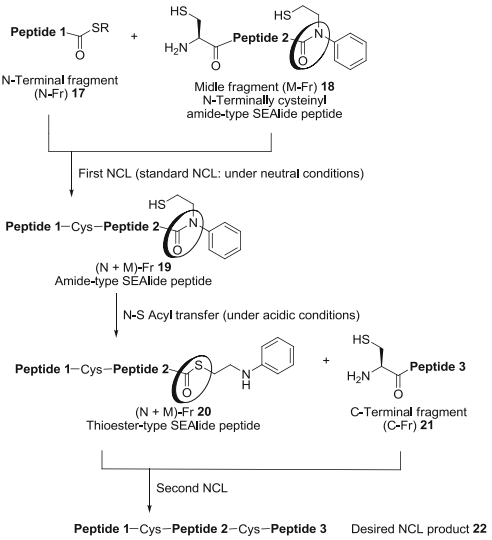


Fig. 6 Synthesis of SEAlide peptide by standard Fmoc chemistry and conversion to corresponding thioester under acidic conditions

Scheme 4 Envisioned N−to−C-directed sequential NCL using SEAlide peptide



but not intramolecularly, to yield a two-fragment-ligated product. Such requirements are achieved by using either kinetically controlled NCL (KCL) [73–75], thioester precursors [71, 72], or N- and/or S-protected N-terminal cysteinyl thioester fragments [67–70]. The first two methods allow us to achieve N−to−C-directed sequential NCL; C−to−N-directed NCL can be performed using N- and/or S-protected thioester fragments. In view of our initial observation that only acidic treatment enabled the conversion of amide-type SEAlide peptides as thioester precursors to the corresponding thioesters, we planned to use the N-terminal cysteinyl SEAlide peptides **18** in N−to−C-directed sequential NCL, as shown in Scheme 4.

We anticipated that the sequential NCL would proceed via the following reactions. (1) The first NCL of thioesters **17** as N-terminal fragments with the cysteinyl SEAlide peptides **18** would afford ligated amide-type SEAlide peptides**19**. (2) The resulting amide linkage in the SEAlide could be converted to the corresponding thioester linkage under acidic conditions. (3) The second NCL, of thioester-type SEAlide peptides **20** with N-terminal cysteinyl peptides **21**, should afford three-fragment ligated peptides **22**. Here, the use of *N*-terminal cysteinyl SEAlide peptides as substituents for C-terminal fragments **21** in step (3) would allow the sequence of reactions to be performed iteratively. Based on these assumptions, we undertook the N−to−C-directed sequential ligation using cysteinyl SEAlide peptides **18**. The first NCL under normal neutral NCL conditions in the presence of phosphate salts proceeded to yield a desired ligated peptide **19**; however, contrary to our initial expectation, a non-negligible amount of cyclic peptide **23**, resulting from intramolecular NCL of the cysteinyl SEAlide peptide **18**, was observed (Fig. 7).

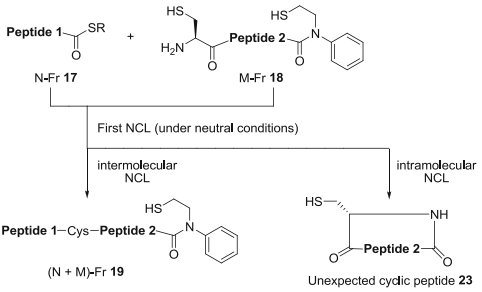


Fig. 7 Unexpected formation of cyclic peptide during NCL of thioester with N-terminal cysteinyl SEAlide peptide under standard NCL conditions

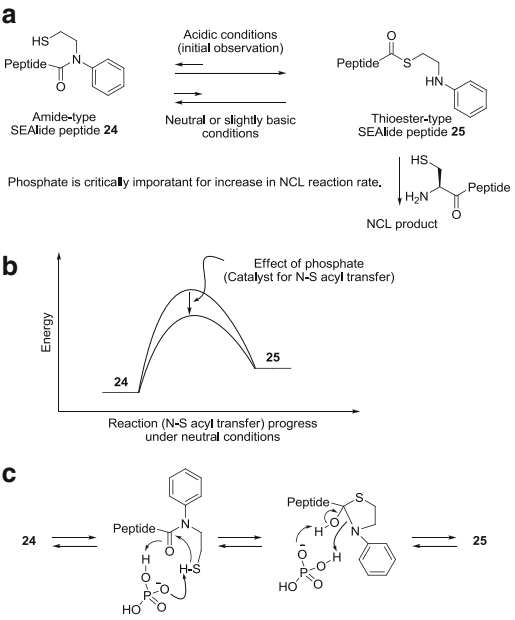


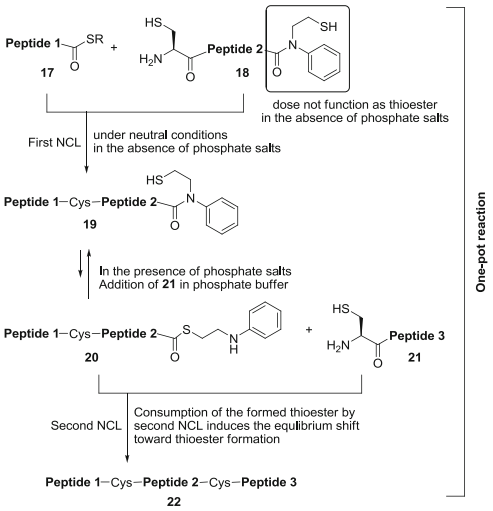
Fig. 8 Effect of phosphate salts on SEAlide peptide. (**a**) Amide-type SEAlide peptide can participate in NCL even under neutral conditions in the presence of phosphate salts. (**b**) Phosphate salts can catalyze N−S acyl transfer. (**c**) Plausible involvement of phosphate salts as an acid−base catalyst

2.7 SEAlide Peptides Function as Thioesters in the Presence of Phosphate Salts

This finding contradicted earlier predictions that amide-type SEAlide peptides could be transformed into "NCL-active" thioesters under acidic conditions. Incubation of the SEAlide peptide without N-terminal cysteine in a neutral buffer did not afford a detectable amount of the corresponding thioester.

We hypothesized that the divergence of the experimental results could be attributable to the presence or absence of an N-terminal cysteinyl moiety in the reaction mixture. The small amount of thioester **25** formed in the equilibrium between amide **24** and thioester **25** under neutral conditions was consumed by NCL with the N-terminal cysteinyl moiety to give the NCL product (Fig. 8a).

Scheme 5 One-pot/N−to−C-directed sequential NCL using SEAlide peptide



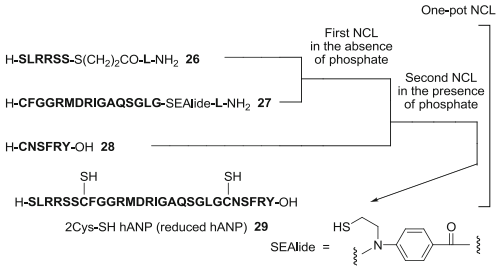
Ligation of model SEAlide peptides with cysteinyl peptides under neutral conditions in the presence of phosphate salts went almost to completion within 12–24 h to yield the desired ligated peptides, without the accompanying epimerized products. It is worth noting that NCL involving SEAlide peptides proceeds efficiently in the presence of acid−base catalysts such as phosphate to yield the NCL product, but not in the absence of phosphate. A plausible explanation for the effect of phosphate salts is that phosphate salts function as acid−base catalysts for proton transfer in the N−S acyl transfer, causing a decrease in the activation energy, as shown in Fig. 8b, c. These properties of SEAlide peptides prompted us to investigate the use of the SEAlide unit in a kinetically controlled manner for a one-pot/N−to−C-directed sequential NCL protocol, different from our initial attempt, as shown in Scheme 5 [76, 77].

The first NCL of thioesters **17** with N-terminal cysteinyl SEAlide peptides **18** in the absence of phosphate salts under neutral conditions should proceed without involvement of the SEAlide moiety in the NCL to yield the ligated amide-type SEAlide peptides **19**. Addition of cysteinyl peptides **21** in phosphate buffer to the reaction mixture should enable the resulting amide-type SEAlide peptides **19** to function as thioesters **20**, to give the desired three-fragment ligated peptides **22** via an N−to−C-directed one-pot reaction. Such a SEAlide-peptide-mediated one-pot/N−to−C-directed sequential NCL operated under kinetically controlled conditions would be very useful in peptide/protein preparations. In the following sections, practical applications of SEAlide peptides to peptide/protein synthesis are described.

2.8 *Synthesis of hANP by One-Pot/N−to−C-Directed Sequential NCL*

The usefulness of SEAlide peptides was first evaluated in the synthesis of hANP by one-pot/N−to−C-directed sequential NCL. The 28-residue hANP, containing one disulfide bond, was divided into three fragments (Scheme 6).

The N−to−C-directed construction of the hANP molecule required the N-terminal cysteinyl thioester corresponding to residues 7–22 as the middle fragment. This middle fragment has a glycine thioester moiety, which shows high reactivity in NCL coupling. KCL [73–75], which was developed by Kent's group, the success of which depends on the large difference between the reactivities of aryl and alkyl thioesters, was one potential option for the one-pot synthesis of hANP. However, there was concern regarding retention of kinetic selectivity, i.e., whether intermolecular NCL of the cysteine residue in the middle fragment with the aryl thioester in the N-terminal fragment proceeded more efficiently than intramolecular NCL, because the glycine alkyl thioester involved in KCL is among the most reactive alkyl thioesters [78]. Insufficient kinetic selection was observed in the KCL protocol for one-pot synthesis of hANP. However, one-pot/sequential NCL using SEAlide peptides as the middle fragment for hANP was successful (Scheme 6). The first NCL of thioester **26** with N-terminal cysteinyl SEAlide peptide **27** in 6 M Gn·HCl-0.2 M HEPPS in the presence of 30 mM TCEP and 30 mM MPAA [79], pH 7.3, at 37 °C, proceeded almost to completion within 3 h, to yield the desired ligated peptide without a detectable amount of the intramolecular NCL product. Under the first NCL conditions, the SEAlide moiety remained unreactive and the cysteinyl residue in the middle fragment reacted preferentially with the alkyl thioester of N-terminal fragment **26** in a kinetically controlled manner. Then addition of the C-terminal fragment **28** in 1 M phosphate buffer to the reaction mixture initiated the second NCL, yielding the desired reduced form of hANP **29** after reaction for 24 h. This experiment proved that SEAlide peptides are useful thioester equivalents, and their reactivities in the NCL protocol can be tuned by the addition of an appropriate buffer salt. We next attempted to perform an unprecedented one-pot/four-segment ligation, using a combination of SEAlide chemistry and the KCL protocol.



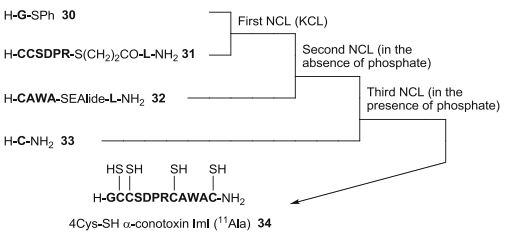
Scheme 6 One-pot/N−to−C-directed sequential NCL for synthesis of reduced hANP

2.9 One-Pot/Four-Segment Ligation

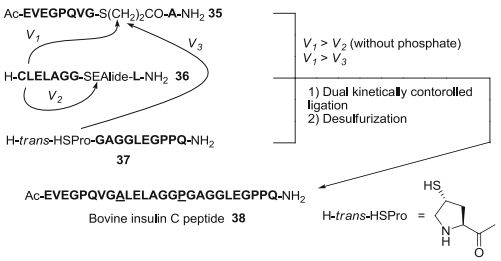
Although the KCL protocol showed insufficient kinetic selectivity in NCL using the highly reactive glycine alkyl thioester, as mentioned above, KCL is still an indispensable one-pot/N−to−C-directed sequential ligation method if appropriate alkyl thioesters are selected. We expected that a combination of KCL and the SEAlide-mediated method would result in an unprecedented one-pot/N−to−C-directed four-segment coupling. This possibility was confirmed by the synthesis of the reduced form of α-conotoxin ImI (^{11}Ala) as a model peptide (Scheme 7). The entire sequence of the conotoxin was divided into four segments. The first NCL of glycyl aryl thioester **30** with leucyl alkyl thioester **31** in the presence of HEPPS buffer proceeded under kinetically controlled conditions to yield the ligated alkyl thioester (KCL conditions). Addition of N-terminal cysteinyl SEAlide peptide **32** to the reaction mixture initiated the second NCL, to afford the desired amide-type SEAlide peptide. This ligation step also operated under kinetically controlled conditions because of the absence of phosphate salts, which are essential catalysts for activation of the SEAlide moiety. The subsequent NCL was started by the addition of cysteine amide **33** in phosphate buffer to the reaction mixture, allowing the amide-type SEAlide unit to function as a thioester, yielding 4Cys-SH α-conotoxin ImI (^{11}Ala) **34**. In this scheme, a sequence of three NCLs was conducted in a one-pot manner. To the best of our knowledge [70], this is the first example of a one-pot/N−to−C-directed four-segment sequential ligation.

2.10 Dual Kinetically Controlled Ligation

In a different project from the SEAlide project, we developed protocols for proline site ligation [80–84], including an NCL-like reaction using HSPro followed by desulfurization. Although the NCL-like reaction of HSPro with thioesters proceeds more slowly than that of cysteine, the ligation of HSPro with thioesters using *trans*-HSPro isomer, not the *cis*-isomer, has been independently reported by our group and Danishefsky's group. In addition, we found that *trans*-HSPro could be involved in ligation with SEAlide peptides in the presence of phosphate salts. On the basis of the successful reaction of HSPro with SEAlide peptides, and the difference between the NCL reactivities of cysteine and HSPro, we anticipated that an ordered



Scheme 7 Unprecedented one-pot/ four-segment ligation for synthesis of α-conotoxin ImI(^{11}Ala). See Scheme 6 for structure of SEAlide in **32**



Scheme 8 Dual kinetically controlled ligation for synthesis of bovine insulin C peptide using a combination of SEAlide peptide and sulfanyl proline

sequence of ligations of three peptide fragments simultaneously present in the same reaction vessel would be possible using a combination of HSPro and SEAlide chemistry [80]. This hypothesis was tested by the synthesis of the 26-residue bovine insulin C peptide, as shown in Scheme 8. Cysteine and *trans*-HSPro were used for ^{9}Ala and ^{16}Pro, respectively. Initially, three peptide fragments (N-terminal fragment (N-Fr **35**), middle Fr (M-Fr **36**), and C-terminal Fr (C-Fr **37**)) were simultaneously dissolved in a ligation buffer without phosphate salts, to perform the kinetically selective ligation of the glycine alkyl thioester in N-Fr **35** with cysteine in M-Fr **36**, not with *trans*-HSPro in C-Fr **37**, to give the desired (N + M)-Fr. Under these reaction conditions, one kinetic selection was achieved because of the difference between the NCL reactivities (alkyl thioester − Cys vs alkyl thioester − HSPro; $V_1 > V_3$). Additionally, the absence of phosphate salts in the ligation buffer guaranteed another kinetic selection based on the non-engagement of the SEAlide moiety in the NCL ($V_1 > V_2$). Because the successful first ligation step affording the (N + M)-Fr required the achievement of two kinetically controlled reactions, this reaction is referred to as dual kinetically controlled NCL. Subsequent NCL of *trans*-HSPro in C-Fr **37** with the SEAlide moiety in (N + M)-Fr was initiated by the addition of phosphate salts to yield the ^{9}Cys, ^{16}HSPro C peptide. The C peptide precursor obtained was successfully converted to the desired bovine insulin C peptide **38** via desulfurization [85].

2.11 Chemical Synthesis of Proteins Using SEAlide Peptides

As mentioned in previous sections, one-pot/N−to−C-directed sequential NCLs were achieved using SEAlide peptides, and SEAlide peptides can be used in ligation with *trans*-HSPro, which is less reactive than standard cysteine. These findings prompted us to use SEAlide peptides in a convergent chemical synthesis of a protein (for recent achievement in chemical synthesis of proteins, see [86–96]). We planned to use SEAlide chemistry in both the one-pot synthesis of a large peptide fragment and the convergent coupling of large peptide fragments [97].

The monoglycosylated GM2-activator protein (GM2AP) analog **48** containing 162 amino acid residues was synthesized using the SEAlide protocol (Fig. 9). Native

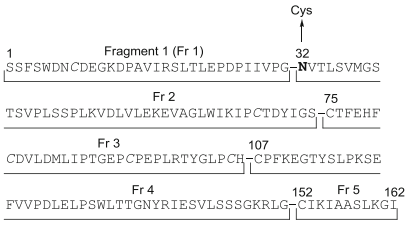
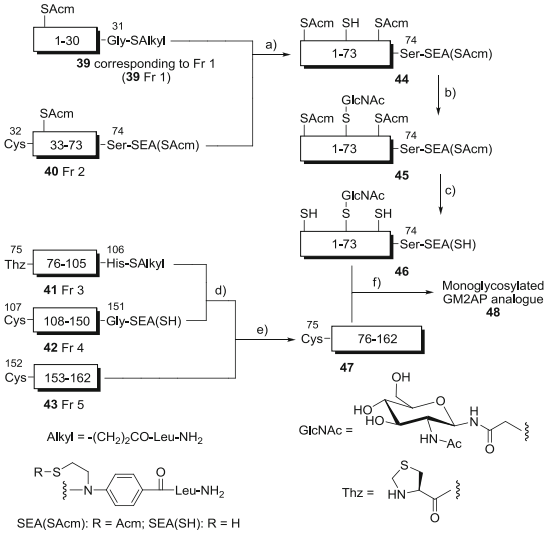


Fig. 9 Amino acid sequence of monoglycosylated GM2AP. Cysteine residues represented by italic characters are other potential ligation sites; however, prolyl-cysteine is not suitable for NCL



Scheme 9 Scheme for convergent synthesis of monoglycosylated GM2AP analog: (**a**) NCL; (**b**) *S*-glycosylation using iodoacetyl-*N*-acetylglucosamine; (**c**) removal of Acm group with AgOTf; (**d**) NCL in absence of phosphate salts; (**e**) NCL in presence of phosphate salts and then opening of thiazolidine ring; (**f**) convergent NCL in presence of phosphate salts and then folding

GM2AP has eight cysteine residues, at positions 8, 68, 75, 81, 94, 105, 107, and 152, and one *N*-glycosylation site at ^{32}Asn [98]. Synthesis of the N-half fragment of the native form needed the preparation of a 67-residue thioester, corresponding to positions 8–74. However, anticipating the difficulty of straightforward SPPS of a fragment containing over 50 residues, we planned to overcome this difficulty by substitution of cysteine for ^{32}Asn at the *N*-glycosylation site. Cysteine substituent was used both for the NCL site and the *S*-alkylation-mediated incorporation of the *N*-acetyl glucosamine moiety with iodoacetyl-*N*-acetyl glucosamine. A convergent synthetic strategy based on cysteine replacement in the GM2AP analog is shown in Scheme 9. It was planned to perform the convergent coupling between the 74-residue N-half and 88-residue C-half fragments (**46** and **47**, respectively). Standard NCL of the ^{8}Cys(Acm)-containing alkyl thioester **39** with N-terminal cysteinyl SEAlide peptide **40**, with *S*-Acm protection on

the SEAlide moiety, was conducted to give the desired 74-residue non-glycosylated N-half fragment precursor **44**, with one sulfanyl group only for *S*-alkylation at position 32. The *S*-alkylation-mediated incorporation of an *N*-acetyl glucosamine unit with iodoacetyl-*N*-acetyl glucosamine afforded *S*-Acm-protected monoglycosylated SEAlide peptide **45**. The resulting protected peptide was subjected to AgOTf-mediated removal of the Acm group in TFA, to yield the monoglycosylated SEAlide N-half fragment **46** needed for the following convergent coupling between N- and C-half fragments. The C-half fragment **47** was constructed by one-pot/N−to−C-directed sequential NCL using SEAlide peptide**42**. NCL of the N-terminal Thz alkyl thioester **41** with SEAlide peptide **42** in 6 M Gn·HCl−0.1 M HEPPS, 50 mM TCEP, and 30 mM MPAA, pH 7.0, at 37 °C, in the absence of phosphate salts, proceeded under kinetically controlled conditions to yield the desired amide-type SEAlide peptide (**41** + **42**) with high chemoselectivity. Subsequent addition of N-terminal cysteine peptide **43** in 0.4 M phosphate buffer to the reaction mixture allowed the SEAlide moiety to function as a thioester, to afford the 88-residue Thz peptide (**41** + **42** + **43**). Ring opening of the 1,3-thiazolidine ring in Thz was achieved by the addition of NH_2OMe·HCl to the reaction, to afford the requisite C-half fragment **47** for the convergent synthesis. One-pot/N−to−C-directed sequential NCL using SEAlide peptide **42**, followed by the ring opening, enabled the efficient one-pot preparation of the 88-residue peptide fragment **47**. Convergent coupling of the N-half SEAlide peptide fragment **46** with the C-half N-terminal cysteine fragment **47** in the presence of phosphate salts, as essential additive, went almost to completion within 24 h to yield the desired 8Cys-SH GM2AP. Exposure of the resulting reduced form of GM2AP to a folding buffer in the presence of reduced and oxidized forms of glutathione gave the folded monoglycosylated GM2AP analog **48**, which assisted the hydrolysis of GM2 to GM3 by HexA with comparable potency to a recombinant GM2AP protein. The successful synthesis of the GM2AP analog involving use of a SEAlide moiety in one-pot/sequential and convergent couplings clearly proves that SEAlide peptides are of great use in peptide/protein synthetic chemistry.

Although tremendous utility of the *N*-sulfanylethyl moiety on *N*-peptidyl units have been disclosed by application of SEAlide peptides to chemical protein synthesis, similar observations have also been independently reported by several research groups, including the use of bis(2-sulfanylethyl)amido (SEA) [99–103], bis(2-mercaptoethyl)amido (BMEA) [104], thioethylalkylamido (TEA) [105], and α-methylcysteine [106]. In comparison with this research, it is worth noting that the function of the SEAlide unit can be controlled just by the presence or absence of phosphate salts as essential additive.

3 Summary, Conclusions, and Outlook

Attempts to develop efficient synthetic methodologies for peptide thioesters based on the N−S acyl transfer chemistry seen in the intein−extein system initially led to the *N*-peptidyloxazolidinone system. The exploration of solutions to the problems

encountered in the use of the oxazolidinone system in Fmoc chemistry generated the idea that *N*-peptidyl-*N*-sulfanylethylanilines (SEAlide peptides) could be superior alternatives to the oxazolidinone. An initial survey of the chemical nature of SEAlide peptides indicated that amide-type SEAlide peptides could be routinely synthesized by Fmoc protocols and converted to the corresponding thioesters only under acidic conditions. These early findings led to research on the use of SEAlide peptides in sequential NCL using the SEAlide moiety as protected thioesters. Contrary to our expectations, this attempt revealed an unexpected property of SEAlide peptides, namely that SEAlide peptides, which apparently seem to be inactive under NCL conditions, can function as peptide thioesters in the presence of phosphate salts as an acid−base catalyst. Tuning of the NCL reactivity using phosphate salts enabled the use of SEAlide peptides in one-pot/N−to−C-directed sequential ligation, in which KCL played a key role. One-pot/sequential ligation using SEAlide peptides was successfully applied to the hANP synthesis and α-conotoxin ImI (^{11}Ala) synthesis by combining with Kent's KCL protocol. One-pot synthesis of bovine insulin C peptide by dual kinetic NCL using HSPro was also achieved. Furthermore, the use of SEAlide peptides in both the one-pot synthesis of an 88-residue peptide fragment and convergent coupling between 74- and 88-residue fragments achieved the chemical synthesis of a 162-residue monoglycosylated GM2AP analog. We believe that the developed SEAlide peptides can be of great importance in protein synthetic chemistry. The incorporation of a functional group on the aromatic ring of the SEAlide moiety is expected to modulate the reactivity of the N−S acyl transfer. Our preliminary examination of the modification of the SEAlide structure has been revealing that "modified SEAlides" should have great potential as alternative N−S acyl transfer devices. The development of other innovative applications of SEAlide moieties in chemical biology is under way, and further applications of SEAlide peptides in protein chemical synthesis will be presented in due course.

References

1. Dawson P, Muir T, Clark-Lewis I, Kent S (1994) Synthesis of proteins by native chemical ligation. Science 266(5186):776–779. doi:10.1126/science.7973629
2. Dawson PE, Kent SBH (2000) Synthesis of native proteins by chemical ligation. Annu Rev Biochem 69:923–960. doi:10.1146/annurev.biochem.69.1.923
3. Kent S (2004) Novel forms of chemical protein diversity – in nature and in the laboratory. Curr Opin Biotech 15(6):607–614. doi:10.1016/j.copbio.2004.10.003
4. Hackenberger CPR, Schwarzer D (2008) Chemoselective ligation and modification strategies for peptides and proteins. Angew Chem Int Ed 47(52):10030–10074. doi:10.1002/anie.200801313
5. Kent SBH (2009) Total chemical synthesis of proteins. Chem Soc Rev 38(2):338–351. doi:10.1039/b700141j
6. Hojo H, Aimoto S (1991) Polypeptide synthesis using the S-alkyl thioester of a partially protected peptide segment. Synthesis of the DNA-binding domain of c-Mybprotein (142–193)–NH_2. Bull Chem Soc Jpn 64(1):111–117

7. Hackeng TM, Griffin JH, Dawson PE (1999) Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. Proc Natl Acad Sci USA 96 (18):10068–10073. doi:10.1073/pnas.96.18.10068
8. Aimoto S (1999) Polypeptide synthesis by the thioester method. Peptide Sci 51(4):247–265. doi:10.1002/(sici)1097-0282(1999)51:4<247::aid-bip2>3.0.co;2-w
9. Bang D, Pentelute BL, Gates ZP, Kent SB (2006) Directon-resin synthesis of peptide-α-thiophenylesters for use in native chemical ligation. Org Lett 8(6):1049–1052. doi:10.1021/ol052811j
10. Izumi M, Murakami M, Okamoto R, Kajihara Y (2014) Safe and efficient Boc-SPPS for the synthesis of glycopeptide-α-thioesters. J Pept Sci 20(2):98–101. doi:10.1002/psc.2608
11. Mende F, Seitz O (2011) 9-Fluorenylmethoxycarbonyl-based solid-phase synthesis of peptide alpha-thioesters. Angew Chem Int Ed 50(6):1232–1240. doi:10.1002/anie.201005180
12. Backes BJ, Virgilio AA, Ellman JA (1996) Activation method to prepare a highly reactive acylsulfonamide "safety-catch" linker for solid-phase synthesis. J Am Chem Soc 118 (12):3055–3056. doi:10.1021/ja9535165
13. Backes BJ, Ellman JA (1999) Analkane sulfonamide "safety-catch" linker for solid-phase synthesis. J Org Chem 64(7):2322–2330. doi:10.1021/jo981990y
14. Ingenito R, Bianchi E, Fattori D, Pessi A (1999) Solid phase synthesis of peptide C-terminal thioesters by Fmoc/*t*-Bu chemistry. J Am Chem Soc 121(49):11369–11374. doi:10.1021/ja992668n
15. 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(50):11684–11689. doi:10.1021/ja992881j
16. Quaderer R, Hilvert D (2001) Improved synthesis of C-terminal peptide thioesters on "safety-catch" resins using LiBr/THF. Org Lett 3(20):3181–3184. doi:10.1021/ol016492h
17. Flavell RR, Huse M, Goger M, Trester-Zedlitz M, Kuriyan J, Muir TW (2002) Efficient semisynthesis of a tetraphosphorylated analogue of the type I TGF beta receptor. Org Lett 4 (2):165–168. doi:10.1021/ol016859i
18. Wehofsky N, Koglin N, Thust S, Bordusa F (2003) Reverse proteolysis promoted by in situ generated peptide ester fragments. J Am Chem Soc 125(20):6126–6133. doi:10.1021/ja0344213
19. Mezzato S, Schaffrath M, Unverzagt C (2005) An orthogonal double-linker resin facilitates the efficient solid-phase synthesis of complex-type *N*-glycopeptide thioesters suitable for native chemical ligation. Angew Chem Int Ed 44(11):1650–1654. doi:10.1002/anie.200461125
20. Mende F, Seitz O (2007) Solid-phase synthesis of peptide thioesters with self-purification. Angew Chem Int Ed 46(24):4577–4580. doi:10.1002/anie.200700356
21. Alsina J, Yokum TS, Albericio F, Barany G (1999) Backbone amide linker (BAL) strategy for *N*-alpha-9-fluorenylmethoxycarbonyl (Fmoc) solid-phase synthesis of unprotected peptide *p*-nitroanilides and thioesters. J Org Chem 64(24):8761–8769. doi:10.1021/jo990629o
22. Brask J, Albericio F, Jensen KJ (2003) Fmoc solid-phase synthesis of peptide thioesters by masking as trithioortho esters. Org Lett 5(16):2951–2953. doi:10.1021/ol0351044
23. Botti P, Villain M, Manganiello S, Gaertner H (2004) Native chemical ligation through in situ O to S acyl shift. Org Lett 6(26):4861–4864. doi:10.1021/ol0481028
24. 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(21):6576–6578. doi:10.1021/ja0491836
25. Zheng J-S, Cui H-K, Fang G-M, Xi W-X, Liu L (2010) Chemical protein synthesis by kinetically controlled ligation of peptide O-esters. Chem Biochem 11(4):511–515
26. Zheng J-S, Xi W-X, Wang F-L, Li J, Guo Q-X (2011) Fmoc-SPPS chemistry compatible approach for the generation of (glyco)peptide aryl thioesters. Tetrahedron Lett 52(21):2655–2660. doi:10.1016/j.tetlet.2011.03.064

27. Liu F, Mayer JP (2013) An Fmoc compatible, O to S shift-mediated procedure for the preparation of C-terminal thioester peptides. J Org Chem 78(19):9848–9856. doi:10.1021/jo4015112
28. Camarero JA, Hackel BJ, de Yoreo JJ, Mitchell AR (2004) Fmoc-based synthesis of peptide alpha-thioesters using an aryl hydrazine support. J Org Chem 69(12):4145–4151. doi:10.1021/jo040140h
29. 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 (36):6851–6855. doi:10.1002/anie.200705471
30. Raz R, Rademann J (2011) Fmoc-based synthesis of peptide thioesters for native chemical ligation employing a tert-butyl thiol linker. Org Lett 13(7):1606–1609. doi:10.1021/ol1029723
31. Sharma I, Crich D (2011) Direct Fmoc-chemistry-based solid-phase synthesis of peptidyl thioesters. J Org Chem 76(16):6518–6524. doi:10.1021/jo200497j
32. Mahto SK, Howard CJ, Shimko JC, Ottesen JJ (2011) A reversible protection strategy to improve Fmoc-SPPS of peptide thioesters by the *N*-acylurea approach. Chem Biochem 12 (16):2488–2494. doi:10.1002/cbic.201100472
33. Okamoto R, Morooka K, Kajihara Y (2012) A synthetic approach to a peptide alpha-thioester from an unprotected peptide through cleavage and activation of a specific peptide bond by *N*-acetylguanidine. Angew Chem Int Ed 51(1):191–196. doi:10.1002/anie.201105601
34. Anraku Y, Mizutani R, Satow Y (2005) Protein splicing: its discovery and structural insight into novel chemical mechanisms. Iubmb Life 57(8):563–574. doi:10.1080/15216540500215499
35. Camarero JA (2008) Recent developments in the site-specific immobilization of proteins onto solid supports. Peptide Sci 90(3):450–458. doi:10.1002/bip.20803
36. Flavell RR, Muir TW (2008) Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. Acc Chem Res 42(1):107–116. doi:10.1021/ar800129c
37. Chattopadhaya S, Abu-Bakar FB, Yao SQ (2009) Use of intein-mediated protein ligation strategies for the fabrication of functional protein arrays. Methods in Enzymol Non-natural Amino Acids 462:195. doi:10.1016/S0076-6879(09)62010-3
38. Vila-Perello M, Muir TW (2010) Biological applications of protein splicing. Cell 143 (2):191–200. doi:10.1016/j.cell.2010.09.031
39. Binschik J, Mootz HD (2013) Chemical bypass of intein-catalyzed N–S acyl shift in protein splicing. Angew Chem Int Ed 52(15):4260–4264. doi:10.1002/anie.201208863
40. Kawakami T, Sumida M, Nakamura K, Vorherr T, Aimoto S (2005) Peptide thioester preparation based on an N–S acyl shift reaction mediated by a thiol ligation auxiliary. Tetrahedron Lett 46(50):8805–8807. doi:10.1016/j.tetlet.2005.09.184
41. Ollivier N, Behr JB, El-Mahdi O, Blanpain A, Melnyk O (2005) Fmoc solid-phase synthesis of peptide thioesters using an intramolecular N, S-acyl shift. Org Lett 7(13):2647–2650. doi:10.1021/ol050776a
42. Ohta Y, Itoh S, Shigenaga A, Shintaku S, Fujii N, Otaka A (2006) Cysteine-derived S-protected oxazolidinones: potential chemical devices for the preparation of peptide thioesters. Org Lett 8(3):467–470. doi:10.1021/ol052755m
43. Nagaike F, Onuma Y, Kanazawa C, Hojo H, Ueki A, Nakahara Y, Nakahara Y (2006) Efficient microwave-assisted tandem *N*- to *S*-acyl transfer and thioester exchange for the preparation of a glycosylated peptide thioester. Org Lett 8(20):4465–4468. doi:10.1021/ol0616034
44. Nakamura KI, Sumida M, Kawakami T, Vorherr T, Aimoto S (2006) Generation of an S-peptide via an N–S acyl shift reaction in a TFA solution. Bull Chem Soc Jpn 79 (11):1773–1780. doi:10.1246/bcsj.79.1773
45. Hojo H, Onuma Y, Akimoto Y, Nakahara Y, Nakahara Y (2007) *N*-Alkyl cysteine-assisted thioesterification of peptides. Tetrahedron Lett 48(1):25–28. doi:10.1016/j.tetlet.2006.11.034

46. Kawakami T, Aimoto S (2007) Sequential peptide ligation by using a controlled cysteinyl prolyl ester (CPE) autoactivating unit. Tetrahedron Lett 48(11):1903–1905. doi:10.1016/j.tetlet.2007.01.086
47. Ki N, Mori H, Kawakami T, Hojo H, Nakahara Y, Aimoto S (2007) Peptide thioester synthesis via an auxiliary-mediated N–S acyl shift reaction in solution. Int J Pept Res Ther 13(1–2):191–202. doi:10.1007/s10989-006-9065-9
48. Hojo H, Murasawa Y, Katayama H, Ohira T, Nakaharaa Y, Nakahara Y (2008) Application of a novel thioesterification reaction to the synthesis of chemokine CCL27 by the modified thioester method. Org Biomol Chem 6(10):1808–1813. doi:10.1039/b800884a
49. Ozawa C, Katayama H, Hojo H, Nakahara Y, Nakahara Y (2008) Efficient sequential segment coupling using *N*-alkylcysteine-assisted thioesterification for glycopeptide dendrimer synthesis. Org Lett 10(16):3531–3533. doi:10.1021/ol801340m
50. Kang J, Reynolds NL, Tyrrell C, Dorin JR, Macmillan D (2009) Peptide thioester synthesis through N≥S acyl-transfer: application to the synthesis of a beta-defensin. Org Biomol Chem 7(23):4918–4923. doi:10.1039/b913886b
51. Kang J, Richardson JP, Macmillan D (2009) 3-Mercaptopropionic acid-mediated synthesis of peptide and protein thioesters. Chem Commun 4:407–409. doi:10.1039/b815888f
52. Kawakami T, Aimoto S (2009) The use of a cysteinyl prolyl ester (CPE) autoactivating unit in peptide ligation reactions. Tetrahedron 65(19):3871–3877. doi:10.1016/j.tet.2009.03.008
53. Nakamura KI, Kanao T, Uesugi T, Hara T, Sato T, Kawakami T, Aimoto S (2009) Synthesis of peptide thioesters via an N–S acyl shift reaction under mild acidic conditions on an *N*-4,5-dimethoxy-2-mercaptobenzyl auxiliary group. J Pept Sci 15(11):731–737. doi:10.1002/psc.1164
54. Erlich LA, Kumar KSA, Haj-Yahya M, Dawson PE, Brik A (2010) *N*-Methylcysteine-mediated total chemical synthesis of ubiquitin thioester. Org Biomol Chem 8(10):2392–2396. doi:10.1039/c000332h
55. Katayama H, Hojo H, Shimizu I, Nakahara Y, Nakahara Y (2010) Chemical synthesis of mouse pro-opiomelanocortin(1-74) by azido-protected glycopeptide ligation via the thioester method. Org Biomol Chem 8(8):1966–1972. doi:10.1039/b927270d
56. Richardson JP, Chan C-H, Blanc J, Saadi M, Macmillan D (2010) Exploring neoglycoprotein assembly through native chemical ligation using neoglycopeptide thioesters prepared via N -> S acyl transfer. Org Biomol Chem 8(6):1351–1360. doi:10.1039/b920535g
57. Eom KD, Tam JP (2011) Acid-catalyzed tandem thiol switch for preparing peptide thioesters from mercaptoethyl esters. Org Lett 13(10):2610–2613. doi:10.1021/ol2007204
58. Hojo H, Kobayashi H, Ubagai R, Asahina Y, Nakahara Y, Katayama H, Ito Y, Nakahara Y (2011) Efficient preparation of Fmoc-aminoacyl-*N*-ethylcysteine unit, a key device for the synthesis of peptide thioesters. Org Biomol Chem 9(19):6807–6813. doi:10.1039/c1ob05831b
59. Macmillan D, De Cecco M, Reynolds NL, Santos LFA, Barran PE, Dorin JR (2011) Synthesis of cyclic peptides through an intramolecular amide bond rearrangement. Chem Biochem 12(14):2133–2136. doi:10.1002/cbic.201100364
60. Zheng J-S, Chang H-N, Wang F-L, Liu L (2011) Fmoc synthesis of peptide thioesters without post-chain-assembly manipulation. J Am Chem Soc 133(29):11080–11083. doi:10.1021/ja204088a
61. Adams AL, Cowper B, Morgan RE, Premdjee B, Caddick S, Macmillan D (2013) Cysteine promoted C-terminal hydrazinolysis of native peptides and proteins. Angew Chem Int Ed 52(49):13062–13066. doi:10.1002/anie.201304997
62. Adams AL, Macmillan D (2013) Investigation of peptide thioester formation via N → Se acyl transfer. J Pept Sci 19(2):65–73. doi:10.1002/psc.2469
63. Romanelli A, Shekhtman A, Cowburn D, Muir TW (2004) Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the *N*-extein-intein junction. Proc Natl Acad Sci USA 101(17):6397–6402. doi:10.1073/pnas.0306616101

64. Tsuda S, Shigenaga A, Bando K, Otaka A (2009) N - > S acyl-transfer-mediated synthesis of peptide thioesters using anilide derivatives. Org Lett 11(4):823–826. doi:10.1021/ol8028093
65. Sakamoto K, Sato K, Shigenaga A, Tsuji K, Tsuda S, Hibino H, Nishiuchi Y, Otaka A (2012) Synthetic procedure for *N*-Fmoc amino acyl-*N*-sulfanylethylaniline linker as crypto-peptide thioester precursor with application to native chemical ligation. J Org Chem 77(16):6948–6958. doi:10.1021/jo3011107
66. Hibino H, Nishiuchi Y (2011) 4-Methoxybenzyloxymethyl group as an Nπ-protecting group for histidine to eliminate side-chain-induced racemization in the Fmoc strategy. Tetrahedron Lett 52(38):4947–4949. doi:10.1016/j.tetlet.2011.07.065
67. Villain M, Vizzavona J, Rose K (2001) Covalent capture: a new tool for the purification of synthetic and recombinant polypeptides. Chem Biol 8(7):673–679. doi:10.1016/s1074-5521(01)00044-8
68. Bang D, Kent SBH (2004) A one-pot total synthesis of crambin. Angew Chem Int Ed 43(19):2534–2538. doi:10.1002/anie.200353540
69. Ueda S, Fujita M, Tamamura H, Fujii N, Otaka A (2005) Photolabile protection for one-pot sequential native chemical ligation. Chem Biochem 6(11):1983–1986. doi:10.1002/cbic.200500272
70. Boerema DJ, Tereshko VA, Kent SBH (2008) Total synthesis by modern chemical ligation methods and high resolution (1.1 Å) X-ray structure of ribonuclease A. Peptide Sci 90(3):278–286. doi:10.1002/bip.20800
71. Shigenaga A, Sumikawa Y, Tsuda S, Sato K, Otaka A (2010) Sequential native chemical ligation utilizing peptide thioacids derived from newly developed Fmoc-based synthetic method. Tetrahedron 66(18):3290–3296. doi:10.1016/j.tet.2010.03.016
72. Tsuji K, Shigenaga A, Sumikawa Y, Tanegashima K, Sato K, Aihara K, Hara T, Otaka A (2011) Application of N-C- or C-N- directed sequential native chemical ligation to the preparation of CXCL14 analogs and their biological evaluation. Bioorg Med Chem 19(13):4014–4020. doi:10.1016/j.bmc.2011.05.018
73. Bang D, Pentelute BL, Kent SBH (2006) Kinetically controlled ligation for the convergent chemical synthesis of proteins. Angew Chem Int Ed 45(24):3985–3988. doi:10.1002/anie.200600702
74. Durek T, Torbeev VY, Kent SBH (2007) Convergent chemical synthesis and high-resolution X-ray structure of human lysozyme. Proc Natl Acad Sci USA 104(12):4846–4851. doi:10.1073/pnas.0610630104
75. Torbeev VY, Kent SBH (2007) Convergent chemical synthesis and crystal structure of a 203 amino acid "covalent dimer" HIV-1 protease enzyme molecule. Angew Chem Int Ed 46(10):1667–1670. doi:10.1002/anie.200604087
76. Sato K, Shigenaga A, Tsuji K, Tsuda S, Sumikawa Y, Sakamoto K, Otaka A (2011) *N*-Sulfanylethylanilide peptide as a crypto-thioester peptide. Chem Biochem 12(12):1840–1844. doi:10.1002/cbic.201100241
77. Otaka A, Sato K, Ding H, Shigenaga A (2012) One-pot/sequential native chemical ligation using *N*-sulfanylethylanilide peptide. Chem Rec 12(5):479–490. doi:10.1002/tcr.201200007
78. Bang D, Lee J, Kwon Y, Pentelute BL (2011) Use of model peptide reactions for the characterization of kinetically controlled ligation. Bioconjugate Chem 22(8):1645–1649. doi:10.1021/Bc2002242
79. Johnson ECB, Kent SBH (2006) Insights into the mechanism and catalysis of the native chemical ligation reaction. J Am Chem Soc 128(20):6640–6646. doi:10.1021/ja058344i
80. 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(20):5588–5591. doi:10.1021/ol202316v
81. Shang S, Tan Z, Dong S, Danishefsky SJ (2011) Anadvance in proline ligation. J Am Chem Soc 133(28):10784–10786. doi:10.1021/ja204277b
82. Townsend SD, Tan Z, Dong S, Shang S, Brailsford JA, Danishefsky SJ (2012) Advances in proline ligation. J Am Chem Soc 134(8):3912–3916. doi:10.1021/ja212182q

83. Raibaut L, Seeberger P, Melnyk O (2013) Bis(2-sulfanylethyl)amidopeptides enable native chemical ligation at proline and minimize deletion side-product formation. Org Lett 15 (21):5516–5519. doi:10.1021/ol402678a
84. Nakamura T, Shigenaga A, Sato K, Tsuda Y, Sakamoto K, Otaka A (2014) Examination of native chemical ligation using peptidyl prolyl thioesters. Chem Commun 50(1):58–60. doi:10.1039/c3cc47228k
85. 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(48):9248–9252. doi:10.1002/anie.200704195
86. Fang G-M, Wang J-X, Liu L (2012) Convergent chemical synthesis of proteins by ligation of peptide hydrazides. Angew Chem Int Ed 51(41):10347–10350. doi:10.1002/anie.201203843
87. 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(12):5428–5431. doi:10.1021/ja2109079
88. 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(15):3567–3572. doi:10.1002/anie.201109034
89. Hojo H, Tanaka H, Hagiwara M, Asahina Y, Ueki A, Katayama H, Nakahara Y, Yoneshige A, Matsuda J, Ito Y, Nakahara Y (2012) Chemoenzymatic synthesis of hydrophobic glycoprotein: synthesis of saposin C carrying complex-type carbohydrate. J Org Chem 77:9437–9446. doi:10.1021/jo3010155
90. Boutureira O, Bernardes GJL, Fernandez-Gonzalez M, Anthony DC, Davis BG (2012) Selenenylsulfide-linked homogeneous glycopeptides and glycoproteins: synthesis of human "hepatic Se metabolite A". Angew Chem Int Ed 51(6):1432–1436. doi:10.1002/anie.201106658
91. Ullmann V, Rädisch M, Boos I, Freund J, Pöhner C, Schwarzinger S, Unverzagt C (2012) Convergent solid-phase synthesis of *n*-glycopeptides facilitated by pseudoprolines at consensus-sequence Ser/Thr residues. Angew Chem Int Ed 51(46):11566–11570. doi:10.1002/anie.201204272
92. Siman P, Karthikeyan SV, Nikolov M, Fischle W, Brik A (2013) Convergent chemical synthesis of histone H2B protein for the site-specific ubiquitination at Lys34. Angew Chem Int Ed 52(31):8059–8063. doi:10.1002/anie.201303844
93. Wilson RM, Dong S, Wang P, Danishefsky SJ (2013) The winding pathway to erythropoietin along the chemistry–biology frontier: a success at last. Angew Chem Int Ed 52(30):7646–7665. doi:10.1002/anie.201301666
94. Kawakami T, Akai Y, Fujimoto H, Kita C, Aoki Y, Konishi T, Waseda M, Takemura L, Aimoto S (2013) Sequential peptide ligation by combining the Cys-Pro ester (CPE) and thioester methods and its application to the synthesis of histone H3 containing a trimethyllysine residue. Bull Chem Soc Jpn 86(6):690–697. doi:10.1246/bcsj.20130026
95. Deng F-K, Zhang L, Wang Y-T, Schneewind O, Kent SBH (2014) Total chemical synthesis of the enzyme sortase AΔN59 with full catalytic activity. Angew Chem Int Ed 53(18):4662–4666. doi:10.1002/anie.201310900
96. 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(20):5188–5193. doi:10.1002/anie.201310574
97. Sato K, Shigenaga A, Kitakaze K, Sakamoto K, Tsuji D, Itoh K, Otaka A (2013) Chemical synthesis of biologically active monoglycosylated GM2-activator protein analogue using *N*-sulfanylethylanilide peptide. Angew Chem Int Ed 52(30):7855–7859. doi:10.1002/anie.201303390

98. Kolter T, Sandhoff K (2005) Principles of lysosomal membrane digestion: stimulation of sphingolipid degradation by sphingolipid activator proteins and anionic lysosomal lipids. Ann Rev Cell Develop Biol 21:81–103. doi:10.1146/annurev.cellbio.21.122303.120013
99. Mhidia R, Beziere N, Blanpain A, Pommery N, Melnyk O (2010) Assembly/disassembly of drug conjugates using imide ligation. Org Lett 12(18):3982–3985. doi:10.1021/ol101049g
100. Ollivier N, Dheur J, Mhidia R, Blanpain A, Melnyk O (2010) Bis(2-sulfanylethyl)amino native peptide ligation. Org Lett 12(22):5238–5241. doi:10.1021/ol102273u
101. Dheur J, Ollivier N, Vallin A, Melnyk O (2011) Synthesis of peptide alkylthioesters using the intramolecular N, S-acyl shift properties of bis(2-sulfanylethyl)amido peptides. J Org Chem 76(9):3194–3202. doi:10.1021/jo200029e
102. Ollivier N, Vicogne J, Vallin A, Drobecq H, Desmet R, El Mahdi O, Leclercq B, Goormachtigh G, Fafeur V, Melnyk O (2012) A one-pot three-segment ligation strategy for protein chemical synthesis. Angew Chem Int Ed 51(1):209–213. doi:10.1002/anie.201105837
103. Raibaut L, Ollivier N, Melnyk O (2012) Sequential native peptide ligation strategies for total chemical protein synthesis. Chem Soc Rev 41(21):7001–7015. doi:10.1039/c2cs35147a
104. Yang R, Hou W, Zhang X, Liu C-F (2012) N-to-C Sequential ligation using peptidyl *N*, *N*-bis (2-mercaptoethyl)amide building blocks. Org Lett 14(1):374–377. doi:10.1021/ol2031284
105. Taichi M, Hemu X, Qiu Y, Tam JP (2013) A thioethylalkylamido (TEA) thioester surrogate in the synthesis of a cyclic peptide via a tandem acyl shift. Org Lett 15(11):2620–2623. doi:10.1021/ol400801k
106. Burlina F, Papageorgiou G, Morris C, White PD, Offer J (2014) In situ thioester formation for protein ligation using α-methylcysteine. Chem Sci 5(2):766–770. doi:10.1039/C3SC52140K

Top Curr Chem (2015) 363: 57–102
DOI: 10.1007/128_2014_594
© Springer International Publishing Switzerland 2014
Published online: 7 February 2015

Postligation-Desulfurization: A General Approach for Chemical Protein Synthesis

Jimei Ma, Jing Zeng, and Qian Wan

Abstract Native chemical ligation, involving regioselective and chemoselective coupling of two unprotected peptide segments, enabled the synthesis of polypeptide with more than 200 amino acids. However, cysteine was indispensable in this synthetic technique in its initial format, which limited its further application. Thus, considerable effort has been put into breaking the restriction of cysteine-containing ligation. As a consequence, postligation-desulfurization, concerning thiol-mediated ligation followed by desulfurization, was developed. This review describes the development and recent progress on the chemical synthesis of peptides and proteins encompassing postligation-desulfurization at alanine, valine, lysine, threonine, leucine, proline, arginine, aspartic acid, glutamate, phenylalanine, glutamine, and tryptophan.

Keywords Chemical ligation · Metal-free desulfurization · Peptide · Protein · Radical

Contents

J. Ma
Department of Chemistry, College of Science, Huazhong Agricultural University, Wuhan 430070, China

J. Zeng
School of Pharmacy, Huazhong University of Science and Technology, Wuhan 430030, China

Q. Wan (✉)
School of Pharmacy, Huazhong University of Science and Technology, Wuhan 430030, China

Institute of Brain Research, Huazhong University of Science and Technology, Wuhan, Hubei 430030, China
e-mail: wanqian@hust.edu.cn

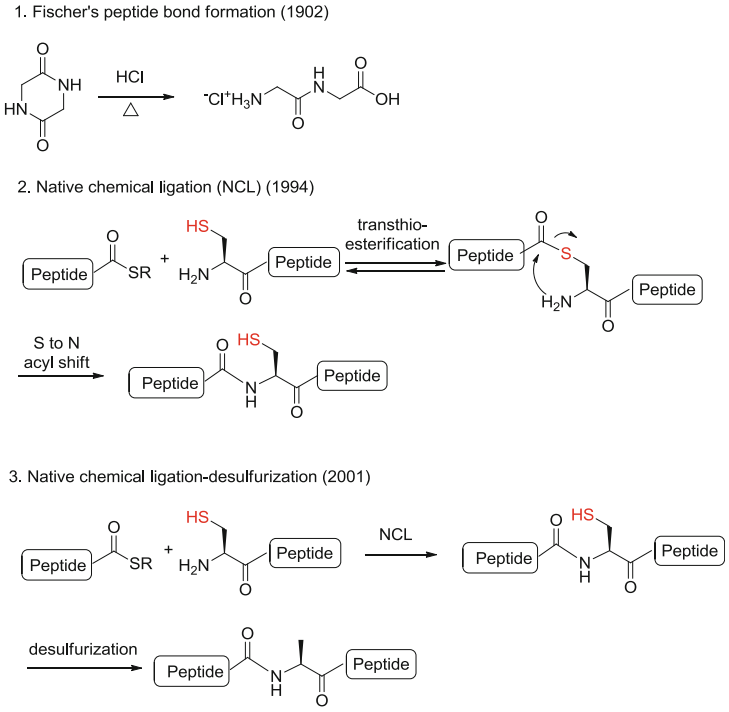
Abbreviations

Ac	Acetyl
AFGP	Antifreeze glycoprotein
Ala	Alanine
Alloc	Allyloxycarbonyl
Arg	Arginine
Asp	Aspartic acid
Boc	*tert*-Butyloxycarbonyl
CXCR	Chemokine receptor
Cys	Cysteine
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DIAD	Diisopropyl azodicarboxylate
DIBAL-H	Diisobutylaluminum hydride
DIPEA	Ethyldiisopropylamine
DMAP	4-Dimethylaminopyridine
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethyl sulfoxide
EDC	*N*-(3-Dimethylanino propyl)-*N*′-ethylcarbodiimide
Fmoc	Fluorenylmethyloxycarbonyl
Gln	Glutamine
Glu	Glutamic acid
Gn	Guanidine
HEPES	2-[4-(2-Hydroxyethyl)-1-peperazinyl]ethanesulfonic acid
HMDS	Bis(trimethylsilyl)amide
HOOBt	Hydroxy-3,4-dihydro-4-oxo-1,2,3-benzotriazine
hPTHrP	Human parathyroid hormone-related protein
KHMDS	Potassium 1,1,1,3,3,3-hexamethyldisilazane
LC-MS	Liquid chromatography-mass spectrometry
Leu	Leucine
Lys	Lysine
*m*CPBA	*meta*-Chloroperoxybenzoic acid
MESNa	Sodium-2-mercaptoethane sulfonate

Met	Methionine
MMTS	(*S*)-Methyl methanethiosulfonate
MPAA	Mercaptophenylacetic acid
Ms	Methanesulfonyl
NCL	Native chemical ligation
NMM	*N*-Methylmorpholine
NVOC	*o*-Nitroveratryloxycarbonyl
Pen	Penicillamine
Phe	Phenylalanine
PhFl	9-Phenylfluroenyl
Pro	Proline
PyBOP	Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
rt	Room temperature
Ser	Serine
SPPS	Solid phase peptide synthesis
*t*Bu	*tert*-Butyl
TBAF	Tetra-*n*-butylammonium fluoride
TBS	*tert*-Butyldimethylsilyl
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TFE	Trifluoroethanol
TFET	Trifluoroethanethiol
THF	Tetrahydrofuran
Thr	Threonine
TIS	Triisopropylsilane
TMSE	Trimethylsilylethyl
Tris	Tri(hydroxymethyl)aminomethane
Trp	Tryptophan
Trt	Trityl
Ts	4-Toluenesulfonyl
Ub	Ubiquitin
UV	Ultraviolet
VA-044	2,2′-Azobis[2-(2-imidazolin-2-yl)propane] dihydrochloride
Val	Valine

1 Introduction

The chemical synthesis of peptides and proteins has gained widespread attention of chemists and biochemists because of the great importance of and huge necessity for homogeneous proteins [1–3]. The past century has witnessed a major advance in the development of peptide assembly techniques [4–12] since the first peptide bond formation via chemical synthesis by Fisher in 1901 [13]. Among these great strides,



Scheme 1 The evolution of chemical ligation-desulfurization

perhaps the most innovative and central advances was the introduction of solid phase peptides synthesis (SPPS) by Merrifield in 1963, which has revolutionized the whole of peptide and protein chemistry [14–16]. Owing to Merrifield's SPPS technology, the chemical synthesis of peptides and proteins has become a routine pipeline production. Unfortunately, SPPS is generally limited to peptides with chain lengths less than ca. 50 amino acids because of the linear nature of the technique, although there are some examples for the synthesis of larger peptides, even proteins such as ribonuclease [17–19].

Another important milestone in peptide synthesis is the development of native chemical ligation (NCL) by Dawson and Kent in 1994 [20], inspired by Kemp's earlier feasibility demonstrations [21, 22]. NCL involved a native peptide bond formation between an unprotected C-terminal peptide thioester and a second unprotected N-terminal cysteinyl peptide (Scheme 1). The occurrence of thiol group on cysteine initiated the ligation by a reversible transthioesterification, and subsequent irreversible S → N acyl shift via a five-membered transition state furnished the native peptide bond [23]. The great potential of NCL has been demonstrated by the total chemical synthesis of a series of proteins [12, 24, 25]. However, the strict requirement of cysteine residue, which is actually relatively rare in naturally-occurring peptides and proteins, dramatically limited its

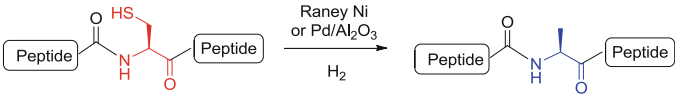
application. In many proteins there are no cysteine residues or the cysteine residue isn't in a position suitable for the NCL.

Great efforts have been made to overcome this limitation. One solution is to introduce a removable thiol-containing auxiliary as a cysteine mimic. Albeit with some successful examples, these strategies are limited by the typically sluggish coupling rate and the requirement of relatively unhindered ligation sites [26–31]. Another solution is to transform the thiol group of cysteine to a hydroxyl group after ligation, which allowed the conversion of cysteine to serine [32–35]. Very recently, Li et al. also developed a salicylaldehyde ester-induced chemoselective serine and threonine ligation, which is different from NCL in mechanism [36–39]. Among these efforts, the most important and widely applied method is to reduce the thiol group of cysteine after ligation, which converts the cysteine to more abundant alanine. In the past few years, this strategy has been further expanded to other natural amino acid sites through the ligation of peptides containing synthetic thiol-derived unnatural amino acids and subsequent desulfurization. In this review we focus on the development of postligation-desulfurization strategy and its application in protein synthesis [4, 5, 7, 9, 11, 40].

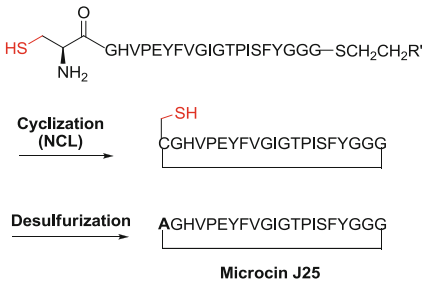
2 Postligation-Desulfurization (Alanine Ligation)

2.1 Metal-Based Desulfurization

In 2001, Yan and Dawson first combined NCL and desulfurization to convert the cysteine-containing peptide to alanine-containing peptide [41]. The hydrogenolytic desulfurization of cysteine residue was achieved through the action of either Raney nickel or Pd/Al_2O_3 (Scheme 2). It was found by the Kent group in 2007 that these metal-based desulfurization protocols could effectively accommodate both methionine and acetamidomethyl (Acm) functionality [42]. The applicability of this strategy was demonstrated by the syntheses of cyclic antibiotic microcin J25 (Scheme 3), streptococcal protein G B1 domain, and an analogue of baranase. Despite these successes, the drawbacks, such as requirement of a large excess of nickel, cause epimerization of secondary alcohols, not accommodated with Trp, thiols, thioethers, thioesters, and thiazolidine (Thz) limiting the generality of this approach. Nevertheless, this conceptual approach opens a new window for the application of NCL on peptide and protein synthesis. It provides new possibilities to extend cysteine-based peptide ligation to other amino acids.



Scheme 2 Metal-based desulfurization



Scheme 3 Synthesis of Microcin J25 by postligation-metal-based desulfurization. Conditions: NCL: 0.1 M Tris-HCl, 6.0 M Gn·HCl, pH 8.5, 2% BnSH, 2% PhSH, overnight, rt, 50%; Desulfurization: 20% HOAc, Pd/Al_2O_3, or Pd/C, or Raney nickel, H_2, quantitative

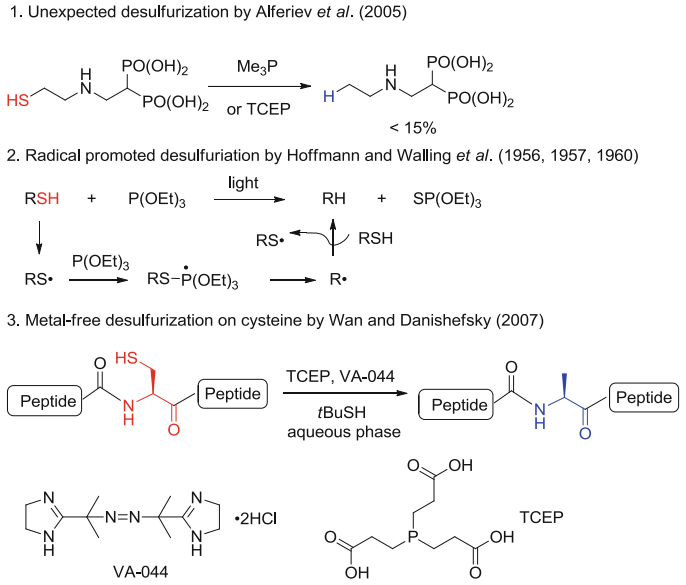
2.2 *Metal-Free Desulfurization*

Given the limitations of metal-based desulfurization, development of new conditions was urgently needed to make the postligation-desulfurization strategy more general. These conditions must fulfill the following requirements:

1. Mild reaction conditions with a tolerance of a range of functional groups including carbohydrates and various amino acids, especially thiol-containing functionalities, such as methionine, Cys(Acm), biotin, Thz, thioether and thioesters
2. High efficiency with less side reactions and perfect yields
3. Tolerance in aqueous phase
4. Easy purification

In 2005, Alferiev et al. reported an unexpected desulfurization reaction when trimethylphosphine or tris(2-carboxyethyl)phosphine (TCEP) was used to reduce disulfide bonds in aqueous phase (Scheme 4) [43]. This side reaction possibly passed through a free-radical process under basic conditions in the presence of air. Inspired by this unwanted side reaction, Wan and Danishefsky came to consider the possibility of radical-based reduction of cysteine to alanine [44].

Inspiringly, Hoffmann et al. described a desulfurization reaction promoted by trialkylphosphite derivatives under thermal or photochemical conditions [45] via a radical reaction mechanism, proposed by Walling and Rabinowitz soon thereafter (Scheme 4) [46, 47]. These radical chain processes involved an alkylthiyl radical attack on phosphite to generate a phosphoranyl radical intermediate, subsequent β scission to provide an alkyl radical which rapidly abstracted hydrogen from the



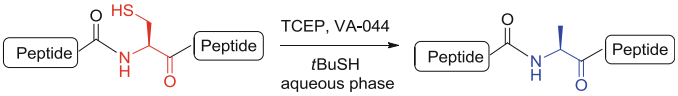
Scheme 4 Radical promoted desulfurization

parent thiol to furnish the desulfurized alkane product and propagate the chain. Walling et al. also extended this reaction to trialkylphosphines [47]. In addition, Valencia and co-workers also reduced cysteine to alanine through radical reaction initiated by triethylphosphite and borane [48–50]. Unfortunately, these conditions were unsuccessful when used to reduce a cysteine-containing peptide.

Encouraged by these reports, Wan and Danishefsky sought to develop mild, metal-free, radical-based desulfurization conditions for cysteine reduction in peptide settings. TCEP, which was widely used as disulfide reducing reagent in peptide and glycoprotein chemistry [51], was chosen as phosphine source for the benefits from its good tolerance to a range of functionality and easy manipulation in air and aqueous solution over a wide pH range. In addition, water-soluble 2,2′-azobis [2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) was used as radical initiator. With these reagents, the desulfurization was carried out in water in the presence of *t*BuSH. As expected, nearly quantitative desulfurization of cysteine in model peptides was observed in a few hours (Scheme 4). More importantly, these reaction conditions accommodated various functional groups including methionine, Cys(Acm), Thz and biotin (Table 1) [44].

The applicability was further examined by synthesis of a glycopeptide containing an N-terminal Thz group, Acm-protected Cys residue, N-linked glycan, and C-terminal thioester. It was a pleasure to see that the postligation and subsequent desulfurization under the established TCEP-mediated conditions furnished the desired glycopeptide, which incorporated alanine residue in place of the cysteine group, in good overall yields. Wan and Danishefsky further applied this strategy

Table 1 Metal-free desulfurization



Entry	Cysteinyl peptide	Alanyl peptide	Yield (%)
1	Fmoc-RYKDSG**C**AHPRG-OH	Fmoc-RYKDSG**A**AHPRG-OH	82
2	H-LRHKDS**C**RWKITR-OH	H-LRHKDS**A**RWKITR-OH	81
3	H-VETRFP**C**RNYEK-OH	H-VETRFP**A**RNYEK-OH	71
4	H-RFDS**C**RPMHWR-OH	H-RFDS**A**RPMHWR-OH	74
5	H-VRYT**C**KLSCys(Acm)WR-OH	H-VRYT**A**KLSCys(Acm)WR-OH	89
6	Fmoc-Thz-YTRG**C**AKG-OH	Fmoc-Thz-YTRG**A**AKG-OH	75
7	Biotin-KWRITN**C**EHR-OH	Biotin-KWRITN**A**EHR-OH	91

to the synthesis of cyclic peptide crotogossamide. It was also demonstrated that the protocol was able to reduce selenocysteine and other seleno-amino acid residue [17, 44, 52–54].

2.3 *Application of Postligation-Metal-Free Desulfurization (Ala Ligation)*

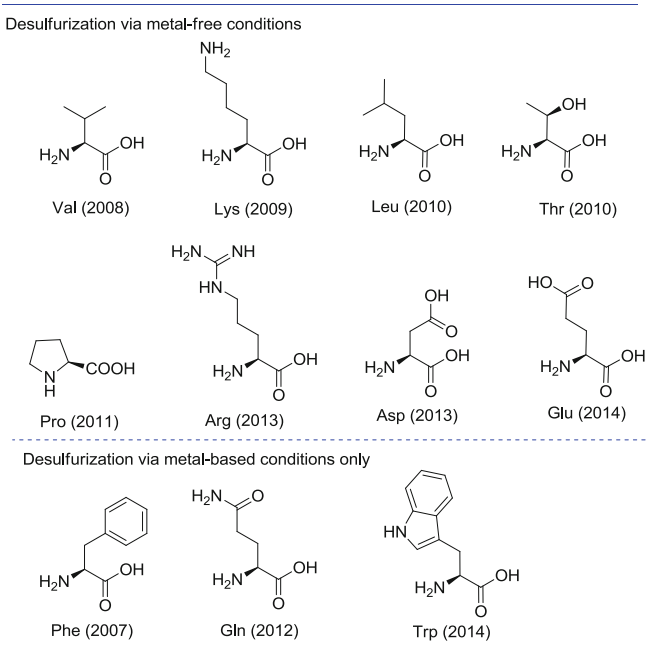
Using this highly versatile free-radical cysteine reduction protocol, a serious of proteins with complex structures and up to 306 amino acid residues have been synthesized, including human glycosyl-interferon-β [55], orf virus entry fusion proteins ORFV036 and 049 [56], enzyme sortase $A_{\Delta N59}$ [57], antifreeze proteins (AFPs) and antifreeze glycoproteins (AFGPs) [58, 59], ribosomal protein S25 (RpS25) [60], α-synuclein (α-syn) [61, 62], granulocyte-macrophage colony-stimulating factor (GM-CSF) [57], etc.

Among these protein syntheses, one of the most fantastic examples is the total synthesis of homogeneous erythropoietin (EPO). EPO is a glycoprotein containing polysaccharide chains and plays a vital role and mediates a range of crucial biological processes. Many researchers have provided impressive contributions to the total synthesis of EPO and, finally, Danishefsky's group achieved the full synthesis with all carbohydrates at all native sites [63, 64]. The synthesis involved quadruple cysteine NCLs from four glyco-containing peptide fragments, one global metal-free desulfurization, and a final folding step. This synthetic EPO was proved to possess in vitro activity. The success of the total synthesis of such a challenging molecule illustrated a promising future for the application of this postligation-metal-free desulfurization strategy in complex protein synthesis.

3 Postligation-Desulfurization at Other Amino Acid Sites

The successful application of native chemical ligation-desulfurization strategy (alanine ligation) – especially metal-free desulfurization strategy – in chemical total synthesis of proteins has stimulated the extension of this strategy to peptide ligation at other amino acid sites. To date, the chemical synthesis of peptides or proteins via postligation-metal-free-desulfurization strategy has been achieved at valine, lysine, leucine, threonine, proline, arginine, aspartic acid, and glutamate sites (Table 2). Ligation at phenylalanine, valine, leucine, glutamine, and tryptophan sites has also been carried out, accompanied by metal-based desulfurization. The key point of this triumph is to introduce suitable thio-derived proteinogenic amino acids as Cys surrogates which were able to facilitate the ligation reaction in a similar manner to Cys residue and, more importantly, could be subsequently desulfurized via either metal-free or metal-based conditions. A common strategy to synthesize these Cys surrogates is to install thiol groups into α or β positions of amino groups of corresponding natural amino acids.

Table 2 Postligation-desulfurization achieved at amino acids other than Cys



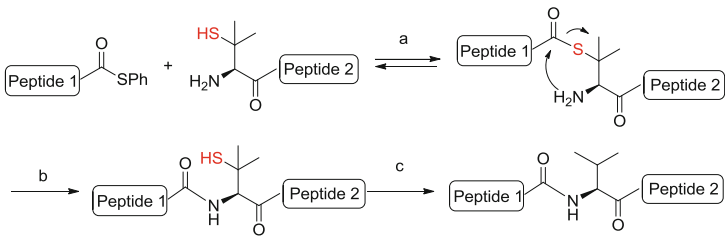
3.1 Metal-Free Desulfurization

3.1.1 Ligation at Valine Site

In 2008, the Seitz group described that commercially available β-thiol valine (penicillamine) could be utilized in ligation-desulfurization strategy (Scheme 5) [65]. The model peptide Pen-Arg-Ala-Glu-Tyr-Ser-NH_2 containing a Boc/Trt-protected Pen was prepared by Fmoc-SPPS. The ligation of **1** with less sterically hindered C-terminal peptide thioester was remarkably fast under optimal condition (6 M Gn·HCl, 100 mM NaH_2PO_4, 50 mM TCEP, 5% PhSH, pH 8.5, 37°C). The ligation of Pen-Arg-Ala-Glu-Tyr-Ser-NH_2 with more sterically hindered peptide thioester also went well, to give target product, albeit with lower reaction rate and a small amount of epimerized ligation product. However, these shortcomings could be avoided by increasing pH and loading excess thioester. These tactics were successfully applied in the synthesis of peptide **5**, segment of STAT-1, and the 22-mer Syk kinase peptide **6**. Interestingly, subsequent desulfurization was not smooth under metal-induced conditions, but proceeded successfully under metal-free conditions (an aqueous 100 mM phosphate buffer adjusted to pH 6.5 which contained 250 mM TCEP, 200 mM VA-044, 40 mM glutathione, and 3 M Gn·HCl) at 65°C (Table 3).

Almost at the same time, Wan and Danishefsky also demonstrated that β-thiol valine (penicillamine) could assist in the oxo-ester mediated ligation step [66]. β-Thiol valine was readily ligated with peptides **9** and **10** with nitro phenyl ester terminal in pH 6.3–6.5 buffer (6.0 M Gn·HCl, 188.8 mM Na_2HPO_4, 7.2 mM *p*-NO_2PhOH) with TCEP at 30°C for 2–3 h. The desulfurization proceeded under standard condition (6.0 M Gn·HCl, 188.8 mM Na_2HPO_4, 18.8 mM TCEP·HCl, pH 6.3–6.5, TCEP, VA-044, 37°C, 2–3 h) to give target peptides (Scheme 6). It should be noted that this oxo-ester-mediated NCL strategy exhibited high efficiency compared to the traditional reaction of the corresponding thioester, which afforded trace amount of product indicated by LC-MS.

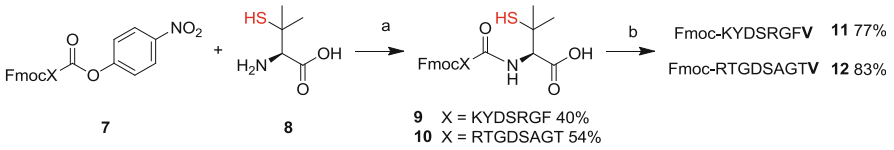
However, when the β-thiol valine was incorporated into the N-terminal peptide as the acyl acceptor, the ligation was dilatory because of the tertiary property of the thiol group. Considering the less steric hindrance and greater reactivity of a primary thiol, the Danishefsky group tried to install a thiol group at the γ-position on Val to extend further the Val-promoted NCL. The γ-mercapto valine was synthesized



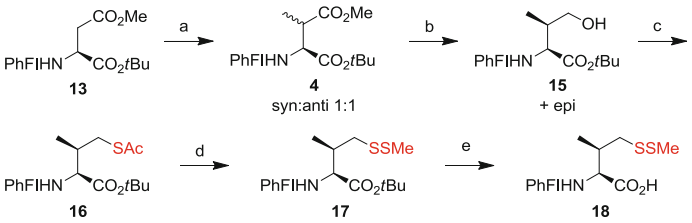
Scheme 5 Native chemical ligation at Val: (**a**) trans-thioesterification; (**b**) S- to N-acyl transfer; (**c**) desulfurization

Table 3 Ligation at Val site and desulfurization

Ligation		Desulfurization yield (%)	
Product/yield (%)/reaction time (h)	Desulfurization product	Metal-based	Metal-free
LYKAG**Pen**RAEYS **1**/87/12	LYKAG**V**RAEYS	61	98
LYKAH**Pen**RAEYS **2**/70/24	LYKAH**V**RAEYS	–	93
LYKAM**Pen**RAEYS **3**/65/24	LYKAM**V**RAEYS	–	77
LYKAL**Pen**RAEYS **4**/82/48	LYKAL**V**RAEYS	–	79
TLQNREHETNG**Pen**AKSDQKQEQL **5**/78/24	TLQNREHETNG**V**AKSDQKQEQL	0	72
LKKPFNRPQG**Pen**QPKTGPFEDLK **6**/87/24	LKKPFNRPQG**V**QPKTGPFEDLK	54	91



Scheme 6 Reagents and conditions: (a) 6.0 M Gn·HCl, 188.8 mM Na_2HPO_4, 7.2 mM *p*-NO_2PhOH, pH 6.3–6.5, TCEP, 30°C, 2–3 h; (b) 6.0 M Gn·HCl, 188.8 mM Na_2HPO_4, 18.8 mM TCEP·HCl, pH 6.3–6.5, TCEP, VA-044, 37°C, 2–3 h



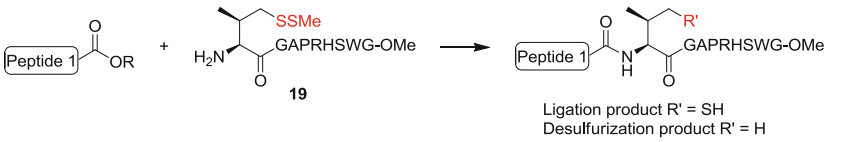
Scheme 7 Synthetic route to compound **18**. Reagents and conditions: (a) KHMDS, MeI, THF, −78°C, 3 h, 97%; (b) DIBAL-H, THF, −35°C, 1 h, 83%; (c) (1) MsCl, Et_3N, CH_2Cl_2, 0 °C, 1 h; (2) AcSH, DBU, DMF, rt, 16 h, 73%; (d) (1) 1 N NaOH, MeOH, 0°C, 10 min; (2) MMTS, Et_3N, CH_2Cl_2, rt, 30 min, 86%; (e) HCl in EtOAc, rt, 82%

from Fmoc-Asp-O*t*Bu in several steps [67]. Initially, methylation at the β position of **13** produced compound **14** in syn/anti mixture with the ratio of 1:1. Diastereopure alcohol **15** was obtained by selective reduction of **14** and chromatographic separation. Then the acetylated thiol was introduced to replace the hydroxyl group. Further transformation led to the target γ-thiol valine derivative **18**, which was then employed in the ligation of a range of peptides (Scheme 7). γ-Thiol valine displayed much higher efficiency than that of β-thiol valine residue, especially in the case of extremely sterically hindered amino acids ligation. The versatility of this approach was further certified by the effective ligation of glycopeptide **20** and peptide **19** containing γ-thiol valine moiety (Table 4).

3.1.2 Ligation at Lysine Site

Lysine is a common amino acid in proteins and both amino groups of lysine provide a wide platform for post-translational protein modifications. Because of the two amino groups in lysine, the mercapto group could be introduced either in the γ or δ position. In 2009, the Liu X.-W. and Liu C.-F. group first reported a native chemical ligation at Lys residue [68]. They introduced a single thiol group on the γ-carbon of lysine derivative **22** (Scheme 8), which was prepared starting from (*S*)-Boc-Asp-O*t*Bu **21**. Treatment of **22** with TBAF, MsCl, and CH_3COSK in turn furnished **23** in good yields. The thiol was unmasked and further protected in disulfide form using MMTS to access **24**. Subjecting **24** to 95% TFA and following protection with

Table 4 NCL at Val site



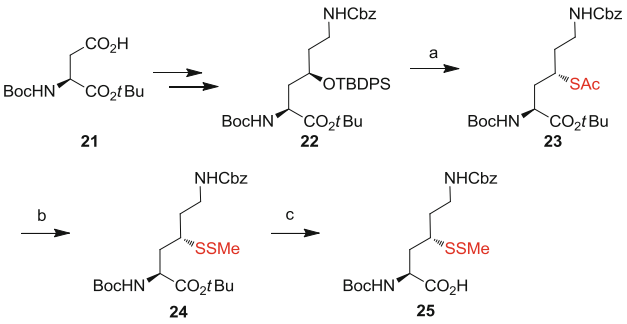
Peptide 1-CO_2R	Ligation[a] yield/time	Desulfurization[b] product/yield/time
FmocThzRGDSCysRPGQ(Acm)-OPh-*m*SSEt	78%/1 h	FmocThzRGDSCysRPGQ**V**GAPRHSWG-OMe (Acm) 84%/3 h
FmocKYDSRGF-OPh-*p*CN	80%/1 h	FmocKYDSRGF**V**GAPRHSWG-OMe 81%/3 h
FmocRTGDSAGT-SPh	87%/4 h	FmocRTGDSAGT**V**GAPRHSWG-OMe 89% /3 h
FmocVRSYTAGP-OPh-*p*CN	55%/1 h	FmocVRSYTAGP**V**GAPRHSWG-OMe 98%/3 h
Fmoc-RLGNSTAGQ-OPh*m*-SSEt **20**	90%/1 h	Fmoc-RLGNSTAGQ**V**GAPRHSWG-OMe 89%/3 h

[a]pH 6.5 buffer, TCEP, rt, 30°C
[b]TCEP, VA-044, 37°C

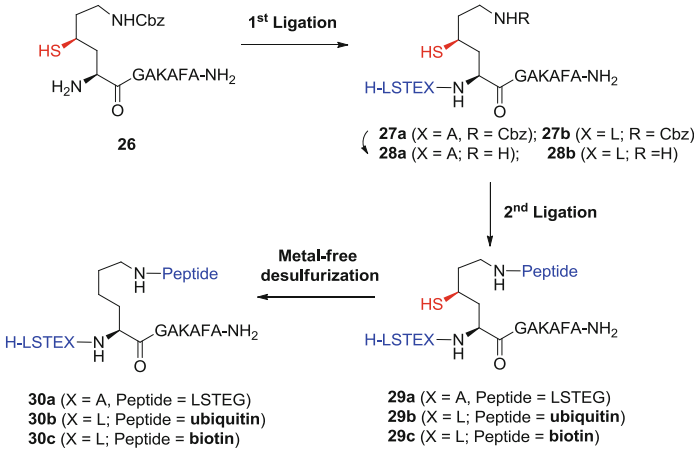
Boc_2O led to the protected form of 4-mercaptolysine **25**, which was then introduced to the N-terminal peptide **26**.

After unmasking the thiol group in situ, peptide **26** was successfully ligated with H-LSTEA-COSR and H-LSTEL-COSR, respectively. A cascade ligation on the ε-amino group followed by desulfurization could give ubiquitin (Scheme 9). A control reaction indicated that the γ-SH of 4-mercoptolysine derivatives played a vital role in mediating ligation at both α- and ε-NH_2. Furthermore, the results of ligation reactions revealed that the bulkiness of Leu did not affect the ligation rate or yield.

The γ-mercapto group is extremely important, allowing chemical ligation at both α and ε amines in a so-called unique "one-stone-two-birds fashion." However, the harsh deprotecting condition for Cbz group limited the application of this strategy. Thus, Cbz group on the ε-amino group was replaced by a photolabile protecting group *o*-nitroveratryloxycarbonyl (NVOC) to facilitate its application [69]. The protected form of 4-mercaptolysine **31** was synthesized using the similar procedures as **25**. Ligation of peptide thioester **32** with C-terminal segment **33** including 4-mercaptolysine proceed efficiently in a pH 7.5 buffer (6 M Gn·HCl, 0.2 M



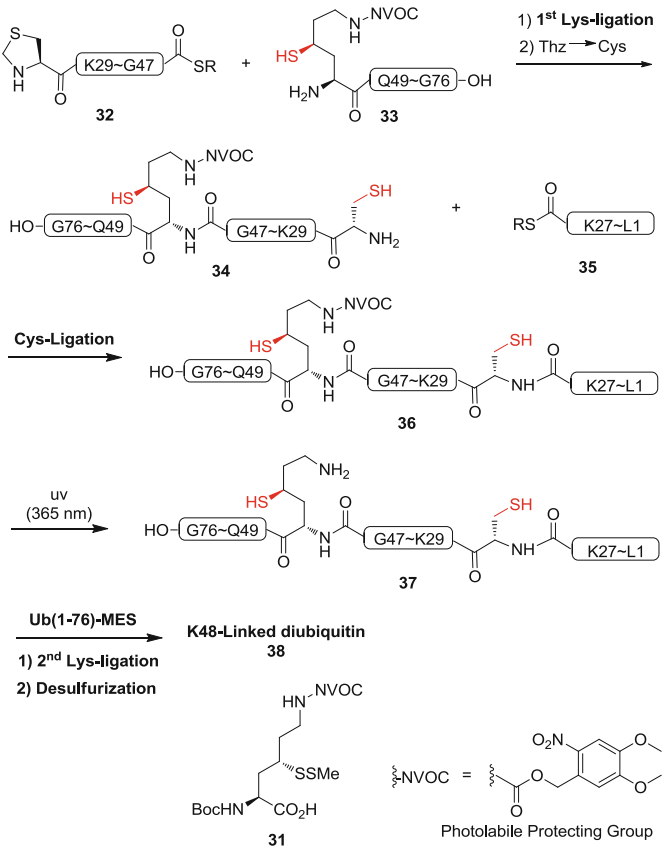
Scheme 8 Synthetic route to compound **25**. Reagents and conditions: (a) (1) TBAF, THF, 0°C, 77%; (2) MsCl, DIPEA, 0 °C; (3) AcSK, DMF, 40 °C, 70% over two steps; (b) (1) NaOH, MeOH, rt; (2) (*S*)-methyl methanethiosulfonate (MMTS), Et_3N, CH_2Cl_2, rt, 50% over two steps; (c) (1) 95% TFA, H_2O, rt; (2) Boc_2O/TEA. MeOH, rt, 78% over two steps



Scheme 9 Synthesis of peptides by dual ligation

phosphate, 20 mM TCEP) and spontaneously completed the conversion of Thz (1,3-thiazolidine-4-carboxo group) to Cys to afford **34**. Further ligation of **34** with peptide thioester **35** in pH 7.5 buffer (6 M Gn·HCl, 0.2 M phosphate, 0.2 M MESNa, 20 mM TCEP) furnished **36**. Desulfurization was conducted under metal-free conditions to provide monoubiquitin **37**. Ligation of **37** with Ub(1–76)-MES in pH 8.0 (6 M Gn·HCl, 0.1 M phosphate, 40 mM TCEP, 1 vol.% benzyl mercaptan) and desulfurization afforded the K48(4-SH)-containing ubiquitin **38** (Scheme 10). All the intermediates and the final product were purified by C18 semi-preparative HPLC. Moreover, this approach executed two consecutive ligation steps via the same thiol to synthesize many proteins containing modified lysine residues.

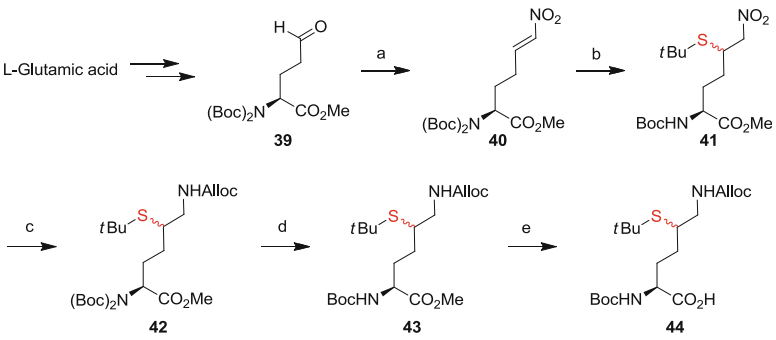
In 2009, Brik et al. also investigated the Lys-ligation using δ-mercaptolysine [70, 71]. The synthesis of modified lysine started from L-glutamic acid, which was



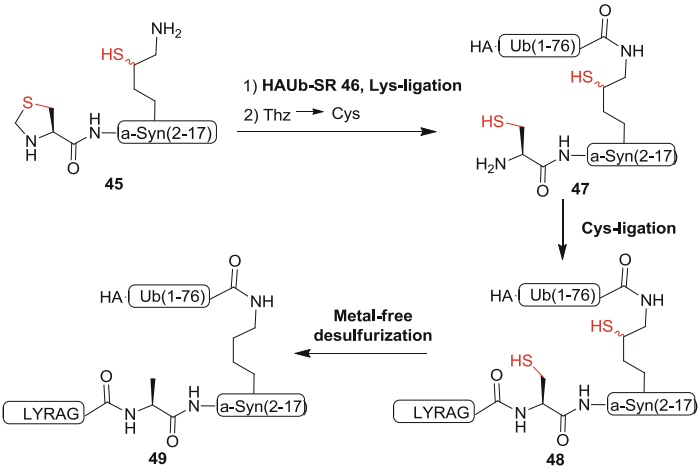
Scheme 10 Synthesis of K48-diubiquitin

converted into aldehyde **39** in three steps. A Henry reaction of **39** with nitromethane followed by dehydration resulted in the E/Z mixture of conjugated nitro olefin **40**. A Michael reaction of *t*BuSLi with **40** at low temperature afforded a diastereomeric mixture of compound **41**. The reduction of nitro group followed by masking with allyloxycarbonyl group gave the alloc-protected *tert*-butyl mercaptolysine **42**, which was converted to **43** and further saponified to produce the desired δ-(*R*,*S*)-mercaptolysine **44** in quantitative yield (Scheme 11).

This δ-mercaptolysine building block was efficiently utilized in the ubiquitylation of peptides. Thz-a-syn(2–17) **45** installed with the δ-mercaptolysine was ligated with HAUb-SR **46** and, following unmasking of the N-terminal cysteine, resulted in peptide **47**. A further ligation of **47** with LYRAF-SR (6 M Gn·HCl, pH 7.5 and 2 vol.% of benzyl mercaptan and thiophenol for 4 h) and desulfurization using Danishefsky's method formed ubiquitylated protein **49** (Scheme 12). In this methodology, they demonstrated that the configuration of the δ-carbon on the δ-mercaptolysine did not affect the ligation. Subsequently, the Brik



Scheme 11 Synthesis of δ-(*R*, *S*)-mercaptolysine **44**. Reagents and conditions: (a) (1) $MeNO_2$, TBAF, 97%; (2) Ac_2O, DMAP, Et_2O, 0°C to rt, 4 h, 71%; (b) *t*BuSH, BuLi, −10 to −78°C, 45 min, 85%; (c) (1) $NaBH_4$, $NiCl_2{\cdot}6H_2O$, THF/MeOH(1:1), −20 to −15°C, 20 min; (2) AllocCl, Et_3N, THF, 0°C to rt, 1 h, 69%; (d) (1) HCl, EtOAc, −20°C, 1 h; (2) Boc_2O, Et_3N, MeOH, 0°C to rt, 2 h, 96%; (e) LiOH, THF/H_2O, 0°C, 1 h, 95%



Scheme 12 Utilize δ-mercaptolysine building block in the ubiquitylation of peptides

group used this method to synthesize a series of di-ubiquitin and tetraUb chains. They indicated that it is also possible to introduce unnatural elements such as specific chromophores [72, 73].

Ovaa et al. prepared both δ- and γ-lysine starting from δ-hydroxylysine and L-lysine [74, 75]. These two kinds of modified lysine were applied in the synthesis of K48-Ub and K43-Ub and the ligation efficiencies were compared. It was found that both δ- and γ-lysine showed nearly equal efficiency in terms of reaction rate and yields. These lysine-assisted ligation strategies contribute greatly to the synthesis of Ub and to the understanding of all their structural and biological properties [76].

3.1.3 Ligation at Threonine Site

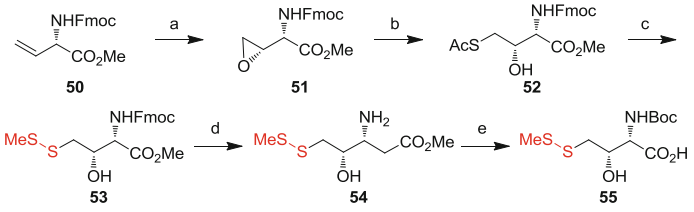
Compared to cystine, threonine is relatively abundant in nature. Chen et al. found an efficient way to merge two different glycopeptides by incorporating γ-thiol threonine at the N-terminus of peptide, which was ligated with C-terminal peptide thioester and subsequently desulfurized to provide the desired peptide [77]. The γ-thiol Thr building block was prepared from D-vinylglycine through modification of Rapoport's route. Epoxidation of vinylglycine **50** gave *syn*-**51** and *anti*-**51** in a ratio of 5:1. The major *syn*-epoxide **51** was separated and opened by thioacetate to afford acetylated thiol **52**, which was then changed to disulfide form **53**. A routinely transformation of **53** could provide the γ-thiol Thr variants **54** and **55** (Scheme 13).

The feasibility of this protocol was evaluated in the assembly of the threonine variant **54** with various peptides. As expected, the ligation with peptides bearing a less sterically hindered C-terminus (Ala, Gln, Phe, Trp, Tyr) proceeded quickly and gave good yields; peptides with a more sterically hindered C-terminus (Val, Leu, Ile, Pro) suffered slow reaction rate but provided a reasonable yield. Subsequent desulfurization under standard radical-based conditions furnished the corresponding threonine extension products in good yields. It should be noted that a variety of C-terminal esters including thiophenyl ester, *ortho*-thiophenolic ester, and *para*-thiophenyl ester were efficient in the ligation step. In particular, *ortho*-thiophenolic ester is compatible with glycopeptide ligation.

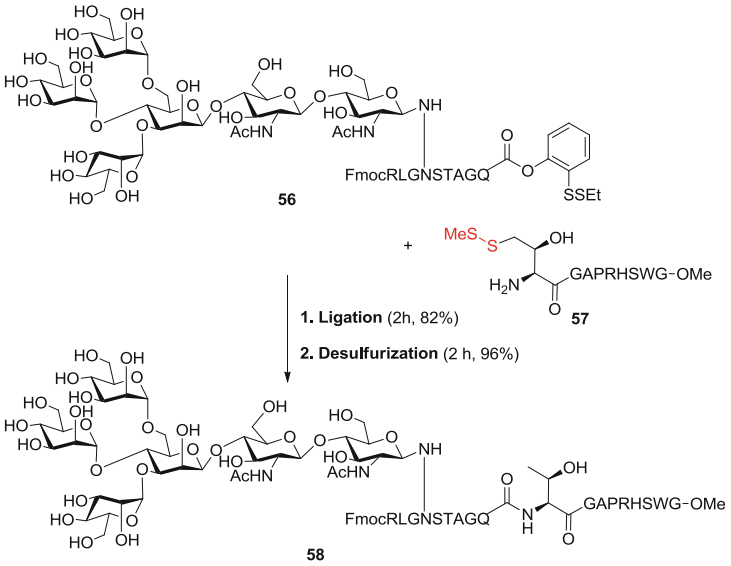
The versatility of the threonine ligation/desulfurization approach was further explored by the coupling of two different peptides. Peptide **57** incorporating the γ-thiol surrogate at N-terminus, prepared from **55**, was successfully ligated with its partner **56** under TCEP in Gn·HCl buffer. Subsequent radical-based desulfurization readily provided the target products **58** (Scheme 14).

3.1.4 Ligation at Leucine Site

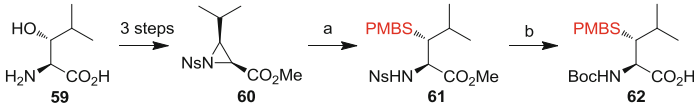
In 2010, the Brik group extended the ligation at Leu residue and demonstrated its use in the synthesis of HIV-1Tat protein [78]. The key building block



Scheme 13 Synthesis of γ-thiol threonine **55**. Reagents and conditions: (a) *m*CPBA, CH_2Cl_2, 0°C to rt, 24 h; (b) AcSH, NaOAc, toluene, DMF, rt, 2 h; (c) (1) 0.2 M NaOH, MeOH, 0°C, 20 min; (2) MMTS, DIPEA, CH_2Cl_2, rt, 2 h; (d) Et_2NH, DMF, rt, 2 h; (e) (1) Boc_2O, THF, MeOH, Et_3N; (2) 1 N NaOH, H_2O, THF

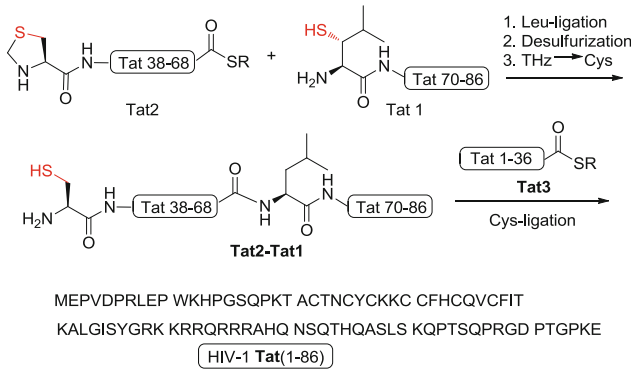


Scheme 14 Glycoprotein synthesis via Thr-ligation

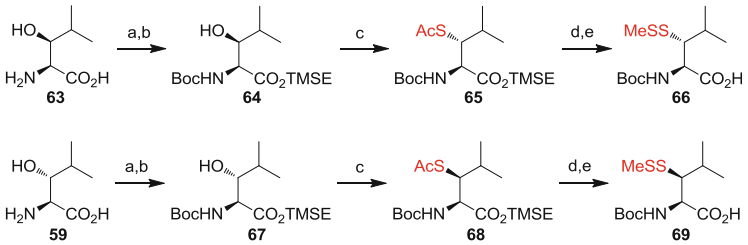


Scheme 15 Synthesis of β-mercaptoleucine **62**. Reagents and conditions: (a) *p*-MeOPhCH$_2$SH, $BF_3{\cdot}OEt_2$, CH_2Cl_2, 0°C, 40% (+60% isomer); (b) (1) *p*-MeOPhSH, K_2CO_3, rt, MeCN/DMSO; (2) Boc_2O, DIPEA, CH_2Cl_2, 0°C; (3) 1 M LiOH, THF/H_2O, 0°C, 40% over three steps

β-mercaptoleucine was achieved from β-hydroxy-L-leucine **59**, which was converted to **60** in three steps. Ring opening of **60** with PMB-SH afforded a mixture of **61** and its regio isomer, which could be separated by flash chromatography. Compound **61** was transformed to the protected β-mercaptoleucine **62** by changing the protective group on amino group and hydrolysis of ester (Scheme 15). The efficacy of ligation at Leu was then illustrated by model reaction of C-terminal peptide of HIV-1 Tat(69–86) fused with β-mercaptoleucine and various peptide thioesters (Scheme 16). The ligation and desulfurization under metal-free conditions furnished the desired product in excellent yields, albeit with slower ligation rate compared to Cys. This is to be expected because of the intrinsic steric hindrance of β-mercaptoleucine. The β-mercaptoleucine-assisted ligation was then utilized in the synthesis of HIV-1Tat. It was noticeable that the close location of several Cys residues hindered the use of NCL at Cys and their presence required selective desulfurization. Thus, Brik et al. utilized a β-mercaptoleucine promoted ligation of Tat 1 and Tat 2 followed by metal-free desulfurization, and Cys-NCL with Tat 3 to accomplish the synthesis of HIV-1 Tat.



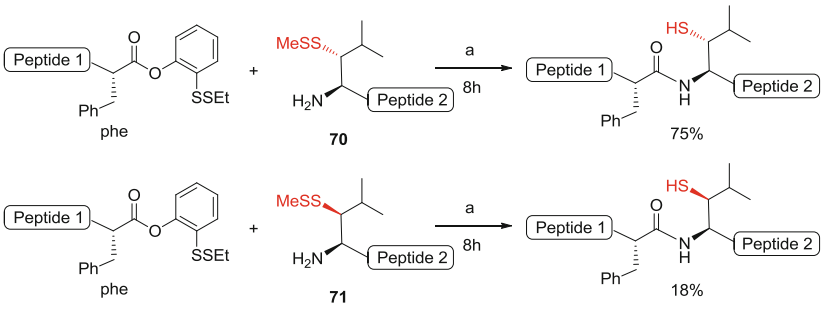
Scheme 16 Synthetic strategy of HIV-1 Tat



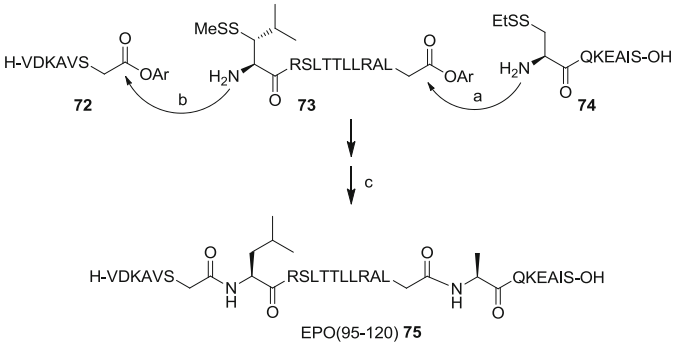
Scheme 17 Synthesis of compound **66** and **69**. Reagents and conditions: (a) Boc_2O, Na_2CO_3, THF/H_2O, rt, 91%; (b) TMSE-OH, DCC, DMAP, CH_2Cl_2, 0°C to rt, 99%; (c) (1) MsCl, Et_3N, CH_2Cl_2, 0°C; (2) AcSK, DMF, rt, 40–60°C, 82%; (d) (1) NaOH, MeOH, 0°C; (2) MMTS, DIPEA, CH_2Cl_2, rt, 79%; (e) TBAF, THF, rt, 98%

At the same time, Tan et al. exploited synthetic routes to epimeric Leu building blocks and studied their ligation efficiency separately [79]. Two different diastereomers of β-thioleucine surrogate, **66** and **69**, were prepared from (2*S*,3*S*)-3-hydroxyleucine and (2*S*,3*R*)-3-hydroxyleucine, respectively (Scheme 17). Sulfyl moiety was introduced by substitution of β-hydroxyl group with thiol acetate and then transferred to disulfide form. Compounds **66** and **69** merged with N-terminal peptides to give **70** and **71**, which were ligated with C-terminal peptide under standard conditions. The validation results showed that configuration of β-C on Leu significantly influence the ligation efficacy (Scheme 18). Peptide **70** was readily coupled with C-terminal peptides under standard conditions, whereas peptide **71** gave poor reactivity. The capacity of this Leu-mediated ligation combined with metal-free desulfurization was further demonstrated in the synthesis of EPO (95–120) (Scheme 19).

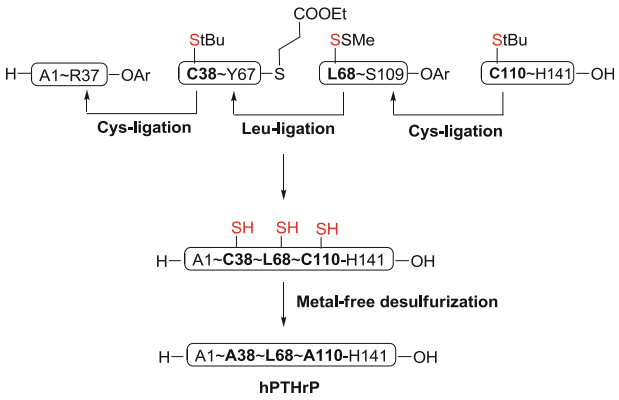
Danishefsky's group then applied this approach to the chemical synthesis of hPTHrP and ATAd2 [80, 81]. hPTHrP is a protein widespread in human tissues. The synthesis was encompassed by iterative ligations of four component peptides and desulfurization in a convergent manner (Scheme 20). The two Cys-ligations



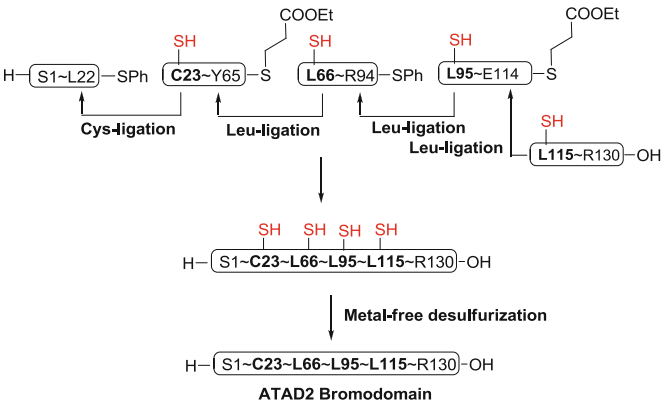
Scheme 18 Leucine ligation with two leu(SSMe) diastereomers. Reagents and conditions: (a) 6 M Gn·HCl, 100 mM NaH_2PO_4, 50 mM TCEP, pH 7.5. Peptide 1: GKHLNSAERVE-; Peptide 2: -RKKLQDVHNFVALG-OMe



Scheme 19 Synthesis of EPO(95–120). Reagents and conditions: (a) 6 M Gn·HCl, 100 mM NaH_2PO_4, 50 mM TCEP, pH 7.5, 0.5 h; (b) 6 M Gn·HCl, 100 mM NaH_2PO_4, 50 mM TCEP, pH 7.5, 0.5 h; MESNa, H_2O/MeCN, 1 min, 61% over two steps; (c) TCEP, VA-044, *t*BuSH, 1 h, 82%



Scheme 20 Synthetic strategy to hTPHrP



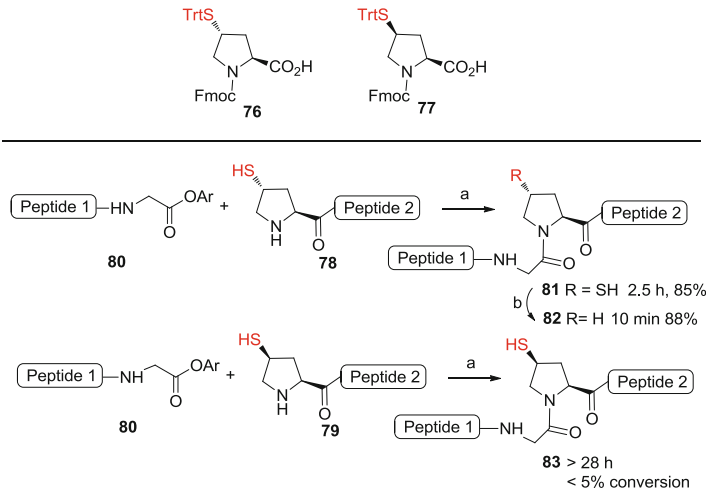
Scheme 21 Synthetic strategy to ATAD2 bromodomain

and one Leu-ligation followed by subsequent global desulfurization under metal-free conditions furnished hPTHrP(1–141) in a 16% total yield with high purity [80].

ATAD2, a bromodomain protein related to gene expression and transcriptional regulation, could also be synthesized by means of the Cys-ligation and Leu-ligation (Scheme 21). This 130-mer peptide was assembled from five fragments via one Cys-ligation and three Leu-ligations in a highly convergent manner with only three total RP-HPLC events. The global desulfurization was also achieved under metal-free conditions [81].

3.1.5 Ligation at Proline Site

Proline ligation had been a problematic issue indicated by the lower reactivity of C-terminal proline thioester. Danishefsky group and Otaka group exploited HSPro-mediated ligation by attaching proline moiety to N-terminal peptides [82–84]. Commercially available diastereomeric 4-thioprolines **76** and **77** were appended to peptides **78** and **79** by Fmoc-SPPS, respectively. The ligation efficiencies of **78** and **79** with peptide **80** bearing a 2-(ethyldithiophenyl)ester were then evaluated. Both **78** and **79** went smoothly during trans-thioesterification. Interestingly, peptide **78** readily underwent the S → N acyl transfer at room temperature, whereas peptide **79** failed (Scheme 22). This was thought to be caused by the highly hindered peptide **77** moiety restraining the S → N acyl transformation. The ligation of **78** with peptides bearing less sterically hindered C-termini, such as Gln and Phe, was readily implemented under standard conditions and desulfurized to give the target products. Those peptides with Val and Pro C-termini showed poor reactivities. It was obvious that the efficiency of Pro-ligation was significantly influenced by the size of the C-terminal peptide. A further investigation indicated that the ligation efficiency could be improved by replacing thiol-Pro moiety with seleno-Pro and



Scheme 22 Ligation with two Pro(SH) diastereomers. Reagents and conditions: (a) 6 M Gn·HCl, 100 mM NaH_2PO_4, 50 mM TCEP, pH 7.5; (b) TCEP, VA-044, *t*BuSH, MeCN/H_2O, 37°C, 10 min, 88%. Peptide 1 = ALLVNSS-; Peptide 2: -WEPLN; Ar = 2-(ethyldithio)phenyl

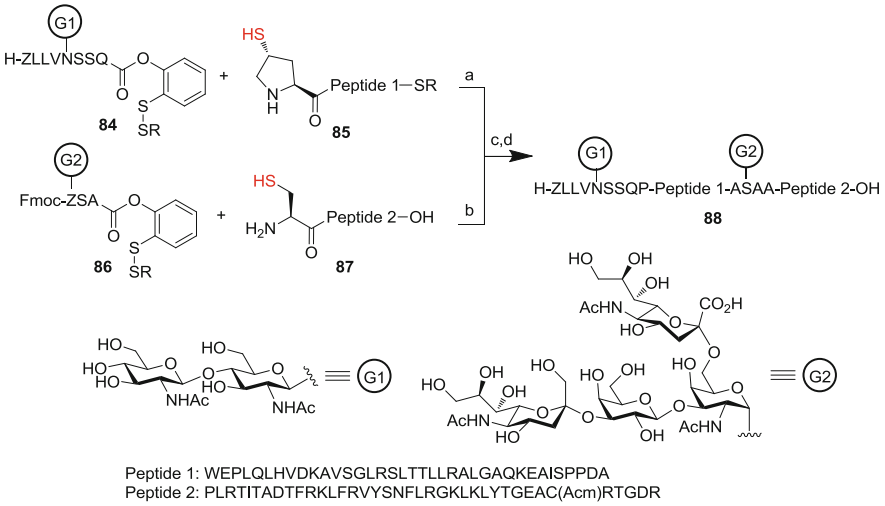
deselenation could be achieved under dithiothreitol, 6 M Gn · HCl, 100 mM NaH_2PO_4, followed by TCEP (pH 5–6) [83].

The Pro ligation was later employed in the synthesis of hEPO(79–166) (Scheme 23) [83]. The main skeleton was constructed by a Pro-ligation and two Cys-ligations. A subsequent radical-based desulfurization could provide the target molecule.

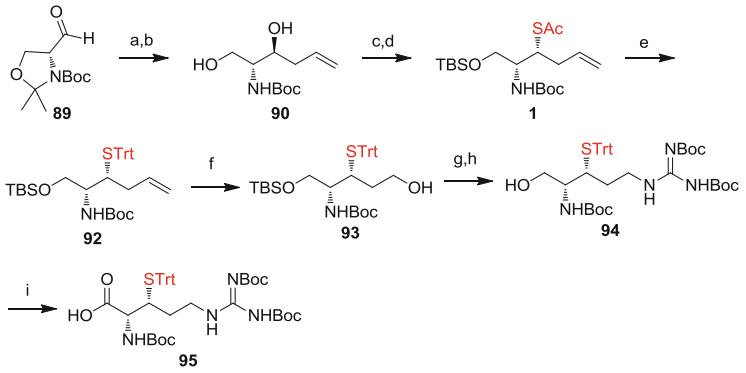
3.1.6 Ligation at Arginine Site

To expand the ligation-desulfurization methodology, Payne et al. developed a practical route to synthesize protected β-thiol arginine and applied it to peptide synthesis [85]. Building block **95** was synthesized starting from Garner's aldehyde **89**. Introduction of an allyl group on the carbonyl group and deprotection of the hemiaminal provided compound **90**. Selective protection of primary alcohol and subsequent import of the sulfur moiety afforded compound **91**. To facilitate incorporation into peptides through SPPS, the acetyl group on thiol was converted to an *S*-trityl (Trt) group to access compound **92**. The terminal alkene of **92** was oxidized and subsequently reduced to give alcohol **93**. A Mitsunobu reaction of the *syn* diastereomer of **93** with globally protected guanidine followed by deprotection of the TBS ether resulted in alcohol **94**, which was oxidized to the protected thiol arginine **95** (Scheme 24).

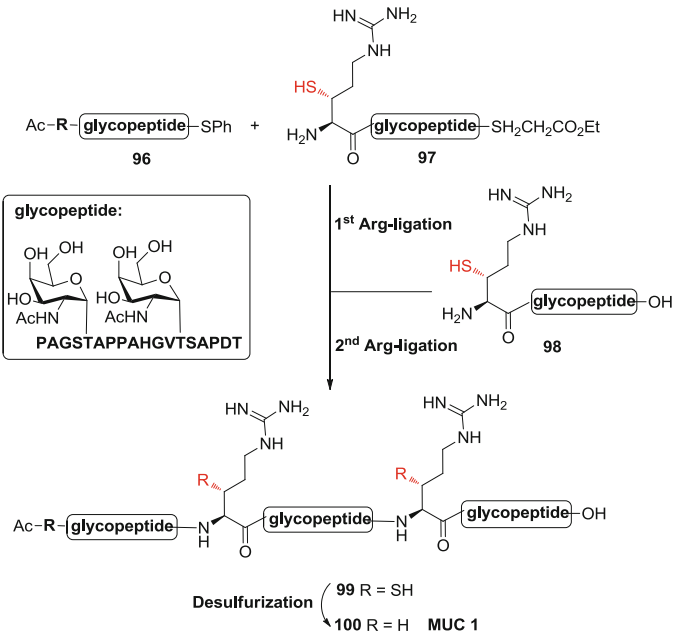
Incorporation of **95** into the pentapeptide through SPPS and successful ligation with a range of peptide thioesters demonstrated the efficiency of this building block



Scheme 23 Synthesis of hEPO(79–166) glycopeptide. Reagents and conditions: (a) (1) 6 M Gn·HCl, 100 mM NaH_2PO_4, 50 mM TCEP, pH 7.5, 67%; (2) piperidine, DMSO, 61%; (3) 0.2 M $MeONH_2$, 60%; (b) 6 M Gn·HCl, 100 mM NaH_2PO_4, 50 mM TCEP, pH 7.5, 23%; (c) 6 M Gn·HCl, 100 mM NaH_2PO_4, 50 mM TCEP, 200 mM MPAA, pH 7.8, 40%; (d) TCEP, VA-044, *t*BuSH. R = $CH_2CH_2CO_2Et$



Scheme 24 Synthesis of β-thiol Arg. Reagents and conditions: (a) allyltributyltin, $BF_3{\cdot}OEt_2$, CH_2Cl_2, −78°C, 3 h, 80%; (b) *p*-TsOH, 1,4-dioxane, rt, 3 h, 82%; (c) TBSCl, Et_3N, DMAP, CH_2Cl_2, rt, 16 h, 87%; (d) (1) MsCl, Et_3N, CH_2Cl_2, 0°C, 20 min; (2) KSAc, DMF, 50°C, 65% over two steps; (e) (1) NaOMe, MeOH, rt, 5 min; (2) Trt-OH, $BF_3{\cdot}OEt_2$, Et_2O, rt, 45 min, 76% over two steps; (f) (1) OsO_4, $NaIO_4$, 2,6-lutidine, H_2O/1,4-dioxane (3:1 v/v), rt, 2 h; (2) $LiBH_4$, THF, rt, 40 h, 50% over two steps; (g) *N*,*N*′,*N*″-*tri*-Boc-guanidine, PPh_3, DIAD, 30°C, 10 min, 80%; (h) TBAF, THF, rt, 1.5 h, 96% (i) (1) DMP, CH_2Cl_2, rt, 2 h; (2) $NaClO_2$, NaH_2PO_4, 1-methylcyclohexane, *t*BuOH/THF/H_2O (1:7:2 v/v/v), rt, 20 min, 37% over two steps



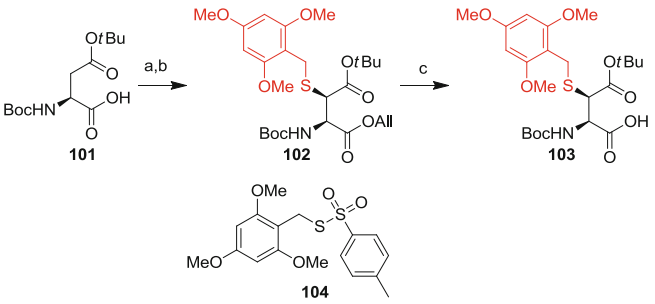
Scheme 25 Assembly of MUC1 glycopeptide

in ligation. A kinetic study indicated that the reaction rate of sterically hindered peptide thioester was much slower than that of unhindered counterparts. However, the overall trend was consistent with those studied for NCL at Cys. Desulfurization reaction at Arg was also sluggish compared to that at Cys, which could be because of the guanidine moiety of the Arg. However, the final yields were satisfactory.

This synthetic methodology was further applied in the synthesis of a fragment of the extracellular domain of mucin 1 (MUC 1). A glycopeptide **96** containing a C-terminal thiophenyl thioester, bifunctional glycopeptides **97** including an N-terminal β-thiol Arg and C-terminal thioester, and glycopeptides **98** with an N-terminal β-thiol Arg, prepared through SPPS, were successfully formed MUC1 by two ligations and desulfurization in one pot (Scheme 25).

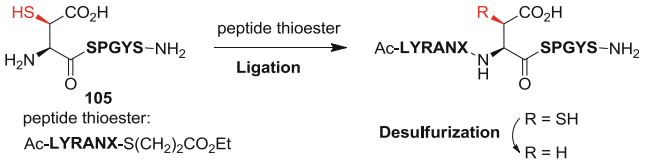
3.1.7 Ligation at Aspartic Acid Site

The Payne group developed a feasible route to a protected β-mercapto aspartate (Asp) building block and applied it in ligation chemistry [86]. The protected β-mercapto aspartate **103** was synthesized from Boc-Asp(O*t*Bu)-OH **101** in three steps. The mercapto moiety on β-C of Asp was introduced by a novel sulfenylating reagent **104** (Scheme 26).



Scheme 26 Synthesis of β-mercapto Asp building block **103**. Reagents and conditions: (a) allyl bromide, iPr$_2$EtN, DMF, 16 h, 94%; (b) LiHMDS (2 equiv.), **104** (1.4 equiv.), THF, −78°C, 56%; (c) Pd(PPh$_3$)$_4$, *N*-methylaniline, THF, 30 min, 80%

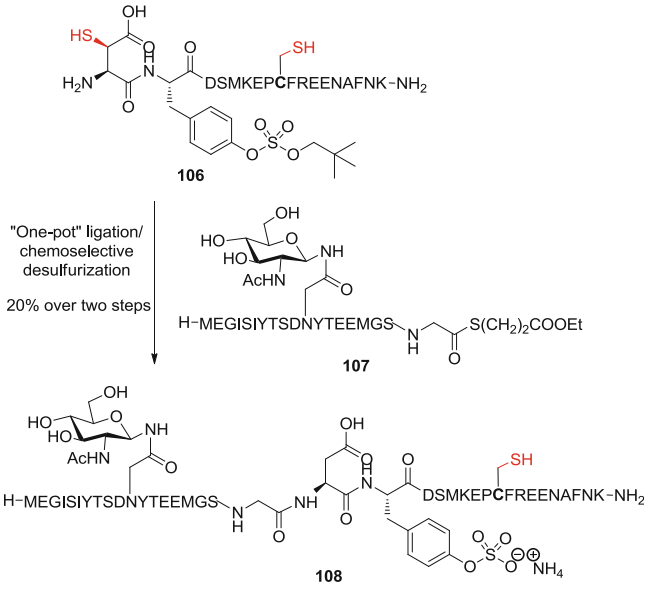
Table 5 Ligation-desulfurizations at Asp site



Entry	X	Ligation[a] yield (%)	Desulfurization[b] yield (%)
1	Gly	80	75
2	Ala	82	71
3	Met	71	63
4	Phe	78	76
5	Val	75	71

[a]6 M Gn·HCl, 200 mM HEPES, 50 mM TCEP, pH 7.3–7.5, PhSH, 37°C, 24 h
[b]6 M Gn·HCl, 200 mM HEPES, 250 mM TCEP, 40 mM reduced glutathione, 20 mM VA-044, pH 6.5–7.0, 37°C, 16 h

β-Mercapto Asp **103** was incorporated into the peptide using standard Fmoc-SPPS. The applicability was then demonstrated by the successful ligation between **105** and a few peptide thioesters (Table 5). It was shown that the ligation rate at β-mercapto Asp was comparable to those of Cys residue and the configuration at the β-carbon did not affect the ligation efficiency. Based on inference and computation, the Payne group further displayed a formal one-pot ligation-desulfurization reaction with β-mercapto Asp residue; however, the thiophenol used in the ligation step must be removed before desulfurization. This strategy was further successfully applied in the assembly of the chemokine receptor CXCR4 (Scheme 27). During the investigation it was observed that pH 3 was optimal for desulfurization of β-mercapto Asp residue. Increase of pH value resulted in sluggish desulfurization rate and worse chemoselectivity. This was reasoned to be because the C-S bond on



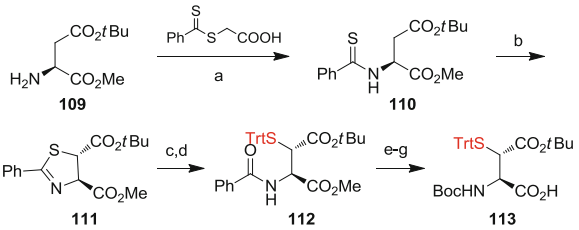
Scheme 27 Synthesis of CXCR4(1–38)

β-mercapto Asp was much weaker when the neighboring carboxylate was protonated. It is a good example showing that desulfurization of β-mercapto Asp residue within ligation products is possible in the presence of unprotected Cys residues.

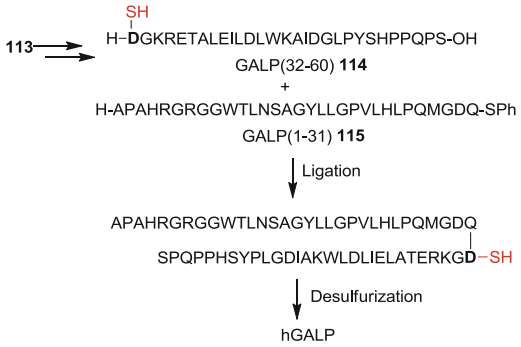
The Tan group prepared β-thioaspartic acid **113**, a similar diastereomer of **103**, through a *trans*-oxazoline intermediate [87] (Scheme 28). To this end, the thiobenzamidomalonic ester **110** was first obtained from Asp, and further iodination of dianion followed by an acidic workup formed the *trans* thiazoline **111**. Unmasking and succeeding protection of the mercapto group gave compound **112**. β-Thioaspartic acid was finally obtained after a successive deprotection and protection. The feasibility and versatility of the β-thioaspartic acid-assisted ligation was then demonstrated by a series of peptide assemblies and the synthesis of hGALP (Scheme 29).

3.1.8 Ligation at the Glutamate Site

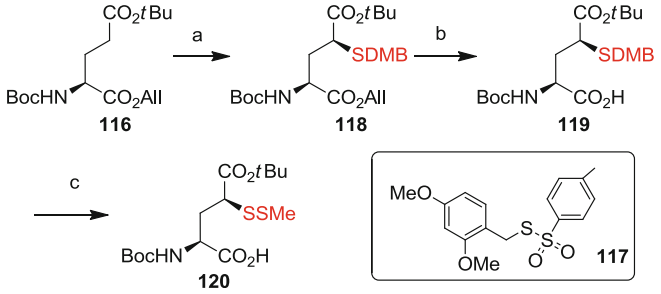
In 2014, the Payne group described an alternative ligation at glutamate [88]. First, the γ-thiol-Glu building block **119** was prepared from Boc-Glu(O*t*Bu)-OAll **116** by installation of a 2,4-dimethoxybenzyl (DMB) thiol at the γ-position and deprotection of allyl ester. Subsequently, γ-thiol-Glu **119** was transformed into disulfide form to give **120** (Scheme 30), which was readily integrated into N-terminal peptide **121** by SPPS. Then ligation of **121** with several peptide thioesters was performed in pH 7.2–7.4 buffer (6 M Gn·HCl, 100 mM Na_2HPO_4,



Scheme 28 Synthesis of β-mercapto Asp building block **113**. Reagents and conditions: (a) Et_3N, pyridine, 97%; (b) (1) LiHDMS, THF, 0°C; (2) I_2, THF, −78°C, 38%; (c) 2 M HCl, THF, 45°C, 54%; (d) Trt-Cl, CH_2Cl_2, 76%; (e) Boc_2O, DMAP, THF, 88%; (f) NH_2NH_2, THF, MeOH; (g) Me_3SnOH, 80°C, DCE, 97%



Scheme 29 Synthesis of hGALP. Ligation condition: 6 M Gn·HCl, 300 mM Na_2HPO_4, 200 mM MPAA, 20 mM TCEP, pH 7.9, 2 h, 56%; desulfurization condition: 3 M Gn·HCl, 100 mM Na_2HPO_4, 250 mM TCEP, 40 mM glutathione, UV (365 nm), pH 6.5, 80%



Scheme 30 Synthesis of γ-mercapto Glu building block. Reagents and conditions: (a) (1) LiHMDS, THF, −78°C, 1 h, (2) **117**, −78°C, 30 min, 83%; (b) $Pd(PPh_3)_4$, $PhSiH_3$, THF, rt, 30 min, quant.; (c) dimethyl(methylthio)sulfonium tetrafluoroborate, MeOH/H_2O, 45 min, rt, 55%

50 Mm TCEP, 2 vol.% PhSH) to afford the product in excellent yields (Table 6). After HPLC purification and lyophilization, the ligation products were subjected to reduced glutathione, TCEP, and VA-044 in pH 6.5–6.8 buffer to provide the native peptide products in excellent yields. Streamlining of this protocol by avoiding

Table 6 Ligation-desulfurization at Glu site

Entry	Thioester (X=)	Ligation yield (%)	Desulfurization yield (%)	"One-pot" yield (%)
1	Gly	72	89	73
2	Ala	77	91	67
3	Met	83	98	72
4	Phe	80	84	74
5	Val	68	98	56

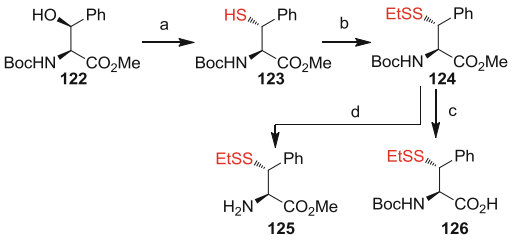
HPLC purification was also attempted. It was achieved by removing excess thiophenol through extraction after the ligation and immediate exposure of the ligation mixture to desulfurization conditions. Although a small amount of byproduct emanating from VA-044 and γ-thiol was observed in the case of peptide thioesters containing C-terminal Ala, Met, Phe, and Val, the overall efficiency and yield is still satisfactory. The asset of "one-pot" ligation-desulfurization reaction at γ-thiol-Glu was further highlighted by the synthesis of teriparatide, a drug for glucocorticoid-induced osteoporosis.

3.2 Metal-Based Desulfurization

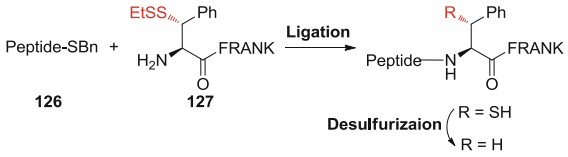
Despite the metal-free desulfurization strategy showing great advantages, such as mildness, user friendliness, and wide functional group tolerance, etc., application of this method on several Cys surrogates such as β-mercaptophenylalanine, β-mercapto tryptophan, and γ-mercapto glutamine were unsuccessful or unreported. The desulfurization was achieved by metal-based conditions on these surrogates.

3.2.1 Ligation at the Phenylalanine Site

Crich and Banerjee reported a feasible synthetic route to the β-mercaptophenylalanine building block and its application in the synthesis of two decapeptides [89]. *threo*-β-Hydroxy-L-phenylalanine derivative **122** was converted to the *erythro*-thiol **123** in accordance with Easton's bromination protocol. Thiol **123** was transformed to disulfide **124**, which was deprotected with acid to



Scheme 31 Synthesis of β-mercapto Phe building block. Reagents and conditions: (a) (1) MsCl, Et_3N, CH_2Cl_2; (2) AcSH, DBU, DMF; (3) 1 M NaOH, MeOH; 55–60% over three steps; (b) EtSS(O)Et, Et_3N, CH_2Cl_2; (c) LiOH, THF, 40% over two steps; (d) TFA, CH_2Cl_2, 76%



Peptide	Ligation yield	Desulfurization yield
126a LYRMG	72%	60%
126b LYRAM	74%	74%

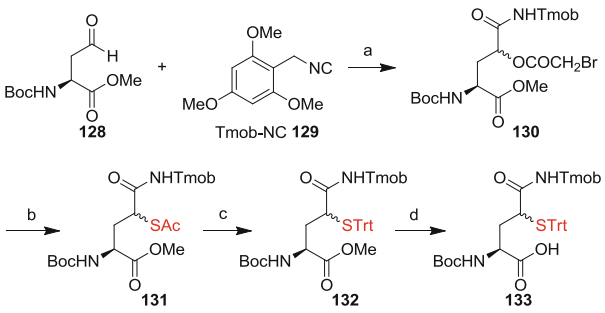
Scheme 32 Decapeptide synthesis. Ligation condition: MESNa, TCEP, 0.1 M Tris–HCl, pH 8; desulfurization condition: $NiCl_2{\cdot}6H_2O$, $NaBH_4$, 0.1 M phosphate buffer, pH 7

give β-mercaptophenylalanine derivative **125** or hydrolyzed under basic condition to form **126** (Scheme 31).

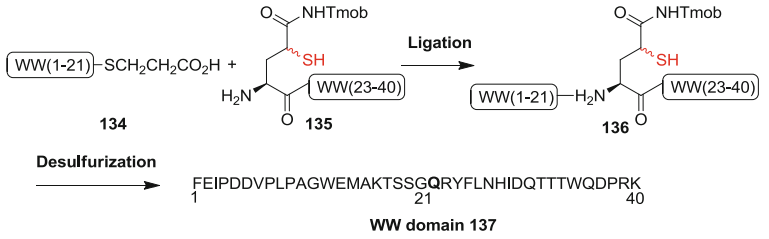
Model ligation of **125** with N-Cbz glycine thioesters and N-Boc-L-methionine thioesters, respectively, in the presence of MESNa in MeCN/Tris buffer (pH 7.5–8.0) indicated that the sterically hindered thioester slows down the ligation rate. Selective desulfurization was fulfilled with the combination of $NaBH_4$ and $NiCl_2$. An ulterior application of this protocol was the ligation of β-(SSEt)-FRANK peptide **127** and thioesters **126a** and **126b**. The reaction was performed in the presence of MESNa and TCEP in Tris buffer (pH 8). After purification, desulfurization was accomplished with $NiCl_2$ and $NaBH_4$ in pH 7 phosphate buffer, providing LYRMGFRANK and LYRAMFRANK in reasonable yields (Scheme 32). It is notable that the desulfurization condition is compatible with the existence of methionine and ACM-protected cysteine.

3.2.2 Ligation at the Glutamine Site

In 2012, the Brik group reported NCL at the Glu site [90]. Racemic γ-mercapto-L-glutamine (mGln) **133** was prepared from aldehyde **128** derived from L-aspartic acid. Aldehyde **128** was converted to **130** by the Passerini three-component reaction. Thiol moiety was then introduced by removal of bromoacetyl group on the



Scheme 33 Synthesis of γ-mercapto Gln building block. Reagents and conditions: (a) $BrCH_2COOH$, CH_2Cl_2, rt, 2 h, 89%; (b) (1) thiourea, $NaHCO_3$, THF/H_2O, 50°C, 1.5 h, 94%; (2) MsCl, Et_3N, CH_2Cl_2, 0°C, 15 min, 99%; (3) CH_3COSH, DBU, DMF, rt, 14 h, 98%; (c) (1) 1 M NaOH, MeOH, 0°C; (2) Trt-Cl, Et_3N, CH_2Cl_2, rt, 1 h, 70%; (d) 0.3 M LiOH, THF/H_2O 4:1, 0°C, 1 h, 92%



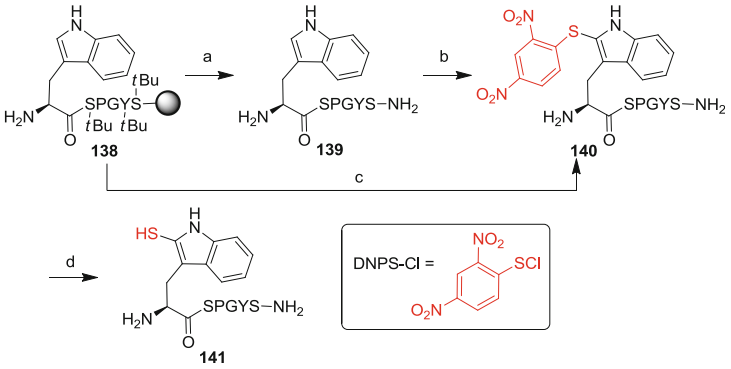
Scheme 34 Synthesis of WW domain. Ligation condition: 6 M Gn·HCl, 200 mM Na_2HPO_4, 100 mM MPAA, 20 mM TCEP, pH 7.0, 37°C, 50%; desulfurization condition: Ni/H_2, 20% AcOH

γ-hydroxyl group, mesylation, and substitution with thioacetic acid. Changing the protecting group on thiol and hydrolysis of ester provided γ-(*R*,*S*)-mercapto-L-glutamine **133** (Scheme 33).

Glutamine **133** was readily incorporated into mGln-WW-(23–40) **135** through SPPS. Ligation of peptide **135** with various peptide thioesters was successful and gave good yields. The chirality on both the α- and γ-carbons of **133** did not affect the ligation-desulfurization efficiency at all. The subsequent desulfurization step was conducted under $NiBr_2$, which was circumvented by routine metal-free conditions in this case. It was possibly imputed to the strong propensity of forming a radical on the γ-thiol. Subsequently, WW domain **137** was successfully synthesized by employing this Gln ligation (Scheme 34). It was believed that this strategy is beneficial to access proteins bearing repeated polyQ units, which are related to several inherited neurodegenerative diseases.

3.2.3 Ligation at the Tryptophan Site

Many more ligation strategies other than Cys were developed, but they are not widely used because of the tedious synthetic procedures of building blocks. Most



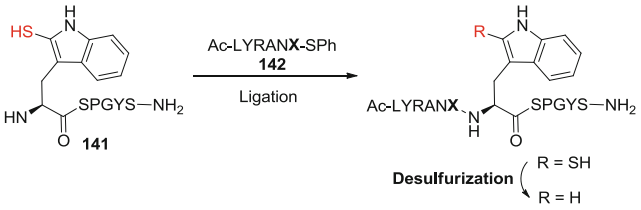
Scheme 35 Synthesis of 2-thiol Trp peptide **141**. Reagents and conditions: (a) TFA/iPr$_3$SiH/H$_2$O (90:5:5 v/v/v), rt, 2 h, 88%; (b) DNPS-Cl, AcOH, rt, 16 h, 56%; (c) (1) DNPS-Cl, AcOH, DMF, rt, 2 h; (2) TFA/iPr$_3$SiH/H$_2$O (90:5:5 v/v/v), rt, 2 h, 38%; (d) PhSH, 6 M Gn·HCl, 0.1 M Na_2HPO_4, pH 8.0, rt, 3 h, 84%

amino acid surrogates required numerous synthetic steps except Val and Pro in current reports. The Payne group developed 2-thiol Trp-mediated ligation in a simple and rapid way [91]. Peptide **138** was prepared by standard Fmoc-SPPS on Rink amide resin. 2-Thiol moiety was directly introduced by DNPS-Cl in the form of resin. Acidic cleavage and further thiolysis generated 2-thiol Trp containing peptide **141**, which could only be isolated by lyophilization because of the propensity for oxidation to the disulfide form (Scheme 35).

A few C-terminal peptide thioesters were synthesized and ligated with peptide **141**. It was found that peptide with alkyl thioesters could not facilitate the ligation under standard NCL conditions (6 M Gn·HCl/0.1 M Na_2HPO_4/100 mM TCEP, 5 mM concentration with respect to **141**, pH 7.3–7.5, 2 vol.% PhSH, 37°C) because of the exogenous aryl thiol attaching at the 2-position of indole. However, reaction of activated thiophenyl ester **142a** with **141** could generate the ligation product at a lower concentration. It was found that reaction of **141** with increased sterically hindered **142b** could form the ligation product, along with a small amount of epimerized thiophenyl ester. This shortcoming was surmounted in optimal conditions (4 mM concentration with respect to peptide **141**, pH 6.5–6.7). The first attempt to access the native peptide product conducted under standard radical-based desulfurization conditions failed, probably because of the intrinsic strength of the C–S bond in the 2-thiol indole fragment. Finally, the thiol auxiliary was removed by standard metal-based desulfurization conditions (Pd on Al_2O_3 in pH 5.8 buffer in the presence of H_2 at 0°C) (Table 7).

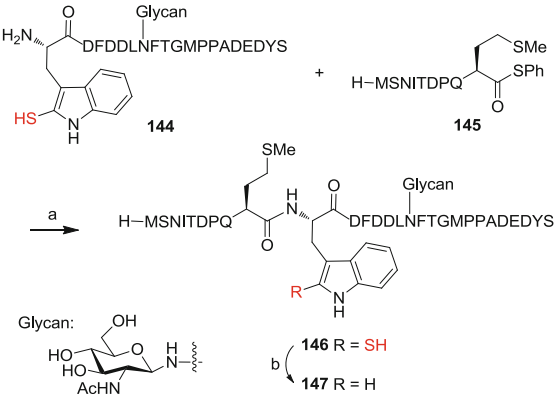
The ligation of a peptide possessing an N-terminal 2-thiol Trp residue and C-terminal Met residue with thiophenyl ester **142d** and Ac-LYRC(Acm)NG-SPh **143** followed by reductive desulfurization generated the desired product without Met desulfurization. This methodology was again used in the synthesis of CXCR1 (1–28) (Scheme 36). CXCR1(10–28) **144** was obtained by standard Fmoc-SPPS, including incorporation of both sugar and thio moiety. The peptide **144** and

Table 7 Ligation of 2-thiol Trp peptide with peptide thioester and subsequent desulfurization



Entry	X=	Ligation yield (%)	Desulfurization yield
1	Gly (**142a**)	71	71
2	Ala (**142b**)	81	89
3	Met (**142c**)	80	61
4	Phe (**142d**)	65	72
5	Pro (**142e**)	58	82

Ligation conditions: 6 M Gn·HCl, 0.1 M Na_2HPO_4, 100 mM TCEP, pH 6.5–6.7, 37°C, 24 h. Desulfurization conditions: Pd/Al_2O_3, 6 M Gn·HCl, 0.1 M Na_2HPO_4, pH 5.8, H_2, 0°C, 4 h



Scheme 36 Synthesis of CXCR1 (1–28). Reagents and conditions: (a) 6 M Gn·HCl, 0.1 M Na_2HPO_4, 100 mM TCEP, pH 6.6, 37°C, 5 h, 91%; (b) (1) Pd/Al_2O_3, 6 M Gn·HCl, 0.1 M Na_2HPO_4, pH 5.8, H_2, 0°C, 5 min; (2) thiourea, 79%

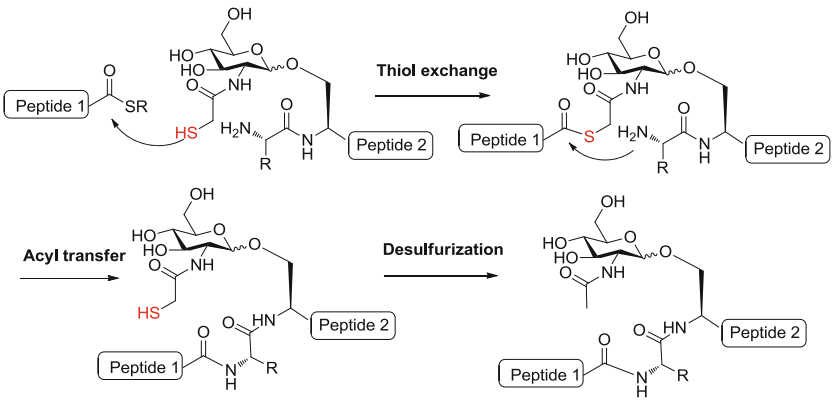
thiophenyl ester **145** were ligated under optimal conditions to give the target product in excellent yield. Subsequent desulfurization was fulfilled under the above-mentioned condition in a shorter time (5 min) to avoid the side reaction of Met desulfurization. This 2-thiol Trp ligation-desulfurization approach was highly efficient in the assembly of peptides. Moreover, the 2-thiol moiety on Trp could be introduced through the solid-phase in simple way. It is an elegant alternative to be used in the synthesis of complex peptides with Trp residue.

3.3 *Sugar-Assisted Ligation (SAL)*

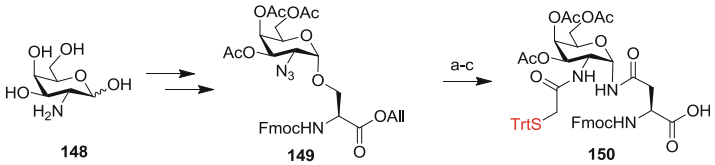
To extend the application of NCL in the synthesis of proteins, Cys-free ligation has been developed during the past few years. In addition to the above-mentioned amino acids, other removable thiol-based auxiliaries were also exploited to function as cysteine. Among them (thiol-containing N^{α}-linked auxiliaries, 4,5,6-trimethoxy-2-mercaptobenzyl auxiliaries, etc.), a carbohydrate auxiliary with a mercaptoacetate at 2-position reported by the Wong group was quite attractive because sugars are moieties of glycoproteins (Scheme 37) [92]. It was speculated that the restricted conformation of the sugar plays a crucial role in the S → N acyl transfer by closing the proximity between the N-amino group of glycopeptides and the carbonyl group of the thioesters to facilitate the intermolecular rearrangement via 14- or 15-membered rings. Subsequent desulfurization under metal-based conditions could provide the target glycopeptides in good yields. This approach was successfully applied in the synthesis of β-O-linked glycopeptides, α-O-linked glycopeptides, and N-linked glycopeptides [93, 94].

Later, Yang et al. extended this method to the synthesis of α-O-linked antibacterial glycoprotein diptericin ϵ [95]. The total synthesis contained an SAL and subsequent NCL. Since diptericin ϵ contains two GalNAc moieties, which are α-linked to serine and threonine, respectively, N^{α}-Fmoc-Thr[Ac_3-α-GalNAc(SH)] **150** was prepared as the building block (Scheme 38).

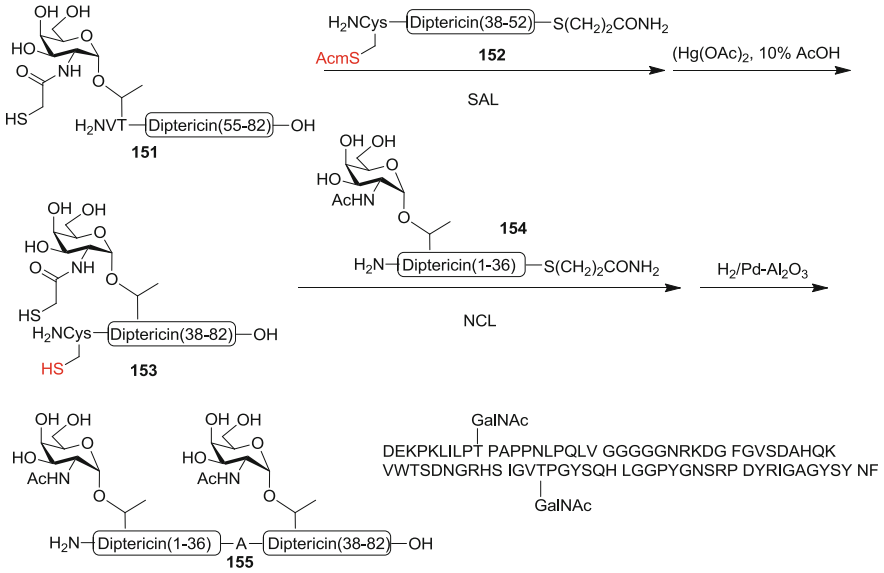
The C-terminal glycopeptides Val^{53}-Phe^{82} **151**, prepared from N^{α}-Fmoc-Thr [Ac_3-α-GalNAc(SH)] by Fmoc-SPPS on 2-ClTrt resin, was ligated with the peptide thioester $Cys(Acm)^{37}$-Gly^{52} **152** in 6 M Gn·HCl, 200 mM phosphate, pH 8.5 at 37°C. During the reaction, pH was adjusted to 7.2 because the pH dropped after mixing both starting materials. An extra 0.5 equiv. of peptide thioester was added to provide the desired product in 36% yield after 48 h. The other glycopeptide thioester Asp^{1}-Asn^{36} **154** was prepared using the side-chain anchoring strategy by Fmoc-SPPS. After removal of Acm, the N-terminal glycopeptide Cys^{37}-Phe^{82} was



Scheme 37 Sugar assisted ligation



Scheme 38 Synthesis of building block **150**. Reagents and conditions: (a) Zn/AcOH; (b) $TrtSCH_2CO_2H$, HBTU, DIPEA, DMF, 84%; (c) $Pd(PPh_3)_4$, *N*-methylalanine, THF, 95%



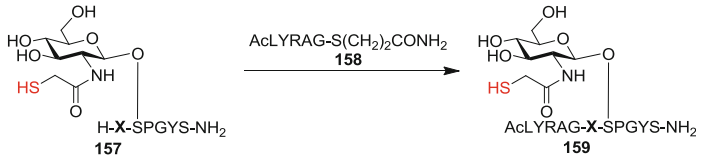
Scheme 39 The synthesis of dipericin and its sequence

ligated with glycopeptide thioester Asp^{1}-Asn^{36} **154** in 6 M Gn·HCl, pH 7.9 containing 2% PhSH and 2% BnSH at 37°C for 16 h and gave 47% yield after HPLC purification. Subsequent metal-based desulfurization reduced both thiols on GalNAc and Cys^{37} to afford diptericin **155** in 54% yield (Scheme 39).

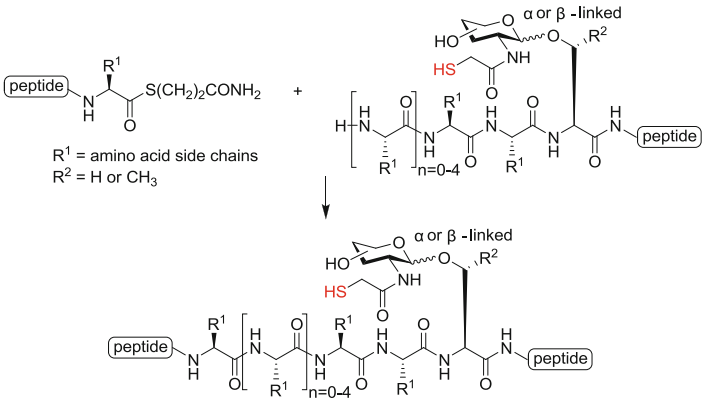
Payne et al. aimed to apply the SAL strategy to the total synthesis of native glycoproteins [96]. However, direct utilization of the SAL method was unsuitable for 75% of glycoproteins, indicated by the screening of over 200 *O*-linked glycoproteins in O-GlycBase v6.00. As a consequence, Payne et al. modified the SAL method to extend its application in the synthesis of a variety of glycoproteins (Table 8).

Building block **156**, similar to **150**, was prepared from glucosamine by a known method. The N-terminal glycopeptides **157a–f** of various lengths were achieved from **156** by Fmoc-SPPS to investigate the ligation efficiency with thioester **158**. It was found that the extended SAL afforded the highest yield (86%) compared to double-extended (70%), triple-extended (60%), quadruple-extended (38%), and

Table 8 Extended sugar-assisted ligation



Ligation method	X	Isolated yield (%)
SAL	**a**: Gly	91
exSAL	**b**: GlyVal	86
dexSAL	**c**: GlyValLeu	70
texSAL	**d**: GlyArgValLeu	60
qexSAL	**e**: GlySerArgValLeu	38
pexSAL	**f**: GlyAlaSerArgValLeu	49



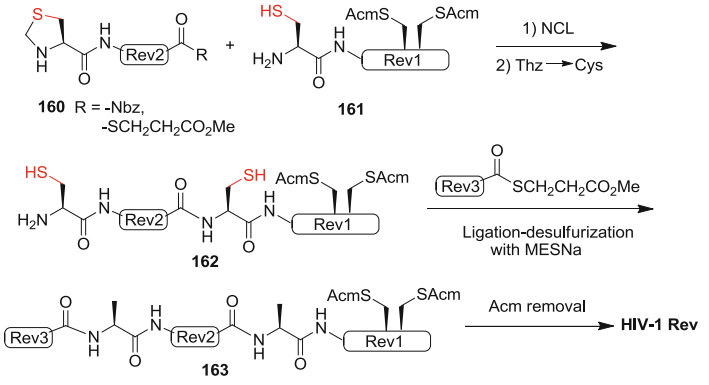
Scheme 40 The synthetic route to MUC1 repeating unit employing the extended SAL

penta-extended (49%) SAL in NMP and HEPES buffer (4:1 v/v NMP/6 M Gn·HCl, 2% PhSH, 1 M HEPES, pH 8.5, 37°C). Both kinetic study and molecular dynamics simulations showed that the ring size of the proposed transition state during S → N acyl transfer influenced the ligation efficiency. As the ring size increased, the distances between the N-terminus and the carbonyl carbon of the thioester increased, resulting in slower ligation rate. However, it was found that 95% of all O-glycoproteins in O-GlycBase v6.00 could be synthesized by SAL and extended SAL (extended up to six amino acid residues) methodology. As a preliminary application, the MUC1 repeating unit was effectively synthesized by employing the extended SAL (Scheme 40).

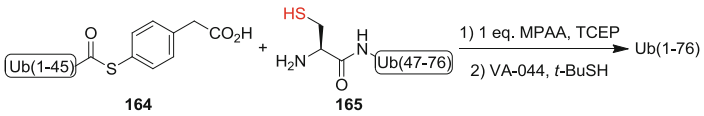
3.4 One-Pot Ligation-Desulfurization

The combination of native chemical ligation and desulfurization has been proved to be a powerful strategy in protein synthesis. Chemists further considered the possibility of carrying out the ligation and desulfurization in a one-pot manner to simplify incrementally the purification procedure and improve the reaction efficiency. However, aryl thiols commonly used in the ligation step are radical scavengers which hamper the homogenous desulfurization. To realize the one-pot ligation and desulfurization, the additional aryl thiol must be removed from the reaction system before proceeding with desulfurization [86, 88, 97]. Interestingly, the Brik group examined several thiol additives and found that MESNa not only afforded the ligation at a fast rate but also allowed the rapid and efficient desulfurization, which promoted the one-pot ligation-desulfurization method with good efficiency [98]. Having used MESNa promoted NCL to obtain HIV-1 Rev protein, a 13 kDa protein plays an important role in the HIV replication cycle. Rev 1 (residues 68–116) bearing a Cys N-terminal, thioester segments Rev 2 (residues 37–67), and Rev 3 (1–36) were prepared by Fmoc-SPPS. Ligation between Rev 1 and Rev 2 followed by unmasking the Cys generated Rev 2-Rev 1, which was ligated with Rve 3. MESNa was added as an alternative for aryl thiols during the ligation process. Subsequent desulfurization with free radical conditions and transformation of Cys to Ala in one pot afforded the desired Rev in 15% overall yield (Scheme 41).

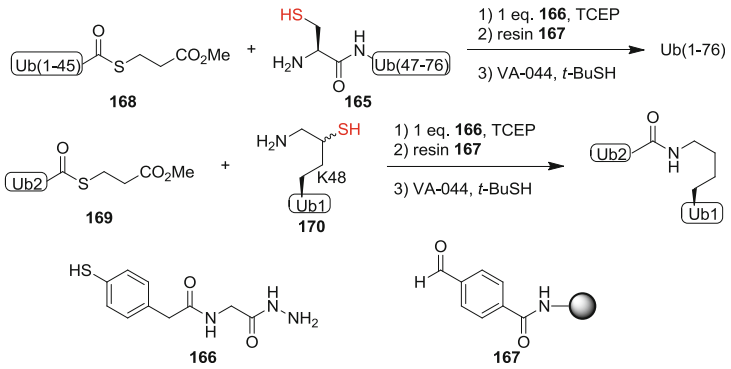
The Brik group [99] further demonstrated that a highly activated aryl thioester (MPAA) could proceed with NCL in the absence of additional aryl thiol, which would produce only 1 equiv. of aryl thiol and hardly influence the desulfurization step. The peptide thioesters LYRAGLYRAG-MPAA, LYRAGLYRAA-MPAA, and LYRAGLYRAV-MPAA were ligated with CYRAGLYRAG in 6 M Gn·HCl, TCEP, pH 7, 37°C. Then the desulfurization was successfully completed in 3 h after



Scheme 41 NCL-based synthesis of HIV-1 Rev and its sequence



Scheme 42 Synthesis of Ub(1–76) applying one-pot ligation and desulfurization

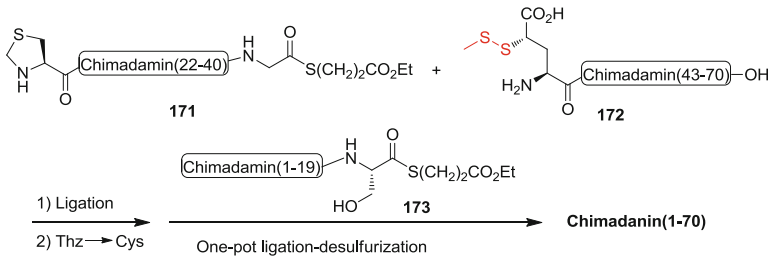


Scheme 43 Synthesis of Ub(1–76) and di-Ub employing bifunctional reagent in one-pot ligation and desulfurization

addition of VA-044 and *t*BuSH. It indicated that 1 equiv. of MPAA released from the ligation step would not affect desulfurization. However, the addition of an extra 1–2 equiv. of MPAA significantly obstructed the desulfurization, indicated by low conversion with a long reaction time. The protocol was next tested in the synthesis of Ub by the ligation of Ub(1–45)-MPAA **164** and Cys-Ub(47–76) **165** and subsequent desulfurization (Scheme 42). The desired product was obtained with 60% yield in 11 h in total, much better than with conventional synthesis.

A further application of this protocol in the synthesis of di-Ub in the chemical ligation of Ub2-MPAA and Ub1(K48K*) (K* = δ-mercaptolysine) was not smooth, which was solved by the utilization of a bifunctional reagent. The modified synthesis of Ub was accomplished within 1.5 h by ligation of Ub(1–45)-methyl mercaptopropionate and Cys-Ub(47–76) by the presence of reagent **166** bearing a hydrazide functionality, capture with a resin bearing an aldehyde functionality, and desulfurization (Scheme 43). This strategy was successfully applied to synthesize di-Ub chain.

Payne et al. also developed an alkyl thiol TFET as an effective thiol additive in NCL [100]. This strategy features high efficiency compared to MPAA and simple operation. TFET effectively promoted the ligation reactions because its pK_a (7.3) is similar to other alkyl thiols. It should be noted that the intrinsic volatility simplifies the intermediate purification, as TFET could be removed by degassing after the ligation. The 70 amino acid thrombin inhibitory protein chimadanin and the 60 amino acid protein madanin-1 were successfully synthesized by this one-pot ligation and desulfurization methodology, employing TFET as an additive.



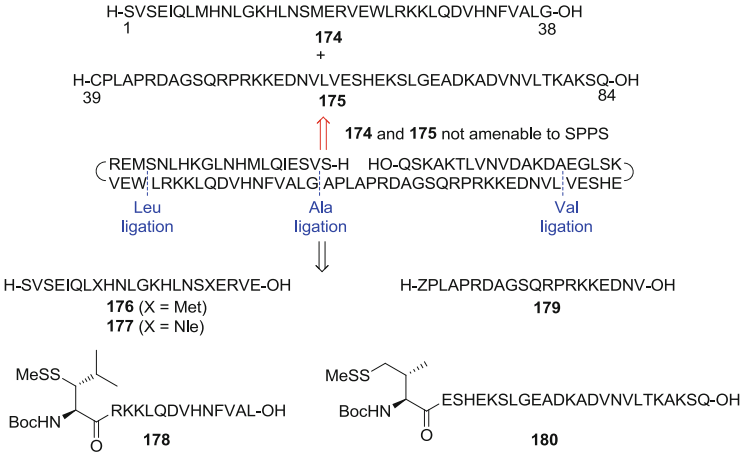
Scheme 44 One-pot synthesis of Chimadanin. Ligation: 6.0 M Gn·HCl, 100 mM Na_2HPO_4, 25 mM TCEP, pH 6.8, 2 vol.% TFET, 30°C, 2 h. Thiazolidine deprotection: 0.2 M $MeONH_2$ (to pH 4.2) 30°C, 3 h. One-pot ligation-desulfurization: for ligation, **173**, pH 6.8, 2 vol.% TFET, 30°C, 18 h; for desulfurization, 500 mM TCEP, 40 mM reduced glutathione, argon, pH 6.2, VA-044, 37°C, 5 h; 35% total yield

Chimadanin was assembled by three fragments in the C → N sequence (Scheme 44). N-terminal γ-thiol Glu residue chimadanin(42–70) **172** was reacted with C-terminal thioester chimadanin(21–41) **171** in the presence of TFET, followed by treatment with methoxyamine at pH 4.2 to furnish an intermediate. Without any purification, the intermediate was directly subjected to N-terminal chimadanin(1–20) **173** and TFET at pH 6.8 to complete the ligation. The reaction mixture was degassed and then treated with TCEP, glutathione, and VA-044 to afford chimadanin in 35% yield.

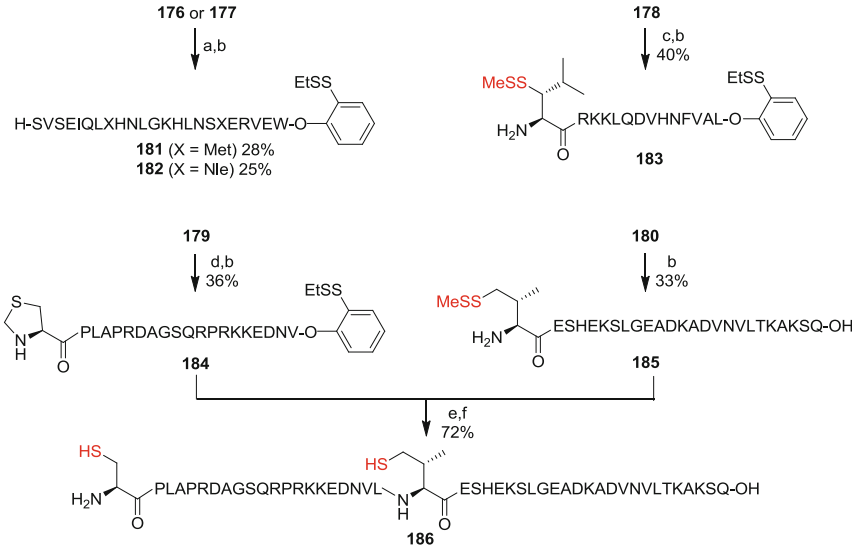
4 Total Synthesis of hPTH

With the endeavor of many scientists, the NCL has been expanded to encompass a series of Cys-free amino acids to extend its influence. It has been proven to be powerful in the synthesis of a range of complex biological active polypeptides and proteins. A fantastic example is the synthesis of human parathyroid hormone [hPTH(1–84)], a polypeptide contains 84 amino acids, based on Cys-free native chemical ligations [101]. The initial plan was to synthesize hPTH(1–84) by NCL between Ser1-Gly38 and Ala39-Gln84. However, the requisite peptides fragments **174** and **175** were unable to be prepared by standard SPPS. Then the synthesis was modified to iterative leucine, alanine, and valine ligations starting from peptide segments **176**, **178**, **179**, and **180** (Scheme 45).

The ligation substrates **181**–**185** were synthesized from peptides **176**–**180**, which were prepared through Fmoc-SPPS (Scheme 46). The assembly of fragments **184** and **185** was accomplished by Val ligation. The terminal Cys of **186** was unveiled to give peptide **187**. An attempt of a one-pot Cys ligation between **183** with **187** and following combination with **181** through Leu ligation failed because of the self-cyclization of **183**. Thus, peptide **178** was converted to **187** with less reactive alkyl thioester at the C-terminal. Peptide **187** was merged with **181** or **182** by Leu ligation to afford **188** and **189**, respectively. NCL of **188** with **186** in the presence of MPAA

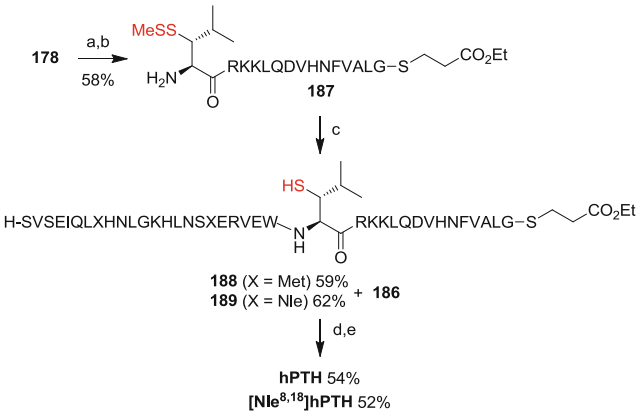


Scheme 45 Retrosynthetic analysis of hPTH(1–84)



Scheme 46 Preparation of peptide segments: (a) HCl·H-Trp-O(EtSS)Ph, EDC, HOOBt, TFE/$CHCl_3$ (1:3), 4 h; (b) TFA/TIS/H_2O (95:2.5:2.5), 45 min; (c) HCl·H-Gly-O(EtSS)Ph, EDC, HOOBt, TFE/$CHCl_3$ (1:3), 4 h; (d) HCl·H-Leu-O(EtSS)Ph, EDC, HOOBt, TFE/$CHCl_3$ (1:3), 4 h; (e) 6.0 M Gn·HCl, 100 mM Na_2HPO_4, 50 mM TCEP, pH 7.5, 15 h; (f) MeO-NH_2·HCl, pH 4, 3 h

followed by metal-free desulfurization provided full-length hPTH. [$Nle^{8,18}$]hPTH (1–84) was obtained from **189** and **186** in a similar way (Scheme 47). This protocol was demonstrated to be effective in preparing a few analogues of hPTH by varying peptides chain.



Scheme 47 Synthesis of hPTH(1–84) and [$Nle^{8,18}$] hPTH(1–84): (a) HCl·H-Gly-$SCH_2CH_2CO_2Et$, EDC, HOOBt, TFE/$CHCl_3$ (1:3), 4 h; (b) TFA/TIS/H_2O (95:2.5:2.5), 45 min; (c) **188** or **189**, 6.0 M Gn·HCl, 100 mM Na_2HPO_4, 50 mM TCEP, pH 7.2, 3 h; (d) 6.0 M Gn·HCl, 300 mM Na_2HPO_4, 200 mM MPAA, 20 mM TCEP, pH 7.3, 3 h; (e) VA-044, *t*BuSH, TCEP, H_2O, MeCN, 37°C, 2 h

5 Conclusion

In the past two decades, native chemical ligation has become the most popular method for chemoselective synthesis of polypeptides and proteins. To overcome the restriction of NCL requiring N-terminal cysteine to promote ligation, postligation-desulfurization involving thiol-mediated ligation followed by desulfurization has been developed. So far, peptide ligation has been realized at amino acid sites other than Cys, including Ala, Val, Lys, Thr, Leu, Pro, Arg, Asp, Glu, Phe, Gln, and Trp. The design and synthesis of these thiol-containing amino acids for native chemical ligation at non-Cys sites have been described in this review. After the target sequence assembly was furnished, the thiol moieties of the peptides were removed by metal-based or radical-based desulfurization. As well as the above-mentioned amino acids, other removable thiol-based auxiliaries such as carbohydrate auxiliary were also exploited to function as Cys surrogates, providing an effective way to the synthesis of glycoproteins.

To improve the efficiency of postligation-desulfurization strategy, synthesis of polypeptides and proteins in a one-pot manner was also developed. The strength of postligation-desulfurization has been demonstrated by the synthesis of a few complex polypeptides and proteins. The present drawback of the powerful postligation-desulfurization is that most of the building blocks for ligation are not commercially available and require tedious synthesis. It is expected that the building blocks should become more easily available and more generally applicable. Anyhow, the postligation-desulfurization strategy facilitates the chemical synthesis of plentiful homogeneous polypeptides and proteins, providing great help for further investigation of protein structures and functions.

Acknowledgments We are grateful for financial support by the National Natural Science Foundation of China (21272082, 21402055, 21402056, 21472054), the Specialized Research Fund for the Doctoral Program of Higher Education (20120142120092), the Recruitment Program of Global Youth Experts of China, the State Key Laboratory of Bio-organic and Natural Products Chemistry (SKLBNPC13425), and Huazhong University of Science and Technology (2014ZZGH015), the Fundamental Research Funds for the Central Universities (Project 2014QC007).

References

1. Walsh G (2010) Post-translational modifications of protein biopharmaceuticals. Drug Discov Today 15(17–18):773–780
2. Walsh G (2010) Biopharmaceutical benchmarks. Nat Biotechnol 28(9):917–924
3. Walsh G, Jefferis R (2006) Post-translational modifications in the context of therapeutic proteins. Nat Biotechnol 24(10):1241–1252
4. Chalker JM (2013) Prospects in the total synthesis of protein therapeutics. Chem Biol Drug Des 81(1):122–135
5. Kent S, Sohma Y, Liu S, Bang D, Pentelute B, Mandal K (2012) Through the looking glass—a new world of proteins enabled by chemical synthesis. J Pept Sci 18(7):428–436
6. 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(4):265–269
7. Wong CTT, Tung CL, Li X (2013) Synthetic cysteine surrogates used in native chemical ligation. Mol BioSyst 9(5):826–833
8. Payne RJ, Wong C-H (2010) Advances in chemical ligation strategies for the synthesis of glycopeptides and glycoproteins. Chem Commun 46(1):21–43
9. Hemantha HP, Narendra N, Sureshbabu VV (2012) Total chemical synthesis of polypeptides and proteins: chemistry of ligation techniques and beyond. Tetrahedron 68(47):9491–9537
10. Flavell RR, Muir TW (2009) Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. Acc Chem Res 42(1):107–116
11. Monbaliu J-CM, Katritzky AR (2012) Recent trends in Cys- and Ser/Thr-based synthetic strategies for the elaboration of peptide constructs. Chem Commun 48(95):11601–11622
12. Kent SBH (2009) Total chemical synthesis of proteins. Chem Soc Rev 38(2):338–351
13. Fisher EFE (1901) Ueber einige derivate des glykocolis. Ber 34:10
14. Merrifield RB (1963) Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. J Am Chem Soc 85(14):2149–2154
15. Merrifield RB, Stewart JM, Jernberg N (1966) Instrument for automated synthesis of peptides. Anal Chem 38(13):1905–1914
16. Merrifield RB (1986) Solid phase synthesis. Science 232(4748):341–347
17. Malins LR, Mitchell NJ, Payne RJ (2014) Peptide ligation chemistry at selenol amino acids. J Pept Sci 20(2):64–77
18. Gutte B, Merrifield RB (1969) Total synthesis of an enzyme with ribonuclease A activity. J Am Chem Soc 91(2):501–502
19. Kent SB (1988) Chemical synthesis of peptides and proteins. Annu Rev Biochem 57 (1):957–989
20. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Synthesis of proteins by native chemical ligation. Science 266(5186):776–779
21. Kemp DS, Carey RI (1993) Synthesis of a 39-peptide and a 25-peptide by thiol capture ligations—observation of a 40-fold rate acceleration of the intramolecular O, N-acyl-transfer reaction between peptide-fragments bearing only cysteine protective groups. J Org Chem 58 (8):2216–2222

22. Kemp DS (1981) The amine capture strategy for peptide bond formation—an outline of progress. Biopolymers 20(9):1793–1804
23. 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(1):129–149
24. Dawson PE, Kent SB (2000) Synthesis of native proteins by chemical ligation. Annu Rev Biochem 69(1):923–960
25. Nilsson BL, Soellner MB, Raines RT (2005) Chemical synthesis of proteins. Annu Rev Biophys Biomol Struct 34:91
26. Botti P, Carrasco MR, Kent SB (2001) Native chemical ligation using removable N-alpha-(1-phenyl-2-mercaptoethyl) auxiliaries. Tetrahedron Lett 42(10):1831–1833
27. Offer J, Dawson PE (2000) N-α-2-Mercaptobenzylamine-assisted chemical ligation. Org Lett 2(1):23–26
28. Offer J, Boddy C, Dawson PE (2002) Extending synthetic access to proteins with a removable acyl transfer auxiliary. J Am Chem Soc 124(17):4642–4646
29. Macmillan D, Anderson D (2004) Rapid synthesis of acyl transfer auxiliaries for cysteine-free native glycopeptide ligation. Org Lett 6(25):4659–4662
30. Wu B, Chen J, Warren J (2006) Building complex glycopeptides: development of a cysteine-free native chemical ligation protocol. Angew Chem Int Ed 45(25):4116–4125
31. Lutsky M-Y, Nepomniaschiy N, Brik A (2008) Peptide ligation via side-chain auxiliary. Chem Commun 10:1229–1231
32. Gross E, Witkop B (1961) Selective cleavage of the methionyl peptide bonds in ribonuclease with cyanogen bromide. J Am Chem Soc 83(6):1510–1511
33. Liu CF, Tam JP (1994) Peptide segment ligation strategy without use of protecting groups. Proc Natl Acad Sci U S A 91(14):6584–6588
34. 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-TN antigens. J Org Chem 74(6):2494–2501
35. Chalker J, Davis B (2010) Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues. Curr Opin Chem Biol 14(6):781–789
36. Li X, Lam H, Zhang Y, Chan C (2010) Salicylaldehyde ester-induced chemoselective peptide ligations: enabling generation of natural peptidic linkages at the serine/threonine sites. Org Lett 12(8):1724–1727
37. Zhang Y, Xu C, Lam H, Lee C, Li X (2013) Protein chemical synthesis by serine and threonine ligation. Proc Natl Acad Sci U S A 110(17):6657–6662
38. Lam H, Zhang Y, Liu H (2013) Total synthesis of daptomycin by cyclization via a chemoselective serine ligation. J Am Chem Soc 135(16):6272–6279
39. Xu C, Lam H, Zhang Y, Li X (2013) Convergent synthesis of MUC1 glycopeptides via serine ligation. Chem Commun 49(55):6200–6202
40. Rohde H, Seitz O (2010) Ligation-desulfurization: a powerful combination in the synthesis of peptides and glycopeptides. Biopolymers 94(4):551–559
41. 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(4):526–533
42. 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(4):687–690
43. Alferiev IS, Connolly JM, Levy RJ (2005) A novel mercapto-bisphosphonate as an efficient anticalcification agent for bioprosthetic tissues. J Organomet Chem 690(10):2543–2547
44. 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(48):9248–9252
45. Hoffmann F, Ess R, Simmons T, Hanzel R (1956) The desulfurization of mercaptans with trialkyl phosphites. J Am Chem Soc 78(24):6414

46. Walling C, Rabinowitz R (1957) The reaction of thiyl radicals with trialkyl phosphites. J Am Chem Soc 79(19):5326
47. Walling C, Basedow O, Savas E (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(9):2181–2184
48. González A, Valencia G (1998) Photochemical desulfurization of L-cysteine derivatives. Tetrahedron Asymmetry 9(16):2761–2764
49. Cuesta J, Arsequell G, Valencia G, González A (1999) Photochemical desulfurization of thiols and disulfides. Tetrahedron Asymmetry 10(14):2643–2646
50. Arsequell G, González A, Valencia G (2001) Visible light promoted organic reaction on a solid support. Tetrahedron Lett 42(14):2685–2687
51. Burns J, Bulter J, Moran J, Whitesides G (1991) Selective reduction of disulfides by tris (2-carboxyethyl)phosphine. J Org Chem 56(8):2648–2650
52. Metanis N, Keinan E, Dawson PE (2010) Traceless ligation of cysteine peptides using selective deselenization. Angew Chem Int Ed 49(39):7049–7053
53. Malins LR, Payne RJ (2012) Synthesis and utility of β-selenol-phenylalanine for native chemical ligation–deselenization chemistry. Org Lett 14(12):3142–3145
54. Quaderer R, Hilvert D (2002) Selenocysteine-mediated backbone cyclization of unprotected peptides followed by alkylation, oxidative elimination or reduction of the selenol. Chem Commun 22:2620–2621
55. 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-β that exhibits potent antitumor activity in vivo. J Am Chem Soc 134(12):5428–5431
56. Yeung H, Harris PWR, Squire CJ, Baker EN, Brimble MA (2014) Preparation of truncated orf virus entry fusion complex proteins by chemical synthesis. J Pept Sci 20(6):398–405
57. Deng F-K, Zhang L, Wang Y-T, Schneewind O, Kent SBH (2014) Total chemical synthesis of the enzyme sortase $A_{\Delta N59}$ with full catalytic activity. Angew Chem Int Ed 53 (18):4662–4666
58. 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(15):3606–3610
59. Garner J, Jolliffe KA, Harding MM, Payne RJ (2009) Synthesis of homogeneous antifreeze glycopeptides via a ligation-desulfurisation strategy. Chem Commun 45:6925–6927
60. Fang G-M, Wang J-X, Liu L (2012) Convergent chemical synthesis of proteins by ligation of peptide hydrazides. Angew Chem Int Ed 51(41):10347–10350
61. Fauvet B, Butterfield SM, Fuks J, Brik A, Lashuel HA (2013) One-pot total chemical synthesis of human alpha-synuclein. Chem Commun 49(81):9254–9256
62. 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 Protoc 8(12):2483–2495
63. Wilson RM, Dong SW, Wang P, Danishefsky SJ (2013) The winding pathway to erythropoietin along the chemistry-biology frontier: a success at last. Angew Chem Int Ed 52 (30):7646–7665
64. Wang P, Dong SW, Shieh JH, Peguero E, Hendrickson R, Moore MAS, Danishefsky SJ (2013) Erythropoietin derived by chemical synthesis. Science 342(6164):1357–1360
65. Haase C, Rohde H, Seitz O (2008) Native chemical ligation at valine. Angew Chem Int Ed 47 (36):6807–6810
66. Wan Q, Chen J, Yuan Y, Danishefsky SJ (2008) Oxo-ester mediated native chemical ligation: concept and applications. J Am Chem Soc 130(47):15814–15816
67. 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(44):8521–8524
68. Yang R, Pasunooti KK, Li F, Liu X-W, Liu C-F (2009) Dual native chemical ligation at lysine. J Am Chem Soc 131(38):13592–13593

69. Yang R, Pasunooti KK, Li F, Liu X-W, Liu C-F (2010) Synthesis of K48-linked diubiquitin using dual native chemical ligation at lysine. Chem Commun 46(38):7199–7201
70. Kumar KSA, Haj-Yahya M, Olschewski D, Lashuel HA, Brik A (2009) Highly efficient and chemoselective peptide ubiquitylation. Angew Chem Int Ed 48(43):8090–8094
71. Kumar KSA, Spasser L, Erlich LA, Bavikar SN, Brik A (2010) Total chemical synthesis of di-ubiquitin chains. Angew Chem Int Ed 49(48):9126–9131
72. Kumar KSA, 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 (27):6137–6141
73. Bavikar SN, Spasser L, Haj-Yahya M, Karthikeyan SV, Moyal T, Kumar KSA, Brik A (2012) Chemical synthesis of ubiquitinated peptides with varying lengths and types ubiquitin chains to explore the activity of deubiquitinases. Angew Chem Int Ed 51(3):758–763
74. 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(52):10149–10153
75. 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(12):4494–4498
76. Virdee S, Kapadnis PB, Elliott T, Lang K, Madrzak J, Nguyen DP, Riechmann L, Chin JW (2011) Traceless and site-specific ubiquitination of recombinant proteins. J Am Chem Soc 133(28):10708–10711
77. Chen J, Wang P, Zhu J, Wan Q, Danishefsky SJ (2010) A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides. Tetrahedron 66 (13):2277–2283
78. Harpaz Z, Siman P, Kumar KSA, Brik A (2010) Protein synthesis assisted by native chemical ligation at leucine. ChemBioChem 11(9):1232–1235
79. Tan Z, Shang S, Danishefsky SJ (2010) Insights into the finer issues of native chemical ligation: an approach to cascade ligations. Angew Chem Int Ed 49(49):9500–9503
80. Li J, Dong S, Townsend SD, Dean T, Gardella TJ, Danishefsky SJ (2012) Chemistry as an expanding resource in protein science: fully synthetic and fully active human parathyroid hormone-related protein (1–141). Angew Chem Int Ed 51(49):12263–12267
81. Creech GS, Paresi C, Li Y-M, Danishefsky SJ (2014) Chemical synthesis of the ATAD2 bromodomain. Proc Natl Acad Sci U S A 111(8):2891–2896
82. Shang S, Tan Z, Dong S, Danishefsky SJ (2011) An advance in proline ligation. J Am Chem Soc 133(28):10784–10786
83. Townsend SD, Tan Z, Dong S, Shang S, Brailsford JA, Danishefsky SJ (2012) Advances in proline ligation. J Am Chem Soc 134(8):3912–3916
84. 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(20):5588–5591
85. Malins LR, Cergol KM, Payne RJ (2013) Peptide ligation-desulfurization chemistry at arginine. ChemBioChem 14(5):559–563
86. Thompson RE, Chan B, Radom L, Jolliffe KA, Payne RJ (2013) Chemoselective peptide ligation-desulfurization at aspartate. Angew Chem Int Ed 52(37):9723–9727
87. Guan X, Drake MR, Tan Z (2013) Total synthesis of human galanin-like peptide through an aspartic acid ligation. Org Lett 15(24):6128–6131
88. Cergol KM, Thompson RE, Malins LR, Turner P, Payne RJ (2014) One-pot peptide ligation-desulfurization at glutamate. Org Lett 16(1):290–293
89. Crich D, Banerjee A (2007) Native chemical ligation at phenylalanine. J Am Chem Soc 129 (33):10064–10065
90. Siman P, Karthikeyan SV, Brik A (2012) Native chemical ligation at glutamine. Org Lett 14 (6):1520–1523

91. Malins LR, Cergol KM, Payne RJ (2014) Chemoselective sulfenylation and peptide ligation at tryptophan. Chem Sci 5(1):260–266
92. 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 (36):11945–11952
93. Brik A, Yang YY, Ficht S, Wong CH (2006) Sugar-assisted glycopeptide ligation. J Am Chem Soc 128(17):5626–5627
94. Brik A, Ficht S, Yang Y-Y, Bennett CS, Wong C-H (2006) Sugar-assisted ligation of N-linked glycopeptides with broad sequence tolerance at the ligation junction. J Am Chem Soc 128(46):15026–15033
95. Yang Y-Y, Ficht S, Brik A, Wong C-H (2007) Sugar-assisted ligation in glycoprotein synthesis. J Am Chem Soc 129(24):7690–7701
96. 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 (44):13527–13536
97. Rohde H, Schmalisch J, Harpaz Z, Diezmann F, Seitz O (2011) Ascorbate as an alternative to thiol additives in native chemical ligation. ChemBioChem 12(9):1396–1400
98. 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 (7):1097–1104
99. Moyal T, Hemantha HP, Siman P, Refua M, Brik A (2013) Highly efficient one-pot ligation and desulfurization. Chem Sci 4(6):2496–2501
100. Thompson RE, Liu X, Alonso-Garcia N, Barbosa Pereira PJ, Jolliffe KA, Payne RJ (2014) Trifluoroethanethiol: an additive for efficient one-pot peptide ligation-desulfurization chemistry. J Am Chem Soc 136(23):8161–8164
101. Dong S, Shang S, Li J, Tan Z, Dean T, Maeda A, Gardella TJ, Danishefsky SJ (2012) Engineering of therapeutic polypeptides through chemical synthesis: early lessons from human parathyroid hormone and analogues. J Am Chem Soc 134(36):15122–15129

Top Curr Chem (2015) 363: 103–154
DOI: 10.1007/128_2014_609
© Springer International Publishing Switzerland 2014
Published online: 20 March 2015

Solid Phase Protein Chemical Synthesis

Laurent Raibaut, Ouafâa El Mahdi, and Oleg Melnyk

Abstract The chemical synthesis of peptides or small proteins is often an important step in many research projects and has stimulated the development of numerous chemical methodologies. The aim of this review is to give a substantial overview of the solid phase methods developed for the production or purification of polypeptides. The solid phase peptide synthesis (SPPS) technique has facilitated considerably the access to short peptides (<50 amino acids). However, its limitations for producing large homogeneous peptides have stimulated the development of solid phase covalent or non-covalent capture purification methods. The power of the native chemical ligation (NCL) reaction for protein synthesis in aqueous solution has also been adapted to the solid phase by the combination of novel linker technologies, cysteine protection strategies and thioester or *N*,*S*-acyl shift thioester surrogate chemistries. This review details pioneering studies and the most recent publications related to the solid phase chemical synthesis of large peptides and proteins.

Keywords solid phase, peptide, protein, native chemical ligation, covalent capture, non-covalent capture

Contents

L. Raibaut and O. Melnyk (✉)
UMR CNRS 8161, Université de Lille, Institut Pasteur de Lille, 59021 Lille, France
e-mail: oleg.melnyk@ibl.cnrs.fr

O. El Mahdi
Université Sidi Mohamed Ben Abdellah, Fez, Morocco

1 Introduction

Today, synthetic peptides are intensively used in chemistry, biology, medicine and material sciences. During the last few decades, chemical synthesis has also emerged as an alternative to the use of recombinant techniques for accessing large peptides or small proteins. Chemical methods are particularly useful for producing proteins which are toxic for living systems or whose expression can be difficult. They can also give access to proteins which are modified at specific sites by non-native modifications such as non-coded amino acids or chemical labels such as fluorophores. Moreover, semi or total chemical protein synthesis enables the site-specific introduction of a large variety of post-translational modifications from the simplest ones (acetylation [1–3], phosphorylation [3, 4], methylation [5]) to the more complex ones such as glycosylation [6–9], ubiquitination [10–15] or sumoylation [16, 17], which are often difficult to install using recombinant techniques [18].

Figure 1 gives the size and the year of synthesis for a selection of peptides and proteins produced by chemical synthesis since the establishment of the field which began with the pioneering work of Fisher in 1901 [19]. It shows a significant increase in the size of the peptides or proteins produced by chemical synthesis since the introduction of the solid phase peptide synthesis method by Merrifield in 1963 [20] (SPPS, Fig. 1, blue triangles) and of the native chemical ligation reaction originated by Kent and coworkers in 1994 [21] (NCL, Fig. 1, green squares). This trend has also been sustained by the development of numerous synthetic methodologies and stimulated by the need to access large polypeptides in various fields of research and particularly in biology [21, 22].

Small synthetic peptides are usually produced using the SPPS technique [20], which relies on the iterative coupling of protected amino acids to a solid support. SPPS was automated soon after its discovery [23–25]. The potential of SPPS to provide access to small functional proteins was demonstrated in 1969 with the production of an enzymatically active ribonuclease A enzyme composed of 124 amino acids [26, 27]. Today, Fmoc-SPPS is the most popular technique for producing peptides [28–30]. It has benefited from many improvements such as the design of powerful activating reagents [31], efficient backbone [32] or side-chain

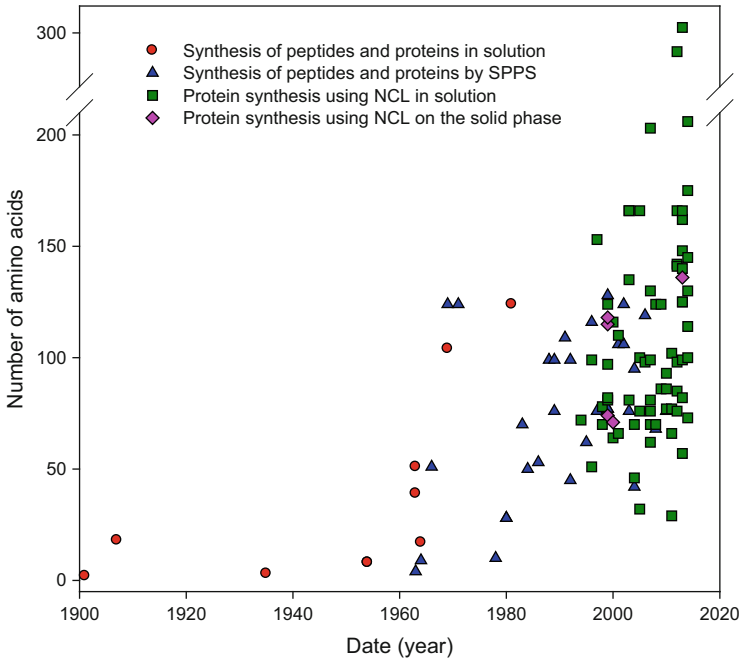


Fig. 1 Size and year of synthesis for a selection of peptides or proteins produced by chemical synthesis since the beginning of the twentieth century

protecting groups [33], various amino acid building blocks which minimize side reactions or the aggregation of the growing peptide chain such as pseudo-prolines [34–36] or isoacyldipeptides [37–44]. Last, but not least, various powerful linker strategies and solid supports have been designed to facilitate the solid phase peptide elongation step and cleavage procedure [45–47]. All these optimizations can give access to peptides composed of up to 50–60 amino acids. Importantly, and as shown in Fig. 1 (blue triangles), extensive optimization of the SPPS protocols can sometimes enable the synthesis of small proteins (>100 AA). The power of SPPS for accessing large peptides is discussed in the second section of this review.

As already discussed above, the synthesis of large peptides (>50 amino acids) is often challenging, despite the numerous improvements of the SPPS method. One obvious reason for this is the inevitable decrease in the yield of the target polypeptide as the number of chemical steps increases, often caused by incomplete couplings and thus the accumulation of capped truncated peptide segments on the solid phase. Major mass losses also occur during the purification process because of the similar chromatographic behaviours of the target polypeptide and of the main impurities, i.e. the capped truncated peptide segments. This observation led to the development of several useful purification methods based on a selective capture step by a complementary solid support (Figs. 2, 3, 4 and 5, see section 3). These methods, which rely on the selective modification of the α-amino group of the target

peptide prior to its detachment from the solid support (Fig. 2, see section 3), are detailed in the third section of this review.

Note that the solid phase fragment condensation (SPFC) approach, which relies on the iterative coupling of protected peptide segments on a solid support, is an alternative to SPPS for accessing long peptides. Importantly, the SPFC technique simplifies the final purification step because the protected peptide segments are purified before use. However, the widespread application of this method was limited by the difficulty in synthesizing, purifying and solubilizing protected peptide segments and by the potential racemization of the C-terminal residue of the peptide segments during the coupling steps. The SPFC strategy is not discussed in detail here because of the availability of excellent reviews on this topic [48, 49]. Nevertheless, important applications of the SPFC method must be highlighted, such as the total synthesis of prothymosin-α (109 amino acids, see entry 16 of Table 1) [74] and of β-amyloid peptide 1–42 [108].

Today, the chemical synthesis of small proteins (>60–70 AA) is usually performed in water by ligating chemoselectively short unprotected peptide segments (Fig. 1, green squares), which can be produced in a pure form using SPPS and high resolutive purification techniques such as HPLC [22, 109]. Alternately, the peptide segments can be produced using living systems, in this case with virtually no limit to the number of amino acid residues composing the peptide chain. The peptide segments are usually ligated using chemoselective amide bond-forming reactions, which must proceed efficiently in water to enable the solubilisation of the unprotected peptide segments [21, 110–114]. Among these, the NCL reaction [21], which involves the reaction of a C-terminal peptide thioester with an N-terminal cysteinyl peptide, is undoubtedly the most popular chemoselective amide bond-forming reaction for protein total synthesis [115]. The concepts underlying NCL were pioneered by Brenner [116] and Wieland [117] in the 1950s. Another important tool is Expressed Protein Ligation (EPL) [118–120], which is an extension of the NCL reaction to the use of recombinant protein thioesters.

For a given protein, the design of the synthetic strategy – and in particular the choice of the ligation junctions and the order of assembly of the different peptide segments – is dictated by many factors. Among these, we can mention the presence and location of specific residues such as cysteines (or alanines after desulfurization), the nature of the junctions, the ease of synthesis of the different segments, their solubility and, of course, the potential occurrence of side-reactions. For these reasons, the chemical synthesis of proteins often requires the chemical ligation of three or more than three peptide segments, either sequentially [121] or by combining sequential and convergent [122–124] assembly schemes. In any case, the assembly strategies require intermediate isolation or purification steps to change the solvent systems, to remove some reagents and to avoid the accumulation of unreacted peptide segments and side-products. Not surprisingly, these isolation steps often result in significant material losses, especially when the intermediates are poorly soluble.

To overcome these limitations, much effort has been focussed on the design of one-pot three peptide segments sequential assembly methods [17, 109, 122, 125–127]. For the assembly of more than three peptide segments, a potential

solution is to combine the power of the solid phase approach and of the NCL reaction. The use of a solid phase enables separation of the growing peptide chain from the excess of reagents and unreacted peptide segment by simple washing steps. Moreover, the solid phase synthesis of proteins can have many other advantages over the solution phase approach, reminiscent of the advantages of the SPPS in comparison with the peptide synthesis in solution. In particular, the use of a water compatible and hydrophilic solid support can improve the solvation of the peptide chain by imposing its solvation properties onto the peptidyl resin, a point which is important for a good accessibility of the reactive ends and for the efficiency of the ligation reactions. In addition, excess of the reagents and of the peptide segments can potentially be used, even if for the latter case the associated cost can be limiting. Last, but not least, the solid phase approach makes it possible to automate the process. The potential of the solid phase approach for the chemical synthesis of proteins by ligating chemoselectively unprotected peptide segments is discussed in the fourth section of this review.

2 Chemical Protein Synthesis by Stepwise Solid Phase Peptide Synthesis

The seminal work of Merrifield published in 1963 described the synthesis of a tetrapeptide using the Boc SPPS method (entry 1, Table 1). Soon after its discovery the potential of SPPS for accessing small proteins was demonstrated by the synthesis of bovine insulin [53] in 1966 (entry 3, Table 1) and of ribonuclease A [26, 55] in 1969 (RNAse A, entry 5, Table 1). The work on RNAse A is outstanding in several respects. It demonstrated the capacity of SPPS to yield large polypeptides, but also to contribute to the understanding of important fundamental phenomena, such as the mechanism of protein folding. Indeed, over the same period several studies showed that reduced and denatured RNAse A of biological origin could be refolded into an active enzyme by air oxidation, suggesting that the primary structure of a protein determines its tertiary structure [128–132]. However, and as discussed by Merrifield in his seminal paper [26], "the conclusion that this response was determined solely by the primary structure of the protein depended on establishing that the unfolded chain had a completely random form and had lost all of its secondary and tertiary structure", as some remaining secondary or tertiary structure from biological origin might serve as a nucleus for the refolding of the enzyme. As no pre-existing conformation of biological origin was present in the synthetic RNAse polypeptide, the reconstitution of an active enzyme by total synthesis showed that the information contained within the primary structure was sufficient to obtain a folded and functional enzyme.

The SPPS of peptides relies on repetitive deprotection and coupling steps which take place on a solid support. Performing all the chemical reactions on a solid support enables the use of a large excess of reagents for optimizing the yields of

Table 1 Boc or Fmoc SPPS of proteins[a]

Entry	Year	Protein	Number of Aa	SPPS chemistry	Activator	Resin	Yield	Functionality	Remarks	Reference
1	1963	LAGV[b]	4	Boc	DCC	Merrifield resin[c]	76 mg	Model	Capping of free amines using acetic anhydride	[20]
2	1964/1965	Bradykinin	9	Boc	DCC	Merrifield resin	32% 68% yield by automated synthesis [23, 24]	Isolated rat uterus and duodenum assays	Final deprotection and cleavage step: (1) HBr in TFA; (2) hydrogenation using Pd black in MeOH/AcOH.	[23, 24, 50, 51] see also [52]
3	1966	Bovine insulin	21 + 30	Boc	DCC	Merrifield resin	A chain: 37% B chain: 21%	In vivo mouse assay	Intermediate *S*-sulfonate protection of Cys residues. Disulfide bonds formed by air oxidation at pH 10	[53]
4	1967	Bradykinin	9	Boc	DCC	Merrifield resin	66%	Isolated rat uterus contraction assay	Final deprotection and cleavage in anhydrous HF	[54]
5	1969	RNase A	124	Boc	DCC	Merrifield resin	85 mg (0.42 mmol scale)	13–24% specific enzymatic activity	Intermediate *S*-sulfonate protection of Cys residues. Four disulfide bonds formed by air oxidation at pH 8.3	[55]
6	1971	RNase A	124	Boc	DCC	Merrifield resin	2.9%	78% specific enzymaticactivity	See entry 5	[26]

7	1978	ACP65-74	10	Fmoc	Symmetric anhydride	HMP PS [56]	74%	Nd[d]	First report of a Fmoc SPPS	[28]
8	1980	Desacetyl-thymosin α_1	28	Boc	DCC	Pam PS resin [57]	31%	Nd	Use of SulfFmoc affinity purification method [58]	[59]
9	1980	Thymosin α_1	28	Boc	DCC	Pam PS	34%	In vitro cellular rosette inhibition assay		[59]
10	1983	IGF-1	70	Boc	DCC/HOBt [60]	Pam PS	1%	In vitro mitogenic activity	Three disulfide bonds formed in 2 M Gdn.HCl, pH 8.3, 24°C	[61]
11	1984	TGF-I	50	Boc	Nd	Merrifield resin	31%	Receptor binding and cellular assays	Low–high HF method [62–64] Three disulfide bonds formed using GSH/GSSG redox system.	[65]
12	1986	EGF	53	Boc	DCC in CH_2Cl_2 and DCC/HOBt in DMF	Pam PS	10%	Receptor binding and stimulation of DNA synthesis	[^{3}H]Leu at position 52 to facilitate SPPS monitoring and EGF quantitation	[66]
13	1988	HIV-1 PR[e]	99	Boc	DCC in CH_2Cl_2 and DCC/HOBt in DMF	Pam PS	20 mg (0.5 mmol scale)	Enzymatic activity		[67, 68]
14	1989	[Aba^{67,95}] HIV-1 PR.	99	Boc	DCC in CH_2Cl_2 and DCC/HOBt in DMF	Pam PS	Nd	X-Ray, enzymatic activity [69]	For the X-ray structure of HIV-1 PR with a pseudopeptide substrate see: [70]	[71]

(continued)

Table 1 (continued)

Entry	Year	Protein	Number of Aa	SPPS chemistry	Activator	Resin	Yield	Functionality	Remarks	Reference
15	1989	Ub	76	Fmoc	HOBt/DIC	HMP PS	4.3%	Nd		[72, 73]
16	1991	ProTα	109	Fmoc	DCC/HOBt	2-ClTrt PS	11%	Cellular assays	Use of the solid phase fragment condensation approach (SPFC) [48, 49]	[74, 75]
17	1992	L-[$Aba^{67,95-}$] HIVPR	99	Boc	HBTU [76]/DIEA	Pam PS	Nd	Enzymatic activity	Synthesis of a D protein	[77]
18	1992	RbDD	45	Fmoc	Pentafluorophenyl esters	Nd	Nd	CD	Metal-binding protein. Synthesis of L and D proteins	[78, 79]
19	1995	4-OT	62	Boc	HBTU/DIEA	Pam PS	L-4OT: 8% D-4OT:4%	CD, enzymatic activity	In situ neutralization protocol [80] Synthesis of L and D proteins	[81]
20	1996	FIV PR	116	Boc	Nd	Pam PS	Nd	Enzymatic activity	Automated in situ neutralization protocol [80]	[82]
21	1997	[Abu^{26}, Nva^{30}]-Ub	76	Fmoc	HOBt/DCC	HMP PS	Nd	CD, X-ray	TbFmoc capture purification method [83]. Gel filtration then RP-HPLC on C8 column, followed by Tbfmoc removal using piperidine	[84, 85]

22	1999	UBICEP52	128	Fmoc	HOBt/DCC	HMP PS (0.29 mmol/g)	16 mg from 600 mg Fmoc-Lys (Boc)-HMP PS	DUBs cleavage assay	Primary purification using TbFmoc method [83]	[86, 87]
23	1999	Ub-N^{ε}Lys	77	Fmoc	HOBt/DCC	HMP PS (0.29 mmol/g)	Ub-N^{ε}Lys: 4 mg from 600 mg Boc-N^{α}-Lys-HMP PS	DUBs cleavage	Primary purification using TbFmoc method [83]	[86, 87]
24	1999	Ub-Val	77	Fmoc	HOBt/DCC	HMP PS (0.29 mmol/g)	Ub-Val: 5 mg from 800 mg Fmoc-Val-HMP PS	DUBs cleavage assay	Primary purification using TbFmoc method [83]	[86, 87]
25	2001	PrP106	106	Fmoc	HOBt/DCC	MBHA PS	13%	CD	Use of backbone protection (Hmb-Gly) [32]. Capping with Z (2-Cl)-OSu [89]	[88]
26	2002	PrP106	106	Fmoc	HBTU/DIEA	HMP PS	5%	CD, in vitro neurotoxicity	DACFmoc affinity purificationmethod [90]	[91]
27	2002	hGIIA sPLA2	124	Fmoc	HOBt/DCC	HMP PS	1%	CD, enzymatic activity	Seven disulfide bonds	[92]
28	2003	Ub[f]	76	Fmoc	HOBt/DIPC	HMP PS	Nd	CD, ^{19}F NMR, X-ray	(2*S*, 4*S*)-5-Fluoroleucines in the hydrophobic core (positions 50 and 67)	[93]

(continued)

Table 1 (continued)

Entry	Year	Protein	Number of Aa	SPPS chemistry	Activator	Resin	Yield	Functionality	Remarks	Reference
29	2004	Aβ1-42	42	Fmoc	DIPC/HOBt	Trt PS	34%	Aggregation properties	Use of isoacyldipeptides. See also the work of Carpino, L. A [41, 42] and Mutter, M [43, 44] Yield for the standard Fmoc SPPS of Aβ 1–42: 7%	[37–40]
30	2004	FAS	95	Fmoc	HATU [94–96]/ DIEA	Rink AM PS [97]	Nd	Nd	Use of pseudoprolines [34–36]	[98]
31	2006	Id3	119	Fmoc	HBTU/HOBt/ DIEA	HMP PS	8%	CD, dimer formation	No capping	[99]
32	2008	SDF-1α	68	Fmoc	HCTU [100]/ NMM	HMPB Chemmatrix [101]	Nd	Nd	Effect of different resins on crude peptide purity	[102]
33	2010	Ub^{g}	76	Fmoc	PyBOP [103, 104]/ DIEA	Trt PS	5–6%		Ub protected polypeptide was condensed with Gly-AMC or Gly-Rho^{110}-Gly and then deprotected	[105]

34	2010	Ub	76	Fmoc	PyBOP/DIEA	HMP PS	14%	CD, Ub enzymatic ligation experiments	Use of pseudoprolines [34–36] and DMB dipeptides building blocks [32, 106].	[105]

[a]*Abu* α-aminobutyric acid, *Aβ* amyloid-beta peptide, *ACP* acyl carrier protein, *AMC* 7-amino-4-methylcoumarin, *Boc tert*-butyloxycarbonyl, *CD* circular dichroism, *2-ClTrt PS* 2-chlorotrityl polystyrene resin, *DACFmoc* 4-dodecyl aminocarbonyl fluorene-9-ylmethyl oxycarbonyl, *DCC N,N′*-dicyclohexylcarbodiimide, *DIEA N,N*-diisopropylethylamine, *DIC N,N′*-diisopropyl carbodiimide, *DMB* dimethoxybenzyl, *DUB* deubiquitinating enzyme, *EGF* epidermal growth factor, *FAS* ligand of FAS receptor (FasR), *FIV PR* feline immunodeficiency virus aspartyl proteinase, *Fmoc* 9-fluorenylmethoxycarbonyl, *GSH* glutathione, *GSSG* glutathione disulfide, *HATU N*-[(1*H*-1,2,3-triazolo[4,5-b]pyridin-1-yloxy)(dimethylamino)methylene]-*N*-methylmethanaminium hexafluorophosphate, *HBTU N*-[(1*H*-benzotriazol-1-yl)(dimethylamino)methylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide, *HCTU O*-(1*H*-6-chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate, *HIV-1 PR* HIV-1 protease, *hGIIA sPLA2* human group IIA secretory phospholipase A2, *Hmb* 2-hydroxy-4-methoxybenzyl, *HMP PS* 4-hydroxymethylphenoxymethyl polystyrene resin (also known as Wang resin), *HMPB* 4-(4-hydroxymethyl-3-methoxyphenoxy)-butyric acid linker, *HOAt* 1-hydroxy-7-azabenzotriazole, *HOBt* 1-hydroxybenzotriazole, *Id3* inhibitor of DNA-binding/cell-differentiation 3, *IGF-1* insulin-like growth factor I, *MBHA* methylbenzhydrylamine PS resin, *NMM N*-methylmorpholine, *Nva* norvaline, *4OT* 4-oxalocrotonate tautomerase, *Pam* 4-(hydroxymethyl)phenylacetamidomethyl PS resin, *ProTα* prothymosin α, *PrP* prion protein, *PyBOP* benzotriazol-1-yl-*N*-oxy-tris(pyrrolidino)phosphonium hexafluorophosphate, *RbDD* rubredoxin protein from anaerobic bacterium *Desulfovibrio desulfuricans* 27774, *Rho*[110] rhodamine 110, *Rink* 4-[(*R*,*S*)-(2,4-dimethoxyphenyl)aminomethyl]phenoxyacetic acid linker, *Rink AM PS* rink aminomethyl polystyrene resin, *RNase* ribonuclease, *SDF 1α* stromal cell-derived factor 1α, *SulfFmoc* 9-(2-sulfo)fluorenylmethyloxycarbonyl, *TbFmoc* tetrabenzo [a,c,g,i]fluorenyl-17-methoxycarbonyl, *TGF-I* transforming growth factor I, *Trt PS* trityl polystyrene resin, *Ub* ubiquitin, *UBICEP52* ubiquitin-52-amino-acid ribosomal protein fusion

[b]Tetrapeptide Leu-Ala-Gly-Val

[c]Merrifield, R. B. used either nitrated or brominated chloromethyl polystyrene resin as starting solid support for the syntheses. The best results were obtained with the nitrated resin

[d]*Nd* not determined (or not reported)

[e]For the SPPS of HIV-2 protease see Wu et al. [107]

[f]Fluorinated derivatives of Ub

[g]C-terminally labelled Ub derivatives

peptide bond formation. The reagents are subsequently removed by performing simple washings procedures. The potential of the solid phase method for automating the synthesis of peptides was illustrated soon after its introduction in 1963 by the automated SPPS of bradykinin in 1965 (entry 2, Table 1) [23–25].

Since that time, a huge amount of work has been devoted to the optimization of the SPPS method. Several significant advances in the field of SPPS are mentioned in Table 1. These include the development of efficient carboxylic acid activators and coupling additives such as HOBt [60], HBTU [103], BOP or PyBOP [103, 104] and HATU [94–96]. The improvement of Boc SPPS by the development of anhydrous HF procedures for the final deprotection and cleavage steps must be mentioned (HF method [54]: entry 4 of Table 1, low–high HF method [62–64]: entry 11 of Table 1). Significant advances were the introduction of the Fmoc protecting group by Carpino and coworkers in 1970 [133–135] and the development of the Fmoc SPPS method by Sheppard and coworkers in 1978 [28], now the most popular method for the solid phase synthesis of peptides (entry 7, Table 1). The solid supports have also been optimized to minimize the aggregation of the protected peptide chain during the peptide elongation step and to enable the swelling of the peptidyl resin in a large variety of organic solvents. Polyethylene glycol-based solid supports such as PEGA [136], Tentagel [137] or ChemMatrix [16, 101] resins are particularly useful for accessing large polypeptides. Other strategies for minimizing the aggregation of the growing peptide chain consist of incorporating pseudo-prolines [34–36] at Ser, Thr or Cys positions (entries 30 and 34, Table 1), and/or Hmb [32, 88] or DMB [32, 106] protecting groups for the peptide backbone (entries 25 and 34, Table 1). Both types of modifications reduce the number of NH amide groups within the peptidyl resin, which can induce the aggregation of the protected peptide through hydrogen bond formation. Pseudo-proline residues also enable the formation of turns in the peptide chain by adopting the *cis* peptide bond conformation, thereby preventing the aggregation of the peptide which often occurs through β-sheet formation (composed of all *trans* peptide bonds). The recent application of pseudo-proline and DMB backbone modifications for the Fmoc SPPS of ubiquitin analogs illustrates the power of these approaches for the production of large peptides (entry 34, Table 1).

3 Purification by Selective Capture on a Solid Support

The numerous improvements of SPPS since its introduction in 1963 make possible the stepwise synthesis of large polypeptides or small proteins the length of which can exceed 80–100 amino acids. Nevertheless, these achievements (some of which are listed in Table 1) must be considered as exceptions, because the synthesis of peptides composed of more than 50 amino acids is often challenging. Besides the yield of the target polypeptide which decreases inevitably with the number of coupling steps, a significant source of mass loss is the difficulty in separating the target polypeptide from the large number of side-products formed during the

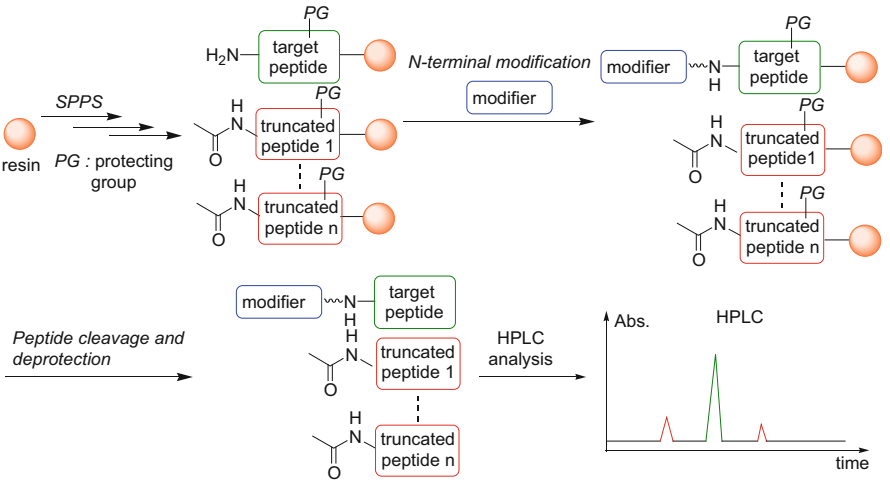


Fig. 2 The SPPS technique results in the formation of truncated and capped peptide contaminants (*red*) which have to be separated from the target peptide (*green*). The covalent or non-covalent capture purification methods require that the target peptide is modified on its N-terminus prior to the peptide deprotection and cleavage steps

synthesis. For example, if a capping step is applied, these are mainly capped peptides formed during the peptide elongation step as a consequence of incomplete couplings.

Several strategies were developed to simplify the purification step. To find a needle in a haystack, the α-amino group of the target polypeptide is first derivatized by a modifier before its deprotection and cleavage from the solid support (Fig. 2). For this, the peptidyl resin must be capped carefully after each coupling step to ensure that only the target and full-length peptide features a free α-amino group. The modifier is used subsequently for capturing the target peptide by selective immobilization on a solid support. Two strategies are possible, depending on whether the immobilization involves the formation of a covalent bond (Figs. 3 and 4, Tables 2 and 3) or not (Fig. 5, Table 4). The capped and truncated peptides are removed after the capture step by performing simple washing procedures or a chromatographic separation. Finally, the detachment of the peptide from the solid support and the cleavage of the bond between the modifier and the peptide yield the target peptide in solution.

3.1 Purification by Covalent Capture on a Solid Support

The purification by covalent capture on a solid phase relies on the selection of a chemoselective and site-specific ligation method between the solid support and the modified peptide (Fig. 3). Merrifield and coworkers pioneered the field in 1976 by

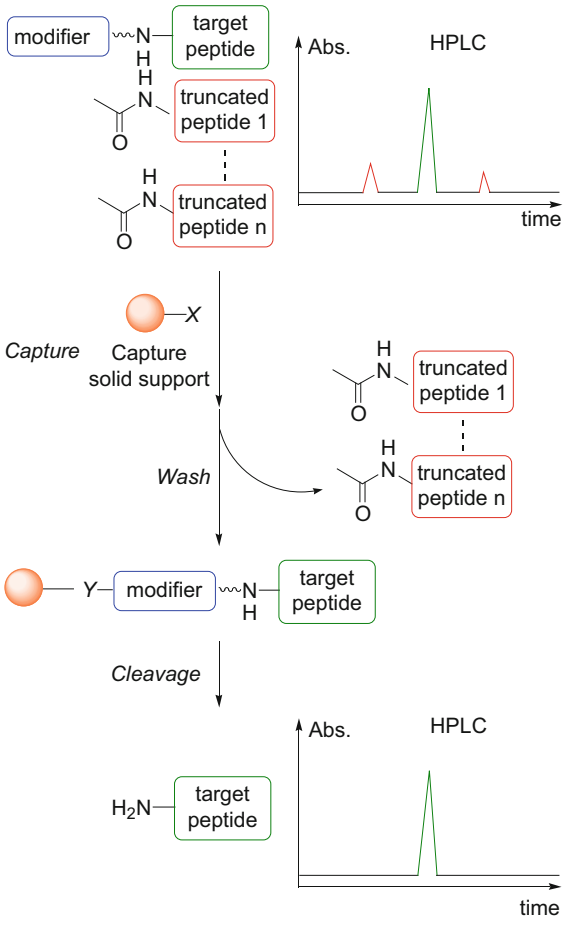
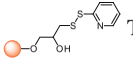
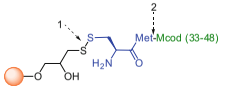
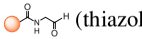
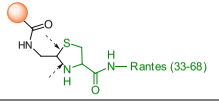
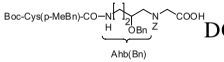
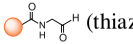
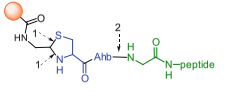
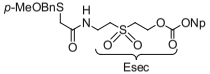
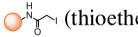
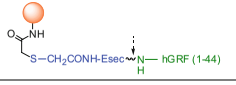
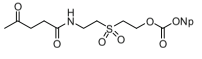
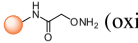
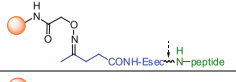
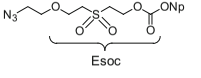
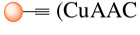
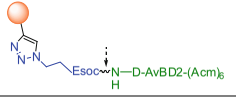


Fig. 3 In the covalent capture purification method, the target peptide is captured chemoselectively by a complementary functionalized solid support and then released in solution in a purified form

using the Cys-Met dipeptide modifier (entry 1, Table 2), which can be easily introduced during the SPPS elongation step. The cysteine thiol was used to capture the target peptide on an organomercurial agarose solid support through the formation of an Hg–S bond. Treatment of the peptidyl resin with an excess of cysteine resulted in the displacement and cleavage of the Hg–S bond, leading to the detachment of the peptide from the solid support. In the last step, the Cys−Met dipeptide modifier was removed in solution by the selective cleavage of the Met-Xaa peptide bond with cyanogen bromide. Two years later, Lindeberg and coworkers applied the same modifier to the immobilization of a synthetic peptide using the thiol-disulfide exchange reaction, a strategy which avoids the use of the toxic organomercurial agarose solid support (entry 2, Table 2) [139].

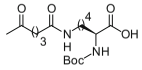
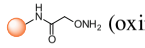
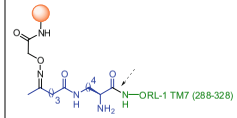
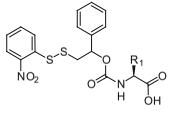
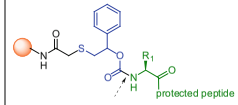
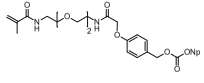
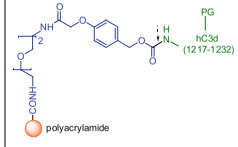
The ease of producing peptides featuring a free N-terminal cysteine residue by the SPPS stimulated the use of chemoselective ligation methods specific for the β-aminothiol group of cysteine for the capture step. A great advantage of these

Table 2 Strategies for peptide purification by covalent capture on a solid support[a]

Entry	Peptide (reference)	Modifier	Solid support –*x* (chemoselective reaction)	Solid support–modifier-target peptide ¦ cleavage site	Cleavage conditions
1	H4 (1–37) [138]	Boc-Cys(*p*-MeOBn)-Met-OH DCC activation	–Hg Organomercurial agarose (thiol-Hg bond)		1: 0.5 M cysteine then 2: CNBr in solution
2[b]	Mcod (33–48) [139]	Boc-Cys(*p*-MeOBn)-Met-OH DCC activation	 Thiopropyl-sepharose 6B (thiol-disulfide exchange)		1: 0.1 M ammonium acetate pH 8.0, 20 mM DTE, 2: CNBr in solution
3	Rantes (34–68) [140]	Boc-Cys(*p*-MeBn)-OH HBTU/DIEA activation [76]	 (thiazolidine ligation) [141]		$MeONH_2$ (400 mM), TCEP (10 mM) in ACN/H_2O (1/1) pH 2.9
4[c, d]	34 AA model peptide [142]	 DCC/HOAt activation [143–145]	 (thiazolidine ligation) [141]		1 (thiazolidine cleavage): $MeONH_2$ (400 mM) in ACN/H_2O (1/1) Then 2 (β-amino alcohol cleavage): $NaIO_4$ (5 equiv.) [146–150]
5	hGRF (1–44) [151]		 (thioether ligation)		5% NH_4OH
6	12 AA model peptide [152]		 (oxime ligation)		NaOH pH 12
7	D-AvBD2-$(Acm)_6$ [153], see also [154]		 (CuAAC [155, 156])		50 mM CAPS pH 11.7

(continued)

Table 2 (continued)

Entry	Peptide (reference)	Modifier	Solid support —x (chemoselective reaction)	Solid support–modifier-target peptide ¦ cleavage site	Cleavage conditions
8	ORL-1 TM7 (288–328) [157]	 1: Boc-Lys(Fmoc)-OH, HBTU/HOBt activation 2: piperidine 3: 5-oxohexanoic acid, DIPC/HOSu activation	 (oxime ligation)		1 (transamination [158–165]): $CHOCO_2H$, Ni^{2+} in acetate buffer pH 5 Then 2 (cleavage of α-ketoamide): *o*-phenylenediamine [159] in acetate buffer pH 4.0
9[e]	9 AA model protected peptide [166]	 DCC or BOP/DIEA activation	 PS beads (thioether ligation)		TFA, water, 1,2-ethanedithiol The peptide is released in the deprotected form.
10	hC3d (1,217–1,232) [167]		(Free radical polymerization of the modifier) *N*,*N*-dimethyl-acrylamide, *N*,*N*′-methylene-bis(acrylamide) $(NH_4)_2S_2O_8$, TMEDA H_2O/DMF/MeCN		TFA, scavenger cocktail. The peptide is released in the deprotected form.

[a]*Ahb* 4-amino-2-hydroxy-butyryl, *Acm* acetamidomethyl, *Bn* benzyl, *CAPS* *N*-cyclohexyl-3-aminopropanesulfonic acid, *CuAAC* copper-catalyzed alkyne azide cycloaddition, *D-AvBD2* D enantiomer of the chicken beta-defensin AvBD2, *DCC* *N*,*N*′-dicyclohexylcarbodiimide, *DIPC* diisopropylcarbodiimide, *DTE* dithioerythritol, *Esec* ethylsulfonyl-2-ethoxycarbonyl, *Esoc* 2-ethoxy-ethylsulfonyl-2-ethoxycarbonyl, *H4* histone H4, *HATU* *N*-[(1*H*-1,2,3-triazolo [4,5-b]pyridin-1-yloxy)(dimethylamino)methylene]-*N*-methylmethanaminium hexafluorophosphate, *HBTU* *N*-[(1*H*-benzotriazol-1-yl)(dimethylamino)methylene]-*N*-methylmethanaminium hexafluorophosphate-*N*-oxide, *hC3d* human complement C3 domain, *HOAt* 1-hydroxy-7-azabenzotriazole, *HOBt* 1-hydroxybenzotriazole, *HOSu* *N*-hydroxysuccinimide, *Mcod* allergen M from cod, *ONp* *p*-nitrophenoxy, *ORL-1TM7*: 7th transmembrane domain of opioid receptor-like 1, *TMEDA* *N*,*N*,*N*′,*N*′-tetramethylethylenediamine

[b]The sequence of Mcod (33–48) is CysMetValGlyLeuAspAlaPheSerAlaAspGluLeuLysLysLeuPheLys

[c]When present, cysteine thiols were protected by a 2-pyridylsulfenyl group during the Boc SPPS. 2-Pyridylsulfenyl groups were removed after the periodate oxidation step by treatment with TCEP (10 mM) in MES buffer (100 mM, pH 6.4)

[d]For a recent review on the oxidative cleavage of β-amino alcohols by periodate see El-Mahdi and Melnyk [168]

[e]Asp(*t*Bu)AlaGlu(*t*Bu)PheArg(Pmc)His(Trt)Asp(*t*Bu)Ser(*t*Bu)Gly-OH

approaches is their compatibility with the presence of internal cysteine residues. In particular, thiazolidine [141, 184] ligation was used for capturing N-terminal cysteinyl peptides with an aldehyde solid support (entry 3, Table 2) [142]. The peptide was released subsequently by reversing the thiazolidine in the presence of *O*-methylhydroxylamine. The method has been extended to peptides featuring any kind of N-terminal amino acid residue by inserting a β-amino alcohol linker between the cysteine and the target peptide (entry 4, Table 2). In this case, the removal of the modifier required the oxidative cleavage of the β-amino alcohol linker with sodium periodate, a method reminiscent of the synthesis of α-oxo aldehyde peptides by oxidative cleavage of N-terminal serine or threonine residues with periodate (for reviews see [168, 185]) [146–150].

The rate of oxidative cleavage of β-amino alcohols is 10^2–10^4 times that of the vicinal diol analogues in sugar series [147]. Therefore, this method is potentially compatible with the purification of glycopeptides. Periodate oxidation of β-amino alcohols is a fast reaction. However, some amino acids such as cysteine [186], tryptophan [187], tyrosine [187, 188] and particularly methionine [150, 188, 189] can also be oxidized by periodate, thereby potentially limiting the scope of the method.

Other studies used non-oxidative methods to detach the captured peptide from the solid support. In particular, several linkers have been developed which are reminiscent of the methylsulfonylethyloxycarbonyl group (Msc) originally developed in 1975 by Tesser and coworkers (entries 5–7, Table 2) [190]. In these approaches, the ethylsulfonyl-2-ethyloxycarbonyl (Esec) linker connects the chemical group used for the capture step to the N-terminal amino acid of the target peptide. They differ in the chemistry used for the immobilization step. The method described in entry 5 of Table 2 relies on the formation of a thioether bond by reaction of a thiol modifier with an iodoacetamide-functionalized solid support. The chemoselective formation of an oxime bond is used in entry 6 of Table 2, whereas the copper(I)-catalyzed alkyne azide cycloaddition reaction (CuAAC) [155, 156] is used in entry 7 of Table 2. At the end, the target peptide is released in solution by exposing the peptidyl resin to an aqueous base, which triggers a β-elimination process. Note that the cleavage of the Esec linker requires strong basic conditions (pH >11). Moreover, the alkylation of cysteine thiols by the vinylsulfone formed by β-elimination of the Esec moiety has been mentioned as a potential limitation [191].

The method described in entry 8 of Table 2 makes use of oxime ligation for the capture step as in the work of Canne and coworkers (see entry 6, Table 2) [157]. In this case, the cleavage of the amide bond between the modifier and the target peptide relies on the work of Dixon and coworkers on the selective transamination of the free α-amino group in the presence of glyoxylic acid and catalytic nickel (II) ions [158–164]. The intermediate α-keto acyl moiety produced by transamination of the N-terminal lysine residue is removed subsequently by reaction with *o*-phenylenediamine or related derivatives [159]. The great interest of this method is the mildness of the experimental conditions and its compatibility with large peptides or proteins. However, the authors mentioned that the transamination

reaction in the presence of glyoxylic acid and nickel(II) used in catalytic amounts was incomplete and required stoichiometric amounts of the metal ion to proceed efficiently.

The method described in entry 9 of Table 2 was developed with the aim of purifying hydrophobic peptides such as protected peptide segments [166]. The capture step proceeds in an organic solvent and relies on chemoselective thioether ligation. The structure of the modifier is reminiscent of the benzyloxycarbonyl protecting group and is removed in concentrated TFA in the presence of appropriate scavengers.

The last method described in entry 10 of Table 2 relies on the polymerization of an acrylamide modifier, which is attached to the N-terminus of the peptide through a Wang linker [167]. Because the peptidyl resin is formed by polymerization of the modified peptide, this method does not correspond exactly to the general strategy depicted in Fig. 3. To be applicable, the modified peptide must be produced in a protected form, otherwise the acrylamide function might react with the nucleophilic groups naturally present on polypeptides. The target peptide was detached subsequently from the polyacrylamide polymer by treatment with trifluoroacetic acid (TFA) in the presence of the appropriate scavengers.

3.2 Purification by Covalent Internal Resin Capture

For all the examples listed in Table 2, the peptide was assembled on a solid support by SPPS, released in solution together with the capped and truncated contaminants and then purified by covalent capture using another solid support. This strategy usually requires two different solid supports because the SPPS is carried out in organic solvents, while the covalent capture step is usually performed in water using deprotected peptides. Another limitation of this strategy resides in the need to set up an efficient chemoselective capture step/release procedure to recover the target peptide at the end. A way to simplify the overall process is to keep the protected peptide produced by SPPS on the resin for introducing the modifier and performing the covalent capture step on the same solid support (Fig. 4). In this case, various chemical reactions operating in organic solvents can potentially be used for linking the modifier to the solid support because chemoselectivity is no longer mandatory. In practice, all the examples published to date rely on the formation of an amide bond between selectively deprotected amine and carboxylic acid functionalities (Table 3). It should be noted that the covalent capture step results in this case in the formation of a tail-to-head cyclic peptide. The selective cleavage of the linker between the C-terminus and the resin, i.e. linker 1, results in the reversal of the peptide chain and in the removal of the capped peptides. The cleavage of the second linker formed in the capture step between the N-terminus and the solid support finally yields the purified peptide in solution.

The principle of the method was pioneered in the early 1990s by several developments in the field of combinatorial peptide synthesis [192]. The production

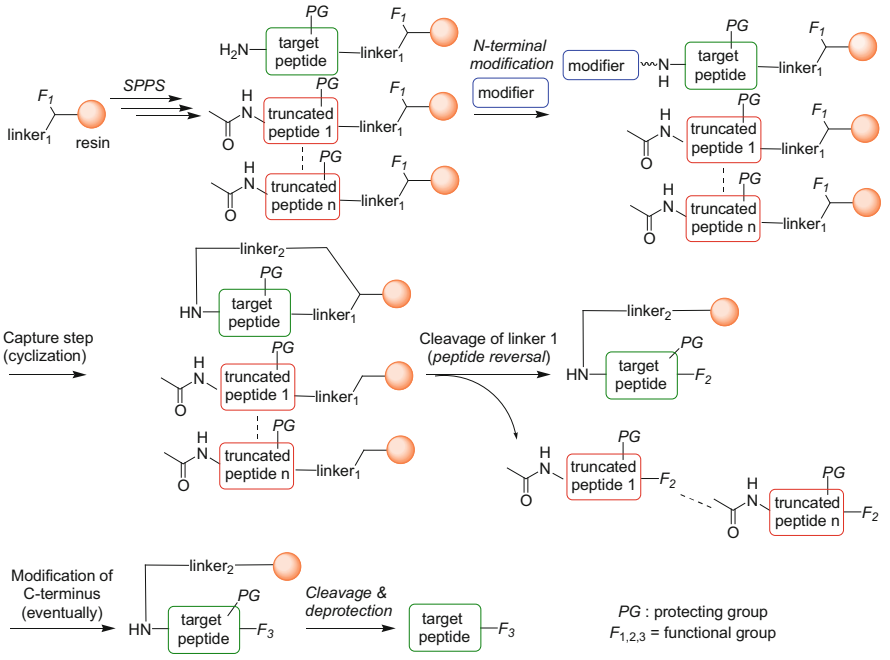
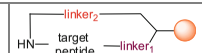
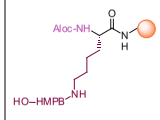
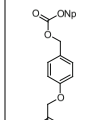
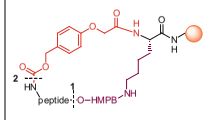
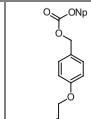
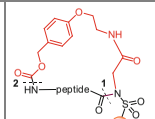
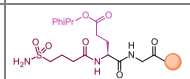
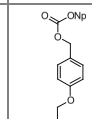
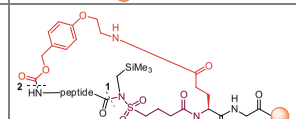


Fig. 4 The target peptide is synthesized by SPPS on the solid support using linker 1 as a starting point. It is subsequently cyclized by reaction of the modifier introduced on the N-terminus with the functional handle F_1present on the solid support (formation of linker 2). This step is usually preceded by the selective activation of the modifier and of F_1 group (not shown for clarity). The selective cleavage of linker 1 results in the release of the truncated peptides and in the reversal of the peptide chain with the exposure of F_2 functionality at the C-terminus. Eventually, F_2 can be converted into F_3 prior to the cleavage of linker 2 which results in the release of the target peptide in solution

of on-bead peptide libraries enables one to take advantage of the one-bead one-compound concept [193, 194] and to use various screening formats for peptide selection. However, for peptides produced by the classical SPPS method, only the N-terminus is available for binding as the C-terminus is linked to the solid support. Moreover, peptides modified at the C-terminus have many applications, for example as substrates or inhibitors of proteolytic enzymes. The production of on-bead peptide libraries displaying a free C-terminus or a C-terminal modification requires the reversal of the peptide orientation after the SPPS. Pioneering studies in this direction were reported by Lebl [195, 196], Holmes [197] and Kania [198]. In these studies, the linker formed in the cyclization step (linker 2 in Fig. 4) was stable in the conditions used for deprotecting the peptide chain, because the goal was to keep the peptide attached to the solid phase for screening purposes.

The potential of the method for the self-purification of peptides produced by SPPS was reported in 1997 by Bradley and coworkers (entry 1, Table 3) [169, 170]. In this work, a lysine residue was used as the starting point of the synthesis.

Table 3 Synthesis and purification of peptides by combining the Fmoc SPPS method and an internal resin capture strategy[a]

Entry	Reference		Modifier		Cleavage 1 (reagent/F_2)	Cleavage 2 (reagent)	F_3
1	[169, 170]				1% TFA/OH	Conc. TFA	OH, various alkyl or aryl amides
2	[171]				EtSH or BnSH, NaSPh, DMF/SEt or SBn	Conc. TFA	SEt or SBn
3	[172]				EtSH or BnSH, NaSPh, DMF/SEt or SBn	Conc. TFA	SEt or SBn

[a]*Aloc* allyloxycarbonyl, *Bn* benzyl, *HO-HMPB* (4-(4-hydroxymethyl-3-methoxyphenoxy)butanoic acid, *Mtt* 4-methyltrityl, *ONp* *p*-nitrophenyloxy, *PhiPr* 2-phenyl-isopropyl

The ε-amino group was derivatized by the highly acid labile 4-(4-hydroxymethyl-3-methoxyphenoxy) butanoic acid (HMPB) linker which plays the role of linker 1 in Fig. 4. The N-terminus of the peptidyl resin was modified by a handle reminiscent of the Wang linker. Then the allylcarbamate and allyl ester protecting groups used for masking the α-amino group of the lysine directly appended to the solid support and the carboxylic acid function of the modifier, respectively, were removed simultaneously using [$Pd(PPh_3)$] in the presence of excess dimedone. The cyclative lactamization step was carried out with PyBrop/DIEA/4-dimethylaminopyridine. The HMPB linker was cleaved first with 1% TFA, thereby allowing the reversal of the peptide chain with formation of a C-terminal carboxylic acid function. In the original work [169], the resin was further treated with neat TFA to release the deprotected and purified peptide in solution. Later on, Bradley and coworkers exploited the C-terminal carboxylic acid function of the reverted and protected peptidyl resin to produce various C-terminal amides derived from 4-*tert*-butoxycarbonylaminobutylamine, serinol(OBn), benzylamine and *p*-nitroaniline [170]. The activation of the C-terminal carboxylic acid function was carried out using PyBrop in the presence of DIEA. Of course, in this case a major issue is the potential racemization of the C-terminal residue through 5(4*H*)-oxazolone formation, which was found to be less than 10%. However, a detailed characterization of the C-terminal peptide amides produced in this study in regard to the racemization problem was not reported.

More recently, the concept of internal resin capture was applied to the synthesis of C-terminal peptide thioesters, which are important building blocks for protein total synthesis using NCL reaction (entries 2 and 3, Table 3). The synthesis of large peptide thioesters using Fmoc SPPS is a recognized limitation to the chemical synthesis of large proteins and has stimulated numerous developments [199]. One important contribution to the field is certainly the application of the safety-catch sulfonamide linker (Kenner linker [200–202]) to the synthesis of peptide thioesters [203–206].

The Kenner linker allows the assembly of the peptide chain by standard Fmoc SPPS. According to this method, the protected peptide chain is C-terminally linked to the solid support through an *N*-acylsulfonamide bond, which is stable in the presence of nucleophiles such as piperidine. Upon alkylation, the *N*-alkyl-*N*-acylsulfonamide can be displaced by a variety of nucleophiles such as hydroxide ion, amines or thiols. The application of the Kenner linker to the synthesis of peptide thioesters by displacement of the *N*-alkyl-*N*-acylsulfonamide with a nucleophilic thiol has been reported independently by the Pessi [203] and Bertozzi [204] groups in 1999. The protected peptide thioesters detached from the solid support during the thiolysis step were subsequently deprotected in TFA.

In the example described in entries 2 and 3 of Table 3, the Kenner linker plays the role of linker 1 (Fig. 4), meaning that the reversal of the peptide chain and the removal of the truncated peptides occur during the thiolysis step leading to the formation of a C-terminal thioester functionality [171, 172]. In both cases, linker 2 is reminiscent of the Wang linker which is labile in concentrated TFA. The difference between entries 2 and 3 of Table 3 resides in the strategy used for

alkylating and activating the *N*-acylsulfonamide moiety. In entry 2 of Table 3 [171], the *N*-acyl sulfonamide was alkylated with the allyl ester of iodoacetic acid. The allyl ester was deprotected and the resulting *N*-carboxymethyl group was used subsequently for lactam formation. At the end, several atoms of the alkylating reagent are incorporated in linker 2 during the cyclization process. In the second approach (entry 3, Table 3) [172], the alkylation and lactamization steps are disconnected. This allows a simplification of the method and the possibility of using milder alkylating reagents such as trimethylsilyldiazomethane for the activation of the *N*-acyl sulfonamide linker.

Internal resin capture strategies have a great potential. However, the paucity of reports in the field highlights the difficulty in setting up internal resin capture/self-purification strategies which require the combination of orthogonal linker and protecting group strategies and the performance of a cyclization step on the solid phase.

3.3 Purification by Selective and Non-Covalent Adsorption on a Solid Support

The preceding section introduced the different methods developed for the purification of peptides by covalent capture on a solid support. An alternative is to use a modifier which induces significant changes in the physico-chemical or binding properties of the modified peptide relative to the capped peptide contaminants. In this case, the modified peptide is trapped selectively by a complementary solid support without involvement of a covalent bond and separated from the impurities by performing simple washing steps (method A, Fig. 5) or by using resolutive chromatographic methods (method B, Fig. 5).

Here, again, Merrifield and coworkers pioneered the field in 1978 by developing the 9-(2-sulfo)fluorenylmethyloxycarbonyl group (SulFmoc, entry 1, Table 4).[58] The SulFmoc group is a modification of the Fmoc amino-protecting group. It is thus removed in the presence of a base such as piperidine. The negatively charged sulfonate group is exploited for separating the target peptide from impurities using ion-exchange chromatography. Merrifield illustrated the method with short peptides. Its usefulness for the purification of large peptides has not been demonstrated and may be complicated if negatively charged residues (Asp, Glu) are present both in the target peptide and in the capped peptide segments.

Another modifier derived from the Fmoc group is the tetrabenzo[a,c,g,i] fluorenyl-17-methoxycarbonyl (TbFmoc) group developed by Ramage and coworkers in the 1990s (entries 2 and 3, Table 4) [173]. The TbFmoc group is highly hydrophobic and enables the isolation of the modified target peptide by adsorption on porous graphitized carbon (PGC, entry 2, Table 4) or by RP-HPLC (entry 3, Table 4). In the former case, the target peptide is released from the PGC solid support by treatment with piperidine. In the latter case, the Tbfmoc-modified

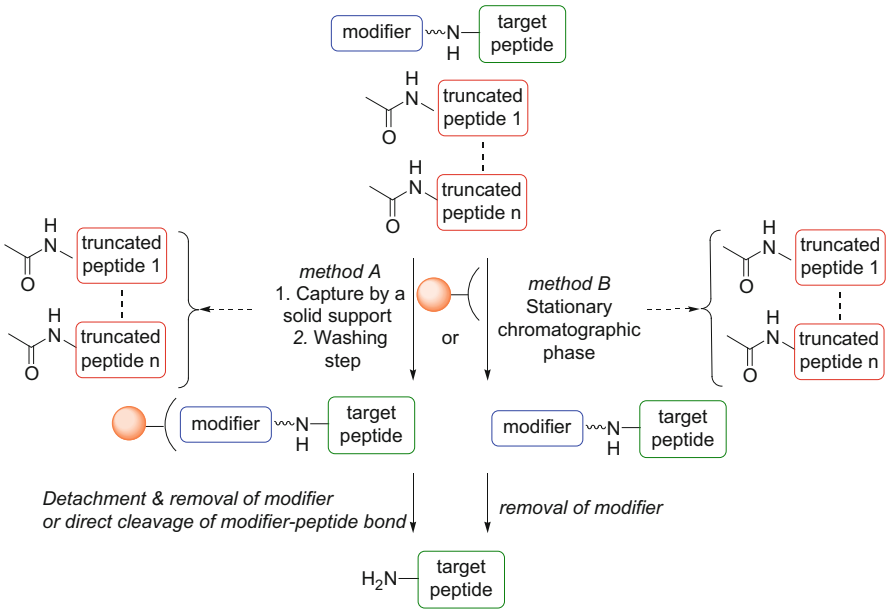


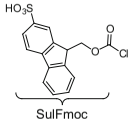
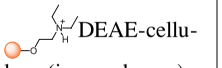
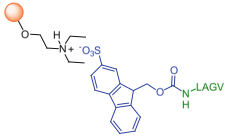
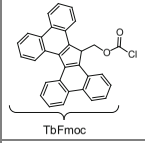
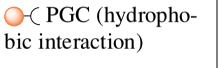
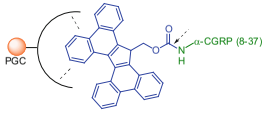
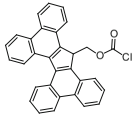
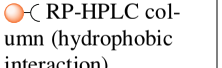
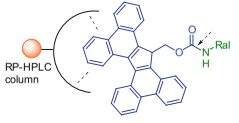
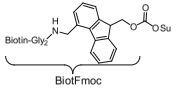
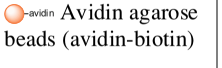
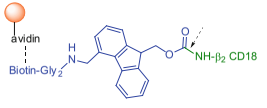
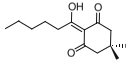
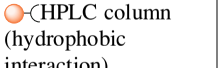
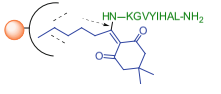
Fig. 5 Purification of peptides by selective adsorption on a solid support

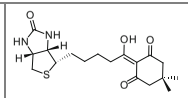
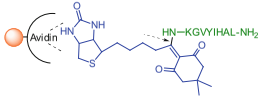
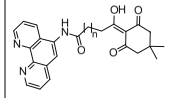
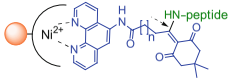
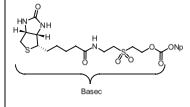
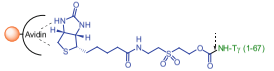
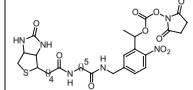
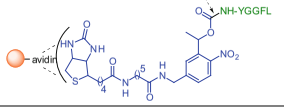
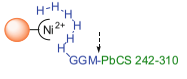
peptide is first purified by RP-HPLC and deprotected subsequently in aqueous solution at pH 8.5.

The combination of the Fmoc group and of biotin-avidin technology led to the development of the BiotFmoc group described at entry 4 of Table 4 [174]. The biotinylated peptide was captured specifically using avidin agarose beads. The target peptide was released subsequently in solution by treating the solid support with aqueous triethylamine.

Another amine protecting group which stimulated the development of several non-covalent capture purification methods is the 1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl (Dde) group introduced by Bycroft and coworkers in 1993 [207, 208]. As with the examples derived from the Fmoc group, the Dde group was modified by a hydrophobic tail (entry 5, Table 4) or a biotin moiety (entry 6, Table 4) to enable the isolation of the target peptide using RP-HPLC or avidin technology, respectively [176]. The last example based on the Dde group exploits the affinity of a 1,10-phenanthroline handle for metal ions for the separation of the modified target peptide by immobilized-metal affinity chromatography (IMAC, entry 7, Table 4) [177]. In any case, the removal of the Dde-based modifiers was performed in the presence of aqueous hydrazine. It should be noted that the Dde group cannot be used for the protection of secondary amines. Consequently, the Dde-based non-covalent purification strategies discussed above cannot be used for the purification of peptides featuring an N-terminal proline residue.

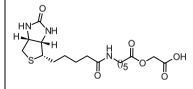
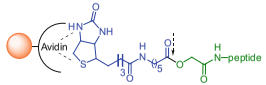
Table 4 Strategies for peptide purification by non-covalent capture[a]

Entry	Peptide (reference)	Modifier	Solid support –x (adsorption principle)	Solid support modifier-target peptide ¦ cleavage site	Method, adsorption and cleavage conditions
1	LAGV (4 AA) [58], for another example see: [59]				Method B: 1. Elution with a gradient of formic acid/ammonium formate in formic acid. 2. Cleavage in solution with piperidine.
2	α-CGRP (8–37) (29 AA) [173], for another example see [83]				Method A: 1. Adsorption on PGC. 2. Cleavage of the PGC-bound peptide with 10% piperidine in 6 M Gdn.HCl/isopropanol:1/1).
3	Ral (41 AA) [173]				Method B: 1. RP-HPLC purification. 2. Cleavage in solution with 6 M Gdn. HCl, 0.1 M TRIS pH 8.5, dithiothreitol, 37°C (4 h).
4	β_2 CD18 (46 AA) [174], see also [175]				Method A: 1. Capture by avidin beads. 2. Cleavage of the immobilized peptide with 5% triethylamine in water (15 min).
5	AgII (8 AA) [176]				Method B: 1. Separation by RP-HPLC. 2. Cleavage in solution with 5% NH_2NH_2 in H_2O.

6	AgII (8 AA) [176]		avidin Avidin agarose beads (avidin-biotin)		Method A: 1. Capture by avidin beads. 2. Cleavage of the immobilized peptide with 5% NH_2NH_2 in H_2O.
7[b]	Model peptide (25 AA) [177]		Ni^{2+} Ni-NTA resin (IMAC) [178, 179]		Method A: 1. Capture by Ni-NTA resin. 2. Cleavage of the immobilized peptide with 10% NH_2NH_2 in ACN/H_2O (1/1).
8[c]	Tγ (1–67) (67 AA) [180]		avidin Avidin beads (avidin-biotin)		Method A: 1. Capture by avidin beads. 2. Cleavage of the immobilized peptide with 5% NH_4OH.
9	YGGFL (5 AA) [181]		avidin Avidin beads (avidin-biotin)		Method A: 1. Capture by avidin beads. 2. Cleavage of the immobilized peptide by photolysis (365 nm) at pH 7.4.
10[d]	PbCS 242–310 (69 AA) [182]	HHHHHHGGM	Ni^{2+} Ni-NTA resin (IMAC)		Method A: 1. Detachment with 250 mM imidazole. 2. CNBr cleavage in solution.

(continued)

Table 4 (continued)

Entry	Peptide (reference)	Modifier	Solid support —*x* (adsorption principle)	Solid support modifier-target peptide ¦ cleavage site	Method, adsorption and cleavage conditions
11[e]	Model peptide (16 AA) [183]		avidin Avidin beads (avidin-biotin)		Method A: 1. Capture by avidin beads. 2. Direct cleavage with CAPSO buffer pH 9.5 (24 h)

[a]*Ag II* angiotensin II receptor binding protein fragment. Octapeptide having the following sequence: KGVYIHAL-NH_2; *CAPSO* 3-(cyclohexylamino)-2-hydroxypropanesulfonic acid, *α-CGRP* α-calcitonin gene related peptide, *DACFmoc* 4-dodecyl aminocarbonyl fluoren-9-ylmethyl oxycarbonyl, *DEAE* diethylaminoethyl, *IMAC* immobilized-metal affinity chromatography, *LAGV* LeuAlaGlyVal tetrapeptide, *Ni-NTA resin* nickel-nitrilotriacetic acid resin, *PbCS Plasmodium berghei* CS protein, *PGC* porous graphitized carbon, *Ral* restriction alleviation protein, *Su* succinimidyl, *SulFmoc* 9-(2-sulfo)fluorenyl-methyloxycarbonyl, *Tγ* transducin γ-subunit, *TbFmoc* tetrabenzo[a,c,g,i]fluorenyl-17-methoxycarbonyl

[b]STKKTQLQLEHLLLDLQMILNGINN-NH_2

[c]A Cys residue was introduced as an Acm protected derivative

[d]Eventually a third step may be necessary to reduce Met(O) into Met: 10% β-mercaptoethanol in 8 M urea, 0.1 M Na_2HPO_4, 0.01 M Tris, pH 8.0

[e]N-terminal glycolyl peptide having the following sequence: $HOCH_2CO$-FSRSDELAKLLRLHAG-NH_2

The example described at entry 8 of Table 4 combines the biotin-avidin technology and the usefulness of the Msc group for the protection of the amino group [180]. Msc-based strategies have attracted a lot of attention for the purification of peptides by covalent capture as already discussed in the previous section (see Esec-based linkers, entries 5–7, Table 2). In this case, the capture of the biotinylated peptide by avidin beads was followed by the treatment of the solid support with ammonia, which triggered the cleavage of the Esec linker and the separation of the target peptide from the solid support.

Note that the methods based on the Fmoc, Dde or Msc protecting groups require basic conditions and/or strong nucleophiles for removing the modifier. The photochemical method described at entry 9 of Table 4 was developed with the aim of generating the target peptide using very mild conditions [181]. The modifier is derived from the 1-(2-nitrophenyl)ethyloxycarbonyl (NPEOC) photolabile amine protecting group [209, 210], which is usually cleaved upon irradiation at 365 nm. In the example described at entry 9 of Table 4, the NPEOC scaffold was decorated with a biotin handle to enable the capture of the target peptide using avidin technology.

Another method, described at entry 10 of Table 4, makes use of the polyhistidine tag, which is often used for the purification of peptides or proteins using IMAC technology. In this work, the His_6 tag was separated from the target peptide by a glycyl–glycyl-methionyl tripeptide linker, which was cleaved by treatment with cyanogen bromide using a method similar to that developed by Merrifield and coworkers (see entry 1, Table 2) [182].

The last example described in Table 4 is a case apart because the target peptide is N-terminally modified by a non-native glycolyl group [183]. The ester bond linking the glycolyl residue to the biotin handle is cleaved at pH 9.5 after performing the capture step with avidin beads.

4 Chemical Protein Synthesis by the Solid Phase Sequential Chemoselective Ligation of Unprotected Peptide Segments

4.1 Advantages of the Solid Phase Approach

This section discusses the emergence of novel chemical methods and strategies for the assembly of large polypeptides by the solid phase sequential chemoselective ligation of unprotected peptides segments. Up to now, the majority of proteins produced by chemical synthesis were assembled in solution using the NCL reaction (see Fig. 1, green squares). In most cases, small proteins with appropriately spaced cysteine residues were assembled by ligating two or three peptide segments sequentially. Efficient one-pot three peptide segments assembly procedures working either in the N-to-C or C-to-N direction have been developed for this purpose

[109]. They allow one isolation step to be skipped, thus saving time and yield. With some exceptions [121], the assembly of more than three peptide segments is usually carried out using convergent approaches, the aim of which is to minimize the number of chemical and isolation steps needed to obtain the target protein [122–124].

The synthesis of large proteins inevitably requires a significant number of chemical steps. Besides the chemoselective ligations themselves, the synthetic schemes can include several activation steps (e.g. for converting latent thioesters into active thioesters), deprotection steps (for example for unmasking cysteine residues) or a desulfurization step to convert cysteines or other thiol-modified amino acids into proteinogenic residues. The experimental conditions used for a given chemical transformation are frequently poorly compatible with the next one. Consequently, a large number of chemical transformations often imply several intermediate isolation steps. Moreover, each chemical or isolation step can be dramatically complicated by the poor solubility of the segments or their tendency to form aggregates.

Not surprisingly, and stimulated by the numerous advantages of the SPPS over solution methods for peptide synthesis, performing sequential NCLs on a solid phase was early viewed as a potential solution to the limitations encountered in solution [211–213]. As already discussed in the introduction, performing the synthesis on a solid phase allows the removal of excess reagents by simple washing and filtration steps. The immobilized peptide cannot precipitate and, by adjusting the loading of the solid support, the aggregation of the peptide can be minimized to ensure the highest accessibility of the reactive sites. The solid support matrix can sometimes assist the folding of the immobilized polypeptide [213–215]. Moreover, the use of a solid support opens the possibility to automate the process, and therefore to produce chemical protein libraries in different formats with the highest reliability. These proteins can be assayed in solution or still attached to the solid support [213]. In principle, most of the methods which were developed for generating combinatorial peptide libraries by SPPS can be extended to the stepwise ligation of unprotected peptide segments on a solid phase, and thus potentially be used for producing chemical protein libraries. Given the high potential of the solid phase approach for protein synthesis, it is surprising to see that the number of papers published in this area can be counted on the fingers of one hand (pink diamonds in Fig. 1).

The first reason which might explain the reluctance to start a solid phase approach is that this method is best suited for the assembly of at least four peptide segments. This is because several highly efficient one-pot three peptide segment strategies working either in the C-to-N [125, 216, 217] or N-to-C [17, 122, 218–220] direction have been developed in solution. The solid phase approach is thus best adapted to highly challenging targets.

The second reason which can be invoked to explain the paucity of reports in the field is the difficulty in developing efficient and simple linker methodologies for attaching the first unprotected peptide segment to the solid support. This point is discussed in a recent report from Jbara and coworkers [221]. Indeed, the method used to immobilize the first segment must be compatible with the native functional

groups present on the peptide segment, but also with those introduced for extending the peptide chain using NCL such as thioester group or thioester surrogates. Moreover, the resulting linker must be stable during the elongation steps and enable the detachment of the target peptide under mild conditions. The paucity of reports in the field is also probably because of the difficulty in adapting the methods developed for protein synthesis in solution to the solid phase.

4.2 Solid Phase Protein Synthesis in the C-to-N Direction

Most of the protein total syntheses in solution reported to date were achieved by the sequential NCL reaction of unprotected peptide segments in the C-to-N direction [109]. Not surpringly, the majority of the solid phase approaches published up to now performed the assembly in the C-to-N direction too. Note that the solid phase approach has been used successfully for the semi-synthesis of Crk-II protein in the C-to-N direction [222]. Although of high significance, this work is not discussed in detail in this review, which is focused on proteins produced by using chemical methods only. Likewise, the Raines's solid phase approach to the semisynthesis of RNAse is not discussed here, although of importance because it is one of the rare examples combining two orthogonal native peptide ligation methods, i.e. NCL and the traceless Staudinger ligation [112, 113], as well as chemical and biological sources for the individual peptide segments [223].

4.2.1 N-Terminal Cysteine Protection Strategies

The assembly of the peptide chain in the C-to-N direction requires temporary protection for N-terminal cysteines and the use of C-terminal thioester peptide segments (Fig. 6). If N-terminal cysteines are left unprotected, the incoming peptide segment can potentially cyclize or oligomerize and thus lead to a loss of expensive material, to incomplete couplings or, even worse, to the insertion of multiple copies of the peptide segments into the growing polypeptide chain. A large variety of protecting groups (*PG*) have been designed for cysteine mainly to be used in SPPS [33]. They are usually removed in organic solvents using specific reagents or during the final cleavage and deprotection step in strong acids such as anhydrous HF or TFA. In contrast, the cysteine protecting groups needed for the C-to-N elongation strategies must be removed rapidly in aqueous solution using mild conditions and, if possible, non-toxic reagents. Such protecting groups were not available when the first solid phase NCL-based sequential ligation approaches were attempted in the late 1990s.

The acetamidomethyl (Acm) group is a classical protecting group for cysteine thiol, which can be removed in the presence of mercury or silver salts and an excess of a thiol such as β-mercaptoethanol or dithiothreitol (Fig. 7). This protection strategy has been used successfully by several groups for protein synthesis in

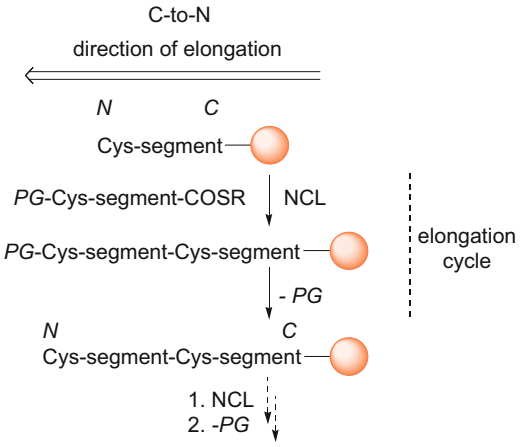


Fig. 6 The C-to-N solid phase sequential NCL strategy requires the temporary protection of N-terminal cysteine residues (*PG* protecting group)

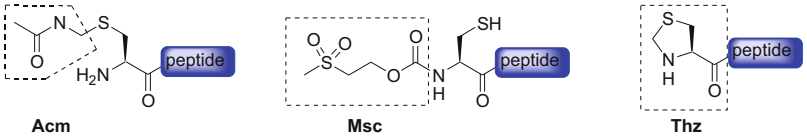


Fig. 7 Structure of acetamidomethyl (Acm), methylsulfonylethyloxycarbonyl (Msc) and thiazolidine (Thz) protecting groups for cysteine

solution [224–226]. Not surprisingly, the first report of a C-to-N NCL-based solid phase elongation process also used Acm protecting group for the cysteine thiol (Fig. 8a) [212]. Soon after the seminal report of Canne and coworkers in 1999, Brik and coworkers reported the solid phase synthesis of vMIP-I-(1–71) chemokine by a three-segment approach using Acm protection strategy too (Fig. 8b) [211]. The authors also evaluated the Msc amine protecting group which, as discussed several times in this review, is removed with aqueous base at pH ~11.

The harsh conditions required for removing Acm or Msc protecting groups have stimulated the development of other methods such as thiazolidine protection (Thz), which was introduced in 2004 (Fig. 7) [125]. It is perhaps today the most popular protection strategy for synthesis of proteins in solution using NCL [8, 228]. Part of this success is because of the ease of introducing Boc-protected 1,3-thiazolidine-4-*R*-carboxylic acid during the last stage of SPPS using either Boc or Fmoc SPPS protocols, but also the mild conditions for its removal which is carried out at pH 4 in the presence of *O*-methylhydroxylamine. The potential of Thz for protein total synthesis was first evaluated in solution by the total synthesis of [V15A] crambin [125]. This protein was also produced by a mixed solution-solid phase approach as shown in Fig. 9a [215]. In this example, the first two chemical steps – the ligation of crambin peptide segments (16–31) and (32–46)-His_6 and the subsequent unmasking

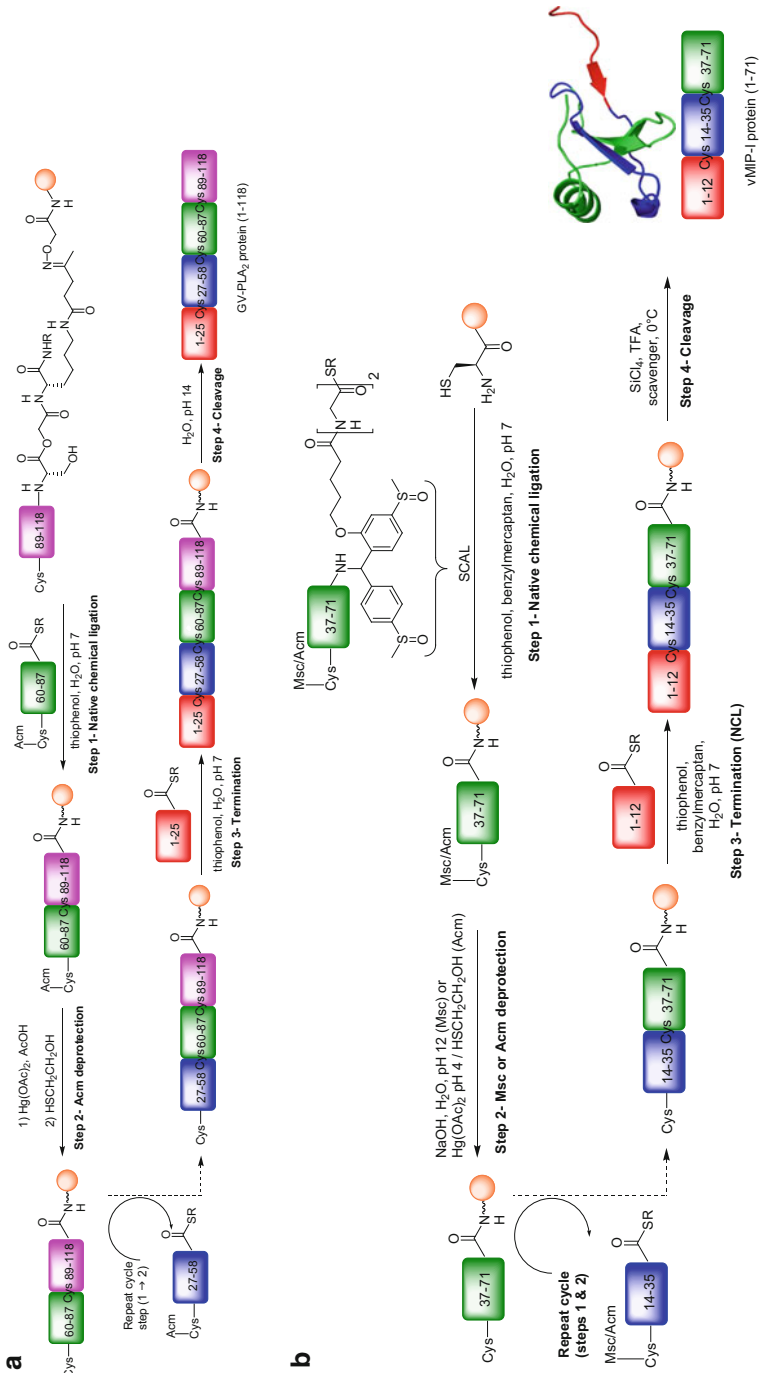


Fig. 8 C-to-N solid phase protein synthesis using NCL reaction and Acm or Msc protection strategy for N-terminal cysteine residues. (**a**) Acm protection strategy [212]. (**b**) Acm or Msc protection strategy [211] (the structure of vMIP-1 protein was generated using pdb entry 1ZXT [227])

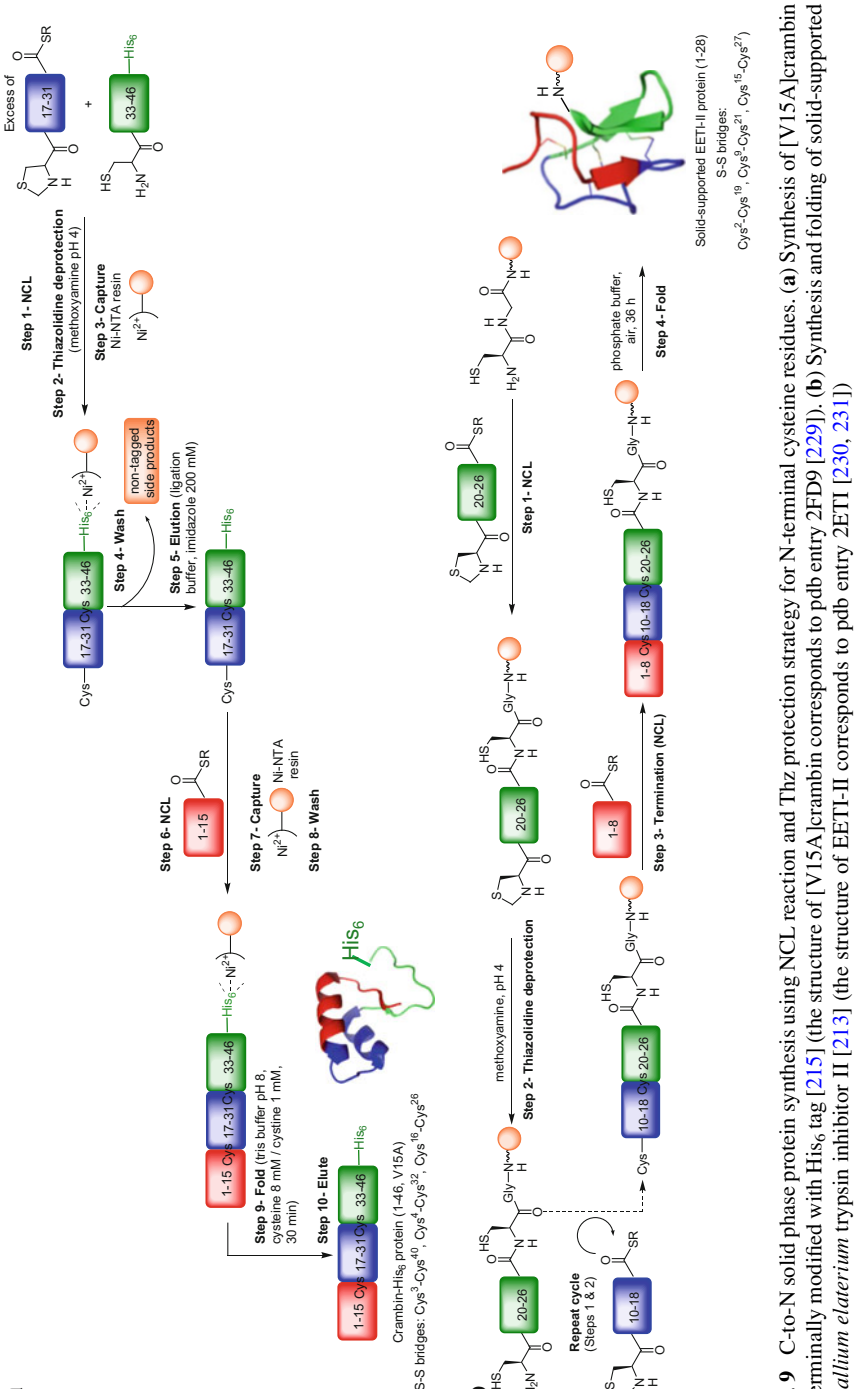


Fig. 9 C-to-N solid phase protein synthesis using NCL reaction and Thz protection strategy for N-terminal cysteine residues. (**a**) Synthesis of [V15A]crambin C-terminally modified with His_6 tag [215] (the structure of [V15A]crambin corresponds to pdb entry 2FD9 [229]). (**b**) Synthesis and folding of solid-supported *Ecballium elaterium* trypsin inhibitor II [213] (the structure of EETI-II corresponds to pdb entry 2ETI [230, 231])

of the N-terminal cysteine – were performed in solution. The presence of His_6 tag on the C-terminus of the crambin (16–46)-His_6 intermediate enabled its selective capture and purification by Ni-NTA affinity chromatography. The product of the second NCL reaction between crambin thioester peptide (1–15, V15A) and crambin (16–46)-His_6 was captured similarly and folded on the Ni-NTA agarose gel. Overall, the isolation of His_6 tagged products by Ni-NTA affinity column purification enabled a significant reduction in the handling losses and the time required to produce the target protein.

The first application of Thz protection strategy to the C-to-N solid phase chemical synthesis of a protein was reported in 2006 by Johnson and coworkers (Fig. 9b) [213]. In this work, the 28-amino acid trypsin inhibitor EETI-II was assembled in 3 pieces. All the chemical steps were performed on a water-compatible resin. After assembly, the peptide was folded and assayed, still attached to the resin. Finally, a very recent application of Thz protection strategy to the solid phase synthesis of histone H2B was reported by Jbara and coworkers [221].

4.2.2 Linker Strategies for C-to-N Solid Phase Elongation Strategies

The recent publication of Jbara and coworkers discusses the difficulty in designing efficient linker strategies for assembling unprotected peptide segments on a solid support as already mentioned at the beginning of this section [221]. This is a serious bottleneck for the development of solid phase methods in the field of protein total synthesis. The seminal work of Kent's group relied on oxime ligation for attaching the first segment to the solid support (Fig. 8a) [212]. An ester linker was used for allowing the cleavage of the peptide from the resin using highly acidic conditions. Alternately, Dawson's group used a highly stable, safety catch amide linker (SCAL [211]) originally developed by Patek and coworkers (Fig. 8b) [232]. SCAL linker is reminiscent of the benzhydrylamine-type linkers which are very popular for peptide amide synthesis using SPPS [233]. The benzhydrylamine core of the SCAL linker is substituted by electron-withdrawing methylsulfinyl groups. SCAL linker is highly stable in strong acids such as TFA or anhydrous HF. It is also stable in aqueous base (pH 13) or in the presence of $Hg(OAc)_2$, which are used for removing Msc or Acm cysteine protecting groups, respectively. The release of the peptide chain from the solid support proceeds through the reduction of the methylsulfinyl groups into methylsulfanyl moieties by treatment with $SiCl_4$/TFA.

Jbara and coworkers proposed a simple linker strategy consisting of a cysteine residue attached to a Rink linker (Fig. 10) [221]. The cysteine residue was introduced by coupling Fmoc-protected 1,3-thiazolidine-4-*R*-carboxylic acid to the Rink PEGA solid support. The solid support was treated with piperidine to remove the Fmoc group and then with *O*-methylhydroxylamine at pH 4 in the presence of TCEP to unmask the cysteine residue. The free cysteine residue was used as a starting point for the assembly of HA-tagged histone H2B polypeptide by assembling four thioester peptide segments sequentially using the NCL reaction. Because

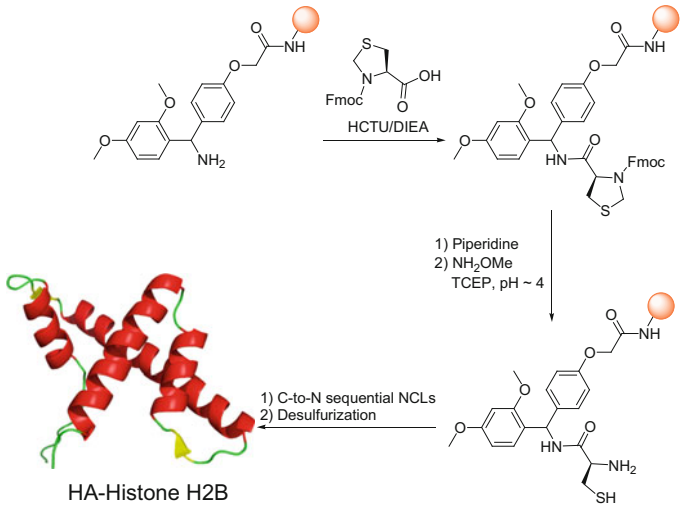


Fig. 10 Preparation of a cysteinyl Rink PEGA solid support for sequential NCL reactions in the C-to-N direction [221]. HA-Histone H2B corresponds to the human histone H2B protein labelled on the N-terminus with the HA tag (the structure of histone H2B protein was generated using pdb entry 1AOI [234])

histone H2B is devoid of cysteine residues, four appropriately spaced alanine residues were replaced by cysteine residues to enable the ligations. The cysteine residues were subsequently desulfurized on the solid phase using Danishefsky conditions [235]. Finally, the target HA-H2B polypeptide was separated from the solid support in concentrated TFA.

4.3 N-to-C Solid Phase Chemical Protein Synthesis

The reports describing N-to-C solid phase elongation strategies are rare with only two papers published in 1999 [212] and 2013 [154]. The difficulty in developing N-to-C elongation methods arises from the challenge in designing latent thioester surrogates which can be activated on demand. Indeed, in the N-to-C direction, each incoming peptide segment must feature a C-terminal blocked thioester group to avoid its cyclization or oligomerization during the NCL reaction (Fig. 11).

In 1999, Kent's group [212] published a seminal paper describing an N-to-C solid phase sequential elongation process relying on the chemical properties of

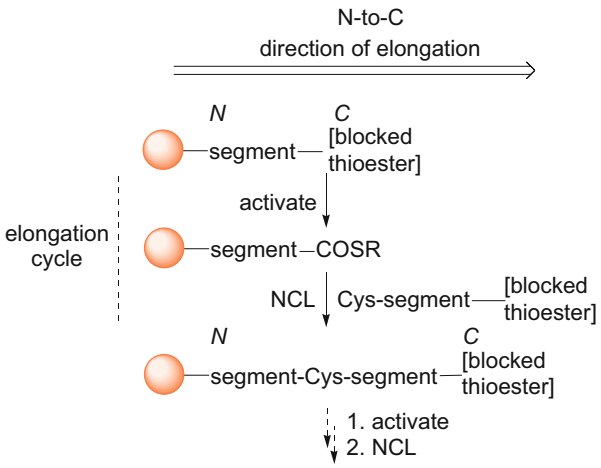


Fig. 11 Solid phase sequential NCLs in the N-to-C direction require the temporary masking of the C-terminal thioester group

C-terminal peptide thioacids (Fig. 12) [237]. Owing to its low pKa, the thioacid group is mainly deprotonated at neutral pH and therefore reacts slowly with the arylthiol catalyst, cysteine thiols or other nucleophiles during the NCL reaction. Interestingly, the thiocarboxylate group can be activated in a subsequent step by alkylation to give an alkylthioester. This transformation was performed selectively at pH 5 in the presence of bromoacetic acid. The chemoselectivity of the alkylation step is governed by the greater nucleophilicity of the thiocarboxylate group in comparison with the other nucleophilic groups naturally present on peptides. Indeed, at this pH cysteine thiols are mainly in the SH form, whereas α- or ε-amino groups are protected by protonation. The method described in Fig. 12 involves, as a first step, the chemoselective attachment of the first segment to the solid support through its N-terminus. The linker used in this study is again derived from the base-labile Msc group. It is modified by a levulinoyl group to enable the formation of a ketoxime bond with an aminooxyacetyl derivatized solid support. In practice, only one activation/NCL elongation cycle was performed on the solid phase as the first segment was introduced with a C-terminal thioester functionality already present. Besides the alkylation step, which can be difficult to control on the solid phase, the method is limited by the capacity of the thiocarboxylate group to react with amine nucleophiles as discussed by the authors themselves and more recently by others [238–240].

The design of a true latent thioester surrogate which could be validated by the N-to-C sequential assembly of five peptide segments was described only recently in the work of Raibaut and coworkers, who exploited the chemical properties of the bis(2-sulfanylethyl) amido group (SEA) [154]. The SEA group was introduced in 2010 as a new member of the *N*,*S*-acyl shift system family [111, 241]. The field has

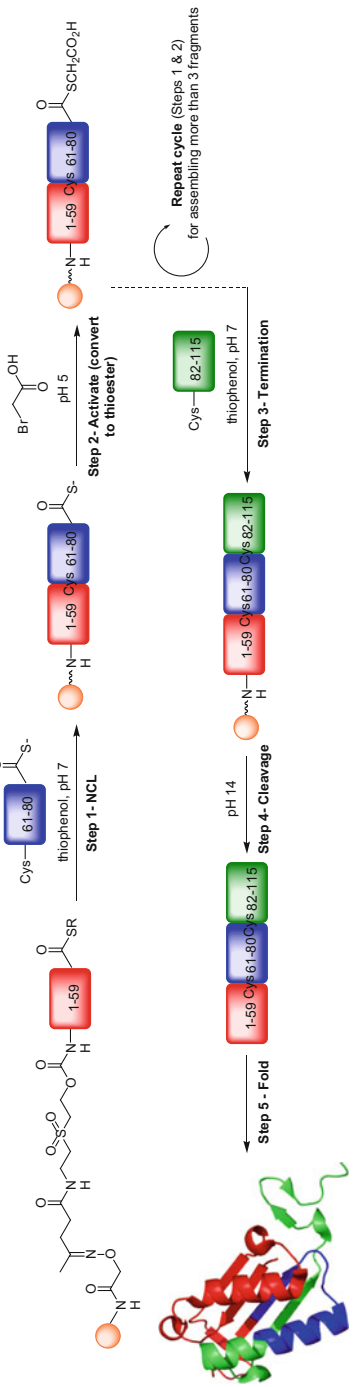


Fig. 12 Protein synthesis by N-to-C solid phase sequential ligation of unprotected peptide segments using NCL reaction. Thioacid functionality is used as a latent thioester surrogate during the NCL reaction. It is activated into an alkylthioester upon alkylation with bromoacetic acid. The structure of human macrophage migration inhibitory factor (MIF) was generated using pdb entry 1GD0 [236]

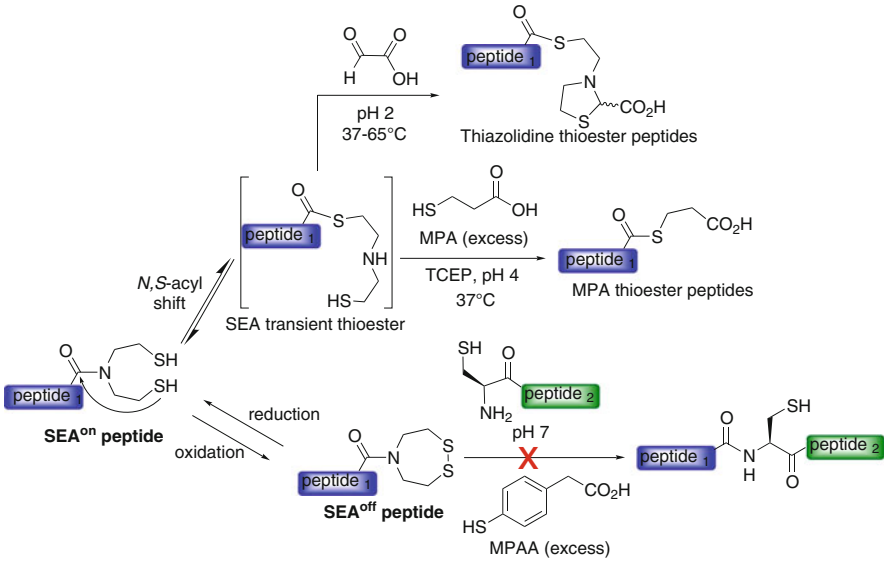


Fig. 13 Main chemical properties of SEA^{on} and SEA^{off} groups exploited for the solid phase N-to-C sequential assembly of unprotected peptide segments

been reviewed recently [242, 243]. The SEA chemical properties that are essential for the solid phase elongation process are highlighted in Fig. 13 [244].

First, the SEA group is a tertiary amide which in the dithiol form called SEA^{on} rearranges spontaneously at neutral to mildly acidic pH into a SEA transient thioester. This transient thioester can react with an exogeneous alkylthiol such as 3-mercaptopropionic acid (MPA) to give a stable alkylthioester through a thiol-thioester exchange reaction [245–247]. This reaction proceeds efficiently in water at pH 4 and constitutes the activation step of the elongation cycle described later on. Another method for converting the SEA^{on} group into an alkylthioester involves trapping the SEA transient thioester with an excess of glyoxylic acid [247]. Indeed, the SEA transient thioester features a β-aminothiol functionality which enables the formation of a stable thiazolidine, thereby displacing the SEA^{on}/SEA transient thioester equilibrium toward thiazolidine thioester peptide derivatives. Although this mode of activation has not yet been illustrated on the solid phase, the high reactivity of thiazolidine thioester peptides might be of interest for forming difficult junctions such as Val-Cys [247] or Pro-Cys [248] peptide bonds.

The second important property is the ease of interconverting the SEA^{on} group and the corresponding cyclic disulfide form called SEA^{off} by oxidation/reduction. Interestingly, the SEA^{off} group is inert in the presence of 4-mercaptophenylacetic acid (MPAA), which is used to catalyse the NCL reaction, even when this aromatic thiol is used in large excess (200 mM) as is usually the case to work in the best kinetic conditions [16, 17, 109, 126, 249]. Thanks to this property, the SEA^{off} group acts as a latent thioester surrogate during the NCL reaction.

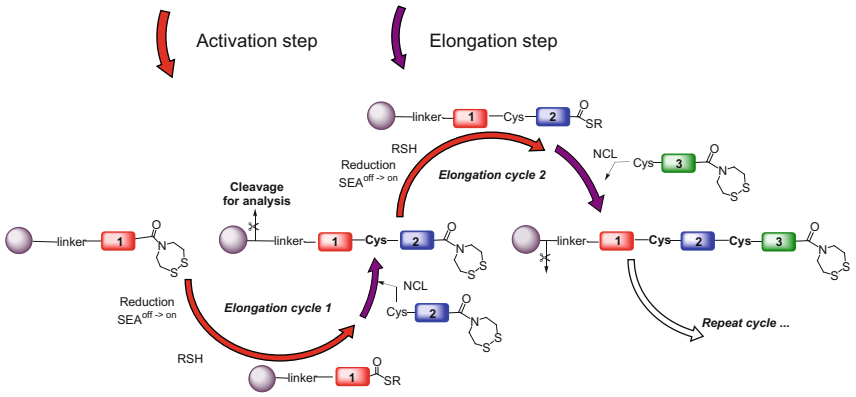


Fig. 14 Principle of the N-to-C elongation cycle based on SEA chemistry

In summary and as illustrated in Fig. 14, the N-to-C elongation cycle based on SEA chemistry consists of (1) converting the SEA^{off} group into an MPA thioester by treating the SEA^{off} peptidyl resin with a strong disulfide bond reducing agent such as TCEP in the presence of MPA at pH 4, (2) washing the resin to remove the excess of MPA and TCEP and (3) performing the NCL reaction with the incoming SEA^{off} peptide segment in the presence of MPAA [154].

The method was illustrated with the synthesis of a 94- or 135-amino acid polypeptides by ligating sequentially 4 or 5 peptide segments respectively. The synthesis of the 94-amino acid polypeptide is depicted in Fig. 15. The first SEA^{off} segment was anchored chemoselectively to the solid support through its N-terminus using the copper-catalyzed or the strain-promoted alkyne azide cycloaddition reaction, i.e. CuAAC [155, 156] and SPAAC [250], respectively. For this, the first SEA^{off} peptide segment was modified on the N-terminus with an azide-functionalized Esec handle using the method developed by Aucagne and coworkers ([153]; see entry 7 of Table 2), while the solid support was modified by a terminal alkyne or a cyclooctyne derivative. The ethylsulfonyl-2-ethyloxycarbonyl moiety of Esec linker can be cleaved with aqueous base (pH ~11) as already discussed in the previous sections. It is stable in the neutral or mildly acidic conditions used for the elongation cycle, while its lability in basic conditions is exploited for monitoring the elongation process or for cleaving off the target polypeptide by treatment of the peptidyl resin with aqueous base. The overall isolated yield of the target 94-amino acid polypeptide was 6.5% including the HPLC purification step. Nine chemical steps were performed on the solid phase, meaning an average yield per step of 74%. A similar average yield per step was reported for the assembly of five peptide segments.

As illustrated in Fig. 15, the method has also been used for isolating a large SEA^{off} peptide corresponding to the first 60 amino acids, thanks to the stability of the SEA^{off} group in basic conditions. The synthesis of large peptide thioesters or thioester surrogates is a known limitation for accessing large proteins using the

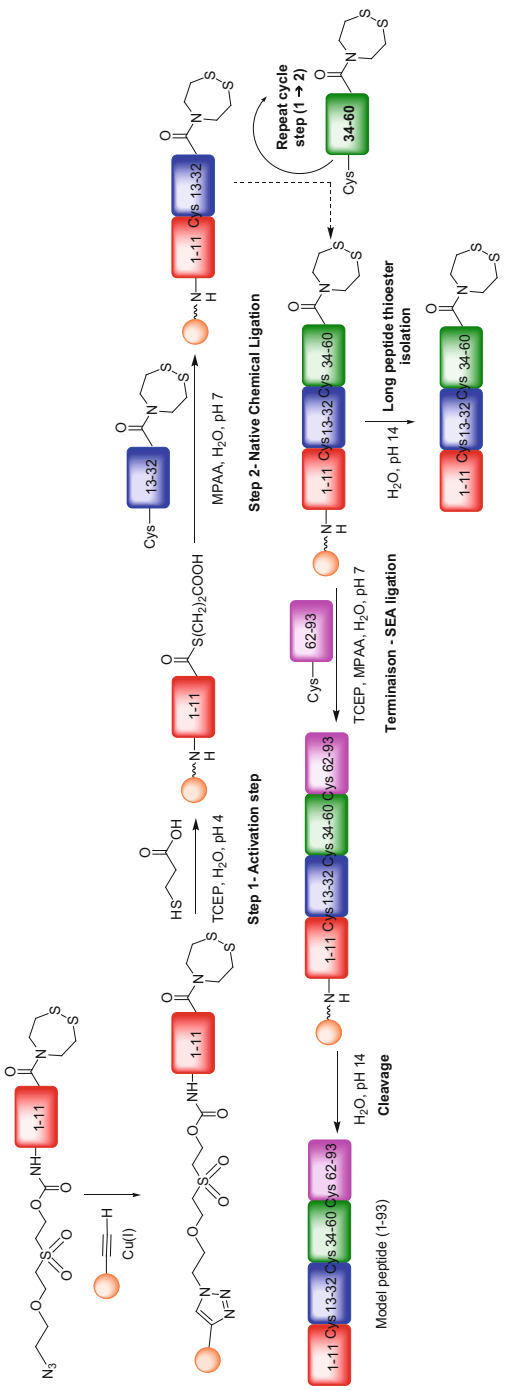


Fig. 15 Protein synthesis by N-to-C solid phase sequential ligation of unprotected peptide segments using NCL reaction. C-terminal *N*-peptidyl-perhydro-1,2,5-dithiazepine functionality (C-terminal bis(2-sulfanylethyl)amido group in its cyclic disulfide form) [244] is used as a latent thioester surrogate. It is activated upon reduction with TCEP and exchange by an alkylthiol such as 3-mercaptopropionic acid

NCL reaction. The method described in Fig. 15 is a potential solution to this highly challenging problem.

5 Conclusion

About 50 years after the introduction of the SPPS method by Merrifield, the field of solid phase chemical protein synthesis is still a dynamic and highly challenging field of research. The SPPS method is continuously improving and enables the synthesis of small proteins. The purification of large peptides produced by SPPS can be simplified by using one of the several capture purification methods which were developed over the past 30 years. The SPPS technique also gives access to large unprotected peptide segments which can be used for assembling proteins using the NCL reaction. The adaptation of NCL chemistry to the solid phase might extend considerably the limits of protein chemical synthesis and allow the automation of the process. Recent developments have shown the great potential of the method in the N-to-C and C-to-N directions. We expect automated solid phase protein synthesis relying on the sequential ligation of unprotected peptide segments to become a popular technique for accessing large polypeptides in the near future.

References

1. Shimko JC, North JA, Bruns AN, Poirier MG, Ottesen JJ (2011) Preparation of fully synthetic histone H3 reveals that acetyl-lysine 56 facilitates protein binding within nucleosomes. J Mol Biol 408(2):187–204
2. Manohar M, Mooney AM, North JA, Nakkula RJ, Picking JW, Edon A, Fishel R, Poirier MG, Ottesen JJ (2009) Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding. J Biol Chem 284(35):23312–23321
3. Chiang KP, Jensen MS, McGinty RK, Muir TW (2009) A semisynthetic strategy to generate phosphorylated and acetylated histone H2B. Chembiochem 10(13):2182–2187
4. Hejjaoui M, Butterfield S, Fauvet B, Vercruysse F, Cui J, Dikiy I, Prudent M, Olschewski D, Zhang Y, Eliezer D, Lashuel HA (2012) Elucidating the role of C-terminal post-translational modifications using protein semisynthesis strategies: alpha-synuclein phosphorylation at tyrosine 125. J Am Chem Soc 134(11):5196–5210
5. He S, Bauman D, Davis JS, Loyola A, Nishioka K, Gronlund JL, Reinberg D, Meng F, Kelleher N, McCafferty DG (2003) Facile synthesis of site-specifically acetylated and methylated histone proteins: reagents for evaluation of the histone code hypothesis. Proc Natl Acad Sci U S A 100(21):12033–12038
6. Unverzagt C, Kajihara Y (2013) Chemical assembly of N-glycoproteins: a refined toolbox to address a ubiquitous posttranslational modification. Chem Soc Rev 42(10):4408–4420
7. Wang P, Dong S, Brailsford JA, Iyer K, Townsend SD, Zhang Q, Hendrickson RC, Shieh J, Moore MA, Danishefsky SJ (2012) At last: erythropoietin as a single glycoform. Angew Chem Int Ed 51(46):11576–11584
8. Piontek C, Varon Silva D, Heinlein C, Pohner C, Mezzato S, Ring P, Martin A, Schmid FX, Unverzagt C (2009) Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 2. Angew Chem Int Ed 48(11):1941–1945

9. Piontek C, Ring P, Harjes O, Heinlein C, Mezzato S, Lombana N, Pohner C, Puttner M, Varon Silva D, Martin A, Schmid FX, Unverzagt C (2009) Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 1. Angew Chem Int Ed 48(11):1936–1940
10. Kumar KS, Spasser L, Erlich LA, Bavikar SN, Brik A (2011) Total chemical synthesis of di-ubiquitin chains. Angew Chem Int Ed 49(48):9126–9131
11. 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 (27):6137–6141
12. 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(2):405–409
13. Fierz B, Chatterjee C, McGinty RK, Bar-Dagan M, Raleigh DP, Muir TW (2011) Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. Nat Chem Biol 7(2):113–119
14. Yang R, Pasunooti KK, Li F, Liu X-W, Liu C-F (2010) Synthesis of K48-linked diubiquitin using dual native chemical ligation at lysine. Chem Commun 46(38):7199–7201
15. McGinty RK, Chatterjee C, Muir TW (2009) Semisynthesis of ubiquitylated proteins. Methods Enzymol 462:225–243
16. 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
17. Boll E, Drobecq H, Ollivier N, Blanpain A, Raibaut L, Desmet R, Vicogne J, Melnyk O (2014) One-pot chemical synthesis of small ubiquitin-like modifier (SUMO) protein-peptide conjugates using *bis*(2-sulfanylethyl)amido peptide latent thioester surrogates. Nat Protoc 10:269–292
18. Pellois JP, Muir TW (2006) Semisynthetic proteins in mechanistic studies: using chemistry to go where nature can't. Curr Opin Chem Biol 10(5):487–491
19. Fischer E, Fourneau E (1901) Ueber einige Derivate des Glykokolls. Ber Dtsch Chem Ges 34:2868–2879
20. Merrifield RB (1963) Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. J Am Chem Soc 85(14):2149–2154
21. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Synthesis of proteins by native chemical ligation. Science 266(5186):776–779
22. Kent SB (2009) Total chemical synthesis of proteins. Chem Soc Rev 38(2):338–351
23. Merrifield RB (1965) Automated synthesis of peptides. Science 150(3693):178–185
24. Merrifield RB, Stewart JM (1965) Automated peptide synthesis. Nature 207(996):522–523
25. Merrifield RB, Stewart JM, Jernberg N (1966) Instrument for automated synthesis of peptides. Anal Chem 38(13):1905–1914
26. Gutte B, Merrifield RB (1971) The synthesis of ribonuclease A. J Biol Chem 246 (6):1922–1941
27. Kresge N, Simoni RD, Hill RL (2006) The solid phase synthesis of ribonuclease A by Robert Bruce Merrifield. J Biol Chem 281(26):e21
28. Atherton E, Fox H, Harkiss D, Logan CJ, Sheppard RC, Williams BJ (1978) A mild procedure for solid phase peptide synthesis: use of fluorenylmethoxycarbonylamino-acids. J Chem Soc Chem Commun 13:537–539
29. Fields GB, Noble RL (1990) Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. Int J Pept Protein Res 35(3):161–214
30. Coin I, Beyermann M, Bienert M (2007) Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. Nat Protoc 2(12):3247–3256
31. El-Faham A, Albericio F (2011) Peptide coupling reagents, more than a letter soup. Chem Rev 111(11):6557–6602

32. Johnson T, Quibell M, Owen D, Sheppard RC (1993) A reversible protecting group for the amide bond in peptides. Use in the synthesis of 'difficult sequences'. J Chem Soc Chem Commun 4:369–372
33. Isidro-Llobet A, Ãlvarez M, Albericio F (2009) Amino acid-protecting groups. Chem Rev 109(6):2455–2504
34. Mutter M, Nefzi A, Sato T, Sun X, Wahl F, Wohr T (1995) Pseudo-prolines (psi Pro) for accessing "inaccessible" peptides. Pept Res 8(3):145–153
35. Haack T, Mutter M (1992) Serine derived oxazolidines as secondary structure disrupting, solubilizing building blocks in peptide synthesis. Tetrahedron Lett 33(12):1589–1592
36. Wôhr T, Mutter M (1995) Pseudo-prolines in peptide synthesis: direct insertion of serine and threonine derived oxazolidines in dipeptides. Tetrahedron Lett 36(22):3847–3848
37. Sohma Y, Hayashi Y, Skwarczynski M, Hamada Y, Sasaki M, Kimura T, Kiso Y (2004) O-N intramolecular acyl migration reaction in the development of prodrugs and the synthesis of difficult sequence-containing bioactive peptides. Biopolymers 76(4):344–356
38. Sohma Y, Hayashi Y, Kimura M, Chiyomori Y, Taniguchi A, Sasaki M, Kimura T, Kiso Y (2005) The '*O*-acyl isopeptide method' for the synthesis of difficult sequence-containing peptides: application to the synthesis of Alzheimer's disease-related amyloid beta peptide (Abeta) 1–42. J Pept Sci 11(8):441–451
39. Sohma Y, Taniguchi A, Skwarczynski M, Yoshiya T, Fukao F, Kimura T, Hayashi Y, Kiso Y (2006) *O*-Acyl isopeptide method for the efficient synthesis of difficult sequence-containing peptides: use of *O*-acyl isodipeptide unit. Tetrahedron Lett 47(18):3013–3017
40. Sohma Y, Sasaki M, Hayashi Y, Kimura T, Kiso Y (2004) Design and synthesis of a novel water-soluble Aβ1-42 isopeptide: an efficient strategy for the preparation of Alzheimer's disease-related peptide, Aβ1-42, via O-N intramolecular acyl migration reaction. Tetrahedron Lett 45(31):5965–5968
41. Carpino LA, Krause E, Sferdean CD, Schümann M, Fabian H, Bienert M, Beyermann M (2004) Synthesis of difficult peptide sequences: application of a depsipeptide technique to the Jung-Redemann 10- and 26-mers and the amyloid peptide Aβ(1–42). Tetrahedron Lett 45 (40):7519–7523
42. Coin I, Dölling R, Krause E, Bienert M, Beyermann M, Sferdean CD, Carpino LA (2006) Depsipeptide methodology for solid-phase peptide synthesis: circumventing side reactions and development of an automated technique via depsidipeptide units. J Org Chem 71 (16):6171–6177
43. Mutter M, Chandravarkar A, Boyat C, Lopez J, Dos Santos S, Mandal B, Mimna R, Murat K, Patiny L, Saucede L, Tuchscherer G (2004) Switch peptides in statu nascendi: induction of conformational transitions relevant to degenerative diseases. Angew Chem Int Ed 43 (32):4172–4178
44. Dos Santos S, Chandravarkar A, Mandal B, Mimna R, Murat K, Saucede L, Tella P, Tuchscherer G, Mutter M (2005) Switch-peptides: controlling self-assembly of amyloid beta-derived peptides *in vitro* by consecutive triggering of acyl migrations. J Am Chem Soc 127(34):11888–11889
45. Guillier F, Orain D, Bradley M (2000) Linkers and cleavage strategies in solid-phase organic synthesis and combinatorial chemistry. Chem Rev 100(6):2091–2158
46. Boas U, Brask J, Jensen KJ (2009) Backbone amide linker in solid-phase synthesis. Chem Rev 109(5):2092–2118
47. James IW (1999) Linkers for solid phase organic synthesis. Tetrahedron 55(16):4855–4946
48. Lloyd-Williams P, Albericio F, Giralt E (1993) Convergent solid-phase peptide synthesis. Tetrahedron 49(48):11065–11133
49. Benz H (1994) The role of solid-phase fragment condensation (SPFC) in peptide synthesis. Synthesis 1994(04):337–358
50. Merrifield RB (1964) Solid phase peptide synthesis. II. The synthesis of bradykinin. J Am Chem Soc 86(2):304–305
51. Merrifield RB (1964) Solid-phase peptide synthesis. III. An improved synthesis of bradykinin. Biochemistry 3(9):1385–1390

52. Stewart JM, Woolley DW (1965) Importance of the carboxyl end of bradykinin and other peptides. Nature 207(5002):1160–1161
53. Marglin B, Merrifield RB (1966) The synthesis of bovine insulin by the solid phase method. J Am Chem Soc 88(21):5051–5052
54. Lenard J, Robinson AB (1967) Use of hydrogen fluoride in Merrifield solid-phase peptide synthesis. J Am Chem Soc 89(1):181–182
55. Gutte B, Merrifield RB (1969) Total synthesis of an enzyme with ribonuclease A activity. J Am Chem Soc 91(2):501–502
56. Wang S-S, Kulesha ID (1975) Preparation of protected peptide intermediates for a synthesis of the ovine pituitary growth hormone sequence 96–135. J Org Chem 40(9):1227–1234
57. Mitchell AR, Erickson BW, Ryabtsev MN, Hodges RS, Merrifield RB (1976) tert-Butoxycarbonylaminoacyl-4-(oxymethyl)phenylacetamidomethyl-resin, a more acid-resistant support for solid-phase peptide synthesis. J Am Chem Soc 98(23):7357–7362
58. Merrifield RB, Bach AE (1978) 9-(2-Sulfo)fluorenylmethyloxycarbonyl chloride, a new reagent for the purification of synthetic peptides. J Org Chem 43(25):4808–4816
59. Wong TW, Merrifield RB (1980) Solid-phase synthesis of thymosin α_1 using *tert*-butyloxycarbonylaminoacyl-4-(oxymethyl)phenylacetamidomethyl-resin. Biochemistry 19 (14):3233–3238
60. König W, Geiger R (1970) Eine neue Methode zur Synthese von Peptiden: Aktivierung der Carboxylgruppe mit Dicyclohexylcarbodiimid unter Zusatz von 1-Hydroxy-benzotriazolen. Chem Ber 103(3):788–798
61. Li CH, Yamashiro D, Gospodarowicz D, Kaplan SL, Van Vliet G (1983) Total synthesis of insulin-like growth factor I (somatomedin C). Proc Natl Acad Sci U S A 80(8):2216–2220
62. Tam JP, Heath WF, Merrifield RB (1982) Improved deprotection in solid phase peptide synthesis: removal of protecting groups from synthetic peptides by an SN2 mechanism with low concentrations of HF in dimethylsulfide. Tetrahedron Lett 23(43):4435–4438
63. Tam JP, Heath WF, Merrifield RB (1982) Improved deprotection in solid phase peptide synthesis: quantitative reduction of methionine sulfoxide to methionine during HF cleavage. Tetrahedron Lett 23(29):2939–2942
64. Tam JP, Heath WF, Merrifield RB (1983) An SN2 deprotection of synthetic peptides with a low concentration of hydrofluoric acid in dimethyl sulfide: evidence and application in peptide synthesis. J Am Chem Soc 105(21):6442–6455
65. Tam JP, Marquardt H, Rosberger DF, Wong TW, Todaro GJ (1984) Synthesis of biologically active rat transforming growth factor I. Nature 309(5966):376–378
66. Heath WF, Merrifield RB (1986) A synthetic approach to structure-function relationships in the murine epidermal growth factor molecule. Proc Natl Acad Sci U S A 83(17):6367–6371
67. Darke PL, Nutt RF, Brady SF, Garsky VM, Ciccarone TM, Leu CT, Lumma PK, Freidinger RM, Veber DF, Sigal IS (1988) HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins. Biochem Biophys Res Commun 156(1):297–303
68. Nutt RF, Brady SF, Darke PL, Ciccarone TM, Colton CD, Nutt EM, Rodkey JA, Bennett CD, Waxman LH, Sigal IS, Anderson PS, Veber DF (1988) Chemical synthesis and enzymatic activity of a 99-residue peptide with a sequence proposed for the human immunodeficiency virus protease. Proc Natl Acad Sci U S A 85(19):7129–7133
69. Schneider J, Kent SB (1988) Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease. Cell 54(3):363–368
70. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SB, Wlodawer A (1989) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. Science 246(4934):1149–1152
71. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SB (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science 245(4918):616–621
72. Ramage R, Green J, Ogunjobi OM (1989) Solid phase peptide synthesis of ubiquitin. Tetrahedron Lett 30(16):2149–2152

73. Ramage R, Green J, Muir TW, Ogunjobi OM, Love S, Shaw K (1994) Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. Biochem J 299:151–158
74. Barlos K, Gatos D, Schäfer W (1991) Synthesis of prothymosin α (ProTα)—a protein consisting of 109 amino acid residues. Angew Chem Int Ed 30(5):590–593
75. Barlos K, Chatzi O, Gatos D, Stavropoulos G (1991) 2-Chlorotrityl chloride resin. Studies on anchoring of Fmoc-amino acids and peptide cleavage. Int J Pept Protein Res 37(6):513–520
76. Dourtoglou V, Ziegler J-C, Gross B (1978) L'hexafluorophosphate de O-benzotriazolyl-N, N-tetramethyluronium: un reactif de couplage peptidique nouveau et efficace. Tetrahedron Lett 19(15):1269–1272
77. Milton RC, Milton SC, Kent SB (1992) Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. Science 256 (5062):1445–1448
78. Zawadzke LE, Berg JM (1992) A racemic protein. J Am Chem Soc 114(10):4002–4003
79. Green LM, Berg JM (1989) A retroviral Cys-Xaa2-Cys-Xaa4-His-Xaa4-Cys peptide binds metal ions: spectroscopic studies and a proposed three-dimensional structure. Proc Natl Acad Sci U S A 86(11):4047–4051
80. Schnolzer M, Alewood P, Jones A, Alewood D, Kent SB (1992) In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. Int J Pept Protein Res 40(3–4):180–193
81. Fitzgerald MC, Chernushevich I, Standing KG, Kent SBH, Whitman CP (1995) Total chemical synthesis and catalytic properties of the enzyme enantiomers L- and D-4-oxalocrotonate tautomerase. J Am Chem Soc 117(45):11075–11080
82. Schnolzer M, Rackwitz HR, Gustchina A, Laco GS, Wlodawer A, Elder JH, Kent SB (1996) Comparative properties of feline immunodeficiency virus (FIV) and human immunodeficiency virus type 1 (HIV-1) proteinases prepared by total chemical synthesis. Virology 224 (1):268–275
83. Ramage R, Raphy G (1992) Design on an affinity-based Nα-amino protecting group for peptide synthesis: tetrabenzo[a, c, g, i]fluorenyl-17-methyl urethanes (Tbfmoc). Tetrahedron Lett 33(3):385–388
84. Love SG, Muir TW, Ramage R, Shaw KT, Alexeev D, Sawyer L, Kelly SM, Price NC, Arnold JE, Mee MP, Mayer RJ (1997) Synthetic, structural and biological studies of the ubiquitin system: synthesis and crystal structure of an analogue containing unnatural amino acids. Biochem J 323:727–734
85. Alexeev D, Bury SM, Turner MA, Ogunjobi OM, Muir TW, Ramage R, Sawyer L (1994) Synthetic, structural and biological studies of the ubiquitin system: chemically synthesized and native ubiquitin fold into identical three-dimensional structures. Biochem J 299 (Pt 1):159–163
86. Layfield R, Franklin K, Landon M, Walker G, Wang P, Ramage R, Brown A, Love S, Urquhart K, Muir T, Baker R, Mayer RJ (1999) Chemically synthesized ubiquitin extension proteins detect distinct catalytic capacities of deubiquitinating enzymes. Anal Biochem 274 (1):40–49
87. Franklin K, Layfield R, Landon M, Ramage R, Brown A, Love S, Muir T, Urquhart K, Bownes M, Mayer RJ (1997) Capillary electrophoresis assay for ubiquitin carboxyl-terminal hydrolases with chemically synthesized ubiquitin-valine as substrate. Anal Biochem 247 (2):305–309
88. Ball HL, King DS, Cohen FE, Prusiner SB, Baldwin MA (2001) Engineering the prion protein using chemical synthesis. J Pept Res 58(5):357–374
89. Ball H, Mascagni P (1995) *N*-(2-Chlorobenzyloxycarbonyloxy)-succinimide as a terminating agent for solid-phase peptide synthesis: application to a one-step purification procedure. Lett Pept Sci 2(1):49–57
90. Ball HL, Mascagni P (1992) Purification of synthetic peptides using reversible chromatographic probes based on the Fmoc molecule. Int J Pept Protein Res 40(5):370–379

91. Bonetto V, Massignan T, Chiesa R, Morbin M, Mazzoleni G, Diomede L, Angeretti N, Colombo L, Forloni G, Tagliavini F, Salmona M (2002) Synthetic miniprion PrP106. J Biol Chem 277(35):31327–31334
92. Dong CZ, Romieu A, Mounier CM, Heymans F, Roques BP, Godfroid JJ (2002) Total direct chemical synthesis and biological activities of human group IIA secretory phospholipase A2. Biochem J 365(Pt 2):505–511
93. Alexeev D, Barlow PN, Bury SM, Charrier JD, Cooper A, Hadfield D, Jamieson C, Kelly SM, Layfield R, Mayer RJ, McSparron H, Price NC, Ramage R, Sawyer L, Starkmann BA, Uhrin D, Wilken J, Young DW (2003) Synthesis, structural and biological studies of ubiquitin mutants containing (2S, 4S)-5-fluoroleucine residues strategically placed in the hydrophobic core. Chembiochem 4(9):894–896
94. Carpino LA, El-Faham A, Minor CA, Albericio F (1994) Advantageous applications of azabenzotriazole (triazolopyridine)-based coupling reagents to solid-phase peptide synthesis. J Chem Soc Chem Commun 2:201–203
95. Abdelmoty I, Albericio F, Carpino L, Foxman B, Kates S (1994) Structural studies of reagents for peptide bond formation: crystal and molecular structures of HBTU and HATU. Lett Pept Sci 1(2):57–67
96. Carpino LA, Imazumi H, El-Faham A, Ferrer FJ, Zhang C, Lee Y, Foxman BM, Henklein P, Hanay C, Mugge C, Wenschuh H, Klose J, Beyermann M, Bienert M (2002) The uronium/guanidinium peptide coupling reagents: finally the true uronium salts. Angew Chem Int Ed 41 (3):441–445
97. Rink H (1987) Solid-phase synthesis of protected peptide fragments using a trialkoxy-diphenyl-methylester resin. Tetrahedron Lett 28(33):3787–3790
98. White P, Keyte JW, Bailey K, Bloomberg G (2004) Expediting the Fmoc solid phase synthesis of long peptides through the application of dimethyloxazolidine dipeptides. J Pept Sci 10(1):18–26
99. Svobodova J, Cabrele C (2006) Stepwise solid-phase synthesis and spontaneous homodimerization of the helix-loop-helix protein Id3. Chembiochem 7(8):1164–1168
100. Hood CA, Fuentes G, Patel H, Page K, Menakuru M, Park JH (2008) Fast conventional Fmoc solid-phase peptide synthesis with HCTU. J Pept Sci 14(1):97–101
101. Garcia-Martin F, Quintanar-Audelo M, Garcia-Ramos Y, Cruz LJ, Gravel C, Furic R, Cote S, Tulla-Puche J, Albericio F (2006) ChemMatrix, a poly(ethylene glycol)-based support for the solid-phase synthesis of complex peptides. J Comb Chem 8(2):213–220
102. Patel H, Chantell CA, Fuentes G, Menakuru M, Park JH (2008) Resin comparison and fast automated stepwise conventional synthesis of human SDF-1α. J Pept Sci 14(12):1240–1243
103. Castro B, Dormoy JR, Evin G, Selve C (1975) Reactifs de couplage peptidique I (1) - l'hexafluorophosphate de benzotriazolyl *N*-oxytrisdimethylamino phosphonium (B.O.P.). Tetrahedron Lett 16(14):1219–1222
104. Coste J, Le-Nguyen D, Castro B (1990) PyBOP: a new peptide coupling reagent devoid of toxic by-product. Tetrahedron Lett 31(2):205–208
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(52):10149–10153
106. Cardona V, Eberle I, Barthélémy S, Beythien J, Doerner B, Schneeberger P, Keyte J, White PD (2008) Application of Dmb-dipeptides in the Fmoc SPPS of difficult and aspartimide-prone sequences. Int J Pept Res Ther 14(4):285–292
107. Wu JC, Carr SF, Jarnagin K, Kirsher S, Barnett J, Chow J, Chan HW, Chen MS, Medzihradszky D, Yamashiro D, Santit DV (1990) Synthetic HIV-2 protease cleaves the GAG precursor of HIV-1 with the same specificity as HIV-1 protease. Arch Biochem Biophys 277(2):306–311
108. Hendrix JC, Halverson KJ, Lansbury PT (1992) A convergent synthesis of the amyloid protein of Alzheimer's disease. J Am Chem Soc 114(20):7930–7931

109. Raibaut L, Ollivier N, Melnyk O (2012) Sequential native peptide ligation strategies for total chemical protein synthesis. Chem Soc Rev 41(21):7001–7015
110. Bode JW, Fox RM, Baucom KD (2006) Chemoselective amide ligations by decarboxylative condensations of *N*-alkylhydroxylamines and α-ketoacids. Angew Chem Int Ed 45 (8):1248–1252
111. Ollivier N, Dheur J, Mhidia R, Blanpain A, Melnyk O (2010) *Bis*(2-sulfanylethyl)amino native peptide ligation. Org Lett 12(22):5238–5241
112. Saxon E, Armstrong JI, Bertozzi CR (2000) A "traceless" Staudinger ligation for the chemoselective synthesis of amide bonds. Org Lett 2(14):2141–2143
113. Nilsson BL, Kiessling LL, Raines RT (2000) Staudinger ligation: a peptide from a thioester and azide. Org Lett 2(13):1939–1941
114. Pattabiraman VR, Bode JW (2011) Rethinking amide bond synthesis. Nature 480 (7378):471–479
115. Hackenberger CP, Schwarzer D (2008) Chemoselective ligation and modification strategies for peptides and proteins. Angew Chem Int Ed 47(52):10030–10074
116. Brenner M, Zimmermann JP, Wehrmüller J, Quitt P, Photaki I (1955) Eine neue Umlagerungsreaktion und ein neues Princip zum Aufbau von Peptidketten. Experientia 11 (10):397–399
117. Wieland T, Bokelmann E, Bauer L, Lang HU, Lau H (1953) Uber Peptidsynthesen. 8. Mitteilung. Liebigs Ann Chem 583:129–148
118. Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. Proc Natl Acad Sci U S A 95(12):6705–6710
119. Flavell RR, Muir TW (2009) Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. Acc Chem Res 42(1):107–116
120. Evans TC, Benner J, Xu M-Q (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. Protein Sci 7(11):2256–2264
121. Wintermann F, Engelbrecht S (2013) Reconstitution of the catalytic core of F-ATPase $(\alpha\beta)_3\gamma$ from *Escherichia coli* using chemically synthesized subunit γ. Angew Chem Int Ed 52 (4):1309–1313
122. Bang D, Pentelute BL, Kent SB (2006) Kinetically controlled ligation for the convergent chemical synthesis of proteins. Angew Chem Int Ed 45(24):3985–3988
123. Fang G-M, Wang J-X, Liu L (2012) Convergent chemical synthesis of proteins by ligation of peptide hydrazides. Angew Chem Int Ed 51(41):10347–10350
124. Li J, Dong S, Townsend SD, Dean T, Gardella TJ, Danishefsky SJ (2012) Chemistry as an expanding resource in protein science: fully synthetic and fully active human parathyroid hormone-related protein (1–141). Angew Chem Int Ed 51(49):12263–12267
125. Bang D, Kent SB (2004) A one-pot total synthesis of crambin. Angew Chem Int Ed 43 (19):2534–2538
126. Ollivier N, Vicogne J, Vallin A, Drobecq H, Desmet R, El-Mahdi O, Leclercq B, Goormachtigh G, Fafeur V, Melnyk O (2012) A one-pot three-segment ligation strategy for protein chemical synthesis. Angew Chem Int Ed 51(1):209–213
127. Li J, Li Y, He Q, Li Y, Li H, Liu L (2014) One-pot native chemical ligation of peptide hydrazides enables total synthesis of modified histones. Org Biomol Chem 12 (29):5435–5441
128. White FH Jr (1960) Regeneration of enzymatic activity by airoxidation of reduced ribonuclease with observations on thiolation during reduction with thioglycolate. J Biol Chem 235:383–389
129. Anfinsen CB, Haber E (1961) Studies on the reduction and re-formation of protein disulfide bonds. J Biol Chem 236:1361–1363
130. Anfinsen CB, Haber E, Sela M, White FH Jr (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A 47:1309–1314

131. Haber E, Anfinsen CB (1961) Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. J Biol Chem 236:422–424
132. White FH Jr (1961) Regeneration of native secondary and tertiary structures by air oxidation of reduced ribonuclease. J Biol Chem 236:1353–1360
133. Carpino LA, Han GY (1970) 9-Fluorenylmethoxycarbonyl function, a new base-sensitive amino-protecting group. J Am Chem Soc 92(19):5748–5749
134. Carpino LA, Han GY (1972) 9-Fluorenylmethoxycarbonyl amino-protecting group. J Org Chem 37(22):3404–3409
135. Carpino L, Han G (1973) Correction. The 9-fluorenylmethoxycarbonyl amino-protecting group. J Org Chem 38(24):4218
136. Meldal M (1992) Pega: a flow stable polyethylene glycol dimethyl acrylamide copolymer for solid phase synthesis. Tetrahedron Lett 33(21):3077–3080
137. Quarrell R, Claridge TW, Weaver G, Lowe G (1996) Structure and properties of TentaGel resin beads: implications for combinatorial library chemistry. Mol Div 1(4):223–232
138. Krieger DE, Erickson BW, Merrifield RB (1976) Affinity purification of synthetic peptides. Proc Natl Acad Sci U S A 73(9):3160–3164
139. Lindeberg G, Tengborn J, Bennich H, Ragnarsson U (1978) Purification of a synthetic peptide with the aid of covalent chromatography. J Chromatogr A 156(2):366–369
140. Villain M, Vizzavona J, Rose K (2001) Covalent capture: a new tool for the purification of synthetic and recombinant polypeptides. Chem Biol 8(7):673–679
141. Liu CF, Tam JP (1994) Peptide segment ligation strategy without use of protecting groups. Proc Natl Acad Sci U S A 91(14):6584–6588
142. Vizzavona J, Villain M, Rose K (2002) Covalent capture purification of polypeptides after SPPS via a linker removable under very mild conditions. Tetrahedron Lett 43(48):8693–8696
143. Carpino LA (1993) 1-Hydroxy-7-azabenzotriazole. An efficient peptide coupling additive. J Am Chem Soc 115(10):4397–4398
144. Carpino LA, El-Faham A, Albericio F (1995) Efficiency in peptide coupling: 1-hydroxy-7-azabenzotriazole vs 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine. J Org Chem 60 (11):3561–3564
145. Carpino LA, Imazumi H, Foxman BM, Vela MJ, Henklein P, El-Faham A, Klose J, Bienert M (2000) Comparison of the effects of 5- and 6-HOAt on model peptide coupling reactions relative to the cases for the 4- and 7-isomers. Org Lett 2(15):2253–2256
146. Nicolet BH, Shinn LA (1939) The action of periodic acid on α-amino alcohols. J Am Chem Soc 61(6):1615–1615
147. Barlow CB, Guthrie RD, Prior AM (1966) Periodate oxidation of amino-sugars. J Chem Soc Chem Commun 9:268–269
148. Cantley M, Hough L (1963) Malonaldehyde derivatives as intermediates in the periodate oxidations of amino- and acetamido-sugars. J Chem Soc 2711–2716
149. Clamp JR, Hough L (1965) The periodate oxidation of amino acids with reference to studies on glycoproteins. Biochem J 94:17–24
150. Dixon HB, Weitkamp LR (1962) Conversion of the N-terminal serine residue of corticotrophin into glycine. Biochem J 84:462–468
151. Funakoshi S, Fukuda H, Fujii N (1991) Chemoselective one-step purification method for peptides synthesized by the solid-phase technique. Proc Natl Acad Sci U S A 88 (16):6981–6985
152. Canne LE, Winston RL, Kent SBH (1997) Synthesis of a versatile purification handle for use with Boc chemistry solid phase peptide synthesis. Tetrahedron Lett 38(19):3361–3364
153. Aucagne V, Valverde IE, Marceau P, Galibert M, Dendane N, Delmas AF (2012) Towards the simplification of protein synthesis: iterative solid-supported ligations with concomitant purifications. Angew Chem Int Ed 51(45):11320–11324
154. 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

155. Tornoe CW, Christensen C, Meldal M (2002) Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. J Org Chem 67(9):3057–3064
156. Rostovtsev VV, Green LG, Fokin VV, Sharpless KB (2002) A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. Angew Chem Int Ed 41(14):2596–2599
157. Hara T, Tainosho A, Nakamura K, Sato T, Kawakami T, Aimoto S (2009) Peptide purification by affinity chromatography based on alpha-ketoacyl group chemistry. J Pept Sci 15 (5):369–376
158. Dixon HB (1964) Transamination of peptides. Biochem J 92(3):661–666
159. Dixon HB, Moret V (1965) Removal of the N-terminal residue of a protein after transamination. Biochem J 94:463–469
160. Van Heyningen S, Dixon HB (1967) Scission of the N-terminal residue from a protein after transamination. Biochem J 104(3):63P
161. Van Heyningen S, Dixon HB (1969) Catalysis of transamination by acetate. Biochem J 114 (4):70P–71P
162. Dixon HBF, Fields R (1972) Specific modification of NH_2-terminal residues by transamination. Methods Enzymol 25:409–419
163. Dixon HB, Moret V (1964) Removal of the N-terminal residue of corticotrophin. Biochem J 93(3):25C–26C
164. Stevens J, Dixon HB (1995) The removal of 2-oxoacyl residues from the N-terminus of peptides and cystatin in non-denaturing conditions. Biochim Biophys Acta 1252(2):195–202
165. Mix H, Wilcke FW (1960) Nonenzymatic reactions between α-amino- and α-keto acids. II. Transformations between α-amino acids and pyruvates catalyzed by copper(II) ions and pyridine. Hoppe-Seylers Z Physiol Chem 318:148–158
166. Sucholeiki I, Lansbury PT (1993) An affinity chromatographic method for the purification of water-insoluble peptides. J Org Chem 58(6):1318–1324
167. Zhang M, Pokharel D, Fang S (2014) Purification of synthetic peptides using a catching full-length sequence by polymerization approach. Org Lett 16(5):1290–1293
168. El-Mahdi O, Melnyk O (2013) Alpha-oxo aldehyde or glyoxylyl group chemistry in peptide bioconjugation. Bioconjug Chem 24(5):735–765
169. Davies M, Bradley M (1997) C-terminally modified peptides and peptide libraries—another end to peptide synthesis. Angew Chem Int Ed 36(10):1097–1099
170. Davies M, Bradley M (1999) Internal resin capture-A self purification method for the synthesis of C-terminally modified peptides. Tetrahedron 55(15):4733–4746
171. Mende F, Seitz O (2007) Solid-phase synthesis of peptide thioesters with self-purification. Angew Chem Int Ed 46(24):4577–4580
172. Mende F, Beisswenger M, Seitz O (2010) Automated Fmoc-based solid-phase synthesis of peptide thioesters with self-purification effect and application in the construction of immobilized SH3 domains. J Am Chem Soc 132(32):11110–11118
173. Brown AR, Irving SL, Ramage R, Raphy G (1995) (17-Tetrabenzo[a,c,g,i]fluorenyl) methylchloroformate (TbfmocCl) a reagent for the rapid and efficient purification of synthetic peptides and proteins. Tetrahedron 51(43):11815–11830
174. Ball HL, Bertolini G, Levi S, Mascagni P (1994) Purification of synthetic peptides with the aid of reversible chromatographic probes. J Chrom A 686(1):73–83
175. Ball HL, Bertolini G, Mascagni P (1995) Affinity purification of 101 residue rat cpn10 using a reversible biotinylated probe. J Pept Sci 1(5):288–294
176. Kellam B, Chan WC, Chhabra SR, Bycroft BW (1997) Transient affinity tags based on the Dde protection/deprotection strategy: synthesis and application of 2-biotinyl-and 2-hexanoyldimedone. Tetrahedron Lett 38(30):5391–5394
177. Frank H-G, Casaretto M, Knorr K (2006) Method for solid-phase peptide synthesis and purification. WO 2006/056443, 1 June 2006

178. Porath J, Carlsson J, Olsson I, Belfrage G (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. Nature 258(5536):598–599
179. Block H, Maertens B, Spriestersbach A, Brinker N, Kubicek J, Fabis R, Labahn J, Schafer F (2009) Immobilized-metal affinity chromatography (IMAC): a review. Methods Enzymol 463:439–473
180. Funakoshi S, Fukuda H, Fujii N (1993) Affinity purification method using a reversible biotinylating reagent for peptides synthesized by the solid-phase technique. J Chromatogr 638:21–27
181. Olejnik J, Sonar S, Krzymanska-Olejnik E, Rothschild KJ (1995) Photocleavable biotin derivatives: a versatile approach for the isolation of biomolecules. Proc Natl Acad Sci U S A 92(16):7590–7594
182. Roggero MA, Servis C, Corradin G (1997) A simple and rapid procedure for the purification of synthetic polypeptides by a combination of affinity chromatography and methionine chemistry. FEBS Lett 408(3):285–288
183. Shogren-Knaak MA, Imperiali B (1998) A reversible affinity tag for the purification of *N*-glycolyl capped peptides. Tetrahedron Lett 39(45):8241–8244
184. Zhang L, Tam JP (1996) Thiazolidine formation as a general and site-specific conjugation method for synthetic peptides and proteins. Anal Biochem 233(1):87–93
185. Melnyk O, Fehrentz JA, Martinez J, Gras-Masse H (2000) Functionalization of peptides and proteins by aldehyde or keto groups. Biopolymers 55(2):165–186
186. Gardlik S, Rajagopalan KV (1991) The mechanisms of inactivation of sulfite oxidase by periodate and arsenite. J Biol Chem 266(25):16627–16632
187. Geoghegan KF, Dallas JL, Feeney RE (1980) Periodate inactivation of ovotransferrin and human serum transferrin. J Biol Chem 255(23):11429–11434
188. Penner MH, Yamasaki RB, Osuga DT, Babin DR, Meares CF, Feeney RE (1983) Comparative oxidations of tyrosines and methionines in transferrins: human serum transferrin, human lactotransferrin, and chicken ovotransferrin. Arch Biochem Biophys 225(2):740–747
189. Geoghegan KF, Stroh JG (1992) Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine. Bioconjug Chem 3(2):138–146
190. Tesser GI, Balvert-Geers IC (1975) The methylsulfonylethyloxycarbonyl group, a new and versatile amino protective function. Int J Pept Protein Res 7(4):295–305
191. Canne L, Kent SBH, Simon R (1998) Solid phase native chemical ligation of unprotected or N-terminal cysteine protected peptides in aqueous solution US 6326468, Dec 4, 2001
192. Rinnova M, Lebl M (1996) Molecular diversity and libraries of structures: synthesis and screening. Collect Czechoslov Chem Commun 61:171–231
193. Lam KS, Lebl M, Krchnak V (1997) The "one-bead-one-compound" combinatorial library method. Chem Rev 97(2):411–448
194. Lam KS, Liu R, Miyamoto S, Lehman AL, Tuscano JM (2003) Applications of one-bead one-compound combinatorial libraries and chemical microarrays in signal transduction research. Acc Chem Res 36(6):370–377
195. Lebl M, Krchnak V, Sepetov NF, Nikolaev V, Stierandova A, Safar P, Seligmann B, Strop P, Thorpe D, Felder S, Lake DF, Lam KS, Salmon SE (1994) One bead-one structure libraries. In: Epton R (ed) Innovation and perspectives in solid phase synthesis. Mayflower Worldwide Ltd, Birmingham, pp 233–238
196. Lebl M, Krchnak V, Sepetov NF, Seligmann B, Strop P, Felder S, Lam KS (1995) One-bead-one-structure combinatorial libraries. Biopolymers 37(3):177–198
197. Holmes CP, Rybak CM (1994) Peptide reversal on solid supports: a technique for the generation of C-terminal exposed peptide libraries. In: Hodges RS, Smith JA (eds) Peptides: chemistry, structure and biology. ESCOM, Leiden, pp 992–994
198. Kania RS, Zuckermann RN, Marlowe CK (1994) Free C-terminal resin-bound peptides: reversal of peptide orientation via a cyclization/cleavage protocol. J Am Chem Soc 116 (19):8835–8836

199. Mende F, Seitz O (2011) 9-Fluorenylmethoxycarbonyl-based solid-phase synthesis of peptide α-thioesters. Angew Chem Int Ed 50(6):1232–1240
200. Kenner GW, McDermott JR, Sheppard RC (1971) The safety catch principle in solid phase peptide synthesis. J Chem Soc D Chem Commun 12:636–637
201. Heidler P, Link A (2005) *N*-Acyl-*N*-alkyl-sulfonamide anchors derived from Kenner's safety-catch linker: powerful tools in bioorganic and medicinal chemistry. Bioorg Med Chem 13(3):585–599
202. Backes BJ, Ellman JA (1999) An alkanesulfonamide "safety-catch" linker for solid-phase synthesis. J Org Chem 64(7):2322–2330
203. Ingenito R, Bianchi E, Fattori D, Pessi A (1999) Solid phase synthesis of peptide C-terminal thioesters by Fmoc/*t*-Bu chemistry. J Am Chem Soc 121(49):11369–11374
204. Shin Y, Winans KA, Backes BJ, Kent SBH, Ellman JA, Bertozzi CR (1999) Fmoc-based synthesis of peptide-α thioesters: application to the total chemical synthesis of a glycoprotein by native chemical ligation. J Am Chem Soc 121(50):11684–11689
205. Ollivier N, Behr JB, El-Mahdi O, Blanpain A, Melnyk O (2005) Fmoc solid-phase synthesis of peptide thioesters using an intramolecular N, S-acyl shift. Org Lett 7(13):2647–2650
206. Quaderer R, Hilvert D (2001) Improved synthesis of C-terminal peptide thioesters on "safety-catch" resins using LiBr/THF. Org Lett 3(20):3181–3184
207. Bycroft BW, Chan WC, Chhabra SR, Hone ND (1993) A novel lysine-protecting procedure for continuous flow solid phase synthesis of branched peptides. J Chem Soc Chem Commun 9:778–779
208. Augustyns K, Kraas W, Jung G (1998) Investigation on the stability of the Dde protecting group used in peptide synthesis: migration to an unprotected lysine. J Pept Res 51(2):127–133
209. Rajasekharan Pillai VN (1980) Photoremovable protecting groups in organic synthesis. Synthesis 1980(01):1–26
210. Klan P, Solomek T, Bochet CG, Blanc A, Givens R, Rubina M, Popik V, Kostikov A, Wirz J (2013) Photoremovable protecting groups in chemistry and biology: reaction mechanisms and efficacy. Chem Rev 113(1):119–191
211. Brik A, Keinan E, Dawson PE (2000) Protein synthesis by solid-phase chemical ligation using a safety catch linker. J Org Chem 65(12):3829–3835
212. Canne LE, Botti P, Simon RJ, Chen Y, Dennis EA, Kent SBH (1999) Chemical protein synthesis by solid phase ligation of unprotected peptide segments. J Am Chem Soc 121 (38):8720–8727
213. Johnson EC, Durek T, Kent SB (2006) Total chemical synthesis, folding, and assay of a small protein on a water-compatible solid support. Angew Chem Int Ed 45(20):3283–3287
214. Stempfer G, Holl-Neugebauer B, Rudolph R (1996) Improved refolding of an immobilized fusion protein. Nat Biotechnol 14(3):329–334
215. Bang D, Kent SB (2005) His6 tag-assisted chemical protein synthesis. Proc Natl Acad Sci U S A 102(14):5014–5019
216. Tan Z, Shang S, Danishefsky SJ (2010) Insights into the finer issues of native chemical ligation: an approach to cascade ligations. Angew Chem Int Ed 49(49):9500–9503
217. Ueda S, Fujita M, Tamamura H, Fujii N, Otaka A (2005) Photolabile protection for one-pot sequential native chemical ligation. Chembiochem 6(11):1983–1986
218. Zheng JS, Cui HK, Fang GM, Xi WX, Liu L (2010) Chemical protein synthesis by kinetically controlled ligation of peptide O-esters. Chembiochem 11(4):511–515
219. Sato K, Shigenaga A, Tsuji K, Tsuda S, Sumikawa Y, Sakamoto K, Otaka A (2011) *N*-Sulfanylethylanilide peptide as a crypto-thioester peptide. Chembiochem 12 (12):1840–1844
220. Okamoto R, Morooka K, Kajihara Y (2012) A synthetic approach to a peptide α-thioester from an unprotected peptide through cleavage and activation of a specific peptide bond by *N*-acetylguanidine. Angew Chem Int Ed 51(1):191–196
221. Jbara M, Seenaiah M, Brik A (2014) Solid phase chemical ligation employing a rink amide linker for the synthesis of histone H2B protein. Chem Commun 50:12534–12537

222. Cotton GJ, Muir TW (2000) Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. Chem Biol 7(4):253–261
223. Nilsson BL, Hondal RJ, Soellner MB, Raines RT (2003) Protein assembly by orthogonal chemical ligation methods. J Am Chem Soc 125(18):5268–5269
224. Becker CF, Hunter CL, Seidel R, Kent SB, Goody RS, Engelhard M (2003) Total chemical synthesis of a functional interacting protein pair: the protooncogene H-Ras and the Ras-binding domain of its effector c-Raf1. Proc Natl Acad Sci U S A 100(9):5075–5080
225. 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 SB, Bradburne JA (2003) Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. Science 299 (5608):884–887
226. Bang D, Chopra N, Kent SB (2004) Total chemical synthesis of crambin. J Am Chem Soc 126 (5):1377–1383
227. Luz JG, Yu M, Su Y, Wu Z, Zhou Z, Sun R, Wilson IA (2005) Crystal structure of viral macrophage inflammatory protein I encoded by Kaposi's sarcoma-associated herpesvirus at 1.7A. J Mol Biol 352(5):1019–1028
228. Sohma Y, Pentelute BL, Whittaker J, Hua Q-X, Whittaker LJ, Ma W, Kent SBH (2008) Comparative properties of insulin-like growth factor 1 (IGF-1) and [Gly7D-Ala]IGF-1 prepared by total chemical synthesis. Angew Chem Int Ed 47:1102–1106
229. Bang D, Tereshko V, Kossiakoff AA, Kent SB (2009) Role of a salt bridge in the model protein crambin explored by chemical protein synthesis: X-ray structure of a unique protein analogue, [V15A]crambin-alpha-carboxamide. Mol Biosyst 5(7):750–756
230. Chiche L, Gaboriaud C, Heitz A, Mornon JP, Castro B, Kollman PA (1989) Use of restrained molecular dynamics in water to determine three-dimensional protein structure: prediction of the three-dimensional structure of Ecballium elaterium trypsin inhibitor II. Proteins 6 (4):405–417
231. Heitz A, Chiche L, Le-Nguyen D, Castro B (1989) 1H 2D NMR and distance geometry study of the folding of *Ecballium elaterium* trypsin inhibitor, a member of the squash inhibitors family. Biochemistry 28(6):2392–2398
232. Patek M, Lebl M (1990) A safety-catch type of amide protecting group. Tetrahedron Lett 31 (36):5209–5212
233. Matsueda GR, Stewart JM (1981) A *p*-methylbenzhydrylamine resin for improved solid-phase synthesis of peptide amides. Peptides 2(1):45–50
234. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389(6648):251–260
235. 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(48):9248–9252
236. Orita M, Yamamoto S, Katayama N, Aoki M, Takayama K, Yamagiwa Y, Seki N, Suzuki H, Kurihara H, Sakashita H, Takeuchi M, Fujita S, Yamada T, Tanaka A (2001) Coumarin and chromen-4-one analogues as tautomerase inhibitors of macrophage migration inhibitory factor: discovery and X-ray crystallography. J Med Chem 44(4):540–547
237. Sasaki K, Aubry S, Crich D (2011) Chemistry with and around thioacids. Phosphorus Sulfur Silicon 186:1005–1018
238. Pira SL, Boll E, Melnyk O (2013) Synthesis of peptide thioacids at neutral pH using *bis* (2-sulfanylethyl)amido peptide precursors. Org Lett 15(20):5346–5349
239. Fecourt F, Delpech B, Melnyk O, Crich D (2013) *Se*-(9-Fluorenylmethyl) selenoesters; preparation, reactivity, and use as convenient synthons for selenoacids. Org Lett 15 (14):3758–3761
240. Wang P, Danishefsky SJ (2010) Promising general solution to the problem of ligating peptides and glycopeptides. J Am Chem Soc 132(47):17045–17051

241. Hou W, Zhang X, Li F, Liu CF (2011) Peptidyl N, N-bis(2-mercaptoethyl)-amides as thioester precursors for native chemical ligation. Org Lett 13:386–389
242. Zheng J-S, Tang S, Huang Y-C, Liu L (2013) Development of new thioester equivalents for protein chemical synthesis. Acc Chem Res 46:2475–2484
243. Melnyk O, Agouridas V (2014) From protein total synthesis to peptide transamidation and metathesis: playing with the reversibility of N, S-acyl or N, Se-acyl migration reactions. Curr Opin Chem Biol 22:137–145
244. Melnyk O, Agouridas V (2014) Perhydro-1,2,5-dithiazepine. e-ROS. doi:10.1002/9780470842898.rn9780470801723
245. Ollivier N, Raibaut L, Blanpain A, Desmet R, Dheur J, Mhidia R, Boll E, Drobecq H, Pira SL, Melnyk O (2014) Tidbits for the synthesis of bis(2-sulfanylethyl)amido (SEA) polystyrene resin, SEA peptides and peptide thioesters. J Pept Sci 20(2):92–97
246. Dheur J, Ollivier N, Vallin A, Melnyk O (2011) Synthesis of peptide alkylthioesters using the intramolecular N, S-acyl shift properties of bis(2-sulfanylethyl)amido peptides. J Org Chem 76(9):3194–3202
247. Dheur J, Ollivier N, Melnyk O (2011) Synthesis of thiazolidine thioester peptides and acceleration of native chemical ligation. Org Lett 13(6):1560–1563
248. Raibaut L, Seeberger P, Melnyk O (2013) Bis(2-sulfanylethyl)amido peptides enable native chemical ligation at proline and minimize deletion side-product formation. Org Lett 15 (21):5516–5519
249. Raibaut L, Vicogne J, Leclercq B, Drobecq H, Desmet R, Melnyk O (2013) Total synthesis of biotinylated N domain of human hepatocyte growth factor. Bioorg Med Chem 21 (12):3486–3494
250. Agard NJ, Prescher JA, Bertozzi CR (2004) A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. J Am Chem Soc 126 (46):15046–15047

Top Curr Chem (2015) 363: 155–192
DOI: 10.1007/128_2014_599
© Springer International Publishing Switzerland 2014
Published online: 24 February 2015

New Methods for Chemical Protein Synthesis

Xiaoyang Guan, Patrick K. Chaffey, Chen Zeng, and Zhongping Tan

Abstract Chemical protein synthesis is a useful tool to generate pure proteins which are otherwise difficult to obtain in sufficient amounts for structure and property analysis. Additionally, because of the precise and flexible nature of chemical synthesis, it allows for controllable variation of protein sequences, which is valuable for understanding the relationships between protein structure and function. Despite the usefulness of chemical protein synthesis, it has not been widely adopted as a tool for protein characterization, mainly because of the lack of general and efficient methods for the preparation and coupling of peptide fragments and for the folding of polypeptide chains. To address these issues, many new methods have recently been developed in the areas of solid-phase peptide synthesis, peptide fragment assembly, and protein folding. Here we review these recent technological advances and highlight the gaps needing to be addressed in future research.

Keywords Fragment condensation · Metal-free desulfurization · Native chemical ligation · Protein-folding · Solid-phase peptide synthesis

Contents

X. Guan, P.K. Chaffey, C. Zeng, and Z. Tan (✉)
Department of Chemistry and Biochemistry, BioFrontiers Institute, University of Colorado, Boulder, CO 80303, USA
e-mail: zhongping.tan@colorado.edu

1 Introduction

Chemical synthesis is a valuable tool for studying the structure and function of proteins, particularly proteins with post-translational modifications [1, 2]. Compared to recombinant DNA-based protein biosynthesis, chemical synthesis is a more flexible and precise method to introduce natural and unnatural amino acid mutations. Chemical coupling is less sensitive to large variations in the size, structure, and sequence of peptide building blocks [3]. Condensation of easily characterized fragments ensures purity of final protein products. Moreover, chemical synthesis enables unparalleled control over the preparation of protein isoforms, allowing systematic but well-defined structural variations. Access to such isoforms greatly facilitates the development of a more comprehensive and deeper understanding of protein structure–function relationships and leads to better protein engineering strategies [4].

Given the great scientific significance of chemical synthesis, it is not surprising that considerable research effort has been devoted to developing more effective chemical methods for preparing proteins. This effort has successfully led to technologies which are widely used today: solid-phase peptide synthesis (SPPS), native chemical ligation (NCL), and metal-free desulfurization (MFD) [5–7]. Together with the development of new folding methods, the application of SPPS, NCL, and MFD has revolutionized the preparation of chemically pure proteins. Current peptide synthesis methods allow for the relatively routine construction of many kinds of proteins, including those without Cys residues and proteins with diverse post-translational modifications.

As shown in Fig. 1, a protein molecule can be prepared by chemically ligating two or more peptide fragments together. SPPS, developed by Merrifield [8, 9], provides a fast and effective way to prepare peptide fragments. NCL [10] and MFD [11], developed by Kent and Danishefsky, are the methods of choice for joining the fragments. NCL is the ligation of an activated peptide fragment, in most cases a peptide thioester (**2**), and a peptide fragment containing an N-terminal thio-substituted amino acid (**3**). The ligation is initiated by a rapid thioester exchange between the N-terminal thiol group and the C-terminal thioester. An $S \rightarrow N$ acyl transfer immediately follows, leading to the formation of a native peptide linkage between fragments **2** and **3**. This process can be repeated many times to complete the synthesis of longer protein chains [10]. After complete assembly of the full-length target protein, global desulfurization under metal-free conditions can remove

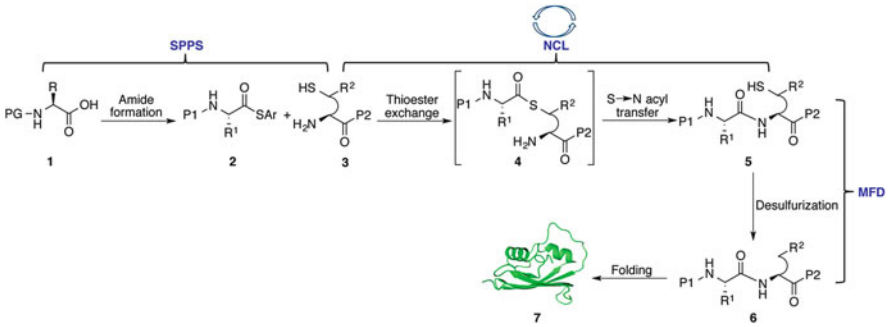


Fig. 1 General procedures for the synthesis of proteins: SPPS, NCL, and MFD. R, R^1, and R^2 can be varied with the amino acid used. *PG* protecting group, *P1* and *P2* peptides

any thiol auxiliaries used to facilitate the ligation of individual peptide fragments [11]. In the final step, the target protein folds into a well-defined tertiary structure through several methods of disulfide formation [12, 13].

Despite many recent advances in chemical protein synthesis, there are still practical issues to be addressed before chemical synthesis can be generally applied by protein biochemists. The main problems limiting the wider use of chemical protein synthesis include synthesis of long peptides [14], synthesis and handling of hydrophobic peptides [15], side reactions in peptide synthesis [16, 17], preparation of peptide thioesters [18], ligation at new amino acid sites [19], and in vitro folding of chemically synthesized proteins [12]. These problems, separately or together, render the chemical synthesis of many proteins a challenge which can only be tackled by specialists.

To overcome the limitations of the existing methods, many studies have been performed with the aim of achieving facile synthesis of proteins. Early efforts in this direction have already been extensively reviewed [20–27]. To avoid redundancy, the present review focuses on recent developments in this field, mostly since 2007. To enable readers to locate relevant information quickly, this review is organized according to the general procedure for chemical protein synthesis (Fig. 1). As listed in the table of contents, the three major topics covered here are (1) peptide fragment synthesis, (2) assembly of peptide fragments, and (3) protein folding.

2 Synthesis of Peptide Fragments

Linking individual amino acids is necessarily the first step in the synthesis of proteins. Since Merrifield invented the technology in 1963, SPPS has become the primary approach for the routine preparation of peptides [28–32]. The general scheme of solid phase peptide synthesis is illustrated in Fig. 2. The peptide is built from the C-terminus to the N-terminus. The first amino acid is attached to a

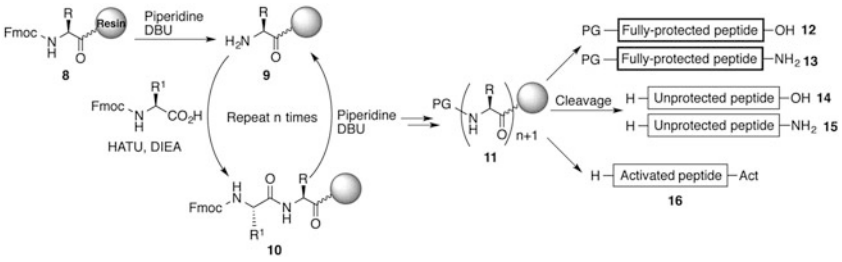


Fig. 2 General scheme of Fmoc-based solid phase peptide synthesis. *Act* activating group

polymer support (resin) via a linker moiety. In addition to acting as the protecting group for the C-terminal carboxyl group, the insoluble solid resin permits a rapid and clean separation of the growing peptide product from complex reaction mixtures during synthesis.

The elongation of the peptide chain begins with the removal of the N^{α}-protecting group of the first amino acid. In SPPS, two protecting groups – *tert*-butoxycarbonyl (Boc) and 9-fluorenylmethoxycarbonyl (Fmoc) – are commonly used for N^{α}-protection. In the Boc protection approach, trifluoroacetic acid (TFA) is typically used for N^{α}-deprotection and highly corrosive and toxic hydrofluoric acid (HF) is used for the final global deprotection and release of the assembled peptide from the resin [33]. For the Fmoc protection approach, the repetitive deprotection is performed in a piperidine solution or a mixture of piperidine and 1,8-diazabicyclo [5.4.0]undec-7-ene (DBU) and the final deprotection and cleavage of the peptide chain is performed in TFA and appropriate scavengers [34]. Because of the milder conditions and less toxic reagents used in the Fmoc approach, it is often preferred over the Boc approach. Complete N^{α}-Fmoc removal can be achieved by treating the peptide-resin with piperidine/DBU for 30 min.

After N^{α}-deprotection, subsequent amino acids are coupled to the free N-terminus of the growing peptide chain (Fig. 2). Each coupling reaction is driven to completion in two ways – using large excesses of amino acid and activation reagents and employing a long reaction time, typically 1 h. When coupling is complete, a thorough washing of the resin-bound peptide removes excess coupling reagents and uncoupled amino acids. N^{α}-deprotection is then repeated and, following another thorough washing step, the peptide is subjected to the next coupling cycle. These steps are repeated until the peptide chain is complete, at which point the desired peptide is cleaved from the resin support. Fully protected, unprotected, and/or activated forms of peptides can be obtained through proper use of different cleavage cocktails.

Since the establishment of the original procedures for solid phase synthesis of oligopeptides, all aspects of SPPS, including resins, linkers, protecting groups, coupling reagents, and synthesis protocols, have been optimized and improved, making the preparation of large quantities of peptides of up to 30 amino acids feasible and reliable [35]. Nevertheless, peptide synthesis is not without challenges. Aggregation, incomplete coupling, racemization, aspartimide formation, and

methionine oxidation are all common complications which may arise during SPPS and pose important product quality concerns [36]. Additionally, synthesis of activated peptide fragments using the Fmoc approach presents a significant challenge because of the nucleophilic attack by piperidine at the activated acyl group [18]. To develop more efficient and reliable procedures for preparing structurally diverse proteins, numerous new methods have been introduced. In the following paragraphs, we discuss a few such methods including those for forming amide bonds, controlling side reactions, and activating C-terminal carboxyl groups.

2.1 New Methods for Peptide Synthesis

2.1.1 Amide Bond Formation

The major chemical reaction in peptide synthesis is amide bond formation. Currently, the most prevalent methods for forming amide bonds were developed based on the general framework of directly coupling carboxylic acids with amines in the presence of coupling reagents, bases, and additives (Fig. 3). The first step in these methods involves the activation of the carboxylic acid. In the second step, nucleophilic displacement of the activated carboxylic acid derivative by the free amine of an amino acid generates the desired amide bond. Coupling reagent mediated methods are highly effective and widely regarded as the methods of choice for SPPS. However, as with all other methods in organic chemistry, these methods have inherent limitations. Racemization of amino acid residues and expensive or wasteful procedures are common problems associated with the current coupling reagent-mediated methods [37]. In the search for solutions to these issues, many new methods have recently been published.

Several newly developed methods are based on the use of catalysts, both organo- and metal catalysts [38, 39]. Catalysis has the potential to provide an atom-economical and cost-effective route to amide bond formation and has already shown considerable promise. However, most of these methods are still in their infancy, and while they may provide useful solutions for existing problems, a

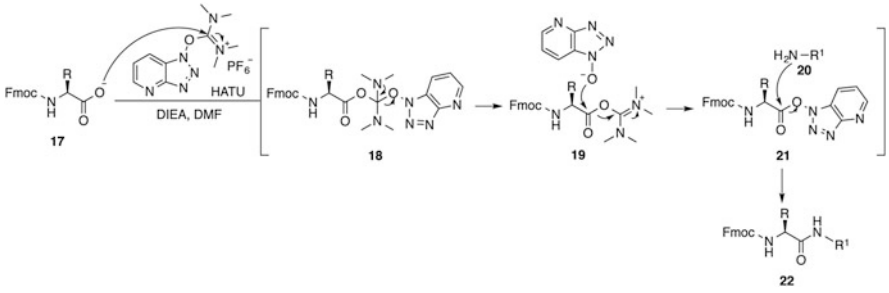


Fig. 3 Mechanism of amide bond formation using HATU as coupling reagent

number of technological challenges need to be overcome before catalysis can be generally applied in the solid phase synthesis of peptides.

Other groups have focused on the chemistry of functional groups not previously used in peptide coupling reactions in order to improve amide bond formation in SPPS. One of these new methods which has been applied in the context of SPPS is the isonitrile-mediated amidation developed by Danishefsky and co-workers, which highlights the usefulness of isonitriles in the synthesis of peptidyl and glycopeptidyl amide bonds [40]. Early experiments during the development of this approach used free carboxylic acids **23** as the acylating agents [41–44]. As shown in Fig. 4a, this microwave-induced reaction is a two-component coupling reaction between carboxylic acids and isonitriles. According to the proposed mechanism, the combination of an isonitrile **24** and a carboxylic acid **23** at elevated temperature leads to the formation of a high-energy formimidate carboxylate mixed anhydride **26** (FCMA), possibly through an Alder-ene reaction, or, alternatively, through a step-wise pathway of protonation and nucleophilic addition. The FCMA intermediate subsequently undergoes a 1,3-*O*-to-*N* acyl transfer to give rise to an N-formyl amide product **27**. If necessary, the N-formyl group can be selectively converted to a variety of functionalities, including the natural N-H, but also N-homoallyl or N-Me. It was also observed that, in the presence of external nucleophilic trapping agents (R^2XH), the putative FCMA intermediate could be intercepted by the nucleophile to give a mixture of rearrangement **27** and acylation **28** products. Because no external acids or bases are used in this new isonitrile-based coupling method, racemization issues could be well addressed.

Danishefsky and co-workers then investigated the possibility of replacing carboxylic acids with thioacids, which were found to allow for much milder conditions [45–47]. Thioacids readily react with isonitrile substrates at room temperature to generate dipeptide products **32** in good yields. The reaction is proposed to proceed through a similar mechanism to that of carboxylic acids (Fig. 4b). Again, in the presence of a nucleophile, thio FCMA intermediates can undergo the bimolecular acylation reaction to produce the acylation adduct **33**.

Importantly, Danishefsky has successfully translated the isonitrile method from proof-of-concept experiment to successful synthesis. In addition to a solution-phase

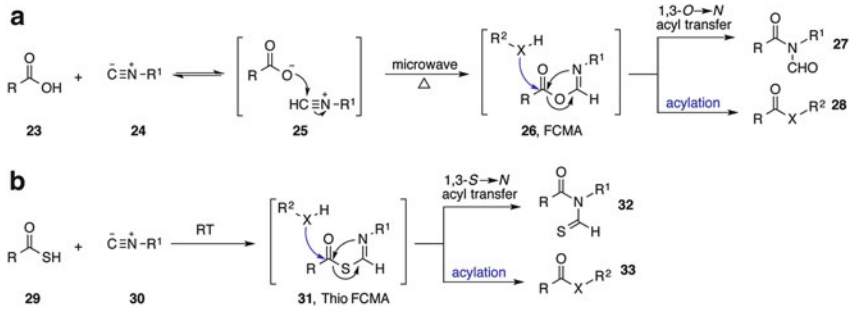


Fig. 4 Isonitrile-mediated amidation. $R^2XH = ROH$ or RNH_2

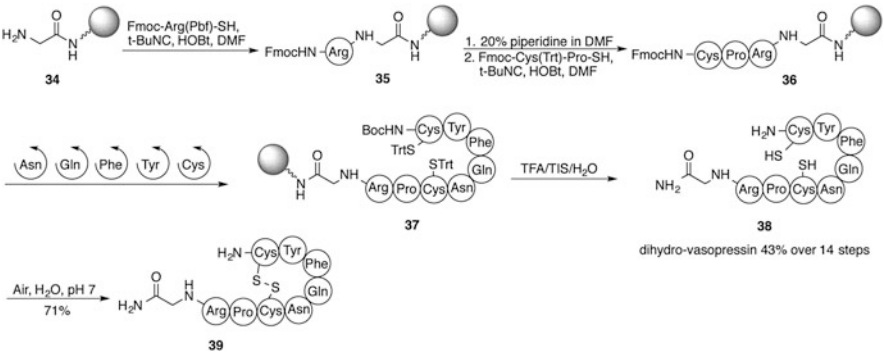
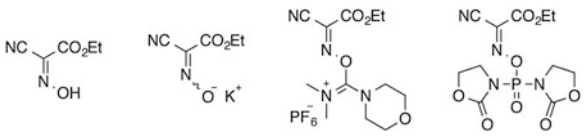


Fig. 5 Application of isonitrile-mediated thioacid amidation in the solid phase synthesis of vasopressin

synthesis of a highly N-methylated cyclic peptide, cyclosporine A [48], the method has also been applied to the solid-phase synthesis of vasopressin **39**, a cyclic peptide hormone involved in conservation of water by the kidney [49]. As shown in Fig. 5, the synthesis of vasopressin was achieved through a sequence of isonitrile-mediated thioacid amidation steps on solid support. It began with Arg-derived thioacid and solid support linked Gly **34**. The resulting dipeptide was deprotected and reacted with a Cys-Pro-derived thioacid to provide the resin-bound tetrapeptide **36**. Following this, Asn-, Gln-, Phe-, Tyr-, and Cys-derived thioacids were added through the same two-step sequence of Fmoc deprotection/isonitrile-mediated acylation to afford the peptide backbone **37** of vasopressin. Cleavage of the nonapeptide from the solid support and removal of all the protecting groups provides the linear vasopressin **38** in 43% overall yield. Finally, formation of the intramolecular disulfide bond via air oxidation gave vasopressin **39** in 71% isolated yield.

Of course, rather than entirely new chemistry, improvement of amide bond synthesis could be achieved through the optimization of existing methods. As described above, the classic methods rely on the use of coupling reagents, bases, and/or additives to promote amide bond formation (Fig. 3), and a potential solution to the limits of these classical methods could be to improve those coupling reagents, bases, and/or additives. With this approach in mind, Albericio and co-workers introduced several derivatives of ethyl 2-cyano-2-(hydroxyimino)acetate (Oxyma, **40**) for use in amide bond formation [50]; see Fig. 6.

Oxyma **40**, first reported in the early 1970s, is an oxime with a similar acidity (pKa 4.60) to those of the most widely used additives, HOBt (pKa 4.60) and HOAt (pKa 3.28) [51]. By systematically testing these three additives, Albericio et al. found that Oxyma is more effective than either HOBt or HOAt at both suppressing racemization during peptide coupling and coupling sterically hindered amino acids [50]. They also showed that Oxyma decomposed at a slower rate than HOBt and HOAt, thus having a lower risk of explosion. These properties make Oxyma a practical replacement for HOBt and HOAt. In order to suppress side reactions further, they also prepared a new formulation, *K*-Oxyma **41**, a potassium



40, Oxyma **41**, K-Oxyma **42**, COMU **43**, Bop-Oxy

Fig. 6 Derivatives of Oxyma for amide bond formation

salt of Oxyma. In addition to preserving most of the beneficial properties of Oxyma, the new *K*-Oxyma formulation was found to inhibit premature cleavage of peptides from acid labile resins, such as 2-chlorotrityl resin [52].

That same year, Albericio and co-workers went on to develop a safer and more effective coupling reagent, a uronium salt derived from Oxyma, COMU **42** [53]. Because of the excellent leaving group ability of the Oxyma moiety [50], the great solubility of the morpholino-containing iminium moiety [54], and the high reactivity of the formed uronium salt [55], COMU shows superior performance in the synthesis of many different types of amide bonds in comparison with HATU and HBTU [53]. Unfortunately, COMU degrades more rapidly than HATU or HBTU in DMF, although there has been discussion that this may not be because of the inherent properties of COMU and can easily be overcome by using purer COMU [56]. They also developed another derivative by combining Oxyma with bis (2-oxo-3-oxazolidinyl) phosphorodiamidic chloride (BOP-Cl). Through their tests, they found that the use of BOP-Oxy **43** is advantageous in most cases because of the higher coupling efficiency and more efficient suppression of racemization [57].

Precise heating has also emerged as a powerful method for the synthesis of amide bonds. Heating during coupling substantially accelerates the rates of amide formation [29]. Unfortunately, any heat used can also accelerate competing side reactions, including racemization. By determining the ratio of D-amino acids incorporated into a model peptide, VYWTSPFMKLIHEQCNRADG-NH_2, under different microwave conditions, Collins and co-workers suggested that Cys, His, and Asp were more susceptible to racemization at 80°C than other amino acids [58]. Kappe and co-workers also reported similar racemization results based on the study of peptides prepared under microwave and conventionally heated conditions [59]. Lowering the temperature for the coupling of these three problematic amino acids is currently the recommended solution to minimize their racemization [29].

2.1.2 Methods for the Synthesis of Long/Difficult Peptides

The chemical synthesis of full-length proteins usually requires the joining together of several peptide fragments (Fig. 1). The longer the fragments, the fewer fragment-joining steps required. Since the isolated yield of each fragment-joining step is usually quite low (<50%), fewer synthetic steps generally significantly improve the overall yield. Fewer synthetic steps almost always result in less time being spent on the synthesis as well. For these reasons, it is highly desirable to use the longest peptide fragments possible in the condensation reactions of a protein chemical

synthesis. Currently, the synthesis of peptides longer than 50 amino acids is inefficient, giving many side products which are, in most cases, difficult to separate from the desired peptide [60]. Such limitations are most often caused either by aggregation or by various side reactions which occur under the coupling and/or deprotection reaction conditions. The most problematic side reactions are premature cleavage of peptides from the solid support, aspartimide formation, and Met oxidation [36].

Growing peptide chains can form intra-/intermolecular hydrophobic interactions and/or hydrogen bonds with themselves or neighboring peptides. The cumulative effect of many of these interactions is peptide aggregation [61]. Long peptides, or peptides containing clusters of hydrophobic residues, generally have a higher tendency to aggregate [62]. Aggregation during synthesis can lead to the blocking of the reacting N-terminus, which can cause incomplete coupling and/or deprotection. By-products resulting from such incomplete reactions tend to have chromatographic properties very similar to those of the desired products, making the isolation and purification of the target peptide a difficult task. Additionally, many peptide fragments en route to full-length proteins are especially prone to aggregation, even if the full-length construct is not, thereby further complicating the synthesis of proteins [63]. To overcome the aggregation issue, many methods have been developed, including new resins, solvents, building blocks, coupling reagents or times, and heating the reaction mixture. When synthesizing particularly difficult sequences, more than one of these methods can be combined for more effective aggregation suppression [64, 65].

Aggregation is less pronounced on a few solid supports [66]. The newly developed ChemMatrix® (CM) resin **44**, a 100% PEG (polyethylene glycol)-based solid support, is one of them (Fig. 7a) [67]. As compared to previously popular

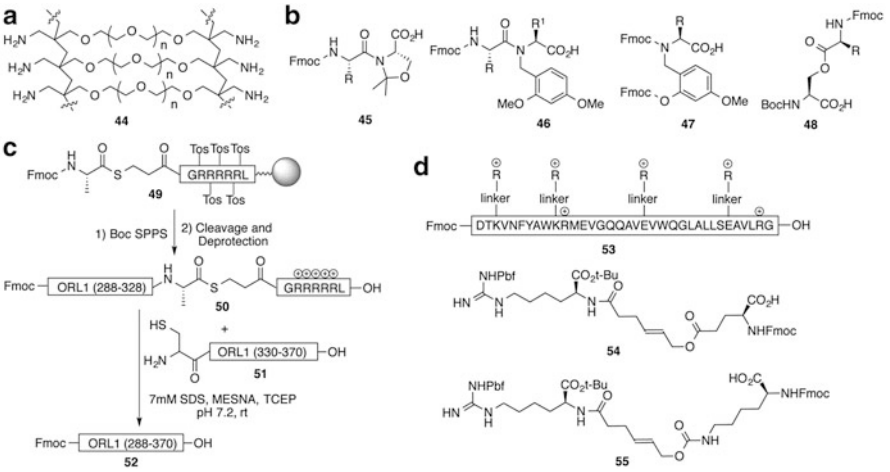


Fig. 7 Methods to overcome the aggregation problem in SPPS: (**a**) ChemMatrix® (CM) resin; (**b**) dipeptide analogs; (**c**) thioester Arg_5 tag; (**d**) cleavable Arg tag

polystyrene-based resins, the CM resin is more polar and contains both hydrogen bond donor and acceptor functionality. These chemical properties mean CM resins swell more effectively in polar solvents and can form aggregation-disrupting intramolecular interactions with the growing peptides. This makes the CM resin a valuable resin for preventing peptide chains from forming the ordered secondary structures which lead to severe aggregation [68]. Another simple and effective way to suppress aggregation is to use rigid amino acid building blocks which disfavor highly ordered or aggregation prone structures during peptide synthesis. Such building blocks are commonly used in difficult-to-synthesize peptides and include pseudoproline dipeptides **45**, Dmb (2,4-dimethoxybenzyl) dipeptides **46**, Dmb/Hmb (2-hydroxy-4-methoxy-benzyl) amino acids **47**, and isoacyl dipeptides **48** (Fig. 7b). The combined use of CM resin and rigid building blocks has greatly facilitated the synthesis of long and difficult sequences. Notable syntheses using these approaches include HIV protease, Rantes (1–68), CCL4-L1 (chemokine C–C motif ligand 4-like 1), and β-amyloid (1–42) [69, 70].

Heat has also been utilized to prepare long and difficult peptides successfully, typically through the use of heating baths or microwave energy. In 2014, Collins and co-workers developed a microwave irradiation-based high-efficiency-SPPS method [71]. Critical to the success of the high efficiency microwave method was a novel internal temperature probe which allowed for extremely careful monitoring of the resin temperature and made it possible to achieve temperatures of 90°C in about 20 s without significant temperature overshoot. This enabled high temperatures to be used without long ramp times while simultaneously minimizing the side reactions which can happen during temperature overshoot. Through aggressive use of carefully controlled microwave irradiation and elimination of unnecessary washing steps, the average total cycle time for deprotection, coupling, and all washes was reduced to only 4 min. This is a significant improvement over standard Fmoc SPPS cycle times without microwave irradiation, which are typically in excess of 60 min.

A couple of strategies were also developed to increase the solubility and/or decrease the aggregation propensity of peptides once cleaved from solid supports [63, 72, 73]. The first strategy relies upon the use of a C-terminal $(Arg)_{5/6}$ to increase the solubility of peptide fragments (Fig. 7c), while the second strategy introduced Arg residues via linkers to the side chains of peptides (Fig. 7d). The first strategy was developed by Aimoto and co-workers, building upon earlier work by Deber [74] and Muir [75] on the synthesis of hydrophobic proteins. Aimoto and co-workers incorporated five Arg residues into the thiol moiety of the thioester fragment [72]. This allowed them to synthesize the C-terminal 83 residues of ORL1 (opioid receptor-like 1). With the Arg_5 sequence in place, the thioester fragment, which contains one entire transmembrane domain of the receptor, was easily able to undergo NCL in the presence of 7 mM SDS and MESNA with the C-terminal intercellular domain of ORL1 to give the target peptide. The Arg residues used to enhance solubility were removed along with the thiol moiety as a result of the NCL reaction. In 2011, Danishefsky and co-workers developed a slightly different method to prevent the aggregation of a synthetic fragment, hEPO (43–77)

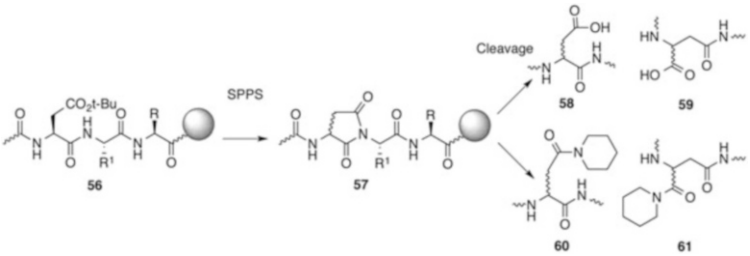


Fig. 8 Aspartimide formation and subsequent by-products

[63]. Rather than attaching the Arg residues to the thioester moiety, they chose to attach the residues to the side chains of the problematic peptide fragment via cleavable linkers. This method has the advantage of keeping the aggregation-suppressing Arg residues attached during long syntheses which require more than one NCL reaction. The Arg residues can be removed after all fragment coupling steps are completed by cleaving the allylic ester in the presence of palladium(0).

Similar to aggregation, aspartimide formation leads to the generation of inseparable impurities during the synthesis of long peptides, and can occur under both acidic and basic conditions [76]. In Fmoc-based SPPS, the initially formed five-membered aspartimide ring **57** can be opened by a variety of nucleophiles, such as water, methanol, and piperidine, to produce numerous by-products **58–61** (Fig. 8). Because of the repetitive use of piperidine and/or DBU during each N^{α}-deprotection step, aspartimide formation becomes a particularly serious problem in the synthesis of long peptides containing multiple Asp residues or C-terminal Asp residues. Many factors, including peptide sequence (Asp-X), conformation, base, acid, β-carboxyl protecting groups, solid support, temperature, and solvent, affect the severity of the aspartimide problem.

Many different methods have been developed to circumvent the aspartimide problem. Among all the methods, the addition of dinitrophenol or HOBt to the deprotection solution and the use of Fmoc-Asp(OMpe)-OH and Fmoc-Asp(OtBu)-(Dmb)Gly-OH during SPPS are employed most frequently because of their effectiveness and convenience [76]. In 2007, Collins and co-workers demonstrated that the use of piperazine for Fmoc removal can significantly lower the level of aspartimide [58]. However, because the basicity of piperazine is lower than that of piperidine (pK_a 9.8 vs 11.1, respectively), the rate of Fmoc-deprotection is slower when using piperazine. This issue can be addressed by elevating the temperature of the deprotection reactions. With the assistance of microwave irradiation, the deprotection can be complete in 3 min. Using microwave-SPPS and piperazine (5% piperazine with 0.1 M HOBt in DMF), the amount of aspartimide by-products observed during the synthesis of H-VYWTSPFMKLIHEQCNRADG-NH_2 was reduced from 32% to 3% [58]. More recently, it was demonstrated that the addition of Oxyma [77] and acids such as formic acid [78] to the Fmoc cleavage solution could also suppress the formation of aspartimide by-products. For example, in the presence of formic acid (5 vol.% to piperidine), aspartimide formation

during the synthesis of the C-terminal peptide fragment of parathyroid hormone (PTH), H-AGSQPRRKKEDNVLVESHEKSLGEADKADVNVLTKAKSQ-NH_2, PTH (46–84), was reduced by around 90% [78].

Met oxidation is another side reaction complicating the purification of long peptide fragments [79] and can happen at almost any time during the SPPS process, the global deprotection step, or the handling and storage steps. It is highly likely that the peptide sequence and conformation strongly influence Met oxidation, but the exact connection is not well understood. Conventional methods to alleviate this problem have involved the use of special cleavage cocktails, for example cocktails composed of TFA/EDT/water/TIS or cocktails containing different Met sulfoxide reducing agents such as *N*-methylmercaptoacetamide [80–82] and NH_4I/DMS [83]. Recently, Danishefsky and co-workers demonstrated that the attachment of cleavable arginine tags to peptide side chains inhibits the oxidation of Met, which may become a new way to advance the synthesis of long/difficult peptides [63].

2.2 New Methods for Activated Peptide Synthesis

Easily synthesized activated peptides are the key to the chemical synthesis of proteins [18, 84]. Many types of activating groups have been used for peptide ligations, including thioesters [85], acyl azides [86–90], acid chlorides [91], acyl isoureas [92], acyl imidazoles [93], and aromatic esters [94]. Of all the possibilities, however, peptide thioesters have become the most widely used (Fig. 1).

Since the discovery of NCL, a great deal of effort has been invested in developing easier and more reliable ways to synthesize peptide thioesters for all types of peptide fragments [84]. At present, peptide thioesters can be readily obtained either by SPPS [95–97] or by direct coupling of fully protected peptides with an amino acid thioester under non-racemization conditions [98–101]. They can also be produced in situ by either $O \rightarrow S$ or $N \rightarrow S$ acyl shift from the corresponding esters [102–105] or amides [106–118]. Each method has its inherent strengths and weaknesses [119], and we discuss those of the newer methods here.

One of the most important methods developed to improve access to peptide thioesters is the *O*-mercaptoaryl ester rearrangement (OMER) process developed by Danishefsky and co-workers [120, 121]. As shown in Fig. 9, an inert ortho-thiophenolic ester is first installed on the C-terminus of the peptide. After the reduction of the disulfide bond, the intermediate undergoes a spontaneous intramolecular $O \rightarrow S$ migration to afford the fully active peptide thioester. This method has been extended to *O*-alkylesters as well by Liu and co-workers [105] and has been widely used in the chemical syntheses of proteins [7]. An 11-residue glycopeptide containing two N-glycans (**72**) was the first peptide synthesized using the OMER methodology [7]. As shown in Fig. 9, 2-(ethylsulfinothioyl)phenol **64** was first synthesized from 2-mercaptophenol **62** via oxidization and subsequent exchange with excess ethyl disulfide. The resulting phenol derivative **64** was then coupled with Boc-Phe-OH, followed by removal of the Boc protecting group, to afford a

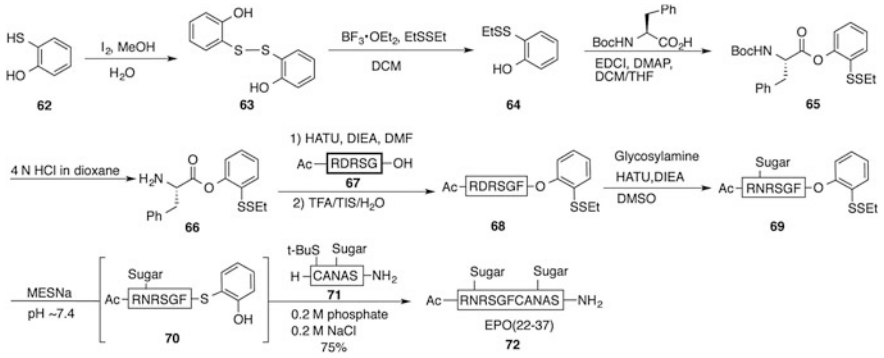


Fig. 9 Synthesis of a glycopeptide with use of the *O*-mercaptoaryl ester rearrangement (OMER) methodology

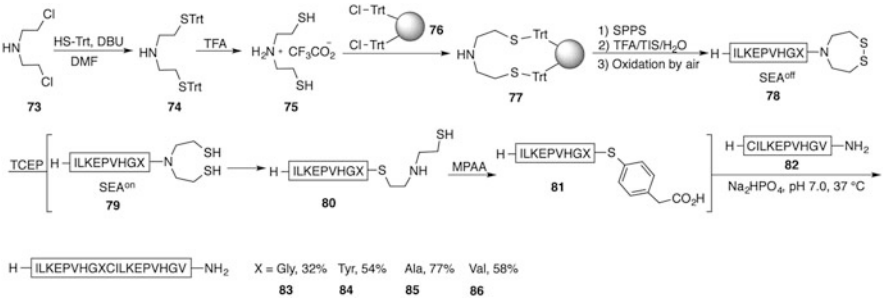


Fig. 10 Application of SEA peptide thioesters

phenylalanine ester **65**. The preparation of the peptide ester **68** was accomplished by condensing the fully protected peptide acid **67** with the phenylalanine ester, followed by TFA-deprotection. The peptide ester **68** is stable under the conditions for the synthesis of N-linked glycopeptide and enabled the production of a complex glycopeptide **72** in 75% yield.

In 2010, Melnyk and co-workers developed a new $N \rightarrow S$ acyl shift-based method for the synthesis of peptide thioesters [122–128] which utilizes a C-terminal bis(2-sulfanylethyl)amide (SEA) group. As depicted in Fig. 10, the SEA moiety is introduced to the C-terminus of the peptide via standard SPPS on a slightly modified resin support. After peptide cleavage, the open dithiol form **79** undergoes a reversible intramolecular thiol-exchange reaction to form a transient thioester intermediate **80**. The convenience of the SEA method has been illustrated by the synthesis of many peptides. For example, at slightly elevated temperature, peptides H-ILKEPVHGX-SEA (X=G, A, Y, or V) **78** were successfully ligated with peptide H-CILKEPVHGV-NH_2 **82** to afford peptides H-ILKEPVHGX-CILKEPVHGV-NH_2 **83–86** in 32–77% yield (Fig. 10) [122].

Peptide thioesters can also be manufactured by thiolyzing activated acyl groups. This idea was explored and successfully applied by Dawson and co-workers

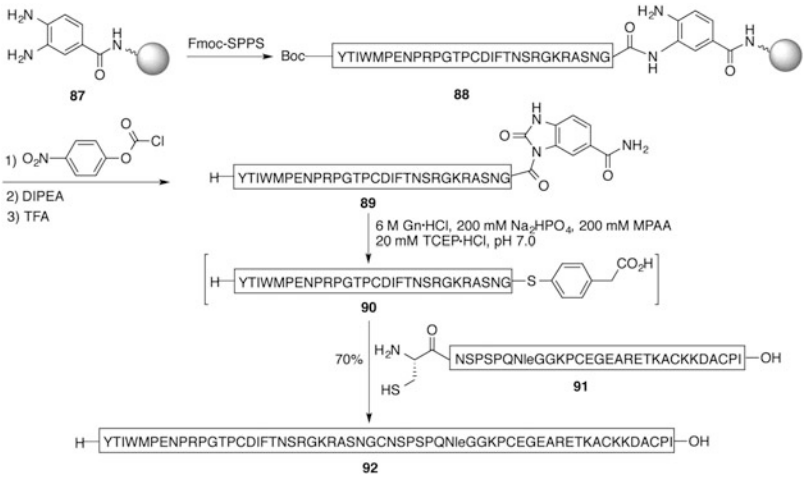


Fig. 11 Application of Nbz-activated peptides for peptide-thioester generation

[129]. Dawson's method uses a 3-(Fmoc-amino)-4-aminobenzoyl AM resin and, as shown in Fig. 11, the peptide chain is assembled on one of the aniline groups. After the peptide synthesis is completed, the 3,4-diaminobenzoyl (Dbz) linker is activated with *p*-nitrophenylchloroformate and mild base to generate a cyclic *N*-acylurea moiety (Nbz) **89**. This Nbz group is mildly activating, but stable to the acid-catalyzed cleavage methods used in standard Fmoc SPPS protocols; it can also be rapidly thiolyzed in the presence of thiols such as 4-mercaptophenylacetic acid (MPAA) to give peptide thioesters.

Several groups have reported that problems with over-acylation during coupling cycles of SPPS can occur in certain cases while using the Dbz resin-based method. For example, glycine rich sequences tend to lead to acylation at the second amino group of the Dbz linker, leading to branched Nbz derivatives of varying length. Recently, Ottesen and co-workers have proposed using Alloc as a protecting group for the second amino group of the Dbz resin [130], potentially eliminating almost all problems with branched peptide derivatives and over-acylation. The Alloc group is orthogonal to almost all common Fmoc SPPS protecting groups, stable under SPPS conditions, and can be easily removed with palladium(0) before Dbz activation to form the peptide Nbz derivative.

Similar to Dawson's Nbz-peptides, peptide acyl azides can be readily thiolyzed to generate peptide thioesters. Peptide acyl azides are also fairly well known and are easily generated from peptide hydrazides after oxidation with nitrous acid (Fig. 12) [131–135]. Early work in the field of chemical peptide synthesis by Hofmann and others relied heavily on acyl azides as a way to couple peptide fragments [87–89]. These methods were problematic for several reasons, including the favorable Curtius rearrangement which acyl azides readily undergo. In an attempt to overcome these limitations, Liu and co-workers have investigated extensively peptide hydrazines and their conversion to acyl azides [134]. They found that the conversion of hydrazide **96** to acyl azide **97** could be completed nearly

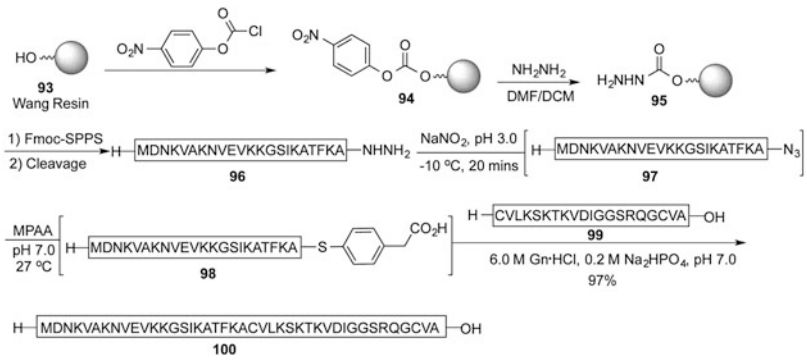


Fig. 12 Use of peptide hydrazines for peptide thioester synthesis

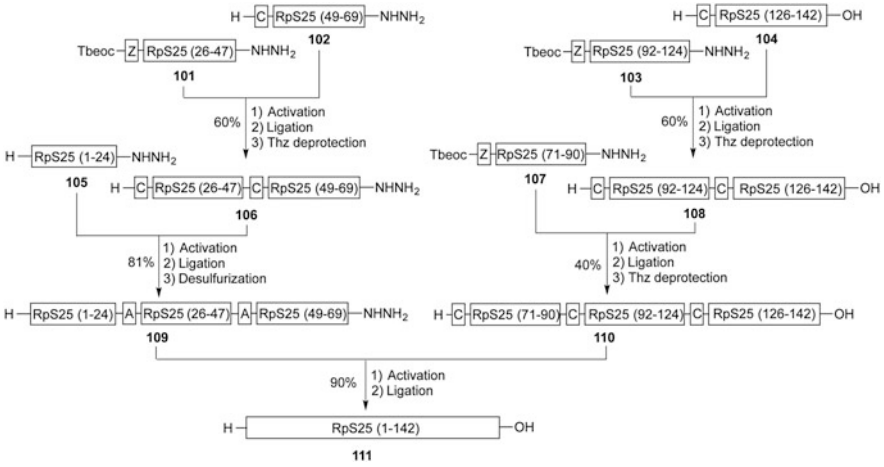


Fig. 13 Synthesis of ribosomal protein S25 using peptide hydrazines as thioester precursors

quantitatively in about 20 min at a pH of 3–4 at −10°C. Subsequent addition of thiols, such as MPAA, to the reaction mixture, followed by an increase in pH to 7.0, reliably generated the required thioester **98** in a convenient one-pot procedure. Usually, the resulting thioester was used directly in an NCL reaction without further purification. Remarkably, there was no evidence for oxidation of Met or Trp residues in the peptide fragments and very little racemization at C-terminal amino acids as a result of this procedure. Because of the rapid internal cyclization of Gln, Asp, and Asn acyl azide derivatives, this method was not successful at generating thioesters with these three residues; but all 17 other proteinogenic amino acids are compatible with this method.

Besides being easy to synthesize and convert to reactive thioesters, peptide hydrazides are also stable to standard NCL conditions. This opens a new door for designing convergent syntheses of proteins, as demonstrated by Liu and co-workers with the preparation of a 142-residue protein, the ribosomal protein S25 (RpS25)

[132]. As shown in Fig. 13, the full-length protein was divided into two halves, **109** and **110**, each to be assembled from three fragments, **101**, **102**, **105** and **103**, **104**, **107**, in the *N*-to-*C* direction. Except for the C-terminal one, each fragment was synthesized as a peptide hydrazide and converted to the thioester via acyl azide immediately before NCL.

3 Assembly of Peptide Fragments

Recent advances in solid-phase peptide synthesis have enabled the synthesis of peptides of around 30 residues to become relatively straightforward. Proteins in living systems, however, average 250 residues in length, making SPPS alone an impractical means of obtaining full-length, fully functional, and relevant proteins for study [6]. Ligation of shorter peptide fragments is currently the most convenient way to access longer protein constructs. Early examples of peptide fragment assembly revolved around the use of acyl azides as active partners in coupling reactions. This kind of "azide coupling" strategy culminated in the 1981 synthesis of the 124-residue bovine pancreatic ribonuclease A [90]. Yajima and co-workers successfully coupled a total of 30 peptide fragments together using peptide azides to form the full-length protein. That same year saw the advance of two other significant approaches to peptide fragment assembly. Sakakibara et al. proposed their "maximum protection" strategy of ligating fully protected peptide fragments together with the carbodiimide coupling reagent EDCI [136, 137]. This approach allowed ligation at many different residues and was validated through the successful synthesis of an 84-residue protein, human parathyroid hormone. The other method proposed in 1981, called the prior thiol-capture method, was put forward by Kemp and co-workers [138]. Prior thiol-capture involves the attachment of an ester linked auxiliary group to the C-terminus of one fragment. The auxiliary group contains a free thiol, which initially forms a disulfide bond to the thiol side chain of an N-terminal cysteine on the other fragment. A spontaneous acyl transfer reaction then occurs, releasing the thiol-containing auxiliary and forming an amide bond between the two fragments. Although powerful, Kemp's method was found to be overly sensitive to steric bulk at the ligation site.

Several years later, in 1992, a new method for coupling unprotected fragments in aqueous solutions was introduced by Kent and Schnolzer [139]. The new coupling took advantage of the unique ability of thioacids to perform nucleophilic substitutions on alkyl bromides to create thioester linkages between two peptide fragments. Even though the ligation was efficient, did not require side chain protection, and proceeded readily in aqueous solution, the introduction of a non-native bond into the protein structure severely limited application of this kind of chemical ligation. It was 2 years later that Kent's group published a reliable way to form native amide bonds between unprotected peptide fragments in aqueous solutions. The method was called native chemical ligation (NCL) to contrast with Kent's previous chemical ligation strategy which did not form a native, amide bond. Rather than a nucleophilic thioacid, NCL used a thioester as an electrophile. At the

time, it was well known that thioester exchange reactions took place rapidly in aqueous solutions. By placing a thioester at the C-terminus of one peptide fragment and a thiol-containing cysteine residue at the N-terminus of the other peptide fragment, a thioesterification reaction could take place linking the two fragments via thioester. Unlike previous thioester-forming ligation strategies, however, the free amine of the cysteine residue would be well spaced to undergo an *S*-to-*N* acyl shift. Such a shift occurs rapidly via a favorable five-membered ring-containing cyclic intermediate and results in the formation of a natural amide bond between the two peptide fragments (Fig. 1).

Since the original development of NCL, many groups have discovered ways of extending the method. For example, thio-substituted amino acids were developed to get around the requirement of an N-terminal cysteine in the classic NCL design [140–158]. Although most of the thio-amino acids are not commercially available and can be somewhat complicated to synthesize, they have significantly expanded the scope of NCL. Figure 14 shows the 17 new thio-substituted amino acids which have been synthesized and, together with mild desulfurization conditions, allow for ligations at almost all proteogenic amino acids.

Extensive synthetic studies have confirmed the viability of a Cys-free NCL approach utilizing thio-amino acid building blocks. In a seminal piece of work, Danishefsky and co-workers synthesized human parathyroid hormone (PTH) using Cys, thio-Val **115**, and thio-Leu **121** [159]. As shown in Fig. 15, PTH, a Cys-lacking protein containing 84 amino acids, was assembled from 4 fragments: PTH (1–23) **129**, (24–38) **130**, (39–59) **131**, and (60–84) **132**. Under this synthetic

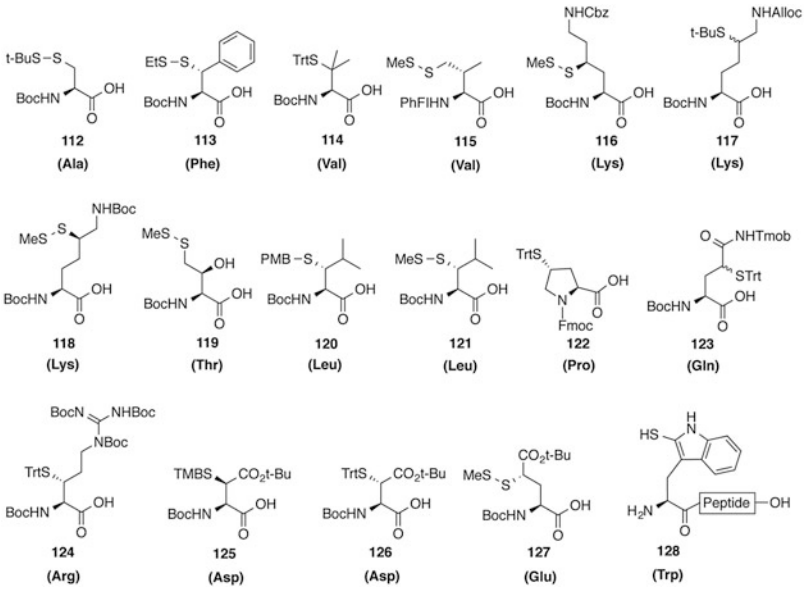


Fig. 14 Thiol-substituted amino acids for use in NCL

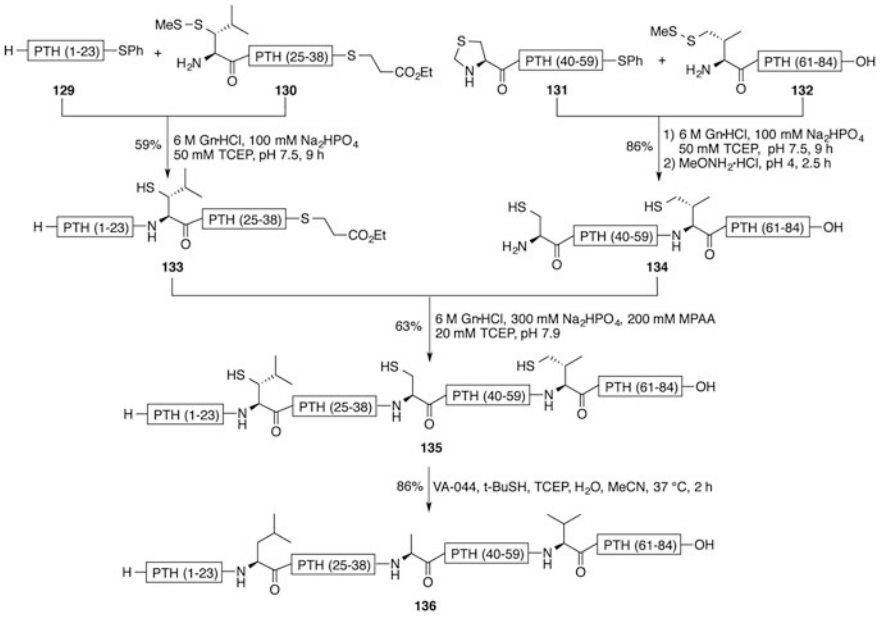


Fig. 15 Chemical synthesis of human PTH using only Cys-free NCL

route, thio-Leu, Thz, and thio-Val, which are located at the N-termini of **130**, **131**, and **132**, served as surrogates for Leu, Ala, and Val to facilitate NCL of the four peptide fragments. Following ligation, all three thiol auxiliaries were simultaneously removed through exposure to the MFD conditions. This work, for the first time, demonstrated the power and broad scope of the two-step, Cys-free NCL/MFD for the total synthesis of Cys-poor proteins.

In addition to new thio-substituted amino acids, novel alternatives to NCL have surfaced in recent years. Efforts to expand upon NCL have also led to reliable ways to ligate multiple peptide fragments in one-pot, opening up more strategies for chemical synthesis of proteins.

3.1 New Methods for Peptide Ligation

3.1.1 KAHA Method

Unprotected hydroxylamines and α-ketoacids are known to undergo condensation and form new amide bonds. Similar to classic NCL, this α-ketoacid-hydroxylamine (KAHA) ligation requires no reagents and works even in the presence of unprotected amino acid side chains (Fig. 16a). This attractive reaction was investigated as a means to ligate peptide fragments together, but was found to have severe limitations [160]. Further development, however, revealed that, by replacing the

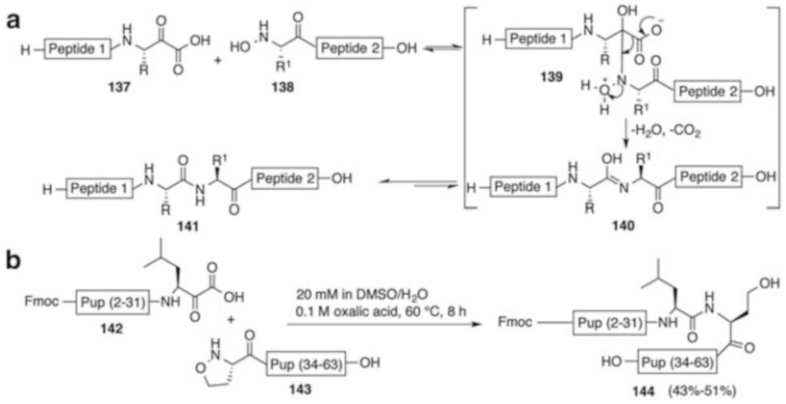


Fig. 16 The KAHA ligation. (**a**) General reaction and proposed mechanism. (**b**) Synthesis of Pup 2-63 via KAHA ligation

unprotected N-terminal hydroxylamine with a 5-oxaproline non-natural amino acid, KAHA ligation could be very successful [161]. Unlike the *N*-hydroxyamino acids used initially, 5-oxaproline is stable to peptide coupling conditions, a distinct advantage which allows attachment to proceed exactly as does any standard amino acid used in Fmoc SPPS. C-terminal peptide α-ketoacids are also easily obtained by SPPS with the use of a recently developed cyanosulfur-ylide linker and Rink amide MBHA resin. Cleavage of the linker from the resin gives a free cyanosulfur-ylide which is readily and selectively oxidized to the α-ketoacid upon treatment with oxone [162]. After ligation, the 5-oxaproling ring is opened to form a homoserine residue. While not a natural amino acid, homoserine can be used as a substitute for several natural amino acids with minimal disruption of the final protein structure.

Bode has demonstrated the usefulness of this KAHA ligation method through the synthesis of Pup, the prokaryotic analog of ubiquitin (Fig. 16b) [161]. The 62-residue protein was assembled from 2 unprotected fragments. The C-terminal, α-ketoacid fragment Pup (2–32) **142** was prepared using standard Fmoc SPPS and the cyanosulfur-ylide linker system mentioned previously. The N-terminal fragment Pup (33–63) **143** was also synthesized with standard Fmoc SPPS chemistry, followed by manual coupling of *N*-Boc-protected 5-oxaproline with HCTU. Treatment with TFA-based cleavage cocktail liberated the N-terminal oxaproline fragment from the resin and removed all side chain protecting groups. The two fragments were coupled successfully in 24 h in a solution of 6:4 DMSO:H_2O and 0.1 M oxalic acid to give the full-length Pup (2–63) protein **144** in 43–51% isolated yield. Despite the ligation being run at 50°C, no evidence of asparagine hydrolysis was observed.

3.1.2 Fragment Condensation

In addition to the Cys-free NCL approach mentioned previously, fragment condensation is a common means of overcoming the difficulty associated with synthesizing Cys-scarce peptides and proteins [163]. Danishefsky and co-workers recently developed a new fragment condensation method based on the previously established *O*-mercaptoaryl ester rearrangement (OMER). This new method is known as phenolic ester directed amide coupling (PEDAC) [164]. In this coupling reaction, peptide fragments with *O*-mercaptoaryl esters are coupled to peptide fragments with standard amino acids at the N-terminus, removing the need for any thiol auxiliary groups at the ligation site. The PEDAC reaction can be either TCEP or AgCl assisted (Fig. 17a). In the presence of TCEP or AgCl and DIEA in DMSO, the *O*-mercaptoaryl ester is in equilibrium with the thioester form. The reactive thioester peptide reacts with HOOBt, also in solution, to give a high energy acyl donor which is easily attacked by the peptidyl amine of a second fragment to afford the ligation product.

Peptide esters bearing C-terminal Gly or Pro residues were found to react smoothly with peptidyl amines to give excellent yields. However, peptide esters bearing C-terminal Ala or Phe residues were found to undergo significant amounts of racemization under these conditions. Danishefsky and co-workers screened different conditions in an attempt to suppress racemization and found that replacing DIEA with the sterically hindered base DBDMAP (2,5-di-*tert*-butyl-*N*,*N*-dimethylaminopyridine, $pK_a = 9$) decreases racemization of Ala from 12–15% to 7–9% and Phe from 40–50% to 20–25%. Furthermore, only slightly lower coupling rates and yields are observed when DBDMAP was used in place of DIEA.

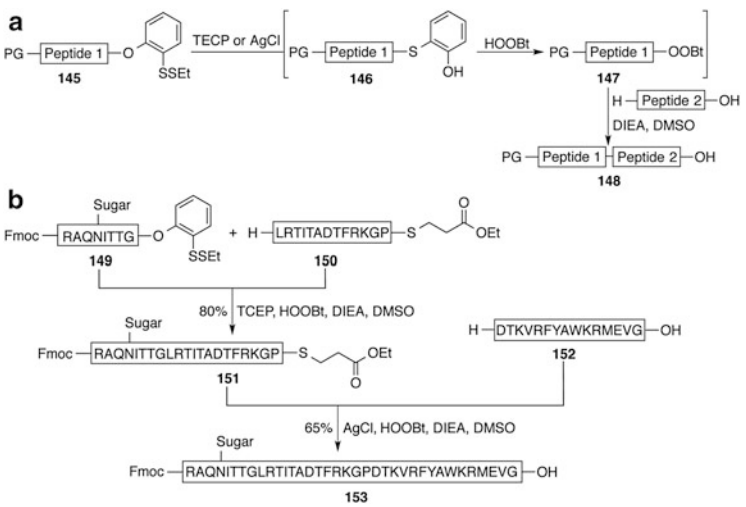


Fig. 17 Phenolic ester directed amide coupling (PEDAC). (**a**) General scheme of PEDAC reaction. (**b**) Selective PEDAC cascade for the synthesis of a short glycopeptide

While either TCEP or AgCl can be used in this new fragment condensation, they each have unique features. AgCl was found to activate either aryl acyl donors or alkyl thioesters at the C-terminus of peptide fragments. TCEP, however, could activate only aryl acyl donors, and did so without disruption of an alkyl thioester at the C-terminus of the non-cysteine acyl acceptor fragment. This difference in activation ability enables the cascade assembly of peptide fragments under strategic TECP or AgCl activation, a feature that was exploited to synthesize a short glycopeptide. Under PEDAC-TCEP conditions, peptide **149** was coupled with peptide **150** without sacrificing the C-terminal alkyl thioester. Resulting intermediate peptide **151** was then joined with another fragment, **152**, under AgCl conditions to furnish the final adduct **153** in 65% yield (Fig. 17b) [164].

Later, Hojo and co-workers developed a post-SPPS thioesterification reaction to synthesize reactive peptide aryl thioesters using Fmoc SPPS [116]. Peptide fragments are synthesized on the CLEAR amide resin, and the N-protected amino acid Fmoc-(Et)Cys(Trt)-OH was introduced as the first amino acid. After Fmoc SPPS, TFA cleavage, and deprotection, the precursor peptide fragment is mixed with 2% MPAA in 30% aqueous acetonitrile. Under these conditions, an *N*-to-*S* acyl transfer occurs to form an intermediate thioester from the N-protected Cys residue introduced as the first amino acid during SPPS. This intermediate alkyl thioester then undergoes transthioesterification with MPAA in the solution to form a more reactive aryl thioester (Fig. 18a).

By combining this thioester synthesis and fragment condensation, Hojo's group was able to synthesize a 95-residue chemokine, CCL27 **163** (Fig. 18b). In their synthetic plan for CCL27, they chose ligation sites such that Gly would be the

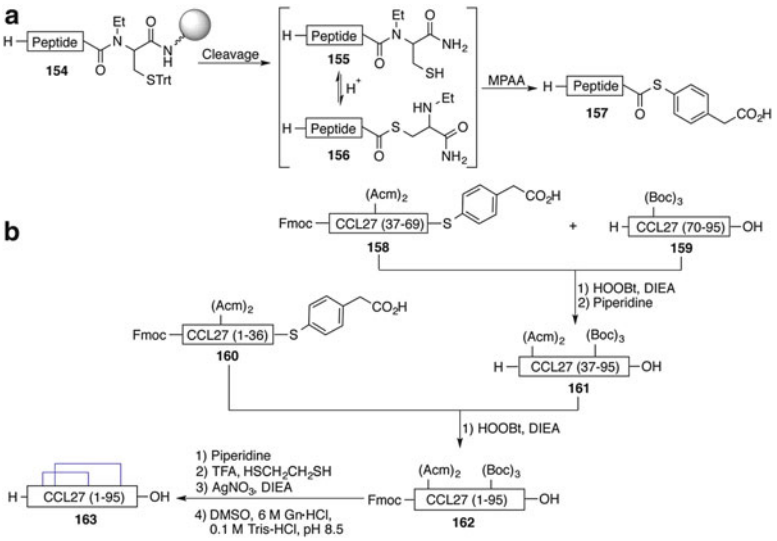


Fig. 18 Synthesis of CCL27 using a thioester synthesis and fragment condensation approach

C-terminal thioester residue for each fragment. Additionally, they protected the thiol of Cys side chains and the ε-amine of Lys side chains with Acm and Boc groups, respectively. Under HOOBt/DIEA-promoted coupling conditions, peptides **158** and **159** were successfully linked. Standard N^{α}-Fmoc removal freed the peptidyl amine for another coupling and intermediate peptide **161** was ligated to peptide **160** to afford full-length CCL27.

3.1.3 Ser and Thr Ligation

The total frequency of serine and threonine residues is up to 12.7% in natural polypeptides and proteins [6]. Therefore, it is highly desirable to develop the ligations at Ser and Thr. In 2010, Li and co-workers described an efficient protocol which featured the salicylaldehyde (SAL) ester-mediated ligation of unprotected peptides at Ser and Thr residues (Fig. 19a) [165]. First, the amine group of the N-terminal Ser and Thr is reversibly ligated with the aldehyde group of the C-terminus to form a cyclic *N*,*O*-acetal intermediate **166**, which spontaneously undergoes an irreversible acyl transfer to provide the amide bond. Second, the acetal group is hydrolyzed under acidic conditions to release the serine or threonine residues.

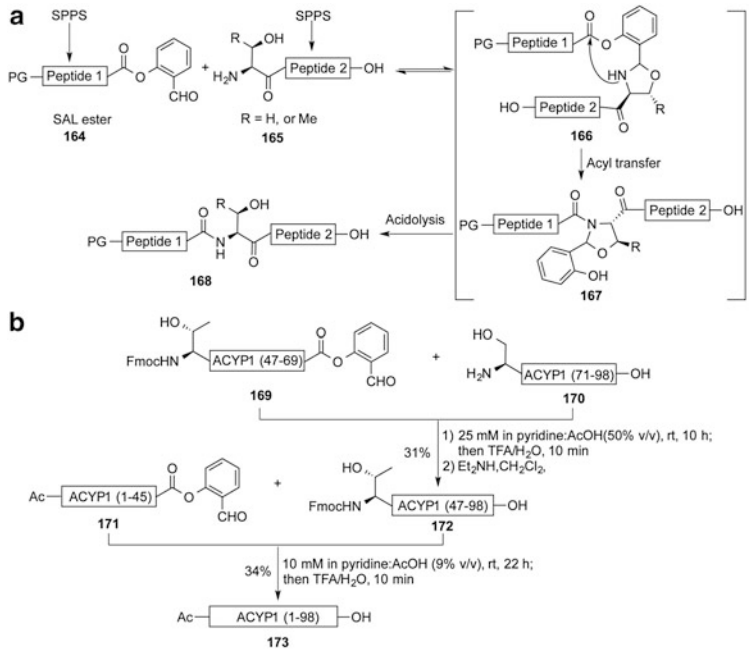


Fig. 19 Ligation at serine and threonine. (**a**) General example of salicylaldehyde ester-mediated ligation. (**b**) Convergent synthesis of human erythrocyte acylphosphatase via serine ligation

To demonstrate the applicability of this strategy, they used this ligation method to synthesize the 98-residue human erythrocyte acylphosphatase (ACYP1) in a convergent manner (Fig. 19b) [166]. The protein was assembled from three fully protected segments, each prepared with Fmoc SPPS. The C-termini of segments (46–69) **169** and (1–45) **171** were converted into salicylaldehyde esters by DCC-promoted condensation with salicylaldehyde. They first performed the ligation of peptide segments (46–69) **169** and (70–98) **170** at Gly69-Ser70. The mixture was stirred in a pyridine acetate solution for 10 h at room temperature, and then treated with TFA cocktail to deliver the ligated peptide corresponding to segment (46–98) **172** in 31% yield. Next, the resultant peptide **172** was ligated with fragment (1–45) **171** at Gly45-Thr46. After exposure to TFA cocktail, the full-length polypeptide **173** was synthesized in 34% isolated yield. After folding, the chemically synthesized protein exhibited the reported activity.

3.2 One-Pot Strategies for the Assembly of Peptide Fragments

Consecutive assembly of multiple peptide fragments allows for the synthesis of long protein constructs. However, traditional NCL methods involve the tedious purification of intermediate ligation products after each sequential reaction. These purification steps often result in significant product loss and can have a discouraging effect on the overall synthetic yield. Removing, or at the very least minimizing, these purification steps, has the potential to boost substantially the synthetic efficiency of large proteins. For this reason, one-pot strategies where multiple ligation reactions can take place selectively and reliably without the need to purify intermediate products have been proposed and optimized in recent years [167]. Currently, one-pot fragment assembly processes have been developed for ligation in either the *C*-to-*N* or *N*-to-*C* direction.

3.2.1 *C*-to-*N* Direction

In the *C*-to-*N* strategy, the first ligation is between two fragments which end up at the C-terminus of the final protein sequence (Fig. 20). The resulting crude peptide, without purification, is directly subject to ligation with the fragment on its N-terminal side. This process is repeated until the desired peptide is synthesized.

Early *C*-to-*N* assembly strategies used different Cys protecting groups to control the reactivity of each fragment, making the protecting groups for the N-terminal Cys residues of the middle fragments especially critical. The conditions required for the removal of these protecting groups must be compatible with the conditions for the subsequent ligation. Otherwise, isolation and purification steps would be necessary following each deprotection. Currently, the protecting group most widely

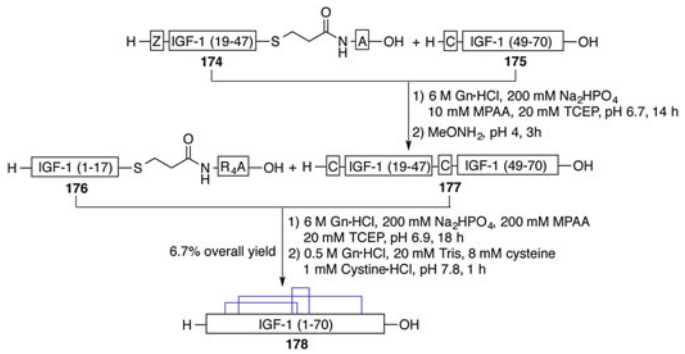


Fig. 20 *C*-to-*N* one-pot NCL for the construction of IGF-1 (1–70)

used for this purpose is the thiazolidine (Thz) group [168]. The Thz group can be easily introduced to the N-termini of middle fragments by using the commercially available Boc-Thz-OH building block. Thz residues are also easily transferred to Cys residues by treatment with an excess of *O*-methylhydroxylamine ($MeONH_2$) under slightly acidic conditions (typically at pH 4). Importantly, these deprotection conditions do not interfere with subsequent ligation steps, which are often carried out at neutral or slightly basic pH.

Kent and co-workers took advantage of the *C*-to-*N* one-pot strategy in their 2008 synthesis of insulin-like growth factor 1 (IGF-1) and its diastereomeric analogue [Gly7D-Ala]IGF-1 [169]. As shown in Fig. 20, the IGF-1 sequence was divided into three fragments. Both a Thz residue and a C-terminal thioester were installed on the middle fragment, IGF-1 (Thz18-47) **174**. It was found that the most N-terminal fragment, IGF-1 (1–17) **176**, was not water soluble, so a special Arg_4-tag was introduced into the thioester moiety of the fragment to increase its solubility. Fragments IGF-1 (Thz18-47) **174** and IGF-1 (48–70) **175** were assembled first under general NCL conditions to furnish fragment IGF-1 (Thz18-70). The Thz residue was subsequently deprotected with 0.2 mM methoxylamine · HCl at pH 4 to yield IGF-1 (18–70) **177**. After solid-phase extraction and lyophilization, ligation of fragment IGF-1 (1–17) **176** and crude fragment IGF-1 (Cys18-70) **177** was performed. A high concentration of MPAA was required to complete this step because of the low reactivity of Val-Cys ligation.

The above example of one-pot fragment assembly was dependent on the use of Cys protecting groups. However, it is also possible to control the order of ligations by exploiting the different reactivities of N-terminal thio-substituted amino acids. Such a strategy would mean fragments could be assembled without N-terminal thiol protecting groups on the middle fragments. Danishefsky and co-workers first demonstrated such a kinetically controlled cascade coupling strategy in 2010 [148]. As shown in Fig. 21, EPO fragment (95–120) **184** was obtained from three individual fragments through sequential cysteine and thioleucine ligations in one pot with no intervening purification or isolation steps. The first ligation connected peptides **179** and **180** by standard Cys-based NCL. The resulting intermediate

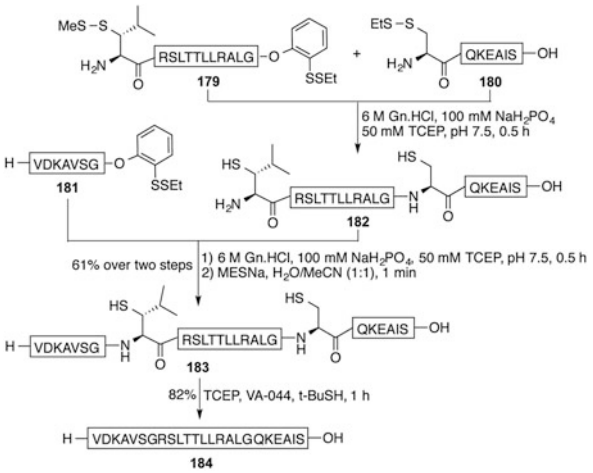


Fig. 21 *C*-to-*N* one-pot ligation using a kinetically-controlled coupling cascade

peptide **182** then underwent a second ligation to peptide **181** through a slower thioleucine ligation. The desired double ligation product **183** was isolated in 61% yield. This product was found to correspond to the expected sequence: **181** + **179** + **180**. The single ligation product corresponding to the self-assembly of peptide **179** could not be found during the first ligation and neither was the double ligation product with the sequence: **179** + **179** + **180**. These results illustrate a novel way of controlling the order of sequential ligations by fine-tuning the reaction kinetics at each ligation site.

3.2.2 *N*-to-*C* Direction

Several methods have also been developed to ligate multiple peptide fragments in the *N*-to-*C* direction. The Kinetically Controlled Ligation (KCL) developed by Kent and co-workers is one such method, which takes advantage of the different reactivities of alkyl and aryl thioesters [85, 170]. This method was elegantly applied in the 2012 synthesis of a cyclic, crambin-derived topological analogue protein (topologue) by Kent and co-workers (Fig. 22) [171]. The topologue **189** consisted of the natural sequence of crambin, but the ε-amino group of Lys10 is acylated by the α-carboxyl group of Asn46 though an iso-peptide bond. An aryl thioester was installed at the N-terminus of fragment **185**, while an alkyl thioester was introduced to the C-terminus of fragment **186**. The first NCL was conducted under KCL reaction conditions, in the absence of an exogenous thiol catalyst (MPAA or thiophenol), to give the intermediate fragment **187**. After deprotection (Thz40 to Cys40 conversion), the ring was closed through a second NCL in the presence of MPAA. Acm removal and folding gave the final cyclic peptide construct **189**.

The SEA system mentioned previously can also be used in the one-pot, *N*-to-*C* direction assembly of three peptide fragments [122]. In 2012, Melnyk and

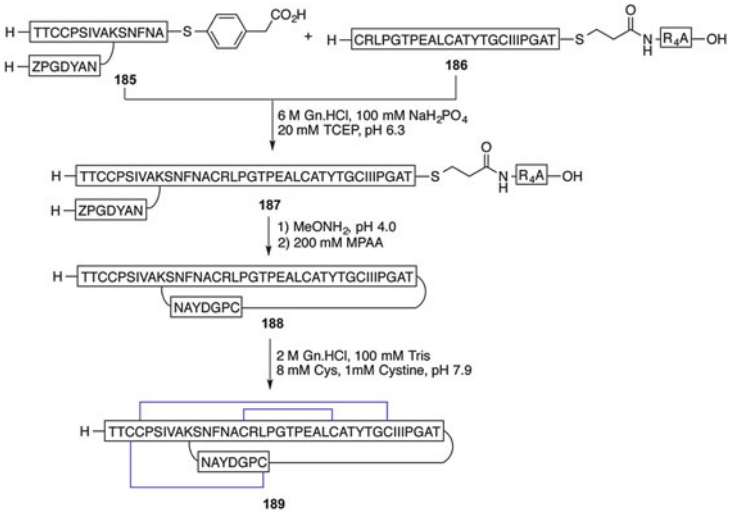


Fig. 22 Kinetically controlled ligation in the synthesis of crambin topologue

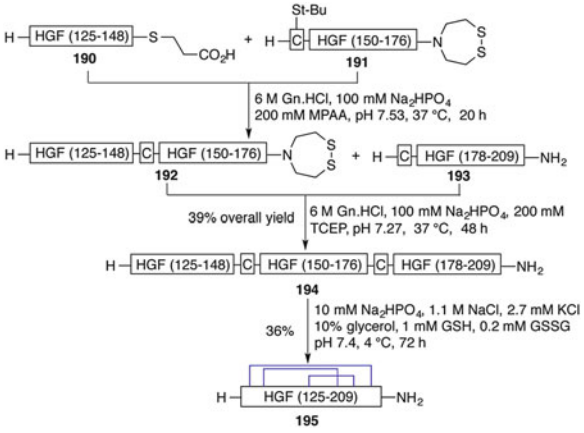


Fig. 23 One-pot *N*-to-*C* assembly of HGF(125–209) using the $SEA^{on/off}$ approach

co-workers synthesized the biologically active K1 domain from the hepatocyte growth factor (HGF) using this strategy [124]. The K1 domain, HGF (125–209) **195**, was divided into three fragments: fragment (125–148) **190**, fragment (149–176) **191**, and fragment (177–209) **193**. The SEA^{off} amide moiety was installed to the C-terminus of fragment **191**. The ligation between fragment **190** and **191** proceeded in the presence of MPAA to afford peptide **192**. Addition of TCEP to the reaction "turned on" the SEA moiety by reducing the disulfide bond and allowing an *N*-to-*S* acyl transfer to occur. MPAA, still present in the reaction solution, exchanged with the newly created thioester to activate the intermediate

peptide **192** towards attack by the N-terminal Cys of peptide **193** and trigger the second ligation. This one-pot method led to the formation of the full-length, linear K1 domain **194** in 39% yield and obviated the need for purification of any intermediate peptide fragments (Fig. 23).

4 Folding of Synthetic Polypeptide Chains

Once the full-length primary sequence is assembled, it must be folded into a biologically active structure. Two distinct approaches exist for the formation of disulfide bonds and folding proteins into proper tertiary structures. Individual disulfide bonds can be sequentially formed using an orthogonal protecting group for each pair of cysteine residues and irreversible oxidation steps, or a reversible redox buffer system can be used to form all disulfide bonds simultaneously in a thermodynamically driven random process [172, 173]. Neither of these methods is without flaws and the success of either method is highly sequence dependent.

Forming disulfide bonds irreversibly using orthogonal thiol protecting groups allows for complete control over the resulting disulfide network, making it a very appealing method. It is, however, inherently limited by the number of orthogonal thiol protecting groups that are stable during peptide synthesis [174]. Four protecting groups are most commonly used for Fmoc-based peptide synthesis and subsequent step-wise, selective disulfide bond formation: *tert*-butyl mercapto (*S*tBu), removed with mild reduction by thiols such as DTT; acetamidomethyl (Acm), removed by treatment with I_2; monomethoxytrityl (Mmt), removed by acid; and trityl (Trt), also removed by acid but considerably less labile than Mmt, allowing for selective removal of Mmt groups in the presence of Trt.

The synthesis of human insulin is a classic example of using this totally controlled, step-wise disulfide bond methodology. Insulin consists of two peptide chains, known as the A and B chains, joined by intermolecular disulfide bonds; a third intramolecular disulfide bond is present within the A chain (Fig. 24). Many groups have synthesized insulin and various analogs using a controlled step-wise approach, and their methods have already been reviewed elsewhere [175, 176]. One noteworthy example from the recent literature was accomplished in a remarkable 24% yield, based on the substitution of the A-chain resin [177]. The synthetic plan required use of all four above-mentioned thiol protecting groups for cysteine. In addition, the synthesis used the well documented strategies of 2,2′-dithiobis (5-nitropyridine) (DTNP) activation for rapid disulfide formation [178] and isoacyl dipeptides (Fig. 24, underlined residues) for synthesis of the hydrophobic chains [179]. The 21-residue A chain **197** was first synthesized by standard Fmoc SPPS chemistry with all four of its cysteines protected orthogonally. The intramolecular A chain disulfide bond was formed on resin by first deprotecting CysA6, activating it with DTNB, and then deprotecting CysA11. The resulting free thiol of CysA11 rapidly reacted with the DTNB-activated CysA6 to form the disulfide bond. The A chain **198** was then cleaved from the resin, simultaneously removing the Trt group

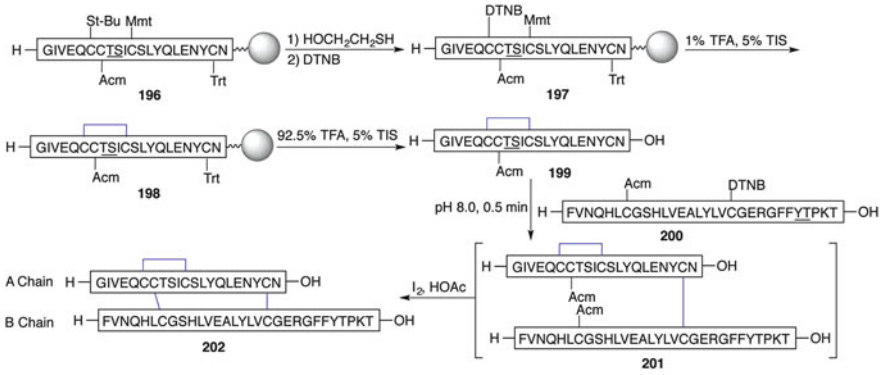


Fig. 24 Synthesis of human insulin in high yield with disulfide bonds formed in a regioselective manner

of CysA20. The B chain **200** contains two Cys residues, and was also synthesized with standard Fmoc SPPS chemistry. CysB7 was protected with Acm group and CysB19 was activated with DTNB after cleavage from the resin. Combining the DTNB-CysB19 activated B chain **200** and free thiol CysA20 A chain **199** rapidly gave the ligated A-B dimer **201** as expected. Simultaneous Acm group removal and oxidation to form the remaining disulfide bond (A7-B7), and subsequent isoacyl dipeptide rearrangement completed the synthesis.

Recent efforts to improve the chemistry involved in the controlled, step-wise formation of disulfide bonds has resulted in new methods of oxidizing free thiols to form disulfides and novel cysteine protecting groups. Albericio and co-workers described *N*-chlorosuccinimide (NCS) as a mild oxidizing reagent for the formation of disulfide bonds on resin [180]. They used this new method to synthesize oxytocin, a 9-residue peptide containing 1 disulfide bond, and SI α-conotoxin, a 13-residue peptide containing 2 disulfide bonds. Furthermore, they report that methionine and tryptophan residues as well as Trt and Mmt cysteine protecting groups were unaffected by the NCS oxidation step. None of those groups are compatible with I_2 oxidation, the most common method for disulfide bond formation, which makes NCS a valuable alternative. The same group also recently reported the trimethoxyphenylthio (STmp) protecting group as an improved substitute for the commonly used *tert*-butylthio (StBu) group [181]. Compared to the StBu group, STmp was shown to be removed by mild DTT reduction both quicker and more completely, resulting in higher peptide purity and fewer side reactions (Fig. 25).

The alternative to a step-wise orthogonal protecting group strategy for disulfide bond formation is to use a redox buffer to form all disulfide bonds simultaneously. This approach has the advantage of requiring fewer steps and being operationally more convenient, although isomeric disulfide species are common, which complicates the purification and can lower yield. A redox buffer contains a free thiol and its oxidized disulfide together in solution [182]. The presence of both oxidized and

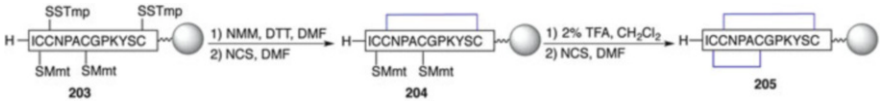


Fig. 25 Use of both NCS oxidation and *S*Tmp thiol protection in the synthesis of SI α-conotoxin. *S*Tmp = trimethoxyphenylthio

reduced thiols initiates a series of random exchanges between free thiols and disulfides, eventually leading to a thermodynamically favored structure for the peptide. The exact mechanism of this folding/disulfide bond formation reaction is not well understood, since the partially folded intermediates and various crossed disulfide species are difficult to characterize. Commonly used redox buffer systems are all small, aliphatic thiols, such as glutathione, cysteine, β-mercaptoethanol (BME), and dithiothreitol (DTT). A detailed examination of the effect of different kinds of thiols on the folding efficiency of lysozyme revealed that symmetric, aromatic dithiols containing quaternary ammonium moieties gave the fastest kinetics and highest yields [183]. The use of thiols containing positively charged groups has also recently been investigated by a different group working on analogs of glutathione. The Hidaka group has done extensive work with the tripeptide Arg-Cys-Gly (RCG) and its use as a redox buffer [184, 185]. They have shown that an RCG-based buffer system resulted in a significantly higher folding recovery of lysozyme as compared to glutathione. The effect, however, was demonstrated to be pH and concentration dependent, thus highlighting the currently murky understanding of the relationship between folding efficiency and the redox buffer.

5 Conclusion

Scientific interest in precisely understanding the fundamental aspects of protein structure and function has led to the development of chemical synthesis of proteins. With the advances in solid phase peptide synthesis, peptide fragment assembly, and protein chain folding, the total chemical synthesis of proteins is now a practical reality. This has greatly facilitated research in biochemistry and molecular biology through the application of carefully designed and chemically synthesized peptides and proteins. Recently, many new and more general synthetic methods to produce proteins have been introduced. By addressing the limitations and inefficiencies associated with existing methods, these new developments have the potential to make chemical protein synthesis more efficient and even easier to exploit. Despite the significant progress made, chemical synthesis of proteins remains an imperfect tool. Depending on the sequence of the synthetic target, unexpected difficulties can pop up at every stage of a potential synthesis. Therefore, much effort is still required to develop more general methods in the future, especially those allowing for the synthesis of long peptide fragments, ligation at any pair of amino acids, and selective folding.

Acknowledgements The authors wish to thank the University of Colorado Boulder and the Butcher seed grant program for financial support. Additionally, we thank all of those who have contributed to the advancement of the science of protein chemical synthesis.

References

1. Kent SB (2009) Total chemical synthesis of proteins. Chem Soc Rev 38:338–351
2. Price JL, Culyba EK, Chen W, Murray AN, Hanson SR, Wong CH, Powers ET, Kelly JW (2012) N-Glycosylation of enhanced aromatic sequons to increase glycoprotein stability. Biopolymers 98:195–211
3. Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. Annu Rev Biochem 79:413–444
4. Chen LQ, Drake MR, Resch MG, Greene ER, Himmel ME, Chaffey PK, Beckham GT, Tan ZP (2014) Specificity of O-glycosylation in enhancing the stability and cellulose binding affinity of Family 1 carbohydrate-binding modules. Proc Natl Acad Sci USA 111:7612–7617
5. Kimmerlin T, Seebach D (2005) 100 Years of peptide synthesis: ligation methods for peptide and protein synthesis with applications to beta-peptide assemblies. J Pept Res 65:229–260
6. Nilsson BL, Soellner MB, Raines RT (2005) Chemical synthesis of proteins. Annu Rev Biophys Biomol Struct 34:91–118
7. Wilson RM, Dong S, Wang P, Danishefsky SJ (2013) The winding pathway to erythropoietin along the chemistry-biology frontier: a success at last. Angew Chem Int Ed Engl 52:7646–7665
8. Merrifield RB (1963) Solid phase peptide synthesis I. The synthesis of a tetrapeptide. J Am Chem Soc 85:2149–2154
9. Mitchell AR (2008) Bruce Merrifield and solid-phase peptide synthesis: a historical assessment. Biopolymers 90:175–184
10. Dawson PE, Muir TW, Clarklewis I, Kent SBH (1994) Synthesis of proteins by native chemical ligation. Science 266:776–779
11. 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 Engl 46:9248–9252
12. Reinwarth M, Nasu D, Kolmar H, Avrutina O (2012) Chemical synthesis, backbone cyclization and oxidative folding of cystine-knot peptides: promising scaffolds for applications in drug design. Molecules 17:12533–12552
13. Narayan M (2012) Disulfide bonds: protein folding and subcellular protein trafficking. FEBS J 279:2272–2282
14. Chandrudu S, Simerska P, Toth I (2013) Chemical methods for peptide and protein production. Molecules 18:4373–4388
15. Tickler AK, Clippingdale AB, Wade JD (2004) Amyloid-beta as a "difficult sequence" in solid phase peptide synthesis. Protein Pept Lett 11:377–384
16. Kaiser T, Nicholson GJ, Kohlbau HJ, Voelter W (1996) Racemization studies of Fmoc-Cys (Trt)-OH during stepwise Fmoc-solid phase peptide synthesis. Tetrahedron Lett 37:1187–1190
17. Nicolás E, Pedroso E, Girald E (1989) Formation of aspartimide peptides in Asp-Gly sequences. Tetrahedron Lett 30:497–500
18. Hackenberger CP, Schwarzer D (2008) Chemoselective ligation and modification strategies for peptides and proteins. Angew Chem Int Ed Engl 47:10030–10074
19. Malins LR, Payne RJ (2014) Recent extensions to native chemical ligation for the chemical synthesis of peptides and proteins. Curr Opin Chem Biol 22C:70–78

20. Allis CD, Muir TW (2011) Spreading chromatin into chemical biology. ChemBioChem 12:264–279
21. Chatterjee C, Muir TW (2010) Chemical approaches for studying histone modifications. J Biol Chem 285:11045–11050
22. Coin I, Beyermann M, Bienert M (2007) Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. Nat Protoc 2:3247–3256
23. Dawson PE, Kent SB (2000) Synthesis of native proteins by chemical ligation. Annu Rev Biochem 69:923–960
24. Dirksen A, Dawson PE (2008) Expanding the scope of chemoselective peptide ligations in chemical biology. Curr Opin Chem Biol 12:760–766
25. Flavell RR, Muir TW (2009) Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. Acc Chem Res 42:107–116
26. Muralidharan V, Muir TW (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. Nat Methods 3:429–438
27. Tiefenbrunn TK, Dawson PE (2010) Chemoselective ligation techniques: modern applications of time-honored chemistry. Biopolymers 94:95–106
28. Made V, Els-Heindl S, Beck-Sickinger AG (2014) Automated solid-phase peptide synthesis to obtain therapeutic peptides. Beilstein J Org Chem 10:1197–1212
29. Pedersen SL, Tofteng AP, Malik L, Jensen KJ (2012) Microwave heating in solid-phase peptide synthesis. Chem Soc Rev 41:1826–1844
30. Robinson NE, Robinson AB (2008) Use of Merrifield solid phase peptide synthesis in investigations of biological deamidation of peptides and proteins. Biopolymers 90:297–306
31. Sabatino G, Papini AM (2008) Advances in automatic, manual and microwave-assisted solid-phase peptide synthesis. Curr Opin Drug Discov Devel 11:762–770
32. Tickler AK, Wade JD (2007) Overview of solid phase synthesis of "difficult peptide" sequences. Chapter 18: Unit 18. Curr Protoc Protein Sci 18. doi:10.1002/0471140864.ps1808s50
33. Merrifield RB (1964) Solid-phase peptide synthesis. III. An improved synthesis of bradykinin. Biochemistry 3:1385–1390
34. Carpino LA, Han GY (1972) 9-Fluorenylmethoxycarbonyl amino-protecting group. J Org Chem 37:3404–3409
35. Bray BL (2003) Large-scale manufacture of peptide therapeutics by chemical synthesis. Nat Rev Drug Discov 2:587–593
36. Cudic M, Fields GB (2008) Solid-phase peptide synthesis. Springer, New York
37. Al-Warhi TI, Al-Hazimi HMA, El-Faham A (2012) Recent development in peptide coupling reagents. J Saudi Chem Soc 16:97–116
38. Pattabiraman VR, Bode JW (2011) Rethinking amide bond synthesis. Nature 480:471–479
39. Allen CL, Williams JM (2011) Metal-catalysed approaches to amide bond formation. Chem Soc Rev 40:3405–3415
40. Wilson RM, Stockdill JL, Wu X, Li X, Vadola PA, Park PK, Wang P, Danishefsky SJ (2012) A fascinating journey into history: exploration of the world of isonitriles en route to complex amides. Angew Chem Int Ed Engl 51:2834–2848
41. Li X, Danishefsky SJ (2008) New chemistry with old functional groups: on the reaction of isonitriles with carboxylic acids–a route to various amide types. J Am Chem Soc 130:5446–5448
42. Li X, Danishefsky SJ (2008) Noncatalytic reaction of isonitriles and carboxylic acids en route to amide-type linkages. Nat Protoc 3:1666–1670
43. Li X, Yuan Y, Berkowitz WF, Todaro LJ, Danishefsky SJ (2008) On the two-component microwave-mediated reaction of isonitriles with carboxylic acids: regarding alleged formimidate carboxylate mixed anhydrides. J Am Chem Soc 130:13222–13224
44. Li X, Yuan Y, Kan C, Danishefsky SJ (2008) Addressing mechanistic issues in the coupling of isonitriles and carboxylic acids: potential routes to peptidic constructs. J Am Chem Soc 130:13225–13227

45. Rao Y, Li XC, Danishefsky SJ (2009) Thio FCMA intermediates as strong acyl donors: a general solution to the formation of complex amide bonds. J Am Chem Soc 131:12924–12926
46. Wu X, Li X, Danishefsky SJ (2009) Thio-mediated two-component coupling reaction of carboxylic acids and isonitriles under mild conditions. Tetrahedron Lett 50:1523–1525
47. Wu X, Yuan Y, Li X, Danishefsky SJ (2009) Thio-mediated synthesis of derivatized N-linked glycopeptides using isonitrile chemistry. Tetrahedron Lett 50:4666–4669
48. Wu X, Stockdill JL, Wang P, Danishefsky SJ (2010) Total synthesis of cyclosporine: access to N-methylated peptides via isonitrile coupling reactions. J Am Chem Soc 132:4098–4100
49. Wang T, Danishefsky SJ (2013) Solid-phase peptide synthesis and solid-phase fragment coupling mediated by isonitriles. Proc Natl Acad Sci USA 110:11708–11713
50. Subiros-Funosas R, Prohens R, Barbas R, El-Faham A, Albericio F (2009) Oxyma: an efficient additive for peptide synthesis to replace the benzotriazole-based HOBt and HOAt with a lower risk of explosion. Chemistry 15:9394–9403
51. Itoh M (1973) Peptide. IV. Racemization suppression by the use of ethyl 2-hydroximino-2-cyanoacetate and its amide. Bull Chem Soc Jpn 46:2219–2221
52. Cherkupally P, Acosta GA, Nieto-Rodriguez L, Spengler J, Rodriguez H, Khattab SN, El-Faham A, Shamis M, Luxembourg Y, Prohens R, Subiros-Funosas R, Albericio F (2013) K-Oxyma: a strong acylation-promoting, 2-CTC resin-friendly coupling additive. Eur J Org Chem 2013:6372–6378
53. El-Faham A, Subiros Funosas R, Prohens R, Albericio F (2009) COMU: a safer and more effective replacement for benzotriazole-based uronium coupling reagents. Chemistry 15:9404–9416
54. El-Faham A, Albericio F (2007) Novel proton acceptor immonium-type coupling reagents: application in solution and solid-phase peptide synthesis. Org Lett 9:4475–4477
55. Carpino LA, Imazumi H, El-Faham A, Ferrer FJ, Zhang C, Lee Y, Foxman BM, Henklein P, Hanay C, Mugge C, Wenschuh H, Klose J, Beyermann M, Bienert M (2002) The uronium/guanidinium peptide coupling reagents: finally the true uronium salts. Angew Chem Int Ed Engl 41:441–445
56. Subiros-Funosas R, Nieto-Rodriguez L, Jensen KJ, Albericio F (2013) COMU: scope and limitations of the latest innovation in peptide acyl transfer reagents. J Pept Sci 19:408–414
57. El-Faham A, Khattab SN, Subiros-Funosas R, Albericio F (2014) BOP-OXy, BOP-OBt, and BOP-OAt: novel organophosphinic coupling reagents useful for solution and solid-phase peptide synthesis. J Pept Sci 20:1–6
58. Palasek SA, Cox ZJ, Collins JM (2007) Limiting racemization and aspartimide formation in microwave-enhanced Fmoc solid phase peptide synthesis. J Pept Sci 13:143–148
59. Bacsa B, Horvati K, Bosze S, Andreae F, Kappe CO (2008) Solid-phase synthesis of difficult peptide sequences at elevated temperatures: a critical comparison of microwave and conventional heating technologies. J Org Chem 73:7532–7542
60. 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
61. Krchnak V, Flegelova Z, Vagner J (1993) Aggregation of resin-bound peptides during solid-phase peptide-synthesis - prediction of difficult sequences. Int J Pept Prot Res 42:450–454
62. Pawar AP, Dubay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CM (2005) Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. J Mol Biol 350:379–392
63. Tan Z, Shang S, Danishefsky SJ (2011) Rational development of a strategy for modifying the aggregatibility of proteins. Proc Natl Acad Sci USA 108:4297–4302
64. Bacsa B, Bosze S, Kappe CO (2010) Direct solid-phase synthesis of the beta-amyloid (1-42) peptide using controlled microwave heating. J Org Chem 75:2103–2106
65. Finneman JI, Pozzo MJ (2012) Novel approach for optimization of a 'difficult' peptide synthesis by utilizing quantitative reaction monitoring assays. J Pept Sci 18:511–518

66. Delgado M, Janda K (2002) Polymeric supports for solid phase organic synthesis. Curr Org Chem 6:1031–1043
67. Côté S (2005) New polyether based monomers and highly cross-linked amphiphile resins. WO2005012277
68. Martin FG, Albericio F (2008) Solid supports for the synthesis of peptides - from the first resin used to the most sophisticated in the market. Chim Oggi 26:29–34
69. Garcia-Martin F, White P, Steinauer R, Cote S, Tulla-Puche J, Albericio F (2006) The synergy of ChemMatrix resin and pseudoproline building blocks renders RANTES, a complex aggregated chemokine. Biopolymers 84:566–575
70. Garcia-Martin F, Quintanar-Audelo M, Garcia-Ramos Y, Cruz LJ, Gravel C, Furic R, Cote S, Tulla-Puche J, Albericio F (2006) ChemMatrix, a poly(ethylene glycol)-based support for the solid-phase synthesis of complex peptides. J Comb Chem 8:213–220
71. Collins JM, Porter KA, Singh SK, Vanier GS (2014) High-efficiency solid phase peptide synthesis (HE-SPPS). Org Lett 16:940–943
72. Sato T, Saito Y, Aimoto S (2005) Synthesis of the C-terminal region of opioid receptor like 1 in an SDS micelle by the native chemical ligation: effect of thiol additive and SDS concentration on ligation efficiency. J Pept Sci 11:410–416
73. Zheng JS, Yu M, Qi YK, Tang S, Shen F, Wang ZP, Xiao L, Zhang L, Tian CL, Liu L (2014) Expedient total synthesis of small to medium-sized membrane proteins via Fmoc chemistry. J Am Chem Soc 136:3695–3704
74. Liu L-P, Deber CM (1998) Guidelines for membrane protein engineering derived from de novo designed model peptides. Pept Sci 47:41–62
75. Valiyaveetil FI, MacKinnon R, Muir TW (2002) Semisynthesis and folding of the potassium channel KcsA. J Am Chem Soc 124:9113–9120
76. Subirós-Funosas R, El-Faham A, Albericio F (2011) Aspartimide formation in peptide chemistry: occurrence, prevention strategies and the role of N-hydroxylamines. Tetrahedron 67:8595–8606
77. Subiros-Funosas R, El-Faham A, Albericio F (2012) Use of Oxyma as pH modulatory agent to be used in the prevention of base-driven side reactions and its effect on 2-chlorotrityl chloride resin. Biopolymers 98:89–97
78. Michels T, Dolling R, Haberkorn U, Mier W (2012) Acid-mediated prevention of aspartimide formation in solid phase peptide synthesis. Org Lett 14:5218–5221
79. Huang H, Rabenstein DL (1999) A cleavage cocktail for methionine-containing peptides. J Pept Res 53:548–553
80. Houghten RA, Li CH (1978) Studies on pituitary prolactin. 41. Reaction of ovine prolactin with 2-(2-nitrophenylsulfenyl)-3-methyl-3′-bromoindolemine. Int J Pept Prot Res 11:345–352
81. Houghten RA, Li CH (1979) Reduction of sulfoxides in peptides and proteins. Anal Biochem 98:36–46
82. Houghten RA, Choh HL (1983) Reduction of sulfoxides in peptides and proteins. Method Enzymol 91:549–559
83. Yajima H, Fujii N, Funakoshi S, Watanabe T, Murayama E, Otaka A (1988) New strategy for the chemical synthesis of proteins. Tetrahedron 44:805–819
84. Mende F, Seitz O (2011) 9-Fluorenylmethoxycarbonyl-based solid-phase synthesis of peptide alpha-thioesters. Angew Chem Int Ed Engl 50:1232–1240
85. Johnson EC, Kent SB (2006) Insights into the mechanism and catalysis of the native chemical ligation reaction. J Am Chem Soc 128:6640–6646
86. Curtius T (1904) Verkettung von amidosäuren I. Abhandlung. J Prakt Chem 70:57–72
87. Yanaihara N, Yanaihara C, Dupuis G, Beacham J, Camble R, Hofmann K (1969) Polypeptides. XLII. Synthesis and characterization of seven fragments spanning the entire sequence of ribonuclease T1. J Am Chem Soc 91:2184–2185
88. Felix AM, Merrifield RB (1970) Azide solid phase peptide synthesis. J Am Chem Soc 92:1385–1391

89. Romovacek H, Dowd SR, Kawasaki K, Nishi N, Hofmann K (1979) Studies on polypeptides. 54. The synthesis of a peptide corresponding to positions 24-104 of the peptide chain of ribonuclease T1. J Am Chem Soc 101:6081–6091
90. Fujii N, Yajima H (1981) Total synthesis of bovine pancreatic ribonuclease A. Part 6. Synthesis of RNase A with full enzymic activity. J Chem Soc Perkin Trans 1:831–841
91. Fischer E (1905) Synthese von polypeptiden. IX. Chloride der aminosäuren und ihrer acylderivate. Ber Dtsch Chem Ges 38:605–619
92. Sheehan JC, Hess GP (1955) A new method of forming peptide bonds. J Am Chem Soc 77:1067–1068
93. Fang G-M, Cui H-K, Zheng J-S, Liu L (2010) Chemoselective ligation of peptide phenyl esters with N-terminal cysteines. ChemBioChem 11:1061–1065
94. Wan Q, Chen J, Yuan Y, Danishefsky SJ (2008) Oxo-ester mediated native chemical ligation: concept and applications. J Am Chem Soc 130:15814–15816
95. Li XQ, Kawakami T, Aimoto S (1998) Direct preparation of peptide thioesters using an Fmoc solid-phase method. Tetrahedron Lett 39:8669–8672
96. Clippingdale AB, Barrow CJ, Wade JD (2000) Peptide thioester preparation by Fmoc solid phase peptide synthesis for use in native chemical ligation. J Pept Res 6:225–234
97. Bu XZ, Xie GY, Law CW, Guo ZH (2002) An improved deblocking agent for direct Fmoc solid-phase synthesis of peptide thioesters. Tetrahedron Lett 43:2419–2422
98. Manabe S, Sugioka T, Ito Y (2007) Facile peptide thioester synthesis via solution-phase tosylamide preparation. Tetrahedron Lett 48:849–853
99. Yamamoto N, Tanabe Y, Okamoto R, Dawson PE, Kajihara Y (2008) Chemical synthesis of a glycoprotein having an intact human complex-type sialyloligosaccharide under the Boc and Fmoc synthetic strategies. J Am Chem Soc 130:501–510
100. Flemer S (2009) Efficient method of circumventing insolubility problems with fully protected peptide carboxylates via in situ direct thioesterification reactions. J Pept Sci 15:693–696
101. Kan C, Trzupek JD, Wu B, Wan Q, Chen G, Tan ZP, Yuan Y, Danishefsky SJ (2009) Toward homogeneous erythropoietin: chemical synthesis of the Ala1−Gly28 glycopeptide domain by "alanine" ligation. J Am Chem Soc 131:5438–5443
102. Botti P, Villain M, Manganiello S, Gaertner H (2004) Native chemical ligation through in situ O to S acyl shift. Org Lett 6:4861–4864
103. Chiang KP, Jensen MS, McGinty RK, Muir TW (2009) A semisynthetic strategy to generate phosphorylated and acetylated histone H2B. ChemBioChem 10:2182–2187
104. George EA, Novick RP, Muir TW (2008) Cyclic peptide inhibitors of staphylococcal virulence prepared by Fmoc-based thiolactone peptide synthesis. J Am Chem Soc 130:4914–4924
105. Zheng JS, Cui HK, Fang GM, Xi WX, Liu L (2010) Chemical protein synthesis by kinetically controlled ligation of peptide O-esters. ChemBioChem 11:511–515
106. Hou W, Zhang X, Li F, Liu CF (2011) Peptidyl N, N-bis(2-mercaptoethyl)-amides as thioester precursors for native chemical ligation. Org Lett 13:386–389
107. Richardson JP, Chan CH, Blanc J, Saadi M, Macmillan D (2010) Exploring neoglycoprotein assembly through native chemical ligation using neoglycopeptide thioesters prepared via N → S acyl transfer. Org Biomol Chem 8:1351–1360
108. Katayama H, Hojo H, Shimizu I, Nakahara Y, Nakahara Y (2010) Chemical synthesis of mouse pro-opiomelanocortin(1-74) by azido-protected glycopeptide ligation via the thioester method. Org Biomol Chem 8:1966–1972
109. Erlich LA, Kumar KSA, Haj-Yahya M, Dawson PE, Brik A (2010) N-Methylcysteine-mediated total chemical synthesis of ubiquitin thioester. Org Biomol Chem 8:2392–2396
110. Tsuda S, Shigenaga A, Bando K, Otaka A (2009) N → S acyl-transfer-mediated synthesis of peptide thioesters using anilide derivatives. Org Lett 11:823–826
111. Nakamura K, Kanao T, Uesugi T, Hara T, Sato T, Kawakami T, Aimoto S (2009) Synthesis of peptide thioesters via an N-S acyl shift reaction under mild acidic conditions on an N-4,5-dimethoxy-2-mercaptobenzyl auxiliary group. J Pept Sci 15:731–737

112. Kawakami T, Aimoto S (2009) The use of a cysteinyl prolyl ester (CPE) autoactivating unit in peptide ligation reactions. Tetrahedron 65:3871–3877
113. Kang J, Richardson JP, Macmillan D (2009) 3-Mercaptopropionic acid-mediated synthesis of peptide and protein thioesters. Chem Commun 407–409
114. Kang J, Reynolds NL, Tyrrell C, Dorin JR, Macmillan D (2009) Peptide thioester synthesis through N → S acyl-transfer: application to the synthesis of a beta-defensin. Org Biomol Chem 7:4918–4923
115. Ozawa C, Katayama H, Hojo H, Nakahara Y, Nakahara Y (2008) Efficient sequential segment coupling using N-alkylcysteine-assisted thioesterification for glycopeptide dendrimer synthesis. Org Lett 10:3531–3533
116. Hojo H, Murasawa Y, Katayama H, Ohira T, Nakaharaa Y, Nakahara Y (2008) Application of a novel thioesterification reaction to the synthesis of chemokine CCL27 by the modified thioester method. Org Biomol Chem 6:1808–1813
117. Kawakami T, Aimoto S (2007) Sequential peptide ligation by using a controlled cysteinyl prolyl ester (CPE) autoactivating unit. Tetrahedron Lett 48:1903–1905
118. Hojo H, Onuma Y, Akimoto Y, Nakahara Y, Nakahara Y (2007) N-Alkyl cysteine-assisted thioesterification of peptides. Tetrahedron Lett 48:25–28
119. Nagalingam AC, Radford SE, Warriner SL (2007) Avoidance of epimerization in the synthesis of peptide thioesters using Fmoc protection. Synlett 2517–2520
120. 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
121. Chen G, Warren JD, Chen JH, 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
122. Ollivier N, Dheur J, Mhidia R, Blanpain A, Melnyk O (2010) Bis(2-sulfanylethyl)amino native peptide ligation. Org Lett 12:5238–5241
123. Dheur J, Ollivier N, Vallin A, Melnyk O (2011) Synthesis of peptide alkylthioesters using the intramolecular N, S-acyl shift properties of bis(2-sulfanylethyl)amido peptides. J Org Chem 76:3194–3202
124. Ollivier N, Vicogne J, Vallin A, Drobecq H, Desmet R, El Mahdi O, Leclercq B, Goormachtigh G, Fafeur V, Melnyk O (2012) A one-pot three-segment ligation strategy for protein chemical synthesis. Angew Chem Int Ed Engl 51:209–213
125. Pira SL, Boll E, Melnyk O (2013) Synthesis of peptide thioacids at neutral pH using bis (2-sulfanylethyl)amido peptide precursors. Org Lett 15:5346–5349
126. 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
127. Raibaut L, Seeberger P, Melnyk O (2013) Bis(2-sulfanylethyl)amido peptides enable native chemical ligation at proline and minimize deletion side-product formation. Org Lett 15:5516–5519
128. Ollivier N, Raibaut L, Blanpain A, Desmet R, Dheur J, Mhidia R, Boll E, Drobecq H, Pira SL, Melnyk O (2014) Tidbits for the synthesis of bis(2-sulfanylethyl) amido (SEA) polystyrene resin, SEA peptides and peptide thioesters. J Pept Sci 20:92–97
129. 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 Engl 47:6851–6855
130. Mahto SK, Howard CJ, Shimko JC, Ottesen JJ (2011) A reversible protection strategy to improve Fmoc-SPPS of peptide thioesters by the N-acylurea approach. ChemBioChem 12:2488–2494
131. 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 Engl 50:7645–7649

132. Fang GM, Wang JX, Liu L (2012) Convergent chemical synthesis of proteins by ligation of peptide hydrazides. Angew Chem Int Ed Engl 51:10347–10350
133. Li YM, Yang MY, Huang YC, Li YT, Chen PR, Liu L (2012) Ligation of expressed protein alpha-hydrazides via genetic incorporation of an alpha-hydroxy acid. ACS Chem Biol 7:1015–1022
134. Zheng JS, Tang S, Huang YC, Liu L (2013) Development of new thioester equivalents for protein chemical synthesis. Acc Chem Res 46:2475–2484
135. Zheng JS, Tang S, Qi YK, Wang ZP, Liu L (2013) Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. Nat Protoc 8:2483–2495
136. Kimura T, Takai M, Masui Y, Morikawa T, Sakakibara S (1981) Strategy for the synthesis of large peptides: an application to the total synthesis of human parathyroid hormone [hPTH(1–84)]. Biopolymers 20:1823–1832
137. Sakakibara S (1995) Synthesis of large peptides in solution. Biopolymers 37:17–28
138. 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
139. Schnolzer M, Kent S (1992) Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. Science 256:221–225
140. Crich D, Banerjee A (2007) Native chemical ligation at phenylalanine. J Am Chem Soc 129:10064–10065
141. Chen J, Wan Q, Yuan Y, Zhu J, Danishefsky SJ (2008) Native chemical ligation at valine: a contribution to peptide and glycopeptide synthesis. Angew Chem Int Ed Engl 47:8521–8524
142. Haase C, Rohde H, Seitz O (2008) Native chemical ligation at valine. Angew Chem Int Ed Engl 47:6807–6810
143. Ajish Kumar KS, Haj-Yahya M, Olschewski D, Lashuel HA, Brik A (2009) Highly efficient and chemoselective peptide ubiquitylation. Angew Chem Int Ed Engl 48:8090–8094
144. Yang R, Pasunooti KK, Li F, Liu XW, Liu CF (2009) Dual native chemical ligation at lysine. J Am Chem Soc 131:13592–13593
145. Chen J, Wang P, Zhu J, Wan Q, Danishefsky SJ (2010) A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides. Tetrahedron 66:2277–2283
146. 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 Engl 49:10149–10153
147. Harpaz Z, Siman P, Kumar KS, Brik A (2010) Protein synthesis assisted by native chemical ligation at leucine. ChemBioChem 11:1232–1235
148. Tan Z, Shang S, Danishefsky SJ (2010) Insights into the finer issues of native chemical ligation: an approach to cascade ligations. Angew Chem Int Ed Engl 49:9500–9503
149. Yang RL, Pasunooti KK, Li FP, Liu XW, Liu CF (2010) Synthesis of K48-linked diubiquitin using dual native chemical ligation at lysine. Chem Commun 46:7199–7201
150. 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
151. Shang S, Tan Z, Dong S, Danishefsky SJ (2011) An advance in proline ligation. J Am Chem Soc 133:10784–10786
152. Siman P, Karthikeyan SV, Brik A (2012) Native chemical ligation at glutamine. Org Lett 14:1520–1523
153. Van de Vijver P, Scheer L, van Beijnum J, Griffioen A, Hackeng TM (2012) Application of an omonasteine ligation strategy for the total chemical synthesis of the BRD7 bromodomain. Chem Commun 48:9403–9405
154. Guan XY, Drake MR, Tan ZP (2013) Total synthesis of human galanin-like peptide through an aspartic acid ligation. Org Lett 15:6128–6131
155. Malins LR, Cergol KM, Payne RJ (2013) Peptide ligation-desulfurization chemistry at arginine. ChemBioChem 14:559–563

156. Thompson RE, Chan B, Radom L, Jolliffe KA, Payne RJ (2013) Chemoselective peptide ligation-desulfurization at aspartate. Angew Chem Int Ed Engl 52:9723–9727
157. Cergol KM, Thompson RE, Malins LR, Turner P, Payne RJ (2014) One-pot Peptide ligation-desulfurization at glutamate. Org Lett 16:290–293
158. Malins LR, Cergol KM, Payne RJ (2014) Chemoselective sulfenylation and peptide ligation at tryptophan. Chem Sci 5:260–266
159. 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 USA 108:5986–5989
160. Bode JW, Fox RM, Baucom KD (2006) Chemoselective amide ligations by decarboxylative condensations of N-alkylhydroxylamines and alpha-ketoacids. Angew Chem Int Ed Engl 45:1248–1252
161. Pattabiraman VR, Ogunkoya AO, Bode JW (2012) Chemical protein synthesis by chemoselective alpha-ketoacid-hydroxylamine (KAHA) ligations with 5-oxaproline. Angew Chem Int Ed Engl 51:5114–5118
162. Ju L, Bode JW (2009) A general strategy for the preparation of C-terminal peptide alpha-ketoacids by solid phase peptide synthesis. Org Biomol Chem 7:2259–2264
163. Aimoto S (2001) Contemporary methods for peptide and protein synthesis. Curr Org Chem 5:45–87
164. Chen G, Wan Q, Tan ZP, Kan C, Hua ZH, Ranganathan K, Danishefsky SJ (2007) Development of efficient methods for accomplishing cysteine-free peptide and glycopeptide coupling. Angew Chem Int Ed Engl 46:7383–7387
165. 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
166. Zhang Y, Xu C, Lam HY, Lee CL, Li X (2013) Protein chemical synthesis by serine and threonine ligation. Proc Natl Acad Sci USA 110:6657–6662
167. Raibaut L, Ollivier N, Melnyk O (2012) Sequential native peptide ligation strategies for total chemical protein synthesis. Chem Soc Rev 41:7001–7015
168. Bang D, Kent SB (2004) A one-pot total synthesis of crambin. Angew Chem Int Ed Engl 43:2534–2538
169. Sohma Y, Pentelute BL, Whittaker J, Hua QX, Whittaker LJ, Weiss MA, Kent SBH (2008) Comparative properties of insulin-like growth factor 1 (IGF-1) and [Gly7D-Ala]IGF-1 prepared by total chemical synthesis. Angew Chem Int Ed Engl 47:1102–1106
170. Bang D, Pentelute BL, Kent SB (2006) Kinetically controlled ligation for the convergent chemical synthesis of proteins. Angew Chem Int Ed Engl 45:3985–3988
171. Mandal K, Pentelute BL, Bang D, Gates ZP, Torbeev VY, Kent SBH (2012) Design, total chemical synthesis, and X-ray structure of a protein having a novel linear-loop polypeptide chain topology. Angew Chem Int Ed Engl 51:1481–1486
172. Tam JP, Wu CR, Liu W, Zhang JW (1991) Disulfide bond formation in peptides by dimethylsulfoxide - scope and applications. J Am Chem Soc 113:6657–6662
173. Sieber P, Kamber B, Hartmann A, Johl A, Riniker B, Rittel W (1977) Total synthesis of human insulin. IV. Description of the final steps (author's transl). Helv Chim Acta 60:27–37
174. Isidro-Llobet A, Álvarez M, Albericio F (2009) Amino acid-protecting groups. Chem Rev 109:2455–2504
175. Mayer JP, Zhang F, DiMarchi RD (2007) Insulin structure and function. Pept Sci 88:687–713
176. Belgi A, Hossain MA, Tregear GW, Wade JD (2011) The chemical synthesis of insulin: from the past to the present. Immunol Endocr Metab Agents Med Chem 11:40–47
177. Liu F, Luo EY, Flora DB, Mezo AR (2014) A synthetic route to human insulin using isoacyl peptides. Angew Chem Int Ed Engl 53:3983–3987
178. Rabanal F, DeGrado WF, Dutton PL (1996) Use of 2,2′-dithiobis(5-nitropyridine) for the heterodimerization of cysteine containing peptides. Introduction of the 5-nitro-2-pyridine-sulfenyl group. Tetrahedron Lett 37:1347–1350

179. Sohma Y, Sasaki M, Hayashi Y, Kimura T, Kiso Y (2004) Novel and efficient synthesis of difficult sequence-containing peptides through O-N intramolecular acyl migration reaction of O-acyl isopeptides. Chem Commun 124–125
180. Postma TM, Albericio F (2013) N-Chlorosuccinimide, an efficient reagent for on-resin disulfide formation in solid-phase peptide synthesis. Org Lett 15:616–619
181. Postma TM, Giraud M, Albericio F (2012) Trimethoxyphenylthio as a highly labile replacement for tert-butylthio cysteine protection in Fmoc solid phase synthesis. Org Lett 14:5468–5471
182. Gough JD, Lees WJ (2005) Effects of redox buffer properties on the folding of a disulfide-containing protein: dependence upon pH, thiol pKa, and thiol concentration. J Biotechnol 115:279–290
183. Patel AS, Lees WJ (2012) Oxidative folding of lysozyme with aromatic dithiols, and aliphatic and aromatic monothiols. Bioorgan Med Chem 20:1020–1028
184. Okumura M, Saiki M, Yamaguchi H, Hidaka Y (2011) Acceleration of disulfide-coupled protein folding using glutathione derivatives. FEBS J 278:1137–1144
185. Okumura M, Shimamoto S, Hidaka Y (2012) A chemical method for investigating disulfide-coupled peptide and protein folding. FEBS J 279:2283–2295

Top Curr Chem (2015) 363: 193–226
DOI: 10.1007/128_2015_629
© Springer International Publishing Switzerland 2015
Published online: 12 April 2015

Chemical and Biological Tools for the Preparation of Modified Histone Proteins

Cecil J. Howard, Ruixuan R. Yu, Miranda L. Gardner, John C. Shimko, and Jennifer J. Ottesen

Abstract Eukaryotic chromatin is a complex and dynamic system in which the DNA double helix is organized and protected by interactions with histone proteins. This system is regulated through a large network of dynamic post-translational modifications (PTMs) which ensure proper gene transcription, DNA repair, and other processes involving DNA. Homogenous protein samples with precisely characterized modification sites are necessary to understand better the functions of modified histone proteins. Here, we discuss sets of chemical and biological tools developed for the preparation of modified histones, with a focus on the appropriate choice of tool for a given target. We start with genetic approaches for the creation of modified histones, including the incorporation of genetic mimics of histone modifications, chemical installation of modification analogs, and the use of the expanded genetic code to incorporate modified amino acids. We also cover the chemical ligation techniques which have been invaluable in the generation of complex modified histones indistinguishable from their natural counterparts. We end with a prospectus on future directions.

Keywords Chemical ligation • protein chemistry • expanded genetic code • post-translational modifications • histones • nucleosomes • chromatin

Contents

C.J. Howard, R.R. Yu, M.L. Gardner, J.C. Shimko, and J.J. Ottesen (✉)
Department of Chemistry and Biochemistry and The Ohio State Biochemistry Program, The Ohio State University, Columbus, OH 43210, USA
e-mail: ottesen.1@osu.edu

1 Introduction

Eukaryotic chromatin is a complex and dynamic system in which the DNA double helix is organized and protected through interactions with histone proteins to form nucleosomes. These further interact to form higher order chromatin structures. This serves to stabilize and sequester DNA, while also regulating interactions with biologically relevant functional partners. At the core of this regulatory system are the dynamic post-translational modifications of histone proteins which help control gene transcription, DNA repair, and a host of other cellular functions. The nucleosome is the unit structure of chromatin (Fig. 1). In a canonical nucleosome, there are four primary histone proteins – H3, H4, H2A, and H2B. Two copies of each of H3 and H4 form the $H3_2/H4_2$ tetramer; whereas one copy each of H2A and H2B form the H2A/H2B dimer. One tetramer and two dimers together form the histone octamer, around which is wrapped ~147 bp of DNA [2]. This structure is elaborated by incorporation of histone variants. For example, there are three major H3 variants (H3.1, H3.2, and H3.3) in human chromatin, whereas the H3 variant CENP-A is found only in centromeric chromatin [3]. Similarly, the dimer may contain variant histones such as H2A variants H2A.X, which plays a role in DNA repair, or H2A.Z, which is implicated in regulation of a variety of cellular functions [4–6]. Specific incorporation of these histone variants is one mechanism by which chromatin function may be dynamically regulated.

Conceptually, the nucleosome may be subdivided into two distinct functional areas: the highly structured histone core, which forms the primary binding surface for DNA, and the histone tails, which project out from the core and are typically unstructured in the context of a mononucleosome. Histone octamers are deposited to form nucleosomes in arrays along the DNA molecule, which along with linker histone H1 can compact into higher order chromatin structures. Taken as a whole, the nucleosome core packages and protects DNA, counteracting the negative charge of the phosphate backbone with the positive charge of the basic histones,

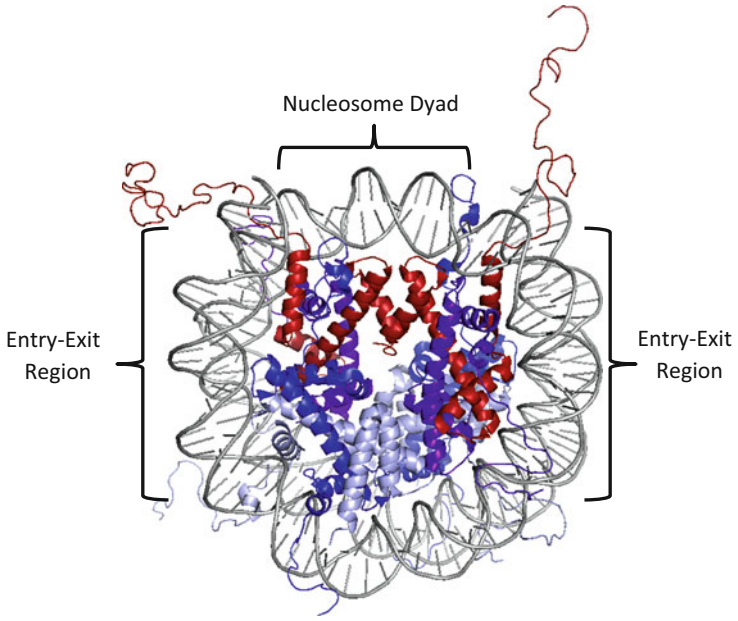


Fig. 1 The structure of the nucleosome [1]. Wrapped DNA is depicted in *gray*. Histone H3 is shown in *dark red*, histone H4 in *purple*, histone H2A in *dark blue*, and histone H2B in *light blue*. The nucleosome dyad and entry-exit regions are labeled for clarity

and the act of wrapping physically occludes the DNA from interaction with cellular partners. The histone tails are poised to coordinate interactions between the nucleosomes and to recruit binding partners to regulate biological activity through complex patterns of modifications. All these functions are dynamically mediated by combinations of histone post-translational modifications which, with a focus on the histone tails, have been described as the "histone code" [7, 8].

Histone proteins are extensively and specifically modified throughout the tail and core regions. Our knowledge of the nature and number of histone modifications is constantly expanding, and an exhaustive list of modifications is an ever-moving target [9, 10]. In part, lysines are commonly mono-, di-, or tri-methylated, acetylated, ubiquitinated, sumoylated, biotinylated, formylated, or crotonylated [11–13]. Arginine can be methylated or converted to citrulline. Ser, Thr, Tyr, and His may be phosphorylated, and Ser and Thr are glycosylated [14]. In vivo, these modifications are coordinated by an elaborate interplay of regulatory enzymes. In vitro, histones have served as high-value targets to develop protein chemistry tools to generate homogenous and precisely-modified histone proteins because of the number and importance of PTMs in human health and disease. These modified proteins can then be refolded back into functional nucleosomes and nucleosome arrays to determine the molecular mechanisms by which these numerous modifications function.

Here, we discuss sets of chemical and biological tools for the creation of modified histones, focusing on the appropriate choice of tool for a given target. There have been several excellent and detailed reviews on the rapidly expanding field of chemical ligation chemistry as applied to histone proteins [15–18]. We hope to find a unique niche with a general overview designed for those new to the field, with a focus on practical aspects of design and selected case studies, rather than an exhaustive survey. We discuss genetic approaches to modified histones, including the incorporation of genetic mimics of histone modifications, chemical installation of residue analogs, and the use of expanded genetic code techniques to incorporate modified amino acids. Next, we cover the chemical ligation techniques which have been invaluable in the generation of complex modified histones which are indistinguishable from the natural counterparts. We discuss a variety of ligation approaches developed for the production of these designer histones and chromatin. We end with a perspective on future directions of synthetic chromatin in living systems.

2 Genetic Approaches for Modified Histones

2.1 Genetic Mimics of Histone Modifications

Techniques in chemical biology allow unparalleled control over each residue of a protein, leading in the ideal case to the generation of a chemically modified protein that is otherwise indistinguishable from the native counterpart. However, the simplest approach to studying a histone modification is the introduction of one of the 20 natural amino acids that mimics the features of the modified amino acid of interest (Fig. 2). For example, glutamine has been used to substitute for acetylated lysine, and arginine for constitutively unmodified lysine. Glutamate or aspartate is often substituted for phosphorylated serine, threonine, or tyrosine, whereas alanine is used to replicate the unmodified residue.

These approaches have two primary advantages. First, using amino acids that are naturally available allows access to incredibly powerful genetic tools to screen for phenotypic effects of a histone modification [19]. Many early leads on functionally significant histone residues emerged out of large-scale mutational screens in yeast, and were later confirmed by mass spectrometry or other studies [20]. Of note, because these mimics introduce a completely different chemical moiety than either the modified or unmodified states, the sites of mimic incorporation are incapable of undergoing dynamic modification by histone modifying enzymes such as histone acetyltransferases, methyltransferases, or histone deacetylases, even in a cellular context. As such, observed changes may be caused by a substitution mimicking either the modified or unmodified state, or through restriction of dynamic modification at the static residue. Second, nucleosomes are easily refolded and reconstituted from recombinant histone proteins expressed in and purified from *E. coli* [21]. Any laboratory with expertise in recombinant protein expression can

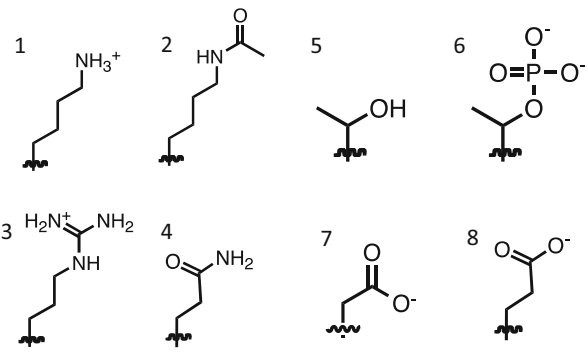


Fig. 2 Side chains relevant to genetic mimics of modifications. (1) Unmodified lysine. (2) Acetylated lysine. (3) Arginine, used to mimic constitutively unmodified lysine. (4) Glutamine, commonly used to mimic constitutively acetylated lysine. (5) Threonine. (6) Phosphorylated threonine. (7) Aspartate. (8) Glutamate, commonly used to mimic constitutively phosphorylated residues

generate large quantities of histones bearing mimics of acetylation or phosphorylation using standard techniques. This opens up the use of biochemical or biophysical techniques which require milligram quantities of a histone, such as crystallography, to assess the role of a modification [22–24].

When employing genetic substitution to mimic a modified amino acid, it is important to confirm that the mimic alters the same properties of a nucleosome as the modification that is being studied. In some cases, such as the effect of lysine acetylation in the H3 and H4 tails on chromatin compaction, Gln often appears to replicate the effects of lysine acetylation, suggesting that neutralization of the positive charge of an unmodified lysine, coupled with potential hydrogen bonding capabilities of the Gln amide, is sufficient to replicate acetylation [25, 26]. These results are contradicted by other studies suggesting that Gln mimics do not replicate the effect of precisely acetylated histone tails in compaction [27]. The study of histone modifications located in structured histone-DNA interfaces in the core of the nucleosome has offered a more nuanced look at these mimics. In the dyad region, H3-K115ac and H3-K122ac have been shown to destabilize the nucleosome structure [28]. However, H3-K115Q and H3-K122Q do not replicate this effect and, in fact, may stabilize the nucleosome slightly as assessed by competitive nucleosome reconstitution. Phosphorylation of H3-T118, also in the nucleosome dyad region, significantly destabilizes the nucleosome [29] 40-fold relative to unmodified nucleosome standards and, in fact, can support a stable altered nucleosome structure [30]. H3-T118E as a mimic of phosphorylation does not have these effects, suggesting that negative charge is insufficient to replicate phosphorylation in the context of these functions.

Within the nucleosome entry-exit region, where DNA begins to contact the histone surface, the effectiveness of mimics is also ambiguous. Acetylation of H3-K56 has been demonstrated to increase DNA unwrapping from the histone octamer and to enhance transcription factor binding [31–33]. Incorporation of Gln

as H3-K56Q qualitatively enhances DNA unwrapping, but does not quantitatively reproduce the effect of acetylation. However, as Gln does enhance DNA unwrapping, it may be sufficient to mimic this effect, depending on the precision required. These studies suggest that the ease of production of these natural amino acid mimics must be balanced against the elements of the modified amino acid that are essential for function. Further, mimics are a crude tool for some modifications, as there are no good mimics to distinguish, for instance, mono-, di-, and/or tri-methylation of lysine. These considerations are crucial to choose the tool appropriate to the task. Because each of the more elaborate methods developed for more precise replication of a modified residue also requires extra chemical steps or processing, it is reassuring to know that the extra effort and expense of modified histone preparation is justified by the improved accuracy and reliability of results.

2.2 Codon Suppression: Expanded Genetic Code Approaches to Modified Histones

Expanded genetic code approaches have been developed to site specifically insert unnatural amino acids at the amber codon (UAA) using orthogonal aminoacyl-tRNA synthetases (aaRS) and tRNA (Fig. 3). These provide enhanced ability to use molecular biology approaches to generate modified histone proteins containing a wide range of modifications from naturally occurring PTMs to PTM mimics, fluorescent amino acids, and amino acids modified to alter their intrinsic chemistry. These subjects have been extensively discussed elsewhere [34, 35]; here, we do not attempt an exhaustive review but highlight a few topics which span the diversity of applications in histone proteins.

2.2.1 Encoded Lysine Modifications

Given the number of histone modifications that occur on lysine, the discovery of the pyrrolysine incorporation machinery in methanogens and subsequent development of artificial pyrrolysyl-tRNA synthetase (PylRS)/tRNA pairs for efficient incorporation of modified lysine residues has had a tremendous impact on the histone modification field [36, 37]. The synthetase has been effective at incorporating several lysine variants chemically similar to pyrrolysine, for example with a modification that includes an amide bond at the ε-amine of the lysine side chain (Fig. 3, compounds 1–4). This is exemplified by the classic work of Neumann and coworkers in which they demonstrated the genetic incorporation of acetyllysine into H3-K56ac, located in the entry-exit region of the nucleosome [32]. Milligram quantities of uniformly acetylated histone H3 were generated using this genetic encoding system, which enabled bulk and single molecule FRET experiments that demonstrated an increase of DNA unwrapping and SWI/SNF-dependent chromatin

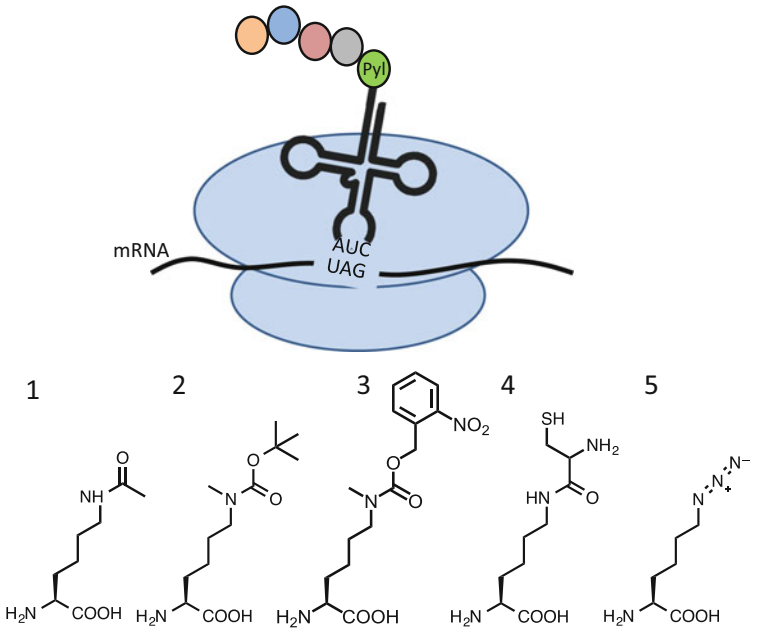


Fig. 3 *Top*: Schematic for expanded genetic code incorporation of lysine mimics by modified pyrrolysine incorporation machinery. *Bottom*: Representative modified amino acids incorporated using expanded genetic code techniques. (1) Acetylated lysine. (2) Boc-protected N^{ε}-methyllysine. (3) Photocaged N^{ε}-methyllysine. (4) N^{ε}(Cys)-Lysine. (5) Azidonorleucine

remodeling upon acetylation without destabilizing the nucleosome as assessed by salt-induced dissociation studies, or affecting higher order chromatin structures in nucleosome arrays. Genetically incorporated acetyllysine has been used to probe the role of lysine acetylation at several positions in the nucleosome core. In one key study, H3 acetylated at K122 was used to elucidate the role of H3-K122ac in transcriptional activation, partially through chromatin assembled in vitro with recombinantly expressed acetylated histone H3 [38]. An additional study of note used genetic incorporation of acetyllysine at H3-K64 to demonstrate that acetylation at this residue increased chromatin remodeling by Chd1, but not RSC, and additionally caused destabilization of the nucleosome as assessed by salt dependence of nucleosome dissociation and competitive nucleosome reconstitution [39]. In each case, the ability to easily generate sufficient quantities of uniformly acetylated histone proteins for nucleosome reconstitution was essential for the success of the study.

The genetic incorporation system used for acetylated lysines has been further developed for site-specific incorporation of methylated lysines. The development of a synthetase with specificity for methylated lysine over unmodified lysine is challenging because of their structural similarity. Therefore, methylated lysines are typically introduced as a protected derivative that is deprotected to reveal the methylated species. For example, Chin and coworkers generated H3

monomethylated at K9 through introduction of a Boc-protected monomethyllysine that could be deprotected with 2% TFA after purification (Fig. 3, compound 2) [40]. A particularly elegant example is the genetic encoding of a photocaged N^{ε}-methyllysine (Fig. 3, compound 3) by Liu and coworkers which can be deprotected by photolysis under mild conditions [41]. This opens the intriguing possibility of the dynamic introduction of methyllysine in living cells. More esoteric lysine modifications, including propionyl-, butyryl-, or crotonyl-lysine have been found in histone proteins [42] as markers of active regions of chromatin [11]. PylRS-tRNA pairs have been found to insert each of these modified lysines into histone proteins [43, 44].

Expanded genetic code approaches in combination with other protein chemistry tools have been used to probe site-specific ubiquitination of histones and other proteins. Of note here are two interesting modified lysine variants that can be genetically incorporated as intermediates to ubiquitinated proteins. Lysine in which cysteine has been coupled to the ε-amine (Fig. 3, compound 5) can be genetically incorporated, and then used as an avenue for native chemical ligation onto the lysine side chain [45, 46]. Similarly, azidonorleucine (Fig. 3, compound 5) can serve as a protected lysine derivative, allowing chemistry specifically at this side chain position to introduce diubiquitin [47]. With the rapid pace of ongoing discovery in this area, many more interesting derivatives of lysine suitable for genetic incorporation may be anticipated in the future.

2.2.2 Combined Genetic and Chemical Approaches: Modifications Introduced Through Dehydroalanine

Several researchers have exploited genetic approaches to install chemical moieties that may be converted to the bioorthogonal, reactive dehydroalanine, an entry point for the introduction of a wide variety of modified residues (Fig. 4). Schultz and coworkers introduced phenylselenocysteine, which is susceptible to oxidative elimination to yield a reactive dehydroalanine moiety, at H3-K9. Michael addition of an appropriate thiol reagent results in a thioether analog of a post-translational modification [48]. Liu and coworkers evolved the pyrrolysine incorporation machinery to accept N^{ε}-Cbz-lysine for site-specific incorporation at H3-K9, with protein yields reported at 100 mg/L. Mild oxidation again resulted in conversion to the reactive dehydroalanine, which was further converted into thioether analogs of methylated lysine, acetylated lysine, and phosphocysteine [49].

The site-specific incorporation of dehydroalanine is not restricted to expanded genetic code approaches. Davis and coworkers developed gentle chemical approaches for the conversion of a cysteine residue to dehydroalanine by treatment with 2,5-dibromohexanediamide [50]. Through this approach, they introduced PTM mimics at single cysteine residues introduced at H3-K4, H3-K9, or H3-K79 [51]. Intriguingly, they also demonstrated that cysteine residues could be introduced simultaneously at H3-K4 and H3-K79, which resulted in two dehydroalanine moieties and therefore two PTM mimics installed at separate locations within the

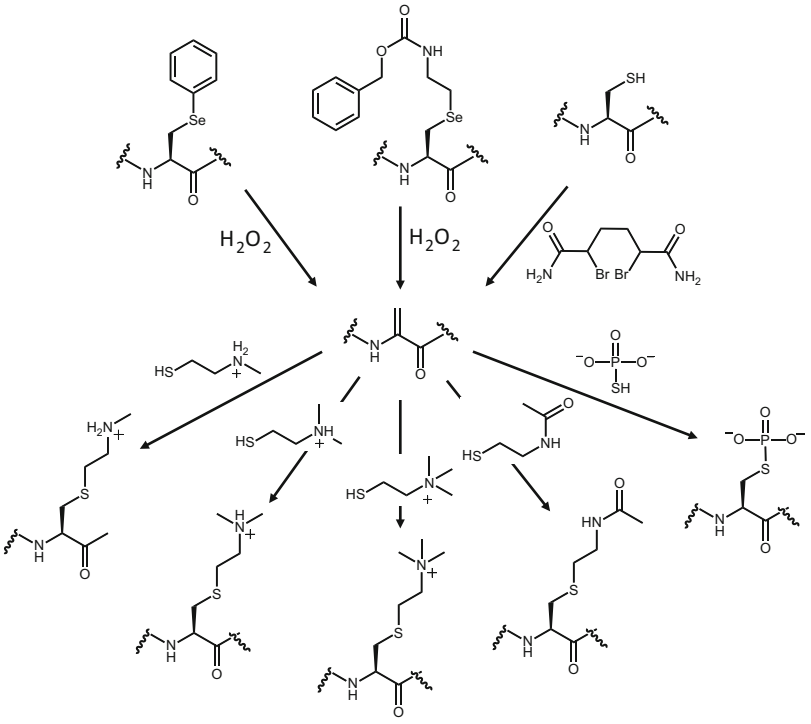


Fig. 4 Schematic for installation of PTM-mimics through a dehydroalanine intermediate. Alkylated selenocysteine or cysteine is converted to the reactive dehydroalanine with loss of stereochemistry. Michael addition of thiol reagent results in the corresponding post-translational modification mimic

histone. However, it should be noted that conversion of cysteine or selenocysteine to dehydroalanine eliminates the chirality of the α-carbon. The PTM mimics are therefore not fully chirally resolved, although there have been suggestions that the inherent chirality of the protein molecule may lead to enrichment of the L-form PTM mimic.

2.2.3 Encoded Phosphoserine

Serine phosphorylation in histones is essential to regulation of several cellular events. Phosphorylation of H3-S10, in particular, is thought to be interdependent with acetylation and methylation of surrounding lysine residues in a network of modification switches [52]. Park and coworkers developed an expanded genetic code approach to the introduction of phosphoserine into proteins in *E. coli*, but with poor expression yields [53]. However, they used H3-S10 as an ideal platform to refine and improve the level of expression for genetically incorporated phosphoserine in the context of a physiologically relevant substrate [54]. They

were able to improve yields to 3 mg/L of culture, a 3,000-fold improvement over previous work, bringing expression to levels useful for production of designer nucleosomes. With the increased expression, they were able to demonstrate the importance of context in histone H3 modification by carrying out histone acetyltransferase (HAT) assays using Gcn5 and the Saga complex on H3 protein alone, octamers, and nucleosome arrays with and without H3-S10ph. They found that H3 alone demonstrated a decrease in acetylation compared to wild-type, histone octamers demonstrated similar levels with phosphorylated and unmodified substrates, whereas acetylation increased with H3-S10ph in nucleosome arrays. These results clearly demonstrate the importance of precisely characterized and regulated substrates in determining the function of histone modifications.

Expanded genetic code systems are clearly very powerful for the generation of modified proteins, particularly in laboratories that are most comfortable with molecular biology approaches to protein production [55]. They expand the reach of precisely regulated modified histone proteins, particularly with acetylated lysine, to a wider audience of chromatin researchers to determine the interplay between histone modifications and other cellular functions. The primary limitation is that without advanced techniques which are only beginning to be explored, only one modification at a time may be introduced into the histone sequence because of the limitation of the available matched stop codon pairs, and of the reduced yield often observed with each incorporation event. This is likely to change as designer organisms that lack amber codons in their genome become more common [56]. As an example, amber codon (UAG) suppression is one method used in nature by select organisms to expand their own genetic code, for example to include either selenocysteine (Sec) or pyrrolysine (Pyl) [57]. Because the native pyrrolysine tRNA/aaRS pair performs amber codon suppression, this is often used as the starting point for the directed co-evolution of new noncanonical lysine analog incorporation systems for efficient and specific amino-acyl charging of a noncanonical amino acids (reviewed in Liu et al. 2010 [34]). The use of tRNA that read quadruplet codons [58] was originally discovered in *Salmonella typhimurium* as a suppressor of a +1 frameshift mutation, and it has been used as a way to make changes to translation without mutations to the ribosomal complex, which are often lethal [59, 60]. This has even given rise to suppression systems in which two modified amino acids can be inserted into a protein.

3 Chemical Installation of PTM Analogs at Single Cysteine Sites

Among the different functional groups presented by the 20 common amino acids, the cysteine sulfhydryl has unique chemical reactivity and is relatively rare in natural protein sequences. With careful selection of a modification reagent, cysteine makes an attractive target for the site-specific incorporation of mimics of post-

translational modifications. Cysteine modification can be extremely powerful due to the ease of use of site-directed mutagenesis for the incorporation of a nonnative cysteine, and the relatively high yields of recombinant histone proteins with these single site substitutions. This same reactivity is commonly used to introduce moieties which allow direct characterization of a modified biomolecule or complex including spin labels [61, 62], fluorophores [63], cross-linkers [64], and foot printing reagents [65] allowing for biophysical characterization of nucleosomes. In this section we discuss methods developed to exploit the reactivity of the cysteine sulfhydryl to introduce PTM mimics that resemble native modifications. We discuss methods that alkylate cysteine directly to generate modified lysine mimics, or that use the unique properties of cysteine for a disulfide "staple" to reversibly link a modification such as ubiquitin to a histone. Methods that exploit the reactivity of cysteine are particularly appropriate for use in histone proteins, because there are no indispensable cysteine residues in the nucleosome. The only native cysteines in a typical *Homo sapiens* nucleosome composed of standard histone variants are found in histone H3 – at position C110 in histone variants H3.1, H3.2, and H3.3, and at position 96 in histone variant H3.1 [66]. These cysteines are commonly replaced by Ser at position 96 and Ala at C110, with minimal to no perturbation of nucleosome structure, function, or dynamics [67, 68]. This allows introduction of a single cysteine elsewhere in the nucleosome without cross-reactivity with native residues.

3.1 MLAs: Methyllysine Analogs

Shokat and coworkers revolutionized the study of methylated lysine residues in histones through the development of simple and elegant cysteine alkylation techniques which are accessible to a wide range of research groups [69]. The aminoethylation of cysteine to generate a lysine analog in which the γ-methylene is replaced with sulfide, with sufficiently similar properties to allow cleavage by trypsin, had been known for decades [70]. Shokat and coworkers made the key recognition that commercially available derivatives of the aminoethylation reagents could be exploited to produce the corresponding mono-, di-, or tri-methylated lysine analogs (MLA) (Fig. 5a). Substituting the methylene with a sulfide decreases the pKa of the residue by 1.1, and increases the length of the sidechain by only 0.28 Å. The method is sufficiently high-throughput that, in the initial study, specific methylated lysine variants were incorporated at H3-K4, H3-K9, H3-K36, H3-K79, and H4-K20. The proteins readily refolded into histone octamer for reconstitution into nucleosomes, and the MLA-modified nucleosomes were able to recapitulate modification-specific recognition by natural binding partners, such as interaction of H3-K9me2 with HP1α. MLAs were directly tested against the native modification in synthetic peptides to assess activity of the methyltransferase SUV39H1, which targets substrates methylated at H3-K9; the MLAs demonstrated equivalent activity

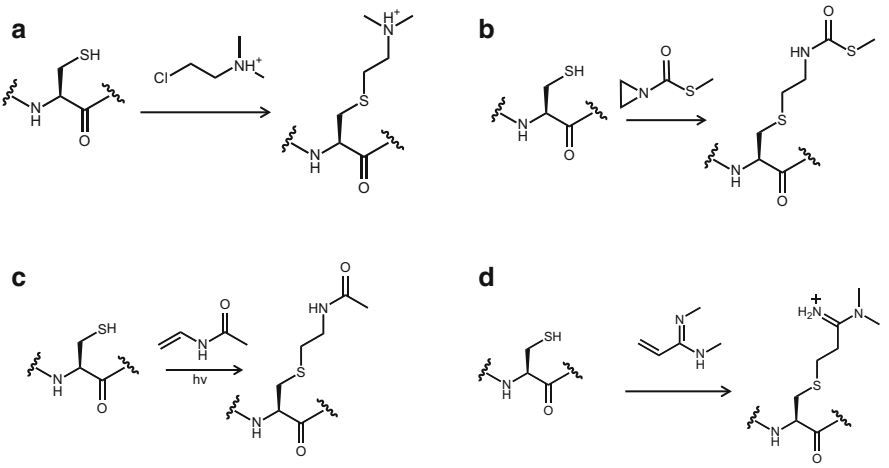


Fig. 5 Modification of unique cysteine sites. (**a**) MLA approach, illustrated for dimethyllysine. (**b**) Generation of the thio-methyl MTCTK acetylation mimic. (**c**) Generation of acetyllysine mimic by thiol-ene chemistry (**d**) Generation of methylated arginine mimic

to the native modification. Further, antibodies raised against the natural H3-K9me, H3-K9me2, H3-K9me3, H4-K20me, H3-K4me3, H3-K36me3, and H3-K79me2 also recognized the MLA equivalents of each of these nucleosomes. The one caveat was that the H3-K9me2 antibody showed fivefold lower affinity, which suggests that the functional equivalency of the MLA is somewhat context-dependent. A survey of the literature suggests that this theme continues; in the majority of cases, MLAs are accessible and cost-efficient mimics of methylated histone function that allow the incorporation of methylated lysine mimics throughout the histone sequence, although concerns are raised in rare cases regarding the impact of the thioether moiety on specific interactions with the MLA.

The elegance and simplicity of the MLA approach (together with the commercial availability of methylated histones prepared using this approach) has enabled discovery and characterization of important functional histone interactions [71]. A few selected examples are as follows. Crystallography requires large quantities of homogenous modified protein to explore the role of a histone modification in the nucleosome. MLAs were used to prepare nucleosomes which included H3-Kc79me2, H4-K20me3, or unmodified histone cores. Analysis of the crystal structures suggested that H3-K79 methylation alters local sidechain structure to partially reveal a hydrophobic pocket on the nucleosome surface. The ability to generate large quantities of nucleosome arrays with these pure MLA-modified histones also allowed experimental characterization of these modifications by analytical ultracentrifugation, which revealed an influence of the tail modification, but not the core, in chromatin compaction [72]. Similarly, MLAs have enabled structural characterization of methylated histone binding partners by nuclear magnetic resonance, for example low affinity binding of methylated H3-K36 by the

histone deacetylase complex Rpd3S [73] or high affinity binding of nucleosomes methylated at H3-K36 to the PWWP domain of LEDGF, which plays an important role in HIV integration [74]. In this study, the ability to prepare homogenous MLA-nucleosomes was essential to the identification of cooperativity between a hydrophobic H3-K36me3 binding site and a basic surface patch that interacts with DNA [73]. While modified peptide pull-down studies were able to identify interactions of H3-K4me3 with the ING4 PHD finger, which is important to tumor suppression, only the ability to prepare MLA-nucleosomes identified that this interaction mediates acetylation on histone H3 within the nucleosome; the ability to prepare these well-defined samples is essential to understanding similar crosstalk between histone modifications [75]. The ability to rapidly generate different modification states of each lysine allows rapid screening across large numbers of differentially modified nucleosomes [76]. To sum up, MLAs have proven to be a valuable chemical tool enabling biochemical and biophysical characterization of histone methylation important for biological function.

3.2 Acetyllysine Analogs via Cysteine Alkylation

Alkylation of cysteine to form MLAs is rapid and high yielding using methylated 2-bromoethylamine derivatives. Unfortunately, reaction with corresponding alkylated reagents provide poor kinetics and poor yields [77]. The Cole Laboratory developed the methylthiocarbonyl-aziridine (MTCA) reagent as an alternate approach to introduce methylthiocarbonyl-thiaLys (MTCTK) acetylation mimics at cysteine sites (Fig. 5b), where the side chain includes the γ-sulfide as well as methylthiocarbonyl in place of the acetyl group. While the additional sulfur does add considerable steric bulk to the modification, specifically modified peptides could be recognized by interaction partners, including the Brdt bromodomain and specific antibodies, although with two- to fourfold lower affinity than the precisely modified acetyllysine, and could stimulate Rtt109 HAT activity, although to a lesser extent than acetyllysine. Although MTCTK may not fully recapitulate the effect of lysine acetylation, the ease of installation may make it a valuable tool in the nucleosome context.

3.3 Thiol-ene Chemistry to Introduce Modification Analogs

An alternate approach to acetyllysine uses free radical induced thiol-ene (or "thiol click" [78]) chemistry to add *N*-vinyl-acetamide at a single cysteine site [79] to generate the corresponding acetyllysine mimic (Fig. 5c). This mimic replicated lysine deacetylation qualitatively, although not fully quantitatively. The mimic was further reconstituted into nucleosome arrays where it appeared to reproduce fully the impact of acetylation at H4-K16 on chromatin compaction. Fujimori and

coworkers used a similar approach to introduce several methylarginine analogs into histone proteins (Fig. 5d) [80]. In the methylated arginine mimics, the ε-nitrogen of the guanidyl group is replaced with a methylene, which perturbs both geometry and polarity of the arginine side chain. However, the mimic was functional across a wide range of methylarginine binding partners, indicating that reproduction of the terminal groups of methylarginine may dominate these interactions.

3.4 Disulfide Stapling

The unique ability of cysteine to form disulfide bonds has been exploited for chemical ligation in histone proteins through the disulfide stapling approach, exemplified by the Muir Laboratory in preparation of a disulfide-linked histone-ubiquitin library [81]. To enable this approach, ubiquitin was expressed as a fusion with an intein domain (Fig. 6). Thiolysis with the 1,2-aminothiol reagent 2-mercaptoethylamine generates a transient intermediate that undergoes rearrangement to form the stable, amide-linked thiol derivative, which can then be activated and incubated with a histone protein with a single cysteine to generate the disulfide-linked ubiquitylated histone. Although this does not recapitulate the native isopeptide linkage of a ubiquitylated protein, nucleosomes with ubiquitin at H2B-K120C were able to stimulate the H3-K79 methyltransferase activity of hDot1L, suggesting the mimic was sufficiently similar for recognition. Further, this linkage enables the dynamic removal of a protein modification via reduction of the disulfide bond, mimicking the dynamic behavior of protein modifications under controlled conditions. The Muir Laboratory exploited this approach to determine the position-dependent effect of ubiquitin in histones H2A and H2B while also demonstrating that H2B ubiquitylation perturbs chromatin compaction [81, 82]. A similar approach was used to prepare histone H4 with sumoylation at H4-K12 in the N-terminal tail, and demonstrated that SUMO-3 inhibits higher order chromatin structure required for chromatin compaction [13].

Each of the approaches described in this section benefits from the overexpression of a recombinant histone protein with a cysteine substitution at the site or sites of interest, with diverse yet elegant methods to elaborate these cysteines into a modification mimic. The power of these techniques lies in the ability to easily generate milligram quantities of modified histones. Consequently, an important limitation of these techniques is that only a single class of modification may be easily introduced within a single molecule because of the requirement for complete reaction of all cysteine residues to yield a homogenous modified sample. Highly modified histone proteins often contain varying numbers and types of post-translational modifications within a single molecule. Cysteine modification approaches are therefore unlikely to reach this level of diversity.

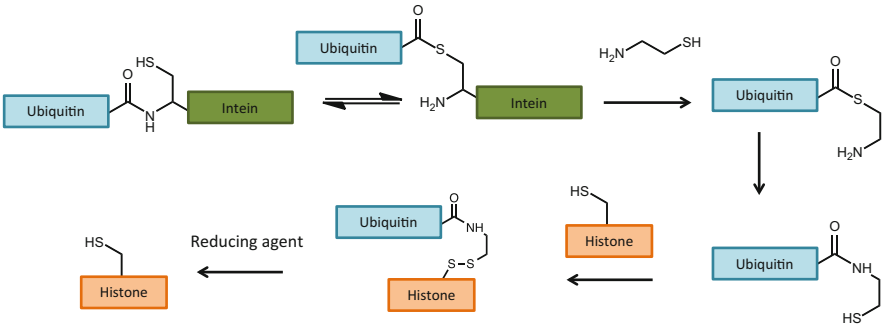


Fig. 6 Ubiquitin-histone disulfide stapling. Ubiquitin-intein fusion is thiolized by 2-mercaptoethylamine, which undergoes rearrangement to form the stable thiol derivative. Activation and disulfide bond formation results in the stapled ubiquitin-histone conjugate for study

4 Chemical Ligation for the Preparation of Modified Histone Proteins

Although each method of preparing modified histones has advantages, to date only chemical ligation has allowed access to histone proteins bearing the complex mixtures of multiple, different modifications representative of native modified proteins. In the general case, chemical ligation is any specific reaction of two peptide or protein segments via a bioorthogonal, chemoselective mechanism to generate a joined product. The most common type of chemical ligation used is native chemical ligation (NCL). In the original form, a peptide with an α-thioester at the C-terminus is reacted with a peptide that contains a 1,2-aminothiol, typically a cysteine residue, at the N-terminus to form a product peptide with a native peptide bond at the juncture [83]. If carried out at a site with a native Cys, the reaction generates a product indistinguishable from the parent sequence. If a peptide segment is generated synthetically, this approach allows full chemical control over each amino acid within the sequence. This allows the introduction of chemical moieties which are challenging to incorporate biologically – for example, amino acids bearing precise post-translational modifications at any desired location in the component peptide sequence. The utility of such an approach to chemically modified histones is immediately apparent.

4.1 Histone Semi-Synthesis by Expressed Protein Ligation: Modifications Near the Histone N-Terminus

Of course, the preparation of component peptides is limited by the typical parameters of solid phase peptide synthesis, such that peptide segments longer than ~40–50 amino acids are challenging through standard techniques. The core histone

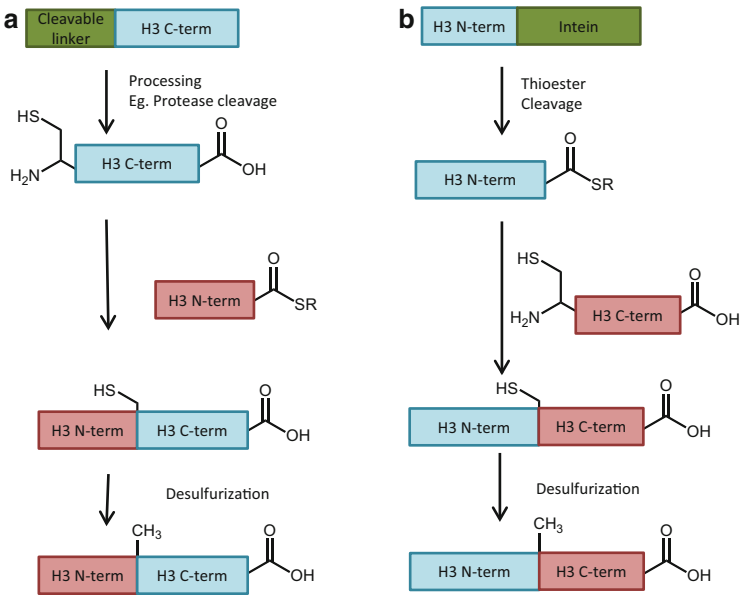


Fig. 7 Expressed protein ligation to generate (**a**) H3 with modifications clustered near the N-terminus or (**b**) H3 with modifications clustered near the C-terminus. Recombinant histone fragments are in *blue*, synthetic peptide fragments are in *red*. Similar schemes apply for other histone proteins, and desulfurization is optional depending on selection of ligation site

proteins are approximately 100–140 amino acids in length, and a single ligation step is not sufficient to generate a full-length protein from fully synthetic peptide segments. However, many interesting histone modifications occur in the N-terminal histone tails. To access modifications in these regions, only the tails need be prepared synthetically to permit the controlled introduction of histone modifications. The remainder, and largest portion, of the protein sequence may be expressed recombinantly in *E. coli*, without any modifications along the sequence, as depicted in Fig. 7a. In general, native chemical ligation carried out with at least one recombinant protein segment has been termed expressed protein ligation (EPL).

A main challenge in employing EPL to generate histones with modifications in the N-terminal tails is the expression of the C-terminal portion of the protein with an N-terminal cysteine. Several methods have been used in histone proteins. The simplest approach is to introduce an initiator Met prior to a Cys, such that removal of the Met by aminopeptidase results in an N-terminal Cys [84]. This expressed protein can then be used directly in a ligation reaction. This approach does not work in all situations; N-terminal Cys is thought to be alkylated by pyruvate in a number of cases, which renders it incapable of acting as a chemoselective ligation site. Another common approach is to follow the initiator Met with a specific proteolysis site such that processing of the expressed protein results in the active N-terminal Cys. This method also allows for the inclusion of an affinity tag N-terminal to the protease sequence. Proteases including Factor Xa [85], TEV protease [86], or

SUMO protease [87] have been successfully used to reveal an N-terminal Cys on recombinant histone fragments with high yield and purity.

Historically, EPL has been used by many groups to study modifications in the N-terminal tails of histones [69, 84, 85, 88–93]. Shogren-Knaak and coworkers were the first to exploit this technique to generate histone H3 containing a phosphorylated serine residue (H3-S10ph, T32C). They demonstrated that the semi-synthetic histone could be refolded into octamer and reconstituted into homogeneous nucleosome arrays, resulting in an increased histone acetyltransferases activity with Gcn5 over WT [85]. However, the physiological Gcn5 complex, SAGA, does not show the same increase in acetyltransferases activity as recombinant Gcn5 and the phosphate modification did not introduce a disruption in the higher-order chromatin structure [88]. Later that same year, He et al. 2003 reported the production of semi-synthetic histone H3 containing a methylated lysine residue (H3-K9me3) and histones H3 and H4 with multiple acetylated lysine residues (H3-K4ac,9ac,14ac,18ac,22ac; H4-K5ac,8ac; and H4-K5ac,8ac,12ac). This group showed that both semi-synthetic modified histones H3 and H4 were able to form tetramer and serve as active substrate for histone-modifying enzymes [84].

In a landmark early study, Shogren-Knaak and coworkers prepared histone H4-K16ac with an R23C substitution by ligating N-terminal peptide H3(1–22)-K16ac to the C-terminal histone fragment H3(23–102, R23C). They found that acetylation of lysine 16 affected higher-order chromatin structure by preventing chromatin compaction in nucleosome arrays, and interrupted protein-histone interaction with adenosine triphosphate-utilizing chromatin remodeling and assembly factor (ACF) [89]. This study clearly demonstrated the potential of semi-synthetic histones to address questions in chromatin structure and function. In another early study, Ferreira and coworkers prepared semi-synthetic, tetra-acetylated H3 (H3-K9ac,14ac,18ac,23ac,S28C) and H4 proteins (H4-K5ac,8ac,12ac,16ac,V21C) to investigate the effects of these modification on chromatin remodeling by Snf2, and found that combinations of acetylation impacted chromatin remodeling rates differentially. H3 tetra-acetylation increased thermal mobility of nucleosomes twofold over unmodified nucleosomes, and increased recruitment of chromatin structure remodeling (RSC) complex 16-fold. H4 tetra-acetylation does not alter thermal mobility of nucleosomes and slightly reduced remodeling activity of both chromodomain helicase DNA binding protein 1 (Chd1) and Isw2 [90]. These results suggested a complex network of unique mechanisms which, although dissimilar, are interrelated to allow for multiple modes of regulation.

Approaches for the semi-synthesis of histones with modified N-terminal tails have reached near-ubiquity in the field, such that semi-synthetic modified histones have even become commercially available (Active Motif). As such, an exhaustive listing of studies with histones prepared using these approaches would be somewhat impractical. However, one recent advance of note is the use of sequential EPL, in which the N-terminal synthetic segment is split into two distinct segments. This allows the preparation of short peptide segments to achieve modifications of histone residues either closer to the core domain (such as H3-R42me2, prepared from recombinant H3(47–135) and synthetic H3(1–28) and H3(29–46)) [94], or

somewhat synthetically challenging such as simultaneous modification of H3-K4me3 and H3-K27me3 prepared from recombinant H3(29–135) and synthetic H3(1–20) and H3(21–28) [95].

4.2 Considerations for Selection of Appropriate Ligation Sites

One challenge in the design of ligation strategies is the selection of appropriate ligation sites selected for the modification of interest. Many different ligation sites have been successfully used to generate modified histone proteins, and selection of an ideal ligation site depends on several factors. (1) The peptides must be synthetically accessible, of a reasonable length, and maintaining desirable solubility properties. (2) The kinetics of the ligation reaction are dependent on the C-terminal residue of the thioester peptide such that β-branched residues should be avoided if possible, and proline is unsuitable [96]. (3) The classical NCL reaction results in a cysteine at the ligation junction. EPL of histone H3 with a synthetic C-terminal peptide carried out by Manohar and coworkers resulted in a cysteine at H3-C110 [28], which is the only native cysteine in nucleosomes in *X. laevis*. For all other EPL schemes it is important to select a ligation site that can permit residue substitution by cysteine if no further chemical steps are planned. H3 residues 25 [97], 28 [90], and 31 [85], as well as H4 residues 23 [98], and 20 [90] have been used as ligation sites resulting in cysteine substitutions.

Because cysteine is rare in histone proteins, these cysteine substitutions result in a non-native histone sequence for most ligation schemes. Methods to convert cysteine to alanine through desulfurization dramatically increase the number of potential ligation sites that can result in native sequences, eliminating any potential effects of the substitution on histone function. Yan and Dawson introduced cysteine desulfurization by treatment with Raney nickel [99]. Early histone ligation protocols including those by the McCafferty Laboratory [84] exploited these approaches successfully to generate native histone sequences. Free radical desulfurization was introduced by Wan and Danishefsky in 2007 [100], using mild reagents fully compatible with ligation conditions. This alternate technique rapidly gained wide-spread adoption within the field and is currently the method of choice for most groups carrying out histone ligation chemistry. Free radical desulfurization requires tris(2-carboxy)phosphine (TCEP) and the water soluble radical initiator 2,2′-azobis[2-(2-imaidazolin-2-yl)propane]dihydrochloride (VA-044US), with a thiol proton source [100]. Several different thiols are compatible with both NCL and with free radical desulfurization, including mercaptoethanesulfonic acid (MESNA), which allows desulfurization to be carried out directly on crude ligation mixtures. However, aromatic thiols such as mercaptophenylacetic acid (MPAA) which are often used for improved ligation kinetics appear to quench

desulfurization, requiring complete removal from a reaction mixture prior to desulfurization [101].

Although the ability to use alanine ligation sites does expand the choices of ligation sites available, not all modifications are conveniently accessible through ligation schemes with these limited sites. The ability to carry out desulfurization, when combined with non-native 1,2-aminothiol derivatives of amino acids, has allowed the expansion of ligation sites to many more residues, reviewed in [102]. For example, penicillamine is a β-thiol derivative of valine commercially available in Fmoc- and Boc-protected forms, including the thiazolidine derivative required for complex ligation schemes discussed in Sect. 4.4. Ligation with an N-terminal penicillamine followed by desulfurization thus generates a valine residue at the ligation site valine [103]. Similar approaches have allowed ligation at an expanding number of amino acid sites including phenylalanine [104], leucine [47], and lysine [45], one of the most common residues in histone proteins, which will likely become more commonly used as the reagents become more accessible. Of course, using these expanded ligation sites requires that the peptide segment bearing the 1,2-aminothiol occur in a synthetic histone fragment; they are not suitable for a one-step EPL reaction scheme to introduce modifications into the N-terminal tail. Another consideration is that any methods for modification of cysteine, described in Sect. 2, can be used to mask a residual ligation site. For example, alkylation of a ligation site Cys by 2-bromoethylamine or derivatives yields thiolysine analogs [27].

4.3 Histone Semi-Synthesis by Expressed Protein Ligation: Modifications Near the Histone C-Terminus

Although modifications near the N-terminus of histone proteins are easily accessible by the methods described above, several interesting modifications near the C-terminus of histone proteins require a different approach. Modifications of interest within ~30 residues of the histone C-terminus include PTM sites in the C-terminal tail of histone H2A, as well as modifications at key interfaces in the structured nucleosome core in histones H3 and H4 [105, 106]. Access to these modified residues requires a synthetic C-terminal histone peptide that can incorporate the modified residues, and a recombinant N-terminal histone fragment that remains unmodified (Fig. 7b). This has most commonly been accomplished by fusion of the N-terminal protein fragment with an intein domain. Folded intein domains are capable of transferring the fused protein segment, or "extein," to a thiol side chain to form an intramolecular thioester [107]. This can be intercepted by free thiol to form a reactive thioester in solution. The Mxe GyrA intein has historically been the most commonly used because of its early discovery and commercial availability as the pTXB1 plasmid (New England Biolabs), the capability to refold

the functional protein, and robust cleavage conditions including the presence of denaturant and high ionic strength [108].

The behavior of histone proteins fused to an intein is somewhat unpredictable. Muir and coworkers found that histone H2B(1–116) expresses as a soluble protein when fused to a GyrA intein-chitin binding domain construct (GyrA-CBD) [109] and may be purified over chitin columns to easily generate the functional intein derivative, but that histone H2A(1–113)-GyrA-His6 expressed into inclusion bodies, which is typical of most recombinant histone proteins. Similarly, our own laboratory has found that histone H3(1–109)-GyrA-CBD and H4(1–75)-GyrA-CBD express into inclusion bodies and, further, interact nonspecifically with chitin columns such that the chitin binding domain is not appropriate for purification purposes [28, 110]. As such, these proteins must be resolubilized from inclusion bodies and refolded to the active intein prior to thiolysis to generate the functional thioester for ligation. At this point, ligation may proceed with standard considerations.

Several modifications in the structured nucleosome core are accessible only via EPL with a C-terminal synthetic peptide. This approach was used to explore the impact of acetylation in the histone–DNA interface at the nucleosome dyad region, at H3-K115ac,K122ac. In this approach, a native cysteine (H3-C110) was used as a ligation site, such that ligation generated the native histone sequence with the modification of interest. These studies demonstrated that acetylation in the nucleosome dyad reduces histone-DNA affinity and increases thermal repositioning, suggesting a destabilization of the histone-DNA interface [28]. Interestingly, Gln substitution for acetylated lysine does not replicate these effects. In addition, acetylation at the dyad increases nucleosome disassembly in the context of mechanically unwrapped DNA [111] and in the context of chromatin remodeling by the hMSH2/hMSH6 DNA repair complex [112]. Subsequent work by Tropberger and coworkers using H3-K122ac prepared using an expanded genetic code approach suggested a role for H3-K122ac in transcriptional activation which is consistent with the biophysical work carried out with the semi-synthetic histones [38]. EPL followed by desulfurization was used to determine the impact of acetylation at H4-K77ac,K79ac in the histone-DNA interface in the Loss of Ribosomal Silencing (LRS) region of the nucleosome [110], resulting in the generation of the native H4 sequence with an alanine at the ligation site. Acetylation in this region of the histone-DNA interface was found to increase DNA unwrapping and site exposure for transcription factor binding, with no effect on nucleosome disassembly. Taken together with studies of the nucleosome dyad region described above and for acetylation of H3-K56ac in the entry-exit region [31, 32], these studies suggest that different regions of the histone-DNA interface play distinct structural and functional roles in the regulation of nucleosome unwrapping and disassembly [111].

Histone phosphorylation remains best accessed by ligation approaches. EPL has enabled study of H3-T118ph, which places a phosphate group into the histone-DNA interface at the nucleosome dyad. This site had been highlighted in genetic screens in yeast as likely to be important for transcriptional regulation and DNA repair [20,

113, 114]. Preparation by EPL and incorporation into mononucleosomes revealed a significant impact. H3-T118ph dramatically decreases histone-DNA affinity by 2 kcal/mol, increases nucleosome mobility in thermal repositioning assays 28-fold, and increases DNA accessibility near the dyad 6-fold. Consistent with this picture of a destabilized histone-DNA interface, this modification also dramatically increased nucleosome disassembly by hMSH2/hMSH6 and by the SWI/SNF chromatin remodeling complex [29]. Reconstitution of nucleosomes with H3-T118ph also resulted in formation of a second defined histone-DNA construct which was structurally distinct from nucleosomes [30]. Extensive characterization of this species suggested that on short segments of DNA, a nucleosome duplex is formed in which two histone octamers are wrapped by two segments of DNA. On long stretches of DNA up to 3 kb in length, structures form in which two histone octamers are wrapped by a single piece of DNA. These structures were reminiscent of the altosomes described during SWI/SNF chromatin remodeling [115, 116]. This study demonstrated an important role for H3-T118 in regulating DNA wrapping, and was the first to demonstrate that a single histone modification could result in large-scale alterations of the nucleosome structure. In both studies, reconstitution of nucleosomes with glutamate substitution for the phosphorylated threonine failed to recapitulate the effects of the modifications, which is a strong argument for the necessity of precise chemical modifications to understand the role of PTMS in chromatin structure and function.

The Muir Laboratory has extensively studied ubiquitylation of histones H2A and H2B using EPL schemes with synthetic C-terminal peptides, and optimizations to their ligation approaches over time have resulted in dramatically improved product yields. Their first chemical strategy to site-specifically introduce ubiquitin into a histone produced ubiquitylated H2B (uH2B) via a three-piece ligation scheme in which the N-terminal fragment of H2B and the first 75 residues of ubiquitin were each expressed as intein fusions to generate the reactive thioester. H2B(117–125)-A117C was generated synthetically, with a photocleavable ligation auxiliary linked to the ε-amine of the lysine side chain. After ligation, desulfurization of the Cys to Ala (see Sect. 3.3) and photolysis of the ligation auxiliary resulted in a traceless ligation. Significantly, this study found that ubiquitylation of H2B-K120 enhanced methylation of H3-K79 by the methyltransferase Dot1, a clear example of histone modification crosstalk [109]. Subsequent ligation approaches eliminated the synthetically costly photocleavable auxiliary in favor of a cysteine linked to the ε-amine of the lysine through an isopeptide bond, such that after ligation and desulfurization, the product maintained a native histone sequence but with a G76A mutation in the ubiquitin domain [117]. Preparation of ubiquitylated H2A added an additional twist, in that no suitable Cys or Ala ligation sites were available. Instead, H2A(1–113) was expressed as a thioester, and H2A(114–128) was synthesized with the thiazolidine derivative of penicillamine such that, after ligation and desulfurization, the histone ligation junction was converted to the native Val residue [101].

Although several laboratories have refined EPL approaches for the preparation of modified histone proteins, the study that perhaps best illustrates the full power of

this methodology is the development of high-throughput libraries of nucleosomes, prepared with semi-synthetic modified histones, marked by reconstitution onto barcoded DNA for identification [118]. These libraries, currently prepared on the order of tens of nucleosomes, could be assembled rapidly with different targeted modification marks in the N- or C-terminal histone tails of all four histone proteins, either individually or simultaneously incorporated within the same nucleosome to allow the identification of individual, synergistic, or antagonistic effects of different histone modifications in the physiologically relevant context of a full nucleosome. This study found robust evidence of crosstalk among modification marks on different histone proteins within the same nucleosome across several histone marks "readers" or "writers."

4.4 Total Synthesis of Histone Proteins by NCL

While EPL grants access to modifications sequestered near the N- or C-terminus of a histone protein, histone proteins contain modifications throughout the entire sequence. Some of the modifications are centrally located such that a single synthetic peptide does not span the required region. Other modification schemes require the simultaneous incorporation of modifications in the N-terminal tail of a histone and in the folded core domain. These cases require the development of elaborate ligation schemes that enable total synthesis of a histone protein, enabling complete chemical control of the entire histone sequence. This field is rapidly advancing, such that any survey is outdated almost upon publication. However, some trends are immediately apparent by assessing current progress and directions.

The first total synthesis of a histone protein was carried out by the Ottesen Laboratory to prepare H3-K56ac via a sequential native chemical ligation scheme (Fig. 8a). In a typical sequential native chemical ligation scheme, C-terminal peptide is synthesized with a reactive 1,2-aminothiol. The central peptide segment, in this case bearing the acetylated lysine of interest, is prepared with the reactive thioester at the C-terminus as well as a protected 1,2-aminothiol at the N-terminus, often introduced as a thiazolidine derivative which remains masked through the first ligation step. After the first ligation step, the reactive 1,2-aminothiol can be revealed to take part in the final ligation step to generate the full-length target histone protein. Interestingly, the first generation synthesis of H3-K56ac serves as an example of the need to carefully select ligation sites and chemistry to avoid distorting results. The first generation ligation scheme used peptides split with ligation junctions at H3-R40C based on sequence alignment to a divergent histone from *Cairina moschata* and H3-S96C based on the human histone variant H3.1. However, analysis of the first generation product H3-R40C,S96C with and without acetylated K56 revealed that these semiconservative cysteine substitutions destabilized the histone-DNA interface to a greater extent than the modification of interest. An improved second generation ligation scheme (Fig. 8a) identified alanine ligation sites at residues H3-47 and H3-91. Of note, all peptides in this first

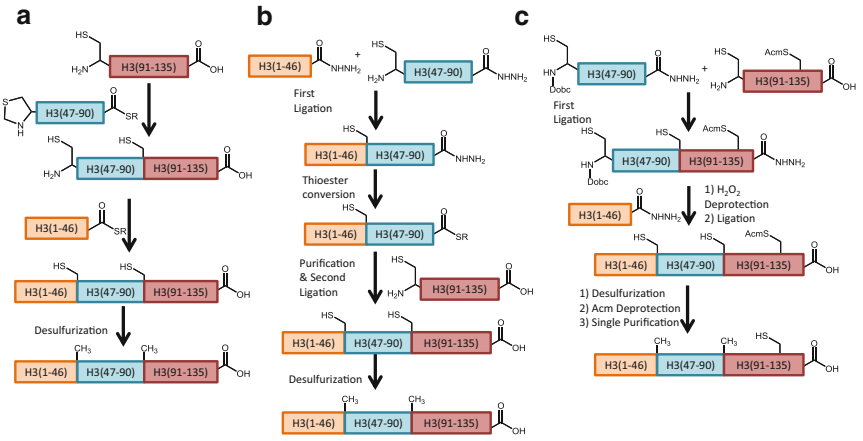


Fig. 8 Total chemical synthesis of histones. (**a**) Sequential NCL of peptide thioesters with Thz-Cys protection. Purification is carried out after each ligation/deprotection step. (**b**) Sequential NCL of peptide hydrazides. Purification is carried out after each ligation step. (**c**) One-pot NCL of peptide hydrazides with Dobz-Cys protection. A single purification step is carried out after ligation, deprotection, and desulfurization

generation synthesis were prepared by solid phase peptide synthesis using Boc chemistry, which is in use only among synthesis-intensive laboratories. Conversion of the synthesis to Fmoc chemistry, which is more amenable to automation, required development of improved techniques for the preparation of the Fmoc-thioester peptide segments [119].

Synthesis of these segments followed by sequential ligation and desulfurization resulted in a native-like histone sequence acetylated at H3-K56ac. Importantly, the use of desulfurization in histone H3 requires the use of the H3-C110A substitution, similar to any cysteine-modification approach. This substitution has been widely used in studies requiring cysteine modification, and no effect of this substitution on nucleosome structure, dynamics, or function has been detected. Studies using this synthetic histone revealed that H3-K56ac increased transcription factor binding in the histone-DNA interface threefold, consistent with results from the Chin laboratory using H3-K56ac prepared by codon suppression techniques (Sect. 2.2) [32]. Direct comparison of H3-K56ac prepared synthetically by sequential native chemical ligation and through genetic incorporation of the H3-K56ac species demonstrates the utility of each approach. For a single acetylation site such as H3-K56ac in the center of the histone sequence, exploitation of the pyrrolysine incorporation machinery resulted in multimilligram quantities of protein and is more suitable for repeated expression. Total synthesis by sequential native chemical ligation provided 7% overall yield of the total histone protein, and each repetition of synthesis is as labor-intensive as the first. The true power of the total synthesis approach, then, lies in the ability to prepare histones with combinations of modifications throughout the sequence, such as H3-Y41ph,K56ac (unpublished data from the Ottesen Laboratory).

Sequential native chemical ligation using masked Cys protection is limited to ligation in a C-to-N direction. In an exciting advance, Liu and coworkers introduced peptide hydrazides as a stable masked thioester which could be converted to a reactive thioester *in situ* for use in native chemical ligation, allowing sequential ligation in the reverse direction (Fig. 8b) [120]. Peptide hydrazides are stable under ligation conditions, but can be converted to a reactive thioester by treatment with $NaNO_2$ and an external ligation-compatible thiol, typically MPAA. They applied this methodology to carry out the sequential ligation of histone H3 with the same split sites as the Ottesen scheme, but in reverse order, such that the N-terminal and central segments were ligated to generate an initial ligation product, followed by conversion of the central hydrazide to a reactive thioester for ligation to the C-terminal peptide segment. Selective sidechain protection of H3-C110 allowed desulfurization followed by ligation to yield Ala at the ligation sites, while also maintaining the native H3-C110 residue. However, similar to the C-to-N synthesis above, they found that yields were limited by the challenges of purification of the final product from the component segments.

To eliminate these handling problems, they sought to develop a one-pot chemical ligation approach (Fig. 8c). The key advance in this study was the identification of an N-terminal cysteine protection strategy compatible with the chemistry required for conversion of the peptide hydrazide, because the Thz group commonly used for N-terminal cysteine protection is labile to the $NaNO_2$ conversion step. These requirements were satisfied by *p*-boronobenzyloxycarbonyl (Dobz) protection of the central segment, which could then be reversed by reaction with H_2O_2 to regenerate the active cysteine. Overall, this one-pot approach resulted in 20% isolated yield of histone H3-K4me3, a substantial improvement over the Ottesen approach. The Liu group extended the one-pot chemical ligation approach with peptide hydrazides and Dobz protection to the total synthesis of histone H4 [121]. Using the same approaches, they achieved an average 18% overall isolated yield of H4-K16ac. Although these percentage yields are not commensurate with an EPL strategy for N-terminal tail modification, they are highly respectable for a total synthesis which can be applied in the future to combinations of modifications throughout the histone sequence, and generated multimilligram quantities of each modified histone. These amounts are sufficient for nearly all biochemical and biophysical assays for characterization of modified nucleosomes.

The Brik Laboratory has introduced two alternate solutions to the challenges posed by sequential native chemical ligation in the context of histone H2B. In 2013, Siman and coworkers developed a convergent ligation strategy that exploits the regulated reactivity of peptide hydrazides used by Liu and coworkers (Fig. 9) [122]. In this strategy, H2B is split into four synthetically accessible peptide fragments. In the first ligation step, H2B(1–20) is ligated to H2B(22–56)-hydrazide and purified; separately Thz-H2B(59–96) is ligated to H2B(98–125), the N-terminal Cys is revealed by treatment with methoxylamine, and the segment is purified. The ligated H2B(1–56)-hydrazide segment is then converted to reactive thioester by treatment with $NaNO_2$, and ligated to the H2B(59–125) segment to generate the full length histone sequence. Desulfurization and purification yields

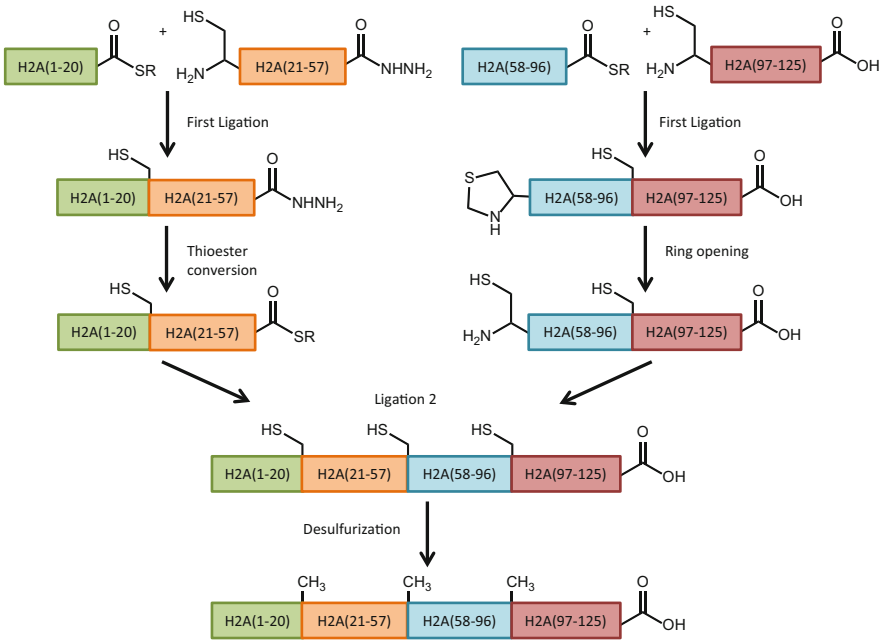


Fig. 9 Total chemical synthesis of histone H2A by convergent ligation. Purification is carried out after synthesis of each two-fragment ligation product

the final product. While the synthesis of H2B alone is a significant achievement, the potential of total histone synthesis was further demonstrated by modification of the reaction scheme to incorporate ubiquitin at H2B-K34 [123]. Brik synthetically introduced a δ-mercaptolysine residue at position H2B-K34 in the H2B(22–56) peptide hydrazide, which enabled orthogonal ligation to the ubiquitin thioester to generate the H2B(22–56, K34Ub)-hydrazide fragment. This peptide could then be plugged into the convergent ligation scheme to generate the full-length H2B-K34Ub in low yields, but sufficient for incorporation into 12-mer nucleosome arrays suitable for biochemical characterization.

Solid phase synthesis chemistry has had profound implications for the preparation of complex molecules, including individual peptide segments. Some elements of solid phase chemistry would appear ideal for use with the assembly of histone proteins via sequential solid phase ligation reaction. The Brik Laboratory explored this concept for an improved total synthesis of histone H2B (Fig. 10) [123]. A key element of any solid phase ligation approach is the use of a chemical linker to connect the growing protein chain to a solid support that is stable for all ligation conditions, but labile to orthogonal cleavage conditions after synthesis. Here, researchers selected the acid-labile Rink linker commonly used for synthesis of peptide amides. They first assembled a solid support connected to a Rink linker with an N-terminal cysteine suitable for ligation. The prepared four peptide segments using similar split sites to those exploited in the convergent synthesis to assemble

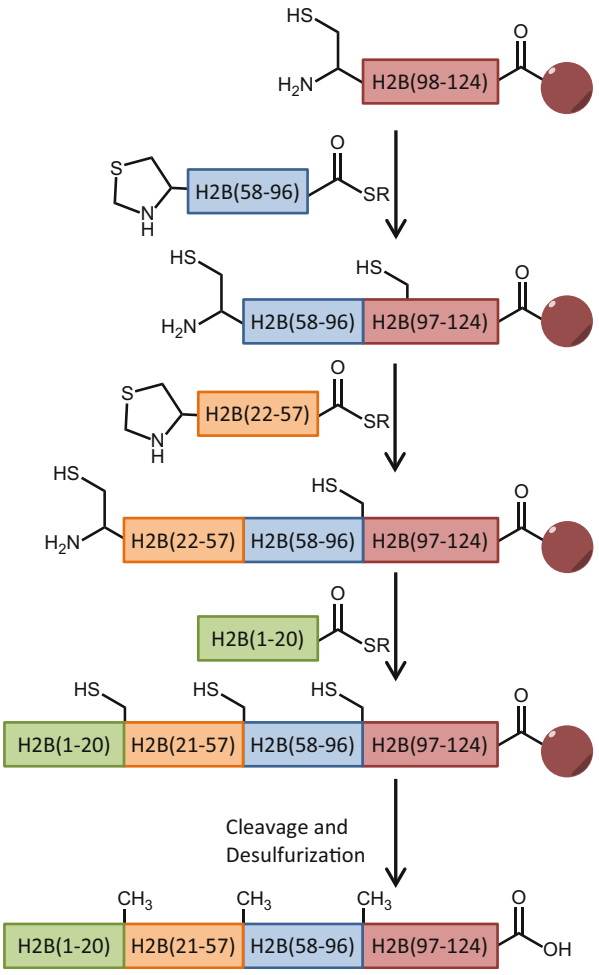


Fig. 10 Total chemical synthesis of histone H2B by solid phase native chemical ligation on PEGA resin with a Rink linker. A single purification step is carried out after on-resin desulfurization and cleavage

the final product: H2B(1–20)-thioester, Thz-H2B(22–57)-thioester, Thz-H2B(59–96)-thioester, and Thz-H2B(98–124)-thioester. Each round of solid phase ligation then consisted of a repeated cycle of ligation, wash, Cys deprotection, and wash steps. After chain assembly, free radical desulfurization was used to convert all Cys residues to Ala on the solid phase. Treatment with trifluoroacetic acid then revealed the full-length product in 10% isolated yield, which is significantly improved over the convergent approach for this histone. Of note, the final protein product resulted in an H2B-K125A substitution, and generation of the C-terminal amide rather than acid derivative. Because the C-terminal tail of histone H2B is not folded into the nucleosome core, these substitutions are likely to be permitted. However, care would be required to consider alternate resin attachment strategies for proteins such as H3 and H4 in which the C-terminal tail forms interactions within the structured nucleosome core.

Total protein synthesis by NCL offers the greatest potential level of chemical control over every residue within a histone sequence. These methods offer the possibility of complex combinations of chemically precise modifications installed throughout the nucleosome. However, because of the challenges inherent to these synthetic routes, this comes as a trade-off with effort and yield in synthesis. Continual rapid advances in the field are likely to reduce the barriers to the preparation of fully synthetic histones for researchers who are not specialists in chemical protein synthesis.

5 Prospects: Synthetic Histone Proteins in the Eukaryotic Cell

The previous sections have considered different ways to prepare precisely modified histone proteins with single or multiple PTMs. We have available extremely powerful tools covering the full range from fast, simple, and large scale preparation of modification mimics, to careful, precise chemical control over the full histone sequence. These amazing toolkits have enabled biochemical and biophysical investigation of the structural, dynamic, and functional properties of modified chromatin. However, carefully modified histones, for the most part, have been restricted to *in vitro* studies, isolated from the full biological complexity of the cellular environment. Even though major advances have been made in understanding histone modification cross-talk [124] and the local chromatin environment through preparation of complex synthetic chromatin arrays [16, 125], the next frontier is the introduction of synthetic histones directly into a live eukaryotic cell for incorporation into functional chromatin. If this could be accomplished, the effects of specific sets of modifications could be probed directly.

The most promising lead for this work is the slime mold *Physarum polycephalum*. This fascinating myxomycete has several different growth stages ranging from free-swimming amoeba to micro- and macro-plasmodia, single-celled states which can grow up to 30 cm in diameter, in which each cell contains tens to millions of nuclei synchronized across the cell cycle [126]. In the microplasmodial and macroplasmodial stages, this organism has been shown to spontaneously uptake exogenous histone proteins from media, transport these histones into the nucleus, and incorporate them into active regions of transcription within their chromatin [127, 128]. This myxomycete is easy to grow, and is widely used in the field of bio-computing [129, 130].

Work carried out primarily by Thiriet and Hayes over the past decade has demonstrated that properly folded histone H2A/H2B dimers or H3/H4 tetramers added exogenously to *P. polycephalum* microplasmodia and macroplasmodia for spontaneous uptake, as demonstrated by following the localization of fluorescein- or FLAG-tagged histones. Uptake is most rapid during the end of the G2 phase of the cell cycle, which occurs simultaneously for all nuclei within a single micro- or

macroplasmodium [127, 131, 132]. Both H2A/H2B dimers and H3/H4 tetramers are incorporated into chromatin, although H2A/H2B dimers are deposited at a higher rate than H3/H4 tetramers [133, 134]. Thiriet and coworkers have explored a range of questions in histone transport and chromatin assembly using these techniques. In a recent study, they introduced recombinant histone proteins with Gln mimics of acetylation in the H4 tails at combinations of positions 5, 8, 12, and 16. In general, these acetylation mimics appeared to increase nucleosome exchange, although glutamine substitution at solely H4-K8 and/or H4-K16 abolished uptake into the nuclei. There would appear to be no reason why synthetic and semi-synthetic histones could not be used similarly, to probe directly histone modification cross-talk in the context of the cellular environment.

Although the spontaneous uptake of exogenous histone complexes has been described as unique to *P. polycephalum*, the true target for designer chromatin would be a mammalian or human cell line which could incorporate chemically modified histones into chromatin. The ultimate target for designer chromatin would be a human cell line that could take up histones and incorporate them into chromatin. The N-terminal tails of histones have stretches of highly positive charge resembling cell-penetrating peptides [135, 136] and, in fact, researchers have demonstrated that histone-derived CPPs can be used to carry protein cargo such as bovine serum albumin into *Leishmania tarentolae* and into protoplasts of petunia cells [137, 138]. Initial reports suggested that unfolded histones might penetrate HeLa and Colo-205 cells to enter the nucleus, possibly through direct translocation across the plasma membrane rather than through endocytosis, although several questions remain unresolved [139]. Histones have also been proposed to increase the uptake of plasmids into human cell lines in a process denoted "histonefection" [140] through a poorly-understood, non-endosome-mediated pathway [141]. If any of these methods are validated they could enable the power of chemistry to be used to insert precisely modified histone proteins into the complexity of the biological test tube that is the cell.

References

1. Davey CA et al (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. J Mol Biol 319:1097–1113
2. Luger K et al (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389:251–260
3. Lochmann B, Ivanov D (2012) Histone H3 localizes to the centromeric DNA in budding yeast. PLoS Geneti 8:e1002739
4. Redon C et al (2002) Histone H2A variant H2AX and H2AZ. Curr Opin Genet Dev 12:162–169
5. Ward IM et al (2003) Accumulation of checkpoint protein 53BP1 at DNA breaks involves its binding to phosphorylated histone H2AX. J Biol Chem 278:19579–19582
6. Park YJ et al (2004) A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. J Biol Chem 279:24274–24282

7. Jenuwein T (2001) Translating the histone code. Science 293:1074–1080
8. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403:41–45
9. Schwammle V et al (2014) Large scale analysis of co-existing post-translational modifications in histone tails reveals global fine structure of cross-talk. Mol Cell Proteomics 13:1855–1865
10. Lin S, Garcia BA (2012) Examining histone posttranslational modification patterns by high-resolution mass spectrometry. Methods Enzymol 512:3–28
11. Tan M et al (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell 146:1016–1028
12. Singh MP, Wijeratne SSK, Zempleni J (2013) Biotinylation of lysine 16 in histone H4 contributes toward nucleosome condensation. Arch Biochem Biophys 529:105–111
13. Dhall A et al (2014) Sumoylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions. J Biol Chem 289:33827–33837
14. Sakabe K, Wang Z, Hart GW (2010) Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. Proc Natl Acad Sci U S A 107:19915–19920
15. Fierz B, Muir TW (2012) Chromatin as an expansive canvas for chemical biology. Nat Chem Biol 8:417–427
16. Pick H, Kilic S, Fierz B (2014) Engineering chromatin states: chemical and synthetic biology approaches to investigate histone modification function. Biochim Biophys Acta 1839:644–656
17. Fierz B (2014) Synthetic chromatin approaches to probe the writing and erasing of histone modifications. ChemMedChem 9:495–504
18. Frederiks F et al (2011) A modified epigenetics toolbox to study histone modifications on the nucleosome core. ChemBioChem 12:308–313
19. Matsubara K et al (2007) Global analysis of functional surfaces of core histones with comprehensive point mutants. Genes Cells 12:13–33
20. Hyland EM et al (2005) Insights into the role of histone H3 and histone H4 core modifiable residues in Saccharomyces cerevisiae. Mol Cell Biol 25:10060–10070
21. Luger K, Rechsteiner TJ, Richmond TJ (1999) Preparation of nucleosome core particle from recombinant histones. Methods Enzymol 304:1–19
22. Watanabe S et al (2010) Structural characterization of H3K56Q nucleosomes and nucleosomal arrays. Biochim Biophys Acta 1799:480–486
23. Muthurajan UM et al (2004) Crystal structures of histone Sin mutant nucleosomes reveal altered protein-DNA interactions. EMBO 23:260–270
24. Iwasaki W et al (2011) Comprehensive structural analysis of mutant nucleosomes containing lysine to glutamine (KQ) substitutions in the H3 and H4 histone-fold domains. Biochemistry 50:7822–7832
25. Yu Q et al (2011) Differential contributions of histone H3 and H4 residues to heterochromatin structure. Genetics 188:291–308
26. Wang X, Hayes JJ (2008) Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure. Mol Cell Biol 28:227–236
27. Allahverdi A et al (2010) The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. Nucleic Acids Res 39:1680–1691
28. Manohar M et al (2009) Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding. J Biol Chem 284:23312–23321
29. North JA et al (2011) Phosphorylation of histone H3(T118) alters nucleosome dynamics and remodeling. Nucleic Acids Res 39:6465–6474
30. North JA et al (2014) Histone H3 phosphorylation near the nucleosome dyad alters chromatin structure. Nucleic Acids Res 42:4922–4933
31. Shimko JC et al (2011) Preparation of fully synthetic histone H3 reveals that acetyl-lysine 56 facilitates protein binding within nucleosomes. J Mol Biol 408:187–204

32. Neumann H et al (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Mol Cell 36:153–163
33. Masumoto H et al (2005) A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. Nature 436:294–298
34. Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. Annu Rev Biochem 79:413–444
35. Chin JW (2014) Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem 83:379–408
36. Hao B et al (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science 296:1462–1466
37. Tarrant MK, Cole PA (2009) The chemical biology of protein phosphorylation. Annu Rev Biochem 78:797–825
38. Tropberger P, Schneider R (2013) Scratching the (lateral) surface of chromatin regulation by histone modifications. Nat Struct Mol Biol 20:657–661
39. Di Cerbo V et al. (2014) Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. Elife 3:e01632
40. Nguyen DP et al (2010) Genetically directing ε-N,N-dimethyl-l-lysine in recombinant histones. Chem Biol 17:1072–1076
41. Wang Y-S et al (2010) A genetically encoded photocaged N^{ε}-methyl-l-lysine. Mol BioSyst 6:1557
42. Xie Z et al (2012) Lysine succinylation and lysine malonylation in histones. Mol Cell Proteomics 11:100–107
43. Gattner MJ, Vrabel M, Carell T (2013) Synthesis of ε-N-propionyl-, ε-N-butyryl-, and ε-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. Chem Commun 49:379
44. Kim CH et al (2012) Site-specific incorporation of ε-N-crotonyllysine into histones. Angew Chem Int Ed 51:7246–7249
45. Yang R et al (2009) Dual native chemical ligation at lysine. JACS Commun 131:13592–13593
46. Li X et al (2009) A pyrrolysine analogue for site-specific protein ubiquitination. Angew Chem Int Ed 48:9184–9187
47. Yang R et al (2014) Native chemical ubiquitination using a genetically incorporated azidonorleucine. Chem Commun 50:7971
48. Guo J et al (2008) Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. Angew Chem Int Ed 47:6399–6401
49. Wang ZU et al (2012) A facile method to synthesize histones with posttranslational modification mimics. Biochemistry 51:5232–5234
50. Chalker JM et al (2011) Methods for converting cysteine to dehydroalanine on peptides and proteins. Chem Sci 2:1666
51. Chalker JM et al (2012) Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications. Angew Chem Int Ed Engl 51:1835–1839
52. Sawicka A, Seiser C (2014) Sensing core histone phosphorylation — a matter of perfect timing. Biochim Biophys Acta 1839:711–718
53. Park HS et al (2011) Expanding the genetic code of *Escherichia coli* with phosphoserine. Science 333:1151–1154
54. Lee S et al (2013) A facile strategy for selective incorporation of phosphoserine into histones. Angew Chem Int Ed 52:5771–5775
55. Chin JW et al (2003) Progress toward an expanded eukaryotic genetic code. Chem Biol 10:511–519
56. Lajoie MJ et al (2013) Genomically recoded organisms expand biological functions. Science 342:357–360
57. Ambrogelly A, Palioura S, Söll D (2007) Natural expansion of the genetic code. Nat Chem Biol 3:29–35

58. Anderson JC et al. (2004) An expanded genetic code with a functional quadruplet codon. Proc Natl Acad Sci 101:7566–7571
59. Wang K, Schmied WH, Chin JW (2012) Reprogramming the genetic code: from triplet to quadruplet codes. Angew Chem Int Ed 51:2288–2297
60. Riddle DL, Carbon J (1973) Frameshift suppression: a nucleotide addition in the anticodon of a glycine transfer RNA. Nat New Biol 242:230–234
61. Bowman A et al (2010) Probing the (H3-H4)2 histone tetramer structure using pulsed EPR spectroscopy combined with site-directed spin labelling. Nucleic Acids Res 38:695–707
62. Ward R et al (2009) Long distance PELDOR measurements on the histone core particle. J Am Chem Soc 131:1348–1349
63. Tims HS, Widom J (2007) Stopped-flow fluorescence resonance energy transfer for analysis of nucleosome dynamics. Methods (San Diego, Calif.) 41:296–303
64. Dechassa ML et al (2008) Architecture of the SWI/SNF-nucleosome complex. Mol Cell Biol 28:6010–6021
65. Ferreira H et al (2007) Histone tails and the H3 αN helix regulate nucleosome mobility and stability. Mol Cell Biol 27:4037–4048
66. Kurumizaka H et al (2013) Current progress on structural studies of nucleosomes containing histone H3 variants. Curr Opin Struct Biol 23:109–115
67. Flaus A et al (1996) Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals. Proc Natl Acad Sci U S A 93:1370–1375
68. Poirier MG et al (2009) Dynamics and function of compact nucleosome arrays. Nat Struct Mol Biol 16:938–944
69. Simon MD et al (2007) The site-specific installation of methyl-lysine analogs into recombinant histones. Cell 128:1003–1012
70. Kenyon G, Bruice TW (1977) Novel sulfhydryl reagents. Methods Enzymol 47:407–430
71. Lauberth SM et al (2013) H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell 152:1021–1036
72. Lu X et al (2008) The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. Nat Struct Mol Biol 15:1122–1124
73. Xu C et al (2008) Structural basis for the recognition of methylated histone H3K36 by the Eaf3 subunit of histone deacetylase complex Rpd3S. Structure 16:1740–1750
74. Eidahl JO et al (2013) Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. Nucleic Acids Res 41:3924–3936
75. Hung T et al (2009) ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. Mol Cell 33:248–256
76. Margueron R et al (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461:762–767
77. Huang R et al (2010) Site-specific introduction of an acetyl-lysine mimic into peptides and proteins by cysteine alkylation. J Am Chem Soc 132:9986–9987
78. Hoyle CE, Bowman CN (2010) Thiol-ene click chemistry. Angew Chem Int Ed 49:1540–1573
79. Li F et al (2011) A direct method for site-specific protein acetylation. Angew Chem Int Ed 50:9611–9614
80. Le D et al. (2013) Site-Specific and Regiospecific Installation of Methylarginine Analogues into Recombinant Histones and Insights into Effector Protein Binding. J Am Chem Soc 135:2879–2882.
81. Chatterjee A et al (2010) Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. Nat Chem Biol 6:267–269
82. Whitcomb SJ et al (2012) Histone monoubiquitylation position determines specificity and direction of enzymatic cross-talk with histone methyltransferases Dot1L and PRC2. J Biol Chem 287:23718–23725
83. Dawson PE et al (1994) Synthesis of proteins by native chemical ligation. Science 266:776–779

84. He S et al. (2003) Facile synthesis of site-specifically acetylated and methylated histone proteins: reagents for evaluation of the histone code hypothesis. Proc Natl Acad Sci 100:12033–12038
85. Shogren-Knaak MA, Fry CJ, Peterson CL (2003) A native peptide ligation strategy for deciphering nucleosomal histone modifications. J Biol Chem 278:15744–15748
86. Fierz B et al (2011) Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. Nat Chem Biol 7:113–119
87. Nguyen DP et al (2014) Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. J Am Chem Soc 136:2240–2243
88. Fry CJ, Shogren-Knaak MA, Peterson CL (2004) Histone H3 amino-terminal tail phosphorylation and acetylation: synergistic or independent transcriptional regulatory marks? Cold Spring Harb Symp Quant Biol 69:219–226
89. Shogren-Knaak M (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science 311:844–847
90. Ferreira H, Flaus A, Owen-Hughes T (2007) Histone modifications influence the action of Snf2 family remodelling enzymes by different mechanisms. J Mol Biol 374:563–579
91. Liu Y et al (2011) Influence of histone tails and H4 tail acetylations on nucleosome-nucleosome interactions. J Mol Biol 414:749–764
92. Chiang KP et al (2009) A semisynthetic strategy to generate phosphorylated and acetylated histone H2B. Chembiochem 10:2182–2187
93. Casadio F et al (2013) H3R42me2a is a histone modification with positive transcriptional effects. PNAS 110:14894–14899
94. Kim J et al (2013) The n-SET domain of Set1 regulates H2B ubiquitylation-dependent H3K4 methylation. Mol Cell 49:1121–1133
95. Chen Z, Gryzbowski AT, Ruthenburg AJ (2014) Traceless semisynthesis of a set of histone 3 species bearing specific lysine methylation marks. ChemBioChem 15:2071–2075
96. Hackeng TM, Dawson PE (1999) Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. Proc Natl Acad Sci U S A 96:10069–10073
97. Li S, Shogren-Knaak MA (2008) Cross-talk between histone H3 tails produces cooperative nucleosome acetylation. Proc Natl Acad Sci 105:18243–18248
98. Fry CJ et al (2006) The LRS and SIN domains: two structurally equivalent but functionally distinct nucleosomal surfaces required for transcriptional silencing. Mol Cell Biol 26:9045–9059
99. Zhu Y, van der Donk W (2001) Convergent synthesis of peptide conjugates using dehydroalanines for chemoselective ligations. Org Lett 3:1189–1192
100. 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
101. Fierz B et al (2012) Stability of nucleosomes containing homogenously ubiquitylated H2A and H2B prepared using semisynthesis. J Am Chem Soc 134:19548–19551
102. Wong CTT et al (2014) Realizing serine/threonine ligation: scope and limitations and mechanistic implication thereof. Front Chem 2:28
103. Haase C, Rohde H, Seitz O (2008) Native chemical ligation at valine. Angew Chem Int Ed 47:6807–6810
104. Crich D, Banerjee A (2007) Native chemical ligation at phenylalanine. JACS Commun 129:10064–10065
105. Mersfelder EL, Parthun MR (2006) The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. Nucleic Acids Res 34:2653–2662
106. Jack AM, Hake S (2014) Getting down to the core of histone modifications. Chromosoma 123:355–371
107. Wood DW, Camarero JA (2014) Intein applications: from protein purification and labeling to metabolic control methods. J Biol Chem 289:14512–14519

108. Ayers B et al (1999) Introduction of unnatural amino acids into proteins using expressed protein ligation. Pept Sci 51:343–354
109. McGinty RK et al (2008) Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. Nature 453:812–816
110. Shimko JC et al. (2013) Preparing semisynthetic and fully synthetic histones H3 and H4 to modify the nucleosome core. Methods Mol Biol 981:177–192
111. Simon M et al (2011) Histone fold modifications control nucleosome unwrapping and disassembly. Proc Natl Acad Sci U S A 108:12711–12716
112. Javaid S et al (2009) Nucleosome remodeling by hMSH2-hMSH6. Mol Cell 36:1086–1094
113. Kruger W et al (1995) Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. Genes Dev 9:2770–2779
114. Hurd PJ et al (2009) Phosphorylation of histone H3 Thr-45 is linked to apoptosis. J Biol Chem 284:16675–16683
115. Ulyanova NP, Schnitzler GR (2005) Human SWI/SNF generates abundant, structurally altered dinucleosomes on polynucleosomal templates. Mol Cell Biol 25:11156–11170
116. Schnitzler GR, Sif S, Kingston RE (1998) Human SWI/SNF interconverts a nucleosome between its base state and a stable remodeled state. Cell 94:17–27
117. McGinty RK et al (2009) Structure–activity analysis of semisynthetic nucleosomes: mechanistic insights into the stimulation of dot1l by ubiquitylated histone H2B. ACS Chem Biol 4:958–968
118. Nguyen UTT et al (2014) Accelerated chromatin biochemistry using DNA-barcoded nucleosome libraries. Nat Methods 11:834–840
119. Mahto SK et al (2011) A reversible protection strategy to improve Fmoc-SPPS of peptide thioesters by the N-acylurea approach. ChemBioChem 12:2488–2494
120. Fang GM, Wang JX, Liu L (2012) Convergent chemical synthesis of proteins by ligation of peptide hydrazides. Angew Chem Int Ed Engl 51:10347–10350
121. Li J et al (2014) One-pot native chemical ligation of peptide hydrazides enables total synthesis of modified histones. Org Biomol Chem 12:5435
122. Siman P et al (2013) Convergent chemical synthesis of histone H2B protein for the site-specific ubiquitination at Lys34. Angew Chem Int Ed 52:8059–8063
123. Jbara M, Seenaiah M, Brik A (2014) Solid phase chemical ligation employing a Rink amide linker for the synthesis of histone H2B protein. Chem Commun 50:12534–12537
124. Linghu C et al (2013) Discovering common combinatorial histone modification patterns in the human genome. Gene 518:171–178
125. Zheng C, Hayes JJ (2003) Intra- and inter-nucleosomal protein-DNA interactions of the core histone tail domains in a model system. J Biol Chem 278:24217–24224
126. Mohberg J, Rusch HP (1969) Isolation of the nuclear histones from the myxomycete, *Physarum polycephalum*. Arch Biochem Biophys 134:577–589
127. Thiriet C, Hayes JJ (1999) Histone proteins in vivo: cell-cycle-dependent physiological effects of exogenous linker histones incorporated into *Physarum polycephalum*. Methods 17:140–150
128. Prior CP et al (1980) Incorporation of exogenous pyrene-labeled histone into Physarum chromatin: a system for studying changes in nucleosomes assembled in vivo. Cell 20:597–608
129. Adamatzky A (2013) Slimeware: engineering devices with slime mold. Artificial Life 19:317–330
130. Taylor B et al (2014) *Physarum polycephalum*: towards a biological controller. Biosystems 127C:42–46
131. Ejlassi-Lassallette A et al (2010) H4 replication-dependent diacetylation and Hat1 promote S-phase chromatin assembly in vivo. Mol Biol Cell 22:245–255
132. Thiriet C, Hayes JJ (2001) A novel labeling technique reveals a function for histone H2A/H2B dimer tail domains in chromatin assembly in vivo. Genes Dev 15:2048–2053

133. Studitsky VM, Clark DJ, Felsenfeld G (1994) A histone octamer can step around a transcribing polymerase without leaving the template. Cell 76:371–382
134. Kireeva ML et al (2002) Nucleosome remodeling induced by RNA polymerase II: loss of the H2A-H2B dimer during transcription. Mol Cell 9:541–552
135. Nishiyama A et al (2008) Intracellular delivery of acetyl-histone peptides inhibits native bromodomain-chromatin interactions and impairs mitotic progression. FEBS Lett 582:1501–1507
136. Heo K et al (2013) Cell-penetrating H4 tail peptides potentiate p53-mediated transactivation via inhibition of G9a and HDAC1. Oncogene 32:2510–2520
137. Keller AA et al (2014) Transduction of proteins into Leishmania tarentolae by formation of non-covalent complexes with cell-penetrating peptides. J Cell Biochem 115:243–252
138. Rosenbluh J et al (2004) Non-endocytic penetration of core histones into petunia protoplasts and cultured cells: a novel mechanism for the introduction of macromolecules into plant cells. Biochim Biophys Acta 1664:230–240
139. Hariton-Gazal E et al (2003) Direct translocation of histone molecules across cell membranes. J Cell Sci 116:4577–4586
140. Balicki D et al. (2002) Structure and function correlation in histone H2A peptide-mediated gene transfer. Proc Natl Acad Sci 99:7467–7471
141. Kaouass M, Beaulieu R, Balicki D (2006) Histonefection: Novel and potent non-viral gene delivery. J Control Release 113:245–254

Index

,