

## Single amino acid Gly-tag enables metal-free protein purification

L. Purushottam,<sup>af</sup> Unnikrishnan V. B.,<sup>af</sup> D. G. Rawale,<sup>a</sup> M. Gujrati,<sup>b</sup> S. D. Mishra,<sup>ab</sup> Sajeev T. K.,<sup>b</sup>  
Neelesh C. Reddy,<sup>a</sup> S. R. Adusumalli,<sup>a</sup> R. K. Mishra,<sup>b</sup> and V. Rai<sup>\*a</sup>

<sup>a</sup>Department of Chemistry, <sup>b</sup>Department of Biological Sciences, Indian Institute of Science Education  
and Research (IISER) Bhopal, India.

vrai@iiserb.ac.in

### Contents

|   |     |
|---|-----|
| <b>1. General information</b> .....   | S2  |
| <b>2. General procedures</b> .....  | S3  |
| 2a. Expression of SUMO1 protein with N-terminus glycine .....                                 | S3  |
| <i>Bacterial transformation</i> .....   | S3  |
| <i>Protein purification</i> .....   | S3  |
| 2b. Immunofluorescence .....  | S4  |
| 2c. Western blotting .....  | S4  |
| 2d. SUMOylation reaction .....  | S4  |
| 2e. Procedure for purification of insulin and myoglobin .....                                 | S5  |
| 2f. Procedure for purification of expressed SUMO1 from cell lysate .....                      | S5  |
| 2g. Procedure for purification of insulin from mixture of proteins and cell lysate .....      | S6  |
| 2h. Labeling of protein in solution phase <sup>1</sup> .....                                  | S6  |
| 2i. Calculation of concentration of sample using spectrophotometer .....                      | S6  |
| <b>3. Synthesis and characterization data</b> .....   | S7  |
| <b>4. Experimental data for protein purification</b> .....                                    | S11 |
| 4a. Immobilization of reagent (2a) on NHS sepharose resin .....                               | S11 |
| 4b. Immobilization of N-terminus Gly containing proteins on modified sepharose resin 5a ..... | S12 |
| 4c. Release of the bound protein from modified sepharose resin .....                          | S13 |
| 4d. CD data of released insulin 1a .....  | S14 |
| 4e. Purification of protein of interest from a mixture of proteins and cell lysate .....      | S15 |
| 4f. In vitro activity of purified insulin.....  | S17 |
| 4g. Activity analysis of purified SUMO1 .....   | S17 |
| <b>5. Spectral data</b> .....   | S18 |
| <b>6. References</b> .....  | S23 |

## 1. General information

The reagents, proteins, and enzymes were purchased from Sigma-Aldrich, Alfa Aesar and Merck Novabiochem. Hydrazide agarose beads were purchased from Thermo Scientific. Boronic acid (polymer bound) was purchased from Sigma Aldrich. The organic solvents used were reagent grade. Aqueous buffers were prepared freshly using Millipore Grade I water (Resistivity > 5 M $\Omega$  cm, Conductivity < 0.2  $\mu$ S/cm, TOC <30 ppb). Mettler Toledo (FE20) pH meter was used to adjust the final pH. The reaction mixture for the small molecules was stirred (Heidolph, 500-600 rpm). Proteins were either vortexed or incubated in incubator-shaker Thermo Scientific MaxQ 8000 (350 rpm, 25-37 °C). Amicon® Ultra-0.5 mL 3-kDa or 10-kDa MWCO Centrifugal Filters from Merck Millipore was used to remove small molecules from protein mixture, desalting and buffer exchange. Organic solvents were removed by BUCHI rotavapor R-210/215 whereas aqueous samples were lyophilized by CHRiST ALPHA 2-4 LD plus lyophilizer. Circular Dichroism (CD) measurements were recorded on JASCO J-815 CD spectropolarimeter equipped with Peltier temperature controller. All the spectra were measured with a scan speed of 50 nm/min, spectral band width 1 nm using 1 mm path length cuvette at 25 °C. UV-Vis spectra were recorded in Agilent Carry-100 UV-Vis Spectrophotometer connected with peltier temperature controller.

**Chromatography:** Thin-layer chromatography (TLC) was performed on silica gel coated aluminium TLC plates (Merck, TLC Silica gel 60 F254). The compounds were visualized using a UV lamp (254 nm) and stains such as iodine, ninhydrin, 2,4-diphenylhydrazine. The flash column chromatography of reagents was carried out on Combiflash Rf 200 or gravity columns using 230-400 or 100-200 mesh silica gel from Merck.

**Nuclear magnetic resonance spectra:**  $^1\text{H}$ ,  $^{13}\text{C}$  and spectra were recorded on Bruker Avance III 400 and 500 MHz NMR spectrometer.  $^1\text{H}$  NMR spectra were referenced to TMS (0 ppm)  $\text{CDCl}_3$  (7.26 ppm), whereas  $^{13}\text{C}$  NMR spectra were referenced to  $\text{CDCl}_3$  (77.16 ppm). Peak multiplicities are designated by the following abbreviations: s, singlet; bs, broad singlet; d, doublet; t, triplet; q, quartet; m, multiplet; dd, doublet of doublets; ddd, doublet of doublet of doublets. Spectra were recorded at 298 K.

**Mass spectrometry:** Agilent Technologies 1200 series HPLC paired to Agilent 6130 mass spectrometer (ESI/APCI) was used for ESI-MS data. HRMS data were recorded on Bruker Daltonics MicroTOF-Q-II with electron spray ionization (ESI). Matrix assisted laser desorption/ionisation time of flight mass spectrometry was performed with Bruker Daltonics UltrafleXtreme Software-Flex control version 3.4, using sinapic acid and  $\alpha$ -cyano-4-hydroxycinnamic acid (HCCA) matrix. Data analysis was performed using flex analysis. Peptide mass and fragment ion calculator (<http://db.systemsbio.net:8080/proteomicsToolkit/FragIonServlet.html>) were used for peptide mapping and sequencing.

Acetonitrile and H<sub>2</sub>O were buffered with 0.01% formic acid.

**Method A** (Column: Agilent, Poroshell 300 SB-C18 5 μm 2.1 × 75 mm, flow rate 0.4 ml/min)

| Time (min) | Acetonitrile (%) | H <sub>2</sub> O (%) |
|------------|------------------|----------------------|
| 0          | 10               | 90                   |
| 1          | 10               | 90                   |
| 8          | 60               | 40                   |
| 12         | 90               | 10                   |
| 15         | 90               | 10                   |

Method A was used to record the protein data.

## 2. General procedures

### 2a. Expression of SUMO1 protein with N-terminus glycine

#### *Bacterial transformation*

*E. coli* strain [(DH5α for plasmid replication and BL21 (DE3) for protein expression] was used for transformation. The plasmid (1 μl) was added to the competent cells (50-100 μl) and was incubated on ice for 20 min. Subsequently, the heat shock was given at 42 °C for 40 seconds. The cells were kept on ice for 1 min, and 1 ml of LB was added to cells for recovery. The cells were incubated at 37 °C, 180 rpm for 45 min. The recovered cells were plated on LB plates containing desired antibiotics. The plates were incubated at 37 °C for 12-16hrs.

#### *Protein purification*

Primary culture was grown in LB overnight at 37 °C. 1% of primary culture was sub-cultured into desired volume of LB media as secondary culture. At approximately 0.6-0.8 OD (600 nm), the secondary culture was induced with IPTG (200 μM) for 4 h at 30 °C for SUMO1. The induced culture was spun at 8000 rpm for 10 min to pellet down cells and the pellet was stored at -80 °C.

For lysis, the cells were thawed on ice and resuspended in lysis buffer [20 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA, 5 mM β-ME]. Subsequently, 50 μg/ml lysozyme, 0.2% Triton X-100, 1X protease inhibitors 1 mM PMSF, Leupeptin, Pepstatin and Aprotinin mix, were added to facilitate cell lysis and protein stability. Lysate was incubated for 10-15 min in ice with constant shaking in between. This was followed by sonication (45% Amplitude, 10 sec ON 10 sec OFF cycle) till the suspension became clear. The supernatant was collected after spinning for 30 min at 11000 rpm, 4 °C.

For protein binding and elution, the supernatant was transferred to column containing washed GSH beads. The protein bead binding was facilitated at 4 °C on the tumbler for 1 h. The beads were washed thrice with wash buffer [20 mM Tris (pH 7.5), 400 mM NaCl, 1 mM EDTA, 5 mM β-ME]. The protein was eluted in elution buffer [20 mM Tris (pH 8.0), 150 mM NaCl, 1 mM EDTA, 20 mM glutathione] and the concentration of eluted protein was determined using Bradford assay.

For clipping, protein-bound beads were washed thrice with prescission protease buffer [50 mM Tris (pH 7.5), 1 mM EDTA, 1 mM DTT, 150 mM NaCl, 0.1% triton]. The bead bound protein was quantified

by Bradford method. In Prescission protease buffer, protein was clipped on beads using Prescission protease while maintaining the Prescission protease to total protein ratio 1:50. The clipping reaction proceeded at 4 °C for 18 h. The clipped proteins with N-terminus glycine were collected as supernatant, quantified and analyzed for their purity/stability on SDS-PAGE.

## **2b. Immunofluorescence**

HEK293T cells were grown on coverslips in a six-well plate format in Dulbecco's modified Eagle's medium (DMEM) and serum starved for 24 hrs in 1% serum containing DMEM medium. Subsequently, cells were treated with 100 nM Wortmannin for 30 minutes. After Wortmannin treatment, cells were washed twice with PBS and treated with either native insulin or purified insulin (3µg each) in 1 ml of 1% FBS containing DMEM media for 30 minutes. Post-treatment, cells were again washed twice with PBS and fixed using 100% chilled methanol for 15 min at 4 °C. The cells were then rehydrated and permeabilized with rehydration buffer (10 mM Tris, 150 mM NaCl, 0.1% TritonX-100) for 10 min. Cells were blocked with 5% Normal Goat Serum (NGS) for 30 min at 4 °C after rehydration. The cells were stained overnight with pAkt antibodies (1:200, anti pAKT-S473, CST #4060) at 4 °C. After primary antibody incubation, cells were washed three times with PBS-T (5 minutes each), and incubated with Alexa Fluor-568 conjugated goat anti-rabbit IgG (1:800, Life Technologies) for 1h. Cells were washed thrice with PBS-T, and mounted on slides using DAPI containing mounting medium (SIGMA F6057). Fluorescence signals were captured on Zeiss LSM 780 confocal microscope and images were analysed using Image J software.

## **2c. Western blotting**

Cells were cultured and treated as mentioned in immunofluorescence section. Post-treatment, cells were washed twice with PBS and lysed directly in the 1X Laemmli buffer (containing 1 mM Sodium orthovanadate and 1mM PMSF) by boiling at 100 °C for 10 minutes. The clear lysate were obtained by centrifugation and were run on 8% SDS-PAGE. Following wet transfer protocols, proteins were transferred onto the methanol activated 0.45µM PVDF membrane (GE Healthcare, Cat #10600029) using 1X transfer buffer (2.5 mM Tris-HCl pH 7.5, 19.2 mM glycine). The membrane was blocked for 1 h in 5% BSA. Further, the membrane was incubated overnight with pAkt antibodies (1:2000, CST #4060) and washed TBS-T buffer (20 mM Tris HCl pH 7.5, 150 mM NaCl, 0.1% Tween-20; 3 times for 10 minutes each). Membrane was incubated with Alexa fluor Plus 680 secondary antibodies (Invitrogen# A32734) for 1 hour and was washed with TBS-T buffer (3 times for 10 minutes each). Images were taken using LI-COR IR system. Anti-Crm1 antibodies were used for loading control detection.

## **2d. SUMOylation reaction**

SUMOylation reactions were carried out at 37 °C in the reaction buffer containing 50 mM Tris pH 7.5, 5 mM MgCl<sub>2</sub>, 5 mM DTT, 5 mM ATP. 0.25 µg of GST-SAE1/2 and 1 µg of (His)<sub>6</sub>-Ubc9. Reactions were initiated by the addition of 4µg each of SUMO1 and the standard substrate (GST-tagged

SUMOylation sequence bearing 11 a.a. long peptide) in the reaction. Samples were removed at 2, 4, and 6 hours intervals and mixed with laemmli buffer. Processed samples were run on 14% SDS-PAGE and detected by Coomassie staining and also by immunoblotting with anti-SUMO1 antibodies.

### **2e. Procedure for purification of insulin, myoglobin, and SUMO1**

N-hydroxy succinimidyl sepharose beads **4** (500  $\mu$ l, resin loading: 23  $\mu$ mol/ml) were taken in a 5 ml fritted polypropylene chromatography column with end tip closures. Sodium bicarbonate buffer (0.1 M, pH 7.8, 3 x 1 ml) was used to wash the beads and were re-suspended (sodium bicarbonate buffer, 360  $\mu$ l, 0.1 M, pH 7.8). To this solution, N-(3-(2-(2-(3-aminopropoxy)ethoxy)ethoxy)propyl)-2-(2-formylphenoxy)acetamide **2a** (13.8  $\mu$ M) in DMSO (40  $\mu$ l) from a freshly prepared stock solution was added and vortexed at 25 °C. The progress of the immobilization of the reagent on sepharose resin was monitored (2 h) by UV-absorbance of the supernatant. Subsequently, the supernatant was removed and the beads were washed with aqueous buffer (0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0, 3 x 1 ml; 0.1 M acetate/ 0.5 M NaCl pH 4.0, 3 x 1 ml) and H<sub>2</sub>O (3 x 1 ml) to remove the adsorbed reagent from resin (store at 4 °C). The sepharose beads **5a** were washed with aqueous buffer (0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0, 3 x 1 ml) and re-suspended (aqueous buffer, 375  $\mu$ l of 0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0). To this solution, insulin (20 nmol), myoglobin (20 nmol) and SUMO1 (20 nmol) dissolved in aqueous buffer (25  $\mu$ l, 0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0) was added and vortexed at 25 °C for 18 h. Binding was ensured using UV-Vis analysis. The beads were washed thoroughly with aqueous buffer (0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0, 3 x 1 ml), 1 N KCl (3 x 1 ml) and H<sub>2</sub>O (3 x 1 ml) to remove any non-specifically bound protein from the resin. This was confirmed by analysing the final wash fraction using LC-MS. For eluting out the bound protein, pyridoxal 5'-phosphate **11i** (50 equiv.) in 0.1 M NaHCO<sub>3</sub> buffer, pH 7.8) was added to the resin and vortexed for 2 h at 25 °C. Hydroxyl amine was added for rapid removal of the excess PLP bound to the proteins through imine formation before a final round of spin concentration. The eluted protein was analysed by using ESI-MS and SDS-PAGE.

### **2f. Procedure for purification of expressed SUMO1 from cell lysate**

Gly-tag purification resin was synthesized according the procedure previously described (section 2b). The sepharose beads **5a** were washed with aqueous buffer (0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0, 3 x 1 ml) and re-suspended (aqueous buffer, 375  $\mu$ l of 0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0). To this solution, 500  $\mu$ l of cell lysate overexpressed with SUMO1 (the cell lysate was prepared according to the procedure described in section 2a, protein concentration was determined by Bradford assay and adjusted to 1 mg/ mL) was added and vortexed at 25 °C for 24 h. The beads were washed thoroughly washed with aqueous buffer (0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0, 3 x 1 ml), 1 N KCl (3 x 1 ml) and H<sub>2</sub>O (3 x 1 ml) to remove any non-specifically bound protein from the resin. For eluting out the bound protein, pyridoxal 5'-phosphate **11i** (50 equiv.) in 0.1 M NaHCO<sub>3</sub> buffer (pH 7.8) was added to the resin and vortexed for 2 h at 25 °C. Hydroxyl amine was added for rapid removal of the excess PLP bound to the

proteins through imine formation, before a final round of spin concentration. The eluted protein was analysed by using SDS-PAGE.

## **2g. Procedure for purification of insulin from mixture of proteins and cell lysate**

Gly-tag purification resin was synthesized according the procedure previously described (section 2b). The sepharose beads **5a** were washed with aqueous buffer (0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0, 3 x 1 ml) and re-suspended (aqueous buffer, 375 µl of 0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0). To this solution, 20 nmol of each protein (insulin, aprotinin, ubiquitin, cytochrome C, lysozyme C, β-lactoglobulin, and chymotrypsinogen A), dissolved in aqueous buffer (25 µl, 0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0) was added and vortexed at 25 °C for 18 h. In case of cell lysate (cell lysate was prepared according to the procedure described in section 2a) premixed native insulin (20 nmol) in cell lysate (500 µg) dissolved in aqueous buffer (300 µl, 0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0) was added and vortexed at 25 °C for 18 h. The beads were washed thoroughly with aqueous buffer (0.1 M NaHCO<sub>3</sub>/ 0.5 M NaCl pH 8.0, 3 x 1 ml), 1 N KCl (3 x 1 ml) and H<sub>2</sub>O (3 x 1 ml) to remove any non-specifically bound protein from the resin. This was confirmed by analyzing the final wash fraction using LC-MS. For eluting out the bound protein, pyridoxal 5'-phosphate **11i** (50 equiv.) in 0.1 M NaHCO<sub>3</sub> buffer, pH 7.8) was added to the resin and vortexed for 2 h at 25 °C. The eluted protein was analysed by using ESI-MS and SDS-PAGE.

## **2h. Labeling of protein in solution phase<sup>1</sup>**

In a 1.5 ml Eppendorf tube, protein **1a** (3 nmol) was mixed with sodium bicarbonate buffer (120 µl, 0.1 M, pH 7.8). To this solution, N,N'-(((oxybis(ethane-2,1-diyl))bis(oxy))bis(propane-3,1-diyl))bis(2-(2-formylphenoxy)acetamide) **S8** (1500 nmol) in DMSO (30 µl) from a freshly prepared stock solution was added and vortexed at 25 °C. The overall concentration of protein and N,N'-(((oxybis(ethane-2,1-diyl))bis(oxy))bis(propane-3,1-diyl))bis(2-(2-formylphenoxy)acetamide) **S8** was 20 µM and 10 mM respectively. After 24-48 h, the reaction mixture was diluted with acetonitrile:water (10:90, 3000 µl). Unreacted 2-(2-formylphenoxy)acetic acid and salts were removed by using Amicon® Ultra-0.5 mL 3-kDa or 10-kDa MWCO centrifugal filters spin concentrator. The protein mixture was further washed with Millipore Grade I water (5 × 0.4 ml). The sample was analyzed by ESI-MS. The aqueous sample was concentrated by lyophilization before subjecting it to digestion, peptide mapping, and sequencing by MS-MS.

## **2i. Calculation of concentration of sample using spectrophotometer**

The spectrophotometer works on the principle of Beer-Lambert law which states that the absorbance of a material sample is directly proportional to the path length and the concentration of the material sample. Mathematically, it is given by the equation

$$A = \epsilon bc$$

A - Absorbance of the sample

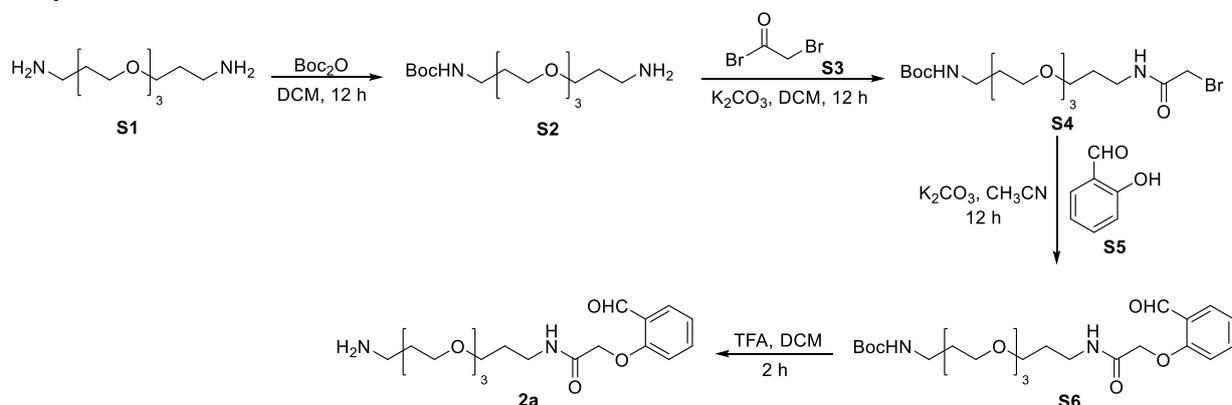
ε - Molar absorptivity co-efficient

b - Path length (length of cuvette)

c - Concentration of the sample

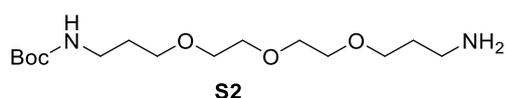
Thus, by keeping the path length constant (by using a 1 cm cuvette), concentration of the material sample can be calculated.

### 3. Synthesis and characterization data



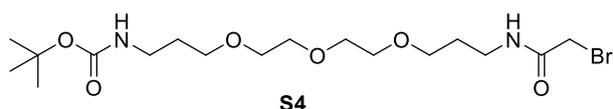
**Figure S1:** Synthesis of N-(3-(2-(2-(3-aminopropoxy)ethoxy)ethoxy)propyl)-2-(2-formylphenoxy)acetamide **2a**<sup>1</sup>

#### Synthesis of tert-butyl (3-(2-(2-(3-aminopropoxy)ethoxy)ethoxy)propyl)carbamate (**S2**)



A solution of 4,7,10-trioxa-1,13-tridecanediamine **S1** (34.1 mmol, 7.50 g) in 250 ml round bottom flask and dissolved in DCM (100 mL) followed by slow addition of Boc anhydride (16.9 mmol, 3.70 g). The reaction mixture was stirred at room temperature overnight. The solvent was removed under reduced pressure and the residue was purified by silica gel chromatography (MeOH: CHCl<sub>3</sub> 3:97) to afford tert-butyl (1-bromo-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate **S2** (34% yield, 1.86 g). TLC (MeOH: CHCl<sub>3</sub> 10:90), <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 3.63-3.60 (m, 4H), 3.61-3.56 (m, 4H), 3.56-3.52 (m, 4H), 3.22 (d, *J* = 6.0 Hz, 2H), 2.79 (t, *J* = 6.7 Hz, 2H), 1.79-1.69 (m, 4H), 1.45 (s, 9H) ppm. <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 156.0, 78.7, 70.5, 70.5, 70.2, 70.1, 69.5, 69.4, 39.5, 38.4, 33.3, 29.5, 28.4 ppm. MS (ESI) [M+H]<sup>+</sup> calcd. C<sub>15</sub>H<sub>32</sub>N<sub>2</sub>O<sub>5</sub> 321.2, found 321.1.

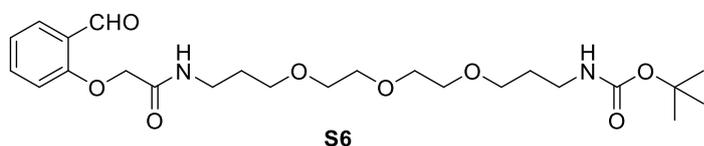
#### Synthesis of tert-butyl (1-bromo-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate (**S4**)



Tert-butyl (3-(2-(2-(3-aminopropoxy)ethoxy)ethoxy)propyl)carbamate **S2** (4.7 mmol, 1.5g) was dissolved in DCM (3 ml), in a 50 ml round bottom flask and K<sub>2</sub>CO<sub>3</sub> (7 mmol, 1 g) in 3 ml of H<sub>2</sub>O was added to it. Bromoacetyl bromide **S3** (7 mmol, 1.4 g), dissolved in DCM (3 ml), was added drop wise

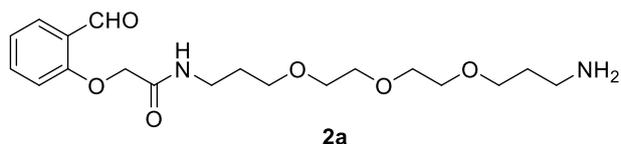
to the mixture at 0-5 °C. The reaction mixture was stirred for 12 h and the progress of the reaction was analyzed by using thin layer chromatography. Upon completion, the reaction mixture was extracted with DCM and the solution was concentrated under vacuum. The product was purified using silica gel column chromatography (MeOH: CHCl<sub>3</sub> 3:97) to afford tert-butyl (1-bromo-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate **S4** (75% yield, 1.5 g). TLC (MeOH : DCM 10:90), <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 3.84 (s, 2H), 3.68-3.65 (m, 2H), 3.64-3.56 (m, 8H), 3.54 (t, *J* = 6.0 Hz, 2H), 3.45-3.37 (m, 2H), 3.26-3.03 (m, 2H), 1.78-1.83 (m, 2H), 1.77-1.71 (m, 2H), 1.42 (s, 9H) ppm. <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 165.5, 156.0, 78.7, 70.5, 70.5, 70.2, 70.1, 69.5, 69.4, 39.5, 38.4, 33.3, 29.5, 28.4 ppm (One aliphatic carbon overlap). MS (ESI) [M+H]<sup>+</sup> calcd. C<sub>17</sub>H<sub>33</sub>Br<sup>79</sup>N<sub>2</sub>O<sub>6</sub> 441.1, found 441.0 and calcd. C<sub>17</sub>H<sub>33</sub>Br<sup>81</sup>N<sub>2</sub>O<sub>6</sub> 443.1, found 443.0.

### Synthesis of tert-butyl (1-(2-formylphenoxy)-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate (**S6**)



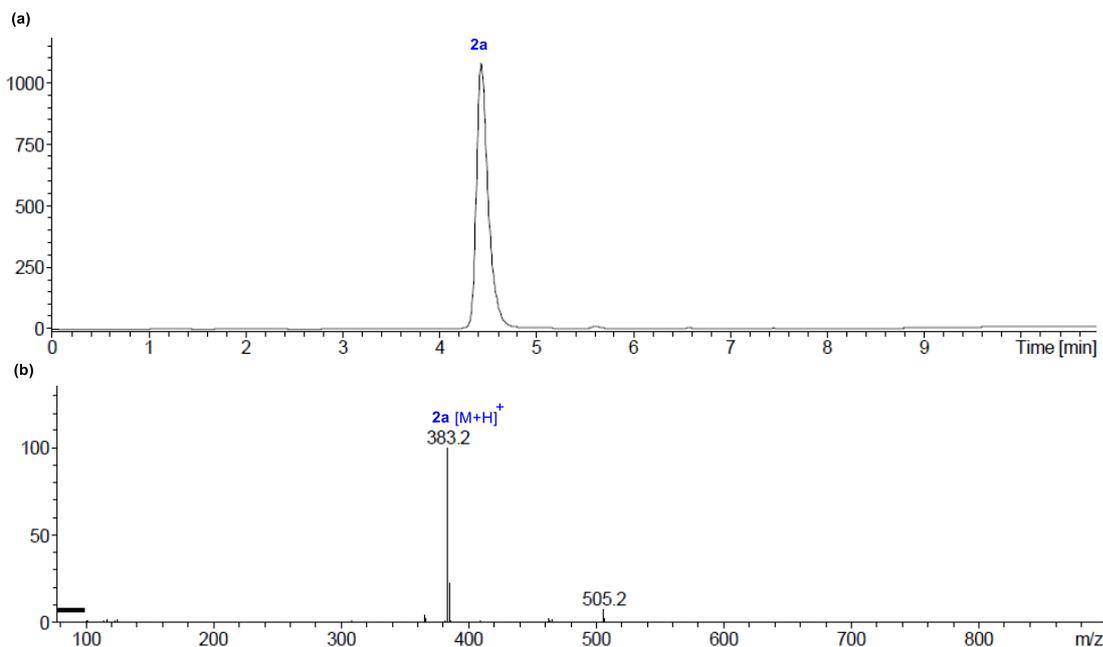
In a 50 ml round bottom flask, 2-hydroxybenzaldehyde **S5** (4.1 mmol, 500 mg) was dissolved in acetonitrile (8 ml). To this solution, K<sub>2</sub>CO<sub>3</sub> (5.2g, 37.7 mmol) and tert-butyl (1-bromo-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate **S4** (6 mmol, 828 mg) were added and the reaction mixture was allowed to reflux for 12 h. The reaction was monitored using thin layer chromatography. Upon completion, the reaction mixture was filtered to remove potassium carbonate. The solution was concentrated under vacuum and the product was purified using silica gel column chromatography (MeOH: DCM 5:95) to afford tert-butyl (1-(2-formylphenoxy)-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate **S6** (53% yield, 1.0 g). TLC (MeOH : DCM 10:90), <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 10.25 (s, 1H), 7.79 (dd, *J* = 7.6, 1.8 Hz, 1H), 7.66 (bs, 1H), 7.52-7.55 (m, 1H), 7.19-7.13 (m, 1H), 6.97-6.90 (m, 1H), 4.57 (s, 2H), 3.61-3.57 (m, 8H), 3.55-3.53 (m, 2H), 3.51-3.46 (m, 4H), 3.30-3.12 (m, 2H), 1.90-1.85 (m, 2H), 1.78-1.68 (m, 2H), 1.42 (s, 9H) ppm. <sup>13</sup>C NMR (125MHz, CDCl<sub>3</sub>) δ 190.1, 167.5, 158.4, 156.2, 136.2, 133.0, 125.2, 122.1, 113.2, 79.0, 70.6, 70.6, 70.4, 70.3, 69.7, 69.5, 67.8, 38.7, 37.3, 29.8, 29.4, 28.6 ppm. MS (ESI) [M+H]<sup>+</sup> calcd. C<sub>24</sub>H<sub>38</sub>N<sub>2</sub>O<sub>8</sub> 483.2, found 483.1.

### Synthesis of N-(3-(2-(2-(3-aminopropoxy)ethoxy)ethoxy)propyl)-2-(2-formylphenoxy)acetamide **2a**

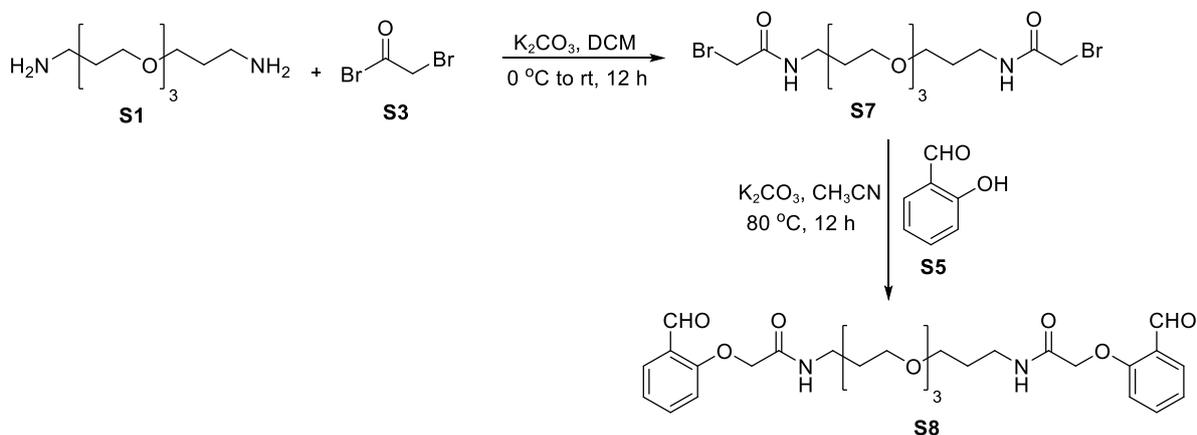


**Procedure:** In a 25 ml round bottom flask, tert-butyl (1-(2-formylphenoxy)-2-oxo-7,10,13-trioxa-3-azahexadecan-16-yl)carbamate **S6** (500 mg, 1.3 mmol), was mixed with dichloromethane (3 ml). To this solution trifluoro acetic acid (1 ml) was added drop wise at 0-5 °C. The reaction mixture was

allowed to stir for 2 h. The reaction was monitored using thin layer chromatography and upon completion of the reaction, the solution was concentrated under vacuum to afford N-(3-(2-(2-(3-aminopropoxy)ethoxy)ethoxy)propyl)-2-(2-formylphenoxy)acetamide **2a** (90% yield, 450 mg). Note: The oligomeric imine formation in deuterated solvent leads to complex NMR spectra. However LC and MS confirms the purity. HRMS (ESI)  $[M+H]^+$  calcd.  $C_{19}H_{30}N_2O_6$  383.2182, found 383.2192.

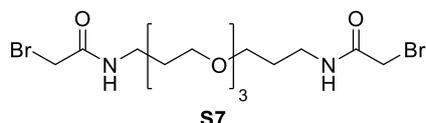


**Figure S2:** (a) HPLC spectrum of **2a**. (b) ESI-MS spectrum of **2a**.



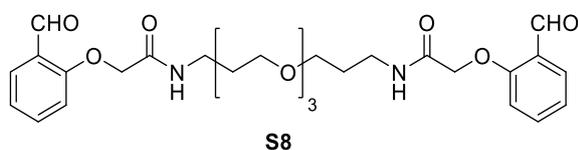
**Figure S3:** Synthesis of N,N'-(((oxybis(ethane-2,1-diyl))bis(oxy))bis(propane-3,1-diyl))bis(2-(2-formylphenoxy)acetamide) **S8**<sup>1</sup>

### Synthesis of N,N'-(((oxybis(ethane-2,1-diyl))bis(oxy))bis(propane-3,1-diyl))bis(2-bromoacetamide) (S7)



**Procedure:** 4,7,10-Trioxa-1,13-tridecanediamine **S1** (9 mmol, 2 g) was dissolved in DCM (50 ml), in a 250 ml round bottom flask and  $K_2CO_3$  (30 mmol, 4.1 g) in 20 ml of  $H_2O$  was added to it. Bromoacetyl bromide **S3** (30 mmol, 6 g), dissolved in 20 ml of DCM was added drop wise to the mixture at 0-5 °C. The reaction mixture was stirred for 12 h and the reaction progress was analyzed using thin layer chromatography. On completion of the reaction, reaction mixture was extracted with DCM. The collected organic fractions were dried over anhydrous sodium sulfate and filtered, the filtrate was concentrated under reduced pressure to afford N,N'-(((oxybis(ethane-2,1-diyl))bis(oxy))bis(propane-3,1-diyl))bis(2-bromoacetamide) **S7**. Yield 72%; TLC (MeOH: DCM 10:90),  $^1H$  NMR (500 MHz,  $CDCl_3$ )  $\delta$  3.85 (s, 4H), 3.68-3.64 (m, 4H), 3.64-3.56 (m, 8H), 3.41-3.39 (m, 4H), 1.81-1.78 (m, 4H) ppm.  $^{13}C$  NMR (125 MHz,  $CDCl_3$ )  $\delta$  165.4, 70.5, 70.3, 70.3, 38.9, 29.3, 28.5 ppm. HRMS (ESI)  $[M+H]^+$  calcd.  $C_{14}H_{27}Br_2N_2O_5$  463.0287, found 463.0267.

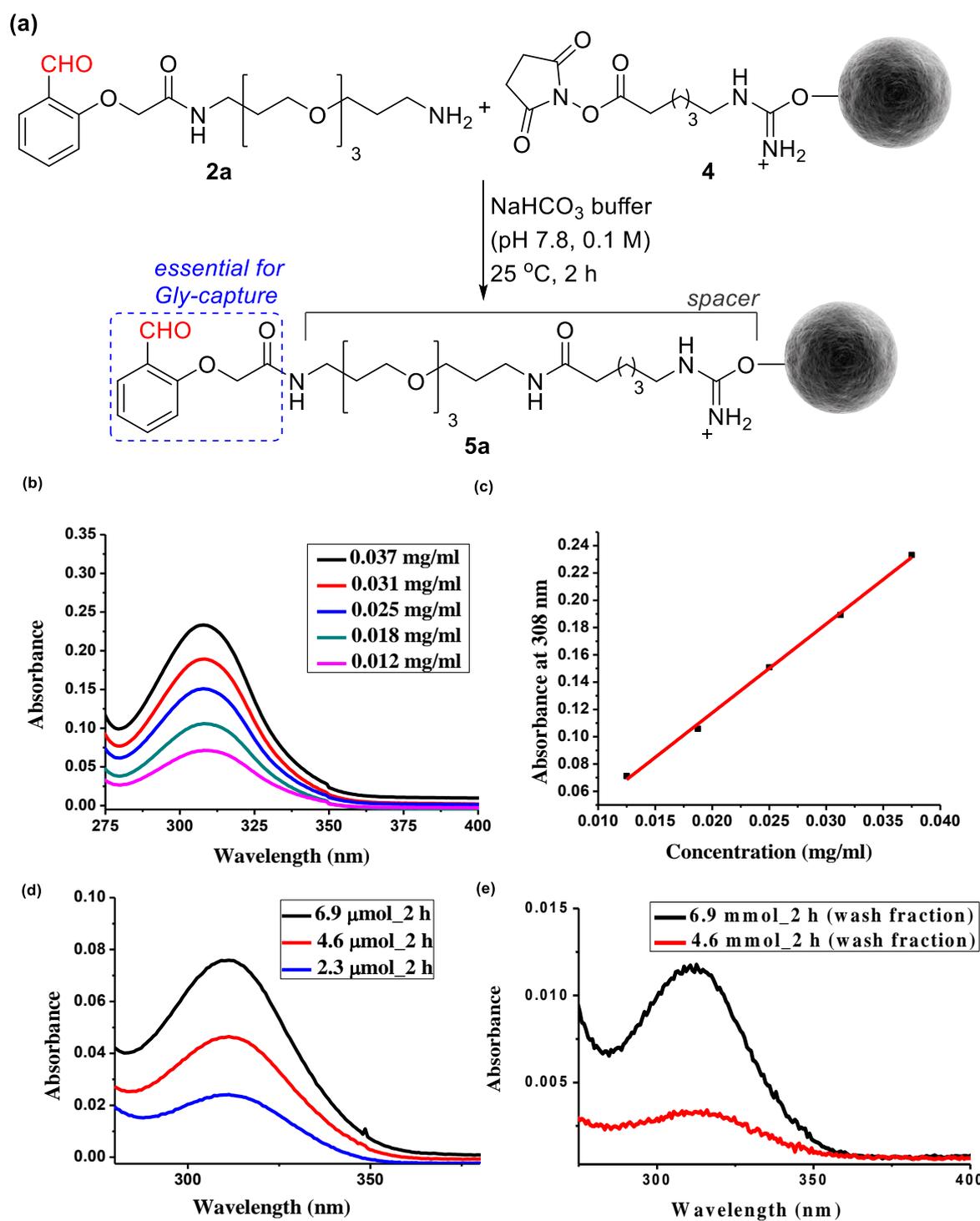
### Synthesis of N,N'-(((oxybis(ethane-2,1-diyl))bis(oxy))bis(propane-3,1-diyl))bis(2-(2-formylphenoxy)acetamide) (S8)



**Procedure:** In a 250 ml round bottom flask, 2-hydroxy benzaldehyde **S5** (4.6 g, 37.7 mmol), N,N'-(((oxybis(ethane-2,1-diyl))bis(oxy))bis(propane-3,1-diyl))bis(2-bromoacetamide) **S7** (3 g, 7 mmol) and  $K_2CO_3$  (5.2 g, 37.7 mmol) were dissolved in acetonitrile (75 ml). The reaction mixture was allowed to reflux for 12 h. The reaction was monitored using thin layer chromatography and upon completion, the reaction mixture was filtered to remove potassium carbonate. The solution was concentrated under vacuum and the product was purified using flash column chromatography (MeOH:DCM 3:97) to afford N,N'-(((oxybis(ethane-2,1-diyl))bis(oxy))bis(propane-3,1-diyl))bis(2-(2-formylphenoxy)acetamide) **S8**. Yield 73%; MeOH : DCM 10:90,  $^1H$  NMR (500 MHz,  $CDCl_3$ )  $\delta$  10.25 (s, 2H), 7.77-7.75 (m, 2H), 7.66 (bs, 2H), 7.59-7.53 (m, 2H), 7.12-7.09 (m, 2H), 6.92-6.88 (m, 2H), 4.55 (s, 4H), 3.56-3.51 (m, 12H), 3.48-3.42 (m, 4H), 1.87-1.80 (m, 4H) ppm.  $^{13}C$  NMR (125 MHz,  $CDCl_3$ )  $\delta$  190.0, 167.3, 158.4, 136.1, 132.6, 125.0, 121.9, 113.0, 70.4, 70.2, 69.3, 67.6, 37.0, 29.2 ppm. HRMS (ESI)  $[M+H]^+$  calcd.  $C_{28}H_{36}N_2O_9$  545.2499, found 545.2508.

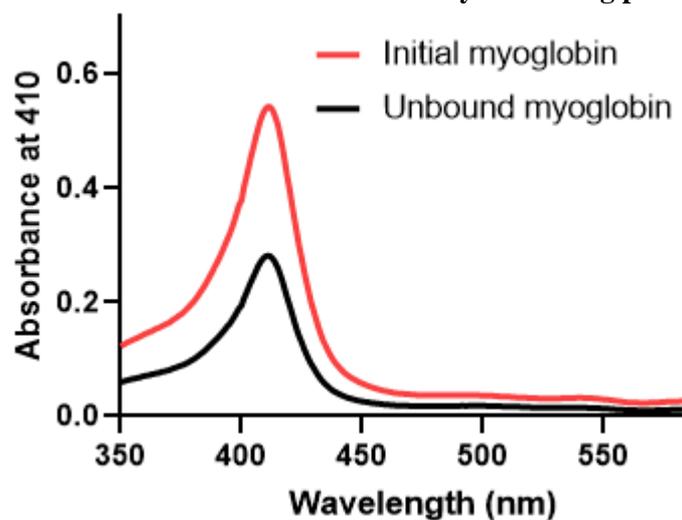
## 4. Experimental data for protein purification

### 4a. Immobilization of reagent (2a) on NHS sepharose resin

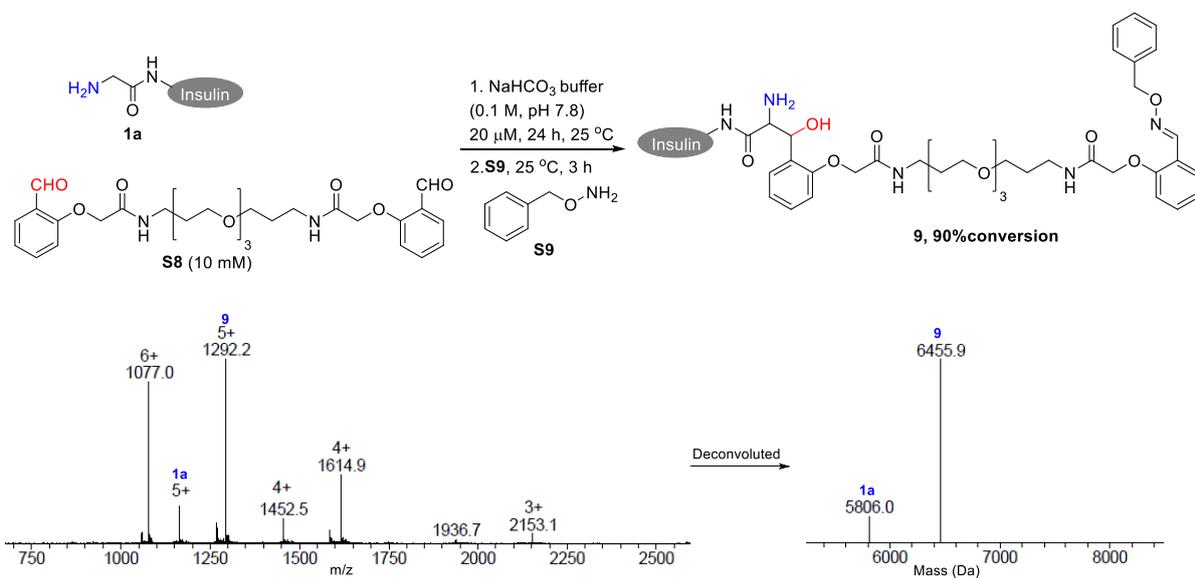


**Figure S4:** (a) Immobilization of reagent (2a) on NHS sepharose resin; (b) UV spectra of reagent 2a at different concentrations; (c) Determination of molar extinction coefficient for the reagent 2a; (d) UV spectra of the eluted fraction containing unbound reagent after 2 h; (e) UV spectra of first wash fraction which has the adsorbed reagent.

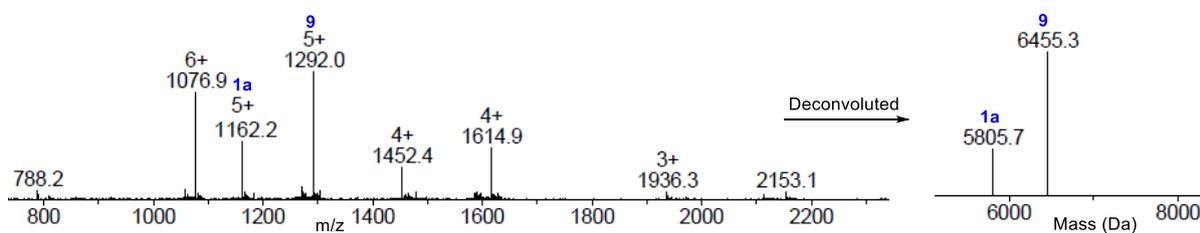
#### 4b. Immobilization of N-terminus Gly containing proteins on modified sepharose resin 5a



**Figure S5:** UV spectrum of myoglobin before and after immobilization on functionalized sepharose resin.

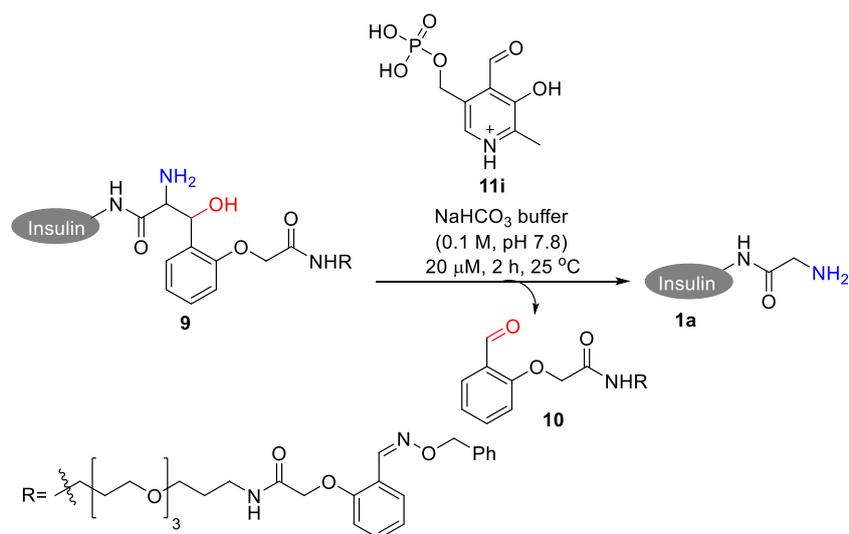


**Figure S6:** ESI-MS spectra of N-terminus Gly labeling.



**Figure S7:** ESI-MS spectra of released insulin with TMG (**11h**) reagent.

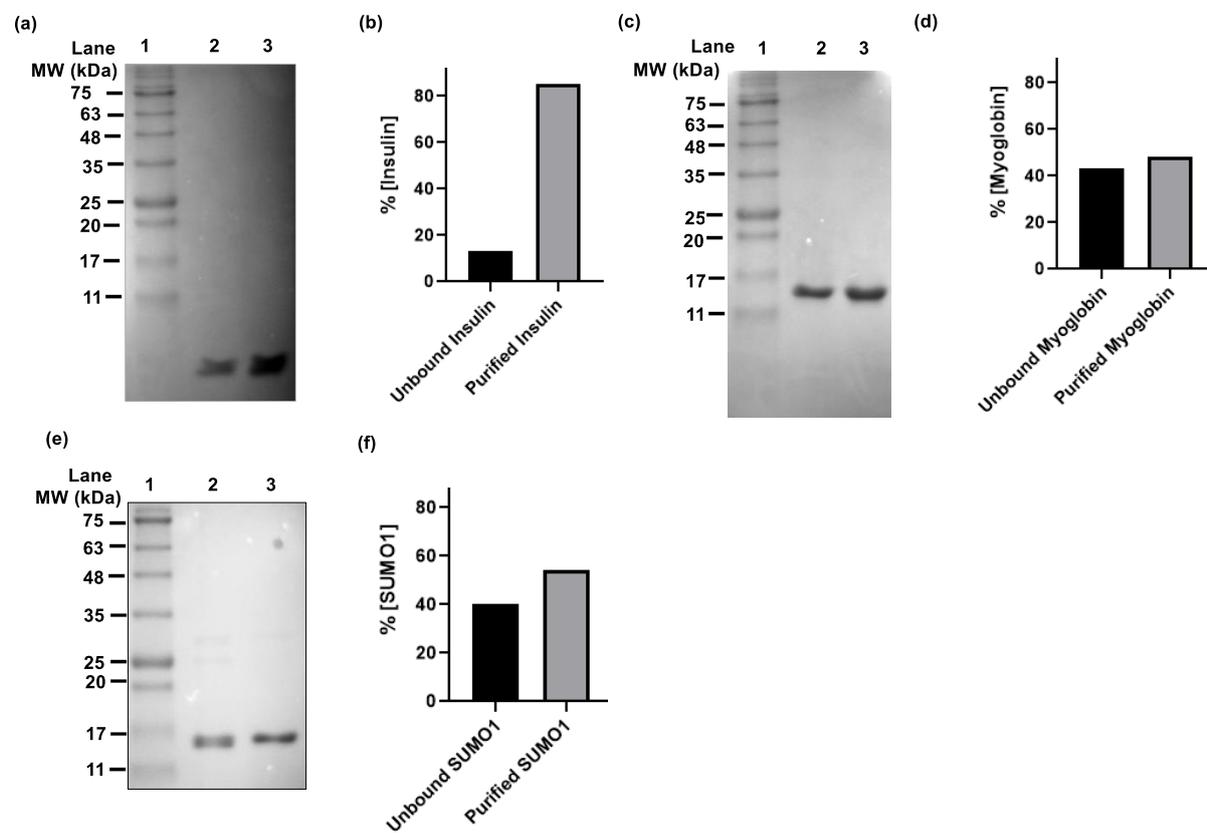
**Table S1.** Optimization of retro-aldol reaction in solution phase with PLP **11i**

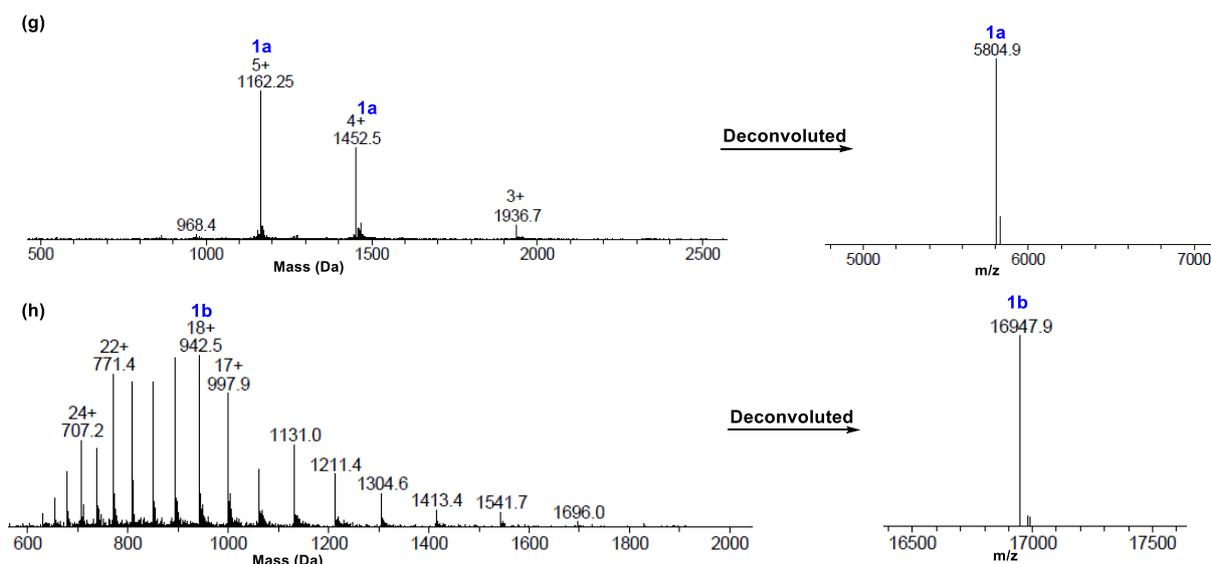


| Entry | PLP <b>11i</b> (equiv.) | %Conversion <b>1a</b> <sup>a</sup> |
|-------|-------------------------|------------------------------------|
| 1     | 10                      | 50                                 |
| 2     | 20                      | >99                                |
| 3     | 50                      | >99                                |

<sup>a</sup> %Conversions were monitored by ESI-MS.

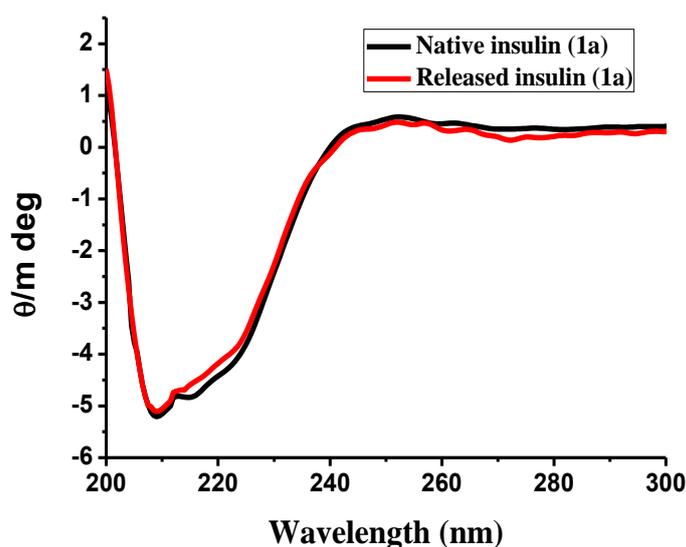
#### 4c. Release of the bound protein from modified sepharose resin





**Figure S8:** (a) Purified insulin visualized on 14% SDS-PAGE; lanes 1:2:3 are ladder, unbound insulin, and purified insulin. (b) Relative quantification of unbound and purified insulin determined from SDS-PAGE using Gel Quant. (c) Purified myoglobin visualized on 12% SDS-PAGE; lane 1:2:3 are ladder, unbound myoglobin, and purified myoglobin respectively. (d) Relative quantification of unbound myoglobin and purified myoglobin determined from SDS-PAGE using Gel Quant. (e) Purified SUMO1 visualized on 14% SDS-PAGE; lane 1:2:3 are ladder, unbound SUMO1, and purified SUMO1 respectively. (f) Relative quantification of unbound SUMO1 and purified SUMO1 determined from SDS-PAGE using Gel Quant. (g) ESI-MS spectrum of released insulin **1a**. (h) ESI-MS spectrum of released myoglobin **1b**.

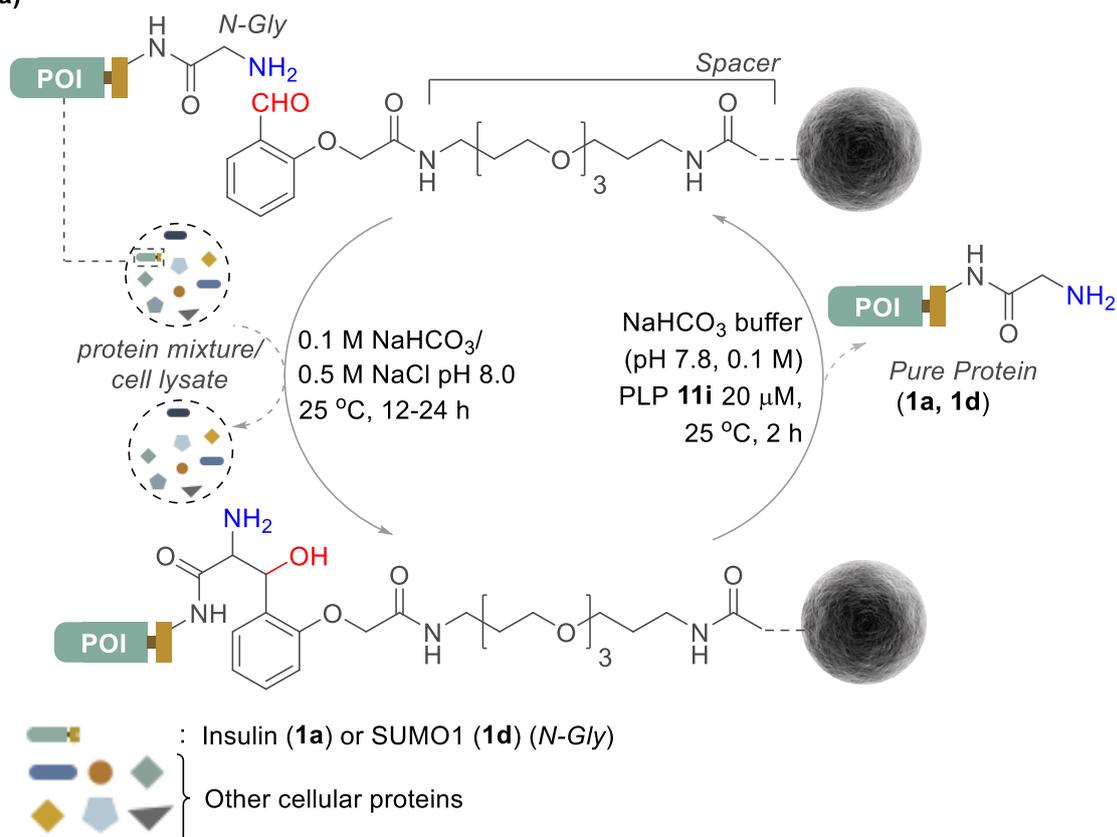
#### 4d. CD data of released insulin **1a**



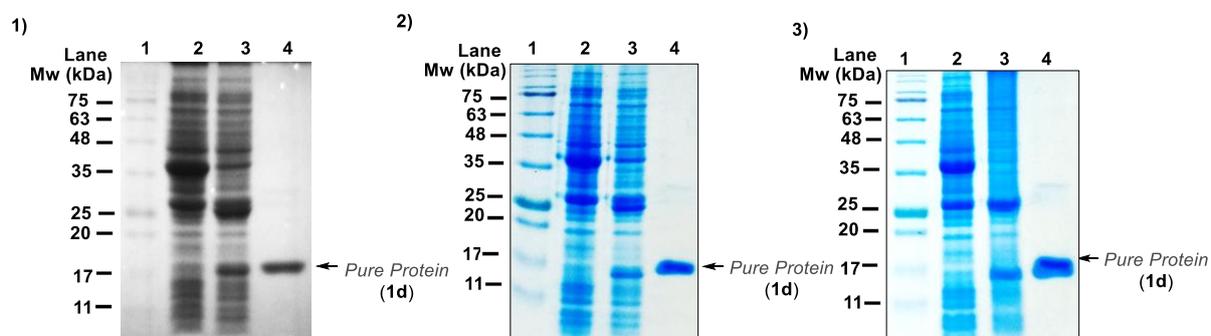
**Figure S9:** Circular dichroism (CD) spectra of Insulin **1a** (black line), released insulin **1a** in water at at concentration 0.1 mg/ml. The concentration of released protein is kept constant with respect to the native protein.

#### 4e. Purification of protein of interest from a mixture of proteins and cell lysate

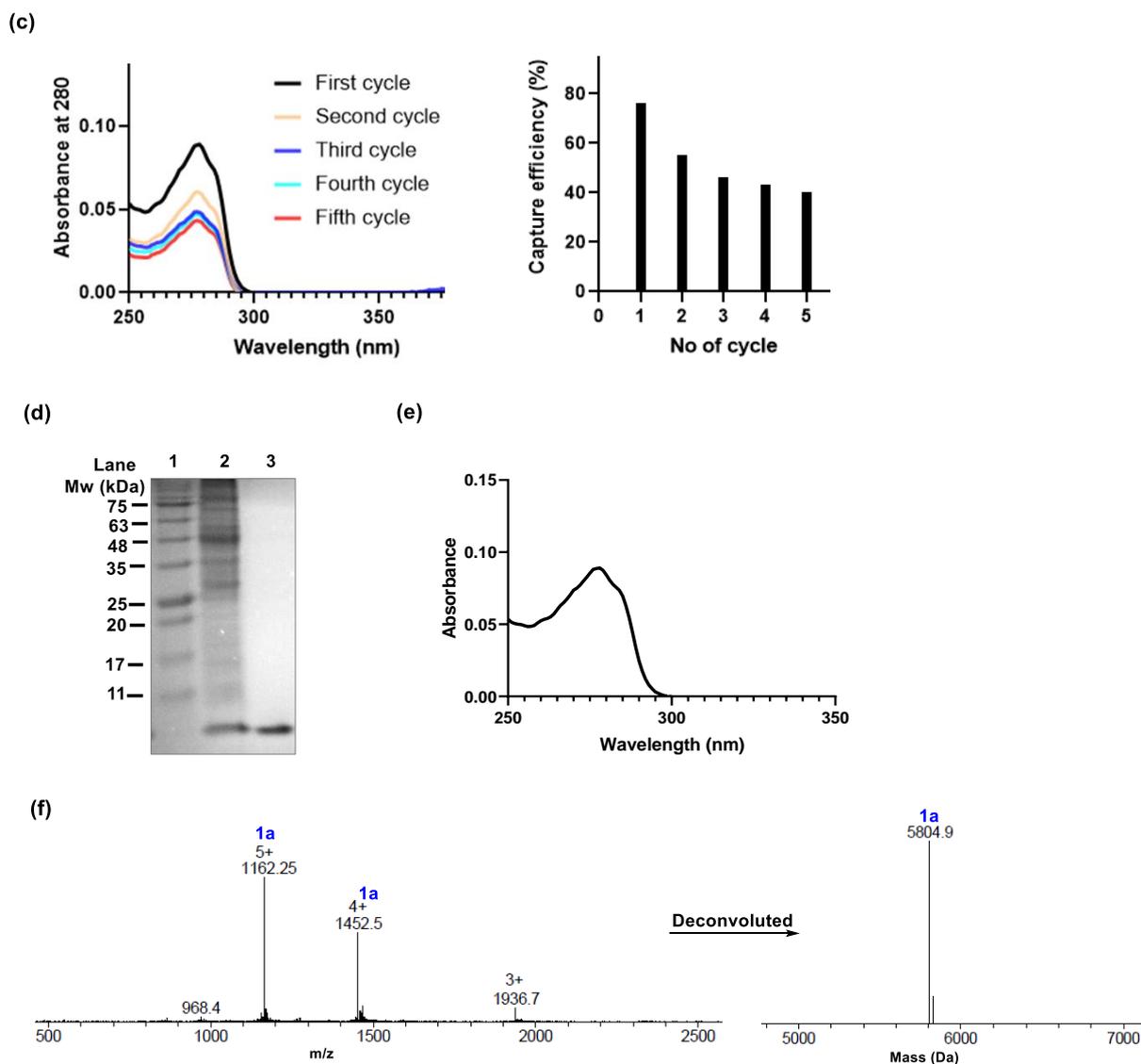
(a)



(b)

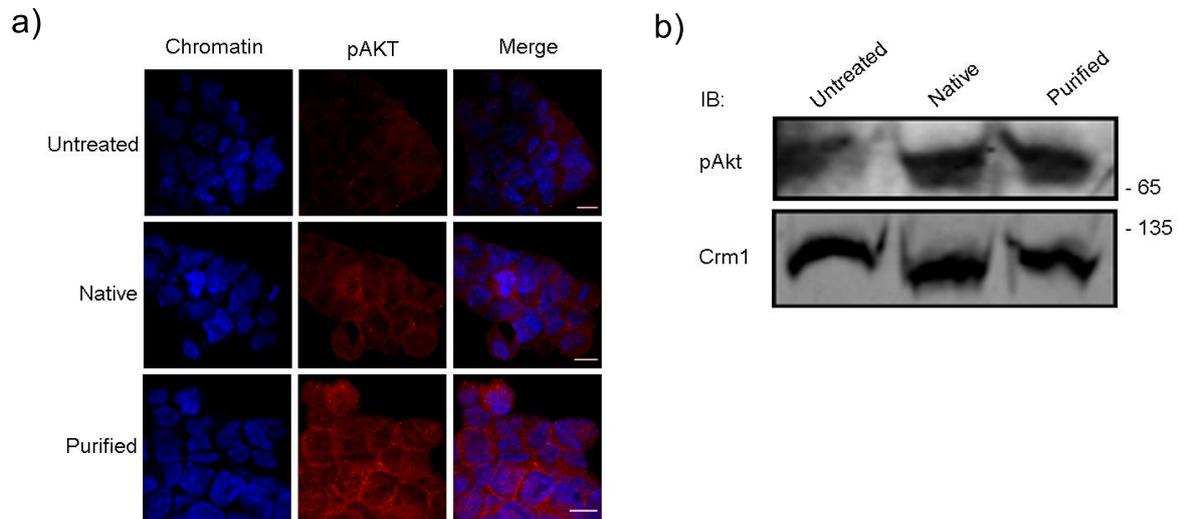


| Trial | % SUMO1<br><i>in cell lysate</i> | % SUMO1<br><i>purified by Gly-tag resin</i> | Fold Enrichment | % Recovery Yield |
|-------|----------------------------------|---|-----------------|------------------|
| a     | 1.58                             | 100   | 63              | 39               |
| b     | 3.38                             | 100   | 30              | 43               |
| c     | 5.60                             | 100   | 18              | 50               |



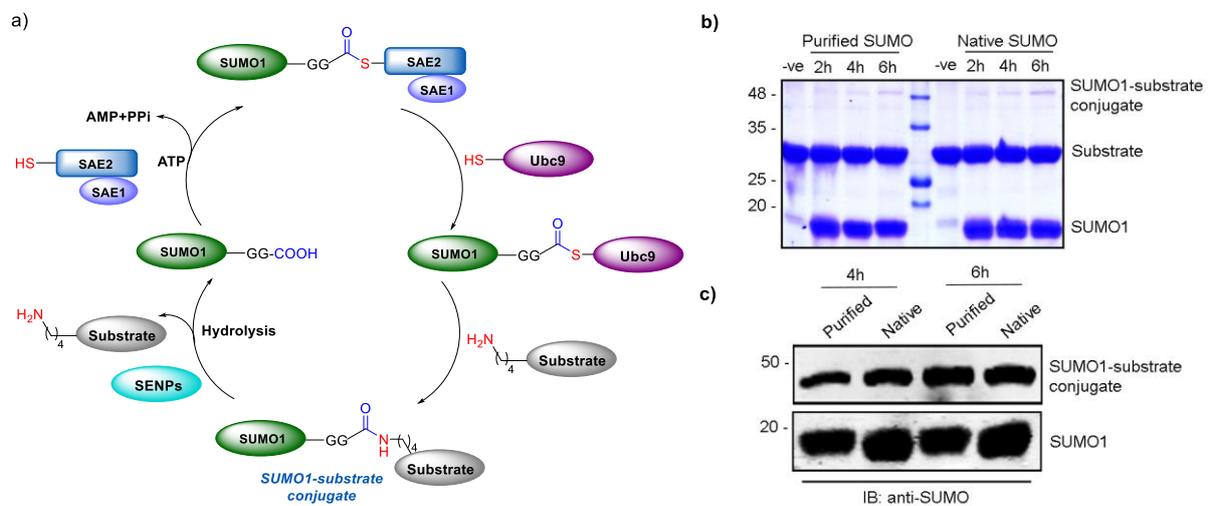
**Figure S10:** (a) Purification of protein of interest from a mixture of proteins and cell lysate. (b) Purified of over-expressed recombinant SUMO1 from cell lysate visualized on 14% SDS-PAGE; lanes 1:2:3:4 are ladder: GST-SUMO1: GST-clipped SUMO1 in cell lysate: purified SUMO1, respectively. (c) Recycling efficiency of the Gly-tag purification resin from cell lysate. (d) Purified insulin from cell lysate visualized on 12% SDS-PAGE; 1:2:3 are ladder, insulin in cell lysate, and purified insulin respectively. (e) UV-visible spectrum of cleaved insulin **1a** from the cell lysate. (f) ESI-MS spectrum of cleaved insulin **1a** from the mixture of proteins.

#### 4f. In vitro activity of purified insulin



**Figure: S11** In vitro activity of purified insulin. a) Enhancement of insulin receptor-mediated signaling in cells treated with native and purified insulin. Immunofluorescence mediated detection of subcellular pAkt (red) in cells. b) pAkt levels by western blotting (scale bar: 10  $\mu$ m).

#### 4g. Activity analysis of purified SUMO1

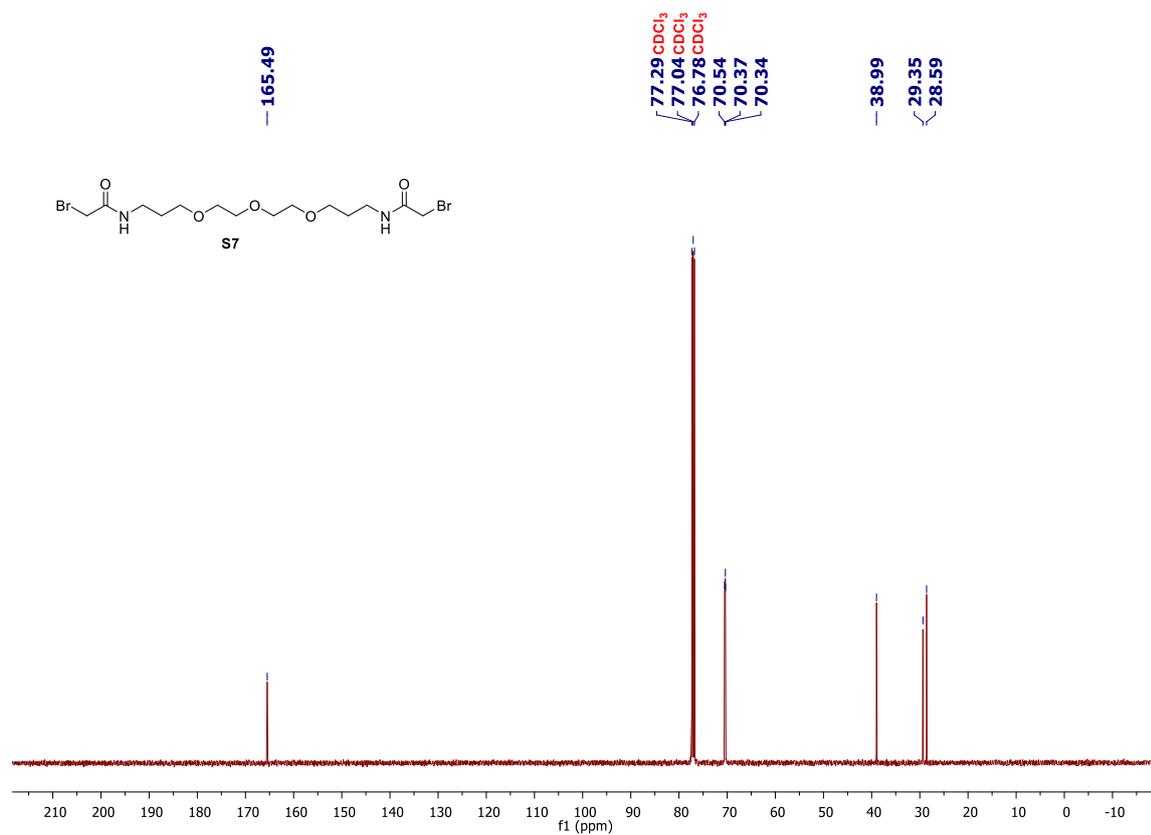
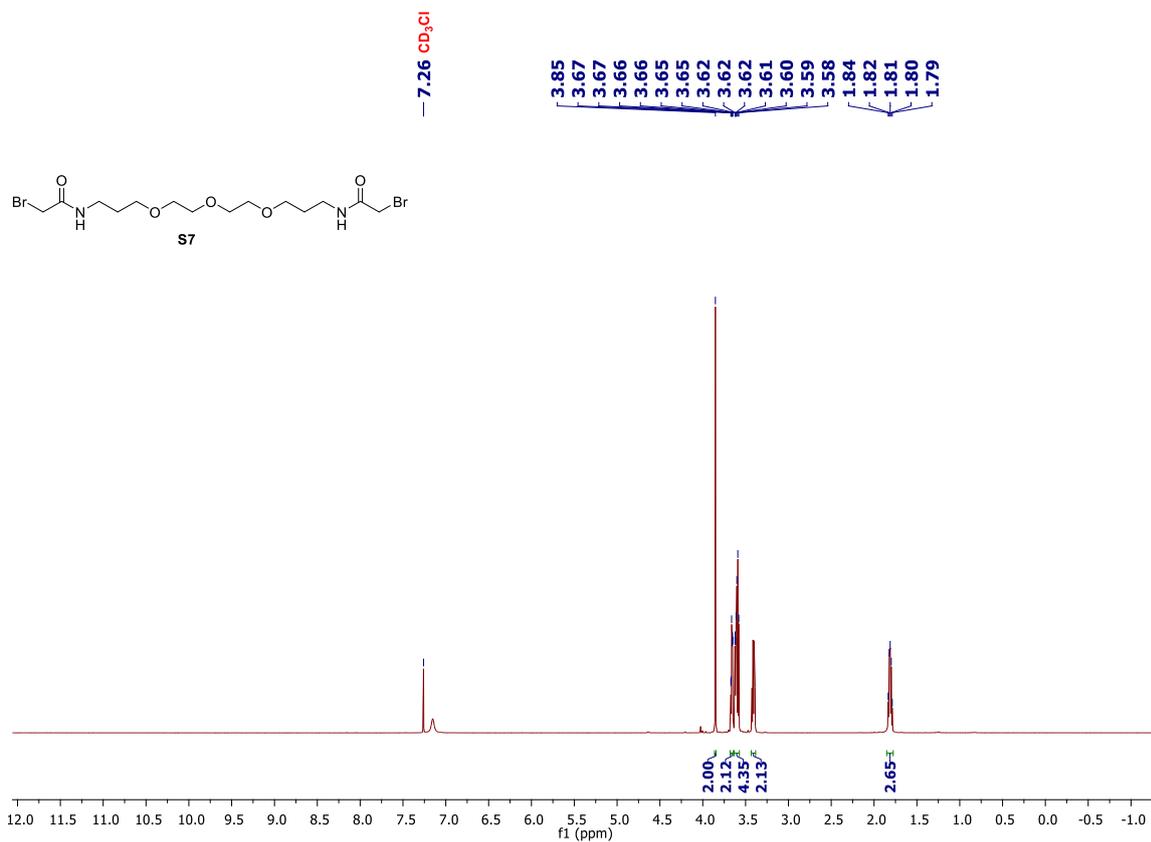


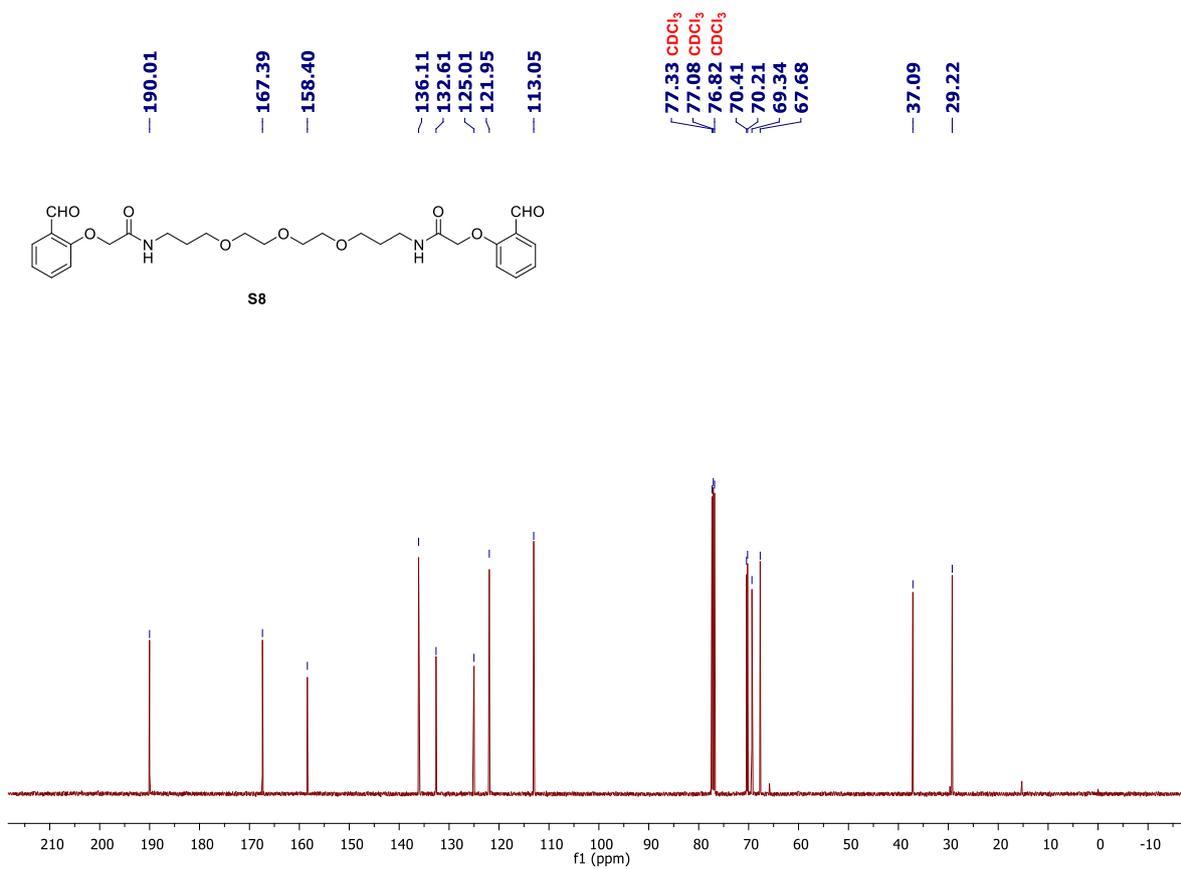
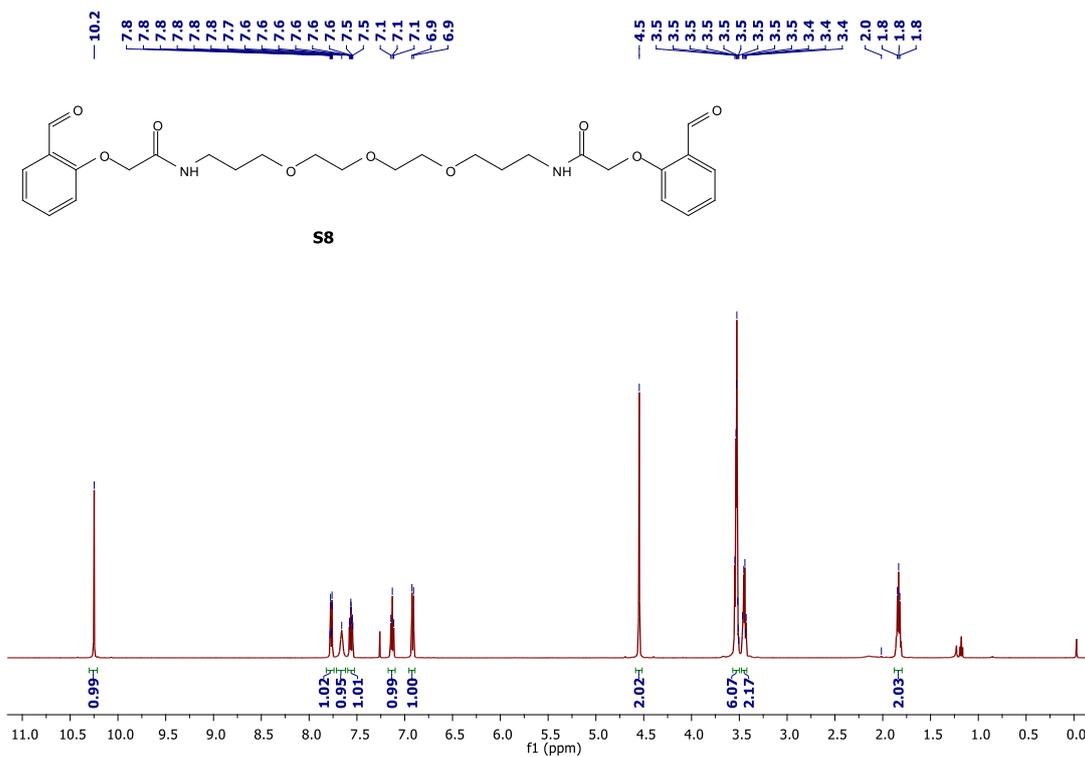
**Figure. S12** Activity analysis of purified SUMO1. a) Schematic representation of the steps in the multiprotein SUMOylation cycle. b) Activity analysis of purified SUMO1 in an *in vitro* SUMOylation reaction with purified components. c) Western blot analysis of SUMOylation reaction as in (b), where bands ~50 kDa in the upper panel indicate SUMO1-substrate conjugate, and the lower panel ~18 kDa bands indicate unconjugated SUMO1.











## 6. References

- 1 L. Purushottam, S. R. Adusumalli, U. Singh, V. B. Unnikrishnan, D. G. Rawale, M. Gujrati, R. K. Mishra, V. Rai, *Nat. Commun.*, 2019, **10**, 2539.
- 2 S. R. Adusumalli, D. G. Rawale, K. Thakur, L. Purushottam, N. C. Reddy, N. Kalra, S. Shukla, V. Rai, *Angew. Chem. Int. Ed.* **2020**, *59*, 10332-10336.