Cytosolic Delivery of Proteins by Bioreversible Esterification

Kalie A. Mix,[†,‡] Jo E. Lomax,[§] and Ronald T. Raines*[,†,‡,#]

[†]Department of Biochemistry, [§]Program in Cellular and Molecular Biology, and [#]Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, United States; [‡]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

*rtraines@mit.edu

Content	Page

General Experimental

Materials. Silica gel (40 μm, 230–400 mesh) was from SiliCycle. Reagent chemicals were obtained from commercial sources and used without further purification. Dichloromethane and tetrahydrofuran were dried by passage over a column of alumina. The progress of reactions was monitored by thin-layer chromatography using plates of 250-μm silica 60-F_{254} from EMD Millipore.

Conditions. All procedures were performed in air at ambient temperature (~22 °C) and pressure (1.0 atm) unless indicated otherwise.

Solvent removal. The phrase "concentrated under reduced pressure" refers to the removal of solvents and other volatile materials using a rotary evaporator at water aspirator pressure (<20 torr) while maintaining a water bath below 40 °C. Residual solvent was removed from samples at high vacuum (<0.1 torr).

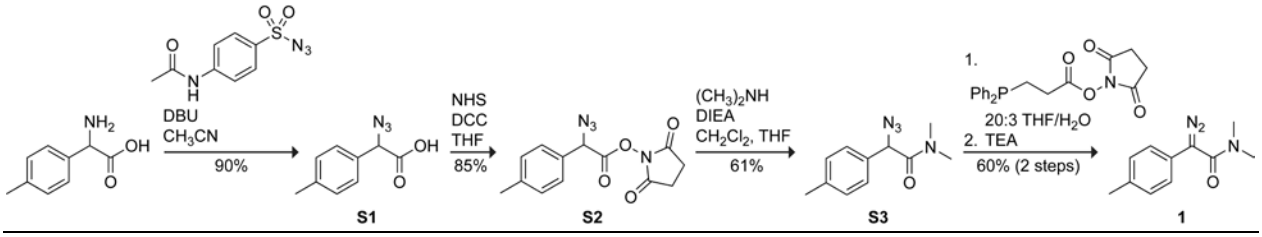
NMR spectroscopy. ^{1}H and ^{13}C NMR spectra for all compounds were acquired with Bruker spectrometers in the National Magnetic Resonance Facility at Madison operating at 500 MHz. Chemical shift data are reported in units of δ (ppm) relative to an internal standard (residual solvent or TMS).

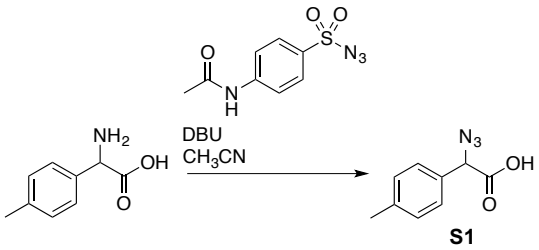
Mass spectrometry. Mass spectra were acquired at the Mass Spectrometry Facility in the Department of Chemistry at the University of Wisconsin–Madison. Electrospray ionization (ESI) mass spectra for small-molecule characterization were acquired with an LCT instrument from Waters. Atmospheric solids analysis probe (ASAP) mass spectra for small-molecule characterization were acquired with a Thermo Q Exactive Plus instrument from Thermo Fisher Scientific. Matrix-assisted laser desorption-ionization–time-of-flight (MALDI–TOF) mass spectra for protein characterization were acquired with a microflex LRF instrument form Bruker.

Melting point. The melting point of diazo compound **1** was determined with an Optimelt automated melting point system from Stanford Research Systems.

Chemical Synthesis

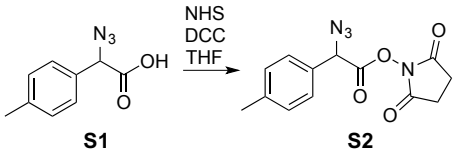
Scheme S1. Route for the synthesis of diazo compound **1.** Overall yield: 28% (unoptimized).





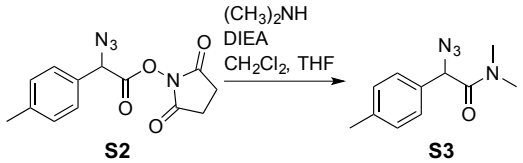
2-Amino-2-(4-methylphenyl)acetic acid (5.0 g, 30.3 mmol) from Matrix Scientific was dissolved in acetonitrile (50 mL). 1,8-Diazabicyclo(5.4.0)undec-7-ene (DBU; 13.8 g, 90.9 mmol) and *p*-

aminobenzenesulfonyl azide (8.0 g, 33.3 mmol) were added, and the resulting solution was stirred overnight. The solution was concentrated under reduced pressure. The resulting residue was dissolved in EtOAc (50 mL) and washed twice with 1 M HCl(aq). The organic layer was dried over anhydrous Na_2SO_4(s) and concentrated under reduced pressure to afford α-azido acid S1 (5.2 g, 90%) as a white solid. **^{1}H NMR** (500 MHz, $CDCl_3$, δ): 7.31 (d, 2H, J = 8.1 Hz), 7.24 (d, 2H, J = 7.9 Hz), 5.01 (s, 1H), 2.38 (s, 3H). **^{13}C NMR** (125 MHz, $CDCl_3$, δ): 173.0, 139.7, 130.2, 129.9, 127.6, 64.8, 21.3. **HRMS–ESI** (m/z): $[M - H]^-$ calcd for $C_9H_9N_3O_2$, 190.0622; found 190.0622.



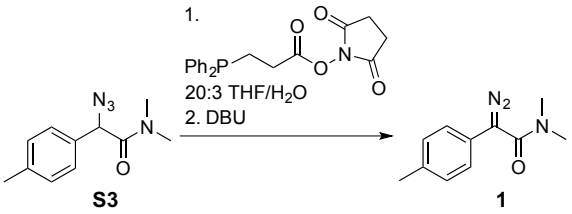
S1 **S2**

α-Azido acid **S1** (5.2 g, 27.4 mmol) was dissolved in THF (100 mL). *N*-Hydroxysuccinimide (3.1 g, 27.4 mmol) and DCC (6.2 g, 30.1 mmol) were added, and the resulting solution was stirred overnight. The slurry was removed by filtration, and the solution was concentrated under reduced pressure. The residue was dissolved in EtOAc (50 mL) and washed with saturated aqueous $NaHCO_3$. The organic layer was dried over anhydrous Na_2SO_4(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes to afford α-azido *N*-hydroxysuccinimide ester **S2** (6.7 g, 85%) as a white solid. **^{1}H NMR** (500 MHz, $CDCl_3$, δ): 7.36 (d, 2H, J = 8.1 Hz), 7.28 (d, 2H, J = 8.0 Hz), 5.25 (s, 1H), 2.83 (s, 4H), 2.38 (s, 3H). **^{13}C NMR** (125 MHz, $CDCl_3$, δ): 168.4, 165.2, 140.2, 130.1, 128.8, 127.9, 63.2, 25.6, 21.3. **HRMS–ESI** (m/z): $[M + NH_4]^+$ calcd for $C_{13}H_{12}N_4O_4$, 306.1197; found, 306.1191.

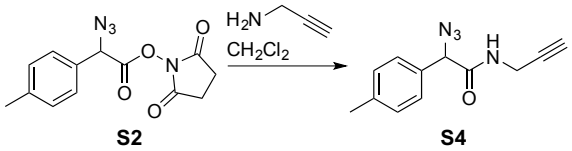


S2 **S3**

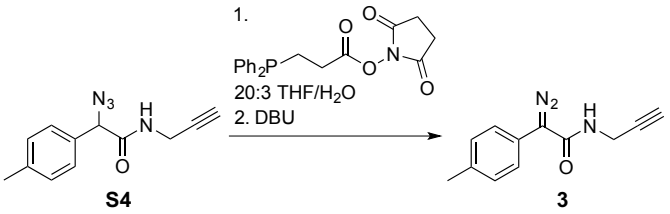
α-Azido *N*-hydroxysuccinimide ester **S2** (417 mg; 1.4 mmol) was dissolved in CH_2Cl_2 (15 mL). A solution of dimethylamine (0.8 mL; 2.0 M in THF) and DIEA (361 mg; 2.8 mmol) were added, and the resulting solution was stirred overnight. The solution was concentrated under reduced pressure. The residue was dissolved in EtOAc and washed twice with 1 M HCl(aq) and saturated aqueous $NaHCO_3$ (2 × 10 mL). The organic layer was dried over anhydrous Na_2SO_4(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford α-azido dimethylamide **S3** (192 mg; 61%) as a white solid. **^{1}H NMR** (500 MHz, $CDCl_3$, δ): 7.28 (d, 2H, J = 8.2 Hz), 7.23 (d, 2H, J = 8.0 Hz), 4.91 (s, 1H), 3.01 (s, 3H), 2.81 (s, 3H), 2.37 (s, 3H). **^{13}C NMR** (125 MHz, $CDCl_3$, δ): 169.0, 139.2, 130.7, 130.0, 127.9, 63.5, 36.9, 36.1, 21.2. **HRMS–ESI** (m/z): $[M + H]^+$ calcd for $C_{11}H_{14}N_4O$, 219.1240; found, 219.1235.



α-Azido dimethylamide **S3** (100 mg, 0.46 mmol) was dissolved in 20:3 THF/H_2O (4.6 mL). *N*-Succinimidyl 3-(diphenylphosphino)propionate (179 mg, 0.50 mmol) was added, and the resulting solution was stirred for 3 h under N_2(g). 1,8-Diazabicycloundec-7-ene (DBU; 140 mg, 0.92 mmol) was added, and the solution was stirred for 1 h. The solution was diluted with brine (10 mL) and extracted with CH_2Cl_2 (2 × 10 mL). The organic layer was dried over anhydrous Na_2SO_4(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes to afford α-diazo dimethylamide **1** (56 mg, 60%) as an orange solid with mp 57.2–61.6 °C. **^{1}H NMR** (500 MHz, $CDCl_3$, δ): 7.19 (d, 2H, J = 8.1 Hz), 7.11 (d, 2H, J = 8.2 Hz), 2.95 (s, 6H), 2.34 (s, 3H). **^{13}C NMR** (125 MHz, $CDCl_3$, δ): 166.1, 135.7, 129.9, 124.7, 124.4, 37.7, 21.0. **HRMS–ESI** (*m*/*z*): $[M + H]^+$ calcd for $C_{11}H_{13}N_3O$, 204.1131; found, 204.1128.

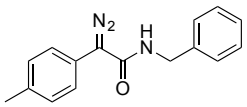


α-Azido *N*-hydroxysuccinimidyl ester **S2** (1.1 g, 3.7 mmol) was dissolved in CH_2Cl_2 (20 mL). Propargylamine (0.2 g, 4.0 mmol) was added, and the resulting solution was stirred overnight. The solution was concentrated under reduced pressure. The resulting residue was dissolved in EtOAc and washed with 1.0 M HCl (2 × 10 mL), followed by saturated aqueous $NaHCO_3$ (2 × 10 mL). The organic layer was dried with anhydrous Na_2SO_4(s) and concentrated under reduced pressure to afford α-azido propargylamide **S4** (0.6 g, 75%) as an off-white solid. **^{1}H NMR** (500 MHz, $CDCl_3$, δ): 7.25 (d, J = 6.3 Hz, 2H), 7.21 (d, J = 8.1 Hz, 2H), 6.64 (s, 1H), 5.03 (s, 1H), 4.08 (dd, J = 2.5, 5.25 Hz, 2H), 2.36 (s, 3H), 2.26 (t, J = 2.4 Hz, 1H); **^{13}C NMR** (125 MHz, $CDCl_3$, δ): 167.8, 139.3, 131.6, 129.8, 127.7, 78.8, 72.1, 67.0, 29.4, 21.2; **HRMS–ESI** (*m*/*z*): $[M–N_2 + H]^+$ calcd for $C_{12}H_{12}N_4O$, 229.1084; found, 229.1085.



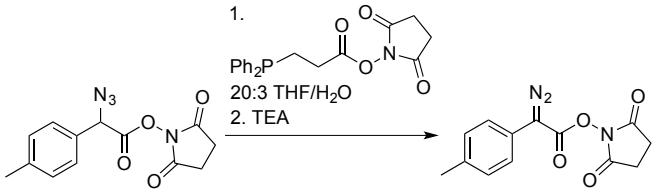
α-Azido propargyl amide **S4** (0.6 g, 2.7 mmol) was dissolved in 16 mL of 20:3 THF/H_2O. *N*-Succinimidyl 3-(diphenylphosphino)propionate (1.1 g, 3.1 mmol) was added under N_2(g), and the resulting solution was stirred for 5 h. 1,8-Diazabicycloundec-7-ene (DBU; 0.8 g, 5.5 mmol) was added, and the solution was stirred for 1 h. The solution was diluted with brine (20 mL) and extracted with CH_2Cl_2 (2 × 10 mL). The organic layer was dried with anhydrous Na_2SO_4(s) and

concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes to afford α-diazo propargylamide **3** (0.176 g, 30%) as a red solid. **^{1}H NMR** (500 MHz, $CDCl_3$, δ): 7.28–7.24 (m, 4H), 5.52 (s, 1H), 4.15–4.14 (dd, J = 2.5, 5.4 Hz, 2H), 2.38 (s, 3H), 2.23 (s, 1H); **^{13}C NMR** (125 MHz, $CDCl_3$, δ): 164.9, 138.3, 130.5, 128.0, 122.6, 79.6, 71.6, 64.0, 29.7, 21.2; **HRMS–ESI** (m/z): $[M + H]^+$ calcd for $C_{12}H_{11}N_3O$, 214.0975; found, 214.0975.



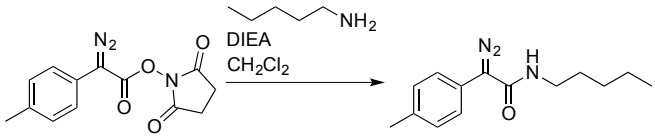
4

α-Diazo benzylamide **4** was prepared as described previously[1]. Yields and spectral data replicated those reported previously.



S2 **S5**

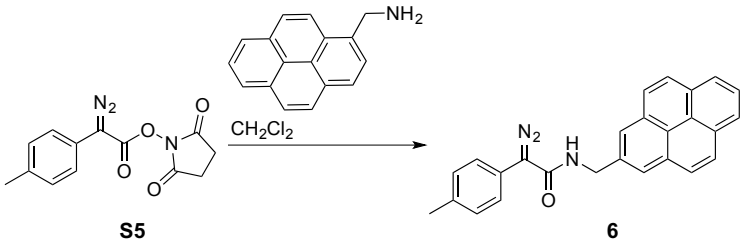
α-Azido *N*-hydroxysuccinimidyl ester **S2** (3.4 g, 11.6 mmol) was dissolved in 50 mL of 20:3 THF/H_2O. *N*-Succinimidyl 3-(diphenylphosphino)propionate (4.5 g, 12.8 mmol) was added under N_2(g), and the resulting solution was stirred for 5 h. Triethylamine (TEA; 2.3 g, 23.2 mmol) was added, and the solution was stirred for 1 h. The solution was diluted with brine (20 mL) and extracted with CH_2Cl_2 (2 × 10 mL). The organic layer was dried with anhydrous Na_2SO_4(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes to afford α-diazo *N*-hydroxysuccinimidyl ester **S5** (0.31 g, 10%) as an orange solid. **^{1}H NMR** (500 MHz, $CDCl_3$, δ): 7.32 (d, J = 8.3 Hz, 2H), 7.22 (d, J = 8.1 Hz, 2H), 2.88 (s, 4H), 2.35 (s, 3H); **^{13}C NMR** (125 MHz, $CDCl_3$, δ): 169.4, 160.5, 137.1, 129.9, 124.6, 119.8, 25.6, 21.1; **HRMS–ASAP** (m/z): $[M–N_2 + H]^+$ calcd for $C_{13}H_{11}N_3O_4$, 246.0761; found, 246.0764.



S5 **5**

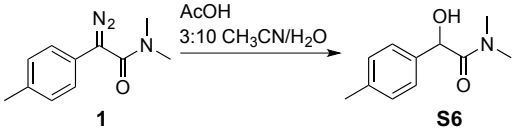
α-Diazo *N*-hydroxysuccinimidyl ester **S2** (100 mg, 0.37 mmol) was dissolved in CH_2Cl_2 (37 mL). *n*-Pentylamine (35.4 mg, 0.41 mmol) and DIEA (143 mg, 1.1 mmol) were added, and the resulting solution was stirred overnight. The solution was concentrated under reduced pressure, and the residue was dissolved in EtOAc. The residue was purified by chromatography on silica gel, eluting with 1:4 EtOAc/hexanes to afford α-diazo pentylamide **5** (65 mg, 72%) as a red solid. **^{1}H NMR** (500 MHz, $CDCl_3$, δ): 7.26–7.23 (m, 4H), 5.37 (s, 1H), 3.36–3.32 (q, J = 7.0 Hz, 2H), 2.38 (s, 3H), 1.53–1.49 (m, 2H), 1.33–1.28 (m, 4H), 0.90–0.88 (t, 3H, J = 6.9 Hz); **^{13}C**

NMR (125 MHz, $CDCl_3$, δ): 164.9, 137.9, 130.4, 127.8, 123.3, 63.8, 40.3, 29.6, 29.0, 22.3, 21.2, 14.0; **HRMS–ESI** (*m*/*z*): $[M–N_2 + H]^+$ calcd for $C_{14}H_{19}N_3O$, 218.1539; found, 218.1541.



S5 **6**

α-Diazo *N*-hydroxysuccinimidyl ester **S5** (55 mg, 0.20 mmol) was dissolved in CH_2Cl_2 (20 mL). 1-Pyrene methylamine (47.0 mg, 0.2 mmol) was added, and the solution was stirred overnight. The solution was concentrated under reduced pressure, and the residue was dissolved in EtOAc. The residue was purified by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford α-diazo pyrenylamide **6** (18 mg, 23%) as an orange solid. **^{1}H NMR** (500 MHz, $CDCl_3$, δ): 8.35 (d, 1H, J = 9.2 Hz), 8.28–8.22 (m, 3H), 8.17 (d, 1H, J = 7.8 Hz), 8.13–8.06 (m, 3H), 8.00 (d, 1H, J = 7.8 Hz), 7.21 (d, 2H, J = 8.2 Hz), 7.15 (d, 2H, J = 8.1 Hz), 5.78 (t, 1H, J = 4.8 Hz), 5.31 (d, 2H, J = 5.4 Hz), 2.29 (s, 3H); **^{13}C NMR** (125 MHz, $CDCl_3$, δ): 164.7, 137.9, 131.2, 131.17, 130.9, 130.7, 130.4, 129.0, 128.2, 127.7, 127.5, 127.2, 127.1, 126.1, 126.0, 125.4, 125.3, 125.0, 124.8, 124.6, 122.8, 64.0, 42.6, 21.0; **HRMS–ASAP** (*m*/*z*): $[M + H]^+$ calcd for $C_{26}H_{19}N_3O$, 390.1601; found, 390.1596.



1 **S6**

α-Diazo dimethylamide **1** (80 mg, 0.39 mmol) was dissolved in 10 mL of 3:10 CH_3CN/H_2O. Acetic acid (10 μL, 0.17 mmol) was added, and the resulting solution was stirred for 1 h. The solution was concentrated under reduced pressure, and the residue was dissolved in EtOAc. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes. The eluate was purified further by recrystallization from DCM and hexanes to afford α-hydroxy dimethylamide **S6** (6 mg, 14%) as a white solid. **^{1}H NMR** (500 MHz, $CDCl_3$, δ): 7.21 (d, 2H, J = 8.2 Hz), 7.17 (d, 2H, J = 8.1 Hz), 5.17 (d, 1H, J = 6.4 Hz), 4.71 (d, 1H, J = 6.4 Hz), 3.03 (s, 3H), 2.78 (s, 3H), 2.34 (s, 3H); **^{13}C NMR** (125 MHz, $CDCl_3$, δ): 172.5, 138.3, 136.2, 129.7, 127.4, 71.3, 36.4, 36.3, 21.2; **HRMS–ASAP** (*m*/*z*): $[M + H]^+$ calcd for $C_{11}H_{15}NO_2$, 194.1176; found, 194.1176.

Protein Preparation

Preparation of Green Fluorescent Protein (GFP)

The "superfolder" variant of GFP was prepared as described previously[2]. The protein was dialyzed into 10 mM Bis-Tris buffer, pH 6.5, prior to esterification.

Preparation of Super-Charged GFP

A gene encoding enhanced GFP was amplified from a mammalian expression vector (Promega) and inserted into a novel vector derived from a pET vector (Novagen). The following

substitutions were introduced using site-directed mutagenesis to recapitulate those in the "superfolder"[3] and "cell-penetrating"[4] variants: F64L, S65T, F99S, M153T, V163A, S30R, Y145F, I171V, A106V, Y39I, N105K, I128T, K166T, I167V, S205T, L221H, F223Y, T225N, E17R, D19R, D21R, V111R, and E124R. The expression vector was transformed into BL21(DE3) electrocompetent *E. coli* cells (New England Biolabs) and plated on LB agar containing ampicillin (amp; 200 µg/mL). The resulting plates were incubated overnight at 37 °C. A single colony was added to 50 mL of LB–amp (which contained 200 µg/mL amp) and incubated overnight at 37 °C in a shaking incubator. On the following day, 10 mL of starter culture was added to each of 4 L of Terrific Broth–amp medium (which contained 200 µg/mL amp). Cultures were grown at 37 °C in a shaking incubator until the cell density reached an OD_{600} = 0.6–0.8. Cultures were incubated for 20 min at 20 °C, and then induced by the addition of IPTG (to 1.0 mM). Cells were grown overnight at 20 °C in a shaking incubator.

Cells were harvested by centrifugation at 5,000 rpm for 20 min at 4 °C. Cell pellets were collected and resuspended in 15 mL of lysis buffer per 1 L liquid growth. (Lysis buffer was 50 mM Tris–HCl buffer, pH 7.0, containing 100 mM NaCl, 30 mM imidazole, 1% v/v Triton X-100, and 20% w/v sucrose.) The resuspended cells were stored frozen overnight at –20 °C. On the following day, cells were thawed and lysed by mechanical disruption using a cell disruptor (Constant Systems) at 22 kPsi. The lysate was cleared by centrifugation at 11,000 rpm for 1 h at 4 °C. The supernatant was collected and filtered through a 0.2-µm PES filter (GE Healthcare).

Super-charged GFP in the filtered cell lysate was purified by chromatography with a HisTrap nickel column (GE Healthcare). The binding (wash) buffer was 20 mM sodium phosphate buffer, pH 7.4, containing NaCl (0.50 M) and imidazole (30 mM). The elution buffer was 20 mM sodium phosphate buffer, pH 7.4, containing NaCl (0.50 M) and a linear gradient of imidazole (30 mM–0.50 M). Eluted fractions were pooled and dialyzed against 4 L of 20 mM Tris–HCl buffer, pH 7.4, containing EDTA (1.0 mM).

Dialyzed material was purified further by chromatography with a HiTrap SP HP cation-exchange column (GE Healthcare). The binding (wash) buffer was 20 mM Tris–HCl buffer, pH 7.4, containing EDTA (1.0 mM). The elution buffer was 20 mM Tris–HCl buffer, pH 7.4, containing EDTA (1.0 mM) and NaCl (1.0 M). Upon elution, colored fractions were pooled and dialyzed against PBS overnight and concentrated as needed. *m*/*z*, 29,547; expected: 29,539 (Supplementary Figure 1a).

Preparation of GFP Containing a Nuclear Localization Sequence (nlsGFP)

A vector containing the gene that encodes "superfolder" GFP gene was reported previously.[2] This vector was modified to install a nuclear localization sequence[3] at the N-terminus of the encoded protein by using the primers:

5′-AAGAAACGCAAGGTACTGGTCCCGGTGGCGACAGTGAGCAAGGGCGAGGAGC-3′

5′-CGGGACCAGTACCTTGCGTTTCTTCTTCGGCATATCTATATCTCCTTCTTAAGGTAAA-3′

In addition, the His_6 tag was moved from the N terminus to the C terminus, and the TEV protease recognition sequence was removed. The ensuing nlsGFP has the amino-acid sequence:

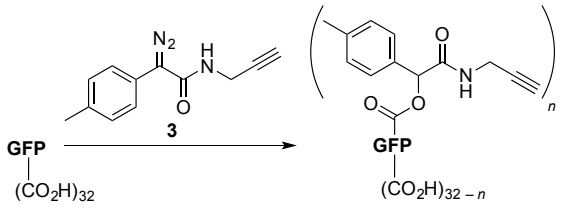
```
 1                                                  50
MPKKKRKVLVPVATVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDA
 TIGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSA
 MPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNIL
 GHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTP
 IGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMVLHEYVNAAGITLGMDELY
 KAVDKLAAALEHHHHHH
```

The expression vector was transformed into BL21(DE3) electrocompetent *E. coli* cells (New England Biolabs). The resulting cells were plated on LB agar containing amp (200 µg/mL), and the plates were incubated overnight at 37 °C. A single colony was added to 50 mL of LB–amp (which contained 200 µg/mL amp), and the resulting culture was incubated overnight at 37 °C in a shaking incubator. On the following day, 10 mL of starter culture was added to each of 4 L of Terrific Broth–amp medium (which contained 200 µg/mL amp). Cultures were grown at 37 °C in a shaking incubator until the cell density reached an OD_{600} of 0.6–0.8. Cultures were incubated for 20 min at 20 °C, and then induced by the addition of IPTG (to 1.0 mM). Cells were grown overnight at 20 °C in a shaking incubator.

Cells were harvested by centrifugation at 5,000 rpm for 20 min at 4 °C. Cell pellets were collected and resuspended in 15 mL of lysis buffer per 1 L of liquid growth. (Lysis buffer was 50 mM Tris–HCl buffer, pH 7.0, containing 100 mM NaCl, 30 mM imidazole, 1% v/v Triton X-100, and 20% w/v sucrose.) The resuspended cells were stored frozen overnight at –20 °C. On the following day, cells were thawed and lysed by mechanical disruption using a cell disruptor (Constant Systems) at 22 kPsi. The lysate was cleared by centrifugation for 1 h at 11,000 rpm at 4 °C. The supernatant was collected and filtered through a 0.2-µm PES filter (GE Healthcare), and nlsGFP was purified by chromatography using a HisTrap nickel column, dialysis, and anion-exchange chromatography, as described above *m*/*z*, 29,945; expected: 29,943 (Supplementary Figure 1b).

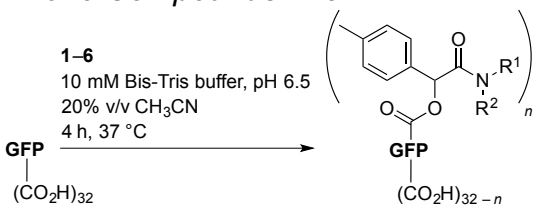
Protein Esterification

Optimization of Solvent Conditions for GFP Esterification



A solution of diazo compound **3** (1.2 mg, 0.054 µmol) in acetonitrile was added to a solution of GFP (0.0017 µmol) in 10 mM Bis-Tris buffer (pH 6.0, 6.5, or 7.0). Additional Bis-Tris buffer was added so that the final composition of the solution ranged from 5–40% v/v acetonitrile. The esterification reaction was incubated for 4 h at 37 °C. Precipitated protein was removed by filtration through a 0.2-µm PES syringe filter (GE Healthcare), and the number of esters per protein was assigned from the mass of the peak with the highest relative intensity in the MALDI–TOF mass spectrum. The mildest condition that enabled a high level of esterification was 10 mM Bis-Tris buffer, pH 6.5, containing acetonitrile (20% v/v) (Supplementary Figure 2).

*Esterification of GFP with Diazo Compounds **1–6***



A solution of diazo compound **1**–**6** (34.1 µmol) in acetonitrile (400 µL) was added to a solution of GFP (0.341 µmol) in 1600 µL of 10 mM Bis-Tris buffer, pH 6.5, and incubated for 4 h at 37 °C. Precipitated protein was removed by filtration through a 0.2-µm PES syringe filter (GE Healthcare). The number of esters per protein was assigned from the mass of the peak with the highest relative intensity in the MALDI–TOF mass spectrum (Supplementary Figure 3). Protein in each mixture was then purified and exchanged into PBS buffer using PD10 desalting columns (GE Healthcare). Protein was concentrated as needed by centrifugation, and the protein concentration was determined with a Bradford assay (Thermo Fisher Scientific).

*Identification of GFP Carboxyl Groups Esterifed with Diazo Compound **1***

For each digest, a 10-µg aliquot of protein solution was diluted with water to 100 µL, and 1 µL of 10% v/v aqueous formic acid was added. Immobilized pepsin (Thermo Fisher Scientific product 20343) was washed with 0.1% v/v formic acid according to the manufacturer's instructions and resuspended as a 50% slurry. A 50-µL aliquot of the immobilized pepsin slurry was added to each sample. Samples were placed on a shaking incubator at 37 °C at 200 rpm and incubated for 2, 5, 10, or 20 min. Upon removal from the incubator, samples were subjected to centrifugation, and the supernatant was removed to quench the digestion.

Data were acquired on an Orbitrap Elite mass spectrometer equipped with a Thermo EasySpray column (15 cm × 75 µm, packed with 3-µm PepMap C18 resin) and eluted over a 45-min gradient using solvents of 0.1% v/v formic acid in water (A) and 0.1% v/v formic acid in acetonitrile (B). A top-20 method was used to acquire MS/MS spectra on the 20 highest abundance precursors in each MS scan with dynamic exclusion of precursors that had been selected already within the preceding 30 s for MS/MS analysis.

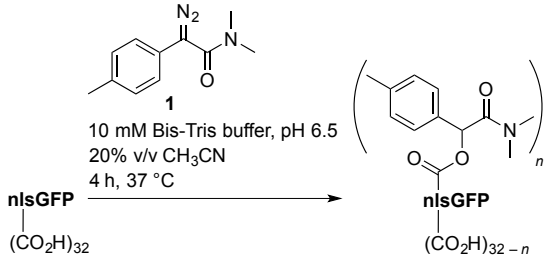
Data were searched against an e.coli database to which was added the sfGFP sequence, and the +175 modification was allowed as a variable modification. Pepsin was used as the enzyme specificity with up to 4 missed cleavages per peptide. Precursor tolerance was set at 15 ppm, and MS/MS fragment ion tolerance was set to 0.5 Da (MS/MS data collected in the linear ion trap portion of the Orbitrap Elite).

Residues identified as being esterified are indicated in red:

```
 1                                                  50
MHHHHHHSSGVDLGTENLYFQGMVSKGEELFTGVVPILVELDGDVNGHKFS
 VRGEGEGDATIGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHM
 KQHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGT
 DFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQL
 ADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMVLHEYVNAAG
 ITLGMDELYK
```

Note: This GFP variant contains additional residues at its N and C termini relative to the protein used to calculate the electrostatic surface in Table S1; thus, 3 esterified carboxyl groups are not listed in Table S1.

*Esterification of nlsGFP with Diazo Compound **1***



A solution of diazo compound **1** (3.5 mg, 17.1 μmol) in acetonitrile (400 μL) was added to a solution of nlsGFP (0.341 μmol) in 1.6 mL of 10 mM Bis-Tris buffer, pH 6.5, and the resulting solution was incubated for 4 h at 37 °C. Precipitated protein was removed by filtration through a 0.2-μm PES syringe filter (GE Healthcare). Protein was purified and exchanged into PBS buffer using a PD10 desalting column (GE Healthcare). The number of esters per protein was assigned from the peak with the highest relative intensity in the MALDI–TOF mass spectrum (Supplementary Figure 4). Protein was concentrated as needed, and the protein concentration was determined with a Bradford assay (Thermo Fisher Scientific).

Mammalian Cell Culture

Chinese hamster ovary- (CHO-) K1 cells were from the American Tissue Culture Collection and cultured according to recommended protocols. Cells were grown in F12K nutrient medium supplemented with fetal bovine serum (10% v/v), penicillin (100 units/mL) and streptomycin (100 μg/mL). Cells were grown in T75 sterile culture flasks in a cell culture incubator at 37 °C under CO_2 (5% v/v). Cells were counted to determine seeding density using a hemacytometer.

Flow Cytometry

Cells were seeded at a density of 50,000 cell/well in a sterile 8-well dish (Ibidi) 24 h prior to treatment. Cells were incubated with either unmodified GFP or GFP esterified with compounds **1–6** (15 μM) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 μg/mL) for 2 h at 37 °C. Cells were rinsed twice with DPBS, and released from the plate with 250 μL of 0.25% v/v trypsin–EDTA mix. Tryspin was quenched by the addition of 500 μL of medium, and cells were then subjected to centrifugation for 5 min at 130*g*. Cells were resuspended in 300 μL of DPBS supplemented with fetal bovine serum (10% v/v). 7AAD stain (10 μL of a 1.0 mg/mL solution) was added to each sample, and cells were kept on ice until the time of analysis. The fluorescence intensity of at least 10,000 events was measured by flow cytometry with an Accuri C6 flow cytometer (BD Biosciences). The median fluorescence intensity of live, single cells is reported.

Time-Dependence of GFP–1 Internalization

Cells were seeded at a density of 100,000 cells/well in a sterile 12-well dish (CellStar) 24 h prior to treatment. The cells were then incubated with GFP–**1** (4 μM) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 μg/mL) for 30, 120, or 240 min at 37 °C.

Cells were rinsed with DPBS, and released from the plate with 250 μL of 0.05% trypsin–EDTA. Trypsin was quenched by the addition of 250 μL of medium. Propidium iodide was added to each sample (final concentration: 10 μg/mL), and cells were kept on ice until the time of analysis. The fluorescence intensity of at least 10,000 events was measured by flow cytometry with a FACS Canto II HTS flow cytometer (BD Biosciences). The median fluorescence intensity of live, single cells is reported (Figure S3).

Confocal Microscopy

*Internalization of GFP–**1***

Cells were seeded at a density of 50,000 cell/well in a sterile 8-well dish (Ibidi) 24 h prior to treatment. Cells were incubated with either unmodified GFP or GFP–**1** (15 μM) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 μg/mL) for 2 h at 37 °C. Cells were rinsed twice with DPBS, and nuclei were stained by incubation with Hoechst 33342 dye (2 μg/mL) for 5 min at 37 °C. Cell membranes were stained by incubation with wheat germ agglutinin (WGA)–Alexa Fluor® 647 dye (5 μg/mL) for 15 min on ice. Cells were then washed twice and kept in medium on ice until the time of analysis. Live cells were examined using a Nikon A1R+ scanning confocal microscope. The results are shown in Figure 2. Image acquisition and processing settings were maintained between all samples.

*Internalization of GFP, Super-Charged GFP, and GFP–**1** at 37 °C and 4 °C*

Cells were seeded at a density of 50,000 cell/well in a sterile 8-well dish (Ibidi) 24 h prior to treatment. Cells were incubated with unmodified GFP, super-charged GFP, or GFP–**1** (15 μM) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 μg/mL) for 2 h at either 37 or 4 °C. Cells were rinsed twice with DPBS, and nuclei were stained by incubation with Hoechst 33342 dye (2 μg/mL) for 5 min at 37 °C. Cell membranes were stained by incubation with WGA–Alexa Fluor® 647 dye (5 μg/mL) for 15 min on ice. Cells were then washed twice and kept in medium on ice until the time of analysis. Live cells were examined using a Nikon A1R+ scanning confocal microscope. Image acquisition and processing settings were maintained between all samples. The results are shown in Figure 2.

*Internalization of nlsGFP and nlsGFP–**1***

Cells were seeded at a density of 50,000 cell/well in a sterile 8-well dish (Ibidi) 24 h prior to treatment. Cells were incubated with either unmodified GFP or nlsGFP–**1** (15 μM) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 μg/mL) for 2 h at 37 °C. Cells were rinsed twice with DPBS, and nuclei were stained by incubation with Hoechst 33342 dye (2 μg/mL) for 5 min at 37 °C. Cell membranes were stained by incubation with WGA–Alexa Fluor® 647 dye (5 μg/mL) for 15 min on ice. Cells were then washed twice and kept in medium on ice until the time of analysis. Live cells were examined using a Nikon A1R+ scanning confocal microscope. Image acquisition and processing settings were maintained between all samples. The results are shown in Figure 3. Pearson's correlation coefficient (r) was calculated with the PSC colocalization plugin in ImageJ software.

Esterification Reversibility

Unlike GFP, human angiogenin is a small protein (15.3 kDa) that maintains its structure after incubation with a detergent-containing cell lysate and produces a well-resolved peak in a MALDI–TOF spectrum. Moreover, a FLAG-tagged variant of angiogenin[5] binds to an anti-

FLAG antibody with extremely high affinity, thus enabling high recovery of this protein from a cell lysate. Thus, we used angiogenin for a rigorous assessment of the bioreversibility of protein esterification with diazo compound **1**.

Diazo compound **1** (437 μg, 2.2 μmol) in acetonitrile (40 μL) was added to a solution of FLAG–angiogenin (0.043 μmol) in 160 μL of 10 mM Bis-Tris buffer, pH 6.5. The resulting solution was incubated at 37 °C for 4 h, and the number of esters was determined with MALDI–TOF mass spectrometry. Acetonitrile was removed using a Vivaspin filtration column (5,000 MWCO) from GE Life Sciences. The resulting solution was added to a CHO-cell extract (50 μL), which was prepared by using CelLytic M lysis reagent (Sigma–Aldrich product C2978) supplemented with 1× protease inhibitor (Thermo Fisher Scientific product 78430). The solution was incubated at 25 °C overnight. FLAG–angiogenin was reisolated using anti-FLAG magnetic beads from Sigma–Aldrich and analyzed again with MALDI–TOF mass spectrometry (Figure S7).

Cytotoxicity Assay

The cytotoxicity of compound **S4** was measured with a CellTiter96® AQueous One Cell Proliferation (MTS) Assay from Promega according to the manufacturer's instructions. Cells were plated at a density of 50,000 cells/well in a sterile 96-well plate 24 h prior to treatment. Cells were treated with either vehicle (1% v/v DMSO in medium) or compound **S6** (50–500 μM in 1% v/v DMSO in medium) for 2½ h. The medium was replaced, and 20 μL of CellTiter96® AQeous One Solution Reagent was added to each well. Cells were incubated for 1 h, and absorbance at 490 nm was measured with a Tecan Infinite M1000 plate reader. Cell viability is expressed relative to vehicle control (Figure S8).

References

(1) Mix, K. A.; Raines, R. T. *Org. Lett.* **2015**, *17*, 2358–2361.

(2) Andersen, K. A.; Smith, T. P.; Lomax, J. E.; Raines, R. T. *ACS Chem. Biol.* **2016**, *11*, 319–323.

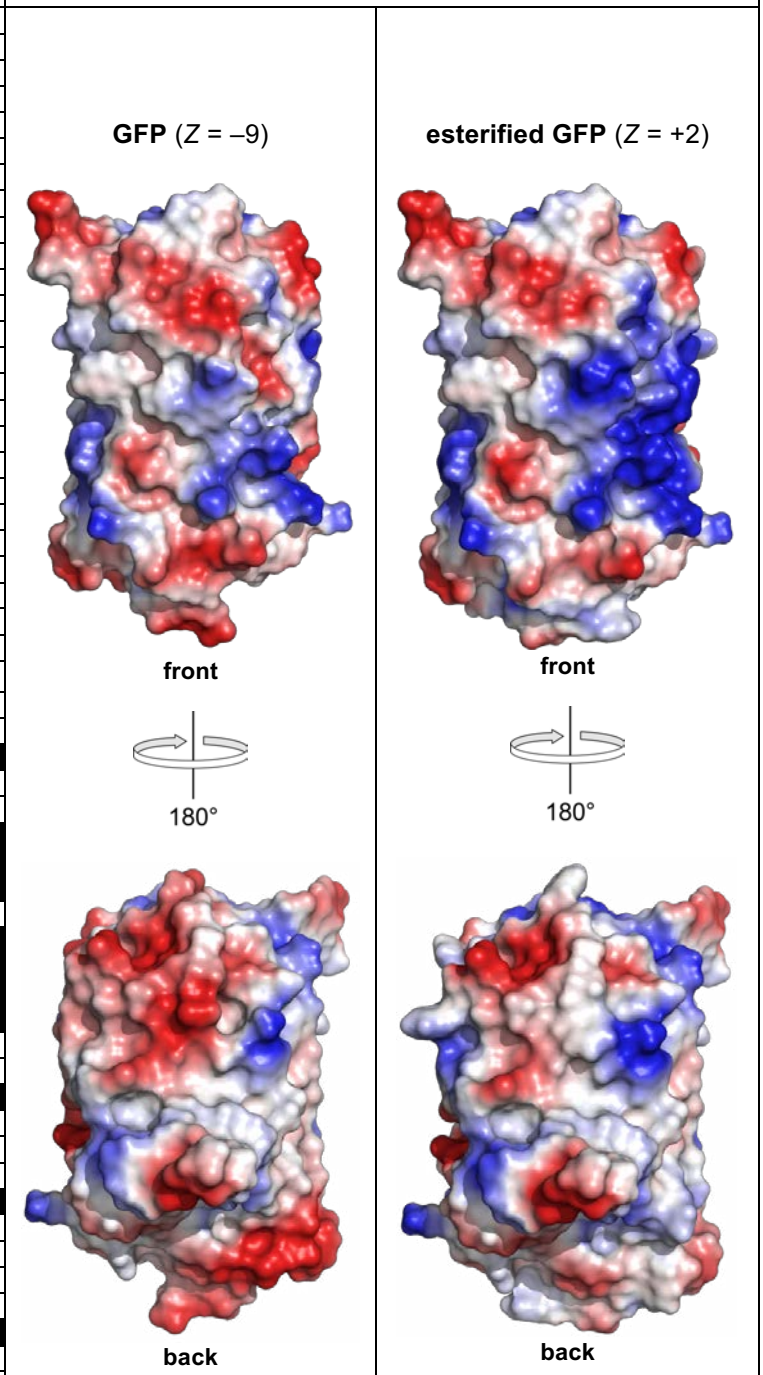
(3) Pédelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2006**, *24*, 79–88.

(4) Fuchs, S. M.; Raines, R. T. *ACS Chem. Biol.* **2007**, *2*, 167–170.

(5) Hoang, T. T.; Raines, R. T. *Nucleic Acids. Res.* **2017**, *45*, 818–831.

Table S1. Notional effect of esterification on the electrostatic surface of GFP.

Residue type	Residue number	Calculated pK_a[a]	Esterified?[b]	Calculated electrostatic surfaces[c]
Lys	140	12.55		
Lys	45	12.53		
Lys	209	12.47		
Lys	79	12.43		
Lys	3	12.41		
Lys	131	12.31		
Lys	162	12.19		
Lys	113	12.07		
Lys	101	12.06		
Lys	41	12.05		
Lys	105	11.96		
Lys	26	11.93		
Lys	126	11.86		
Lys	107	11.79		
Lys	85	11.10		
Glu	222	8.64		
His	139	7.54		
His	77	7.43		
His	217	7.41		
His	25	7.30		
His	81	7.15		
His	231	7.13		
His	221	7.12		
Asp	82	6.44		
His	148	6.44		
His	199	5.65		
His	181	5.57		
Asp	103	4.88		
Glu	124	4.75	■	
Asp	155	4.71		
Glu	115	4.67		
Glu	172	4.61	■	
Glu	6	4.59	■	
Glu	90	4.58	■	
Glu	132	4.50		
Glu	142	4.47	■	
Glu	95	4.42	■	
Glu	34	4.41	■	
Glu	17	4.34	■	
Glu	5	4.27		
Glu	213	4.24		
Asp	190	4.17	■	
Glu	32	4.16		
Asp	129	4.13		
Asp	133	4.09		
Asp	117	4.03	■	
His	169	3.76		
Asp	197	3.76		
Asp	36	3.75		
Asp	180	3.71		
Asp	19	3.69	■	
Asp	21	3.54		
Asp	76	3.53		
Asp	102	3.34		
Asp	216	2.72		
Asp	210	2.62		
Asp	173	2.51		



[a]pK_a values were calculated with the program Depth (Tan, K.P.; Nguyen, T.B.; Patel, S.; Varadarajan, R.; Madhusudhan, M. S. *Nucleic Acids Res.* **2013**, *41*, W314–W321. [b]Esterification with diazo compound **1**, as detected by MS/MS analysis. [c]Electrostatic surfaces were calculated with the program PyMOL from Schrödinger (New York, NY). Glu222 and Asp82 (red highlight in table) are not solvent accessible. To generate the "esterified GFP" image, the other eleven aspartate and glutamate residues (green highlight in table) with the highest calculated pK_a values were replaced with phenylalanine residues. Net charge (*Z*) = Arg + Lys – Asp – Glu.

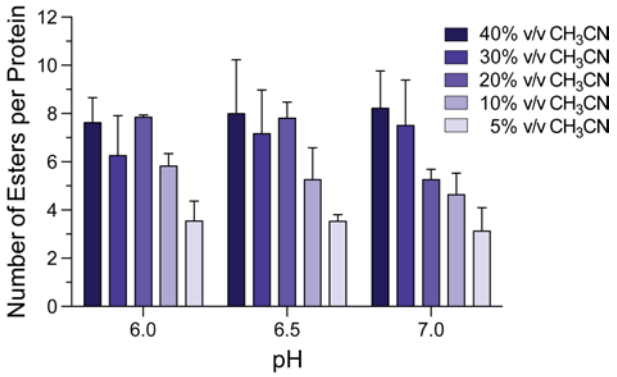


Figure S1. Bar graph showing the extent of esterification of GFP with diazo compound **3** under different solvent conditions.

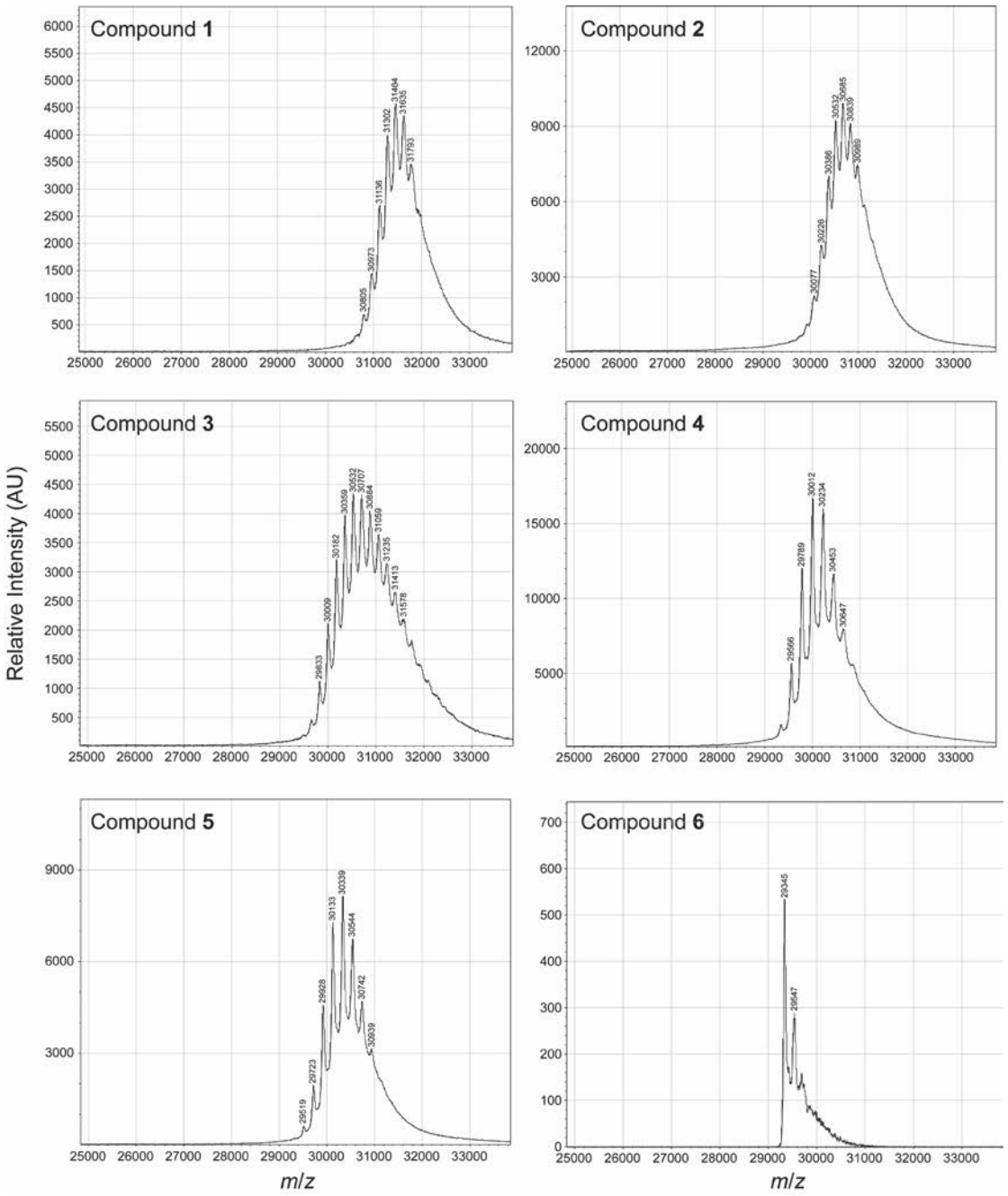


Figure S2. Representative MALDI–TOF spectra of GFP esterified with diazo compounds **1**–**6** (100 equiv, 3 equiv per carboxyl group) in 10 mM Bis-Tris buffer, pH 6.5, containing CH_3CN (20% v/v). Expected *m*/*z*: 29,343 + 175 per ester group.

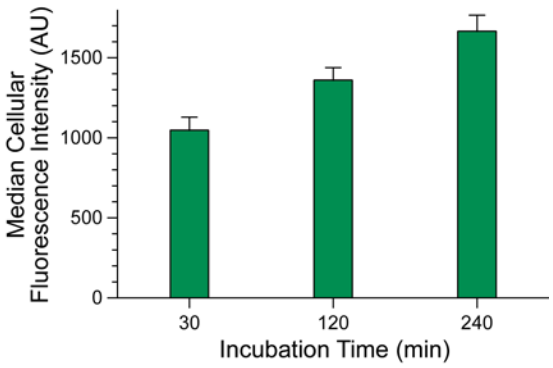
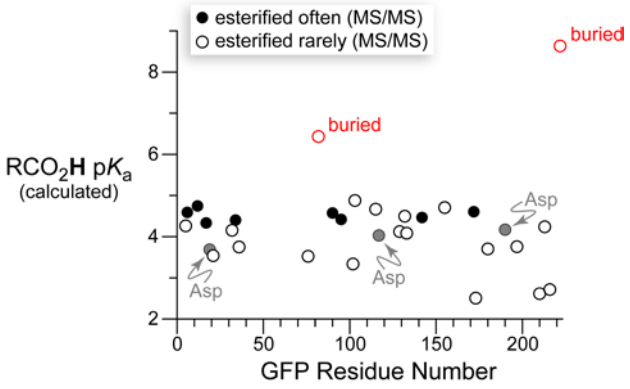


Figure S3. Time-course for the cellular internalization of GFP–**1**. CHO-K1 cells were incubated with GFP–**1** (4 μM) at 37 °C, and internalization was quantified by flow cytometry after 30, 120, and 240 min.



Peptide	Peptidic Residues	Modified Residue sfGFP Number	Modified Residue 2b3p Number
MHHHHHHSSGVDLGTENL	1–18	Glu16	NA
YFQGMVSKGEEL	14–20	Glu29	Glu6
VELDGDVNGHKFSVRGEGEGDATIGKLTLKF	39–69	Glu40	Glu17
VELDGDVNGHKFSVRGEGEGDATIGKLTLKF	39–69	Asp42	Asp19
VELDGDVNGHKFSVRGEGEGDATIGKL	39–65	Glu57	Glu34
KSAMPEGYVQERTISF	108–123	Glu113	Glu90
KSAMPEGYVQERTISF	108–123	Glu118	Glu95
KDDGKYKTRAVVKFEGDTLVNRIEL	124–148	Asp140	Asp117
KDDGKYKTRAVVKFEGDTLVNRIEL	124–148	Glu147	Glu124
KGTDFKEDGNILGHKLEYNF	149–168	Glu165	Glu142
TVRHNVEDGSVQL	189–201	Glu195	Glu172
ADHYQQNTPIGDGPVL	202–217	Asp213	Asp190
GMDELYK	255–261	Asp257	NA
HEYVNAAGITLGMDELYK	244-261	Glu258	NA

Figure S4. Graph showing the often-esterified carboxyl groups in sfGFP as identified by tandem mass spectrometry. Values of calculated carboxyl group pK_a are for PDB entry 2b3p (Table S1). GFP residue number also refers to PDB entry 2b3p.

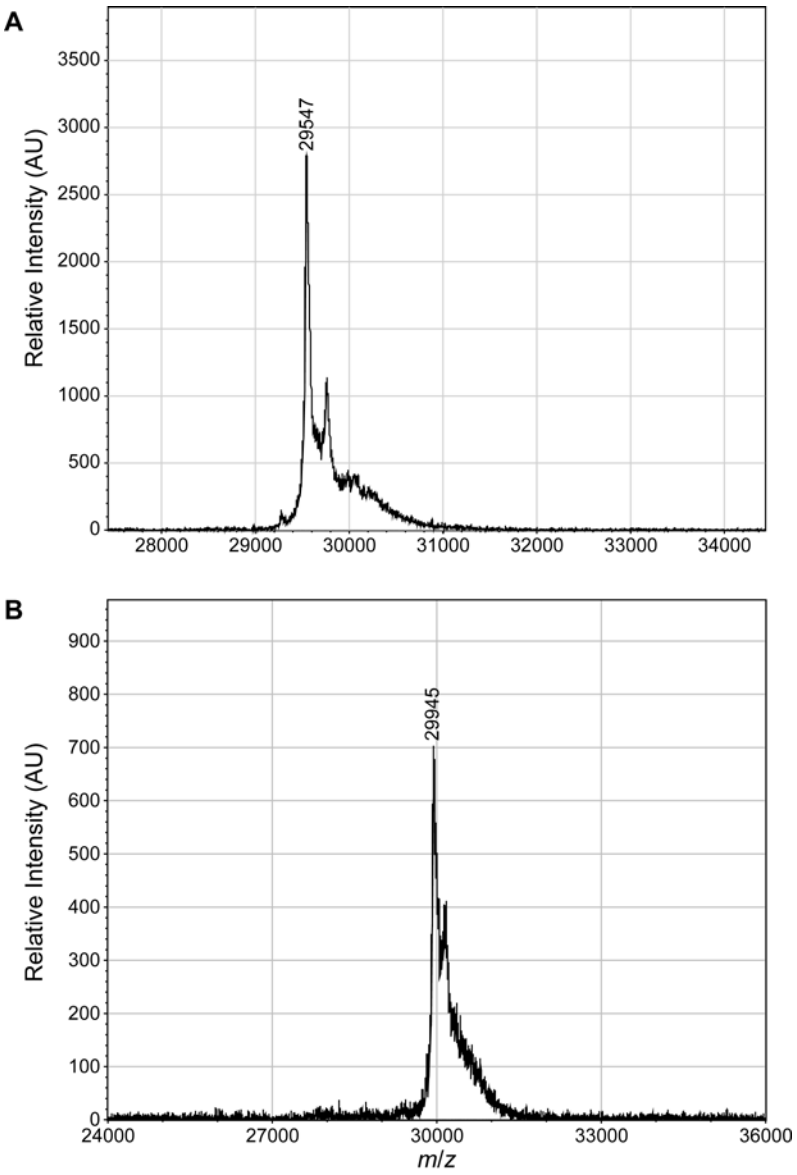


Figure S5. MALDI–TOF mass spectra of purified super-charged GFP and nlsGFP. (A) Super-charged GFP. *m*/*z*, 29,547; expected: 29,536 without an N-terminal methionine residue. (B) nlsGFP. *m*/*z*, 29,945; expected: 29,940 without an N-terminal methionine residue.

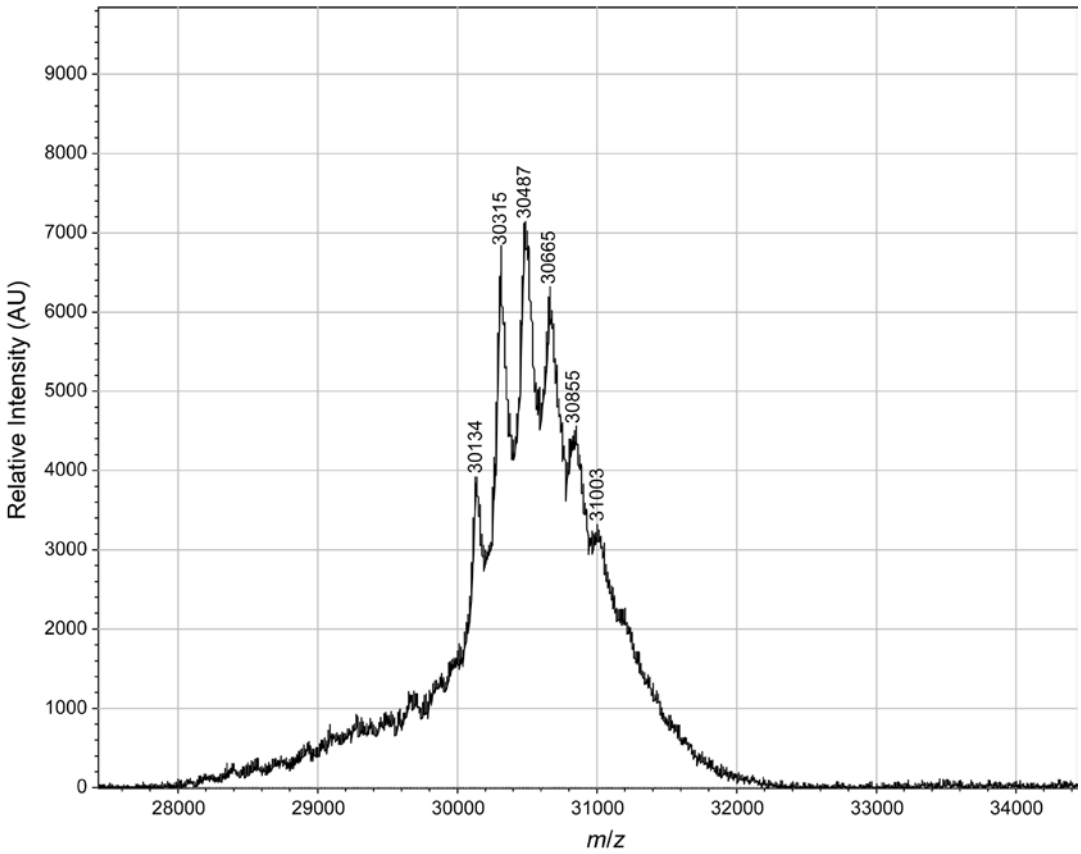


Figure S6. MALDI–TOF spectrum of nlsGFP esterified with diazo compound **1**. Expected *m*/*z*: 29,943 + 175 per ester group.

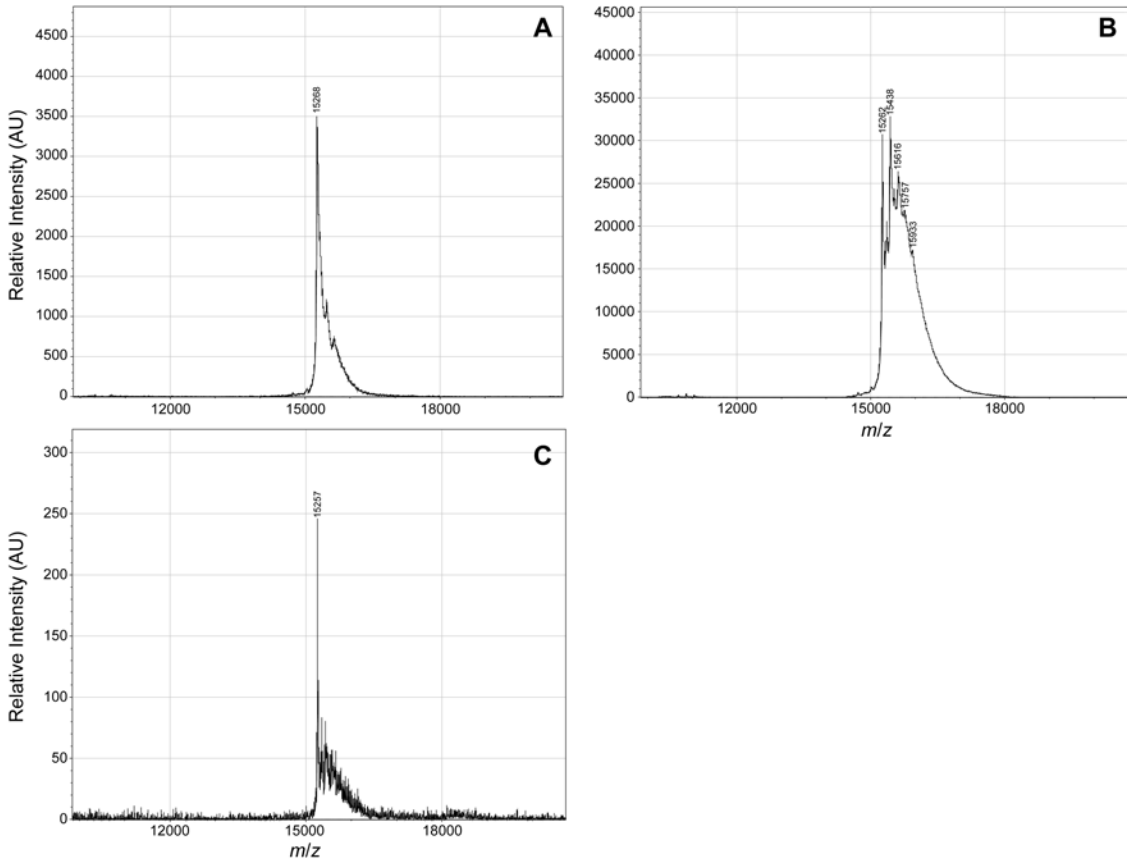


Figure S7. MALDI–TOF spectra to assess the reversibility of protein esterification with diazo compound **1**. (A) FLAG–angiogenin. (B) FLAG–angiogenin after treatment with diazo compound **1**. (C) FLAG–angiogenin after treatment with diazo compound **1** and subsequent incubation with a CHO-cell extract. Expected *m*/*z*: 15,270 + 175 per ester group.

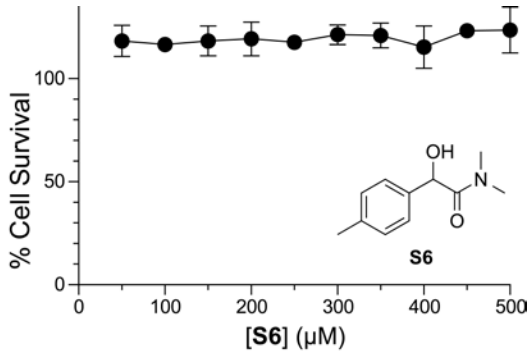
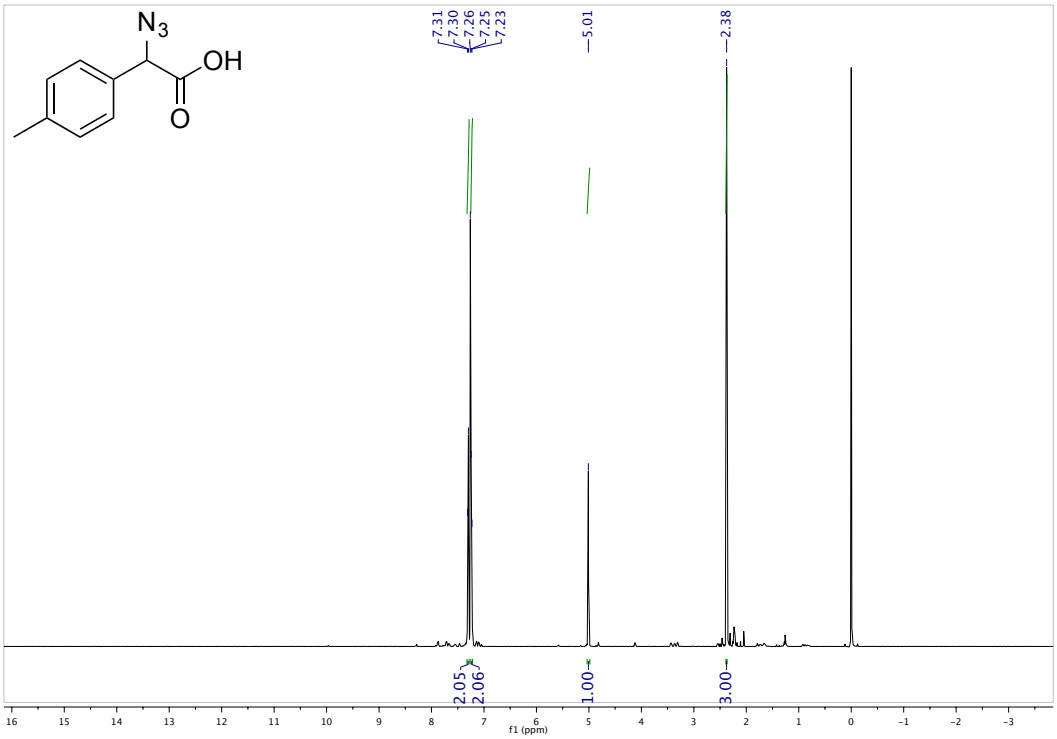


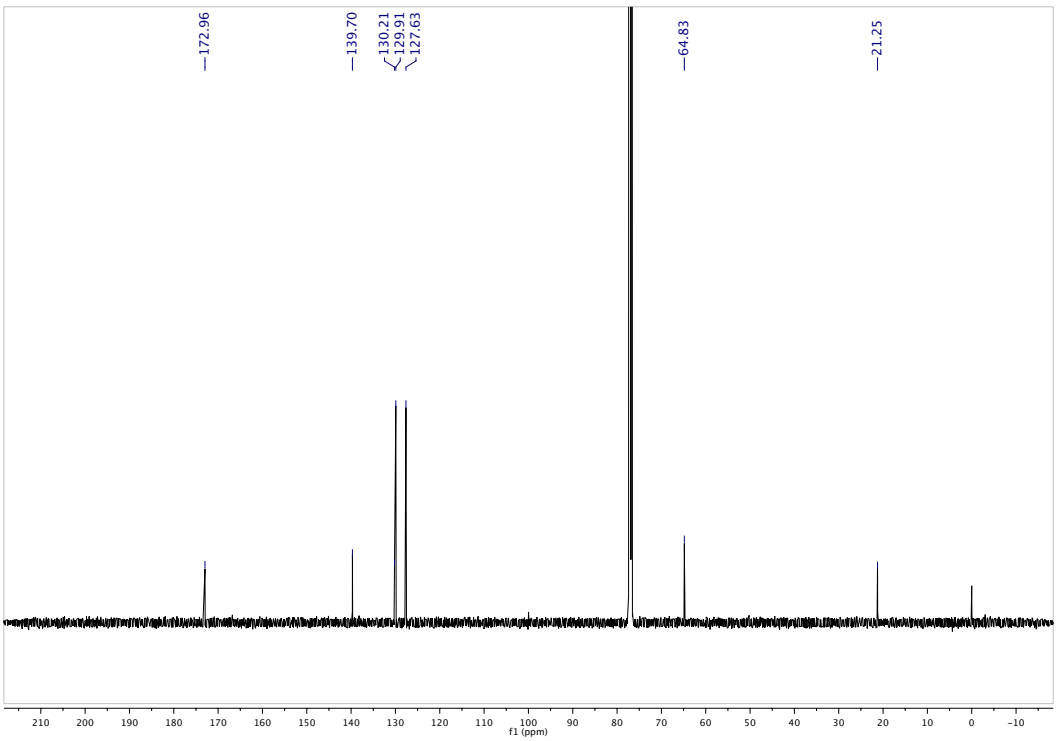
Figure S8. Graph of the viability of CHO-K1 cells treated with α-hydroxy dimethylamide **S6**.

NMR Spectra

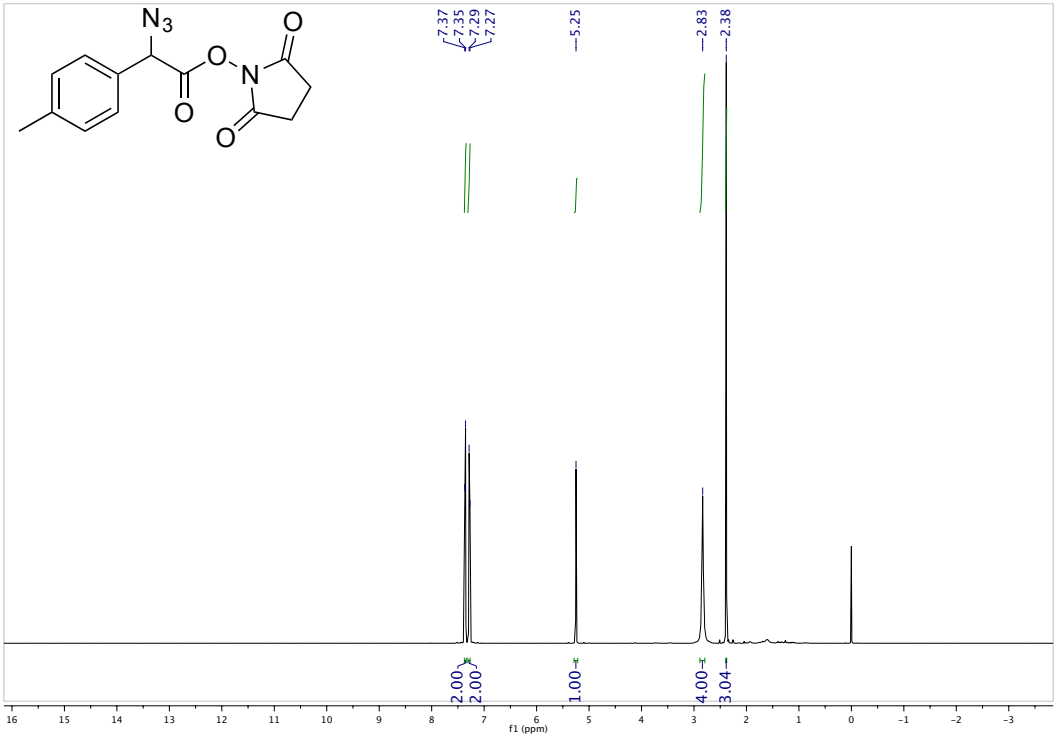
^{1}H NMR spectrum of compound **S1** in $CDCl_3$ (500 MHz):



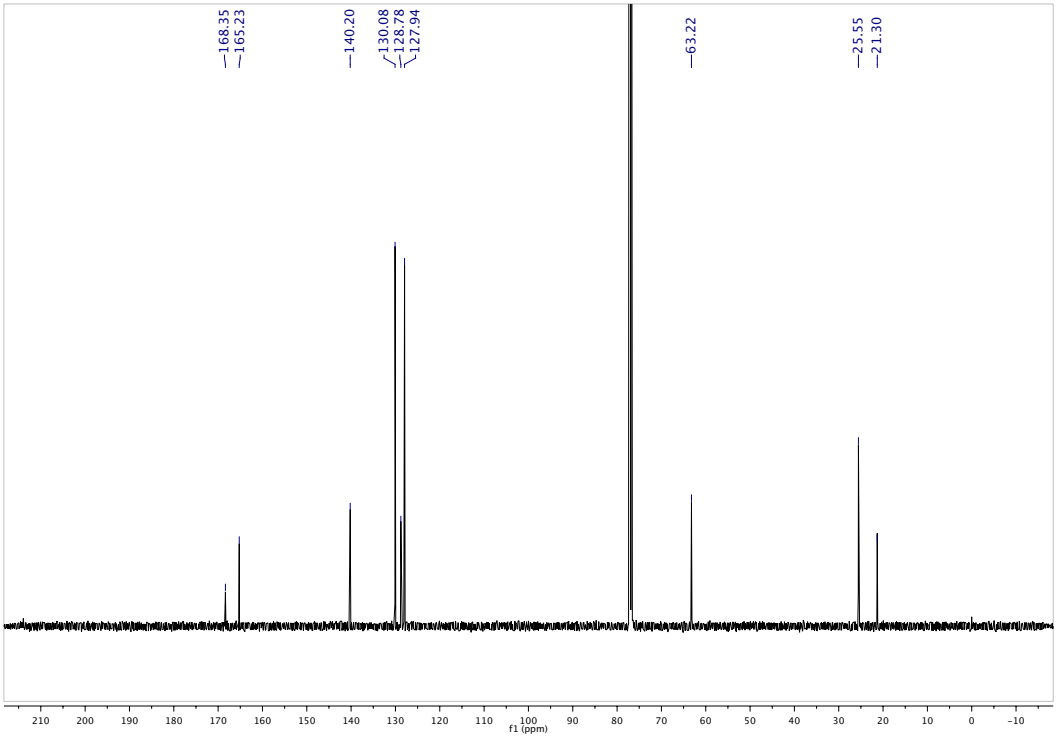
^{13}C NMR spectrum of compound **S1** in $CDCl_3$ (125 MHz):



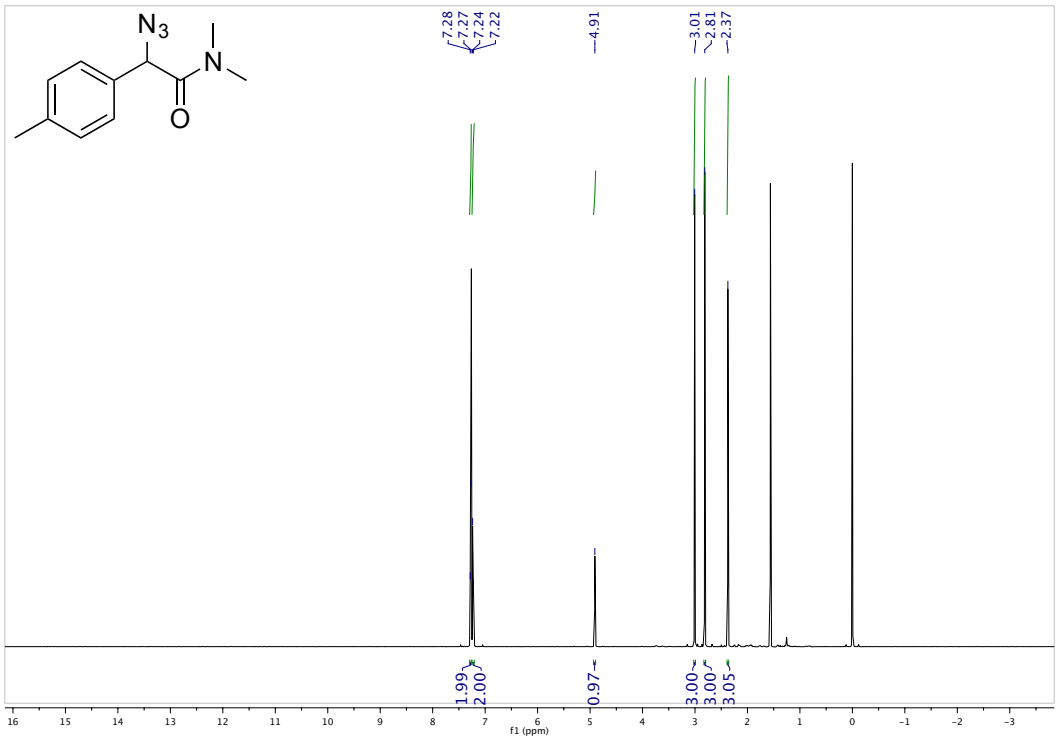
^{1}H NMR spectrum of compound **S2** in $CDCl_3$ (500 MHz):



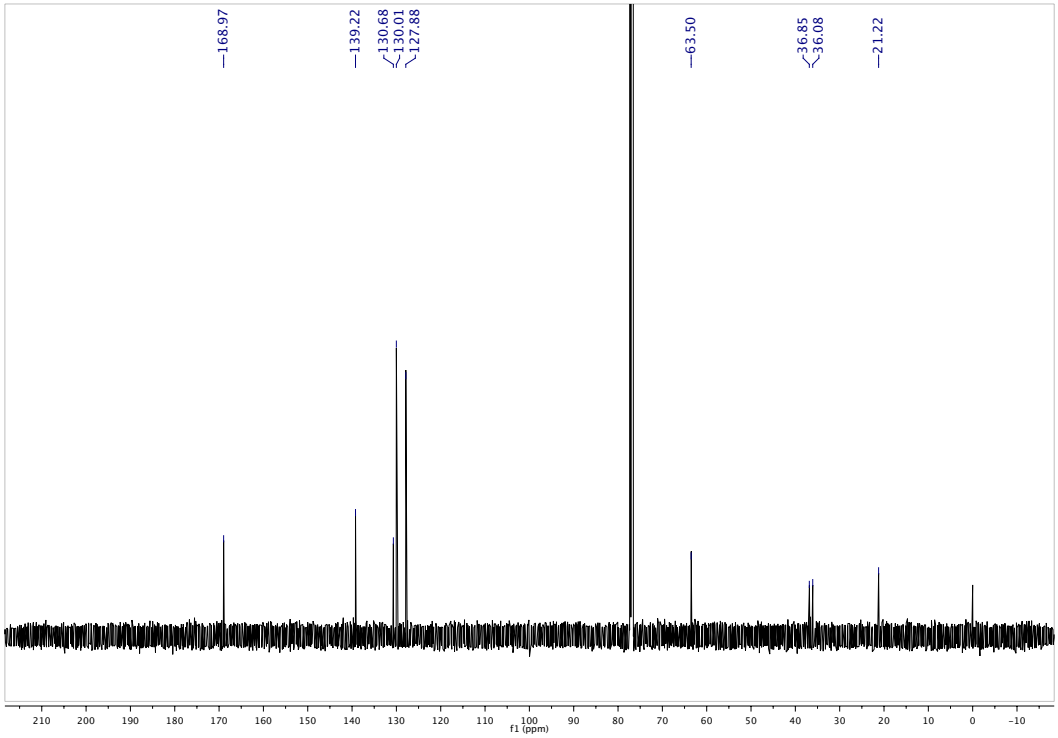
^{13}C NMR spectrum of compound **S2** in $CDCl_3$ (125 MHz):



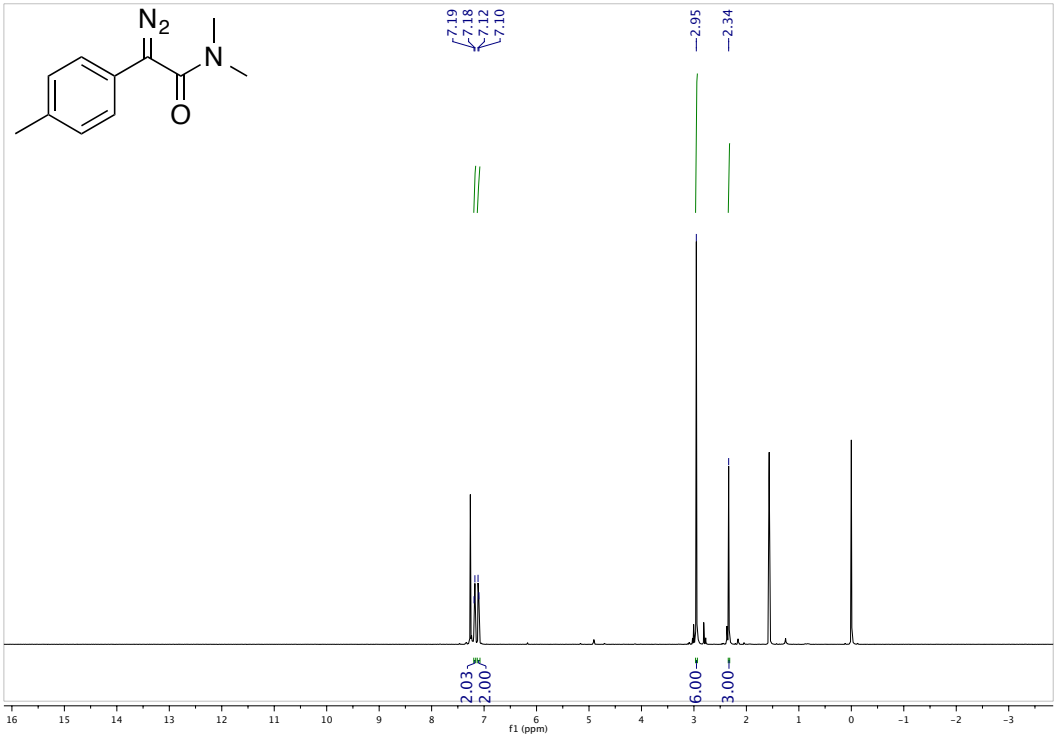
^{1}H NMR spectrum of compound **S3** in $CDCl_3$ (500 MHz):



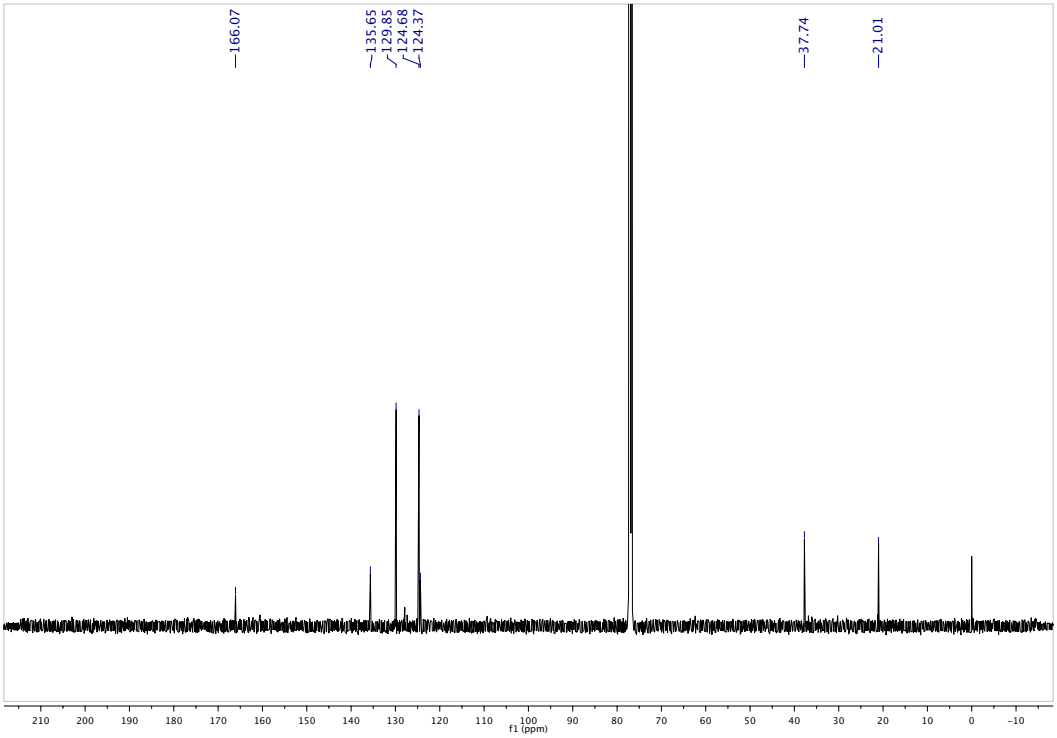
^{13}C NMR spectrum of compound **S3** in $CDCl_3$ (125 MHz):



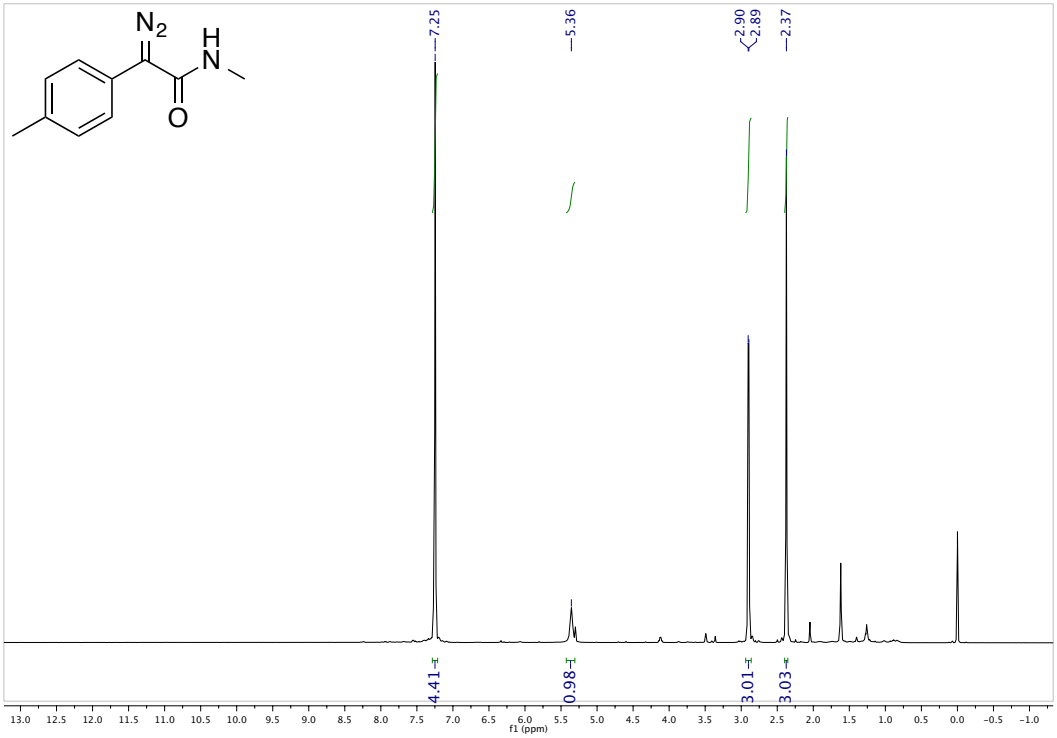
^{1}H NMR spectrum of compound **1** in $CDCl_3$ (500 MHz):



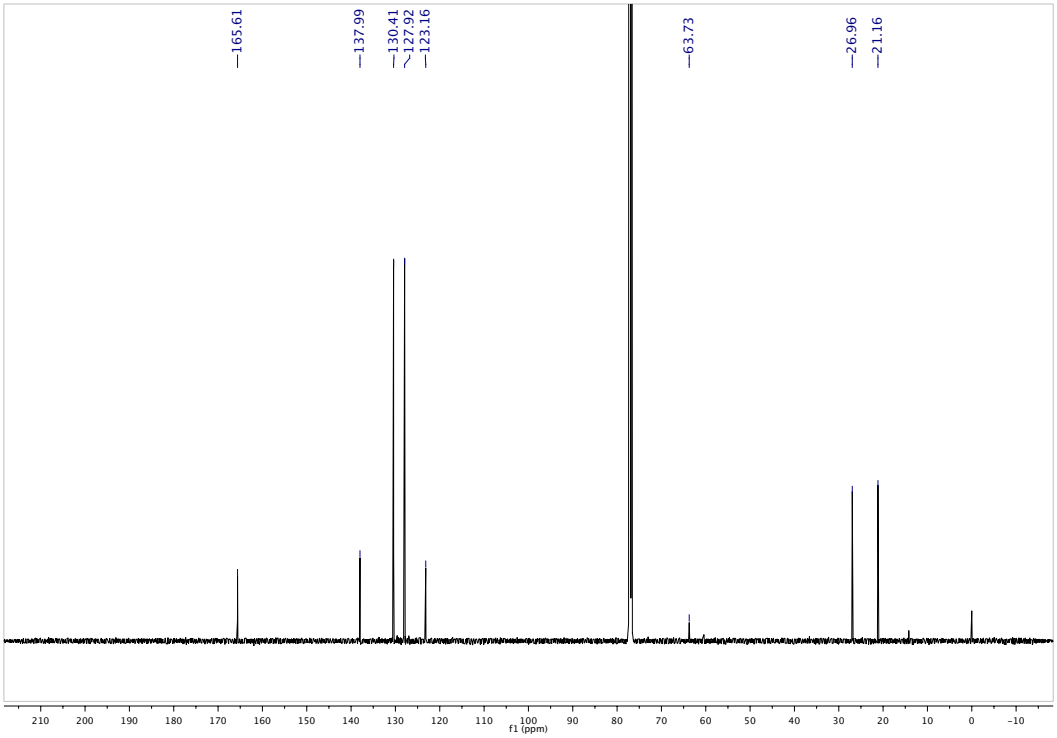
^{13}C NMR spectrum of compound **1** in $CDCl_3$ (125 MHz):



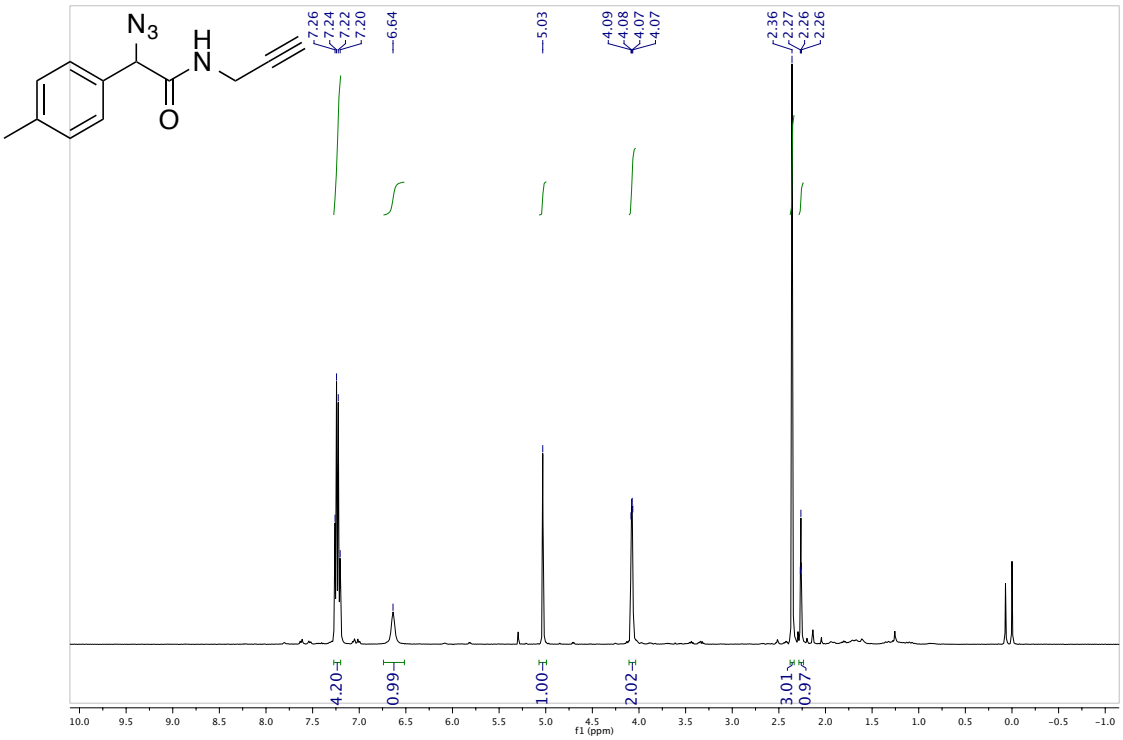
^{1}H NMR spectrum of compound **2** in $CDCl_3$ (500 MHz):



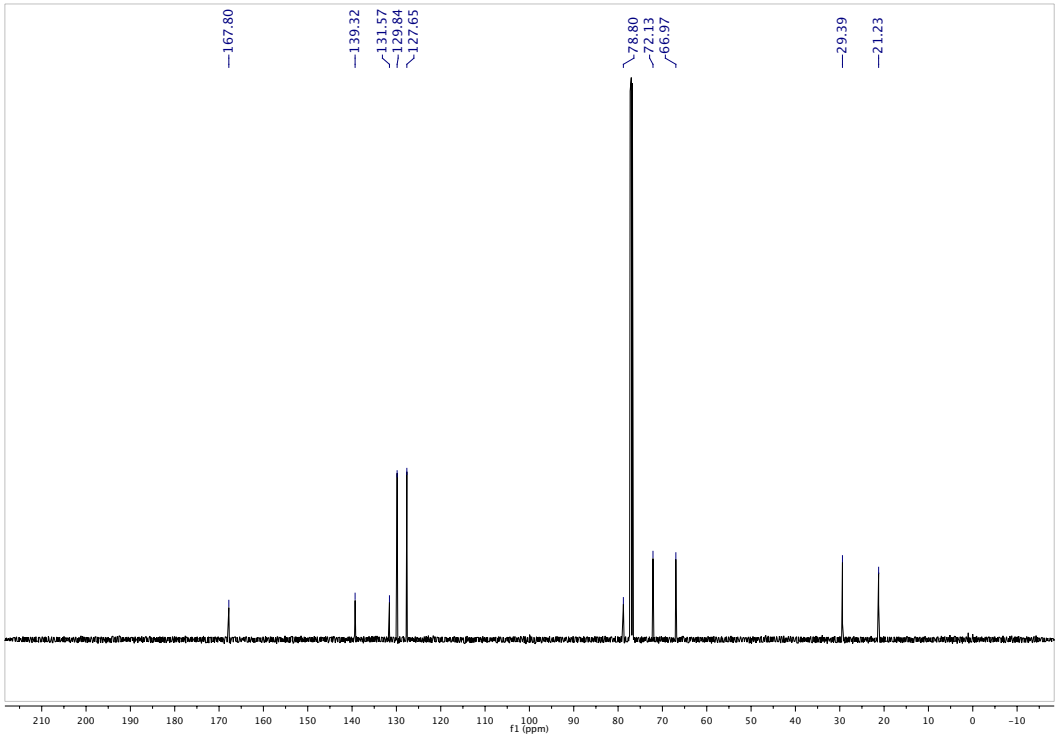
^{13}C NMR spectrum of compound **2** in $CDCl_3$ (125 MHz):



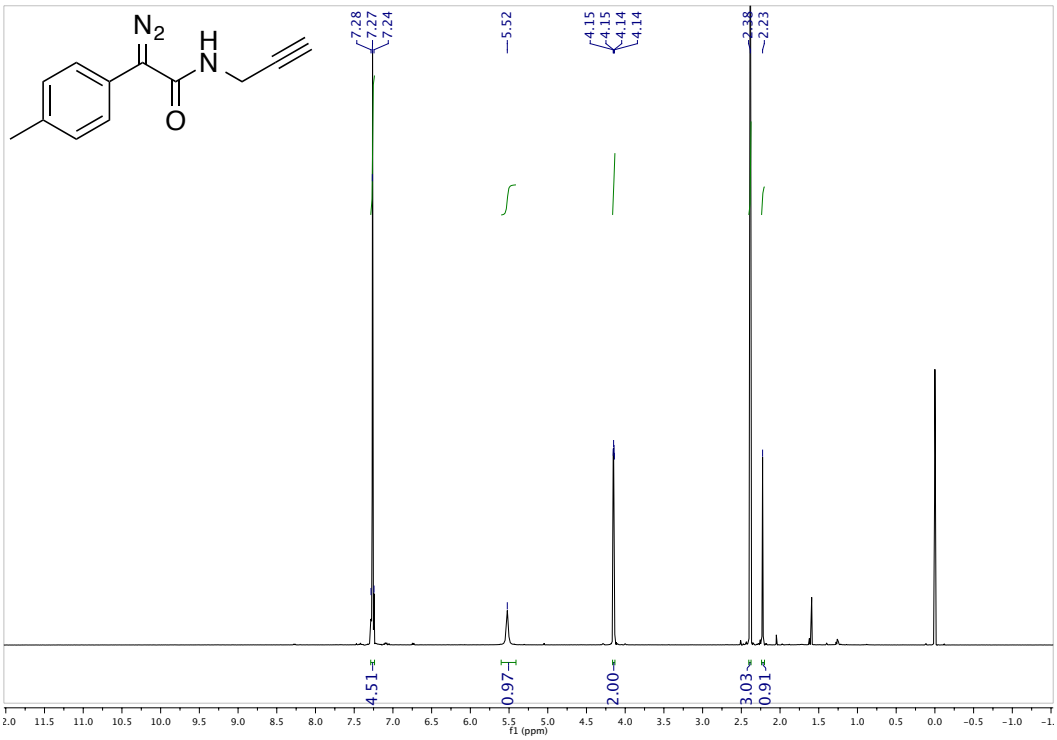
^{1}H NMR spectrum of compound **S4** in $CDCl_3$ (500 MHz):



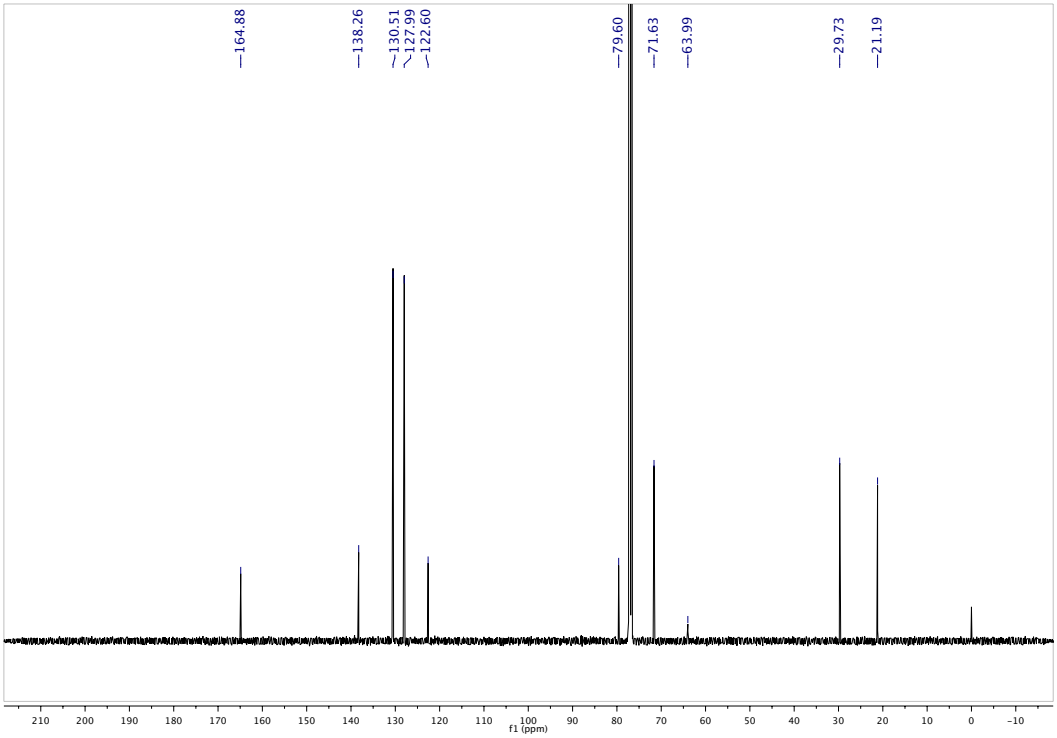
^{13}C NMR spectrum of compound **S4** in $CDCl_3$ (125 MHz):



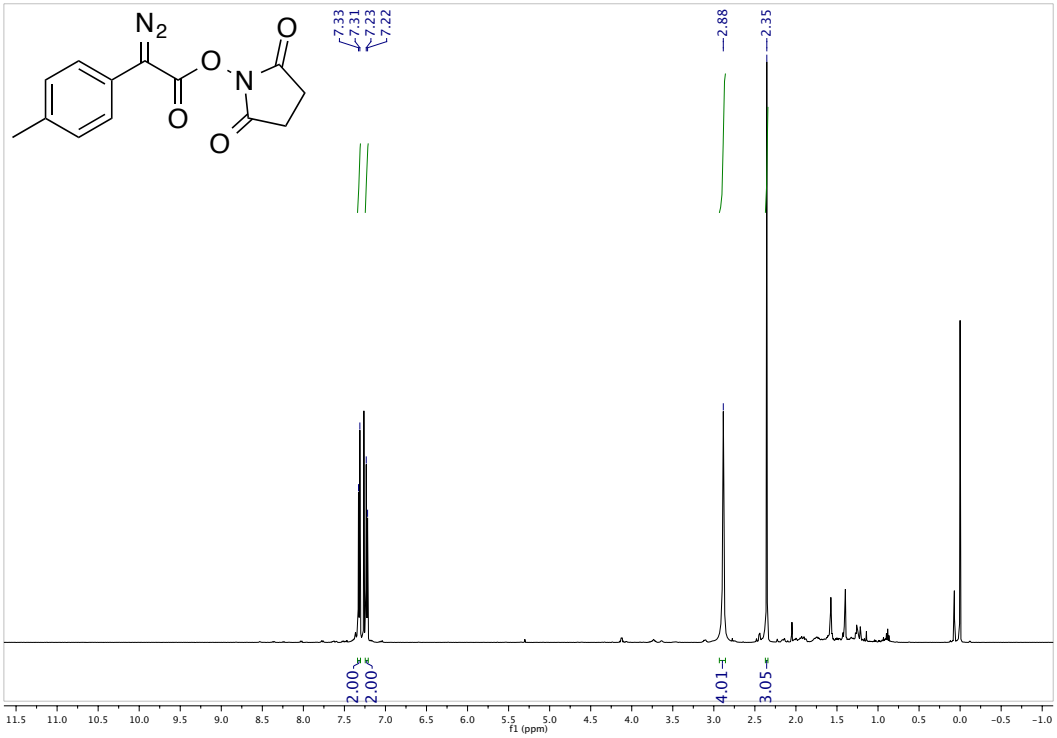
^{1}H NMR spectrum of compound **3** in $CDCl_3$ (500 MHz):



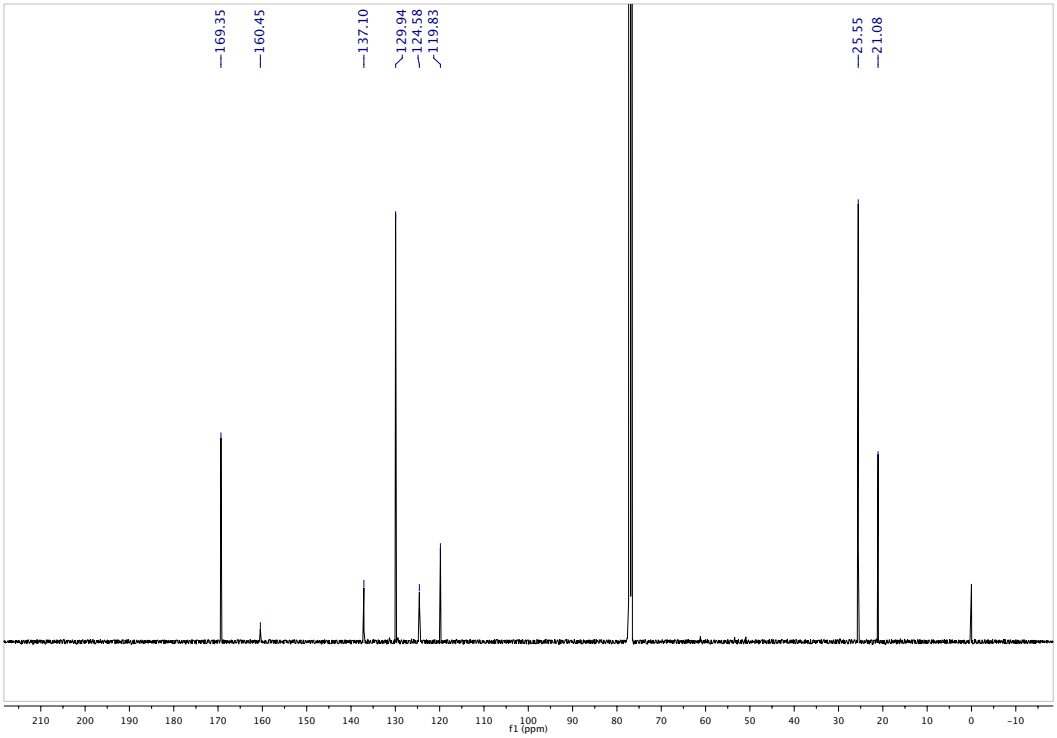
^{13}C NMR spectrum of compound **3** in $CDCl_3$ (125 MHz):



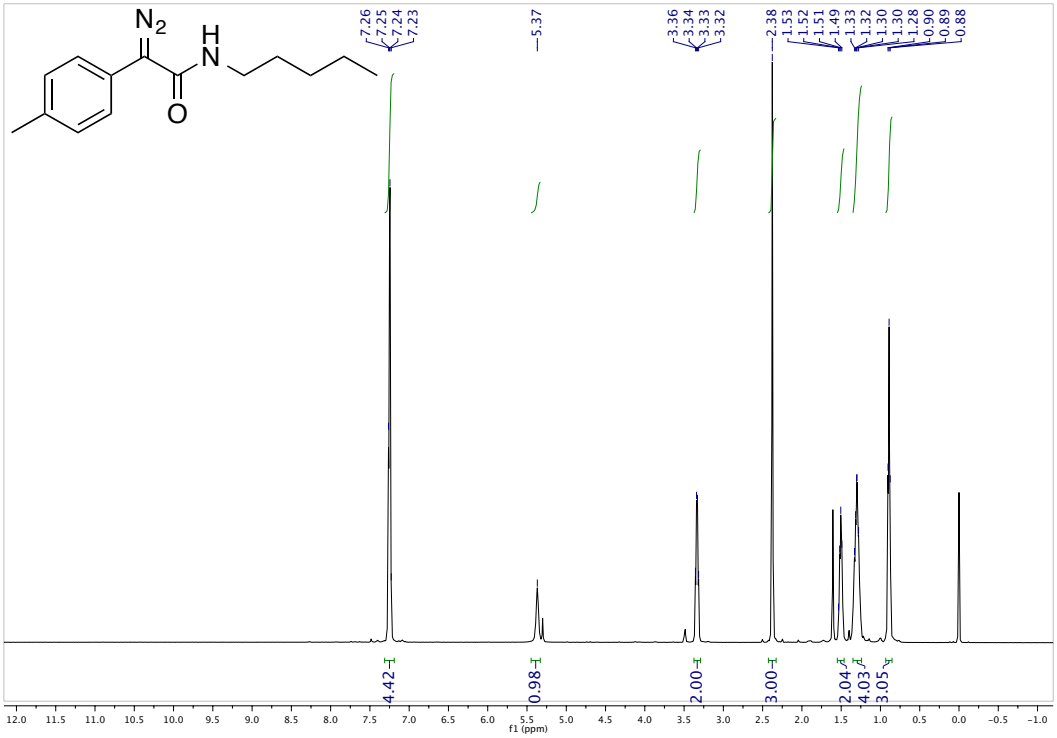
^{1}H NMR spectrum of compound **S5** in $CDCl_3$ (500 MHz):



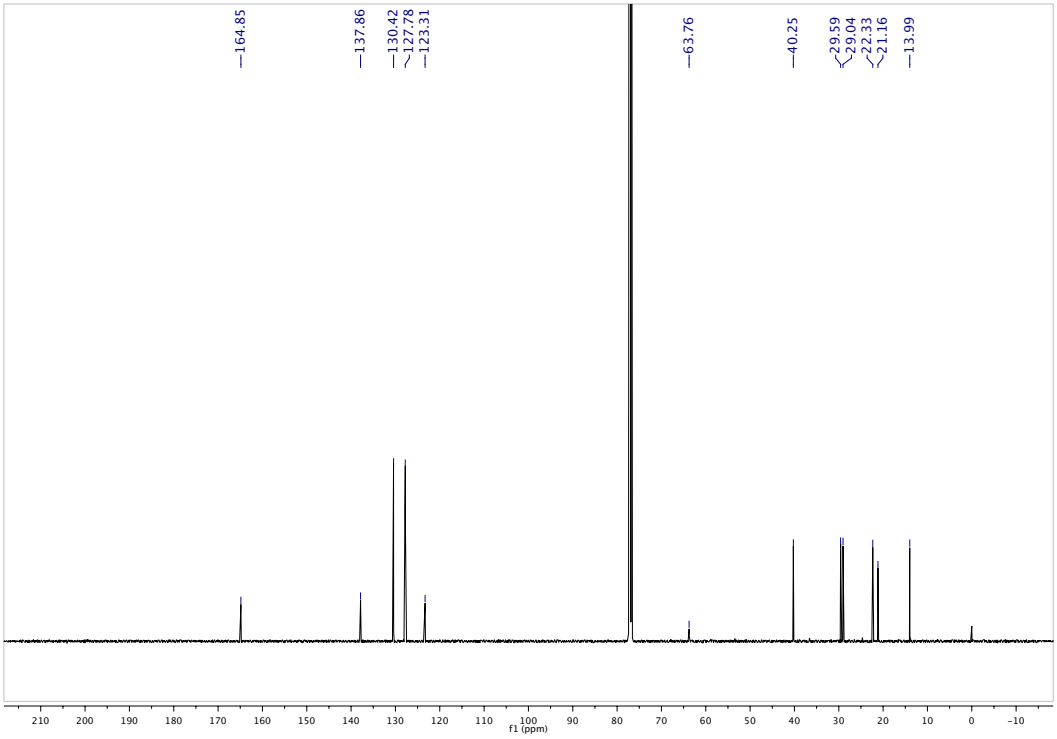
^{13}C NMR spectrum of compound **S5** in $CDCl_3$ (125 MHz):



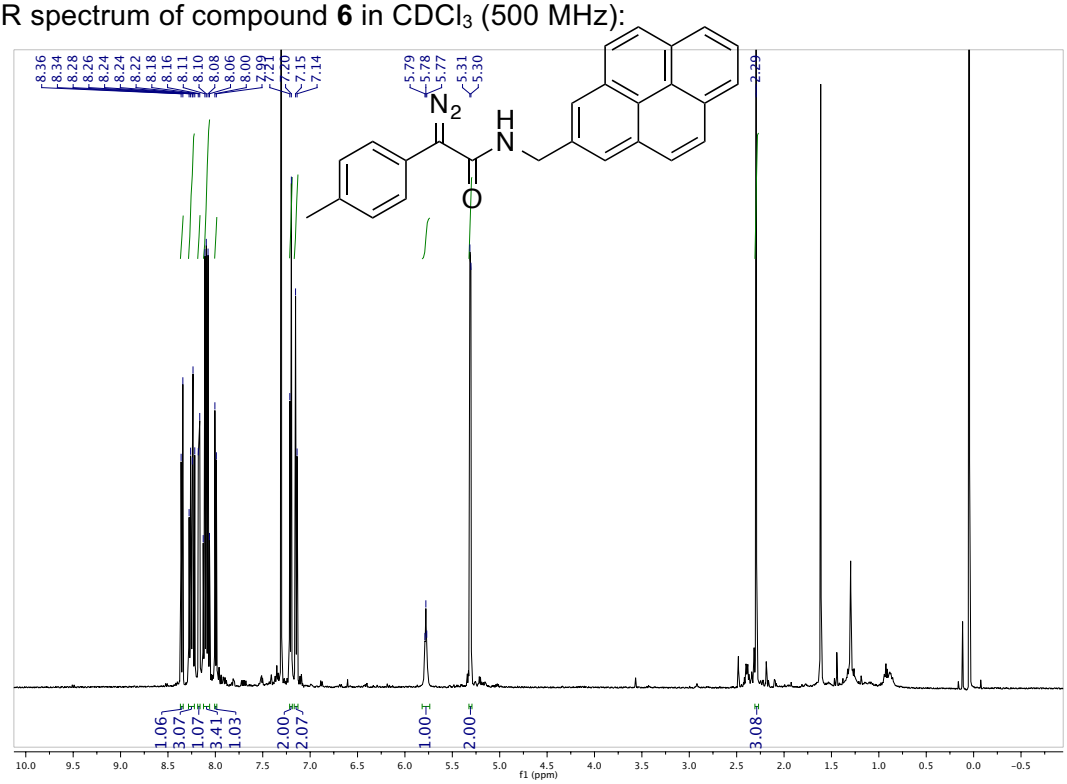
^{1}H NMR spectrum of compound **5** in $CDCl_3$ (500 MHz):



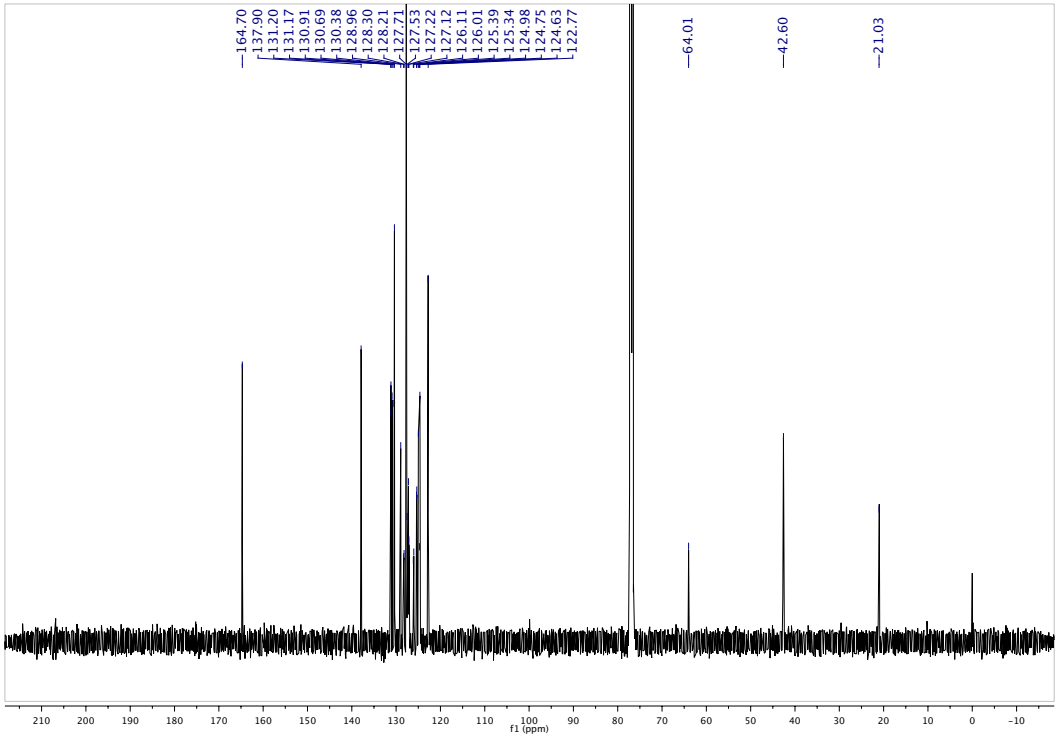
^{13}C NMR spectrum of compound **5** in $CDCl_3$ (125 MHz):



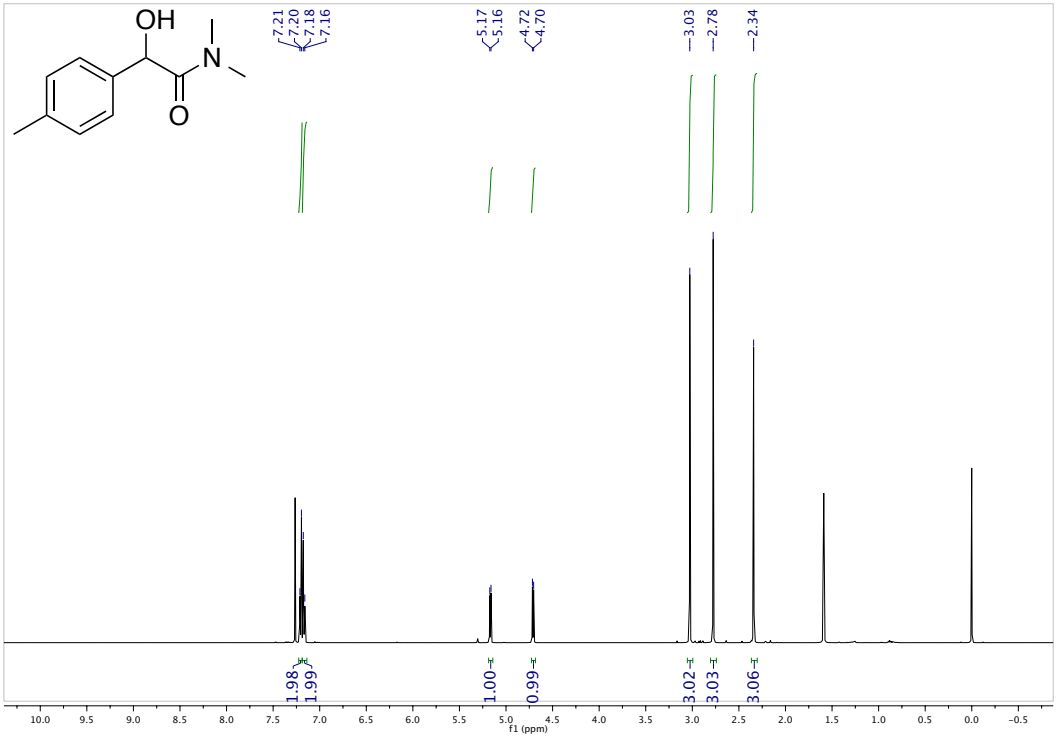
^{1}H NMR spectrum of compound **6** in $CDCl_3$ (500 MHz):



^{13}C NMR spectrum of compound **6** in $CDCl_3$ (125 MHz):



^{1}H NMR spectrum of compound **S6** in $CDCl_3$ (500 MHz):



^{13}C NMR spectrum of compound **S6** in $CDCl_3$ (125 MHz):