Supplementary Material

A Genetically Encoded Aldehyde for Rapid Protein Labelling

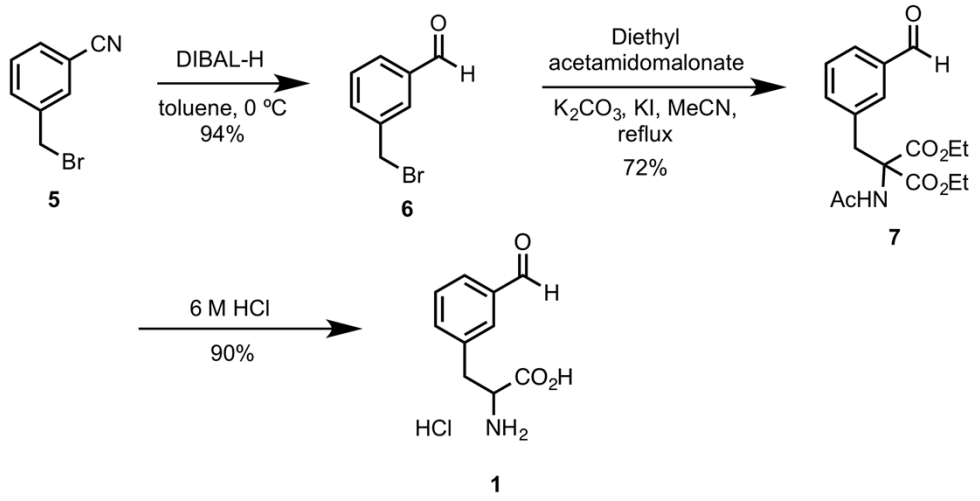
Alfred Tuley,[a] Yan-Jiun Lee,[a] Bo Wu,[a] Zhiyong U. Wang,[b] and Wenshe R. Liu*[a]

[a]Department of Chemistry, Texas A&M University, College Station, TX 77843
[b]Department of Chemistry and Physics, Troy University, Troy, Alabama, 36081
E-mail: wliu@chem.tamu.edu

Table of Contents

General Experimental Conditions and Equipment:
Organic synthesis was performed in oven-dried glassware under an argon atmosphere, and reagents were purchased and used without further purification. NMR spectra were obtained using Inova-300, Mercury-300, and Inova-500 instruments. Centrifugation was performed using a Sorvall Superspeed RC2-B automatic refrigerated centrifuge. Cell lysis was achieved with a VWR Scientific Branson Sonifier 450. Fluorometer data was obtained on an instrument purchased from Photon Technology International, Inc.; data from fluorometric studies were processed using Kaleidograph. Gel imaging was achieved with a Bio-Rad ChemiDoc™ XRS+ gel imager using ImgaeLab™ software. Protein yield was determined using a Shimadzu UV-1800 spectrophotometer, using the sfGFP molar extinction coefficient of 83300 $M^{-1}cm^{-1}$.[1]

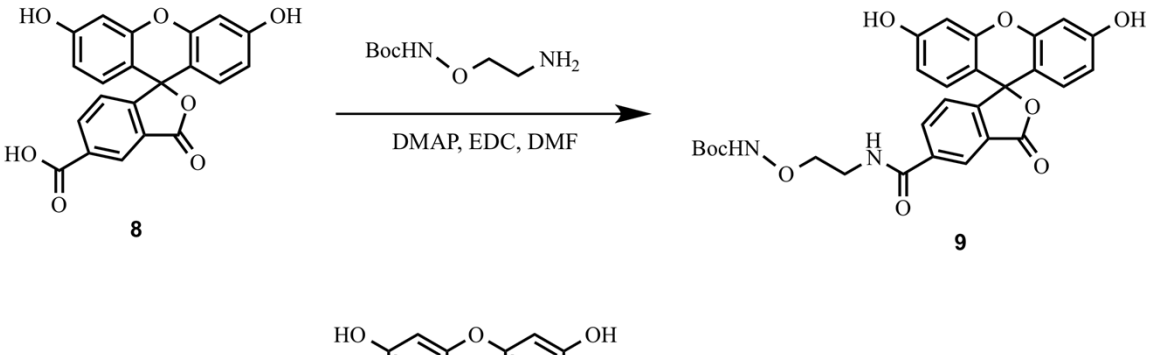
1. Organic Synthesis



1. 1. 1. 3-(bromomethyl)benzaldehyde (6):
Commercially available alpha-bromo *m*-toluinitrile **5** (5 g, 25.51 mmol) was dissolved in toluene and cooled to 0 ºC under an argon atmosphere. Diisobutylaluminum hydride (35.96 ml, 35.96 mmol, 1 M/Hexanes) was added dropwise to the solution, upon which the white opaque solution became clear and yellow. After 2 hrs, 66 ml of $CHCl_3$ was added to the solution, followed by 200 ml 10% HCl. After stirring for an additional hour, the organic layer was separated and washed with distilled water, then dried with Na_2SO_4 and concentrated to yield a clear oil, which crystallized overnight. The white crystals were filtered and washed with ice-cold hexanes, then allowed to dry to yield the desired compound in 94% yield (4.77 g). Spectroscopic data was in agreement with literature values.[2] ^{1}H NMR (300 MHz, $CDCl_3$) δ 10.03 (s, 1H), 7.91 (s, 1H), 7.83 (d, J = 7.5 Hz, 1H), 7.68 (d, J = 7.5 Hz, 1H), 7.54 (t, J = 7.5 Hz, 1H), 4.55 (s, 2H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 191.6, 138.9, 136.8, 134.9, 129.7, 32.1.

1. 1. 2. Diethyl 2-acetamido-2-(3-formylbenzyl)malonate (7):
Alpha-bromo aldehyde **6** (4.52 g, 22.72 mmol), diethyl acetamidomalonate (4.44 g, 20.45 mmol), K_2CO_3 (6.28 g, 45.44 mmol), and KI (3.77 g, 22.72 mmol) were suspended in 142 ml of anhydrous MeCN and heated to reflux. After 15 hrs, the reaction was cooled to room temperature, filtered, and concentrated. The resulting crude residue was dissolved in boiling hexanes with minimal amounts of ethyl acetate; a yellow solid was obtained upon cooling, which was washed with ice-cold hexanes and dried *in vacuo* to yield **7** in 72% yield (5.4 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 9.97 (s, 1H), 7.76 (d, *J* = 7.5 Hz, 1H), 7.54 (s, 1H), 7.44 (t, *J* = 7.5 Hz, 1H), 7.29 (d, *J* = 7.5 Hz, 1H), 6.53 (s, 1H), 4.28 (q, *J* = 6.9, 7.2 Hz, 4H), 4.12 (q, *J* = 7.2 Hz, 1H), 3.75 (s, 2H), 2.05 (s, 3H), 1.31 (t, *J* = 7.2 Hz, 6H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 191.9, 169.3, 167.3, 136.5, 135.9, 130.3, 129.2, 129.0, 67.0, 62.9, 37.4, 23.0, 14.0.

1. 1. 3. 2-amino-3-(3-formylphenyl)propanoic acid hydrochloride (1):
m-Formyl malonate **7** from the previous reaction (5 g, 14.91 mmol) was dissolved in 51.4 ml of 6 M HCl, and the resulting suspension was refluxed for 33 hrs. The resulting solution was cooled to room temperature and the solvent was removed *in vacuo*, then the solid product was filtered and washed with diethyl ether to remove any organic impurities. The resulting off-white solid was lyophilized overnight to yield the title compound in 90% yield (3.08 g). ^{1}H NMR (300 MHz, D_2O) δ 9.79 (s, 1H), 7.78 (dt, *J* = 1.8, 6.9 Hz, 1H), 7.72 (s, 1H), 7.46-7.55 (m, 2H), 4.19 (t, *J* = 6.9 Hz, 1H), 3.25 (dd, *J* = 6.3, 14.7 Hz, 3H). ^{13}C NMR (75 MHz, D_2O) δ 195.9, 171.2, 136.1, 136.0, 135.1, 130.3, 129.8, 129.7, 53.9, 35.2.



8 9 4

1. 2. 1. *tert*-butyl (2-(3',6'-dihydroxy-3-oxo-3*H*-spiro[isobenzofuran-1,9'-xanthene]-5-carboxamido)ethoxy)carbamate (9). To a solution of 5-carboxyfluorescein (0.10 g, 0.27 mmol), 4-(dimethylamino)pyridine (10 mg, 0.08 mmol) and *tert*-butyl (2-

aminoethoxy)carbamate[3] (53 mg, 0.30 mmol) in anhydrous DMF (2 mL) was added *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC hydrochloride, 78 mg, 0.41 mmol), and the mixture was stirred at room temperature overnight. The mixture was diluted in ethyl acetate (50 mL), washed with sodium hydroxide (0.5 M, 10 mL), hydrochloric acid (0.5 *N*, 10 mL) and brine (10 mL), dried (Na_2SO_4), evaporated, and chromatographed (EtOAc/hexanes, 1:3 to 1:1) to give **9** (0.10 g, 70%) as a yellow oil. ^{1}H NMR (CD_3OD, 500 MHz) δ 8.95 (t, 1 H, J = 5.5 Hz), 8.49 (s, 1 H), 8.24 (dd, 1 H, J = 8.0, 1.5 Hz), 7.30 (d, 1 H, J = 8.0 Hz), 6.69 (d, 2 H, J = 2.0 Hz), 6.59 (d, 2 H, J = 9.0 Hz), 6.53 (dd, 2 H, J = 9.0, 2.5 Hz), 3.99 (t, 2 H, J = 5.2 Hz), 3.67 (t, 2 H, J = 5.0 Hz), 1.47 (s, 9 H); ^{13}C NMR (CD_3OD, 75 MHz) δ 170.7, 168.4, 161.5, 159.8, 156.9, 154.1, 137.8, 135.7, 130.3, 128.8, 125.8, 125.0, 113.8, 111.0, 103.8, 82.7, 75.9, 39.9, 28.7; HRMS (ESI) calcd for $C_{28}H_{27}N_2O_9$ ($[M+H]^+$) 535.1717, found 535.1234; calcd for $C_{28}H_{26}N_2O_9Na$ ($[M+Na]^+$) 557.1536, found 557.1028.

1. 2. 2. *N*-(2-(aminooxy)ethyl)-3',6'-dihydroxy-3-oxo-3*H*-spiro[isobenzofuran-1,9'-xanthene]-5-carboxamide hydrochloride (4). To a solution of **9** (70 mg, 0.13 mmol) in 1,4-dioxane (1.0 mL) was added hydrogen chloride in dioxane (4.0 M, 0.3 mL, 1.2 mmol), and the mixture was stirred at room temperature for 4 h. The solvent was evaporated under high vacuum to afford **4** (62 mg, quant.) as a yellow oil. ^{1}H NMR (CD_3OD, 500 MHz) δ 8.78 (s, 1 H), 8.35 (s, 1 H), 7.57 (s, 1 H), 7.32 (bs, 2 H), 7.25 (bs, 2 H), 7.08 (bs, 2 H), 4.30 (bs, 2 H), 3.81 (bs, 2 H); HRMS (ESI) calcd for $C_{23}H_{19}N_2O_7$ ($[M+H]^+$) 435.1192, found 435.1292.

2. Superfolder Green Fluorescent Protein (sfGFP) Expression:

2. 1. sfGFPS2X Protein Sequence:

MAXKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKL
PVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYK
TRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIK
ANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRD
HMVLLEFVTAAGITHGMDELYKGSHHHHHH

X Denotes an amber stop codon where the non-canonical amino acid is incorporated.

2. 2. sfGFPN149X Protein Sequence:

MVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKL
PVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYK
TRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHXVYITADKQKNGIK
ANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRD
HMVLLEFVTAAGITHGMDELYKGSHHHHHH

X Denotes an amber stop codon where the non-canonical amino acid is incorporated.

2. 3. sfGFPN149TAG Primers Sequences:
sfGFPN149TAG-F: 5' - TAG GTG TAT ATT ACC GCC GAT AAA CAG AAA AAT GG - 3'
sfGFPN149TAG-R: 5' - ATG GCT GTT GAA ATT ATA TTC CAG TTT ATG - 3'

2. 4. Construction of pBAD-sfGFPN149TAG:
pBAD-sfGFP was a generous gift from Dr. Ryan Mehl at Franklin & Marshall College. sfGFP149TAG-F and sfGFP149TAG-R were used to perform the quick-change PCR on pBAD-sfGFP to afford pBAD-sfGFPN149TAG. This construct was cotransformed with pEVOL-PylT-PylRS-N346A/C348A into Top10 *E. coli* cells and used for sfGFP expression.

2. 5. Genetic Incorporation of *m*-Formyl Phenylalanine Into Superfolder GFP:
The procedure for expressing sfGFPS2TAG using BL21(DE3) *E. coli* has been reported previously.[4,5] For expressing sfGFPN149TAG, Top10 *E. coli* cells containing pBAD-sfGFPN149TAG and pEVOL-PylT-PylRS-N346A/C348A were used to inoculate 5 ml of LB media supplemented with Cm (34 μg/ml) and Amp (100 μg/ml), then grown overnight, followed by transfer to 500 ml of LB media containing Cm (34 μg/ml) and Amp (100 μg/ml). When the cells had reached $O.D._{600}$ 1.3-1.6, the culture was spun down for 20 min at 4,000 rpm, then the resulting pellet was resuspended with water and spun down once more. The pellet from this centrifugation was resuspended in 15 ml ddH_2O, and 5ml of this suspension was added to a flask containing 45 ml minimal media (33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 9.4 mM NH_4Cl, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 1% glycerol) supplemented with 2 mM non-canonical amino acid (NCAA). Induction occurred via addition of 0.2% arabinose and 1 mM IPTG. After 12 hrs, the culture was centrifuged for 20 min at 4,000 rpm. The resulting pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8) and sonicated, then centrifuged for 1 hr at 4,000 rpm. The supernatant treated with imidazole to a final concentration of 10 mM, then incubated at 4 ºC for 1 h with Ni^{2+}-NTA resin. The resin was washed with 10 mM imidazole (3 x column volume) and 20 mM imidazole (3 x column volume), and then eluted with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 8). The protein was dialyzed three times with 20 mM Tris buffer, pH 8.1, and concentrated using Amicon Ultra-cel 10k centrifugal filter units. Protein purity was determined via SDS-PAGE analysis and subjected to ESI-MS.

<u>Incorporation Using Autoinduction Media</u>:

Proteins were expressed in a manner similar to the protocol described above, the exception being the use of autoinduction media developed by Mehl and co-workers.[6] The recipe used was identical to the referenced protocol except a different amino acid solution was utilized (400 μg/ml each of the following: glutamic acid sodium salt, aspartic acid, lysine hydrochloride, arginine hydrochloride, histidine hydrochloride monohydrate, alanine, proline, glycine, threonine, serine, glutamine, asparagine monohydrate, valine, leucine, isoleucine, tryptophan, and methionine). In other words, this 17 x amino acid

solution differs from the reported recipe in that phenylalanine is removed to eliminate potential misincorporation. The cultures were allowed to grow overnight until a constant O.D. was obtained (1.82-1.83), upon which the cells were spun down, resuspended in lysis buffer, and sonicated. After centrifugation for 1 hr at 4,000 rpm, the lysate was purified and analyzed as described above, yielding full-length sfGFP.

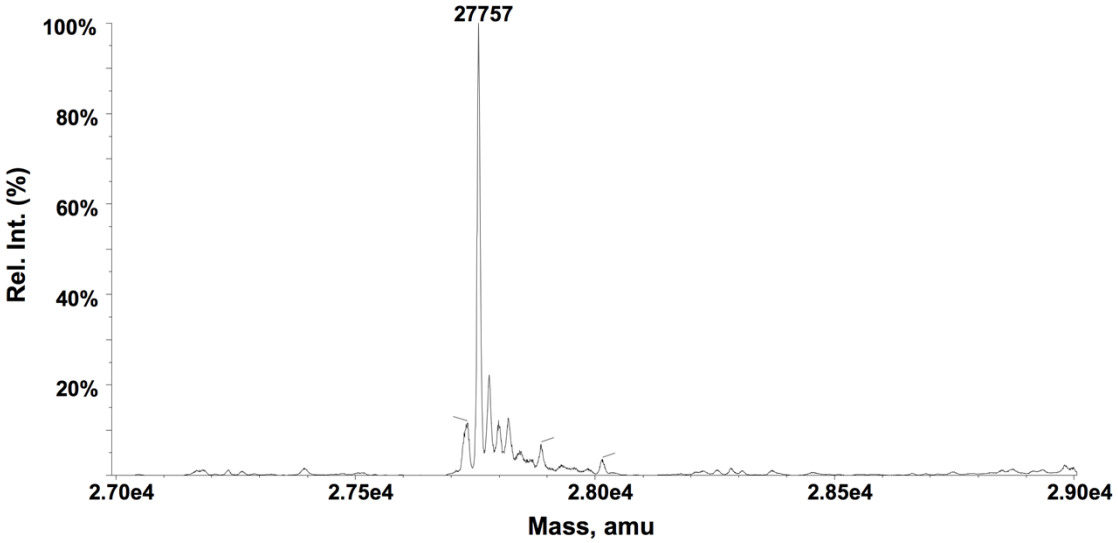


Figure S1. ESI-MS data for sfGFPS2X-**1**, M9 minimal media. Predicted mass: 27756. Observed mass: 27757.

3. *In Vitro* Aniline-Catalyzed Labelling of sfGFP

3. 1. Determination of Second-Order Rate Constant:

A stock solution of sfGFP was added to a solution of 100 mM aniline and 100 mM sodium phosphate, pH 7, to a total volume of 2 ml. To this solution was added coumarin-based dye **3**, and the reaction rate was monitored with an excitation wavelength of 417nm and emission of 510 nm for 2 hours. This reaction was repeated with various concentrations of hydroxylamine dye while keeping the sfGFP concentration constant, and each resulting scatterplot was fitted to an exponential trendline. The observed rate constants obtained from these runs were plotted, and the second-order rate constant was derived as described in the manuscript.

3. 2. Time-Based Fluorescence Assay:

An aliquot of sfGFP-**1** was diluted with a 2X stock of labelling buffer (200 mM aniline, 200 mM sodium phosphate, pH 7), with a final concentration of 8 μM sfGFP, 100 mM aniline, and 100 mM sodium phosphate. To this was added 10 μM of fluorescein-based hydroxylamine dye. The reaction was quenched with 2 mM benzaldehyde and immediately treated with SDS-PAGE loading buffer, and the resulting samples were

subjected to 12% SDS-PAGE analysis. Quenching occurred after 1, 5, 25, and 125 minutes to give the appropriate samples for the assay. The above procedure was also used for sfGFP-**2**.

Fluorescence assay without aniline catalyst:

To a solution of sfGFP-**1** in PBS buffer (8 µl, 3.6 µM) was added dye **4** (1.8 µl, 20 µM) to a final concentration of 2.9 µM sfGFP and 3.6 µM dye **4**. After the designated amount of time had elapsed (1 h, 2 h, 7 h, and 18 h), the solution was precipitated using trichloroacetic acid, then the resulting pellet was washed twice with ice-cold acetone. After allowing the pellet to dry, it was resuspended in 20 µl H_2O and prepared for SDS-PAGE analysis using conventional methods. An identical procedure was performed for sfGFP-**2**, resulting in a total of eight reactions.

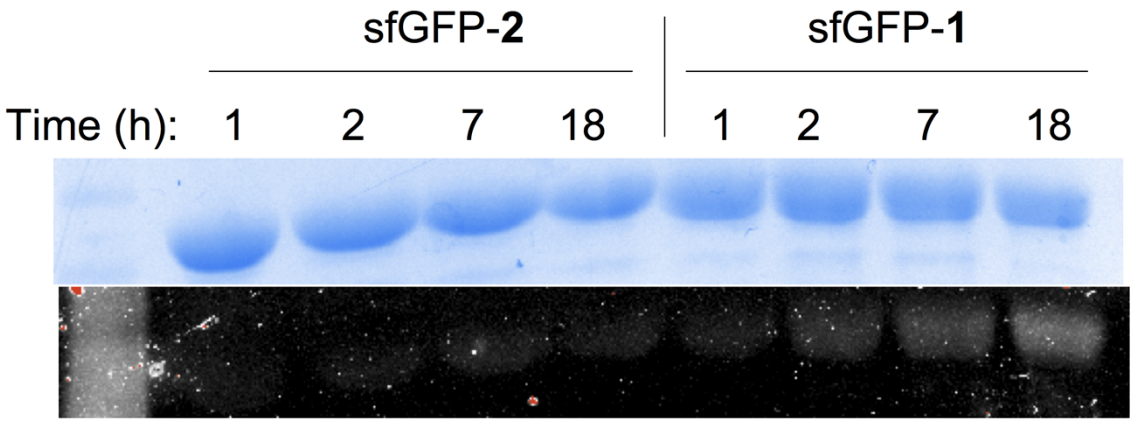


Figure S2. Bioconjugation of aldehyde in the absence of aniline catalyst at netural pH. No appreciable labelling was observed with ketone-containing sfGFP-**2**.

4. *In vivo* labelling of OmpX

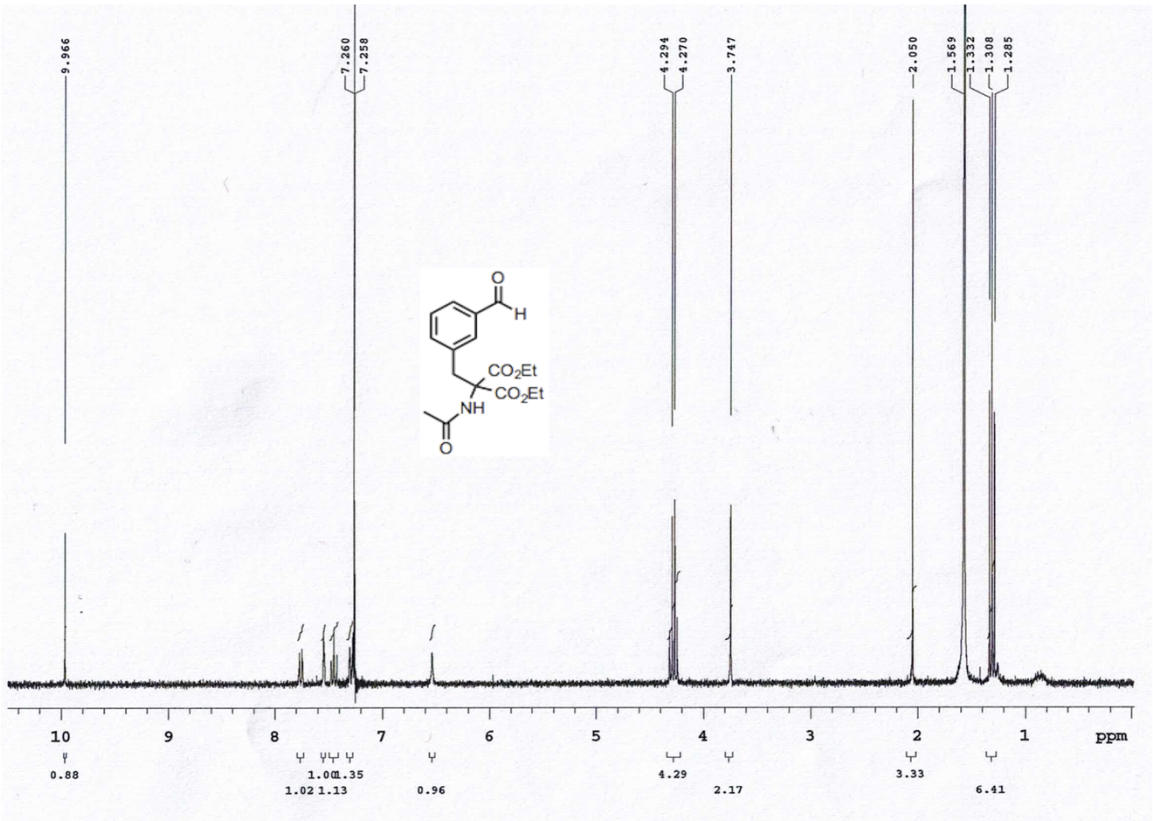
4.1. OmpX expression:

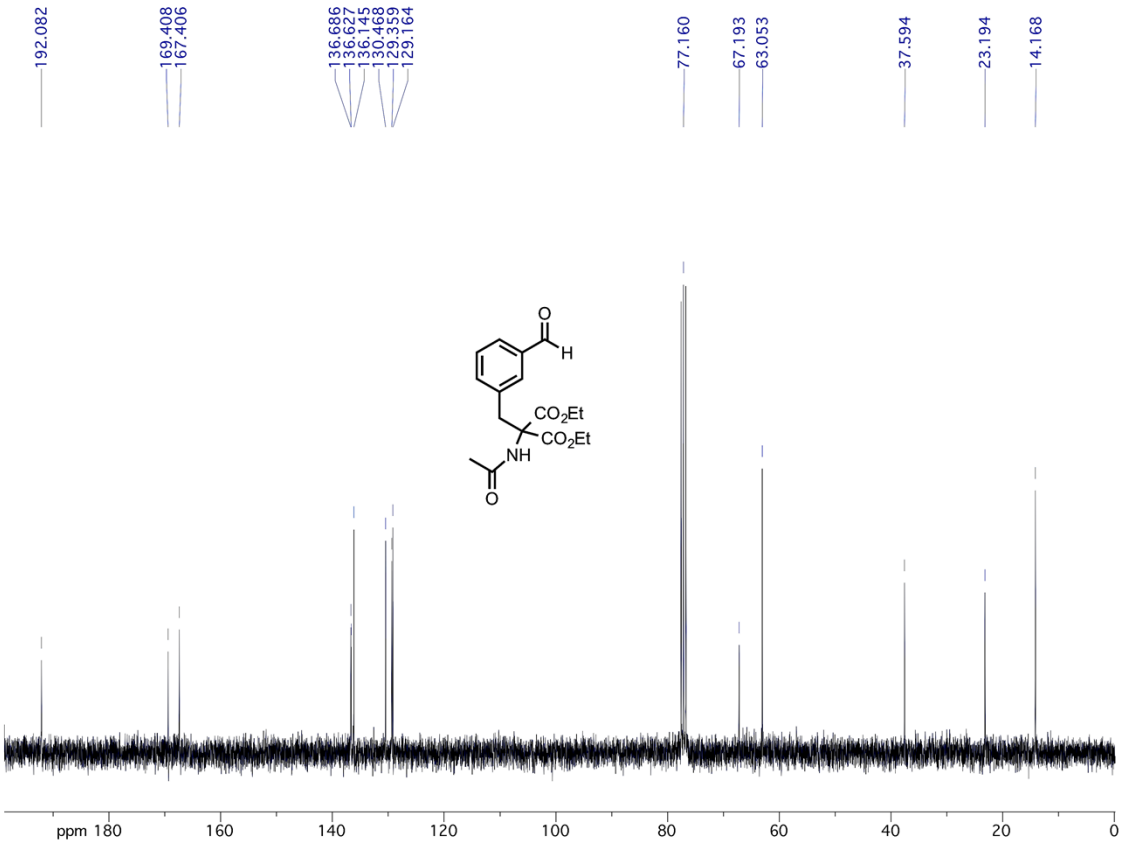
Expression of the membrane protein OmpX was performed as described previously.[7] Briefly, BL21(DE3) cells cotransformed with pEVOL-pylT-N346A/C348A and pETDuet-OmpXTAG were grown in 500 mL LB media until the O.D. reached 1.0-1.3, then the cells were pelleted and washed three times with PBS buffer. The cells were resuspended and transferred to 45 mL of minimal media supplemented with 2 mM NCAA, 1 mM IPTG, and 0.2% arabinose. Protein was expressed at 23 ºC for 10 h, and then the cells were pelleted and labeled as described below.

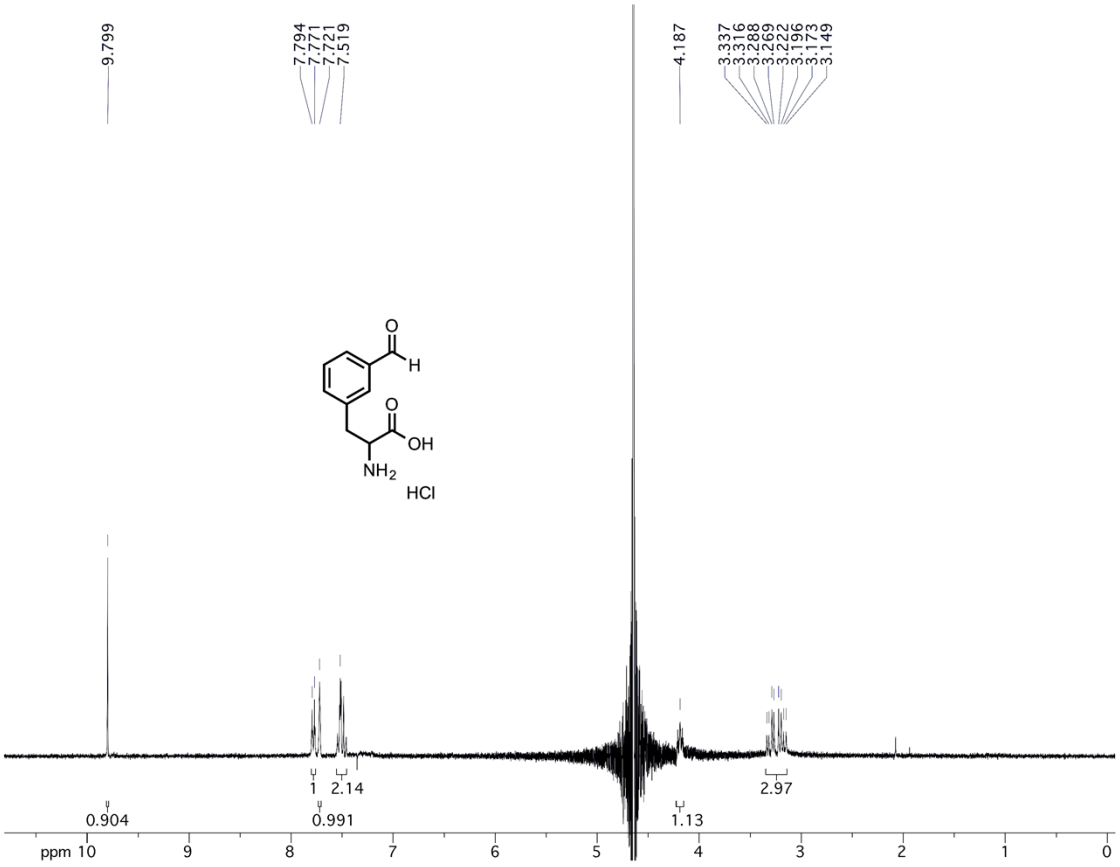
4. 2. *In vivo* labelling:

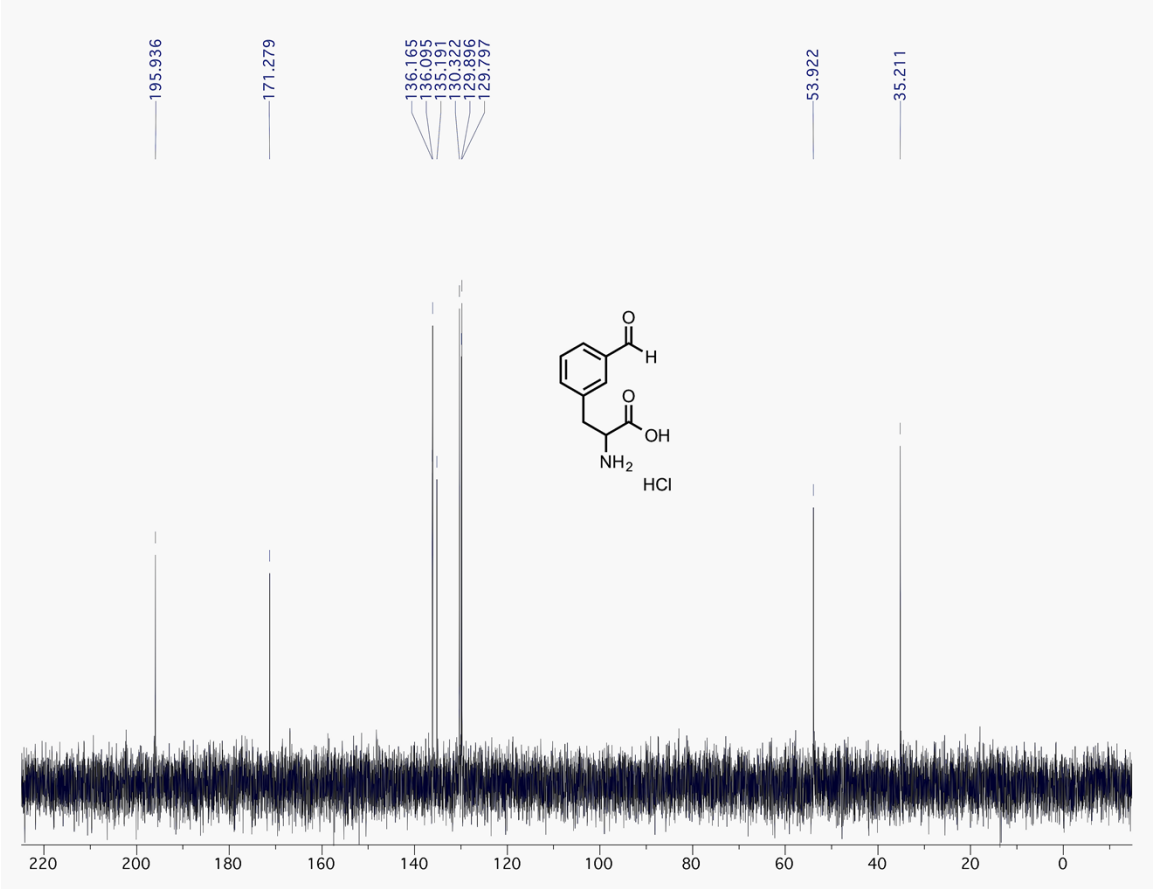
A 50 μL suspension of BL21(DE3) *E. coli* cells from procedure **4.1** was spun down for 5 min at 14,000 r.p.m., then resuspended in 100 μL labelling solution (100 mM aniline, 100 mM sodium phosphate, pH 7). Labelling was achieved by supplementing the suspension with 2 mM Dye **4**, and the reaction was left at 23 ºC for 1 h. When complete, the cells were spun down for 5 min at 14,000 r.p.m., then resuspended in PBS buffer and centrifuged again. The previous step was repeated for a total of six PBS washes, and the pellet was resuspended once more in PBS buffer and visualized via fluorescent microscopy.

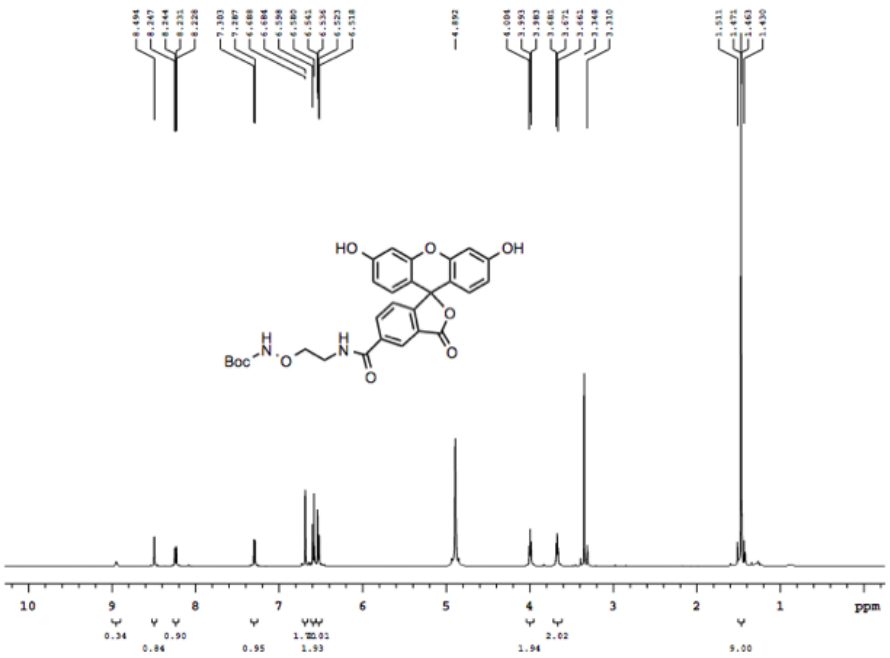
5. NMR and Mass Spectra:

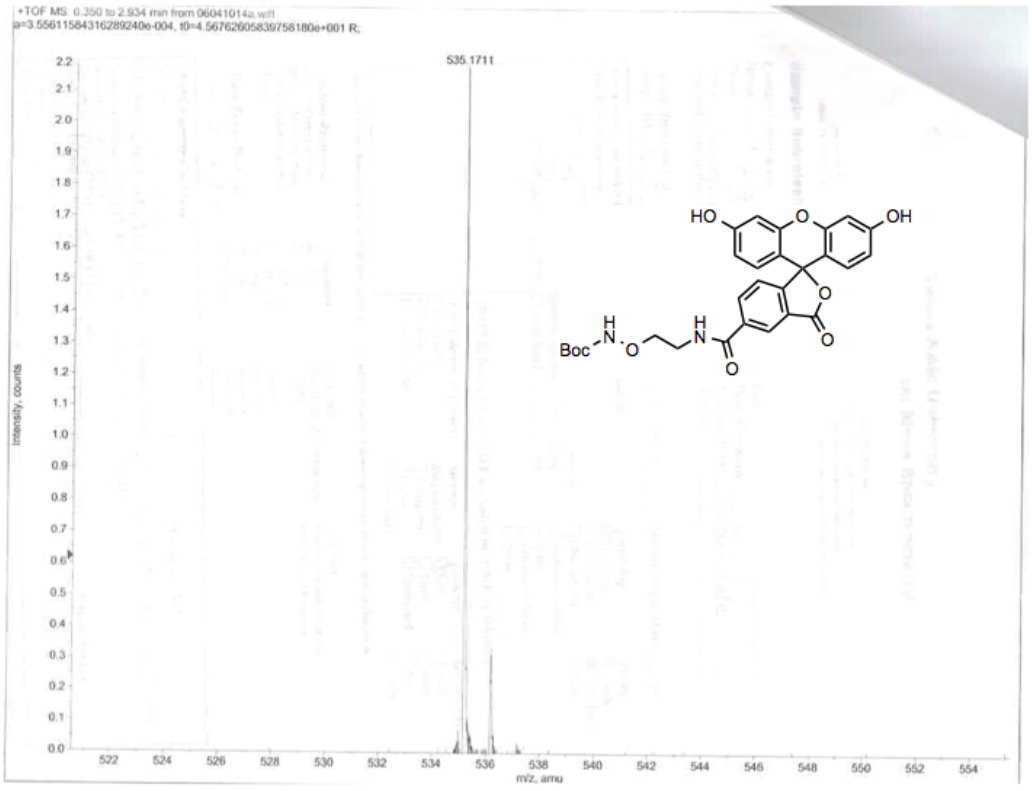


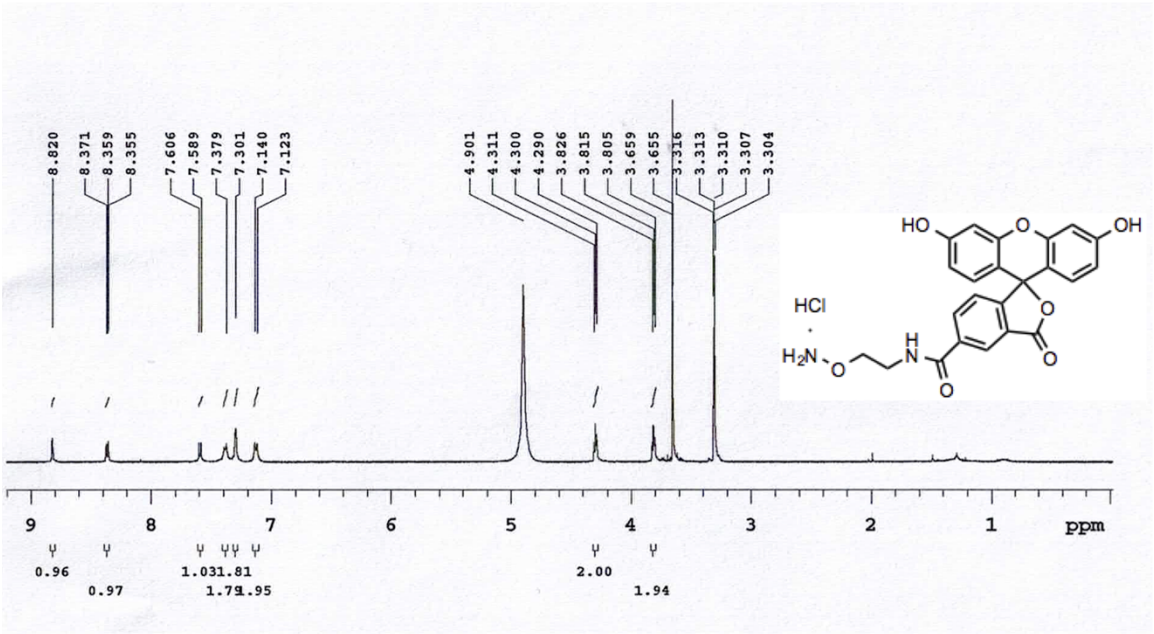

192.082
169.408
167.406
136.686
136.627
136.145
130.468
129.359
129.164
77.160
67.193
63.053
37.594
23.194
14.168
O
H
CO2Et
CO2Et
NH
O
ppm 180
160
140
120
100
80
60
40
20
0

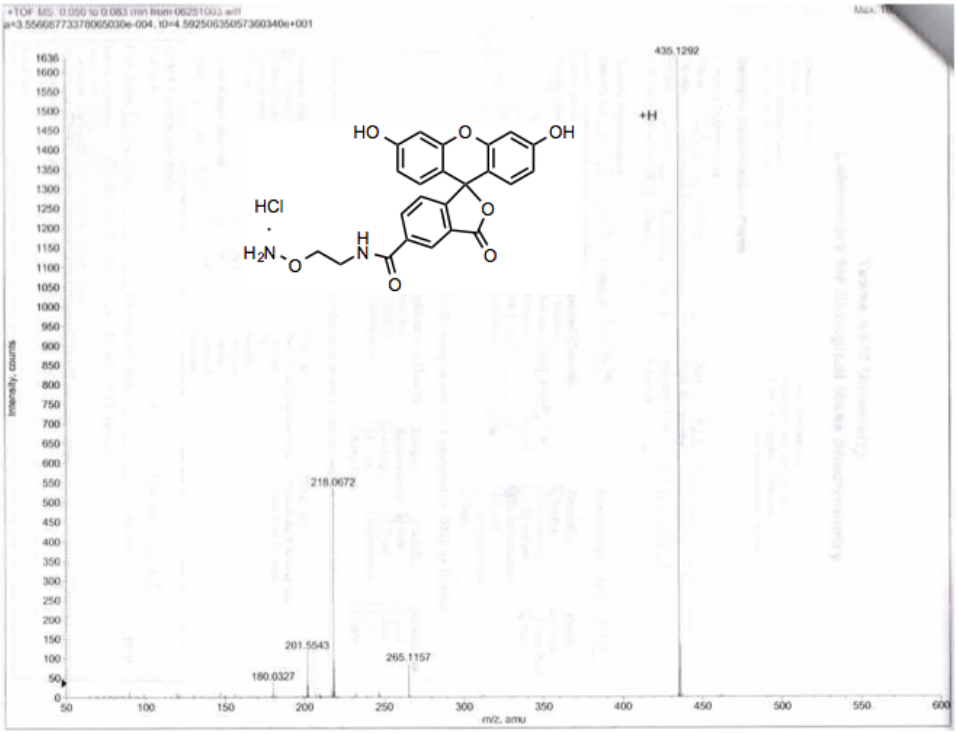

9.799
7.794
7.771
7.721
7.519
4.187
3.337
3.316
3.288
3.269
3.222
3.196
3.173
3.149
O
H
O
OH
NH2
HCl
1
2.14
2.97
0.904
0.991
1.13
ppm 10
9
8
7
6
5
4
3
2
1
0


195.936
171.279
136.165
136.095
135.191
130.322
129.896
129.797
53.922
35.211
O
H
O
OH
NH2
HCl
220
200
180
160
140
120
100
80
60
40
20
0


HO
O
OH
O
O
Boc
H
N
O
H
N
O
10
9
8
7
6
5
4
3
2
1
ppm


+TOF MS: 0.350 to 2.934 min from 06041014a.wiff
a=3.556115843162892 40e-004, t0=4.567626058397581 80e+001 R,
535.1711
Intensity, counts
m/z, amu
HO
OH
Boc
H
N
O
O
O
O


8.820
8.371
8.359
8.355
7.606
7.589
7.379
7.301
7.140
7.123
4.901
4.311
4.300
4.290
3.826
3.815
3.805
3.659
3.655
3.316
3.313
3.310
3.307
3.304
HO
OH
HCl
H2N
O
O
O
O
9
8
7
6
5
4
3
2
1
ppm
0.96
0.97
1.03
1.81
1.79
1.95
2.00
1.94



References:

1. Pedelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2005**, *24*, 79-88.
2. Bookser, B. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1991**, *113*, 4208-4218.
3. Wu, B.; Wang, Z.; Huang, Y.; Liu, W. R. *ChemBioChem* **2012**, *13*, 1405-1408.
4. Y.-S. Wang, X. Fang, A. L. Wallace, B. Wu, and W. R. Liu, *J. Am. Chem. Soc.* 2012, **134**, 2950-2953.
5. Y.-S. Wang, X. Fang, H.-Y. Chen, B. Wu, Z. U. Wang, C. B. Hilty, and W. R. Liu, *ACS Chem. Biol.*, 2012, **8**, 405-415.
6. J. T. Hammill, S. Miyake-Stoner, J. L. Hazen, J. C. Jackson, and R. A. Mehl, *Nature Protocols*, 2007, **2**, 2601-2607.
7. Y.-J. C. Lee, B. Wu, J. E. Raymond, Y. Zeng, X. Fang, K. L. Wooley, and W. R. Liu, *ACS Chem. Biol.* **2013**, *8*, 1664-1670.