Catalyst-free site-selective cross-aldol bioconjugations

Nicholas D. Yates,[†a] Saeed Akkad,[†a] Amanda Noble,[a] Tessa Keenan,[a] Natasha E. Hatton,[a] Nathalie Signoret[*b] and Martin A. Fascione[*a]

a. *Department of Chemistry, University of York, York, UK, YO10 5DD*
b. Hull York Medical School, University of York, York, UK, YO10 5DD
† These authors contributed equally to the work.
* Corresponding author. *Email:* martin.fascione@york.ac.uk; nathalie.signoret@york.ac.uk

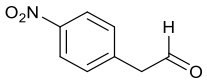
Supporting information

Contents

Synthesis and characterisation of small molecule probes

(4-nitrophenyl)acetaldehyde 10

To a suspension of (methoxymethyl)triphenylphosphonium chloride (16.5 g, 48 mmol) in anhydrous THF (80 mL) was treated with sodium hydride (1.92 g of 60 % dispersion in mineral oil, 48 mmol) and potassium tert-butoxide (0.54 g, 4.8 mmol) at 0 °C under an N_2 atmosphere. The resultant solution was stirred at 0 °C for 2 hours, after which time 4-nitrobenzaldehyde (2.41 g, 16 mmol) in anhydrous THF (5 mL) was added dropwise such that the temperature of the reaction did not exceed 0 °C. The resultant solution was then stirred at 0 °C for 2 hours and then stirred overnight at room temperature. The reaction solution was then concentrated *in vacuo*, and EtOAc was added (80 mL). The resultant suspension was washed with water (50 mL), brine (50 mL) and dried over $MgSO_4$. The organic extraction was then concentrated *in vacuo* and purified via flash column chromatography (hexane → 20 % EtOAc in hexane) to yield 1-(2-methoxyethenyl)-4-nitrobenzene (8.43 g, 47 mmol) as an E/Z mixture. The 1-(2-methoxyethenyl)-4-nitrobenzene was then dissolved in THF (80 mL). 1 M HCl (aq) was then added (160 mL) and the resultant solution refluxed until TLC showed the reaction to be complete. NaCl was then added and phase-separation occurred. The organic layer was taken, dried over $MgSO_4$ and concentrated in vacuo. Purification via flash silica column chromatography (hexane → 20% EtOAc in hexane) yielded (4-nitrophenyl)acetaldehyde as a yellow solid (3.15 g, 19 mmol, 40%).



^{1}H-NMR (400 MHz, $CDCl_3$): δ_H 9.82 (t, *J* = 1.70 Hz, 1H), 8.26-8.21 ($m^{AA'BB'}$, 2H), 7.42-7.37 ($m^{AA'BB'}$, 2H), 3.87 (d, *J* = 1.70 Hz, 2H).

^{13}C-NMR (100 MHz, $CDCl_3$): δ_C 197.33, 147.53*, 139.41, 130.73, 124.19, 50.17.

(ESI)HRMS: Found $[2M+Na]^+$ 353.0741, $C_{16}H_{14}N_2NaO_6$ requires 353.0744.

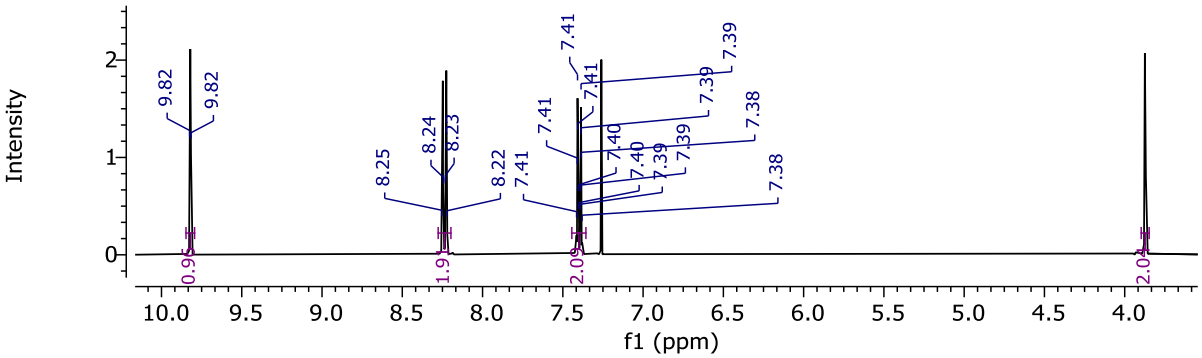


Figure S1. ^{1}H-NMR spectrum of **10**.

*The intensity of this signal is very low relative to the other signals. This phenomenon has previously been observed in the literature.[1]

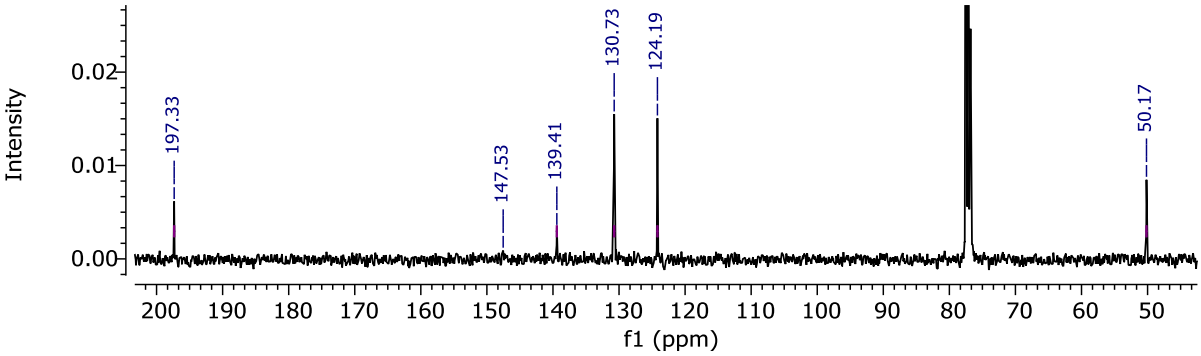


Figure S2. ^{13}C-NMR spectrum of **10**.

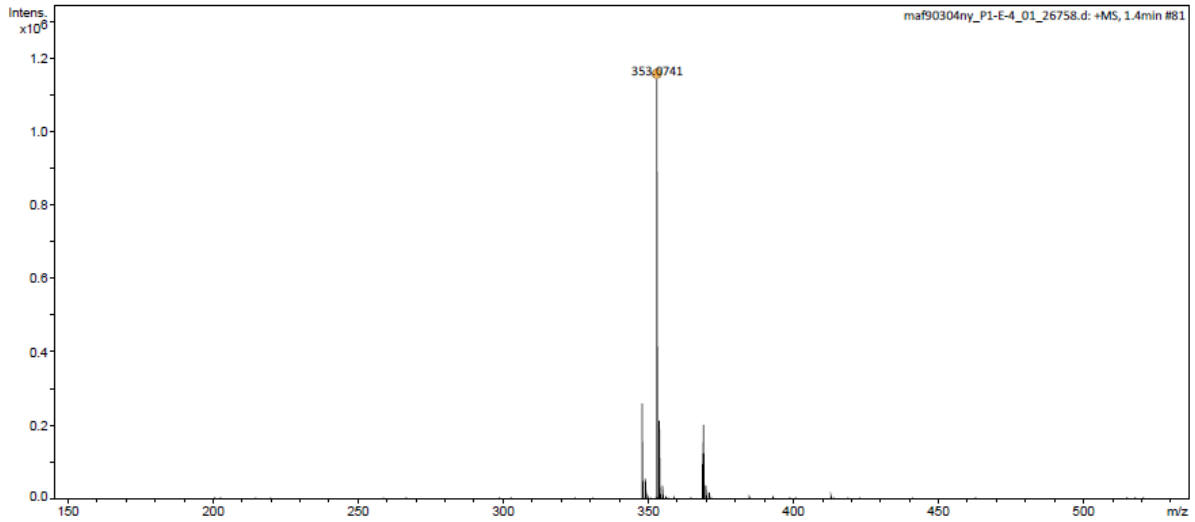
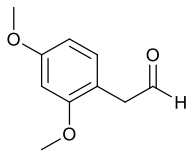


Figure S3. (ESI)HRMS of **10**.

(2,4-dimethoxyphenyl)acetaldehyde 13

This compound was prepared using a method based on a literature procedure.[2] To a suspension of (methoxymethyl)triphenylphosphonium chloride (823 mg, 2.4 mmol) in anhydrous THF (1 mL) was treated with lithium bis(trimethylsilyl)amide (0.435 g, 2.6 mmol) at 0 °C under an N_2 atmosphere. The resultant solution was stirred at 0 °C for 15 min, after which 2,4-dimethoxybenzaldehyde (306 mg, 2.0 mmol) in anhydrous THF (2 mL) was added slowly at 0 °C. The resultant solution was then warmed to 25 °C and stirred for 3 hours, after which time the reaction was quenched via the addition of saturated aqueous NH_4Cl. The product was then extracted with EtOAc (20 mL × 2), and the organic extractions dried by washing with brine (20 mL) and drying over $MgSO_4$. The organic extraction was then concentrated *in vacuo* and then dissolved in 3 mL THF. Add 3 mL of 1 M aqueous HCl was then added, and the resultant solution heated to 80 °C for 4 hours. EtOAc (20 mL) was then added and the solution washed with saturated aqueous $NaHCO_3$. The $NaHCO_3$ (aq) extraction was back-extracted with EtOAc (20 mL × 2). All organic extractions were then combined, washed with brine (20 mL) and dried over $MgSO_4$. Purification via flash silica column chromatography (10% EtOAc in hexane → 30% EtOAc in hexane) yielded (2,4-dimethoxyphenyl)acetaldehyde as a pale green oil that solidified into a white

solid in the freezer (39 mg, 0.22 mmol, 11%). Characterisation data is consistent with that reported in the literature.[2]



^{1}H-NMR (400 MHz, $CDCl_3$): δ_H 9.65 (t, *J* = 2.20 Hz, 1H), 7.04 (d, *J* = 7.94 Hz, 1H), 6.50-6.45 (m, 2H), 3.81 (s, 3H), 3.80 (s, 3H), 3.57 (d, *J* = 2.20 Hz, 2H).

^{13}C-NMR (100 MHz, $CDCl_3$): δ_C 200.87, 160.71, 158.71, 131.71, 113.60, 104.52, 99.77, 55.56, 55.54, 44.92.

(ESI)HRMS: Found $[M+H]^+$ 181.0858, $C_{10}H_{13}O_3$ requires 181.0859.

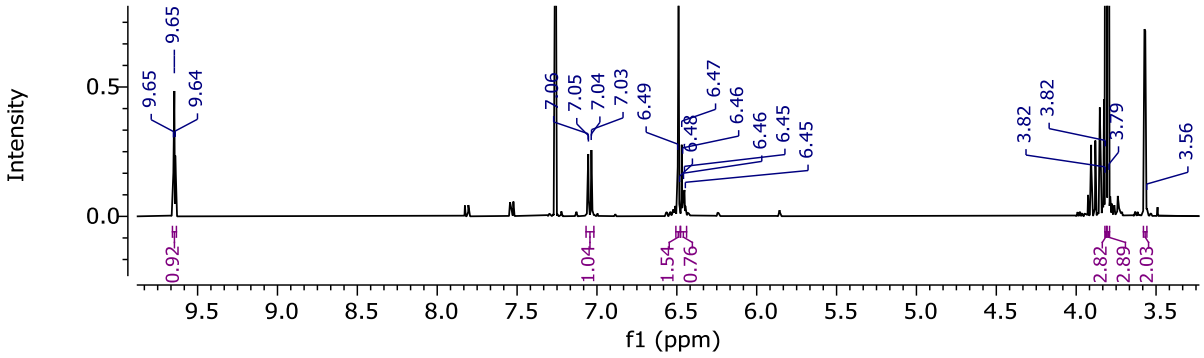


Figure S4. ^{1}H-NMR spectrum of **13**.

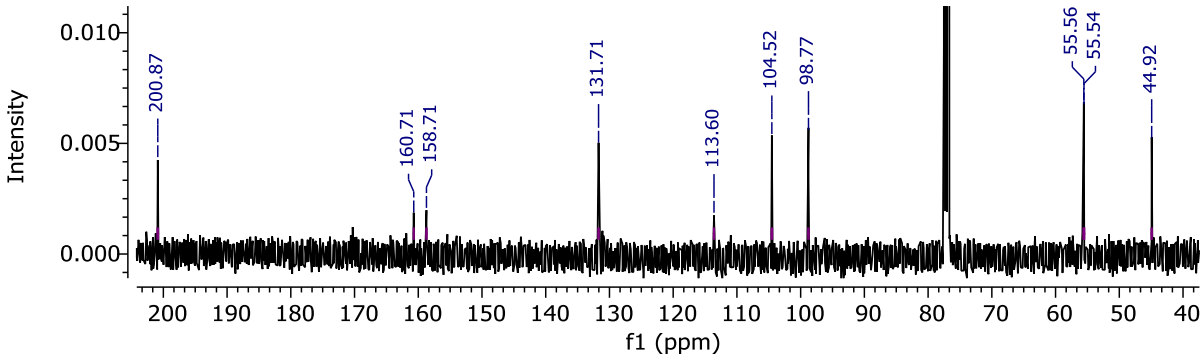
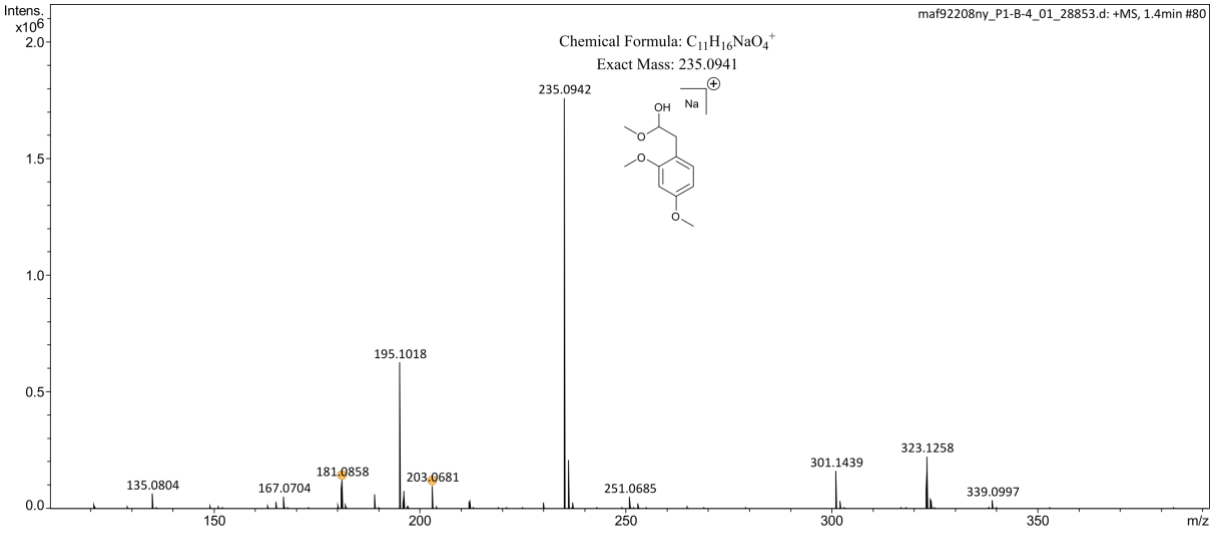


Figure S5. ^{13}C-NMR spectrum of **13**.

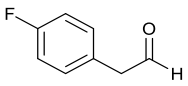


Meas. m/z	#	Ion Formula	m/z	err [ppm]	err [mDa]	mSigma	Mean err [ppm]
181.0858	1	C10H13O3	181.0859	0.8	0.1	6.3	-3.1
203.0681	1	C10H12NaO3	203.0679	-1.3	-0.3	6.9	-1.2

Figure S6. (ESI)HRMS of **13**.

(4-fluorophenyl)acetaldehyde 14

To 4-fluorophenylacetaldehyde dimethyl acetal (0.1 g, 0.543 mmol) was added 5 mL of a 1:1 mixture of TFA and water. The resultant solution was stirred at room temperature until TLC indicated reaction completion. DCM (20 mL) was then added and the organic extraction was dried over brine and $MgSO_4$. The resultant solution was then dry-loaded onto silica and purified via flash column chromatography (hexane → ethyl acetate). Concentration in vacuo yielded a colourless oil (49.5 mg, 0.358 mmol, 66%) that solidified into a white solid in the freezer.



^{1}H-NMR (400 MHz, $CDCl_3$): δ_H 9.75 (t, *J* = 2.12 Hz, 1H), 7.21-7.15 (m, 2H), 7.09-7.03 (m, 2H), 3.68 (d, *J* = 2.12 Hz, 2H).

^{19}F-NMR (376 MHz, $CDCl_3$): δ_F -115.01 (m).

^{13}C-NMR (100 MHz, $CDCl_3$): δ_C 199.18 (s), 162.35 (d, *J* = 246.15 Hz), 131.31 (d, *J* = 8.19 Hz), 127.65 (d, *J* = 3.00 Hz), 116.04 (d, *J* = 21.53 Hz), 49.79 (s).

(ESI)HRMS: Found $[M-H]^-$ 137.0409, C_8H_6FO requires 137.0408.

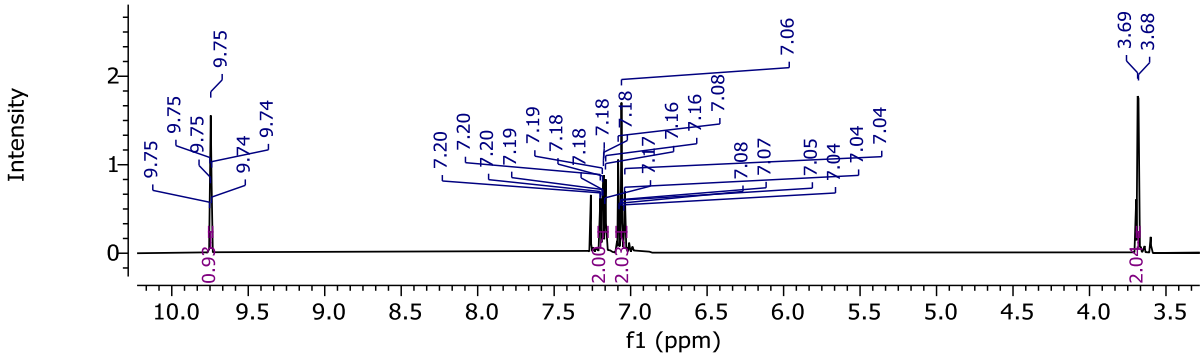


Figure S7. [1]H-NMR spectrum of **14**.

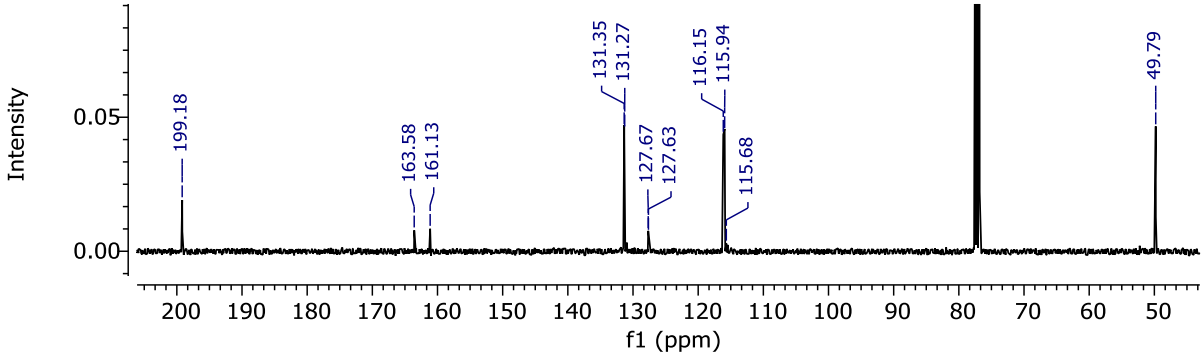


Figure S8. [13]C-NMR spectrum of **14**.

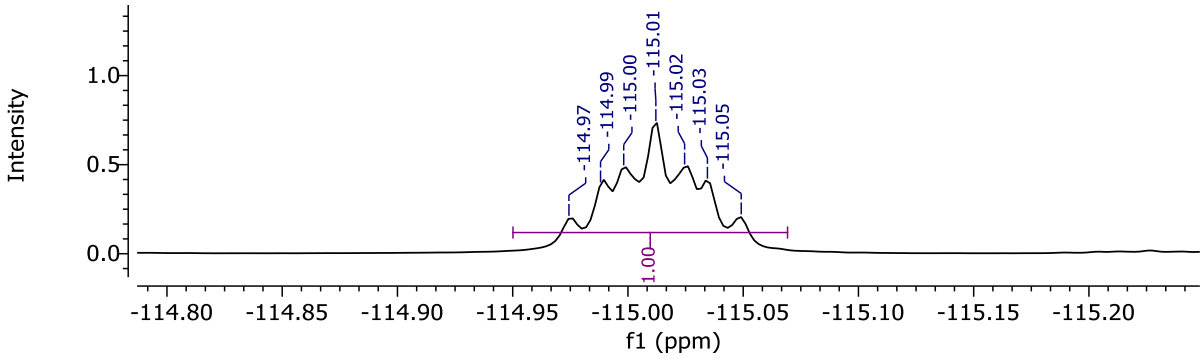


Figure S9. [19]F-NMR spectrum of **14**.

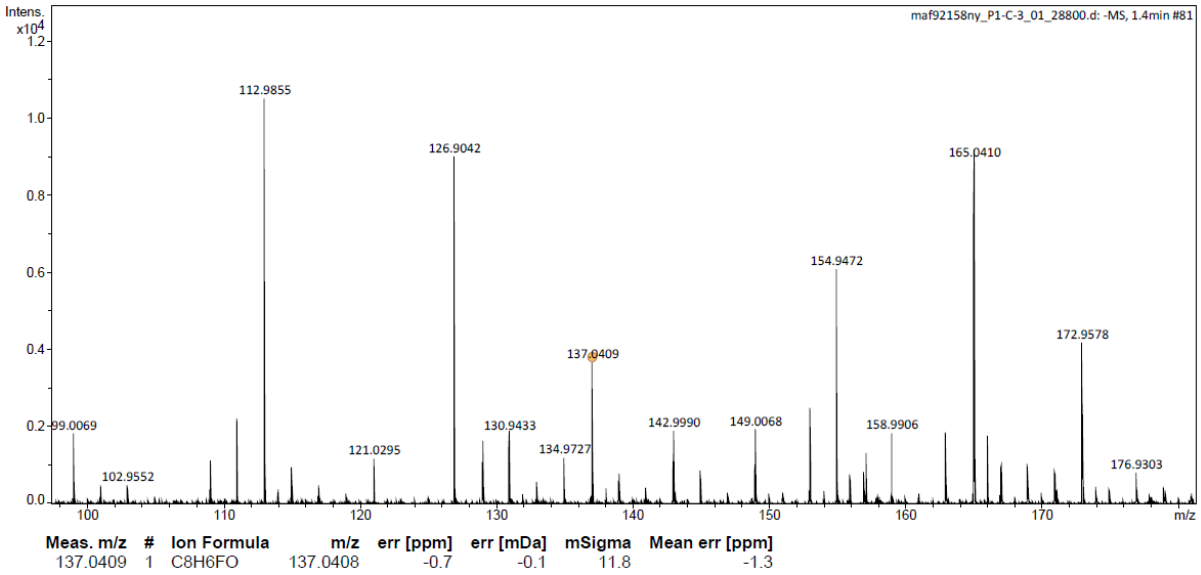
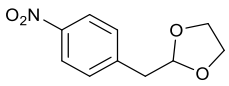


Figure S10. (ESI)HRMS of **14**.

2-(4-nitrobenzyl)-1,3-dioxolane S1

10 (0.194 g, 1.17 mmol) was dissolved in anhydrous toluene (5 mL) and 1 mL of ethylene glycol and p-toluenesulfonic acid monohydrate (28 mg, 0.16 mmol) added. The resultant solution was refluxed for 3 hours, after which time TLC showed the reaction to be complete. The solution was cooled to room temperature and transferred to a separating funnel. DCM (30 mL) and saturated $NaHCO_3$(aq) were added (30 mL) and shaken. The organic extraction was taken, washed with water (30 mL) and brine (30 mL) and dried over $MgSO_4$. Concentration in vacuo yielded **S1** as a dull pale-yellow solid (0.20 g, 0.96 mmol, 82%).



^{1}H-NMR (400 MHz, $CDCl_3$): δ_H 8.18-8.12 ($m^{AA'BB'}$, 2H), 7.47-7.40 ($m^{AA'BB'}$, 2H), 5.10 (t, *J* = 4.43 Hz, 1H), 3.93-3.81 (m, 4H), 3.07 (d, *J* = 4.43 Hz, 2H).

^{13}C-NMR (100 MHz, $CDCl_3$): δ_C 147.00, 143.90, 130.89, 123.52, 103.63, 65.23, 40.55.

(APCI)HRMS: Found $[M+H]^+$ 210.0752, $C_{10}H_{12}NO_4$ requires 210.0760.

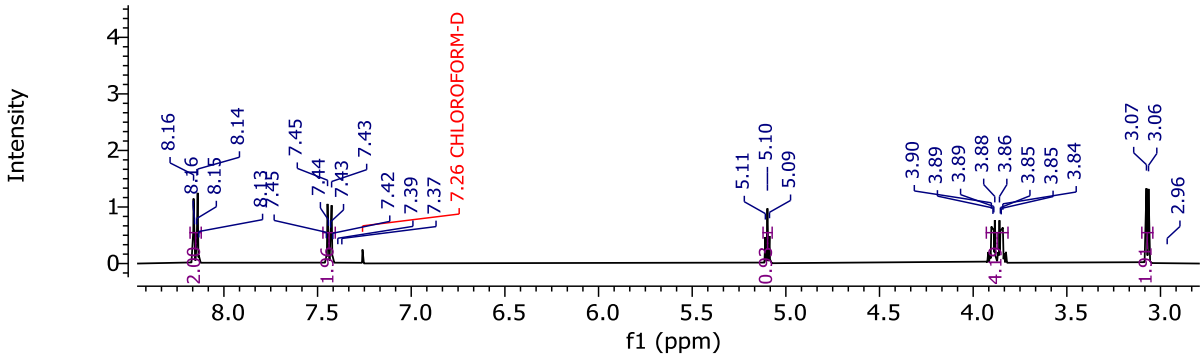


Figure S11. ^{1}H-NMR spectrum of **S1**.

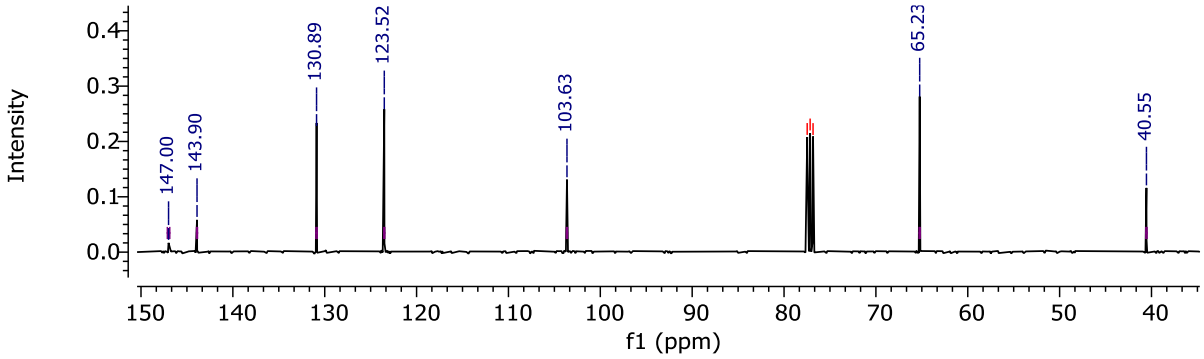


Figure S12. ^{13}C-NMR spectrum of **S1**.

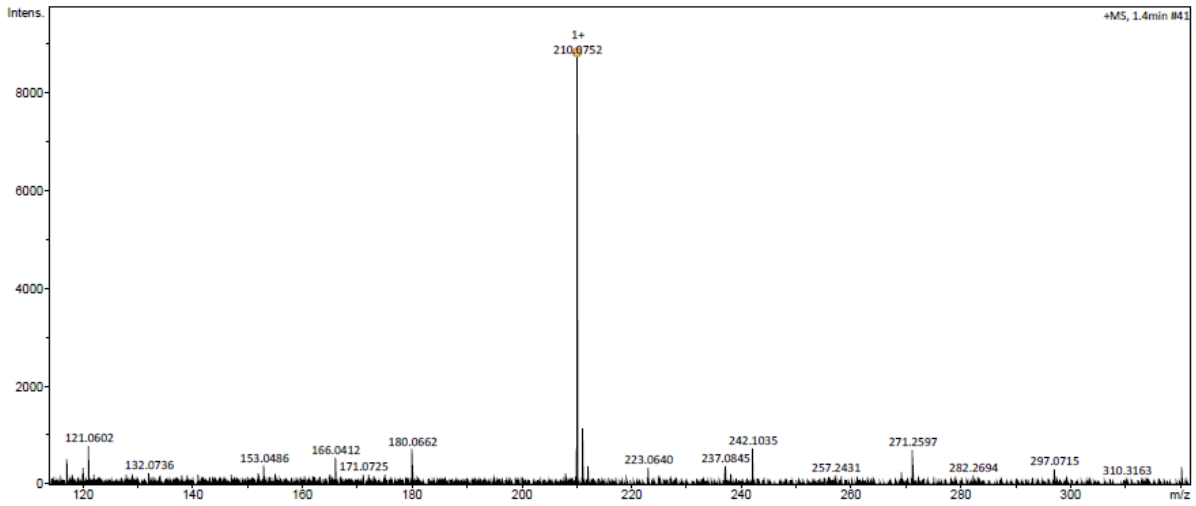
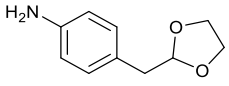


Figure S13. (ESI)HRMS of **S1**.

2-(4-aminobenzyl)-1,3-dioxolane S2

To **S1** (0.199 g, 0.956 mmol) dissolved in EtOAc (20 mL) was added a spatula-end of 10 % palladium on activated carbon. The system was then placed under N_2 before being placed under 1 atm of H_2. The resultant mixture was stirred overnight whereupon TLC showed the reaction to be complete. Pd/C was removed via filtration and the filtrate concentrated *in vacuo* to yield **S2** (0.13 g, 0.72 mmol, 76%) as a dull orange oil that solidified upon storage at -20 °C.



^{1}H-NMR (400 MHz, $CDCl_3$): δ_H 7.10-7.05 ($m^{AA'BB'}$, 2H), 6.71-6.65 ($m^{AA'BB'}$, 2H), 5.00 (t, *J* = 4.61 Hz, 1H), 3.96-3.89 (m, 2H), 3.88-3.80 (m, 2H), 2.86 (d, *J* = 4.61 Hz, 2H).

^{13}C-NMR (100 MHz, $CDCl_3$): δ_C 144.19, 130.65, 126.82, 115.80, 105.09. 65.09, 40.03.

(ESI)HRMS: Found $[M+H]^+$ 180.1021, $C_{10}H_{14}NO_2$ requires 180.1019.

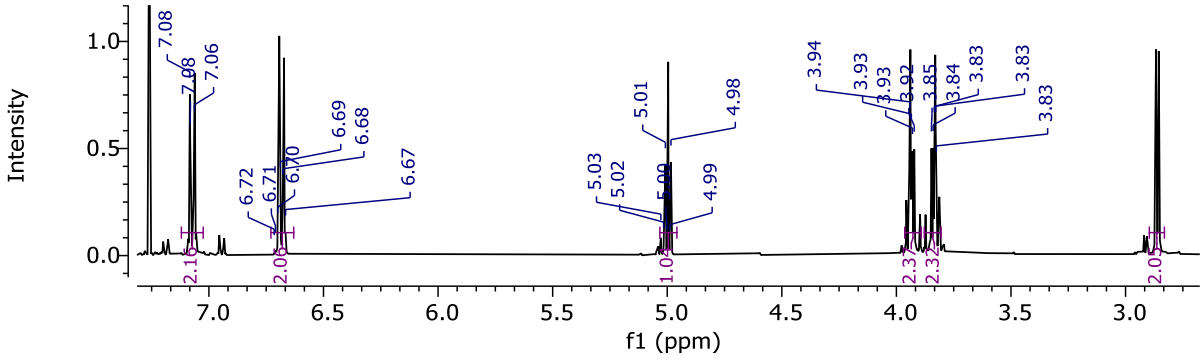


Figure S14. ^{1}H-NMR spectrum of **S2**.

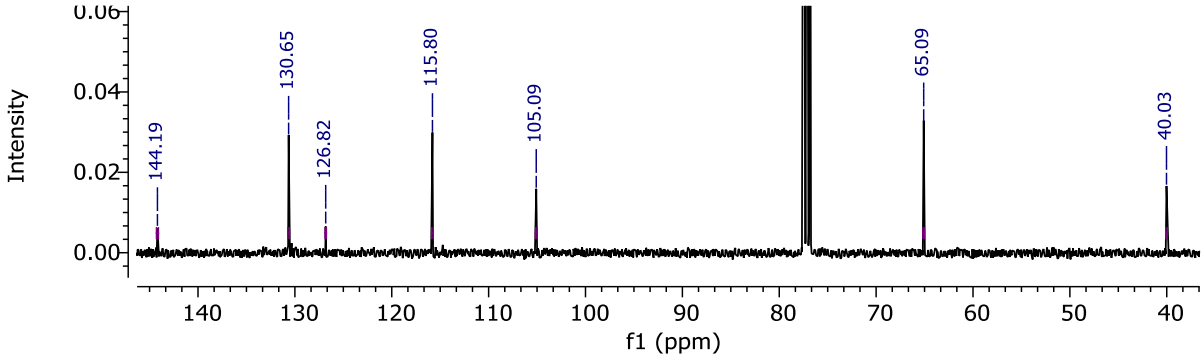


Figure S15. ^{13}C-NMR spectrum of **S2**.

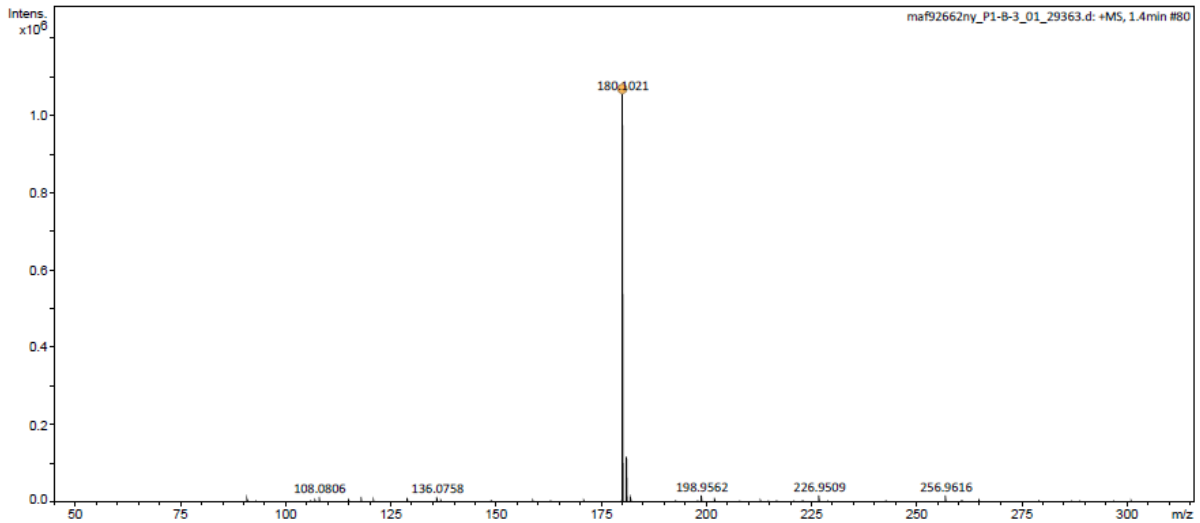
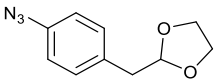


Figure S16. (ESI)HRMS of **S2**.

2-(4-azidobenzyl)-1,3-dioxolane S3

Nitrosonium tetrafluoroborate (0.24 g, 2.0 mmol) was dissolved in anhydrous acetonitrile (0.5 mL) under nitrogen at -30 °C. To the resultant solution was added **S2** (0.36 g, 2.0 mmol) dissolved in anhydrous acetonitrile (0.5 mL). The resultant solution was stirred for 1 hour at -30 °C before being allowed to warm to 0 °C. Complete diazotization was at this point assumed to have occurred. A solution of sodium azide (0.156 g, 2.4 mmol) dissolved in water (5 mL) was then prepared. Diethyl ether was added to the sodium azide solution, and the resultant mixture stirred at 0 °C. The aqueous sodium azide + diethyl ether mixture was then added dropwise to the diazonium solution at 0 °C. The mixture was then allowed to warm to room temperature and was stirred for 1 hour. Diethyl ether (30 mL) was then added and the organic layer taken, washed with brine (20 mL) and dried over $MgSO_4$. The crude 2-(4-azidobenzyl)-1,3-dioxolane sample was then purified via flash column chromatography (Hexane → 30% EtOAc in hexane) to yield **S3** as a yellow oil (145 mg, 0.71 mmol, 35%).



^{1}H-NMR (400 MHz, $CDCl_3$): δ_H 7.31-7.25 ($m^{AA'BB'}$, 2H), 7.02-6.96 ($m^{AA'BB'}$, 2H), 5.06 (t, *J* = 4.69 Hz, 1H), 3.99-3.81 (m, 4H), 2.97 (d, *J* = 4.69 Hz, 2H).

^{13}C-NMR (100 MHz, $CDCl_3$): δ_C 138.41, 132.93, 131.16, 118.95, 104.41, 65.04, 40.08.

FT-IR (ATR) (υmax/cm^{-1}): 2954 (C-H stretch, aromatic), 2884 (C-H stretch, alkyl), 2111 (N=N=N stretch, azide), 1507 (C-C stretch, aromatic), 1286, 1128.

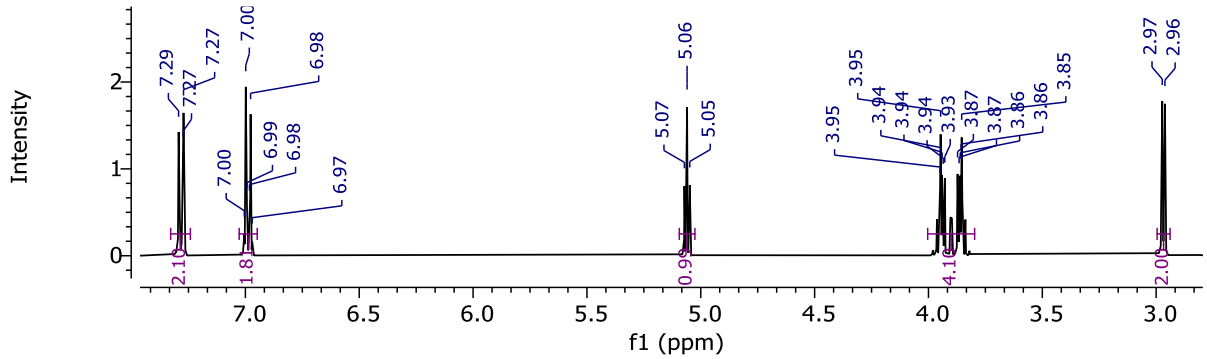


Figure S17. ^{1}H-NMR spectrum of **S3**.

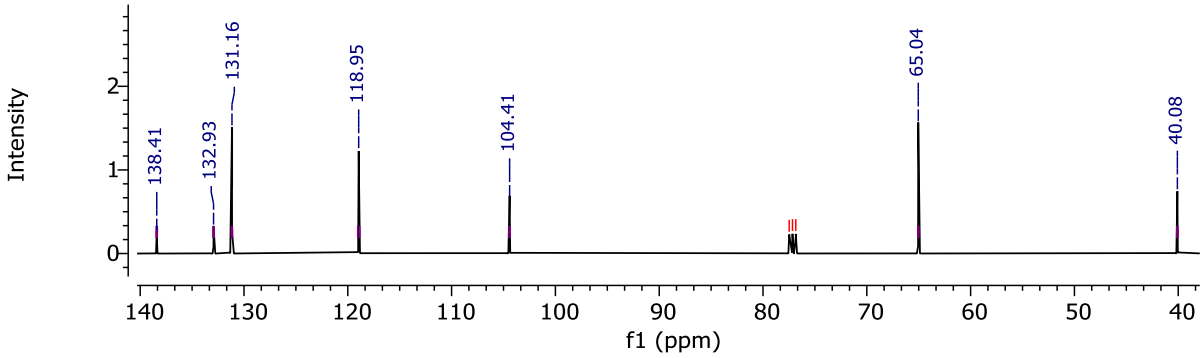


Figure S18. ^{13}C-NMR spectrum of **S3**.

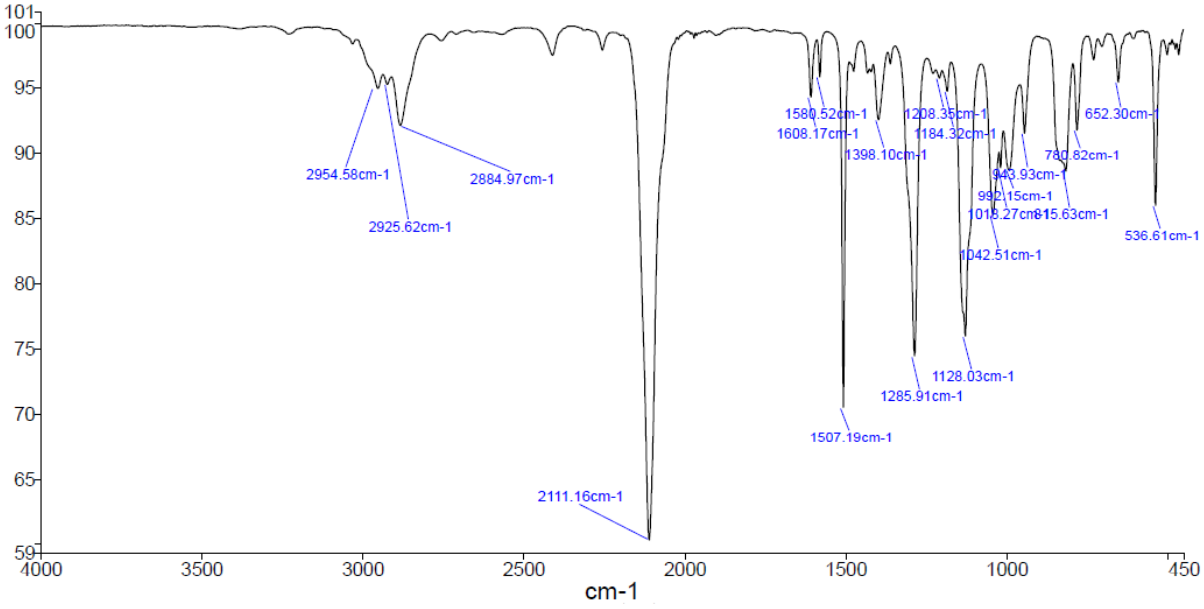
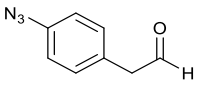


Figure S19. FT-IR (ATR) spectrum of **S3**.

(4-azidophenyl)acetaldehyde 15

S3 (0.130 g, 0.633 mmol) was dissolved in 10 mL of a 1:1 mixture of TFA and water. The resultant solution was stirred at room temperature until TLC showed that reaction to be complete. After this time DCM was added (30 mL) and the mixture was transferred to a separating funnel. The organic layer was washed with brine and dried over $MgSO_4$. Purification via flash column chromatography (Hexane → 10% EtOAc in hexane) yielded **15** as a yellow oil (67 mg, 0.42 mmol, 66%).



^{1}H-NMR (400 MHz, $CDCl_3$): δ_H 9.74 (t, *J* = 2.16 Hz, 1H), 7.23-7.18 ($m^{AA'BB'}$, 2H), 7.06-7.01 ($m^{AA'BB'}$, 2H), 3.69 (d, *J* = 2.16 Hz, 2H).

^{13}C-NMR (100 MHz, $CDCl_3$): δ_C 199.02, 139.54, 131.17, 128.56, 119.72, 49.98.

FT-IR (ATR) (υmax/cm^{-1}): 2826 (C-H stretch, alkyl), 2098 (N=N=N stretch, azide), 1721 (C=O stretch, aldehyde), 1505 (C-C stretch, aromatic), 1284.

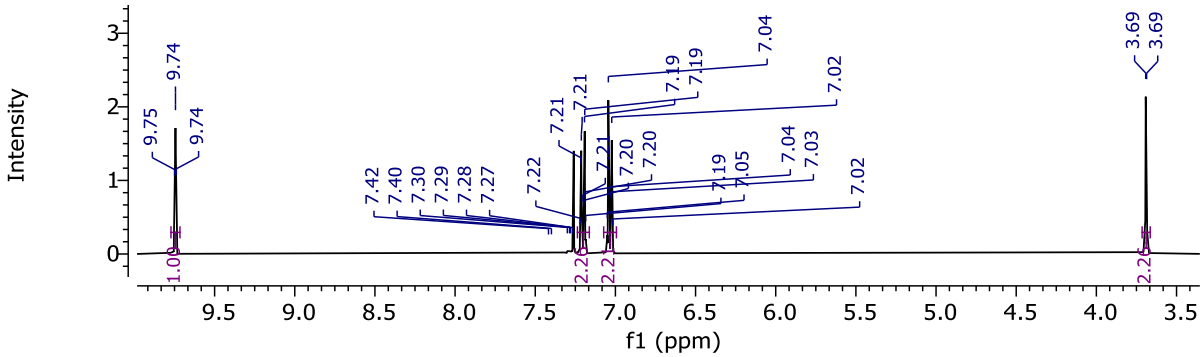


Figure S20. ^{1}H-NMR spectrum of **15**.

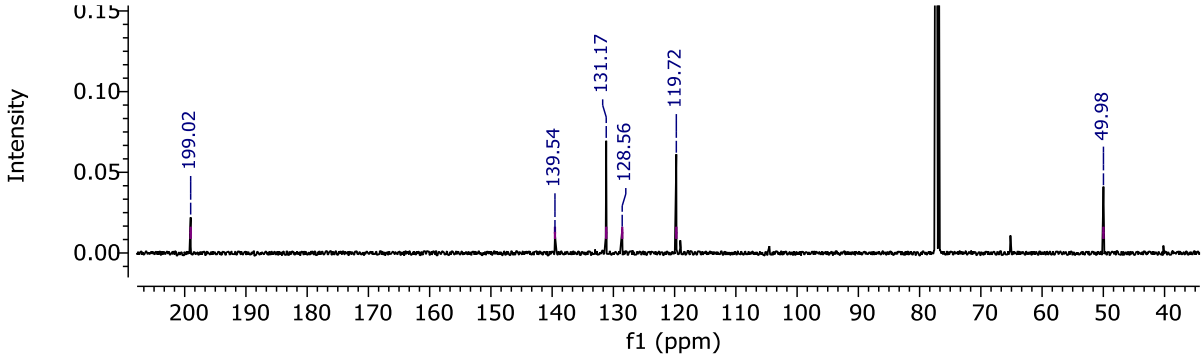


Figure S21. ^{13}C-NMR spectrum of **15**.

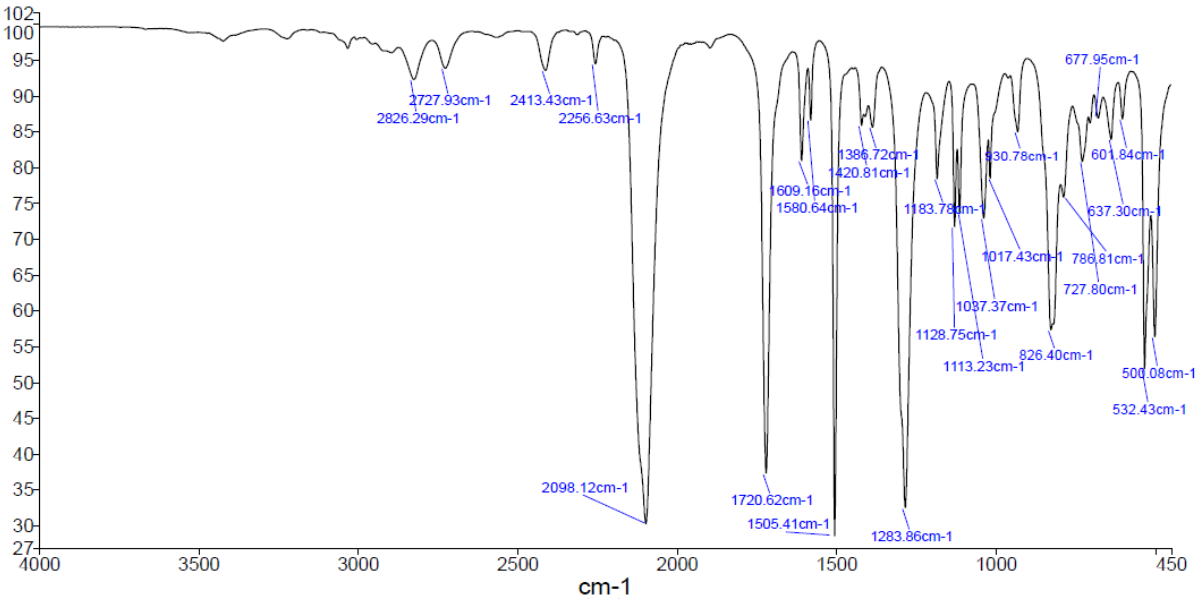
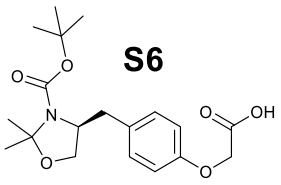


Figure S22. FT-IR (ATR) spectrum of **15**.

Biotinylated SCALP-donor probe S5 (via S4)

S4 was synthesised using solid-phase peptide synthesis (SPPS). **S5** was obtained via treatment of **S4** with $NaIO_4$. Protected small molecule SCALP warhead (S)-2-(4-((3-(tert-butoxycarbonyl)-2,2-dimethyloxazolidin-4-yl)methyl)phenoxy)acetic acid **S6** was synthesised as described previously.[3]



To synthesise **S4**, H-Gly-2-ClTrt resin (bearing 0.051 mmol of glycine functionality) was weighed out into an SPPS cartridge fitted with a PTFE stopcock, swollen in DMF for 30 minutes and then filtered. For each amino acid coupling during the SPPS of **S4** the following method was used:

DIPEA (98 µL, 0.56 mmols, 11 eq) was added to a solution of Fmoc-protected amino acid (0.26 mmols, 5 eq) and HTCU (107 mg, 0.26 mmols, 5 eq) dissolved in the minimum volume of DMF. The resultant solution was then immediately added to the resin. The reaction mixture was gently agitated by rotation for 1 h and the resin was filtered off and washed with DMF (3 x 2 minutes with rotation). A solution of 20 % piperidine in DMF was added to the resin and the mixture gently agitated by rotation for 2 minutes. The resin was filtered off and this piperidine treatment process was repeated a further four more times. The resin (now bearing deprotected amine functionality) was then washed using DMF (5 x 2 minutes with rotation) prior to coupling on the next amino acid.

The order of the amino acids used to prepare the peptide backbone of **S4** were as follows: N_α-Biotinyl-N_ε-Fmoc-L-lysine, Fmoc-O-tert-butyl-L-serine, Fmoc-glycine, Fmoc-O-tert-butyl-L-serine, Fmoc-glycine, Fmoc-O-tert-butyl-L-serine, Fmoc-glycine.

After deprotecting the final glycine residue and washing with DMF (5 x 2 minutes with rotation), the peptide was *N*-terminally capped using the protected small molecule SCALP warhead **S6**. DIPEA (98 µL, 0.56 mmols, 11 eq) was added to a solution of **S6** (102 mg, 0.26 mmols, 5 eq) and HTCU (107 mg, 0.26 mmols, 5 eq) dissolved in the minimum volume of DMF. The resultant solution was then immediately added to the resin. The reaction mixture was gently agitated by rotation for 1 h and the resin was filtered off and washed with DMF (5 x 2 minutes with rotation).

The resin was washed with DCM (3 x 2 minutes with rotation) and MeOH (3 x 2 minutes with rotation). The resin was dried on a vacuum manifold and further dried on a high vacuum line overnight. A solution of cleavage cocktail (95:2.5:2.5 TFA:H_2O:triisopropylsilane) was then added to the resin and the mixture gently agitated by rotation for 1 h. The reaction mixture was drained into cold Et_2O (incubated at -20 °C prior to use) and centrifuged at 4000 rpm at 4 °C until pelleted (ca 5-10 minutes). The supernatant was carefully decanted. The pellet was subsequently resuspended in cold Et_2O and the centrifugation and supernatant decantation process repeated a further three more times. The precipitated peptide pellet was then dissolved in water and lyophilised to obtain **S4** (**Figure S23**) as a powder.

(ESI)MS: Found $[M+H]^+$ 1069.4711, $C_{44}H_{69}N_{12}O_{17}S^+$ requires 1069.4624. Found $[M+2H]^{2+}$ 535.1945, $C_{44}H_{70}N_{12}O_{17}S^{2+}$ requires 535.2351.

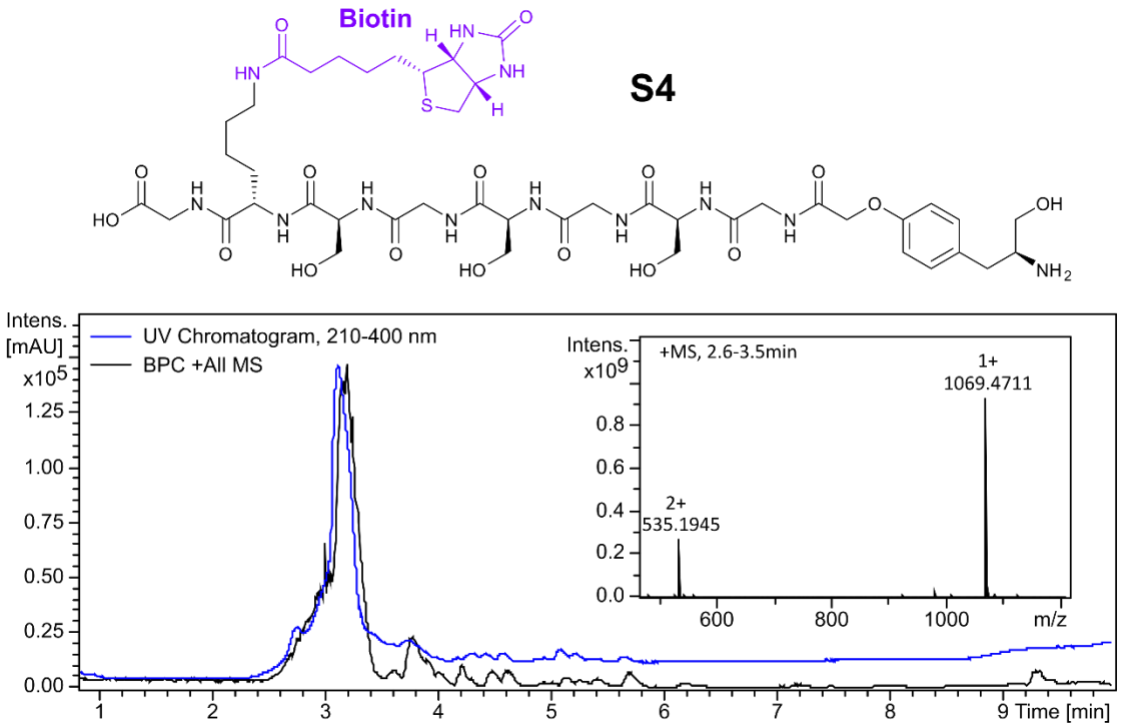


Figure S23. The structure and LC-MS trace of biotinylated molecule **S4**.

To a solution of **S4** (500 µL, 9.4 mM, 5 mg, in 0.1 M PB, 0.1 M NaCl, pH = 7.0) was added a solution of $NaIO_4$ (42 µL, 112 mM in 0.1 M PB, 0.1 M NaCl, pH = 7.0) in 6 × 7 µL additions. The reaction was mixed thoroughly and allowed to sit for 3 minutes on ice in darkness. The solution was then loaded onto a solid phase extraction cartridge (Grace Davison Extract Clean, 8 mL reservoir, Fisher Scientific) equilibrated with water/acetonitrile. After initial washing with water, the product was eluted over a gradient of acetonitrile. The product was then diluted with water, and subsequently lyophilised to give **S5** (**Figure S24**) as a pale-yellow fluffy powder which was used crude (4 mg, ≤ 82%)

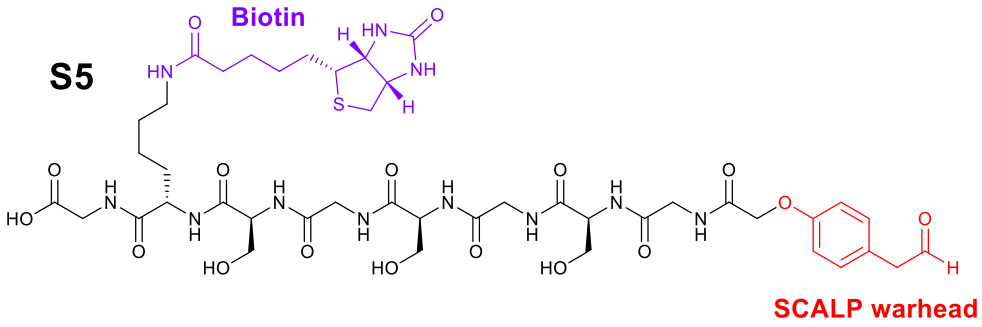


Figure S24. The structure of biotinylated SCALP-donor probe **S5**.

Heterobiofunctional terminal alkyne-DBCO linker S7

S7 was synthesised using SPPS. H-Gly-2-ClTrt resin (bearing 0.020 mmol of glycine functionality) was weighed out into an SPPS cartridge fitted with a PTFE stopcock, swollen in DMF for 30 minutes and then filtered. For each amino acid coupling during the SPPS of **S7** the following method was used:

DIPEA (37 µL, 0.21 mmols, 11 eq) was added to a solution of Fmoc-protected amino acid (0.10 mmols, 5 eq) and HTCU (41 mg, 0.10 mmols, 5 eq) dissolved in the minimum volume of DMF. The resultant solution was then immediately added to the resin. The reaction mixture was gently agitated by rotation for 1 h and the resin was filtered off and washed with DMF (3 x 2 minutes with rotation). A solution of 20 % piperidine in DMF was added to the resin and the mixture gently agitated by rotation for 2 minutes. The resin was filtered off and this piperidine treatment process was repeated a further four more times. The resin (now bearing deprotected amine functionality) was then washed using DMF (5 x 2 minutes with rotation) prior to coupling on the next amino acid.

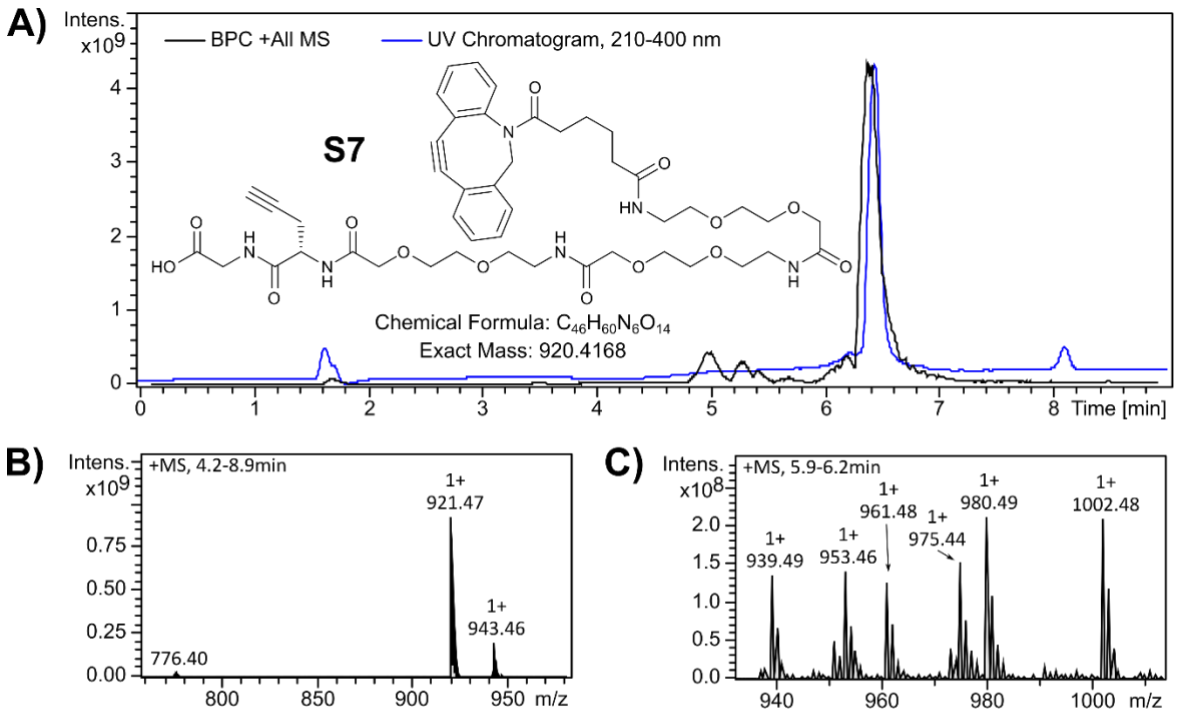
The order of the amino acids used to prepare the linker backbone of **S7** were as follows: Fmoc-L-propargylglycine, Fmoc-8-amino-3,6-dioxaoctanoic acid, Fmoc-8-amino-3,6-dioxaoctanoic acid, Fmoc-8-amino-3,6-dioxaoctanoic acid.

After deprotecting the final 8-amino-3,6-dioxaoctanoic acid residue and washing with DMF (5 x 2 minutes with rotation), the peptide was *N*-terminally capped using dibenzocyclooctyne-acid. DIPEA (11 µL, 0.062 mmols, 3.3 eq) was added to a solution of dibenzocyclooctyne-acid (10 mg, 0.030 mmols, 1.5 eq) and HTCU (12.3 mg, 0.030 mmols, 1.5 eq) dissolved in the minimum volume of DMF. The resultant solution was then immediately added to the resin. The reaction mixture was gently agitated by rotation for 24 h and the resin was filtered off and washed with DMF (5 x 2 minutes with rotation).

The resin was washed with DCM (3 x 2 minutes with rotation) and MeOH (3 x 2 minutes with rotation). The resin was dried on a vacuum manifold and further dried on a high vacuum line overnight. A solution of 1,1,1,3,3,3-hexfluoropropan-2-ol in DCM (1:4 v/v) was added to the resin and the mixture gently agitated by rotation for 1 h, mildly cleaving **S7** from the resin via solventolysis[4] while leaving the DBCO unit intact. The reaction mixture was drained into cold Et_2O (incubated at -20 °C prior to use) and centrifuged at 4000 rpm at 4 °C until pelleted (ca 5-10 minutes). The supernatant was carefully decanted. The pellet was subsequently resuspended in cold Et_2O and the centrifugation and supernatant decantation process repeated a further three more times. The precipitated peptide pellet was then dissolved in water and lyophilised to obtain **S7** (**Figure S25**) as a yellow solid (14 mg, 78%).

(ESI)HRMS: Found $[M+H]^+$ 941.4271, $C_{46}H_{61}N_6O_{14}$ requires 921.4240.

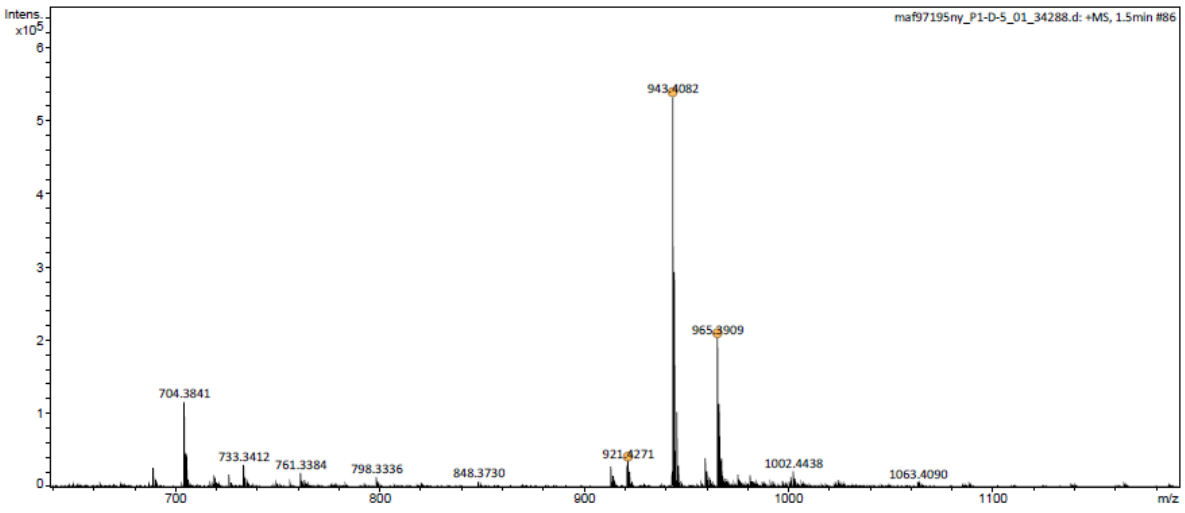
^{1}H-NMR (400 MHz, $CDCl_3$): δ_H 8.38 (t, *J* = 5.87 Hz, 1H), 7.79 (d, *J* = 8.64 Hz, 1H), 7.72 – 7.23 (m, 10H), 5.01 (d, *J* = 14.01 Hz, 1H), 4.52 (td, *J* = 8.08, 5.28 Hz, 1H), 3.93 (d, *J* = 2.95 Hz, 2H), 3.87 – 3.82 (m, 4H), 3.74 (d, *J* = 5.69 Hz, 2H), 3.63 – 3.06 (m, CH_2 environments, integral cannot be discerned due to the H_2O peak), 2.82 (t, *J* = 2.42 Hz, 1H), 2.67 – 2.50 (m, 3H), 2.20 – 2.10 (m, 1H), 1.83 (t, *J* = 6.62 Hz, 2H), 1.79 – 1.66 (m, 1H), 1.49 – 1.05 (m, 6H).



D)

m/z found	+ H_2O	+ MeOH	+ Na^+	+ H^+	+ HCOOH	Expected m/z	Diff
921.47	0	0	0	1	0	921.42	0.05
943.46	0	0	1	0	0	943.42	0.04
939.49	1	0	0	1	0	939.42	0.07
953.46	0	1	0	1	0	953.42	0.04
961.48	1	0	1	0	0	961.42	0.06
975.44	0	1	1	0	0	975.42	0.02
980.49	2	0	1	0	0	979.42	1.07
1002.48	2	0	0	0	1	1002.42	0.06

Figure S25. LC-MS analysis of the terminal alkyne + DBCO functionalised peptide **S7**. **A)** LC-MS chromatogram. **B)** MS trace for elution times 4.2 – 8.9 min showing **S7** (for which the theoretical m/z values of $[M+H]^+$ and $[M+Na]^+$ are 921.42 and 931.41 respectively) to be the dominant species. **C)** MS trace for elution time 5.9-6.2 min, showing species derived from acid-catalysed hydrolysis, or alcoholysis, of the DBCO motif of **S7**. **D)** Theoretical and calculated m/z values for the species detected during the LC-MS analysis of **S7**.



Meas. m/z	#	Ion Formula	m/z	err [ppm]	err [mDa]	mSigma	Mean err [ppm]
921.4271	1	C46H61N6O14	921.4240	-3.3	-3.1	37.1	-2.7
943.4082	1	C46H60N6NaO14	943.4060	-2.4	-2.3	18.5	-2.3
965.3909	1	C46H59N6Na2O14	965.3879	-3.1	-3.0	21.7	-4.3

Figure S26. (ESI)HRMS of **S7**.

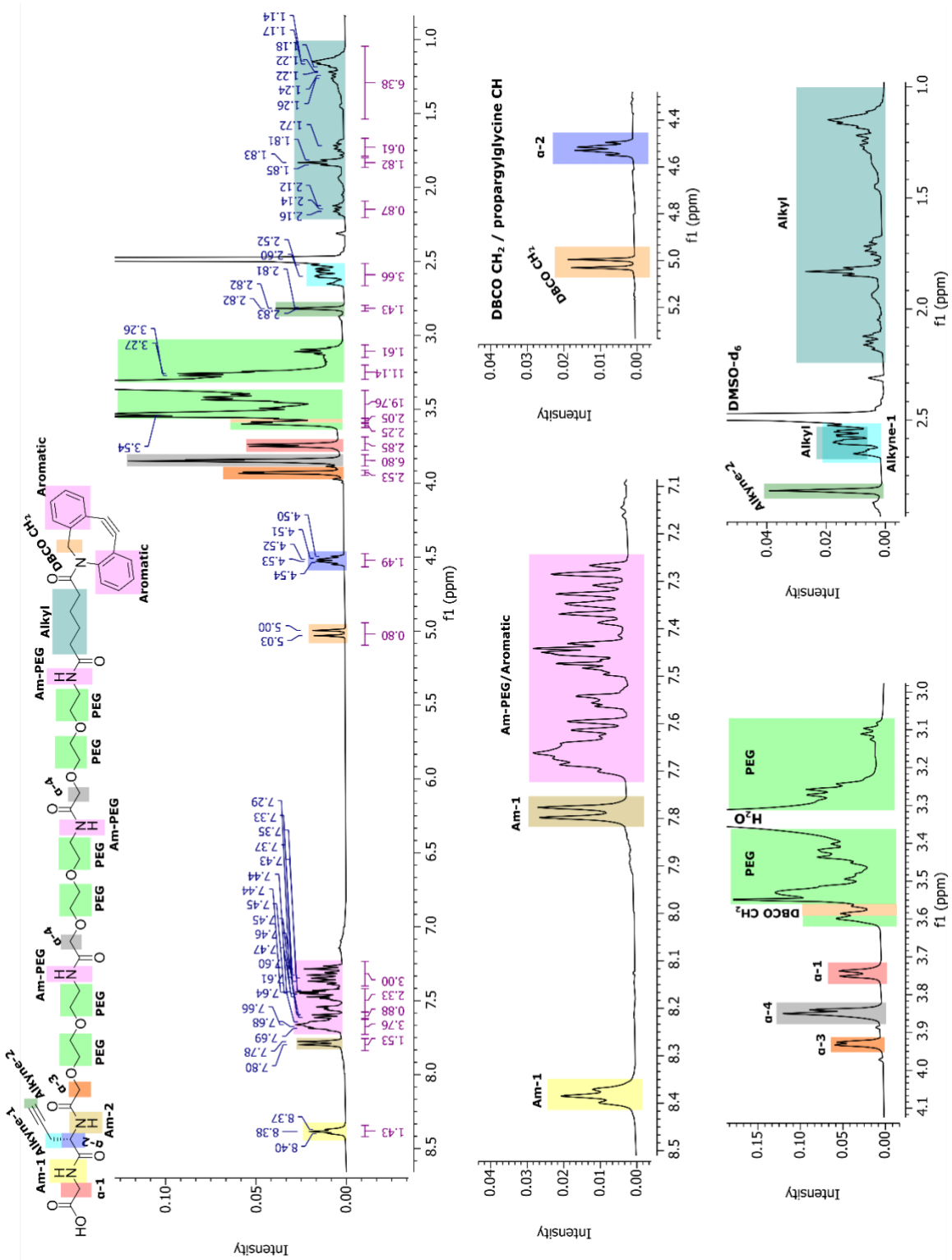
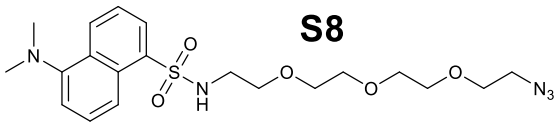


Figure S27. ^{1}H-NMR spectrum of **S7**.

***N*-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)-5-(dimethylamino)naphthalene-1-sulfonamide S8**

To a solution of dansyl chloride (27 mg, 0.10 mmol) in DCM (1 mL) was added 11-azido-3,6,9-trioxaundecan-1-amine (25 mg, 23 µL, 1.10 mmol) and triethylamine (10 mg, 14 µL, 0.10 mmol). The resultant solution was stirred overnight at room temperature. The reaction mixture was then dry-loaded onto silica, and purified via flash column chromatography (hexane → EtOAc) to yield **S8** as a viscous yellow oil (35 mg, 73%). Characterisation was consistent with that present in the literature.[5]



^{1}H-NMR (400 MHz, $CDCl_3$): δ_H 8.55 (d, *J* = 8.43 Hz, 1H), 8.32 (d, *J* = 8.62 Hz, 1H), 8.25 (dd, *J* = 7.27, 1.05 Hz, 1H), 7.57 (dd, *J* = 8.62, 7.51 Hz, 1H), 7.53 (dd, *J* = 8.43, 7.27 Hz, 1H), 7.20 (d, *J* = 7.51 Hz, 1H), 5.47 (t, *J* = 5.88 Hz, 1H), 3.72 – 3.64 (m, 4H), 3.64 – 3.59 (m, 2H), 3.51 (m, 2H), 3.39 (m, 6H), 3.11 (m, 2H), 2.90 (s, 6H).

^{13}C-NMR (100 MHz, $CDCl_3$): δ_C 149.91, 135.45, 130.00, 129.82, 129.54, 128.41, 123.41, 119.23, 115.40, 70.81, 70.78, 70.61, 70.43, 70.21, 69.36, 50.81, 45.61, 43.24 (×2).

(ESI)HRMS: Found $[M+Na]^+$ 474.1787, $C_{20}H_{29}N_5NaO_5S$ requires 474.1782.

FT-IR (ATR) (υmax/cm^{-1}): 2938 (C-H stretch, aromatic), 2871 (C-H stretch, alkyl), 2106 (N=N=N stretch, azide), 1145.

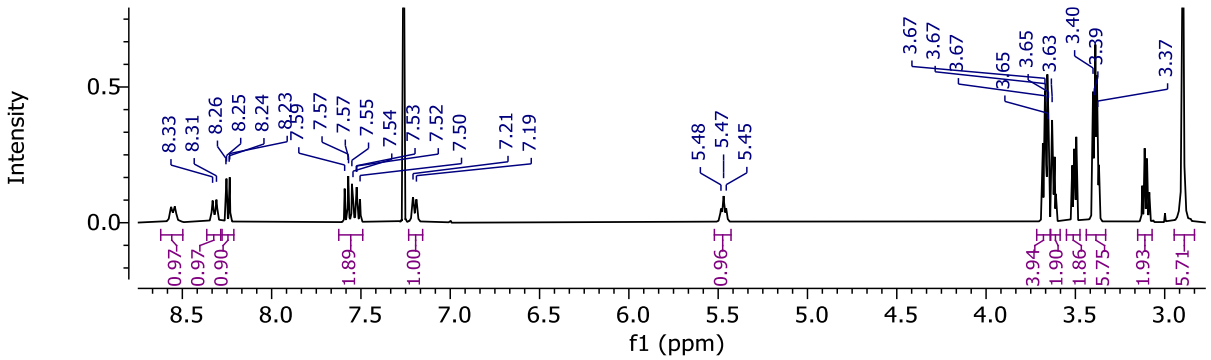


Figure S28. ^{1}H-NMR spectrum of **S8**.

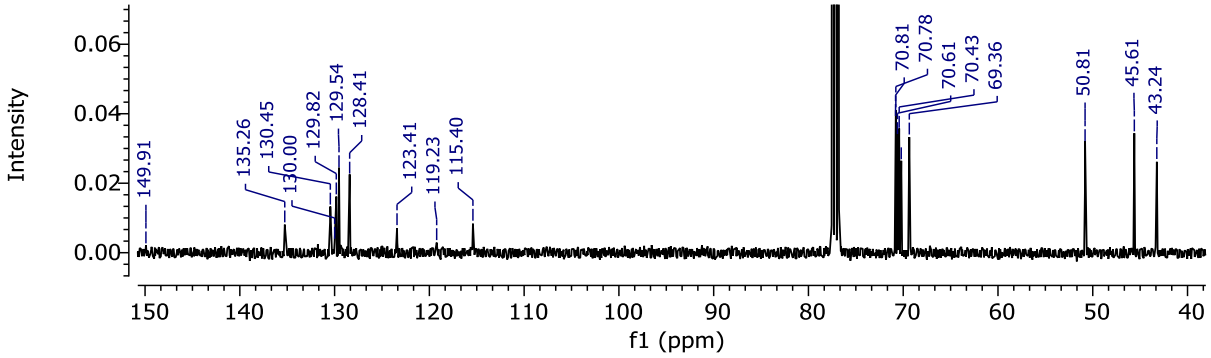
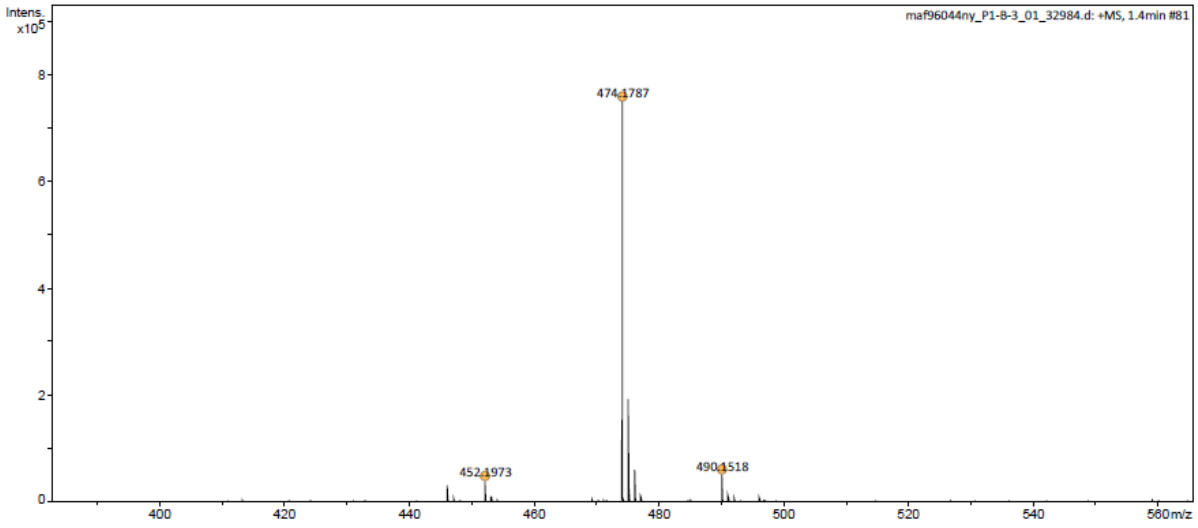


Figure S29. ^{13}C-NMR spectrum of **S8**.



Meas. m/z	#	Ion Formula	m/z	err [ppm]	err [mDa]	mSigma	Mean err [ppm]
452.1973	1	C20H30N5O5S	452.1962	-2.4	-1.1	12.5	-0.9
474.1787	1	C20H29N5NaO5S	474.1782	-1.1	-0.5	5.3	-1.6
490.1518	1	C20H29KN5O5S	490.1521	0.7	0.3	11.9	1.0

Figure S30. (ESI)HRMS of **S8**.

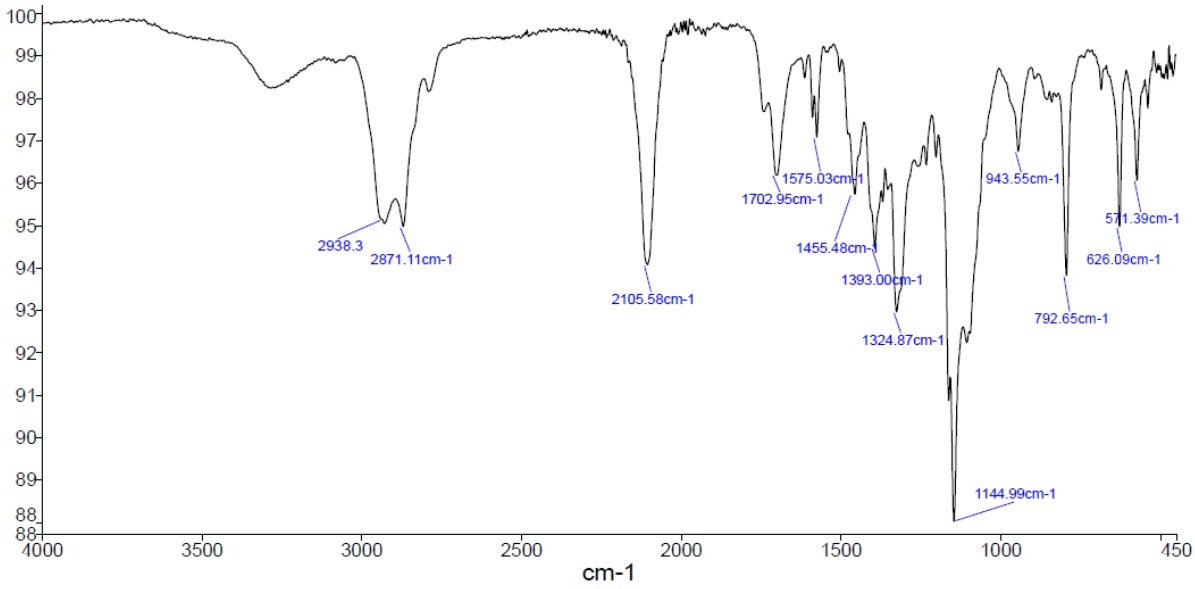


Figure S31. FT-IR (ATR) spectrum of **S8**.

Protein preparation, SDS page gel and western blotting protocols

Analysis of proteins via Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

15% SDS-PAGE gels were often used for the analysis of protein expression, purification, or bioconjugation experiments. These 15% SDS-PAGE gels were poured in-house using a specialised kit, and were made using the following protocol: H_2O (2.4 mL) was mixed with 2.5 mL of resolving buffer (1.5 M Tris-HCl, 0.4% SDS, pH 8.8). To the resultant mixture was added 30% acrylamide (5 mL), 20% ammonium persulfate solution (50 µL) and tetramethylethylenediamine (10 µL). The resultant mixture was briefly gently agitated to ensure mixing, and then poured into the mould and allowed to set. Once the gel had set a stacking gel was allowed to set on top of the main gel, with a comb inserted to create sample lanes. The protocol for mixing the stacking gel is as follows: H_2O (3.2 mL) was mixed with 1.3 mL of resolving buffer. To the resultant mixture was added 30% acrylamide (0.5 mL), 20% ammonium persulfate solution (12.5 µL) and tetramethylethylenediamine (8 µL).

Once the SDS-PAGE gels were fully prepared, complete with stacker gel and sample lanes, the SDS-PAGE gels were ready to be loaded with samples. Unless otherwise stated, samples were mixed with a 5 × concentrated reducing buffer (10% SDS, 10 mM 2-mercaptoethanol, 20% glycerol, 200 mM Tris-HCl pH 6.8, 0.05% bromophenol blue) and boiled for 5 min prior to running on the SDS-PAGE gel. The molecular weight markers used were SDS-PAGE Molecular Weight Standards, Low Range (Bio-Rad). Each gel was run at 200 V for 30-80 min in SDS running buffer (25 mM Tris, 192 mM Gly, not pH adjusted). In cases where concentrations could be easily determined, typically $\sim$ 2 µg of purified protein was loaded into each gel lane when conducting SDS-PAGE analyses.

After being run, SDS-PAGE gels were fixed via gel emersion in a fixing solution (40% water, 50% MeOH, 10% AcOH) which was gently rocked for 60 minutes. For experiments in which the SDS-PAGE gel was subsequently stained with Coomassie stain, the fixed gels were then immersed in a solution of 0.1% Coomassie Brilliant Blue R-250 (in 40% water, 50% MeOH, 10% AcOH), and the solution brought to the boil in a microwave, before being gently rocked at room temperature for a further 20 min. The Coomassie stained gels would then be destained via emersion in a detaining solution (50% water, 40% MeOH, 10% AcOH), which was periodically replaced with dye-free destaining solution as the destaining process progressed. Images of the resulting gels were captured and analysed using a Syngene G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera, with GeneSys software (Version 1.5.7.0). Note that for experiments where fluorescently-labelled proteins were used, imaging of fluorescence was performed after fixing the SDS-PAGE gel in fixing solution, but prior to Coomassie staining.

Analysis of proteins using western blots

The blots were assembled using 12 layers of blotting paper and one 15% SDS PAGE gel, all of which had previously soaked in transfer buffer (25 mM Tris, pH 8.3, 192 mM glycine, 20% MeOH) for 10 min. A further layer of polyvinylidene difluoride membrane was used which had been soaked in methanol (30 seconds), water (30 seconds) and transfer buffer (25 mM Tris-HCl pH = 8.3, 192 mM glycine, 20% MeOH) (10 min).

The blots were assembled in the following order: 6 layers blotting paper, PVDF membrane, 15% SDS PAGE gel, 6 layers of blotting paper. The gels were transferred to the PVDF membrane using a Trans-Blot®Turbo™ transfer system running at 1.3 A constant; up to 25 V for 30 minutes. The membranes were transferred to a 1 x Tris buffered saline + 5% milk solution and incubated for 1 h at 4 °C with gentle agitation. The membranes were then incubated with 1:4000 anti-Biotin antibody (HRP) GTX77581 (GeneTex) for 1 h in a 1% milk solution. The membranes were washed twice with 1 x PBS 0.1% tween® 20 (Sigma Aldrich) and once with 1 x PBS, all washes were performed for five minutes with gentle agitation. The blots were incubated with 4 mL of Amersham (Cytiva) for five minutes and the blots were imaged using a Syngene G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera, with GeneSys software (Version 1.5.7.0).

JVZ007 production and purification

A plasmid encoding the anti-PSMA JVZ007 nanobody[6] with an *N*-terminal serine and His-SUMO tag at the *N*-terminus cloned into the NcoI and BamHI restriction sites of pET15b was ordered from GenScript. JVZ007 was produced as previously described.[7] Briefly, the plasmid was introduced into chemically competent *E.Coli* BL21 (Gold) cells by heat shock and selected on LB agar with ampicillin (100 μg mL^{-1}) at 37 °C for 16 h. Starter cultures were prepared by picking single clones into LB with ampicillin (100 μg mL^{-1}) and grown at 37 °C overnight with shaking (180 rpm). 3 mL of started culture was added to 500 mL of LB with ampicillin (100 μg mL^{-1}) (x4). The flasks were shaken at 180 rpm, 37 °C until the cultures had reached an OD_{600} of 0.6-0.7. IPTG was added to a final concentration of 0.1 mM and the cultures were grown at 20 °C overnight with shaking at 190 rpm. Cells were harvested by centrifugation (6000 x g, 6 °C, 10 min) and the pellet was resuspended in lysis buffer (0.1 M Tris pH 7.4, 150 mM NaCl, 30 mM imidazole, protease inhibitor, benzonase 1 U mL^{-1}). The cells were lysed by sonication on ice and the lysate was clarified by centrifugation (28,000 x g, 4 °C, 40 min). The lysate was loaded onto a 5 mL HisTrap FF column (GE Healthcare) pre-equilibrated with binding buffer (0.1 M Tris pH 7.4, 150 mM NaCl, 30 mM imidazole). After washing the column with 20 column volumes (CVs) of binding buffer the *N*-terminal His_6tag JVZ007 nanobody was eluted with elution buffer (0.1 M Tris pH 7.4, 150 mM NaCl, 0.5 M imidazole). The protein was then exchanged into 25 mM Tris pH 7.4, 150 mM NaCl by dialysis.

The His-SUMO tag was cleaved from the nanobody by adding DTT to a final concentration of 5 mM and sumo protease (10 μg per mg of protein) at 20 °C overnight with shaking at 110 rpm. Imidazole was added to a final concentration of 30 mM and loaded onto a 5 mL HisTrap FF column (GE Healthcare) pre-equilibrated with binding buffer (0.1 M Tris pH 7.4, 150 mM NaCl, 30 mM imidazole). The JVZ007 nanobody was collected in the flow through and dialysed into potassium-free PBS overnight at 4 °C, giving a yield of 111 mg from 2 L of culture. The nanobody was assessed by SDS-PAGE (**Figure S32**) and shown to be consistent with the predicted molecular mass of 12.1 kDa.

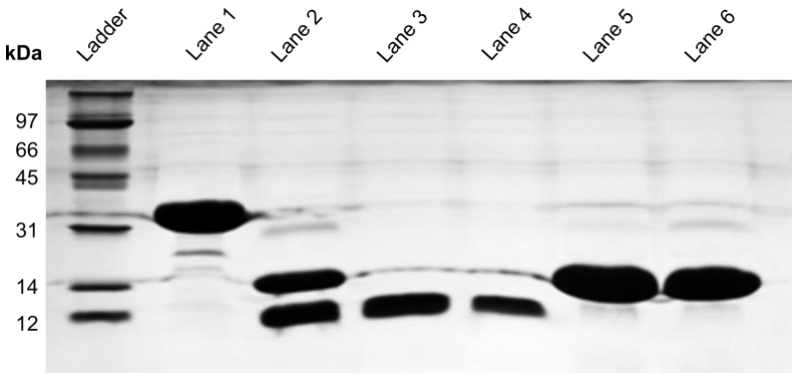


Figure S32. SDS-PAGE analysis of JVZ007 constructs / affinity tags at various stages of the preparation of JVZ007.[7] Ladder: molecular weight marker. Lane 1: Pre-SUMO cleavage. Lane 2: Pre-purification by nickel affinity chromatography. Lane 3: Flow-through from nickel affinity chromatography containing JVZ007. Lane 4: Wash from nickel affinity chromatography containing JVZ007. Lanes 5 and 6: Elution of His-SUMO tag during nickel affinity chromatography.

Supplementary Figures and Tables

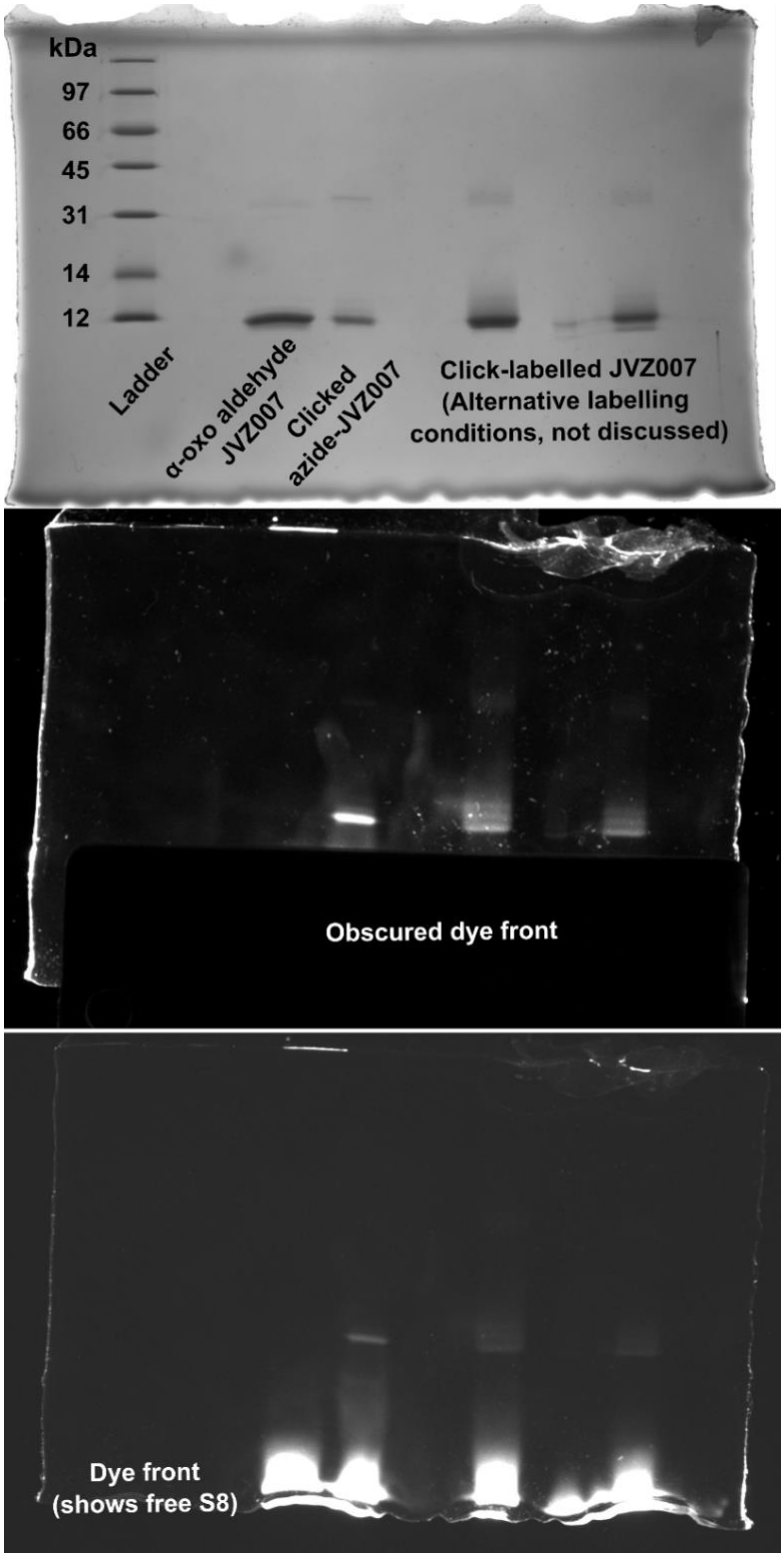


Figure S33. SDS-PAGE analysis of α-oxo aldehyde JVZ007 and clicked azide-JVZ007.

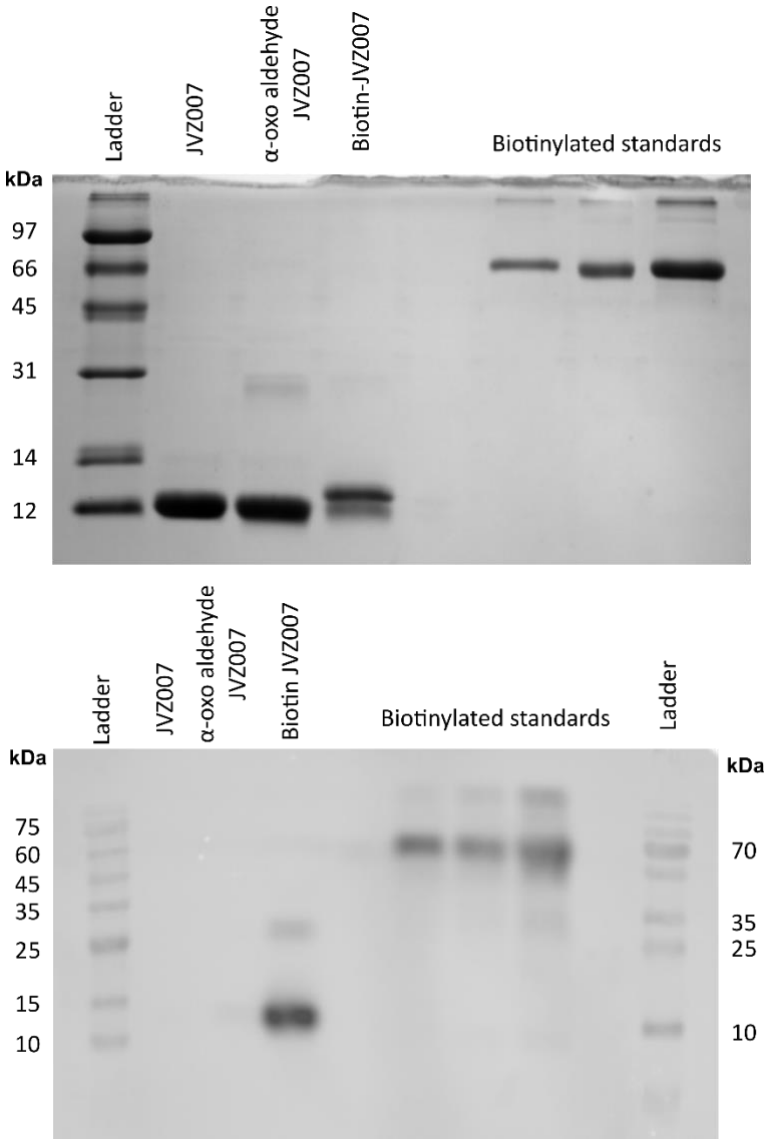


Figure S34. Top) SDS-PAGE analysis of α-oxo aldehyde JVZ007 and biotin-JVZ007. **Bottom)** Western blot analysis of α-oxo aldehyde JVZ007 and biotin-JVZ007.

Figure S35. The preparation of α-oxo aldehyde-LYRAG from SLYRAG using previously reported conditions.[3]

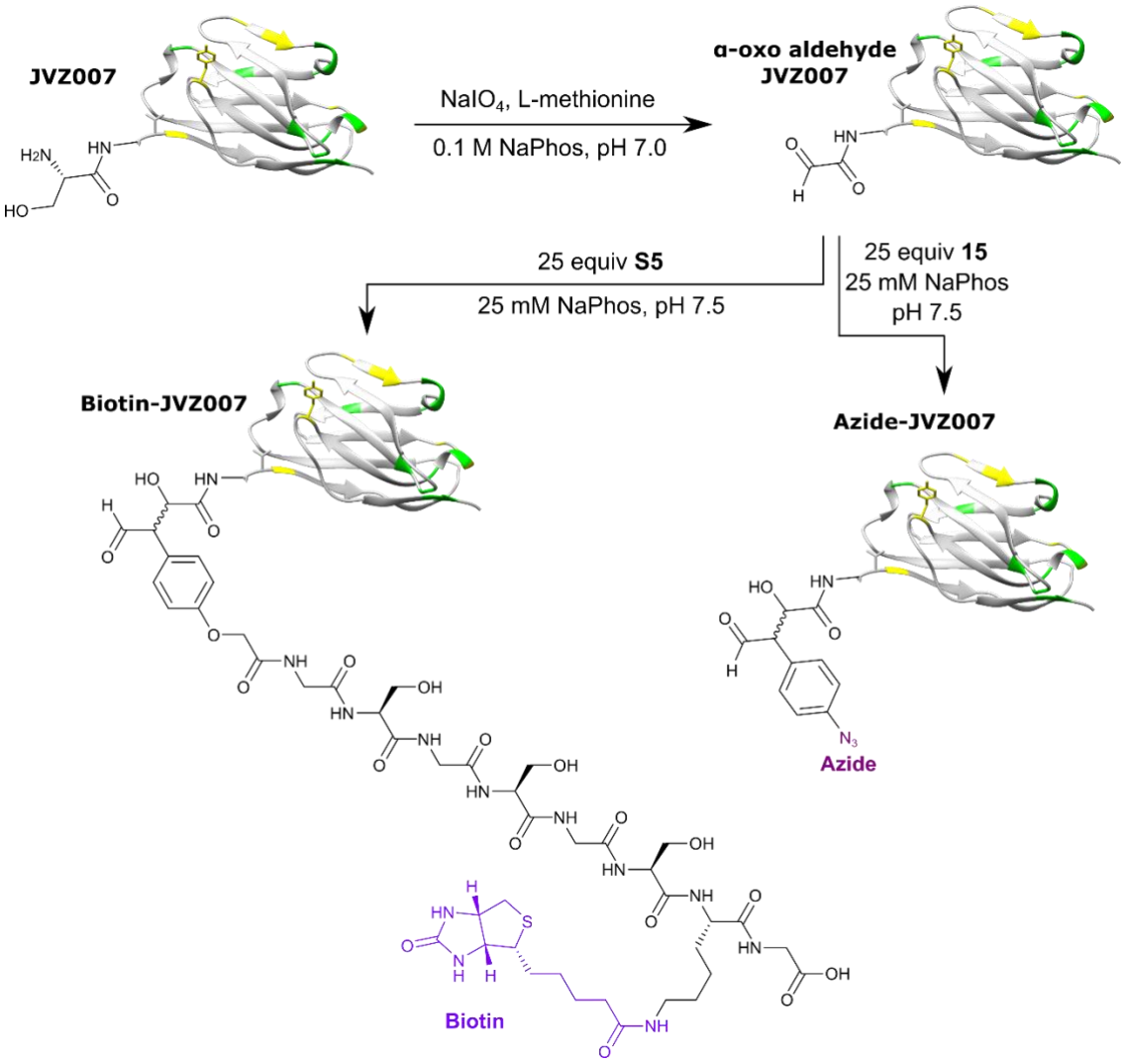


Figure S36. Top) The preparation of α-oxo aldehyde JVZ007 from JVZ007 via the oxidative cleave of the *N*-terminal serine residue with $NaIO_4$ using previously reported conditions.[7] **Bottom)** The preparation of biotin-JVZ007 and azide-JVZ007 from α-oxo aldehyde JVZ007.

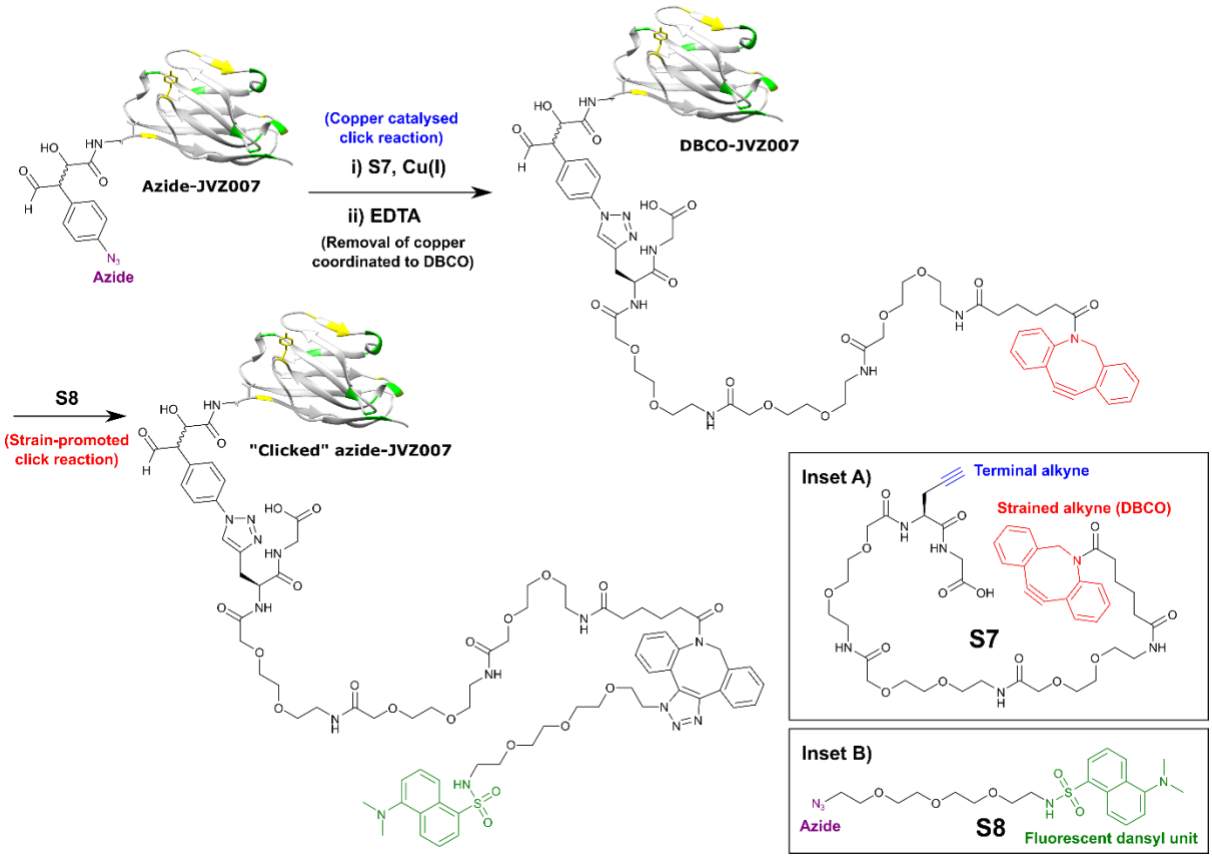
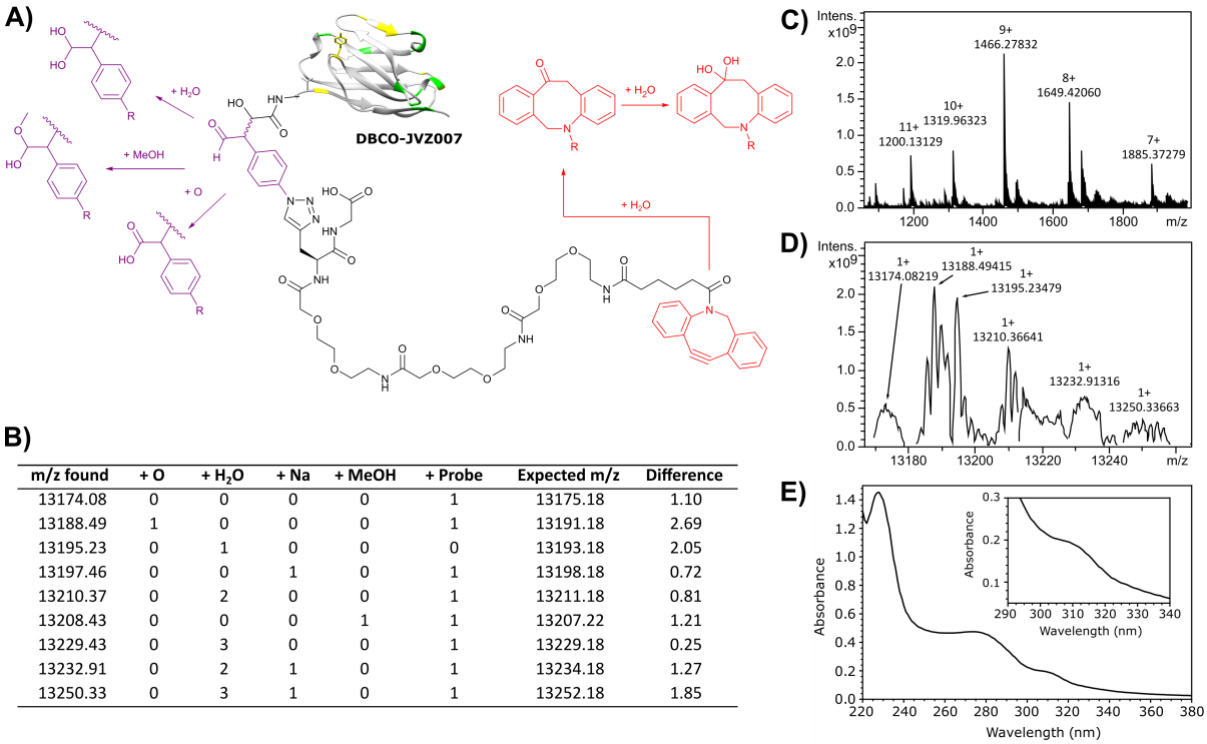


Figure S37. Main) The preparation of "clicked" azide-JVZ007 via sequential copper-catalysed click and copper-free click reactions. **Inset A)** The structure of heterobiofunctional terminal alkyne-DBCO linker **S7**. **Inset B)** The structure of azide-functionalised dansylated probe **S8**.



B)

m/z found	+ O	+ H_2O	+ Na	+ MeOH	+ Probe	Expected m/z	Difference
13174.08	0	0	0	0	1	13175.18	1.10
13188.49	1	0	0	0	1	13191.18	2.69
13195.23	0	1	0	0	0	13193.18	2.05
13197.46	0	0	1	0	1	13198.18	0.72
13210.37	0	2	0	0	1	13211.18	0.81
13208.43	0	0	0	1	1	13207.22	1.21
13229.43	0	3	0	0	1	13229.18	0.25
13232.91	0	2	1	0	1	13234.18	1.27
13250.33	0	3	1	0	1	13252.18	1.85

Figure S38. Characterisation of DBCO-JVZ007. Mass spectrometry analysis reveals peaks consistent with DBCO-JVZ007 bearing either free aldehyde, geminal diol, or hemiacetal functionality. The mass spectrometry analysis also reveals peaks consistent with hydrolysed DBCO motifs (both as ketones or geminal diols), which can be attributed to acid-catalysed addition of water across the strained alkyne of DBCO occurring during mass spectrometric analysis. A peak can also be observed that could be attributed to the oxidation of the aldehyde motif to a carboxylic acid, possibly via the action of copper. **A)** The various forms that DBCO-JVZ007 may be found in during mass spectrometry. **B)** Deconvoluted m/z values found for DBCO-JVZ007 and peak assignments. **C)** Raw mass spectrometric data showing the protein charge-state ladders. **D)** Deconvoluted mass spectrometric data. **E)** The UV-vis spectrum of DBCO-JVZ007 shows the characteristic absorbance at 280 nm attributable to proteins, and an additional absorbance at 310 nm, which is characteristic of an intact DBCO motif.[8]

Table S1. Theoretical monoisotopic masses for α-oxo aldehyde LYRAG derived products of SCALP bioconjugation.

Ligating species	Mass of ligating species / Da	α-oxo aldehyde LYRAG mass / Da	Expected product mass / Da	Expected $[M+H]^+$ m/z
H_2O	18.0105	634.3075	652.318	653.318
4	120.0575	634.3075	754.365	755.372
5	164.0837	634.3075	798.3912	799.3982
6	150.0681	634.3075	784.3756	785.3826
7	134.0732	634.3075	768.3807	769.3877
8	196.0888	634.3075	830.3963	831.4033
9	150.0681	634.3075	784.3756	785.3826
10	165.0426	634.3075	799.3501	800.3571
11	106.0419	634.3075	740.3494	741.3564
13	180.0786	634.3075	814.3861	815.3931
14	138.0481	634.3075	772.3556	773.3626
15	161.0589	634.3075	795.3664	796.3734

LC-MS traces for SCALP donor conversion and competition experiments

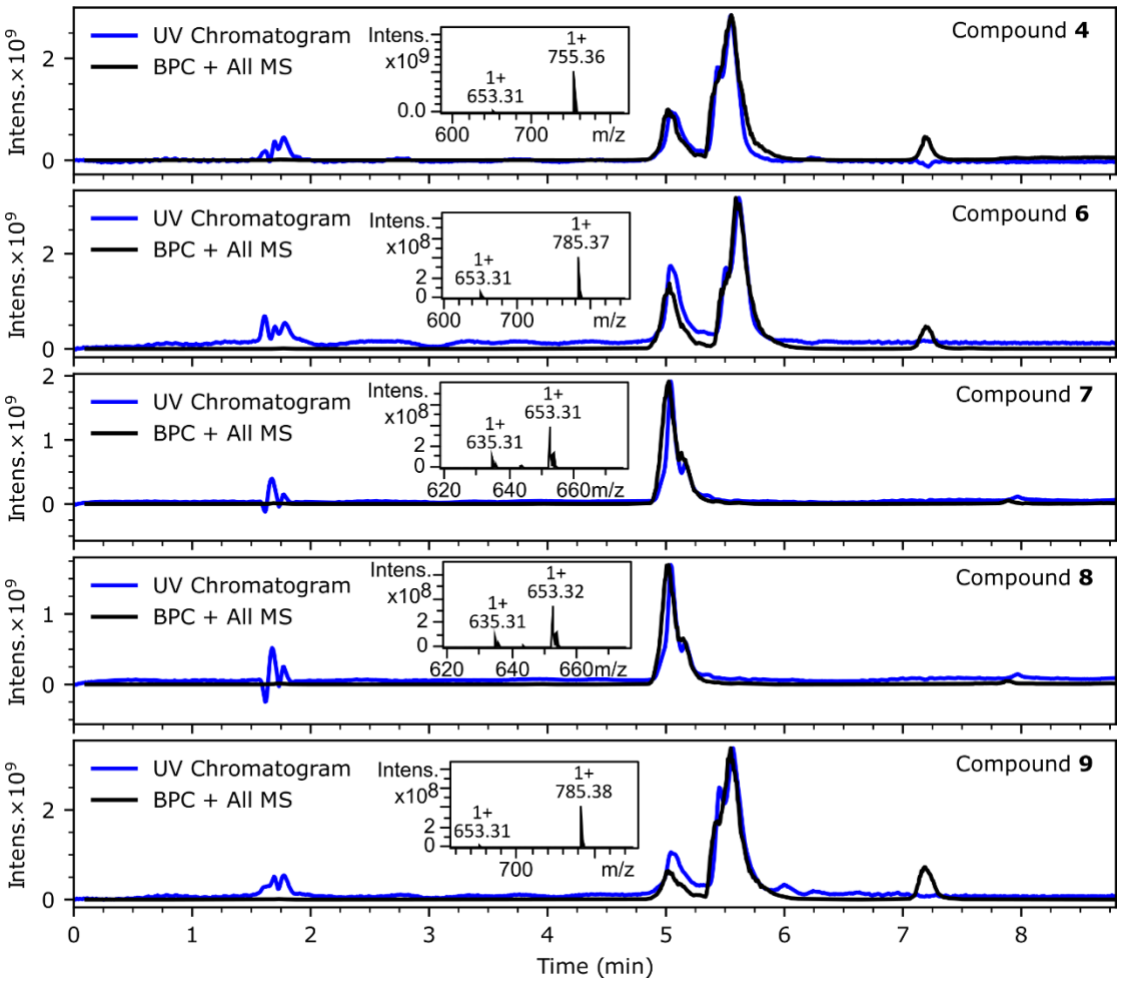


Figure S39. LC-MS traces for the reactions of of α-oxo aldehyde LYRAG **2** with various SCALP donors **4**-**9**. **Insets**) Mass spectra taken between 4 and 8 min.

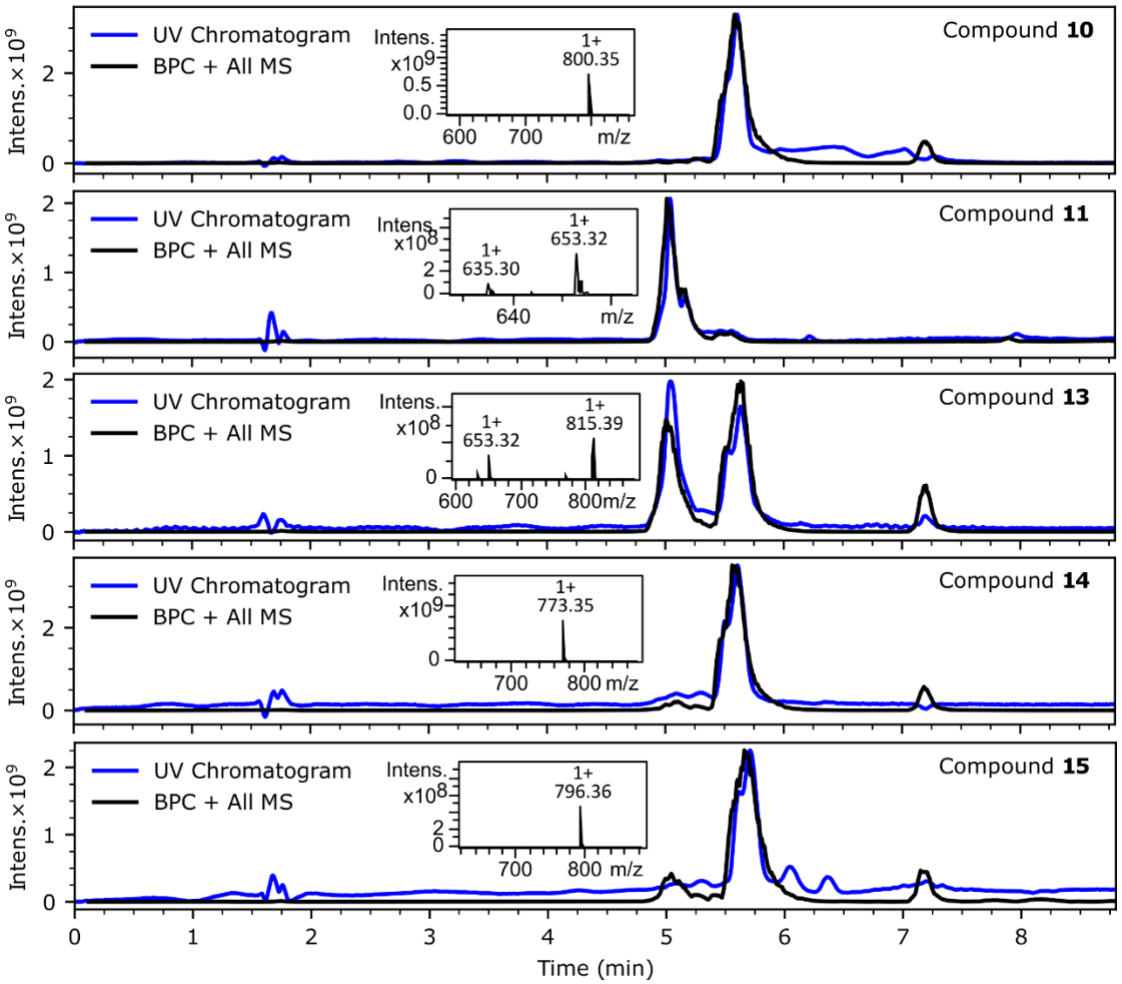


Figure S40. LC-MS traces for the reactions of α-oxo aldehyde LYRAG **2** with various SCALP donors **10**-**15**. **Insets**) Mass spectra taken between 4 and 8 min.

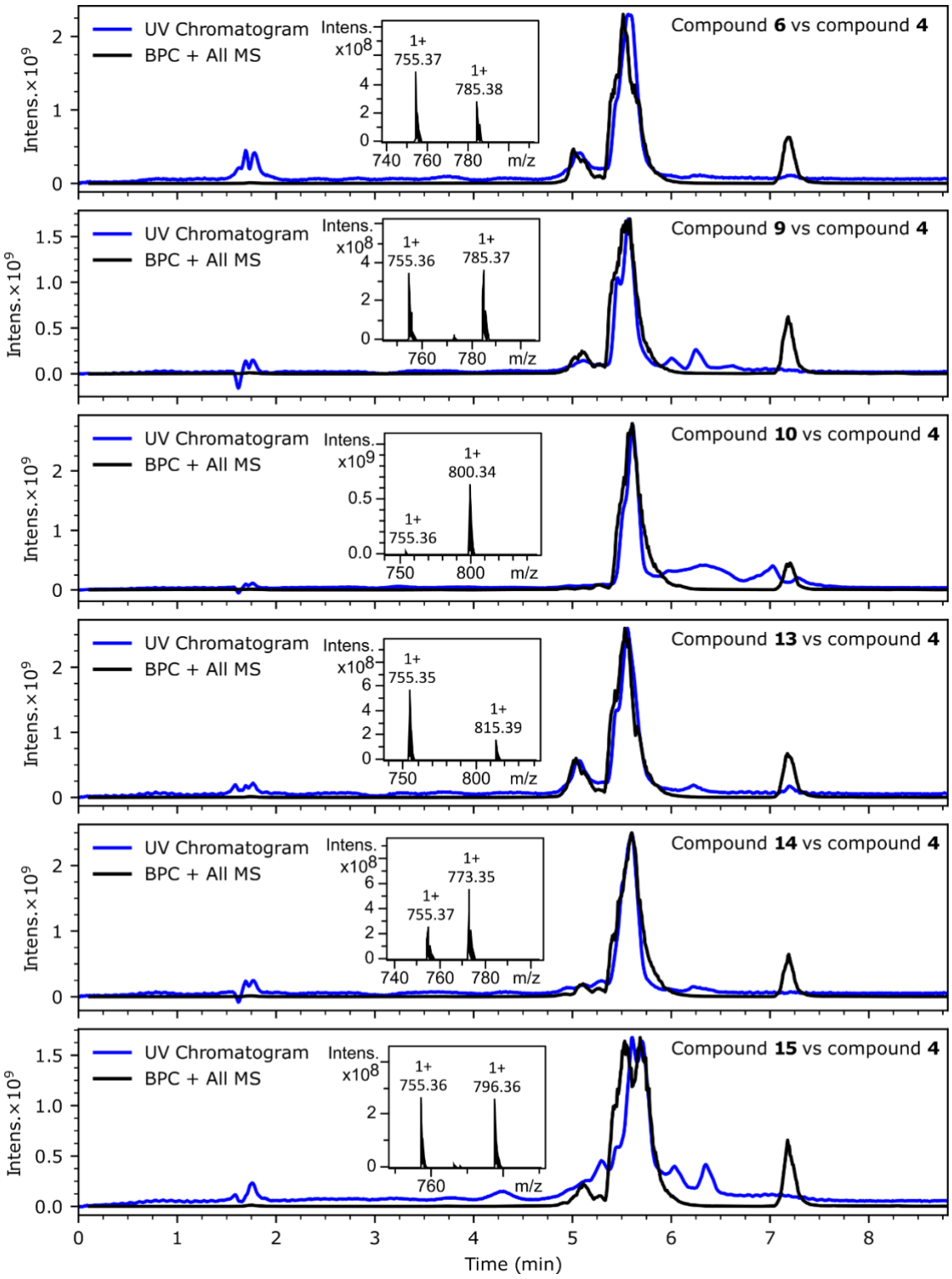


Figure S41. LC-MS traces for competition experiments involving α-oxo aldehyde LYRAG **2** and various SCALP donors. **Insets**) Mass spectra taken between 4 and 8 min.

Trypsin digestion of azide-JVZ007

Trypsin digestion was performed following in-solution following reduction with DTT and alkylation with iodoacetamide. Resulting peptides were analysed by LC-MS/MS using a Data Dependent Acquisition (DDA) approach on a maXis qTOF using a nano-ESI source. Peptides were eluted over a 45 min gradient from a 50 cm micropillar column. Resulting spectra were peak picked to .mgf format and searched against the sequence of JVZ007 using the Byonic search engine. Potential mass additions were included as variable modifications. Peptide identifications were filtered to 1D q-values of 0.01 or lower.

Good coverage is seen across the JVZ007 protein:

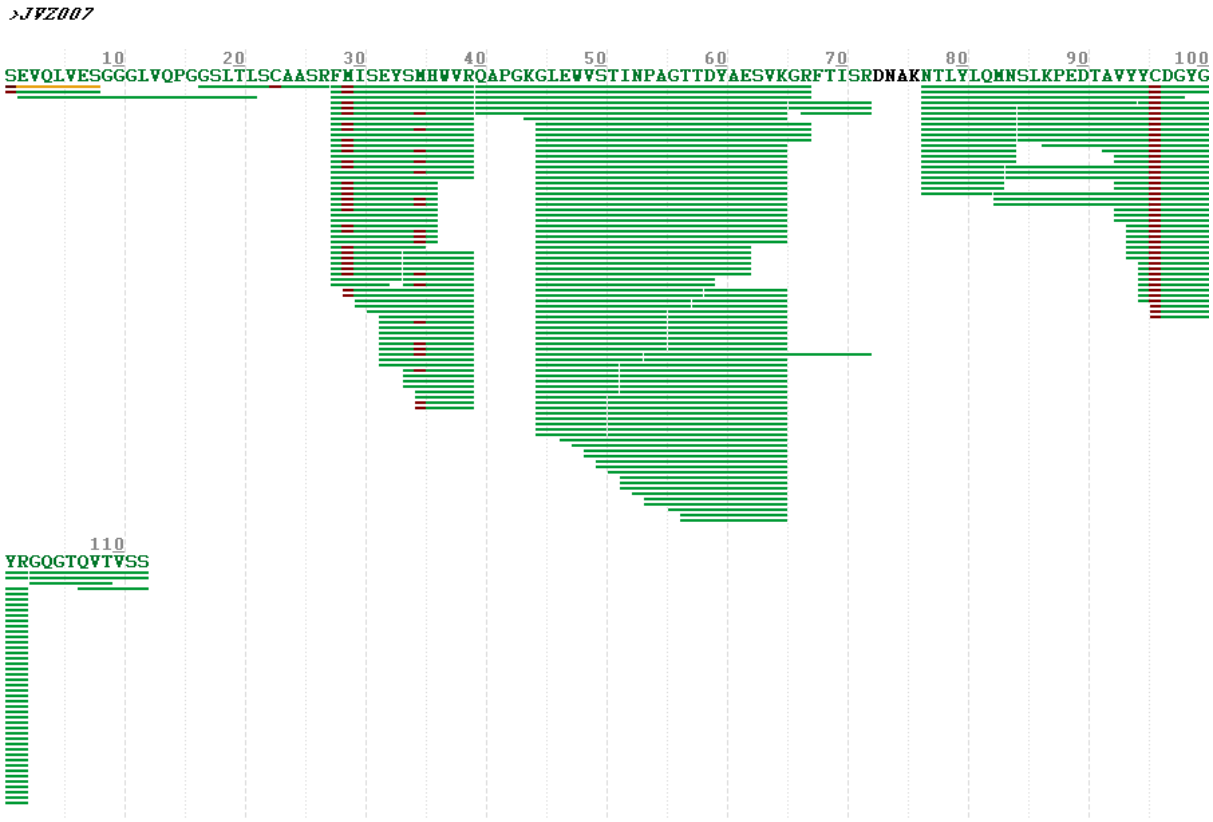


Figure S42. Sequence coverage of JVZ007 achieved via trypsin digestion.

Both typtic and semi-tryptic peptides (where cleavage at one terminus of the peptide is not C-terminal to Arg or Lys) were detected.

Mass spectrometry of tryptic peptides of azide-JVZ would be expected to reveal a modification at the *N*-terminus. Azides can fragment during mass spectrometry, and β-hydroxy aldehyde motifs can undergo dehydration reactions, so a variety of modifications could theoretically be detected. A few possible motifs are presented in **Figure S43.**

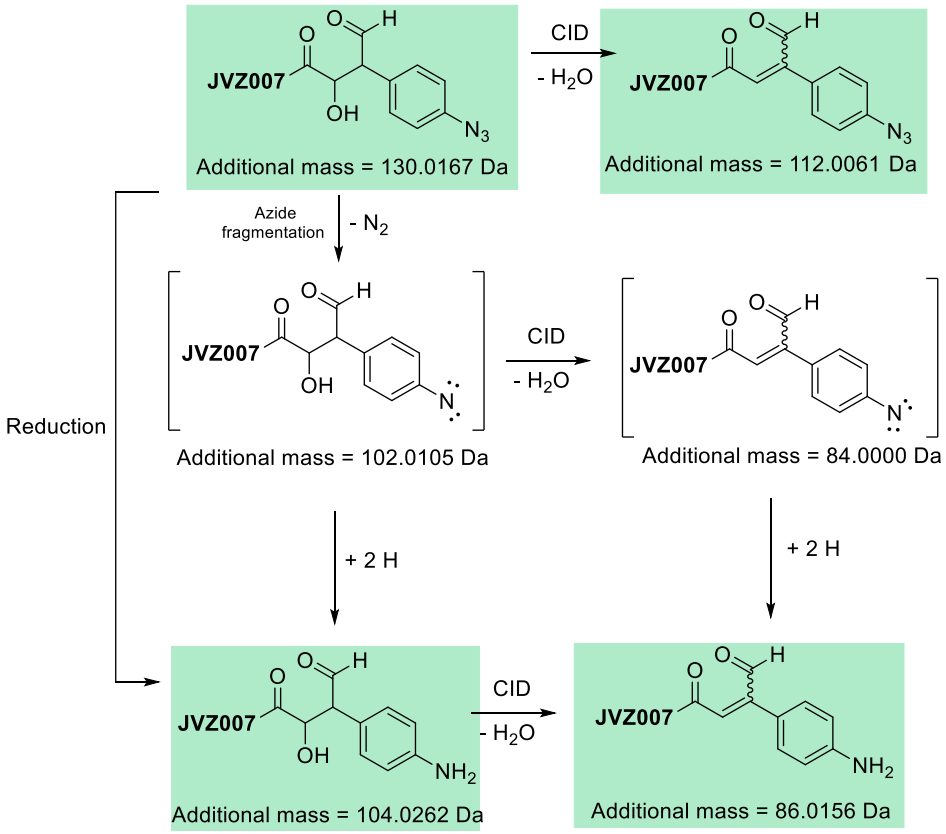


Figure S43. Additional masses may be detected the N-terminus of azide-JVZ007. Motifs marked in green have been detected during the analysis of tryptic peptides derived from azide-JVZ007.

Looking at modified peptides, *N*-terminal non-tryptic peptide is detected with a +130 Da mass addition: [+130.01670]SEVQLVES (**Figure S44**). The mass of this peptide is consistent with the expected mass of the N-terminal peptide fragment SEVQLVES +130.01670. The b-ions match to the internal sequence QLVE, although these are 18 Da smaller than the +130 Da increase expected to proceed them, suggesting the 130 Da addition contains a moiety that rapidly loses water on collision induced dissociation (CID), which is likely due to elimination of the hydroxyl group of the β-hydroxy aldehyde motif.

Sequence	Mods	PEP	PEP	|Log Prob	Score	Delta	Delta	z	Obs.	Calc.
-.[+130.01670].SEVQLVES.G	NTerm(OPAL_130 / 130.0167)	0.00027	0.00027	3.57	92.4	92.4	92.4	1	1020.461	1020.463

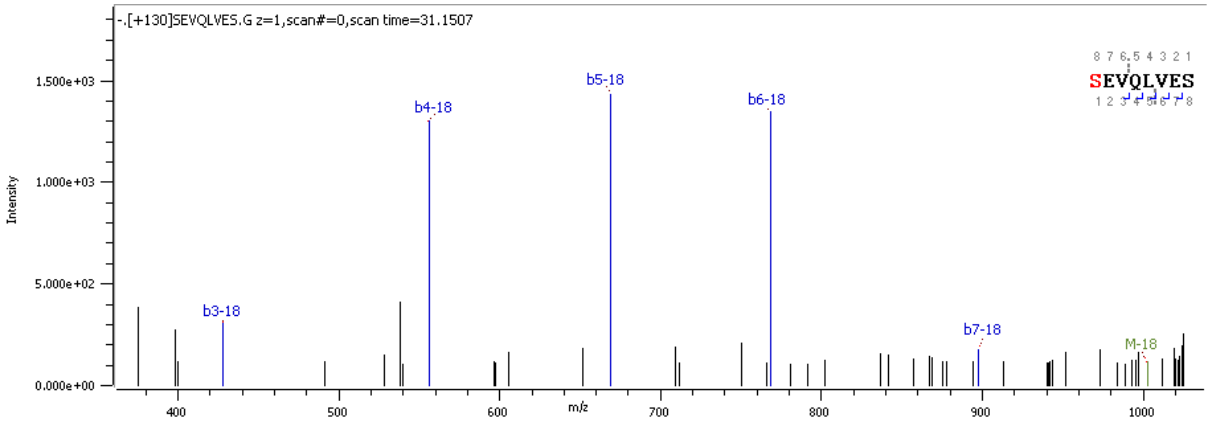


Figure S44. MS data showing the successful detection of a [+130.01670]SEVQLVES tryptic peptide, which is consistent with α-oxo aldehyde JVZ007 being modified with **15** at its *N*-terminus.

Opening up the analysis more widely to consider wildcard modifications a good match is seen to a 104.025 Da addition at the *N*-terminus (**Figure S45**). This is a match for a species that could be generated via fragmentation of the azide motif to a nitrene during ionisation[9] followed by hydrogen atom abstraction/protonation, or otherwise via reduction during preparation of the trypsin digestion sample or during mass spectrometry analysis.[9]

Sequence	Mods	PEP	PEP	|Log Prob	Score	Delta	Delta	z	Obs.	Calc.
-.S{+104.02490}EVQLVESGGGLVQPGGSL.T	S1(Wildcard / 104.0249)	0.0037	0.0038	2.43	189.7	68.9	0	2	958.983	958.983

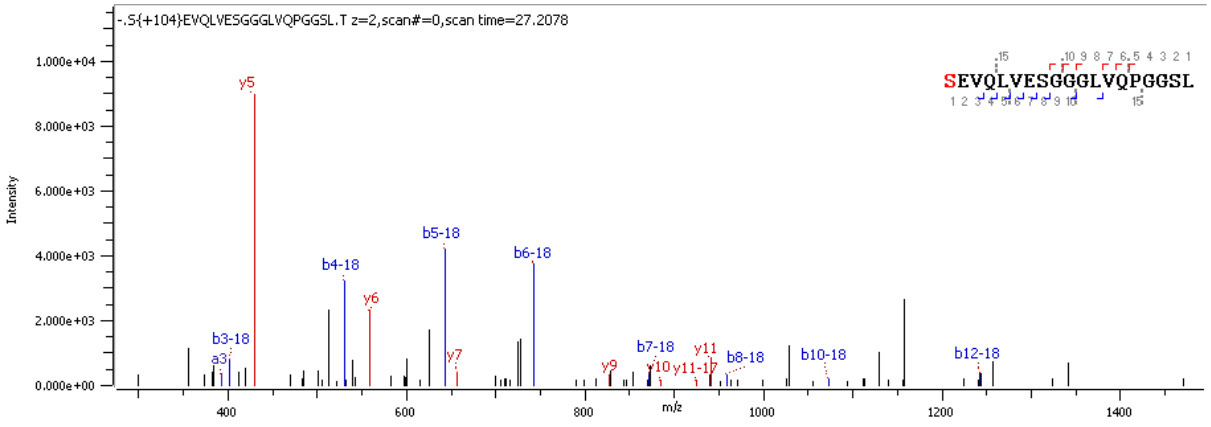


Figure S45. MS data showing the successful detection of a [+104.02490]SEVQLVESGGGLVQPGGSL tryptic peptide, which is consistent with α-oxo aldehyde JVZ007 being modified with **15** at its *N*-terminus, then undergoing conversion of the azide motif to an amine mass spectrometry.

The sequence is longer here and the y-ions help exclude modification to the C-terminus of this peptide. Once again, the observed b-ions show loss of water (-18 Da) suggesting the 104 Da mass addition is also liable to the same loss of the hydroxyl group of the β-hydroxy aldehyde under CID.

Calculation of E-factors/Reaction mass efficiencies for the preparation of azide JVZ007

Molecular mass of α-oxo aldehyde JVZ007: = 12093.47 g mol^{-1}

Molecular mass of 15: = 161.16 g mol^{-1}

Molecular mass of azide JVZ007: = 12254.64 g mol^{-1}

Monobasic sodium phosphate molecular mass: = 119.98 g mol^{-1}

Dibasic sodium phosphate molecular mass: = 141.96 g mol^{-1}

Ratio of Monobasic sodium phosphate to dibasic sodium phosphate at pH 7.5: = 1: 4.3

Averaged molecular mass of NaPhos at pH 7.5: = [(1 × 119.98) + (4.3 × 141.96)] g mol^{-1} / 5.3

= 137.82 g mol^{-1}

Molecular mass of (S)-(–)-5-(2-Pyrrolidinyl)-1H-tetrazole: = 139.16 g mol^{-1}

$$\textbf{Reaction mass efficiency}\ (\%) = 100 \times \frac{Mass\ of\ product}{Mass\ of\ all\ reaction\ components}$$

$$\textbf{Efactor} = \frac{Mass\ of\ waste}{Mass\ of\ product}$$

Classic OPAL conditions:

25 mM Na Phos, pH 7.5, 50 µM α-oxo aldehyde protein, 25 mM S)-(–)-5-(2-Pyrrolidinyl)-1H-tetrazole, 500 µM aldehyde donor.

= 1 equiv α-oxo aldehyde protein, 250 equiv averaged NaPhos, 250 equiv (S)-(–)-5-(2-Pyrrolidinyl)-1H-tetrazole, 10 equiv aldehyde donor.

$$Efactor = \frac{[(137.82 \times 250) + (250 \times 139.16) + (10 \times 161.16)]\ g\ mol^{-1}}{12254.64\ g\ mol^{-1}}$$

$\boldsymbol{Efactor} =$ **5.78**

$Reaction\ mass$ efficiency (%)

$$= 100\ \times \frac{12254.64\ g\ mol^{-1}}{[(137.82 \times 250) + (250 \times 139.16) + (10 \times 161.16) + 12093.47]g\ mol^{-1}}$$

$\boldsymbol{Reaction\ mass}$ **efficiency** $(\%) = \mathbf{14.8}\ \%$

SCALP conditions:

25 mM Na Phos, pH 7.5, 50 µM α-oxo aldehyde protein, 1.25 mM aldehyde donor.

= 1 equiv α-oxo aldehyde protein, 250 equiv averaged NaPhos, 25 equiv aldehyde donor.

$$Efactor = \frac{[(137.82 \times 250) + (25 \times 161.16)]\ g\ mol^{-1}}{12254.64\ g\ mol^{-1}}$$

$\boldsymbol{Efactor} =$ **3.14**

$Reaction\ mass$ efficiency (%)

$$= 100\ \times \frac{12254.64\ g\ mol^{-1}}{[(137.82 \times 250) + (25 \times 161.16) + 12093.47]g\ mol^{-1}}$$

$\boldsymbol{Reaction\ mass}$ **efficiency** $(\%) = \mathbf{24.2}\ \%$

SCALP conditions vs OPAL conditions:

% Improvement in E-factor: = [1 - (3.14/5.78)] × 100
= 45.7%

% Improvement in Reaction mass efficiency of SCALP relative to OPAL: = [(24.2 – 14.8)/14.8] × 100
= 63.5%

References

1. A. A. N. de Souza, N. S. Silva, A. V. Müller, A. S. Polo, T. J. Brocksom and K. T. de Oliveira, *J. Org. Chem.*, 2018, **83**, 15077-15086.
2. Z. Zhu, C. M. Glinkerman and D. L. Boger, *J. Am. Chem. Soc.*, 2020, **142**, 20778-20787.
3. R. J. Spears, R. L. Brabham, D. Budhadev, T. Keenan, S. McKenna, J. Walton, J. A. Brannigan, A. M. Brzozowski, A. J. Wilkinson, M. Plevin and M. A. Fascione, *Chem. Sci.*, 2018, **9**, 5585-5593.
4. R. Bollhagen, M. Schmiedberger, K. Barlos and E. Grell, *Chem. Soc., Chem.*, 1994, DOI: 10.1039/C39940002559, 2559-2560.
5. H. Egami, S. Kamisuki, K. Dodo, M. Asanuma, Y. Hamashima and M. Sodeoka, *Org. Biomol. Chem.*, 2011, **9**, 7667-7670.
6. K. L. S. Chatalic, J. Veldhoven-Zweistra, M. Bolkestein, S. Hoeben, G. A. Koning, O. C. Boerman, M. de Jong and W. M. van Weerden, *J. Nucl. Med.*, 2015, **56**, 1094.
7. T. Keenan, R. J. Spears, S. Akkad, C. S. Mahon, N. E. Hatton, J. Walton, A. Noble, N. D. Yates, C. G. Baumann, A. Parkin, N. Signoret and M. A. Fascione, *ACS Chem. Biol.*, 2021, **16**, 2387-2400.
8. K. Barker, S. K. Rastogi, J. Dominguez, T. Cantu, W. Brittain, J. Irvin and T. Betancourt, *J. Biomater. Sci. Polym. Ed.*, 2016, **27**, 22-39.
9. R. A. Abramovitch, E. P. Kyba, and E. F. V. Scriven, *J. Org. Chem.*, 1971, **36**, 3796-3803.