New aldehyde tag sequences identified by screening formylglycine generating enzymes *in vivo* and *in vitro*

Jason S. Rush and Carolyn R. Bertozzi

Departments of Chemistry and Molecular and Cell Biology and Howard Hughes Medical Institute, University of California, Berkeley, California 94720

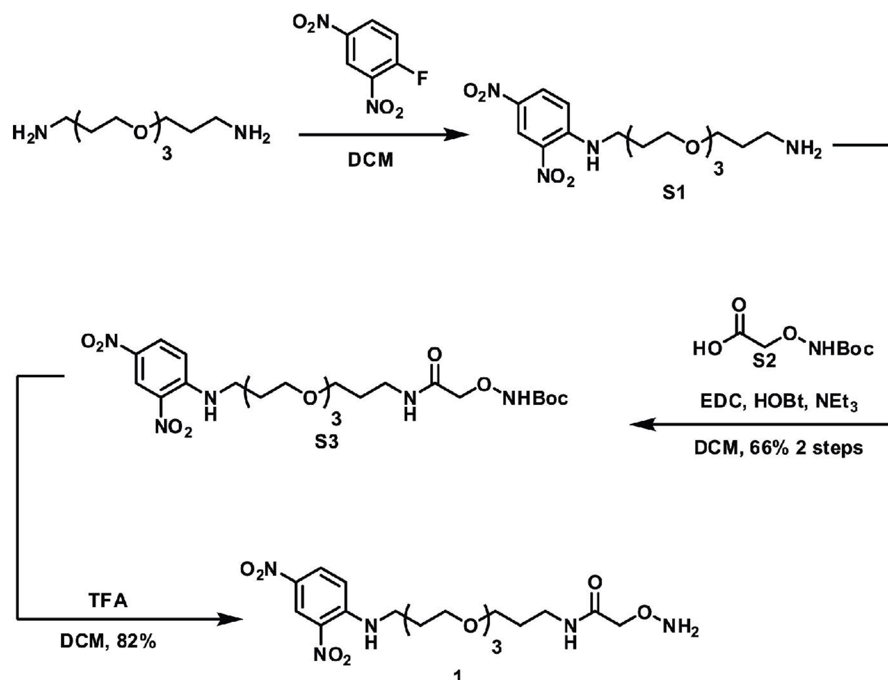
Table of Contents

- General methods S2
- Synthesis of **1** S3
- Experimental procedures S3-S8
- MS of FGE reactions S10-S13
- NMR spectra of **S3** and **1** S13-S17
- References S18

General methods

All reagents and solvents were obtained from commercial suppliers and used without further purification unless otherwise noted. Dichloromethane (DCM) was dried by passage over a column of activated alumina under a N_2 atmosphere. Reactions were monitored via thin-layer chromatography on Analtech uniplate® silica gel plates and visualized using UV absorption and Hanessian's ceric ammonium

molybdate stain. All solvents were removed using rotary evaporation under reduced pressure followed by static high vacuum. All ^{1}H and ^{13}C NMR were recorded on a Bruker AM-400® machine. Shifts are reported in δ values and referenced to the solvent peaks, either 7.26 ppm and 77.0 ppm for $CDCl_3$, or 1.94 ppm and 118.7 ppm for CD_3CN (^{1}H and ^{13}C signals, respectively). Coupling constants (*J*) are reported in Hz. Fast atom bombardment (FAB) spectra were obtained from the UC Berkeley Mass Spectrometry Laboratory. Electrospray mass spectrometry (ESI-MS) was performed on a Hewlett-Packard 1100 mass spectrometer. Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) was performed on a Applied Biosystems Voyager DE Pro mass spectrometer. Quantitative amino acid analysis (QAAA) was performed at the Molecular Structure Facility at UC Davis. High-pressure liquid chromatography (HPLC) was performed on a Rainin Dynamax SD-200 HPLC system using Microsorb and Dynamax C18 reversed-phase columns (analytical: 4.6 x 250 mm, 1 mL/min; semi-preparative: 10 x 250 mm, 3 mL/min) and UV detection was performed with a Rainin Dynamax UV-1 detector. *S. coelicolor* and *M. tuberculosis* FGEs were expressed in *E. coli* as N- and C-terminal His_6 fusions, respectively, and purified as previously described.[1] Site directed mutagenesis was accomplished using a modified QuikChange (Stratagene) protocol[2] using previously disclosed plasmids.[1] DNA sequencing was used to confirm the fidelity of gene products. In-gel fluorescence was imaged using a Typhoon 9410 scanner (GE Healthcare).



Scheme S1. Synthesis of compound **1**.

Experimental procedures

***N*-(3-(2-(2-(3-Aminopropoxy)ethoxy)ethoxy)propyl)-2,4-dinitroaniline (S1)**. A flame dried flask under N_2 was charged with 4,7,10-Trioxa-1,13-tridecanediamine (1.07 g, 4.84 mmol) and DCM (15 mL). 1-Fluoro-2,4-dinitrobenzene (150 mg, 0.8 mmol) in DCM (15 mL) was added dropwise over 15 min via an addition funnel. After 2.5 h the reaction was transferred to a separatory funnel, washed with H_2O (3 x 20 mL) and brine (2 x 20 mL), then dried (K_2CO_3), filtered and concentrated to afford a yellow syrup which was used without further purification.

***tert*-Butyl 16-(2,4-dinitrophenylamino)-2-oxo-7,10,13-trioxa-3-azahexadecyloxycarbamate (S3).** A flame dried flask under a N_2 atmosphere was charged with **S2** (229 mg, 1.20 mmol), *N*-Ethyl-*N*′-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, 230 mg, 1.20 mmol), 1-hydroxybenzotriazole hydrate (HOBt, 162 mg, 1.20 mmol) triethylamine (NEt_3, 0.22 mL, 1.6 mmol), and DCM (1.6 mL). After

stirring for 20 min at rt, the solution was added to another flame-dried flask under N_2 containing **S1** and the yellow solution was stirred overnight. The reaction was then diluted with DCM (20 mL) and washed with 0.1 M HCl (2 x 10 mL), satd. $NaHCO_3$ (2 x 10 mL), and brine (1 x 10 mL), then dried (Na_2SO_4), filtered and concentrated. The crude material was purified on silica with two successive columns. The first with EtOAc and the second with DCM/EtOH (30:1) to afford **S3** (294 mg, 66% over 2 steps). Rf = 0.2 (EtOAc); ^{1}H NMR(400 MHz, $CDCl_3$) δ: 9.06 (1H, d, *J* = 2.8), 8.89 (1H, s), 8.21 (1H, dd, *J* = 10.0, 2.4), 8.19 (1H s), 7.95 (1H, s), 6.95 (1H, d, *J* = 9.6), 4.26 (2H, s), 3.67-3.5 (14H, m), 3.36 (2H, q, *J* = 6.4), 2.00 (2H, m), 1.78 (2H, m), 1.43 (9H, s); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 168.7, 157.5, 148.3, 135.6, 130.2, 130.1, 124.2, 113.8, 82.5, 75.7, 70.5, 70.4, 70.3, 70.0, 69.4, 69.1, 41.8, 36.8, 29.0, 28.5, 28.0; HRMS (FAB+) calcd. for $C_{23}H_{38}N_5O_{11}$ $(MH)^+$: 560.2572, found: 560.2568.

2-(Aminooxy)-*N*-(3-(2-(2-(3-(2,4-dinitrophenylamino)propoxy)ethoxy)ethoxy)propyl)acetamide (1). A flask under a N_2 atmosphere was charged with **S3** (82 mg, 0.15 mmol) , DCM (3 mL) and TFA (1 mL) then stirred at rt for 4 h. The reaction was concentrated and the crude material was purified by silica gel chromatography eluting with DCM/EtOH (20:1) to afford **1** (55 mg, 82%). Rf = 0.1 (20:1 DCM/EtOH); ^{1}H NMR(400 Mhz, CD_3CN) δ: 8.92 (1H, d, *J* = 0.7), 8.81 (1H, s), 8.21 (1H, dd, *J* = 2.5, 0.7), 7.09 (1H, d, *J* = 2.4), 6.95 (1H, s), 5.95 (2H, s), 3.94 (2H, s), 3.62-3.50 (12H, m), 3.48 (2H, t, *J* = 1.5), 3.26 (2H, q, *J* = 1.5), 1.97 (2H, m), 1.70 (2H, m); ^{13}C NMR (100 MHz, CD_3CN) δ: 171.1, 150.0, 136.7, 131.4, 131.3, 125.1, 116.0, 76.0, 71.53, 71.51, 71.4, 71.3, 70.5, 70.1, 42.9, 37.7, 30.7, 29.7; HRMS (FAB+) calcd. for $C_{18}H_{30}N_5O_9$ $(MH)^+$: 460.2044, found: 460.2043.

Library synthesis

The library members were synthesized on Fmoc-Lys(biotin) Wang resin (0.03 mmol) using N^{α}-Fmoc protected amino acids and DIC/HOBt ester activation in NMP. A five-fold excess of amino acid was used in each coupling except for the Thr-Pro step, which was accomplished using Fmoc-Thr(OtBu)OH (0.3 mmol) , HATU (0.28 mmol), and diisopropylethylamine (0.56 mmol) followed by agitation for 10 h. Fmoc removal was accomplished using 30% piperdine in NMP. The peptides was terminally acylated with pyridine/acetic anhydride (2:1) for 1 h at rt. Peptide cleavage/deprotection was accomplished by treatment with TFA/H_2O/EDT/TIS (94:2.5:2.5:1) for 3 h at rt. The peptides were precipitated from cold MTBE, dissolved in H_2O and lyophilized. The crude peptides were purified by semi-preparative RP-HPLC using CH_3CN/H_2O gradients with 0.1% TFA. Peptide concentrations were determined via QAAA. Additionally, an authentic FGly containing sequence was synthesized using previously reported methods.[3]

Functionality of *in vitro* FGE assay

Reactions containing FGly-containing peptide, or the C→A variant, (333 uM) and compound **1** (1.67 mM) in H_2O with 33 mM KOAc pH = 4.6, at a final volume of 6 µL or 16.5 µL (FGly and C→A, respectively) were incubated at 37 ^{o}C for 3 h. The FGly peptide containing reaction was then diluted to 200 µL while the alanine peptide containing reaction was diluted to 550 µL, both with H_2O. Varying ratios of the two reactions were mixed and 50 µL of the mixtures were then loaded onto a 96-well NeutrAvidin plate which had been washed 3x with 200 µL wash buffer (25 mM Tris, 150 mM NaCl, 0.5% BSA (w/v), and 0.05% Tween20 (v/v), pH = 7.2), then incubated with 50 µL of 100 uM $MeONH_2$ in 0.1 M KOAc, pH = 4.6, for 15 min at rt, then washed 3x again with 200 µL wash buffer. After 1 h at rt in the absence of light, the wells were washed 3x with 200 µL wash buffer and 100 µL of α-DNP-AlkPhos (Sigma) diluted 1:5000 in wash buffer was added. The plate was incubated 1 h at rt in the absence of light, washed 3x with 200 µL

wash buffer, after which 200 µL of 1 mg/mL *p*-nitrophenyl phosphate (pNPP) in reaction buffer (10% diethanolamine, 0.5 mM $MgCl_2$, pH = 9.8) was added. After 5-10 min at rt in the absence of light, the reaction was quenched with 50 µL of 3 M KOH and Abs_{405} was read using a UV/VIS spectrophotometric microtiter plate reader.

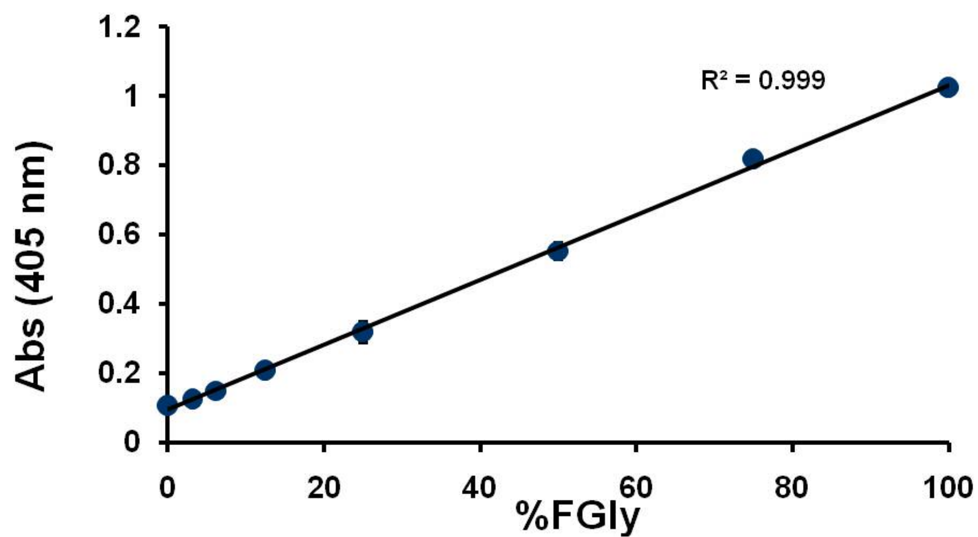


Figure S1. Assay response to varying ratios of FGly. Values and error represent the average and standard deviation of the 3 replicates.

Reaction of peptides with FGE

Reactions containing 125 µM biotinylated peptide, 780 nM FGE, 50 mM Tris, 68 mM NaCl, 0.2 mg/mL BSA and 2 mM DTT, pH = 9 were assembled in the following manner. A solution of biotinylated peptide (2.5 µL, 1 mM) and DIH_2O (12.5 µL) was warmed to 30°C. The reaction was initiated by addition of a mixture of 10X buffer (2 µL, 0.5 M Tris, 680 mM NaCl, 2 mg/mL BSA, 20 mM DTT, pH = 9), FGE (0.04 µL 14.5 mg/mL) and H_2O (2.96 µL). Control reactions lacking FGE contained 10X buffer (2 µL) and H_2O (3 µL) in addition to peptide. After 20 min at 30 ^{o}C the reactions were quenched by the addition of 10 µL 1M KOAc pH = 4.6.

Analysis of FGE-modified peptides

Reactions containing 7.2 µL of the quenched enzymatic reactions and 1 µL of 5 mM compound **1** were incubated at 37 ^{o}C for 3 h then diluted to a final volume of 60 µL. 50 µL were then loaded onto a 96-well NeutrAvidin plate which had been pre-treated in the following manner. The plate was rinsed 3x with 200 µL of wash buffer, then incubated with 50 µL of 100 uM $MeONH_2$ in 0.1 M KOAc, pH = 4.6, for 15 min at rt, then rinsed again 3x with 200 µL of wash buffer. After 1 h at rt in the absence of light, the plate was rinsed 3x with 200 µL of wash buffer and 100 µL of α-DNP-AlkPhos (Sigma) diluted 1:5000 in wash buffer was added. The plate was incubated 1 h at rt in the absence of light, rinsed 3x with 200 µL of wash buffer, after which, 200 µL of 1 mg/mL pNPP in reaction buffer was added. After 5-10 min at rt in the absence of light, the reaction was quenched with 50 µL of 3 M KOH and Abs_{405} was read using a UV/VIS spectrophotometric microtiter plate reader. Signal from reactions without FGE was subtracted as background and % conversion was calculated based on the wild-type sequence for the enzyme.

Expression and chemoselective labeling of ald_6MBP and its related mutants

Mutations to the previously described ald_6MBP gene were performed as described in the general methods using the oligonucleotides in supplementary table 1. A plasmid containing ald_6MBP, or a mutant thereof, were transformed into BL21(DE3) cells (Invitrogen). Clonal popµlations were used to seed 5 mL cultures of LB media containing kanamycin at 37 °C. When OD_{600} = 0.5 the cultures were cooled to rt and MBP expression was induced with 1 mM IPTG. The cultures were agitated at rt overnight, then lysed using the BugBuster reagent (Novagen). MBP was isolated using nickel spin columns (Qiagen) according to the manufacturer's instructions. Protein concentration was normalized using the DC Protein Assay (Bio-Rad) and labeling using Alexa Fluor 647 C5-aminooxyacetamide was accomplished as previously reported.[4]

Supplementary table 1

Primer	Sequence (5'→3')
MBP LCTASR 5'	GCACAGCATCGCGGTGAG
MBP LCTASR 3'	GCGATGCTGTGCACAGGG
MBP LATASR 5'	GCCACAGCATCGCGGTGAGCG
MBP LATASR 3'	GCGATGCTGTGGCCAGGGATC
MBP LCTASA 5'	GCATCGGCGTGAGCGGCCGCAC
MBP LCTASA 3'	GCCGCTCACGCCGATGCTGTGCACAGG
MBP LATASA 5'	CACAGCATCGGCGTGAGCGGCCGCAC
MBP LATASA 3'	CGCTCACGCCGATGCTGTGGCCAGGG

Mass spectrometry confirmation of FGly formation and reactivity

Enzymatic reactions were performed as above but quenched by the addition of 1 µL 10% TFA. The reactions were desalted on C18 ZipTips and eluted with 3 µL CH_3CN with 0.1% TFA. 1 µl of this was mixed with 1 µL of 10 mM $MeONH_2$ and the reaction was incubated at rt for 30 min. The reaction was then mixed 1:1 with matrix solution (10 mg/mL α-cyano-4-hydroxy-cinnamic acid in 50% CH_3CN, 0.1%

TFA) and analyzed by MALDI-TOF MS. Samples of the mass spectra of the reactions are shown in Figs. S2-S4.

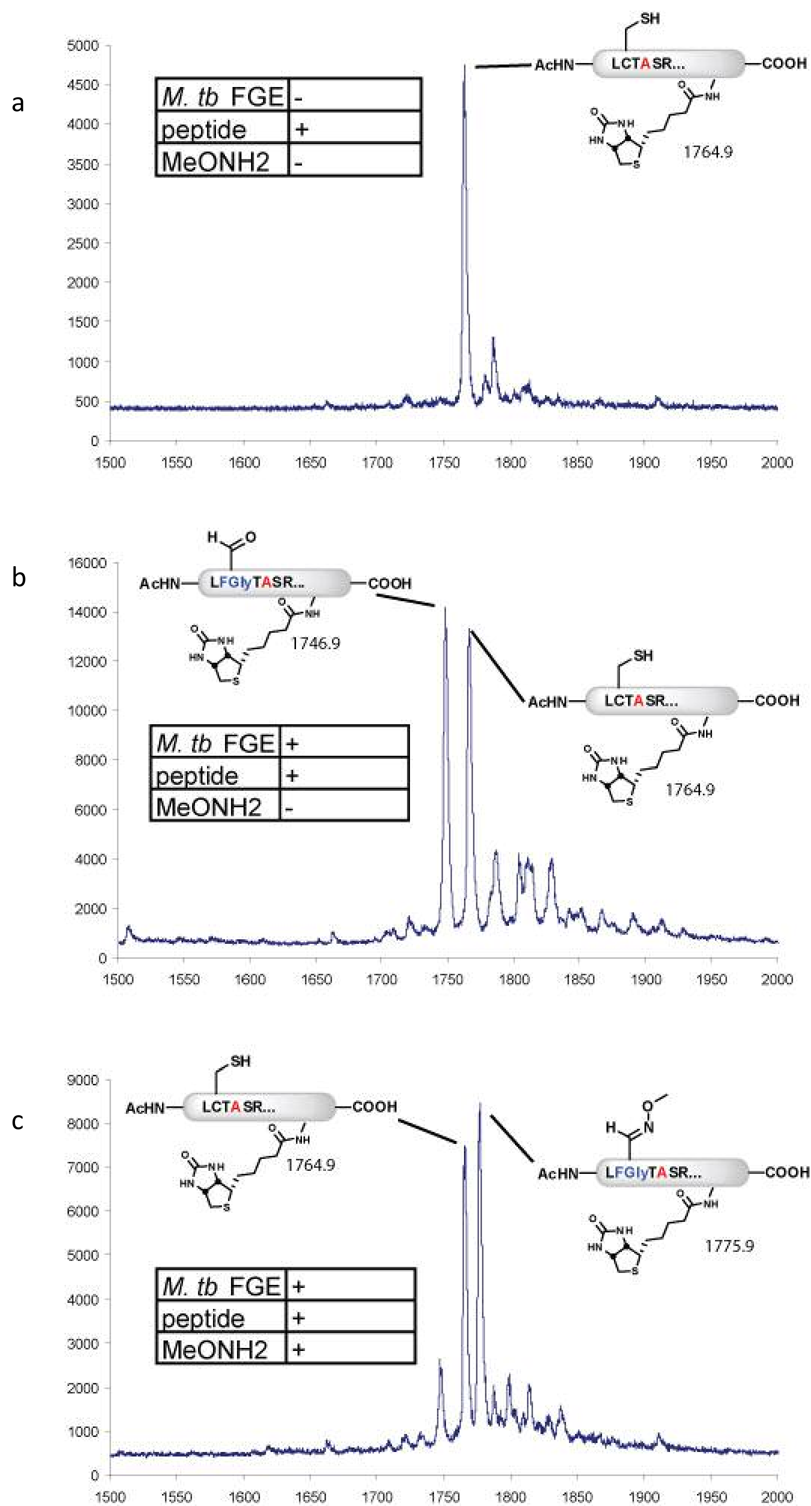


Figure S2. Mass spectra of LCTASRGSLFTGR from FGE reactions before or after treatment with $MeONH_2$. (a) Peptide in the absence of FGE. (b) Peptide in the presence of FGE. (c) Peptide after reaction with $MeONH_2$

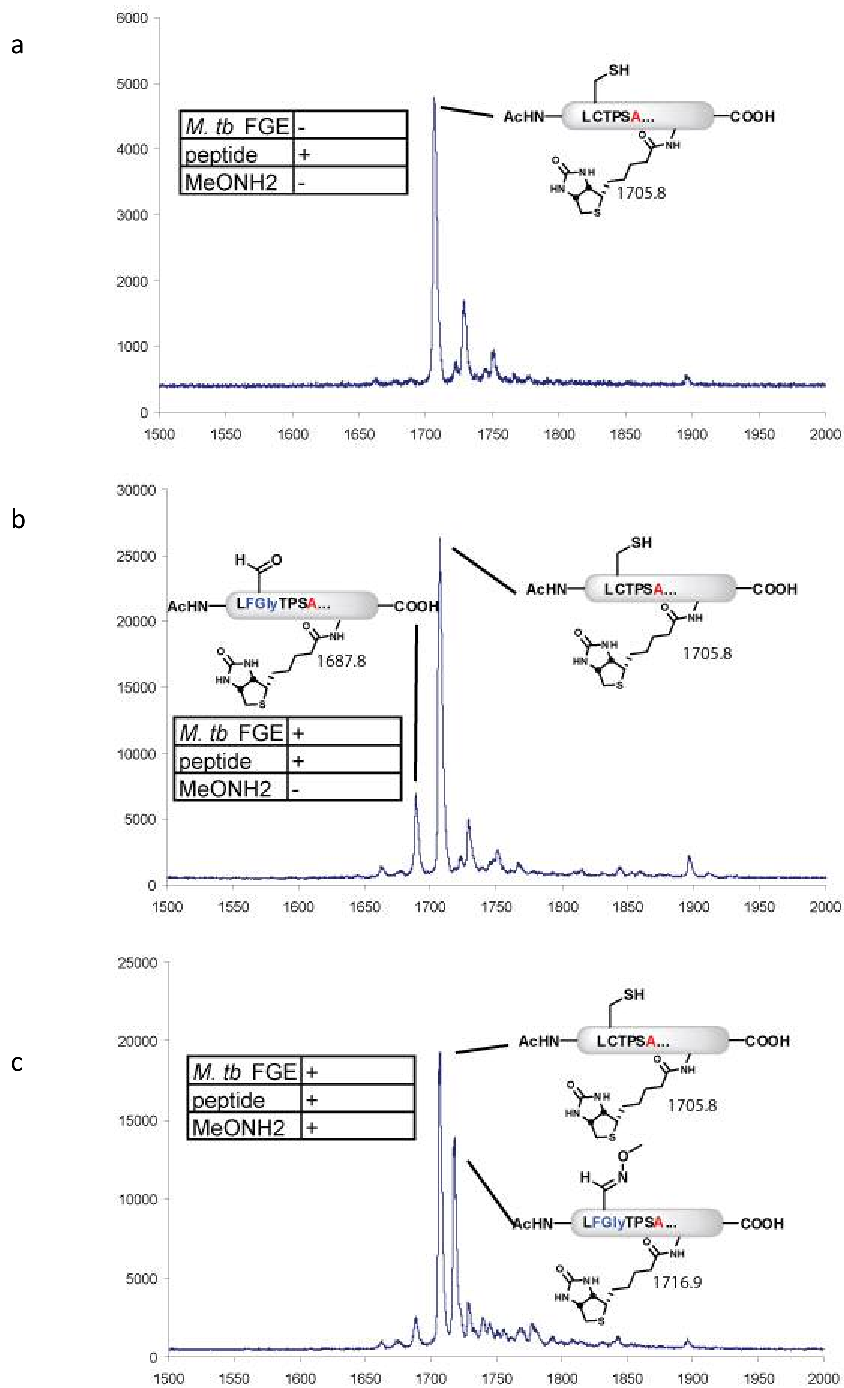


Figure S3. Mass spectra of LCTPSAGSLFTGR from FGE reactions before or after treatment with $MeONH_2$. (a) Peptide in the absence of FGE. (b) Peptide in the presence of FGE. (c) Peptide after reaction with $MeONH_2$

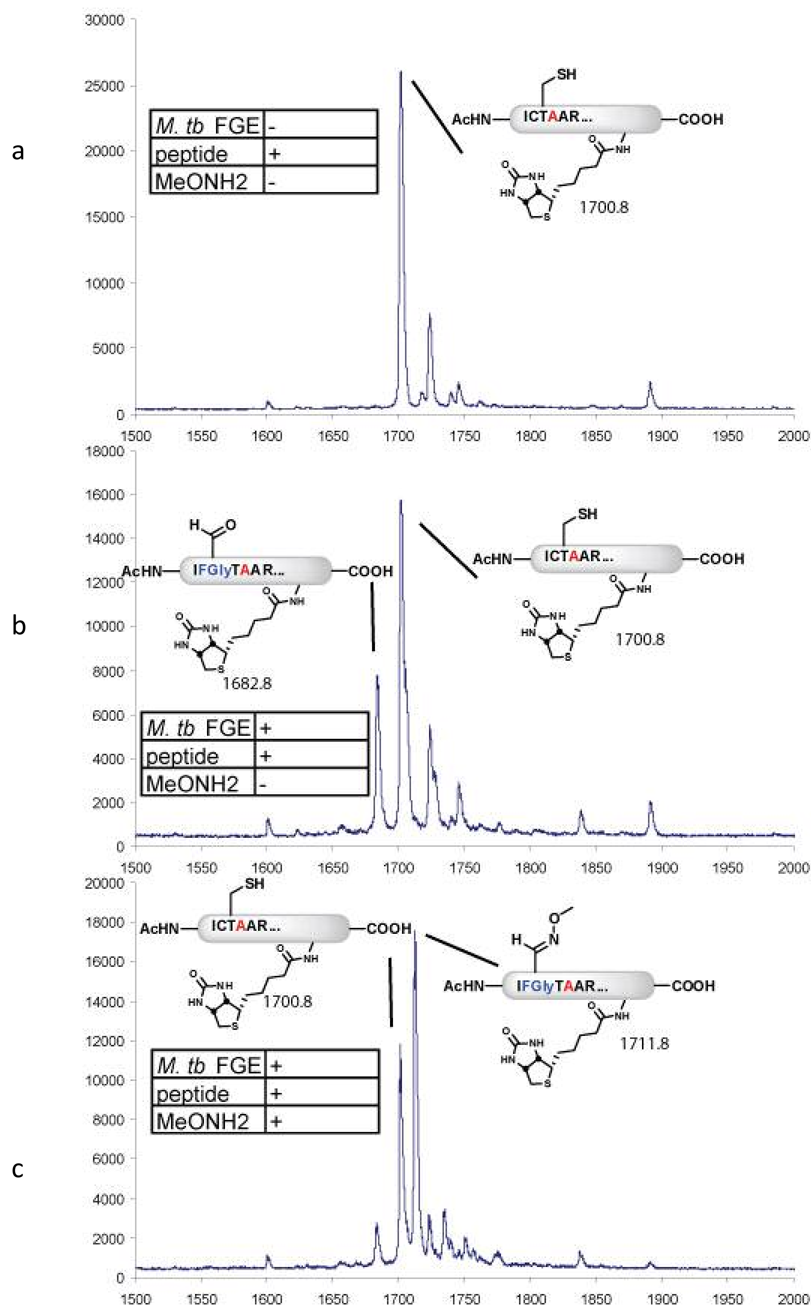
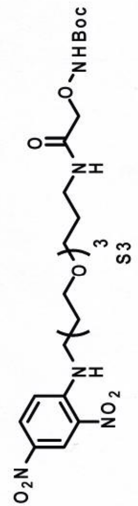
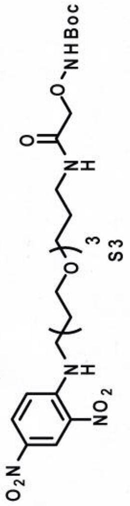
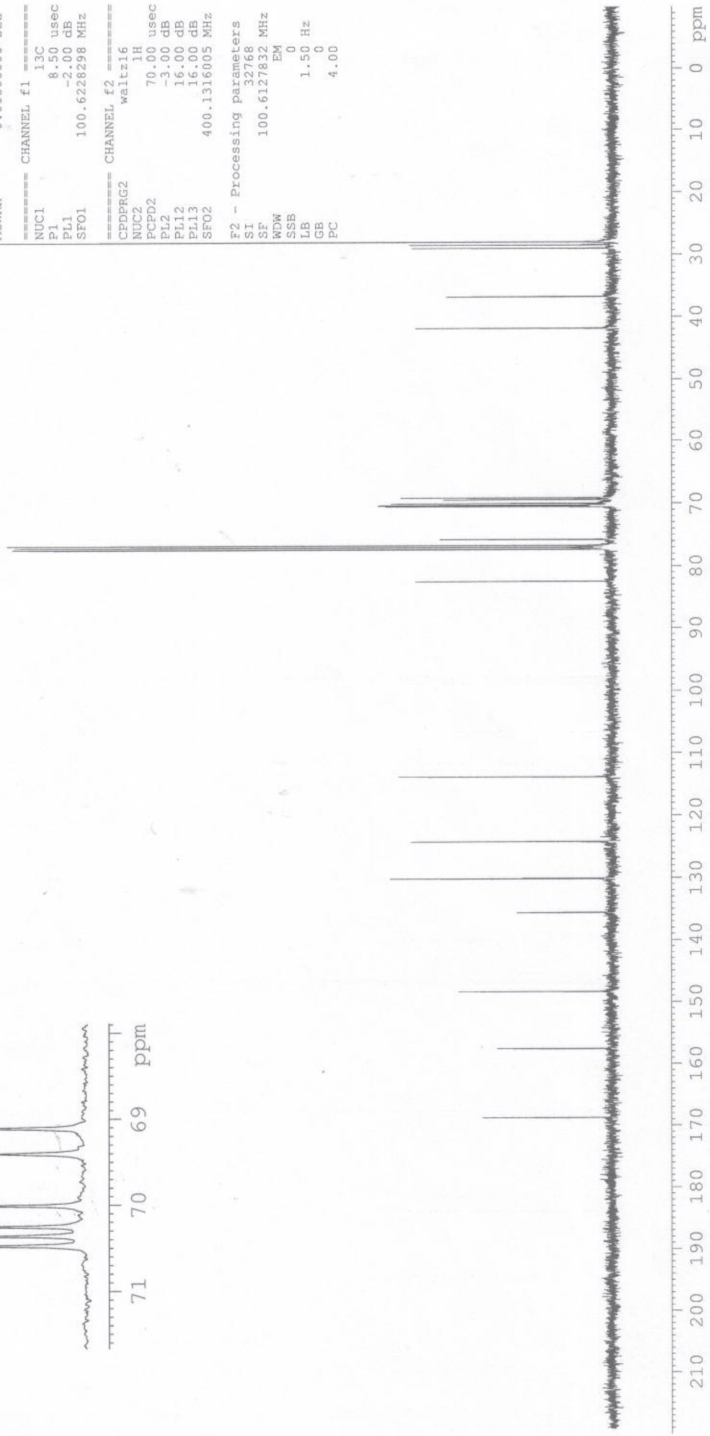


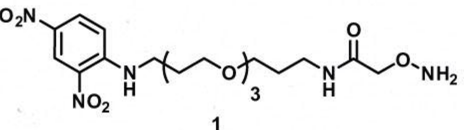
Figure S4. Mass spectra of ICTAARASLLTGQ from FGE reactions before or after treatment with $MeONH_2$. (a) Peptide in the absence of FGE. (b) Peptide in the presence of FGE. (c) Peptide after reaction with $MeONH_2$

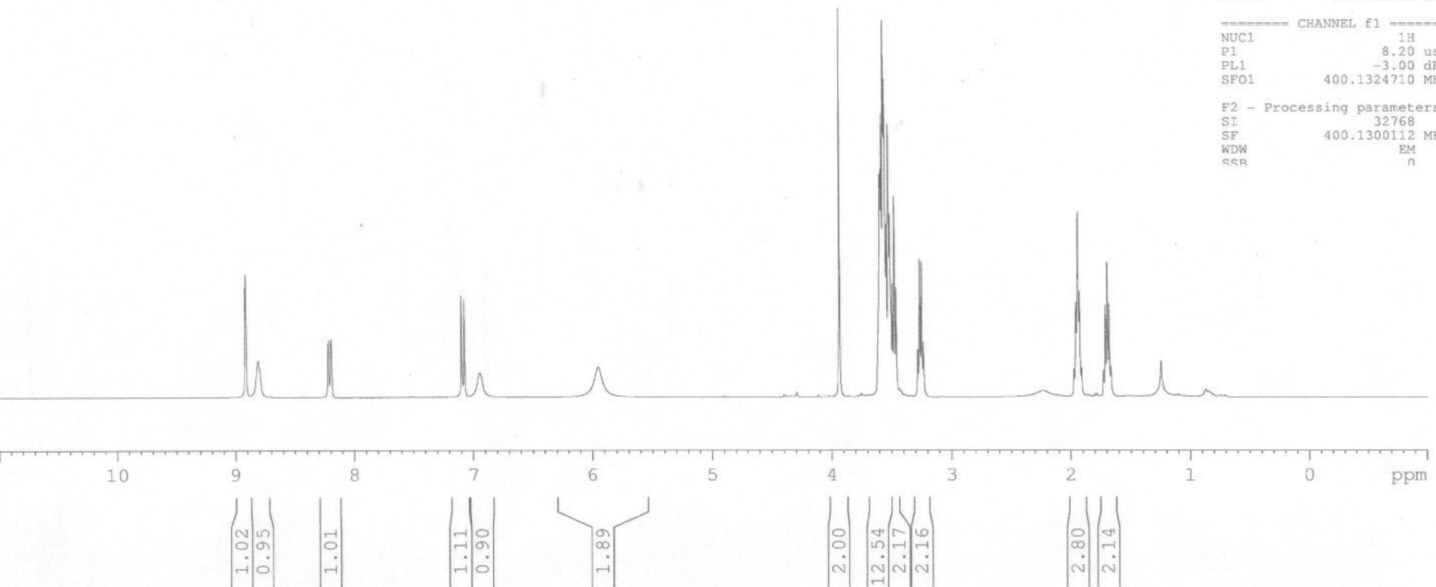
NMR spectra

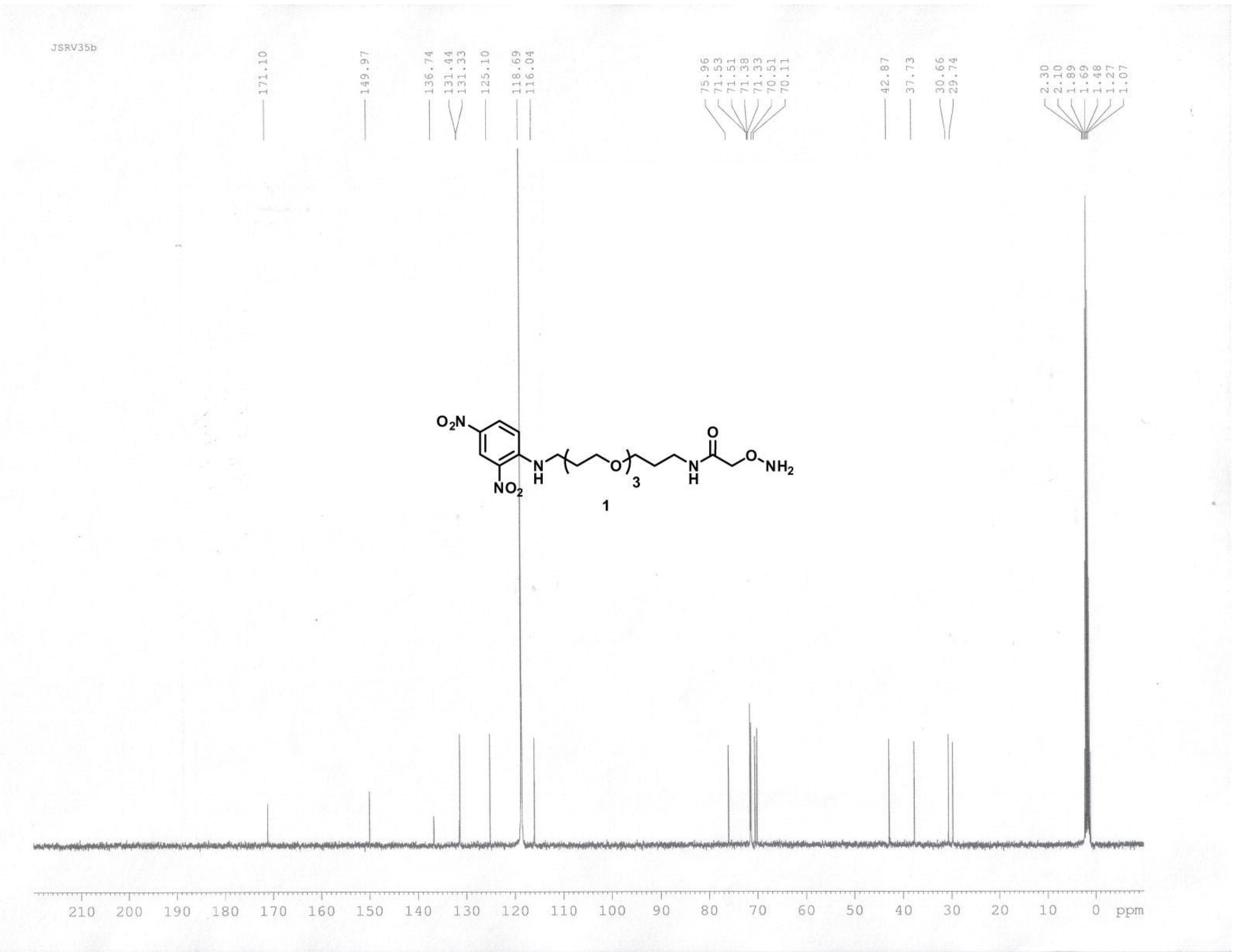







O2N
NO2
N
H
O
3
N
H
O
O
NH2
1


JSRV35b
8.921 8.914 8.810 8.225 8.218 8.200 8.194 7.101 7.077 6.945 6.560 5.952 4.901 4.401 4.378 4.296 4.116 4.024 3.937 3.854 3.754 3.679 3.611 3.596 3.590 3.582 3.569 3.557 3.541 3.526 3.512 3.491 3.476 3.461 3.435 3.419 3.350 3.281 3.264 3.249 3.233 3.075 2.233 1.971 1.955 1.940 1.926 1.911 1.877 1.860 1.846 1.817 1.798 1.783 1.728 1.712 1.696 1.680 1.665 1.536 1.375 1.248 1.107 0.869 0.852 0.745 0.707
Current Data Parameters
USER JSR
NAME JSRV34b
EXPNO 1
PROCNO 1
F2 - Acquisition Parameters
Date_ 20080109
Time 16.36
INSTRUM AVB-400
PROBHD 5 mm PABBO BB-
PULPROG zg30
TD 65536
SOLVENT CDCl3
NS 8
DS 0
SWH 8278.146 Hz
FIDRES 0.126314 Hz
AQ 3.9584243 sec
RG 45.3
DW 60.400 use
DE 6.00 use
TE 294.1 K
D1 1.00000000 sec
MCREST 0.00000000 sec
MCWRK 0.01500000 sec
======== CHANNEL f1 =======
NUC1 1H
P1 8.20 use
PL1 -3.00 dB
SFO1 400.1324710 MHz
F2 - Processing parameters
SI 32768
SF 400.1300112 MHz
WDW EM
SSB 0
10 9 8 7 6 5 4 3 2 1 0 ppm
1.02 0.95 1.01 1.11 0.90 1.89 2.00 12.54 2.17 2.16 2.80 2.14



JSRV35b
171.10
149.97
136.74
131.44
131.33
125.10
118.69
116.04
75.96
71.53
71.51
71.38
71.33
70.51
70.11
42.87
37.73
30.66
29.74
2.30
2.10
1.89
1.69
1.48
1.27
1.07
O_2N
N
H
O
3
N
H
O
O
NH_2
NO_2
1
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

Supplemental references

1. Carlson, B. L.; Ballister, E. R.; Skordalakes, E.; King, D. S.; Breidenbach, M. A.; Gilmore, S. A.; Berger, J. M.; Bertozzi, C. R. *J. Biol. Chem.* **2008**, *283*, 20117-20125.
2. Zheng, L.; Baumann, U.; Reymond, J. L. *Nucleic Acids Res.* **2004**, *32*, e115.
3. Rush, J.; Bertozzi, C. R. *Org. Lett.* **2006**, *8*, 131-134.
4. Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. *Nat. Chem. Biol.* **2007**, *3*, 321-322.