A straightforward methodology to overcome solubility challenges for N-terminal cysteiny1 peptide segments used in native chemical ligation†

Skander A. Abboud, El hadji Cisse, Michel Doudeau, Hélène Bénédetti and Vincent Aucagne

One of the main limitations encountered during the chemical synthesis of proteins through native chemical ligation (NCL) is the limited solubility of some of the peptide segments. The most commonly used solution to overcome this problem is to derivatize the segment with a temporary solubilizing tag. Conveniently, the tag can be introduced on the thioester segment in such a way that it is removed concomitantly with the NCL reaction. We herein describe a generalization of this approach to N-terminal cysteiny1 segment counterparts, using a straightforward synthetic approach that can be easily automated from commercially available building blocks, and applied it to a well-known problematic target, SUMO-2.

Introduction

The advent of the native chemical ligation (NCL) reaction has revolutionized the field of chemical protein synthesis by offering a simple strategy to assemble unprotected peptide segments bearing mutually-reactive C-terminal thioesters and N-terminal cysteines with exquisite chemo- and regioselectivity. Twenty-six years after its discovery, NCL is still the gold standard reaction in the field. The continuous development of many related synthetic methodologies allowed to extend and simplify its applicability, and led to impressive applications in the total synthesis of functional proteins of more than 300 residues.

One of the major current limitations of NCL-based protein synthesis is the low solubility or tendency to aggregate of some of the segments. If this is anticipated when synthesizing a very hydrophobic target such as a transmembrane protein, segments from soluble hydrophilic proteins frequently prove to be problematic, while being often very hard to predict. NCL is usually conducted under denaturing conditions, typically 6 M guanidinium chloride, and is tolerant to the addition of organic solvents and detergents, thus substantially minimizing these solubility/aggregation problems during the reaction itself. Nevertheless, the purification and characterization of the segments prior to NCL regularly remains a severe bottleneck. Many synthetic strategies were developed to overcome this critical problem. For example, modification of the segment through N-methylation of backbone amides or fusion with a hydrophilic peptide, can dramatically increase its solubility; however, these modifications will permanently remain in the synthesized protein. Traceless approaches include the use of acid-labile N-2-hydroxy-4-methoxybenzyl (Hmb) groups on backbone amides, or Ser/Thr O-acyl isopeptide known to inhibit aggregation, but the most widely used strategy is the incorporation of a temporary hydrophilic “solubilizing tag”.

Such temporary tags are generally composed of lysines or arginines that bear a cationic charge side chain, often as homooligomers, and can be introduced either on a backbone amide, the C-terminus of the segment or the side chain of an Asn, Asp, Cys, Gln, Glu, Lys or Thr residues.

A large variety of linkers used to attach the tag to the peptide segment and designed to be cleaved in an additional step after the NCL (Scheme 1A) have been reported. Cleavage conditions include treatments with acids, bases, pH 4.5 buffer, sodium nitrite, metal catalysis, UV irradiation and autoproteolysis. An advantageous alternative is the introduction of the solubilizing tag on the C-terminal thioester moiety (Scheme 1B). In this case, the tag is cleaved in the course of the NCL reaction without needing an additional synthetic step, after playing its solubilizing role during the purification, characterization and handling of the problematic segment. This approach was pioneered by Aimoto and Kent for Boc-SPPS-based thioester synthesis, and later extended by Tietze to Fmoc-SPPS using a β-mercaptoproester precursor converted in situ into a thioester during NCL through an O → S acyl shift (a so-called peptide crypto-thioester). If this one-pot NCL/tag cleavage approach is clearly not suited to extreme situations where the ligation product remains insoluble or prone to aggregation, it is expected to be widely applicable in the many cases when the low-solubility/aggregation behavior is associated with a single isolated segment of the protein. Indeed, the additional segment ligated to

Centre de Biophysique Moléculaire, CNRS UPR 4301, Rue Charles Sadron, 45071 Orléans Cedex 2, France. E-mail: aucagne@cnrs-orleans.fr
† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0sc06001a
the problematic one can further play the role of solubility tag, strikingly demonstrated by the synthesis of fragments of transmembrane proteins using these techniques.\textsuperscript{14,23} However, the strategy is inherently limited to thioester segments and not suited for N-terminal cysteinyl counterparts.\textsuperscript{23}

In a closely related context, Valiyaveetil\textsuperscript{24} proposed the idea to introduce a solubility tag linked through a disulfide to the N-terminal Cys of a cysteiny1 segment synthesized by Boc-SPPS, the disulfide bond being cleaved in a first additional reduction step prior to NCL. In this case, the tag (Arg–Arg–Arg–Cys–NH\textsubscript{2})\textsuperscript{25} was introduced in solution through air oxidation-mediated formation of a mixed disulfide with the crude segment. This resulted in a very low yield in the tagged segment due to both insolubility of the non-tagged segment and the non-directed formation of the disulfide leading to complex mixtures. Nevertheless, the authors succeeded in the highly demanding semi-synthesis of ion channels through expressed protein ligation\textsuperscript{26} (EPL) with a recombinant thioester, demonstrating the feasibility of the approach.

We thought that devising a straightforward methodology for the introduction of such a disulfide-linked solubilizing tag on N-terminal cysteinyl segments, and extending the concept of concomitant NCL/tag cleavage could be extremely valuable and generally applicable in chemical protein synthesis (Scheme 1C). Indeed, disulfides are readily cleaved under standard NCL conditions generally including reducing agents (e.g. TCEP) and a large excess of aromatic thiols like 4-mercaptophenylacetic acid (MPAA).\textsuperscript{26}

We herein describe a straightforward methodology for the introduction of such a tag through Fmoc-SPPS, which can be easily automated on a standard peptide synthesizer, and exemplifies the utility of this method through the synthesis of SUMO-2, a previously reported difficult target.\textsuperscript{23c}

Results and discussion

The disulfide-based linker bridging the N-terminal cysteinyl segment and the solubilizing tag is the cornerstone of the strategy. Ideally, this linker should be (1) introduced through an automation-friendly solid phase step using commercially available materials, (2) bear a primary aliphatic amine or a suitable precursor for further Fmoc-SPPS elongation of a hydrophilic peptide sequence and (3) be stable to the Fmoc-SPPS conditions, including elongation and TFA-based cleavage.

Considering these requirements, we reasoned that incorporation of the N-terminal cysteinyl residue through the coupling of Boc-Cys(Npys)-OH would be ideal. The Npys group\textsuperscript{27} (S-3-nitro-2-pyridinesulfonyl) is classically used for the directed formation of a mixed disulfide on an N-terminal cysteine either on solid support\textsuperscript{28} or in solution, through reaction with a thiol. The simplest solution for automated synthesis would be a reaction on solid support with an amino-thiol. We selected two commercially available candidates: cysteamine (2-amino-ethane-1-thiol, 1), and 2-amino-1,1-dimethyl-ethane-1-thiol (2). Preliminary experiments with a model tripeptide (see ESI p. S5–S11\textsuperscript{f}) showed that disulfide formation was quantitative through simple incubation in NMP for 1 h with an excess (10 equiv.) of 2 as its hydrochloride salt. We also demonstrated that the disulfide was stable to a long piperidine treatment mimicking the repeated Fmoc deprotection conditions needed for the SPPS elongation of a hydrophilic peptide tag. Contrastingly, reaction with 1 was much less clean, and the resulting disulfide was not stable to the piperidine treatment. These results are in accordance with the known higher stability of tertiary thiol disulfides derivatives of cysteine like S-SiBu towards Fmoc-SPPS conditions as compared to simple non-hindered primary thiol disulfides.\textsuperscript{29}

Having in hand a robust method for the solid phase introduction of the linker (referred hereafter as Ades, 2-amino-1,1-dimethyl-ethyl-1-sulfanyl), we applied it to two different long peptides, followed by the automated introduction of a (Lys\textsubscript{k})\textsuperscript{30} hydrophilic tag through repeated couplings of Fmoc-Lys(Boc)-OH under standard conditions (Scheme 2). We started with a 41 amino acids (aa) model sequence devoid of any solubility problem, derived from the human mucin MUC1\textsuperscript{30} variable
number tandem repeat (VNTR) region made of a duplicated 20 aa sequence (7, Scheme 3). Gratifyingly, 7 was obtained in excellent yield and purity without needing of any further optimization of the synthetic protocol.

To test the concomitant NCL/tag cleavage, we synthesized the 41 aa crypto-thioester segment counterpart 3a, also derived from MUC1. This segment is equipped with an N-(2-hydroxy-4-nitrobenzyl)cysteine-based device (N-Hnb-Cys) capable of forming thioester in situ under NCL conditions (Scheme 3). N-Hnb-Cys crypto-thioesters are straightforward to synthesize through automated Fmoc-SPPS and show fast ligation kinetics owing to internal catalysis by a judiciously placed phenol group (Scheme 3).

To our delight, the ligation of 3a with 7 proceeded very cleanly, the (Lys)6-Ades tag being cleaved within seconds under the NCL conditions, giving the target 82 aa polypeptide 9 in a 62% isolated yield (ESI p. S18–S19). Additionally, we took the opportunity of this work to demonstrate the applicability of the C-terminal thioester solubilization strategy to N-Hnb-Cys crypto-thioesters. Quite expectedly, introduction of a (Lys)6 tag was straightforward, and gave segment 3b in good yields without needing any optimization. As for 3a, NCL with 7 proceeded cleanly, with comparable kinetics (ESI, Fig. S15†), giving 9 in an excellent 76% isolated yield (Scheme 3).

The cysteine residue in 9 was further desulfurized under classical conditions32 to give an alanine (ESI, p S20–S21) such as in the native MUC1 sequence.

Encouraged by these results, we then implemented this approach for the synthesis of the 93 aa SUMO-2. Small ubiquitin-related modifiers33 (SUMO) were first discovered in mammals in 1996.34 To date, five SUMO isoforms have been identified in humans, SUMO-1, 2, 3, 4 and 5. SUMOylation is a post-translational modification (PTM) consisting in the covalent attachment of the C-terminus of SUMO proteins via an isopeptide bond to specific lysine residues in target proteins. An enzymatic cascade controls the attachment, involving activator (E1), conjugating (E2), and sometimes ligase (E3) enzymes. This PTM is reversible, through deSUMOylation by sentrin/SUMO-specific proteases (SENPs).35 In contrast with ubiquitin,36,37 only few examples of the chemical synthesis of SUMO proteins, their dimers and conjugates, have been reported.5,12c,38,39 One illustration is the synthesis by Brik12c of SUMO-2-diubiquitin hybrid chains. In this work, the authors reported the low solubility, tendency to aggregation and unusual HPLC behavior of an N-terminal cysteinyl SUMO-2[46–93]40 segment. To circumvent this issue, they developed an elegant C-terminal solubilizing tag in which a 3,4-diaminobenzoic acid (Dbz) linker41 was employed to attach a poly-Arg tag to the C-terminus of this segment. We thought that it could be interesting to challenge our (Lys)6-Ades methodology with this benchmark target.

We applied our synthetic protocol to (Lys)6-Ades-SUMO-2[48–93] (12),42 which gave a clean and soluble crude product that was purified by standard HPLC. As anticipated from Brik’s report, in sharp contrast with 12, the non-tagged version exhibited anomalous HPLC behavior (ESI, S11, Fig. S11†) and...
the crude peptide was barely soluble in HPLC solvents: 0.05 mg mL⁻¹ in 8 : 2 : 0.01 H₂O/MeCN/TFA, thus further validating the choice of this target for this work. Gratifyingly, NCL of 12 with SUMO-2[1–47] N-Hnb-Cys crypto-thioester 10 proceeded very cleanly as shown in the analytical HPLC monitoring of the reaction (Scheme 4), giving the expected SUMO-2[1–93] 14 in a good 40% isolated yield. As observed with the MUC1-derived model segments, the HPLC chromatograms show nearly instantaneous conversion of 12 into non-tagged 13, concomitantly with the slightly slower conversion of S-SrBu-protected dormant N-Hnb-Cys crypto-thioester 10 into an active form 11.⁴⁴

Note that in accordance with Melnyk’s findings,⁴⁵ 15% of a side product showing a molecular mass of [M – 18] m/z relative to the desired product 14 was observed when performing the reaction at 37 °C for a prolonged time and was attributed to aspartimide formation at one of the Asp-Gly sites. Lowering the temperature to 25 °C nearly abolished this side reaction (ESI p. S22–S25).

Finally, in order to further characterize the synthesized SUMO-2 14 from a biochemical point of view, it was folded by simple solubilization into a neutral buffer as described.³⁹/⁴⁰ Its three-dimensional structure was evaluated by circular dichroism, showing a spectrum essentially identical with the previously reported ones⁴⁶ for recombinant and synthetic SUMO-2 (ESI Fig. S21†).

We then compared 14 to a commercially available recombinant version of SUMO-2 for its ability to act as a substrate of the SUMO E2 conjugating enzyme Ubc9 (ref. 47) and a SUMO E1 activating enzyme (namely a heterodimeric complex of SAE1 and SAE2 proteins)⁴⁸ for conjugation to the ubiquitin-conjugating enzyme Ube2K (also known as E2-25K and Huntington Interacting Protein 2). This target is reported to be among the best in vitro substrates for Ubc9-dependent SUMOylation known thus far, although its SUMOylated lysine residue (Lys14) is not surrounded by a consensus SUMOylation motif.⁴⁹ Using anti-SUMO-2/3 and anti-Ube2K antibodies, Western-blots analysis first showed that synthetic SUMO-2 is recognized as efficiently as the recombinant one. The appearance of an intense band recognized by both antibodies and migrating at the expected molecular mass of a SUMO-modified Ube2K demonstrated the successful SUMOylation of the acceptor protein in both cases. Weaker bands likely corresponding to di-SUMOylated Ube2K and di-SUMO-2 were also observed, consistent with the presence of a SUMOylatable lysine in SUMO-2 (Lys11). As expected, no reactions occurred in the absence of ATP and Mg²⁺, required cofactors for the E1 enzyme. Altogether, these experiments demonstrate the fully functional nature of synthesized 14 (Fig. 1 and ESI p. S27–S29).

**Conclusions**

Poor solubility and aggregation of peptide segments are main bottlenecks for the chemical synthesis of proteins using native chemical ligation. Numerous strategies for the solubilization of problematic segments through temporary modification have been developed such as the introduction of solubilizing tags, but often require complex synthetic strategies, in-house synthesis of building blocks or extra steps to generate the native protein sequence after the NCL-based assembly. In this work, we have introduced a straightforward methodology for the temporary solubilization of N-terminal cysteiny1 segments, based on the introduction of an oligolysine tag through a disulfide linkage with the N-terminal cysteine residue. This [Lys]₉-Ades tag is easily incorporated in the target segment through automated solid-phase synthesis using commercially
available building blocks, is stable during handling, purification and storage of the segment, while being cleaved within seconds under NCL conditions to generate in situ the reactive free cysteine. We exemplified the broad potential of this method through the NCL-based synthesis of a model polypeptide derived from the human mucin MUC1, in addition to a well-known difficult small protein target, SUMO-2. Due to its overall simplicity and efficiency, we believe that this strategy will advantageously complement existing methodologies in the synthesis of other challenging proteins.

Conflicts of interest
There are no conflicts to declare.

Acknowledgements
We warmly thank Jean-Baptiste Madinier for assistance in peptide synthesis and Dr Cyril Colas for HRMS analyses. The University of Orleans and Doctoral School SSBCV (ED549) are gratefully acknowledged for PhD fellowships for SAA and EC, and ANR (EasyMiniprot, ANR-15-CE07-0022) for financial support.

Notes and references


23 Note that, besides cationic solubilization tags, concomitant rearrangement of Ser/Thr O-acyl isopeptides during NCL has been exploited, taking advantage of the aggregation-inhibiting properties of the depsipeptide linkage during HPLC purification of segments: see for example ref. 3d, e and 9a.


27 R. Matsueda, S. Higashida, R. J. Ridge and G. R. Matsueda, Activation of conventional S-protecting groups of cysteine


Note that for practical reasons, in this case the native cysteine Cys49 was mutated to a serine, and that ligation was performed at an Ala48 → Cys site which was desulfurized after the NCL to restore the native alanine residue.


42 SUMO-2 methionines residues [Met1, Met44, Met55 and Met78] were substituted by isosteric norleucines in 10 and 12, following the common use in chemical protein synthesis to avoid air-mediated oxidation of the side chain sulfides into sulfoxides.

43 Note that the HPLC monitoring of the NCL reaction was delicate, due to the *in situ* formation of non-tagged SUMO-2[48–93] S11 that required extensive washes of the column between analyses.

44 S-tert-Butylsulfanyl (S-SrBu) is used to afford the crypto-thioester in a dormant state, unable to rearrange into a reactive thioester during handling, purification and storage. Under classical NCL conditions (including TCEP and MPAA), SrBu is readily cleaved within a few minutes affording active N-Hnb-Cys crypto-thioester species.


50 Note that we only tested the (Lys)_6-Ades solubilizing tag under the most classically used conditions for NCL, including a large excess of TCEP and MPAA. In rare cases, NCL reactions are performed under non-reductive conditions, either making use of pre-formed arylthioesters (such as MPAA thioesters), or using azaheterocycles as catalysts instead of an external arylthiol: the strategy herein described would expectedly not be applicable.