Electronic Supplementary Information (ESI)

Programming interchangeable and reversible heterooligomeric protein self-assembly using a bifunctional ligand

Soyeun Son[a] and Woon Ju Song*[a]

[a]Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 08826, Republic of Korea

*Email: woonjusong@snu.ac.kr

TABLE OF CONTENTS

1. Biochemical Experimental Procedures

1.1. General considerations

The following reagents were purchased from LPS Solution (Korea): Luria-Bertani broth (LB), Terrific broth (TB), bacteriological agar, kanamycin sulfate (Kan), isopropyl-ß-D-thiogalactopyranoside (IPTG), tris(hydroxymethyl)aminomethane hydrochloride (tris-HCl), 4-morpholinepropanesulfonic acid (MOPS), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), guanidine hydrochloride, N-[tris(hydroxymethyl)methylamino]propane-sulfonic acid (TAPS), bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane (bis-tris), boric acid, citric acid, sodium chloride (NaCl), ethylenediamine tetraacetic acid (EDTA), imidazole, dithiothreitol (DTT) and ammonium sulfate (AmSol). Buffers were prepared using doubly distilled water (ddH_2O) from Milli-Q purification system (18.2 MΩ). Zinc sulfate heptahydrate ($ZnSO_4 \cdot 7H_2O$), tris(2-carboxyethyl)phosphine (TCEP), ammonium bicarbonate (NH_4HCO_3), sodium sulfate decahydrate ($Na_2SO_4 \cdot 10H_2O$), *p*-nitrophenyl acetate (*p*NPA), dimethyl sulfoxide (DMSO), copper chloride ($CuCl_2$), phenylmethylsulfonyl fluoride (PMSF), and 4-carboxybenzenesulfonamide (CBzS) were purchased from Alfa Aesar. L-cysteine and Coomassie Brilliant Blue G-250 was purchased from Tokyo Chemical Industries (TCI) chemicals. Uranyl acetate was purchased from Electron Microscopy Sciences (USA). DNase I was obtained from TaKaRa Bio (Japan). Lysozyme and β-mercaptoethanol (βME) were purchased from Bio Basic Inc. (Canada). PCR enzyme and kit including KOD-Plus-Neo were purchased from Toyobo (Japan). Human carbonic anhydrase I (CA1) and human carbonic anhydrase XII (CA12) isoforms were purchased from NKMAXBio (Korea). Protein purification steps described below were all conducted at either 4 °C or in an ice bath unless otherwise noted.

1.2. Plasmids construction

The nucleotide sequences encoding four proteins, human carbonic anhydrase II (CA), 2-keto-3-deoxygluconate aldolase from *Sulfolobus solfataricus* (KDGA), aminoglycoside acetyltransferase from *Bacillus anthracis* (AT), and encapsulin from *Thermotoga maritima* (Enc) were optimized for *E. coli* heterologous expression (Tables S1 and S2). After gene synthesis (Gene Universal), the fragments were inserted into pET-26b(+) vector for CA and pET-28a(+) vector for KDGA, AT, and Enc, using Nde I/Not I cut-sites for CA, Nde I/Xho I cut-sites for KDGA and AT, and Nco I/Xho I cut-sites for Enc. Plasmids were transformed into DH5α competent cells and validated by sequencing (Bionics or Macrogen).

1.3. Protein expression and purification

1.3.1. Expression and purification of CA and CA double mutants

The wild-type protein and its variants were prepared as reported previously[1] with slight modifications (Fig. S1). Briefly, after the site-directed mutagenesis to insert the TEV cut site at the C-terminus of the protein using the custom-ordered primers (Table S3), the plasmids were transformed into BL21(DE3) competent cells and plated on LB agar plates containing 50 μg/mL Kan. An individual colony was selected and grown in a 20 mL TB media with 50 μg/mL of Kan at 37 °C overnight in a shaking flask (140 rpm). Subsequently, this culture was inoculated into 1 L TB media by a 100-fold dilution. When the optical density of the culture (A_{600}) reached 0.8–1.0, a final concentration of 0.42 mM IPTG and 0.62 mM $ZnSO_4$ solution was added to the flask. The cells were then grown for 12–16 h at 28 °C in an orbital shaker (180 rpm) and harvested by centrifugation at 4000 rpm and 4 °C for 20 min. The pellets were frozen in liquid nitrogen and stored at −80 °C.

Before cell lysis, the harvested cell pellets were resuspended at a ratio of 1 g pellet per 20 mL of a loading buffer (20 mM tris-HCl, pH 8.0, 200 mM NaCl) supplemented with 10 mM βME,

20 μg/mL lysozyme, 1 mM PMSF, and 1 μg/ml DNase I. After sonication with on/off cycles of 2.0/2.5 s for 30 min, the solution was centrifuged at 13,000 rpm for 1 h. The supernatant was filtered through a 0.45 μm hydrophilic membrane filter and loaded onto a His-tag column (Cytiva Biosciences, HisTrap FF 5 mL). After washing with the loading buffer supplemented with 2.5 mM imidazole, the protein was eluted using a linear gradient from 2.5 mM to 500 mM imidazole in the loading buffer (Fig. S1a).

After exchanging with TEV protease buffer (50 mM tris-HCl, pH 8.0, 200 mM NaCl, 0.5 mM EDTA, 1 mM DTT) using 10k spin columns, 10 units/mL of TEV protease was added to the protein (15 μM) and incubated overnight at 30 °C to remove the C-terminal His-tag. The protein was then concentrated using 30k spin columns and subjected to size exclusion chromatography (Cytiva Biosciences, 16/60 PG75 column) using SEC buffer (10 mM tris-HCl, pH 8.0, 200 mM NaCl) to estimate size and oligomeric state (Fig. S1b). The purified proteins were supplemented with 10% glycerol, snap-frozen in liquid nitrogen, and stored in a -80 °C freezer until further use. Protein purity was confirmed by SDS-PAGE analysis (Fig. S1c). Protein concentration was determined by measuring the absorbance at 280 nm with 800–900 nm baseline correction (Agilent 8453 UV-vis spectrophotometer with a ChemStation software). The molar absorptivity of the wild-type protein CA ($\varepsilon_{280\ nm}$ = 50,420 M^{-1} cm^{-1}) was calculated from the ExPasy tool with the amino acid sequences (Table S2).

1.3.2. Expression and purification of KDGA variants

KDGA*-E288C mutant was prepared through site-directed mutagenesis of KDGA using the custom-ordered primers (Table S4). The purification procedure for KDGA and KDGA*-E288C mutant were similar to those of CA, with a difference in the induction step; after a 100-fold dilution inoculation into a 1 L TB media with 50 μg/mL of Kan at 37 °C, 1.0 mM IPTG was added when the A_{600} reached 0.3–0.5. The cell culture was then grown for 24 h at 37 °C.

The harvested cell pellets were resuspended in a loading buffer (20 mM tris-HCl, pH 8.0, 200 mM NaCl) at a ratio of 1 g pellet per 20 mL of buffer, supplemented with 10 mM βME, 20 μg/mL lysozyme, 1 mM PMSF, and 5 μg/L DNase I. Following sonication and centrifugation as described above, the solution was filtered through a 0.45 μm membrane filter and loaded onto a His-tag column (Fig. S2a and c). The protein was then eluted using a linear gradient elution from 2.5 mM to 500 mM imidazole added to the loading buffer. The eluted protein was concentrated using 30k Amicon spin columns and subjected to size exclusion chromatography (Cytiva Biosciences, 16/60 PG200 column) using SEC buffer (10 mM tris-HCl, pH 8.0, 200 mM NaCl) (Fig. S2b and d). For the SEC of the KDGA*-E288C variant, 5 mM DTT was added to the SEC buffer to maintain the reduced state of the surface-exposed cysteine during the purification process. The conjugated proteins were supplemented with 10% glycerol, snap-frozen in liquid nitrogen, and stored in a -80 °C freezer until further use. The purity of the isolated protein was confirmed by SDS-PAGE analysis (Fig. S2f). Protein concentration was determined by measuring the absorbance at 280 nm with 800–900 nm baseline correction. The molar absorptivity of KDGA was determined from the ExPasy tool with its amino acid sequences (Table S2), which provided the value for the oxidized form of the monomeric subunit, $\varepsilon_{280\,nm}$ = 23,965 $M^{-1}\cdot cm^{-1}$.

1.3.3. Expression and purification of AT variant

AT*-K44C mutant was prepared by the site-directed mutagenesis using the custom-ordered primers (Table S6). The protein was prepared similarly to the CA except for the induction step; IPTG was added to the 1 L TB culture with 50 μg/mL of Kan at the final concentration of 0.5 mM when A_{600} reached 0.6–1.0, and the cell was grown for 18 h at 25 °C. The harvested cell pellets were resuspended in a loading buffer (50 mM tris-HCl, pH 7.4, 50 mM NaCl) as 1 g pellet/10 mL of buffer aided with 10 mM βME, 20 μg/mL lysozyme, 1 mM PMSF, and 5 μg/L

DNase I. After sonication and centrifugation as described above, the supernatant was filtered through 0.45 μm membrane filter and loaded onto the His-tag column. The protein was eluted using a linear gradient elution from 2.5 mM to 500 mM imidazole added to the loading buffer (Fig. S11b). The eluted protein was concentrated using 30k Amicon spin columns and subjected to size exclusion chromatography (Cytiva Biosciences, 16/60 PG200 column) using SEC buffer (10 mM tris-HCl, pH 8.0, 200 mM NaCl) to estimate size and oligomeric state. For the SEC of AT*-K44C variant, 5 mM DTT was added in the SEC buffer (Fig. S11c). Protein purity was confirmed by SDS-PAGE (Fig. S11d). Protein concentration was determined by measuring the absorbance at 280 nm with 800–900 nm baseline using the molar absorptivity of the monomeric subunit of the wild-type protein AT*-K44C ($\varepsilon_{280\ nm}$ = 65,780 M^{-1} cm^{-1}) calculated from ExPasy tool with its amino acid sequences.

1.3.4. Expression and purification of Enc

Encapsulin (Enc) (Fig. S12a) was prepared according to the previously reported procedures[2] with slight modifications. Briefly, the BL21(DE3) cells were grown at 37 °C in a 1 L TB media with 50 μg/mL of Kan until A_{600} reached 0.3. Then, the culture was cooled down to 16 °C and 0.5 mM IPTG was added at the final concentration. After the cell growth at 16 °C for 16 h, the cells were harvested by the centrifugation at 4 k rpm for 20 min at 4 °C. The cell pellets were resuspended in a loading buffer (50 mM tris-HCl, pH 7.4, 150 mM NaCl) as 1 g pellet/20 mL of buffer with 5 mM βME, 20 μg/mL lysozyme, 1 mM PMSF, and 5 μg/L DNase I. After incubation at room temperature for 15 min, the suspension was subjected to sonication for 30 min at 30% intensity (2.0/2.5 s) and centrifuged at 13 k rpm for an hour. The supernatant was filtered through a 0.45 μm hydrophilic membrane filter and incubated at 80 °C for 3 h. After centrifugation at 13 k rpm for 30 min, 35% (w/v) AmSol was added to the supernatant and incubated at room temperature for 30 min. After centrifugation at 8 k rpm for 20 min, additional

AmSol was added to make the final concentration of 70% (w/v) and incubated at room temperature for another 30 min. The white precipitates were separated by centrifugation at 8 k rpm for 25 min, resuspended with the equal volume to the resuspension buffer (20 mM tris-HCl, pH 7.4, 150 mM NaCl, 5 mM βME), and the resulting solution was concentrated using 30k spin columns. The concentrate was dialyzed to the loading buffer with 5 mM βME overnight at 4 °C. The protein was further purified by size exclusion chromatography (Cytiva Biosciences, 16/60 Sephacryl S-500 HR) pre-equilibrated with the loading buffer. Protein purity was confirmed by SDS-PAGE analysis (Fig. S12b). Protein concentration was determined by measuring the absorbance at 280 nm with 800–900 nm baseline correction using the molar absorptivity of Enc ($\varepsilon_{280\ nm}$ = 34,950 M^{-1} cm^{-1}) calculated from ExPasy tool with its amino acid sequence. Of note, the isolated Enc contained the significant amount of DNA, determined by the relatively high ratio of absorptions at 260 over 280 nm (1.6–1.7).

1.4. Protein bioconjugation with bifunctional linkers

Prior to the thiol**–**ene bioconjugation, purified KDGA*-E288C, AT*-K44C, and Enc-WT were treated with 10 mM DTT at the final concentration in 50 mM MOPS, pH 7.4, 200 mM NaCl buffer at room temperature overnight. Immediately after 10 DG desalting chromatography steps to remove excess reductants, a 5-fold molar excess of each bifunctional linker (BzS5 or BzS3) dissolved in DMSO as a 5.0 mM stock solution was added and incubated in the dark overnight. After washing 6 times using a 10k spin column, bioconjugation was determined by subjecting to HR-ESI-MS (Fig. S3, details in section 1.5.1.) and using thiol detection assay kit (Cayman chemical) (Table S5, details in section 1.5.2.). The conjugated proteins were supplemented with 10% glycerol, snap-frozen in liquid nitrogen, and stored in a −80 °C freezer until further use.

1.5. Biochemical characterizations

1.5.1. HR-ESI-MS analysis

The protein samples (1.5 μg/mL) in 50 mM NH_4HCO_3 buffer (pH 7.2) were analyzed using reversed-phase LC-MS (Orbitrap Exploris 120, Thermo Fisher Scientific) and Hypersil GOLD C18 Selectivity HPLC column (5 μM, 4.6 μM, 150 mm, Thermo Fisher) using LC-MS-grade solvent A (0.1% formic acid (Fujifilm) in water) and solvent B (0.1% formic acid in acetonitrile (LC-MS grade, J. T. Baker)). The injected samples were eluted with a flow rate of 0.5 mL/min using the following program; 15% solvent B for 0.5 min, a linear gradient of 15-100% solvent B for 10.5 min, elution with 100% solvent B for 5 min, linear gradient of 100-15% solvent B for 0.5 min, and wash with 15% solvent B for 4.5 min. Proteins were ionized by heated electrospray ionization (vaporizer temperature 350 °C; 3500 V spray voltage). The instrument was operated in data-dependent mode with Xcalibur software version 4.3 (Thermo Fisher Scientific). MS scan range was 200−3000 m/z. The obtained MS data were processed using Thermo Scientific Freestyle™ 1.8 SP2. Native mass spectra were deconvolved using the program UniDec, as described previously.[3]

1.5.2. Quantification of thiols

Quantification of surface-exposed cysteine was carried out by using a commercially available thiol-selective fluorescent dye (thiol detection assay kit, Cayman Chemical, 700340) and a microplate reader (Synergy H1, Biotek). Briefly, the proteins were diluted in the reaction buffer (50 mM HEPES, pH 7.4, 4 M GdnHCl) and incubated at 37 °C overnight. After the samples were cooled down to room temperature, 10 mM DTT was added by adding 1/100-fold (v/v) of 1 M DTT stock solution in ddH_2O, and they were reduced for 2 hours to activate the free thiols. Samples were washed with reaction buffer at least 6 times using 10k spin columns prior to detection. After adding the dye solution to the samples (20 nM–1.0 μM), samples were

incubated in dark for 5 min, then fluorescence intensities were measured using an excitation wavelength of 390 nm and an emission wavelength of 510 nm. For relative quantifications, standard solution of L-cysteine in the reaction buffer was freshly prepared for each measurement and in case of using different microplates, which were also recorded under the same conditions.

1.5.3. Determination of esterase activities and inhibition constants

The hydrolytic activities of CA and dCA were determined with *p*NPA as described previously.[4] Briefly, protein (1.0 μM) in TAPS buffer (50 mM TAPS, pH 7.4, 100 mM Na_2SO_4) was added to the various concentrations of *p*NPA solution (0–1.2 mM). The *p*NPA stock solution (3.0 mM) was prepared with 3% acetone/ddH_2O, by diluting 0.10 M *p*NPA acetone stock solution into ddH_2O. Spectral changes were detected using a 96-well plate reader (SYNERGY H1 microplate reader, BioTek) with pathlength correction. Absorbance increment at 348 nm was measured to determine initial reaction velocities. Catalytic efficiency (k_{cat}/K_M) values were calculated from the initial velocities, using the extinction coefficient ($\varepsilon_{348\ nm}$ = 5000 $M^{-1}\cdot cm^{-1}$), as a function of *p*NPA concentration applied for the assays (Fig. S7). Specific activities were calculated from initial velocities with *p*NPA at the concentration of 1.2 mM. Background rates were measured with CA and dCA treated with 0.5 mM acetazolamide (AZD).

For experiments with inhibitors, a stock solution of 4-carboxybenzenesulfonamide (CBzS) (10 mM in DMSO 100%) was prepared and diluted to make 1.0, 3.0, 5.0, 10, and 20 eq. of inhibitor to the monomeric KDGA (Fig. 3b). The inhibition constants (K_i) of CBzS and protein bioconjugates, BzS5- and BzS3-attached KDGA proteins, were calculated from the remaining activity compared to the results of the inhibition by CBzS, measuring remaining hydrolytic activities with *p*NPA. The data were analyzed by using the program Dynafit using the script

below, where the K_i value of *p*NPA was set as the K_M value from the Lineweaver-Burk plot analysis of the wild-type CA activity (Fig. S7b).

```
[task]
task = fit
data = equilibria
; S … CBzS or KDGA*-E288C|BzS5 or KDGA*-E288C|BzS3
; P … pNPA substrate
; E … CA
[mechanism]
E + S <==> ES  :  KdS dissoc
E + P <==> EP  :  KdP dissoc
[constants]
KdS = 0.1 ?
KdP = 1e+7
[concentrations (μM)]
E = 1.0
P = 1200
[responses]
ES = 0.5 ?
[data]
variable S
file
[output]
directory ./output_directory
[end]
```

1.5.4. Dynamic light scattering (DLS)

In order to determine the size profiles of each protein component, sample was centrifuged at 13 k rpm, 4 °C for at least an hour prior to DLS experiments. 10 μM protein samples in 50–100 μL scales were used for detection, using a disposable cuvette ZEN0040 (Malvern Instrument). Cuvette was washed three times with corresponding buffers prior to measurements in order to minimize protein aggregation by transferring. For heteroassembly, samples were mixed after centrifugation. CA1 and CA12 samples were prepared in 20 mM tris-H_2SO_4 buffer (pH 8.0) and 50 mM TAPS buffer (pH 8.0) with 100 mM Na_2SO_4, respectively (Fig. S5). For assembly reversion experiments using CBzS and DTT, a freshly prepared stock solution (10 mM, DMSO 100% for CBzS and ddH_2O 100% for DTT) was added as 1/20-fold (v/v) dilution to the protein solution in a cuvette.

DLS experiments were performed on Malvern Instrument Zetasizer Nano ZS and data were collected at 25 °C at a detection angle of 173°. The experimental data were analyzed using the bundle program Zetasizer Software v.8.02 for intensity and volume profiles of protein samples.

1.6. Circular dichroism (CD)

CD experiments were performed with 2.5–5.0 μM protein samples (100 μL in 50 mM borate pH 8.0 buffer) using 0.1 mm pathlength cuvette incubated in a Peltier controller at 25 °C (Fig. S8a). For melting temperature analysis, the changes in ellipticity were measured at 210 nm with an interval of 10 s between data points, along with temperature elevation from 25 °C to 105 °C. The molar ellipticity was fit to a sigmoidal curve using Origin software (Fig. 3c).

1.7. Simulation of heterooligomer structures and hydrodynamic diameters

The size of the resulting inhibitor-induced heterooligomeric assembly was determined using three computational programs. First, AutoDock Vina[5] was used to predict the bound conformation of BzS5 to the active center and hydrophobic pocket of CA-WT. Clusters showing plausible locations and conformations of both the sulfonamide residue and reactive free thiol, E288C, were selected to identify the distance between the alpha carbon of cysteine and the zinc center, and the distance values were used to select the plausible cluster in the next step (Fig. S4a). Next, HADDOCK[6] was used to predict the possible structures of CA associated with the monomeric subunit of the tetramer KDGA. Two clusters among nine were selected based on the possible distance regime deduced from AutoDock Vina results (Fig. S4b). Subsequently, four CA (PDB 3CA2) to the KDGA tetramer (PDB 1W37) were placed manually using PyMOL without removing water molecules in the X-ray structure, and structures with possible steric clashes were omitted for the next step (Fig. S4c). Finally, the

obtained pdb file were subjected to WinHydroPro[7] to calculate the radius of gyration and the hydrodynamic diameter.

1.8. Preparation of CA dimer (dCA)

A double mutant of CA (C206S/E187C) was prepared by the site-directed mutagenesis using the custom-ordered primer (Table S3). Experimental conditions for expression and purification were the same to the conditions used for CA, right through the His-tag step, as denoted above. The purified CA mutant (C206S/E187C, 1.0–5.0 μM in 50 mM tris-HCl, pH 8.0, 200 mM NaCl buffer) was incubated with 1 mM TCEP for 30 min at 4°C, washed with TCEP-free buffer, then subjected to 5.0 equiv. $CuCl_2$ (Fig. S9d). The reaction was quenched by the addition of 50 molar excess of EDTA solution and incubated overnight to fully chelate copper species. After the buffer exchange step using 30k spin column, the sample was loaded to the Superdex 16/60 PG75 gel filtration column to isolate the dimer fractions (red solid line in Fig. S9b). The dimerized protein was stored in the SEC buffer with 10% glycerol at –80 °C.

1.9. Transmission electron microscopy (TEM)

The protein samples for TEM were prepared as described above, except for using NaCl instead of Na_2SO_4 to prevent uranyl acetate precipitation in the staining step. The assembled heterooligomer samples were diluted in a degassed, doubly distilled water to make 50–100 nM solutions (3.0 μL). They were loaded onto a freshly glow-discharged carbon-coated copper grid (Ted Pella) using a side-blot method. After incubation for 2 min, the grid was washed twice with a degassed ddH_2O and stained twice using a freshly prepared 2% aqueous solution of uranyl acetate for 15 s each. Uranyl acetate stain solution was freshly prepared for each experiments in a degassed ddH_2O by filtering through a 0.22 μm PVDF filter and into the 1.5 mL amber ep tube. All sampling procedures were undergone using both degassed ddH_2O and

stain solutions cooled down to 4 °C. TEM data were collected using HITACHI H-7600 (HITACHI–Science & technology) with an acceleration voltage of 100 kV.

2. Synthetic Procedures and Characterization of the Bifunctional Linker

2.1. General considerations

All reagents were purchased from suppliers, such as Alfa Asesar, Acros Organics, TCI chemicals, and Sigma Aldrich, and used as received without further purification unless otherwise noted. Tetrahydrofuran (THF; Alfa Aesar) was dried over 3 Å molecular sieves for 1–2 days and then distilled from sodium/benzophenone under argon (Ar) atmosphere. Triethylamine (NEt_3; Alfa Aesar) was prepared by refluxing with *p*-toluenesulfonyl chloride (*p*-TsCl) for 2 h, followed by distillation under Ar. Diisopropyl ethylamine (DIPEA; TCI chemicals) was dried by distillation from ninhydrin and then potassium hydroxide (KOH) under Ar. Silica for column chromatography was purchased from Merck.

All air-sensitive manipulations were carried out under Ar using standard Schlenk-line techniques. ^{1}H- and ^{13}C-NMR spectra were recorded using either 400 MHz Agilent 400-MR DD2 Magnetic Resonance System or a 500 MHz Varian/Oxford As-500 spectrometer. Chemical shifts were referenced to the residual solvent peaks[8] or the internal standard tetramethylsilane (δ=0.00 ppm). NMR solvents $CDCl_3$ and DMSO-d_6 were purchased from Cambridge Isotope Laboratories Inc (USA). High-resolution electrospray ionization (ESI) mass spectra of the synthesized molecules were obtained using an ESI-Q-TOF mass spectrometer coupled with reversed-phase LC-MS (Orbitrap Exploris 120, ThermoFisher Scientific).

2.2. Synthesis and characterization of bifunctional linkers BzSn (a: n=3, b: n=5)

3,6-epoxy-1,2,3,6-tetrahydrophthalimide (1)

The procedure was carried out following a previously reported scheme with minor modifications.[9] Briefly, in a 100-mL ace round-bottom pressure flask, furan (1.5 mL, 20.0

mmol, 1.0 eq.), maleimide (3.89 g, 40.1 mmol, 2.0 eq.) and toluene (25 mL) were combined with Ar blowing. The reaction mixture was tightly sealed with a threaded Teflon cal and left stirring at 90 °C overnight. After cooling to room temperature, the resulting white precipitate was filtered through a P3 glass pore frit and washed with cold ether (100 mL) to obtain compound **1** as a white solid (2.98 g, 18.0 mmol, yield = 90%).

^{1}H-NMR (400 MHz, $CDCl_3$): δ=7.79 (s, 1H), 6.52 (s, 2H), 5.32 (s, 2H), 2.89 (s, 2H); ^{13}C-NMR (101 MHz, $CDCl_3$): δ=176.12, 136.72, 81.13, 48.85; HRMS (ESI): m/z calcd for $C_8H_7NO_3$-H^+: 164.0353 $[M-H]^-$; found: 164.0354.

2,5-dioxopyrrolidin-1-yl 4-sulfamoylbenzoate (2)

The procedure was followed according to a previously reported scheme with slight modifications.[10] Briefly, a dry Schlenk flask was charged with 4-sulfamoylbenzoic acid (500 mg, 2.5 mmol, 1.0 eq.), and dry DMF (17 mL) was transferred using a syringe. After cooling to 0 °C in an ice bath, *N*-hydroxysuccinimide (320.3 mg, 2.8 mmol, 1.12 eq.) and *N,N*'-dicyclohexylcarbodiimide (DCC; 512.8 g, 2.5 mmol, 1.0 eq.) were added all at once with Ar blowing. The reaction mixture was gradually warmed up to room temperature overnight. The resulting white suspension was cooled to 0 °C and then filtered through a P3 pore frit to remove unwanted urea byproduct. The filtrate was dried in vacuo, washed with hot isopropyl alcohol, and left to sir and cool overnight. Compound **2** was obtained as a white solid through filtration (736.1 mg, 2.5 mmol, quant.). The NMR data closely matched previous reports.[11]

^{1}H-NMR (400 MHz, DMSO-d_6): δ=8.30 (d, J = 1.3 Hz, 2H), 8.07 (d, J = 1.3 Hz, 2H), 7.70 (s, 2H), 2.91 (s, 4H); ^{13}C-NMR (126 MHz, DMSO-d_6): δ=170.48 , 161.38, 150.24, 131.26, 127.59, 127.07, 25.95; HRMS (ESI): m/z calcd for $C_{11}H_{10}N_2O_6S$+H^+ 299.0332 $[M+H]^+$; found: 299.0332.

2-(2-(2-hydroxyethoxy)ethoxy)ethyl 4-methylbenzenesulfonate (3a) or 14-hydroxy-3,6,9,12-tetraoxatetradecyl 4-methylbenzenesulfonate (3b)

The procedure was followed according to the previously reported scheme with slight modifications.[12] In short, a dry 100-mL Schlenk flask was charged with either triethylene glycol (4.0 mL, 29.9 mmol, 1.0 eq.) or pentaethylene glycol (4.0 mL, 17.2 mmol, 1.0 eq.) for n=3 or 5, respectively, and then purged with Ar three times. Dry THF (30 mL for n=3; 18 mL for n=5) was added to the flask using a syringe, and the mixture was cooled to 0 °C in an ice bath. Dry NEt_3 (3.3 M, 9.0 mL for n=3; 5.4 mL for n=5) was added to the cooled mixture using syringe. Another dry Schlenk flask was charged with *p*-TsCl (2.85 g, 14.95 mmol, 0.5 eq. for n=3; 1.63 g, 8.6 mmol, 0.5 eq. for n=5), purged with argon three times, and charged with dry THF (11 mL). The *p*-TsCl solution was transferred to the other flask dropwise using syringe over a 30-min period. The reaction mixture was left stirring in the ice bath and gradually warmed up to room temperature overnight.

The crude product was concentrated under a reduced pressure and dissolved in DCM (100 mL) and water (100 mL). The organic layer was extracted with DCM (100 mL) 3 times. The organic phase was then washed with 5% citric acid (100 mL, 3 times), dried over anhydrous Na_2SO_4, filtered, and purified through silica column chromatography (flash, dry loading; Hex:EtOAc = 1:1 to EtOAc 100%, followed by EtOAc/10% MeOH gradient elution). Both compounds **3a** (3.65 g, 12.0 mmol, yield = 40%) or **3b** (2.36 g, 6.02 mmol, yield=35%) were obtained as transparent oils.

For **3a**: R_f=0.30 (EtOAc 100%); ^{1}H-NMR (400 MHz, $CDCl_3$): δ=7.80 (d, J = 8.3 Hz, 1H), 7.34 (d, J = 8.3 Hz, 2H), 4.16 (t, J = 3.9 Hz, 2H), 3.74–3.53 (m, 10H), 2.44 (s, 3H); ^{13}C-NMR (126 MHz, $CDCl_3$): δ=145.00, 133.12, 129.98, 128.12, 72.57, 70.94, 70.48, 69.30, 68.87, 61.92, 21.79; HRMS (ESI): m/z calcd for $C_{13}H_{20}O_6S+H^+$ 305.1054 $[M+H]^+$; found: 305.1056.

For **3b**: R_f=0.30 (EtOAc+10% MeOH); ^{1}H-NMR (400 MHz, $CDCl_3$): δ=7.81 (d, J = 8.0 Hz, 2H), 7.35 (d, J = 8.0 Hz, 2H), 4.17 (t, J = 4.9 Hz, 2H), 3.74–3.59 (m, 18H), 2.46 (s, 3H); ^{13}C-NMR (126 MHz, $CDCl_3$): δ=144.91, 133.18, 129.95, 128.11, 72.66, 70.87, 70.72, 70.70, 70.66, 70.62, 70.44, 69.37, 68.82, 61.86, 21.76; HRMS (ESI): m/z calcd for $C_{17}H_{28}O_8S+H^+$ 393.1578 $[M+H]^+$; found: 393.1578.

2-(2-(2-azidoethoxy)ethoxy)ethan-1-ol (4a) or 14-azido-3,6,9,12-tetraoxatetradecan-1-ol (4b)

The procedure was followed according to the previously reported scheme with slight modifications.[12] Briefly, to a 0 °C absolute ethanol solution (14 mL for **3a**; 8.0 mL for **3b**) filled either **3a** (3.65 g, 12.0 mmol, 1.0 eq.) or **3b** (2.36 g, 6.01 mmol, 1.0 eq.) in an oven-dried Schlenk tube under Ar, NaN_3 (881 mg, 13.0 mmol, 1.08 eq. for **3a**; 442 mg, 6.49 mmol, 1.08 eq. for **3b**) was added in a single fraction with Ar blowing. After refluxing overnight, the solution was cooled down to room temperature and concentrated under a reduced pressure to remove the solvent. The organic fraction was dissolved and extracted with EtOAc (3×50 mL), dried over anhydrous Na_2SO_4, filtered, and concentrated. The crude product was purified using silica chromatography (flash, dry loading; EtOAc 100% to EtOAc: MeOH = 9:1 for both **3a** and **3b**) to yield either **4a** (2.10 g, 12.0 mmol, quant.) or **4b** (1.58 g, 6.0 mmol, quant.) as a transparent oil.

For **4a**: R_f=0.35 (EtOAc 100%); ^{1}H-NMR (400 MHz, $CDCl_3$): δ=3.77–3.60 (m, 10H), 3.41 (t, J = 5.1 Hz, 2H), 2.22 (t, J = 6.2 Hz, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ=72.42, 70.42, 70.15, 69.80, 61.40, 50.43; HRMS (ESI): m/z calcd for $C_6H_{13}N_3O_3+H^+$ 176.1030 $[M+H]^+$; found: 176.1031.

For **4b**: R_f=0.43 (EtOAc+10% MeOH); ^{1}H-NMR (400 MHz, $CDCl_3$): δ=3.75–3.58 (m, 18H), 3.40 (t, J = 4.7 Hz, 2H), 2.49 (t, J = 6.3 Hz, 1H); ^{13}C-NMR (101 MHz, $CDCl_3$): δ=72.66, 70.84,

70.78, 70.75, 70.71, 70.48, 70.17, 61.90, 50.84; HRMS (ESI): m/z calcd for $C_{10}H_{21}N_3O_5$+H^+ 264.1554 $[M+H]^+$; found: 264.1554.

2-(2-(2-aminoethoxy)ethoxy)ethan-1-ol (5a) or 14-amino-3,6,9,12-tetraoxatetradecan-1-ol (5b)

The procedure was followed according to the previously reported scheme with slight modifications.[13] In short, an oven-dried 50-mL Schlenk flask was charged with either **4a** (1.15 g, 6.57 mmol, 1.0 eq.) or **4b** (1.58 g, 6.00 mmol, 1.0 eq.) and THF:ddH_2O = 9:1 (v/v) (36 mL for **4a**; 30 mL for **4b**, both in total) under Ar. PPh_3 (5.17 g, 19.7 mmol, 3.0 eq. for **4a**; 4.72 g, 18.0 mmol, 3.0 eq. for **4b**) was added in one fraction with Ar blowing and stirred at room temperature overnight. The crude was concentrated under reduced pressure, diluted in ddH_2O (3×30 mL), and washed with Et_2O (3×30 mL). The aqueous layer was evaporated in vacuo to give **5a** or **5b** as a yellow oil. It was carried on for the next step without further purification.
For **5a**: ^{1}H NMR (500 MHz, $CDCl_3$): δ=3.74–3.60 (m, 8H), 3.55 (t, J = 5.1 Hz, 2H), 2.88 (t, J = 5.1 Hz, 2H); ^{13}C NMR (126 MHz, $CDCl_3$): δ=73.34, 72.80, 70.60, 70.35, 61.78, 41.78; HRMS (ESI): m/z calcd for $C_6H_{15}NO_3$+H^+ 150.1125 $[M+H]^+$; found: 150.1124.
For **5b**: ^{1}H-NMR (500 MHz, $CDCl_3$): δ=3.74–3.51 (m, 18H), 2.87 (t, J = 5.0 Hz, 2H); ^{13}C-NMR (126 MHz, $CDCl_3$): δ=73.18, 73.00, 70.76, 70.65, 70.61, 70.59, 70.34, 70.19, 61.56, 41.56; HRMS (ESI): m/z calcd for $C_{10}H_{23}NO_5$+H^+ 238.1649 $[M+H]^+$; found: 238.1647.

tert-butyl (2-(2-(2-hydroxyethoxy)ethoxy)ethyl)carbamate (6a) or tert-butyl (14-hydroxy-3,6,9,12-tetraoxatetradecyl)carbamate (6b)

A 100 mL round-bottom flask was charged with either **5a** or **5b** dissolved in DCM (5 mL for **5a**; 5 mL for **5b**). It was cooled down to 0 °C in an ice bath. Di-*tert*-butyl dicarbonate (Boc_2O;

1.89 mL, 7.88 mmol, 1.2 eq. for **5a**; 1.73 mL, 7.2 mmol, 1.2 eq. for **5b**) was added to the solution dropwise. The solution was warmed up to room temperature and left stirring for an hour. Crude was concentrated under a reduced pressure and purified by silica column chromatography (flash, dry loading; DCM +5% (MeOH with 10% NH_4OH) for **6a**; DCM+ 10% 10% (MeOH with 10% NH_4OH) for **6b**). The desired product, which is either **6a** (852.6 mg, 3.42 mmol 2-step yield=52%) or **6b** (607.3 mg, 1.8 mmol, 2-step yield=30%) was obtained both as a faint yellow oil.

For **6a**: R_f=0.38 (DCM+10% (MeOH with 10% NH_4OH)); ^{1}H-NMR (500 MHz, $CDCl_3$): δ=5.16 (s, 1H), 3.74 (t, J = 4.5 Hz, 2H), 3.67–3.59 (m, 6H), 3.55 (t, J = 5.2 Hz, 2H), 3.32 (q, J = 5.4 Hz, 2H), 1.44 (s, 9H); ^{13}C-NMR (126 MHz, $CDCl_3$): δ=156.14, 79.43, 72.68, 70.51, 70.41, 70.35, 61.84, 40.44, 28.52; HRMS (ESI): m/z calcd for $C_{11}H_{23}NO_5$+H^+ 250.1649 $[M+H]^+$; found: 250.1649.

For **6b**: R_f=0.30 (DCM+10% (MeOH with 10% NH_4OH)); ^{1}H-NMR (500 MHz, $CDCl_3$): δ=5.41 (s, 1H), 3.70 (d, J = 4.4 Hz, 2H), 3.68–3.57 (m, 14H), 3.51 (d, J = 5.2 Hz, 2H), 3.29 (d, J = 5.5 Hz, 2H), 1.42 (s, 9H); ^{13}C-NMR (126 MHz, $CDCl_3$): δ=156.22, 79.13, 72.74, 70.58, 70.55, 70.46, 70.27, 70.24, 61.67, 40.45, 28.52; HRMS (ESI): m/z calcd for $C_{15}H_{31}NO_7$+H^+ 338.2174 $[M+H]^+$; found: 338.2174.

tert-butyl (2-(2-(2-(1,3-dioxo-1,3,3a,4,7,7a-hexahydro-2H-4,7-epoxyisoindol-2-yl)ethoxy)ethoxy)ethyl)carbamate (7a) or tert-butyl (14-(1,3-dioxo-1,3,3a,4,7,7a-hexahydro-2H-4,7-epoxyisoindol-2-yl)-3,6,9,12-tetraoxatetradecyl)carbamate (7b)

A 20-mL vial was loaded with either dry **6a** (140.9 mg, 0.3555 mmol, 1.2 eq.) or **6b** (150 mg, 0.3096 mmol, 1.2 eq.), along with the corresponding amount of **1** (48.9 mg, 0.2962 mmol for **6a**; 42.6 mg, 0.2580 mmol for **6b**), PPh_3 (77.9 mg, 0.2970 mmol, 1.0 eq. for **6a**; 67.7 mg, 0.2580 mmol, 1.0 eq. for **6b**), and a stirrer bar. The vial was purged with Ar three times and

dry THF (3 mL for **6a**; 3 mL for **6b**) was added using a syringe. The solution was cooled down to 0 °C in an ice bath. Diisopropyl azodicarboxylate (DIAD; 60 µL, 0.3047 mmol, 1.03 eq. for **6a**; 60 µL, 0.3047 mmol, 1.18 eq. for **6b**) was added dropwise to the cooled mixture with a syringe. After the complete addition of DIAD, the ice bath was removed, and the reaction was continued overnight at room temperature. The crude product was concentrated under a reduced pressure and purified by silica column chromatography (flash, dry loading; n-hexane: EtOAc = 1:2 for **7a**; EtOAc 100% for **7b**). The product was obtained as a transparent oil both for **7a** (42.1 mg, 0.1422 mmol, yield = 48%) and **7b** (39.7 mg, 0.1032 mmol, yield=40%). Asterisks in NMR spectra indicate residual TPPO.

For **7a**: R_f=0.32 (EtOAc 100%); ^{1}H-NMR (500 MHz, $CDCl_3$): δ=6.50 (s, 2H), 5.25 (s, 2H), 5.14 (s, 1H), 3.68 (t, J = 5.6 Hz, 2H), 3.62 (t, J = 5.7 Hz, 2H), 3.59–3.53 (m, 4H), 3.50 (t, J = 5.2 Hz, 2H), 3.29 (d, J = 5.3 Hz, 2H), 2.86 (s, 2H), 1.43 (s, 9H); ^{13}C-NMR (126 MHz, $CDCl_3$): δ=176.29, 136.68, 133.05, 81.04, 70.42, 70.39, 70.09, 67.31, 47.61, 38.28, 29.83, 28.57, 28.54; HRMS (ESI): m/z calcd for $C_{19}H_{28}N_2O_7$+H^+ 397.1970 $[M+H]^+$; found: 397.1973.

For **7b**: R_f=0.28 (EtOAc+5% MeOH); ^{1}H-NMR (500 MHz, $CDCl_3$): δ=6.52 (s, 2H), 5.27 (s, 2H), 5.10 (s, 1H), 3.69 (t, J = 4.9 Hz, 2H), 3.66–3.58 (m, 14H), 3.54 (t, J = 5.1 Hz, 2H), 3.32 (q, J = 5.4 Hz, 2H), 2.86 (s, 2H), 1.44 (s, 9H); ^{13}C-NMR (126 MHz, $CDCl_3$): δ=176.28, 156.17, 136.69, 132.06, 81.02, 79.28, 70.72, 70.70, 70.67, 70.64, 70.39, 70.20, 67.28, 47.61, 40.50, 38.32, 31.07, 29.83, 28.57; HRMS (ESI): m/z calcd for $C_{23}H_{36}N_2O_9$+H^+ 485.2494 $[M+H]^+$; found: 485.2495.

2-(2-(2-(2-aminoethoxy)ethoxy)ethyl)-3a,4,7,7a-tetrahydro-1H-4,7-epoxyisoindole-1,3(2H)-dione (8a) or 2-(14-amino-3,6,9,12-tetraoxatetradecyl)-3a,4,7,7a-tetrahydro-1H-4,7-epoxyisoindole-1,3(2H)-dione (8b)

A 20-mL vial charged with either dry **7a** (42.1 mg, 0.1422 mmol) or **7b** (39.7 mg, 0.1032 mmol) and a stirrer bar was dissolved in DCM (3 mL for **7a**; 3 mL for **7b**). A 10% TFA solution in DCM (1.2 mL, 4.98 mmol, 35 eq. for **7a**; 1.1 mL, 3.61 mmol, 35 eq. for **7b**) was added dropwise to the solution at room temperature using a glass pipette. The reaction progress was monitored by TLC. Once all the starting materials were consumed, the resulting crude product was concentrated in vacuo overnight. The resulting brown oil was used for the next step without further purification. Asterisks in NMR spectra indicate residual TPPO and TFA-derived byproducts.

For **8a**: ^{1}H-NMR (500 MHz, $CDCl_3$): δ=6.53 (s, 2H), 5.26 (s, 2H), 3.75 (t, J = 5.0 Hz, 2H), 3.71–3.58 (m, 8H), 3.28 (q, J = 5.4 Hz, 2H), 2.93 (s, 2H); ^{13}C-NMR (126 MHz, $CDCl_3$): δ=177.50, 136.62, 133.08, 133.06, 81.15, 70.20, 69.97, 68.07, 66.30, 47.62, 40.24, 38.80; HRMS (ESI): m/z calcd for $C_{14}H_{20}N_2O_5$+H^+ 297.1445 $[M+H]^+$; found: 297.1444.

For **8b**: ^{1}H-NMR (500 MHz, $CDCl_3$): δ=6.52 (s, 2H), 5.25 (s, 2H), 3.89–3.84 (m, 2H), 3.74–3.56 (m, 16H), 3.29 (s, 2H), 2.91 (s, 2H); ^{13}C-NMR (126 MHz, $CDCl_3$): δ=177.11, 136.68, 132.46, 129.88, 81.11, 70.35, 70.16, 70.09, 70.01, 69.85, 69.82, 68.26, 67.17, 47.63, 39.98, 38.65, 29.83; HRMS (ESI): m/z calcd for $C_{18}H_{18}N_2O_7$+H^+ 385.1970 $[M+H]^+$; found: 385.1968.

N-(2-(2-(2-(1,3-dioxo-1,3,3a,4,7,7a-hexahydro-2H-4,7-epoxyisoindol-2-yl)ethoxy)ethoxy) ethyl)-4-sulfamoylbenzamide (9a) or N-(14-(1,3-dioxo-1,3,3a,4,7,7a-hexahydro-2H-4,7-epoxyisoindol-2-yl)-3,6,9,12-tetraoxatetradecyl)-4-sulfamoylbenzamide (9b)

Dry crude of either **8a** or **8b** was dissolved in a dry DMF (0.8 mL) and transferred to a flame-dried 20-mL vial charged with a stirrer bar by using a syringe. The solution was cooled down to 0 °C in an ice bath. **2** (42.0 mg, 0.1408 mmol, 1.0 eq. for **8a**; 31.0 mg, 0.1040 mmol, 1.0 eq. for **8b**) was added to the vial in one fraction under Ar, and the mixture was left stirring to cool down. Dry DIPEA (75 μL, 0.4224 mmol, 3.0 eq. for **8a**; 55 μL, 0.3120 mmol, 3.0 eq. for **8b**)

was added dropwise to the solution using a syringe. Ice bath was removed and the reaction was continued overnight. The crude was concentrated under reduced pressure and purified by using ZORBAX SB-C18 Semi-Preparative 9.4 x 250 mm 5-Micron column applying a linear gradient of ddH_2O:MeCN. After collecting the fraction with the desired product and drying in vacuo, both **9a** (10.1 mg, 0.02109 mmol, yield = 15%) and **9b** (8.8 mg, 0.01551 mmol, yield=15%) were obtained as white solids.

For **9a**: ^{1}H-NMR (500 MHz, DMSO-d_6): δ=8.69 (t, J = 5.6 Hz, 1H), 7.99 (d, J = 8.5 Hz, 2H), 7.89 (d, J = 8.5 Hz, 2H), 7.48 (s, 2H), 6.53 (s, 2H), 5.11 (s, 2H), 3.53–3.40 (m, 12H), 2.92 (s, 2H); ^{13}C-NMR (126 MHz, DMSO-d_6): δ=176.34, 165.27, 146.20, 137.29, 136.46, 127.86, 125.61, 80.32, 69.49, 69.46, 68.78, 66.31, 47.08, 37.58; HRMS (ESI): m/z calcd for $C_{21}H_{25}N_3O_8S+H^+$ 480.1436 $[M+H]^+$; found: 480.1435.

For **9b**: ^{1}H-NMR (400 MHz, DMSO-d_6): δ=8.72 (t, J = 5.6 Hz, 1H), 7.99 (d, J = 8.3 Hz, 2H), 7.89 (d, J = 8.3 Hz, 2H), 7.47 (s, 2H), 6.54 (s, 2H), 5.12 (s, 2H), 3.56–3.43 (m, 20H), 2.93 (s, 2H); ^{13}C-NMR (126 MHz, DMSO-d_6): δ=176.34, 165.26, 146.21, 137.29, 136.48, 127.85, 125.60, 80.33, 69.77, 69.71, 69.65, 69.62, 69.47, 68.76, 66.30, 47.09, 37.58; HRMS (ESI): m/z calcd for $C_{25}H_{33}N_3O_{10}S+H^+$ 568.1960 $[M+H]^+$; found: 568.1962.

BzS3 (10a) or BzS5 (10b)

Either **9a** (10.1 mg, 0.02109 mmol) or **9b** (8.8 mg, 0.01551 mmol) was dissolved in dry anisole: dry MeCN = 1:1 (v/v) (2 mL for both) and transferred to a Schlenk tube purged with Ar, with Ar blowing. It was refluxed for 2 h. After evaporating solvent under a reduced pressure, the product was obtained both as a transparent oil for **10a** (4.8 mg, 0.01687 mmol, yield =80%) or **10b** (6.2 mg, 0.01241 mmol, yield=80%).

For **10a**: ^{1}H-NMR (500 MHz, DMSO-d_6): δ=8.69 (t, J = 5.6 Hz, 1H), 7.98 (d, J = 8.5 Hz, 2H), 7.89 (d, J = 8.5 Hz, 2H), 7.48 (s, 2H), 7.01 (s, 2H), 3.56–3.48 (m, 12H); ^{13}C-NMR (126 MHz,

DMSO-d_6): δ=171.38, 165.73, 146.68, 137.77, 135.02, 128.32, 126.08, 69.97, 69.81, 69.25, 67.41, 37.26; HRMS (ESI): m/z calcd for $C_{17}H_{21}N_3O_7S+H^+$ 412.1173 $[M+H]^+$; found: 412.1171.

For **10b**: ^{1}H-NMR (500 MHz, DMSO-d_6): δ=8.72 (t, J = 5.6 Hz, 1H), 7.99 (d, J = 8.5 Hz, 2H), 7.89 (d, J = 8.4 Hz, 3H), 7.48 (s, 2H), 7.02 (s, 2H), 3.54–3.43 (m, 20H); ^{13}C NMR (126 MHz, DMSO-d_6): δ=170.93, 165.27, 146.21, 137.29, 134.57, 127.86, 125.61, 69.79, 69.76, 69.72, 69.63, 69.39, 68.77, 66.93, 36.80; HRMS (ESI) m/z calcd for $C_{21}H_{29}N_3O_9S+H^+$ 500.1698 $[M+H]^+$, found 500.1703.

3. Supplementary Figures and Tables

Fig. S1. Expression and purification of human carbonic anhydrase II (CA). The representative elution profiles from (a) His-tag and (b) Size exclusion chromatography. (c) 12% SDS-PAGE gel image of the purified CA protein. The asterisk in (a) indicates the elution of the desired protein.

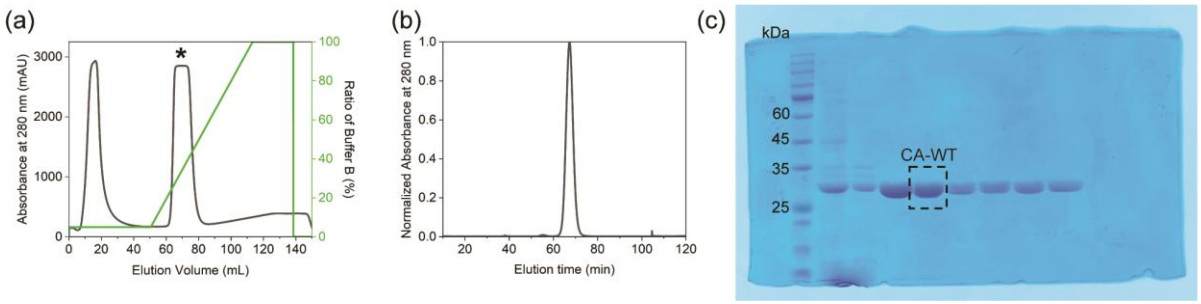


Fig. S2. Expression and purification of 2-keto-3-deoxygluconate aldolase (KDGA) and its variant, KDGA*-E288C. Representative elution profiles of His-tag and Size exclusion chromatography of (a/b) the wild-type KDGA and (c/d) KDGA*-E288C variant. (e) The location of native cysteines (C120 and C150) (PDB 1W37). (f) 12% SDS-PAGE gel of the wild-type protein (WT), C120A/C150A double mutant (AA), and KDGA*-E288C variant (AA E288C). The asterisks in (a) and (c) indicate the elution of the desired proteins.

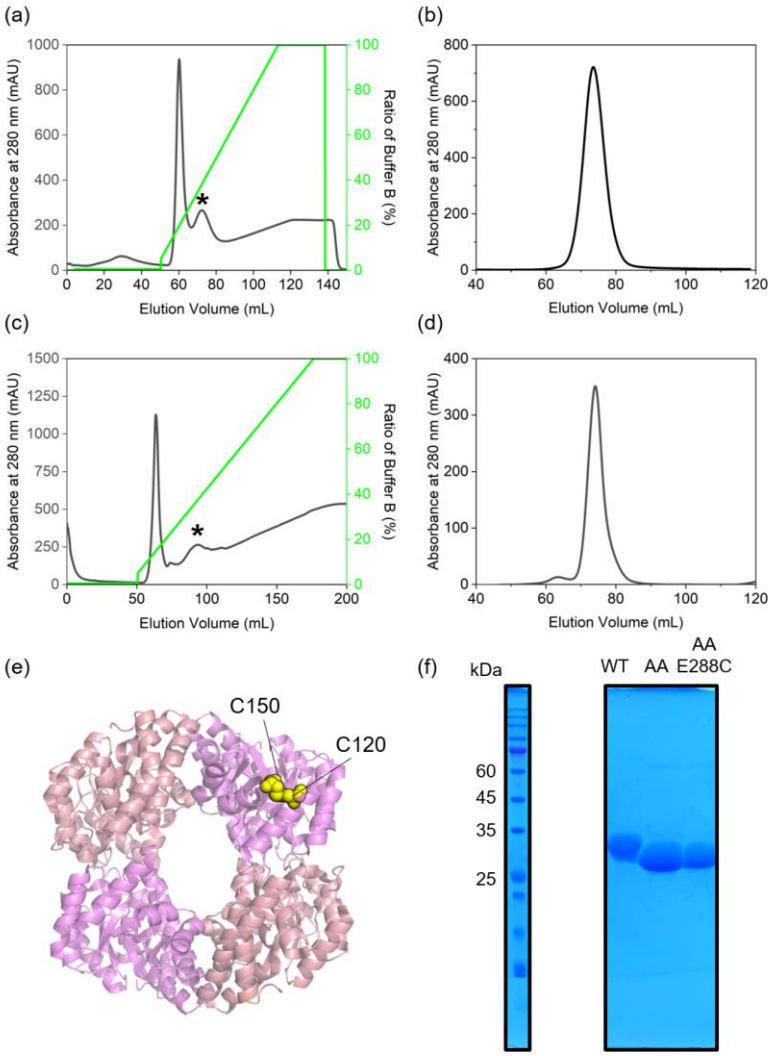


Fig. S3. HR-ESI-MS of KDGA and AT before and after bioconjugation with BzSn. (a) KDGA*-E288C|BzS5 (calculated: 35532, experimental: 35553). (b) KDGA*-E288C|BzS3 (calculated: 35444, experimental: 35468). (c) AT*-K44C|BzS5 (calculated: 48535, experimental: 48379).

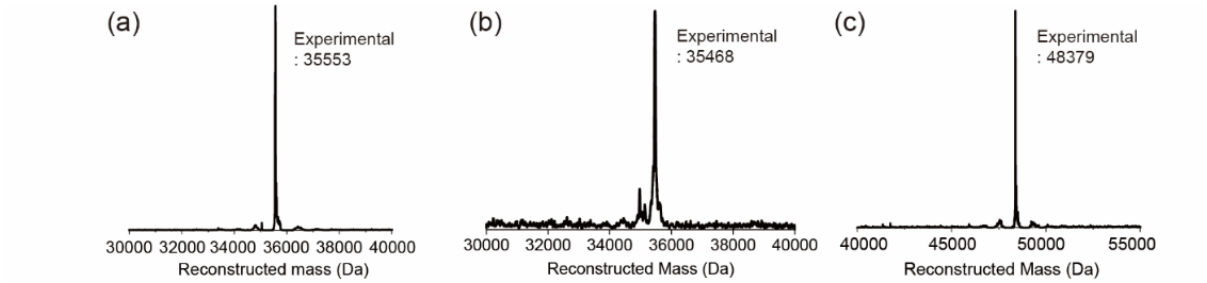


Fig. S4. The structural modeling of CA/KDGA heterooligomer. (a) AutoDock Vina simulation of BzS5 (grey sticks) to CA. The distance of Zn ion and Cα in cysteine-conjugated BzS5 (red arrow) is estimated to be 16–20 Å. (b) Two representative CA and KDGA docking results using HADDOCK. The distance of Zn ion in CA and Cα of E288 position is shown with red arrows. (c) The simulated docking structure of CA and KDGA that satisfy both interactions is shown in (a) and (b).

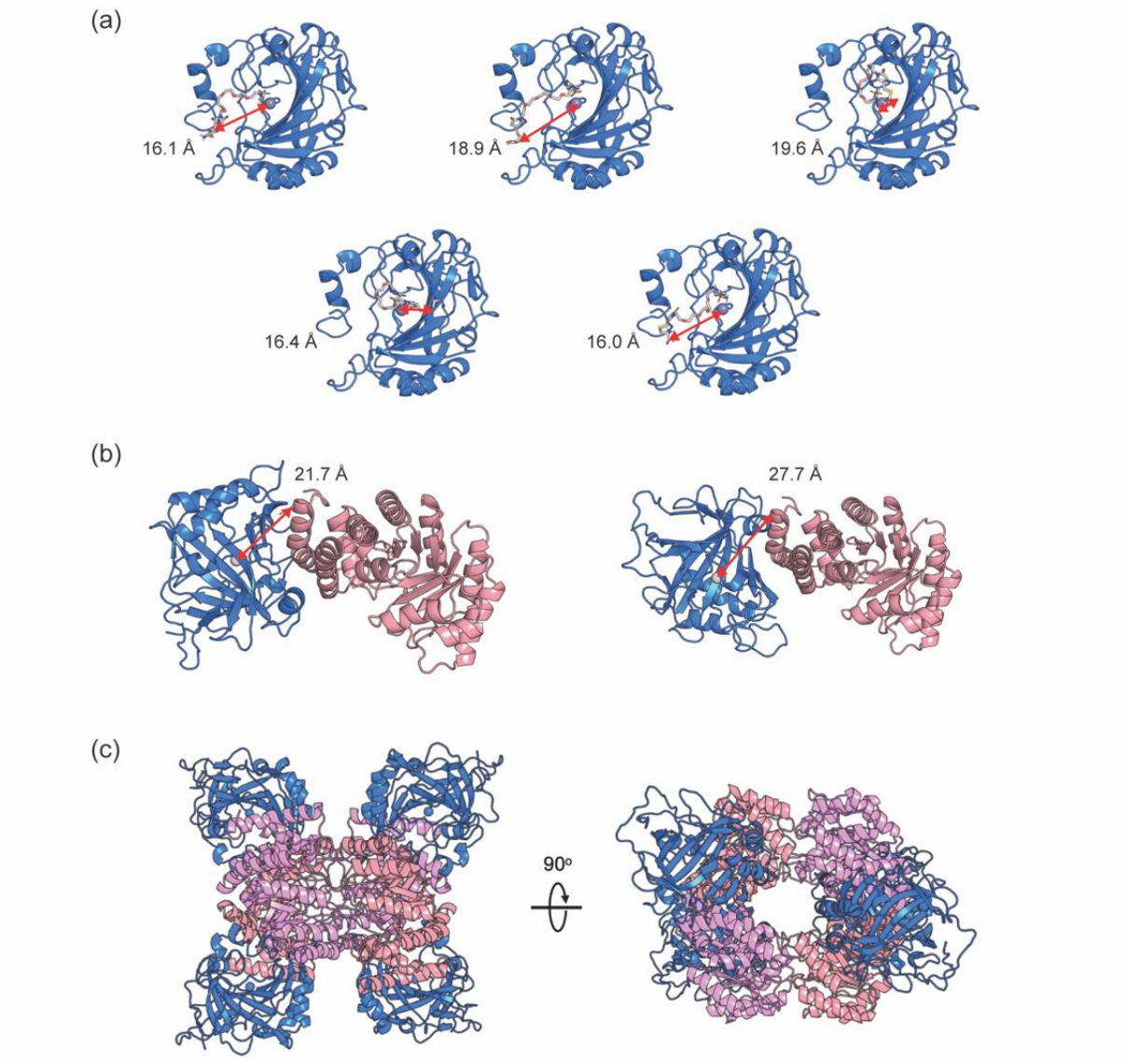


Fig. S5. DLS data of CA before and after adding CBzS. Intensity (left) and volume (right) profile of CA upon the addition of CBzS in (a) 50 mM borate, pH 8.0 buffer (b) 50 mM TAPS, pH 8.0, 100 mM Na_2SO_4 buffer.

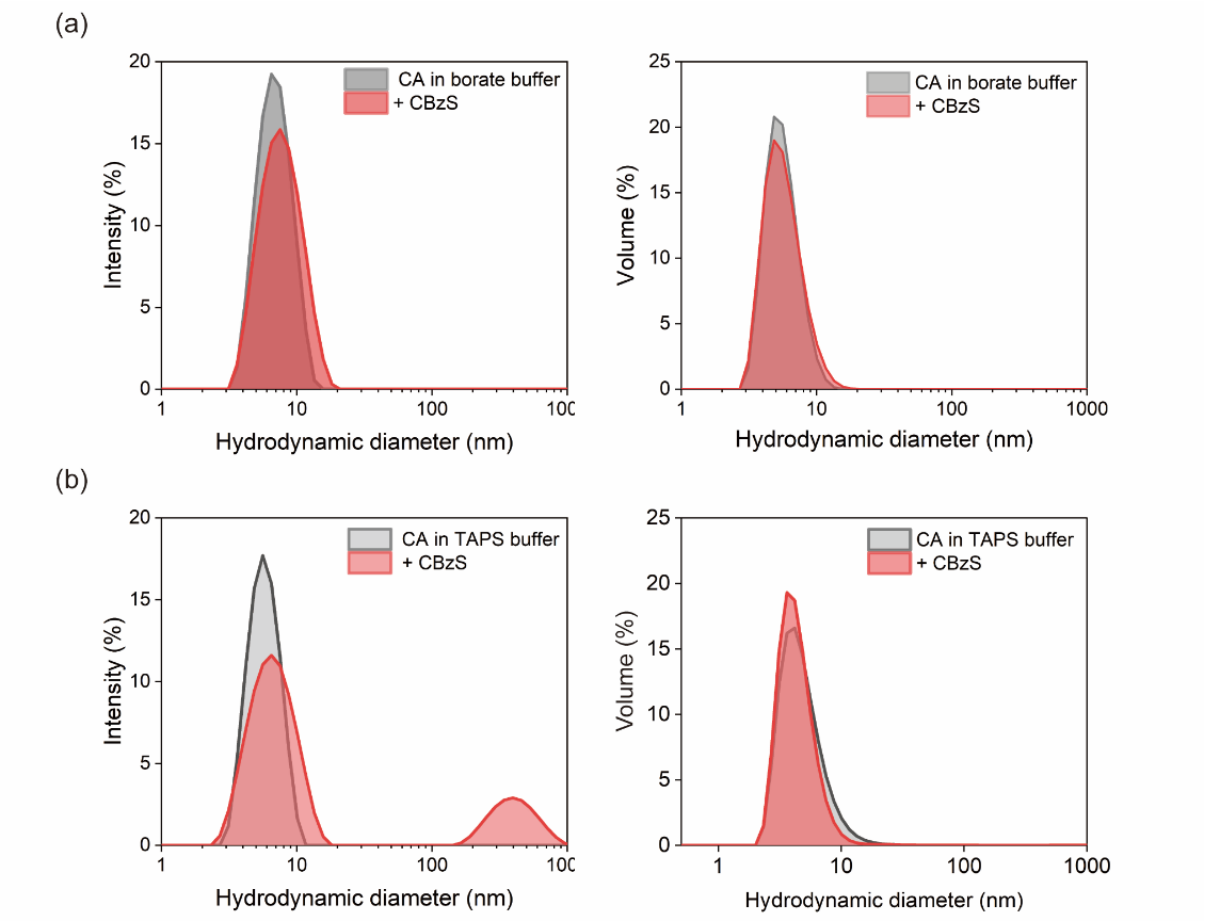
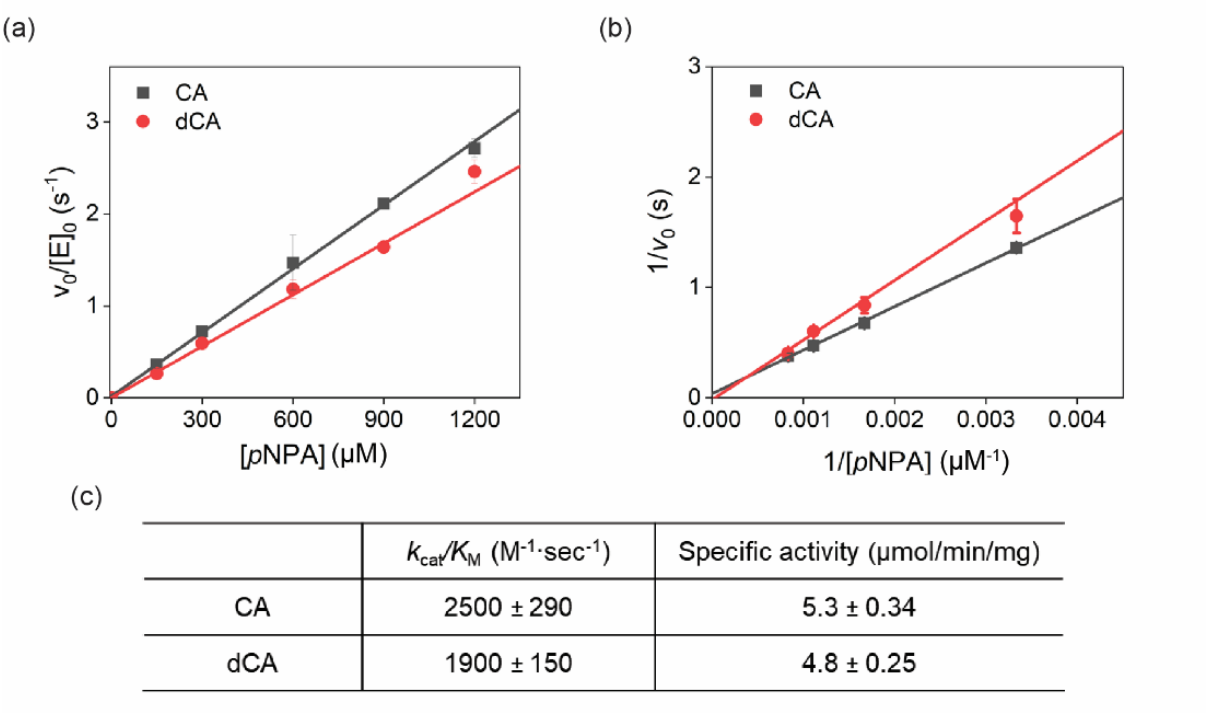


Fig. S6. The hydrolytic activity of CA and dCA with *p*NPA. (a) Catalytic activities upon varying substrate concentrations. (b) The Lineweaver-Burk plot from the data shown in (a). (c) Kinetic parameters from the data in (a) and (b). Standard deviation values were determined from triplicate experiments.



(c)

	k_{cat}/K_M ($M^{-1}\cdot sec^{-1}$)	Specific activity (μmol/min/mg)
CA	2500 ± 290	5.3 ± 0.34
dCA	1900 ± 150	4.8 ± 0.25

Fig. S7. Heteroassembly of CA isoforms and KDGA. (a) DLS data of human carbonic anhydrase isoforms, CA1 (6.0 ± 0.40 nm), CA2 (named as CA in this work; 5.1 ± 0.30 nm), and CA12 (7.4 ± 1.4 nm), where the numbers in parenthesis indicate the determined hydrodynamic diameter. The larger size of CA12 than other isoforms is due to intrinsic dimerization, as reported previously.[14] After mixing the 1:1 ratio of KDGA*-E288C|BzS5 (10 μM monomer) and (b) CA1 and (c) CA12. For comparison, the data for CA and KDGA wild-type protein (mix) and CA and KDGA*-E288C|BzS5 (assembly) were overlaid as the negative and positive control, respectively. The hydrodynamic diameter of assembly in (b) and (c) are detected to be 10 ± 1.6 nm and 13 ± 0.57 nm, respectively.

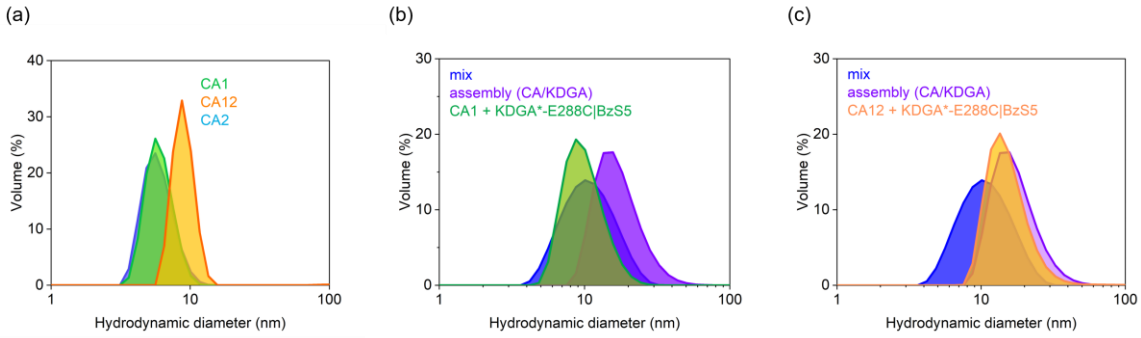


Fig. S8. Biochemical characterizations of CA/KDGA. (a) CD spectra of CA, KDGA*-E288C, and the hetero-assembled complex. (b) Temperature-dependent DLS spectra.

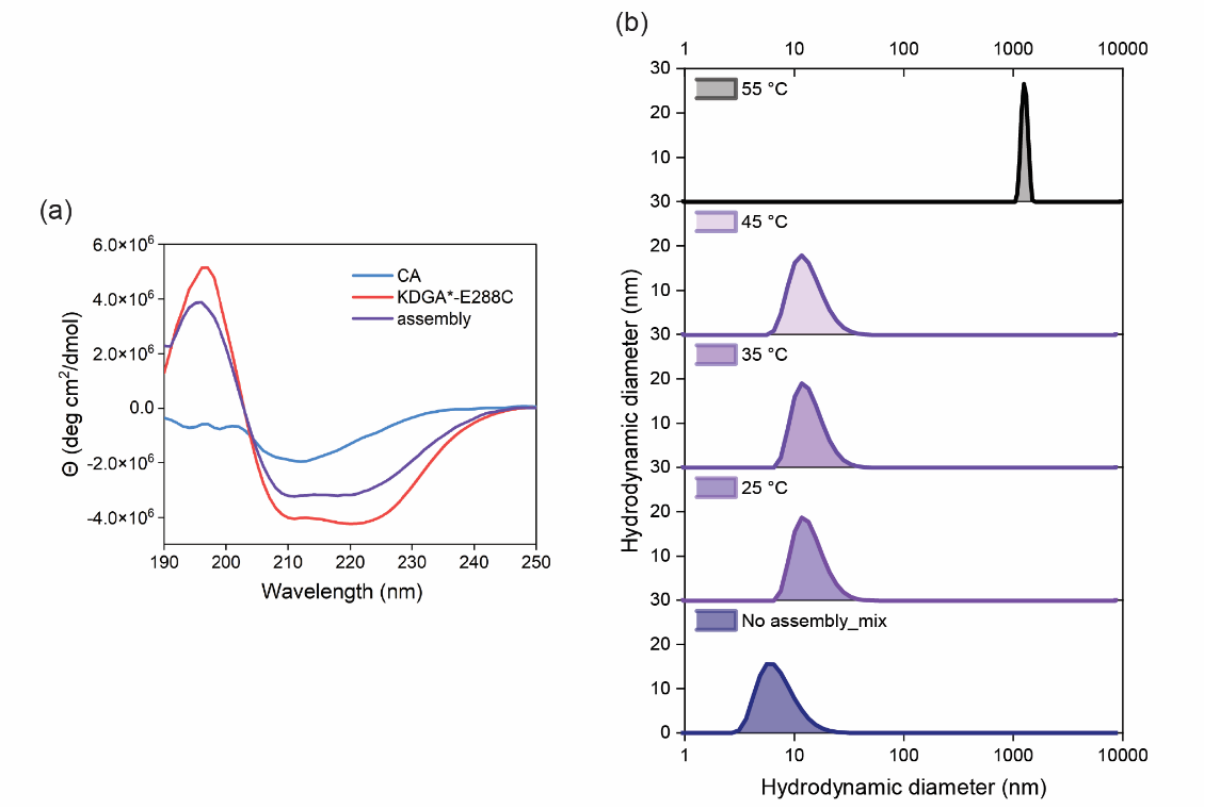


Fig. S9. Expression and purification of dCA. (a) His-tag and (b) size exclusion chromatography. In (b), dimer and monomer were eluted in 57 and 64 min, respectively. 15% SDS-PAGE gel of (c) dCA treated with DTT and (d) $CuCl_2$-mediated dCA formation.

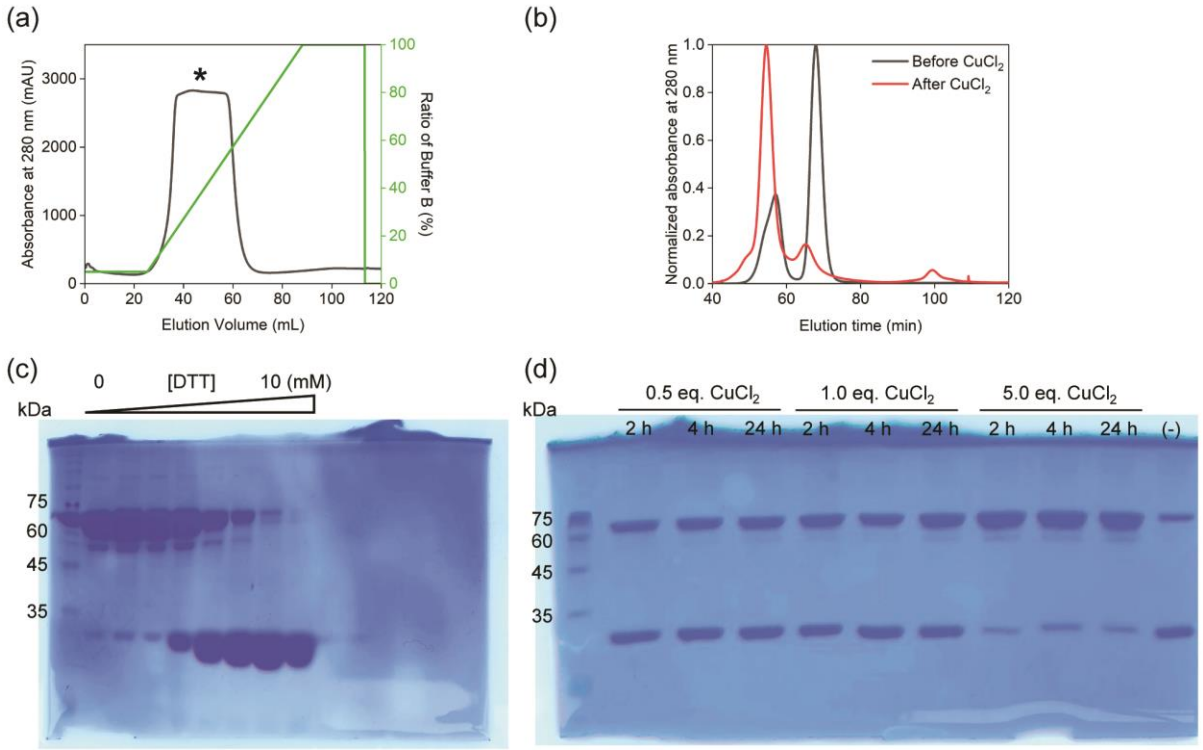


Fig. S10. Hydrolytic activity of dCA with *p*NPA under various conditions. Asterisks indicate the equivalence of added BzS5 protein bioconjugates, KDGA*-E288C|BzS5, calculated as ratio of monomers.

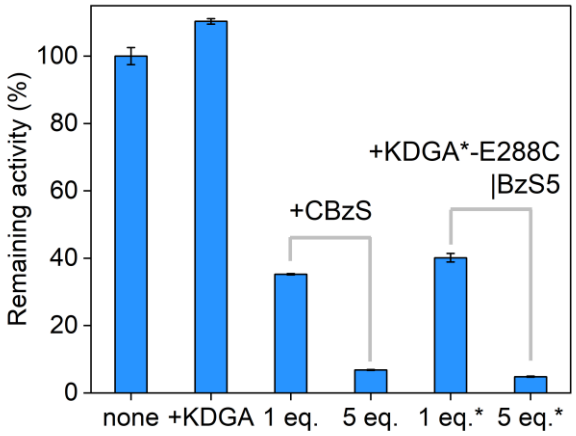


Fig. S11. Preparation of AT*-K44C. (a) The location of the native cysteines (C133, C139, and C244) in the subunit of AT (PDB 3N7Z). The representative elution profiles from (b) His-tag column (c) SEC. (d) 12% SDS-PAGE gel.

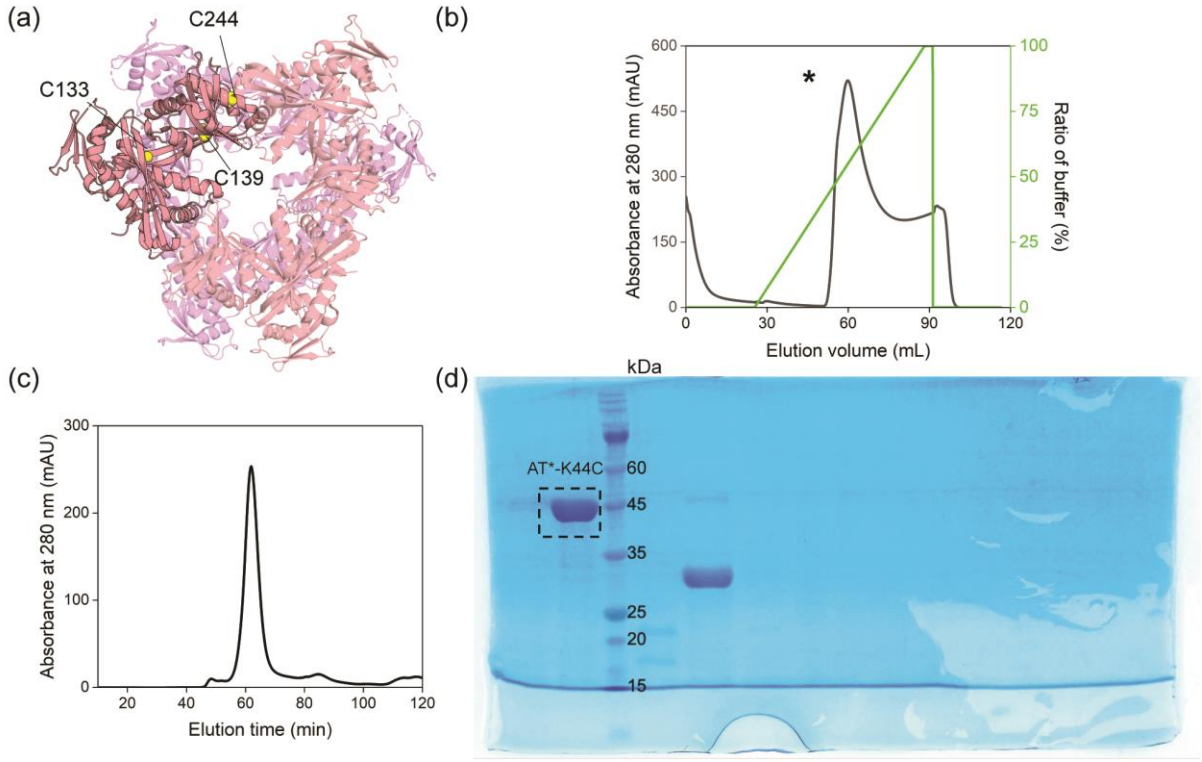


Fig. S12. Preparation of encapsulin (Enc). (a) The location of pre-existing cysteines (C123, C141, C190). C123 is the only surface-accessible cysteine residue. (b) 12% SDS-PAGE gel.

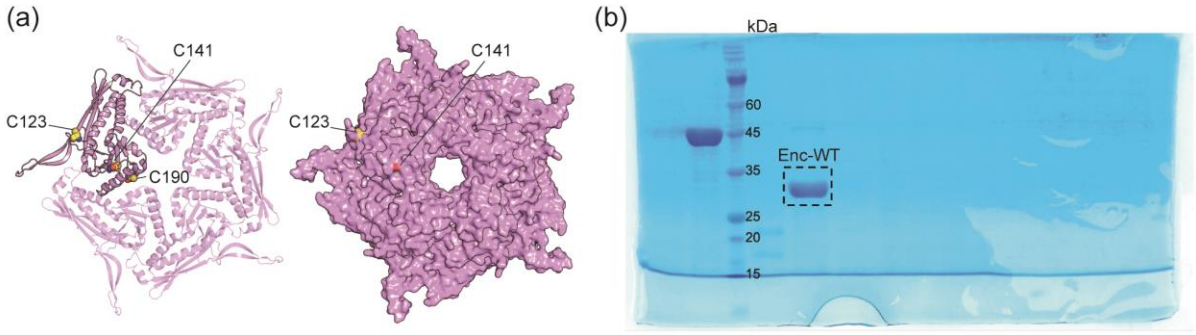


Fig. S13. DLS data of dCA before and after adding (a) wild-type AT and (b) non-conjugated wild-type Enc.

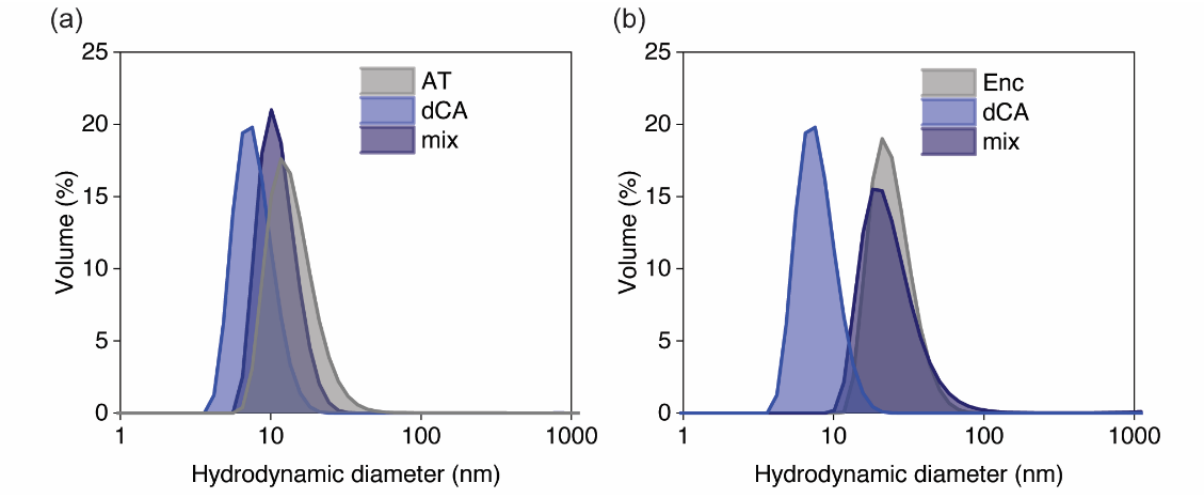


Table S1. Nucleotide sequences of the proteins used in this study. The cut-sites are shown in bold.

Proteins	Nucleotide sequences
Human carbonic anhydrase II (CA)	**CATATG**AGCCATCATTGGGGTTATGGTAAACATAATGGCCCGG AACATTGGCATAAAGATTTTCCGATTGCCAAAGGTGAACGCCA GAGCCCGGTTGATATTGATACCCATACCGCAAAATATGATCCG AGCCTGAAACCGCTGAGTGTGAGCTATGATCAGGCCACCAGTC TGCGTATTCTGAATAATGGCCATGCATTTAATGTTGAATTCGAT GATAGTCAGGATAAAGCAGTGCTGAAAGGCGGTCCGCTGGAT GGTACCTATCGCCTGATTCAGTTTCATTTTCATTGGGGTAGCCT GGATGGTCAGGGCAGTGAACATACCGTTGATAAAAAAAAATA CGCCGCAGAACTGCATCTGGTTCATTGGAATACCAAATATGGC GATTTTGGCAAAGCAGTTCAGCAGCCGGATGGTCTGGCAGTTC TGGGTATTTTTCTGAAAGTGGGTAGTGCAAAACCGGGTCTGCA GAAAGTTGTGGATGTGCTGGATAGCATTAAAACCAAAGGTAA AAGCGCAGATTTTACCAATTTTGATCCGCGTGGTCTGCTGCCG GAAAGTCTGGATTATTGGACCTATCCGGGTAGTCTGACCACCC CGCCGCTGCTGGAATGTGTGACCTGGATTGTTCTGAAAGAACC GATTAGCGTTAGTAGCGAACAGGTTCTGAAATTTCGCAAACTG AATTTTAACGGCGAAGGCGAACCGGAAGAACTGATGGTGGAT AATTGGCGTCCGGCCCAGCCGCTGAAAAATCGTCAGATTAAAG CCAGCTTTAAA**GCGGCCGC**
2-keto-3-deoxygluconate aldolase (KDGA)	**CATATG**CCTGAAATTATCACCCCGATTATTACCCCGTTTACCA AAGATAATCGTATTGATAAAGAGAAGCTGAAAATCCATGCCG AAAATCTGATTCGTAAAGGTATTGATAAGCTGTTTGTTAATGG CACCACCGGCCTGGGTCCGAGTCTGAGTCCGGAAGAAAAACT GGAAAATCTGAAAGCCGTGTATGATGTGACCAATAAGATTATT TTCCAGGTTGGTGGTCTGAATCTGGATGATGCCATTCGTCTGGC AAAACTGAGTAAAGATTTTGATATTGTGGGCATTGCAAGCTAT GCCCCGTATTATTATCCGCGCATGAGCGAAAAACATCTGGTGA AATATTTTAAGACCCTGTGCGAAGTTAGCCCGCATCCGGTTTA TCTGTATAATTATCCGACCGCCACCGGTAAAGATATTGATGCC AAAGTTGCCAAAGAAATTGGCTGTTTTACCGGTGTTAAAGATA CCATTGAAAATATTATCCACACCCTGGATTATAAACGCCTGAA TCCGAATATGCTGGTGTATAGTGGTAGCGATATGCTGATTGCC ACCGTGGCCAGCACCGGCCTGGATGGTAATGTTGCCGCAGGTA GCAATTATCTGCCGGAAGTGACCGTGACCATTAAGAAACTGGC AATGGAACGCAAAATTGATGAAGCACTGAAACTGCAGTTTCTG CATGATGAAGTTATTGAAGCAAGCCGCATTTTTGGTAGTCTGA GTAGCAATTATGTTCTGACCAAATATTTCCAGGGTTATGATCTG GGTTATCCGCGTCCGCCGATTTTTCCGCTGGATGATGAAGAAG AACGCCAGCTGATTAAGAAAGTTGAAGGTATTCGCGCCAAACT GGTTGAACTGAAAATTCTGAAAGAATAA**CTCGAG**
amino-glycoside	**CATATG**AGCAATGCCATGAATGTGATTCGCCTGAAAGAAGAT AAATTTCGCGAAGCACTGCGTCTGAGCGAATATGCATTTCAGT

acetyltransferase (AT)	ATAAAGTGGATGAAGATCGTCTGCAGCAGCAGATTACCAAAA TGAAAGAAAGCCATGAAGTGTATGGCATCATGGAAGGTGAAA ATCTGGCAGCAAAACTGCATCTGATTCCGTTTCATATCTACATC GGCAAAGAAAAATTCAAGATGGGTGGTGTTGCCGGTGTTGCA ACCTATCCGGAATATCGTCGTAGCGGTTATGTTAAAGAACTGC TGCAACATAGCCTGCAGACCATGAAAAAAGATGGTTATACCGT TAGCATGCTGCATCCGTTTGCAGTTAGCTTTTATCGTAAATATG GTTGGGAACTGTGTGCCAATCTGCTGGTTTGTCACATGACCAA AAGCGATCTGGTTATGAAAAAACAGGTTAACGGCACCGTGAA ACGCTTTAACAAAGAAAGTCATCCGGAAGAGGTGGAAAAACT GTATGAAACCTTTGCAGAACTGTTTAGCGGTATGCTGGTTCGT AATGAAAAATGGTGGCTGCAGGCAGTTTATGATGATCTGACCC TGGCAATCTATTATGATGAAAATCAGACCGCAGCAGGCTACAT GCTGTATAAAATCGAGAACTATAAGATGACCGTGGAAGAATTT GTTCCGCTGCATAATGAAGCACGTAATGGTCTGTGGAACTTTA TTTGTCAGCATGATAGCATGATCAAAGATCTGGAAATGACCGT GAGCGAAAATGAACCGCTGCTGTATACCCTGCAAGAACCGCGT GTTAAAACCGAAATTAAACCGTATTTTATGGGTCGCATTGTGG ATGTTGAACAGTTCCTGAAACAGTATGAACTGAATTGGAATAA CGTGCAGCAAGAAGTGATTCTGCATATCACCGATAGCTTTGCA CAGTGGAATAACATTACCGTTCGTATTGCCAACCATGAGATTA CCATTATTGAAGAACCGATCGACAAAGGCATCAAACTGGATAT TAATGCACTGAGCACCATCCTGTTTGGTTATCGTCGTCCGCTGG AACTGAATGAATTAGAACTGATTAGTGGCAGCGAAGAAGAAA TTCGCGCATTTGAAAGCGTTGTTCCGGTTCGTAAACCGTTCATC TATGACTTTTTCTAA**CTCGAG**
encapsulin (Enc)	**CCATGG**AATTCCTGAAACGTAGTTTTGCACCGCTGACCGAAAA ACAGTGGCAGGAAATTGATAATCGCGCACGTGAAATTTTTAAA ACCCAGCTGTATGGCCGTAAATTTGTTGATGTTGAAGGCCCGT ATGGCTGGGAATATGCAGCCCATCCGCTGGGCGAAGTGGAAG TTCTGAGCGATGAAAATGAAGTGGTTAAATGGGGCCTGCGCAA AAGTCTGCCGCTGATTGAACTGCGCGCAACCTTTACCCTGGAT CTGTGGGAACTGGATAATCTGGAACGCGGCAAACCGAATGTG GATCTGAGCAGTCTGGAAGAAACCGTTCGCAAAGTTGCCGAAT TTGAAGATGAAGTGATTTTTCGCGGTTGCGAAAAAAGTGGTGT TAAAGGTCTGCTGAGTTTTGAAGAACGCAAAATTGAATGCGGT AGCACCCCGAAAGATCTGCTGGAAGCAATTGTGCGCGCCCTGA GTATTTTTAGCAAAGATGGCATTGAAGGCCCGTACACCCTGGT TATTAATACCGATCGTTGGATTAATTTCCTGAAAGAAGAAGCA GGTCATTATCCGCTGGAAAAACGCGTTGAAGAATGTCTGCGCG GTGGCAAAATTATTACCACCCCGCGCATTGAAGATGCACTGGT TGTGAGTGAACGCGGCGGTGACTTTAAACTGATTCTGGGCCAG GATCTGAGCATTGGTTATGAAGATCGCGAAAAAGATGCAGTGC GTCTGTTTATTACCGAAACCTTTACCTTTCAGGTGGTTAATCCG GAAGCACTGATTCTGCTGAAATTTTAA**CTCGAG**

Table S2. Amino acid sequences of the proteins used in this study.

Proteins	Amino acid sequences
human carbonic anhydrase II (CA)	MSHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDTHTAKYD PSLKPLSVSYDQATSLRILNNGHAFNVEFDDSQDKAVLKGGPL DGTYRLIQFHFHWGSLDGQGSEHTVDKKKYAAELHLVHWNT KYGDFGKAVQQPDGLAVLGIFLKVGSAKPGLQKVVDVLDSIK TKGKSADFTNFDPRGLLPESLDYWTYPGSLTTPPLLECVTWIV LKEPISVSSEQVLKFRKLNFNGEGEPEELMVDNWRPAQPLKNR QIKASFKAA
2-keto-3-deoxygluconate aldolase (KDGA)	MPEIITPIITPFTKDNRIDKEKLKIHAENLIRKGIDKLFVNGTTGL GPSLSPEEKLENLKAVYDVTNKIIFQVGGLNLDDAIRLAKLSK DFDIVGIASYAPYYYPRMSEKHLVKYFKTLCEVSPHPVYLYNY PTATGKDIDAKVAKEIGCFTGVKDTIENIIHTLDYKRLNPNML VYSGSDMLIATVASTGLDGNVAAGSNYLPEVTVTIKKLAMER KIDEALKLQFLHDEVIEASRIFGSLSSNYVLTKYFQGYDLGYPR PPIFPLDDEEERQLIKKVEGIRAKLVELKILKE
amino-glycoside acetyltransferase (AT)	MSNAMNVIRLKEDKFREALRLSEYAFQYKVDEDRLQQQITKM KESHEVYGIMEGENLAAKLHLIPFHIYIGKEKFKMGGVAGVAT YPEYRRSGYVKELLQHSLQTMKKDGYTVSMLHPFAVSFYRK YGWELCANLLVCHMTKSDLVMKKQVNGTVKRFNKESHPEE VEKLYETFAELFSGMLVRNEKWWLQAVYDDLTLAIYYDENQ TAAGYMLYKIENYKMTVEEFVPLHNEARNGLWNFICQHDSMI KDLEMTVSENEPLLYTLQEPRVKTEIKPYFMGRIVDVEQFLKQ YELNWNNVQQEVILHITDSFAQWNNITVRIANHEITIIEEPIDKG IKLDINALSTILFGYRRPLELNELELISGSEEEIRAFESVVPVRKP FIYDFF
encapsulin (Enc)	MEFLKRSFAPLTEKQWQEIDNRAREIFKTQLYGRKFVDVEGPY GWEYAAHPLGEVEVLSDENEVVKWGLRKSLPLIELRATFTLD LWELDNLERGKPNVDLSSLEETVRKVAEFEDEVIFRGCEKSGV KGLLSFEERKIECGSTPKDLLEAIVRALSIFSKDGIEGPYTLVIN TDRWINFLKEEAGHYPLEKRVEECLRGGKIITTPRIEDALVVSE RGGDFKLILGQDLSIGYEDREKDAVRLFITETFTFQVVNPEALI LLKF

Table S3. Primer sequences used for the site-directed mutagenesis of CA.

Mutation	5′-sequence-3′ (F = forward, R = reverse)
C206S	F: GTC TGC TGC CGT GCA GTC TGG ATT ATT G R: CAA TAA TCC AGA CTG CAC GGC AGC AGA C
E187C	F: CGC TGC TGG AAT CCG TGA CCT GGA TTG R: CAA TCC AGG TCA CGG ATT CCA GCA GCG
TEV cut site	F: GAA AAC CTG TAT TTT CAG AGC CAC CAC CAC CAC CAC CAC TGA GAT CCG GC R: GCT CTG AAA ATA CAG GTT TTC TTT AAA GCT GGC TTT AAT CTG ACG ATT TTT CAG CGG CTG GG

Table S4. Primer sequences used for site-directed mutagenesis of KDGA.

Mutation	5′-sequence-3′ (F = forward, R = reverse)
C120A	F: GAA ATA TTT TAA GAC CCT GGC GGA AGT TAG CCC R: GGG CTA ACT TCC GCC AGG GTC TTA AAA TAT TTC
C150A	F: GCC AAA GAA ATT GGC GCG TTT ACC GGT GTT AAA G R: CTT TAA CAC CGG TAA ACG CGC CAA TTT CTT TGG C
E288C	F: CGC GCC AAA CTG GTT TGC CTG AAA ATT CTG AAA G R: CTT TCA GAA TTT TCA GGC AAA CCA GTT TGG CGC G

Table S5. The number of cysteine residues per monomer. Theoretical and experimental values are derived from protein sequence and free thiol assay, respectively.

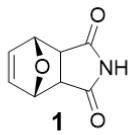
Protein samples	Theoretical	Experimental
KDGA WT	2	0.096 ± 0.041
KDGA*(C120A/C150A)	0	< 0.01
KDGA*-E288C|BzS3	0	< 0.01
KDGA*-E288C|BzS5	0	< 0.01
dCA (C206S/E187C)	0	0.096 ± 0.0058
AT*(C133S/C139S/C244S)-K44C	1	1.0 ± 0.09
AT*-K44C|BzS5	0	< 0.01
Enc-WT	3	1.0 ± 0.09
Enc|BzS5	2	< 0.01

Table S6. Primer sequences used for a site-directed mutagenesis of AT.

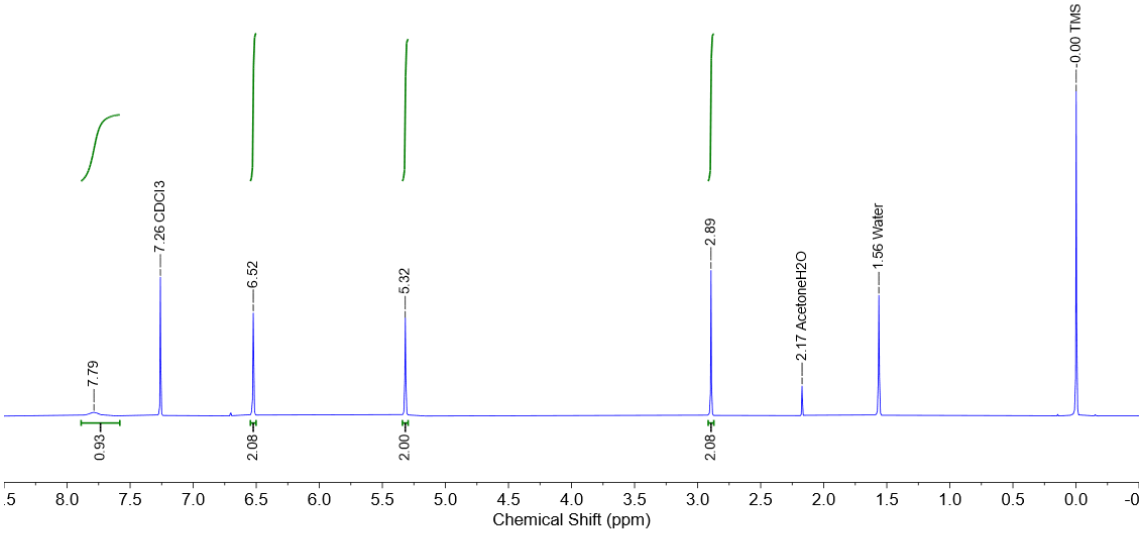
Mutation	5′-sequence-3′ (F = forward, R = reverse)
C133S	F: GTT GGG AAC TGA GCG CCA ATC TGC R: GCA GAT TGG CGC TCA GTT CCC AAC
C139S	F: CTG CTG GTT AGC CAC ATG ACC AAA AG R: CTT TTG GTC ATG TGG CTA ACC AGC AG
C244S	F: GGA ACT TTA TTA GCC AGC ATG ATA GCA TG R: CAT GCT ATC ATG CTG CGT AAT AAA GTT CC
K44C	F: GCA GAT TAC CAA AAT GTG CGA AAG CCA TGA AGT G R: CAC TTC ATG GCT TTC GCA CAT TTT GGT AAT CTG C

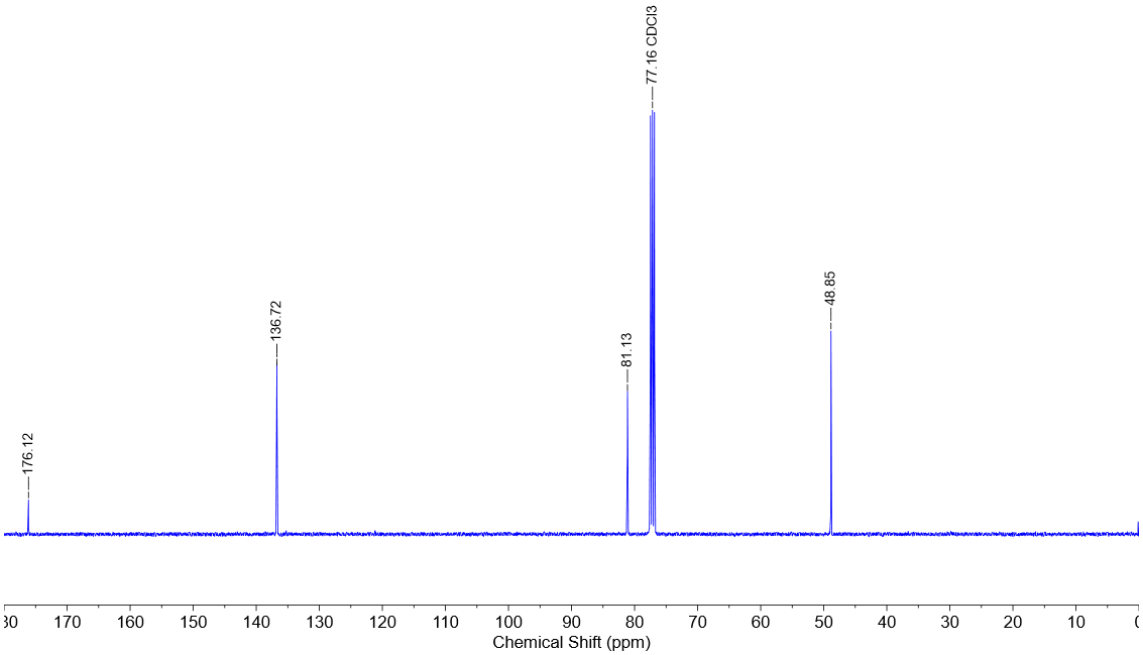
4. NMR Spectra of the Synthesized Molecules

3,6-epoxy-1,2,3,6-tetrahydrophthalimide (1)

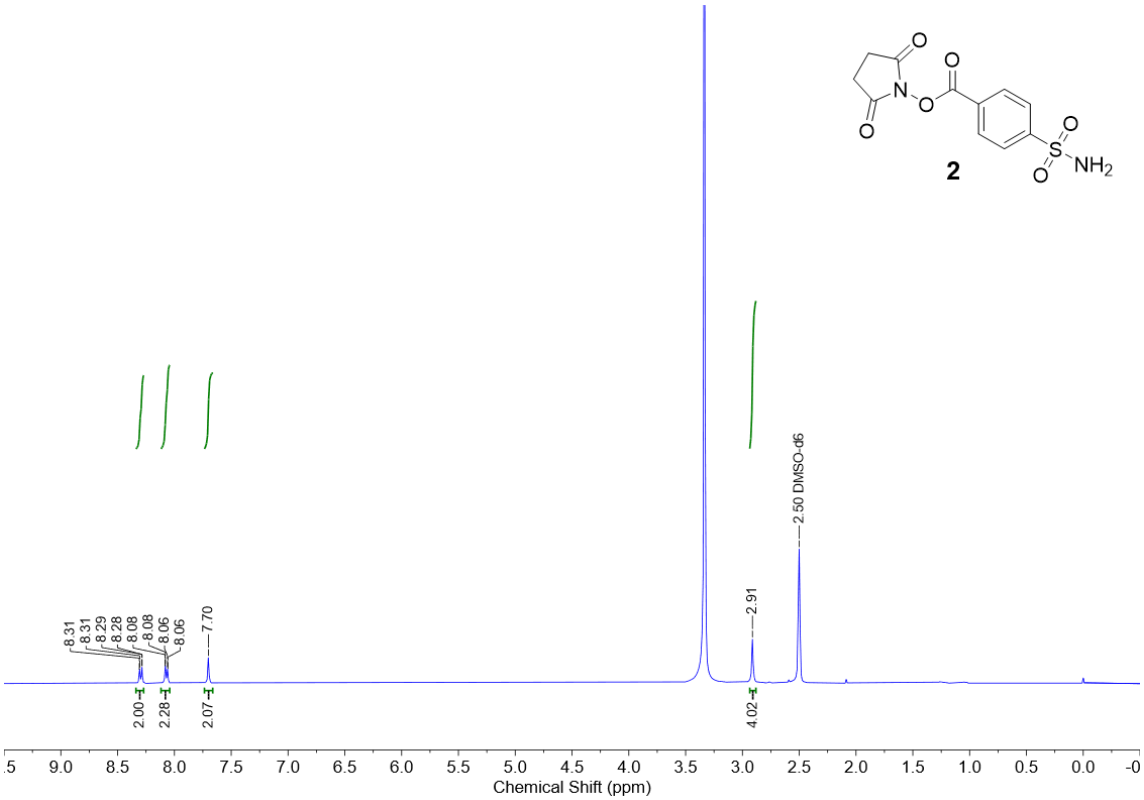


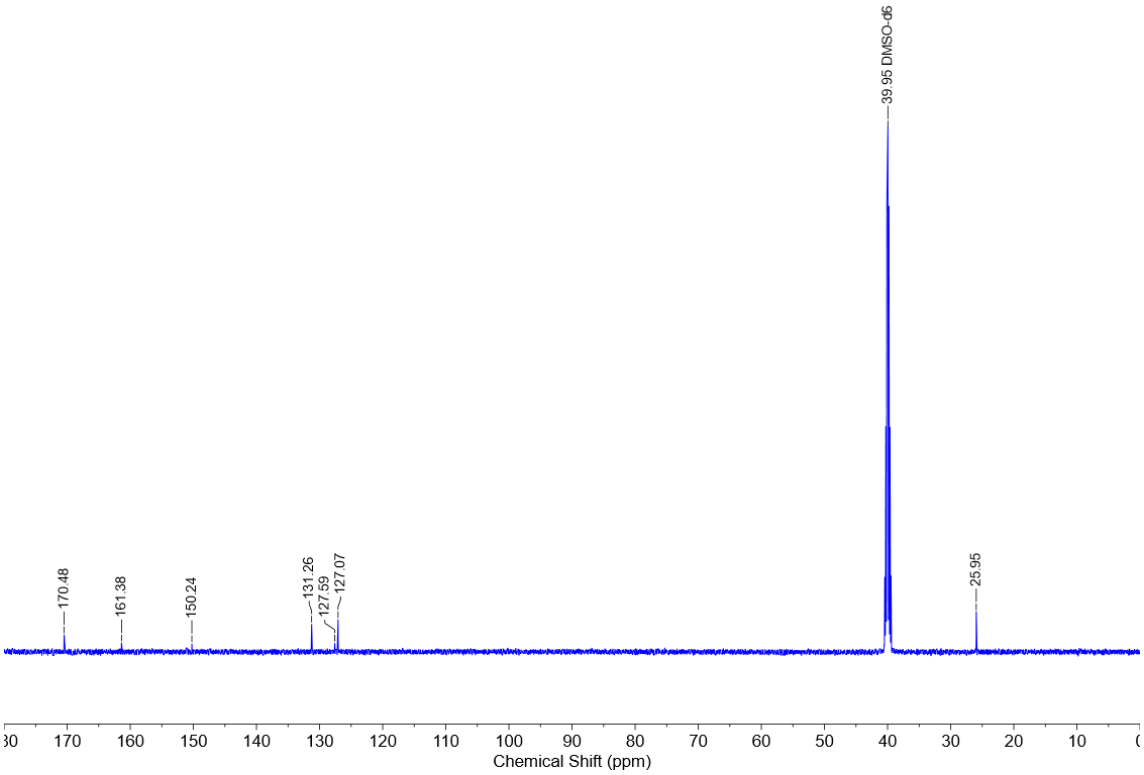
1



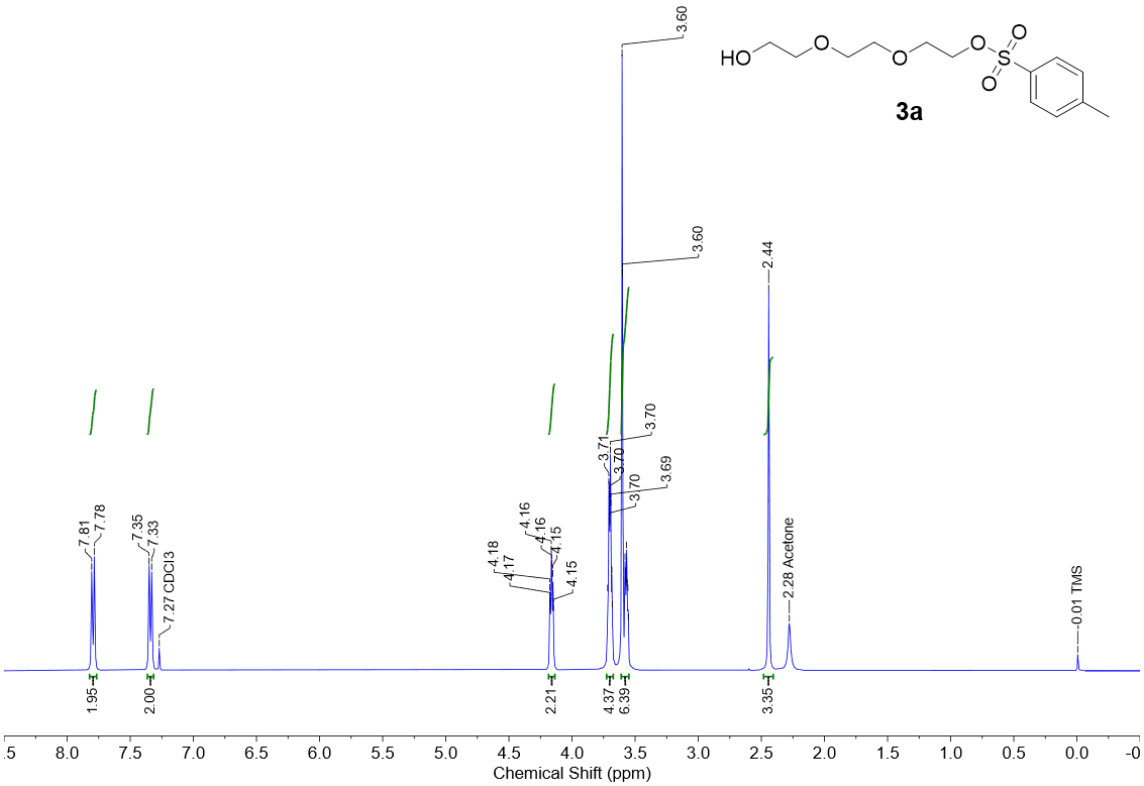


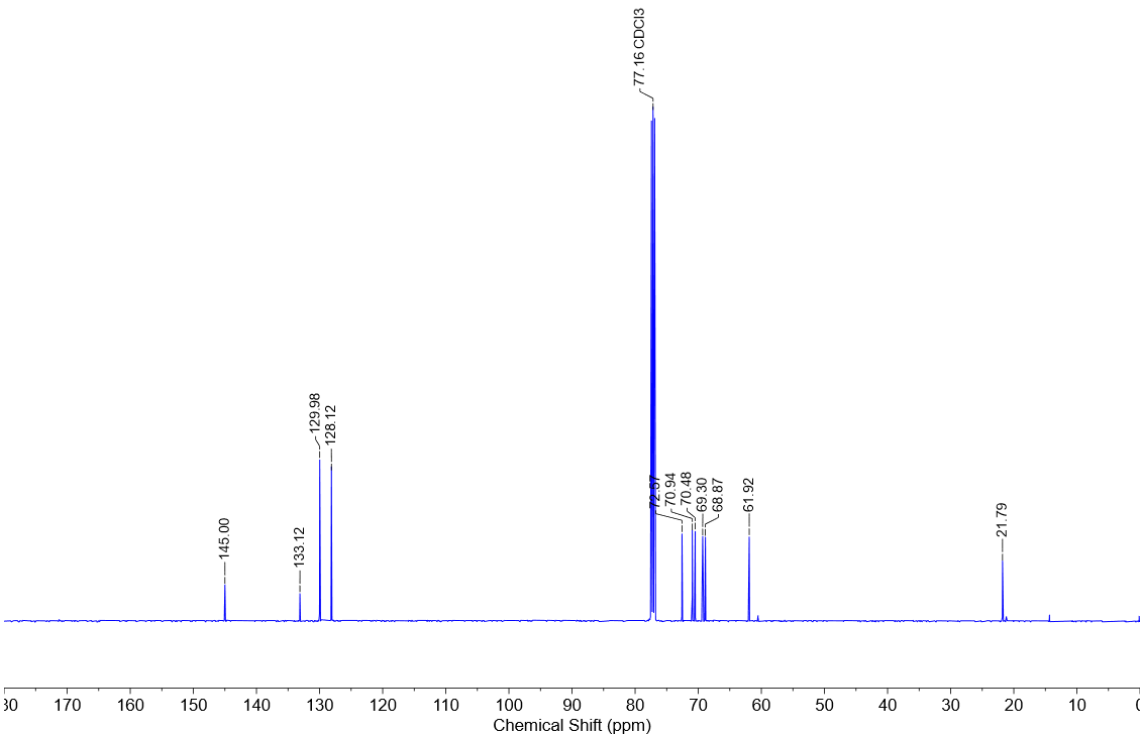
2,5-dioxopyrrolidin-1-yl 4-sulfamoylbenzoate (2)



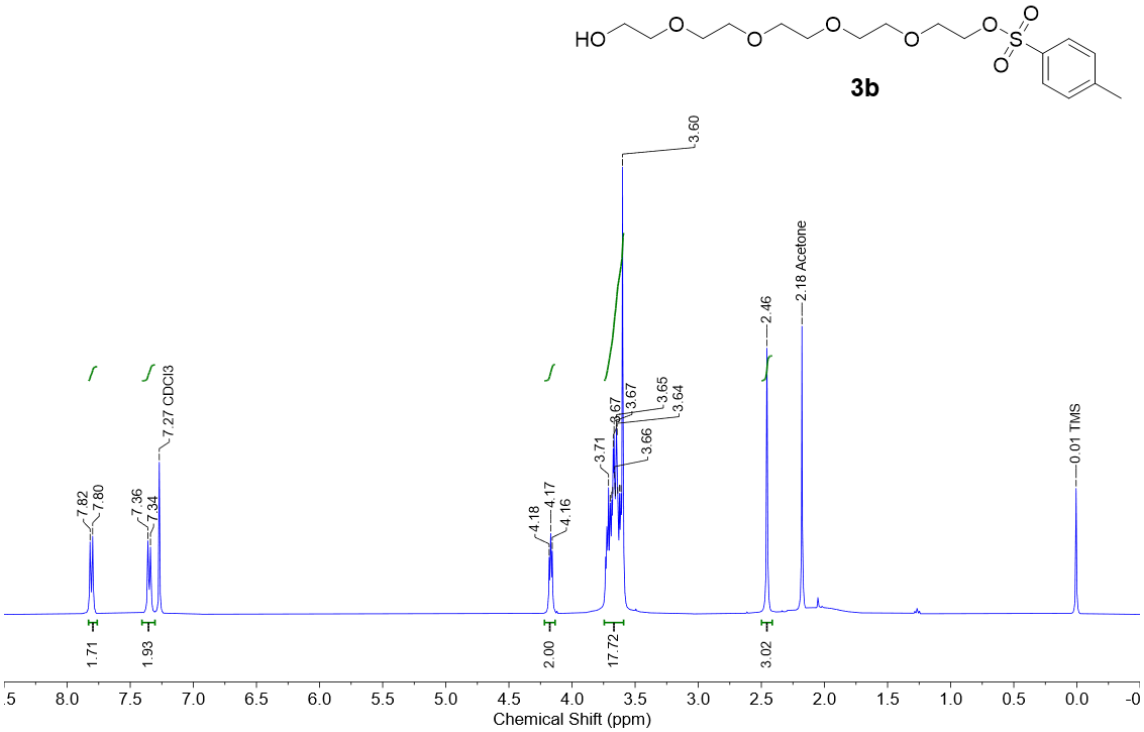


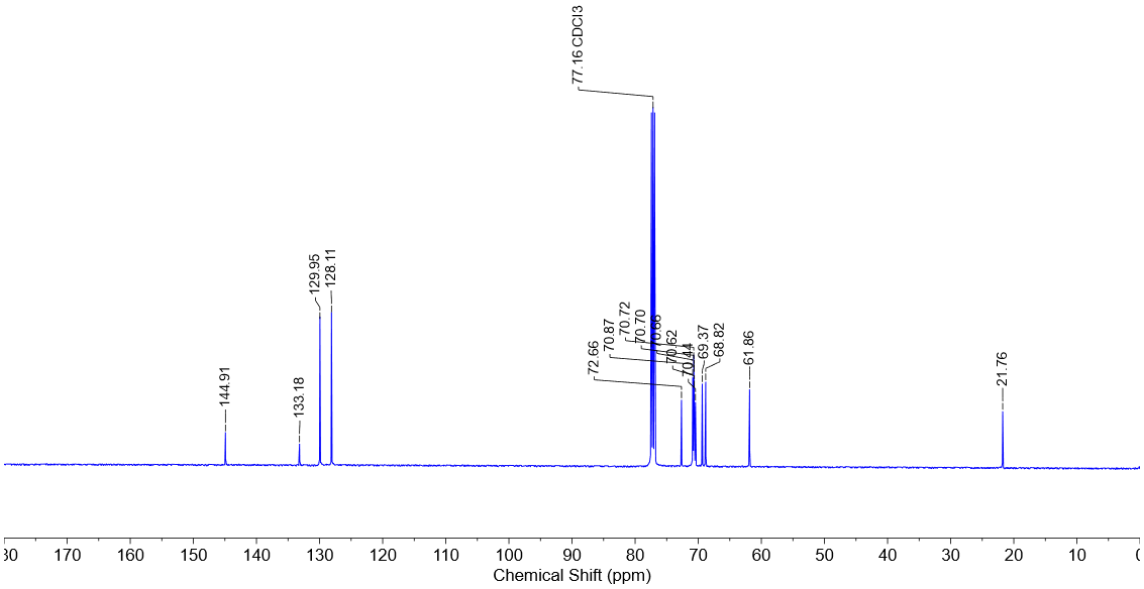
2-(2-(2-hydroxyethoxy)ethoxy)ethyl 4-methylbenzenesulfonate (3a)



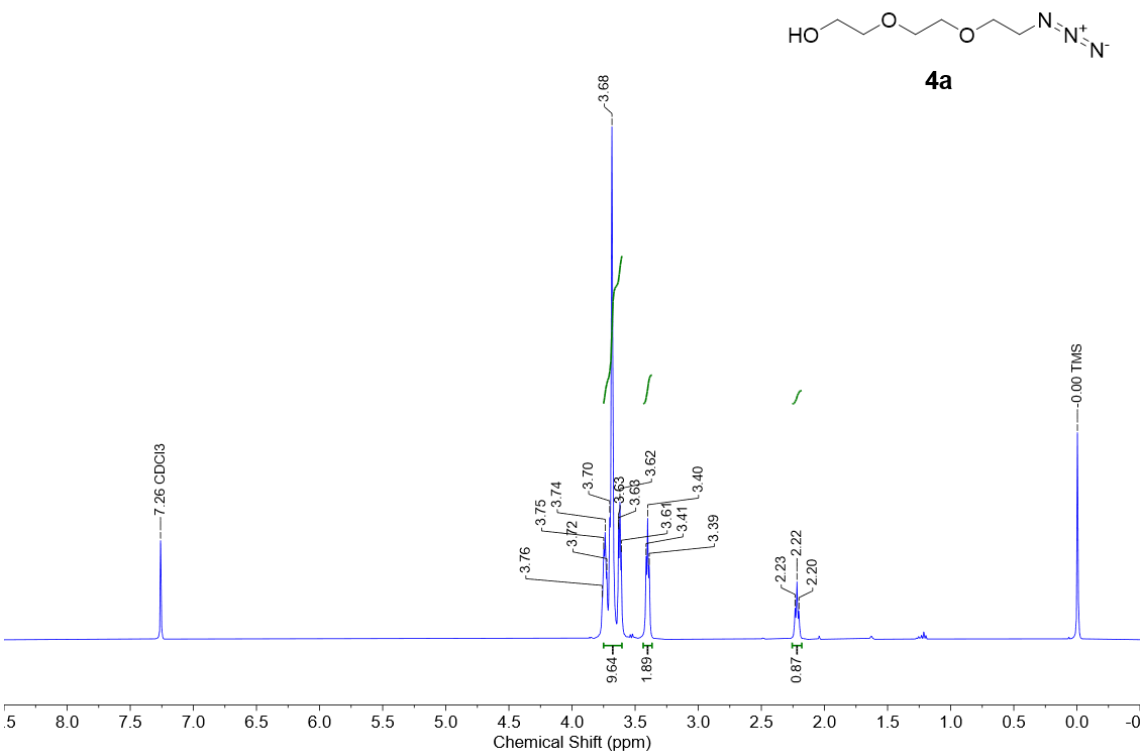


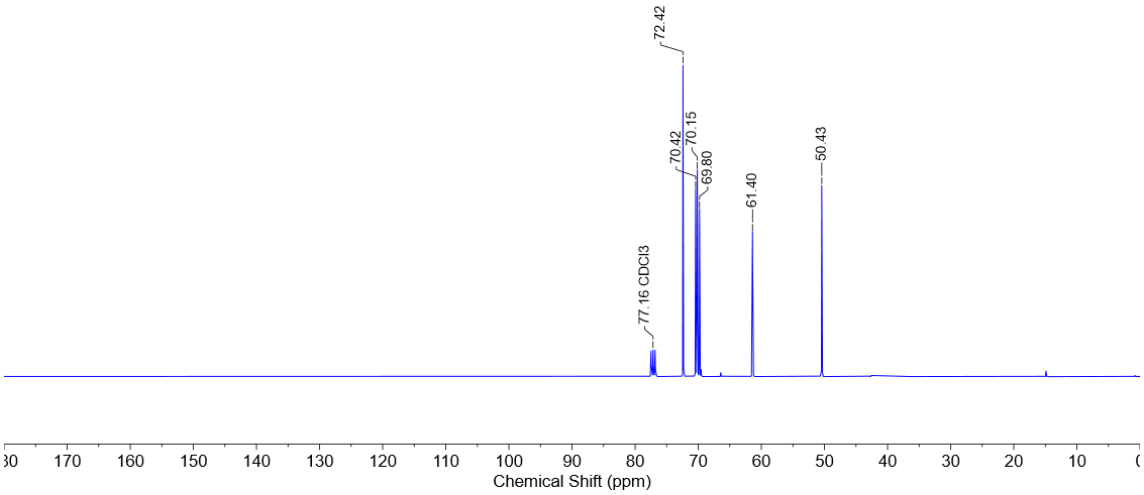
14-hydroxy-3,6,9,12-tetraoxatetradecyl 4-methylbenzenesulfonate (3b)



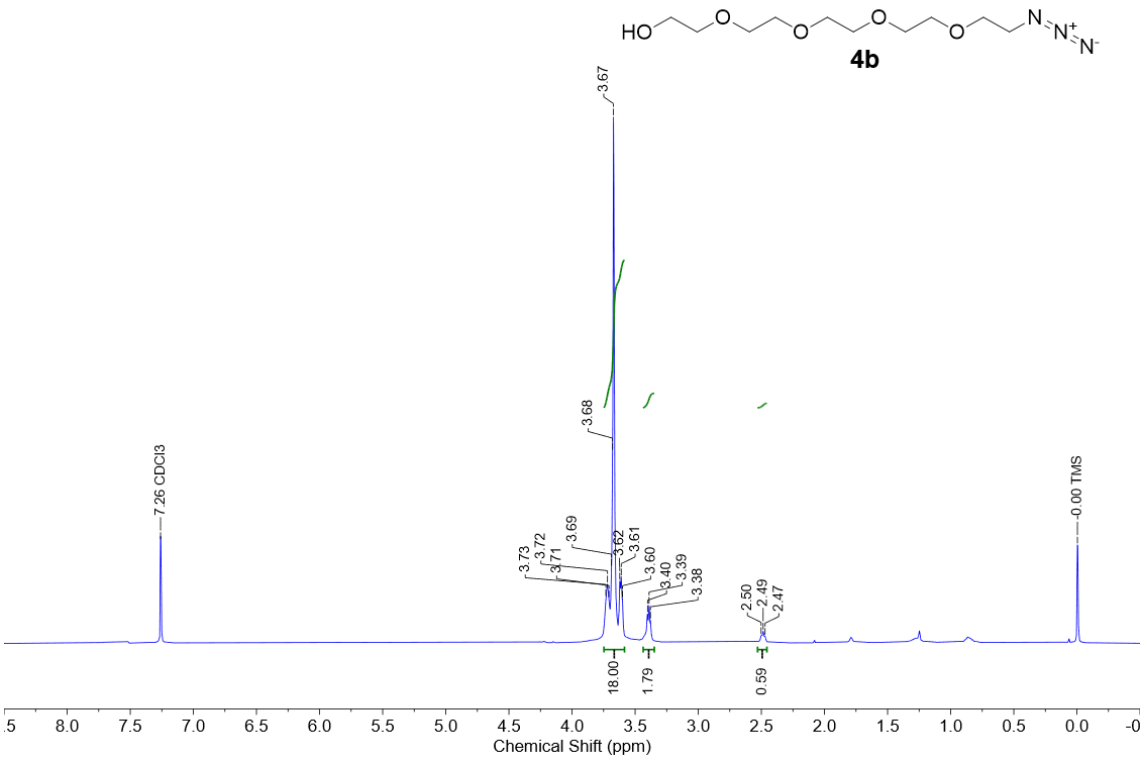


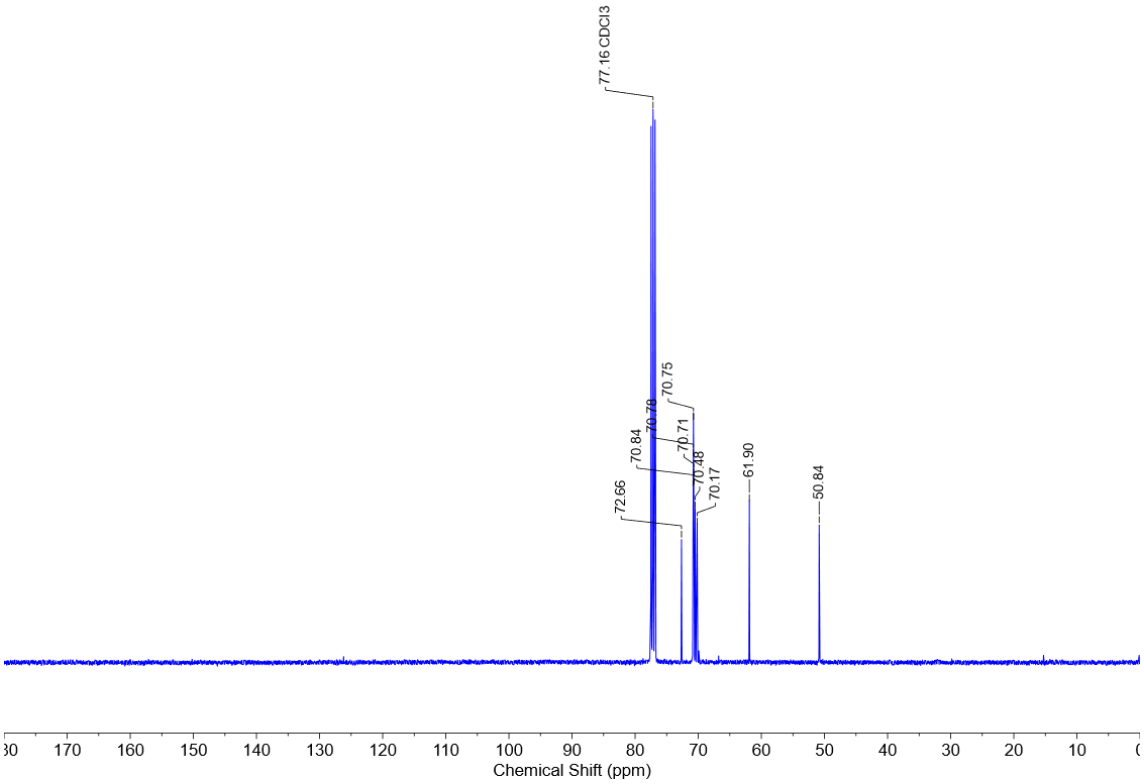
2-(2-(2-azidoethoxy)ethoxy)ethan-1-ol (4a)



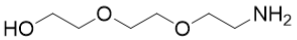


14-azido-3,6,9,12-tetraoxatetradecan-1-ol (4b)

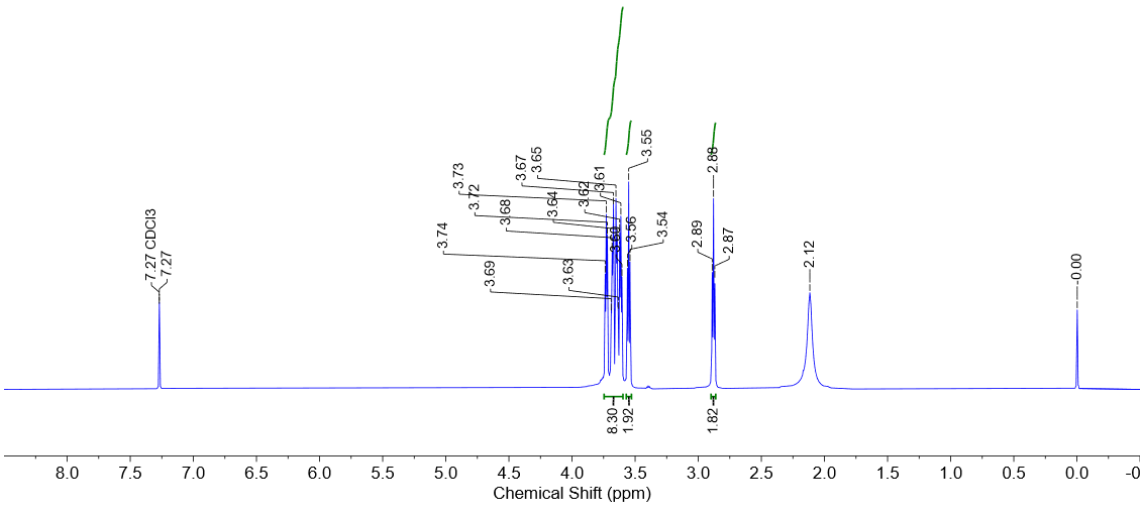


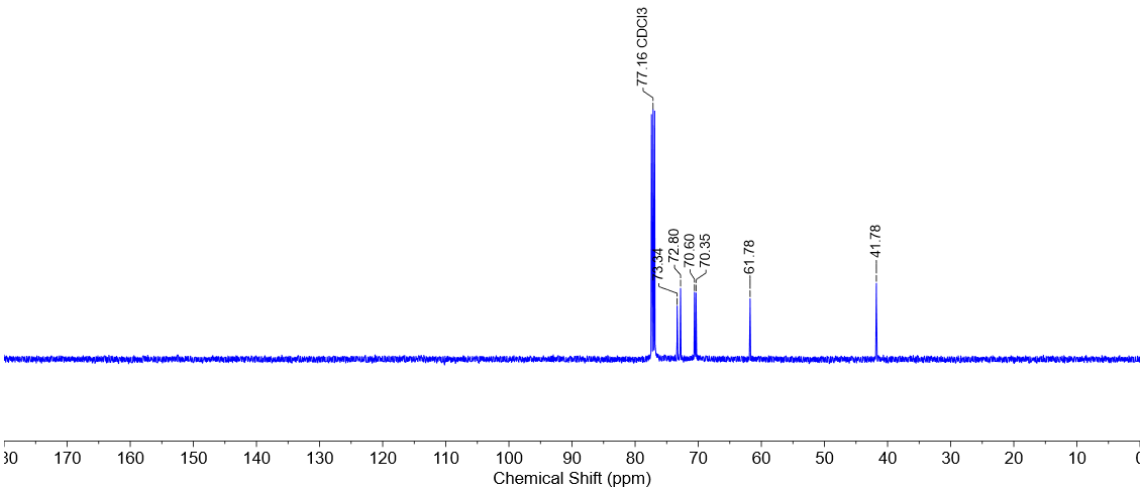


2-(2-(2-aminoethoxy)ethoxy)ethan-1-ol (5a)

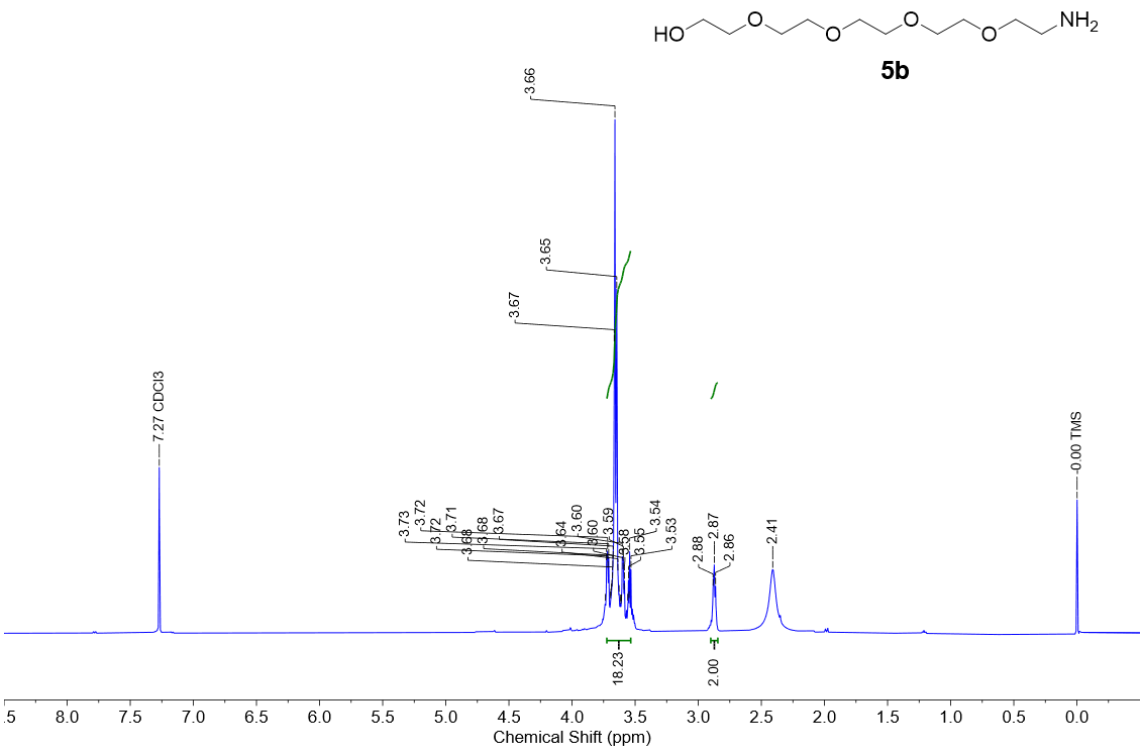


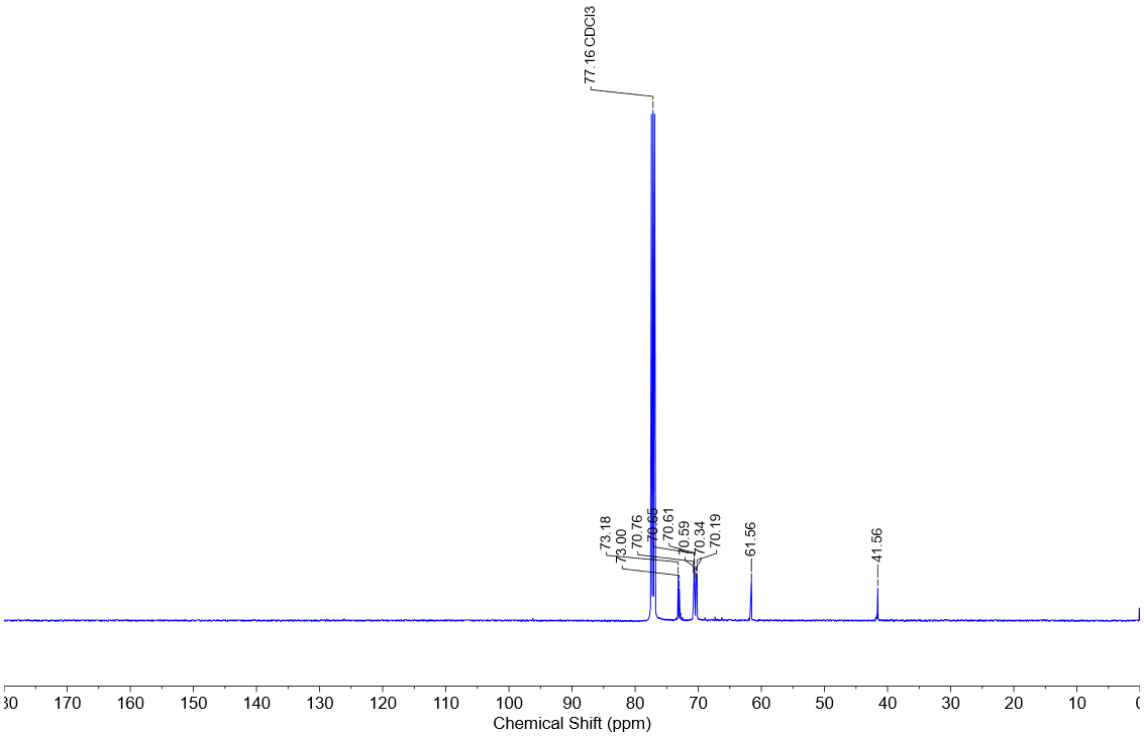
5a



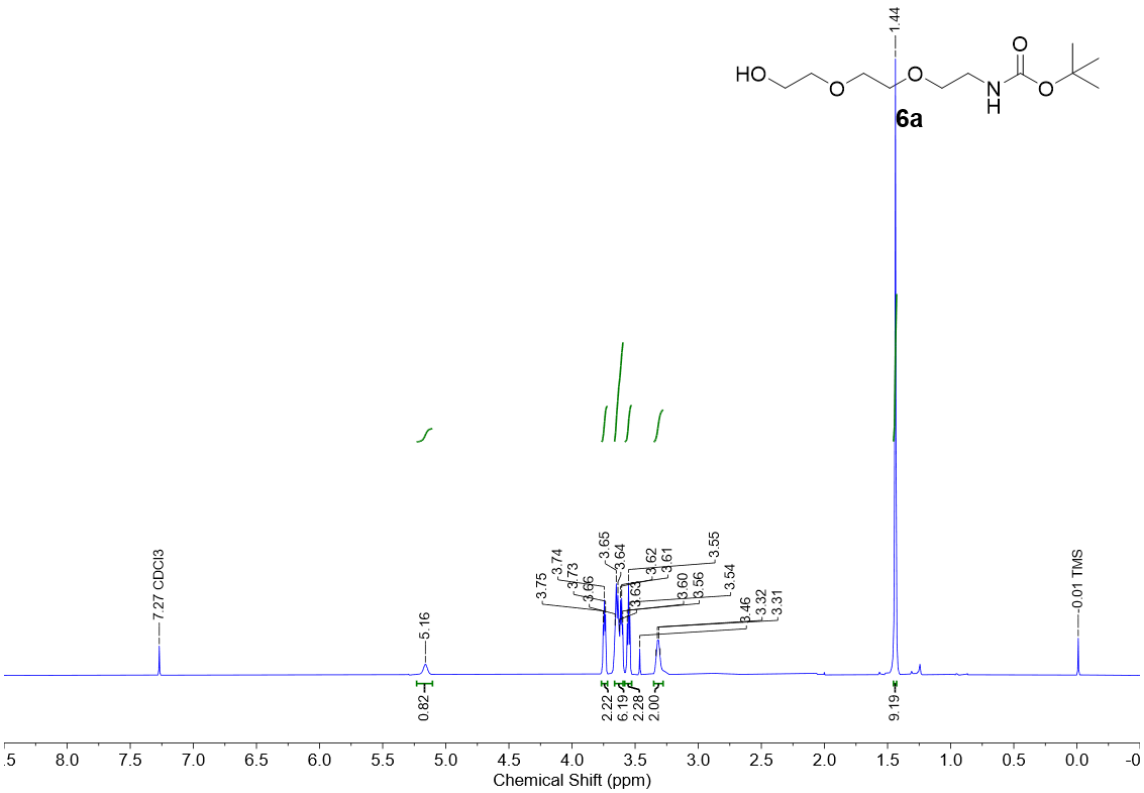


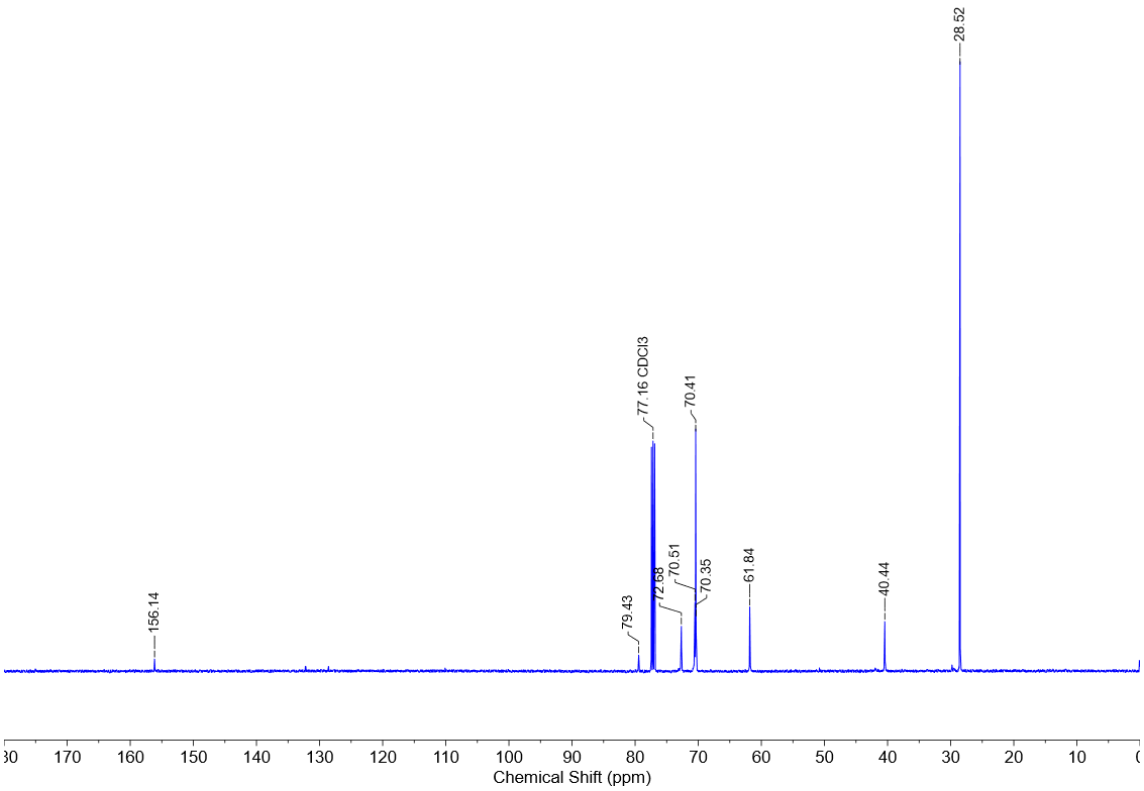
14-amino-3,6,9,12-tetraoxatetradecan-1-ol (5b)



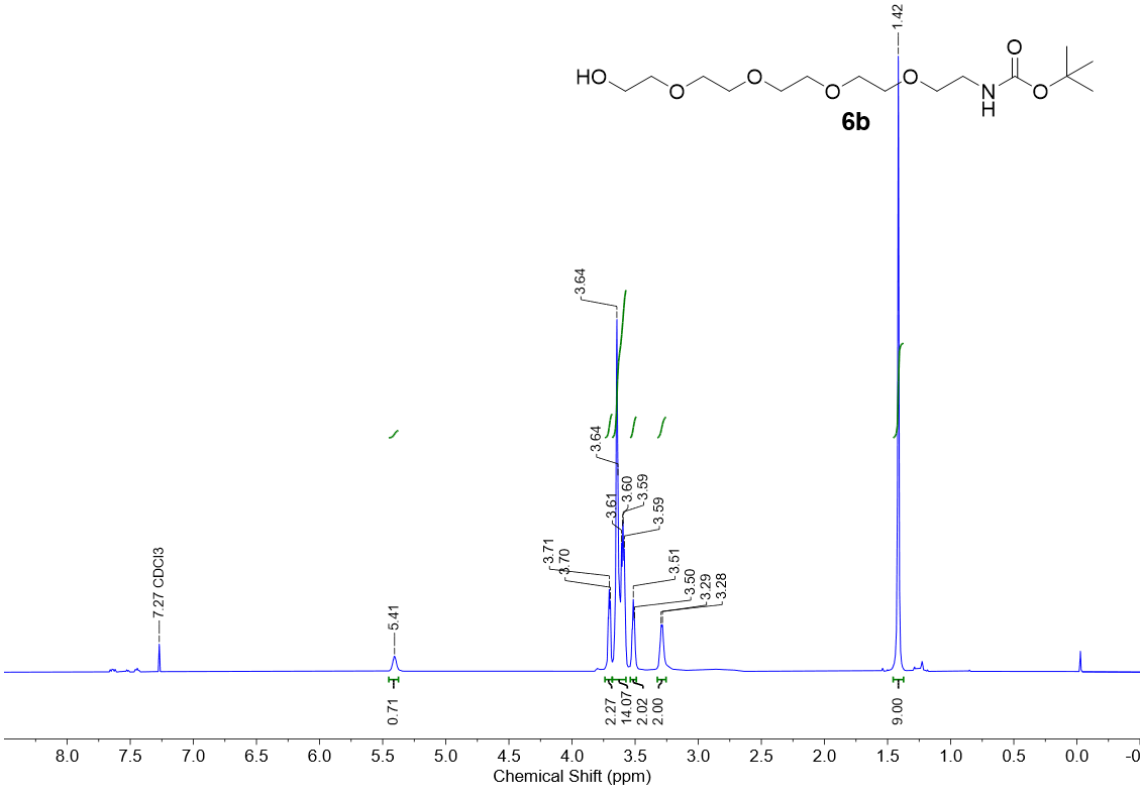


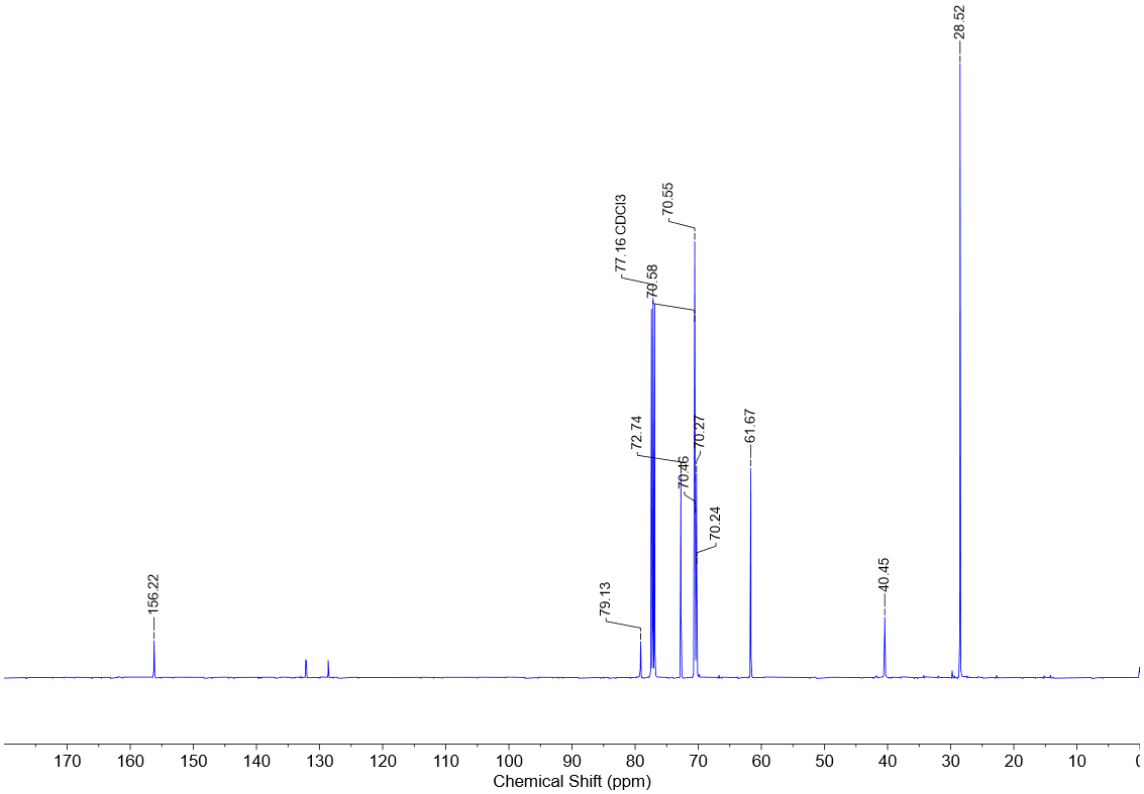
tert-butyl (2-(2-(2-hydroxyethoxy)ethoxy)ethyl)carbamate (6a)



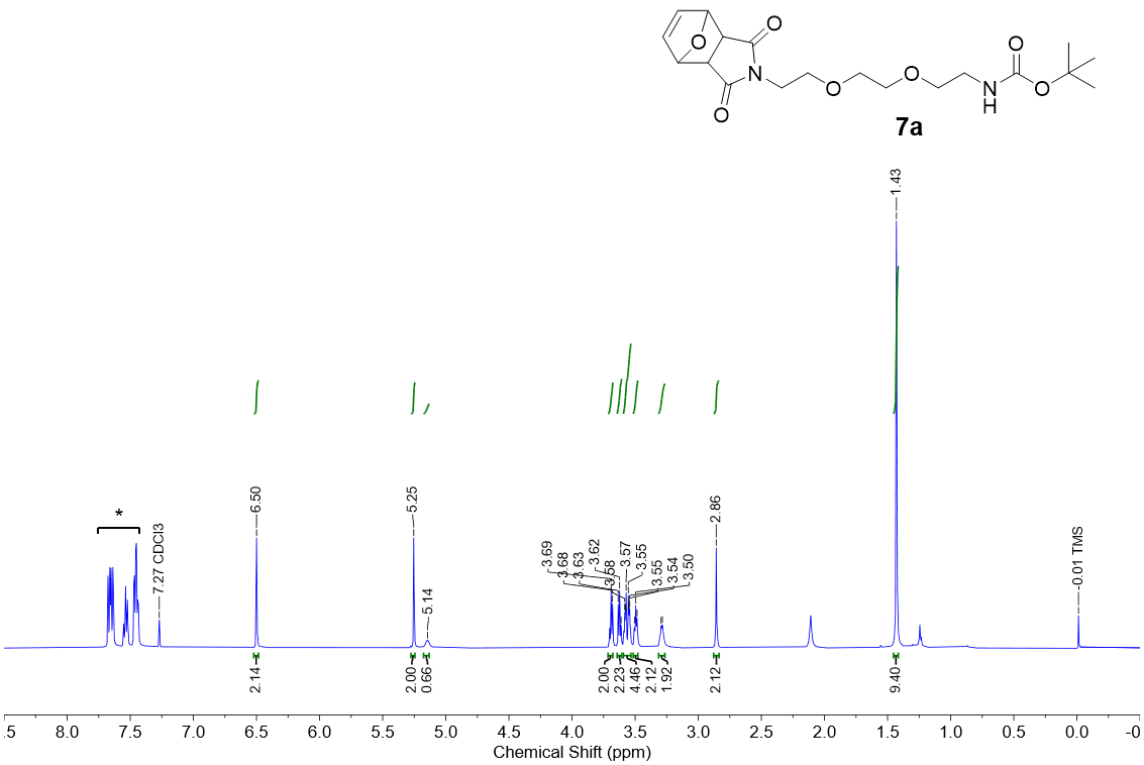


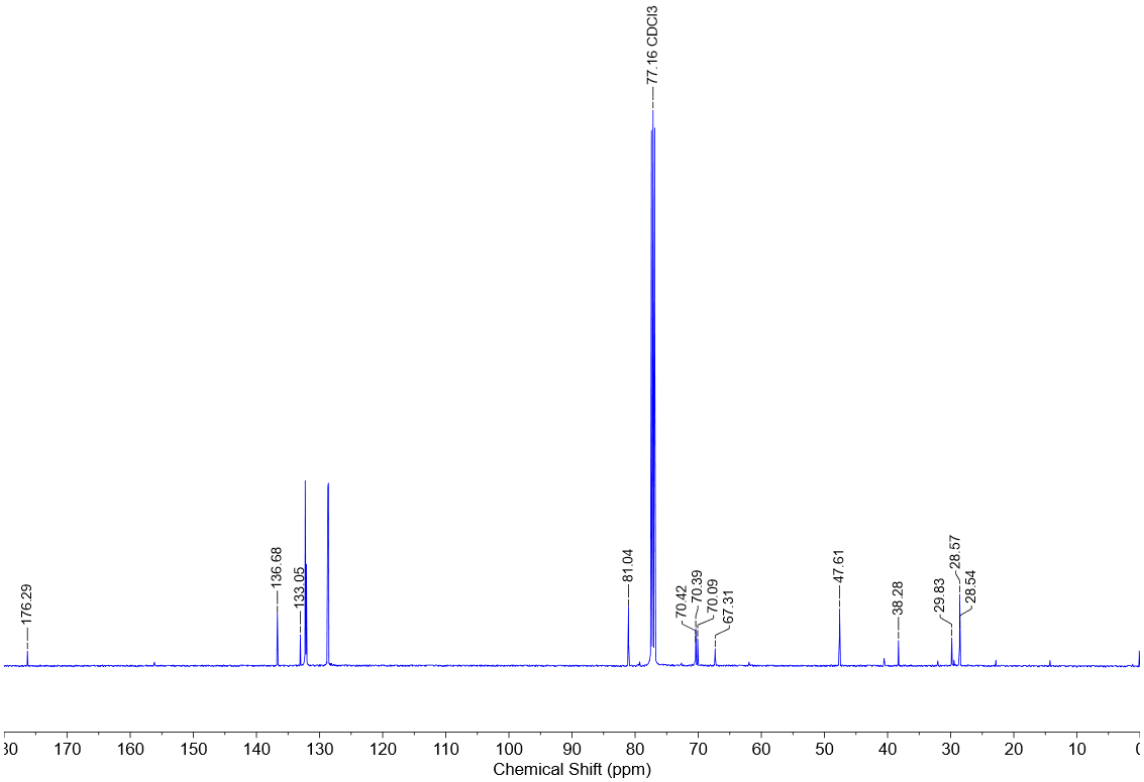
tert-butyl (14-hydroxy-3,6,9,12-tetraoxatetradecyl)carbamate (6b)



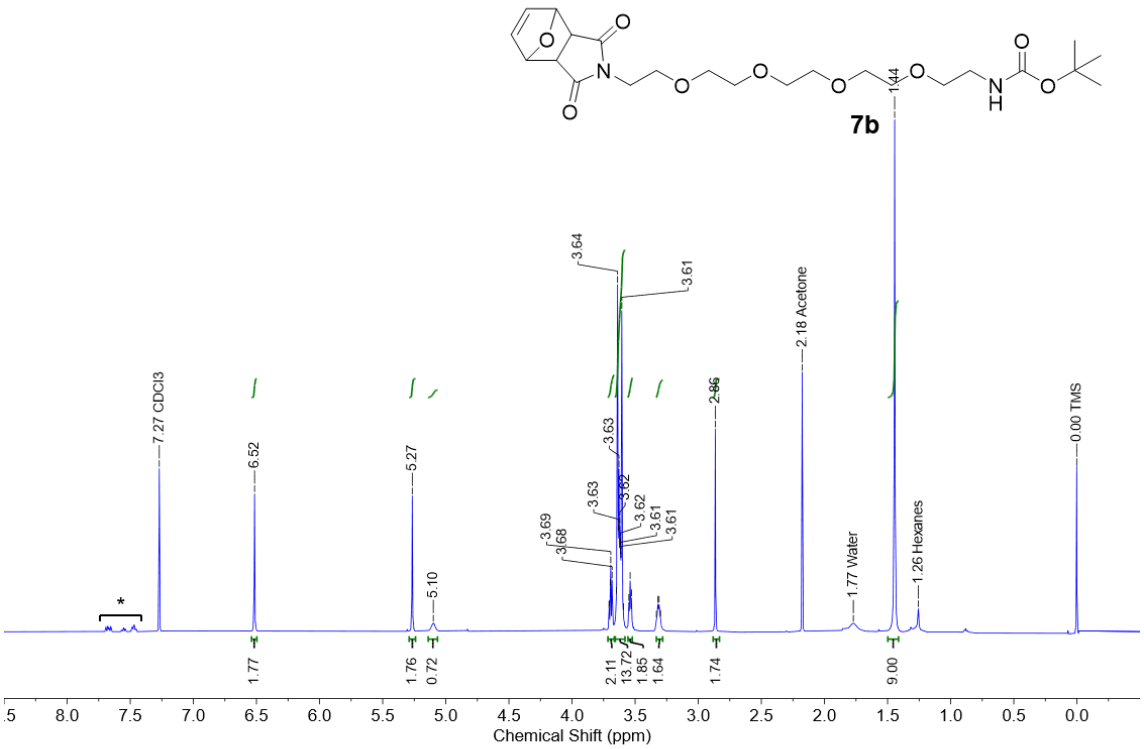


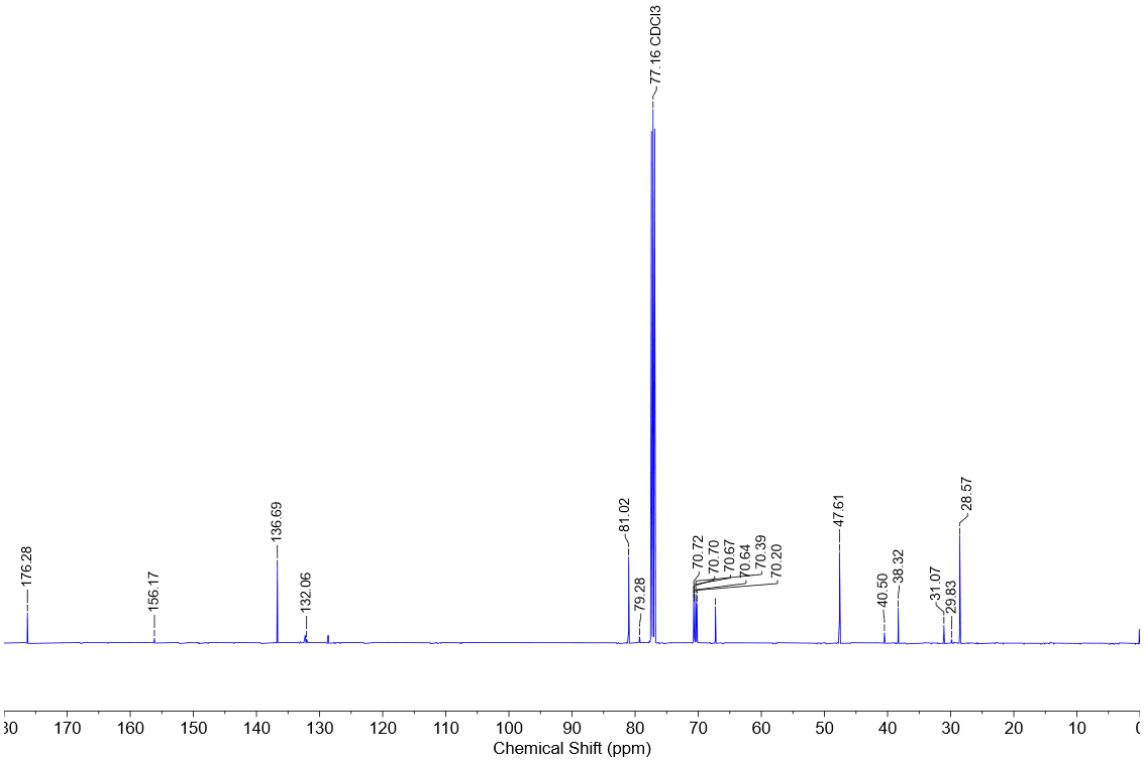
tert-butyl (2-(2-(2-(1,3-dioxo-1,3,3a,4,7,7a-hexahydro-2H-4,7-epoxyisoindol-2-yl)ethoxy)ethoxy)ethyl)carbamate (7a)



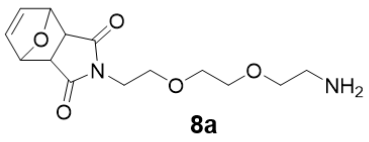


***tert*-butyl (14-(1,3-dioxo-1,3,3a,4,7,7a-hexahydro-2H-4,7-epoxyisoindol-2-yl)-3,6,9,12-tetraoxatetradecyl)carbamate (7b)**

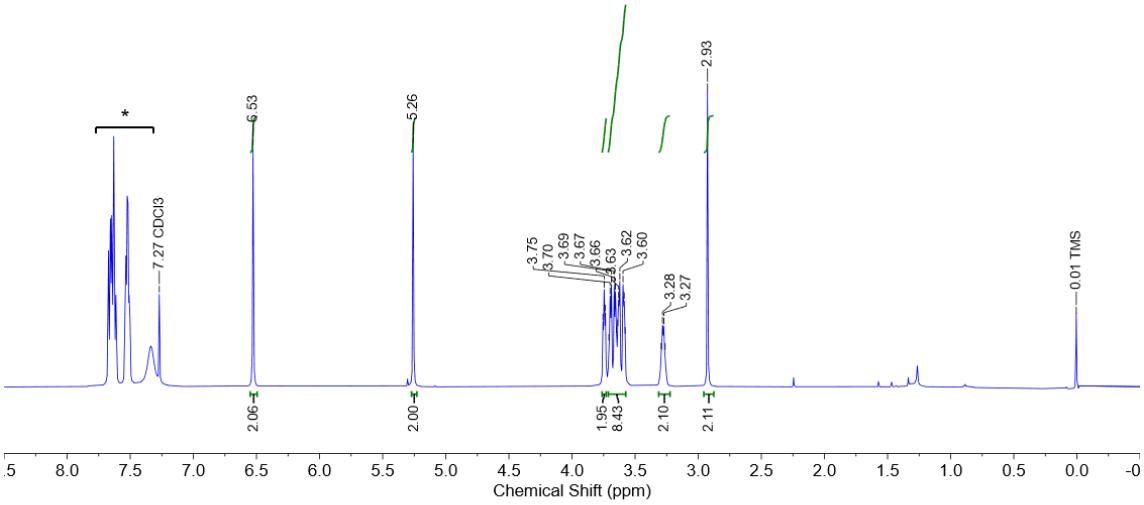


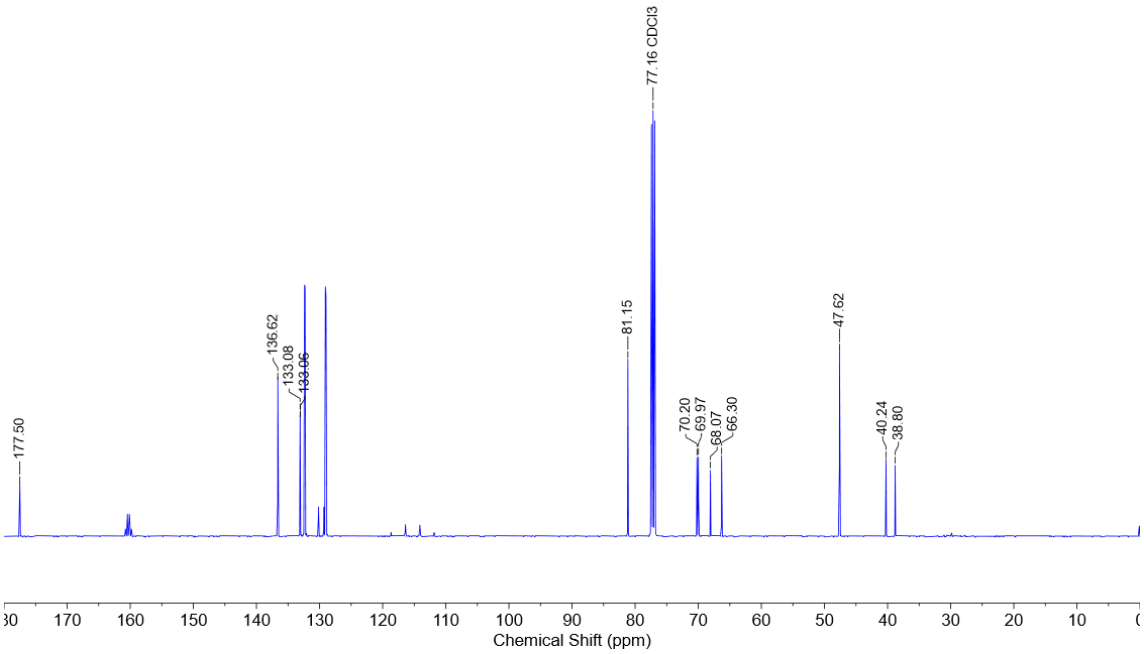


2-(2-(2-(2-aminoethoxy)ethoxy)ethyl)-3a,4,7,7a-tetrahydro-1H-4,7-epoxyisoindole-1,3(2H)-dione (8a)

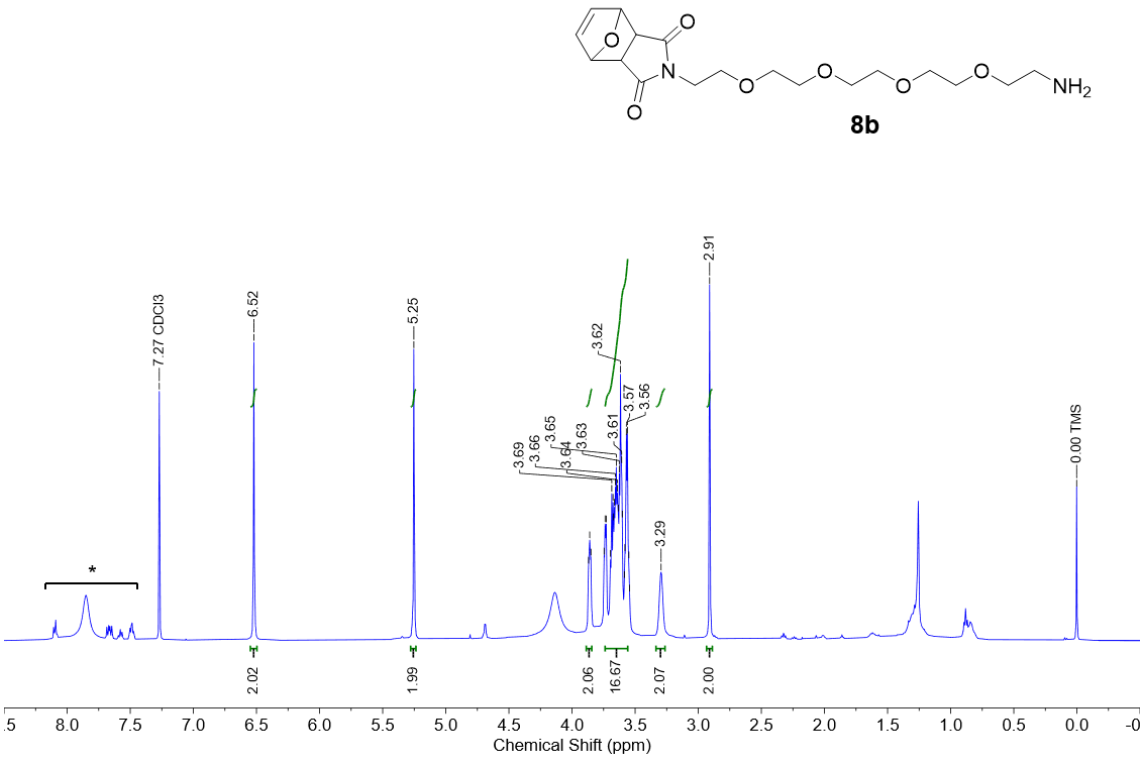


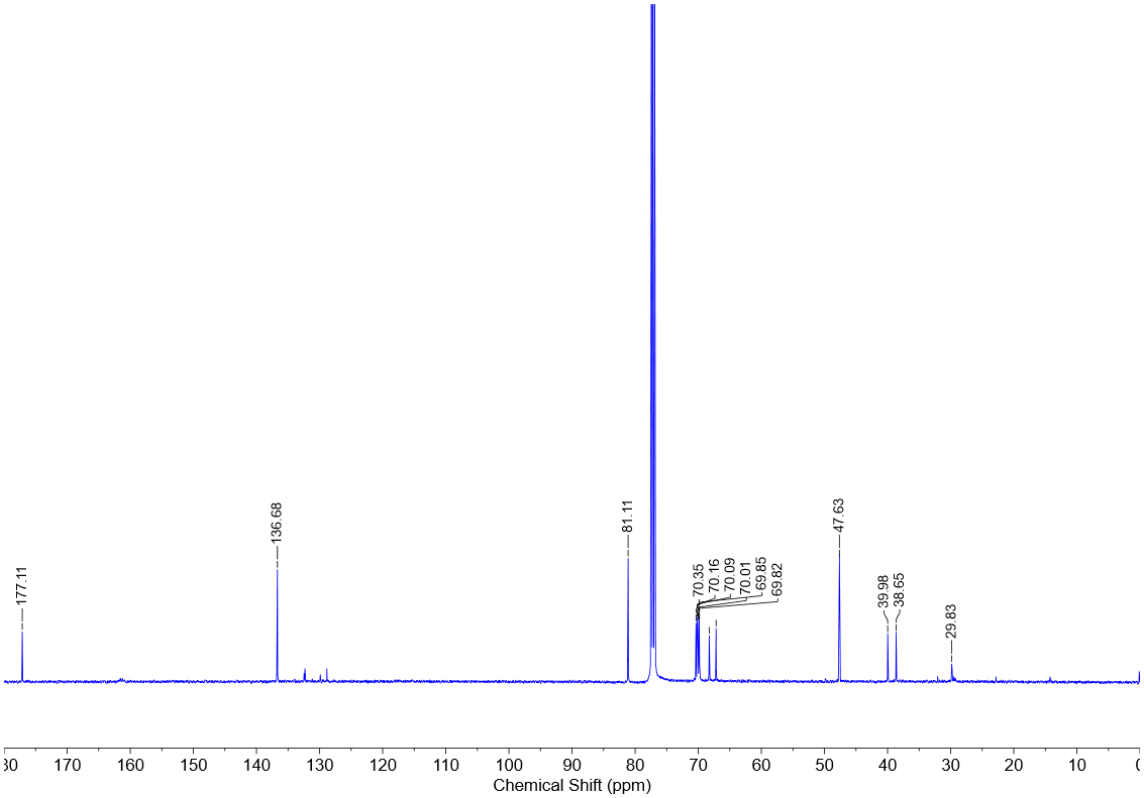
8a



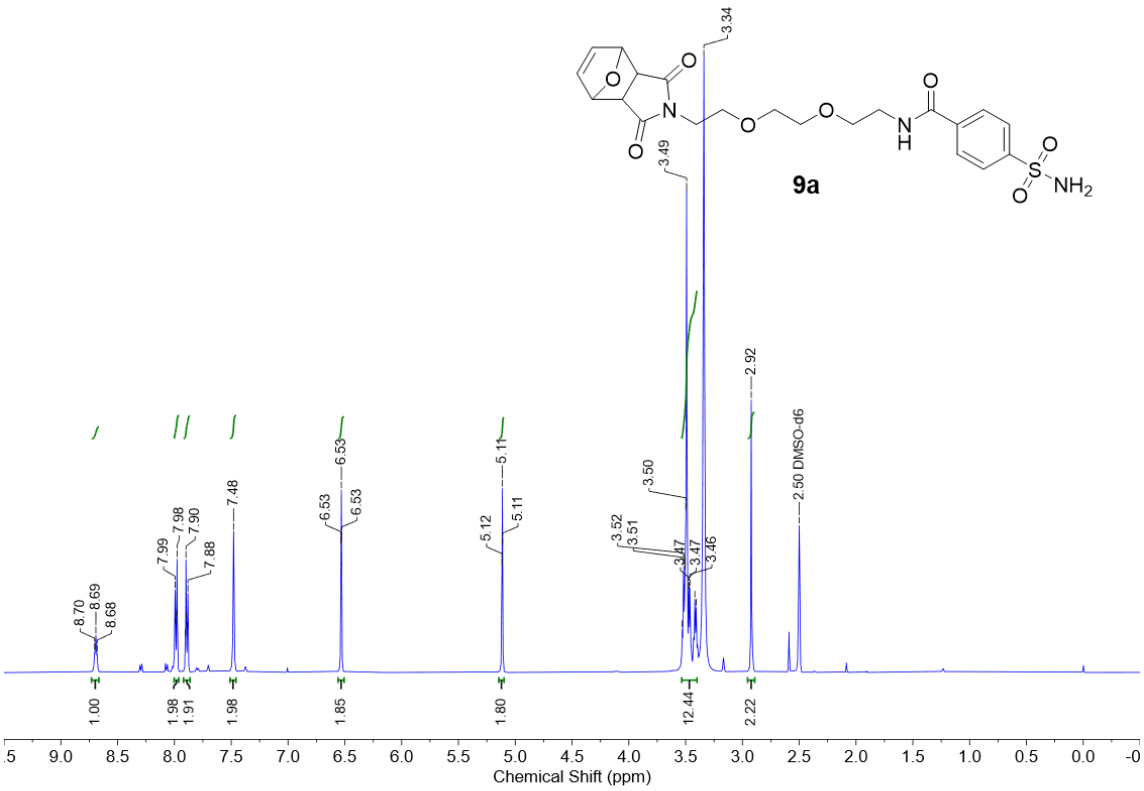


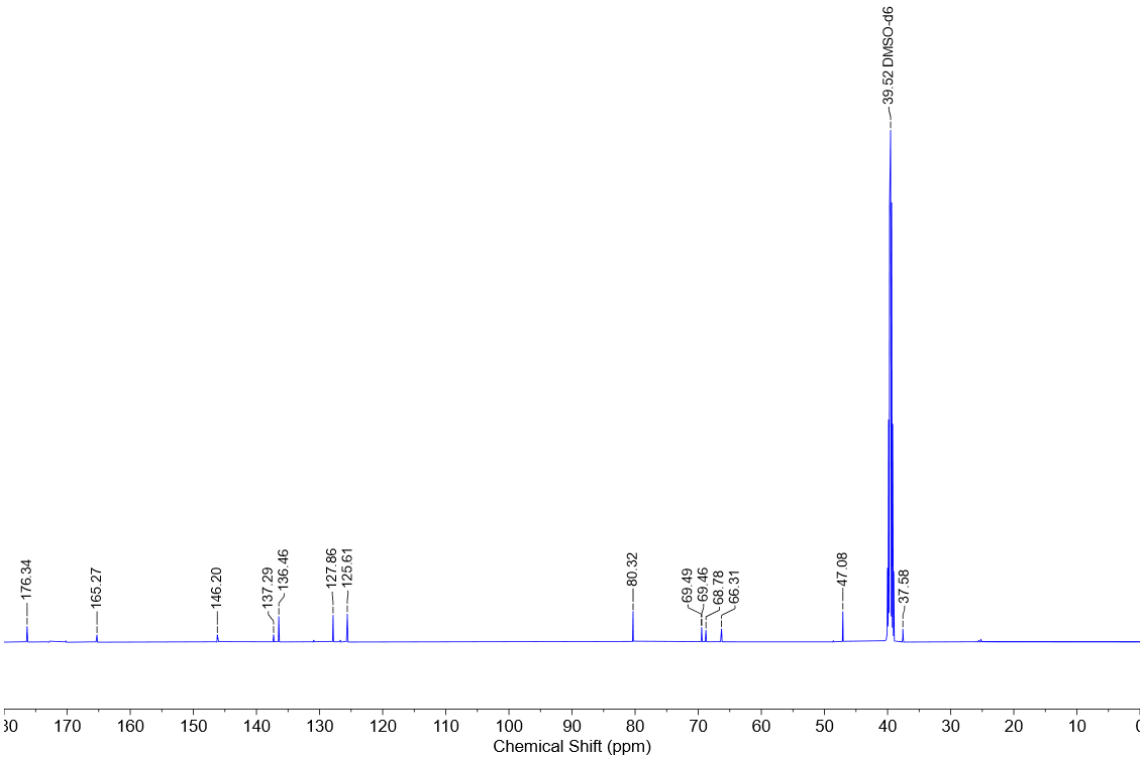
2-(14-amino-3,6,9,12-tetraoxatetradecyl)-3a,4,7,7a-tetrahydro-1H-4,7-epoxyisoindole-1,3(2H)-dione (8b)



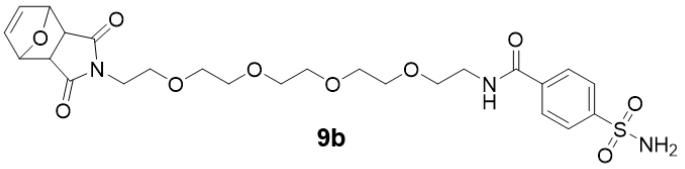


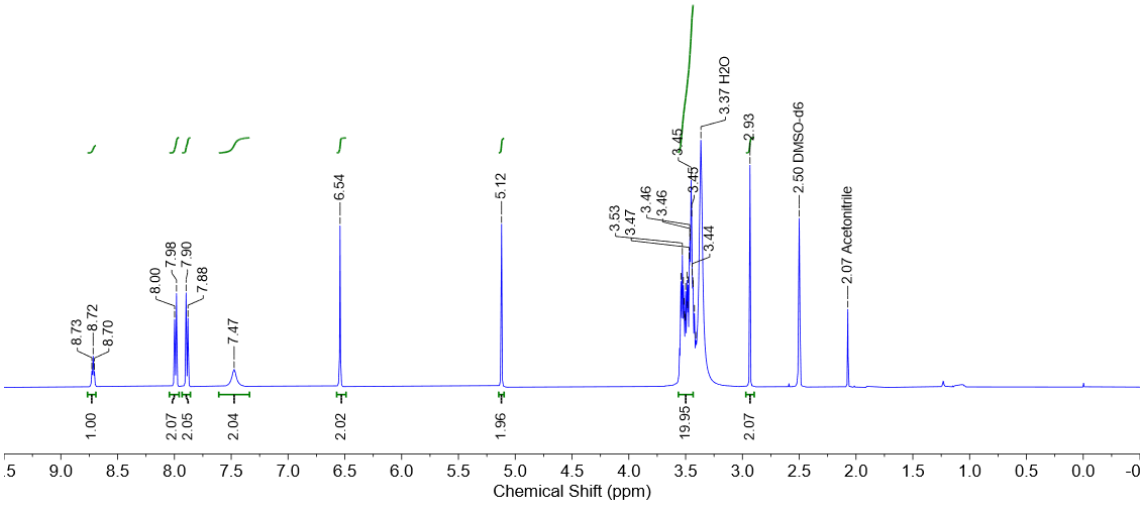
***N*-(2-(2-(2-(1,3-dioxo-1,3,3a,4,7,7a-hexahydro-2H-4,7-epoxyisoindol-2-yl)ethoxy)ethoxy) ethyl)-4-sulfamoylbenzamide (9a)**

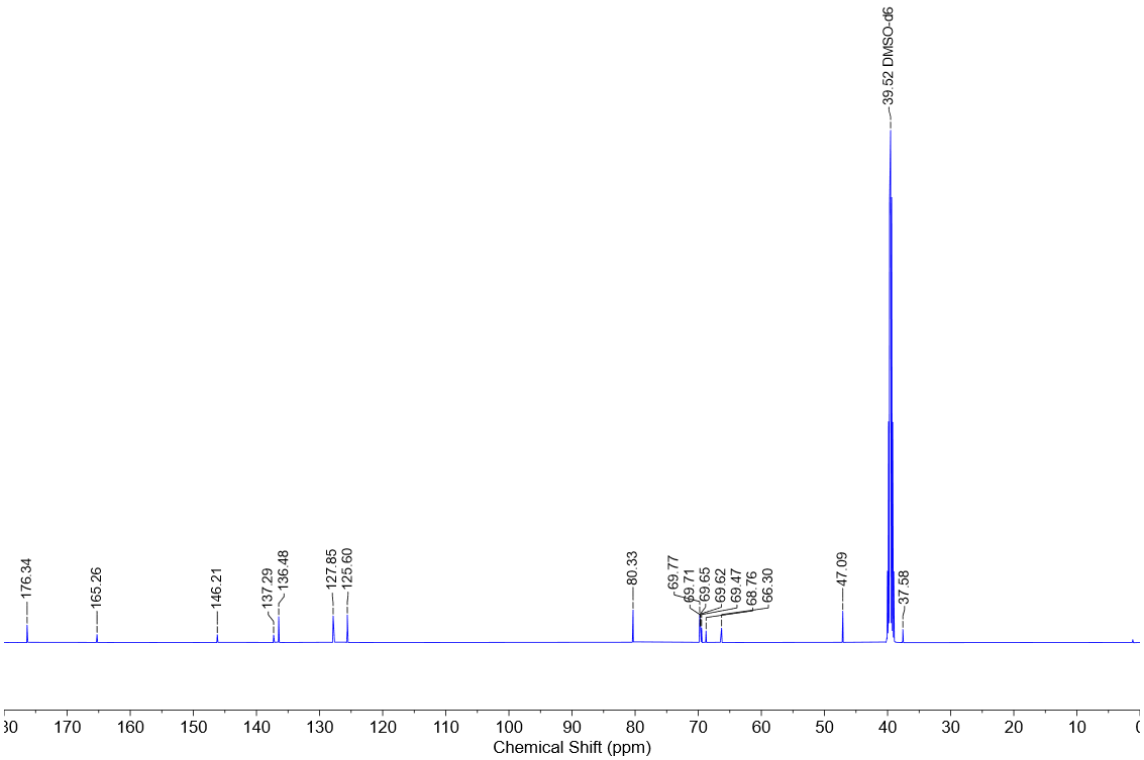




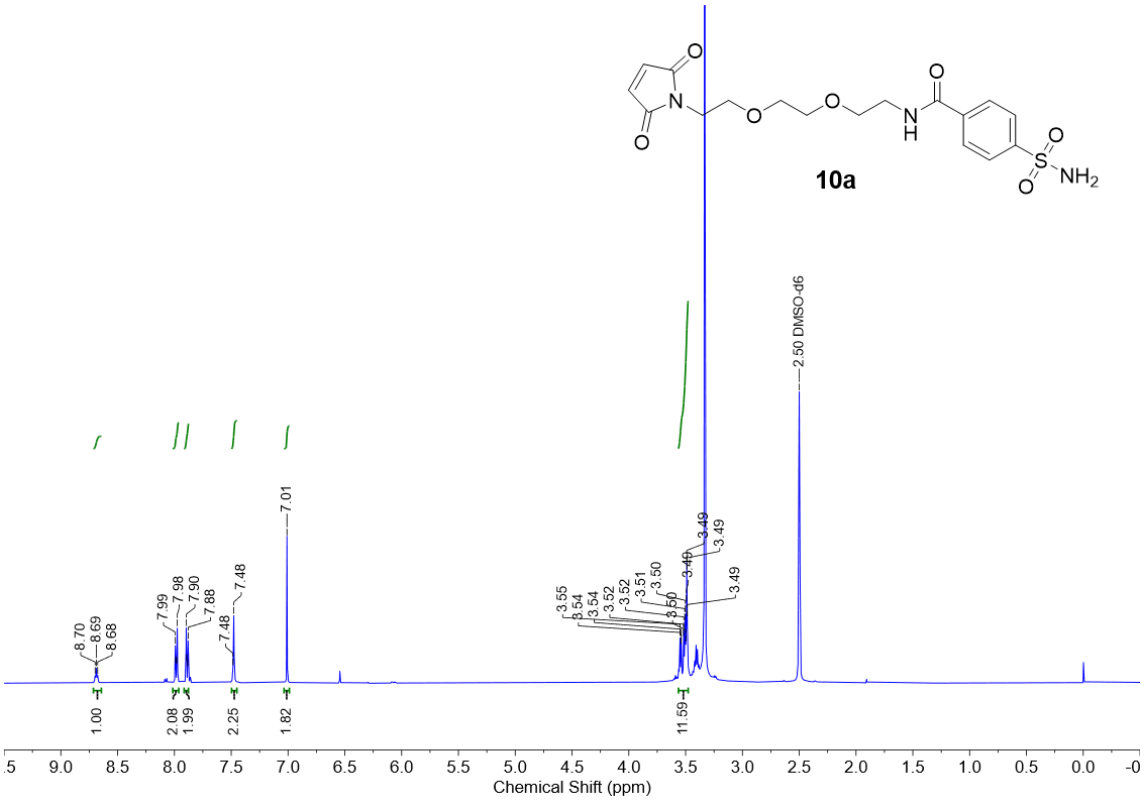
***N*-(14-(1,3-dioxo-1,3,3a,4,7,7a-hexahydro-2H-4,7-epoxyisoindol-2-yl)-3,6,9,12-tetraoxatetradecyl)-4-sulfamoylbenzamide (9b)**

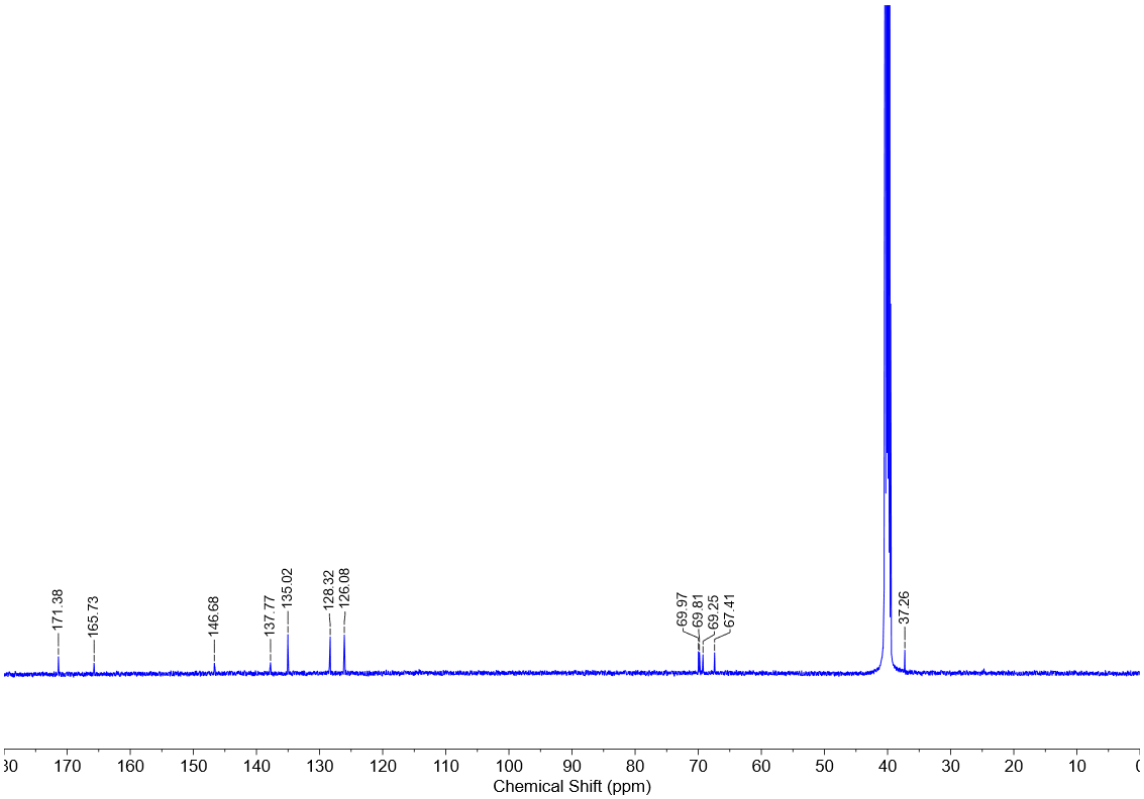




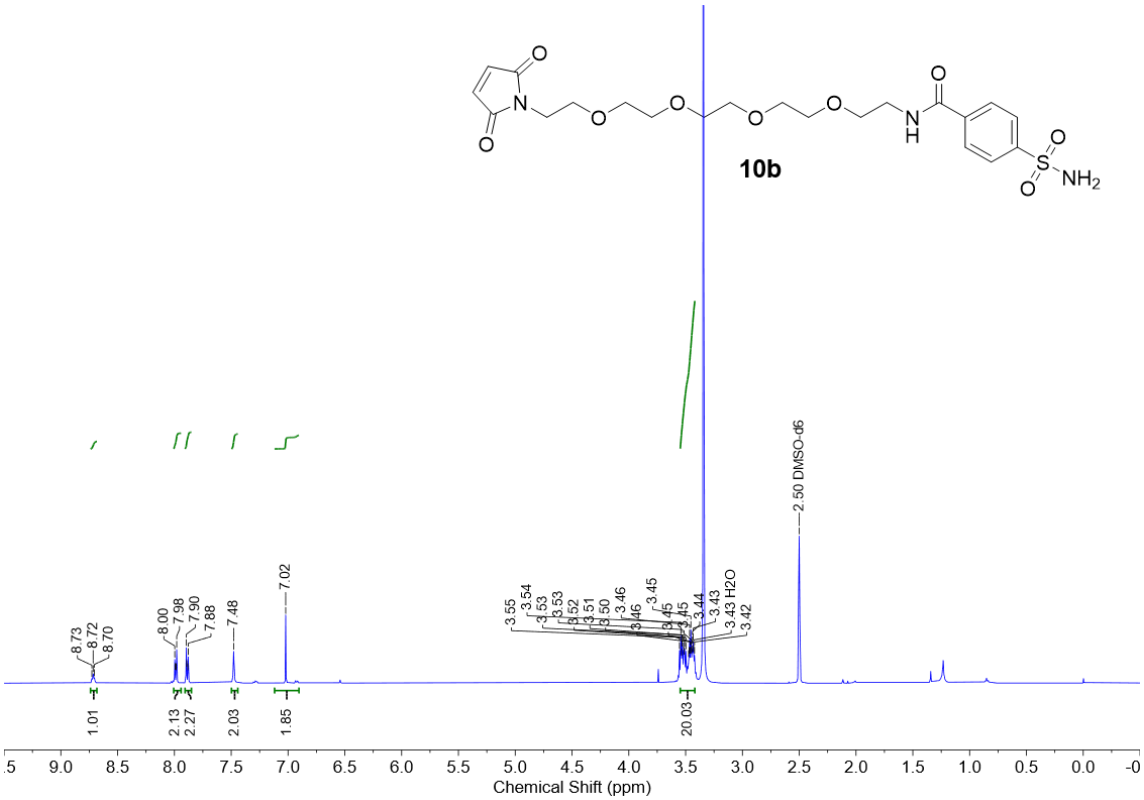


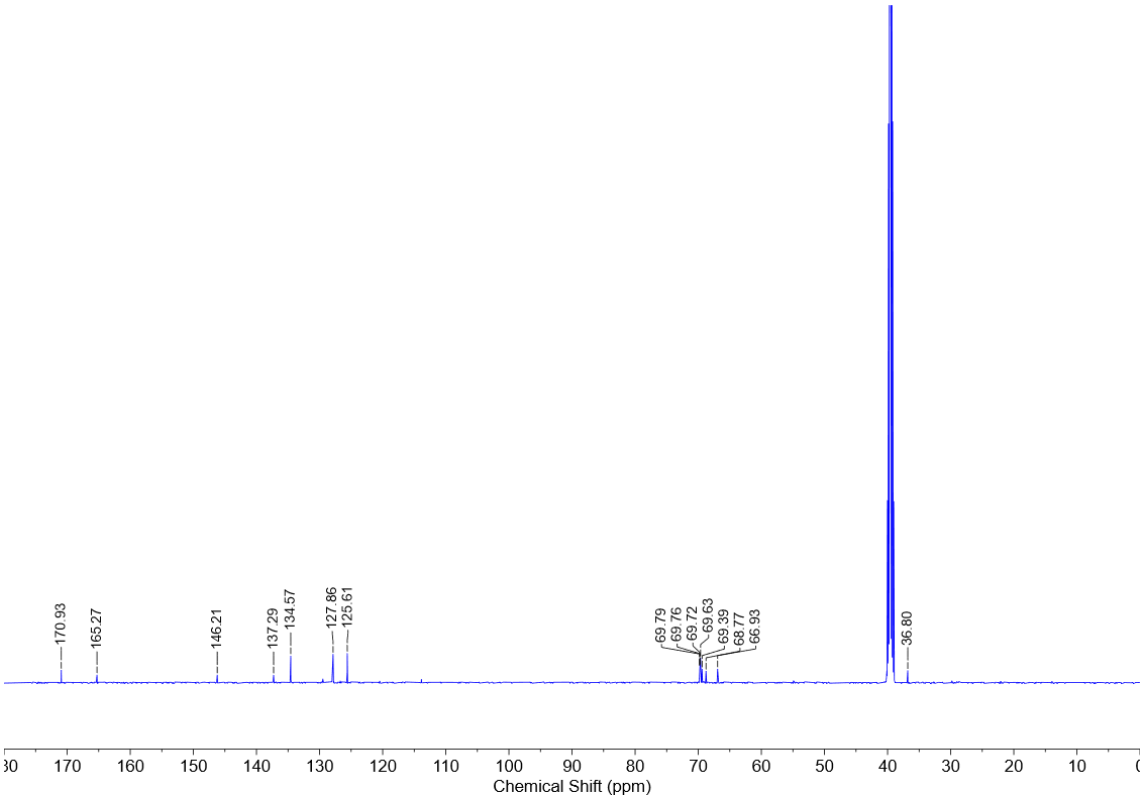
BzS3 (10a)





BzS5 (10b)


10b
8.73
8.72
8.70
8.00
7.98
7.90
7.88
7.48
7.02
3.55
3.54
3.53
3.52
3.51
3.50
3.46
3.45
3.44
3.43
3.43 H2O
3.42
2.50 DMSO-d6
1.01
2.13
2.27
2.03
1.85
20.03
Chemical Shift (ppm)


170.93
165.27
146.21
137.29
134.57
127.86
125.61
69.79
69.76
69.72
69.63
69.39
68.77
66.93
36.80
Chemical Shift (ppm)

References

1. P. Ji, J. Park, Y. Gu, D. S. Clark and J. F. Hartwig, Abiotic reduction of ketones with silanes catalysed by carbonic anhydrase through an enzymatic zinc hydride, *Nat. Chem.*, 2021, **13**, 312-318.
2. B. J. LaFrance, C. Cassidy-Amstutz, R. J. Nichols, L. M. Oltrogge, E. Nogales and D. F. Savage, The encapsulin from *Thermotoga maritima* is a flavoprotein with a symmetry matched ferritin-like cargo protein, *Sci. Rep.*, 2021, **11**, 22810.
3. M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. A. Hochberg, J. L. P. Benesch and C. V. Robinson, Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles, *Anal. Chem.*, 2015, **87**, 4370-4376.
4. J. F. Krebs, J. A. Ippolito, D. W. Christianson and C. A. Fierke, Structural and functional importance of a conserved hydrogen bond network in human carbonic anhydrase II, *J. Biol. Chem.*, 1993, **268**, 27458-27466.
5. S. Forli, R. Huey, M. E. Pique, M. F. Sanner, D. S. Goodsell and A. J. Olson, Computational protein–ligand docking and virtual drug screening with the AutoDock suite, *Nat. Protoc.*, 2016, **11**, 905-919.
6. C. Dominguez, R. Boelens and A. M. J. J. Bonvin, HADDOCK: A protein–protein docking approach based on biochemical or biophysical information, *J. Am. Chem. Soc.*, 2003, **125**, 1731-1737.
7. A. Ortega, D. Amorós and J. G. de la Torre, Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models, *Biophys. J.*, 2011, **101**, 892-898.
8. G. R. Fulmer, A. J. M. Miller, N. H. Sherden, H. E. Gottlieb, A. Nudelman, B. M. Stoltz, J. E. Bercaw and K. I. Goldberg, NMR chemical shifts of trace impurities: common laboratory solvents, organics, and gases in deuterated solvents relevant to the organometallic chemist, *Organometallics*, 2010, **29**, 2176-2179.
9. Z. Huang, J. Zhao, Z. Wang, F. Meng, K. Ding, X. Pan, N. Zhou, X. Li, Z. Zhang and X. Zhu, Combining orthogonal chain-end deprotections and thiol–maleimide Michael coupling: engineering discrete oligomers by an iterative growth strategy, *Angew. Chem., Int. Ed.*, 2017, **56**, 13612-13617.
10. K. Enander, G. T. Dolphin, B. Liedberg, I. Lundström and L. Baltzer, A versatile polypeptide platform for integrated recognition and reporting: affinity arrays for protein–ligand interaction analysis, *Chem. Eur. J.*, 2004, **10**, 2375-2385.
11. M. R. Molla, P. Prasad and S. Thayumanavan, Protein-induced supramolecular disassembly of amphiphilic polypeptide nanoassemblies, *J. Am. Chem. Soc.*, 2015, **137**, 7286-7289.
12. K. Heller, P. Ochtrop, M. F. Albers, F. B. Zauner, A. Itzen and C. Hedberg, Covalent protein labeling by enzymatic phosphocholination, *Angew. Chem., Int. Ed.*, 2015, **54**, 10327-10330.
13. A. H. Antropow, K. Xu, R. J. Buchsbaum and M. Movassaghi, Synthesis and evaluation of agelastatin derivatives as potent modulators for cancer invasion and metastasis, *J. Org. Chem.*, 2017, **82**, 7720-7731.
14. D. A. Whittington, A. Waheed, B. Ulmasov, G. N. Shah, J. H. Grubb, W. S. Sly and D. W. Christianson, Crystal structure of the dimeric extracellular domain of human carbonic anhydrase XII, a bitopic membrane protein overexpressed in certain cancer tumor cells, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 9545-9550.