**Supplementary Material**

# The *De Novo* Engineering of Pyrrolysyl-tRNA Synthetase for Genetic Incorporation of L-Phenylalanine and Its Derivatives

Yane-Shih Wang, William K. Russell, Zhiyong Wang, Wei Wan, Lindsey E. Dodd, Pei-Jing Pai, David H. Russell and Wenshe R. Liu*

Department of Chemistry, Texas A&M University, College Station, TX 77843, USA
E-mail: wliu@chem.tamu.edu

**Table of Contents**



## 1. General Experimental

*p*-Iodo-L-phenylalanine and *p*-bromo-L-phenylalanine were purchased form ChemImpex. 3-(Dansylamino)phenylboronic acid (DaFBA) was purchased from Sigma.

## 2. DNA and Protein Sequences

### 2.1 DNA Sequences

Z Domain:

atgactagtgtagacaactagatcaacaaagaacaacaaaacgccttctatgagatcttcatttacctaacctgaatgaggag
cagcgtgatgccttcatccaaagtttaaaagatgaccaagccaaagcgctaacctttagcagaagctaaaaagctaaatga
tgctcaggcgcctaagggatctgagctccatcaccatcaccatcactaa

$GFP_{UV}$:

atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattgatggtgatgttaatgggcacaaattttctgtc
agtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatg
gccaacacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacagaaaacggcatgacttttcaagagt
gccatgcccgaaggttatgtacaggaacgcacttatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtt
tgaaggtgataccttgttaatcgtatcgagttaaaaggtattgattttaaagaagatggaacattctcggacacaaactcgaat
acaactataactcacacaatgtatacatcacggcagcaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaac
attgaagatggatccgttcaactagcagaccattatcaacaaatactccaattggcgatggccctgtccttttaccagacaacc
attacctgtcgacatagtctgccctttcgaaagatcccaacgaaaagcgtgacccatggtccttcttgagtttgtaactgctgct
gggattacacatggcatggatgaactctacaaagagctccatcaccatcaccatcactaa

pylT:

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

*Methanosarcina mazei* PylRS:

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccagaccggaacaattcataaaataaaaca
ccacgaagtctctcgaagcaaaatctatatgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcagg

actgcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgagatctcaataagttc
ctcacaaaggcaaacgaagaccagacaagcgtaaaagcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgcc
aaaatccgttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggattaaattttcacctg
cgataccggtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggacaactgc
atccgcactggtaaaagggaatacgaaccccattacatcatgtctgcccctgttcaggcaagtgcccccgcacttacgaag
agccagactgacaggcttgaagtcctgttaaacccaaaagatggatttccctgaattccggcaagcctttcagggagcttga
gtccgaattgctctctcgcagaaaaaaagacctgcagcagatctacgcggaagaaagggagaatatctggggaaactcga
gcgtgaaattaccaggttctttgtggacaggggttttctggaataaaatccccgatcctgatccctcttgagtatatcgaaagg
atgggcattgataatgataccgaactttcaaaacagatcttcagggttgacaagaacttctgcctgagaccatgcttgctccaa
acctttacaactacctgcgcaagcttgacagggccctgcctgatccataaaaatttttgaaataggcccatgctacagaaaag
agtccgacggcaaagaacacctcgagagtttaccatgctgaacttctgccagatgggatcgggatgcacacgggaaaatc
ttgaaagcataattacggacttcctgaaccacctgggaattgatttcagatcgtaggcgattcctgcatggtctatggggatac
ccttgatgtaatgcacgggacctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgata
aaccctggataggggcaggtttcgggctcgaacgccttctaaaggttaaacagactttaaaaatatcaagagagctgcaag
gtccgagtcttactataacgggatttctaccaacctgtaa

**2.2 Proteins Sequences**

Z Domain:

MTSVDNXINKEQQNAFYEILHLPNLNEEQRDAFIQSLKDDPSQSALLAEAKKL
NDAQAPKGSELHHHHHH

X represents a noncanonical amino acid.

$GFP_{UV}$:

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDTYGKLTLKFICTTGKLP
VPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAPEGYVQERTISFKDDGNYK
TRAEVKFEGDTLVNRIELKGIDFKEDNILGHKLEYNYNSHNVYITADKQKNGI
KANFKIRHNIEDGSVQLAHYQQNTPIGDGPVLLPDNHYLSTXSALSKDPNEKR
DHMVLLEFVTAAGITHGMDELYKELHHHHHH

X represents a noncanonical amino acid or L-phenylalanine.

*Methanosarcina mazei* PylRS:

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIMACGDHLVVNNSRSS
RTARALRHHKYRKTCKRCRVSDEDLNKFLTKNEDQTSVKVKVVSAPTRTKK
AMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSIST
GATSALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKP
FRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLE
YIERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEI
GPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDKIVGD
SCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWGAGFGLERLLKVK
HDFKNIKRAARSESYYNGISTNL

**3. Construction of plasmids**

**3.1 Constructions of pY+ and pY-**[1]

The plasmid pY+ was derived from the pRep plasmid by replacing the suppres-

sor tRNA in pRep by pylT.[2] The gene of pylT flanked by the lpp promoter at the 5' end and the *rrnC* terminator at the 3' end was amplified from pBK-AcKRS-pylT.[3-4] The plasmid pY- was derived from the pNeg plasmid by replacing the suppressor tRNA with pylT.[5] Similarly, the gene of pylT flanked by the lpp promoter at the 5' end and the *rrnC* terminator at the 3' end was amplified from pBK-AcKRS-pylT.[6,7] pY+ has a tetracycline selection marker, a chloramphenicol acetyltransferase gene with an amber mutation at D112. pY- has an ampicillin selection marker and a barnase gene with two amber mutations at Q2 and D44. The barnase gene is under control of a pBad promoter.

**3.2 Construction of pET-pylT-GFP**

Plasmid pET-pylT-GFP was derived from the plasmid pAcKRS-pylT-GFP1Amber in which $GFP_{UV}$ has an amber mutation at Q204.[6,7] The restriction enzyme *BglII* was used to cut off the ACKRS gene. The digested pAcKRS-pylT-GFP1Amber plasmid was ligated to form pET-pylT-GFP.

**3.3 Construction of pET-pylT-Z**

The pET-pylT-Z plasmid was derived from pET-pylT-GFP. The Z-domain gene was amplified from the pLeiZ plasmid.[8] This gene has an amber mutation at the K7 position. Two restriction sites, *NdeI* at the 5' end and *SacI* at the 3' end, were introduced in the PCR product which was subsequently digested and used to replace $GFP_{UV}$ in pET-pylT-GFP.

**3.4 Constraction of pET-Z**

The pET-Z plasmid was derived from pET-pylT-GFP. Wild type Z domain gene was amplified from the pLeiZ plasmid. The amber mutation at K7 was reversed in the PCR product using a primer that contained an AAA codon at this position. Two restriction sites, *NdeI* at the 5' end and *SacI* at the 3' end, were introduced in the PCR product which was subsequently digested and used to replace $GFP_{UV}$ in pET-pylT-GFP.

**4. Construction of the pRS1 Library**[1]

The plasmid pBK-mmPylRS that encodes wild-type *Methanosarcina mazei* PylRS was derived from a pBK plasmid containing *p*-iodophenylalanyl-tRNA synthetase.[9] The pylRS gene is under the control of *E. coli glnS* promoter and terminator. It was amplified from genomic DNA of *Methanosarcina mazei* strain DSM 3647 (ATCC) by flanking primers, pBK-mmPylRS-NdeI-F and pBK-mmPylRS-PstI~NsiI-R. To construct the pRS1 library, NNK (N=A or C or G or T, K=G or T) mutations were introduced at six sites by overlap extension PCR.[10] The following pairs of primers were used to generate a PylRS gene library with randomization at six sites: (1) pBK-mmPylRS-NdeI-F (5'-gaatcccatatggataaaaaaccactaaacactctg-3') and mmPylRS-Mutlib01-R (5'-ggccctgtcaagcttgcgmnngtagttmnnmnngtttggagcaagca tggg-3'); (2) mmPylRS-Mutlib02-F (5'-cgcaagcttgacagggccctgcctgatcc-3') and mmPylRS-Mutlib03-R (5'-gcatcccgatcccatctgmnngaamnncagcatggtaaactcttc-3'); (3) mmPylRS-Mutlib04-F (5'-cagatgggatcgggatgcacacg-3') and mmPylRS-Mutlib05-R (5'-

ccgaaacctgcccctatmnnngggtttatcaatacccca-3'); (4) mmPylRS-Mutlib06-F (5'-atagggcaggtttcgggctcgaacgcc-3') and pBK-mmPylRS-PstI~NsiI-R (5'-gtttgaaaatgcatttacaggttggtagaaatccc-3'). The gene library was digested with the restriction enzymes *NdeI* and *NsiI*, gel-purified, and ligated back into the pBK vector that was digested by *NdeI* and *PstI* to afford the pRS1 plasmid library. 1 μg of the ligation products were then electroporated into *E. coli* Top10 cells. Electroporated cells were recovered in the SOC medium for 60 min at 37 °C, transferred into a 2 L 2YT medium with kanamycin (25 μg/mL), and then incubated at 37°C to $OD_{600}$ at 1.0. To calculate the library size, 1 μL recovered SOC culture was subjected to serial dilutions in 2YT, plated on LB agar plates with kanamycin (25 μg/mL), and then grown overnight in a 37°C incubator. Based on the colony numbers on these plates, the pRS1 library contains approximately $1.01 \times 10^9$ independent transformants. Sequencing pylRS variants in 20 clones did not reveal any significant bias at the randomization sites.

**5. Selection Procedure for Evolving Pyrrolysyl-tRNA Synthetase[1,7-8]**

The selections followed the scheme shown in **Supplementary Scheme 1**. For the positive selection, the pRS1 library was used to transform *E. coli* TOP10 competent cells harboring pY+ to yield a cell library greater than $1 \times 10^9$ cfu, ensuring complete coverage of the pRS1 library. Cells were plated on LB agar plates containing 12 μg/mL tetracycline (Tet), 25 μg/mL kanamycin (Kan), 68 μg/mL chloramphenicol (Cm) and 1 mM **3**. After incubation at 37°C for 72 h, colonies on the plates were collected. Total plasmids were isolated and separated by 1 % agarose gel electrophoresis. pRS1 plasmids were extracted using a Gel-extraction kit (QIAGEN). The extracted pRS1 plasmids from the positive selection were used to transform *E. coli* TOP10 harboring pY− for the negative selection. After electroporation, the cells were allowed to recover for 1 h at 37°C in SOC media before being plated on LB agar plates containing 50 μg/mL Kan, 200 μg/mL ampicillin (Amp) and 0.2% arabinose. The plates were incubated for 16 h at 37°C. Survived cells were then pooled and pRS1 plasmids were extracted. The selection power to exclude out the mutants that also took endogenous amino acids was tested on LB agar plates containing 50 μg/mL Kan, 200 μg/mL Amp, 0.2% arabinose and 1mM **3**. The plate contains 1 mM **3** showed much fewer colony numbers as the rounds of negative selections increased. Five alternative selections (three positive + two negative) finally yielded many colonies. 16 single colonies after the third positive selection were selected and the plasmids were isolated for sequencing. 4 single colonies from the third positive selection were also chosen for testing their ability to grow on plates with 102 μg/mL chloramphenicol, 25 μg/mL Kan, 12 μg/mL Tet, and 1 mM of **3** or 5 mM of **4**. A plate without NAA supplementary was used as a control. Images of colonies growing on different plates were shown in **Supplementary Figure 7**.

The positive selection of PylRS mutants specific for L-phenylalanine was carried out similarly as that for p-iodo-L-phenylalanine except that no noncanonical amino acid was supplemented into the medium.

## 6. Protein Expression and Purification

### 6.1 $GFP_{UV}$ expression and purification

To express $GFP_{UV}$ incorporated with a NAA, we cotransformed *E. Coli* BL21(DE3) cells with pBK-IFRS1 and pET-pylT-GFP. Cells were recovered in 1 mL of the LB medium for 1 h at 37 ºC before being plated on a LB agar plate containing Kan (50 μg/mL) and Amp (100 μg/mL). A single colony was then selected and grown overnight in a 10 mL culture. This overnight culture was used to inoculate 100 mL of the GMML medium (M9 minimal media supplemented with 1% glycerol, 300 μM leucine, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2% NaCl, 25 μg/mL Kan and 100 μg/mL Amp). Cells were grown at 37ºC in an incubating shaker (300 r.p.m.) and protein expression was induced by adding IPTG to a final concentration of 1 mM and **3** to a final concentration of 1 mM ( or **4** to a final concentration of 5 mM) when $OD_{600}$ reached 0.7. After 6 h induction, cells were harvested, resuspended in a lysis buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 10 mM imidazole, pH 8.0), and sonicated. The cell lysate was clarified by centrifugation (60 min, 11,000 ×g, 4ºC). The supernatant was injected into a 30 mL $Ni^{2+}$-NTA column (Qiagen) on FPLC (ÄKTApurifier™, GE Healthcare Bio-Sciences Corp.) and washed with 60 mL lysis buffer. Protein was finally eluted out by running an imidazole gradient from 40 mM to 250 mM in lysis buffer. Semipure fractions were collected and concentrated to perform hydroxylapatite chromatography in a 30 mL hydroxylapatite column on FPLC. The protein sample was washed with 60 mL wash buffer (10 mM sodium phosphate, pH 6.8). The full-length $GFP_{UV}$ was finally eluted out by running a sodium phosphate gradient from 10 mM to 400 mM in washing buffer. The $GFP_{UV}$ fractions were collected and concentrated. The buffer was later exchanged with 1 mM ammonium bicarbonate using an Amicon Ultra -15 Centrifugal Filter Devices (10,000 MWCO cut) (Millopore). The purified proteins were analyzed by 12% SDS-PAGE. $GFP_{UV}$ proteins incorporated with L-phenylalanine in FRS1 and FRS2 were expressed and purified similarly with the addition of **2** to the GMML medium to a final concentration of 1 mM.

### 6.2 Z domain expression and purification

#### 6.2.1 Z-3 expression and purification

Z-domain proteins incorporated with **3** was expressed similarly to the expression of $GFP_{UV}$ proteins except pET-pylT-Z was used to cotransform *E. coli* BL21(DE3) together with pBK-IFRS1. 1 mM final concentration of **3** was used. The protein purification was performed under denaturing conditions using lysis buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, pH 8.0) supplemented with 6 M guanidinium chloride (GuCl) or 8 M urea. The cell was lysed in the lysis buffer supplemented with 6 M GuCl for 30 min in 37°C. The cell lysate was clarified by centrifugation (60 min, 11,000 ×g, 4ºC). The supernatant was injected into a 30 mL $Ni^{2+}$-NTA column (Qiagen) on FPLC (ÄKTApurifier™, GE Healthcare Bio-Sciences Corp) and washed with 60 mL lysis buffer containing 8M Urea (pH 6.2). Protein was finally eluted out by running a pH gradient from 6.2 to 4.5 in the lysis buffer containing 8M urea. Fractions containing Z-domain were collected and concentrated. Further purification was per-

formed on S200 gel filtration column (GE Healthcare Bio-Sciences Corp) using a running buffer (0.05 M sodium phosphate, 0.15 M NaCl, 4 M urea, pH 7.0). The purified fractions were collected and dialysis against 0.1× PBS (1 mM $Na_2HPO_4$, 0.18 mM $KH_2PO_4$, 13.7 mM NaCl, 0.27 mM KCl, pH 7.4) containing 1 M urea and then 0.1× PBS. The purified Z-domain was concentrated by using Amicon centriplus YM-3 (3,000 MWCO cut) (Millopore). The purified proteins were analyzed by 15% SDS-PAGE.

### 6.2.2 Wild-Type Z-Domain expression and purification

To express wild-type Z-domain proteins (**Z-wt**), we transformed *E. Coli* BL21(DE3) cells with pET-Z. Cells were recovered in 1 mL LB medium for 1 h at 37 ºC before being plated on a LB agar plate containing Kan (50 μg/mL) and Amp (100 μg/mL). A single colony was then selected and grown overnight in a 10 mL culture. This overnight culture was used to inoculate 100 mL LB medium supplemented with 100 μg/mL Amp. Cells were grown at 37ºC in an incubating shaker (300 r.p.m.) and protein expression was induced by adding IPTG to a final concentration of 1 mM when $OD_{600}$ reached 0.7. After 6 h induction, cells were harvested and resuspended in a lysis buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, pH 8.0) supplemented with 6 M guanidinium chloride (GuCl)). Further protein purification was the same as that for **Z-3**. The purified protein was analyzed by 15% SDS-PAGE.

## 7. Suzuki Coupling on the Z-domain Proteins

A water-souble palladium catalyst using 2-amino-4,6-dihydroxyprimidine as its ligand was prepared according to the literature protocol, and the coupling of **Z-3** with 3-(dansylamino)phenylboronic acid (DaFBA) was carried out accordingly with minor modification.[11] Z-wt was used as a control.

To a protein sample of The buffer of 50 μL of **Z-3** (0.5 mg/mL, 80 μL, 4.9 nmol) or **Z-wt** (0.4 mg/mL, 80 μL, 3.9 nmol) in 0.1× PBS solution (1 mM $Na_2HPO_4$, 0.18 mM $KH_2PO_4$, 13.7 mM NaCl, 0.27 mM KCl, pH 7.4) was added the palladium catalyst in water (10 mM, 7 μL, 70 nmol), aqueous formic acid (20 mM, 3 μL, 60 nmol) and DaFBA in DMSO (20 mM, 20 μL, 400 nmol). The mixture was vortexed and then heated in a 35 $^{o}$C water bath for 6 h. After dialysis against 0.1× PBS overnight (2L × 2, 8 h each time) to remove excessive dye and catalyst, the protein samples were lyophilized, redissolved in 10 μL of 8 M urea, and analyzed by the SDS-PAGE (15%) electrophoresis.

## 8. Protein LC-ESI-MS Analysis

An Agilent (Santa Clara, CA) 1200 capillary HPLC system was interfaced to an API QSTAR Pulsar Hybrid QTOF mass spectrometer (Applied Biosystems/MDS Sciex, Framingham, MA) equipped with an electrospray ionization (ESI) source. Liquid chromatography (LC) separation was achieved using a Phenomenex Jupiter C4 microbore column (150 × 0.50 mm, 300 Å) (Torrance, CA) at a flow rate of 10 μL per min. The proteins were eluted using a gradient of (A) 0.1% formic acid versus (B) 0.1% formic acid in acetonitrile. The gradient timetable was as follows: 2% B for 5

min, 2-30% in 3 min, 30-60% in 44 min, 60-95% in 8 min, followed by holding the gradient at 95% for 5 min, for a total run time of 65 min. The MS data were acquired in positive ion mode (500-1800 Da) using spray voltage of +5000 V. BioAnalyst software (Applied Biosystems) was used for spectral deconvolution. For the GFPuv protein analysis, a mass range of m/z 500-1800 was used for deconvolution and the output range was 10000-50000 Da using a step mass of 0.1 Da and a S/N threshold of 20. For the Z-Domain protein analysis, a mass range of m/z 500-2000 was used for deconvolution and the output range was 5000-15000 Da for Z-domain-His6X using a step mass of 0.1 Da and a S/N threshold of 20.

**9. Tandem Mass Spectrametry analysis**

$GFP_{UV}$ variants from the SDS-PAGE gels was cut, dissolved in 25 mM Ammonia bi-carbonate, and denatured at 90 degree for 15 min. Proteinase Asp-N (Roche) was dissolved in 0.01% TFA (pH 3). Proteinase Asp-N solution was added to the substrate protein solution (w/w=1:50), and incubated at 37 degree overnight. Peptides resulting from the proteinase Asp-N digestion were mixed 1:1 (v/v) with matrix (5 mg $mL^{-1}$ α-cyano-4-hydroxycinnamic acid, 50% (v/v) acetonitrile, 10 mM ammonium dihydrogen phosphate, 1% TFA) and 1 μL of the resulting mixture was spotted onto a stainless steel target plate. Mass spectra and tandem MS spectra were collected using an Applied Biosystems 4800 Tof/Tof (Framingham, MA). Collision induced dissociation tandem MS spectra were acquired using air at the medium pressure setting and at 2 kV of collision energy. Tandem MS data was manually interpreted using the Data Explorer™ software package (Applied Biosystems, Framingham, MA).

**10. References**


1 Wang, Y. -S.; Wu, B.; Wang, Z.; Huang, Y.; Wan, W.; Russell, W. K.; Pai, P.-J.; Moe, Y. N.; Russell, D. H.; Liu, W. R. *Mol Biosyst* 2010, **6**, 1557-1560.

2 S. W. Santoro, L. Wang, B. Herberich, D. S. King and P. G. Schultz, *Nat. Biotechnol.*, 2002, **20**, 1044-1048

3 P. Talaga, C. Benezra and J.-L. Stampf, *Bioorg. Chem.,* 1990, **18**, 199-206.

4 S. Wiejek, E. Masiukiewiez, B. Rzeszotarska, *Chem. Pharm. Bull*., 2001, **49**, 1189-1191.

5 Y. Tatsu, Y. Shigeri, A. Ishida, I. Kameshita, H. Fujisawa and N. Yumoto, *Bioorg. Med. Chem. Lett*., 1999, **9**, 1093-1096.

6 Y. Huang, W. K. Russell, W. Wan, P. J. Pai, D. H. Russell and W. Liu, *Mol. Biosyst.*, 2010, **6**, 683-686.

7 Y. Huang, W. Wan, W. K. Russell, P. J. Pai, Z. Wang, D. H. Russell and W. Liu, *Bioorg. Med. Chem.*, 2010, **20**, 878-880.

8 J. Xie, L. Wang, N. Wu, A. Brock, G. Spraggon and P. G. Schultz, *Nat. Biotechnol.*, 2004, **22**, 1297-1301.

9 L. Wang, Z. Zhang, A. Brock and P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 56-61.

10 J. P. Tam, Q. Yu and Z. Miao, *Biopolymers*, 1999, **51**, 311-332.

11 J. M. Chalker, C. S. C. Wood and B. G. Davis, *J. Am. Chem. Soc.* 2009, **131**, 16346-16347.

## 11. Supplementary Tables, Schemes, and Figures

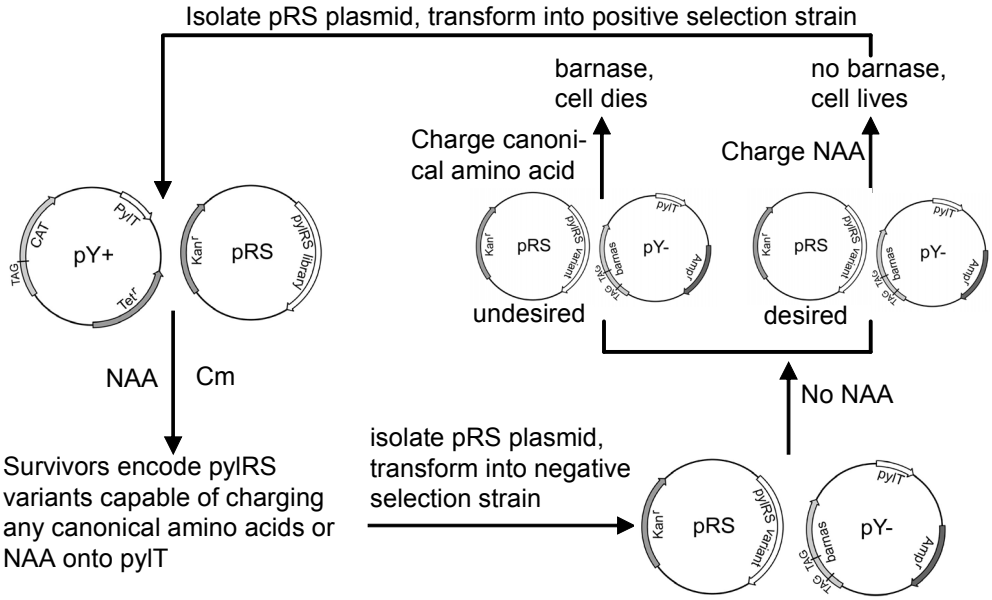
***Supplementary Table 1.*** $GFP_{UV}$ and Z-domain expression yields and MS characterizations

| Proteins | Yield (mg/L)[j] | Calculated Mass (Da) | Detected Mass (Da) |
|---|---|---|---|
| **GFP-3** [a] | 1.0 | 27855[f] | 27855 |
| **GFP-4** [b] | 0.8 | 27808[f] | 27810 |
| **Z-3** [c] | 1.5 | 8281[g] | 8280 |
| | | 8192[h] | 8191 |
| | | 8150[i] | 8149 |
| **GFP-F**[d] | 1.6 | 27729[f] | 27730 |
| **GFP-F**[e] | 2.9 | 27729[f] | 27729 |

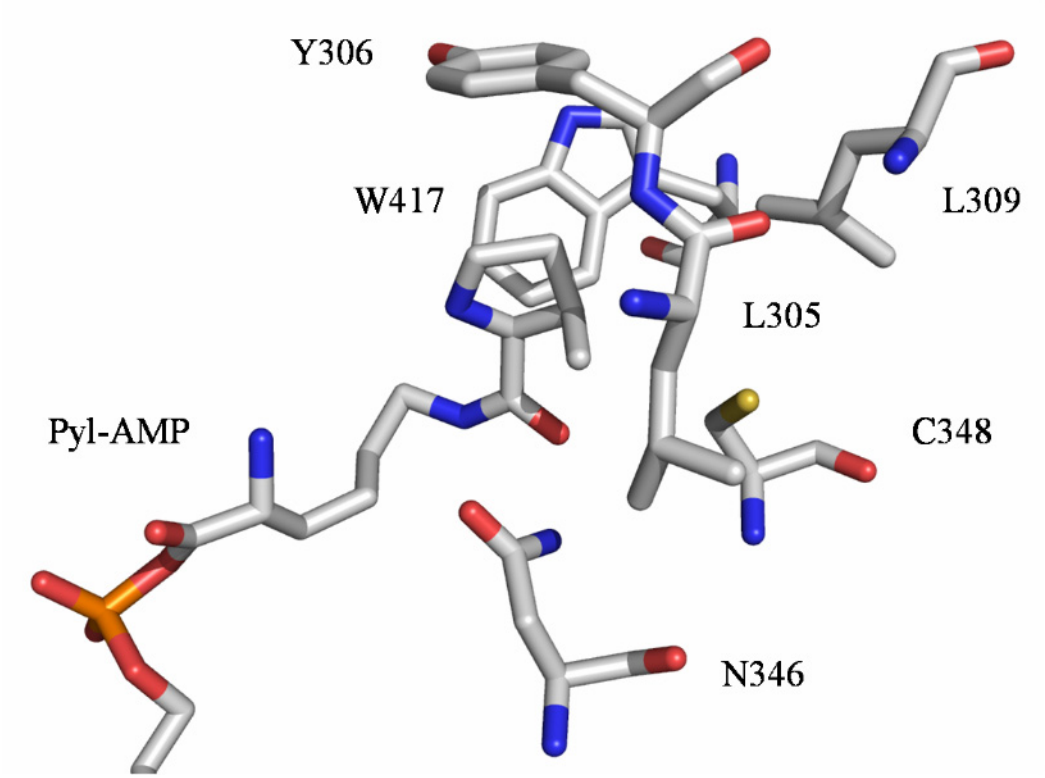
[a]$GFP_{UV}$ incorporated with *p*-iodo-L-phenylalanine at Q204.
[b]$GFP_{UV}$ incorporated with *p*-bromo-L-phenylalanine at Q204.
[c]Z-domain incorporated with *p*-iodo-L-phenylalanine at K7.
[d]$GFP_{UV}$ incorporated with L-phenylalanine at Q204 that was expressed in the LB medium using the FRS1- $tRNA^{Pyl}_{CUA}$ pair.

[e]$GFP_{UV}$ incorporated with l-phenylalanine at Q204 that was expressed in the LB medium using the FRS2- $tRNA^{Pyl}_{CUA}$ pair.

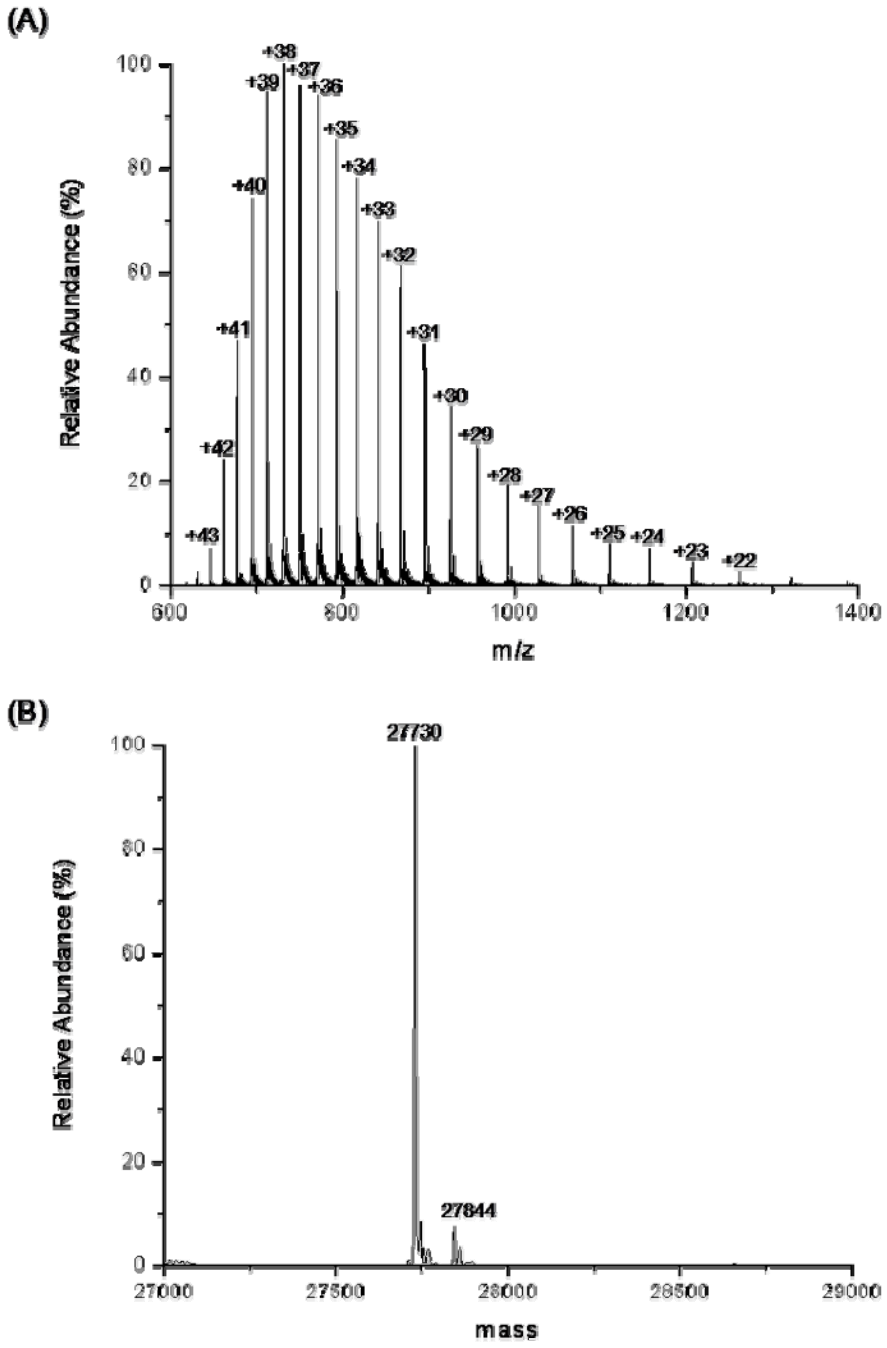
[f]Full-legnth $GFP_{UV}$ proteins without N-terminal methionine.
[g]Full-legnth Z domain proteins.
[h]Full-length Z domain without N-terminal methionine but with an N-terminal acetylation.
[i]Full-length Z-domain without N-terminal methionine.
[j]The yields were determined by the BCA protein assay.



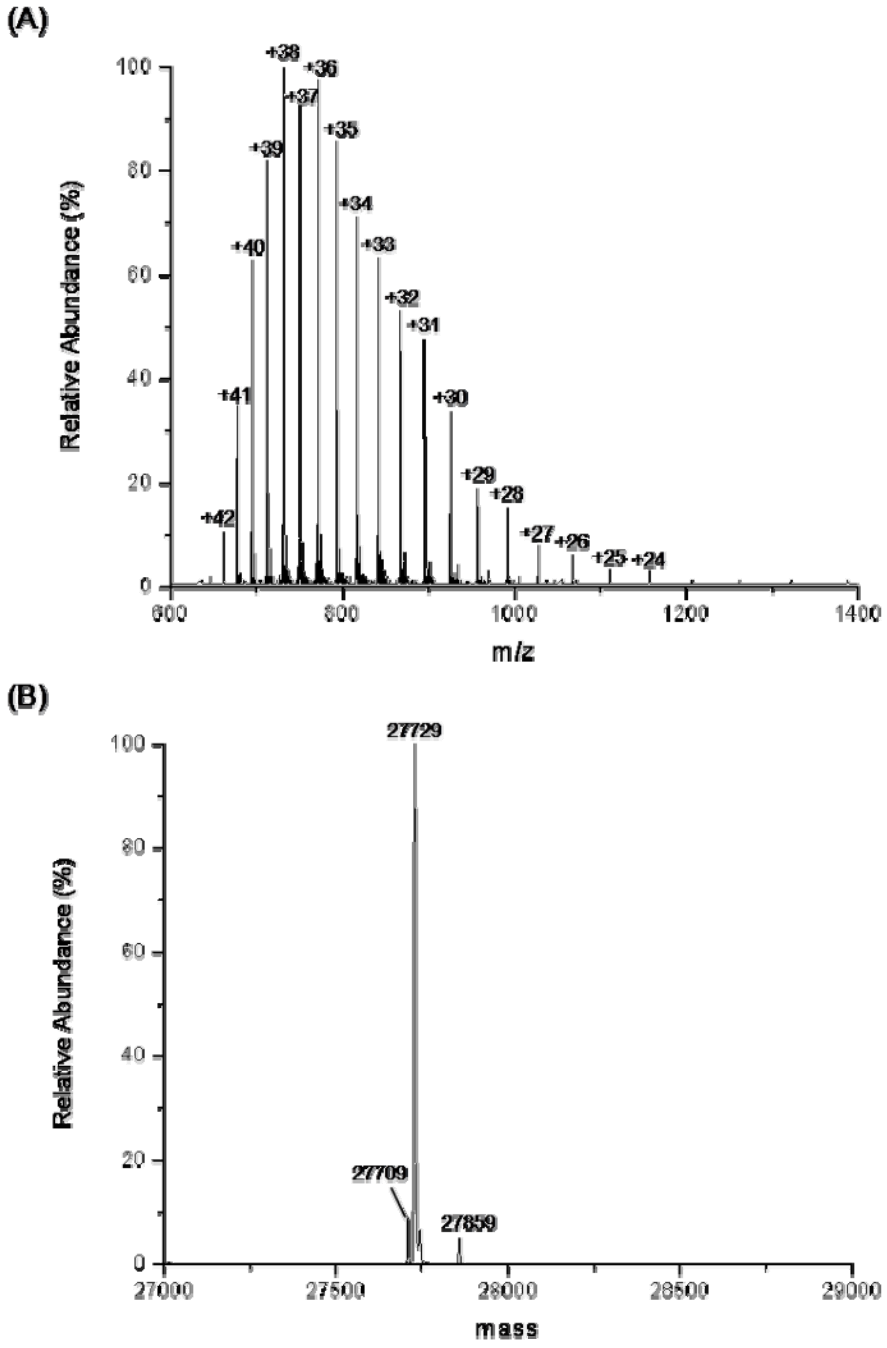

***Supplementary Scheme 1***. The selection scheme to identify PylRS variants specific for a noncanonical amino acid.




***Supplementary Figure1***. The active site of PylRS. The structure was derived from the PDB entry: 2Q7E.

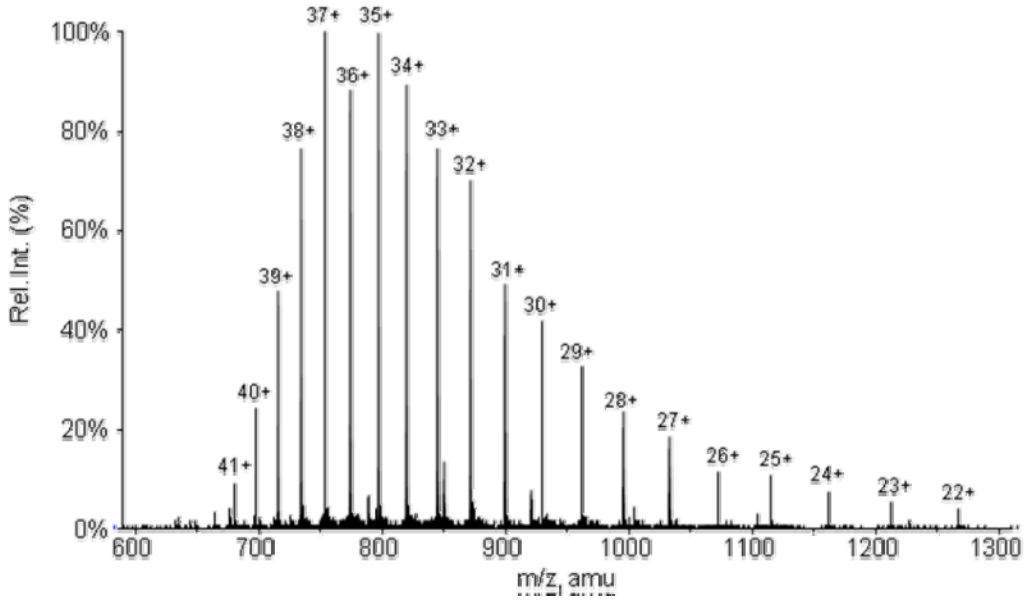



***Supplementary Figure 2***. Mass determination of $GFP_{UV}$ incorporated with L-phenylalanine at Q204 using the FRS1- $tRNA^{Pyl}_{CUA}$ pair. (**A**) The ESI-MS spectrum and (**B**) the deconvoluted ESI-MS spectrum.

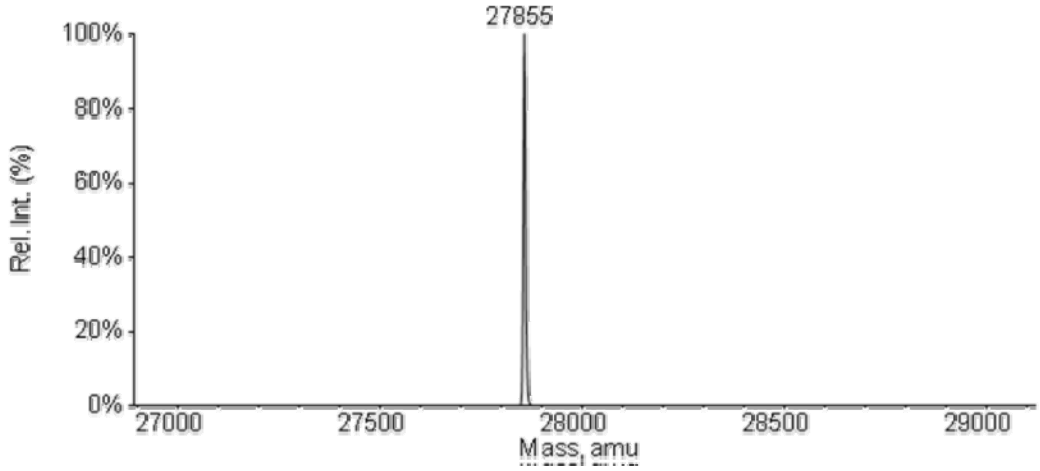



***Supplementary Figure 3***. Mass determination of $GFP_{UV}$ incorporated with L-phenylalanine at Q204 using the FRS2- $tRNA^{Pyl}_{CUA}$ pair. (**A**) The ESI-MS spectrum and (**B**) the deconvoluted ESI-MS spectrum.

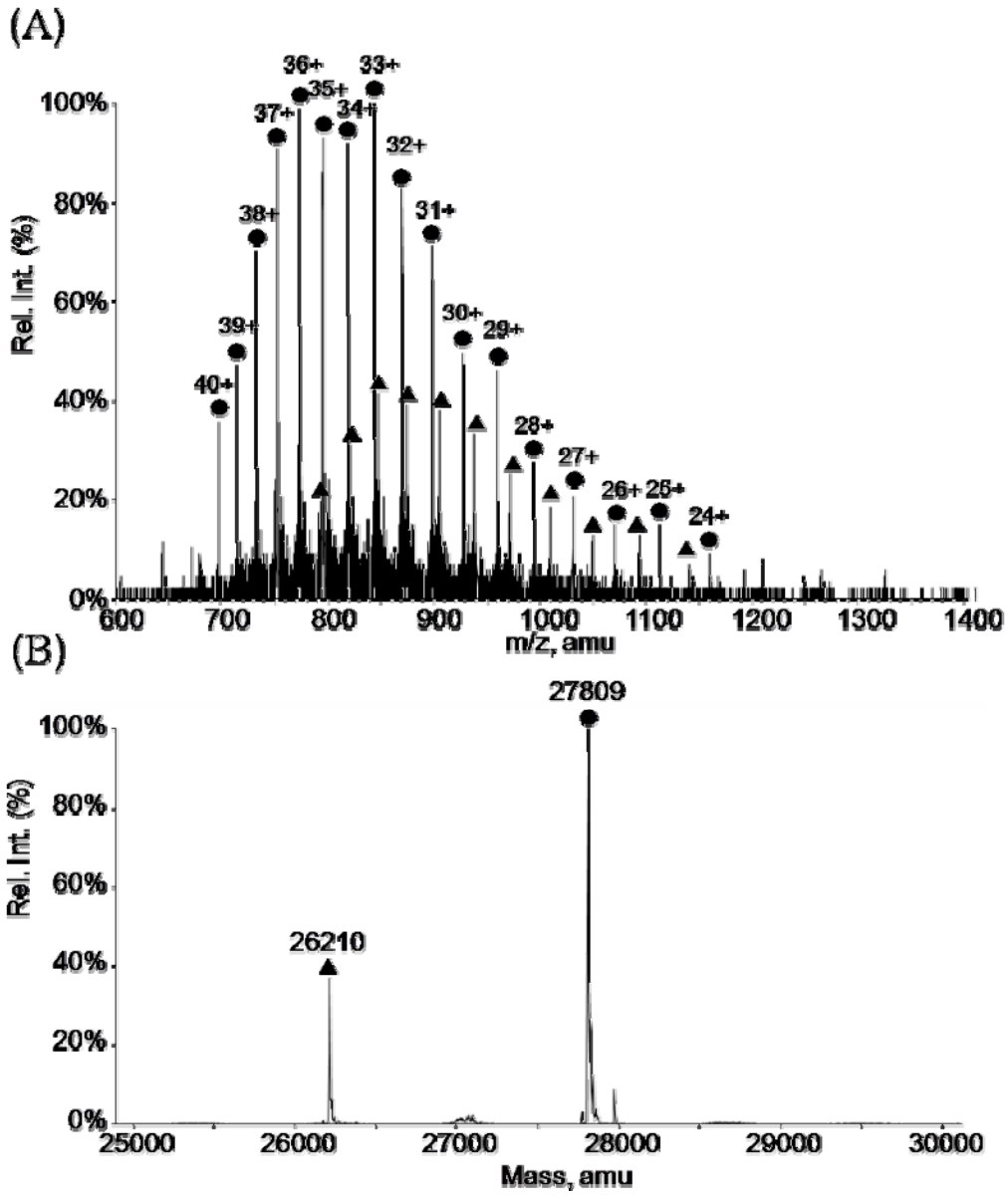
**A**



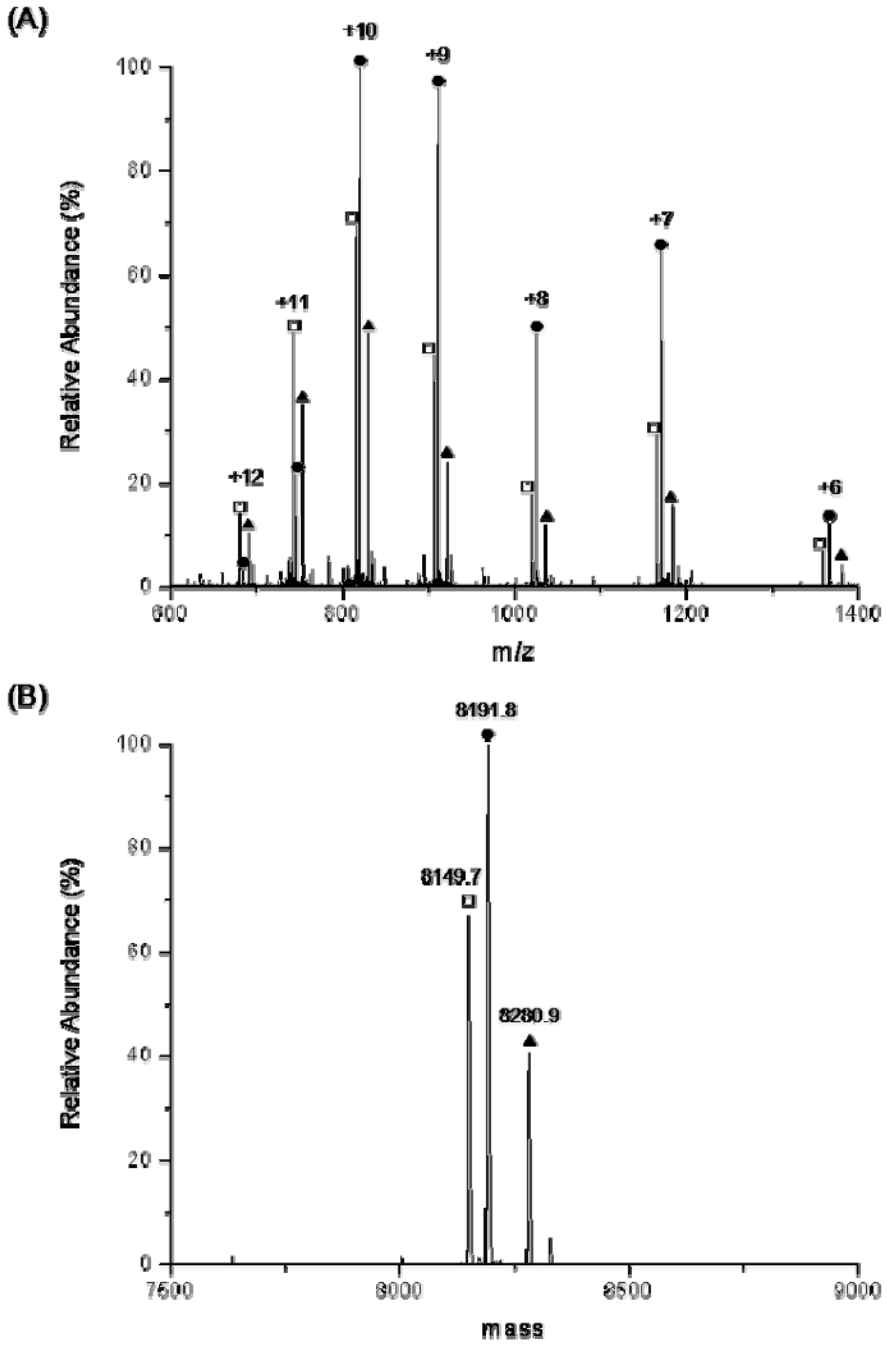

**B**



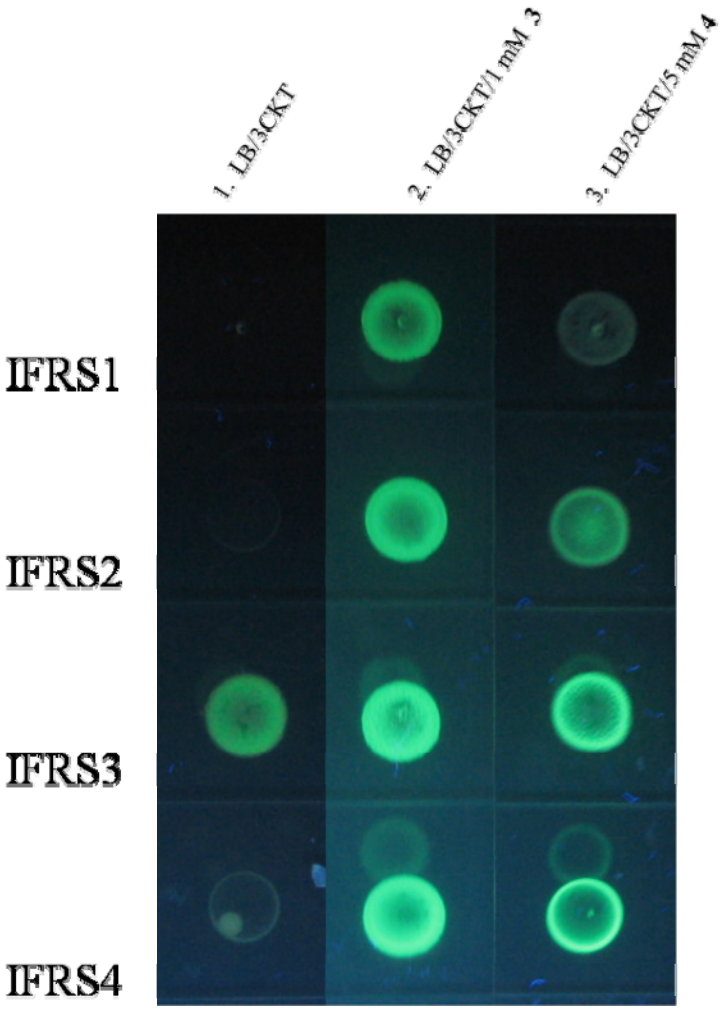

***Supplementary Figure 4***. Mass characterization of **GFP-3**. (**A**) The ESI-MS spectrum and (**B**) the deconvoluted ESI-MS spectrum.



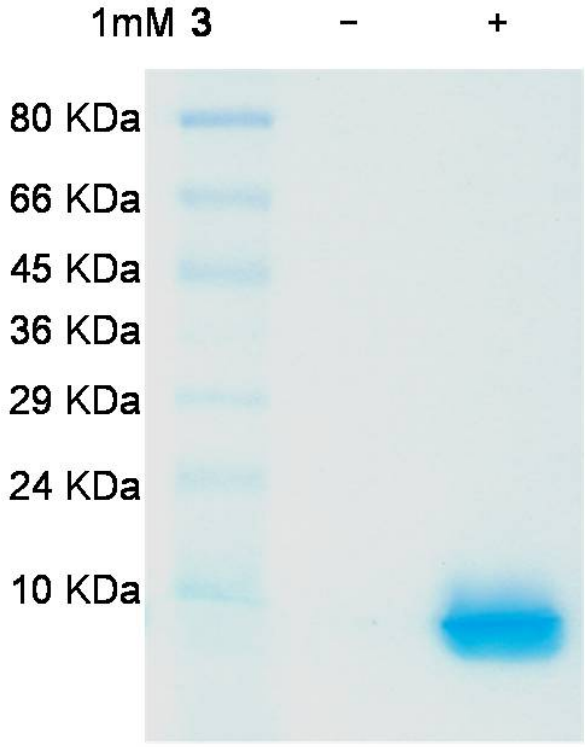

***Supplementary Figure 5***. Mass characterization of **GFP-4**. (**A**) The ESI-MS spectrum and (**B**) the deconvoluted ESI-MS spectrum.



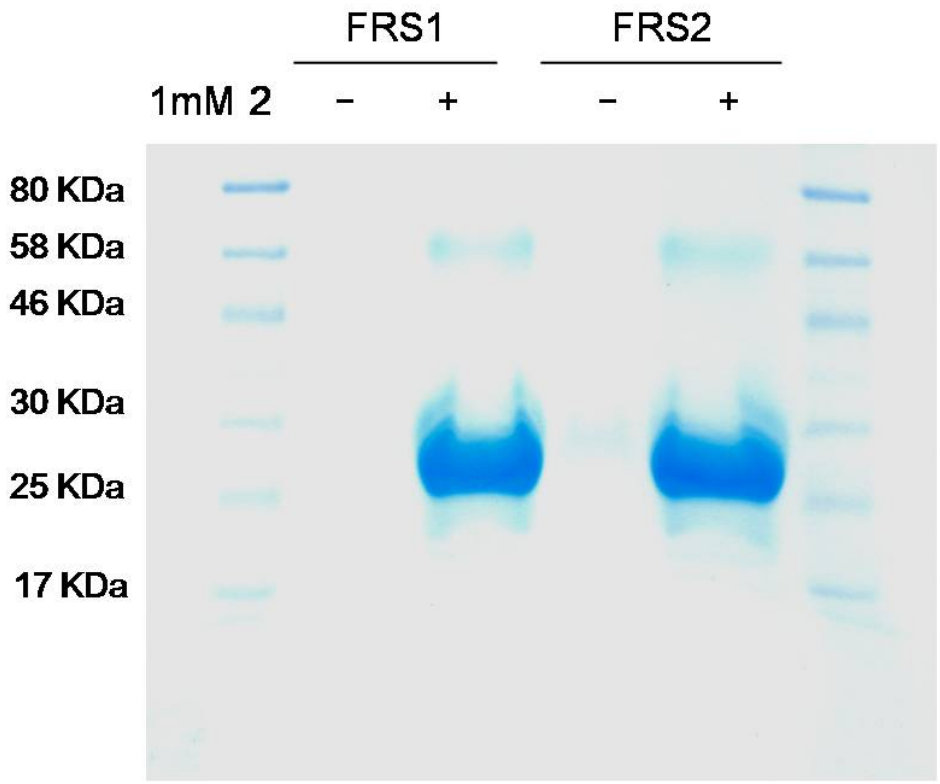

***Supplementary Figure** 6*. Mass characterization of **Z-3**. (**A**) The ESI-MS spectrum and (**B**) the deconvoluted ESI-MS spectrum.



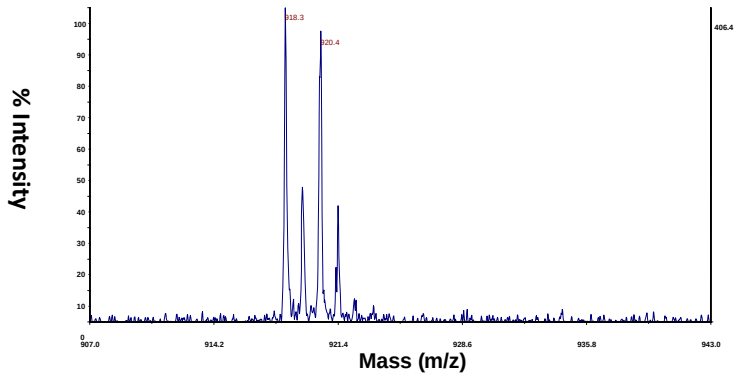

***Supplementary Figure 7***. Growth of 4 selected IFRS mutants from the third positive selection of **3** on LB plates with different supplements. (**1**) Growth on LB/3CKT (LB agar plates containing 102 μg/mL Cm, 25 μg/mL Kan, and 12 μg/mL Tet); (**2**) Growth on plates containing 1 mM **3**, 102 μg/mL Cm, 25 μg/mL Kan, and 12 μg/mL Tet; (**3**) Growth on plates containing 5 mM **4**, 102 μg/mL Cm, 25 μg/mL Kan, and 12 μg/mL Tet. All the colonies were cultured at 37°C for 24 hours. Images were taken under UV 365 nm radiation. The pY+ plasmid has a $GFP_{UV}$ gene under control of a T7 promoter and a T7 RNA polymerase gene that contains two amber mutations at positions 1 and 107. The expression of $GFP_{UV}$ is promoted by the suppression of two amber mutations in the T7 RNA polymerase. The fluorescent intensity of the expression of $GFP_{UV}$ roughly represents the suppression efficiency at amber codons.




***Supplementary Figure 8***. The expression of Z-domain containing an amber mutation at K7. Proteins were expressed in BL21(DE3) cells that grew in minimal media supplemented with 1% glycerol and 1 mM **3**. The proteins were analyzed by SDS-PAGE (15%) gel electrophoresis with gelcole blue staining.



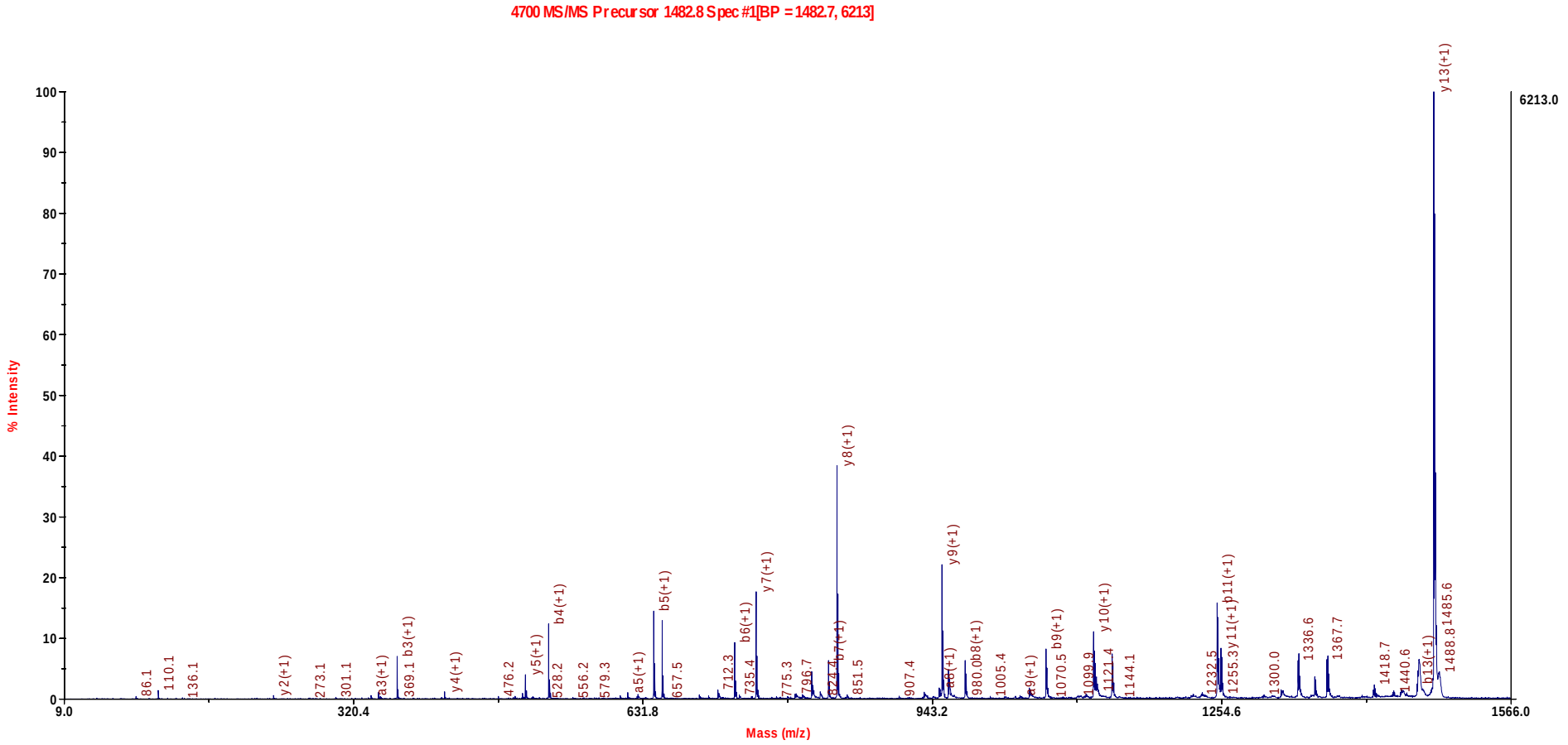

***Supplementary Figure 9***. The expression of $GFP_{UV}$ containing an amber mutation at Q204. Proteins were expressed in BL21(DE3) cells that grew in minimal media supplemented with 1% glycerol and 1 mM **2**. The proteins were analyzed by SDS-PAGE (12%) gel electrophoresis with GelCode blue staining.



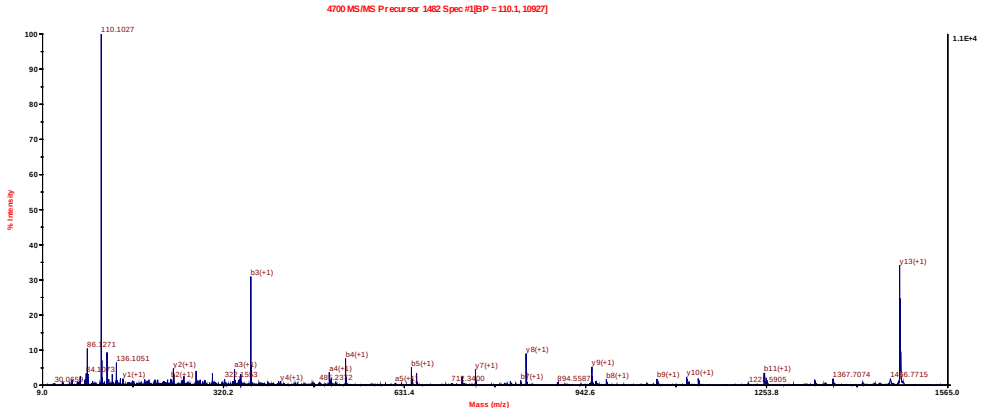

***Supplementary Figure 10***. Evidence for the predicted two bromine isotopes is observed in the y7 fragment of *p*-bromo-L-phenylalanine-containing DNHYLSTF*SALSK (F* denotes *p*-bromo-L-phenylalanine). Spectra were taken on an AB-Sciex 4800 plus mass spectrometer.

## 12. Supplementary Tandem Mass Spectra

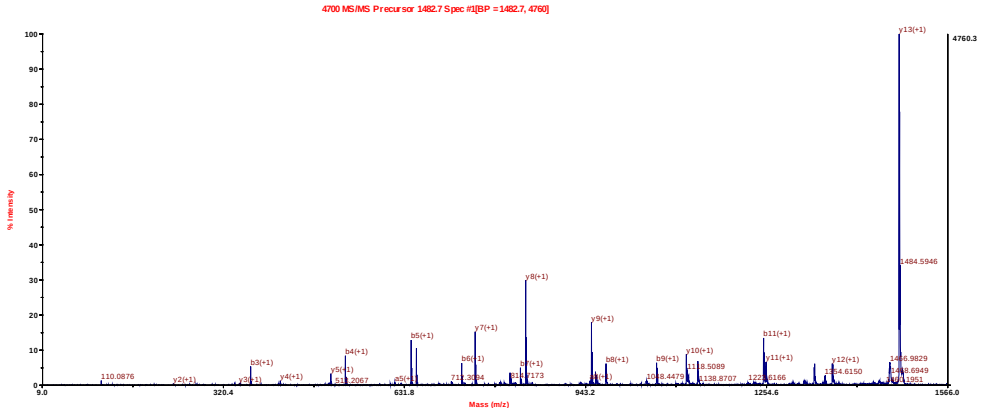
# DNHYLSTF*SALSK (F* denotes L-phenylalanine) from $GFP_{UV}$ expressed using FRS1



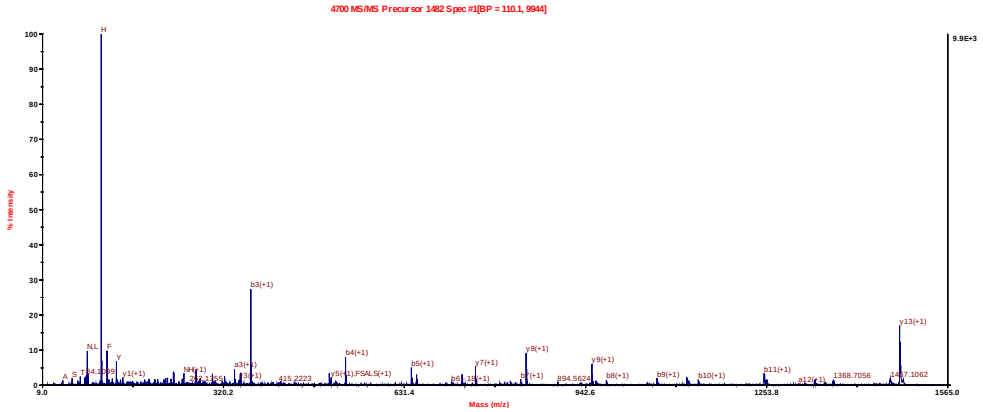
# HIGH GAS mode of DNHYLSTF*SALSK (F* denotes L-phenylalanine) from $GFP_{UV}$ expressed using FRS1



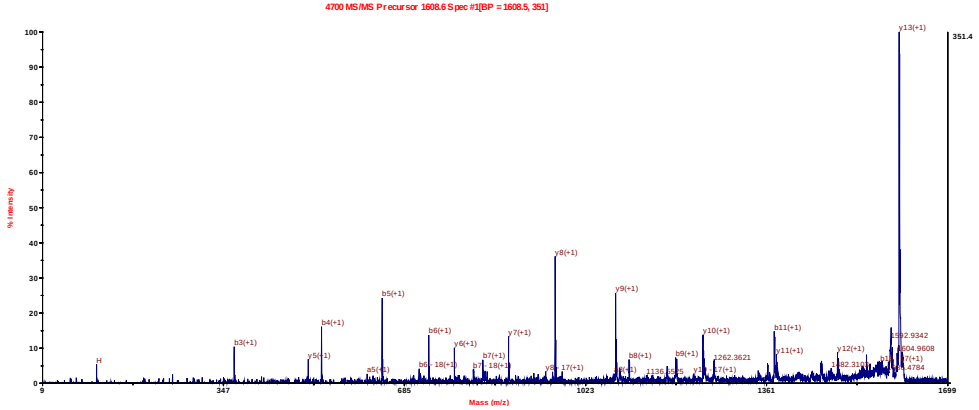
# DNHYLSTF*SALSK (F* denotes L-phenylalanine) from $GFP_{UV}$ expressed using FRS2



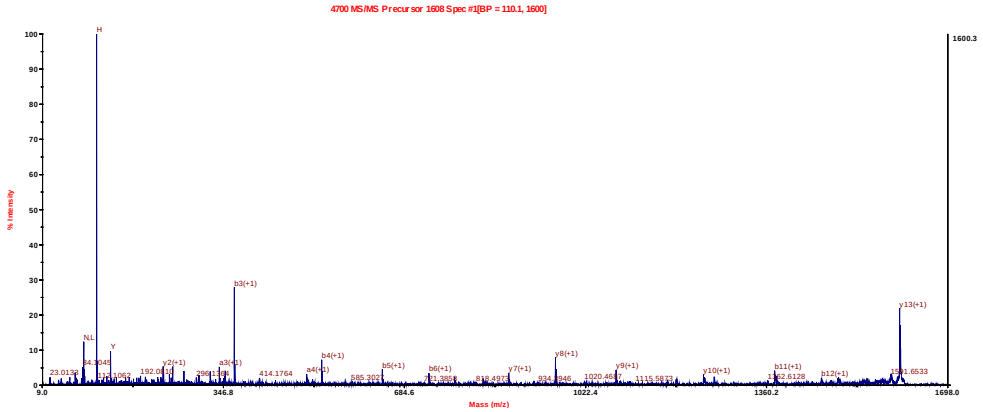
# HIGH GAS mode of DNHYLSTF*SALSK (F* denotes L-phenylalanine) from $GFP_{UV}$ expressed using FRS2



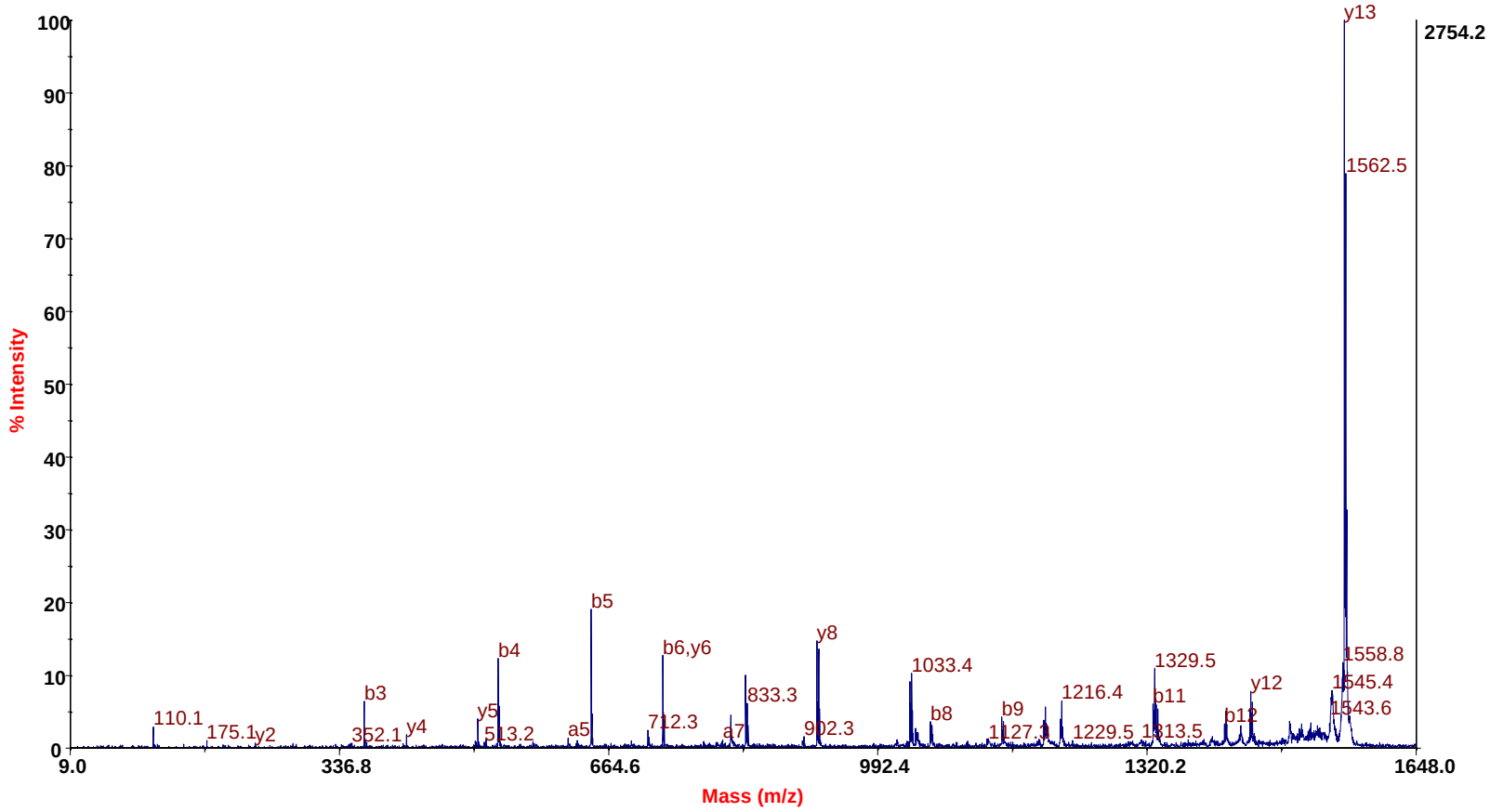
# DNHYLSTF*SALSK (F* denotes *p*-iodo-L-phenylalanine)



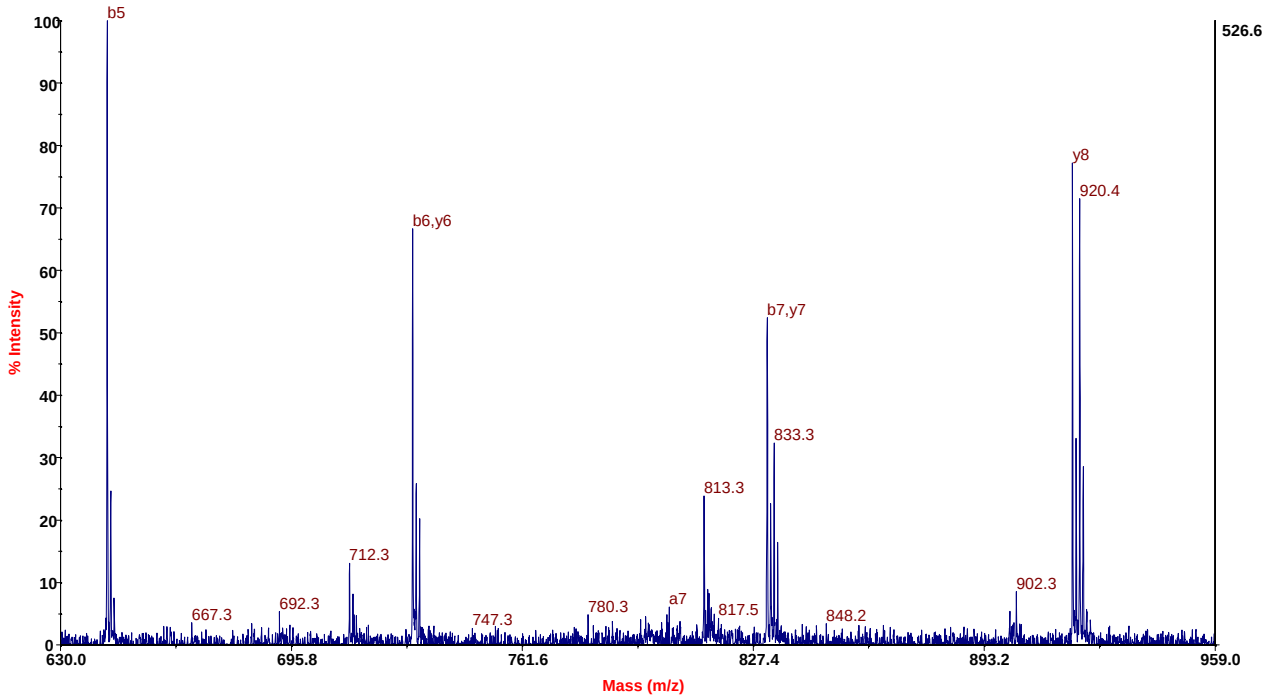
# HIGH GAS mode of DNHYLSTF*SALSK (F* denotes *p*-iodo-L-phenylalanine)



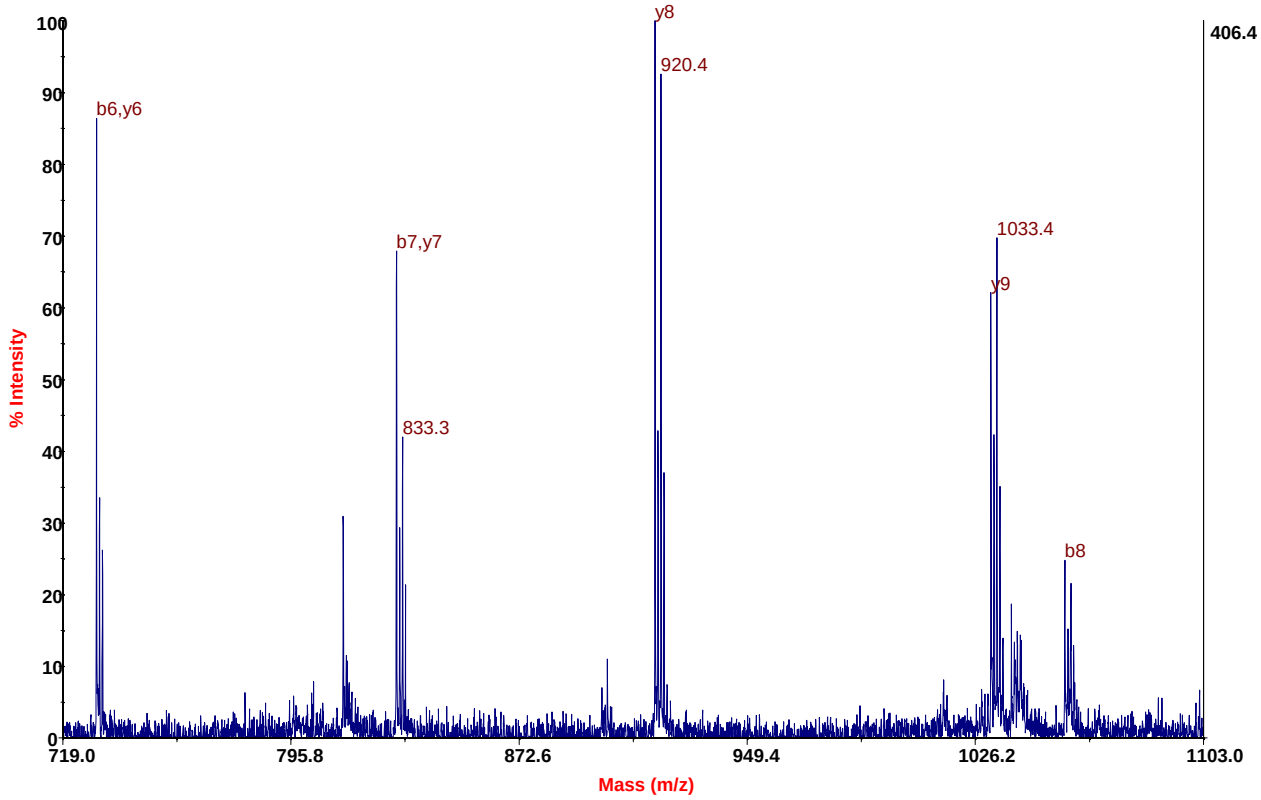
# DNHYLSTF*SALSK (F* denotes *p*-bromo-L-phenylalanine)



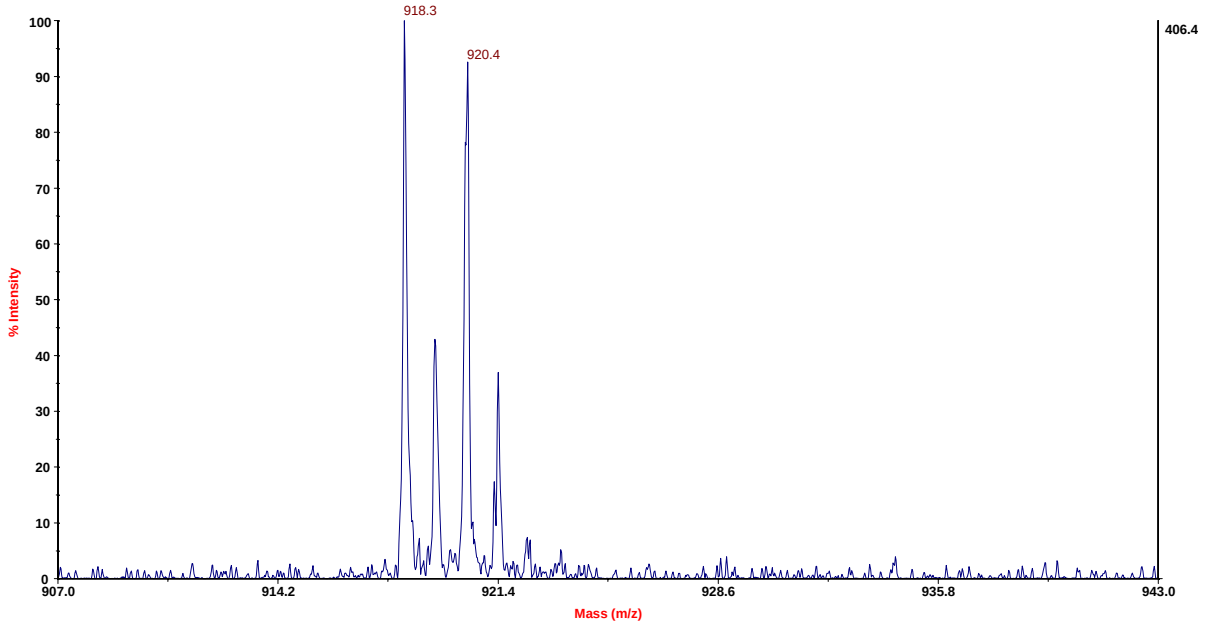
# DNHYLSTF*SALSK (F* denotes *p*-bromo-L-phenylalanine) (zoom in around the region from 600-1000 Da)



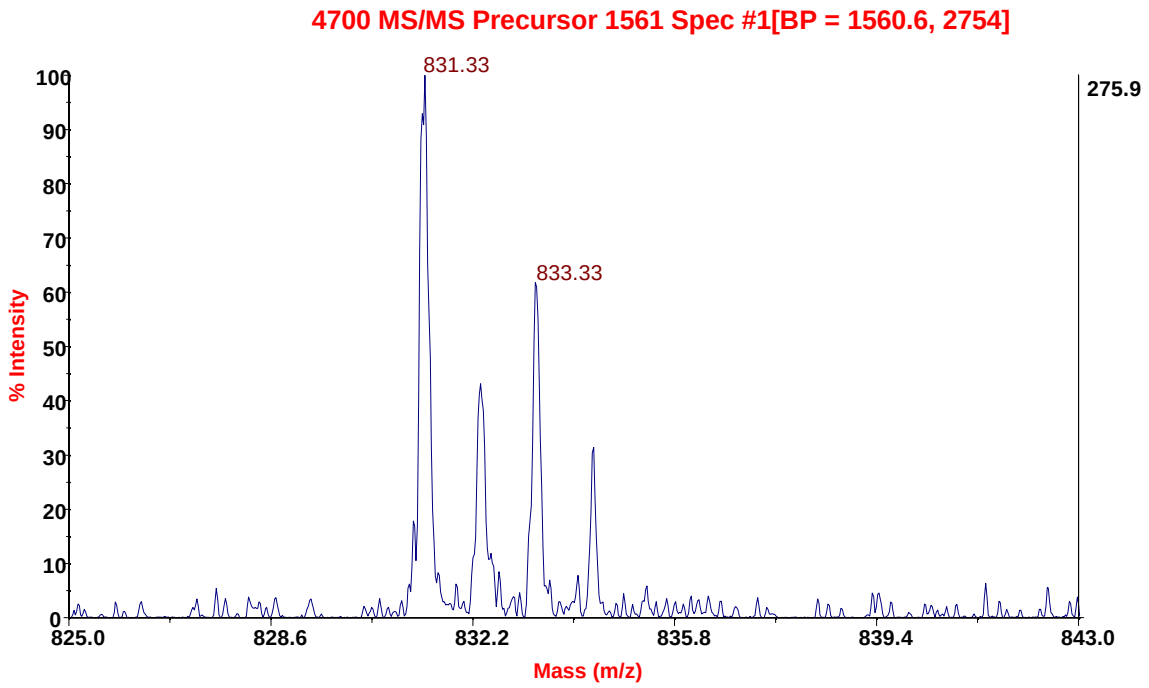
# DNHYLSTF*SALSK (F* denotes *p*-bromo-L-phenylalanine) (zoom in around the region from 700-1100 Da)



# DNHYLSTF*SALSK (F* denotes *p*-bromo-L-phenylalanine) Y8 zoom for isotopes



# DNHYLSTF*SALSK (F* denotes *p*-bromo-L-phenylalanine) zoom on y7 ion




You can clearly see the isotopes of Br have been incorporated