Supporting Information for:

Genetic Incorporation of Twelve *meta*-Substituted Phenylalanine Derivatives Using A Single Pyrrolysyl-tRNA Synthetase

Yane-Shih Wang,[¶] Xinqiang Fang,[¶] Hsueh-Ying Chen, Bo Wu, Zhiyong U. Wang,[†] Christian Hilty and Wenshe R. Liu*

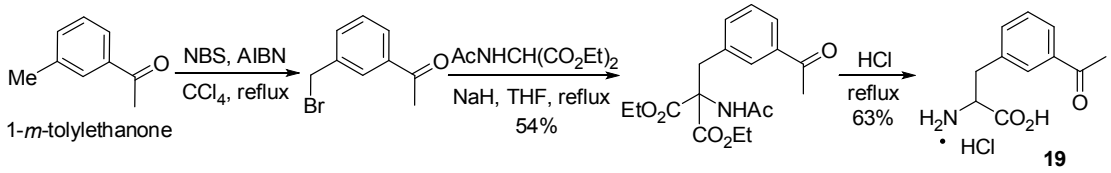
Department of Chemistry, Texas A&M University, College Station, TX 77843

[†]Current address: Department of Chemistry and Physics, Troy University, Troy, AL 36082.

Contents

1. Organic Synthesis

Noncanonical amino acids **19**, **20** and **21** were synthesized via the routes showed below. All other amino acids are commercial available from Chem-Impex International Inc.

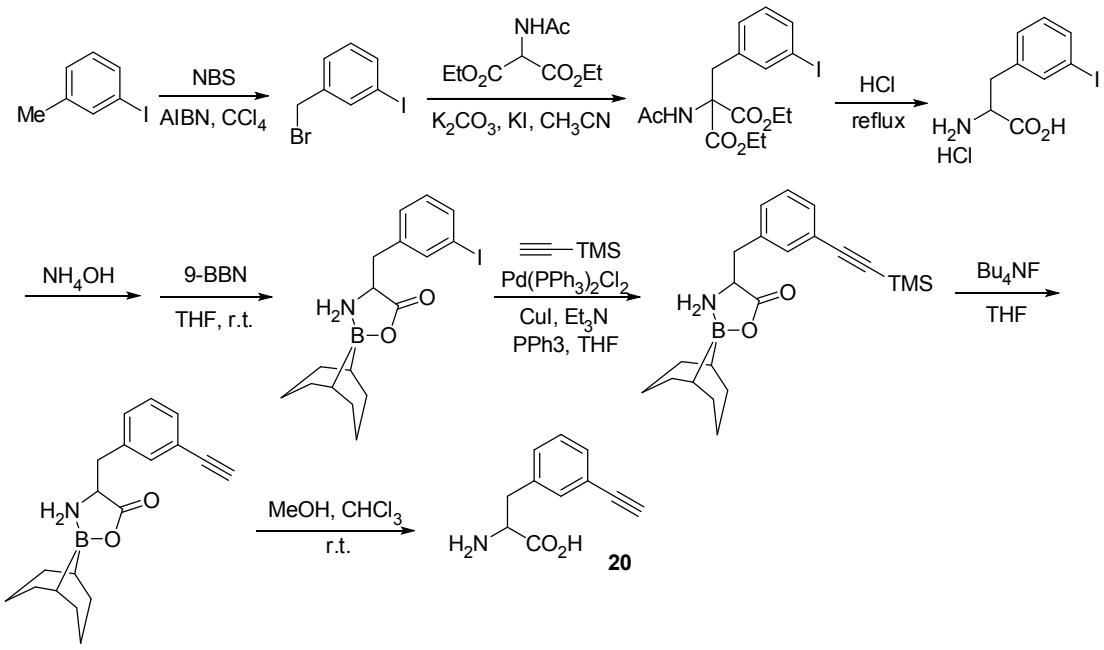


Supplementary Scheme 1. Synthesis of compound **19**.

1-*m*-Tolylethanone (8 g, 0.052 mol) was dissolved in CCl_4 (150 mL) and then NBS (12 g, 1.3 eq.) and AIBN (2.56 g, 30 mol%) were added. The mixture was heated to reflux for 12 h under the protection of Argon. After cooling, the precipitation was removed via filtration and the filtrate was concentrated and dried under vacuum to afford 1-(3-(bromomethyl)phenyl)ethanone, which was used directly to the next step without further purification.

Diethyl 2-acetamidomalonate (10.27 g, 0.04727 mol) was dissolved in dry THF (400 mL) and cooled to 0^{o}C. NaH (2.458 g, 60 % dispersion in mineral oil, 1.3 eq.) was added in portions to the above solution. The mixture was stirred at room temperature for 20 min. 1-(3-(Bromomethyl)phenyl)ethanone (all gotten from the first step) in THF (30 mL) was added dropwise and then heated to reflux for 2 h. After cooling, NH_4Cl aqueous was added slowly and the resulting mixture was extracted with Et_2O (100 mL × 3). The organic layers were collected and dried over anhydrous Na_2SO_4, and concentrated. The residue was purified via re-crystallization in hexane/EtOAc to afford diethyl 2-acetamido-2-(3-acetylbenzyl)malonate as a white solid (8.9 g, 54% yield).

A suspension of diethyl 2-acetamido-2-(3-acetylbenzyl)malonate (8.9 g, 0.02547 mol) in 3M HCl (254.7 mL, 32 eq.) was heated to reflux for 16 h before cooling to room temperature. Water was evaporated under reduced pressure and the solid was collected. The solid was washed with Et_2O (10 mL × 3) and then dried under vacuum to afford 3-(3-acetylphenyl)-2-aminopropanoic acid (**19**) as a white solid (3.9 g, yield 63%). ^{1}H NMR (D_2O, 300 MHz) 2.67 (s, 3H), 3.30-3.44 (m, 2H), 4.33 (dd, 1H, J = 6.0, 10.0 Hz), 7.55-7.58 (m, 2H), 7.88-7.96 (m, 2H); ^{13}C NMR (D_2O, 75 MHz) 26.2, 35.3, 53.9, 128.1, 129.1, 129.5, 134.6, 134.8, 136.8, 171.2, 203.3.



Supplementary Scheme 2. Synthesis of compound **20**.

1-Iodo-3-methylbenzene (6.68 mL, 0.052 mol) was dissolved in CCl_4 (150 mL) under Argon protection, then added NBS (12 g, 1.3 eq.) and AIBN (3 g, 0.35 eq.). The mixture was heated to reflux for 12 h before cooling to room temperature. The solid was removed via filtration, and the filtrate was concentrated to afford 1-(bromomethyl)-3-iodobenzene (9.48g, 62% yield), which was used in the next step without further purification.

To the solution of 1-(bromomethyl)-3-iodobenzene (9.48g, 0.032 mol) in anhydrous acetonitrile (200 mL) was added sequentially diethyl 2-acetamidomalonate (6.935 g, 1.0 eq.), K_2CO_3 (8.84 g, 2.0 eq.) and KI (5.3 g, 1.0 eq.). The resulting mixture was heated to reflux for 12 h under the protection of Argon. Then cooled to room temperature. The solid was filtered, the solvent was removed under reduced pressure, and the residue was purified via column chromatograph with hexanes/ethyl acetate (3:1 v/v) as eluent to give the pure product of diethyl 2-acetamido-2-(3-iodobenzyl)malonate (11.34 g, 82% yield).

Diethyl 2-acetamido-2-(3-iodobenzyl)malonate (5.8 g, 0.0134 mol) was added to concentrated HCl aqueous (37% assay, 100 mL) and the mixture was refluxed overnight. After cooling, the mixture was concentrated under reduced pressure and the resulting solid was washed with CH_2Cl_2 (5 mL) and EtOAc (5 mL), then collected and dried under vacuum to afford 2-amino-3-(3-iodophenyl)propanoic acid hydrochloride as a white solid (3.95 g, yield 90%).

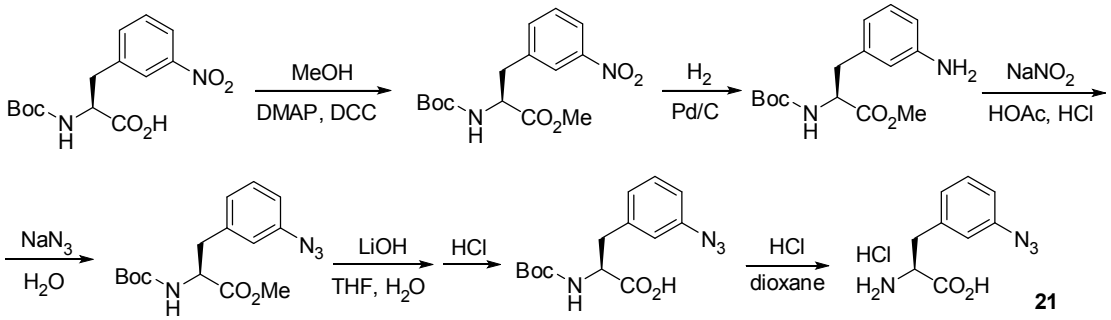
To 2-amino-3-(3-iodophenyl)propanoic acid hydrochloride (1.5 g, 0.0046 mol) was added ammonium hydroxide solution(11 mL, 28% NH_3 basis) at 0^{o}C. The mixture was stirred at 0^{o}C for 30 minutes. Then the solution was concentrated and the resulting solid was dried under high vacuum. The above solid was dissolved in anhydrous THF (5 mL) under Argon and then added 9-BBN (0.5 M in THF, 10.5 mL, 1.14 eq.). The reaction mixture was stirred at room temperature for 20 h. Then the

solvent was removed under reduced pressure and the residue was purified via column chromatography with hexanes-ethyl acetate (3:1 v/v) as eluent to afford the corresponding 3-iodo-phenylalanine boron complex as a white solid (1.2 g, yield 73%).

The above boron complex was added to a 100 mL of round-bottom-flask, followed by Ph_3P (0.076 g, 1 mol%), $Cl_2(Ph_3P)_2Pd$ (0.04 g, 2 mol%) and CuI (0.028 g, 5 mol%). Then the flask was evacuated and refilled with Argon. Then dry THF (15 mL) was added, followed by ethynyltrimethylsilane (0.62 mL, 1.5 eq.) and Et_3N (0.81 mL, 2 eq.). The reaction mixture was stirred at room temperature for 18 h. The solvent was removed and the residue was purified via column chromatography with hexanes-ethyl acetate (3:1 v/v) as eluent to afford the corresponding coupling product (1 g, yield 90%).

The above coupling compound (2.85 g, 0.00747 mol) was dissolved in anhydrous THF (20 mL) at 0°C. Then Bu_4NF (1.0 M in THF, 11.2 mL, 1.5 eq.) was added dropwise and the resulting mixture was stirred at 0°C for 0.5 h. Water (5 mL) was added and stirring continued for another 5 min. The solvent of THF was removed and the residue was extracted with EtOAc (30 mL). The organic layer was collected and dried over anhydrous Na_2SO_4 and concentrated. The residue was purified via column chromatography with hexanes-ethyl acetate (3:1 to 1:1 v/v) as eluent to afford the corresponding de-protected alkyne product (1.2 g, yield 52%).

The above alkyne (1.2 g, 0.00388 mol) was dissolved in MeOH (1.2 mL) and $CHCl_3$ (30 mL). The resulting mixture was stirred at room temperature until all starting material was consumed up. The white precipitation was collected and dried via vacuum to afford 3-ethynylphenylalanine (**20**) as a white solid (0.53 g, yield 72%). ^{1}H NMR (CD_3OD, 300 MHz) 3.00 (dd, 1H, J = 8.4, 14.4 Hz), 3.26 (dd, 1H, J = 4.8, 14.4 Hz), 3.30 (s, 1H), 3.83 (dd, 1H, J = 4.8, 8.4 Hz), 7.30-7.43 (m, 4H); ^{13}C NMR (CD_3OD, 75 MHz) 35.7, 54.9, 77.0, 82.2, 122.4, 128.1, 128.9, 130.1, 132.0, 135.5, 171.2.



Supplementary Scheme 3. Synthesis of compound **21**.

(S)-2-(tert-butoxycarbonylamino)-3-(3-nitrophenyl)propanoic acid (1.6 g, 0.0057 mol) added to MeOH (20 mL). Then DMAP (0.348 g, 0.5 eq.) and *p*-Toluenesulfonic acid monohydrate (0.54 g, 0.5 eq.) were added. DCC (1.176 g, 1 eq.) was dissolved in CH_2Cl_2 (20 mL) and added to the above solution dropwise. The reaction mixture was

stirred overnight. The solvents were rotary evaporated under reduced pressure. The residue was purified via column chromatography with hexanes-ethyl acetate (3:1 v/v) as eluent to afford the (S)-methyl 2-(tert-butoxycarbonylamino)- 3-(3-nitrophenyl) propanoate (1.7 g, yield 99%).

(S)-methyl 2-(tert-butoxycarbonylamino)-3-(3-nitrophenyl)propanoate (1.4 g, 0.0043 mol) was dissolved in MeOH (10 mL) under protection of Argon. Palladium on carbon (0.14 g, 10 wt. % loading, matrix carbon, dry support) was added and the the system was evacuated and then refilled with hydrogen (balloon pressure). The mixture was stirred at room temperature overnight and then filtered through Celite. The filtrate was collected and concentrated, then the resulting (S)-methyl 3-(3-aminophenyl)-2-(tert-butoxycarbonylamino)propanoate (1.1 g, 87% yield) was used directly in the next step.

(S)-methyl 3-(3-aminophenyl)-2-(tert-butoxycarbonylamino)propanoate (1.1 g, 0.0037 mol) was dissolved in the mixture of water (6 mL) and THF (3 mL). HOAc (0.227 mL) and HCl aqueous (1.0 M, 4 mL) were added at 4^{o}C, followed by $NaNO_2$ (0.316 g, 4.58 mmol) in water (3 mL). The mixture was stirred at this temperature for 30 min. Then NaN_3 (0.3 g, 0.046 mol) in water (3 mL) was added and stirring continued for another 30 min. The reaction mixture was extracted with CH_2Cl_2 (10 mL×2), dried over anhydrous Na_2SO_4, and concentrated to afford (S)-methyl 3-(3-azidophenyl)-2-(tert-butoxycarbonylamino)propanoate (1 g, 84% yield).

(S)-methyl 3-(3-azidophenyl)-2-(tert-butoxycarbonylamino)propanoate (1 g, 3 mmol) was dissolved in THF (10 mL) and LiOH aqueous (0.5 M, 13.8 mL, 2.3 eq.) was added. The reaction mixture was stirred at room temperature for 3 h. Then the solution was extracted with Et_2O (10 mL) and the aqueous layer was acidified with 3M HCl to pH = 1-2. Then the mixture was extracted with EtOAc (20 mL × 2). The combined organic layers were dried over anhydrous Na_2SO_4, and concentrated to afford (S)-3-(3-azidophenyl)-2-(tert-butoxycarbonylamino)propanoic acid, and this compound was used directly to the next step.

The above (S)-3-(3-azidophenyl)-2-(tert-butoxycarbonylamino)propanoic acid was dissolved in HCl (4.0 M in dixoane, 2 mL) and stirred overnight. The resulting white solid was collected by filtration and dried under vacuum to afford (S)-2-amino-3-(3-azidophenyl)propanoic acid hydrochloride (**21**) (0.6 g, 82% yield, two steps). ^{1}H NMR (D_2O, 300 MHz) 3.17-3.28 (m, 2H), 4.27-4.32 (m, 1H), 6.95 (d, 1H, J = 1.5 Hz), 7.03-7.08 (m, 2H), 7.32-7.39 (m, 1H); ^{13}C NMR (D_2O, 75 MHz) 35.1, 53.7, 118.3, 119.5, 125.7, 130.4, 135.7, 140.2, 170.1.

2. DNA and Protein Sequences

2.1 DNA Sequences

sfGFP:

atgtagaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgt
ccgtggagagggtgaaggtgatgctacaaacggaaaactcacccttaaatttatttgcactactggaaaactacctgttccgt
ggccaacacttgtcactactctgacctatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaag

agtgccatgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggacctacaagacgcgtgctgaagtc
aagtttgaaggtgataccttgttaatcgtatcgagttaaagggtattgattttaaagaagatggaaacattcttggacacaaac
tcgagtacaactttaactcacacaatgtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcg
ccacaacgttgaagatggttccgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttacc
agacaaccattacctgtcgacacaatctgtcctttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgt
aactgctgctgggattacacatggcatggatgagctctacaaaggatcccatcaccatcaccatcactaa

pylT:

ggaaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

Methanosarcina mazei PylRS:

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaac
accacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagca
ggactgcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataa
gttcctcacaaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggca
atgccaaaatccgttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggatctaaatttt
cacctgcgataccggtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggag
caactgcatccgcactggtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgcccccgcac
ttacgaagagccagactgacaggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcag
ggagcttgagtccgaattgctctctcgcagaaaaaaagacctgcagcagatctacgcggaagaaagggagaattatctgg
ggaaactcgagcgtgaaattaccaggttctttgtggacaggggttttctggaaataaaatccccgatcctgatccctcttgagt
atatcgaaaggatgggcattgataatgataccgaactttcaaaacagatcttcagggttgacaagaacttctgcctgagaccc
atgcttgctccaaacctttacaactacctgcgcaagcttgacagggccctgcctgatccaataaaaatttttgaaataggccca
tgctacagaaaagagtccgacggcaaagaacacctcgaagagtttaccatgctgaacttctgccagatgggatcgggatg
cacacgggaaaatcttgaaagcataattacggacttcctgaaccacctgggaattgatttcaagatcgtaggcgattcctgca
tggtctatggggataccttgatgtaatgcacggagacctggaactttcctctgcagtagtcggacccataccgcttgaccg
ggaatggggtattgataaaccctggataggggcaggtttcgggctcgaacgccttctaaaggttaaacacgactttaaaaat
atcaagagagctgcaaggtccgagtcttactataacgggatttctaccaacctgtaa

2.2 Proteins Sequences

sfGFP:
MXKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFI**C**TTGK
LPVPWPTLVTTLTYGVQ**C**FSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDG
TYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQ
KNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSK
DPNEKRDHMVLLEFVTAAGITHGMDELYKGSHHHHHH

X represents a noncanonical amino acid.

Methanosarcina mazei PylRS:

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRS
SRTARALRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRT
KKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSI
STGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNS

GKPFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPIL
IPLEYIERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIK
IFEIGPCYRKESDGKEHLEEFTML**N**F**C**QMGSGCTRENLESIITDFLNHLGIDFKI
VGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERL
LKVKHDFKNIKRAARSESYYNGISTNL

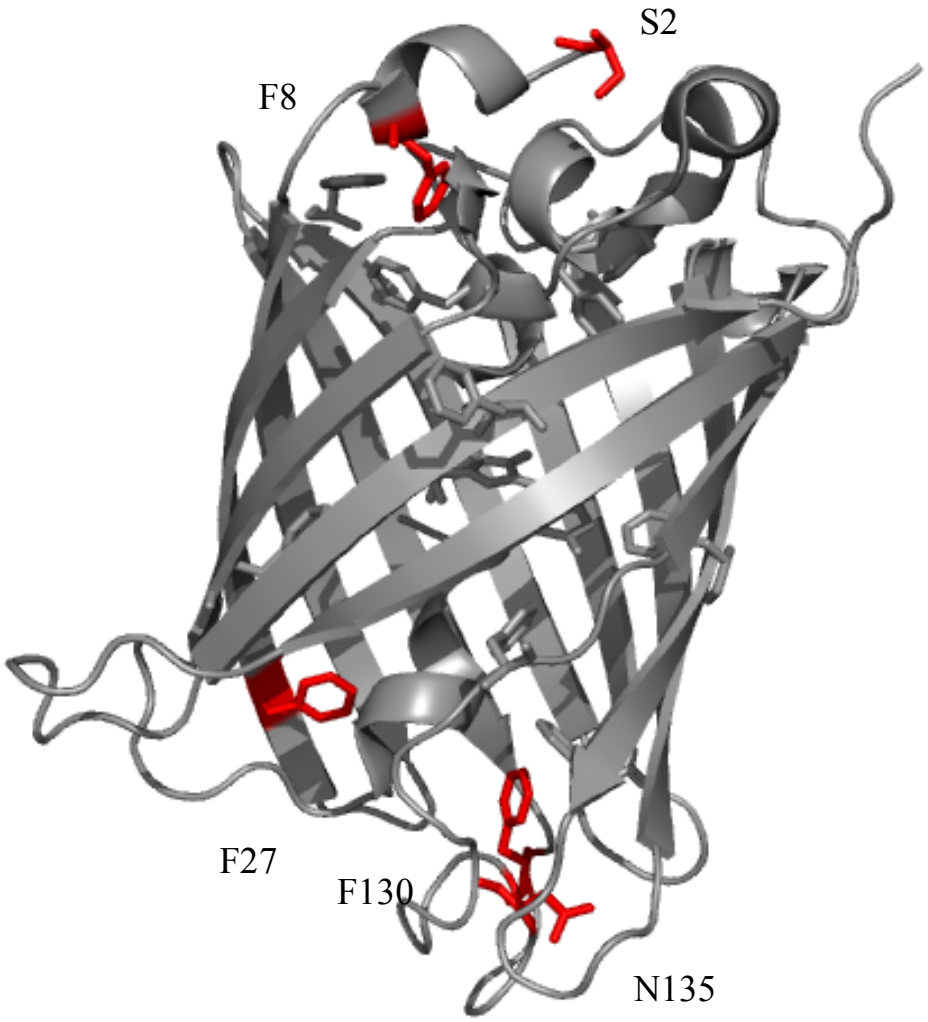
The bold and underline letters indicate the chosen amino acids for mutations

3. Supplementary Tables and Figures

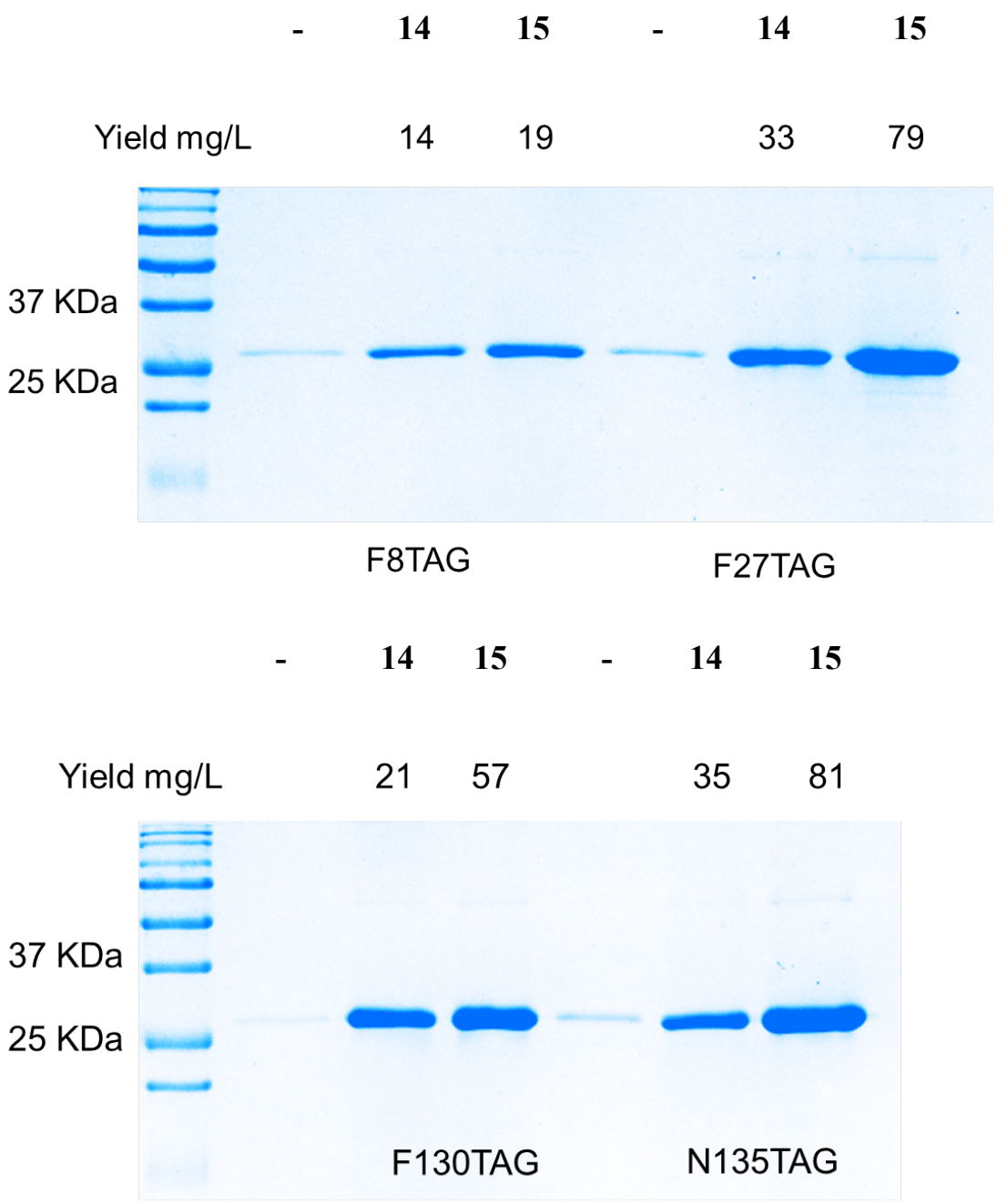
Supplementary Table 1. Growth media, incorporation yields, and potential applications for **10**-**21**

NAA	Incorporation yield in GMML (mg/L)[a]	Feasibility in rich media[b]	Potential applications
10	1	No	
11	16	Yes	Phage display
12	28	Yes	Phage display, Suzuki coupling
13	20	Yes	Phage display, Suzuki coupling, Anomalous diffraction phasing
14	12	Yes	Phage display
15	39	Yes	Phage display, a NMR probe
16	5	Yes	Phage display
17	7	Yes	Phage display, an IR probe
18	10	Yes	Phage display, a fluorescent quencher of tryptophan, a strong immune reaction inducer
19	37	Yes	Phage display, site-selective protein modification
20	10	Yes	Phage display, site-selective protein modification
21	9	Yes	Phage display, site-selective protein modification, an IR probe

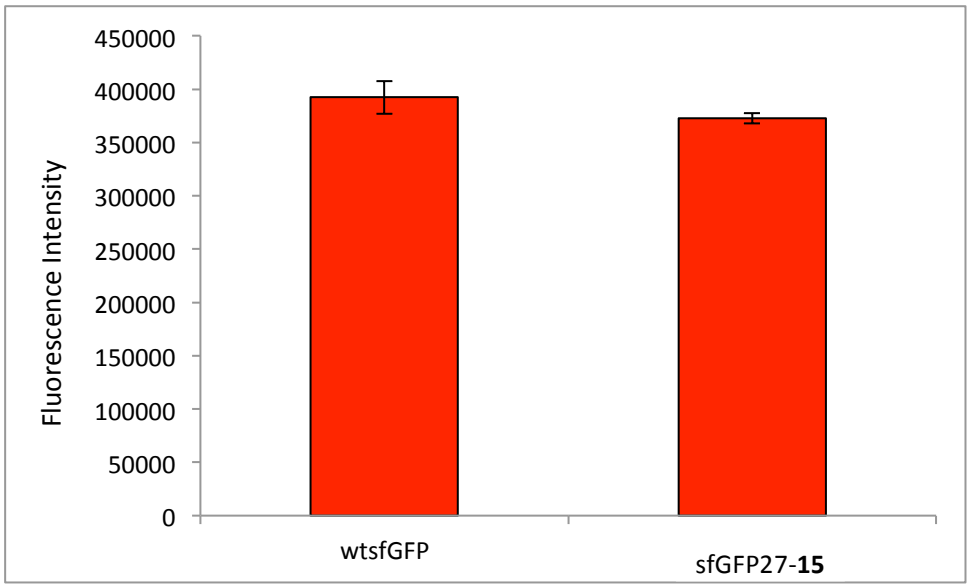
[a]The expression yields of sfGFP with each NAA incorporated at its S2 position.
[b]NAAs from **11**-**21** can be used in both LB and 2YT media.



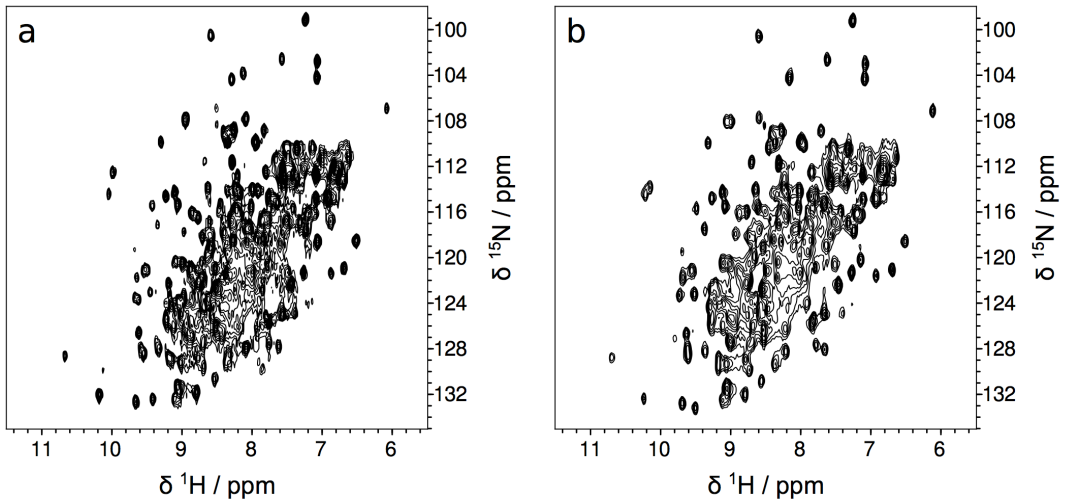
Supplementary Figure 1. sfGFP x-ray crystal structure. The five sites, S2, F8, F27, F130 and N135 were labeled in red color in the structure. PDB code: 2B3P.



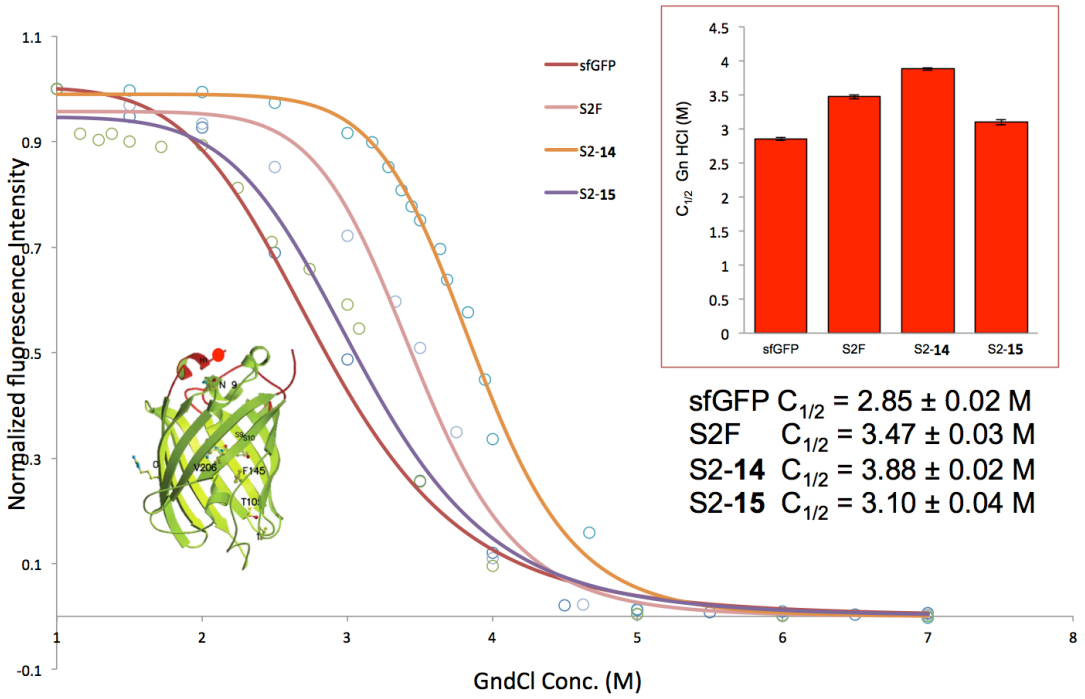
Supplementary Figure 2. The expression of sfGFP containing an amber mutation at F8, F27, F130, and N135. Proteins were expressed in BL21(DE3) cells that grew in LB medium and 1 mM **14** or **15**. The proteins were analyzed by SDS-PAGE (15%) gel electrophoresis with Coomassie blue staining.



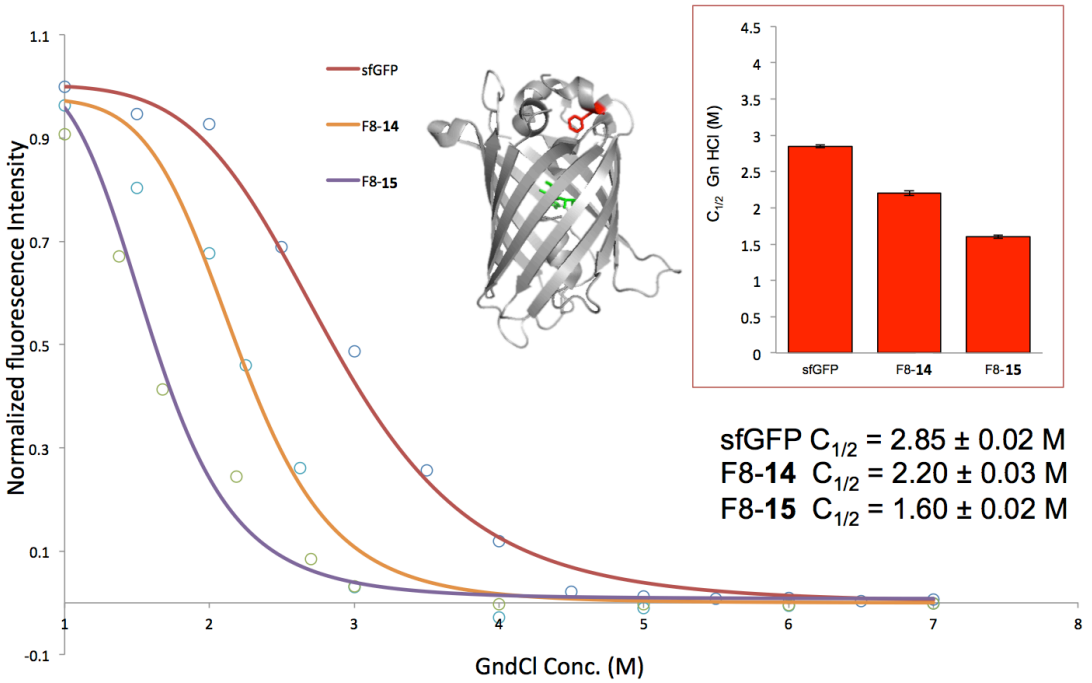
Supplementary Figure 3. The fluorescence intensity of wt-sfGFP and sfGFP27-**15**. The same protein concentration 500 nM was prepared in PBS buffer in pH = 7.0. Excitation and emission wavelength are 450 nm and 520 nm respectively. The fluorescence intensity of sfGFP2-**15** was 95% of wt-sfGFP.



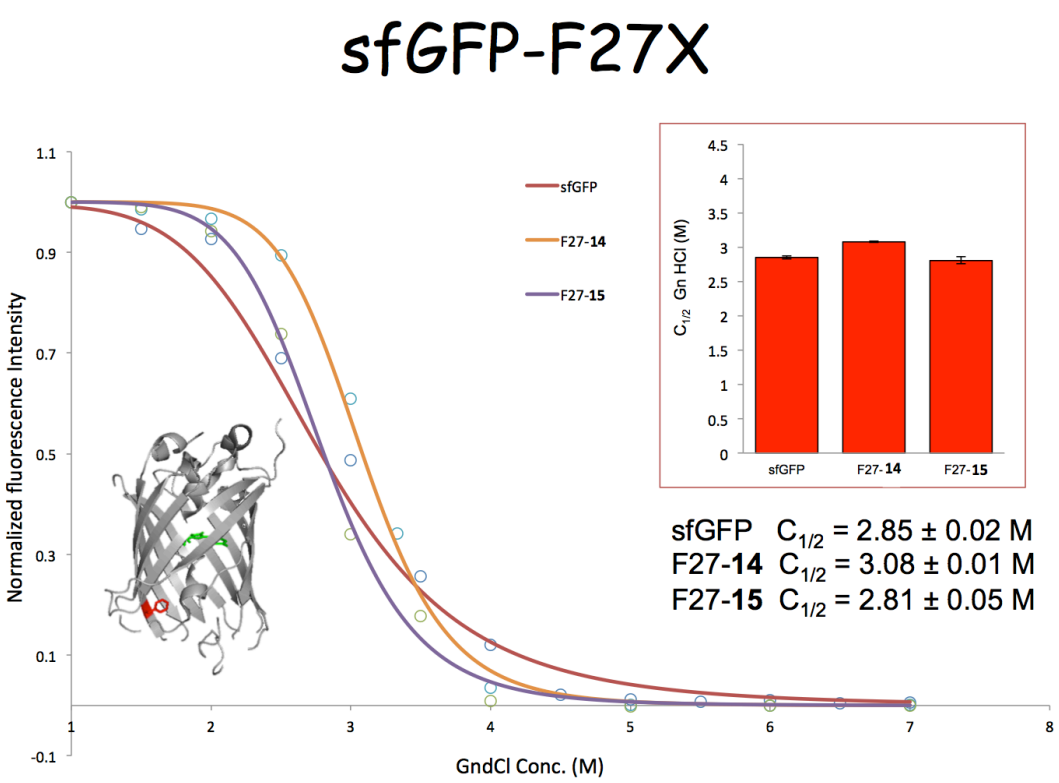
Supplementary Figure 4. [^{15}N,^{1}H]-HMQC spectrum of (**A**)^{15}N-sfGFP27-**15** and (**B**) wild-type sfGFP in 12 mM phosphate buffer (pH 7.0) with 140 mM NaCl and 3 mM KCl, acquired at a temperature of 35 ºC, at 500 MHz. Chemical shifts are referenced against an internal standard of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS).



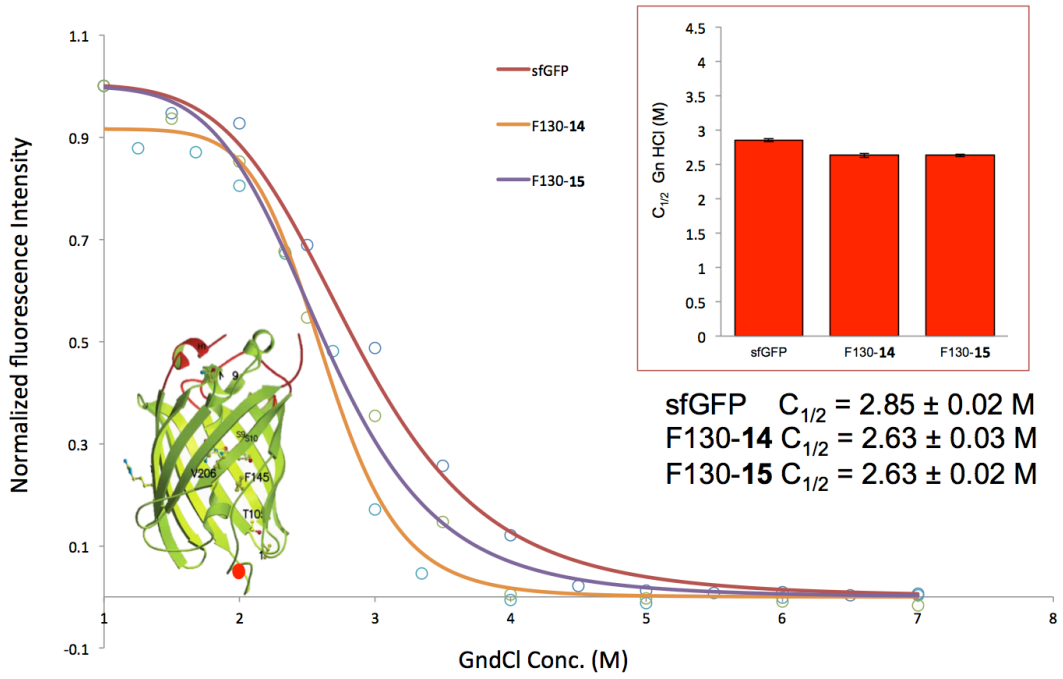
Supplementary Figure 5. Equilibrium unfolding of sfGFP proteins with different amino acids at its S2 position induced by GndCl in the PBS buffer. The plot shows residual fluorescence of each protein at different concentrations of GndCl. Excitation and emission wavelength are 450 nm and 520 nm respectively. Each data set was fit to a standard two-state unfolding mechanism to calculate the midpoint denaturant concentrations that are shown as the figure inlet (ref: van Mierlo et al., Protein Sci. **1998**, 7, 2331-2344). Same unfolding analyses were carried out for other sfGFP proteins. The results are shown in **Supplementary Figures 6-9**.



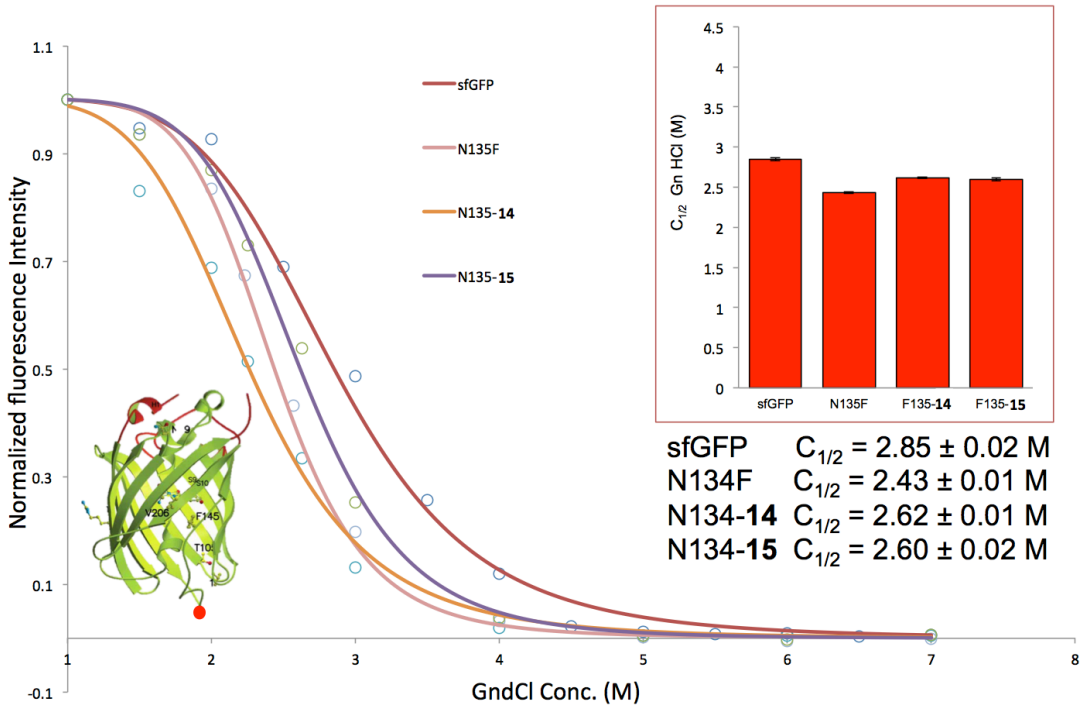
Supplementary Figure 6. Equilibrium unfolding of sfGFP proteins with different amino acids at its F8 position induced by GndCl in the PBS buffer.



Supplementary Figure 7. Equilibrium unfolding of sfGFP proteins with different amino acids at its F27 position induced by GndCl in the PBS buffer.

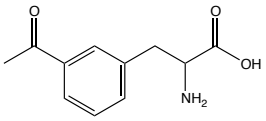


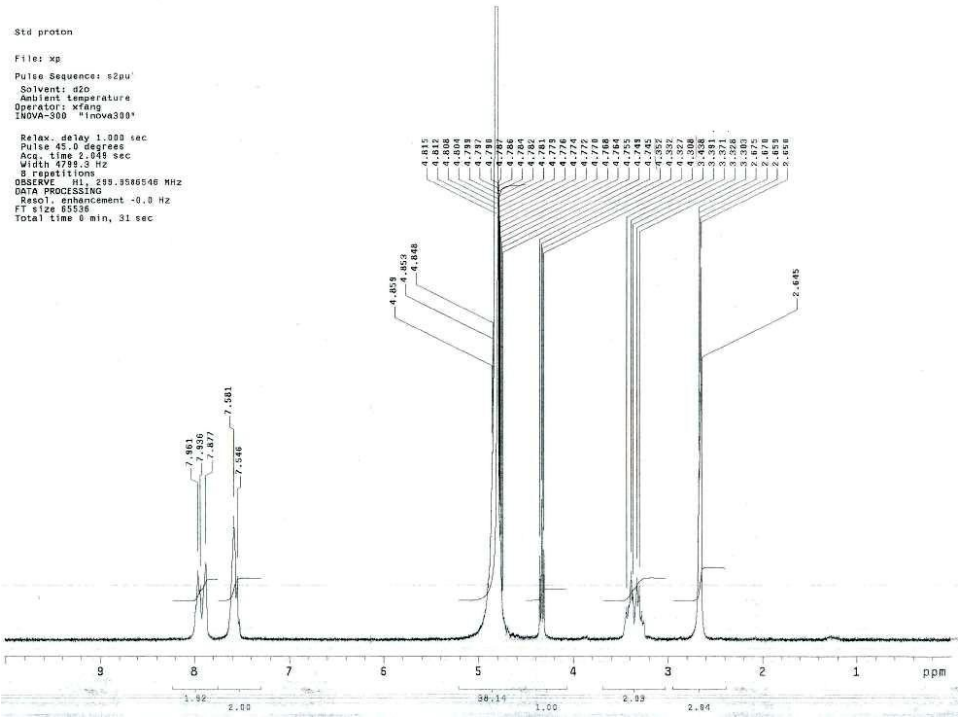
Supplementary Figure 8. Equilibrium unfolding of sfGFP proteins with different amino acids at its F130 position induced by GndCl in the PBS buffer.

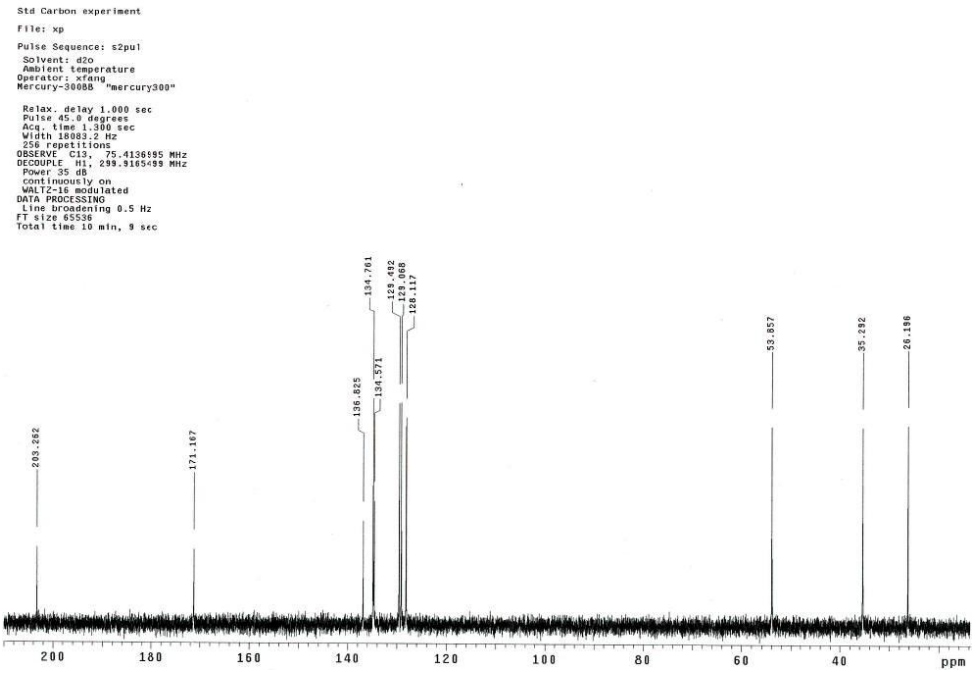


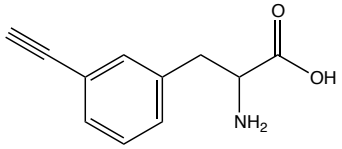
Supplementary Figure 9. Equilibrium unfolding of sfGFP proteins with different amino acids at its N135 position induced by GndCl in the PBS buffer.

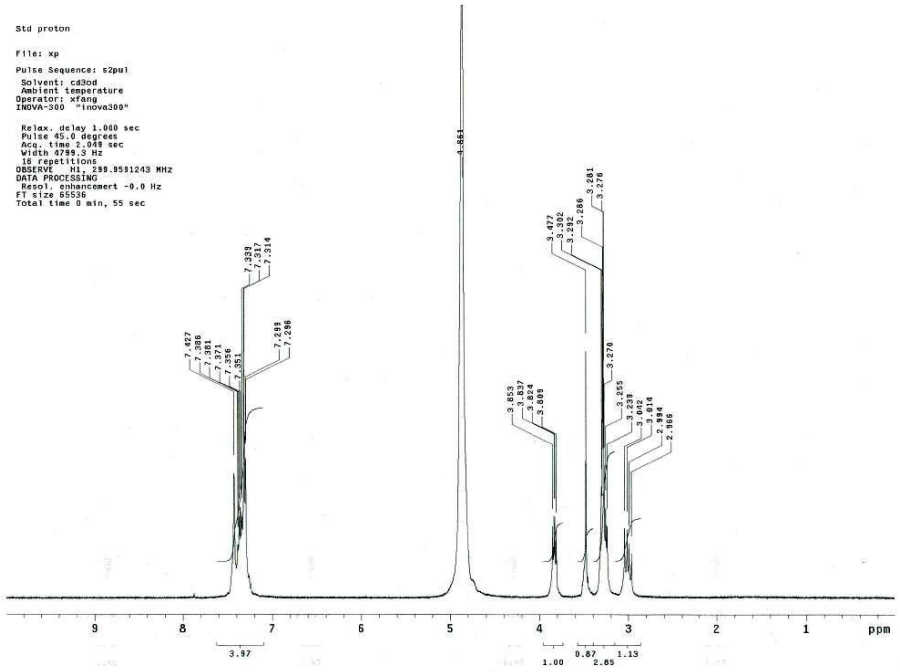
4. Spectra of 19, **20**, and **21**

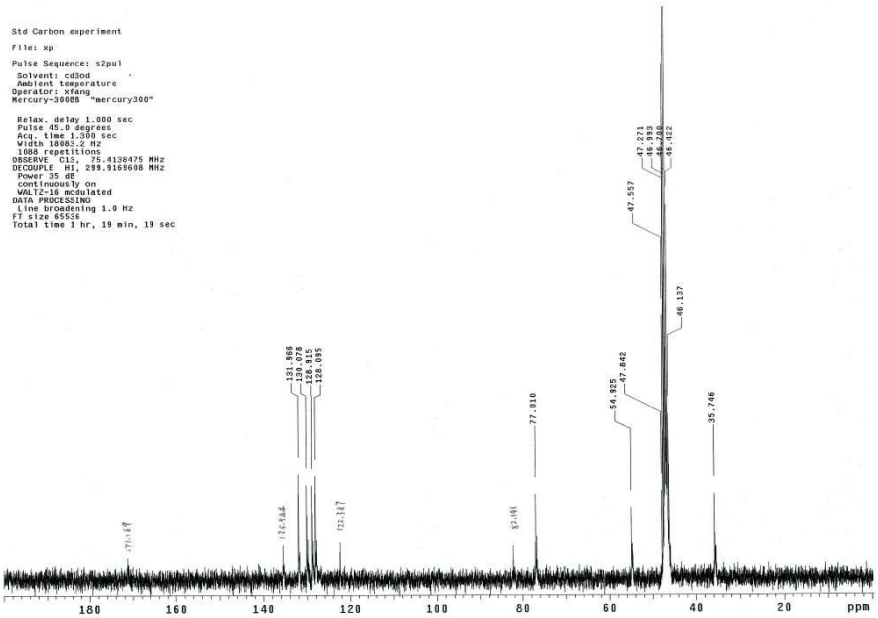


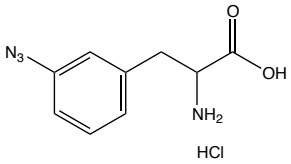




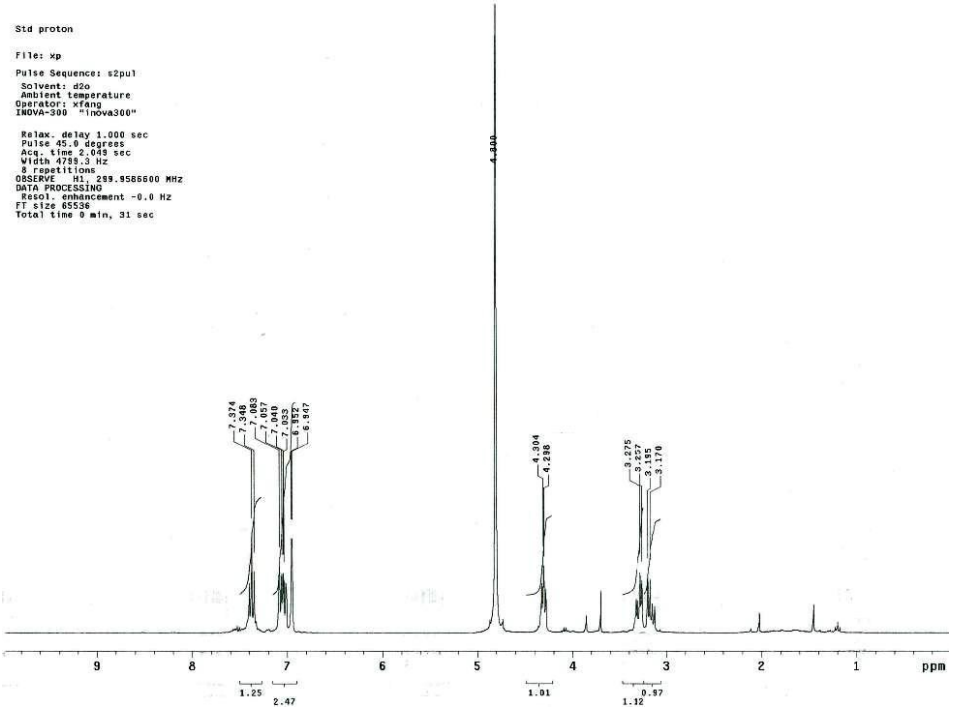

O
OH
NH2

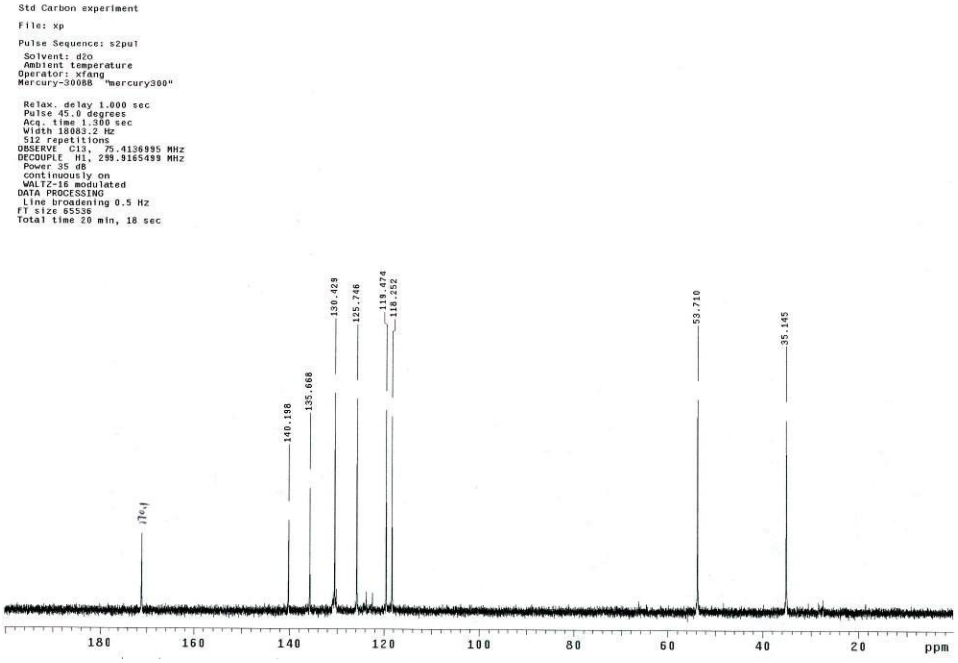

Std proton
File: xp
Pulse Sequence: s2pul
Solvent: cd3od
Ambient temperature
Operator: xfang
INOVA-300 "inova300"
Relax. delay 1.000 sec
Pulse 45.0 degrees
Acq. time 2.048 sec
Width 6799.8 Hz
16 repetitions
OBSERVE H1, 299.9591243 MHz
DATA PROCESSING
Resol. enhancement -0.0 Hz
FT size 65536
Total time 0 min, 55 sec
ppm
3.97
1.00
0.87
1.13
2.85


Std Carbon experiment
File: xp
Pulse Sequence: s2pul
Solvent: cd3od
Ambient temperature
Operator: xfang
Mercury-300BB "mercury300"
Relax. delay 1.000 sec
Pulse 45.0 degrees
Acq. time 1.300 sec
Width 18083.2 Hz
1088 repetitions
OBSERVE C13, 75.4138475 MHz
DECOUPLE H1, 299.9169698 MHz
Power 35 dB
continuously on
WALTZ-16 modulated
DATA PROCESSING
Line broadening 1.0 Hz
FT size 65536
Total time 1 hr, 19 min, 19 sec
ppm


O
N3
OH
NH2

HCl


Std proton
File: xp
Pulse Sequence: s2pul
Solvent: d2o
Ambient temperature
Operator: xfang
INOVA-300 "inova300"
Relax. delay 1.000 sec
Pulse 45.0 degrees
Acq. time 2.049 sec
Width 4799.3 Hz
8 repetitions
OBSERVE H1, 299.9586500 MHz
DATA PROCESSING
Resol. enhancement -0.0 Hz
FT size 65536
Total time 0 min, 31 sec
7.374
7.348
7.083
7.077
7.040
7.033
6.957
6.947
4.800
4.304
4.298
3.275
3.257
3.195
3.170
9
8
7
6
5
4
3
2
1
ppm
1.25
2.47
1.01
0.97
1.12


Std Carbon experiment
File: xp
Pulse Sequence: s2pul
Solvent: d2o
Ambient temperature
Operator: xfang
Mercury-300BB "mercury300"
Relax. delay 1.000 sec
Pulse 45.0 degrees
Acq. time 1.300 sec
Width 18083.2 Hz
512 repetitions
OBSERVE C13, 75.4136995 MHz
DECOUPLE H1, 299.9165499 MHz
Power 35 dB
continuously on
WALTZ-16 modulated
DATA PROCESSING
Line broadening 0.5 Hz
FT size 65536
Total time 20 min, 18 sec
174.1
140.138
135.668
130.424
125.746
119.474
118.252
53.710
35.145
180
160
140
120
100
80
60
40
20
ppm