Electronic Supplemental Information for:

Genetic encoding of the post-translational modification 2-hydroxyisobutyryl-lysine.

William A. Knight [a] and T. Ashton Cropp* [a]

Synthesis

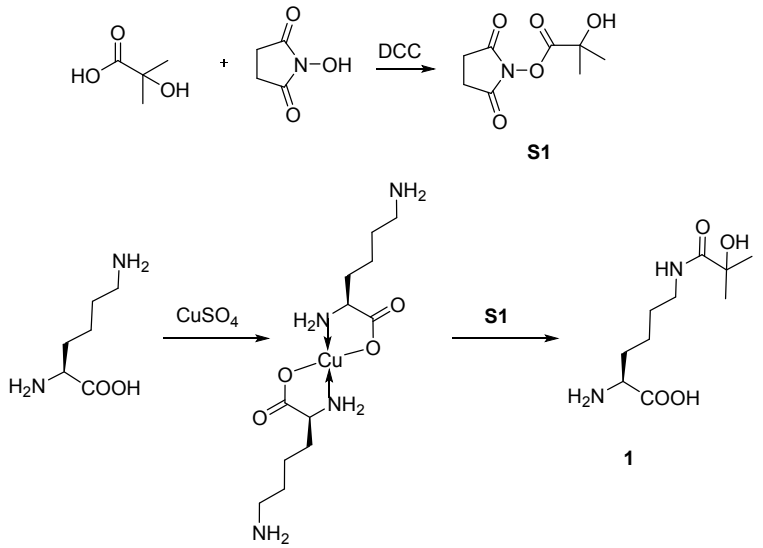


Figure S1

2-hydroxyisobutrylysine (**1**): 2-hydroxyisobutryic acid (3.35g, 32mmol) and N-hydroxysuccinimide (3.75g, 32.6mmol) were combined in 30 mL of CH_3CN. After stirring for 10 minutes, N,N'-dicyclohexylcarbodiimide (DCC) (6.72g, 32.6mmol) was added and allowed to stir. After three hours the solution was filtered and the solvent removed to give 2-hydroxyisobutyryl-OSu (**S1**) as a yellowish oil (3.24g, 50%), that was used directly without further purification. Separately, L-lysine monohydrochloride (980 mg, 5.37mmol) was added to 5mL of saturated aqueous $NaHCO_3$. The mixture was stirred at room temperature for ten minutes and $CuSO_4$ (669mg, 2.68mmol) added. 2-hydroxyisobutyryl-OSu (**S1**) from the first step (3.24g) was re-suspended in 10 mL of acetone, and added drop-wise to the lysine complex over four hours to produce a blue slurry. The reaction was quenched with methanol and evaporated to dryness. The solid was re-suspended in water (50mL) and stirred at room temperature. To the stirring solution was added 8-quinolinol (1.95g, 13.43mmol) and stirred vigorously overnight to produce a green slurry. The green slurry was filtered, and the filtrate washed with ethyl acetate (3 x 30mL). The aqueous layer was then evaporated to dryness to give **1** (967mg, 77%). R_f=0.36 (3:1:1, 1-butanol:acetic acid:water, pink with ninhydrin stain) ^{1}H NMR (300 MHz, D_2O): $\sigma =$ 1.19, (8H, m),1.35, (2H, m), 1.68, (2H, m), 3.01 (2H, t, J = 6Hz), 3.60 (1H, t, J = 6Hz), ^{13}C NMR (75 MHz, D_2O): $\sigma =$ 21.6, 25.2, 26.3, 28.1, 29.9, 38.6, 54.3, 73.0, 176.3, 179.1. MS-ESI (m/z) $[M+H]^+$, calc $C_{10}H_{21}N_2O_4$ 233.1501, found 233.1492.

Synthetase screening

Chemically competent *E.* coli DH10-B were doubly-transformed with pBK_sfGFP151TAG and each of four different synthetase plasmids (pBK_MmPylS, pBK_EV16-5, pBK_EV17, pBK_EV20). A single colony from each plate was picked to inoculate 4mL liquid culture. Once the liquid cultures became saturated, they were streaked on three different plates: plate 1 (LB, Kan 50μg/mL, Tet 15μg/mL, 0.2% arabinose), plate 2 (LB, Kan 50μg/mL, Tet 15μg/mL, 0.2% arabinose, 2mM Nε-Boc-lysine) and plate 3 (LB, Kan 50μg/mL, Tet 15μg/mL, 0.2% arabinose, 1mM 2-hydroxyisobutyryl lysine) (**Figure S2**).

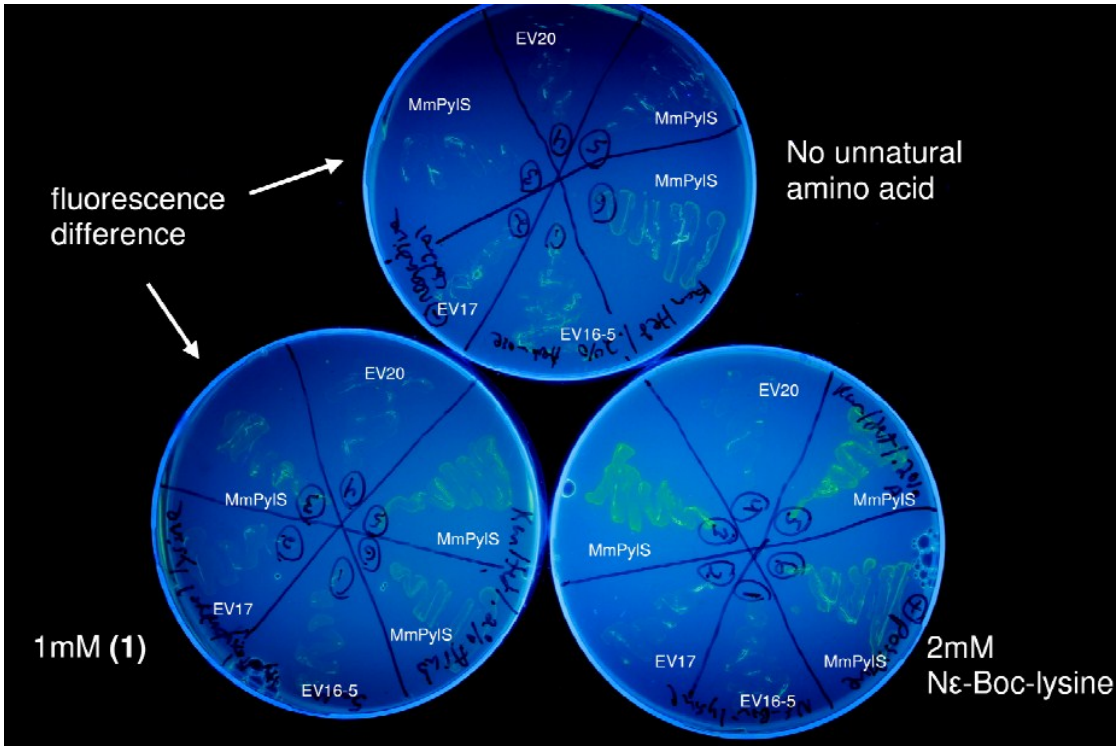


Figure S2. Screening for unnatural amino acid incorporation in sfGFP using four different PylS variants. MmPylS was streaked three times. Plates were photographed on a UV transilluminator.

Protein expression and mass spectrometry

From the 4mL culture containing the pBK_MmPylS synthetase plasmid, 1mL was used to inoculate an expression cultures in 200mL of LB media containing tetracycline and kanamycin. After the culture reached an OD600=0.6, arabinose (final percentage of 0.2%) and 5mM hydroxyisobutyryl lysine (**1**). A negative control culture was treated the same, with no unnatural amino acid added. The cultures were incubated overnight and then the cells pelleted at 6,000 x g, and resuspended in 10mL of lysis buffer (50mM NaH_2PO4, 300mM NaCl, 10mM imidazole, pH=8) with 5mg of lysozyme. After 30 minutes each culture was sonicated 3 x 1 minute at 50% power with a 30 second rest. The resulting pellets were centrifuged at 15,000 x g and the lysate transferred to a clean 15mL Falcon tube. 500μl of Promega His Bind was added to each tube and shaken for 30 minutes. The resin was pelleted by a centrifugation and sequentially washed 3 times with 12mL of lysis/binding buffer. The resin was eluted with 500μl elution buffer (50mM

NaH_2PO_4, 300mM NaCl, 250mM imidazole, pH=8). Samples for analysis by SDS-PAGE were 20μL. Tryptic peptide MS was performed as described using an Agilent 6340 ion trap.[1] Intact protein MS was performed by direct infusion into an Agilent 6340 ion trap, and the resulting spectrum deconvoluted using MagTran software.

Sequence of sfGFP protein. Site of mutation (Y151) shown as X:

MGGSHHHHHHGMASMMSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIEL
KGIDFKEDGNILGHKLEYNFNSHNVXITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLL
PDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

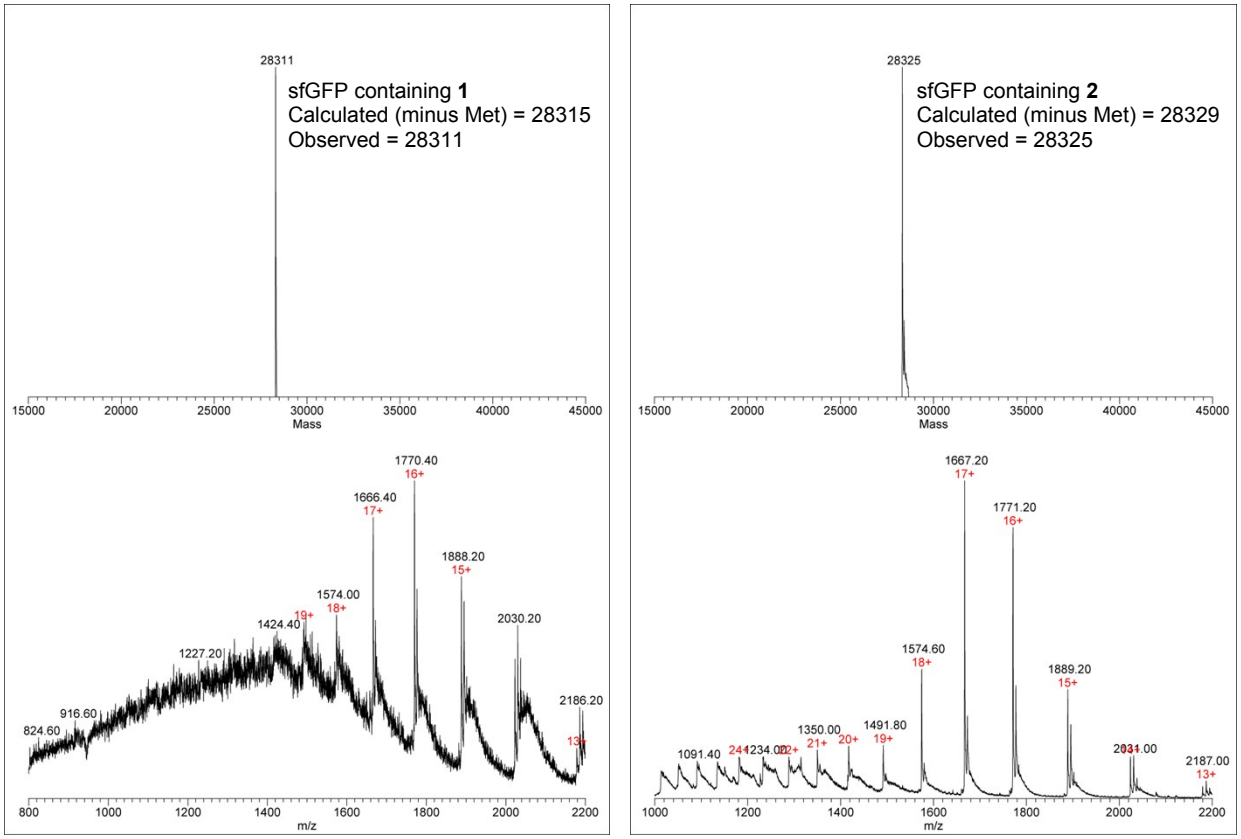
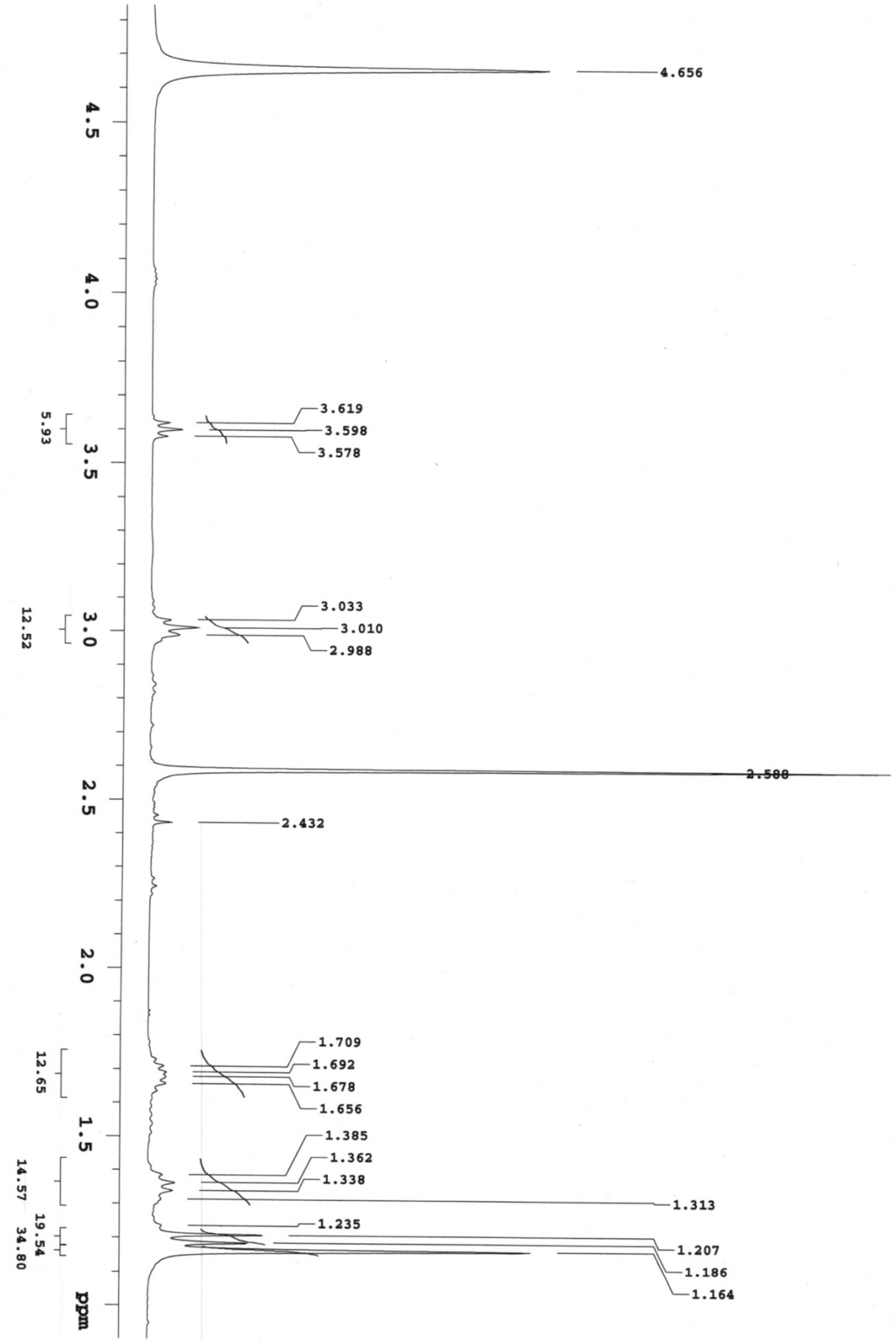
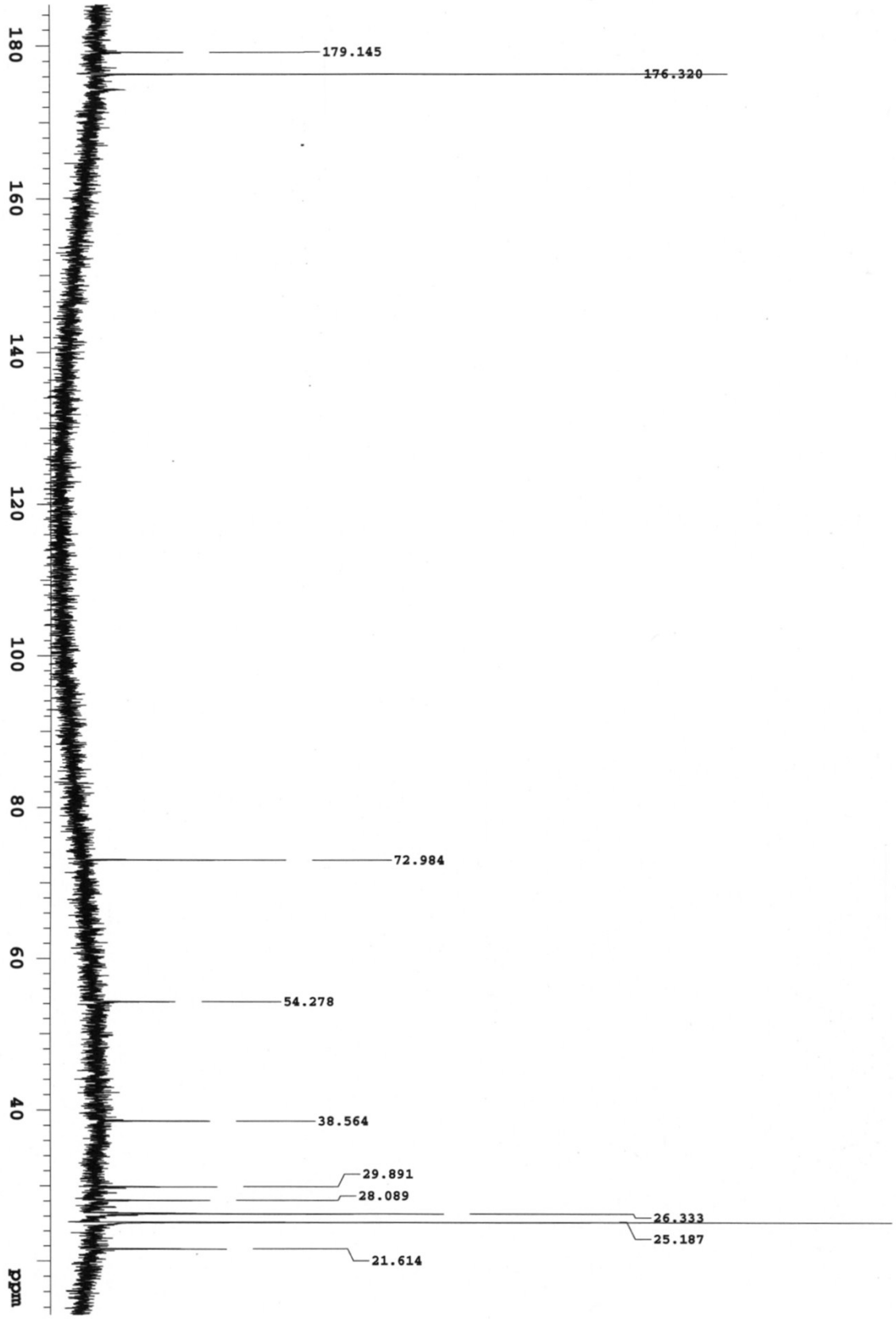


Figure S3. A) Intact protein ESI-MS of sfGFP containing **1** or **2**. Calculated masses correspond to the protein sequence minus the N-terminal methionine (131) and GFP chromophore maturation (20). Calculated mass difference between **1** and **2** is 14, observed mass difference is 14. Minor peaks seen in ESI data correspond to phosphate adducts[2] (M+98), and is seen for both **1** and **2**.

^{1}H NMR of **1**:


4.656
3.619
3.598
3.578
3.033
3.010
2.988
2.588
2.432
1.709
1.692
1.678
1.656
1.385
1.362
1.338
1.313
1.235
1.207
1.186
1.164
5.93
12.52
12.65
14.57
19.54
34.80
4.5
4.0
3.5
3.0
2.5
2.0
1.5
ppm

^{13}C NMR of **1**:


179.145
176.320
72.984
54.278
38.564
29.891
28.089
26.333
25.187
21.614
180
160
140
120
100
80
60
40
ppm

References

1. A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen and M. Mann, *Nat. Protoc.* 2006, **1**, 2856-60.
2. S.K. Chowdhury, V. Katta, R.C. Beavis, B.T. Chait, *J. Am. Soc. Mass. Spectrom.* 1990, **5**, 382-8.