Peptide and protein thioester synthesis *via* N→S acyl transfer†‡

Jaskiranjit Kang and Derek Macmillan*

Received 30th November 2009, Accepted 29th January 2010
First published as an Advance Article on the web 18th February 2010
DOI: 10.1039/b925075a

Peptide and protein thioesters are playing an increasingly prominent role in the chemical toolbox for protein assembly and modification through Native Chemical Ligation (NCL). In this Emerging Area we highlight recent developments in a somewhat surprising route to thioesters: selective disruption of amides, the more stable carboxylic acid derivatives.

Introduction

The peptide and protein thioester is an important tool in chemistry and chemical biology, particularly for the synthesis and modification of peptides and proteins through native chemical ligation. The ability to alter the structure and function of proteins at will has enabled detailed study of protein mechanism and the effect of post translational modifications. The driving force for the NCL reaction is the formation of the amide linkage, which occurs around neutral pH (Scheme 1a).[1–8] Despite the stability of amides several synthetic approaches are emerging which target them for cleavage with concomitant thioester formation in a manner more reminiscent of protein self-splicing, facilitated by distortion of the amide geometry (ground state destabilisation) and participation of a neighbouring β-disposed thiol. In this Emerging Area we highlight new ventures in thioester synthesis employing such $N\rightarrow S$ acyl transfer approaches.

Department of Chemistry, University College London, 20 Gordon Street, London, UK WC1H 0AJ. E-mail: d.macmillan@ucl.ac.uk; Tel: +44 (0)20 7679 4684

† This paper is part of an *Organic & Biomolecular Chemistry* web theme issue on protein labelling and chemical proteomics. Guest editors: Ed Tate and Matthew Bogyo.

‡ The authors gratefully acknowledge funding support from EPSRC and The Royal Society.

1. The occurrence and activity of inteins

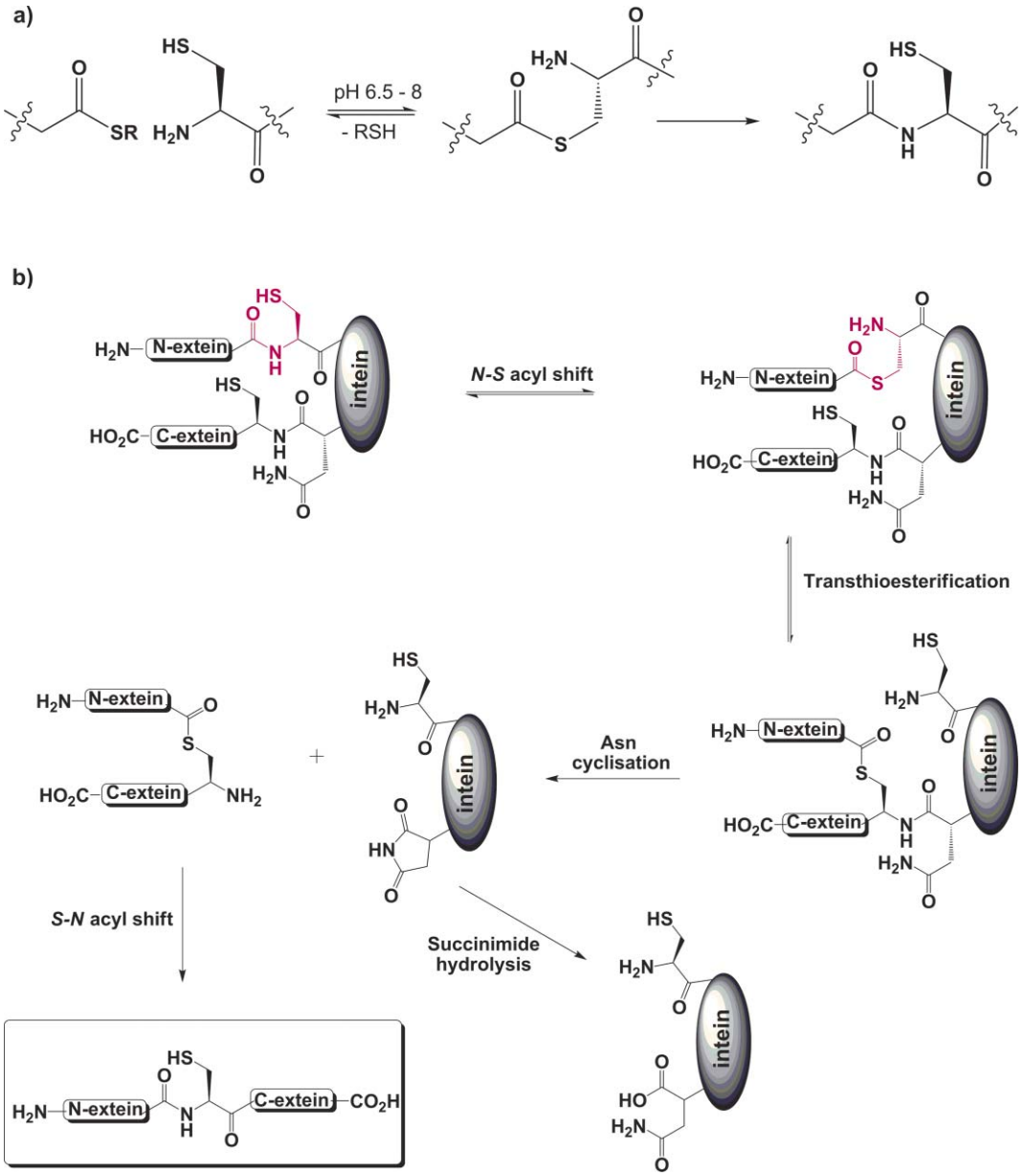
In 1990, a biological $N\rightarrow S$ acyl shift was identified in the mechanism of natural protein splicing elements termed inteins. An intein is named by analogy to an intron in nucleic acids and consists of an internal protein domain that is flanked by two polypeptides, commonly termed the N-extein and the C-extein. Inteins mediate protein splicing through a sequence of highly regulated acyl transfers beginning with a self-catalysed intramolecular rearrangement of the intein N-terminal cysteine residue. This forms a thioester at the N-intein splice junction, which further undergoes transthioesterification by nucleophilic attack of the first residue (cysteine) in the C-extein to form a branched intermediate. The intein is then excised by cyclisation of asparagine (or glutamate) at the C-terminus of the intein splice junction. The resulting thioester-linked exteins ligate by a NCL-like $S\rightarrow N$ acyl shift to form a stable amide bond (Scheme 1b).

Jaskiranjit Kang

Jaskiranjit Kang obtained her initial degree: MSci (First Class Honours) in Medicinal Chemistry from University College London (2003–2007). Since then she has been engaged in her Ph.D. studies in organic and biological chemistry based on the small molecule-mediated synthesis of peptide and protein thioesters through a site-specific intramolecular N→S acyl transfer. The main focus of her research has been the development, optimisation and biological application of an exciting new thioesterification reaction, which has enabled the total chemical synthesis of an antimicrobial and chemoattractant polypeptide. She is currently working towards her doctoral degree under the guidance of Dr Derek Macmillan.

Derek Macmillan

Derek Macmillan obtained his B.Sc. from the University of Edinburgh and subsequently was fortunate to pursue a Ph.D. in Bio-organic chemistry under the supervision of Professor Sabine Flitsch, also at Edinburgh. For postdoctoral studies (1999–2001) he joined the laboratory of Professor Carolyn Bertozzi at UC Berkeley. After returning to Edinburgh as research fellow he was awarded a Royal Society University Research Fellowship in 2003 and, following a short research visit to the laboratory of Professor Frances Arnold (CalTech), the group relocated to the Department of Chemistry at UCL in 2005. The Macmillan group explores the application of organic chemistry to understanding biological processes with particular emphasis on the semi-synthesis of modified proteins.



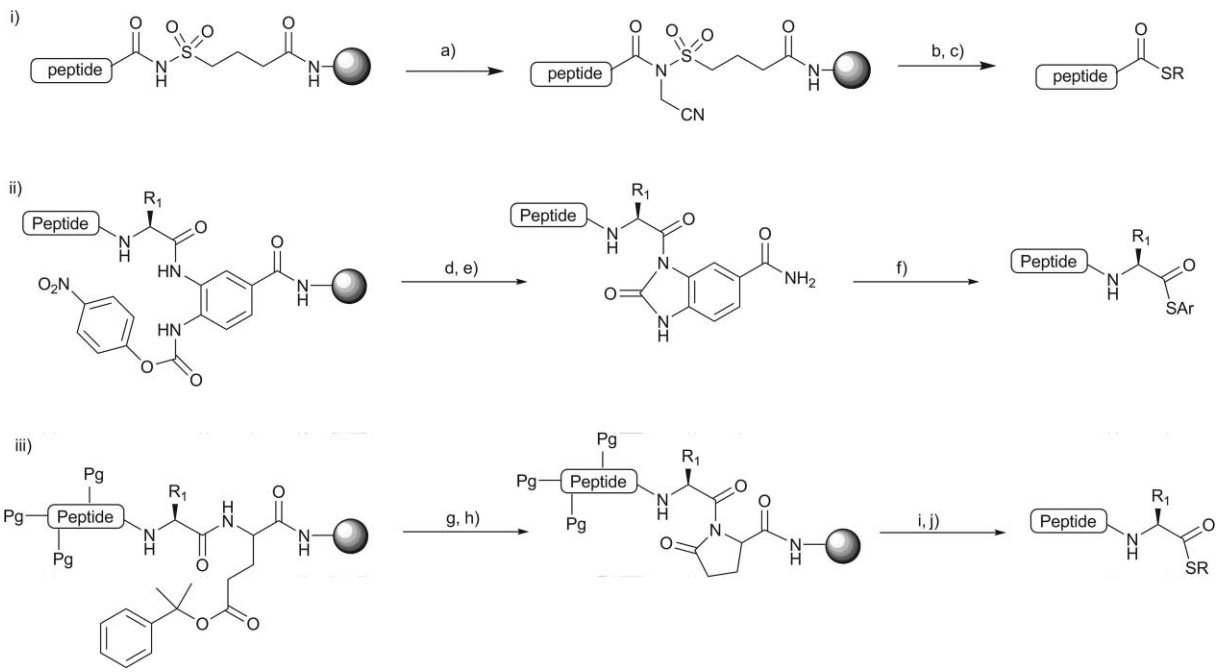
Scheme 1 (a) Native chemical ligation. (b) The mechanism of protein splicing.

Over 350 members of the intein family are known and mutant inteins have been developed that participate only in the initial or final stages of protein splicing and are frequently employed in protein labelling and semi-synthesis. This generally requires the expression of a protein of interest fused at the N- or C-terminus to a mutant intein that is often associated with an affinity tag, such as a chitin binding domain (CBD), for ease of purification. If the protein expressed corresponded to the N-terminal extein, then after the initial $N \rightarrow S$ acyl transfer, subsequent release of the protein from the intein can be achieved by an intermolecular thiol-thioester exchange driven by the addition of a thiol (RSH), such as sodium 2-mercaptoethanesulfonate (MESNa).[9] Alternatively, if the protein corresponded to the C-extein, it would require N-terminal cleavage of the intein, which can be promoted by changing the pH and temperature of the solution to generate a protein with an N-terminal cysteine. Therefore, these intein fusion systems provide a source of recombinant proteins that can be functionalised with either an N-terminal cysteine or a C-terminal thioester to be used as precursors in NCL[10] and are commercially available. In addition to the so-called *cis*-splicing inteins discussed above, protein *trans*-splicing (PTS) occurs through the use of split inteins.[11] These inteins are split (either naturally occurring or engineered) into their complementary "halves" and each pair of halves has a different affinity for subsequent reassociation and this selectivity can be exploited to conduct multiple simultaneous ligations as crossover reactions do not occur in PTS. Commercially available intein technology frequently combines with labelled synthetic peptide fragments, providing biotechnology tools that provide valuable insight into protein structure and function.[12–14] Despite the prominent role for inteins in protein engineering,

relatively little is understood about their own mechanistic details at the atomic level, presumably as a consequence of the practical difficulty in studying an active splicing precursor. Consequently most mechanistic studies have employed inactive mutants. Several mutagenesis, biophysical and biochemical studies have served to elucidate the overall sequence of steps and identify conserved proximal residues (particularly histidine residues) which are essential for the early and late stages of the splicing process. These results have suggested that histidine plays an important role as a general acid and base catalyst while the amide bond is believed to have a non-typical geometry.[15–18] Consequently it has been proposed that protein splicing comes about through a combination of intein-mediated amide distortion, generally termed ground state destabilisation (evidenced by an unusually low $^{1}J_{NC'}$ coupling constant of 12 Hz), and the catalytic features of proximal (mainly histidine) residues.[19]

2 Peptide thioester synthesis

A C-terminal peptide or protein thioester is essential in NCL and, until recently, synthetic thioester preparation has utilised conventional carboxyl activation chemistry for this purpose rather than the $N \rightarrow S$ route so elegantly orchestrated by inteins. Thioesters can be prepared by *tert*-butyloxycabonyl (Boc)-based solid-phase peptide synthesis (SPPS) on a thioester resin.[20] However, the harsh acidic conditions (usually HF) employed in Boc-based SPPS to cleave the peptide thioester from the solid support, despite being a high yielding method with demonstrated application to chemically fragile modifications such as *N*-glycopeptides,[21,22] has led to this chemistry being eclipsed by the more amenable 9-fluorenylmethoxycarbonyl (Fmoc)-based SPPS, which utilises a mild base, usually piperidine, during synthesis.[23] Consequently, several methods to afford peptide thioesters using Fmoc SPPS have been explored. Since thioesters are unstable to piperidine, other non-nucleophilic bases have been investigated for the removal of the N-α-Fmoc group during peptide elongation with reasonable success; these have included 1-methylpyrrolidine/hexamethyleneimine/HOBt[24] or DBU/HOBt mixtures.[25] Use of low nucleophilic bases used during Fmoc SPPS have however been shown to racemise chiral amino acids adjacent to the thioester. A popular route for peptide thioester synthesis uses a modified Kenner's sulfonamide safety-catch linker method and employs standard Fmoc/*t*-Bu SPPS followed by sulfonamide alkylation to aid release of the peptide thioester from solid support (Scheme 2).[23] The resulting peptide thioesters are often isolated in low yield[26] and in the case of post-translationally modified peptides such as glycopeptides the activating alkylation step can also result in the undesirable alkylation of unprotected hydroxyl groups on an attached carbohydrate as well as on methionine residues. Consequently synthetic strategies that release otherwise fully protected peptide acids or amides from solid support prior to conversion to thioesters have been established since the preparation of such species is, in general, higher yielding and easier to monitor (Scheme 2).[27–32] A backbone amide linker (BAL) with Fmoc-SPPS has also been used to prepare peptide thioesters with a masked thioester incorporated as a trithioortho ester.[33,34] Alternatively, peptide thioesters can be introduced after chain assembly to avoid exposure of the thioester to basic conditions during synthesis. This

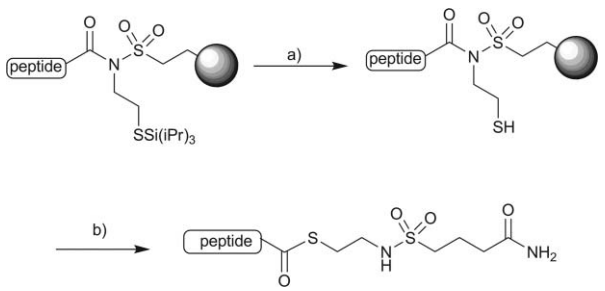


Scheme 2 i) The typical sulfonamide safety-catch approach to peptide thioester synthesis employing Fmoc-based SPPS.[23] *Reagents and conditions*: a) ICH_2CN, DIPEA, DMF, 16 h r.t. b) R-SH (often benzylmercaptan), NaSPh, DMF, 16 h, r.t. c) 95% v/v TFA + scavengers. More recent developments in Fmoc-based peptide thioester formation utilising ii) *N*-acylisoureas: *Reagents and conditions*: d) 0.5 M DIPEA in DMF, 15 min r.t. e) 95% v/v TFA + scavengers. f) 200 mM MPAA (note the thioester is generated *in situ*).[31] iii) Peptide pyroglutamyl imide formation.[32] *Reagents and conditions*: g) 2% v/v TFA + scavengers, DCM. h) PyBrOP, DIEA, NMP, microwave 60 °C, 3 × 1 h. i) RSH, PhSNa, [15]crown-5, MeCN, 40 °C overnight. j) 95% v/v TFA + scavengers. Pg = protecting group.

has been shown with BAL extended by orthogonal allyl protection of the C-terminal α-carboxylic group.[35]

2.1 $N \rightarrow S$ acyl transfer using the sulfonamide-linker

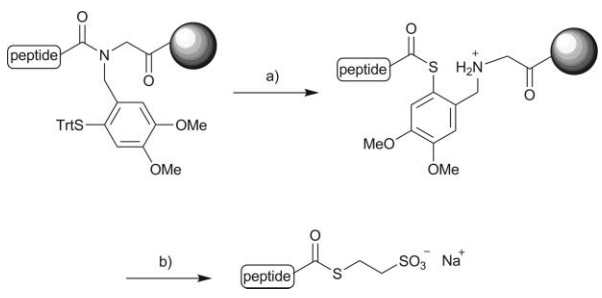
Melnyk and co-workers reported a solid-phase $N \rightarrow S$ acyl transfer for thioester synthesis after peptide chain assembly using Fmoc/*t*-Bu chemistry in combination with the sulfonamide safety-catch linker.[36] Already a widely used platform for thioester synthesis, the acylsulfonamide group of the linker can undergo additional Mitsunobu alkylation with a mercaptoethanol derivative, which allows the introduction of the prerequisite β-amino thiol component that facilitates the subsequent intramolecular $N \rightarrow S$ acyl shift (Scheme 3). The isolated yield of thioester was found to be dependent upon the degree of alkylation, which was independent of the steric influence from the first amino acid due to the remoteness of the alkylation site from Cα. This synthetic methodology also has the advantage of allowing native chemical ligation to potentially proceed on solid-support. Additionally, if 3-carboxypropane sulfonamide is employed to link the target peptide to Rink amide resin then the the $N \rightarrow S$ acyl shift can proceed with concomitant cleavage from the resin such that the thioester can be isolated. A disadvantage is that formation of the resin-bound acylsulfonamide can be inefficient and the subsequent Mitsunobu reaction can be relatively difficult to monitor, especially when using the commercially available 4-sulfamylbutyryl-linked resins.



Scheme 3 Loading Rink amide resin with 3-carboxypropane sulfonamide, prior to peptide assembly and thioesterification, allows the thioester to be recovered. *Reagents and conditions*: a) TBAF/AcOH. b) 95% v/v TFA + scavengers.

2.2 $N \rightarrow S$ shift in acyl-transfer auxiliaries

The natural abundance of cysteine residues in proteins is relatively low (<2%) and cysteine is therefore not always conveniently situated at desired ligation junctions. Strategies to overcome this limitation include the ligation between thioester peptides and N-terminal homocysteinyl or homoselenocysteinyl peptides, followed by methylation to afford Xaa-Met or Xaa-Seleno-Met sequences respectively.[37,38] Similarly, the utilisation of an N-terminal cysteinyl peptide for NCL followed by desulfurisation has been shown to afford an Xaa-Ala ligated junction[39] and analogous strategies have evolved for several additional amino acid residues.[40] The peptide ligation repertoire was further expanded by the use of cleavable auxiliaries that could be appended to potentially any N-terminal amino acid, though most commonly at Gly or Ala, facilitating cysteine-free NCL.[41–46] An important feature within the carbon skeleton of the auxiliary is that a NCL-like β, or γ amino-substituted thiol motif is retained to facilitate ligation *via* a five or pre-organised six-membered ring intermediate. However, in a strategy employed by Danishefsky *et al.*,[47] TFA treatment of an auxiliary-linked peptide (required to remove the auxiliary) after cysteine free NCL yielded a by-product that was assigned as a thioester, formed through intramolecular $N \rightarrow S$ acyl shift. This type of rearrangement was previously documented by Vorherr and Aimoto with the 2-mercapto-4,5-dimethoxybenzyl (Dmmb) auxiliary and has since been developed into a method for thioester synthesis (Scheme 4).[48] The existence of the $N \rightarrow S$ acyl transfer was subsequently verified using ^{13}C NMR spectroscopy, RP-HPLC, and mass spectrometry.[49] This was achieved through a two-part study of model systems: (i) a Dmmb containing dipeptide, Fmoc-Gly(^{13}C-1)-Ala(Dmmb)-OMe and (ii) a Cys containing pentapeptide, Fmoc-Ile-Ala-Gly($^{13}C1$)-Cys-Arg-NH_2, both of which were subjected to acid treatment. Interestingly while both peptides were shown to form thioesters by ^{13}C NMR, the glycinyl thioester formed in (ii) was shown to rapidly revert to starting material upon attempted isolation, whereas the Dmmb-linked thioester could be isolated by RP-HPLC. Dmmb auxiliary-assisted thioester synthesis, in solution or on solid phase, has recently been shown to produce peptide thioesters with up to 41 amino acid residues.[50] An Fmoc-Leu-D,L-[Dmmb(Trt)]-Ala motif was coupled to H-Phe-Alko-PEG resin and then extended by Fmoc SPPS before being treated with Reagent K to release the peptide from the resin. The Dmmb-attached peptide was then treated with dilute acid (0.25 M HCl, pH ~0.6) containing 0.5% TCEP at 37 °C for 3 h, followed by removal of the solvent under reduced pressure. This created the *S*-peptide, which upon treatment with MESNa in an aqueous NaOAc/acetonitrile mixture, at elevated pH (6-7), produced the peptide-$SCH_2CH_2SO_3H$ thioester *via* intermolecular thiol-thioester exchange, with *N*-D,L-(Dmmb)Ala-Phe-OH as the by-product. The isolated yield for a bovine pancreatic trypsin inhibitor thioester BPTI(1-29)-$SCH_2CH_2SO_3H$ was 82% based on the reacting *S*-peptide. When 85% aqueous TFA was used in place of 0.25 M HCl the peptide thioester yield was 44% and side products: [BPTI(1-29)-D,L-Ala-Phe-OH and BPTI(1-29)-D,L-(Dmmb)Ala-OH] were also identified by MALDI-TOF MS. The use of 0.25 M HCl suppressed such side reactions and deamination, which was also observed with 1.0 M HCl, and resulted in the conversion of Asn-24 to Asp-24 within the BPTI(1-29) sequence. The milder acid appears to selectively



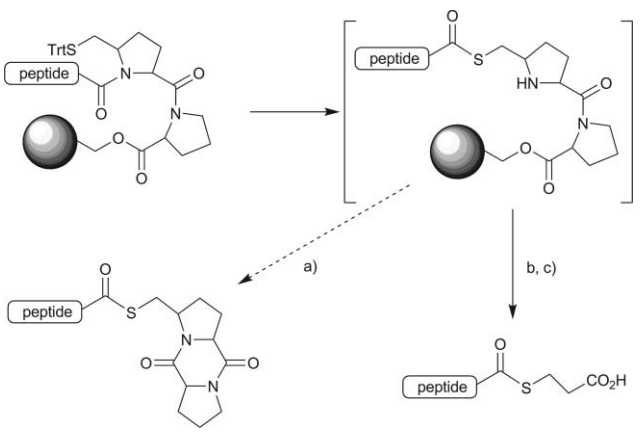
Scheme 4 Thioester synthesis using a solid supported Dmmb auxiliary. The solid-supported thioester is afterwards intercepted with 2-mercaptoethanesulfonic acid. *Reagents and conditions*: a) 88% TFA + scavengers, 1 h b) 0.1 M 2-mercaptoethanesulfonic acid, DMF.

protonate the Dmmb-linked amide which is again attributed to the conjugate acid of an *N*,*N*-dialkylamide having a higher pK_a value than primary amines. A similar yield was reported for the BPTI(1-29)-$SCH_2CH_2SO_3H$ thioester prepared with the $N \rightarrow S$ acyl shift occurring on-resin and no racemisation of the terminal Leu-29 residue was detected. On-resin thioesterification with the Dmmb auxiliary was further used to synthesise a thrombopoietin receptor (TpoR)(461-481)-$SCH_2CH_2SO_3H$ thioester with acid-sensitive linkages such as Trp and the Asp-Pro sequence, in 34% yield based on initial resin loading. A telomere repeat-binding factor 2 (TRF2)(1-41)-$SCH_2CH_2SO_3H$ thioester was prepared in 12% yield. Additional peptides prepared also included modifications such as [Ser$(PO_3H_2)^{10}$]-histone H3(1-12)-$SCH_2CH_2SO_3H$ (31%) and [Lys$(Me_3)^9$]-histone H3(1-33)-$SCH_2CH_2SO_3H$ (19%). Although some deamination of Gln to Glu could be observed in small amounts (<1.5%) for certain peptides, the desired peptide thioesters could be separated by RP-HPLC. These results are highly promising though a drawback is the several synthetic steps required to prepare the auxiliary-linked resin, and then load the first amino acid. Ideally preloaded resins capable of furnishing the amino acid thioester of choice would ultimately become commercially available.

2.3 MPA-mediated thioesterification

Hojo and co-workers described a post-peptide chain assembly thioesterification that was used to produce a 61 amino acid glycopeptide from its 25 amino acid thioester. The method involved a peptide bearing a C-terminal mercaptomethylated proline derivative (Scheme 5)[51] that was converted into a thioester upon treatment with aqueous 3-mercaptopropionic acid (MPA). Originally the peptide was prepared to comply with diketopiperazine (DKP) formation which is known to occur in peptides containing C-terminal Pro ester residues. The reaction occurs when a deprotected amine is able to react with an "activated" carboxylic acid to afford a bicyclic intermediate. However, the formation of the diketopiperazine tricyclic intermediate did not proceed as envisaged. This was attributed to the diketopiperazine formation on resin being slower than the reverse ($S \rightarrow N$) acyl transfer, in weakly acidic or basic solution. The peptide with the C-terminal Pro-Pro ester motif was however successfully thioesterified by intermolecular thioester exchange with excess MPA. At room temperature, the thioester of MPA could be observed within 12 h, using 40% v/v MPA (pH 1.3) without any serious side reactions. Extending the reaction time to 168 h resulted in complete consumption of the starting material. Hydrolysis was reported to be less than 15%, which nevertheless decreases the efficiency of the reaction. The reaction rate increased with microwave irradiation (150 W, 80 °C) and reactions were complete between 0.5–1 h, depending on the MPA concentration used. Decreased MPA concentrations (≤20%) were correlated with increased reaction times, although the thioester purification was more facile. Varying time and MPA concentration suggested that an impressive 70% yield of thioester could be attained in only 10 min, in 40% MPA. The original peptide containing the C-terminal Pro-Pro ester motif was experimentally concluded to be irrelevant because MPA transthioesterified all *S*-peptide formed through the initial $N \rightarrow S$ acyl transfer. The synthesis of an *N*-acetylglucosaminylated peptide thioester of emmprin(34-58) was achieved with 20% MPA and afforded yields that were over twice as high as those previously reported, showcasing the efficiency of this MPA-thioesterification method.

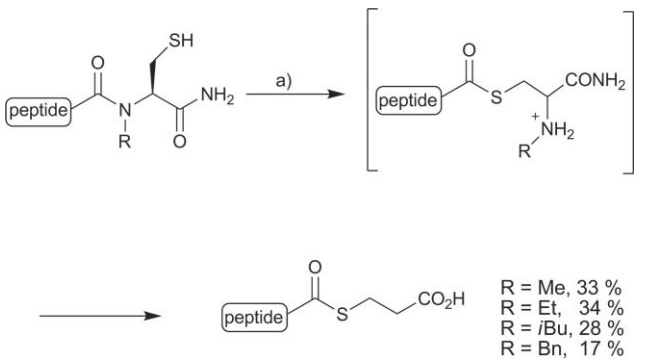
A stated drawback of this method was the lengthy preparation of the initial 5-mercaptomethyl proline to be incorporated at the C-terminus of the peptide. In addition, MPA thioesterification without microwave irradiation was considerably slower and was shown to increase the reaction timescale to a week. Furthermore an Asp-Ser peptide bond was found to be sensitive to treatment with 10–20% MPA though this could likely be solved if a suitable protecting group strategy was employed. The driving force for the initial intramolecular $N \rightarrow S$ acyl transfer was stated as being the lability of the amide bond in an imino acid (Pro).[52]



Scheme 5 Peptide thioesterification employing a C-terminal mercapto-prolyl-prolyl ester motif. *Reagents and conditions*: a) mild H_3O^+ (or mild ^-OH). b) Reagent K, 60 °C, 4 h. c) 10% v/v aqueous 3-mercaptopropionic acid (MPA), microwave, 40 °C, 1 h.

2.4 *N*-Alkyl cysteine-assisted thioesterification

Having accredited the $N \rightarrow S$ acyl shift to the presence of the imino acid proline, the authors diverted their attention to the use of other imino acid "devices" that might promote thioesterification. *N*-Alkylation of a C-terminal cysteine residue was devised as an alternative to facilitate the intramolecular $N \rightarrow S$ acyl migration under acidic conditions, followed by the intermolecular transthioesterification with MPA to afford the peptide thioester (Scheme 6).[52]

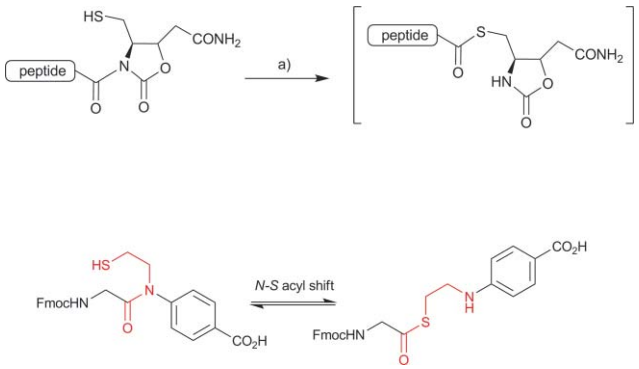


Scheme 6 Substituent effects in *N*-alkyl cysteine mediated thioester formation. *Reagents and conditions*: a) 5% MPA, r.t., 2–3 days.

N-Ethyl cysteine was incorporated into the synthesis of emmprin(49-58) and evidence of an equilibrium between the *N*- and *S*-peptide was presented. Both isomers could subsequently be transthioesterified with 5% MPA without any serious side reactions. The reaction rate with MPA was found to increase with increasing bulk of the R group attached to cysteine in the following order: Me < Et < *i*Bu. The installation of an *N*-alkyl cysteine instead of a mercaptomethylated proline was shown to markedly increase the reaction rate of thioesterification using a lower MPA concentration (5%) and lower temperature (25 °C). However, steric hindrance associated with the bulkier *N*-alkyl cysteine derivatives (*i*Bu and Bn) compromised the coupling of the adjacent glycine residue during synthesis, and hence the overall MPA thioester yield (28% and 17%) compared with *N*-Me Cys (33%) and *N*-Et Cys (34%). The *N*-ethyl derivative produced the highest peptide thioester yield and was relatively simple to prepare, though required HPLC purification. This method was shown to be applicable for the preparation of peptide thioesters with chiral amino acids at the C-terminus in acceptable yields and levels of epimerisation.

N-Alkyl cysteine-mediated thioesterification was rapidly extended to the assembly of a 95-amino acid chemokine, CCL27 utilising the Ag^+ promoted thioester method.[53] Interestingly the more reactive aryl thioesters,[3] rather than the alkyl thioesters derived from MPA, to be used in the fragment coupling reactions were readily obtained after cleavage from the resin with Reagent K followed by exposure to aqueous acetonitrile containing 6 M urea and 2% 4-mercaptophenylacetic acid (MPAA). More recently the *N*-alkyl cysteine method was applied to the assembly of an octabranched glycopeptide dendrimer with a molecular weight greater than 20 kDa.[54] In an attempt to present the tumour associated carbohydrate T-antigen (Gal-β(1,3)-GalNAc) in multi-valent fashion for eventual vaccine development an *N*-ethyl cysteine terminated glycopeptide fragment of MUC 1 was thioesterified upon exposure to 5% MPA for 2 days at room temperature in 20% yield. Reaction with a further *N*-alkylcysteine modified tri-lysine dendrimer core resulted in a tetra-branched species which, after repeating the thioesterification and conjugation to ethylenediamine (using the Ag^+/HOOBt thioester method) gave rise to an octa-branched product. The observed selectivity in the above transformations is once more attributed to a combination of distortion of the planarity of the amide bond through alkylation, and the secondary amine liberated is a stronger base (favouring protonation therefore *S*-peptide formation).

Otaka and co-workers have additionally shown that a C-terminal Cys residue, when in the context of an acyl oxazolidinone, can undergo $N \rightarrow S$ acyl transfer to yield thioesters (Scheme 7, i).[55] In this case the ground state geometry of the exocyclic amide is considered compromised as a consequence of steric factors imposed by the cysteine side chain and electronic factors since the $n_N \rightarrow \pi^*_{CO}$ interaction is additionally weakened by delocalisation of the nitrogen lone pair into the oxazolidinone carbonyl lowering the energy barrier for $N \rightarrow S$ acyl transfer. Despite the relatively lengthy synthesis of the solid-supported oxazolidinone and its sensitivity to piperidine during peptide chain assembly, requiring Aimoto's reagent mixture (1-methylpyrrolidine/hexamethyleneimine/1-hydroxybenzotriazole in NMP/DMSO (1 : 1)) for Fmoc removal, this method was successfully applied to the synthesis of a 9-residue peptide thioester derived from human brain natriuretic peptide-32 (hBNP-32). Native chemical ligation between this peptide and a synthetic *N*-terminal cysteine peptide (hBNP32-NH_2 (10-32)) was conducted in phosphate buffer (pH 7.6) containing 6 M guanidine.HCl in the presence of 1% v/v thiophenol to yield hBNP32-NH_2 in 76% yield for the ligation step. Racemisation of thioesters derived form chiral amino acids occurs with Aimoto's Fmoc removal reagent during synthesis, therefore only the Fmoc-glycyl oxazolidinone has been successfully used.



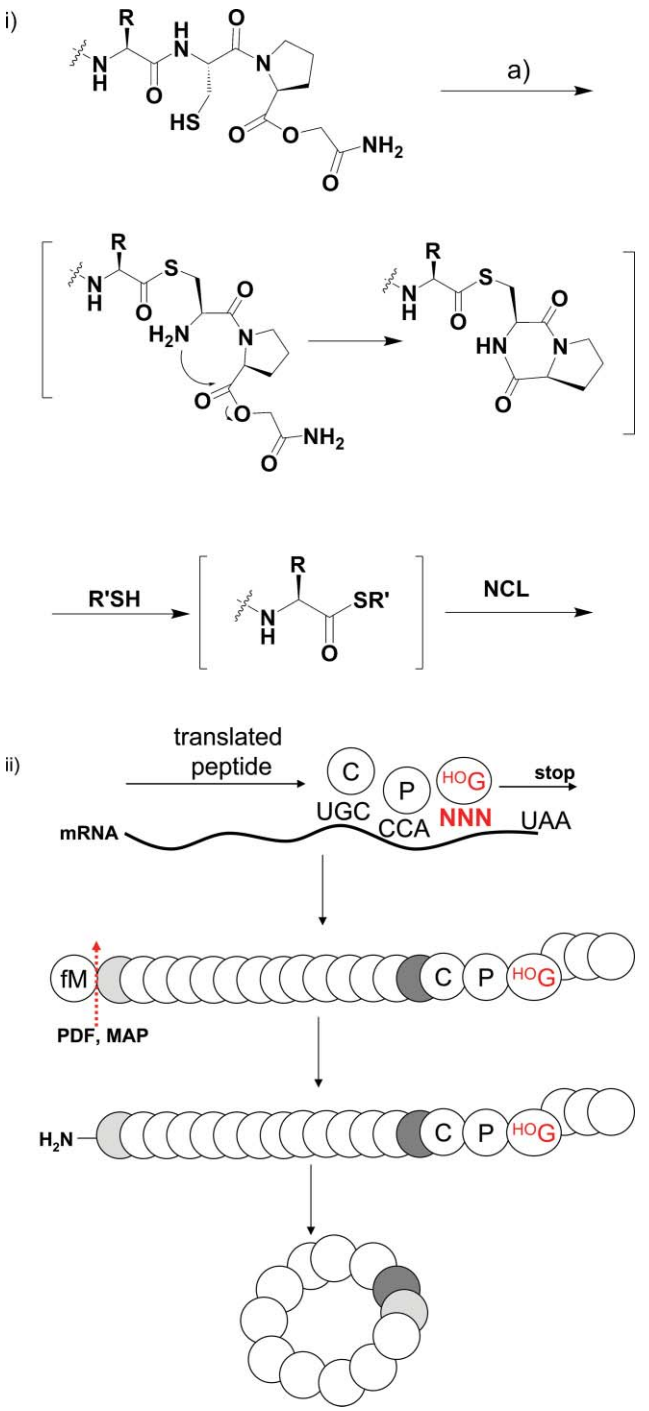
Scheme 7 Ground state destabilisation in i) *N*-acyloxazolidinones confers active-ester like properties and consequently facile thioesterification. *Reagents and conditions*: a) phosphate buffer pH 7.6, r.t. ii) an isolable thioester is produced from the anilide derivative upon exposure to 4 M HCl/DMF, 1% (w/v) TCEP, 8 h, 37 °C.

Otaka *et al.* have recently developed another device to facilitate intramolecular $N \rightarrow S$ acyl shift for peptide thioester synthesis using an *N*-substituted aniline linker (Scheme 7, ii).[56] It was hypothesised that anilide nitrogen rehybridisation could be effected by introduction of two adjacent sp^2 hybridised atoms, ultimately causing activation of the scissile amide bond. The core motif was based on *p*-(2-sulfanylethylamino)benzoic acid rather than the previously used acyloxazolidinone. Consequently ground state destabilisation again facilitated nucleophilic attack by the adjacent thiol at the distorted amide, and the undesired reverse $S \rightarrow N$ acyl shift was only observed in small amounts as the aniline nitrogen is a poor nucleophile allowing the equilibrium to favour the thioester. The synthesis of a pentapeptide thioester (H-HRFAG-$SCH_2CH_2SO_3Na$) was demonstrated using an Fmoc-glycyl aniline derivative with the $N \rightarrow S$ acyl shift achieved by treatment with 4 M HCl/DMF in the presence of 1% (w/v) TCEP for 8 h at 37 °C. The resulting isolable thioester was then finally converted into the MESNa thioester. The aniline linker was also described as allowing an on-resin $N \rightarrow S$ acyl transfer to occur, which was shown to afford ≤9 amino acid residue peptide thioesters upon thiolysis in ~70% yields. A small 9 residue phosphorylated serine-containing peptide thioester was also synthesised using the aniline linker in 67% yield. However, an unexplained partial racemisation occurred upon treating H-FAA-*Ar* peptide with 4 M HCl/DMF, which was suppressed when TFA was used for the $N \rightarrow S$ acyl shift.

2.5 Cysteinyl prolyl esters

Another method that uses a peptide with a C-terminal device includes the autoactivating cysteinyl prolyl ester (CPE) unit as described by Kawakami and Aimoto. Previously, Zanotti

et al.[57] had observed that a *p*-nitrophenyl (Np) ester, $PhCH_2CO$-Cys(S*t*Bu)-Pro-ONp, reacted with tributylphosphine to produce a diketopiperazine thioester, *cyclo*(-Cys($COCH_2Ph$)-Pro-). Kawakami and Aimoto demonstrated a similar process with peptides functionalised with C-terminal CPEs.[58–60] They proposed that an intramolecular $N \rightarrow S$ acyl transfer occurs to form the *S*-peptide, which then allows diketopiperazine formation. The resulting diketopiperazine peptide can then participate in thiol-mediated thioester exchange and ligation with, for example, a cysteinyl peptide under standard native chemical ligation conditions to afford a full-length polypeptide with the liberation of *cyclo*(-Cys-Pro-) (Scheme 8). The Pro residue within the CPE unit may be exchanged with sarcosine (*N*-methylglycine), however the ester is important as a C-terminal -Cys-Pro-carboxylic acid or -Cys-Pro-carboxamide has been shown to not ligate under the same conditions. Whilst the diketopiperazine thioester has not yet been detected, the by-product *cyclo*(-Cys-Pro-) has been observed by mass spectrometry. This method circumvents the need for a peptide thioester building block and potentially offers a higher yielding starting peptide-CPE using Fmoc SPPS, instead relying upon the entrapment of the *S*-peptide generated at equilibrium by a thiol. The reaction is reported to occur at room temperature and at pH ~7.0–8.0, under denaturing conditions, which is ideal for peptides with heat- or acid-sensitive linkages that are prone to aggregation. While a pH < 7.0 would be expected to favour the initial $N \rightarrow S$ acyl transfer, the conditions adopted are reasoned to favour thioester formation at the Xaa-Cys site because the amino group released is then free to form the diketopiperazine in basic solution. The $N \rightarrow S$ acyl transfer reaction rate in Cys-containing peptides appeared to be slow when analysed by RP-HPLC in weakly acidic solution (aq. acetonitrile containing 0.1% TFA), but even if the *S*-peptide was produced it rapidly reverted back to the *N*-peptide upon attempted isolation. Ligation yields recorded when the C-terminal amino acid was varied in a model peptide-CPE from Gly to Ala, Leu, or Val, were similarly good (49–60%), and epimerisation during ligation was found to be negligible though CPE-mediated ligation is also accompanied by a small amount of peptide hydrolysis. A key advantage of the CPE methodology over *N*-alkyl cysteine is that it utilises standard, readily available, materials typically associated with Fmoc/tBu chemistry and methods for assembly of the CPE terminated peptide could be considered as standard. A minor synthetic inconvenience is that the Xaa-Cys motif (in a Xaa-Cys-Pro-glycolyl sequence) needs to be introduced as a pre-prepared dipeptide to prevent diketopiperazine formation during peptide chain assembly. Protecting groups also appear to be retained throughout the process for α-amino groups/lysine ε-amino groups, presumably to prevent cyclisation/intermolecular reactions with thioester intermediates. However this method appears to circumvent many shortcomings in thioester synthesis since the thioester is only produced transiently under the reaction conditions at room temperature and at approximately neutral pH, and the fact that the peptide is assembled as a C-terminal carboxamide also maximises recovery of the thioester precursor. Thioesters can additionally be generated from CPE peptides upon exposure to *e.g.* MESNa at pH < 8.0 (to minimise hydrolysis) and 37 °C for 6 h in good yields. The degree of racemisation is low except in the case of serine thioesters where it was determined to be >20%.



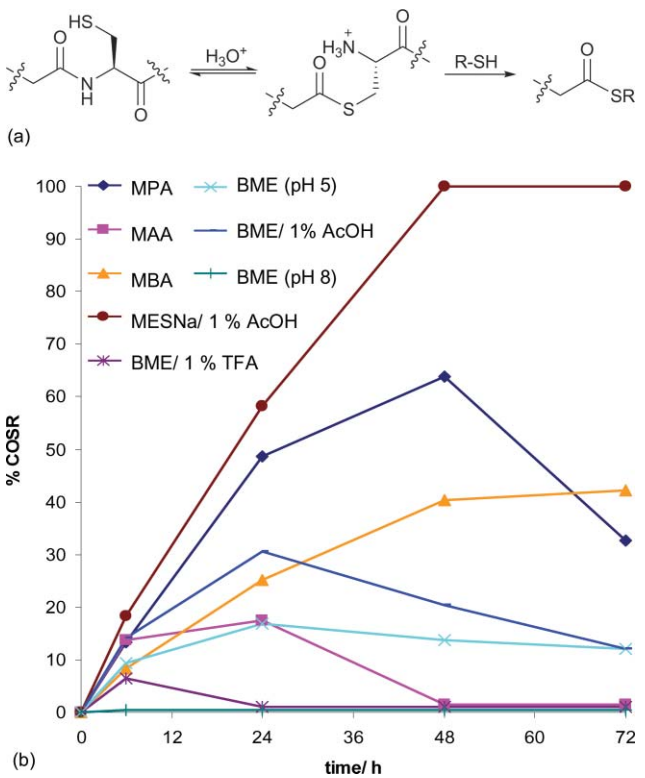
Scheme 8 (i) Cysteine prolinyl ester (CPE) mediated thioesterification/ligation. *Reagents and conditions*: a) 0.1 M buffer, pH 8.2, 6 M guanidine.HCl, 20 mM tris-(hydroxypropyl)phosphine (THP), *N*-cysteinyl peptide, r.t., 24 h. (ii) The bcPURE system allows bacterial production of CPE terminated peptides of interest. Peptide purification is facilitated by incorporation of the flag peptide between the CPE motif and the stop codon which is ultimately removed after purification upon exposure to aqueous buffer; pH 8 and thiophenol.

In a fascinating extension of the methodology Kawakami *et al.*[61] combined the principles of the CPE method with codon reprogramming using an *in vitro* transcription/translation system to develop a generally applicable method for the

recombinant expression then non-enzymatic synthesis of cyclic peptides (Scheme 8ii). The so-called bcPURE method (backbone cyclic peptides using recombinant elements) requires four key elements: (i) The gene of interest containing unused codons at specific sites that can direct the introduction of unnatural amino or hydroxy acids. Unused codons are introduced through withdrawal of particular amino acids from the host. (ii) Suppressor tRNA ($tRNA^{Asn\text{-}E2}_{NNN}$) charged with glycolic acid and capable of utilising the vacant codon (CAC in this case) is then introduced in an *in vitro* transcription/translation system facilitating assembly of a peptide adorned with a C-terminal cysteinylprolyl ester motif. The suppressor tRNA is charged by a ribozyme developed to acylate tRNA with various amino acids since substrate recognition is primarily directed towards a benzylic ester leaving group. Importantly the so-called "Flexizyme" can also charge tRNA with α-hydroxy acids such as glycolic acid. (iii) Recombinant peptide deformylase and (iv) methionine aminopeptidase are added to cleave the N-terminal methionine residue ultimately allowing peptide cyclisation to proceed *via* the *cyclo*(-Cys-Pro)-thioester. Combining these components in a cell-free expression system allows linear peptide CPE's to be produced which cyclise upon cleavage of the formyl methionine residue. When further endogenous amino acids are withdrawn and the cognate tRNA's charged with unnatural amino acids and *N*-methyl amino acids then libraries of novel cyclic peptides can be prepared. The versatility of the method was demonstrated through the production of the cyclic peptides Eptidemnamide, Scleramide, Rhesus-θ-defensin-1, and Sunflower trypsin inhibitor.

2.6 Native chemical thioesterification (NCT): peptide thioesterification in "device-free" systems

During our endeavours towards the semisynthesis of glycoproteins, we experimented with the cysteinyl prolyl ester method.[62] Device free thioesterification was observed as a consequence of the following model experiment. We had devised a parallel set of CPE ligation reactions employing identical peptide sequences (sequence = AENITTGC̲AEA-C̲PE). The difference between the two peptides was that one contained free cysteine sulfhydryls throughout, and the other contained a single acetamidomethyl (Acm) protected internal cysteine. The partially protected peptide was found to ligate to the corresponding cysteinyl peptide (CSLNENIT) to afford the full-length peptide: AENITTGC(*S*-Acm)AEACSLNENIT whereas the fully deprotected peptide did not. We reasoned that the presence, but more importantly the position[63] of the additional free internal cysteine residue may interfere with the desired ligation enough to limit this application of the CPE method. These findings led us to explore alternative strategies to attain peptide thioesters from the fully deprotected CPE peptide. In doing so, we recalled that Hojo and co-workers had reported the use of aqueous 3-mercaptopropionic acid (MPA) to produce peptide thioesters by means of an $N \rightarrow S$ acyl transfer.[51] Upon reaction of our model peptide-CPEs with MPA we found thioesterification occurred selectively at the internal Gly-Cys motif rather than at the expected Ala-Cys sequence. Delving deeper into the profile of favoured Xaa-Cys motifs led us to conclude that, rather surprisingly, peptide thioesters could be selectively prepared upon exposure to MPA at His-Cys, Gly-Cys or Cys-Cys sequences. Coincidentally, peptide thioesters with C-terminal His/Gly/Cys residues are known to accelerate the rate of native chemical ligation,[64] the reverse process. Consequently, installation of this sequence allows us to potentially program protein fragmentation chemoselectively in the presence of aqueous MPA and afford the desired thioester. Our studies demonstrated that this novel method of trapping the $N \rightarrow S$ acyl transfer intermediate, with a small molecular weight thiol has tremendous potential in preparing protein thioesters since the method requires no additional stimulus other than the appropriate sequence. Mechanistically, it is likely that an initial $N \rightarrow S$ acyl shift occurs by reversible nucleophilic attack of the cysteine thiol onto the carbonyl carbon of the amide across the His/Gly/Cys-Cys junction to give a proposed hydroxythiazolidine intermediate,[65] which then releases the protonated amine and undergoes an intermolecular thioester exchange with MPA to give the isolable thioester. However, thioester hydrolysis was observed with MPA so further optimisation of the reaction was undertaken. Reactions of a ^{13}C-labelled model peptide (H-AENITTG(^{13}C-1)C-NH_2), upon treatment with varyious thiols and under different reaction conditions, were followed by ^{13}C NMR spectroscopy. These studies revealed water-soluble MESNa to be a superior transthioesterification reagent to MPA. MESNa fully converted the starting peptide amide into the desired peptide thioester (H-AENITTG(^{13}C-1)-$SCH_2CH_2SO_3H$) within 48 h (Scheme 9),[66] compared to only ~60% conversion with MPA, and no peptide thioester hydrolysis occurred when MESNa (at pH 2.0) was employed as the thiol. The optimal reaction conditions were found be: heating at 60 °C with excess thiol (10%) in aqueous acidic media (pH 2.0-6.0), with 0.5% (w/v) TCEP to reduce any



Scheme 9 Optimisation of NCT employing various water soluble thiols (MAA = 2-mercaptoacetic acid, MBA = 4-mercaptobutyric acid, BME = β-mercaptoethanol). *Reagents and conditions*: 10 % thiol, 60 °C.

potential disulfide bonds. The concentration of the starting peptide is important as lower concentrations ($\leq$ 6 mM) were found to increase the reaction rate, likely as a result of a reduction in unwanted intermolecular disulfide bonds that act to decrease the effective concentration of peptide available to perform the $N \rightarrow S$ acyl transfer. As noted by Hojo with MPA treatment, aspartate residues are sensitive to the aqueous acidic medium of the reaction, even at pH 5.8 (though to a lesser degree). Peptides undergo slow hydrolysis at aspartate which becomes a non-trivial side-reaction when several are present within the sequence, as in a recombinant protein. Solutions to this problem include: exchange of Asp residues for Glu, conducting the reaction at pH 6.0 and at approximately physiological temperatures (40 °C) but this is at the expense of the peptide thioester yield. In synthetic peptides orthogonally protecting the Asp side-chain during the reaction may also prevent peptide hydrolysis at this site.

Despite this obstacle we reported the first application of our optimised reaction conditions to the synthesis of an antimicrobial and chemotactic 45 amino acid β- defensin, HBD3,[66] and broadened the application of our method to a more complex 22 amino acid thioester albeit with a lower thioester yield (~20%). Recently we also applied this method to the assembly of a 33 residue glycopeptide mimetic thioester used in the synthesis of a simply glycosylated glycoform of erythropoietin.[67] Interestingly we also noted that the thioesterification proceeds more rapidly (approximately twice as fast) when the thioester is derived from a C-terminal cysteine residue with a free carboxyl group than when it possesses a C-terminal carboxamide though the reason for this requires further investigation.

3 Conclusion

The increasing demand for chemically modified proteins, whether they serve structure/function studies or find application in biotechnology, will require that several complementary routes to thioesters for use in NCL, are available. This is because no single method will likely be compatible with all envisaged post-translational or non-natural modifications. Thioester production *via* $N \rightarrow S$ acyl shift combines the advantage of generally high yielding isolation of long-chain peptide acids or carboxamides with the lower likelihood of thioester racemisation, since thioesterification can occur at mildly acidic or near-neutral pH. This avoids exposure of thioesters to piperidine and is potentially more compatible with chemically fragile appendages. While there are still significant obstacles to overcome, some methods are currently being investigated in a biological arena, the ultimate test of their generality. Only when they can perform on polypeptides at near physiological pH and temperature will they rival the activity of inteins.

Notes and references

1 S. B. H. Kent, *Chem. Soc. Rev.*, 2009, **38**, 338.
2 A. Dirksen and P. E. Dawson, *Curr. Opin. Chem. Biol.*, 2008, **12**, 760.
3 D. Macmillan, *Angew. Chem., Int. Ed.*, 2006, **45**, 7668.
4 J.-P. Pellois and T. W. Muir, *Curr. Opin. Chem. Biol.*, 2006, **10**, 487.
5 L. Saleh and F. B. Perler, *Chem. Rec.*, 2006, **6**, 183.
6 M. A. Cole, S. E. Tully, A. W. Dodds, J. N. Arnold, G. E. Boldt, R. B. Sim, J. Offer and J. P. Wentworth, *ChemBioChem*, 2009, **10**, 1340.
7 L. S. Wong, F. Khan and J. Micklefield, *Chem. Rev.*, 2009, **109**, 4025.
8 B.-H. Hu, J. Su and P. B. Messersmith, *Biomacromolecules*, 2009, **10**, 2194.
9 T. C. E. Jr, J. Benner and M.-Q. Xu, *Protein Sci.*, 1998, **7**, 2256.
10 C. Piontek, D. V. Silva, C. Heinlein, C. Pöhner, S. Mezzato, P. Ring, A. Martin, Franz X. Schmid and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2009, **48**, 1941.
11 H. D. Mootz, *ChemBioChem*, 2009, **10**, 2579.
12 J.-Y. Yang and W. Y. Yang, *J. Am. Chem. Soc.*, 2009, **131**, 11644.
13 J. H. Appleby, K. Zhou, G. Volkmann and X.-Q. Liu, *J. Biol. Chem.*, 2009, **284**, 6194.
14 S. W. Lockless and T. W. Muir, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10999.
15 Q. Xu, D. Buckley, C. Guan and C.-H. Guo, *Cell*, 1999, **98**, 651.
16 T. Klabunde, S. Sharma, A. Telenti, W. R. Jacobs Jr. and J. C. Sacchettini, *Nat. Struct. Biol.*, 1998, **5**(1), 31.
17 A. Romanelli, A. Shekhtman, D. Cowburn and T. W. Muir, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 6397.
18 V. Muralidharan and T. W. Muir, *Nat. Methods*, 2006, **3**, 429.
19 Z. Du, P. T. Shemella, Y. Liu, S. A. McCallum, B. Pereira, S. K. Nayak, G. Belfort, M. Belfort and C. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 11581.
20 P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. Kent, *Science*, 1994, **266**, 776.
21 E. Haginoya, H. Hojo, Y. Nakahara, Y. Nakahara, K. Nabeshima, B. P. Toole and Y. Watanabe, *Biosci., Biotechnol., Biochem.*, 2006, **70**, 1338.
22 N. Yamamoto, Y. Tanabe, R. Okamoto, P. E. Dawson and Y. Kajihara, *J. Am. Chem. Soc.*, 2008, **130**, 501.
23 (*a*) Y. Shin, K. A. Winans, B. J. Backes, S. B. H. Kent, J. A. Ellman and C. R. Bertozzi, *J. Am. Chem. Soc.*, 1999, **121**, 11684; (*b*) R. Ingenito, E. Bianchi, D. Fattori and A. Pessi, *J. Am. Chem. Soc.*, 1999, **121**(49), 11369; (*c*) F. Mende and O. Seitz, *Angew. Chem., Int. Ed.*, 2007, **46**(24), 4577; (*d*) S. Mezzato, M. Schaffrath and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2005, **44**, 1650.
24 (*a*) X. Li, T. Kawakami and S. Aimoto, *Tetrahedron Lett.*, 1998, **39**, 8669; (*b*) H. Hojo, Y. Matsumoto, Y. Nakahara, E. Ito, Y. Suzuki, M. Suzuki, A. Suzuki and Y. Nakahara, *J. Am. Chem. Soc.*, 2005, **127**, 13720.
25 A. B. Clippingdale, C. J. Barrow and J. D. Wade, *J. Pept. Sci.*, 2000, **6**, 225.
26 R. Quaderer and D. Hilvert, *Org. Lett.*, 2001, **3**, 3181.
27 R. von Eggelkraut-Gottanka, A. Klose, A. G. Beck-Sickinger and M. Beyermann, *Tetrahedron Lett.*, 2003, **44**, 3551.
28 C. P. R. Hackenberger, C. T. Friel, S. E. Radford and B. Imperiali, *J. Am. Chem. Soc.*, 2005, **127**, 12882.
29 D. Swinnen and D. Hilvert, *Org. Lett.*, 2000, **2**, 2439.
30 T. Murase, T. Tsuji and Y. Kajihara, *Carbohydr. Res.*, 2009, **344**, 762.
31 J. B. Blanco-Canosa and P. E. Dawson, *Angew. Chem., Int. Ed.*, 2008, **47**, 6851.
32 A. P. Tofteng, K. S. Kasper, W. C.-F. Kilian, H.-J. Thomas and J. J. Knud, *Angew. Chem., Int. Ed.*, 2009, **48**, 7411.
33 J. Brask, F. Albericio and K. J. Jensen, *Org. Lett.*, 2003, **5**, 2951.
34 U. Boas, J. Brask and K. J. Jensen, *Chem. Rev.*, 2009, **109**, 2092.
35 S. Ficht, Richard J. Payne, Richard T. Guy and C.-H. Wong, *Chem.–Eur. J.*, 2008, **14**, 3620.
36 N. Ollivier, J.-B. Behr, Q. El-Mahdi, A. Blanpain and O. Melnyk, *Org. Lett.*, 2005, **7**, 2647.
37 K. Pachamuthu and R. R. Schmidt, *Synlett*, 2003, 659.
38 G. Roelfes and D. Hilvert, *Angew. Chem.*, 2003, **115**, 2377.
39 B. L. Pentelute and S. B. H. Kent, *Org. Lett.*, 2007, **9**, 687.
40 (*a*) D. Crich and A. Banerjee, *J. Am. Chem. Soc.*, 2007, **129**, 10064; (*b*) R. Okamoto and Y. Kajihara, *Angew. Chem., Int. Ed.*, 2008, **47**, 5402; (*c*) C. Haase, H. Rohde and O. Seitz, *Angew. Chem., Int. Ed.*, 2008, **47**, 6807; (*d*) J. Chen, Q. Wan, Y. Yuan, J. Zhu and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2008, **47**, 8521.
41 P. Botti, M. R. Carrasco and S. B. H. Kent, *Tetrahedron Lett.*, 2001, **42**, 1831.
42 L. E. Canne, S. J. Bark and S. B. H. Kent, *J. Am. Chem. Soc.*, 1996, **118**, 5891.
43 C. Marinzi, J. Offer, R. Longhi and P. E. Dawson, *Bioorg. Med. Chem.*, 2004, **12**, 2749.
44 J. Offer, C. N. C. Boddy and P. E. Dawson, *J. Am. Chem. Soc.*, 2002, **124**, 4642.
45 J. Offer and P. E. Dawson, *Org. Lett.*, 2000, **2**, 23.
46 D. Macmillan and D. W. Anderson, *Org. Lett.*, 2004, **6**, 4659.

47 B. Wu, J. Chen, J. D. Warren, G. Chen, Z. Hua and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2006, **45**, 4116.
48 (*a*) J. Vizzavona, F. Dick and T. Vorherr, *Bioorg. Med. Chem. Lett.*, 2002, **12**, 1963; (*b*) T. Kawakami, M. Sumida, K. Nakamura, T. Vorherr and S. Aimoto, *Tetrahedron Lett.*, 2005, **46**, 8805.
49 K. Nakamura, M. Sumida, T. Kawakami, T. Vorherr and S. Aimoto, *Bull. Chem. Soc. Jpn.*, 2006, **79**, 1773.
50 (*a*) K. Nakamura, H. Mori, T. Kawakami, H. Hojo, Y. Nakahara and S. Aimoto, *Int. J. Pept. Res. Ther.*, 2007, **13**, 191; (*b*) K. Nakamura, T. Kanao, T. Uesugi, T. Hara, T. Sato, T. Kawakami and S. Aimoto, *J. Pept. Sci.*, 2009, **15**, 731.
51 F. Nagaike, Y. Onuma, C. Kanazawa, H. Hojo, A. Ueki, Y. Nakahara and Y. Nakahara, *Org. Lett.*, 2006, **8**, 4465.
52 H. Hojo, Y. Onuma, Y. Akimoto, Y. Nakahara and Y. Nakahara, *Tetrahedron Lett.*, 2007, **48**, 25.
53 H. Hojo, Y. Murasawa, H. Katayama, T. Ohira, Y. Nakahara and Y. Nakahara, *Org. Biomol. Chem.*, 2008, **6**, 1808.
54 C. Ozawa, H. Katayama, H. Hojo, Y. Nakahara and Y. Nakahara, *Org. Lett.*, 2008, **10**, 3531.
55 Y. Ohta, S. Itoh, A. Shigenaga, S. Shintaku, N. Fujii and A. Otaka, *Org. Lett.*, 2006, **8**, 467.
56 S. Tsuda, A. Shigenaga, K. Bando and A. Otaka, *Org. Lett.*, 2009, **11**, 823.
57 G. Zanotti, F. Pinnen and G. Lucente, *Tetrahedron Lett.*, 1985, **26**, 5481.
58 T. Kawakami and S. Aimoto, *Chem. Lett.*, 2007, **36**, 76.
59 T. Kawakami and S. Aimoto, *Tetrahedron Lett.*, 2007, **48**, 1903.
60 T. Kawakami and S. Aimoto, *Tetrahedron*, 2009, **65**, 3871.
61 T. Kawakami, A. Ohta, M. Ohuchi, H. Ashigai, H. Murakami and H. Suga, *Nat. Chem. Biol.*, 2009, **5**, 888.
62 J. Kang, J. P. Richardson and D. Macmillan, *Chem. Commun.*, 2009, 407.
63 C. Haase and O. Seitz, *Eur. J. Org. Chem.*, 2009, 2096.
64 T. M. Hackeng, J. H. Griffin and P. E. Dawson, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 10068.
65 C. Ludwig, D. Schwarzer and H. D. Mootz, *J. Biol. Chem.*, 2008, **283**, 25264.
66 J. Kang, N. L. Reynolds, C. Tyrrell, J. R. Dorin and D. Macmillan, *Org. Biomol. Chem.*, 2009, **7**, 4918.
67 J. P. Richardson, C.-H. Chan, J. Blanc, M. Saadi and D. Macmillan, *Org. Biomol. Chem.*, **8**, 1351.

