A novel PEG-based solid support enables the synthesis of >50 amino-acid peptide thioesters and the total synthesis of a functional SUMO-1 peptide conjugate†

Emmanuelle Boll, Hervé Drobecq, Nathalie Ollivier, Laurent Raibaut, Rémi Desmet, Jérôme Vicogne and Oleg Melnyk*

A bis(2-sulfanylethyl)amino PEG-based resin enabled the synthesis of large (~50 Aa) SEA or thioester peptides using Fmoc-SPPS. These peptide segments permitted the first total synthesis of a 97 amino-acid long SUMO-1-SEA peptide thioester surrogate and of a functional and reversible SUMO-1 peptide conjugate.

Introduction

The synthesis of proteins gives access to large peptides or small proteins with an atom-by-atom control of their structure. Today, proteins are usually assembled by the sequential ligation of unprotected peptide segments. Native chemical ligation (NCL) or expressed protein ligation of a C-terminal peptide thioester with an N-terminal cysteinyl peptide is the most popular amide bond forming reaction used in the field, although other complementary methods are emerging.

The large diversity of chemical methods available now for producing modified peptides in high purity is increasingly used for the total or semisynthesis of modified proteins which cannot be easily produced using biological methods. Proteins featuring well-defined post-translational modifications at specific sites such as phosphorylation or glycosylation belongs to this family. Another post-translational modification which has attracted much attention recently is ubiquitylation, which has been addressed using total or semisynthetic approaches. The preparation of ubiquitin conjugates has also been achieved using non-native chemoselective ligation methods such as oxime or disulfide, or 1,2,3-triazole bond forming reactions.

Recent work described also the total synthesis of the 85 amino-acid long ubiquitin-fold modifier 1 (UFM1) using sequential α-ketoacid–hydroxylamine ligation methods. Besides this, the total synthesis and introduction of other protein modifiers is faced with significant challenges. In particular, the total synthesis of small ubiquitin-like modifier (SUMO) proteins and SUMO conjugates remains to be addressed. C-terminal glycine residue of SUMO modifier is linked to the side chain amino group of a lysine within the target protein through an isopeptide bond. This reversible modification is implicated in many important physiological functions such as transcription and DNA repair. SUMO modifier is thought to act by modulating the interactions of the target protein with its partners. However, the molecular mechanisms implicated in this regulation are still poorly understood mostly because SUMO conjugates are highly labile within cells, making their isolation highly challenging.

Up to now, SUMO-1 or SUMO-2 conjugates were prepared by semisynthesis using recombinant SUMO proteins and non-cleavable 1,2,3-triazole linkers. We show for the first time that protein total synthesis can give access to homogeneous, functional and reversible SUMO-1 conjugates.

The work reported here relies on the chemistry of bis(2-sulfanylethyl)amido (SEA) thioester surrogate (Scheme 1). First, the assembly of SUMO-1 proteins described in this work required to push the limits of current technology by addressing the Fmoc-SPPS synthesis of long (~50 Aa) SEA peptides and peptide thioesters. This was achieved by developing novel SEA PEG-based solid supports which are compatible with standard Fmoc-SPPS. We report also the assembly of SUMO-1 isoform modified at the C-terminus by a latent SEA thioester surrogate (Scheme 1). This protein could be isolated or alternately ligated in situ with a SUMO-1 peptide consensus motif to produce a 12.5 kDa SUMO-1-peptide conjugate in a one-pot process. Importantly, the tertiary structure and functionality of the SUMO-1 modifier within the conjugate as well as its reversible nature was demonstrated by the specific cleavage of the glycyl-cysteinyl peptide bond with a recombinant SUMO-1 deconjugating enzyme.

† Electronic supplementary information (ESI) available: Experimental procedure and characterization data for all compounds. See DOI: 10.1039/c3sc53509f
Results and discussion

The primary sequence of SUMO-1 protein features only one internal cysteine residue in position 51 (Scheme 1). A Cys residue was appended to the target lysine side-chain within the SUMO-1 peptide consensus motif to enable the assembly of the SUMO-1 peptide conjugate in three pieces using the NCL or related reactions. Lys(Cys) peptides have also been used successfully by other groups for accessing to ubiquitinated proteins. An alternative would be to use a target peptide featuring a γ-mercaptolysine or δ-mercaptolysine residue. However, the access to a native SUMO-1 conjugate using this latter strategy would require protecting the Cys51 thiol group during the desulfurization of the mercaptolysine residue.

Given the position of the cysteinyl residues in the target SUMO-1 conjugate, the assembly can start with the SUMO-1 peptide consensus motif or alternately with the SUMO-1 (2–50) peptide segment. The latter approach was selected in this work to enable the isolation of SUMO-1 protein featuring a latent thioester surrogate functionality at the C-terminus. We chose the bis(2-sulfanylethyl)amido (SEA) N-S-acyl system as latent thioester surrogate, which in its cyclic form called SEAox is compatible with the NCL reaction used for assembling the SUMO-1 protein.

This N-to-C assembly strategy resembles that developed by Brik and co-workers for the synthesis of a ubiquitin latent thioester. These authors used a C-terminal S-protected N-methylcysteine as latent thioester, which was deprotected and exchanged by 3-mercaptopropionic acid (MPA) in a subsequent step.

The strategy depicted in Scheme 1 is faced with the challenge of synthesizing 50 amino-acid long peptide segments featuring a C-terminal SEA or alkylthioester group. The synthesis of large peptide thioesters using standard Fmoc-SPPS is one of the major bottlenecks for addressing the synthesis of large proteins or complex scaffolds such as SUMO-1 conjugates. C-terminal bis(2-sulfanylethyl)amido (SEA) peptides can be synthesized by Fmoc-SPPS and used as precursors for the synthesis of peptide alkylthioesters. Up to now, we used a bis(2-sulfanylethyl) amino polystyrene (SEA PS) resin for the Fmoc-SPPS of SEA peptides. Although SEA PS resin permitted the synthesis of various SEA peptides composed of up to 35 amino acids, we anticipated that the synthesis of the large SUMO-1 peptide segments would require a modern PEG-based solid support instead, which is more adapted for the synthesis of long peptides.

Table 1  Fmoc-SPPS of SEA peptides using SEA PEG-based resins

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Protein</th>
<th>Type of automated synthesizer</th>
<th>Resin</th>
<th>nb of Aa</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7a</td>
<td>SUMO-1 (2–51)</td>
<td>Column</td>
<td>5c</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>7b</td>
<td>SUMO-1 (52–97)</td>
<td>Column</td>
<td>5c</td>
<td>46</td>
<td>3</td>
</tr>
<tr>
<td>7c</td>
<td>HGF (31–69)</td>
<td>Batch</td>
<td>5d</td>
<td>40</td>
<td>7</td>
</tr>
<tr>
<td>6d</td>
<td>HGF (125–176)</td>
<td>Column</td>
<td>5c</td>
<td>52</td>
<td>5</td>
</tr>
</tbody>
</table>

a HPLC purified.
SEA PEG-based solid supports were prepared by reacting the trifluoroacetate salt of bis(2-sulfanylthyl)amine 2 with commercially available trityl alcohol PEG-based resins as shown in Scheme 2. We targeted solid supports of type 5 for which both thiol groups are linked to the solid support and protected by a trityl group as for SEA PS resin. However, SEA PS resin was prepared by reacting the bis(2-sulfanylthyl)amine 2 with trityl chloride polystyrene resin in N,N-dimethylformamide (DMF). In this case, an efficient protection of both thiol groups and the formation of cross-links within the PS beads was facilitated by the large excess and high loading of the starting trityl chloride polystyrene resin (~1.4 mmol g⁻¹, 10 equiv.). The loading of commercially available trityl alcohol PEG-based resins (0.25–0.6 mmol g⁻¹) is significantly less than the loading of trityl chloride polystyrene resins. Therefore, partial protection and incomplete cross-linking of the bis(2-sulfanylthyl)amine 2 leading to the formation of solid supports of type 3 was a potential issue. Indeed, the reaction of dithiol 2 with PEG-based solid supports 1a,b in the presence of trifluoroacetic acid yielded solid supports of type 3 as shown by the development of a positive Ellman assay. The presence of free thiols during the Fmoc-SPPS can be troublesome. Therefore, solid supports 3a,b were treated with N,N′,N′,N′-tetramethyl-azodicarboxamide (TMAD) at pH 7 to mask the thiol groups as internal disulfides. Indeed, the resulting solid supports 4a,b furnished a negative Ellman assay, while the chloranil assay was strongly positive. The coupling of Fmoc-Ala-OH using HATU/DIEA activation followed by the quantification of the Fmoc group indicated a loading of 0.15–0.21 mmol g⁻¹. The usefulness of solid supports 4a,b for the Fmoc-SPPS of SEA peptides was verified by the automated synthesis of model peptide ILKEPVHGA-SEA (49–54% crude, 18–21% isolated by HPLC).

Interestingly, solid supports 1c,d furnished directly SEA resins of type 5 as shown by the development of a negative Ellman assay and of a strongly positive chloranil assay, with a loading of ~0.11–0.2 mmol g⁻¹. Solid supports 5c,d proved to be useful for the Fmoc-SPPS of long SEA peptides as shown later. Consequently, the study of solid supports 4a,b was discontinued.

Four different peptides were synthesized to illustrate the usefulness of SEA PEG resins 5c,d for accessing to large SEA peptides (Scheme 3, Table 1). Peptides 7a and 7b correspond to SUMO-1 (2–51) and SUMO-1 (52–97) respectively and were used in this work for the assembly of SUMO-1 proteins. Peptides 6d and 7c are derived from hepatocyte growth factor (HGF). Peptide 7e illustrates the compatibility of SEA PEG resins 5d for the automated batch peptide synthesis, whereas the other peptides including the 52 amino-acid peptide 6d were assembled using an automated column peptide synthesizer. Usually, the oxidation of SEA peptides of type 6 (SEAon form) into cyclic disulfides 7 (SEAoff) is performed prior to the HPLC purification to avoid the potential N,S-acyl shift of SEAoff group during the elution step. The oxidation of peptides 6a–c by iodine yielded successfully peptides 7a–c. In contrast, this procedure induced the aggregation of peptide 6d, which was therefore purified directly by HPLC.
SEA peptide 6d was further ligated with a 72 amino-acid divalent cysteinyl peptide 8 in the presence of 4-mercaptohexanoylacid (MPAA, 200 mM) and tris(2-carboxyethyl)phosphine (TCEP, 200 mM) as shown in Scheme 4 to give the ~20 kDa branched peptide 9 (Fig. 1). Although the reaction was complicated by the partial aggregation of 6d and of the ligation product 9 (see Fig. S29†), thereby explaining the modest yield for the ligation reaction (5%, 22% per step), this example illustrates the utility of SEA ligation for accessing large peptide scaffolds.

The successful SPPS of SUMO-1 peptides 7a,b set the stage for the assembly of SUMO-1-SEAoff protein 11 (Scheme 5). First, the SEAoff group of peptide 7a was exchanged by MPA at pH 4 in the presence of TCEP to produce SUMO-1 (2–50)-MPA thioester segment 10. The exchange proceeded very cleanly (see Fig. S7A†). The conversion reached 87% after 24 h, after which the reaction mixture was purified by HPLC to yield 28% (3.0 mg) of peptide thioester 10. The modest yield is mainly due to the small scale of synthesis. Gratifyingly, ligation of MPA thioester 10 with SEAoff segment 7b in the presence of MPAA (200 mM) at pH 7 furnished SUMO-1-SEAoff protein 11 and the mixed disulfide SUMO-1-SEAoff-MPAA 12 as a minor component after HPLC purification (22%, Fig. 2). The presence of the mixed disulfide 12 is not problematic since the reduction of the Cys51-MPAA disulfide bond is expected to occur during the subsequent activation of the C-terminal SEAoff group in the presence of TCEP.

The successful synthesis of SUMO-1-SEAoff protein 11 set the stage for the one-pot assembly of SUMO-1 conjugate 15, for which SUMO-1 modifier is attached to the side-chain of the target lysine residue through a cysteine linker (Scheme 6). Model peptide 14 features a ΨKXE(D) sumoylation consensus motif where Ψ is a hydrophobic amino acid residue and X any amino acid. These positions are occupied in peptide 14 by Ile and Ala residues, respectively. In the one-pot assembly process, the formation of SUMO-1-SEAoff protein 11 was
followed by the addition of TCEP (200 mM) and of peptide 14 in slight excess. The HPLC trace of the crude reaction mixture (Fig. 3a) shows that the target conjugate 15 is by far the major product, which was purified by HPLC with a 13% overall yield (Fig. 3b). The structure of conjugate 15 was confirmed by ESI (Fig. 3c) and chiefly MALDI-TOF mass spectrometry analysis (Fig. 3d, see also Fig. S21–23† for the in source fragmentation data).

The solution NMR structure of SUMO-1 is available and shows an α-helical content of 10%. Consequently, the α-helical content for conjugate 15 was expected to be ~9% if the target peptide is considered to be random coil. The solubilization of conjugate 15 in water furnished the circular dichroism (CD) spectrum shown in Fig. 4a. The ellipticity at 222 nm indicated an α-helix content of 8.6%, that is a value very close to the expected one. To unambiguously demonstrate the presence of a folded SUMO-1 domain within conjugate 15, we used ubiquitin-like specific protease 1 (Ulp1) as a sensor probe. Ulp1 is a yeast cysteine protease which cleaves specifically the isopeptidic bond linking SUMO-1 to its target proteins. The crystal structure of Ulp1 in complex with a yeast SUMO-1 protein analog has been solved and shows an extensive contact surface area of about 2400 Å² between the two proteins. Thus, Ulp1 requires the SUMO-1 domain to be folded to exert its protease activity on SUMO-1 conjugates. Interestingly, incubation of SUMO-1 conjugate 15 with recombinant Ulp1 enzyme resulted in the specific cleavage of the Gly97-Cys bond within the conjugate (Fig. 4b and c). HPLC analysis of the cleavage mixture and MALDI-TOF analysis of the eluted peaks showed the exclusive formation of SUMO-1 protein and of peptide 14 (Fig. 4b).

Taken together, these data show the functionality of the SUMO-1 domain within conjugate 15 and the efficiency of the synthetic strategy depicted in Scheme 1.

Conclusions

In conclusion, we report a novel bis(2-sulfanylethyl)amino (SEA) PEG resin which enables the Fmoc-SPPS of >50 amino-acid long SEA peptide thioester surrogates. The access to such long peptide thioester surrogates enables the synthesis of complex protein targets and extends the limits of protein total synthesis. In particular, we describe for the first time the total synthesis of a folded, functional and reversible SUMO-1 conjugate. Access to SUMO conjugates proteins should facilitate the biochemical and structural implications of this post-translational modification.

Acknowledgements

This work was supported financially by Cancéropolo Nord Ouest, SIRIC OncoLille, Région Nord pas de Calais and by the European Community.

Notes and references
