Tetrahedron 68 (2012) 9491–9537

Contents lists available at SciVerse ScienceDirect

Tetrahedron

journal homepage: www.elsevier.com/locate/tet

Tetrahedron report number 992

Total chemical synthesis of polypeptides and proteins: chemistry of ligation techniques and beyond☆

Hosahalli P. Hemantha, Narasimhamurthy Narendra, Vommina V. Sureshbabu *

#109, Peptide Research Laboratory, Department of Studies in Chemistry, Central College Campus, Dr. B.R. Ambedkar Veedhi, Bangalore University, Bangalore 560 001, India

ARTICLE INFO

Article history:
Received 13 August 2012
Available online 4 September 2012

Dedicated to the memories of Professor Kadlebal M. Sivanandiah (1931–2011)

Contents

☆ From Emil Fischer's vision lead to accomplishing the chemical assembly of enzymes in a test tube.

* Corresponding author. Tel.: +91 80 2296 1339; e-mail addresses: hariccb@gmail.com, hariccb@hotmail.com, sureshbabuvommina@rediffmail.com (V.V. Sureshbabu).

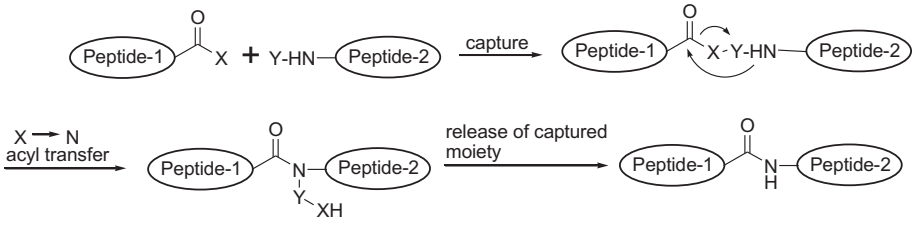
0040-4020/$ – see front matter
http://dx.doi.org/10.1016/j.tet.2012.08.059

1. Introduction

Peptides and proteins are vital components of cells and are crucial for the survival and functioning of cells. They are also important molecules in synthetic, biomedical, pharmaceutical and materials chemistry.[1] An understanding of protein structure and function is key to both the development and sustainability of humankind.[2] Unveiling of the structure–function relationships of peptides and proteins is a key process in the discovery of new therapeutics. The maximum harvesting of proteins as biomolecules enforces a deep knowledge about their post-translational modifications, and the information on how they affect intrinsic biological functions, stability, localization and in vivo 3D structural parameters.[3] An understanding of these phenomena is facilitated by the availability of a facile method of synthesis, which yields proteins of specific sequence in affordable quantities. At present, however, the design of such an adventure at the level of proteome is imaginary due to the mammoth size of the molecules and practical limitations in the present synthetic protocols.

Emil Fischer's disclosure of the synthesis of the first synthetic peptide Gly-Gly (by the acidic hydrolysis of the diketopiperazine of glycine) in 1901 paved the avenue for the era of peptide science in the biological world.[4] Thereafter, the discovery of new protecting groups made the synthesis of a variety of peptide hormones possible,[5] albeit small in size (8–20 amino acid residues). du Vigneaud, in 1953, synthesized an octapeptide hormone, oxytocin.[6] With the following problems are encountered: (a) incomplete coupling and deprotection reactions lead to truncated and deletion sequences, (b) accumulation of byproducts arising from incomplete reactions, impurities from reagents, solvents and protected amino acids, and (c) aggregation of growing peptides.

The efforts in the direction of overcoming the barrier on the length of polypeptides as well as proteins have driven the synthetic peptide chemistry towards a new dimension in the form of ligation techniques.[14] This approach harbours the potential of overcoming the inadequacies of both solid- and solution-phase protocols for the synthesis of proteins. In this technique, previously assembled large-peptide fragments are joined together chemoselectively through the formation of an amide or non-amide linkage.[15] This method has the advantage of enforced proximity between the two segments where the local effective concentration of the reactants is high enough to overcome the entropy barrier and hence facilitate the coupling. The ligation protocol has the advantage of allowing unprotected peptides, choice of reaction medium as organic or aqueous and synthesis on solid or through solution phase. The ligation techniques are based on three processes: an initial capture step, acyl migration and, finally, release of the captured moiety (Scheme 1). Such chemoselective ligations are often based on employing thiol and imine groups as reactive units to join the two fragments either through amidic (peptide bond) or non-amidic linkages such as hydrazone, oxime, thioester and thioether.[16–18]



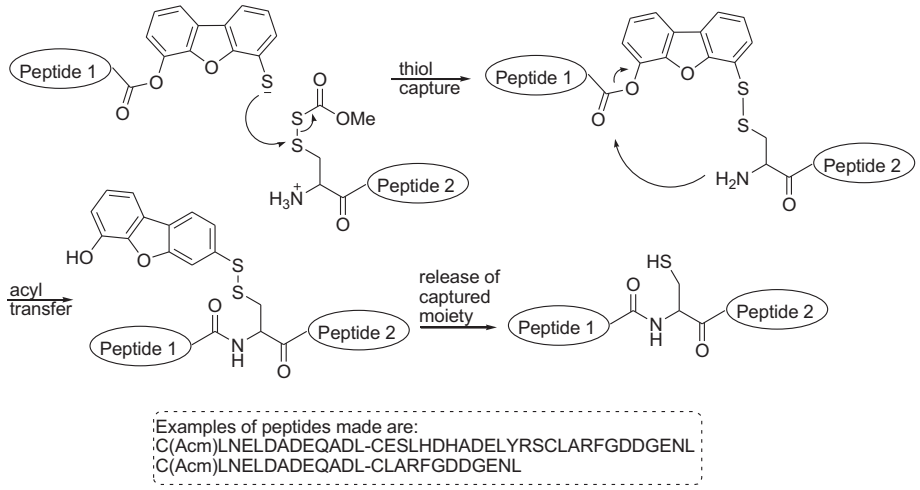
Scheme 1. An outline of the general approach for chemical ligation of peptides.

evolution of the methodology and developments in the chemistry of both protection as well as coupling strategies including the availability of sophisticated instruments (such as RP-HPLC, MPLC and modern HPLC techniques combined with a wide range of mass spectrometers), more and more polypeptides were made.[7]

A breakthrough with respect to the duration of synthesis and length of the peptides was achieved with the discovery of solid-phase peptide synthesis (SPPS) by Merrifield in 1963.[8] In due course, the advancements and automation in the technique of SPPS led to a revolution in the art of peptide syntheses.[9] Such prodigious advancements enabled the preparation of moderate length peptides such as insulin (51 amino acids), ribonuclease A (124 amino acids) and lysozyme (129 amino acids) successfully.[10] However, the range of achievable lengths of peptides with high chemical homogeneity (to the levels detectable by HPLC) lies around 55–60 residues. In other words, the application of SPPS to large polypeptide synthesis did not meet with overwhelming success. The limitation of SPPS[11–13] is that, when a peptide is assembled, the

Kemp's prior thiol-capture strategy, published in 1981, was a seminal contribution to the concept of peptide ligation techniques.[19] It immediately attracted considerable attention from the practitioners of peptide science as well as organic chemistry in those days because the technique was far different from the then prevailing conventional synthetic and semisynthetic methods and, importantly, no enthalpic agents, enzymes, biologicals or protecting groups were involved. In the method, a rigid template, 4-hydroxy-6-mercaptodibenzofuran was used as the capturing moiety to bring the peptide fragments into close proximity via a thiol–disulfide exchange reaction followed by an acyl transfer, thereby establishing a new amide bond between the two fragments. Cleavage of the captured moiety yielded a polypeptide with no racemization at the ligation junction (Scheme 2). This strategy was successfully implemented for the preparation of peptides of 25–40 amino acids in length.[20]

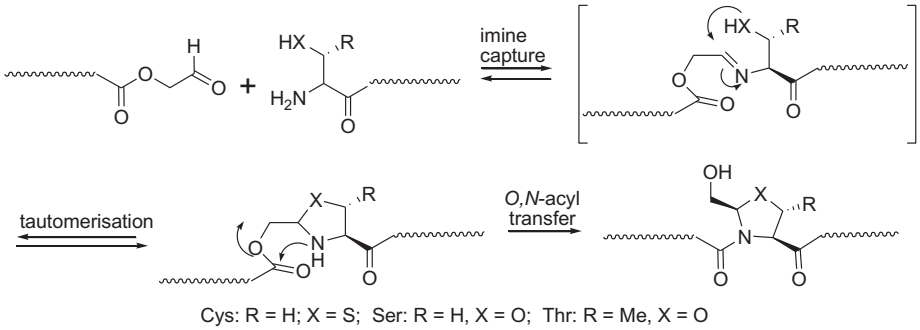
Chemoselective protein ligation began to blossom to its full potential with the introduction of several strategies. Tam et al.



Scheme 2. Template based thiol-capture ligation chemistry.

developed an interesting route based on the reaction between a peptide aldehyde and a Cys- or Ser/Thr-peptide, which led to a stable five-membered 1,3-thiazolidine or -oxazolidine ring through an imine intermediate.[21] In the following step, thiazolidine (or oxazolidine) undergoes a pH-dependent *O*→*N* acyl migration into a peptide bearing a five-membered ring, which mimics proline [thiaproline in the case of Cys (*S*-Pro), oxaproline in the case of Ser/Thr (*O*-Pro)] and the technique was thus aptly named pseudoproline ligation (Scheme 3). Reaction of the peptide aldehyde with other amines present, if any, such as those in the side chain of Lys led to a reversible and unstable Schiff's base.

solvent driven regioselectivity was utilized by Tam's group for tandem ligation of three or more unprotected peptides. In an exemplary three-segment two-ligation experiment, Ser-Leu-Ile-Leu-Asn-Gly-OCH_2CHO and Cys-Phe-Lys-Ile-OH were ligated selectively in an aqueous buffer (pH 5.3) to yield the thiazolidine followed by *O*→*N* acyl migration at pH 6.6, which was complete in 20 h (86% yield). No reaction at the Ser end was reported during this step. The resulting peptide, Ser-Leu-Ile-Leu-Asn-Gly-*S*-*Pro*-Phe-Lys-Ile-OH, after purification, was ligated to Asp-Ser-Phe-Gly-OCH_2CHO in an anhydrous pyridine–acetic acid medium. The reaction was complete in about 35 h and the product, Asp-Ser-Phe-Gly-*O*-*Pro*-Leu-



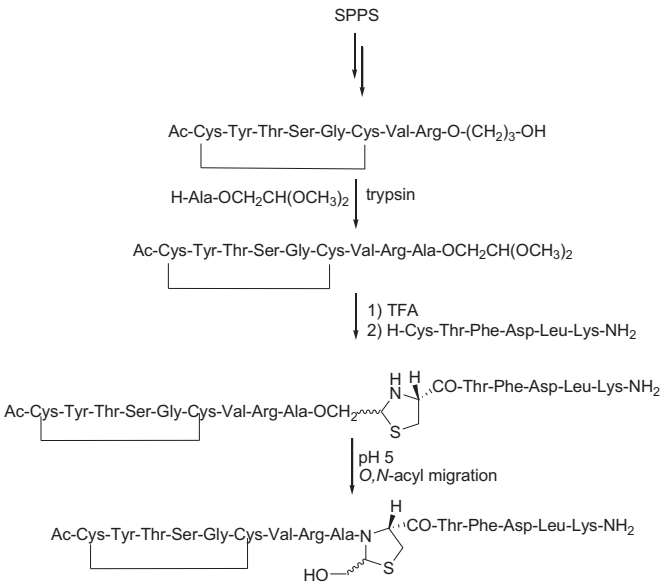
Scheme 3. Pseudoproline mediated ligation.

An exceptional specificity of the reaction obviates the protection of the side chains present in the peptide fragments. The pseudoproline residue thus formed at the ligation site restores the amide backbone, but contains an additional 2-hydroxymethyl moiety. The protocol was first demonstrated for the synthesis of a pentadecapeptide by a 9+6 convergent approach using a nonapeptide aldehyde[22] (Scheme 4). The reaction was smooth and no byproducts and side reactions were observed. Mass spectral analysis (observed: 1747 Da; calculated: 1747.7 Da) confirmed the product. The protocol was then successfully applied to obtain a 99-*mer* HIV-1 protease analogue through a 38+61 approach.[23]

Thiaproline ligation proceeds efficiently in both aqueous and non-aqueous (pyridine–acetic acid) media. However, the oxaproline ligation requires anhydrous conditions and, even in this medium, the thiaproline ligation is about 1000 times faster due to the difference in basicity between –SH and –OH. The potential of Ile-Leu-Asn-Gly-*S*-*Pro*-Phe-Lys-Ile-OH, was isolated in 78% yield (Scheme 5).[21a]

Several peptide ligation strategies took birth with this success, but most of them were confined to the establishment of non-amidic linkages between the peptide segments. Some of the important protocols in this category are those that involved the formation of thioester, oxime, thioether, disulfide and thiazolidine bonds. A list of different ligation reactions yielding native and non-native amide bonds is furnished in Table 1.[24–32]

Interestingly, the thioester forming ligation caught the attention of Kent, who explored this simple reaction to synthesize the two enantiomers of HIV-I protease enzyme employing all L- and all D-amino acids (Fig. 1).[33] The proteolytic activity was determined through a fluorogenic assay in which the enantiomeric enzymes showed reciprocal chiral specificity in the hydrolysis of peptide substrates, i.e., L-enzyme hydrolyzed the L-peptide substrate

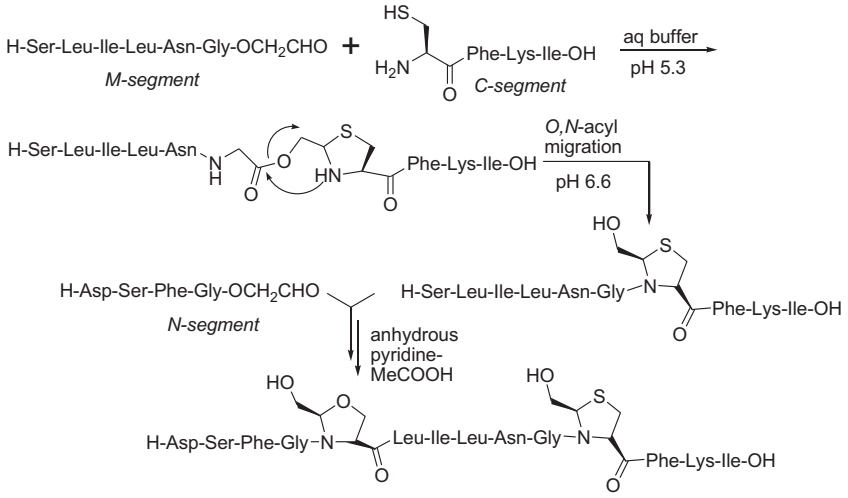


Scheme 4. Synthesis of a 15-*mer* peptide through pseudoproline ligation strategy.

of the thus-far developed chemistries for the chemoselective reaction of two unprotected peptide fragments (Fig. 2).[28] In the first disclosure, they described the reaction of an unprotected synthetic peptide-α-thioester with another unprotected peptide segment containing N-terminal Cys to generate a thioester-linked intermediate, which spontaneously underwent an intramolecular acyl migration to afford a native peptide (Scheme 6).

A platform for the NCL was laid as early as 1953 by Wieland when he observed that Val-SPh reacted with Cys in aqueous buffer to furnish the dipeptide Val-Cys (Scheme 6).[40]

A long time elapsed, however, until it struck the attention of Kent and Dawson[28] to turn it into major protein synthesis tool. A simple reaction is converted into a robust technique, which is a practically viable strategy for ligating two large polypeptide fragments. The reaction can be carried out in neutral or slightly basic pH and in the presence of guanidinium hydrochloride (Gnd·HCl), which acts as an aid to prevent aggregation of the peptide fragments. Although the exact mechanism has not been proved unarguably, based on the available experimental evidence and common observations, it is believed to involve the route shown in Scheme 7, where the first step is a transthioesterification involving the sulfhydryl group of Cys and peptide-α-thioester and the second step is an $S\rightarrow N$ acyl migration resulting in peptide bond formation.[41]



Scheme 5. Solvent driven tandem thiaproline and oxaproline ligations.

exclusively and vice versa. With this landmark success, the chemical synthesis of proteins progressed in leaps and bounds.

In the recent past, the techniques that deliver a native peptide bond at the ligation site have continued to show exponential growth.[34] The present report discusses the ligation techniques with respect to the underlying chemistry, developments and application of each type of technique. Much emphasis has been placed on native chemical ligation (NCL)[35] due to its vast utility as well as its future prospects in protein assembly. Other contemporary ligation methods, expressed protein ligation (EPL)[36] and Staudinger ligation,[37] are also delineated. The recently introduced concept, *O*-acyl isopeptide also known as 'click' or 'switch' peptide method, useful for the chemical assembly of highly aggregation-prone polypeptides, is dealt with.[38] Readers are referred to several reviews and articles on these contemporary approaches for further insights.[24–26,39]

2. Native chemical ligation

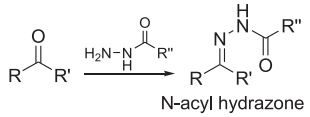
In 1994, Kent et al. developed the innovative idea popularly known as 'native chemical ligation'. It was an ingenious extension

The former step is reversible and rate determining, and depends on the nature of the substituent present on the leaving thiol moiety. In the presence of *endo*-Cys residues in either segment, free ε-amino, phenolic hydroxy, indole, imidazole or guanidine groups are not acylated, since the reaction is too selective for terminal Cys. Consequently, no byproducts are formed. Kent opines that the modest enthalphic activation of the peptide thioester, the use of pH 6.8–7.0 and the stability of the thioester moiety towards hydroxide or hydroxy nucleophiles contribute to such excellent chemoselectivity.[35a] The practical utilization of NCL owes its effectiveness to the unique properties of thioesters, which are stable to hydroxide-catalyzed hydrolysis, but very much susceptible to thiolysis and aminolysis. This stability–lability combination yields an adequate selectivity in NCL. Some important features of NCL are listed below:

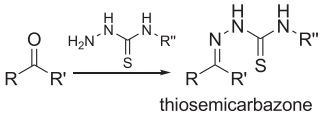
- This robust reaction is so highly chemoselective that it occurs only at the N-terminal Cys residue.
- The thioesters formed at other Cys residues are reversible and unproductive.

Table 1
Selected ligation techniques

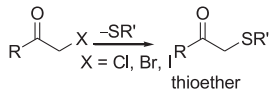
Non-amide bond yielding ligations



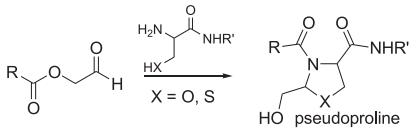
Geoghegen *et al.*, [17c]



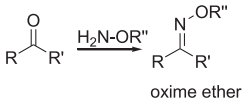
Bertozzi *et al.*, [26b]



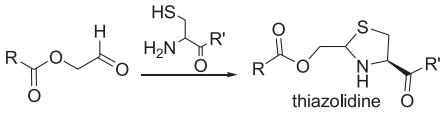
Robey *et al.*, [17d]



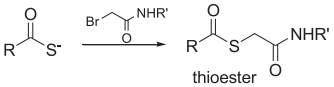
Tam *et al.*, [21a]



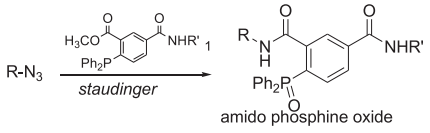
Monaci *et al.*, [17e]



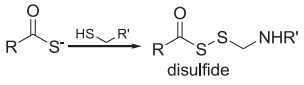
Tam *et al.*, [21b]



Kent *et al.*, [18]

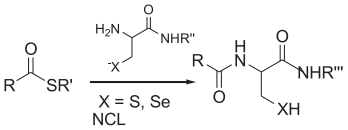


Raines *et al.*, [27b]

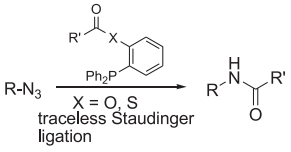


Tam *et al.*, [17f]

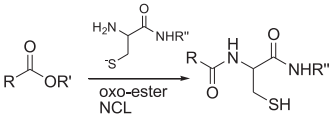
Amide bond yielding ligations



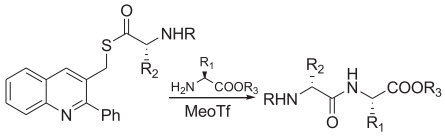
Kent *et al.*, [28]



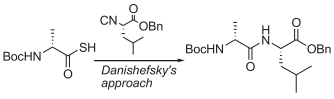
Raines *et al.*, [31b]



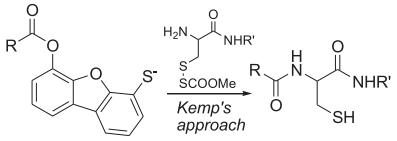
Danishefsky *et al.*, [29]



Levacher *et al.*, [32]



Wu *et al.*, [30]



Kemp *et al.*, [19a]

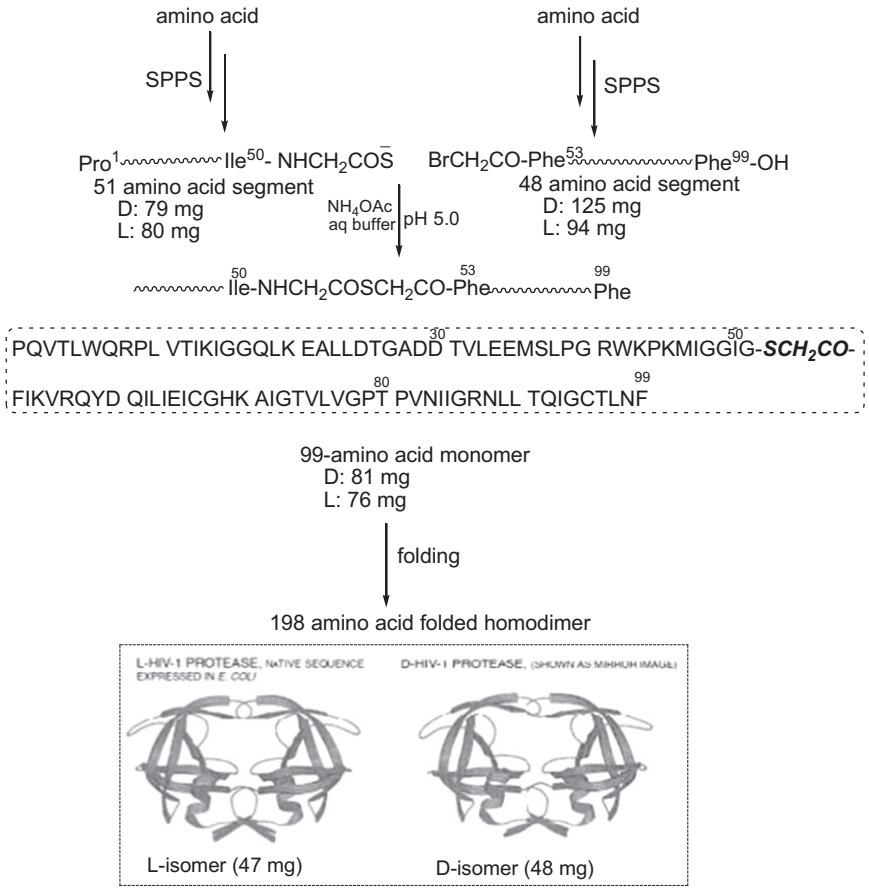


Fig. 1. Chemical assembly of HIV-1 protease.

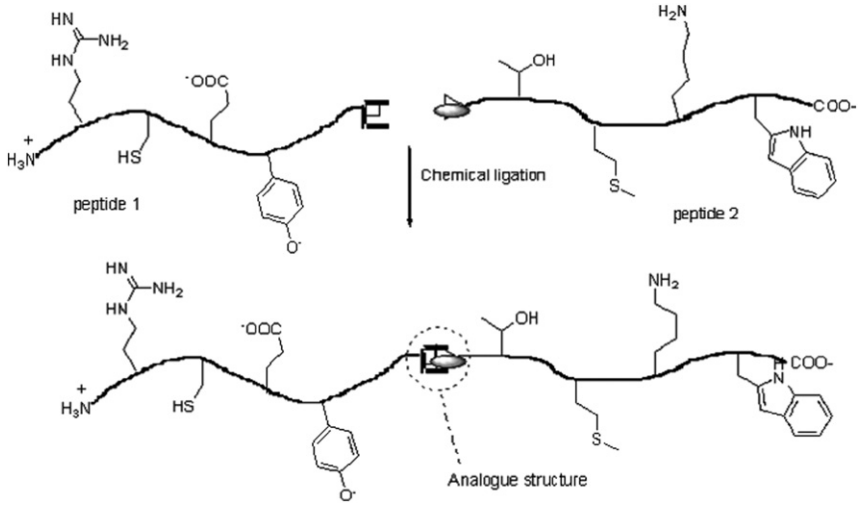
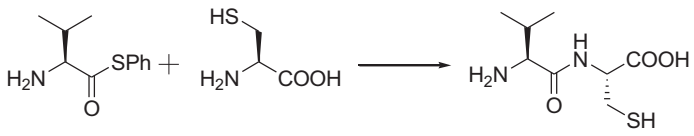
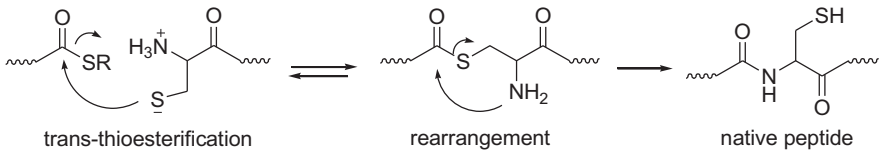


Fig. 2. A schematic representation of peptide ligation.



Scheme 6. Chemoselective reaction between a thioester and unprotected Cys.



Scheme 7. Mechanism of NCL.

- Protection of side chains is not a mandatory requirement. The unprotected segments are ligated under physiological conditions (near neutral pH).
- The mild thioester activation, non-basic reaction conditions and intramolecular mechanism of amide formation make NCL a racemization-free event.
- Thiol additives such as benzyl 2-mercaptoethanesulfonate and thiophenol are added to the reaction mixture to keep other thiol functional groups, if any, under a reduced state and to circumvent any undesired disulfide formation.
- Aryl thioesters are more reactive than alkyl thioesters. Hence, in the case of alkyl thioesters, some exogenous thiols such as benzyl mercaptan and 2-mercaptopyridine are used to promote the in situ formation of 'active thioesters,' thus increasing the kinetics of ligation.
- Ligations at sterically hindered C-terminal residues such as Val or Ile are slower, compared to the less-hindered Ala and Gly. The reaction at the latter occurs rapidly in minutes, while that at the former takes hours and, sometimes, days for completion.

In effect, the Cys-thioester-linked intermediate plays a key role, which after $S \rightarrow N$ acyl migration yields the target protein. However, attempts to trace the formation of the thioester intermediate by using HPLC techniques have not yielded any good results so far.[26] A summary of the protocols for the synthesis of thioesters[42] is given in Scheme 8. Nowadays, NCL is a defining point in establishing chemoselective ligation as a general synthetic route for the chemical assembly of proteins. This approach is enjoying vast attention, because of the development of general access to peptide-α-thioesters. Automation and advancements in SPPS essential for preparing the peptide fragments have strengthened the hands of researchers in achieving the success of NCL to an even greater extent.

proteinases.[43] NCL was the key tool to assemble the active part (6–56) of this peptide. Notably, the protein contains three Cys residues in the sequence at positions 24, 35 and 38, which allow a wide choice for selection of the ligation site (Fig. 3). A ligation of OMTKY3(6–23) with OMTKY3(24–56) was designed at the Cys[24] junction and the required fragments were prepared employing Boc-Cys(*p*-MeBzl)OCH_2-phenylacetamidomethyl (PAM) resin using *tert*-butoxycarbonyl (Boc)/2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) chemistry. Ligation was carried out using 0.25 mmol of each fragment in 6 M Gnd·HCl in the presence of 1% benzyl mercaptan and 3% thiophenol at pH 7.5 for 36 h to afford OMTKY3 in 56% yield (75 mg), the molecular weight of which, determined by ES-MS (5579.8±0.8 Da), matched the calculated value (5579.3 Da).

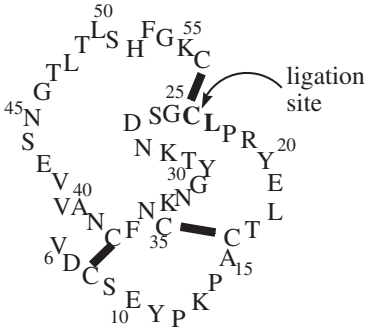
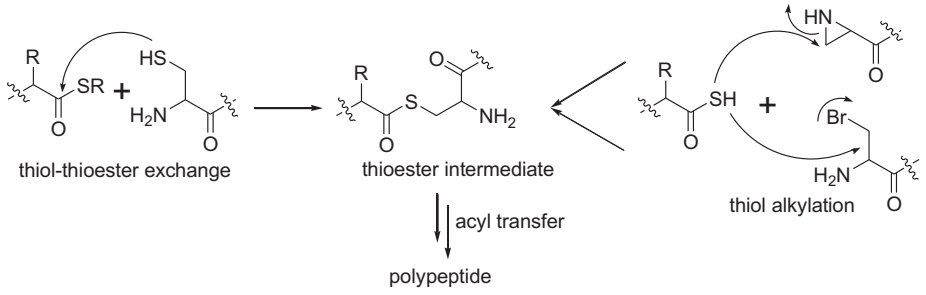


Fig. 3. Sequence of OMTKY3.

In order to determine the possibility of racemization at the C-terminal residue during ligation, a model study of the ligation site Leu[23]-Cys[24] was carried out. OMTKY3(14–23)-αCOSBzl was ligated to OMTKY3(24–33) to afford OMTKY3(14–33). A reference compound, OMTKY3(14–33)-D-Leu[23], was synthesized on 4-



Scheme 8. Pathways to transthioesterification.

2.1. Case studies of protein synthesis via NCL

In the very first report on NCL, Kent et al. demonstrated the total synthesis of human interleukin, a 72-amino acid protein.[28] The presence of Cys at position 34 facilitated this synthesis through a 33+39 convergent approach. This success triggered numerous reports on the application of NCL for the synthesis of proteins.

Kent group reported the synthesis of turkey ovomucoid third domain (OMTKY3), a potent protein inhibitor of most serine

methylbenzhydrylamine (MBHA) resin by the routine SPPS route and used as the standard. Comparison of the HPLC chromatograms of the two diastereomers revealed that no detectable racemization had occurred (<1%).[43]

For the synthesis of human group II secretory phospholipase A_2 ($sPLA_2$), a 124-amino acid residue enzyme found in α-granules of platelets and inflammator sites, three possible ligation junctions are available, at Asp[48]-Cys, Gly[58]-Cys and Thr[76]-Cys. The junction at position 58 was chosen to utilize the advantage of least steric

hindrance offered by the Gly residue (Fig. 4). The synthesis was performed by ligating PLA_2-(1–58) (9 mg/ml, 0.4 μmol) with PLA_2-(59–124) (9 mg/ml, 0.4 μmol).[44] The reaction was complete in about 22 h, as monitored by ES-MS and HPLC. ES-MS analysis of the crude sample confirmed the presence of the 124-amino acid protein chain (13,919.5±1.6 Da; calculated value 13,918 Da).

furnishing chemical and biological information on extracellular matrix and its conventional synthesis is intricate due to the large size (>1000 amino acid residues) and other synthetic problems. In the first step, the repeating-unit peptides equipped with thioester moieties on the C-termini and Cys on the N-termini were prepared and then subjected to intermolecular

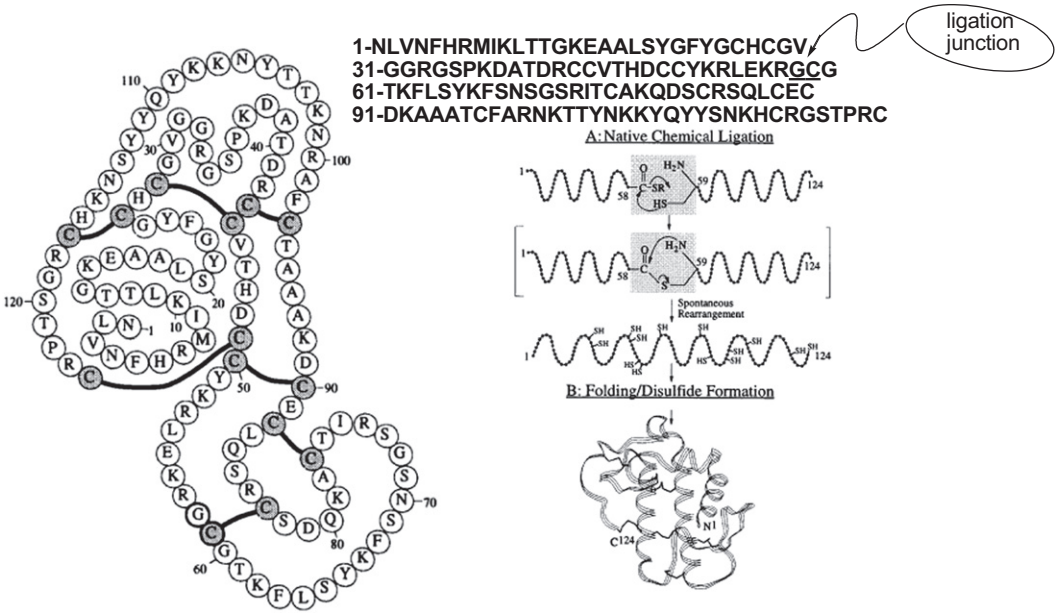


Fig. 4. Sequence of human type II secretory phospholipase A_2.

The product was purified through preparative HPLC and folded to afford the protein $sPLA_2$, which had a mass of 13,904.7+0.8 Da (calculated value 13,904 Da). The decrement of 14 Da in the mass value, as expected, was due to the formation of seven disulfide bridges during folding. The synthetic protein was found to possess both enzymatic and anticoagulant activities.

Lu et al. accomplished the chemical synthesis of a 100-amino acid protein, human psoriasin (S100A7), employing NCL.[45] S100A7 is a member of the S100 family of calcium-binding proteins, expressed abundantly in the keratinocytes of psoriasis patients. The occurrence of a Cys residue at position 46 allows the convenient ligation of two almost equal-length peptide fragments, Ps(1–45)-α-COSR and Ps(Cys^{46}-100). The fragments were prepared on solid phase using Boc chemistry. The NCL carried out on a 14-μmol scale under standard conditions afforded 74 mg of the final product, corresponding to 46% yield (Fig. 5). The molecular mass of the reduced form of psoriasin as analyzed by ESI-MS was 11,367.9±2.0 Da, which is in agreement with the calculated value (11,367.8 Da).

After oxidation and folding, the ESI-MS of the protein was found to be 11,365.7±1.3 Da in agreement with the calculated value of 11,365.8 Da. The difference of 2.2 Da from the reduced form is due to the formation of one disulfide bridge. Psoriasin is known to be a potent destroyer of *Escherichia coli*. The antimicrobial activity of synthetic psoriasin was tested against *E. coli* ATCC 35218 using a standard colony-counting assay. A dose dependent antimicrobial activity was observed with an LD_{90}value of 15 μg/ml or 1.3 μM, which is comparable to the reported value (0.5 μM) for natural psoriasin (Fig. 6; percentage of killing is expressed relative to bacteria incubated with buffer alone for 3 h at 37 °C).

Paramonov et al. synthesized high-molecular-weight collagen-like polymers through self-propagated NCL sequences.[46] In tissue engineering, collagen-like polymers play a pivotal role in native chemical ligation in parallel experiments (each individual peptide 3 mg/ml in 10 mM phosphate buffer, pH: 7.3, 3 mg/ml dithiothritol, 10 h) to afford the high-molecular-weight collagen polymers (Scheme 9). Chain termination was induced as and when desired by hydrolyzing the terminal thioester moiety. Thus, the synthesis of polymers of molecular weight up to 1,660,000 Da, as determined by MALDI-MS, was achieved.

The utility of NCL has also been extended to synthesize multivalent protein and peptide dendrimers.[47] These hyperbranched polymers with a high density of functional groups are attractive scaffolds in drug-delivery systems. Meijer et al. employed NCL to prepare dendrimers of multiple antigen peptides (MAPs) containing 2–8 peptide chains.[48] For this, three different sized poly(-propyleneimine) dendrimers were selected as substrates and the terminal amino groups were functionalized with Cys by reacting with Trt-Cys(Trt)-OSu. Removal of Trt groups opened the dendrimers for NCL with peptide-α-thioesters. Ligation with Ac-Leu-Tyr-Arg-Ala-Gly-mercaptopropionic acid-leucine (Ac-LYRAG-MPAL) afforded a dendrimeric peptide tetramer, octamer and hexadecamer based on the number of Cys residues present on the dendrimer scaffold (Scheme 10).

Fujita et al. assembled highly pure lipopeptides with different peptide antigens over diverse branches of the lipid core peptide system (LPS).[49] The LPS consists of a lipid adjuvant at the C-terminus and conjugated through its N-terminus to an oligolysine multiple copies of one or more peptide antigens. Such molecules are potential multi-epitope vaccine candidates. NCL was effectively used to conjugate different antigens to the Lys tail of LPS through proper selection of protecting groups for the Cys residues (Scheme 11). By employing orthogonal protections on each of the Cys residues, up to four different epitopes can be ligated into the lipid core.

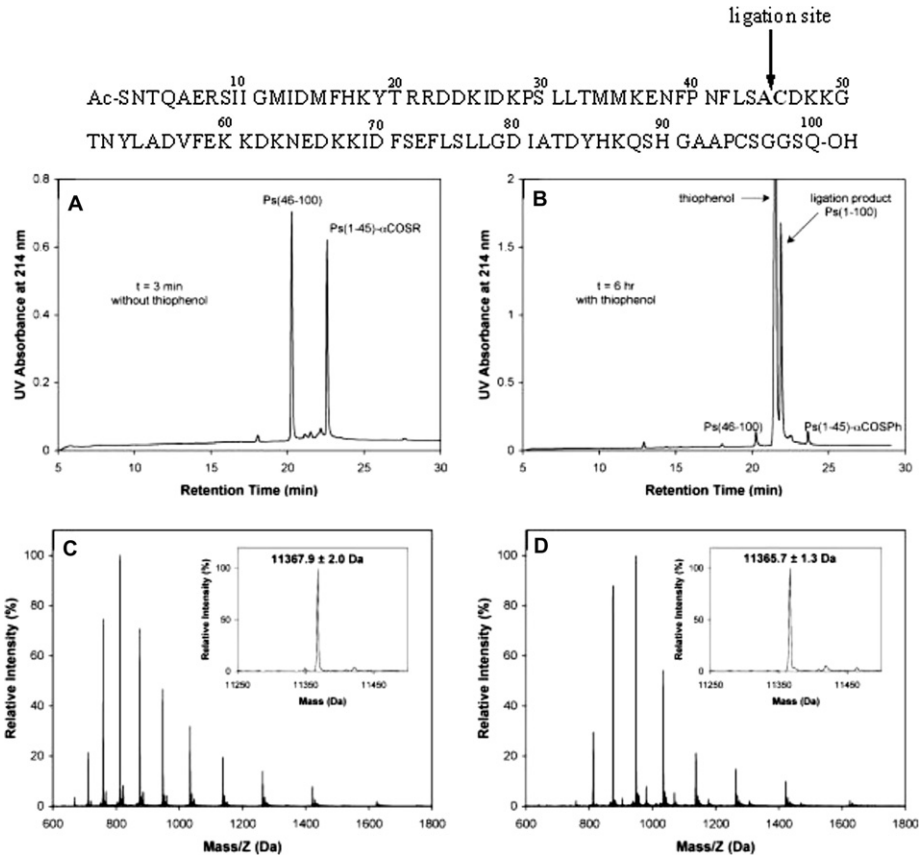


Fig. 5. HPLC chromatograms of ligation mixture: (A) before commencement of reaction and (B) after 6-h ligation. (C) ESI-MS of reduced S100A7. (D) Oxidized and folded form of S100A7.

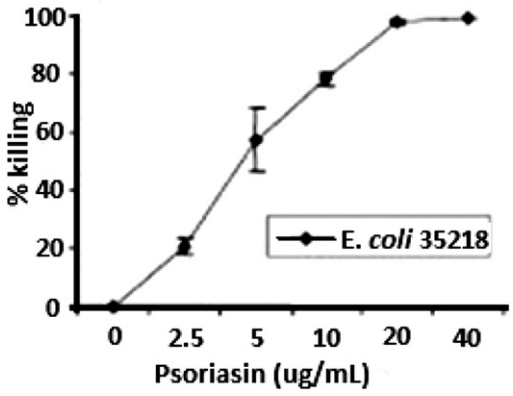


Fig. 6. Antimicrobial activity of synthetic psoriasin.

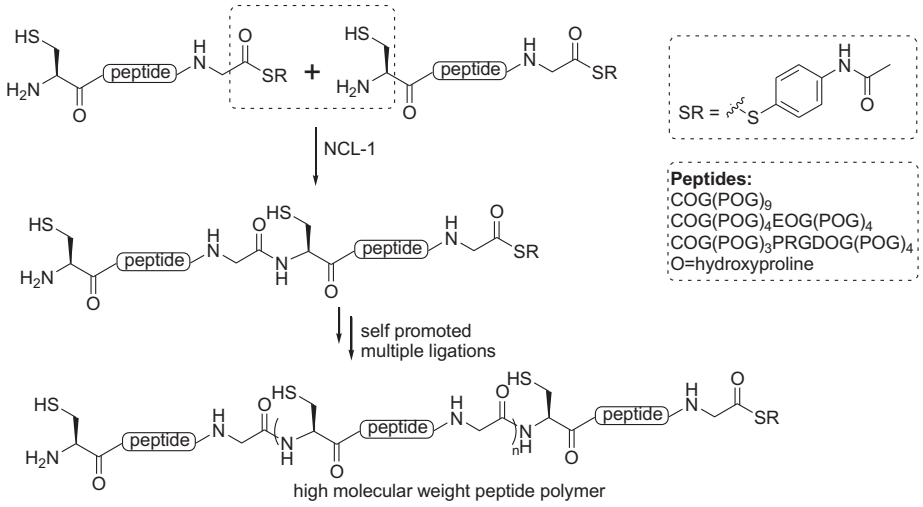
A wide range of molecules such as enzymes, enzyme inhibitor proteins, secretory proteins, adopter domains from intracellular signalling pathways, transcription factors and electron-transfer proteins have been synthesized via the NCL. The inexhaustive repertoire runs to more than 400 examples, some of which are furnished in Table 2.

Reports on the application of NCL to assemble hydrophobic peptides are scarce due to the synthetic challenges and difficulty in the isolation of hydrophobic peptide thioesters. Nevertheless, this area has been addressed by Naider, who reported the synthesis of fragments of G-protein coupled receptors (GPCRs). Determination of structural information on GPCRs has proven to be extremely challenging, due to their resistance to crystallization and the inability to examine them using high-resolution NMR techniques. Hence, short segments of GPCRs are being employed to understand the mechanism of action. NCL was adopted for the synthesis of peptides corresponding to the major portion of sixth and seventh transmembrane domain (TM6 and TM7) and a loop sequence connecting these two domains, the third extracellular loop (EL3).[55]

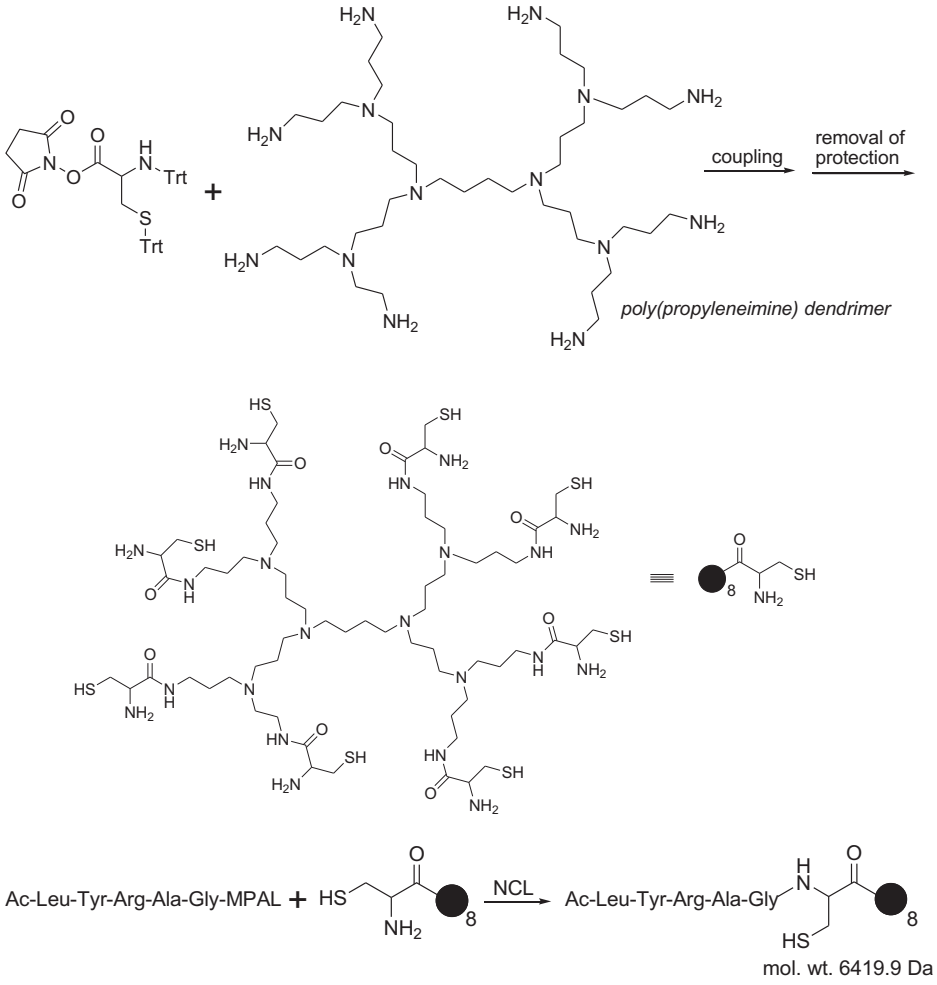
2.2. Synthesis of peptide-α-thioesters

The development of an efficient method for the synthesis of peptide-α-thioesters is critical to the success of NCL.[56] Conventional protocols are based on Boc-SPPS chemistry on a thioester-functionalized resin.[57] However, the repetitive acidolysis step inbuilt in Boc chemistry as well as the harsh cleavage conditions required for the cleavage of peptide-α-thioesters from the solid support (usually anhydrous liquid HF with scavengers) have inherent limitations, especially when acid-sensitive groups such as glycol, phosphono, etc. are present.[58,59] Switching over to Fmoc chemistry, as such, is not a viable route, due to the lability of a thioester linkage to the basic conditions employed for Fmoc removal after each coupling cycle. To address this issue, mild reagents have been employed for the cleavage of Fmoc groups, so as to keep the thioester linkage intact. Reagent systems such as: (a) a cocktail of 2% hexamethyleneimine, 2% 1-hydroxybenzotriazole (HOBt) and 25% 1-methylpyrrolidine in a mixture of *N*-methyl-2-pyrrolidone (NMP) and DMSO,[60] and (b) a non-nucleophilic base, 1,8-diazabicyclo[5,4,0]undeca-7-ene (DBU), in combination with HOBt in *N*,*N*′-dimethylformamide (DMF)[56b] have shown minimal decomposition of the thioester linkage during Fmoc cleavage. Apart from this, a handful of ingenious solid-phase protocols have also been developed, some of which are discussed below.

Bertozzi et al. demonstrated the synthesis of peptide-α-thioesters by Fmoc chemistry employing modified Kenner's sulfonamide

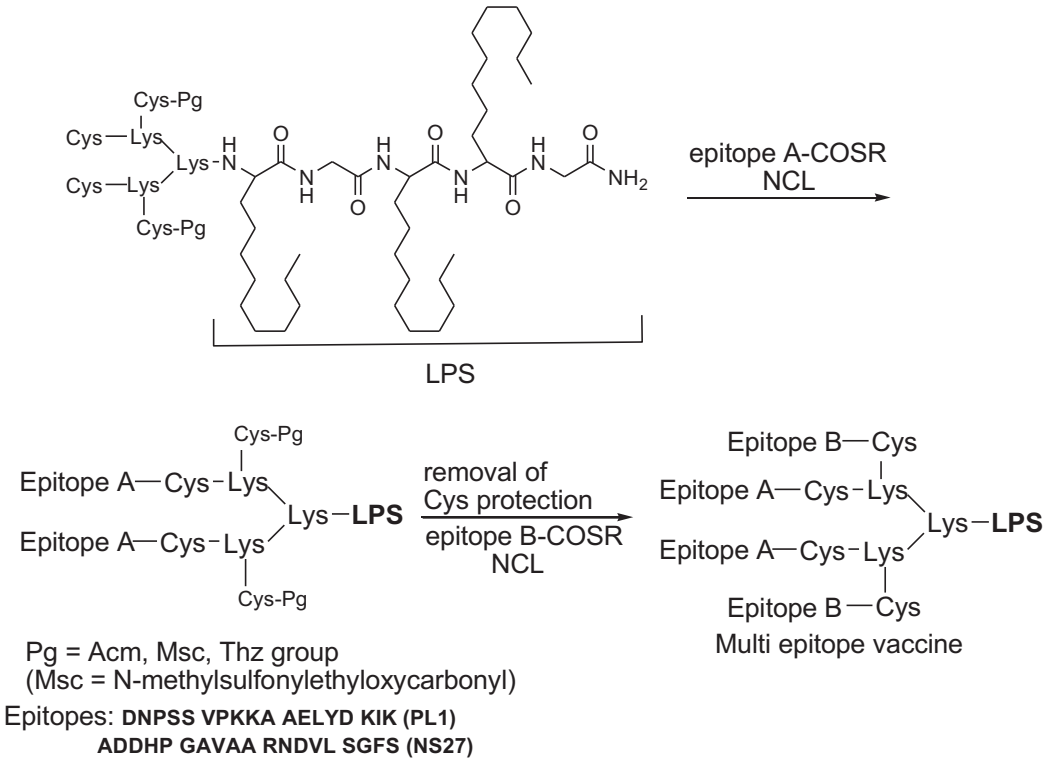


Scheme 9. Synthesis of collagen polymers via repetitive intramolecular NCLs.



Scheme 10. Typical example showing octameric scaffold and NCL.

safety-catch linker.[54] In this route, the carboxy group of the first amino acid is anchored to the resin via an acid- and base-stable *N*-acyl sulfonamide linkage followed by peptide assembly. For the detachment of the assembled peptide from the solid support, the sulfonamide moiety is activated by cyanomethylation and the peptide is released as a thioester by treating with thionucleophiles such as thiophenol. During the synthesis of an analogue of the antibacterial glycoprotein, diptericin, the thioester segment, Thr^{10}(Ac$_3$-α-D-GalNAc)diptericin(1–24)-COSBn, was assembled through this strategy. Initially, Fmoc-Gly^{24} was loaded onto 4-sulfamylbutyryl



Scheme 11. Utility of NCL for tailoring lipopeptides.

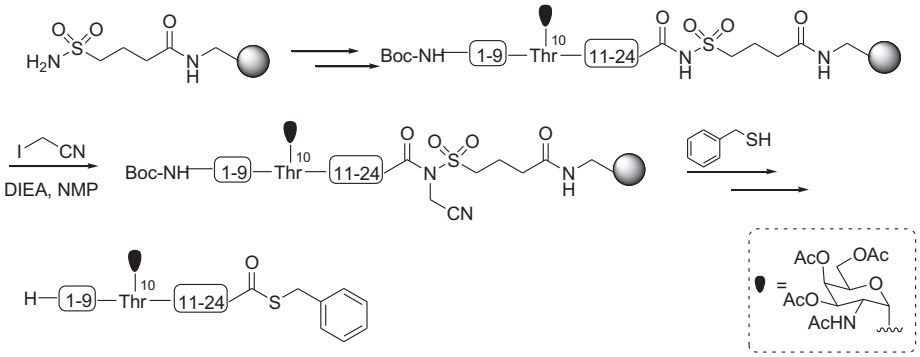
Table 2
Proteins synthesized through NCL

Protein	Chain length	Mol. wt. (kDa)	Ref.
Secretory			
Chemokines	72	8–10	28
Ser PR inhibitors	58–70	6–8	43
Anaphylatoxins	70	8	50a
AGRPs	46–112	7–15	50b
Phospholipase A_2	124	13	35b
Proteases			
HIV-1	2×99	18	18
Flap engineered HIV-1	2×99	18	50c
Influenza A virus M2	97	10–11	50d
Histone	135	14	50e
Zinc finger protein Zif268	81	11	51
Chymotrypsin inhibitor 2	64	7	50f
Enzymes			
Macrophage migration inhibitory factor (MIF)	115	39	52
Barnase	110	12	26b
Bovine ribonuclease	124	12–13	10
Receptors			
β2 Microglobulin	99	12	50g
CC chemokine CL14/HCC-1	66	8	50h
Intracellular			
b/HLH/Z	2×150	16–20	50i
Zn-finger	70	8	50j
Redox			
Rubredoxin	53	6	50k
Cytochrome *b*5	82	9	50l
Cytochrome *b*562	106	12–13	53
Antibacterial			
Diptericin (glycoprotein)	82	7–8	54

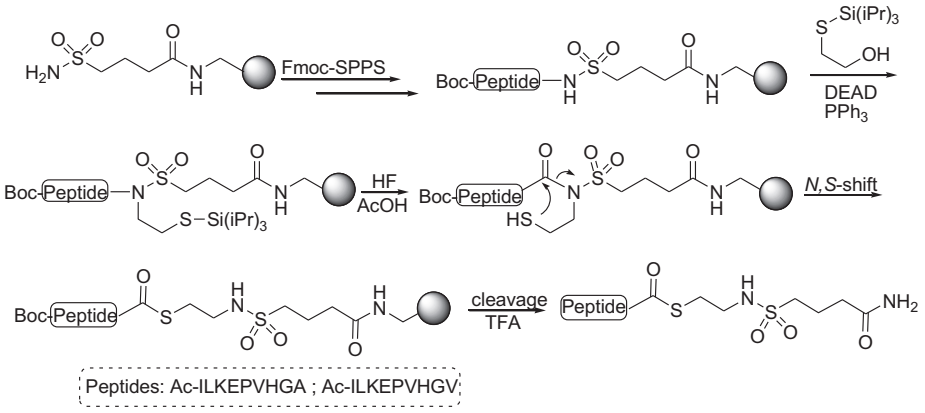
aminomethyl (AM) resin using (1*H*-benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP)/*N*,*N*-diisopropylethylamine (DIEA) and subsequent chain elongation was carried out using Fmoc/DCC/HOBt chemistry. After the final coupling step, the resin-bound peptide was alkylated using iodoacetonitrile to afford the corresponding *N*,*N*-cyanomethylacylalkanesulfonamide active ester. The peptide was cleaved as the peptide-α-thioester by treating with benzyl mercaptan followed by removal of the Boc and other side chain protecting groups using reagent K (82.5% TFA, 5% phenol, 5% water, 5% thioanisole and 2.5% EDT; Scheme 12). The crude product was purified by RP-HPLC to afford the peptide-α-thioester in 21% yield (ES-MS determined: 2843 Da; calculated value 2844.15 Da). The above strategy was also followed by Pessi et al. for the synthesis of peptide-α-thioesters such as MWAGADPA-SR, AEGEFMRASRCN PNCPMWAGADPA-SR, KKKSTWVLVGGVLAA-LAAY-SR [R=$-(CH_2)_2-CO_2Et$].[61]

A strategy involving an intramolecular $N \rightarrow S$ acyl shift was developed for the synthesis of peptide-α-thioesters on sulfonamide safety-catch linker.[62] The peptide was assembled on a 3-carboxypropanesulfonamide linker through Fmoc/*tert*-butyl-SPPS. The acylsulfonamide nitrogen was alkylated using Mitsunobu chemistry with a mercaptoethanol derivative. Removal of the triisopropylsilyl moiety exposed the free thiol group, which triggered the $N \rightarrow S$ acyl shift, furnishing the desired peptide-α-thioester attached to the solid support (Scheme 13). It was either cleaved from the solid support to afford the peptide-α-thioester or used directly for on-resin NCL.

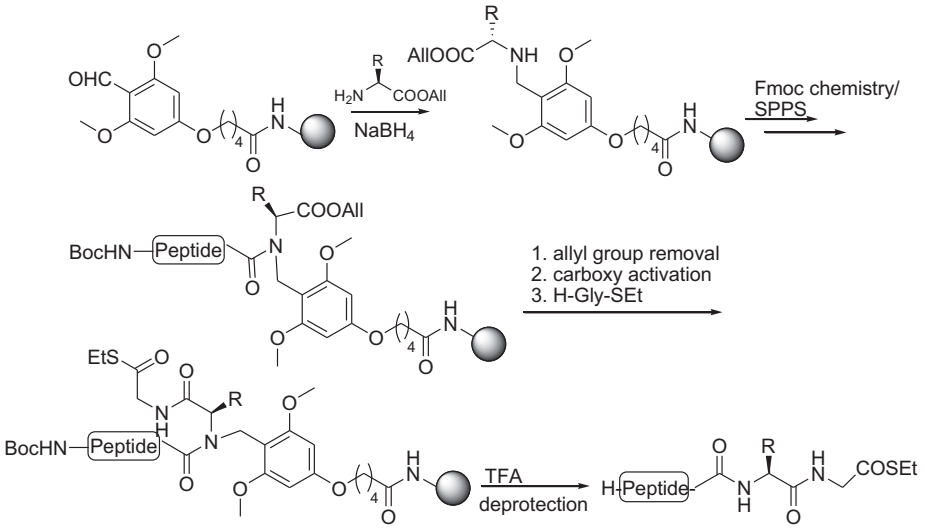
Barany et al. introduced a new route for the solid-phase synthesis of peptide thioesters using a backbone amide linker (BAL).[63] The key aspect in this study is that the growing peptide is attached to the solid support through the amino group of the first amino acid. The carboxy terminus of the first amino acid is kept protected usually as the allyl ester throughout the synthesis. After the assembly of the peptide, the allyl group is removed and the carboxyl is then either directly thioesterified or coupled to an amino acid thioester (Scheme 14).



Scheme 12. SPPS of peptide-α-thioesters employing modified Kenner's sulfonamide safety-catch linker.



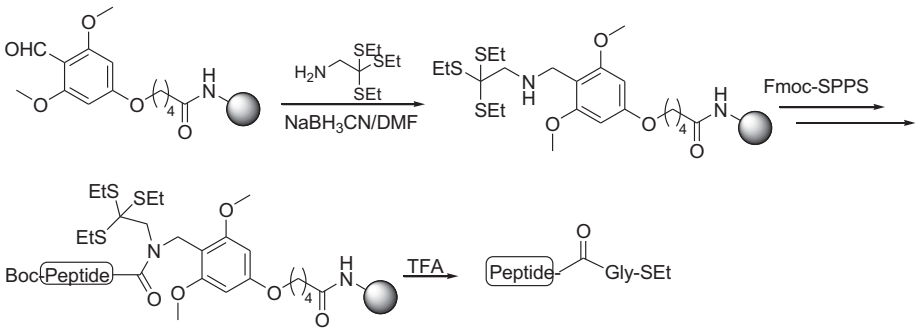
Scheme 13. SPPS of peptide-α-thioester on Kenner's linker via *N*→*S* acyl transfer.



Scheme 14. BAL assisted on-resin synthesis of peptide-α-thioesters.

Another interesting BAL-based approach was reported from Albericio's laboratory.[64] A key step in this protocol is to introduce a Gly-derived amino trithioester into the BAL support and subsequent chain elongation on the N-terminus of the Gly-derived trithioester. The trithio group is resistant to nucleophilic attack and, hence, is compatible with Fmoc chemistry. After the complete assembly of the peptide, a universal cleavage step is carried out employing trifluoroacetic acid (TFA) to afford the free peptide thioester (Scheme 15). In a typical example, the synthesis of Phe-Val-Lys-Glu-Tyr-Ala-Gly-SEt was demonstrated, which was isolated in 42% overall yield.

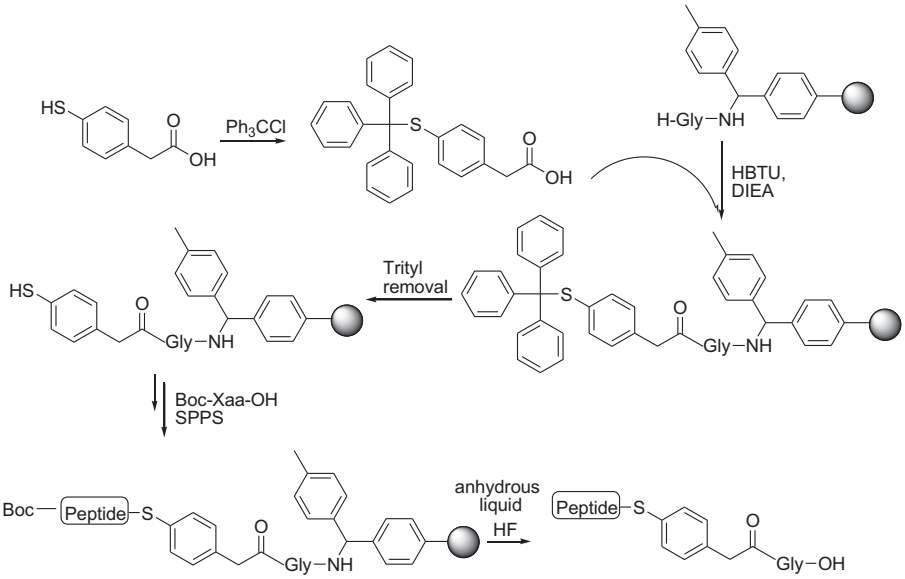
A simple chemistry for on-resin generation of peptide thiophenyl esters was reported by Kent et al. employing a *p*-methylbenzhydrylamine (MBHA) support.[65] *S*-Tritylmercaptoacetic acid,



Scheme 15. BAL-based strategy for synthesis of peptide-α-thioesters.

prepared by the reaction of 4-mercaptoacetic acid with trityl chloride (Trt–Cl), was coupled to the resin-bound Gly. Following the removal of Trt protection employing triisopropylsilane, peptide assembly was carried out employing Boc chemistry. At the end of the procedure, the assembled peptide thiophenyl ester was cleaved from the resin support using anhydrous liquid HF (Scheme 16).

weight protein synthesis has remained a daunting task, due to the limitations of the prevailing protocols (such as SPPS) for the preparation of the large ligatable segments required for the assembling. This problem, however, was addressed to a certain extent by performing the convergent ligation of multiple fragments.[68] Thus, in a typical protocol, two fragments are ligated first. The



Scheme 16. Peptide-α-thioester synthesis via Boc-SPPS on MBHA resin.

Recently, Sharma and Crich developed a novel strategy for the synthesis of peptide-α-thioesters.[66] The Fmoc chemistry-compatible method involved assembling a peptide onto a glycylaminomethyl resin via a thioglycinamide bond. An Fmoc-Xaa-Gly-O^{t}Bu was thionated and the resulting thiopeptide was attached to the resin via an amide bond. After chain elongation, the thioamide was alkylated with benzyl bromide to afford the corresponding thioimide, which on treatment with TFA led to the release of peptide from the resin as the corresponding thioester (Scheme 17).

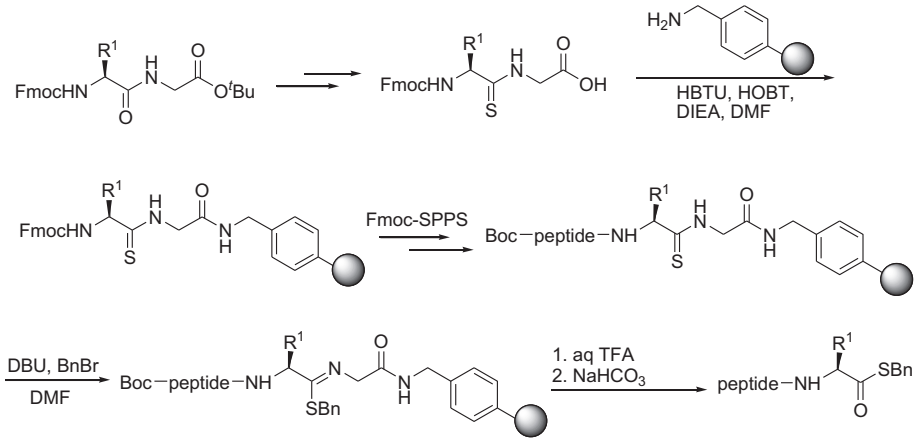
Apart from a variety of chemical methods, recombinant synthesis has also been developed as an alternative and efficient avenue for accessing large polypeptide thioesters.[67] Studies on protein splicing have enabled the scission of large-peptide fragments as the corresponding thioesters through bioengineering.[42]

2.3. Convergent synthesis through multiple ligations

Although NCL is undoubtedly the predominantly practiced protocol for the ligation of two peptide fragments, high-molecular-

resulting intermediate polypeptide is then subjected to ligation with another fragment and so on. Of considerable concern in this strategy is that the middle fragments would bear Cys at the N-terminus and thioester at the C-terminus, which would result in undesired byproducts through intramolecular ligations. To combat this, either of the two termini must be masked temporarily while operating at the other end. Soon, strategies were also developed to carry out the tandem ligations in a one-pot fashion by modulating the reaction conditions, usage of proper protecting groups and controlling the reaction kinetics.

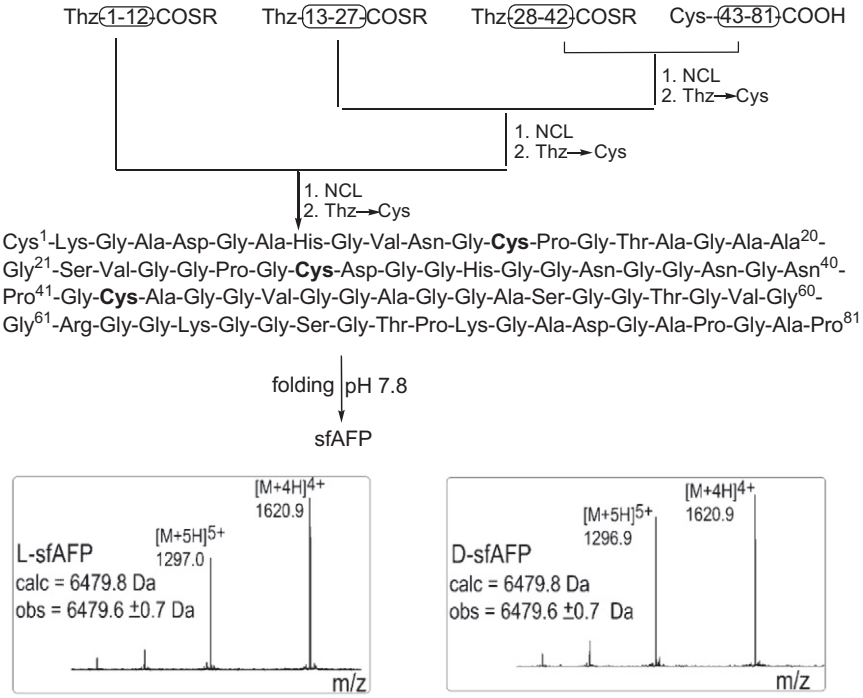
Kent et al. synthesized snow flea antifreeze protein (sfAFP, an 81-amino acid sequence) employing four-segment three sequential ligations.[51] The polypeptide chain of sfAFP contains Cys residues at positions 1, 13, 28 and 43. Its sequence was dissected into four peptide segments, [1–12], [13–27], [28–42] and [43–81], which were prepared through conventional SPPS. During the course of the synthesis, initially the third and the fourth segments were ligated keeping the Cys^{28} of the third segment protected as 1,3-thiazolidine-4-carboxylate (Thz). The resulting fragment, Thz^{28}-Pro^{81}, was purified



Scheme 17. On-resin thioesterification over a pre-defined thioamide on the peptide backbone.

and Thz was opened to generate Cys^{28} by treatment with methoxyamine hydrochloride (pH 4.0) and then ligated to the Thz^{13}-Gly^{27} fragment. Finally, ligation with the first segment, [Cys-Gly^{12}], was carried out (Scheme 18). All the ligations were carried out in Na_2PO_4 buffer in the presence of 6 M Gnd·HCl, pH 6.8 and a thiol. Using a similar strategy, the other enantiomer, D-sfAFP, was also synthesized. All the fragments were assembled on a 0.4 mmol scale. D[Thz^1-Gly^{12}] was obtained in 33% yield (80 mg); D[Thz^{13}-Gly^{27}] (100 mg) and D[Thz^{28}-Gly^{42}] (45 mg) in 22.5% yield; and D[Cys^{43}-Pro^{81}] in 41% yield (300 mg). Both L- and D-isomers were folded separately in phosphate buffer (pH 7.8) in the presence of 8 mM cysteine and 1 mM cystine·HCl at 4 °C. Ice recrystallization inhibition studies confirmed that both the proteins bear similar antifreeze activities.

assemble this 90-amino acid chain. The Cys residues at positions 37 and 65 were chosen as ligation centres and the fragments, Zif(1–36) (MERPYACPVESCDRRFSRSD-ELTRHIRIHTGQKPFQ), Zif(37–64) (CRICMRNFSR SDHLTTHIRTHTGEKPFA) and Zif(65–90) (CDICGRK-FARSDERKRHTKIHLRQKD), were synthesized by manual SPPS. Stepwise ligation was performed, during which, initially, 1.7 mM of Zif(37–64) was ligated to 1.7 mM of Zif(65–90) under the usual ligation conditions by keeping Cys^{37} protected with an Msc (methylsulfonylethoxycarbonyl) group to circumvent undesirable self-ligation. After removing the protection of Cys^{37}, the resulting Zif(37–90) was isolated in 34% yield (3.7 mg), which was then ligated to 0.93 mM of Zif(1–36) to afford 1.75 mg of Zif268 in an overall yield of 30%.



Scheme 18. Synthesis of sfAFP along with mass spectra of L- and D-isomers of folded protein.

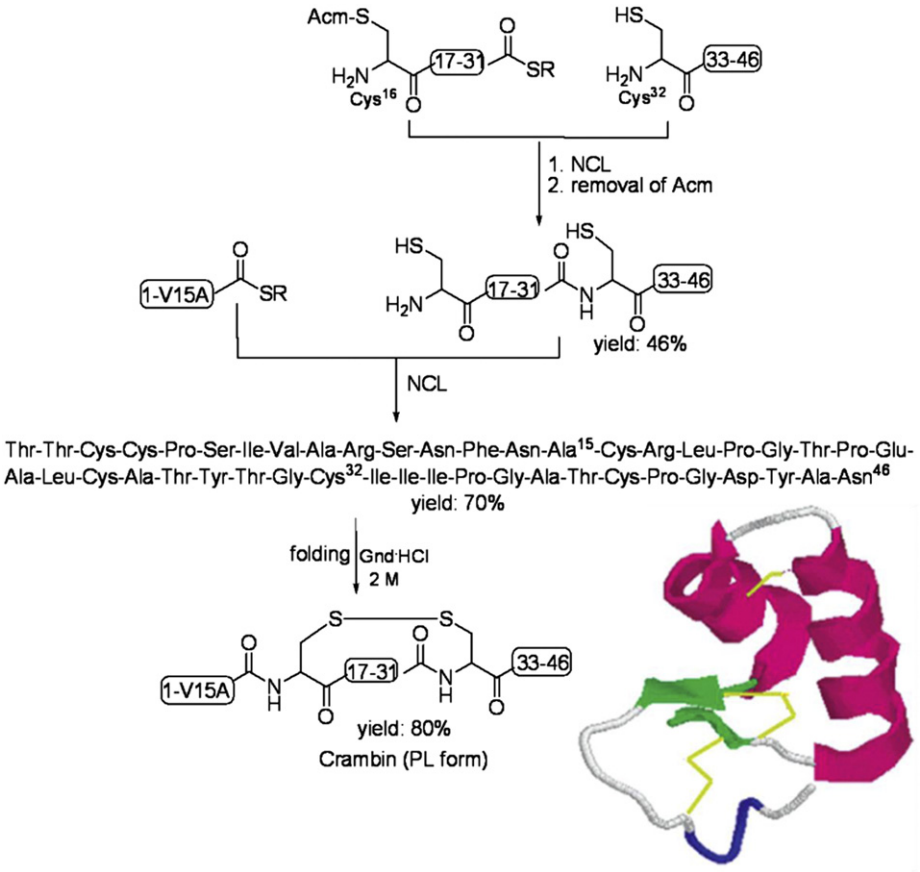
Beligere and Dawson successfully made a transcription factor zinc finger protein, Zif268, employing NCL.[69] Zif268 contains three DNA binding finger domains, due to which it recognizes a broad range of DNA sequences. A two-step, three-segment approach was followed to

During the synthesis of crambin (46-amino acid residue), the usual approach of bisegment coupling between Cram[1–15]+Cram [Cys^{16}-46] turned out to be low yielding. There are two forms of crambin, which differ by the amino acids present at positions 22

and 25. One having Ser[22] and Ile[25] is known as the SI form, while the other with Pro[22] and Leu[25] is called the PL form. When these proteins were assembled through NCL, both products were found to have a major impurity of the analogous peptide with one Thr residue less. Careful studies revealed that this was due to the deletion of Thr[21] during SPPS, the incorporation of which required at least three repeated couplings. Consequently, Kent employed a three-segment two-ligation strategy. It was demonstrated by the assembly of the PL form of crambin [i.e., via (Cram(1-V15A)COSR+-Cram(Cys(Acm)[16]-Gly[31])COSR)+Cram(Cys[32]-Asn[46])].[70] The thiol group of the Cys[16] residue of the middle fragment was blocked employing an acetamidomethyl (Acm) group until the completion of first ligation with Cram(Cys[32]-Asn[46]), the third segment. The ligation of the middle and the third segments was complete in about 6 h. The product was isolated and the Acm group was removed to afford Cram(Cys[16]-46), which, in the last step, was ligated to the first segment, Cram(1-V15A)COSR, to obtain the full-length linear peptide in 70% yield after purification (Scheme 19). In the first segment, V[15] was replaced with Ala to ease the steric hindrance during ligation. The polypeptide was subjected to folding in 2 M Gnd·HCl (about 1 h) and then purified to afford the protein in 80% yield based on the linear peptide (Fig. 7).

isolation, was treated with 0.2 M methoxyamine hydrochloride to convert it into the desired Cram[Cys[16]-Asn[46]] and then ligated directly to Cram[1–16]COSR at pH 7.0 to obtain the full-length peptide. By proper tuning of the reaction conditions, protein folding was also accomplished in the same flask to afford crambin in 45% overall yield (12.5 mg). This protocol reduced the reaction time, eliminated the intermediate purification step and also elevated the overall yield.

Several protecting groups have been proposed for protecting Cys of the middle fragments during multisegment ligations.[72] However, such strategies necessitate the cumbersome isolation of free Cys-peptide from the excess reagents used for deprotection. Interestingly, Otaka et al. developed a photolabile *N*-protector for Cys, which was used in the tandem ligation of multiple segments.[73] After screening several derivatives of 4-(dimethylamino) esters, 4-(dimethylamino) phenacyloxycarbonyl (Mapoc) was found to be useful. Its utility in NCL was demonstrated by synthesizing the 32-amino acid human brain natriuretic peptide (hBNP-32; Scheme 20). The middle fragment, N^{α}-Mapoc-Cys[10]-25, and the third fragment, Cys[26]-32, were ligated in the presence of 0.3% thiophenol in phosphate buffer (pH 7.6) and Gnd·HCl. The fragment, Mapoc-Cys[10]-32, without isolation, was subjected to photoirradiation (100 W, high pressure, Hg lamp, hν>300 nm) for 30 min to cleave the Mapoc group and then the



Scheme 19. A three-segment two-ligation strategy applied for the synthesis of crambin.

In a separate synthesis, Bang and Kent demonstrated a one-pot assembly of crambin[71] via a three-segment two-ligation strategy similar to that described above with the difference that Cys[16] of the middle fragment was protected with an acid-labile Thz moiety instead of Acm. The ligation between Cram[Thz[16]-Gly[31]]COSR and Cram [Cys[32]-Asn[46]] was carried out on a 0.4 mol scale, which took about 20 h to complete. The resulting product, Cram[Thz[16]-Asn[46]], without second ligation was carried out with (Ser[1]-9)-SEt to obtain the complete peptide. Folding was also accomplished in the same flask by diluting with phosphate buffer followed by the addition of DMSO to afford the desired natriuretic peptide in 56% yield.

Bang and Kent developed a novel His_6 tag-based approach for carrying out multiple ligations.[74] The C-terminal-His_6 tag facilitates the isolation and handling of the intermediates. In principle, the

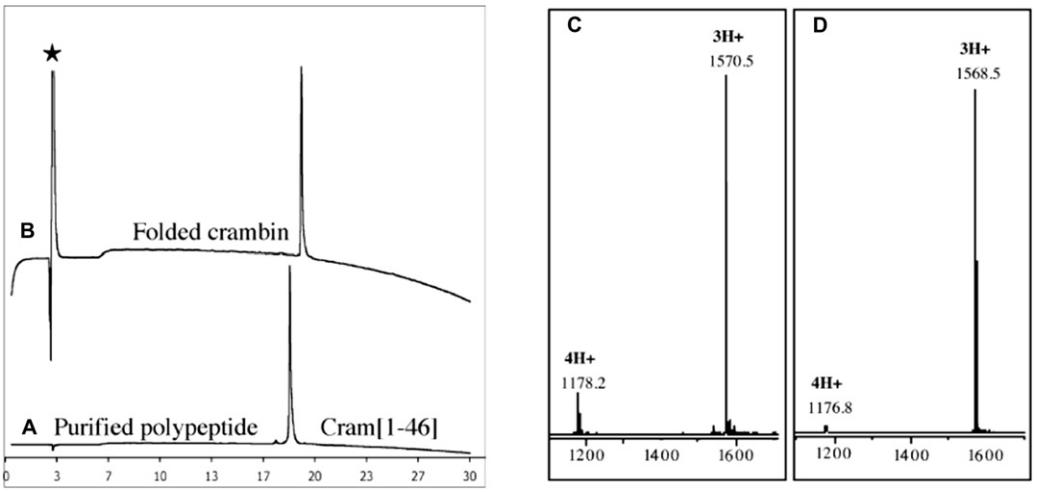
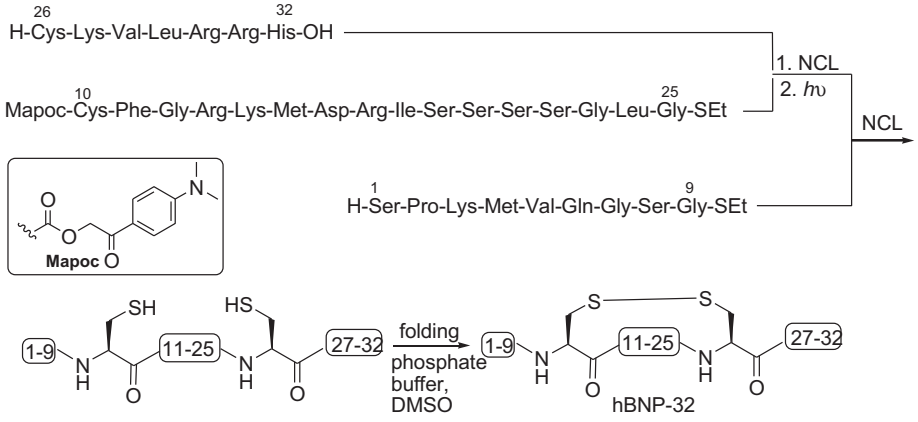


Fig. 7. (A) and (C): HPLC and mass spectra of purified crambin. (B) and (D): HPLC and mass spectra of crambin (PL form) after folding.



Scheme 20. Photolabile protecting group on the N-terminal Cys of the middle fragment facilitates multisegment ligations.

first Cys-peptide-1 fragment is attached to the His_6 tag followed by ligation with a Pg-Cys-peptide-2-COSR. At the end of the reaction, the SH protector is released and the peptide adsorbed on to a Ni–NTA (nickel–nitriloacetic acid) agarose column (the His_6 tags have high affinity towards this column). The undesired materials were washed, and, finally, the peptide-His_6 was eluted from the column using imidazole. At this stage, the product can either be isolated or subjected again for ligation with another Pg-Cys-peptide-COSR and so on. This ligation chemistry was demonstrated by the synthesis of crambin through a three-segment two-ligation approach (Scheme 21).

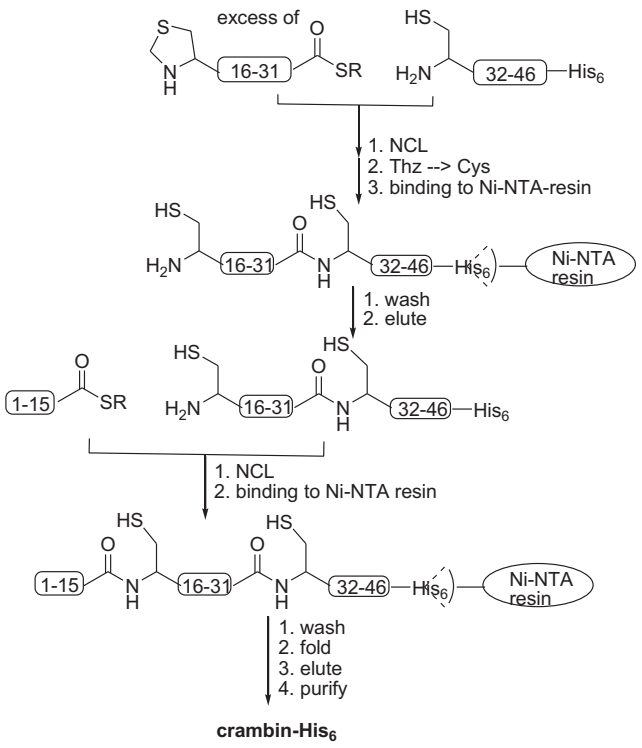
After the final ligation, the full-length peptide adsorbed on to an *N*-agarose support was subjected to folding. It was properly washed and the crambin-His_6 was eluted from the column.

The His_6 tag-assisted NCL was also employed for assembling the repeat protein of a 34-amino acid sequence, Thz-AWYNLGNAYYKQGDYDEAIEYYQKALELDPNNA-thioester, into its trimer (repeat proteins are formed from multiple copies of the same fragment).[74]

2.4. Solid-phase ligation (SP-NCL)

To increase the multiple-segment ligation efficiency and also to minimize repetitive purification steps and lyophilization of intermediates, thus circumventing the handling losses, a solid-phase ligation technique was developed.[72] The prerequisites are a suitable choice of resin linker, solid support and protecting groups. Employing a stable *s*afety-*c*atch *a*cid-*l*abile (SCAL) linker, the synthesis of vMIP-I, a 71-amino acid viral chemokine was accomplished via SP-NCL. To perform the sequential ligation on solid phase, the polypeptide sequence was dissected into three segments as vMIP-I(1–12), vMIP-I(13–35) and vMIP-I(36–71). The C-terminal segment was synthesized on a thioester resin with Gly-Gly spacer and SCAL linker. The other two segments, vMIP-I(1–12)COSR and vMIP-I(13–35)COSR, were separately prepared on solid phase; Cys^{13} and Cys^{36} were protected as the corresponding Msc derivatives. vMIP-Msc-(36–71)-SCAL-Gly-Gly-COSR was first attached to Cys-sepharose through NCL. After the removal of Msc protection on Cys^{36}, the middle segment, vMIP-Msc(13–35)COSR, was ligated. The deprotection and ligation cycle were repeated to attach the vMIP-I(1–12) segment also. The progress of the ligation was monitored through HPLC by following the peptide-α-thioester consumption. The cleavage of the peptide from the linker afforded vMIP-I(1–71). The whole set of ligation sequences was complete within 30 h (Scheme 22).

Kent's group demonstrated the solid-phase ligation of unprotected peptide segments from the *N*→*C* direction (i.e., starting from the polymer-bound N-terminal segment towards the C-terminus of the target polypeptide chain) as well as in the reverse way, i.e., *C*→*N* direction (polymer-bound C-terminal segment towards N-



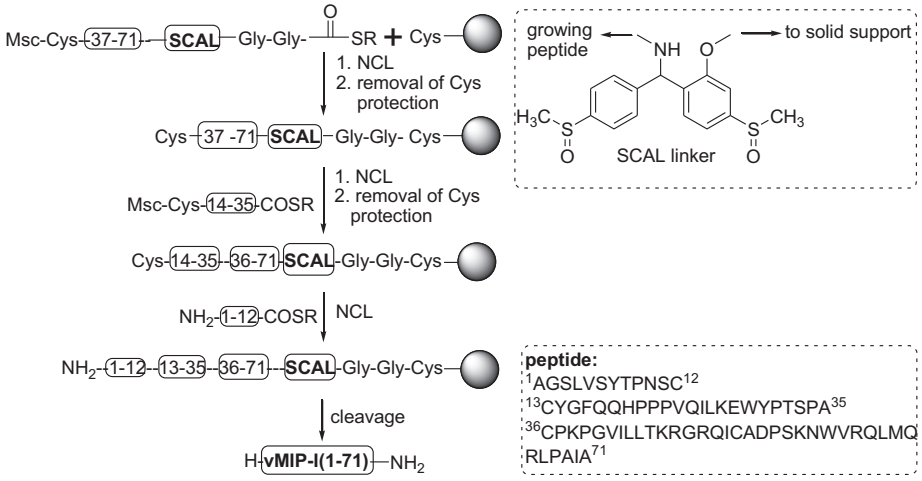
Scheme 21. C-Terminal-His_6 tag facilitates multiple ligations through simpler isolation of the intermediate peptides.

T[1]LQKKIEEIAAKYKHSVVKK[20]-COSH, C[21]CYDGACVNNDETCEQRAAR-ISLGPK[46]-COSH, and C[47]IKAFTECCVVASQLR-ANISHKDMQLG[73]R.

In order to carry out the ligation in the $C \rightarrow N$ direction, Cys-peptide was anchored on to a solid support and a Pg-Cys-peptide thioester was ligated under standard conditions. In the subsequent steps, the thiol protector of Cys was released and unreacted substrates and excess reagents were washed off. The peptide assembly was continued by ligating another Pg-Cys-peptide thioester to the resin-bound Cys-peptide. Cycles of ligation and deprotection were performed to assemble large polypeptides. Importantly, polypeptides obtained after series of ligation reactions were free from stereomutation.

2.5. Kinetically controlled NCL (KCL)

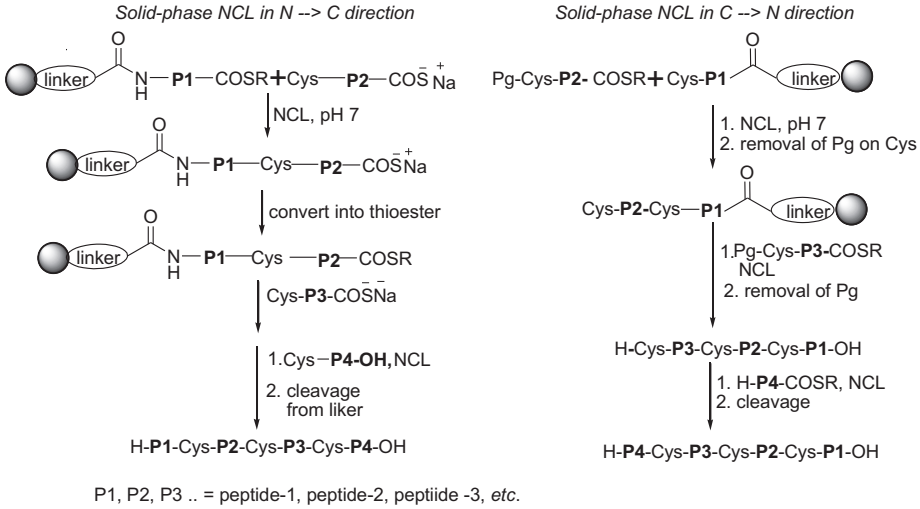
Another remarkable contribution by Kent's group to the convergent synthesis of proteins is the development of KCL.[75] In the tandem multisegment ligations, initially two peptide fragments are introduced to the reaction vessel and ligated. The third peptide fragment is introduced only after the completion of the first ligation. This multistep approach is necessary because of the practical difficulty of controlling the dual reactivity of Cys-peptide-COSR fragments. However, it has been established that the peptide alkylthioesters are relatively unreactive and the addition of thiophenol as catalyst increases their reactivity by forming an 'active' thiophenolic ester.[75] Thus, it was envisaged that, in an uncatalyzed competitive reaction between a peptide thioalkyl ester and a peptide thiophenyl ester, the latter would ligate with a Cys-peptide preferentially leaving the alkylthioester unperturbed. This refined strategy was applied for the assembly of crambin using a six-frag-



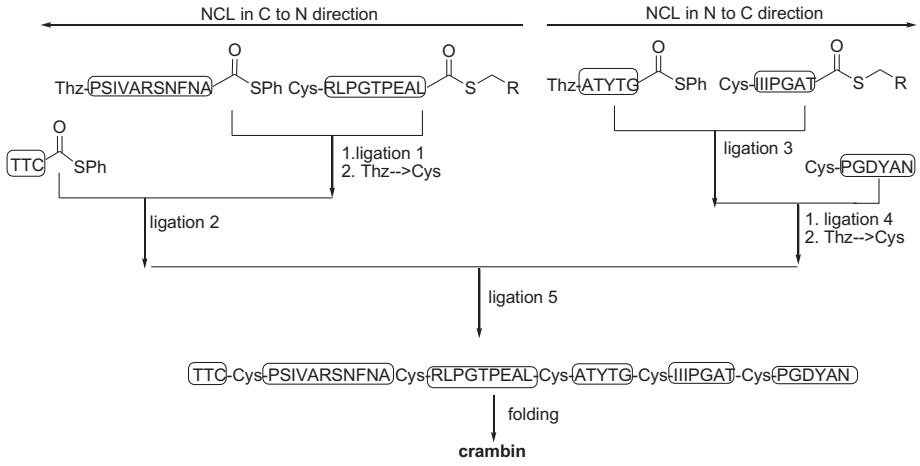
Scheme 22. Sequential ligations on solid support.

terminus of the polypeptide chain).[52] In the $N \rightarrow C$ ligation strategy, the middle segments were prepared as the corresponding Cys-peptide thiocarboxylates. The presence of thiocarboxylic acid in the form of thiocarboxylate ion itself serves as protecting group temporarily and allows for the selective and smooth ligation at the *N*-Cys peptide termini. After ligation, the thiocarboxylate was converted into the reactive thiocarboxylic ester by treating with bromoacetic acid in water and this was then subjected to ligation with another Cys-peptide thiocarboxylate. In the penultimate step, the final ligation was accomplished with Cys-peptide acid and then detached from the resin followed by HPLC purification and folding (Scheme 23). This strategy was demonstrated by assembling human C5a (a 74-amino acid plasma protein with potent chemoattractant and proinflammatory properties) employing three fragments, ment ligation protocol (Scheme 24). The first, second and third ligations were carried out in the absence of the thiophenol catalyst so as to direct the ligations only at the phenolic thioesters, during which the alkylthioester termini remained unreactive. The fourth and fifth ligations were performed in the presence of the thiol additive to activate the alkylthioesters. The yield for the first ligation and subsequent conversion of Thz into Cys was 71%, that for ligation 2 was 50%, 80% for ligation 3 alone but 40% when ligations 3 and 4 were carried out in one pot, and 62% for ligation 5 followed by folding.

Human lysozyme (a 130-amino acid sequence) was also assembled through a four-segment KCL.[76] Lysozymes serve as model systems in various areas of biomedical research including protein folding and enzyme catalysis. It was assembled via the ligation of



Scheme 23. On-resin ligations in N→C as well as C→N directions.



Scheme 24. Synthesis of crambin through KCL.

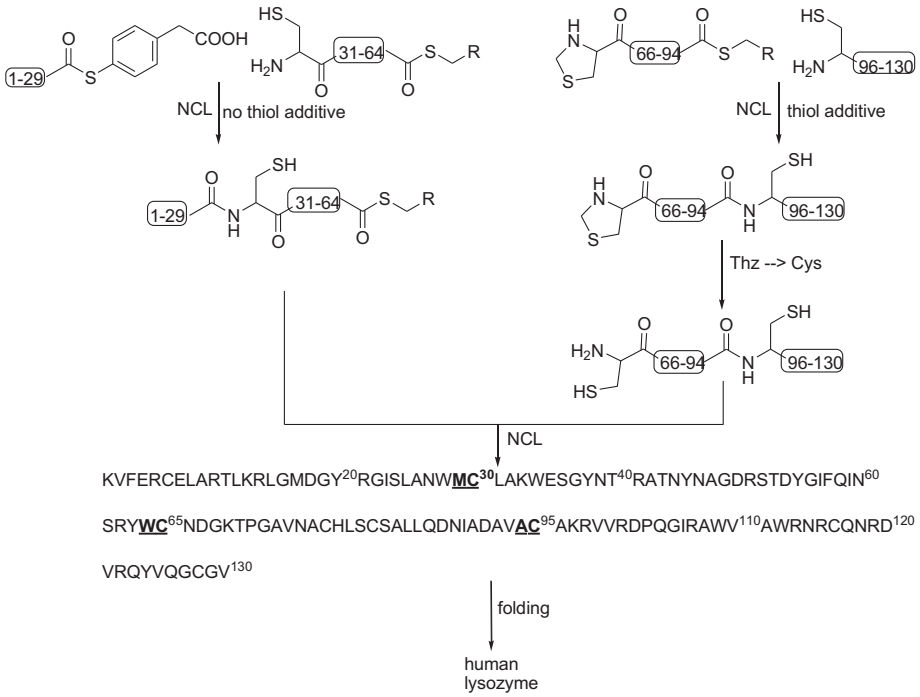
four fragments (Scheme 25). Through retrosynthetic disconnection, the protein was fragmented into four almost equal-length polypeptides—(1–29), (30–64), (65–94) and (95–130). All the fragments were synthesized employing in situ neutralization by Boc chemistry/SPPS. Ligations were carried out in 6 M Gnd·HCl and 0.2 M sodium phosphate buffer. After the completion of four ligations, the final peptide was purified using RP-HPLC (Fig. 8) and was characterized through mass spectroscopy (found: 14,700.2±1.5 Da; calculated: 14,700.7 Da). In the final step, the polypeptide was subjected to folding in a redox system (5 mM oxidized glutathione and 2 mM DTT at pH 8.0). It was lyophilized to afford 2.5 mg of the protein in an overall yield of 40% (Fig. 8).

2.6. Ligation at Glu/Asp-Cys site and limitations

NCL, in general, was considered to be feasible with peptide-α-thioesters having any of the 20 natural amino acids at the C-terminus by which access to any Xaa-Cys sequence at the ligation junction is possible.[35b] However, careful studies revealed that this is not so if Asp/Glu is present at the C-terminus. The presence of a β/γ-carboxy group in proximity to a thio ester results in the migration of the thioester moiety onto the side-chain carboxy group. This leads to an anhydride intermediate and the outcome of this would be a mixture of α-Cys-peptide and ω-Cys peptide (Scheme 26). This necessitates the use of a suitable protector for the side-chain carboxy moiety of Asp/Glu when present at the C-terminus. Botti et al. envisaged that such protecting groups must be easy to introduce and remove, and stable to cleavage and ligation conditions.[77,78] 9-Fluorenylmethyl ester (OFm), (phenylsulfonyl)ethyl ester (OPse) and a couple of protecting groups were thus explored and their utility in NCL was demonstrated. However, Fm and Pse groups are not suitable for Fmoc-SPPS chemistry owing to their base lability. To combat this, Briand et al. developed a photolabile protecting group that was fully stable to Fmoc-SPPS conditions and would be easily removable under photoirradiation without damaging the peptide backbone and other sensitive functionalities.[79] In this regard, the {7-[bis(carboxymethyl)amino]coumarin-4-yl}methyl (BCMACM) group was introduced and successfully employed to protect the γ-carboxy group of Glu during a model reaction and the protector was released post ligation to afford the target molecule.

2.7. Thiol auxiliary-mediated NCL

The necessity of Cys at the ligation site was a bottleneck for the wide utility of NCL and its application for the synthesis of non-Cys peptides was very much limited. For instance, during the synthesis



Scheme 25. Human lysozyme assembled via KCL.

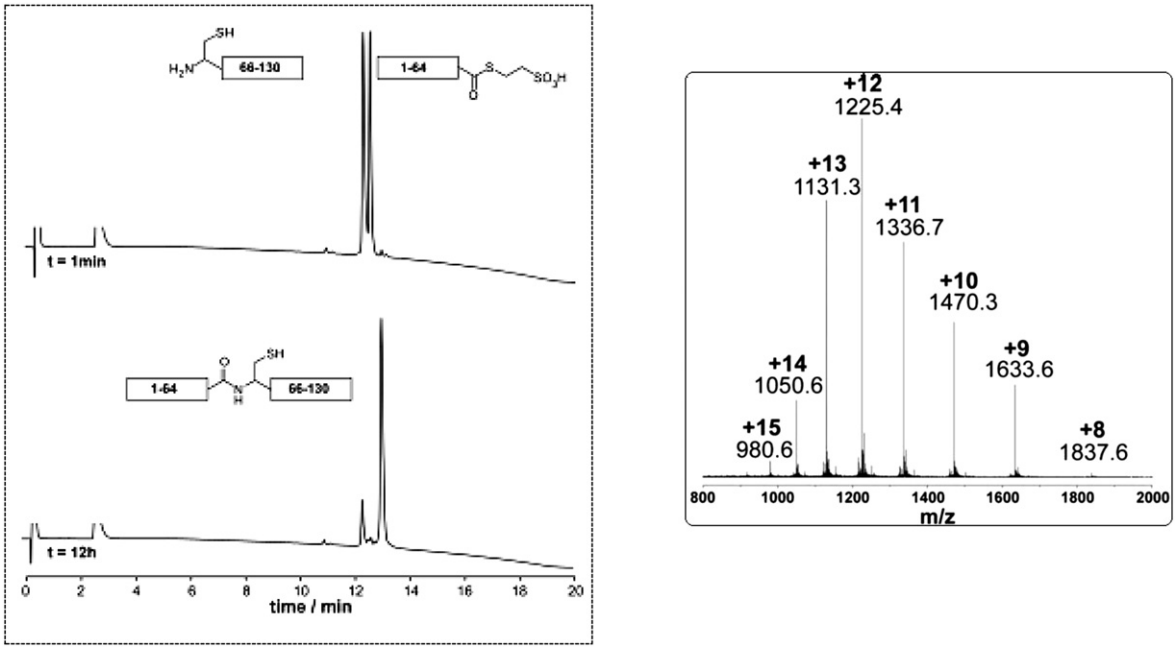
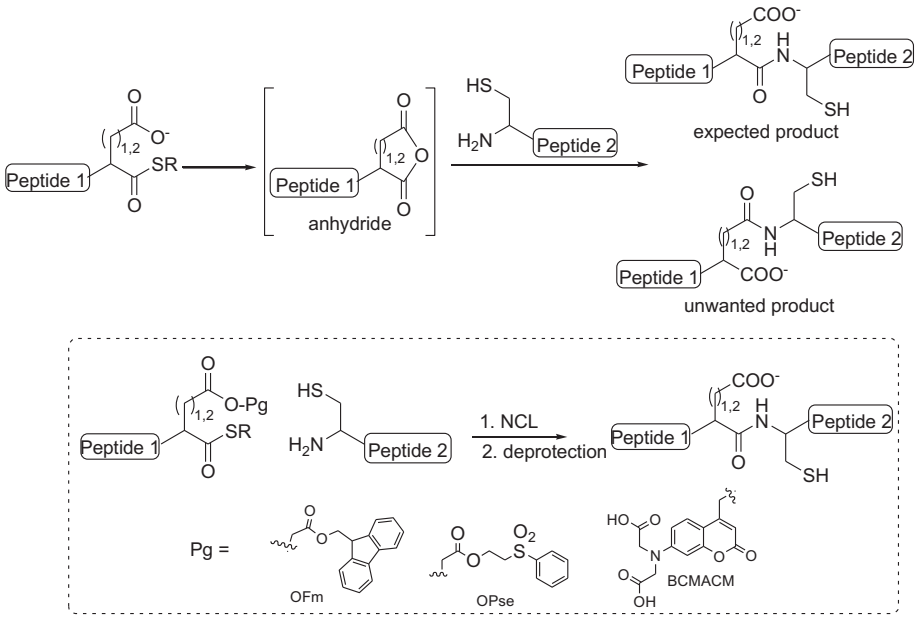


Fig. 8. HPLC profile of reactants and ligation product of human lysozyme. Inset depicts its ESI-MS (deconvoluted value: 14,693.4±0.7 Da; calculated value: 14,692.7 Da).

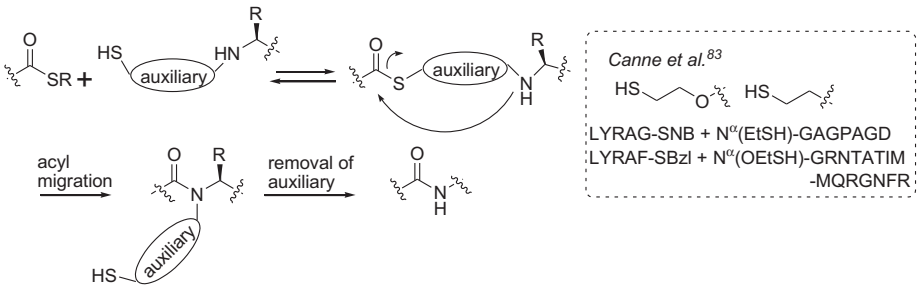
of a protein, barnase, a non-critical amino acid Lys was replaced with Cys to make the peptide segments suitable for ligation.[26b] As Cys is a less-common amino acid in proteins and comprises only 1.7% of the total count, the development of methods alleviating the need for the Cys residue has grown as an important sub-area of research. Consequently, the use has been reported of removable thiol auxiliaries that are attached to the N-termini of the peptide, which perform the functions of Cys during the ligation and can be chopped at a later stage to obtain, ultimately, the native peptides (Scheme 27).[80,81]

A useful auxiliary has to be introduced efficiently through a simple method, should be compatible with SPPS, must facilitate the ligation rapidly at ~1 mM concentration in denaturing aqueous buffer and needs to be removed selectively under mild conditions. The steric bulk of the auxiliary should not affect the rate of ligation.

Canne et al. developed ethanethiol and oxyethanethiol as N-auxiliaries and demonstrated their use in the Cys-free ligation of a few model peptides (Scheme 27).[82] Ligations were carried out in either 8 M urea/0.1 M Na_2HPO_4 at pH 7.0 or in 6 M Gnd·HCl/0.1 M Na_2HPO_4 at pH 7.5. Appreciable yields were obtained in each study.



Scheme 26. Ligation at Asp/Glu leads to a mixture of peptides.



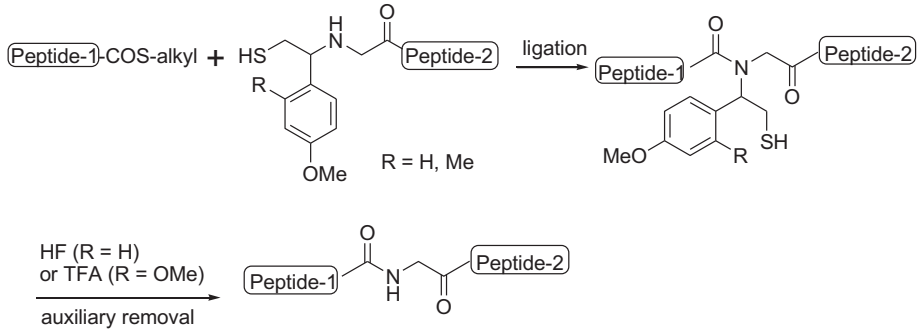
Scheme 27. Thiol auxiliary-mediated ligation.

After the ligation, the oxyethanethiol auxiliary was completely removed by using Zn/H^+.

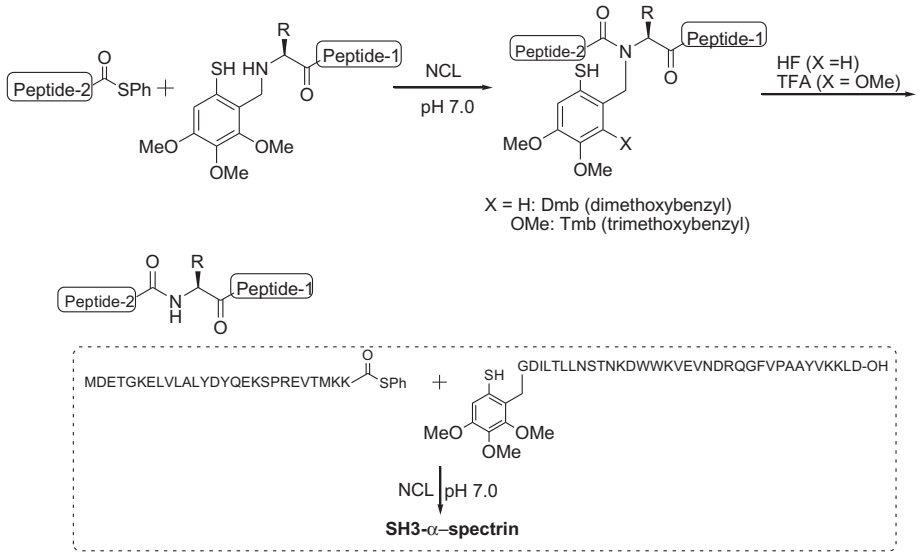
The use of the N^{α}-(1-phenyl-2-mercaptoethyl) group as auxiliary was successfully demonstrated for Cys-free NCL,[80] which can be removed under mildly acidic conditions in a post ligation step. A model peptide, N^{α}-(Aux)GlySerTyrArgPheLeu, was ligated to few peptide-α-thioesters including PheGlyGly, TBP-A 1-His[67], Mouse Larc 1-Ala[31], and MCPI 1-Lys[35] employing either N^{α}-(1-(4-methoxyphenyl)-2-mercaptoethyl) or N^{α}-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) as the auxiliary. After ligation, the auxiliary was removed via the acid treatment to afford a solitary yield of the target peptide (Scheme 28).

Dawson et al.[81] arrived at an improved class of 2-mercaptobenzyl-based auxiliaries, which were designed to be more acid labile and possess improved ligation efficacy by modulating the electron density. Model experiments were carried out to demonstrate the competence of each auxiliary to ligate at different junctions viz., Gly/Gly, Lys/Gly, Gly/Ala and Ala/Ala. A 62-amino acid sequence SH3 domain from α-spectrin was synthesized using a trimethoxy-substituted 2-mercaptobenzyl auxiliary (Scheme 29).

The ligation of SH3 α-spectrin[1–27] thiophenol thioester with SH3 α-spectrin Tmb[28–62] in 6 M Gnd·HCl/200 mM NaH_2PO_4 at pH



Scheme 28. N^{α}-(1-Phenyl-2-mercaptoethyl) group as removal auxiliary.

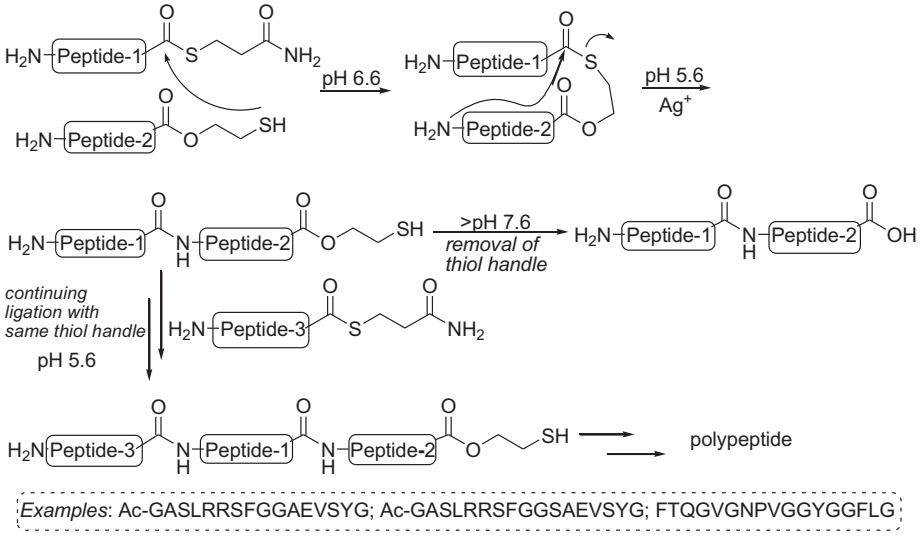


Scheme 29. 2-Mercaptobenzyl based auxiliaries for Cys-free NCL.

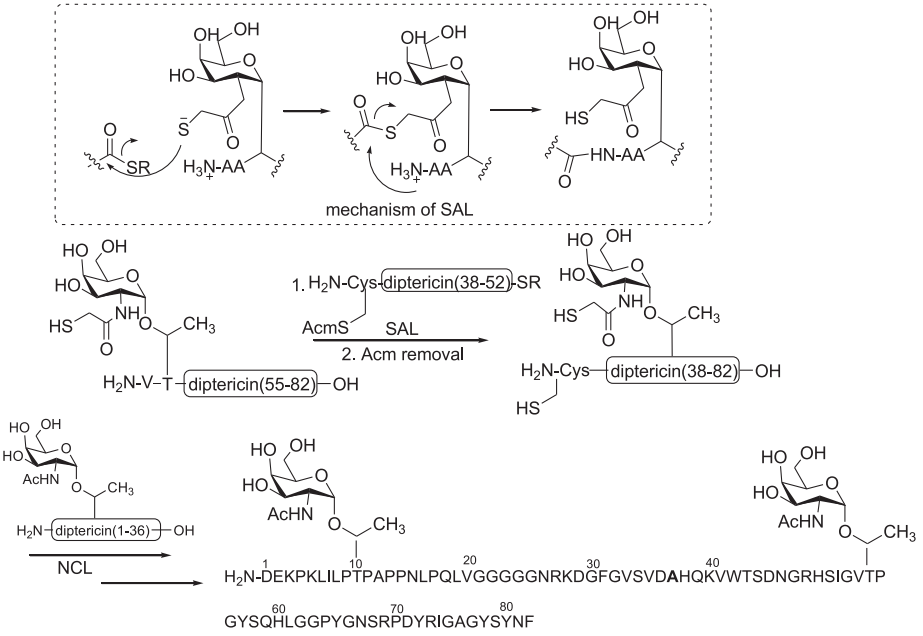
7.0 afforded SH3 α-spectrin 28-*N*-Tmb (Tmb=4,5,6-trimethoxy-2-mercaptobenzyl). After the removal of the auxiliary through treatment of the lyophilized crude ligation product with 5% triisopropylsilane in TFA for 2 h at 23 °C, the end product was obtained in 66% yield.

In an altogether different approach, Lu and Tam designed a reversible and reusable C-terminal mercaptoethyl ester as a thiol handle for NCL (Scheme 30).[83] Initially, the peptide mercaptoethyl ester (thiol handle) reacts with a peptide-α-thioester through transthioesterification to generate an ester–thioester intermediate. Ag^+ activation triggers $S\rightarrow N$ acyl migration to form a peptide bond and subsequent regeneration of the mercaptoethyl ester, which can either be removed under mildly basic conditions or can be retained as thiol handle for a subsequent ligation.

Wong et al. developed *s*ugar-*a*ssisted Cys-free *l*igation (SAL) in which the thiol handle inserted at the C-2 acetamido unit of a sugar moiety attached to the peptide backbone plays the role of Cys during ligation.[84] The sulfhydryl group of the sugar unit participates in transthioesterification with a peptide-α-thioester under NCL conditions and $S\rightarrow N$ acyl migration occurs through a macrocyclic intermediate (Scheme 31). An 82-amino acid antibacterial glycopeptide—diptericin ε—possessing two glycosylation sites at Thr^{10} and Thr^{54} was dissected into three fragments and assembled from C- to N-terminus through SAL followed by NCL. The Gly^{52}-Val^{53} sequence next to the glycosylation site Thr^{54} served as a suitable junction to operate SAL. Thus, the synthesis involved SAL between $Cys(Acm)^{37}$-Gly^{52} and Val^{53}-(Thr^{54}GalNHAc)-Phe^{82} followed by NCL between the fragments Asp^{1}-Asp^{36}-SR and Cys^{37}-Phe^{82}. While synthesizing the middle fragment (37–52), the Ala^{37} was mutated with Cys for the purpose of ligation. The *N*-acetyl galactosamine (GalNAc) unit attached to Thr^{54} was equipped with a thiol handle to facilitate SAL (Scheme 30). Thus, the first ligation between diptericin-[$Cys(Acm)^{37}$-52] and diptericin-[Thr^{54}{GalNAc(SH)}(53–82)] in 6 M Gnd·HCl, 200 mM phosphate (pH=8.5) afforded $Cys(Acm)^{37}$-Phe^{82} with 36% yield in 48 h. In the next step, the Acm group on Cys^{37} was removed by treatment with $Hg(OAc)_2$/AcOH and was subsequently ligated with Asp^{1}-Asp^{36}-SR to afford the full-length glycopeptide. After RP-HPLC purification followed by lyophilization, it



Scheme 30. C-Terminal mercaptoethyl ester as reversible and reusable thiol handle.



Scheme 31. Sugar-assisted native chemical ligation.

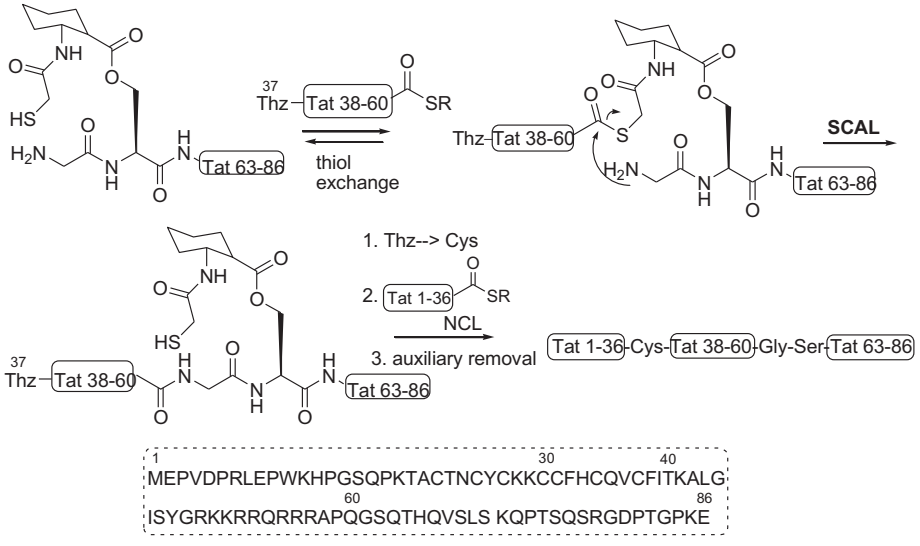
was hydrogenated (H_2/Pd−Al_2O_3) to remove the thiol handle of GalNHAc at Thr[54] as well as to convert Cys[37] into Ala. After final purification and lyophilization, diptericin was isolated in an overall yield of 54% (254 μg; ESI-MS analysis, found: 5408 Da; calculated value: 5409 Da).

Brik and others developed a *s*ide-*c*hain-*a*ssisted *l*igation (SCAL) for Cys-free NCL. In this protocol, a thiol-modified cyclohexyl auxiliary linked to the side chain of Ser/Thr or Asp/Glu was employed to facilitate thiol capture from a peptide-α-thioester.[85] The cyclohexyl moiety was removed post ligation through saponification. The C-terminal 50-*mer* fragment of HIV-I Tat protein was assembled through SCAL, which was later attached to the remaining portion of the peptide through NCL to obtain the full-length protein (Scheme 32). SCAL between Tat(61–86) having a cyclohexyl auxiliary at Ser[62] and Thz[37]-Tat(38–60) was carried out at a concentration of 2–3 mM in 6 M Gnd·HCl, pH 8 (1:1.1 molar ratio of peptide-α-thioester to peptide auxiliary), which took about 72 h at 37 °C. In the next step, Thz was converted into Cys by treatment with methoxylamine (37 °C, 24 h), and the intermediate peptide, after purification, was subjected to NCL with Tat(1–36)-COSR to afford HIV-1 Tat protein.

Several other auxiliaries were also developed for Cys-free NCL and their utility has been successfully demonstrated (Fig. 9).[53,77,86,87]

2.8. NCL at non-Cys amino acids

Development of a Cys-free NCL protocol, to have any amino acid (Xaa-Xaa) at the ligation site, would be practically beneficial, especially when the target protein does not contain Cys, or, if so, not at the desired position. The simplest route to have access to a non-Cys residue at the ligation site is to carry out a post ligation modification such as desulfurization of Cys to an Ala residue. This was first reported by Yan and Dawson.[88] In their approach, the target protein was bisected at the Xaa-Ala junction, but the ligation was carried out by keeping Cys at the N-terminus in place of Ala. After ligation, the Cys residue was



Scheme 32. Side-chain-assisted native chemical ligation.

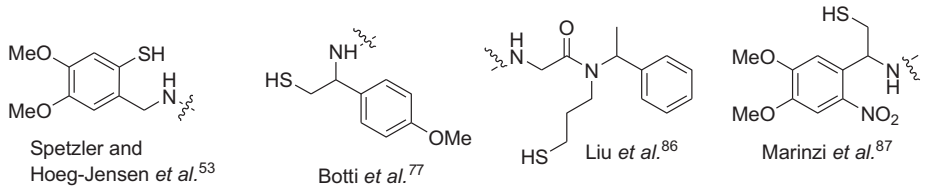


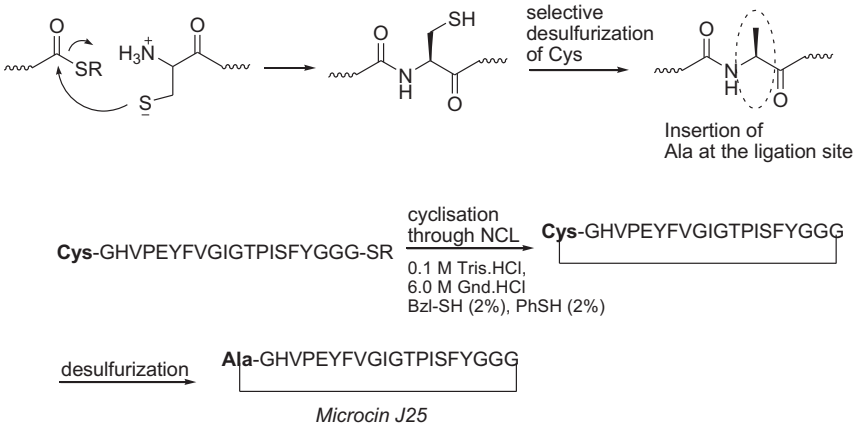
Fig. 9. Auxiliaries for Cys-free NCL.

converted into Ala through desulfurization to obtain the native sequence (Scheme 33). The relative abundance of Ala in proteins makes this protocol attractive, thus opening up plenty of options in the retrosynthetic analysis of a protein to dissect into the ligatable fragments. The utility of such a protocol was demonstrated by the synthesis of the antibiotic, microcin J25, a 21-amino acid cyclic peptide. For the cyclization via NCL, the linear peptide (17.6 mg) was dissolved in buffer (0.1 M Tris·HCl, 6 M Gnd·HCl in the presence of benzyl mercaptan (2%) and thiophenol (2%) additives at pH 8.5 (Tris=tris(hydroxymethyl) aminomethane hydrochloride)). The cyclized product was isolated through semipreparative HPLC in 50% yield (7 mg; ESI-MS: 2139.6±0.3 Da; calculated value: 2139.4 Da). The desulfurization was effected using Pd/Al_2O_3 or Raney nickel. Desulfurization employing Pd/Al_2O_3 in 6 M Gnd·HCl at pH 7.5 led to the formation of trace amounts of disulfide compound along with the desired microcin J25 (high-resolution MALDI-FTMS, found: 2107.0284 Da; calculated value: 2107.0284 Da, Fig. 10). However, when the reaction was carried out at pH 4.5, no dimerization was recorded and the reaction was complete in 2 h with 50% yield. Besides this, other catalysts such as Pd/C, $Pd/BaSO_4$ or Raney nickel also gave appreciable yields of desulfurized product while, with PdO, the reaction was incomplete.

desulfurization[89] and, similarly, *erythro*-β-mercapto-phenylalanine[90] furnished Phe at the ligation site (Fig. 11).

An interesting addition to the non-Cys type of ligation protocols was reported by Yang et al. and others who prepared β-mercapto derivatives of Lys.[91] The thiol handle equipped at the γ-carbon of Lys is β to both α- and ε-amino groups and participates in ligation in a tandem manner for acylation of both the amines of Lys. 4-Mercaptolysine was introduced at the N-terminus of a peptide with its ε-amine protected using a benzyloxycarbonyl (Z) group. Thus, 4-SH-Lys(Z)-GAKAFA-NH_2 was reacted with the thioester peptide, H-LSTEA-COSR, at pH 8.0 and the reaction was found to proceed to completion in about an hour with a yield of over 90%. To facilitate ligation at the ε-amine, the Z group was removed and the peptide segment was treated with H-LSTEG-COSR to afford the desired product in 1.5 h in 92% yield. Desulfurization was effected in 6 M guanidine, 0.2 M phosphate at pH 6.0, employing 100 mM tris(2-carboxyethyl)phosphine (TCEP) to reinstate the unmodified Lys at the ligation site (Scheme 34).

A list of various β-mercapto derivatives available for non-Cys NCL and the corresponding amino acids obtained after desulfurization along with the reaction conditions and yield are delineated in Table 3. A straightforward and high yielding synthesis of γ-(*R*,*S*)-mercapto-



Scheme 33. NCL followed by desulfurization to establish Ala at the ligation centre.

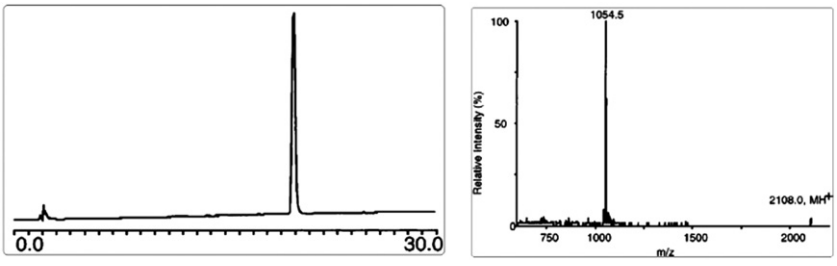


Fig. 10. HPLC and ESI-MS spectra of microcin J25.

β-Mercapto derivatives of several other amino acids were also developed and put to use in Cys-free NCL. Subsequent desulfurization afforded the corresponding amino acid at the ligation site. Penicillamine (Pen) was selected to afford a Val residue after glutamine was reported by Brik et al.[92] The study demonstrated the utility of γ-(*R*,*S*)-mercaptoglutamine in the synthesis of WW domain (1–40) derived from YAP65 WW domain polypeptide.[92] The desulfurization was carried out using nickel boride.[88]

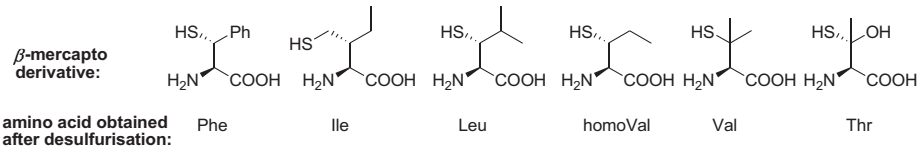
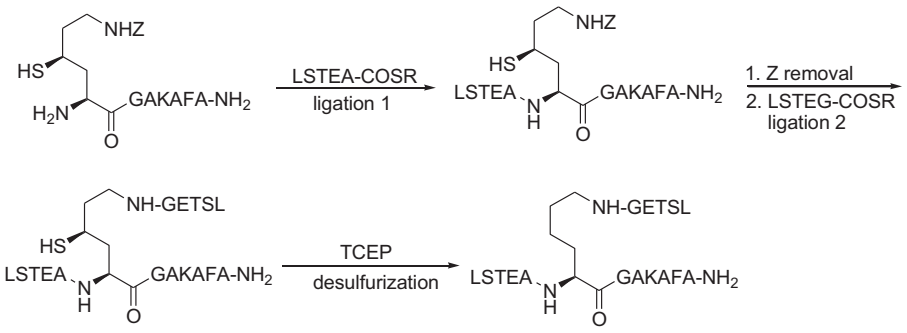


Fig. 11. β-Mercapto derivatives for Cys-free ligation (selected examples).



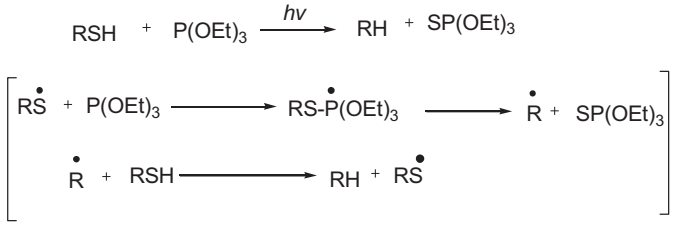
Scheme 34. 4-Mercaptolysine mediated NCL.

Table 3
Few mercaptoamino acids used in Cys-free NCL and their desulfurization condition (selected examples)

No.	β-Mercapto derivative	Amino acid obtained after desulfurization	Desulfurization conditions	Yield (%)	Ref.
1	Cys	Ala	Pd/Al_2O_3	90	88
2	Pen	Val	VA-044,[a] TCEP, EtSH	98	89
3	*erythro*-β-Mercapto-Phe	Phe	$NiCl_4 \cdot 6H_2O$, $NaBH_4$	80	90
4	(2*S*,4*S*)-4-Mercapto-L-Lys	Lys	VA-044,[a] TCEP	40	91a

[a] VA-044=2,2′-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride.

Interestingly, a simple route involving free-radical-based, mild, highly versatile and specific desulfurization of Cys to Ala based on classical organic chemistry (Scheme 35) was developed by Wan and Danishefsky.[93a] The reaction was carried out using (^{t}BUSH), tris(2-carboxyethyl)phosphine (TCEP, phosphine source), 2,2′-azobis[2-(2-imidazolin-2yl)propane]dihydrochloride (VA-044, radical initiator) in water at room temperature for about several hours. This reaction takes place without side reactions. Additional studies in this direction were carried out by Haase and Seitz also.[93b] This metal-free reduction of Cys is tolerant to a range of relevant functional groups. The advantages of this method have been demonstrated by the synthesis of glycooligopeptides[93a] and human parathyroid hormone (hPTH, 84 residues).[93c]

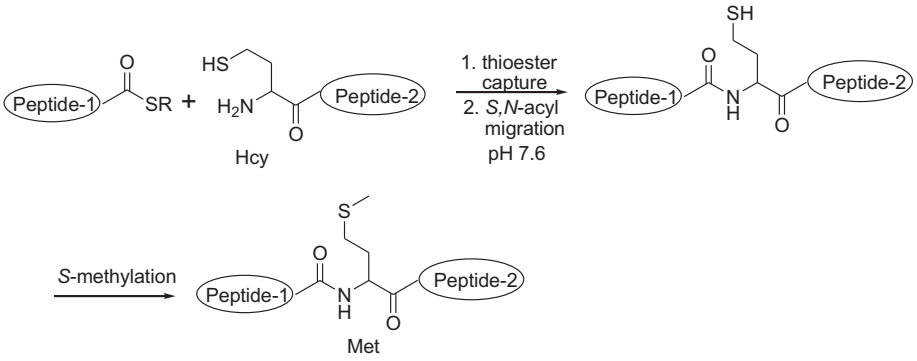


Scheme 35. Proposed mechanism of radical desulfurization reaction.

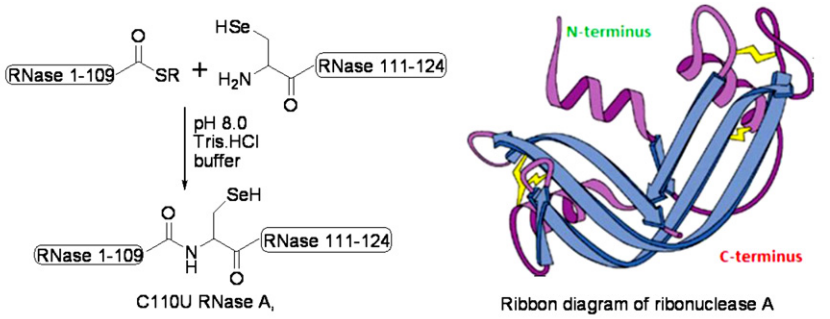
Tam and Yu developed a viable chemistry to establish Xaa-Met as the ligation site. This protocol consisted of a ligation between an N-terminal homocysteinyl peptide and a peptide-α-thioester followed by S-methylation to convert homocysteine (Hcy) into a Met residue (Scheme 36).[94] The ligation was performed at pH 7.6 in phosphate buffer in a highly reductive environment and was generally complete in about 4 h. Selective methylation was carried out employing an excess of methyl *p*-nitrobenzenesulfonate. In a model study, a peptide, KLYG-$SCH_2CH_2CONH_2$, was ligated to Hcy-ARVELKKLQDV and then methylated to obtain the product in 73% yield. Its efficacy was further confirmed through the synthesis of a 34-amino acid human parathyroid hormone fragment, SF34 (SVSEIQLMHNLGKHLNS*M*ERVEWLRKKLQDVHNF). The strategy was also found to be useful for the preparation of cyclic peptides employing unprotected homocysteinyl peptide-α-thioesters.[94]

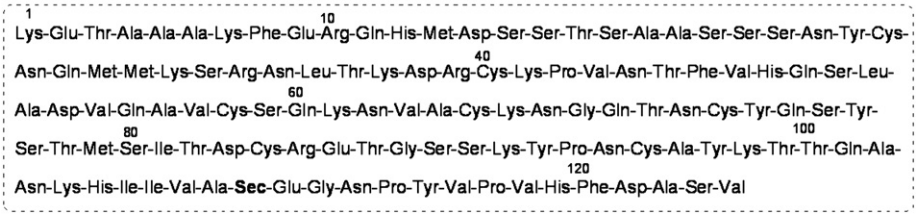
Raines and others demonstrated selenocysteine (Sec)-mediated NCL.[95] A model reaction between Ac-Gly-SR and selenocysteine [$(Sec)_2$] gave the corresponding ligated product $(Ac\text{-}Gly\text{-}Sec\text{-}OH)_2$. The RSe^- is more nucleophilic compared to RS^- and the pK_a of a selenol is much less than its thio analogue. Hence, the ligation reaction carried out employing Sec was 10^3-fold faster than that with Cys, especially at low pH (Scheme 37). To demonstrate the viability of Sec-mediated NCL, ribonuclease A (RNase A) was chosen as a model. Out of the eight Cys residues present in the molecule, Cys^{110} was replaced with Sec and the ligation between RNase A (1–109) with RNase A (Sec^{110}-124) was carried out at pH 8.0 in 0.10 M Tris·HCl buffer containing tris-(2-carboxyethyl)phosphine (TCEP).

In a similar manner to Sec-mediated NCL, Roelfes and Hilvert developed a route for the insertion of selenomethionine (SeMet) at the ligation site.[96] The ligation was performed between a peptide with N-terminal selenohomocysteine (SeHcy) and a peptide-α-thioester. After the ligation, the selenol was methylated to install SeMet at the ligated junction. A model reaction was carried out between a peptide, LYRAG-Set, with SeHcy (1 mM in Tris·HCl buffer, pH 8.5, 6 M Gnd·HCl, 5% thiophenol), which was complete in 20 h. The reaction afforded the product as a diselenide, which was methylated using methyl 4-nitrobenzenesulfonate under reducing conditions (excess thiophenol) to obtain 4.0 mg of the SeMet-containing peptide in 66% yield (Scheme 38). The utility and efficacy of the method were demonstrated by synthesizing seleno bPP,

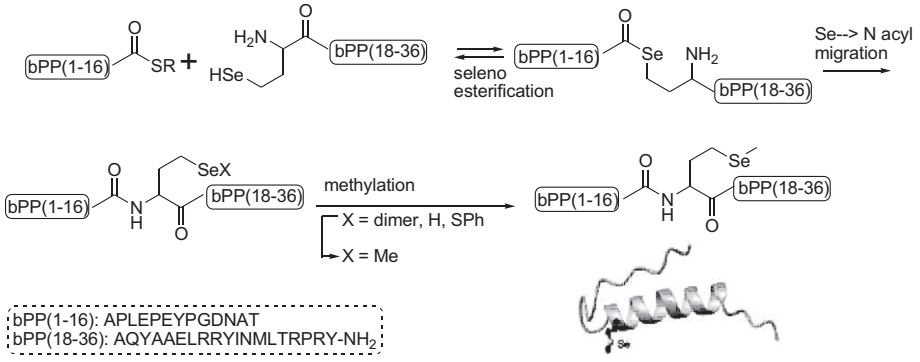


Scheme 36. Homocysteine mediated NCL followed by thiol methylation establishes Met at the ligation junction.





Scheme 37. Selenocysteine mediated NCL.



Scheme 38. Selenohomocysteine mediated NCL followed by methylation establishes SeMet at the ligation junction.

a variant of *b*ovine *p*ancreatic *p*olypeptide containing SeMet at positions 17 and 30. This 36-amino acid polypeptide is a member of the neuropeptide Y family and plays a vital role in the regulation of food uptake. The fragments were assembled on solid phase employing Fmoc/HBTU–HOBt chemistry. The ligation was carried out in excess thiophenol in a buffer (Tris·HCl and Gnd·HCl). The reaction was sluggish and was driven to completion after 5 days by the addition of a fourfold excess of thioester. The crude product containing the diselenide and a mixed selenosulfide with thiophenol was methylated by treatment with Tris·HCl (1 equiv) in sodium phosphate buffer (200 mM, pH 6.0) in the presence of methyl 4-nitrobenzenesulfonate (5 equiv). The crude product was purified using RP-HPLC to afford 6.8 mg of seleno bPP.

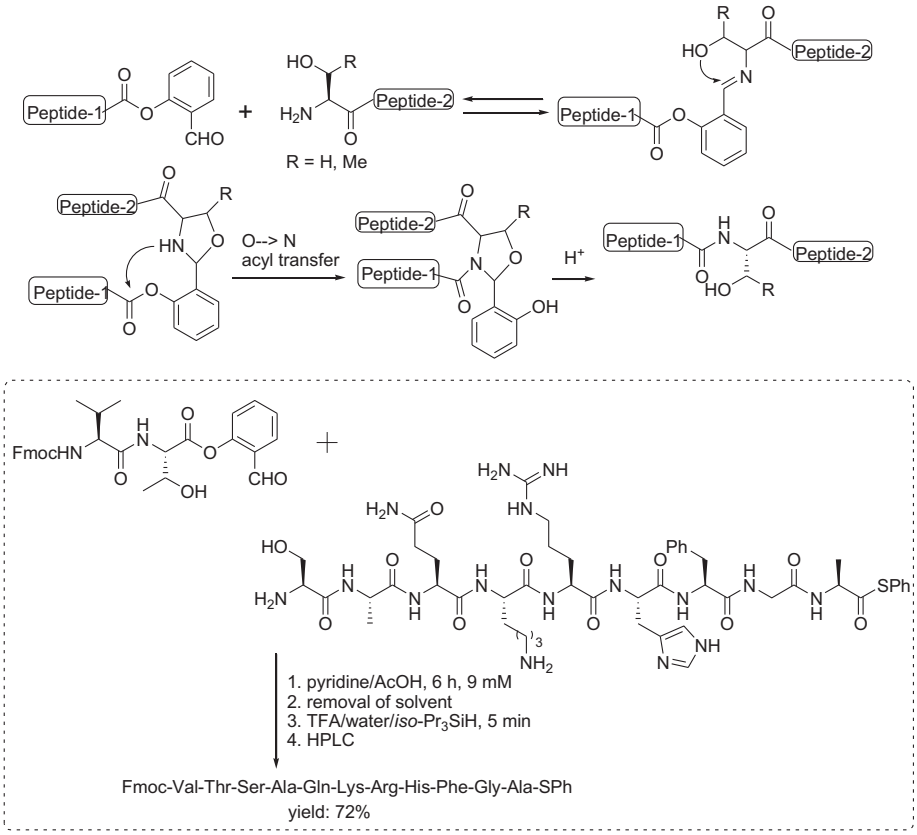
2.9. Salicylaldehyde ester-induced ligation

Li et al. introduced a different concept for Cys-free chemoselective ligation in which a Ser/Thr can be incorporated at the

ligation junction. A reaction between a peptide bearing a C-terminal *O*-salicylaldehyde ester with a peptide having an N-terminal Ser/Thr was carried out.[97] The first step involves a reversible reaction of the α-amino group of Ser/Thr with the aldehyde to form an imine followed by an intramolecular cyclization to an oxazoline ring. Further, $O \rightarrow N$ acyl transfer affords a stable acetal intermediate, which on acidic treatment establishes the peptide bond with concomitant removal of the auxiliary (Scheme 39). NCL at other sites such as pseudocysteine has also been demonstrated, but these methods are yet to gain general applicability.[98]

Danishefsky et al. envisaged that replacement of $S \rightarrow S$ acyl transfer with the more favourable $O \rightarrow S$ acyl transfer would be advantageous.[29] This idea coined the term, oxoester-mediated NCL, which was demonstrated through the ligation of a model *p*-nitrophenyl ester of an Fmoc-peptide with an N-terminal Cys-peptide (Scheme 41).

Interestingly, the oxoester-mediated ligation proceeded smoothly with a series of peptides having hindered C-terminal amino acids including Thr, Val, Ile, D-*allo*-Ile and Pro, while racemization of up to 5–6% was recorded when Leu, Ile and Val were present at the C-terminus.



Scheme 39. N-Terminal Ser/Thr ligate chemoselectively to peptides bearing C-terminal salicyladehyde ester.

2.10. Ligation mediated by quinolinium thioester salts

A novel approach, which relies on conjugated addition to quinolinium salts that aid amine capture and result in the establishment of a peptide bond was developed by Levacher. In this reaction, two steps, namely the acyl transfer and removal of the template, take place simultaneously.[32] The method involves the reaction between a peptide-α-thioester equipped with a quinoline template and the amino terminus of a peptide fragment. The entropic activation of the peptide-α-thioester substrate is achieved via N-methylation of quinoline using methoxytriflate (MeOTf) that develops an electrophilic platform and thus attracts the amine terminus of the peptide. Acyl migration and consequent detachment of the quinoline moiety lead to the insertion of a peptide bond (Scheme 40).

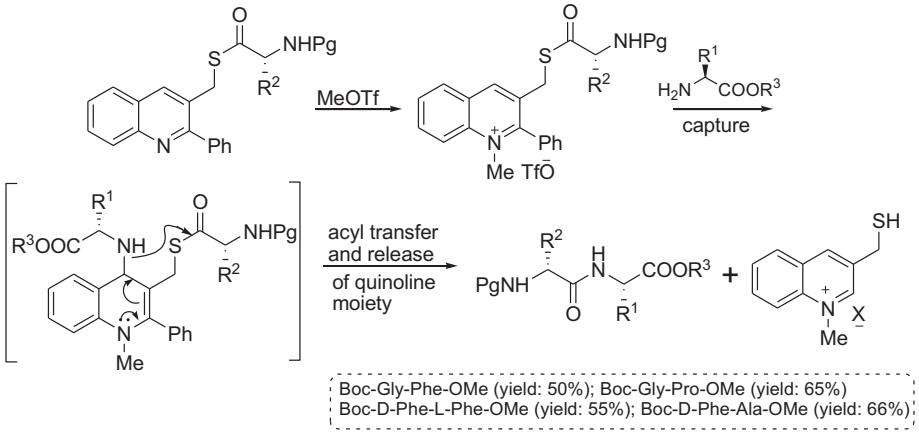
2.11. NCL through peptide phenyl esters

The success of NCL can be attributed to the rate of the thiotransesterification step, i.e., $S \rightarrow S$ acyl transfer, which in turn depends on the steric bulk of the amino acid at the thioester termini.

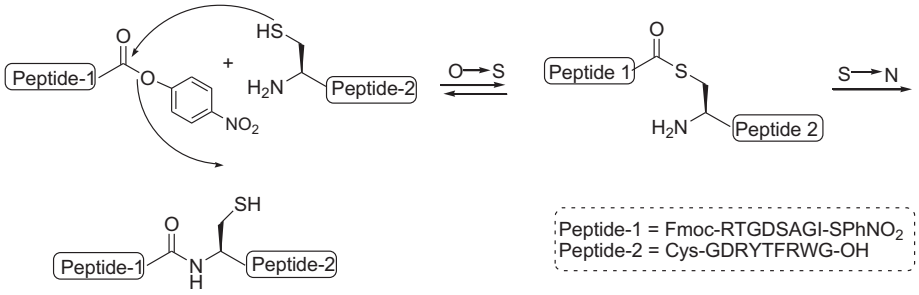
2.12. Synthesis of glycopeptides

Protein glycosylation is known to effect protein folding and play a specific role in improving the in vivo half life. As much as 40% of the proteins are glycosylated.[99] Despite their significance, much advancement in understanding the structure–function of glycoproteins has been retarded, due to the difficulties associated with their isolation in chemically homogeneous forms to both from natural sources as well as from preparation through chemical methods. Although an ultimate and highly convergent protocol would be to merge a matured oligosaccharide unit into a polypeptide domain, the protocol would face solubility, selectivity and other practical considerations. It would be better if an oligosaccharide is first coupled to a more manageable short peptide and then ligated to a polypeptide or protein fragment. Approaches to obtain diverse types of glycopeptides and glycoproteins have been reviewed previously.[100] NCL is one of the latest techniques to be adopted for the assembly of these important classes of biomolecules.[101]

Bertozzi's group assembled an 82-amino acid residue glycoprotein, diptericin, employing NCL.[54] Diptericin consists of an N-terminal Pro rich domain (1–20) and a C-terminal Gly rich domain



Scheme 40. Ligations achieved by employing C-terminal quinolinium esters.

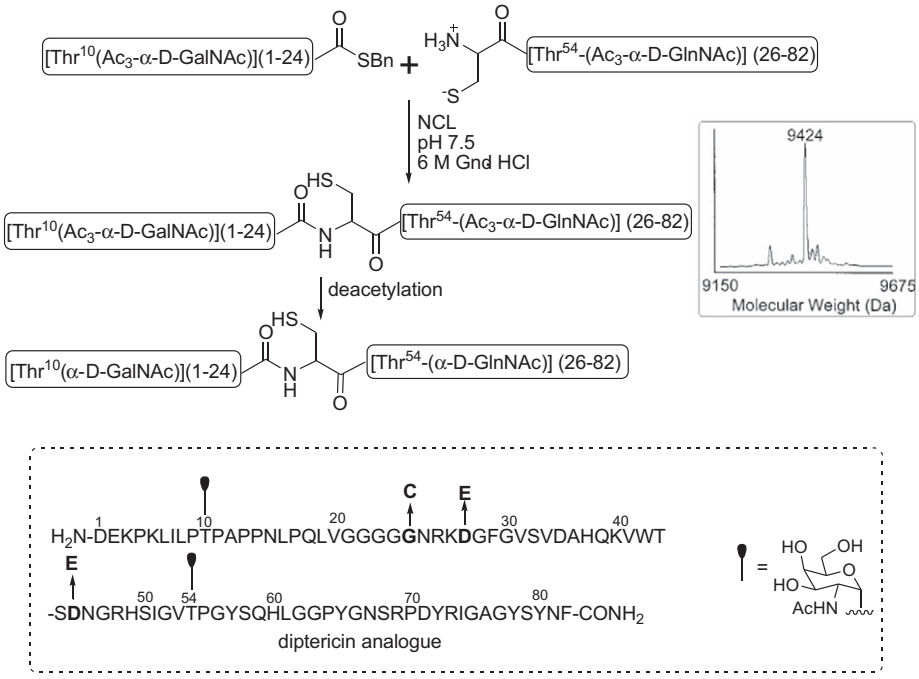


Scheme 41. NCL employing C-terminal arylesters.

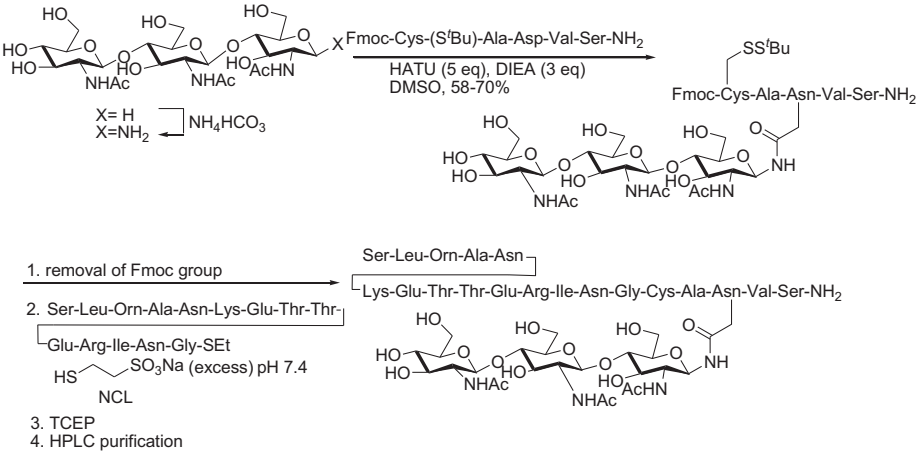
(26–82) bridged by five Gly residues in series. The lack of Cys in the sequence is a technical hurdle for its assembly through NCL. Hence, Gly^{25} was replaced with Cys to make this protein suitable for NCL. Further, to ease the synthesis, two more modifications were made—Asp^{29} and Asp^{45} were replaced with Glu—and all these substitutions were assured not to affect the native biological and structural properties. The sugar unit, Gln(NHAc), was equipped at two positions—Thr^{10} and Thr^{54}—by incorporating the *O*-glycosylated-Thr during the chain assembly. Thus, the two glycopeptide segments selected for ligation were: glycopeptide-α-thioester Thr^{10}(Ac_3-α-D-GalNHAc)diptericin(1–24)-COSBn and *N*-Cys-glycopeptide Cys^{25}-[Thr^{54}(Ac_3-α-D-GalNHAc)]diptericin(25–82). The thioester fragment was synthesized on solid phase using Kenner's modified sulfonamide linker as described previously (Section 2.2; see also Scheme 11). The Cys-peptide fragment was assembled on a Rink resin by SPPS and characterized. The peptide-α-thioester, Thr^{10}-diptericin(1–24)-COSBn (1.4 mg, 0.5 μmol), was ligated with Cys^{25}-diptericin(25–82) (5.0 mg, 0.75 μmol) in 6 M Gnd·HCl, 0.1 M sodium phosphate (pH 7.5) and 4% thiophenol (Scheme 42). The reaction was complete after 18 h, as monitored by HPLC. The product was isolated through preparative HPLC in 55% yield (2.6 mg). Removal of the acetate protections on the sugars by treatment with 5% aqueous hydrazine afforded 1.33 mg of the protein (yield: 53%), which was characterized by ESI-MS (observed: 9172 Da; calculated: 9171.99 Da).

Danishefsky et al. accomplished the assembly of N-linked glycoproteins through NCL.[102a] Starting with the unprotected saccharides, the anomeric carbon was aminated under a Kochetkov reaction using NH_4HCO_3.[102b] In the next step, the glycosylated amines were acylated to the β-COOH group of the Asp residue in the *N*-protected-Cys-pentapeptide pre-activated with 1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium hexafluorophosphate 3-oxide (HATU) in the presence of a tertiary base. In parallel, a tetradecapeptide thioester was assembled on poly(ethylene)glycol (PEG)-supported Wang resin employing Fmoc/*tert*-Bu/SPPS. A ligation of the polypeptide-α-thioester with a Cys-glycopeptide fragment was carried out on a 15 mg scale in phosphate-buffered saline (PBS, 0.2 M), at pH 7.4, in the presence of excess sulfanylethane-2-sulfamate. After the ligation, a global disulfide reduction using water soluble TCEP afforded the fully unprotected glycopeptide in 78% yield (Scheme 43).

Boons et al. demonstrated a liposome mediated NCL to assemble sparingly soluble lipopeptides.[103] A model lipoglycopeptide was assembled through the condensation of three components—a tumour associated glycopeptide derived from mucin-1 (protein targeted in cancerous cells; MUC-1), a known T-cell epitope, YAFKYARHANVGRNAFELFL, and a lipopeptide, *S*-[(*R*)-2,3-dipalmitoyloxy-propyl]-*N*-palmitoyl-(*R*)-cysteine (Pam_3CysSK_4). Initially, the MUC-1 epitope was assembled on solid phase employing Fmoc-amino acids [Thr was introduced as Fmoc-Thr-(α-AcO_3-D-GalNHAc)-OH] on Rink amide resin. After cleaving from the solid support (TFA: 94%, water: 2.5%; ethanedithiol: 2.5% and TIS: 1%), the peptide was treated with 5% hydrazine to deprotect the acetyl groups on galactose. The other two fragments were assembled on solid phase employing a sulfonamide safety-catch linker. In the next stage, MUC-1 and T-cell epitope were ligated under standard conditions (6 M Gnd·HCl, 200 mM $NaHPO_4$, 4% thiophenol), and this was complete in 18 h, as monitored by LC–MS. The product was isolated through semi preparative RP-HPLC in 48% yield. In the following steps, the Acm protector on the N-terminal Cys was removed using Hg(II)OAc and the free amino glycopeptide was subjected to ligation with the lipopeptide, Pam_3CysSK_4. Under standard conditions, the ligation did



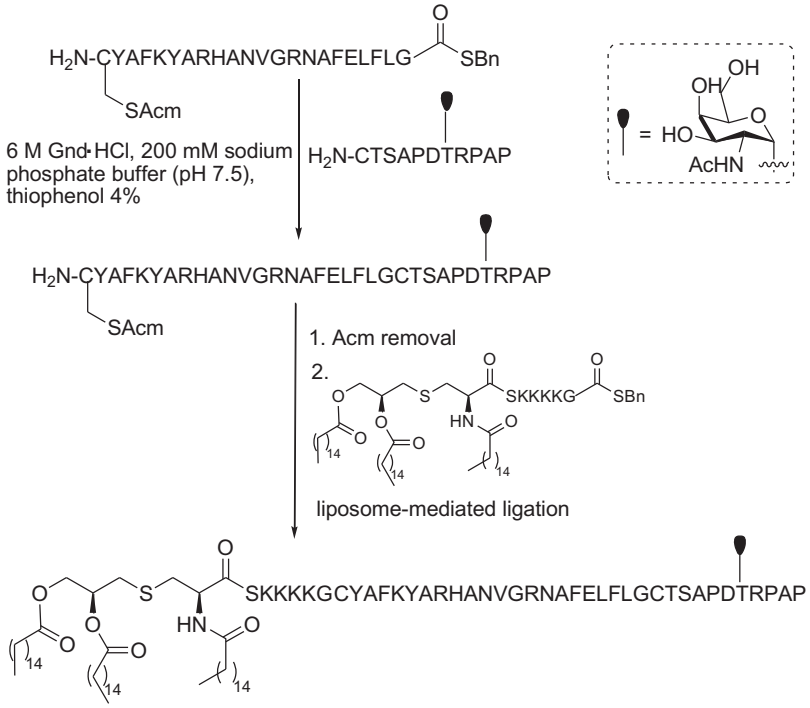
Scheme 42. Synthesis of diptericin analogue via NCL. Inset shows ESI-MS of purified ligation product.



Scheme 43. Utility of NCL in assembling glycopeptides.

not progress, due to the poor solubility of the lipopeptide. Attempts to improve the solubility by adding detergents or through higher temperatures met with limited success. To facilitate the solubilization of the two ligating fragments, the assistance of liposomes was explored. A film of dodecylphosphocholine, glycopeptides and lipopeptide thioester was hydrated by incubation at 37 °C for 4 h in a phosphate buffer (pH 7.5) in the presence of carboxyethylphosphine and ethylenediaminetetraacetic acid (EDTA). The latter two reagents were used to suppress disulfide formation. The mixture was ultrasonicated for 1 min and the resulting vesicles were sized to 1 μm by passing through a polycarbonate membrane filter. The ligation was catalyzed by the addition of sodium 2-mercaptoethanesulfonate. The reaction was complete in 2 h and the target lipoglycopeptide was isolated in 83% yield (3.5 mg) after HPLC isolation, which was confirmed through MALDI-TOF (observed: 5392.9712 Da; calculated value: 5392.0171 Da; Scheme 44).

Danishefsky visualized that a peptide phenolic ester equipped with a thiol moiety at the *ortho* position can be ligated to an N-terminal Cys-peptide.[104] Thus, an ester possessing an adjacent thiol group present in equilibrium with a thioester participates in NCL with a Cys-peptide. Commercially available 2-mercaptophenol was oxidized to its symmetrical disulfide, from which the unsymmetrical alkyl aryl disulfide was prepared by treating with excess diethyl sulfide. The resulting phenol was esterified to Boc-Phe-OH and the Boc group was cleaved by treatment with TFA. It was then coupled to a protected peptide acid under standard coupling conditions using *O*-(7-azabenzotriazolyl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) followed by glycosylation of the aspartyl residue of the peptide with a chitobiose-derived glycosylamine to afford the glycopeptide phenyl ester. Then, the Cys-glycopeptide (1.44 μmol) assembled on solid phase was ligated to the glycopeptide phenyl ester (1.44 μmol) employing strongly reducing conditions [0.2 M phosphate, 0.2 M NaCl and excess sodium 2-mercaptoethanesulfonate, (MES-Na)], under which, the protector on Cys and the phenyl ester were released along with subsequent NCL to afford the corresponding



Scheme 44. NCL of lipopeptides assisted by liposomes.

glycopeptide (Scheme 45). No racemization was observed and the product was confirmed through ESI-MS (calculated: 1108.5 Da; observed: 1108.6 Da) and ^{1}H NMR techniques.

2.13. Combination of NCL with click chemistry

Brimble et al. reported the synthesis of glycopeptidomimetics by combining NCL with Cu(I) catalyzed azide–alkyne cycloaddition.[105a] The click reaction is employed for the regioselective synthesis of 1,4-substituted 1,2,3-triazoles. A one-pot protocol was developed in which a peptide-α-thioester and a Cys-peptide, both equipped with an alkyne functionality, were ligated under normal conditions followed by a click reaction with an azidosugar to afford the triazole-linked glycopeptidomimetics. Boc-L-propargylglycine (Boc-L-Pra-OH) was prepared and inserted into both peptide-α-thioester and *N*-Cys-peptide during the fragment assembly to have a 'clickable' centre. A 3-mM quantity of each segment was dissolved in 6 M Gnd·HCl/0.2 M Na_2HPO_4 along with TCEP and mercaptophenylacetic acid (MPAA; pH 7.0). After the completion of ligation, the click reaction was initiated by adding $CuSO_4$ (20 mM) followed by GalNAc-N_3 (10 mM). The reaction was complete after 6 h (Scheme 46), and the resulting 23-residue neoglycopeptide containing two triazole units was isolated, which had a mass of 3252.0 Da±0.4 Da in agreement with the theoretical value (3252.6 Da).

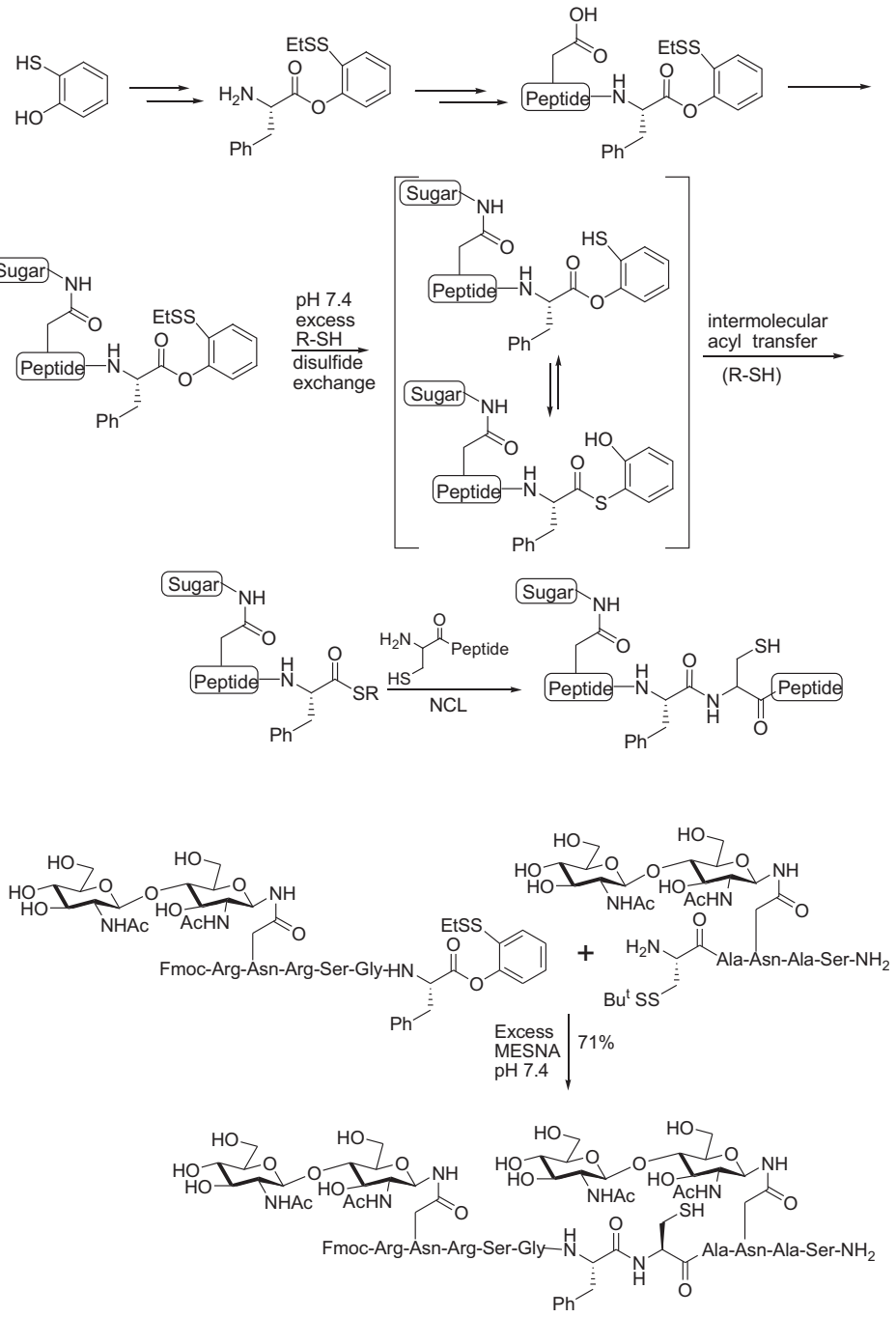
Xiao and Tolbert also employed a click reaction in combination with NCL to assemble N-terminally linked protein dimers and trimers.[105b]

Diversity in both the methodology of protein ligation and design of molecules is being brought about by carrying out the ligation at lipid interfaces.[106] NCL has been applied to the syntheses of polymers, hydrogels, lipopeptides, glycolipopeptides, protein helical bundles and cyclic peptides, and for protein ubiquitination.[107,108] It is likely that NCL would enjoy a still wider utility in chemical biology with the underpinning of the recent developments such as automation of SPPS, easy access to synthetic peptide-α-thioesters and successful removal of the necessity of having the less-common Cys residue at the ligation site. As and when it becomes possible to engineer much larger fragments, the NCL will be at its peak as a tool to mimic nature with respect to protein synthesis. Progress in the field of chemical assembly of polypeptides and proteins has always followed advances in the strategy for preparation of polypeptide-α-thioesters and ligation techniques. Both strands of research have come together with the development of *e*xpressed *p*rotein *l*igation (EPL), which has opened up new avenues for the assembly of large proteins.

3. Expressed protein ligation

EPL,[109,110] one of the milestones in ligation techniques, has widely expanded the scope of NCL. The chemical syntheses of peptide fragments required for NCL is even now a technically demanding process with size constraints. Consequently, the utility of NCL has centred around the assembly of proteins only in the range of 15 kDa. EPL is a biosynthetic version of NCL where a protein thioester generated recombinantly by the thiolysis of an intein fusion protein is ligated to a semisynthetic or recombinant protein with N-terminal Cys to produce a native protein. This technique, also known as *i*ntein-mediated *p*rotein *l*igation (IPL), has found applications in the synthesis of peptide C-terminal thioesters as well as oligopeptide fragments with N-terminal Cys. The question of how to obtain a large protein fragment required for ligation leads to an interesting biological process—protein splicing.[111] This is a post-translational process in which a precursor protein undergoes a series of intramolecular rearrangements, resulting in the precise excision of an internal region referred to as intein and ligation of the two lateral sequences (exteins).

The first step in protein splicing (PS) involves an $N\rightarrow S$ (or $N\rightarrow O$) acyl shift in which the N-extein unit is transferred to the thiol (or hydroxy) group of a Cys (or Ser/Thr) located at the immediate N-terminus of the protein. Although this step is apparently thermodynamically unfavourable, the structure of the intein is supposed to catalyze it through twisting the scissile amide bond into a higher energy conformation thereby guiding the equilibrium to the thioester (ester) side. Then, the entire unit is shifted to a second Cys (Ser/Thr) residue at the intein/C-extein boundary (+1 position) in



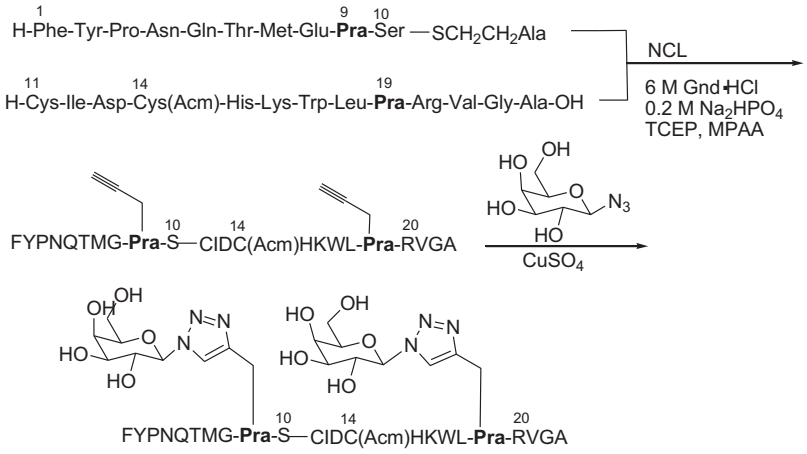
Scheme 45. NCL employing C-terminal peptide possessing an arylester with a thiol handle.

a transesterification step. The branched intermediate, in the next step, undergoes a cyclization reaction involving a conserved Asn residue at the C-terminus of the intein, which will be removed as a C-terminal succinimide derivative. Finally, an $S \rightarrow N$ (or $O \rightarrow N$) acyl shift establishes the amide bond between the two exteins (Scheme 47).

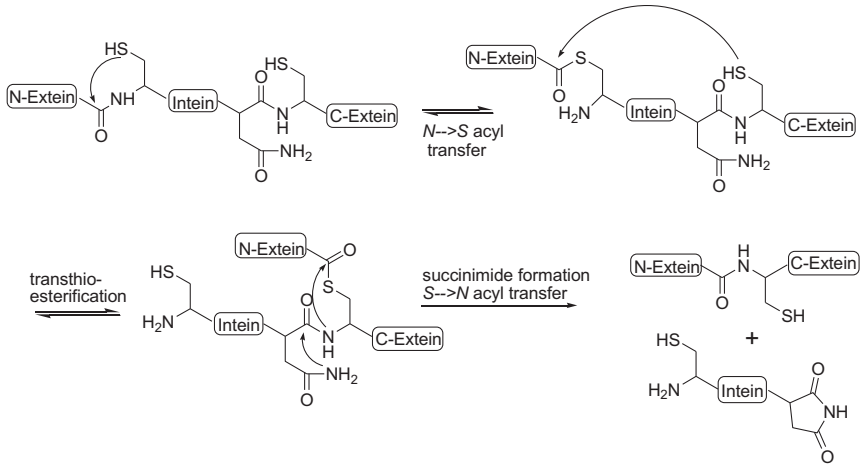
Suitable engineering has been developed in protein splicing to generate protein-α-thioesters.[112] Commercially available pCyB vectors for *E. coli* protein expression can be employed to generate an α-thioester in which a protein of interest is expressed in a frame fused with an intein−chitin binding domain (CBD) sequence. Thiolysis of the intein fusion protein using 2-mercaptoethanol or dithiothreitol (DTT) yields the protein-α-thioester through transesterification (Scheme 48). This highly reactive protein-α-thioester can be ligated instantaneously with an *N*-Cys-peptide to afford the product. Instead, if the thiolysis is carried out with ethyl thiol, less reactive peptide alkyl thioesters are obtained, which can be isolated. The chemistry was fully established to generate *N*-Cys proteins as well through recombinant synthesis. EPL has been applied to the synthesis of several classes of proteins from both prokaryotic and eukaryotic organisms.[113] Over 40 different proteins are addressed with EPL including kinases, phosphatases, transcription factors, polymerases, ion channels, cytoplasmic signalling proteins, histones and antibodies.[114,115]

3.1. Case studies

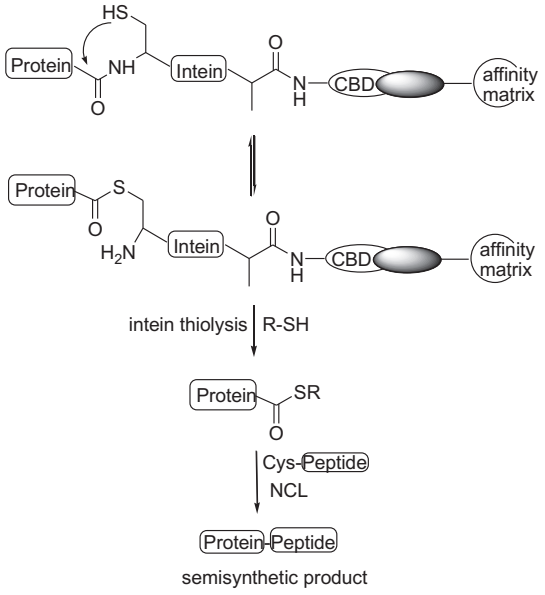
Wang and Cole demonstrated the use of intein fusion constructs as a source of large recombinant protein-α-thioesters in the



Scheme 46. NCL followed by [2+3] azide–alkyne cycloaddition.



Scheme 47. Mechanism of protein splicing.

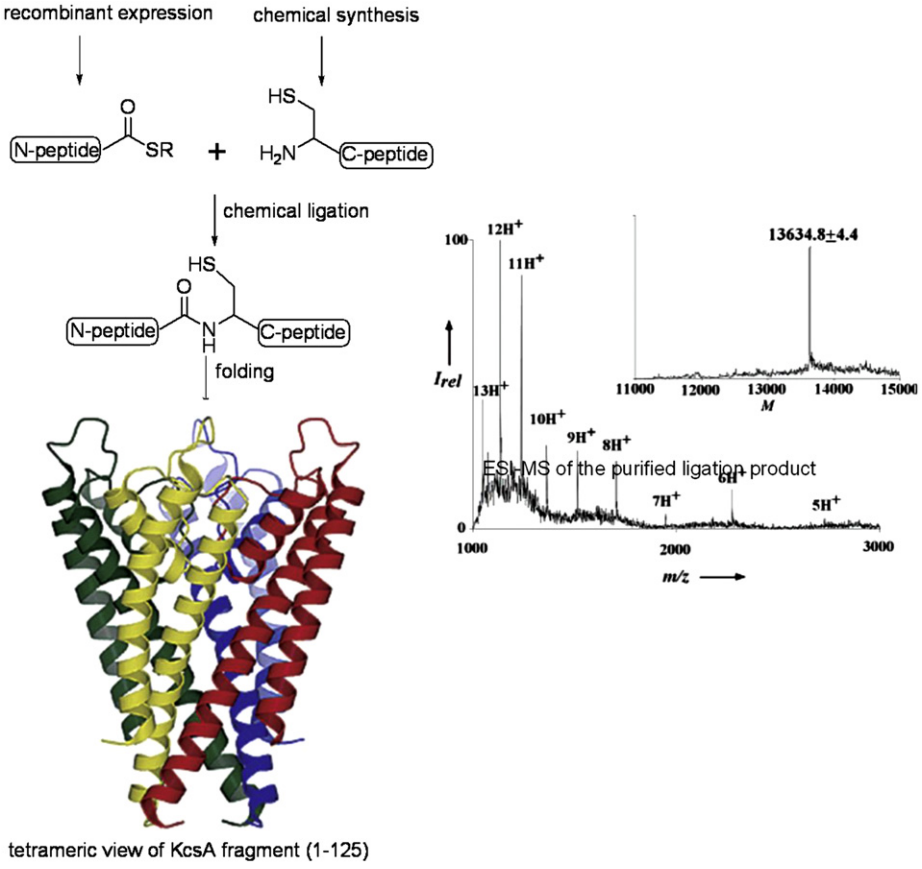


Scheme 48. Protein splicing engineered to generate protein-α-thioesters.

synthesis of a pair of semisynthetic C-terminal Src Kinase (Csk), a 50 kDa protein.[116] This opened up a new era to access large proteins that are otherwise difficult to obtain through chemical synthesis. Csk catalyzes the phosphorylation of a highly conserved tyrosine within the C-terminal tails of Src family members. To study the potential effect of attachment of a phosphotyrosine tail to Csk, its synthesis through EPL was undertaken. Csk protein was recombinantly expressed in *E. coli* as the corresponding thioester. To ligate with this, tyrosine phosphorylated and unphosphorylated forms of the peptide, NH_2-Cys-Asp-Asn-Glu-Tyr-Thr-Ala-Arg-Glu-aminocaprase-Lys-ε-[florescence]-COOH, were synthesized and, in the last step, the recombinant protein-α-thioester and synthetic peptides were ligated at pH 7.0 in the presence of 2% thiophenol.

Bacterial potassium channel (KcsA) is a protein possessing 160 amino acids. Its truncated form corresponding to the first 125 amino acids was synthesized via EPL by Muir through a two fragment ligation of KcsA[1–73] with KcsA[74–125].[117] To facilitate its assembly through NCL, Thr[74] was replaced with Cys and this replacement was confirmed to be unaffected by the properties of the protein. A bacterial expression strategy was used to generate the *N*-peptide-α-thioester fragment {KcsA[1–73]}. The recombinant proteins were directly converted into α-thioesters through thiolysis of intein fusion protein. The KcsA[74–125] segment was synthesized using Boc/HBTU-SPPS. The best ligation results between these

two fragments were obtained at pH 7.6 using 50% polytetrafluoroethylene (PTFE), 1% sodium dodecyl sulfate (SDS) and 2% thiophenol (Scheme 49).



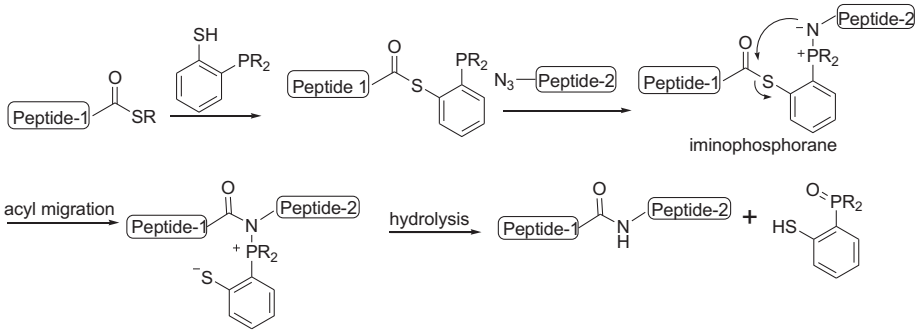
Scheme 49. Application of EPL in assembling KcsA fragment.

Ligation was complete in 2 h at 37 °C and the ESI-MS data of the purified product were found to be 13,634.8±4.4 Da, in agreement with the calculated value (13,633 Da). The folding reaction was carried out concomitantly in the same flask employing lipid buffer (100 mM Tris, pH 7.5, 200 mM NaCl, 15 mM KCl, 10% glycerol, 10 mM DTT).

4. Traceless Staudinger ligation

A recently developed technique for joining two large polypeptide fragments is Staudinger ligation.[31] This originates from the classical Staudinger reaction in which an azide and a phosphinothioester react to form an amide.[27] The reaction was successfully applied as the basis for a ligation technique by the groups of Bertozzi and Rains.[118,119] In the original Staudinger reaction, the products retained the phosphine moiety. However, suitable modifications have now been engineered to ensure that the product is free from any residual components and, hence, it is aptly named as 'traceless Staudinger ligation'. Similarly to NCL, one of the coupling partners is a peptide-α-thioester. The first step would be transthioesterification of a peptide-α-thioester with a phosphinothiol, which then reacts with the alkyl azide terminus of another peptide fragment to form an iminophosphorane intermediate. The iminophosphorane nitrogen attacks the thioester carbonyl to yield an amidophosphonium salt, which on P–N bond hydrolysis affords an amide bond along with a phosphine oxide as co-product (Scheme 50).

Since its discovery, several studies have been devoted to understanding the mechanistic and methodological aspects of traceless Staudinger ligation, the development of water soluble



Scheme 50. Working principle of traceless Staudinger ligation.

phosphinothiols, application to site-specific immobilization of peptides on surfaces, such as amine-derivatized glass slides, and utility to obtain glycopeptide and glycoconjugate systems.[120]

4.1. Case studies

Staudinger ligation, initially, was demonstrated using glycyl azides. Rains et al. reported Staudinger ligation at non-glycyl azides Phe, Asp and Ser as well and further determined that the ligation was racemization free.[119] Several pairs of enantiomeric dipeptides such as Ac-Gly-L-Phe-NHBn and Ac-Gly-D-Phe-NHBn, or Ac-Gly-L-Asp(OMe)-NHBn and Ac-Gly-D-Asp(OMe)-NHBn were made employing Staudinger ligation and were found to be optically homogeneous. Staudinger ligation was adopted to assemble ribonuclease A (RNAse A) through the condensation of three segments, viz., RNAse A(1–109), RNAse A(110–111) and RNAse A(112–124).[121] The RNAse A(1–109) fragment was expressed through a recombinant technique as a C-terminal thioester. The RNAse A(110–111) fragment was synthesized as a C-terminal phosphinothioester (Fmoc-Cys(Trt)Glu(O^{t}Bu)SCH$_2$PPh$_2$) using a sulfonamide linker and isolated in 64% yield. The third fragment, RNAse A(112–124), was synthesized as an N-terminal azide, for which the RNAse A(113–124) fragment [(n−1)peptide] was assembled via the Fmoc-HATU chemistry on a polyethylene glycol (PEG) resin followed by capping with N_3-CH$_2$-COOH (residue 112) to afford N_3CH$_2$CO-Asn(Trt)ProTyr(^{t}Bu)ValProValHis(Trt)-PheAsp(O^{t}Bu)Ala-Ser(^{t}Bu)Val. Staudinger ligation was employed to ligate RNAse A(110–111) with RNAse A(112–124) directly on the poly[acryloyl-bis(aminopropyl)polyethylene glycol] (PEGA) resin. After the ligation, cleavage of the peptide from the acyl resin, deprotection and purification afforded RNAse A(110–124). In the final synthetic step, this fragment was coupled to RNAse A(1–109)-α-thioester through NCL to obtain the product (Scheme 51).

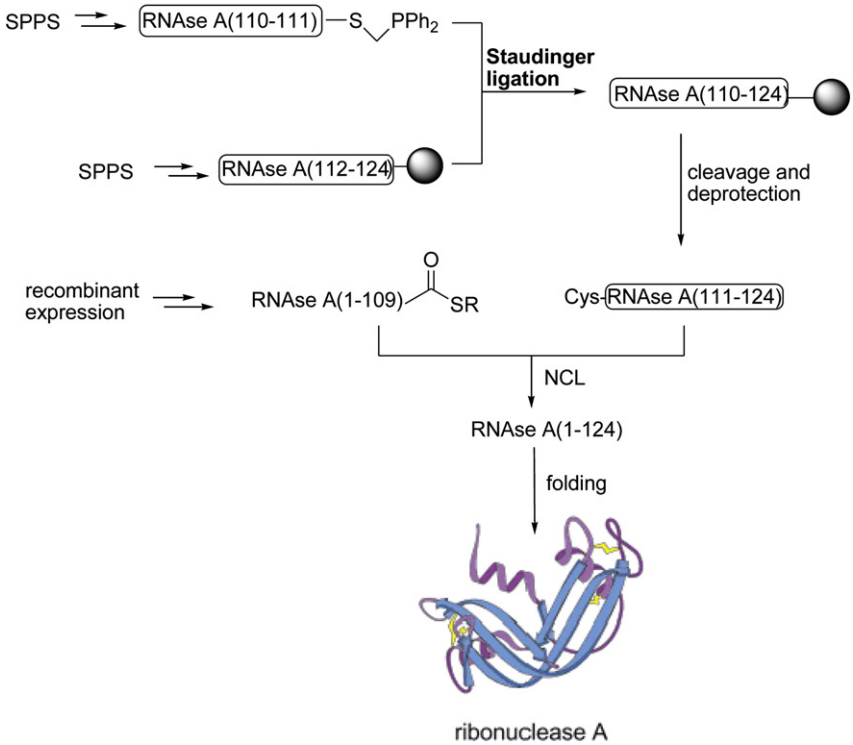
Lee et al. developed a novel core shell resin-supported phosphinothiol for solid phase Staudinger ligation of peptides and also demonstrated that the resin-bound phosphonothiol can be reused.[122] After detachment of the peptide from the resin, the resulting resin-bound phosphine oxide was recovered through filtration and reduced to the original form using trichlorosilane (Scheme 52).

Raines et al. described a water-soluble phosphinothiol for traceless Staudinger ligation. The basic phosphinothiol, bis(*p*-dimethylaminoethyl)phenyl phosphinomethanethiol (Fig. 12) was prepared and its efficacy was demonstrated through the synthesis of peptides.[123]

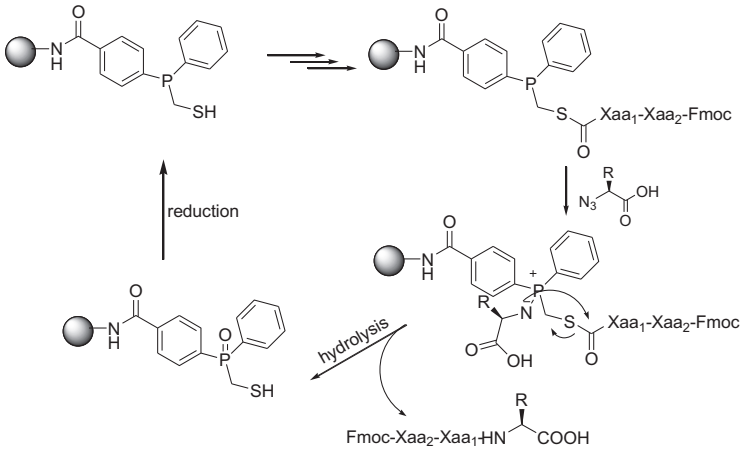
The traceless Staudinger ligation has also been explored for synthesizing cyclic proteins and peptides. In one such example, borane-protected phosphinothiol was employed to install the reactive azide and phosphine moieties in the same molecule.[124] After the removal of borane and other protecting groups, the peptide was subjected to ligation in the presence of 1,4-diazabicyclo[2.2.2]octane (DABCO) or diisopropylethylamine (DIPEA) (Scheme 53). In another interesting report, traceless Staudinger ligation was employed to prepare medium-sized lactams from peptides made of α- and β-amino acids.[125]

5. *O*-Acyl isopeptide method (click or switch peptides)

A major obstacle for the assembly of large peptides is their aggregation tendency that makes coupling and deprotection reactions incomplete and, consequently, their purification, extremely difficult. The synthesis of such 'difficult-sequence' peptides has been the subject of much attention.[126,127] The inter- and the intra-chain interactions, hydrophobic nature, and extensive hydrogen bonding in the molecular matrix extend the secondary structure of the peptide and thus the sequence elongation becomes sturdy. Due to the uncontrollable folding of peptide fragments during their synthesis, a significant proportion of N^{α}-amino groups becomes inaccessible for acylation as well as deprotection. Hence, incomplete segments (truncated and deleted sequence) emerge after cleavage from the resin as major byproducts. To disrupt such disfavourable aggregation in a growing peptide, the hydrophobic interactions must be kept at bay. Current strategies to prevent such



Scheme 51. Staudinger ligation employed during the synthesis of ribonuclease A.



Scheme 52. Reusable resin-bound phosphinothiol employed in Staudinger ligation.

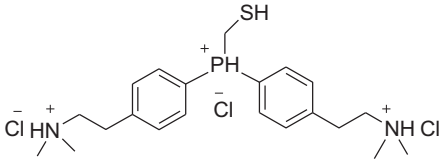
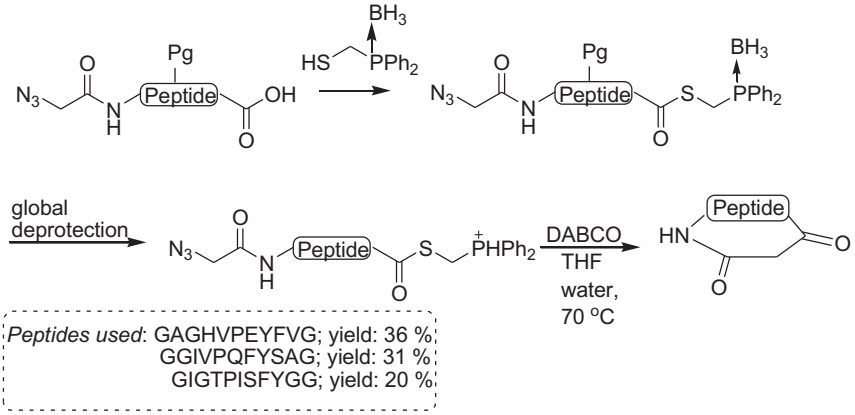


Fig. 12. A water-soluble phosphinothiol.

hydroxy group is acylated with an Fmoc-amino/peptide acid through HBTU activation. The $O\rightarrow N$ acyl migration leads to insertion of the amide bond (Scheme 54).

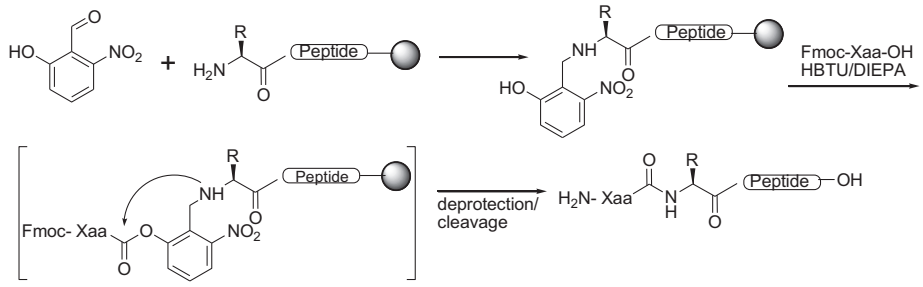
Kiso et al., during their studies on the anticancer prodrugs paclitaxel and docetaxel, found that the presence of an *O*-acyl instead of an *N*-acyl moiety increased the solubility of the intermediate. This facilitated the convenient assembly of the depsipeptide sequences, which in the final stage were subjected to



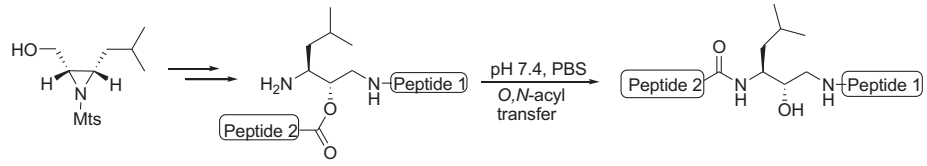
Scheme 53. Cyclic peptides via intramolecular Staudinger ligation.

difficulties have aimed at interfering with secondary structure properties by introducing specific reversible modifications at the peptide backbone using: (a) pseudoproline building blocks[128] and (b) auxiliaries such as a 2-hydroxy-4-methoxybenzyl (Hmb) moiety.[129] In the latter method, the auxiliaries disrupt the secondary structure formed by the inter/intra-chain interactions and facilitate N-acylation of the secondary amine. This is carried out through acyl capture of the activated amino or peptide acid employing the 2-hydroxy moiety of the auxiliary via a base-catalyzed reaction followed by an $O\rightarrow N$ intramolecular acyl transfer. The imposed C- and N-terminal proximity drives this reaction. The foundation for such a type of technique was laid by Brenner et al. more than 60 years ago.[130] Recently, Miranda et al. reported the use of a 2-hydroxy-6-nitrobenzyl auxiliary with the intention of improving the efficiency and yield of peptide synthesis including the couplings involving sterically hindered amino acids.[131] Initially, through reductive alkylation, the auxiliary is attached to a resin-bound peptide. The $O\rightarrow N$ intramolecular acyl migration to afford the unmodified target molecule.[132,133] These results were supported by the observations of Fujii et al. who also demonstrated the synthesis of a peptidomimetic containing a hydroxyethyl amine dipeptide isostere through $O\rightarrow N$ acyl migration (Scheme 55).[134] The $O\rightarrow N$ acyl migration most likely takes place through the formation of five-membered cyclic transition state in case of ligations involving *O*-acyl Ser/Thr residues. However, Katritzky et al. have carried out ligation from *O*-acyl serine sites in which $O\rightarrow N$ acyl migration takes place via 8- and 11-membered cyclic transition states.[135a] The method involves no usage of cysteine or an auxiliary group at the ligation site and causes the shift to one or two amino acid residues away from the *O*-acyl serine site.

Kartritzky's group had also demonstrated the synthesis of cysteine containing native peptides employing *S*-acyl cysteine sites and obtaining $S\rightarrow N$ acyl shift involving 5-, 11- and 14-membered cyclic transition states.[135b]



Scheme 54. Proximity induced *O*→*N* acyl migration.



Scheme 55. Peptidomimetic synthesis using *O*→*N* intramolecular acyl migration.

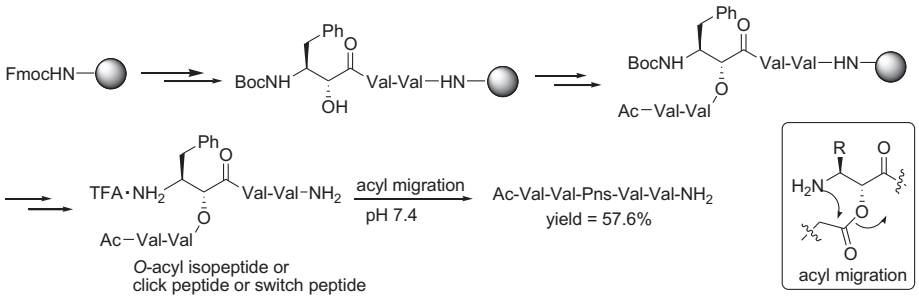
The above two observations triggered a new technique for the assembly of difficult peptide sequences. An interesting discovery of the quick, easy and irreversible transformation of *O*-acyl isopeptide into native peptide was appropriately termed as 'click peptide' method by Kiso[136] and as 'switch peptide' by Mutter and others.[137] The primary requirement is the presence of a β-hydroxy amino component. Proteinogenic amino acids Ser, Thr and unnatural amino acid, phenylnorstatine {(2*R*,3*S*)-3-amino-2-hydroxy-4-phenylbutanoic acid (Pns)}, provide the basic requirement. In a growing peptide on solid phase, an ester bond is installed between the amino acid and a β-hydroxy amino acid such as Ser/Thr and further peptide chain elongation is continued in a regular fashion. After the complete assembly of the desired sequence of amino acids, the peptide is cleaved from the resin, and the protector on Ser/Thr is released to obtain the *O*-acyl isopeptide (click peptide), which under physiological conditions (pH 7.4) in PBS at room temperature undergoes *O*→*N* intramolecular acyl migration to afford the final product.[138] The presence of one single depsi bond [−O−C(=O)−] in place of an amide in the sequence alleviates the unfavourable secondary structural foldings that otherwise would lead to aggregation.

The technique can be well understood by considering the synthesis of a pentapeptide, Ac-ValVal*Pns*ValVal-NH_2, as demonstrated by Kiso.[139] The synthesis of the pentapeptide by the conventional route leads to several byproducts and the solubility problems associated with this peptide posed difficulty during preparative HPLC purification as well, resulting in a very poor yield (6.9%). However, when it was assembled via the *O*-acyl isopeptide method, a satisfactory yield was obtained (18.5 mg, 57.6%, Scheme 56).

During the click approach, Boc-Pns-OH was coupled to the resin-bound Val-Val-NH_2 and subsequent couplings onto the β-hydroxy group of the Boc-Pns were performed using Fmoc chemistry. After assembling the desired sequence, the Fmoc group was cleaved and the free amino group was acetylated. After cleavage from the resin, the Boc protector was released to afford the click peptide. Compared to the native peptide, the 'click peptide' was highly water soluble (nearly 500 fold) and was therefore obtained in a homogeneous form using preparative HPLC. It was storable at 4 °C for 1 month without degradation. Up on stirring in PBS at pH 7.4, at a concentration of 1 mg/ml, the isopeptide underwent rapid acyl migration (<5 min) to afford the desired product quantitatively as a white amorphous solid.

Some advantages of the click method are:

1. the isopeptides are water soluble due to the protonated amino group and therefore can be purified easily.
2. stable in neutral or slightly basic conditions or as a solution in DMSO or as a solid obtainable after lyophilization.
3. the pH-triggered *O*-acyl migration is fast and quantitative, and yields the product in a short duration of time.
4. the presence of a depsipeptide bond helps the peptide sequence to retain a monomeric form and prevents undesired aggregation and fibril formation.
5. the deprotection and coupling efficacy are also enhanced in SPPS due to the presence of the *O*-acyl isopeptide unit.



Scheme 56. *O*-Acyl isopeptide method for the synthesis of 'difficult sequences'.

Carpino et al. demonstrated the advantage of having an ester bond in a peptide sequence for an efficient chain assembly. They chose an extremely difficult peptide—a 37-residue WW domain FBP28 (peroxisome proliferator-activated receptor-binding protein)—which exhibits a triple-stranded antiparallel β-sheet structure (Fig. 13).[140]

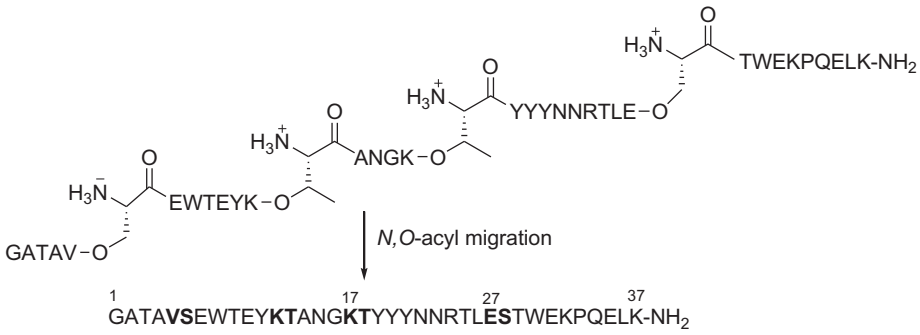
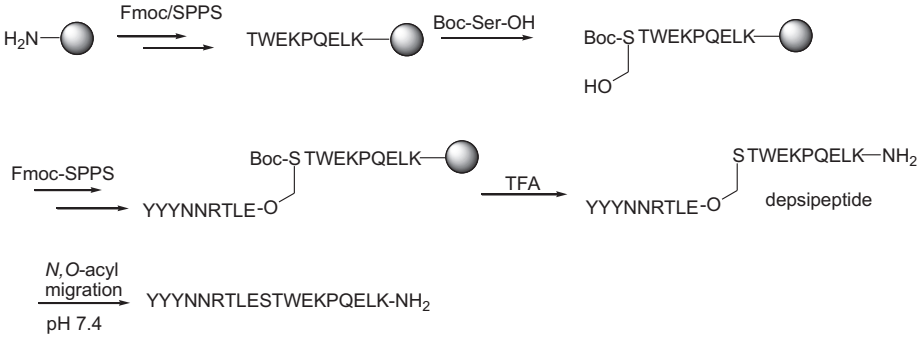


Fig. 13. Synthesis of peptide via multiple 'click' centres.

Even the short segment of this peptide—Y^{19}-K^{37}—was known to be a challenging sequence to assemble. Carpino's attempts to synthesize Y^{19}-K^{37} through SPPS via a TentaGel-SRam resin ended in a very poor yield of 25%. However, when a depsi [(−C═O)−O−] linkage was introduced at the E^{27}-S^{28} position, the aggregation problems were resolved and, after the acyl migration step, the target peptide was isolated in a decent yield. In this description, initially, the fragment, $TWEKPQELKNH_2$, was deposited on a TentaGel-SRam resin using standard Fmoc-SPPS. Boc-Ser-OH was then coupled followed by O-acylation of Ser with Fmoc-Glu(O^tBu)-OH in the presence of *N*-methylimidazole (NMI). The remaining amino acids were then assembled as usual. The resulting depsipeptide was cleaved from the resin using trifluoroacetic acid (TFA), and was subjected to *O*–*N* acyl migration to restore the peptide bond (Scheme 57). The product was confirmed through MALDI-MS (observed value: 2461.12 Da, calculated: 2461.19 Da for $[M+H]^+$).



Scheme 57. On-resin preparation of 'click peptide' followed by acyl migration in solution.

Carpino et al. also investigated the synthesis and use of depsipeptides as units for the automated synthesis of isopeptides.[140] This avoids performing O-acylation on resin, which is cumbersome and racemization prone. Two useful routes were proposed to synthesize a depsidipeptide, Boc-Thr(Fmoc-Ala)-OH (Scheme 58), which was employed for an automated synthesis of the 31-*mer* C-terminal segment of a globular protein, crambin. During the synthesis of Cram(16–46), the depsipeptide unit was introduced at $A^{27}T$ and was later rearranged to the target sequence. In another synthesis, the same peptide sequence was assembled by inserting AT depsipeptide units at two positions (i.e., at 27–28 and 38–39) and then rearranged to Cram(16–46) (ESI-MS $[M+2H]^{2+}$=1570.80; calculated=1571.82 Da; Scheme 59).[140]

5.1. *O*-Acyl isodipeptides

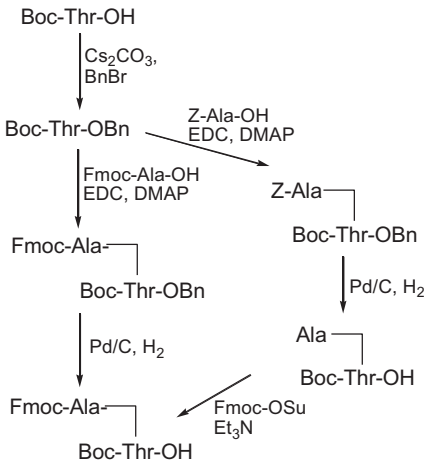
During the assembly of Ac-ValVal*Thr*ValVal-NH_2, Kiso observed the formation of an undesired isopeptide intermediate epimer, Thr(Ac-Val-D-*Val*)ValVal-NH_2·TFA (about 21%), along with the desired sequence.[141] Racemization at Val was due to the slow esterification on solid phase. This led to a new subsidiary approach for the click peptide protocol, namely '*O*-acyl isodipeptide' chemistry, which consists of carrying out the esterification of Boc-Ser/Thr with Fmoc-Xaa-OH in solution and then employing the resulting depsidipeptide unit in the subsequent couplings (Scheme 60).

The *O*-acyl isodipeptide approach for the racemisation-free synthesis of *O*-acyl peptides was a simple alternative, which was utilized for the synthesis of several peptide sequences including Aβ1–42.[142] Kiso's group, in a thorough study, synthesized all the possible combinations of *O*-acyl isodipeptides [Boc-Ser/Thr(Fmoc-Xaa)OH] to avail them readily as building blocks (Scheme 61).[143]

Carpino et al. also investigated the synthesis and use of depsipeptides as units for the automated synthesis of isopeptides.[140] This avoids performing O-acylation on resin, which is cumbersome and racemization prone. Two useful routes were proposed to synthesize a depsidipeptide, Boc-Thr(Fmoc-Ala)-OH (Scheme 58), which was employed for an automated synthesis of the 31-*mer* C-terminal segment of a globular protein, crambin. During the synthesis of Cram(16–46), the depsipeptide unit was introduced at The enantiomeric purity of these depsipeptide units was established flawlessly by synthesizing all the 38 D-isomers [Boc-Ser/Thr(Fmoc-D-Xaa)OH] and comparing the HPLC data.

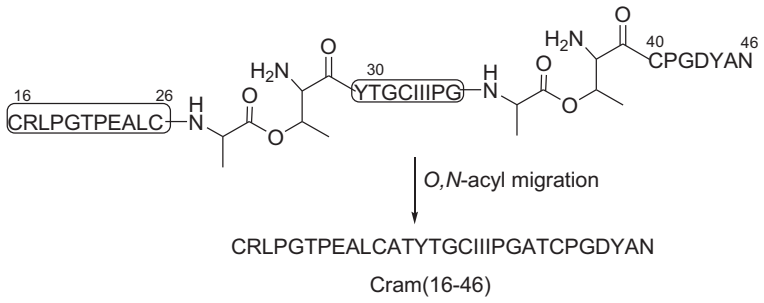
5.2. Synthesis of amyloid β-peptide (Aβ1–42)

A successful application of the *O*-acyl isopeptide method was demonstrated by the preparation of amyloid peptides,[144] which

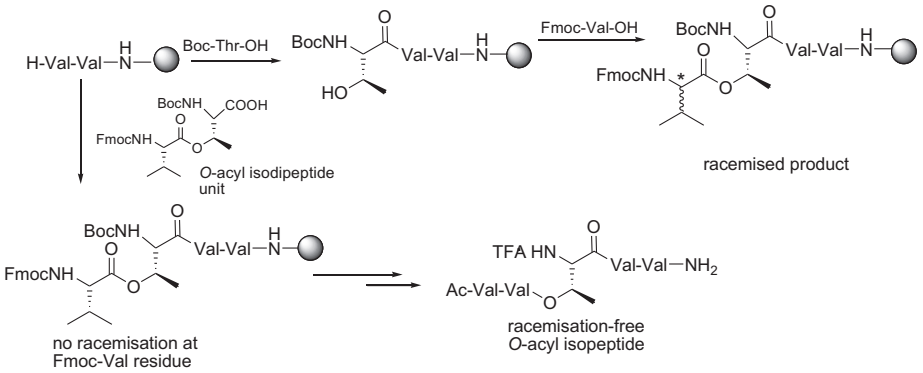


Scheme 58. Synthesis of depsidipeptide Boc-Thr(Fmoc-Ala)-OH.

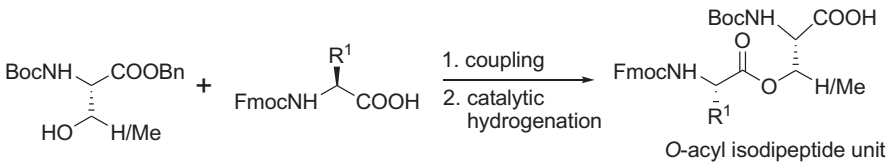
concentration and time-dependent aggregation has been a bottleneck to carry out extensive biological studies. This is because, several properties such as neurotoxicity depend on its β-structure and, if the peptide undergoes structural deformation before reaching the target site, the success of the study will be diminished. It can, however, be overcome by developing a protocol to deliver the peptide on to the biological site in a monomeric form. This was the basis for the synthesis of Aβ1–42, which was accomplished through the *O*-acyl peptide method. Out of the entire length of the peptide, the amide bond between Gly[25] and Ser[26] was isomerized into an ester bond and the rest of the sequence was assembled following standard protocols. After cleavage from the resin followed by deprotection of the Boc group from Ser[26], the resulting click peptide *O*-acyl isoAβ1–42 was easily purified through HPLC owing to its solubility in aqueous media (15 mg/ml). The presence of the ester bond between Gly[25]-Ser[26] in the click peptide blocked the secondary structural properties completely. The 26-*O*-acyl isoAβ1–42 was stable at 4 °C for about 1 month. In PBS, at pH 7.4, the acyl migration was rapid



Scheme 59. Synthesis of cram(16–46) demonstrating the utility of the depsidipeptide Boc-Thr(Fmoc-Ala)-OH.

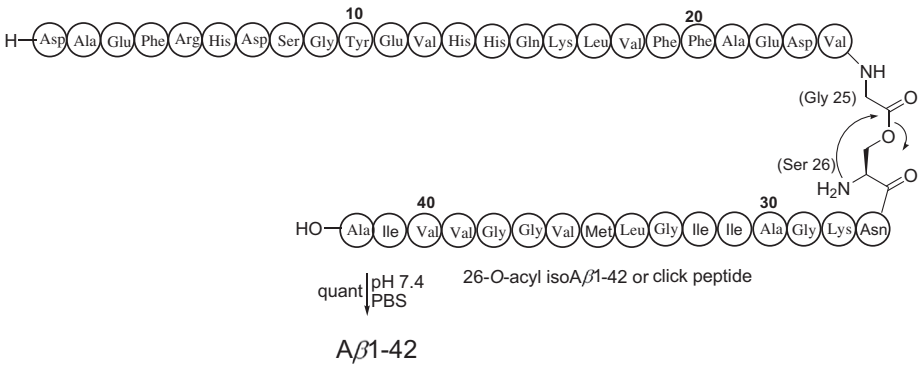


Scheme 60. Use of *O*-acyl isodipeptide circumvents the racemisation.



Scheme 61. Synthesis of *O*-acyl isodipeptide units.

are the components of amyloid plaques found in the brains of patients with Alzheimer's disease. Aβ1–42 is a predominant peptide having 42 residues and is also considered as one of the most difficult peptides to assemble. Its tendency to undergo with a half-life of 2.6 min and quantitative conversion into the final product Aβ1–42 was achieved (22.3 mg; Scheme 62). At pH 4.9, the migration rate was lower with a half-life of 3 h and, at pH 3.5 (acetate buffer), the rearrangement ceased completely.



Scheme 62. Synthesis of 42-*mer* amyloid β-peptide employing *O*-acyl isopeptide approach.

Practical procedures for the solid-phase assembly of difficult peptides by inserting depsilinkages and pseudoproline units have been reported by Coin et al.[140] Stepwise procedures, purification techniques and methods to tackle the synthetic difficulties have been described, which adopt direct application of $O \rightarrow N$ acyl migration or pseudoproline-forming reactions for the efficient synthesis of peptides that otherwise induce unfavourable conformational constraints during chain assembly.[9]

The *O*-acyl isopeptide method has been used for the synthesis, in complete solution phase, of difficult sequences of few oligopeptides such as transmembrane protein fragments.[145]

5.3. Switch peptides

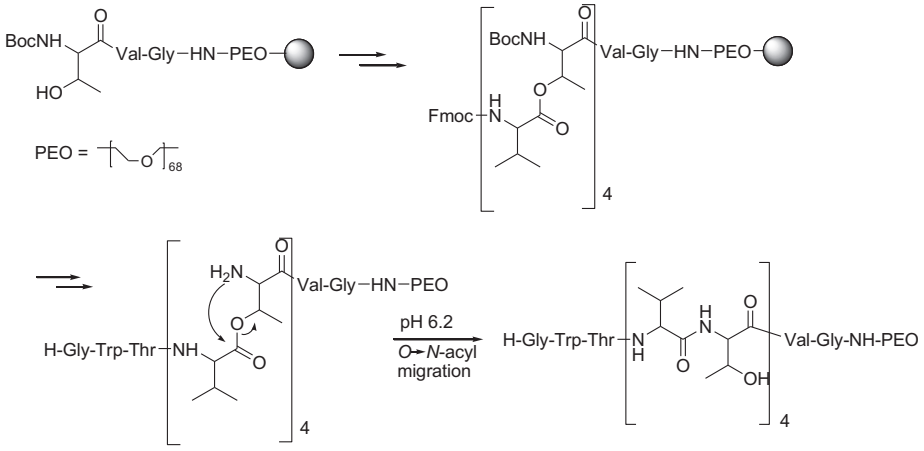
Borner et al. developed a concept known as 'switch peptide' protocol during their studies on peptide guided assembly of polyethylene oxide (PEO)–peptide conjugates.[137b] The authors' interest was in controlling the structure-formation processes in synthetic polymers mediated by peptides. The designed peptide attached to the PEO had five diades of alternating Thr and Val residues [(Thr-Val)$_5$,], the conventional synthesis of which would complicate the chemistry due to their instantaneous secondary structure forming nature. In order to disrupt the aggregating behaviour during the chemical assembly, acyl linkages were introduced between each Val-Thr pair. After the synthesis of peptides followed by detachment from the solid support, the Boc protections on the switch segments were removed and allowed for controlled and simultaneous $O \rightarrow N$ acyl migrations at pH 6.2 for about 1 week (Scheme 63).

5.4. Phototriggered $O \rightarrow N$ acyl migration

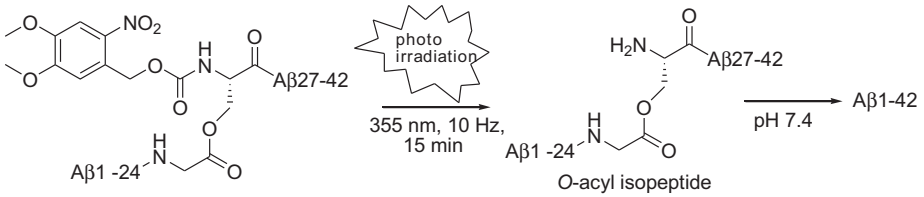
Considering the advantages of phototriggered stimulations in studying biological processes of peptides and proteins, photocleavable protecting groups were envisaged in place of Boc protectors for Ser/Thr residues. By fine tuning the area and duration of irradiation, a spatiotemporal control of the peptide preparation and its functions was enabled. Kiso's group employed 6-nitrovertryloxycarbonyl (Nvoc) as a photocleavable amino-protecting group and demonstrated its utility for the assembly of Aβ1–42. The N-terminus of Ser[26] was protected as Nvoc and the rest of the peptide synthesis was carried out under standard protocols (Scheme 64).[146] The protected acyl peptide was stable, even after incubating at 37 °C in PBS. It was also stable up to 24 h in PBS buffer, and for 3 months either as the solid or as a solution in DMSO at −20 °C. Upon photoirradiation (355 nm, 10 Hz, 5 mJ) for 15 min at 4 °C, the Nvoc group detached completely and the *O*-acyl peptide was obtained in excellent yield.

A few other photocleavable groups employed in click peptide chemistry include 4-(dimethylamino)phenyloxycarbonyl (MaPoc), [7-(diethylamino)coumarin-4-yl]methyloxycarbonyl (DEAC-Moc), (7-aminocoumarin-4-yl)methyloxycarbonyl (Ac-Moc), and {7-[bis(carboxymethyl)amino]coumarin-4-yl}methyloxycarbonyl (BCMAC-Moc).[147]

Mutter et al. reported the controlled self-assembly of amyloid β-derived peptides in vitro by consecutive triggering of acyl migrations.[148] Sequential triggering of $O \rightarrow N$ acyl migrations in amyloid β derived switch peptides has been employed as a tool to understand the initiation and inhibition of polypeptide folding, self-



Scheme 63. Switch peptide guided assembly of polyethylene oxide (PEO)–peptide conjugates.



Scheme 64. Use of photolabile protecting groups in 'click peptide' chemistry.

assembly and aggregation. A β-amyloid peptide derivative was prepared initially and the conformational behaviour was temporarily arrested by inserting switches (Ser/Thr/Cys) at different intervals. The orthogonal N^{α}-protection for each switch ensures selective triggering of migration at the desired *O*-acyl junctions, thereby enabling studying the effect of each segment on the conformational behaviour and molecular assembly after undergoing acyl migration. A representative scheme with two switch positions and protecting groups along with the corresponding triggering system is outlined in Fig. 14. When acyl migration at Y^1 is executed, the depsi linkage at Y^2 remains intact and the resulting peptide shows the conformational behaviour due to that particular peptide segment where the amide bond is newly established. In a similar way, it is possible to introduce several orthogonal switches at the desired positions and study the individual contribution of each segment to the overall aggregation property by selectively triggering the *O*-acyl migration.

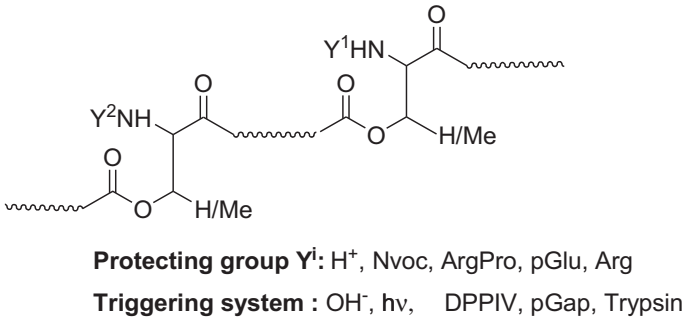


Fig. 14. A model system comprising two orthogonal switch centres.

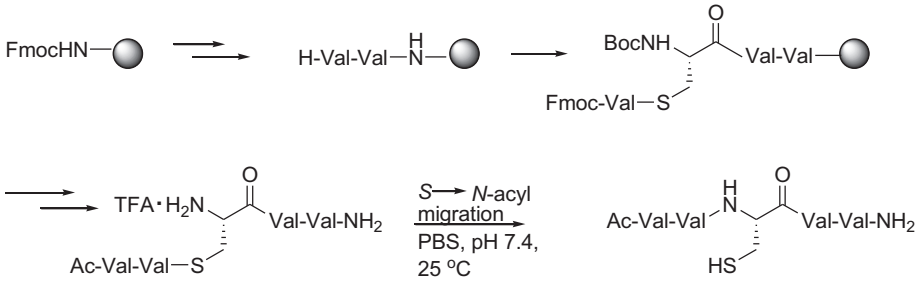
The click peptide protocol has also been demonstrated for difficult peptides with cysteine residues. In place of the *O*-acyl derivative, the *S*-acyl derivative preserves the monomeric integrity and, under physiological conditions, the *S*→*N* acyl migration restores the amide bond.[149] During the synthesis of Ac-ValValCys-ValVal-NH_2, the TFA salt of the *S*-acyl peptide was found to be easily soluble in water (8.5 mg/ml) and therefore was easily purified through HPLC in an overall 10% yield. The *S*-acyl isopeptide was found to be stable (for 1 day in DMSO or methanol and for 3 months when stored as a solid at room temperature) and the acyl migration is quantitative and complete in a very short time (within 1 min; Scheme 65).

6. Recent developments

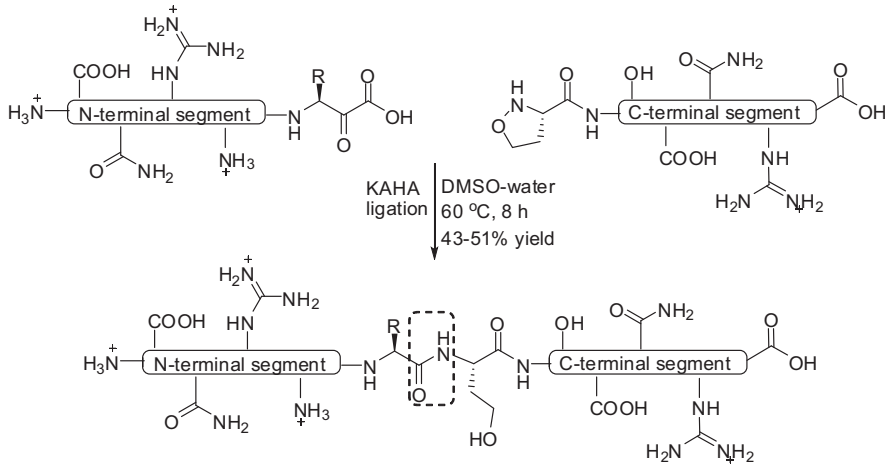
6.1. KAHA ligation

Bode et al. have developed an alternate and general route for the reagentless and native peptide bond forming ligation through the use of α-ketoacid and hydroxylamine (KAHA) bearing peptide fragments.[150] The initial success in this direction was obtained when the peptide with C-terminal α-keto acid was reacted with the peptide bearing N-terminal hydroxylamine leading to chemoselective ligation with the formation of an amide bond at the site of ligation. The main problem of this strategy was the difficulty of performing ligation in aqueous media typically employed for handling medium-sized peptides. This limitation was overcome when the hydroxylamines were replaced by *O*-benzyl hydroxylamines, which provided a clean and rapid ligation in water, but the latter tend to undergo facile elimination leaving them unsuitable for α-peptide derived substrates. Later, it was discovered that 5-oxaproline is a stable and viable alternative to hydroxylamines for KAHA ligations in water. The oxaproline reacts chemoselectively with peptide-α-ketoacids in aqueous medium to afford native peptide, generating homoserine (HSer) residue at the site of ligation (Scheme 66).[150]

The utility of this method was demonstrated by synthesizing two medium-sized proteins from *Mycobacterium*, the prokaryotic-ubiquitin-like protein (Pup, 63 residues) and one of its target proteins, probable cold shock protein A (cspA) having 66 residues. For the preparation of Pup through KAHA ligation, the target peptide was bisected at Leu^{32}-Thr^{33} site. The two fragments were made on an automated SPPS and during the preparation of the C-fragment, the Thr^{33} was replaced with (*S*)-*N*-Boc-5-oxaproline. The ligations were performed in DMSO/water (6:4) mixture containing oxalic acid (0.1 mol) at 60 °C for 8 h. The oxalic acid promoted solubilization as well as increased ligation rate. Importantly, in the isolated protein, it was confirmed through CD spectral studies, that the replacement of Thr with HSer does not affect the protein secondary structure. Hence it was envisioned that HSer can serve as surrogate



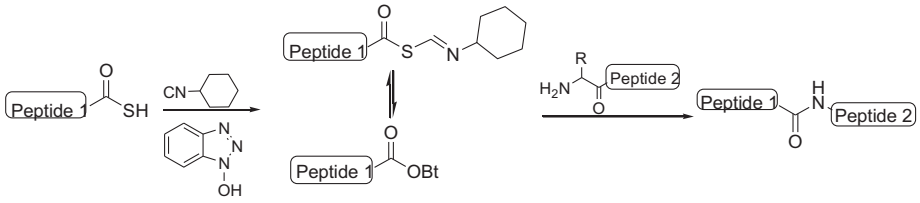
Scheme 65. A click peptide system involving *S*→*N* acyl migration reaction.



Scheme 66. Ligation involving KAHA peptides.

for Thr, Ser, Met and Asp residues, thus validating the KAHA–oxaproline ligation as a general method for peptide bond forming ligations.[150]

polypeptides in 60–82% yield. The reaction proceeds presumably through a thioformimidate carboxylate mixed anhydride intermediate (Scheme 67).



Scheme 67. Mechanism of acylation of N-terminus of a peptide with a peptide thioacid in presence of an isonitrile component.

6.2. Ligation employing peptide thioacids

Danishefsky et al. have developed a general and efficient ligation protocols for assembling polypeptides and glycopolypeptides by exploiting the reactivity of thioacids.[151] They designed a reaction involving a peptide thioacid, cyclohexylisonitrile, N-terminal peptide and HOBt to afford the corresponding ligated peptide. This strategy was successfully applied to synthesize several

Further, the authors conducted a control experiment without the addition of the isonitrile component and, to their surprise, found that the reaction worked even better and the respective peptides were isolated in 77–87% yield in a shorter reaction time (Fig. 15). The procedure could also be used for the preparation of cyclic peptides by the use of N^{α}-unprotected peptide thioacids, albeit the reaction was found to be slow; and for the synthesis of glycopeptides as well (Fig. 16).

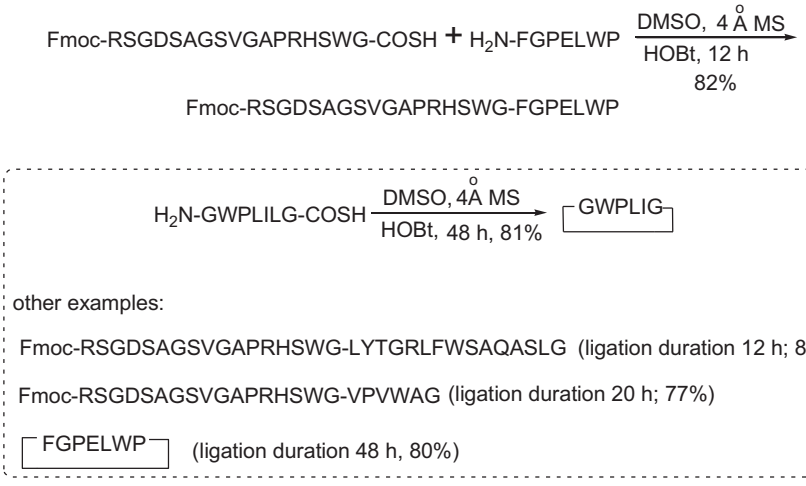


Fig. 15. HOBt mediated peptide ligation.

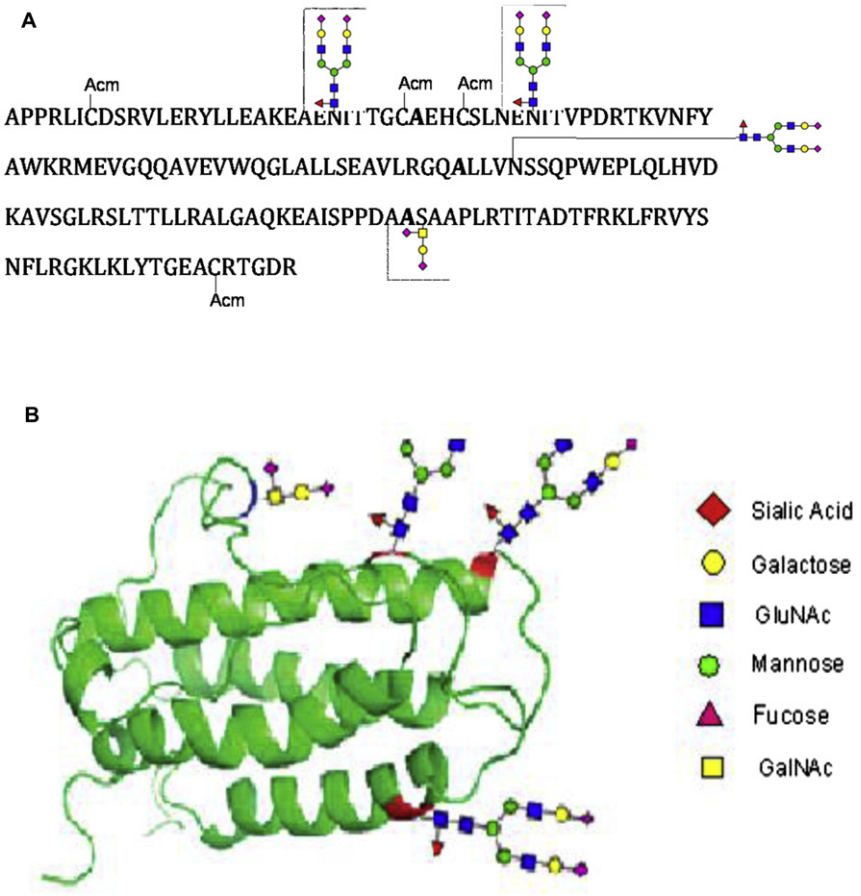
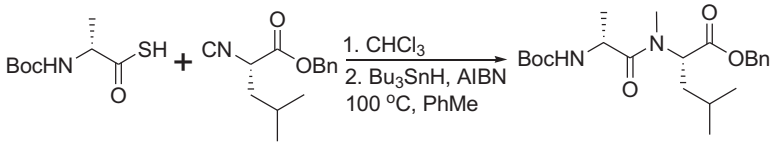


Fig. 16. Structure of EPO. A: amino acid sequence of EPO; B: ribbon diagram with selected oligosaccharides.

Danishefsky et al. have also demonstrated the reaction of thioacid and isonitrile as a possible way to access amides efficiently.[30,152] The chemistry was explored to synthesize *N*-methylated peptides also (Scheme 68).

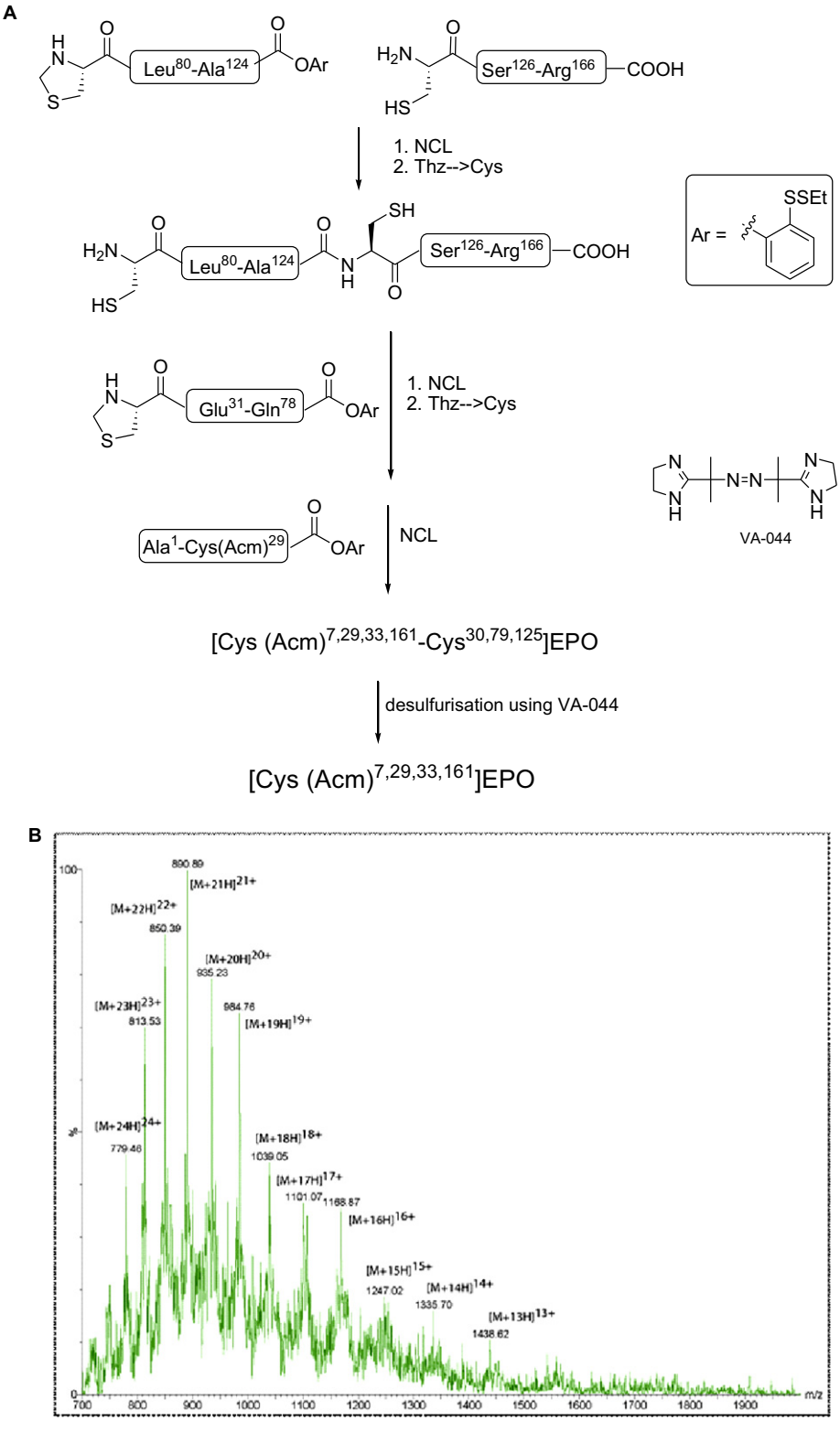


Scheme 68. Reaction of *N*-protected amino thioacid with isocyano ester leading to a peptide ester.

6.3. NCL at its best: total synthesis of non-glycosylated erythropoietin

Recently, in 2012, Bralisford and Danishefsky[153] and Kent et al.[154] independently accomplished the total synthesis of the non-glycosylated form of 166-amino acid long protein hormone, erythropoietin (EPO) using NCL. This marks a landmark achievement in natural product synthesis especially that of proteins with tunable structures, and exemplifies the present state of art of the technique as well as the future potential of NCL as a vital synthetic tool in total synthesis of proteins. The syntheses are built-in with the wise combination of different strategies and methods developed thus far for NCL. EPO is a glycosylated protein hormone, which is secreted by the kidneys and stimulates the production of blood cells. The hormone is used for the treatment of anaemia. The glycan units play an important role in the bio-synthesis as well as the functioning of the hormone. EPO that is produced through recombinant methods is obtained as heterogeneous mixture of glycoforms, which differ with respect to glycan residues, thus complicating the studies on the structure–activity relationship (SAR) of the hormone. For SAR studies, it is strongly desirable that a homogeneous sample of EPO is available and that there is a synthetic control over the amount and the content of glycosylation. The first step in addressing this challenge has been taken through the accomplishment of the total chemical synthesis of the non-glycosylated form of EPO.

Synthesis of EPO via NCL is challenging as the native Cys residues ($Cys^{7,29,33,161}$) are not evenly located. Thus 'Xaa-Cys' NCL cannot be operated with such an uneven distribution of Cys residues. Bralisford and Danishefsky envisioned the synthesis of EPO through NCL at the Ala sites of four EPO fragments of comparable lengths (Scheme 69).[153] Non-native Cys residues, which serve as Ala surrogates were placed at positions 30, 79 and 125 and the EPO was envisaged to arise from the sequential fusion of four fragments Ala^{1}-Cys^{29}, Cys^{30}-Gln^{78}, Cys^{79}-Ala^{124} and Cys^{125}-Arg^{166} through three NCL steps. On the other hand, the native Cys residues were

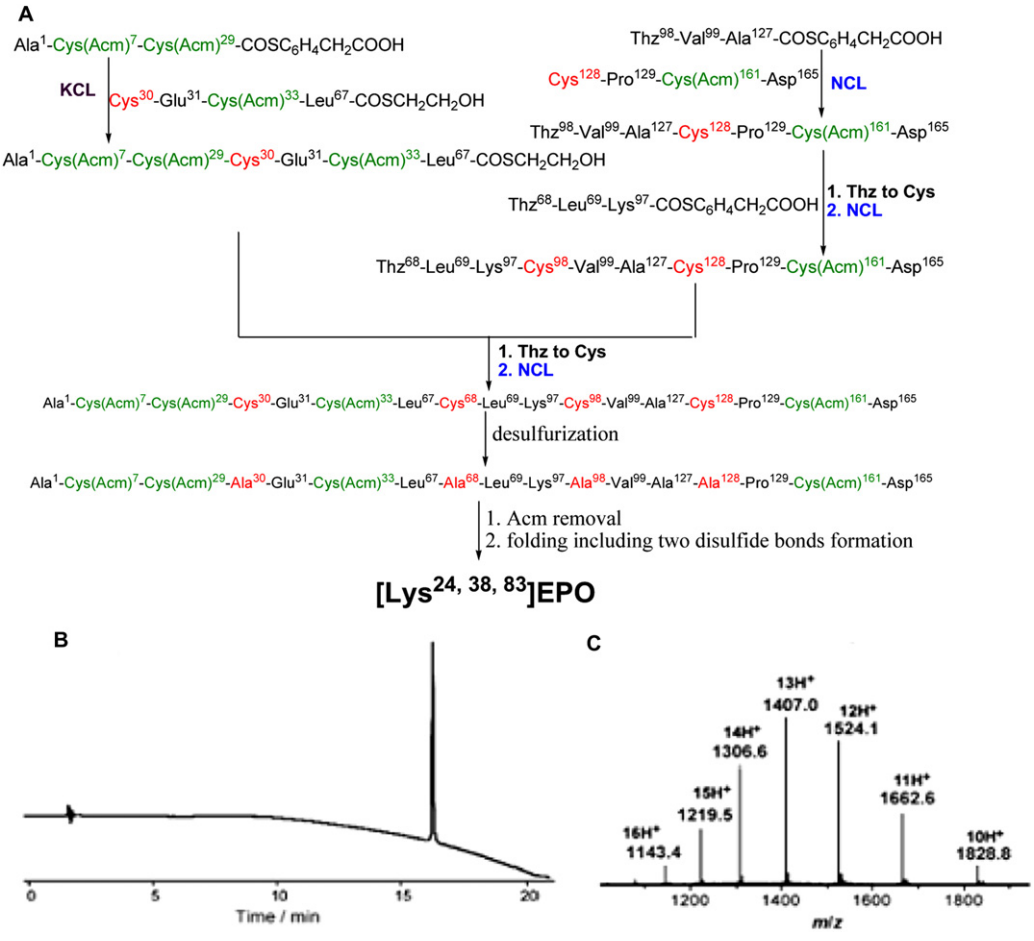


Scheme 69. Synthesis of non-glycosylated erythropoietin possessing four Acm protected Cys residues.[153] A. Synthetic protocol through convergent merger of four peptide fragments via the sequence of NCL. B. ESI-MS of ACM protected EPO: *m/z* calculated for $C_{827}H_{1336}N_{236}O_{246}S_5$ $[M+13H]^{13+}$ 1437.15, found 1438.62; $[M+14H]^{14+}$ 1334.56, found 1335.70; $[M+15H]^{15+}$ 1245.66, found 1247.02; $[M+16H]^{16+}$ 1167.87, found 1168.87; $[M+17H]^{17+}$ 1099.23, found 1101.07; $[M+18H]^{18+}$ 1038.22, found 1039.05; $[M+19H]^{19+}$ 983.63, found 984.76; $[M+20H]^{20+}$ 934.50, found 935.23; $[M+21H]^{21+}$ 890.05, found 890.89; $[M+22H]^{22+}$ 849.63, found 850.39; $[M+23H]^{23+}$ 812.74, found 813.53; $[M+24H]^{24+}$ 778.92, found 779.46.

protected with Acm groups. Innate to the design of the synthesis, the N-terminal Cys residues of the middle fragments were protected as thiazolidine derivatives to circumvent undesired self-ligation. The required four fragments were assembled following Fmoc-based SPPS on NovaSyn TGT resin and peptides Ala[1]-Gly[28], Thz[30]-Gly[77] and Thz[79]-Asp[123], after cleavage as acids, were converted into *ortho*-thiophenolic esters by coupling with the final amino acid bearing thiophenolic ester. These stable thiophenolic esters were introduced to serve as surrogates for the reactive arylthioesters. The C-terminus thio-activations of the peptides

were achieved prior to the commencement of NCL by the reduction of the disulfide bond of the phenolic ester followed by $O \rightarrow S$ acyl migration. The Thz residues at the site of ligation were unmasked into Cys in a sequential manner by hydrolysis with methoxyamine hydrochloride ($MeONH_2 \cdot HCl$) at pH 4.0 before the NCL step. The four EPO fragments were combined through reiterative NCL and Thz hydrolysis sequences. The ligations were executed in a phosphate buffer of pH 7.2–7.4. The products of ligations were purified through HPLC as well as size exclusion centrifugal filtration (in case of fragment Cys^{30}-Arg^{166}, which could not be purified through HPLC). Upon the assembly of the full-length protein, the non-native Cys residues were selectively converted into native Ala residues by reduction via metal-free desulfurization using water-soluble radical initiator VA-044. Thus, 1.4 mg (67% yield) of the Acm protected EPO was obtained, purified through centrifugal filtration and characterized through ESI-MS.

Kent et al.[154] synthesized the full-length and fully folded [$Lys^{24,38,83}$]EPO by the combination of NCL with KCL (Scheme 70). In this EPO mutant, the native $Asn^{24,38,83}$, which are the sites of N-glycosylation are replaced with Lys residues. The [$Lys^{24,38,83}$]EPO produced through recombinant methods has been used for crystallographic studies and shown to retain biological activity. Kent et al.[154] approached the synthesis by the merger of five EPO fragments viz., Ala^1-Cys^{29}, Ala^{30}-Leu^{67}, Ala^{68}-Lys^{97}, Ala^{98}-Ala^{127} and Ala^{128}-Asp^{165} that carry non-native Cys residues (as Ala surrogates) at the ligation sites. The native Cys residues were protected with Acm groups. Although there were three middle fragments, which impose the temporary protection of Cys, only two were protected as thiazolidine derivative, while the Cys^{30} of the other fragment was left unprotected since the C-α-thioalkylester of this fragment was selectively ligated to Ala^1-Cys^{29} through KCL. In KCL, difference in reactivity between aryl- and alkylthioesters is exploited to achieve selectivity in ligation. The larger fragment Cys^{68}-Asp^{165} was obtained by sequential NCL and Thz hydrolysis of fragments Thz^{98}-Ala^{127} and Cys^{128}-Asp^{165} and Thz^{68}-Lys^{97}. Ala^1-Leu^{67}-α-thioalkylester was prepared by performing KCL at pH 6.8 without the addition of thiol catalyst. The full-length chain was obtained by the merger of the thioalkylester to Cys^{68}-Asp^{165} through an NCL step. This ligation was achieved at pH 6.8 in presence of 4-mercaptophenylacetic acid (MPAA) catalyst and took 15 h for completion. In addition, an unusually large amount of Ala^1-Leu^{67}-OH was formed as byproduct due to the hydrolysis of the thioalkyl ester. It is also noteworthy, that during the ligation of Thz^{68}-Lys^{97} fragment, the MPAA mediated removal of Dnp protection on His^{94}, generated a more reactive α-arylthioester, $-COSC_6H_4CH_2COOH$ through transthioesterification. From the full-length polypeptide, the non-native Cys residues at the ligation sites were converted to native Ala through VA-044 mediated metal-free desulfurization. The resulting Acm protected EPO was purified through RP-HPLC and was characterized through ESI-MS (mass observed: 18,566.2±0.7 Da; calculated 18,566.4 Da, average isotopes). The four Acm protections were then cleaved by treating with AgOAc in



Scheme 70. Synthesis of human non-glycosylated [lysine^{24,38,83}]EPO.[154] A. Synthetic protocol employing convergent strategy through NCL and KCL. B. Analytical HPLC trace of folded, synthetic [$Lys^{24,38,83}$]EPO. The chromatographic separation was carried out on an analytical C3 column (Agilent ZORBAX 300SB, 4.6×150 mm; λ=214 nm) using a linear gradient of 5–65% buffer B in buffer A over 15 min (buffer A=0.1% TFA in H_2O; buffer B=0.08% TFA in acetonitrile). C. ESI-MS spectrum of mutant EPO obtained by direct infusion (mass observed: 18,277.9±0.5 Da; calculated 18,278.1 Da).

AcOH/H_2O (1:1) followed by 1,4-dithiothreitol (0.2 M). Thus, 1.9 mg (41%) of pure polypeptide [Ala^1-$Cys^{7,29,33,161}$-$Ala^{30,68,98,128}$-Asp^{165}] EPO was obtained (mass observed: 18,282.6±1.4 Da; calculated 18,282.1 Da, average isotopes). Folding of the fully unprotected EPO was achieved through a three-step sequential dialysis of the guanidine hydrochloride (6 M) solution.

The use of NCL in the total chemical synthesis of polypeptides and proteins of varied length, structure (including post-translationally modified molecules)[155] and activity is fast growing. The recent examples in this context include the synthesis of covalent dimer of HIV-1 protease enzyme (one of the targets in the therapeutic treatment of AIDS, 203 amino acid residues)[68] fully synthetic homogeneous human follicle-stimulating hormone (FSH, α-subunit of human glycoprotein hormone, 92 residues),[156] human glycosyl-interferon β (166 amino acids, antitumour agent)[157] and K48-linked tetraqbiquitin protein[91c] and ubiquitinated peptides (with varying lengths) and its oligomers.[158] The increasing number of successful syntheses of polypeptides using NCL certainly indicates that chemical assembly of transmembrane proteins including GPCR is most likely will be achieved in near future.

7. Conclusions

The need for understanding the biology of proteins and peptides has resulted in several techniques that have made possible the synthesis of large proteins often on a multimilligram scale, thus fulfilling the demands of biochemical research. The quantity of proteins obtainable through existing protocols is just enough to identify, characterize and study a couple of properties and there is a need to increase the quantities. More than a century-old toil in the area of protein research has resulted in several innovations, most of these have the potential restricted to the assembly of small peptides and their oligomers. Nature has engineered the art of making proteins without the assistance of protection for diverse functionalities of amino acids. Emulating this has remained a daunting challenge for the organic chemists. In this regard, a new prospective is the *p*rotecting-group-*f*ree *s*ynthesis (PGFS), which is an emerging area in various subsidiaries of organic synthesis. The strategies that depend on ligation of unprotected peptide fragments either synthesized chemically or engineered biologically have been most successful in assembling proteins. NCL is a robust technique and has achieved a great reputation among the chemical and biochemical research community. The success of NCL is attributed to the development of its biosynthetic version, EPL, which is contributing considerably to the assembly of proteins. The practical utility of protein splicing for recombinant generation of large protein fragments bearing ligatable termini has been a huge leap forward in our exercise to unravel nature's mysteries. Traceless Staudinger ligation is an emerging protocol developed for peptide ligation, which is yet to blossom to its full strength. Similarly, the click (switch) peptide method is a growing field with the potential for assembling difficult peptides through the incorporation of temporary depsi linkages in the peptide backbone. All these chemical ligation methods have made possible the realization of Emil Fischer's century-old dream of chemical synthesis of enzymes and proteins in a test tube. These techniques are contributing to expanding the horizon of our understanding about the fascinating biomolecules.

Acknowledgements

V.V.S. acknowledges Prof. Fred Naider, CUNY, New York for introducing him to the fascinating world of NCL during a post-doctoral assignment in his group. He is grateful to late Prof. K.M. Sivanandaiah, Bangalore University for his guidance in the initial stages of establishment of peptide research group at Central College, Bangalore (CCB) and to late Prof. B.S. Sheshadri for his untiring support in providing laboratory space at CCB in 1998. He is ever thankful to Drs. H.N. Gopi and K. Ananda for their toiling work in establishing the group during its infancy. He is grateful to Prof. M.S. Thimappa, Former Vice-Chancellor of Bangalore University for extending academic support for the sustenance of research group at CCB. The Department of Science and Technology, Council of Scientific and Industrial Research, Department of Biotechnology, University Grants Commission Govt. of India, New Delhi and Board of Research in Nuclear Sciences, Mumbai are thanked for generous financial support. H.P.H. thanks DST-Nano mission project and N.N. acknowledges the support of CSIR for an SRF position.

References and notes

1. (a) Sewald, N.; Jakubke, H. D. *Peptides: Chemistry and Biology*; Wiley-VCH: Darmstadt, 2002.
2. Strasberg, R. L.; Feingold, E. A.; Klausner, R. D.; Collins, F. S. *Science* **1999**, *286*, 455–457.
3. (a) *Synthesis of Peptides and Peptidomimetics (Houben-Weyl)*; Goodman, M., Felix, A., Moroder, L., Toniolo, C., Eds.; Georg Thieme: Stuttgart, Germany, 2003; Vols. E22a–d; and references cited therein; (b) Wieland, T.; Bodanszky, M. *The World of Peptides: A Brief History of Peptide Chemistry*; Springer: Stuttgart, Germany, 1991.
4. Fischer, E.; Fourneau, E. *Ber. Dtsch. Chem. Ges.* **1901**, *34*, 2868–2877.
5. Bergmann, M.; Zervas, L. *Ber. Dtsch.Chem. Ges.* **1932**, *65*, 1192–1201.
6. du Vigneaud, V.; Ressler, C.; Trippet, S. *J. Biol. Chem.* **1953**, *205*, 949–957.
7. Some of the important peptides that are manufactured in chemical industry and marketed as drugs are: fuzeon (HIV fusion inhibitor, 36-amino acid sequence), antagon (GnRH analog, 10-amino acid peptide), pitocin (oxytocin analog, a nonapeptide), firazir (bradykinin analog, decapeptide), angiomax (hirudin analog, 20-amino acid peptide), cetrotide (10-residue GnRH analog), sandostatin (somatostatin analog, 8 amino acids), glypressin (vasopressin analog, chain length: 12 amino acids), angiomax (hirudin analog, 20 amino acids), eptifibatide (integrin antagonist, 7 units) and zadaxin (thymocin, 28 amino acids). See: Gracia, S. R.; Gaus, K.; Sewald, N. *Future Med. Chem.* **2009**, *1*, 1289–1310.
8. (a) Merrifield, B. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154; (b) Merrifield, B. *Science* **1986**, *232*, 341–347; (c) Merrifield, B. *Protein Sci.* **1996**, *5*, 1947–1951.
9. Coin, I.; Beyermann, M.; Bienert, M. *Nat. Protoc.* **2007**, *2*, 3247–3256.
10. (a) Sieber, P.; Eisler, K.; Kamber, B.; Riniker, B.; Ritter, W.; Marki, F.; Degasparo, M. *Hoppe-Seyler's Z. Physiol. Chem.* **1978**, *379*, 113–119; (b) Yajima, H.; Fujii, N. *Biopolymers* **1981**, *20*, 1859–1867; (c) Kenner, G. W. *Proc. R. Soc. London, Ser. A.* **1977**, *353*, 441–447.
11. SPPS of bovine pancreatic ribonuclease A (a 124 amino acid enzyme) was accomplished by R.B. Merrifield on copolystyrene cross-linked with 1% divinyl benzene resin employing DCC/Boc-chemistry. After several purifications, the synthetic enzyme was isolated in 3% overall yield based on the amount of Val residue esterified to the resin. The synthetic enzyme had approximately 80% specific activity compared to the native one.[8b,12]
12. Gutte, B.; Merrifield, R. B. *J. Am. Chem. Soc.* **1969**, *91*, 501–502.
13. An improved protocol namely, 'in situ neutralization Boc-chemistry SPPS' was introduced by Kent et al. Many of the side reactions of conventional SPPS are addressed by this technique and the protocol has been successful in assembling medium sized polypeptides. For details see: Schnolzer, M.; Alewood, P.; Jones, A.; Alawood, D.; Kent, S. B. H. *Int. J. Pept. Protein Res.* **1992**, *40*, 180–193.
14. Wilken, J.; Kent, S. B. H. *Curr. Opin. Biotechnol.* **1998**, *9*, 412–426.
15. Kimmerlin, T.; Seebach, D. *J. Pept. Res.* **2005**, *65*, 229–260.
16. Dirksen, A.; Hackeng, T. M.; Dawson, P. E. *Angew. Chem., Int. Ed.* **2006**, *118*, 7743–7746.
17. (a) Brunel, F. M.; Dawson, P. E. *Chem. Commun.* **2005**, 2552–2554; (b) Dawson, P. E.; Kent, S. B. H. *J. Am. Chem. Soc.* **1993**, *115*, 7263–7266; (c) Geoghegen, K. F.; Stroh, J. G. *Bioconjugate Chem.* **1992**, *3*, 138–146; (d) Robey, F. A.; Fields, R. L. *Anal. Biochem.* **1989**, *177*, 373–377; (e) Neuner, P.; Gallo, P.; Orsatti, L.; Fontana, L.; Monaci, P. *Bioconjugate Chem.* **2003**, *14*, 276–281; (f) Liu, C.-F.; Rao, C.; Tam, J. P. *Tetrahedron Lett.* **1996**, *37*, 933–936.
18. (a) Schnolzer, M.; Kent, S. B. H. *Science* **1992**, *256*, 221–225.
19. (a) Kemp, D. S.; Kerkman, D. J.; Leung, S.-L.; Hanson, G. *J. Org. Chem.* **1981**, *46*, 490–498; (b) Kemp, D. S.; Galakatos, N. G.; Dranginis, S.; Ashton, C.; Fotouhi, N.; Curran, T. P. *J. Org. Chem.* **1986**, *51*, 3320–3324.
20. (a) Kemp, D. S.; Carey, R. I. *J. Org. Chem.* **1993**, *58*, 2216–2222; (b) Kemp, D. S.; Galakatos, N. G.; Bowen, B.; Tan, K. *J. Org. Chem.* **1986**, *51*, 1829–1838.
21. (a) Miao, Z.; Tam, J. P. *J. Am. Chem. Soc.* **2000**, *122*, 4253–4260; (b) Liu, C.-F.; Tam, J. P. *J. Am. Chem. Soc.* **1994**, *116*, 4149–4153; (c) Liu, C.-F.; Tam, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 6584–6588.
22. The peptide fragment bearing C-terminal aldehyde functionality can be synthesized as follows: initially an *N*-protected amino acid is coupled to bromoacetaldehyde dimethyl acetal (2-bromo-1,1-dimethoxyethane) followed by removal of *N*-protector and coupling of the free amino termini with a peptide. Through acid treatment, the aldehyde group is regenerated before executing the ligation.[21a]

23. (a) Liu, C.-F.; Rao, C.; Tam, J. P. *J. Am. Chem. Soc.* **1996**, *118*, 307–312; (b) Tam, J. P.; Miao, Z. *J. Am. Chem. Soc.* **1999**, *121*, 9013–9022.
24. Tam, J. P.; Xu, J.; Eom, K. D. *Biopolymers (Pept. Sci.)* **2001**, *60*, 194–205.
25. Tam, J. P.; Yu, Q.; Miao, Z. *Biopolymers (Pept. Sci.)* **1999**, *51*, 311–332.
26. (a) Dawson, P. E.; Kent, S. B. H. *Annu. Rev. Biochem.* **2000**, *69*, 923–960; (b) Hang, H. C.; Bertozzi, C. R. *Acc. Chem. Res.* **2001**, *34*, 727–736; (c) Liu, L.; Bennett, C. S.; Wong, C.-H. *Chem. Commun.* **2006**, 21–33; (d) Dawson, P. E.; Chrchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. *J. Am. Chem. Soc.* **1997**, *119*, 4325–4339; (e) Aimoto, S. *Biopolymers (Pept. Sci.)* **1999**, *51*, 247–265.
27. (a) Nepomniaschiy, N.; Grimminer, V.; Cohen, A.; DiGiovanni, S.; Lashuel, H. A.; Brik, A. *Org. Lett.* **2008**, *10*, 5243–5246; (b) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2001**, *3*, 9–12.
28. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776–779.
29. Wan, Q.; Chen, J.; Yuan, Y.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2008**, *130*, 15814–15816.
30. Wu, X.; Stockdill, J. L.; Wang, P.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2010**, *132*, 4098–4100.
31. (a) Kohn, M.; Breinbauer, R. *Angew. Chem., Int. Ed.* **2004**, *43*, 3106–3116; (b) Kalia, J.; Abbott, N. L.; Raines, R. T. *Bioconjugate Chem.* **2007**, *18*, 1064–1069.
32. Leleu, S.; Penhoat, M.; Bouet, A.; Dupas, G.; Papamicael, C.; Marsais, F.; Levacher, V. *J. Am. Chem. Soc.* **2005**, *127*, 15668–15669.
33. de Lisle-Milton, R.; Milton, S. C. F.; Schnolzer, M.; Kent, S. B. H. In *Techniques in Protein Chemistry IV*; Angeletti, R. H., Ed.; Academic: New York, NY, 1993; pp 257–267.
34. (a) Nilsson, B. L.; Soellner, M. B.; Raines, R. T. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91–118; (b) Yeo, D. S. Y.; Srinivasan, R.; Chen, G. Y. J.; Yao, S. Q. *Chem.—Eur. J.* **2004**, *10*, 4664–4672.
35. (a) Kent, S. B. H. *Chem. Soc. Rev.* **2009**, *38*, 338–351; (b) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10068–10073; (c) Macmilln, D. *Angew. Chem., Int. Ed.* **2006**, *45*, 7668–7672.
36. (a) Flavell, R. R.; Muir, T. W. *Acc. Chem. Res.* **2009**, *42*, 107–116; (b) Muir, T. W. *Synlett* **2001**, 733–740.
37. (a) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 19–24; (b) Liu, L.; Hong, Z.-Y.; Wong, C.-H. *ChemBioChem* **2006**, *7*, 429–432.
38. Sohma, Y.; Yoshiya, T.; Taniguchi, A.; Kimura, T.; Hayashi, Y.; Kiso, Y. *Biopolymers (Pept. Sci.)* **2007**, *88*, 253–262.
39. (a) Muir, T. W. *Biopolymers (Pept. Sci.)* **2008**, *90*, 743–750; (b) David, R.; Richter, M. P. O.; Beck-Sickinger, A. G. *Eur. J. Biochem.* **2004**, *271*, 663–677; (c) Wallace, C. J. *Curr. Opin. Biotechnol.* **1995**, *6*, 403–410; (d) Coltart, D. M. *Tetrahedron* **2000**, *56*, 3449–3491.
40. Wieland, T.; Bokelmann, E.; Bauer, L.; Lang, H. U.; Lau, H. *Liebigs Ann. Chem.* **1953**, *583*, 129–149.
41. Johnson, E. C. B.; Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, *128*, 6640–6646.
42. Kang, J.; Macmillan, D. *Org. Biomol. Chem.* **2010**, *8*, 1993–2002.
43. Lu, W.; Qasim, M. A.; Kent, S. B. H. *J. Am. Chem. Soc.* **1996**, *118*, 8518–8523.
44. Hackeng, T. M.; Mounier, C. M.; Bon, C.; Dawson, P. E.; Griffin, J. H.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7845–7850.
45. Li, X.; Leeuw, E. d.; Lu, W. *Biochemistry* **2005**, *44*, 14688–14694.
46. Paramonov, S. E.; Gauba, V.; Hartgerink, J. D. *Macromolecules* **2005**, *38*, 7555–7561.
47. Dirksen, A.; Meijer, E. W.; Adriaens, W.; Hackeng, T. M. *Chem. Commun.* **2006**, 1667–1669.
48. van Baal, I.; Malda, H.; Synowsky, S. A.; van Dongen, J. L. J.; Hackeng, T. M.; Merkx, M.; Meijer, E. W. *Angew. Chem., Int. Ed.* **2005**, *44*, 5052–5057.
49. Fujita, Y.; Moyle, P. M.; Hieu, S.; Simerska, P.; Toth, I. *Biopolymers (Pept. Sci.)* **2008**, *90*, 624–632.
50. (a) Dealwis, C.; Fernandez, E. J.; Thompson, D. A.; Simon, R. J.; Siani, M. A.; Lolis, E. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6941–6946; (b) Ueda, H.; Siani, M. A.; Gong, W.; Thompson, D. A.; Brown, G. G.; Wang, J. M. *J. Biol. Chem.* **1997**, *272*, 24966–24970; (c) Baca, M.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11638–11642; (d) Kochendoerfer, G. G.; Salom, D.; Lear, J. D.; Wilk-Orescan, R.; Kent, S. B. H.; DeGrado, W. F. *Biochemistry* **1999**, *38*, 11905–11913; (e) Shogren-Knaak, M. A.; Peterson, C. L. *Methods Enzymol.* **2003**, *357*, 62–76; (f) Deniz, A. A.; Laurence, T. A.; Beligere, G. S.; Dahan, M.; Martin, A. B.; Chemla, D. S.; Dawson, P. E.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5179–5184; (g) Wilken, J.; Kent, S. B. H.; Jackson, M. R. In *Proceedings of the European Peptide Symposium*; Mayflower Scientific Ltd; Edinburgh, Kingswinford, West Midlands, England, Sept. 8–13, 1996, pp 909–910. (h) Escher, S. E.; Kluver, E.; Adermann, K. *Lett. Pept. Sci.* **2001**, *8*, 349–357; (i) Ferre-D'Amare, A. R.; Pognonec, P.; Roeder, R. G.; Burley, S. K. *EMBO J.* **1994**, *13*, 180–189; (j) Futaki, S. .; Tatsuto, K.; Shiraishi, Y.; Sugiura, Y. *Biopolymers* **2004**, *76*, 98–109; (k) Kochendoerfer, G. G.; Kent, S. B. H. *Curr. Opin. Chem. Biol.* **1999**, *3*, 665–671; (l) von Bodman, S. B.; Schuler, M. A.; Jollie, D. R.; Sligar, S. G. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 9443–9447.
51. Pentelute, B. L.; Gates, Z. P.; Dashnau, J. L.; Vanderkooi, J. M.; Kent, S. B. H. *J. Am. Chem. Soc.* **2008**, *130*, 9702–9707.
52. Canne, L. E.; Botti, P.; Simon, R. J.; Chen, Y.; Dennis, E. A.; Kent, S. B. H. *J. Am. Chem. Soc.* **1999**, *121*, 8720–8727.
53. Spetzler, J. C.; Hoeg-Jensen, T. *Bioorg. Med. Chem.* **2007**, *15*, 4700–4704.
54. Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 11684–11689.
55. (a) Balambika, R.; Inui, T.; Sargsyan, H.; Arshava, B.; Cohen, L. S.; Ding, F. X.; Becker, J. M.; Naider, F. *Int. J. Pept. Res. Ther.* **2007**, *13*, 251–263; (b) Naider, F.; Estephan, R.; Englander, J.; Sureshbabu, V. V.; Arevalo, E.; Samples, K.; Becker, J. M. *Biopolymers (Pept. Sci.)* **2004**, *76*, 119–128.
56. (a) Kang, J.; Richardson, J. P.; Macmillan, D. *Chem. Commun.* **2009**, 407–409; (b) Clippingdale, A. B.; Barrow, C. J.; Wade, J. D. *J. Pept. Sci.* **2000**, *6*, 225–234; (c) Blanco-Canosa, J. B.; Dawson, P. E. *Angew. Chem., Int. Ed.* **2008**, *47*, 6851–6855; (d) Escher, S. E.; Kluver, E.; Adermann, K. *Lett. Pept. Sci.* **2002**, *8*, 349–357; (e) Gross, C. M.; Lelievre, D.; Woodward, C. K.; Barany, G. *J. Pept. Res.* **2005**, *65*, 395–410; (f) Johnson, E. C. B.; Kent, S. B. H. *Biopolymers (Pept. Sci.)* **2006**, *88*, 340–349.
57. Conventional Boc-SPPS has inherent disadvantages including strong acid treatment and hazardous operation conditions, which lead frequently to undesired side-reactions. For details see: Bernardi, A.; Manzoni, L. *Chemtracts* **2000**, *13*, 588–591.
58. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 766–779.
59. Haginoya, E.; Hojo, H.; Nakahara, Y.; Nabeshima, K.; Toole, B. P.; Watanabe, Y. *Biosci. Biotechnol. Biochem.* **2006**, *70*, 1338–1349.
60. Li, X.; Kawakami, T.; Aimoto, S. *Tetrahedron Lett.* **1998**, *39*, 8669–8672.
61. Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. *J. Am. Chem. Soc.* **1999**, *121*, 11369–11374.
62. Ollivier, N.; Behr, J.-B.; El-Mahdi, O.; Blanpain, A.; Melnyk, O. *Org. Lett.* **2005**, *7*, 2647–2650.
63. (a) Alsina, J.; Yokum, T. S.; Albericio, F.; Barany, G. *J. Org. Chem.* **1999**, *64*, 8761–8769; (b) Jensen, K. J.; Alsina, J.; Songster, M. F.; Vagner, J.; Albericio, F.; Barany, G. *J. Am. Chem. Soc.* **1998**, *120*, 5441–5452.
64. Brask, J.; Albericio, F.; Jensen, K. J. *Org. Lett.* **2003**, *5*, 2951–2953.
65. Bang, D.; Pentelute, B. L.; Gates, Z. P.; Kent, S. B. H. *Org. Lett.* **2006**, *8*, 1049–1052.
66. Sharma, I.; Crich, D. *J. Org. Chem.* **2011**, *75*, 6518–6524.
67. Perler, F. B.; Adam, E. *Curr. Opin. Biotechnol.* **2000**, 377–383.
68. Torbeev, V. Y.; Kent, S. B. H. *Angew. Chem., Int. Ed.* **2007**, *46*, 1667–1670.
69. Beligere, G. S.; Dawson, P. E. *Biopolymers (Pept. Sci.)* **1999**, *51*, 363–369.
70. Bang, D.; Chopra, N.; Kent, S. B. H. *J. Am. Chem. Soc.* **2004**, *126*, 1377–1383.
71. Bang, D.; Kent, S. B. H. *Angew. Chem., Int. Ed.* **2004**, *43*, 2534–2538.
72. Brik, A.; Keinan, E.; Dawson, P. E. *J. Org. Chem.* **2000**, *65*, 3829–3835.
73. Ueda, S.; Fujita, M.; Tamamura, H.; Fujii, N.; Otaka, A. *ChemBioChem* **2005**, *6*, 1983–1986.
74. Bang, D.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5014–5019.
75. Bang, D.; Pentelute, B. L.; Kent, S. B. H. *Angew. Chem., Int. Ed.* **2006**, *45*, 3985–3988.
76. Durek, T.; Torbeev, Y.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 4846–4851.
77. Low, D. W.; Hill, M. G.; Carrasco, M. R.; Kent, S. B. H.; Botti, P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6554–6559.
78. Villain, M.; Gaertner, H.; Botti, P. *Eur. J. Org. Chem.* **2003**, 3267–3272.
79. Briand, B.; Kotzur, N.; Hagen, V.; Beyermann, M. *Tetrahedron Lett.* **2008**, *49*, 85–87.
80. Botti, P.; Carrasco, M. R.; Kent, S. B. H. *Tetrahedron Lett.* **2001**, *42*, 1831–1833.
81. (a) Offer, J.; Boddy, C. N. C.; Dawson, P. E. *J. Am. Chem. Soc.* **2002**, *124*, 4642–4646; (b) Offer, J.; Dawson, P. E. *Org. Lett.* **2000**, *2*, 23–26.
82. Canne, L. E.; Bark, S. J.; Kent, S. B. H. *J. Am. Chem. Soc.* **1996**, *118*, 5891–5896.
83. Lu, Y.-A.; Tam, J. P. *Org. Lett.* **2005**, *7*, 5003–5006.
84. (a) Yang, Y.-Y.; Ficht, S.; Brik, A.; Wong, C.-W. *J. Am. Chem. Soc.* **2007**, *129*, 7690–7701; (b) Brik, A.; Yang, Y.-Y.; Ficht, S.; Wong, C.-H. *J. Am. Chem. Soc.* **2006**, *128*, 5626–5627.
85. (a) Kumar, K. S. A.; Harpaz, Z.; Haj-Yahya, M.; Brik, A. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3870–3874; (b) Lutsky, M.-Y.; Nepomniaschiy, N.; Brik, A. *Chem. Commun.* **2008**, 1229–1231.
86. Li, J.; Cui, H.-K.; Liu, L. *Tetrahedron Lett.* **2010**, *51*, 1793–1796.
87. Marinzi, C.; Offer, J.; Longhi, R.; Dawson, P. E. *Bioorg. Med. Chem.* **2004**, *12*, 2749–2757.
88. Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc.* **2001**, *123*, 526–533.
89. Haase, C.; Rohde, H.; Seitz, O. *Angew. Chem., Int. Ed.* **2008**, *47*, 6807–6810.
90. Crich, D.; Banerjee, A. *J. Am. Chem. Soc.* **2007**, *129*, 10064–10065.
91. (a) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F. *J. Am. Chem. Soc.* **2009**, *131*, 13592–13593; For related reports regarding the use of δ-mercaptolysine in the synthesis of di-, tri- and tetraubiquitins and ubiquitinylated peptides see: (b) Kumar, K. S. A.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H. A.; Brik, A. *Angew. Chem., Int. Ed.* **2009**, *48*, 8090–8094; (c) Kumar, K. S. A.; Bavikar, S. N.; Spasser, L.; Moyal, T.; Ohayon, S.; Brik, A. *Angew. Chem., Int. Ed.* **2011**, *50*, 6137–6141.
92. Siman, P.; Karthikeyan, S. V.; Brik, A. *Org. Lett.* **2012**, *14*, 1520–1523.
93. (a) Wan, Q.; Danishefsky, S. J. *Angew. Chem., Int. Ed.* **2007**, *46*, 9248–9252; (b) Haase, C.; Seitz, O. *Angew. Chem., Int. Ed.* **2008**, *47*, 1553–1556; (c) Shang, S.; Tan, Z.; Danishefsky, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 5986–5989.
94. Tam, J. P.; Yu, Q. *Biopolymers* **1998**, *46*, 319–327.
95. (a) Hondal, R. J.; Nilsson, B. L.; Raines, R. T. *J. Am. Chem. Soc.* **2001**, *123*, 5140–5141; (b) Quaderer, R.; Sewing, A.; Hilvert, D. *Helv. Chim. Acta* **2001**, *84*, 1197–1206; (c) Gieselman, M. D.; Xie, L.; van der Donk, W. A. *Org. Lett.* **2001**, *3*, 1331–1334.
96. Roelfes, G.; Hilvert, D. *Angew. Chem., Int. Ed.* **2003**, *42*, 2275–2277.
97. Li, X.; Lam, H. Y.; Zhang, Y.; Chan, C. K. *Org. Lett.* **2010**, *12*, 1724–1727.
98. Alves, D. A.; Esser, D.; Broadbridge, R. J.; Beevers, A. P. G.; Chapman, C. P.; Winsor, C. E.; Betley, J. R. *J. Pept. Sci.* **2003**, *9*, 221–228.
99. (a) Seitz, O. *ChemBioChem* **2000**, *1*, 214–246; (b) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720; (c) Richardson, J. P.; Chan, C.-H.; Blanc, J.; Saadi, M.; Macmillan, D. *Org. Biomol. Chem.* **2010**, *8*, 1351–1360.
100. (a) Davis, B. G. *Chem. Rev.* **2002**, *102*, 579–601; (b) Gamblin, D. P.; Scanlan, E. M.; Davis, B. G. *Chem. Rev.* **2009**, *109*, 131–163; (c) Chalker, J. M.; Bernardes, G. J. L.;

Lin, Y. A.; Davis, B. J. *Chem. Asian J.* **2009**, *4*, 630–640; (d) Liu, L.; Bennett, C. S.; Wong, C.-H. *Chem. Commun.* **2006**, 21–33; (e) Peri, F.; Nicotra, F. *Chem. Commun.* **2004**, 623–627.

101. (a) Piontek, C.; Ring, P.; Harjes, O.; Heinlein, C.; Mezzato, S.; Lombana, N.; Pohner, C.; Puttner, M.; Silva, D. V.; Martin, A.; Schmid, F. X.; Unverzagt, C. *Angew. Chem., Int. Ed.* **2009**, *48*, 1936–1940; (b) Piontek, C.; Silva, D. V.; Heinlein, C.; Pohner, C.; Mezzato, S.; Ring, P.; Martin, A.; Schmid, F. X.; Unverzagt, C. *Angew. Chem., Int. Ed.* **2009**, *48*, 1941–1945.
102. (a) Miller, J. S.; Dudkin, V. Y.; Lyon, G. J.; Muir, T. W.; Danishfesky, S. J. *Angew. Chem., Int. Ed.* **2003**, *42*, 431–434; (b) Likhosherstov, L. M.; Novikova, O. S.; Derevitskaja, V. A.; Kochetkov, N. K. *Carbohydr. Res.* **1986**, *146*, C1–C5.
103. Ingale, S.; Buskas, T.; Boons, G.-J. *Org. Lett.* **2006**, *8*, 5785–5788.
104. Warren, J. D.; Miller, J. S.; Keding, S. J.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 6576–6578.
105. (a) Lee, D. J.; Mandal, K.; Harris, P. W. R.; Brimble, M. A.; Kent, S. B. H. *Org. Lett.* **2009**, *11*, 5270–5273; (b) Xiao, J.; Tolbert, T. J. *Org. Lett.* **2009**, *11*, 4144–4147.
106. Otaka, A.; Ueda, S.; Tomita, K.; Yano, Y.; Tamamura, H.; Matsuzaki, K.; Fujii, N. *Chem. Commun.* **2004**, 1722–1723.
107. (a) Hu, B.-H.; Su, J.; Messersmith, P. B. *Biomacromolecules* **2009**, *10*, 2194–2200; (b) Helms, B.; van Baal, I.; Merkx, M.; Meijer, E. W. *ChemBioChem* **2007**, *8*, 1790–1794.
108. (a) Dose, C.; Seitz, O. *Org. Lett.* **2005**, *7*, 4365–4368; (b) Dolphin, G. T. *Chem.—Eur. J.* **2006**, *12*, 1436–1447; (c) Tulla-Puche, J.; Barany, G. *J. Org. Chem.* **2004**, *69*, 4101–4107; (d) Hejjaoui, M.; Haj-Yahya, M.; Kumar, K. S. A.; Brik, A.; Lashuel, H. A. *Angew. Chem., Int. Ed.* **2011**, *50*, 405–409; (e) Kumar, K. S. A.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H. A.; Brik, A. *Angew. Chem., Int. Ed.* **2009**, *48*, 8090–8094; (f) Kumar, K. S. A.; Spasser, L.; Ohayon, S.; Erlich, L. A.; Brik, A. *Bioconjugate Chem.* **2011**, *22*, 137–143; (g) Lu, W.; Starovasnik, M. A.; Kent, S. B. H. *FEBS Lett.* **1998**, *429*, 31–35.
109. (a) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249–289; (b) Hofmann, R. M.; Muir, T. W. *Curr. Opin. Biotechnol.* **2002**, *13*, 297–303.
110. Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705–6710.
111. Shi, J.; Muir, T. W. *J. Am. Chem. Soc.* **2005**, *127*, 6198–6206.
112. Wood, D. W.; Wu, W.; Belfort, G.; Derbyshire, V.; Belfort, M. *Nat. Biotechnol.* **1999**, *17*, 889–892.
113. Muralidharan, V.; Muir, T. W. *Nat. Methods* **2006**, *3*, 429–438.
114. Lovrinovic, M.; Spengler, M.; Deutsch, C.; Niemeyer, C. M. *Mol. Biosyst.* **2005**, *1*, 64–69.
115. (a) Lykke-Anderson, J.; Christiansen, J. *Nucleic Acids Res.* **1998**, *26*, 5630–5635; (b) Mukhopadhyay, J.; Kapanidis, A. N.; Mekler, V.; Kortkhonjia, E.; Ebright, Y. W.; Ebright, R. H. *Cell* **2001**, *106*, 453–463; (c) Xu, R.; Ayers, B.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 388–393.
116. Wang, D.; Cole, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 8883–8886.
117. Valiyaveetil, F. I.; MacKinnon, R.; Muir, T. W. *J. Am. Chem. Soc.* **2002**, *124*, 9113–9120.
118. Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007–2010.
119. (a) Soellner, M. B.; Tam, A.; Raines, R. T. *J. Org. Chem.* **2006**, *71*, 9824–9830; (b) Soellner, M. B.; Nilsson, B. L.; Raines, R. T. *J. Org. Chem.* **2002**, *67*, 4993–4996.
120. (a) Soellner, M. B.; Nilsson, B. L.; Raines, R. T. *J. Am. Chem. Soc.* **2006**, *128*, 8820–8828; (b) Tam, A.; Soellner, M. B.; Raines, R. T. *Org. Biomol. Chem.* **2008**, *6*, 1173–1175; (c) Watzke, A.; Kohn, M.; Gutierrez-Rodriguez, M.; Wacker, R.; Schroder, H.; Breinbauer, R.; Kuhlmann, J.; Alexandrov, K.; Niemeyer, C. M.; Goody, R. S.; Waldmann, H. *Angew. Chem., Int. Ed.* **2006**, *45*, 1408–1412.
121. Nilsson, B. L.; Hondal, R. J.; Soellner, M. B.; Raines, R. T. *J. Am. Chem. Soc.* **2003**, *125*, 5268–5269.
122. Kim, H.; Cho, J. K.; Aimoto, S.; Lee, Y.-S. *Org. Lett.* **2006**, *8*, 1149–1151.
123. Tam, A.; Soellner, M. B.; Raines, R. T. *J. Am. Chem. Soc.* **2007**, *129*, 11421–11430.
124. Kleineweischede, R.; Hackenberger, C. P. R. *Angew. Chem., Int. Ed.* **2008**, *47*, 5984–5988.
125. David, O.; Meester, W. J. N.; Bieraugel, H.; Schoemaker, H. E.; Hiemstra, H.; van Marseveen, J. H. *Angew. Chem., Int. Ed.* **2003**, *42*, 4373–4375.
126. Sheppard, R. *J. Pept. Sci.* **2003**, *9*, 545–552.
127. (a) Carpino, L. A.; Krause, E.; Sferdean, C. D.; Schumann, M.; Fabian, H.; Bienert, M.; Beyermann, M. *Tetrahedron Lett.* **2004**, *45*, 7519–7523; (b) Meutermans, W. F.; Golding, S. W.; Bourne, G. T.; Miranda, L. P.; Dooley, M. J.; Alewood, P. F.; Smythe, M. L. *J. Am. Chem. Soc.* **1999**, *121*, 9790–9796.
128. Wohr, T.; Wahl, F.; Nefzi, A.; Rohwedder, B.; Sato, T.; Sun, X.; Mutter, M. *J. Am. Chem. Soc.* **1996**, *118*, 9218–9227.
129. Johnson, T.; Quibell, M.; Owen, D.; Sheppard, R. C. *J. Chem. Soc., Chem. Commun.* **1993**, 369–372.
130. Brenner, M.; Zimmermann, J. P.; Wehrmuller, J.; Quitt, P.; Hartmann, A.; Schneider, W.; Beglinger, U. *Helv. Chim. Acta* **1957**, *40*, 1497–1517.
131. Miranda, L. P.; Meutermans, W. D. F.; Smythe, M. L.; Alewood, P. F. *J. Org. Chem.* **2000**, *65*, 5460–5468.
132. (a) Skwarczynski, M.; Sohma, Y.; Noguchi, M.; Kimura, M.; Hayashi, Y.; Hamada, Y.; Kimura, T.; Kiso, Y. *J. Med. Chem.* **2005**, *48*, 2655–2666; (b) Hayashi, Y.; Skwarczynski, M.; Hamada, Y.; Sohma, Y.; Kimura, T.; Kiso, Y. *J. Med. Chem.* **2003**, *46*, 3782–3784.
133. Skwarczynski, M.; Sohma, Y.; Kimura, M.; Hayashi, Y.; Kimura, T.; Kiso, Y. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4441–4444.
134. Tamamura, H.; Hori, T.; Otaka, A.; Fujii, N. *J. Chem. Soc., Perkin Trans. 1* **2002**, 577–580.
135. (a) Khatib, M.; Elagawany, M.; Jabeen, F.; Todadze, E.; Bol'shakov, O.; Oliferenko, A.; Khelashvili, L.; El-Feky, S. A.; Asiri, A.; Kartritzky, A. R. *Org. Biomol. Chem.* **2012**, *10*, 4836–4838; (b) Kartritzky, A. R.; Abo-Dya, N. E.; Tala, S. R.; Abdel-Samii, Z. K. *Org. Biomol. Chem.* **2010**, *8*, 2316–2319.
136. (a) Sohma, Y.; Kiso, Y. *ChemBioChem* **2006**, *7*, 1549–1557; (b) Skwarczynski, M.; Sohma, Y.; Noguchi, M.; Kimura, T.; Hayashi, Y.; Kiso, Y. *J. Org. Chem.* **2006**, *71*, 2542–2545.
137. (a) Mutter, M.; Chandravarkar, A.; Boyat, C.; Lopez, J.; Santos, S. D.; Mandal, B.; Mimna, R.; Murat, K.; Patiny, L.; Saucede, L.; Tuchscherer, G. *Angew. Chem., Int. Ed.* **2004**, *43*, 4172–4178; (b) Hentschel, J.; Krause, E.; Borner, H. G. *J. Am. Chem. Soc.* **2006**, *128*, 7722–7723.
138. (a) Sohma, Y.; Hayashi, Y.; Skwarczynski, M.; Hamada, Y.; Sasaki, M.; Kimura, T.; Kiso, Y. *Biopolymers (Pept. Sci.)* **2004**, *76*, 344–356; (b) Wang, H.; Kakizawa, T.; Taniguchi, A.; Mizuguchi, T.; Kimura, T.; Kiso, Y. *Bioorg. Med. Chem.* **2009**, *17*, 4881–4887.
139. Sohma, Y.; Sasaki, M.; Hayashi, Y.; Kimura, T.; Kiso, Y. *Chem. Commun.* **2004**, 124–125.
140. Coin, I.; Dolling, R.; Krause, E.; Bienert, M.; Beyermann, M.; Sferdean, C. D.; Carpino, L. A. *J. Org. Chem.* **2006**, *71*, 6171–6177.
141. Sohma, Y.; Taniguchi, A.; Skwarczynski, M.; Yoshiya, T.; Fukao, F.; Kimura, T.; Hayashi, Y.; Kiso, Y. *Tetrahedron Lett.* **2006**, *47*, 3013–3017.
142. (a) Taniguchi, A.; Yoshiya, T.; Abe, N.; Fukao, F.; Sohma, Y.; Kimura, T.; Hayashi, Y.; Kiso, Y. *J. Pept. Sci.* **2007**, *13*, 868–874; (b) Sohma, Y.; Taniguchi, A.; Yoshiya, T.; Chiyomori, Y.; Fukao, F.; Nakamura, S.; Skwarczynski, M.; Okada, T.; Ikeda, K.; Hayashi, Y.; Kimura, T.; Hirota, S.; Matsuzaki, K.; Kiso, Y. *J. Pept. Sci.* **2006**, *12*, 823–828.
143. Yoshiya, T.; Taniguchi, A.; Sohma, Y.; Fukao, F.; Nakamura, S.; Abe, N.; Ito, N.; Skwarczynski, M.; Kimora, T.; Hayashi, Y.; Kiso, Y. *Org. Biomol. Chem.* **2007**, *5*, 1720–1730.
144. Sohma, Y.; Chiyomori, Y.; Kimura, M.; Fukao, F.; Taniguchi, A.; Hayashi, Y.; Kimura, T.; Kiso, Y. *Bioorg. Med. Chem.* **2005**, *13*, 6167–6174.
145. Lalithamba, H. S.; Hemantha, H. P.; Sureshbabu, V. V. *Protein Pept. Lett.* **2011**, *18*, 848–857.
146. Taniguchi, A.; Sohma, Y.; Kimura, M.; Okada, T.; Ikeda, K.; Hayashi, Y.; Kimura, T.; Hirota, S.; Matsuzaki, K.; Kiso, Y. *J. Am. Chem. Soc.* **2006**, *128*, 696–697.
147. (a) Noguchi, M.; Skwarczynski, M.; Prakash, H.; Hirota, S.; Kimura, T.; Hayashi, Y.; Kiso, Y. *Bioorg. Med. Chem.* **2008**, *16*, 5389–5397; (b) Taniguchi, A.; Skwarczynski, M.; Sohma, Y.; Okada, T.; Ikeda, K.; Prakash, H.; Mukai, H.; Hayashi, Y.; Kimura, T.; Hirota, S.; Matsuzaki, K.; Kiso, Y. *ChemBioChem* **2008**, *9*, 3055–3065.
148. Santos, S. D.; Chandravarkar, A.; Mandal, B.; Mimna, R.; Murat, K.; Saucede, L.; Tella, P.; Tuchscherer, G.; Mutter, M. *J. Am. Chem. Soc.* **2005**, *127*, 11888–11889.
149. Yoshiya, T.; Ito, N.; Kimura, T.; Kiso, Y. *J. Pept. Sci.* **2008**, *14*, 1203–1208.
150. (a) Pattabhiraman, V. R.; Ogunkoya, A. O.; Bode, J. W. *Angew. Chem., Int. Ed.* **2012**, *51*, 5114–5118; (b) Bode, J. W.; Fox, R. M.; Baucom, K. D. *Angew. Chem., Int. Ed.* **2006**, *45*, 1248–1252.
151. Wang, P.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2010**, *132*, 17045–17051.
152. Rao, Y.; Li, X.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2009**, *131*, 12924–12926.
153. (a) Brailsford, J. A.; Danishefsky, S. K. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 7196–7201; (b) For commentary on the synthesis of EPO see: Brown, Z. Z.; Muir, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 7134–7135; For earlier reports on the chemical synthesis toward homogeneous erythropoietin fragments employing NCL and non-NCL protocols see: (c) Tan, Z.; Shang, S.; Halkina, T.; Yuan, Y.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2009**, *131*, 5424–5431; (d) Yuan, Y.; Chen, J.; Wan, Q.; Tan, Z.; Chen, G.; Kan, C.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2009**, *131*, 5432–5437; (e) Kan, C.; Trzupek, J. D.; Wu, B.; Wan, Q.; Chen, G.; Tan, Z.; Yuan, Y.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2009**, *131*, 5438–5443.
154. Liu, S.; Pentelute, B. L.; Kent, S. B. H. *Angew. Chem., Int. Ed.* **2012**, *51*, 993–999.
155. Siman, P.; Brik, A. *Org. Biomol. Chem.* **2012**, *10*, 5684–5697.
156. Aussedat, B.; Fasching, B.; Jphnston, E.; Sane, N.; Nagorny, P.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2012**, *134*, 3532–3541.
157. 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. *J. Am. Chem. Soc.* **2012**, *134*, 5428–5431.
158. Bavikar, S. N.; Spasser, L.; Haj-Yahya, M.; Karthikeyan, S. V.; Moyal, T.; Kumar, K. S. A.; Brik, M. *Angew. Chem., Int. Ed.* **2012**, *51*, 758–763.

Biographical sketch

Hosahalli P. Hemantha was born in Shimoga, Karnataka, India in 1981. He completed M.Sc. in Industrial Chemistry from Kuvempu University, Shimoga in 2004. Presently he is pursuing research in peptides and peptidomimetics under the guidance of Dr. V.V. Sureshbabu at the Department of Chemsitry, Central College, Bangalore.

N. Narendra was born in Bangalore, Karnataka, India in 1981. He completed M.Sc. in Organic Chemistry from Bangalore University, Bangalore in 2005. He is working on peptides and peptidomimetics under the guidance of Dr. V.V. Sureshbabu at the Department of Chemsitry, Central College, Bangalore for his Ph.D.

Prof. Vommina V. Sureshbabu was born in Nellore, Andra Pradesh, India in 1961. He obtained M.Sc. in Chemistry from Sri Krishnadevaraya University, Ananthapur, India in 1983. He was invited by Prof. K.M. Sivanandaiah to join for Ph.D at Central College, Bangalore to work in the area of peptide chemistry. After the completion of Ph.D in 1989, he was appointed as Lecturer at the same department. Later, he went to USA for a post-doctoral assignment to CUNY, New York where he worked on the synthesis of GPCR fragments through native chemical ligation. At present, he is working as a Professor at the Department of Chemistry, Central College, Bangalore. His research interests include development of new reagents for efficient peptide synthesis, design and synthesis of peptidomimetics, incorporation of unnatural linkages into peptide backbone, native chemical ligation, C-terminal versus N-terminal modifications for peptidomimetic synthesis and utility of Fmoc-group in solution-phase synthesis. Recently, V.V.S. has contributed a chapter entitled 'Protection Reactions' to the volume 'Amino acids, Peptides and Proteins in Organic Chemistry' Vol. 4. *Protection Reactions, Combinatorial and Medicinal Chemistry*: (Ed. Andrew Hughes), 2011, Wiley Int. Ltd.