Addition of Isocyanide-containing Amino Acids to the Genetic Code for Protein Labelling and Activation

Yuda Chen[a⊥], Kuan-Lin Wu[a⊥], Juan Tang[a], Axel Loredo[a], Jordan Clements[a], Jingqi Pei[b], Zane Peng[b], Ruchi Gupta[a], Xinlei Fang[a], Han Xiao[abc]*

[a]Department of Chemistry, Rice University, 6100 Main Street, Houston, Texas, 77005, United States

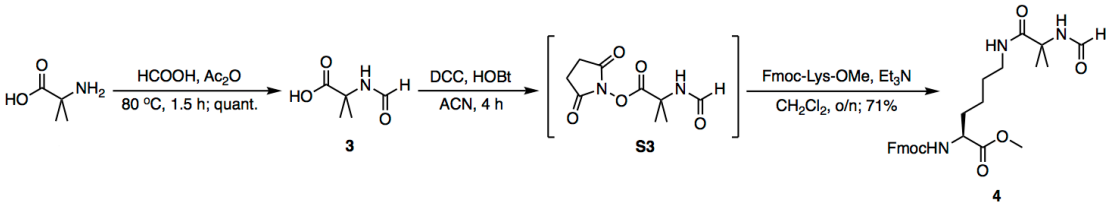
[b]Department of Biosciences, Rice University, 6100 Main Street, Houston, Texas, 77005, United States

[c]Department of Bioengineering, Rice University, 6100 Main Street, Houston, Texas, 77005, United States

Materials

LB agar and 2YT were ordered from BD Difco™. Isopropyl-β-D-thiogalactoside (IPTG) was purchased from Anatrace. 4-12% Bis-Tris gels for SDS-PAGE were purchased from Invitrogen. QuickChange Lightning Multi Site-Directed Mutagenesis Kit was purchased from Agilent Technologies (Cat. 210515). Oligonucleotide primers were purchased from Integrated DNA Technologies and Eurofins Genomics (Supplementary Table S1 lists the oligonucleotides used in this report). Plasmid DNA preparation was carried out with the GenCatch™ Plus Plasmid DNA Miniprep Kit and GenCatch™ Advanced Gel Extraction Kit. BugBuster™ Protein Extraction Reagent was purchased from Novagen (Cat. 70584). Protease inhibitor Cocktail was purchased from Biotool. Pierce™ universal nuclease was purchased from Thermo Scientific (Cat. 88700). Ni-NTA Agarose was obtained from Qiagen (Cat. 30230). Hoechst 33342 was purchased from Life Technologies™ (Cat. H3570). ATP and DRAQ5™ was obtained from Thermo Scientific. PolyJet transfection reagent was from SignaGen Laboratories. Anti-myc antibody was purchased from BioLegend and FITC-conjugated-anti-mouse antibody was from the Jackson Laboratory. 2-aminoisobutyric acid was obtained from Chem-Impex International. Unless otherwise mentioned, all solvents and chemicals for synthesis were purchased from Alfa Aesar and Fisher Chemical and used as received without further purification.

Synthesis of NCibK and NCK

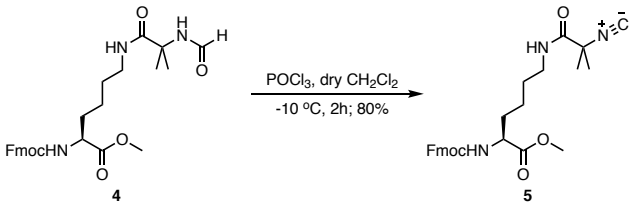


For ease of synthesis, the ester form of NCibK was used in this study. Esterification of the carboxyl group of the lysine derivatives was shown to increase their cellular uptake and intracellular concentration.[1]

2-aminoisobutyric acid (2 g, 19.4 mmol) in formic acid (24 mL) was added to acetic anhydride (12 mL, 127.2 mmol), and the mixture was stirred at 80°C for 2 h. The reaction was cooled down to room temperature and treated with water (10 mL). The reaction mixture was then evaporated, and dried under vacuum overnight to yield 3 as a white solid (2.6 g; quant.). The crude product was used for the next step without further purification. For calculated $[M+H]^+$= 132.1, found $[M+H]^+$= 132.1. To a mixture of N-formyl-2-methyl alanine 3(500 mg, 3.8 mmol), *N,N'*-dicyclohexylcarbodiimide (865 mg, 4.2 mmol) and *N*-hydroxysuccinimide (483 mg, 4.2mmol) were added to dry CH_3CN (25 mL) and stirred at room temperature for 3 h. The insoluble side product was removed by filtration, and the solvent was evaporated. The obtained white solid was then dissolved in CH_2Cl_2(50 mL), followed by the addition of Fmoc-Lys-OMeHCl (1438 mg, 3.4 mmol) and triethylamine (772 mg, 1.1 mL, 7.6 mmol). The solution was stirred at room temperature overnight. The organic solution was extracted with 1N HCl, dried over MgSO4, and evaporated under reduced pressure. The crude product was purified by column chromatography on silica (1-3% MeOH in CH_2Cl_2) to yield a white foam 4(1.3 g; 71%). For calculated $[M+H]^+$= 496.2, found $[M+H]^+$= 496.2.

^{1}H NMR (600 MHz, $CDCl_3$): δ 8.02 (s, 1H), 7.76 (d, J = 7.5 Hz, 2H), 7.61 (t, J = 7.0 Hz, 2H), 7.40 (t, J = 7.4 Hz, 2H), 7.31 (t, J = 7.4, 3H), 6.52 (s, 1H), 6.31 (s, 1H), 5.68 (d, J = 7.5 Hz, 1H), 4.46-4.35 (m, 2H), 4.33 (m, 1H), 4.21 (t, J = 6.9 Hz, 1H), 3.74 (s, 3H), 3.26 (q, J = 6.4 Hz, 2H), 1.84 (m, 1H), 1.71 (m, 1H), 1.57 (s, 6H), 1.53 (m, 2H), 1.37 (m, 1H).

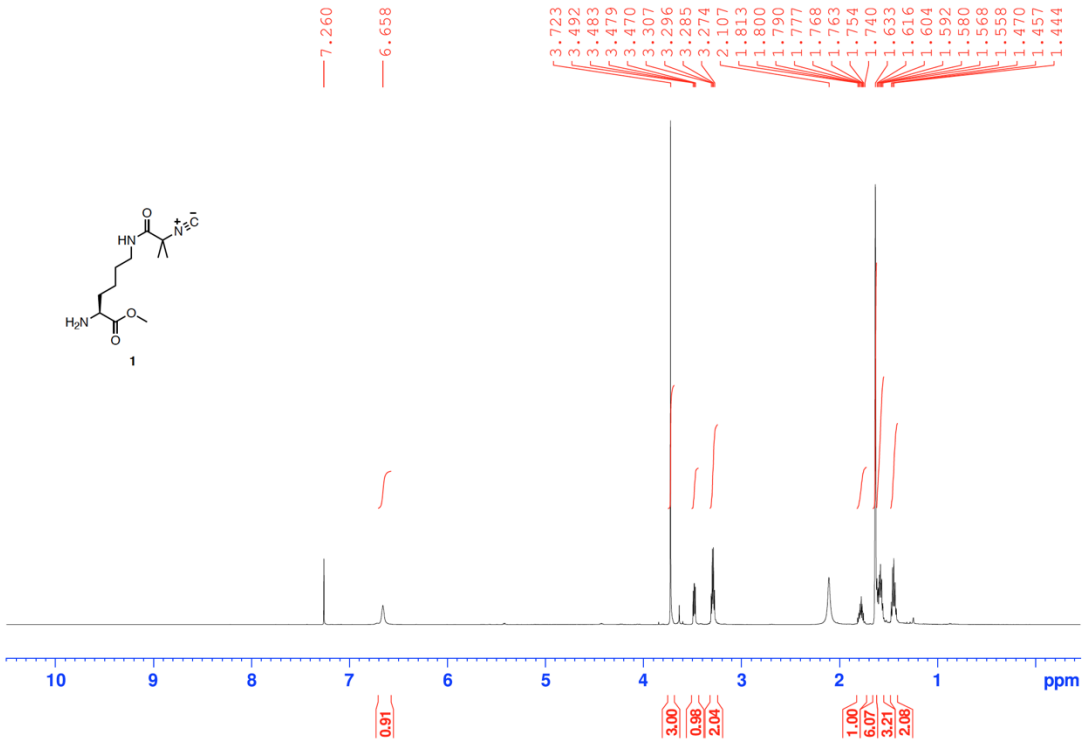
^{13}C-NMR (150 MHz, $CDCl_3$): δ 174.0, 172.9, 161.1, 156.1, 143.9, 143.7, 141.3, 127.7, 127.1, 125.1, 120.0, 66.9, 57.3, 53.7, 52.4, 47.2, 39.1, 31.8, 28.7, 25.2, 22.1.



4 (1.4 g, 2.8 mmol) was dissolved in dry CH_2Cl_2 (50 mL) under N_2 protection, cooled down by NaCl ice bath, and successively treated with triethylamine (1.7 g, 2.4 mL, 17.0 mmol) and phosphorous oxychloride (1.3 g, 0.8 mL, 8.5 mmol). After stirring for 2 h, the organic solution was washed by saturated $NaHCO_{3(aq)}$ (3 x 50 mL), dried over $MgSO_4$, evaporated under reduced pressure, and purified by column chromatography on silica (3% MeOH in CH_2Cl_2) to yield a white foam 5 (1.07 g; 80%). For calculated $[M+H]^+$ = 478.2, found $[M+H]^+$ = 478.2.

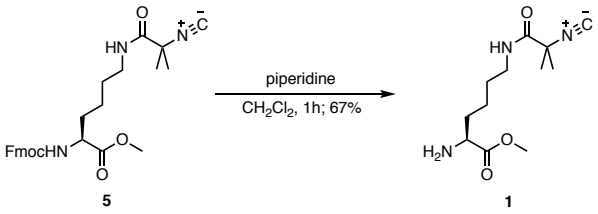
^{1}H NMR (600 MHz, $CDCl_3$): δ 7.77 (d, J = 7.6 Hz, 2H), 7.61 (d, J = 6.4 Hz, 2H), 7.40 (t, J = 7.5 Hz, 2H), 7.32 (t, J = 7.44 Hz, 2H), 6.56 (s, 1H), 5.37 (d, J = 7.9 Hz, 1H), 4.54-4.31 (m, 3H), 4.22 (t, J = 6.9 Hz, 1H), 3.76 (s, 3H), 3.30 (q, J = 6.28 Hz, 2H), 1.88 (m, 1H), 1.74 (m, 1H), 1.63 (s, 6H), 1.58 (m, 2H), 1.39 (m, 2H)

^{13}C-NMR (150 MHz, $CDCl_3$): δ 172.8, 169.1, 160.1, 156.0, 143.9, 143.8, 141.3, 127.7, 127.1, 125.1, 120.0, 67.0, 61.4 (t, J = 5.5 Hz), 53.6, 52.5, 47.2, 39.5, 32.1, 28.9, 27.8 (d, J = 4.3 Hz), 22.3

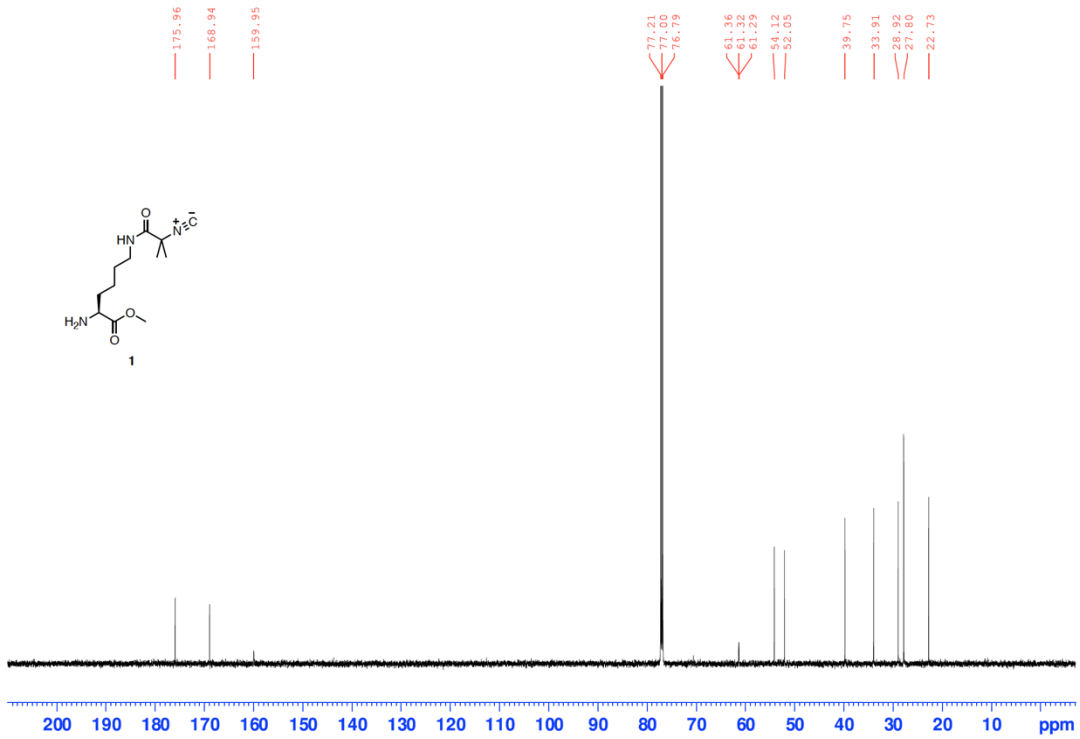


5 (1.06 mg, 2.2 mmol) was treated with 20% piperidine in CH_2Cl_2 (10 mL) for 1 h, and directly dried under reduced pressure. The crude product was purified by column chromatography (5% MeOH in CH_2Cl_2) to yield NCibK 1 as a colorless to slightly yellow oil (380 mg; 67%). For calculated $[M+H]^+$ = 256.2, found $[M+H]^+$ = 256.1.

^{1}H NMR (600 MHz, $CDCl_3$): δ 6.66 (s, 1H), 3.72 (s, 3H), 3.48 (dd, J = 7.6 Hz, J' = 5.5 Hz, 1H), 3.29 (q, J = 6.7 Hz, 2H), 1.78 (m, 1H), 1.67-1.53 (m, 9H), 1.44 (quintet, J = 7.7 Hz, 2H).

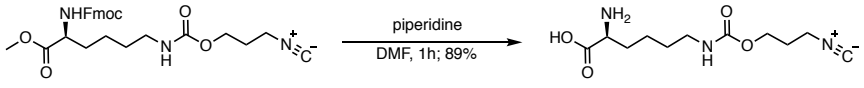


^{13}C-NMR (150 MHz, $CDCl_3$): δ 176.0, 169.0, 160.0, 61.3 (t, J = 5.5 Hz), 54.1, 52.0, 39.8, 33.9, 28.9, 27.8, 22.7.



3-isocyanopropyl-1-(4-nitrophenyl)carbonate 7 was synthesized following Tu, J. et al.'s procedure. The solution of 7 (280 mg, 1.1 mmol) in 3 mL DMF was added to N,N-diisopropylethylamine (542 mg, 0.73 mL, 4.2 mmol) followed by Fmoc-Lys-OMe•HCl 6 (535 mg, 1.3 mmol) and stirred at room temperature for 2 h. The DMF was removed under reduced pressure. The crude product was washed by water to remove residual Fmoc-Lys-OMe. The precipitant was filtered and dried under vacuum to get 8 as a white solid (375 mg; 68%). For calculated $[M+H]^+$ = 494.2, found $[M+H]^+$ = 494.2.

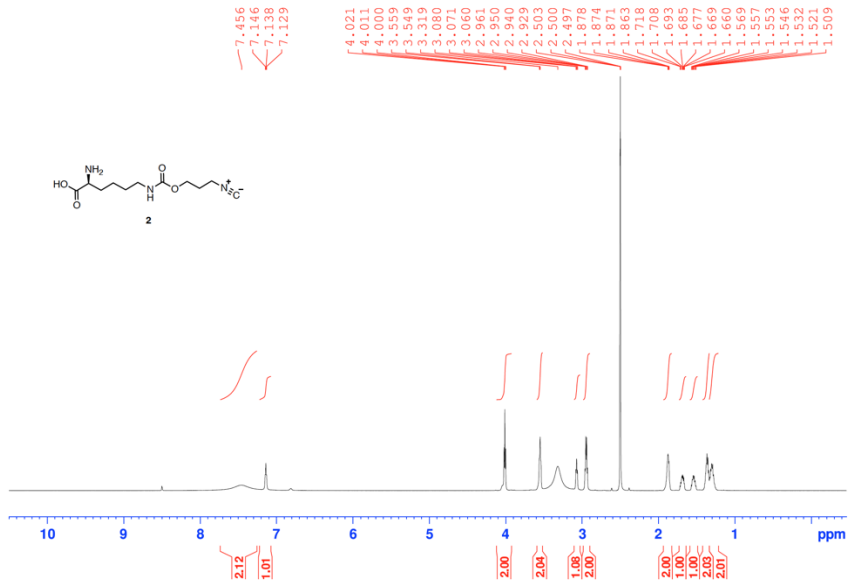
^{1}H NMR (600 MHz, $CDCl_3$): δ 7.77 (d, J = 7.6 Hz, 2H), 7.61 (t, J = 7.6 Hz, 2H), 7.41 (t, J = 7.4 Hz, 2H), 7.32 (t, J = 7.4 Hz, 2H), 4.54-4.33 (m, 3H), 4.27-4.15 (m, 3H), 3.76 (s, 3H), 3.45 (s, 2H), 3.18 (d, J = 6.1 Hz, 2H), 2.04 (s, 1H), 1.96 (m, 2H), 1.87 (m, 1H), 1.70 (m, 1H), 1.58 (m, 1H), 1.54 (m, 2H), 1.39 (m, 2H).



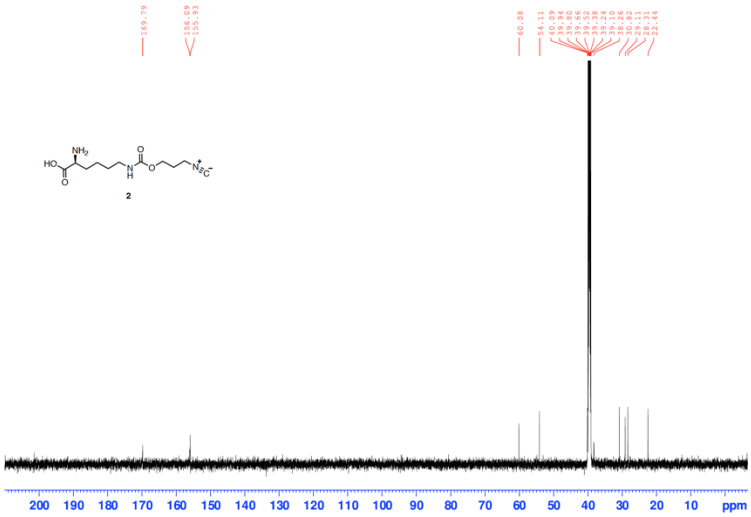
8 **2**

8 (375 mg, 0.76 mmol) was treated with 20% piperidine in DMF (0.5 mL) for 1 h and directly dried under reduced pressure. The crude product was washed by ether, filtered, and dried under vacuum to yield NCK 2 as a white solid (174 mg; 89%). For calculated $[M+H]^+$ = 258.1, found $[M+H]^+$ = 258.1.

^{1}H NMR (600 MHz, DMSO-d_6): δ 7.46 (br, 2H), 7.14 (t, J = 5.1 Hz, 1H), 4.01 (t, J = 6.2 Hz, 2H), 3.55 (t, J = 6.3 Hz, 2H), 3.07 (t, J = 5.9 Hz, 1H), 2.95 (q, J = 6.4 Hz, 2H), 3.55 (t, J = 6.3 Hz, 2H), 1.87 (m, 2H), 1.67 (m, 1H), 1.54 (m, 1H), 1.36 (m, 2H), 1.30 (m, 2H)



^{13}C-NMR (150 MHz, DMSO-d_6): δ 169.8, 156.1, 155.9, 60.1, 54.1, 40.1, 38.3, 30.8, 29.1, 28.3, 22.4.



Plasmid Construction

The ChPylRS(IPYE)/MbPylRS(IPYE)gene was amplified from pTech-chPylRS(IPYE) with forward primer PJ37 and reverse primer PJ38/PJ39, and inserted into pUltra plasmid using the NotI restriction sites to generate pUltra-ChPylRS(IPYE)/ pUltra-MbPylRS(IPYE) via Gibson assembly. To create pAcBac-ChPylRS(IPYE), ChPylRS(IPYE) was generated using PJ40 and PJ41 and inserted into NheI and NcoRI-digested pAcBac plasmid. The pNICE-LAP-neurexin-1beta mutant K6(LAP)TAG was generated by site directed mutagenesis using Da153.

Reaction of isocyanides with BODIPY-Tz

1 mM DMSO solution of BODIPY-*m*-Tz or BODIPY-*p*-Tz and 100 mM NCibK PBS buffer (pH=7.4) was diluted by PBS (pH=7.4) to the final concentration of 100 nM and 1000 uM, respectively. Fluorescence intensity scanning is measured at 0, 1, 30, 60, 120, 180, 240, 300, 360, 480, 540, 720 minutes with triple repetition (Ex 475 nm/ Em 485-600 nm). Turn-on fold is relative to the zero-minute fluorescence intensity. All the fluorescence intensities have been normalized by subtracting the background (cuvette with only PBS).

Rate Determination for NCibK Reaction with dpTz

NCibK & dp-Tz stock solution was prepared in PBS buffer (pH 7.4) and DMSO respectively. The reaction was initiated by mixing and diluting NCibK and dp-Tz stock solution with PBS buffer at a final concentration of 6 mM and 0.6 mM (6% v/v DMSO in PBS), respectively. NCibK at concentration 9 mM and 12 mM was also conducted. The formation of the product was tracked by measuring its absorption peak at 430 nm once per 10 second over the reaction timeframe by using a EVOLUTION 220 UV-Visible spectrophotometer (Thermo Scientific). Samples were measured in a 1 cm pathlength quartz cuvette at room temperature (25 ^{o}C). Reaction rates were obtained by fitting the exponential growth of absorption intensity as a pseudo first order reaction. All the data were performed with GraphPad Prism.

Protein Expression and Purification from *E. coli.*

E. coli. BL21(DE3) cells were co-transformed with protein expression plasmid(pET22b-T5-sfGFP151*/pBad-myo99TAG) and NCK/NCibK incorporation plasmids and grown in LB medium supplemented with or without 1 mM NCK/NCibK. Protein

expression was induced by the addition of isopropyl-β-D-thiogalactoside (IPTG) to a final concentration of 1 mM at OD 0.6 and cells were grown for an additional 20 h at 30 °C. Cells were harvested by centrifugation at 4,750 × g for 10 min. The cell pellets were suspended in lysis buffer (30 mM Tris-HCl buffer with 20 % sucrose, 1 mM EDTA, 0.2 mg/mL lysozyme, and 0.1‰ benzonase, pH 7.4) and lysed at 37 °C. The resulting cell lysate was clarified by centrifugation at 14,000 × g for 30 min, and protein was purified on Ni-NTA resin (Qiagen) following the manufacturer's instructions. LplA(W37V) was expressed and purified according to the Reference.[2]

Expression and Fluorescence Measurement of sfGFP

After sfGFP expression with the same method described above, 0.25 mL cells were harvested by centrifugation at 4,750 × g for 10 min and washed with PH 7.4 phosphate buffer. Then, cells pellets were re-suspended at 1ml PH 7.4 phosphate buffer and 200 µl was used to measure fluorescence (excitation/emission wavelength: 395/509 nm) and OD600.

Expression of EGFP at HeLa Cells

Adherent HeLa cells were cultured in DMEM medium supplemented with 10% FBS and 0.5% antibiotic−antimycotic at 37 °C in a humidified chamber with 5% CO_2. After seeding at 24-well plate with 80% confluency, Hela cells were transfected with pAcBac-EGFP* and pAcBac-chPyl (IPYE) using Polyject transfection reagent in the absence of ncAA. Medium was changed to new media with or without 0.75mM NCK/NCibK/BocK at 12 h post-transfection and images were taken at 60 h post-transfection.

Expression of EGFP at HEK293T Cells

Adherent HEK293T cells were cultured in the same conditions with the Hela cells. After seeding a 24-well plate with appropriate confluency, cells were transfected with pAcBac-EGFP* and pAcBac-chPyl(IPYE) using PolyJet transfection reagent in the presence or absence of ncAAs. Medium was changed to new media with or without 0.75mM ncAA at 12 h post-transfection and images were taken at 36 h post-transfection.

SfGFP-NCK reaction with dpTz

1mM dpTz was added to 100 µM sfGFP-151NCK in a PH 7.4 phosphate buffer at 37°C for 8 hours, followed by removal of excess dpTz using Amicon 5,000 molecular-weight-cutoff protein concentrator.

Decaging of neurexin-1beta at HEK293T Cells

HEK293T cells were transfected with pNICE-LAP-neurexin-1beta or co-transfected with pNICE-LAP-neurexin-1beta-TAG and pAcBac-ChPy1RS (IPYE) for 18 h. The cell culture was changed to new DMEM (10% FBS) containing 0.7mM NCK at 18h post transfection to allow protein expression for another 24 h. Before decaging, the cells were incubated in new DMEM media (10% FBS) at 37°C for 1h to remove a remaining NCK. The cells were treated with 1mM dpTz at PBS7.4 for 20 min, followed by incubation in DMEM media (10% FBS) containing 1mM pyrrolidine for 8 h.[3,4] After decaging, a LAP assay was performed by incubating cells with 10uM ^{W37V}LplA, 50 uM coumarin substrate, 1mM ATP and 2mM $Mg(OAc)_2$in 50mM Na_2HPO_4pH 7.2, followed by cell fixation with pre-cooled methanol. Myc tag immunofluorescence staining was performed by applying 0.1 ug/mL anti-myc antibody to cells at 37°C for 2 h. After washing, cells were incubated with FITC-conjugated secondary antibody (7.5ug/ml) for 1 h. Nucleus staining was performed right before imaging by incubating cells with 5uM DRAQ5 for 10 min.

Oligonucleotide	Sequence (5’-3’)
Da153	ggaggacctgggcttcgagatcgac**TAG**gtgtggcacgacttccccgccacta
PJ37	TTCACAAAGGAGGTGCGGCCGCatggataagaagccgctggatgttctg
PJ38	GAGACCGTTTAAACGCGGCCGCTTACAGGTTGGTAGAAATCCC GTTATAGTAAGAC
PJ39	GAGACCGTTTAAACGCGGCCGCTTACAGGTTGGTAGAGATACC GTTGTAGTAAG
PJ40	AGGGAGACCCAAGCTGGCTAGCGCCACCatggataagaagccgctggatgttctg
PJ4	gtcgacttaacgcgttgaattcTTACAGGTTGGTAGAAATCCCGTTATAGTAAGAC

Table S1. DNA oligomers

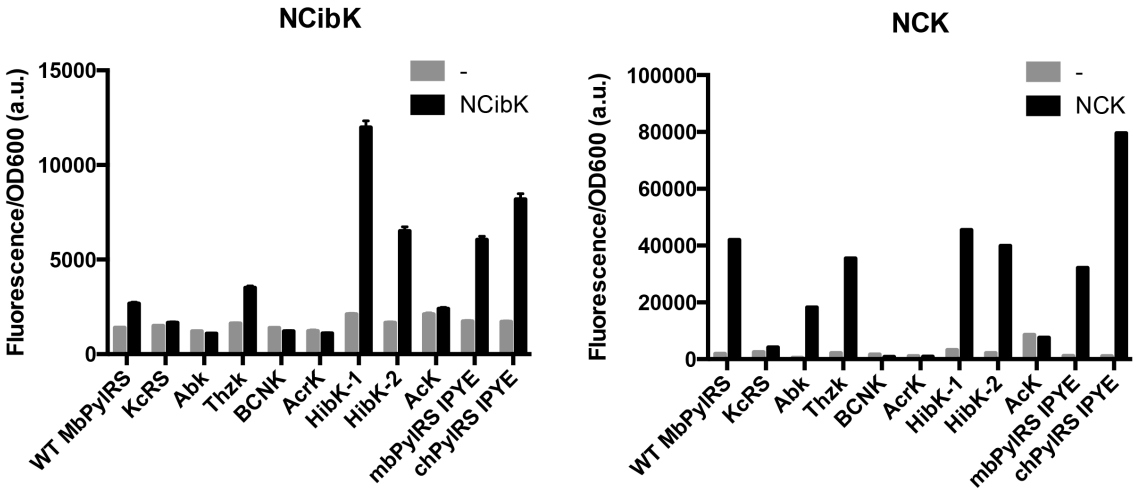


Figure S1. Screening of PylRS mutants for NCibK and NCK incorporation.

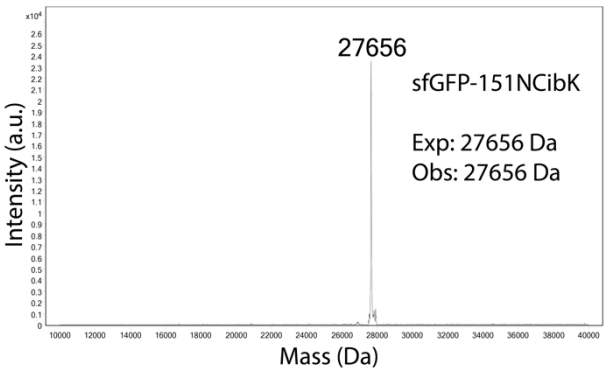


Figure S2. ESI-MS analysis of isolated sfGFP-NCibK expressed from M9-glucose minimal medium.

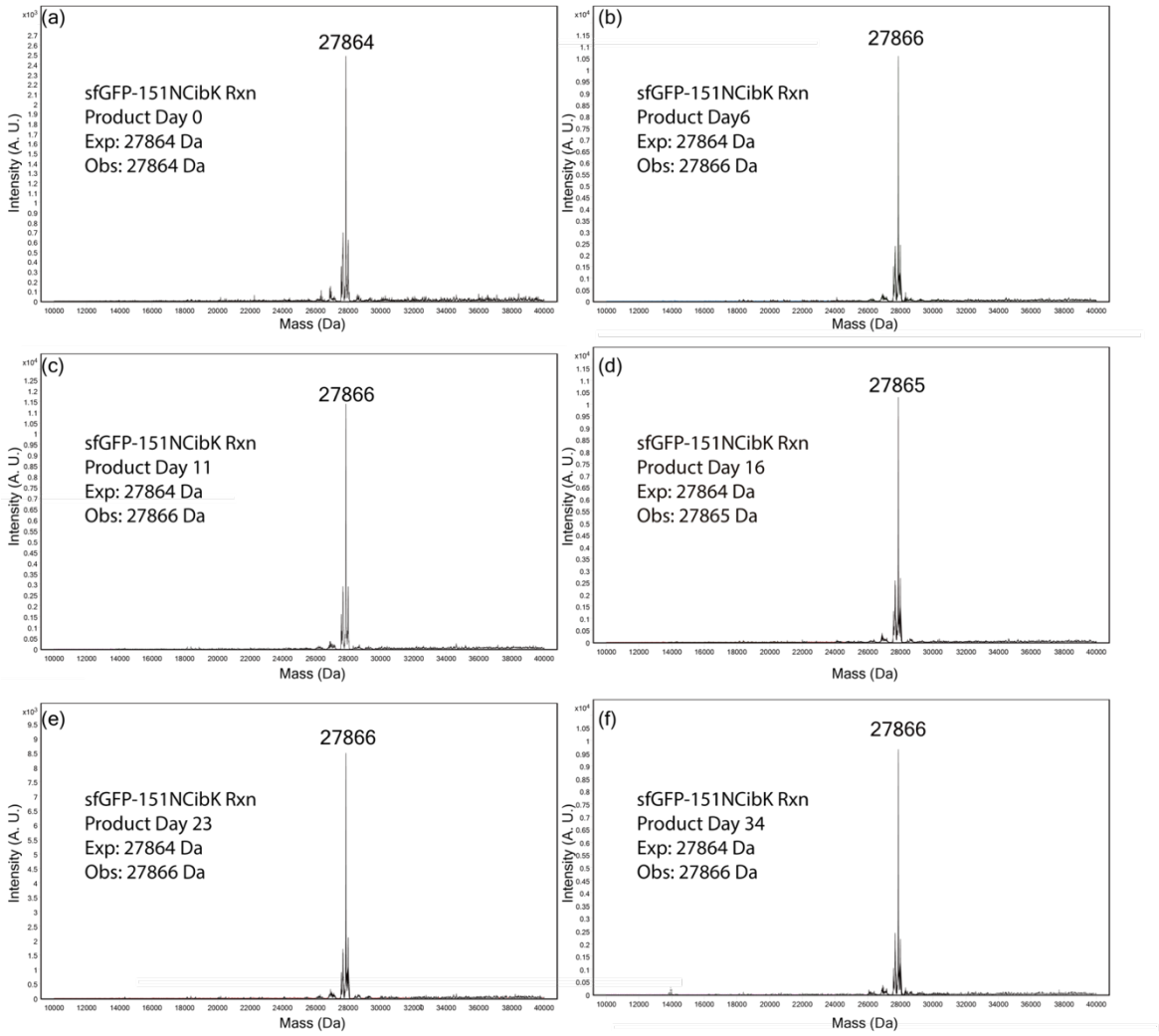


Figure S3. Stability of sfGFP-NCibK conjugate with dpTz.

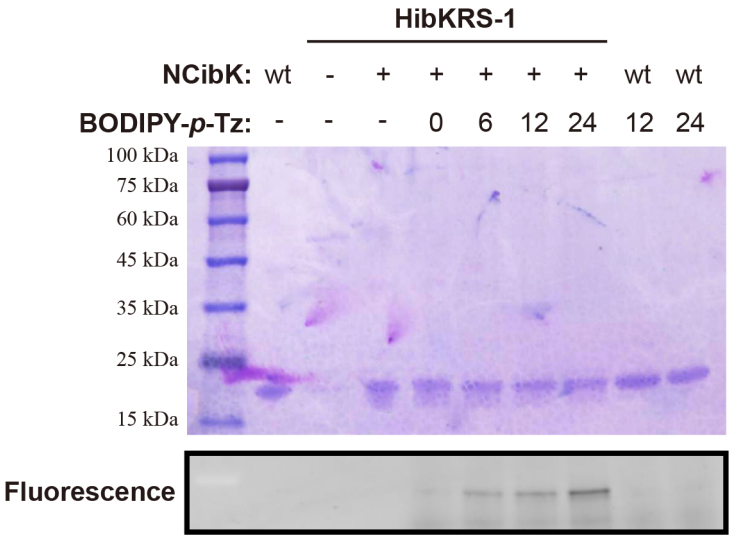


Figure S4. Incorporation of NCibK into Myoglobin.SDS-PAGE analysis of NCibK-Myoglobin labelled with BODIPY-*p*-Tz in different reaction time (0, 6, 12, 24 h). Fluorescence signals showed that BODIPY-*p*-Tz only specifically labelled NCibK-Myoglobin compared to WT-Myoglobin.

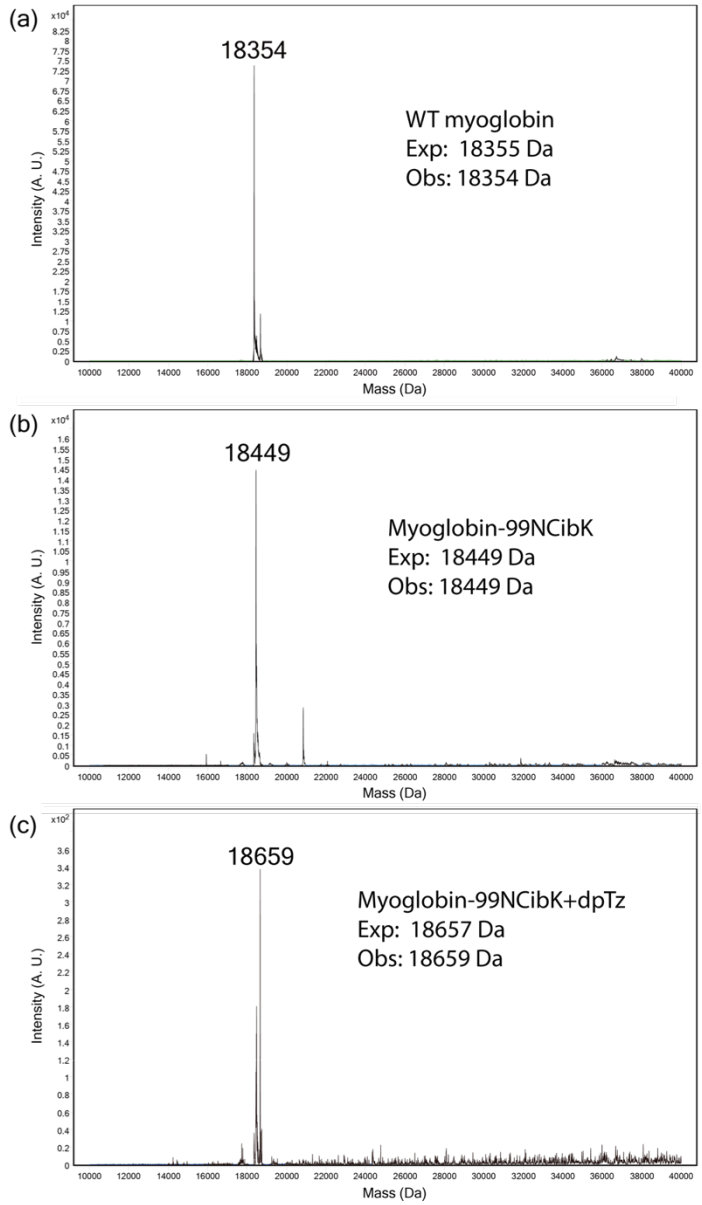


Figure S5. Incorporation of NCibK into Myoglobin

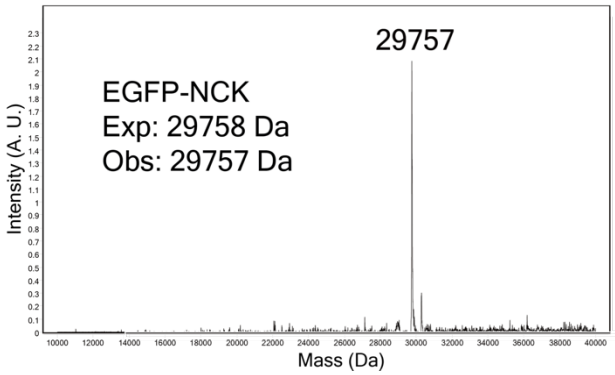


Figure S6. Incorporation of NCK into EGFP in HEK293T cells.

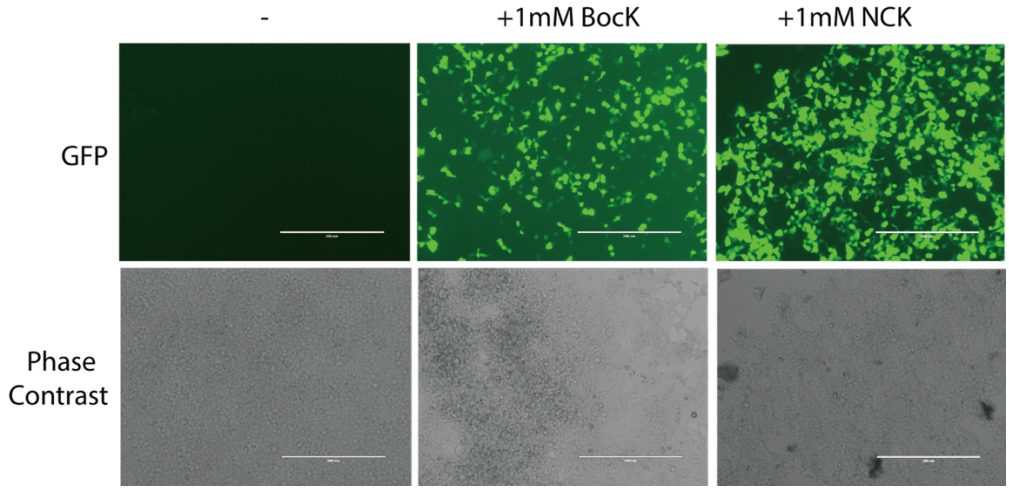


Figure S7. Incorporation of BocK, NCK into EGFP-Y40TAG in HEK293T Cells analyzed by fluorescence microscopy in the presence of or absence (-) of ncAAs.

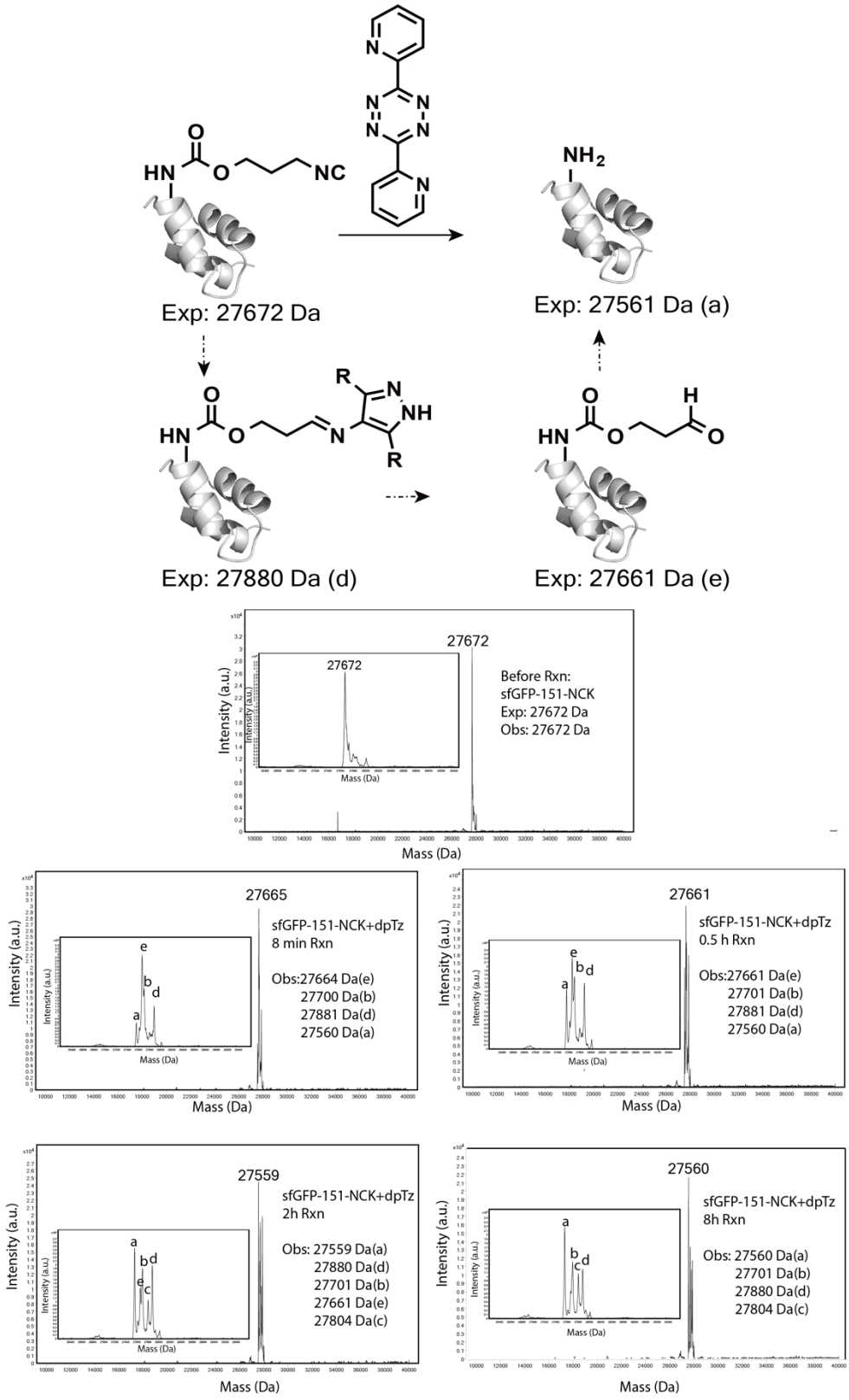


Figure S8. Time-dependent study of tetrazine-mediated sfGFP-NCK decaging. Peaks a, d, and e, represent the final decaged product (27560 Da), pyrazole-imine intermediate (27881 Da), and 3-oxopropyl-caged intermediate (27661 Da), respectively. Peaks c and d are likely additive products derived from the aldehyde intermediate (peak e, 27661 Da).

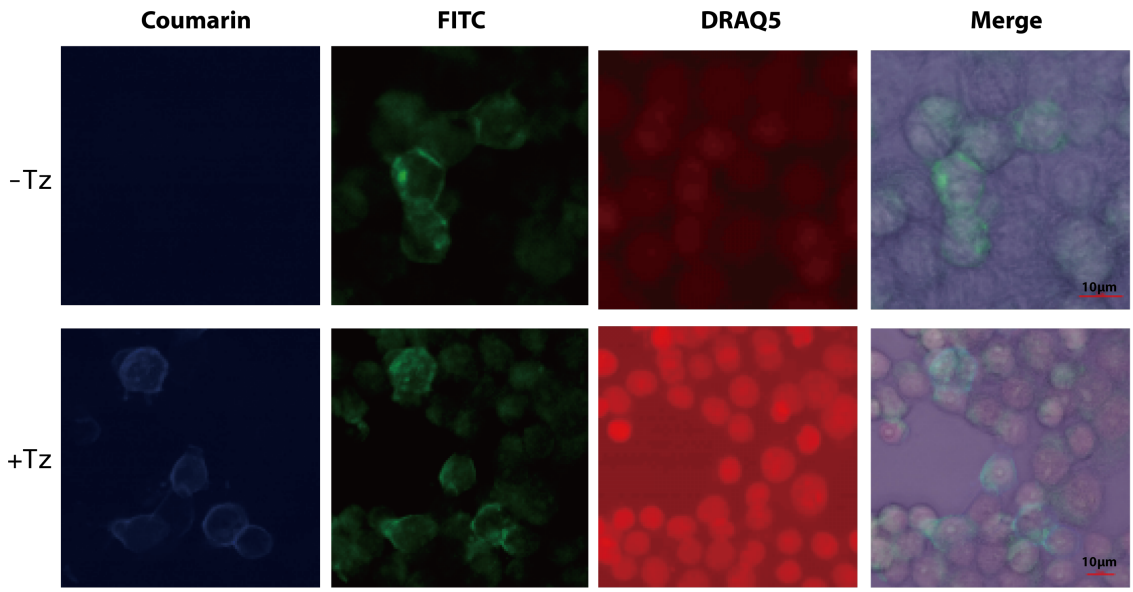


Figure S9. Tetrazine-mediated LAP tag activation in living cells. HEK293T cells expressing NCK-containing NRX-LAP were mixed with PBS with (+Tz) or without (-Tz) 1 mM dpTz in for 20 mins, followed by the incubation with 1 mM pyrrolidine for 8 h. After PRIME labeling (blue) and cell fixation, 0.1 μg/mL anti-Myc antibody for 1 h and 7.5 μg/mL FITC-conjugated secondary antibody (green) were added to the cells to check the expression of NRX-LAP. Nucleus staining was carried out with 5 μM DRAQ5 (red) right before confocal imaging. Scale bars = 10 μm.

References

1 X. Bi, K. K. Pasunooti, A. H. Tareq, J. Takyi-Williams and C.-F. Liu, *Organic & Biomolecular Chemistry*, 2016, **14**, 5282–5285.
2 C. Uttamapinant, K. A. White, H. Baruah, S. Thompson, M. Fernández-Suárez, S. Puthenveetil and A. Y. Ting, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 10914–10919.
3 A. K. Leslie, D. Li and K. Koide, *J. Org. Chem.*, 2011, **76**, 6860–6865.
4 A. L. Garner, C. M. S. Croix, B. R. Pitt, G. D. Leikauf, S. Ando and K. Koide, *Nature Chem*, 2009, **1**, 316–321.