



國立臺灣大學生命科學院生化科學研究所

碩士論文

Graduate Institute of Biochemical Sciences

College of Life Science

National Taiwan University

Master Thesis

吡咯離胺醯-tRNA 合成酶對於掌性非典型胺基酸嵌入  
蛋白之研究

Releasing chiral selectivity of pyrrolysyl-tRNA synthetase  
for encoding D-noncanonical amino acids into proteins

王伊慧

Yi-Hui Wang

指導教授：王彥士 博士

Advisor: Yane-Shih Wang, Ph.D.

中華民國 107 年 7 月

July 2018



## 摘要

在合成生物學的蛋白設計應用中，利用蛋白轉譯系統中重新定義基因密碼，進而在體內將非典型胺基酸嵌入生合成蛋白的特定位置中，是探討酵素催化的機制以及設計含有新穎化學功能蛋白質的重要方法。雖然自然界存在 D-胺基酸，但是核醣體中蛋白胜肽鏈合成的單體皆為 L-胺基酸，且文獻報導中生物正交配對的體內嵌入非典型胺基酸皆為 L-胺基酸。此項研究中，透過演化吡咯-轉核醣核酸合成酶(PyIRS)的 N 端以及與活性中心，演化的 PyIRS，N-DFRS，能在大腸桿菌中分別辨認 D-或 L-非典型胺基酸，六種苯丙胺酸衍生物(D/LFA)，單個及多個位點位置特異嵌入 DFA 於超折疊綠螢光蛋白中，發現 DFA 影響蛋白內部疏水結構，造成單位螢光強度下降。N-DFRS 含活性中心的 C 端蛋白與 D/LFA 六個共結晶的蛋白質 X 光晶體繞射結構顯示 N-DFRS/DFA 的有效活性來自於非穩定的多態鍵結模式，而非 N-DFRS/LFA 的固定鍵結模式。

DFA 於蛋白質功能及蛋白設計的初步測試也近一步瞭解單一位點 DFA 置換在蛋白螢光光學物理，溫度變化結構效應，以及殼蛋白組裝及解構的影響。sfGFP-Y66DmCIPhe 蛋白的螢光光譜顯示 DFA 嵌入增加紅位移(510 nm)的放射波長；此蛋白在圓二色光譜於變溫實驗中在 $\alpha$ 螺旋二級結構標示的波長(215 及 218 nm)顯示反轉趨勢。人類攜鐵蛋白 hFerritin-L56DmCF3Phe 在酸鹼解構實驗中，在 pH = 4 環境下，DLS 分析中如同 hFerritin-L56LmCF3Phe 殼蛋白趨向包裝 (assembling) 結構鬆散而造成再包裝更大粒徑殼蛋白(~100 nm 半徑)。

關鍵字：體內嵌入非典型胺基酸、D-非典型胺基酸、PyIRS•tRNA<sub>CUA</sub><sup>PyI</sup> 配對、蛋白質設計。

## Abstract



Expanding genetic codes research is an emerging field for incorporating noncanonical amino acids (ncAA) into proteins *in vivo* and studying mechanism of enzymatic reactions and designing novel enzymes with innovative chemistry. Yet native protein synthesis machinery precludes D-amino acid (D-aa) incorporation in polypeptide bond formation, protein chemists are assiduously developing the promiscuous amino acyl-tRNA synthetase with polyspecificity for the wide spectrum of ncAA. Here, we report an evolved pyrrolysyl-tRNA synthetase • pyrrolysyl-tRNA (PylRS•tRNA<sup>Pyl</sup><sub>CUA</sub>) pair with collaborated mutations in N-terminus and active site, naming N-DFRS, that can charge D- or L-phenylalanine analogues (DFA or LFA) into superfolder green fluorescent proteins (sfGFP) in *E. coli*. The molecular insight of N-DFRS/DFA and N-DFRS/LFA co-crystal structures reveal multiple and dynamic binding modes for relaxing chiral selective of ncAA sidechain protruding in the enlarged binding pocket. Moreover, sfGFP incorporated with DFA at Y66 position exhibited additional fluorophore in 510 nm emission wavelength. It suggests D-aa may display crucial role in studying protein photochemical property and protein design.

To investigate mirror property and bio-orthogonal chemistry of encoded ncAA, we are exploring DFA incorporated ferritin for protein/drug encapsulation and delivering for therapeutic application. Our research indicated D/LFA incorporated into ferritin at L56 position slightly distort structure at pH 4.0. The establishment of encoding D/LFA into ferritin *C*<sub>2</sub> interface emphasizes its potential to explore as engineered protein cage in therapeutic and diagnostic tools.

**Keywords:** expanding genetic code 、 D-noncanonical amino acids 、 pyrrolysyl-tRNA synthetase•tRNA<sup>Pyl</sup><sub>CUA</sub> pair engineering 、 protein design

## Table of contents

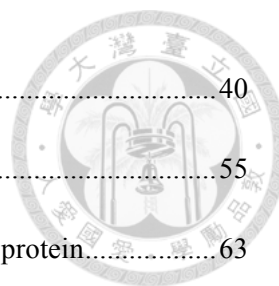


摘要.....	I
Abstract.....	II
Table of contents .....	III
List of figures.....	VI
List of tables.....	X
Abbreviations .....	XI
<b>Chapter 1 Introduction.....</b>	<b>1</b>
<b>1.1 Protein translation mechanism .....</b>	<b>1</b>
<b>1.2 Expanding genetic code research .....</b>	<b>3</b>
1.2.1 Residue-specific incorporation of ncAA.....	3
1.2.2 Site-specific incorporation of ncAA .....	4
<b>1.3 D-amino acid incorporation progress.....</b>	<b>7</b>
1.3.1 Mirror and mixed chiral protein design .....	8
<b>1.4 Properties of Ferritin cage.....</b>	<b>8</b>
1.4.1 Design and application of Ferritin cage .....	9
<b>1.5 Specific aim of thesis .....</b>	<b>12</b>
<b>Chapter 2 Materials and Methods.....</b>	<b>13</b>
<b>2.1 DNA and Protein sequences .....</b>	<b>13</b>

2.1.1	DNA sequences .....	13
2.1.2	Protein sequence .....	15
<b>2.2</b>	<b>Plasmid construction .....</b>	<b>16</b>
2.2.1	Primer list .....	16
2.2.2	Plasmids .....	18
<b>2.3</b>	<b>Screening incorporation efficiency of PylRS variants. ....</b>	<b>23</b>
<b>2.4</b>	<b>Protein productions and purifications.....</b>	<b>24</b>
<b>2.5</b>	<b>Gel analysis .....</b>	<b>27</b>
2.5.1	SDS-PAGE analysis.....	27
2.5.2	Western blot analysis .....	28
2.5.3	Native gel analysis .....	28
<b>2.6</b>	<b>Protein Biophysical characterizations .....</b>	<b>29</b>
2.6.1	X-crystal structural analysis.....	29
2.6.2	Protein ESI-MS and MALDI-TOF-MS/MS analysis .....	30
2.6.3	T <sub>m</sub> analyzed by CD .....	32
2.6.4	DLS analysis .....	32
2.6.5	TEM analysis .....	33
<b>2.7</b>	<b>Emission and Excitation analysis .....</b>	<b>33</b>
2.7.1	UV/Visible absorption spectrum.....	33
2.7.2	Fluoresces spectrum .....	33
<b>Chapter 3</b>	<b>Results .....</b>	<b>35</b>
<b>3.1</b>	<b>PylRS designing for high efficiency incorporating D-aa .....</b>	<b>35</b>
3.1.1	Engineering <i>Mm</i> PylRS and sfGFP.....	35



3.1.2	High throughput screening of pyIRS variants.....	40
3.1.3	Catalytic activity of PyIRS with L/DFA.....	55
3.1.4	Analysis of N-DFRS catalytic activity at different position in protein.....	63
3.1.5	ESI-MS and MALDI-TOF-MS/MS analysis of sfGFP variants.....	68
<b>3.2</b>	<b>Crystallization revealing L/DFA binding mode in N-DFRS .....</b>	<b>83</b>
3.2.1	Purification of N-DFRS binding pocket for crystallization.....	83
3.2.2	D/LFA soaking with N-DFRS binding pocking .....	86
<b>3.3</b>	<b>Studying biophysical property of sfGFP variant with D/LFA .....</b>	<b>93</b>
3.3.1	Incorporating D/LFA into chromophore of sfGFP .....	93
3.3.2	Photophysical characteristics of sfGFP-Y66 with D/LFA.....	101
3.3.3	Thermally induced conformational changes of sfGFP with Y66 D/LFA.....	108
<b>3.4</b>	<b>Application of D/LFA into protein cage design .....</b>	<b>114</b>
3.4.1	Shape and biophysical property of Ferritin with D/LFA .....	121
<b>Chapter 4</b>	<b>Discussion .....</b>	<b>126</b>
<b>Chapter 5</b>	<b>Conclusion .....</b>	<b>131</b>
<b>Reference</b>	<b>.....</b>	<b>132</b>
<b>Appendix</b>	<b>.....</b>	<b>138</b>



## List of figures

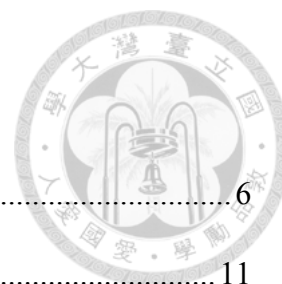


Figure 1. Illustration of protein translation with BTP.....	6
Figure 2. Nanocage structure of h-ferritin. ....	11
Figure 3. The x-ray crystal structure of wild-type sfGFP.....	38
Figure 4. Two plasmid system design for incorporating ncAA into protein in <i>E. coli</i> ...	39
Figure 5. Chemical structures of ncAAs in this study. ....	44
Figure 6. Range of substrate specificities of PylRS with sfGFP-S2TAG.....	45
Figure 7. Range of substrate specificities of N-PylRS with sfGFP-S2TAG. ....	46
Figure 8. Range of substrate specificities of LFRS with sfGFP-S2TAG. ....	47
Figure 9. Range of substrate specificities of DFRS with sfGFP-S2TAG.....	48
Figure 10. Range of substrate specificities of N-DFRS with sfGFP-S2TAG.....	49
Figure 11. Range of substrate specificities of PylRS with sfGFP-F27TAG.....	50
Figure 12. Range of substrate specificities of N-PylRS with sfGFP-F27TAG. ....	51
Figure 13. Range of substrate specificities of LFRS with sfGFP-F27TAG. ....	52
Figure 14. Range of substrate specificities of DFRS with sfGFP-F27TAG.....	53
Figure 15. Range of substrate specificities of N-DFRS with sfGFP-F27TAG.....	54
Figure 16. Incorporation efficiency of PylRS variants with D/LFA. ....	57
Figure 17. Incorporation efficiency of PylRS•tRNA <sub>CUA</sub> <sup>Pyl</sup> pair with twenty D/L-aa. ....	58
Figure 18. Incorporation efficiency of N-PylRS•tRNA <sub>CUA</sub> <sup>Pyl</sup> pair with twenty D/L-aa. ....	59
Figure 19. Incorporation efficiency of DFRS•tRNA <sub>CUA</sub> <sup>Pyl</sup> pair with twenty D/L-aa. ....	60
Figure 20. Incorporation efficiency of LFRS•tRNA <sub>CUA</sub> <sup>Pyl</sup> pair with twenty D/L-aa.....	61
Figure 21. Incorporation efficiency of N-DFRS•tRNA <sub>CUA</sub> <sup>Pyl</sup> pair with twenty D/L-aa...	62
Figure 22. Protein incorporated with ncAA at multiple sites. ....	65

Figure 23. Site-specific incorporation of D/LFA by N-DFRS•tRNA <sup>Pyl</sup> <sub>CUA</sub> pair.....	66
Figure 24. Analyzed amber suppression of D/LFA in sfGFP variants by WB.....	67
Figure 25. Molecular mass determination of the protein sfGFP-3xTAG-2.....	69
Figure 26. Molecular mass determination of the protein sfGFP-3xTAG-3.....	70
Figure 27. Molecular mass determination of the protein sfGFP-3xTAG-6.....	71
Figure 28. Molecular mass determination of the protein sfGFP-3xTAG-7.....	72
Figure 29. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-2 at position S2.....	73
Figure 30. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-2 at position F27.....	74
Figure 31. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-2 at position F130.....	75
Figure 32. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-3 at position S2.....	76
Figure 33. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-3 at position F27.....	77
Figure 34. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-3 at position F130.....	78
Figure 35. MALDI-TOF-MS/MS analysis on of sfGFP-3xTAG-6 at position F27.....	79
Figure 36. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-6 at position F130.....	80
Figure 37. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-7 at position F27.....	81
Figure 38. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-7 at position F130.....	82
Figure 39. Over-expression and purification of N-DFRSc270.....	84
Figure 40. Purification of N-DFRSc-270 with Mono Q column.....	85
Figure 41. Screening for crystallizing condition of N-DFRSc270.....	88
Figure 42. Crystals of N-DFRSc270.....	89
Figure 43. Crystal structure of N-DFRSc270 with ANP and LFA.....	90
Figure 44. Electron density map (2Fo-Fc) around D/LFA in binding pocket.....	91
Figure 45. Crystal structure of N-DFRSc270 with ANP and D/LCImPhe.....	92



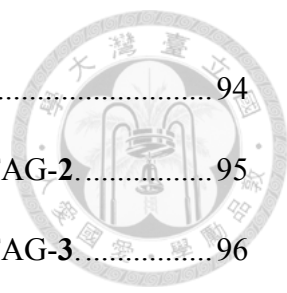


Figure 46. The molecular insight of sfGFP in chromophore.....	94
Figure 47. Molecular mass determination of the protein sfGFP-Y66TAG-2.....	95
Figure 48. Molecular mass determination of the protein sfGFP-Y66TAG-3.....	96
Figure 49. Molecular mass determination of the protein sfGFP-Y66TAG-4.....	97
Figure 50. Molecular mass determination of the protein sfGFP-Y66TAG-5.....	98
Figure 51. Molecular mass determination of the protein sfGFP-Y66TAG-6.....	99
Figure 52. Molecular mass determination of the protein sfGFP-Y66TAG-7.....	100
Figure 53. Photophysical characterization of sfGFP-Y66TAG-2 and 3.....	102
Figure 54. Photophysical characterization of sfGFP-Y66TAG-4 and 5.....	103
Figure 55. Photophysical characterization of sfGFP-Y66TAG-6 and 7.....	104
Figure 56. Intrinsic fluorescence spectrum of sfGFP-Y66TAG-2 and 3.....	105
Figure 57. Intrinsic fluorescence spectrum of sfGFP-Y66TAG-4 and 5.....	106
Figure 58. Intrinsic fluorescence spectrum of sfGFP-Y66TAG-6 and 7.....	107
Figure 59. DSC profile of sfGFP-Y66TAG incorporated with D/LFA.....	109
Figure 60. Thermally induced conformational changes of wt-sfGFP.....	110
Figure 61. Thermally induced conformational changes of sfGFP-Y66TAG-5.....	111
Figure 62. Thermally induced conformational changes of sfGFP-Y66TAG-4.....	112
Figure 63. Thermally induced conformational changes monitored by CD.....	113
Figure 64. The structure of $C_2$ interface of ferritin.....	116
Figure 65. Purification of Ferritin-L56TAG-2 with SEC.....	117
Figure 66. Purification of Ferritin-L56TAG-3 with SEC.....	118
Figure 67. Molecular mass determination of the protein Ferritin-L56TAG-2.....	119
Figure 68. Molecular mass determination of the protein Ferritin-L56TAG-3.....	120

Figure 69. TEM of Ferritin variants.....	123
Figure 70. DLS analysis of Ferritin variants at pH 4.0, 8.0, and 10.0.....	124
Figure 71. Analyzed size distribution of ferritin variants at pH 4.0 by volume.....	125



## List of tables

Table 1 Mutated residues of PylRS variants in this study .....	37
--	----



## Abbreviations



Ala, A	L-alanine
Arg, R	L-arginine
AARS	aminoacyl-tRNA synthetase
Asn, N	L-asparagine
$\beta$ -Me	$\beta$ -mercaptoethanol
CD	circular dichroism
CV	column volume
Cys, C	L-cysteine
D-aa	D-amino acid
Da	dalton
DFA	D-phenylalanine analogues
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DSC	differently scanning calorimetry
DTT	1,4-dithiothreitol
<i>E. coli.</i>	<i>Escherichia coli</i>
EF-G	elongation factor-G
EF-Ts	elongation factor-thermostable
EF-Tu	elongation factor-thermounstable
ESI-MS	electrospray ionization mass spectrometry
FRET	Förster resonance energy transfer



Gln, Q	L-glutamine
Glu, E	L-glutamate
Gly, G	L-glycine
His, H	L-histidine
IPTG	isopropyl $\beta$ -D-1-thiogalactopyranoside
L-aa	L-amino acid
LB	Luria-Bertani broth
LC-ESI-MS/MS	liquid chromatography - electrospray ionization-tandem mass spectrometry
Leu, L	L-leucine
LFA	L-phenylalanine analogues
Lys, K	L-lysine
MALDI-TOF-MS/MS	matrix assisted laser desorption ionization - time of flight - tandem mass spectrometry
Met, M	L-methionine
<i>MmPylRS</i>	<i>Methanosarcina mazei</i> pyrrolysyl-tRNA synthetase
mRNA	messenger ribonucleic acid
ncAA	noncanonical amino acid
NTD	N-terminal domain
PBST	phosphate buffered saline with Tween-20
PBS	phosphate buffered saline
PCR	polymerase chain reaction
Phe, F	L-phenylalanine

PVDF	polyvinylidene ifluoride
Pyl, O	L-pyrrolysine
RT	room temperature
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel
Sec, U	L-seleocysteine
SEC	size exclusion chromatography
Ser, S	L-serine
sfGFP	superfolder green fluorescent protein
Sp	spectinomycin
TEM	transmission electron microscope
TEMED	<i>N,N,N',N'</i> -tetramethylethylenediamine
T <sub>m</sub>	denaturation midpoint
tRNA	transfer ribonucleic acid
Tyr, Y	L-tyrosine
UV	ultraviolet
Val, V	L-valine
WB	western blot
wt	wild-type



## Chapter 1 Introduction

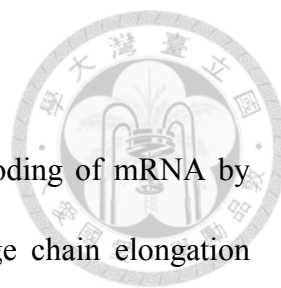


### 1.1 Protein translation mechanism

Protein translation system is a highly and accurate machinery in polyamide bond, protein synthesis from mRNA templates. Protein synthesis needs Three crucial elements: amino acid, tRNA and mRNA to be operated in ribosome chamber. 64 genetic codes based on three letter codes and four building blocks, A, T, G, and C, have been found in the coding and decoding system in genomic DNA and mRNA.<sup>1</sup> With codon degeneracy phenomena in codon decoding, twenty canonical amino acids were found in protein translation process in organisms living in the oxidative environment.

#### Initiation

In protein translation, protein translation starts from ribosome complex 70S assembling that initiated from small subunit (30S) binds with large subunit (50S). Initiation step of protein synthesis requires three initiation factors (IF-1, IF-2, and IF-3).<sup>2</sup> GTP bound IF-3 strongly interacts with 30S subunit and prevents from 50S subunit combining prematurely. Shine-Dalgarno sequence of mRNA was recognized by IF3/30S complex and guide GTP bounded IF-2 brings fMet-tRNA<sup>fMet</sup> to 30S complex and located UAG start site. IF-1 modulates this initiation process and block aminoacyl (A) site before 50S assembling. After fMet-tRNA<sup>fMet</sup> finishing decoding at UAG codon, three initiation factors release from 30S and induce 70S ribosome, 50S/30S complex in prokaryotes. A, peptidyl (P) site and exiting (E) sites are formed in 70S ribosome for further polypeptide bonds synthesis.



## **Elongation**

Peptide chain elongation follows the principle of codon decoding of mRNA by accommodating aminoacyl-tRNA. Three elongation factors manage chain elongation stage, EF-Tu, EF-Ts, and EF-G. EF-Tu • GTP carry aminoacyl-tRNA to A site for matching correct decoding with GTP hydrolysis to EF-Tu • GDP, and EF-Tu leaves from ribosome. EF-Tu • GDP then is reactivated by exchange GDP with EF-Ts and turned into EF-Tu • GTP. While first peptide bond is formed at P site, peptide translocation in A/P site movement was assisted by EF-G in GTP hydrolysis.<sup>3</sup> The ribosome moves one codon toward the 3'end of mRNA and the uncharged tRNA shift from P site to E site. Continually, uncharged tRNA release from E site and then the next codon is recognized by new aminoacyl-tRNA. Therefore, producing nascent peptide chain, the elongation cycles ongoing until meeting stops codon. After the cycles of elongation finish, release factor binds to the stop codon and starts the last stage of translation, termination.

## **Termination**

Termination, the last stage of the polypeptide synthesis, while one of three stop codon (UAA, UAG, and UGA) present in mRNA. When a termination codon occupies the ribosome A site, release factors (RF1, RF2, and RF3) contributing to hydrolysis the peptidyl-tRNA bond, release peptide and the last tRNA on P site, finally, separate the 70S ribosome to 30S and 50S subunit.<sup>4</sup> Definitely, release factors possess a little different with each other, RF1 recognizes the stop codon UAA and UAG, and RF2 recognizes the stop codon UAA and UGA. RF3 is a GTPase giving energy to termination and acting in disassembly of the ribosome termination complex.<sup>5</sup>

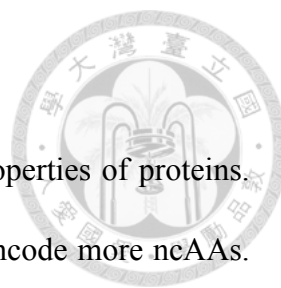




## 1.2 Expanding genetic code research

### 1.2.1 Residue-specific incorporation of ncAA

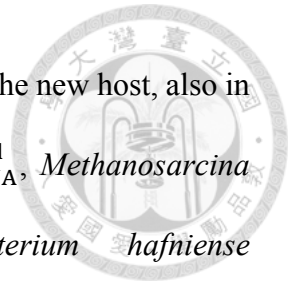
The genetic codes are corresponded to the same specific amino acids in all organism. 20 amino acids are in common with the rare exception of Sec (**9**) and Pyl (**1**). These canonical amino acids build the delicate balance on stability and function of protein under biological condition. But still many things that canonical amino acids cannot do thus limit the development of organisms with new or enhance properties. So, finding the strategy to incorporate new amino acids into protein is important. In 1950s, Cohen and coworkers revealed that the ncAA, selenomethionine **8**, could be recognized by methionyl-tRNA synthetase and incorporated into protein in place of methionine<sup>6</sup>, naming residue-specific incorporation. Residue-specific methodology allows to globally replace canonical amino acids with their non-canonical analogues. The resulting protein illustrate new chemical and physical properties. For example, some chemical biologists focused on residue-specific incorporation which use ncAA with function group such as azide or alkyne. Both of which can be selective modification via click chemistry and applicate in protein labeling *in vivo* or *in vitro*. On the other hand, expanding the scale of residue-specific incorporation was also important, AARS were evolved in the edit domain or amino acid binding pocket to recognize new ncAA that can not be incorporated by wild-type AARS. Residue-specific incorporation of ncAAs into protein with useful properties undoubtedly has reached new milestone in chemistry and biology. However, substitution of a specific site remains unreachable in this method.<sup>7,8</sup>

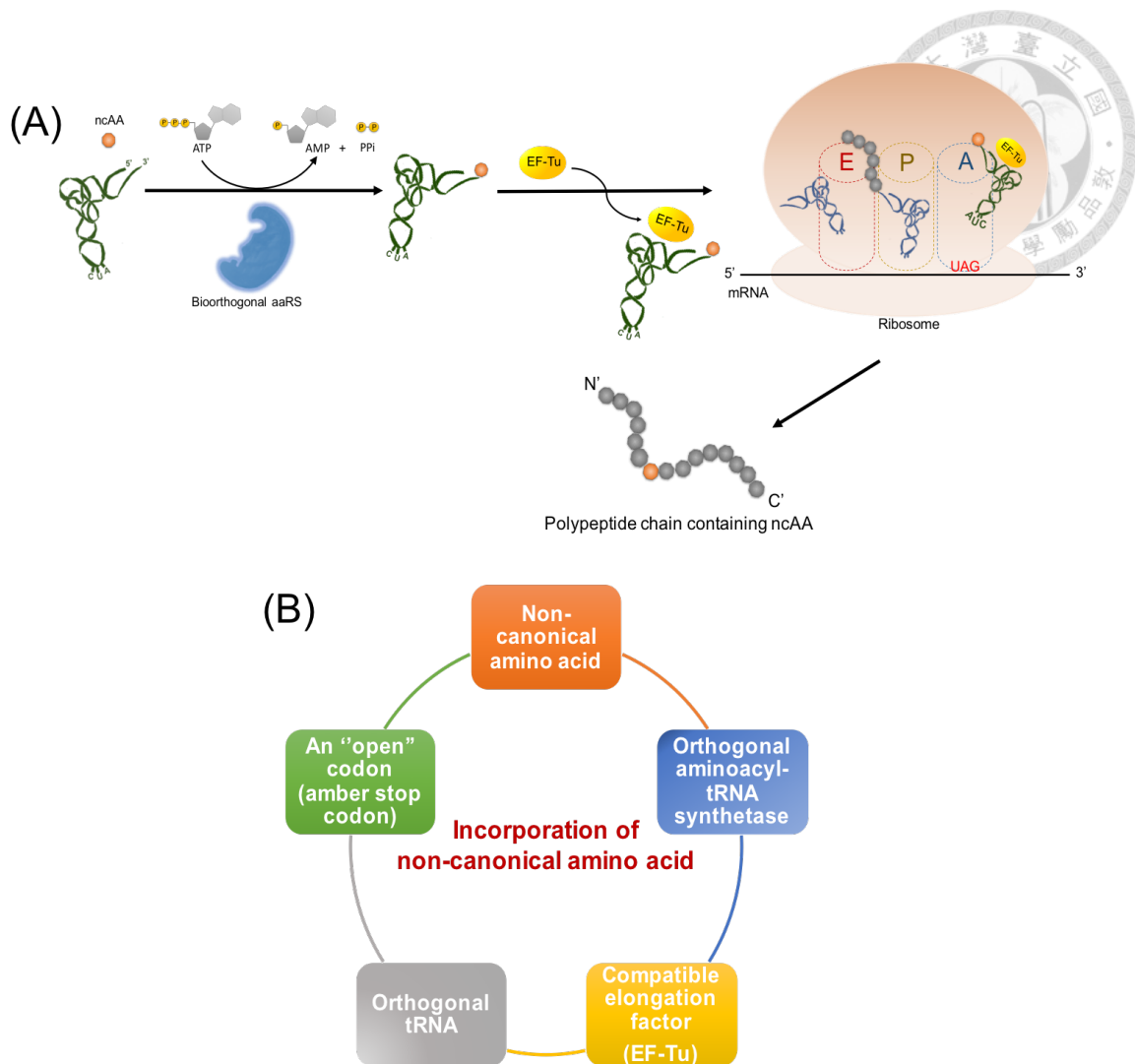


### 1.2.2 Site-specific incorporation of ncAA

To further enhance the ability to control the structure and properties of proteins. The scientists sought to directly create organisms that genetically encode more ncAAs. In 1980s, Peter G. Schultz and coworkers developed new method to make the possibility to site-specifically incorporate ncAAs into proteins. They used a chemically aminoacylated suppressor tRNA that in response to an amber codon, which is not recognized by endogenous tRNA, to insert the ncAA into interested site of protein.<sup>9</sup> However, the application of these techniques was limited by low production, because of the restriction on sustainable aminoacyl-tRNA. Biosynthetic approaches have also been developed for overcoming the barrier. In 2001, Lei Wang who was in Peter G. Schultz's group found out an orthogonal AARS•tRNA pair in *E. coli* which means the pair not interrupt the existing translation system in *E. coli*, for example, orthogonal AARS does not recognize the endogenous *E. coli* tRNAs and 20 amino acids (**Figure 1**). The candidate is *Methanocaldococcus jannaschii* TyrRS•tRNA<sup>Tyr</sup> (MjTyrRS•MjtRNA<sup>Tyr</sup>) pair from archaea.<sup>10</sup> And the author have verified this pair could efficiently incorporate *O*-methyl-L-tyrosine into protein in response to an amber codon (TAG). However, the ability of evolving AARSs to specific incorporating amino acid depends on the structural limitations in the catalytic pocket. Thus, it was necessary to find other orthogonal AARS•tRNA pairs for diverse ncAA. In 2002, the scientists discovered the 22<sup>nd</sup> natural amino acid, Pyl 1, and genetically encoded by UAG in *Methanosarcina barkeri*.<sup>11</sup> And the PylRS•tRNA<sup>Pyl</sup><sub>CUA</sub> pair was utilized to specifically incorporate 1 into protein.<sup>12,13</sup> Soon after, Jason W Chin and coworkers introduced PylRS•tRNA<sup>Pyl</sup><sub>CUA</sub> into expanding genetic codon system and genetically encoded *N*<sup>ε</sup>-acetyllysine in recombinant proteins.<sup>14</sup>

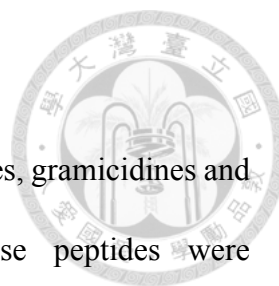
Because the PylRS•tRNA<sup>Pyl</sup><sub>CUA</sub> pair retains excellent orthogonality in the new host, also in mammalian cells, the *Methanosarcina barkeri* (MbPylRS)•tRNA<sup>Pyl</sup><sub>CUA</sub>, *Methanosarcina mazei* PylRS (MmPylRS)• tRNA<sup>Pyl</sup><sub>CUA</sub> and *Desulfitobacterium hafniense* PylRS(DhPylRS)•tRNA<sup>Pyl</sup><sub>CUA</sub> pairs are commonly utilized in site-specific incorporating ncAA into protein.<sup>15</sup> In our study, *Methanosarcina mazei* PylRS (MmPylRS)• tRNA<sup>Pyl</sup><sub>CUA</sub> pair was engineered to incorporate ncAA.<sup>16</sup>





**Figure 1. Illustration of protein translation with BTP.**

(A) Bioorthogonal translation pairs (BTP). The appropriate ncAA was ligated to suppressor tRNA by bio-orthogonal AARS, e.g., *MmPylRS*, an archaeal enzyme. With the assistance of EF-TU, the aminoacyl-tRNA can recognize the corresponding amber codon on mRNA in ribosome A site thus produce protein containing ncAA. (B) Five essential elements for incorporating ncAA. (i) An "open" codon (e.g., amber codon) is used to encode (ii) an ncAA. After (iii) an orthogonal AARS efficiently catalyzes, the ncAA is attached with (iv) Orthogonal tRNA (iv) that can decode the "open" codon and the product can be recognize by (v) compatible elongation factor.<sup>16</sup>



### 1.3 D-amino acid incorporation progress

In 1941, Lipmann firstly discovered that some bacterial peptides, gramicidines and tyrocidines, contain D-aa.<sup>17</sup> These studies showed both these peptides were biosynthesized by large, modular, multifunctional enzyme, known as nonribosomal peptide. 40 years later, a milestone in the history of peptide containing D-aa was the discovery of dermorphin which was isolated from amphibian skin secretions and had high affinity to  $\mu$ -type opiate receptors.<sup>18</sup> The second amino acid is D-alanine in the protein sequence of heptapeptide. Different from above, dermorphin was synthesized from mRNA and the codon at this position was same as L-alanine, however, while the peptide incorporating with L-alanine at position 2, it completely loss biological activity. After more research, the scientists demonstrated the D-aa in the peptide was catalyzed from L-configuration by isomerase acting on an intact peptide rather than on free amino acid.<sup>19,20</sup> However, the D-aa-containing polypeptides were synthesized either by nonribosomal peptide synthetases or post-translational isomerase, it was difficult to incorporating D-aa in specific site or in ideal protein. In 2010s, Sven Klussmann and coworkers introduced an *in vitro* translation system that enables incorporation of D-aa into polypeptide with engineering translational apparatus, EF-Tu and ribosome.<sup>21</sup> In addition, Hiroaki Suga's group logically engineered D-arm and T-stem of tRNA which can be recognized by EF-P and EF-Tu, respective. And the optimized tRNA pre-charged with D-aa drastically enhanced expression level of polypeptide with D-aa.<sup>22</sup> But, these methodologies are still indirectly incorporating D-aa into protein with ribosomal translation system *in vitro*.

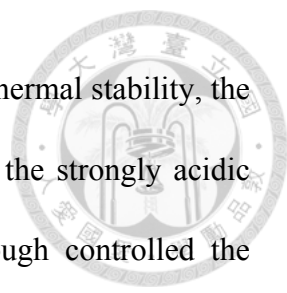


### 1.3.1 Mirror and mixed chiral protein design

D-aa had been detected in variety designing peptides and play an important role in activity, binding affinity,<sup>23</sup> and stability. For example, the chiral mixed peptide with epitope function showed high resistance against proteolytic degradation in diluted human serum and in lysosomal preparation.<sup>24</sup> In addition, the published paper revealed that D-aa antibiotic was hydrolytically cleaved by D-stereospecific peptidases which means the some peptidase possessed stereo-selectivity.<sup>25</sup> The structural effect of protein with D-aa was revealed in 2003. The ubiquitin variants containing an L-aa or the corresponding D-aa at the C-cap G35 position demonstrated that D-aa can accept a left-handed structure without energetic penalty.<sup>26</sup> However, in nature, the protein translation system only used L-aa, the knowledge of D-aa containing protein is rarely understood. Thus, we devoted to realize properties of the chiral mixture protein.

### 1.4 Properties of Ferritin cage

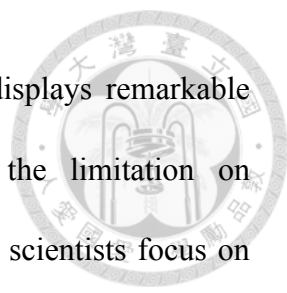
Ferritin was first isolated and crystallized by Laufberger in 1937. The primary role of ferritin is to protect cells from the damage caused by the Fenton reaction, store iron molecules within a hollow core and transport.<sup>27</sup> The process of iron storage is Fe<sup>II</sup> binding, oxidizing to Fe<sup>III</sup> and biomineralization towards ferric oxide clusters in ferritin.<sup>28</sup> The monomer of ferritin is a four-helix bundle protein, totally around 200 amino acids, and composes the 24 subunits in an octahedral symmetry. The size of self-assembled cage is 12 nm diameter with an interior cavity of 8 nm diameter.<sup>29</sup> Ferritin possess three kinds of symmetry axes,  $C_2$ ,  $C_3$ , and  $C_4$  (**Figure 2**). And the center point of  $C_3$  and  $C_4$  symmetry axes on the protein shell compose eight and six channels, respectively, connecting



exterior to the inner cavity.<sup>30</sup> The ferritin cage displays remarkable thermal stability, the  $T_m$  is higher than 85°C, and chemical stability, dissociation under the strongly acidic conditions (pH 2.0).<sup>31</sup> It is particularly able to reconstitution through controlled the (dis)assemble condition. At low pH (pH 2.0), ferritin cages disassemble into subunits allowing the dispersed loading cargoes, daunomycin which has been used for cancer treatment, to encapsulate into the ferritin cage when the pH is changed back to basic (pH 8.5) and protein cage reassembly.<sup>32</sup> The structural and biochemical properties of ferritin protein are used for a wide range of applications, from the synthesis of nanoparticles<sup>33,34</sup> to the design of vaccines in biomedicine.<sup>35</sup>

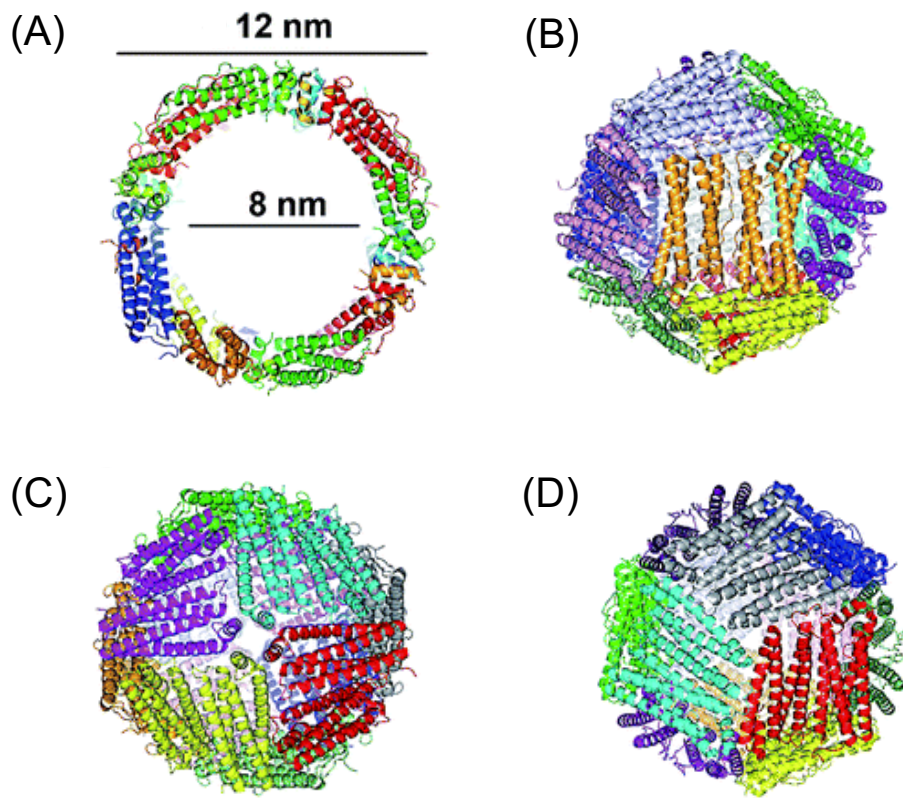
#### **1.4.1 Design and application of Ferritin cage**

Similar to other nanocages, the application of ferritin in biomedicine involve several challenge, for example, lack of suitable release properties. To overcome this challenge, the engineering of protein cage is essential. The decoration can impart to three different cage surfaces, the interior, exterior, and interface.<sup>36</sup> The interior modification of protein cage provides an ideal containment for molecular cargo and increase the encapsulation efficiency, binding affinity, and modulated release profile.<sup>37</sup> And the exterior surface of protein cage usually engineer for decreasing the immune response and targeting the specific cell. However, rare studies focus on engineering exterior surface of ferritin. This is because ferritin exists in almost all organisms, it possesses an outstanding biocompatibility and safety profile in individual species. In addition, the published paper revealed ferritin can recognize tumor cell and be endocytosis via interaction with overexpressed transferrin receptor I on tumor cell.<sup>38</sup> The intersubunit modification of



ferritin have the most research. Because self-assembling ferritin displays remarkable thermal and chemical stability, these properties bring about the limitation on encapsulation and releasing cargo. In order to solve this issue. The scientists focus on modifying at interface of  $C_2$ ,  $C_3$  or  $C_4$  symmetry axes. For example, six pores were composed of the adjacent four E-helix around  $C_4$  axis. And the authors substituted the E-helix of ferritin with GALA (Glu-Ala-Leu-Ala) peptide which transited coli to helix in weak acid condition. After modification, the ferritin variants can completely dissociation at pH6.0 and re-assembled simply by neutralization. Thus, this engineered cage protein can potentially encapsulate weak-acidly tolerant protein or drugs.<sup>39</sup> The  $C_2$  interface is important for ferritin self-assemble. In 2013, Akif Tezcan's group focus on engineer  $C_2$  interface, the residues L56, H60, R63, and E67 that lie across the interface was mutated as histidine. To allow that copper acts as a structural template for ferritin assembly, and disassemble while the environment without metal.<sup>40</sup> In our study,  $C_2$  interface of ferritin was the candidate for engineering protein cage.





**Figure 2. Nanocage structure of h-ferritin.**

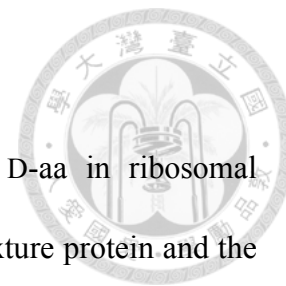
The human-ferritin (h-ferritin) was assembled by 24 h-ferritin monomers. (A) The size of external cage (12 nm) and inner cavity (8 nm). External view of h-ferritin can be grouped in  $C_2$  symmetry axes (B),  $C_4$  symmetry channels (C), and  $C_3$  ion channels (D). The cartoon figures were made by PyMOL. PDB code: 4DAS.<sup>30</sup>

## 1.5 Specific aim of thesis

Until now, many obstacles on synthesizing protein with D-aa in ribosomal translation system still endure. Thus, it's hard to generate chiral mixture protein and the knowledge of protein with D-aa is limited. However, polypeptide/protein incorporating with D-aa play an important role in nature, for example, antibiotic biosynthesis and cell wall in bacteria. In addition, D-aa have proved that is a fundamental character of biological activity and enzymatic stability.

In this thesis, we attempted to investigate the methodology of synthesizing chiral mixture protein in ribosomal translation system. Expand genetic code system which is a powerful tool to incorporate non-proteinogenic amino acids into protein in *E. coli* or mammalian cell was imported. And the enzyme, *MmPylRS*, was engineered to release the chiral selectivity for encoding D-aa into protein and explore the binding mode of D-aa in catalytic pocket. Subsequently, testing to site-specifically incorporate D-aa into different site of protein and studying the biophysical property of protein with D-aa.

On the other hand, utilizing the mirror properties of D-aa to explore D-aa incorporated protein cage, ferritin, for encapsulating and releasing protein drug. Ferritin is a thermally and chemically stable protein, thus, it's have the limitation of disassembly and assembly for encapsulating cargo. According to the structural characteristic of ferritin, designing the protein with D-aa for easier dissociation.



## Chapter 2 Materials and Methods



### 2.1 DNA and Protein sequences

#### 2.1.1 DNA sequences

tRNA<sup>Pyl</sup><sub>CUA</sub> :

*ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggtagattcccggggttccgcca*

*MmPylRS:*

*atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataa  
aacaccacgaagtctctcgaagcaaaatctatattgaaatggcatgctggagaccaccttgttgaacaactccaggagc  
agcaggactgcaagagcgtcagccaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatct  
caataagttctcacaaggcaaacgaagaccagacaagcgtaaaagtcaaggctgttctgcccctaccagaacgaa  
aaaggcaatgccaaaatccgttgcgagagccccgaaacctctgagaatacagaagcggcacaggctcaaccttctgg  
atctaaatfttcacctgcgataccggttccaccaagagtcagtttctgcccggcatctgtttcaacataaatcaagcatt  
tctacaggagcaactgcacccactggtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagt  
gccccgcacttacgaagagccagactgacaggcttgaagtcctgttaaaccctaaagatgagattccctgaattccgg  
caagccttcaggagcttgagtcgaattgctctctcgcagaaaaaaagacctgcagcagatctacgcggaagaaagg  
gagaattatctgggaaaactcgagcgtgaaattaccaggttcttctgtggacaggggtttctggaataaaatccccgatcc  
tgatccctcttgagtatacgaagagatgggcattgataatgataccgaacttcaaacagatcttcaggggttgacaagaa  
cttctgcctgagacctatgcttctccaaacctttacaactacctgcgcaagcttgacagggcctgcctgatccaataaaa  
atfttgaaataggccatgctacagaaaaagatccgacggcaaaagaacacctcgaagagttaccatgctgaactctg  
ccagatgggatcgggatgcacacgggaaaatctgaaagcataattacggacttctgaaccacctgggaattgattca  
agatcgtaggcgattcctgcatggctctatggggatacccttgatgtaatgcacggagacctggaacttctctgcagtagtc*



*ggaccataccgcttgaccgggaatggggattgataaacctggataggggcaggttcgggctcgaacgccttctaaa  
ggttaaacacgacttataaaatatacaagagagctgcaaggctccgagtcttactataacgggatttctaccaacctgtaa*

sfGFP:

*atgagcaagggcgaagaactgtttacgggcgtggcgccgattctggggaactggatggatgtcaatggtcacaaatc  
agcgtgcgcggcgaagggaaggcgatgcaaccaatggtaaacctgacgctgaagttatttgcaccacgggtaaacctgc  
cggttccgtggccgacctggcaccacgctgacgtatgggttcagtgttcagtcgttaccggatcacatgaaacgcca  
cgacttttcaagtccgcgatgccggaagggtatgtccaagaacgtaccatctcatttaaagatgacggcacctacaaaac  
gcgcgccgaagtgaattcgaaggatgatacgtggtaaacctattgaaactgaaagcatcgatttaaggaagacggta  
atattctgggccataaacctggaatataactcaattcgcaaacgtgtacatcaccgcagataagcagaagaacggatc  
aaggctaactcaagatccgccataatgtggaagatggcagcgttcaactggccgaccactateagcaaaaacaccccga  
ttggatggcccggctctgctgccgacaatcattacctgagcacgcagctctgtgctgagtaaagatccgaacgaaaag  
cgtgaccacatggctctgctggaattcgtgaccgcggccggcatcacgcacggatggacgaactgtataaaggctcag  
agctccateaccatcaccatcactaa*

Ferritin:

*atgaccaccgctctacctcacaggtgcgtcagaattatcatcaggatagtgaaagcagcaattaategccagattaatctg  
gaactgtatgcaagctatgtgtatctgtctatgagctatttttgatcgcgatgatgttgccctgaaaaatttgcctaaatattt  
ctgcatcagctctcatgaagaacgcgaacatgccgaaaaactgatgaaattacagaatcagcgtgggtgctgattttcttc  
aagatattaaaaaacggattgtgatgattgggaaagcggcctgaatgcgatggaatgtgccttacatcttgagaaaaat  
gttaatcagtcactgctggaactgcataaacctggcaaccgataaaaatgatccgcacatctgtgtgattttattgaaaccatta  
tctgaatgagcaggttaaagccattaaagaactgggcgatcatgttaccatctccgaaaaatgggcgccccggaaaagt*

ggcttagccgaatatctgttgataaacataccttaggcgatagcgataacgaaagtgagctccatcaccatcaccatcact

aa



### 2.1.2 Protein sequence

*MmPyIRS*:

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSS  
RTARALRHHKYRKTCKRCRVSEDLNKFLTKANEDQTSVKVKVVSAPTRTKK  
AMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSPASVSTSISISTG  
ATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPF  
ELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIE  
RMGIDNDELTKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCY  
RKESDGKEHLEEFMNLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMV  
YGDTLDMHGDLELSSAVVGGPIPLDREWIDKPKWIGAGFGLERLLKVKHDFKN  
IKRAARSESYNGISTNL\*

sfGFP:

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTCLKFICTTGKLP  
VPWPTLVTTLYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYK  
TRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNHNVYITADKQKNGI  
KANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEK  
RDHMLLEFVTAAGITHGMDELYKGSELHHHHHH\*



Ferritin:

MTTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKNFA  
KYFLHQSHREHAEKLMKLNQRGGRIFLQDIKKPDCDDWESGLNAMECAL  
HLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVKAIKELGDHVTNLRK  
MGAPESGLAEYLFDKHTLGSDNESELHHHHHH\*

## 2.2 Plasmid construction

### 2.2.1 Primer list

Name	Sequences(5' to 3')
pET-h-Ferritin-NdeI-F	gatatacatatgaccaccgctctacctcacaggtgc
pET-h-Ferritin-SacI-R	ggtgatggagctcactttcgttatcgctatcgctaaag
pET-PylT-SphI-R	catctccttgcgatgctcgaacttttgcgagttgaag
pET-PylT-SphI-F	gaatggtgatgcccggcgtatgatcttataagcggg
PylT-AGG-anticodon-F	gaacggattaggagtccattcgatctacatgatc
PylT-AGG-anticodon-R	ggactcctaataccggtcagccgggtagattcccgg
PylRS-c270-F2	cacttacgaagagccagactgacaggcttg
pET-PylRS-R2	ggtgatggagctttacaggttgtagaaatcccg
pET-His6X-F1	gagatatacatatgcatcaccatcaccatcac
His6x-PylRS185-R1	ctcttcgtaagtgcgggggcacttgcgtgatggtgatgg
pET-sfGFP-NdeI-F1	gagatatacatatgagcaagggcgaag
pET-sfGFP-sacI-R2	gatggtgatggagctctgagcctttatac
sfGFP-Y66TAG-F2	gtcaccacgctgacgtagggtgttcag
sfGFP-Y66TAG-R1	ctgaacaccctacgtcagcgtggtgac

pET-sfGFP-S2TAG_NdeI-F	gatatacatatgtagaagggcgaagaactgtttacgggcg
pET-sfGFP-F130TAG-R	gtttatggcccagaatattaccgttcttcttaatcgatgcctttcagttc
pET-sfGFP-F130TAG-F	gtaaacctgattgaactgaaaggcatcgattagaaggaagacggtaatttc
pET-sfGFP-F27TAG-R	caccttcgccgcgcacgctctattttgtgacc
pET-sfGFP-F27TAG-F	ggtcacaaatagagcgtgcgcggcgaagggtg
h-Ferritin-L56TAG-F	caaatatttttagcatcagctctcatgaagaacg
h-Ferritin-L56TAG-R	gttcttcatgagactgatgctaaaaatatttggc
h-Ferritin-R63TAG-F	cagctctcatgaagaataggaacatgccgaaaaac
h-Ferritin-R63TAG-R	gttttcggcatgttcttattcttcatgagactgatg
h-Ferritin-H60TAG-F	ctgcatcagctcttaggaagaacgcgaacatg
h-Ferritin-H60TAG-R	catgttcgcgttcttctaagactgatgcag
h-Ferritin-E67TAG-F	cgcgaaatgcctagaaactgatgaaattac
h-Ferritin-E67TAG-R	gtaatttcatcagtttctagggcatgttcgcgttc
pCDF-PylRS-NcoI-F	gtattaacatggataaaaaaccactaac
pCDF-PylRS-EcoRI-F	gattccctgaattccggcaagcctttcaggagc
pCDF-PylRS-EcoRI-R	ggcttgccggaattcagggaatctcatcttttg
PylRS-R61K/H63Y-F	gcaggactgcaagagcgtcaaactataaatacaggaag
PylRS-R61K/H63Y-R	tttgcaggcttctctgtattatagtgttgagcgtcttg
PylRS-S193R-F	caggcaagtgccttcgacttacgaagcgtcagactgacag
PylRS-S193R-R	gactcaagcctgtcagctgacgcttcgtaagtgcg
PylRS-N346V-F	ccatgctggtgttctgccagatggg
PylRS-N346V-R	cccatctggcagaacaccagcatgg

PyIRS-N346G/C348Q-F	ccatgctgggcttccagcagatgggatc
PyIRS-N346G/C348Q-R	gatcccatctgctggaagcccagcatgg
pCDF-PyIRS-V401G-F	cctggaactttcctctgcagggcgtcggaccataaccgcttgacc
pCDF-PyIRS-V401G-R	ggtaagcggatgggtccgacgcctgcagaggaaagtccagg
pCDF-PyIRS-BamHI-R	ggtcgacggatccttacaggttggtagaatcccgttatag

## 2.2.2 Plasmids

### *Construction of pET-pylT*

The plasmid was derived from the pET-22b (+) plasmid having ampicillin (Amp) selection marker, purchasing from Novagen. The gene of *pylT* with anticodon CUA flanked by *lpp* promoter at the 5' end and *rrnC* terminator at 3' end. It was generated by overlapping PCR. The same restriction site, *SphI* at 5' end and 3' end, was introduced in the PCR product which was subsequently digested and used to construct pET-*pylT*.

### *Construction of pCDF-MmPylRS*

pCDF-*MmPylRS* plasmid that encodes wild-type *MmPylRS* was derived from pCDF-1b plasmid containing *Sp* selection marker, purchasing from Novagen. The wild type *MmPylRS* gene was synthesized by MD bio, Inc company. and is under the control of *lpp* promoter at the 5' end and *rrnC* terminator at 3' end. The restriction enzyme *BamHI* and *NcoI* were used to cut off the *MmPylRS* gene. The digested gene was ligated to form pCDF-*MmPylRS*.





### *Construction of pCDF-N-PylRS*

To construct the pCDF-N-PylRS plasmid with R61K, H63Y and S193R mutations, which was based on published paper<sup>41</sup> were introduced by overlap extension PCR from pCDF-*MmPylRS* plasmid. The follow pair of primers were used to generate a N-PylRS gene: (1) pCDF-PylRS-NcoI-F and PylRS-R61K/H63Y-R; (2) PylRS-R61K/H63Y-F and pCDF-PylRS-BamHI-R; (3) pCDF-PylRS-NcoI-F and PylRS-S193R-R (4) PylRS-S193R-F and pCDF-PylRS-BamHI-R. The first gene which was produced with (1) and (2) primers were became the template for primer (3) and (4) to form N-PylRS finally. The gene N-PylRS was digested with restriction enzyme BamHI and NcoI, gel-purified and ligation back into pCDF vector, also digested by BamHI and NcoI, to afford plasmid pCDF-N-PylRS.

### *Construction of pCDF-LFRS*

To construct the pCDF-LFRS plasmid with N346V mutation is similar to above. The follow pairs of primer were used to generate LFRS gene: (1) pCDF-PylRS-EcoRI-F and PylRS-N346V-R (2) PylRS-N346V-F and pCDF-PylRS-BamHI-R. Subsequently, LFRS was digested and gel-purified and ligation back into pCDF vector to form pCDF-LFRS.

### *Construction of pCDF-DFRS*

To construct the pCDF-DFRS plasmid with N346G, C348Q, and V401G mutations, which was based on published paper<sup>42</sup> was similar to above. The follow pairs of primer were used to generate DFRS gene: (1) pCDF-PylRS-EcoRI-F and PylRS-N346G/C348Q-

R; (2) PylRS-N346G/C348Q-F and pCDF-PylRS-BamHI-R (3) pCDF-PylRS-EcoRI-F and pCDF-PylRS-V401G-R; (4) pCDF-PylRS-V401G-F and pCDF-PylRS-BamHI-R.

The first gene which was produced with (1) and (2) primers were became the template for primer (3) and (4) to form DFRS finally. Subsequently, DFRS was digested and gel-purified and ligation back into pCDF vector to form pCDF-DFRS.

#### *Construction of pCDF-N-DFRS*

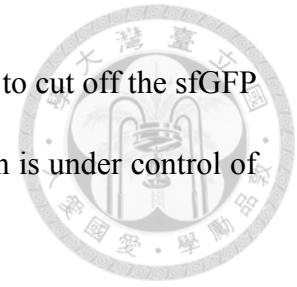
To construct the pCDF-N-DFRS plasmid, which contain the mutation site on both N-PylRS and DFRS, with R61K, H63Y, S193R N346G, C348Q, and V401G mutations were introduced by overlap extension PCR. The follow pairs of primer were used to generate N-DFRS gene: (1) pCDF-PylRS-NcoI-F and pCDF-PylRS-EcoRI-R; (2). pCDF-PylRS-EcoRI-F and pCDF-PylRS-BamHI-R. Importantly, (1) was using pCDF-N-PylRS as template and (2) was using pCDF-DFRS as template. The N-DFRS gene was subsequently digested and gel-purified and ligation back into pCDF vector to form pCDF-N-DFRS.

All ligation products in pCDF-1b plasmid were then transformed into *E. coli* DH5 $\alpha$  cell by chemical procedures. Transformed cells were recovered in LB medium for 60 min at 37°C, then plated on LB agar plates with Sp (100  $\mu$ g/mL), and grown overnight in a 37°C incubator. After incubation, choosing two colonies for DNA sequence.

#### *Construction of pET-pylT-sfGFP variants*

Plasmid pET-pylT-sfGFP was derived from the plasmid pET-pylT<sub>CUA</sub>. This gene was amplified from the optimize superfolder GFP plasmid which was synthesise by MD

bio, Inc. company. The restriction enzyme NdeI and SacI were used to cut off the sfGFP gene. The digested gene was ligated to form pET-pylT-sfGFP which is under control of a T7 promoter and with the hexahistidine at C' terminus.



To construct the four sfGFP variants, each one containing an amber mutation separately at S2, F27, Y66, or F130, mutations were introduced by overlapping extension PCR. The follow pairs of primer were used to generate an amber mutation site in sfGFP variants, respectively: (1) pET-sfGFP-S2TAG\_NdeI-F and pET-sfGFP-sacI-R2. (2) pET-sfGFP-NdeI-F1 and pET-sfGFP-F27TAG-R; pET-sfGFP-F27TAG-F and pET-sfGFP-sacI-R2. (3) pET-sfGFP-NdeI-F1 and sfGFP-Y66TAG-R1; sfGFP-Y66TAG-F2 and pET-sfGFP-sacI-R2. (4) pET-sfGFP-NdeI-F1 and pET-sfGFP-F130TAG-R; pET-sfGFP-F130TAG-F and pET-sfGFP-sacI-R2. The PCR products were digested with restriction enzymes NdeI and SacI-HF, gel-purified and ligated back into pET-pylT vector digested by NdeI and SacI-HF to form pET-pylT-sfGFP-S2TAG, pET-pylT-sfGFP-F27TAG, pET-pylT-sfGFP-Y66TAG, pET-pylT-sfGFP-F130TAG. All gene of sfGFP variants were contained hexahistidine in C-terminus.

To construct the pET-pylT-sfGFP-3xTAG plasmid with three amber mutations at S2, F27, and F130 were introduced by PCR from pET-pylT-sfGFP. The follow pairs of primer were used to generate a sfGFP3xTAG gene: (1) pET-sfGFP-S2TAG\_NdeI-F and pET-sfGFP-F27TAG-R; (2) pET-sfGFP-F27TAG-F and pET-sfGFP-sacI-R2; (3) pET-sfGFP-S2TAG\_NdeI-F and pET-sfGFP-F130TAG-R; (4) pET-sfGFP-F130TAG-F and pET-sfGFP-sacI-R2. The first gene which was produced with (1) and (2) primers were became the template for primer (3) and (4) to form sfGFP-3xTAG finally. The sfGFP-3xTAG gene was subsequently digested and gel-purified and ligation back into pET-pylT

vector forming pET-pylT-sfGFP-3xTAG. The gene was contained hexahistidine in C-terminus.

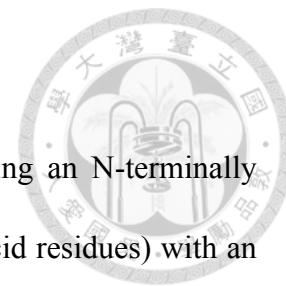


### *Construction of pET-pylT-Ferritin variants*

The pET-pylT-Ferritin plasmid was derived from pET-pylT-sfGFP. This gene was amplified from plasmid which contain the optimize sequence of Ferritin through artificial gene synthesis from MD bio, Inc., by pair of primers, pET-h-Ferritin-NdeI-F and pET-h-Ferritin-SacI-R. Two restriction sites, NdeI at the 5'end and SacI at the 3'end, were introduced in the PCR product which was digested and used to replace sfGFP in pET-pylT-sfGFP. The gene was contained hexahistidine in C-terminus.

To construct the four Ferritin variants, each one containing an amber mutation separately at L56, H60, R63, and E67, mutations were introduced by overlapping extension PCR. The follow pairs of primer were used to generate an amber mutation site in Ferritin variants: (1) pET-h-Ferritin-NdeI-F and h-Ferritin-L56TAG-R; h-Ferritin-L56TAG-F and pET-h-Ferritin-SacI-R. (2) pET-h-Ferritin-NdeI-F and h-Ferritin-H60TAG-R; h-Ferritin-H60TAG-F and pET-h-Ferritin-SacI-R. (3) pET-h-Ferritin-NdeI-F and h-Ferritin-R63TAG-R; h-Ferritin-R63TAG-F and pET-h-Ferritin-SacI-R (4) pET-h-Ferritin-NdeI-F and h-Ferritin-E67TAG-R; h-Ferritin-E67TAG-F and pET-h-Ferritin-SacI-R. The PCR products were digested with restriction enzymes NdeI and SacI-HF, gel-purified and ligated back into pET-pylT vector digested by NdeI and SacI-HF to form pET-pylT-Ferritin-L56TAG, pET-pylT-Ferritin-H60TAG, pET-pylT-Ferritin-R63TAG, pET-pylT-Ferritin-E67TAG. All gene of Ferritin variants were contained hexahistidine in C-terminus.

### *Construction of pET-pylT-N-DFRSc270*

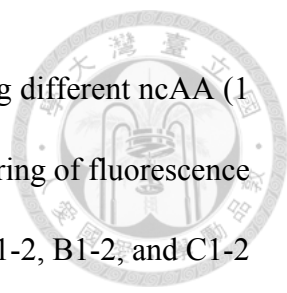


Plasmid pET-pylT-N-DFRSc270, containing the gene encoding an N-terminally truncated *MmPylRS* (residues from 185 to 454; totally 270 amino acid residues) with an N-terminal hexahistidine<sup>43</sup>, was derived from the plasmid pET-pylT-sfGFP. This gene was amplified from the pCDF-N-DFRS plasmid R61H, H63Y, S193R N346G, C348Q, and V401G mutations. The following pairs of primers were used to generate N-DFRSc270 having N-terminal hexahistidine: (1) pET-His6X-F1 and His6x-PylRS185-R1; (2) PylRS-c270-F2 and pET-PylRS-R2. After overlapping PCR, the product was digested with restriction enzyme NdeI and SacI, gel-purified and ligated back into pET-pylT vector to form pET-pylT- N-DFRSc270.

All ligation products in pET-22b (+) plasmid were then transformed into *E. coli* DH5 $\alpha$  cell by chemical procedures. Transformed cells were recovered in LB medium for 60 min at 37°C, then plated on LB agar plates with Amp (100  $\mu$ g/mL), and grown overnight in a 37°C incubator. After incubation, choosing two colonies for DNA sequence.

### **2.3 Screening incorporation efficiency of PylRS variants.**

For substrate screening of PylRS variants, the library containing 359 ncAAs were used. *E. coli* BL21(DE3) were co-transformed with plasmid pET-pylT-sfGFP-F27TAG or pET-pylT-sfGFP-S2TAG and pCDF-PylRS variants. Cells were cultured in LB medium (25 mL) supplemented with Amp (100  $\mu$ g/mL) and Sp (100  $\mu$ g/mL) at 37°C until OD<sub>600</sub> reached 0.8-1.0. The cells were harvested and washed twice with M9 medium, and suspended in M9 medium supplemented with 1 mM IPTG. Aliquots (50  $\mu$ L) of cell

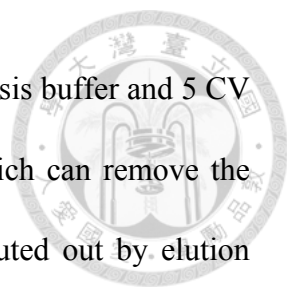


suspension were loaded into wells of 384-well plate which containing different ncAA (1 mM) in each well. Cells were shaken for 12 hrs at 37°C, with monitoring of fluorescence intensity (excitation 535 nm and emission 595 nm). Twelve wells (A1-2, B1-2, and C1-2 without ncAA and IPTG; D1-2, E1-2, and F1-2 without ncAA) were used as controls to measure background signals. The fluorescence intensity of sfGFP incorporating with ncAA was divided with OD<sub>600</sub>, deducted the signal of controls (without IPTG) as *Relative fluorescence intensity* showing in figure.<sup>44,45</sup> To compare fluorescence intensity of different sfGFP variants with ncAA, the screening data was divided with OD<sub>600</sub> and controls (without IPTG and ncAAs), as *Normalized fluorescence intensity*.

## 2.4 Protein productions and purifications

### *Recombinant protein expression and purification*

To express sfGFP protein, *E. coli* BL21(DE3) competent cell was transformed with pET-pyIT-sfGFP. After adding 1 µL plasmid into thawed competent cell and incubating on ice a least 15 min, the competent cell was heated shock at exactly 42°C for 1 min and place on ice 10 min. Cells were recovered in 1 mL of LB for 1 hr at 37°C before being plated on LB agar plate containing Amp (100 µg/mL). A single colony was then selected and grown overnight in a 5 mL culture. The culture was then used to inoculate 500 mL of LB media supplement with 100 µg/mL Amp. Cell were grown at 37°C incubator (200 r.p.m.) and protein induction was induced when OD<sub>600</sub> reached 0.4-0.6 by adding 1 mM IPTG. After induction for 4 hrs, cell was harvest, suspended in the lysis buffer (1xPBS, 400 mM NaCl, pH 8.0) and sonicated. The cell lysate was clarified by centrifugation at 20,000 × g for 30 min in 4°C. The supernatant was collected and transferred into open



column with 3 mL Ni<sup>2+</sup>-NTA resin (Roche) and washed by 10 CV lysis buffer and 5 CV wash buffer (1xPBS, 400 mM NaCl, 5 mM imidazole, pH8.0) which can remove the unbound and non-specific binding protein. Finally, protein was eluted out by elution buffer (1xPBS, 400 mM NaCl, 200 mM imidazole, pH8.0). The solution was immediately changed to 1xPBS using Amicon Ultra-15 Centrifuge Filter Devices (10,000 MWCO cut) (Millopre). The purified protein was analyzed by 12% SDS-PAGE, distributed into Eppendorf (each one 200 μL) and stored in -80°C. Ferritin proteins were expressed and purified similarly.

#### *Expression and purification of protein incorporated with ncAA*

To express sfGFP incorporated with ncAA, *E. coli*. BL21(DE3) competent cell was co-transformed with pET-pylT-sfGFPS-2TAG and pCDF-PylPS-N-DFRS. Cell were recovered in 1 mL LB medium for 1hr at 37°C. After that, the cell was plated on LB agar plate containing Sp (100 μg/mL) and Amp (100 μg/mL). A single colony was then selected and grown overnight in 5ml culture. The 1mL of culture was then used to inoculate 100 mL of LB media supplemented with 100 μg/mL Amp and 100 μg/mL SP. Cell culture was incubated at 37°C incubator (200 r.p.m.) until OD<sub>600</sub> reaching 0.8-1.0. The cell was centrifuged at 10,000 x g for 10 min and removed the supernatant. The *E. coli* cell was washed by M9 minimal medium twice and expressed protein in 500 mL of M9 minimal media supplemented with 1% glycerol, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 0.2% NaCl, 100 μg/mL Sp, 100 μg/mL Amp, 1 mM IPTG, and 1 mM ncAA. After 12 hrs induction at 37°C, cell was harvest. The purification of protein incorporated ncAA was performed by similar method in pervious paragraph. Ferritin proteins incorporated with

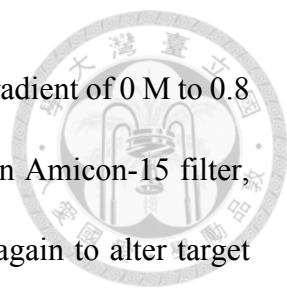
other ncAAs were expressed and purified similarly except the supplemented ncAA was changed. For all ncAAs, 1 mM final concentration was used.



#### *N-DFRSc270 protien expression and purification for crystallization*

To express N-DFRS C' terminal domain (totally 270 amino acids), *E. coli* BL21(DE3) competent cell was transformed with pET-pyIT-N-DFRSc270. Cell was recovered in 1 mL LB medium for 1 hr at 37°C. The method of transform and re-culture were similar to paragraph of *Recombinant protein expression and purification*. The 10 mL of overnight cell culture was used to inoculate 1 L LB medium supplement with 100 µg/ mL Amp. Cell culture was incubated in 37°C and protein induction was induced when OD<sub>600</sub> 0.4-0.6 by adding 0.5 mM IPTG in 25°C overnight. Cell was harvest, suspended in buffer A (50 mM KH<sub>2</sub>PO<sub>4</sub>/K<sub>2</sub>HPO<sub>4</sub>, 500 mM NaCl, 25 mM imidazole, 5 mM β-Me, 10%glycerol, and 0.1 mM PMSF, pH 7.4) and sonicated. The cell lysate was clarified by centrifugation at 20,000 × g for 30 min in 4°C. The supernatant was collected and transferred into open column with 10 mL Ni<sup>2+</sup>-NTA resin (Roche) and washed by 10 CV lysis buffer. Protein was eluted out by buffer A contain 300 mM imidazole. The sample was concentrated with an Amicon-15 filter, 10,000 MWCO (Millipore) and injected into HiLoad 16/60 Superdex 75 pg preparative size exclusion chromatography column (GE healthcare life sciences) equilibrated with buffer B (25 mM KH<sub>2</sub>PO<sub>4</sub>/K<sub>2</sub>HPO<sub>4</sub>, 1 mM DTT, 10% glycerol, and 0.1 mM PMSF, pH 7.4). The SEC column (HiLoad 16/600 superdex 75 pg, GE Healthare) was used to separate N-DFRSc270 with non-specific protein and changed protein from buffer A to buffer B. The N-DFRS c270 fractions were applied onto a mono Q column (GE healthcare life sciences)



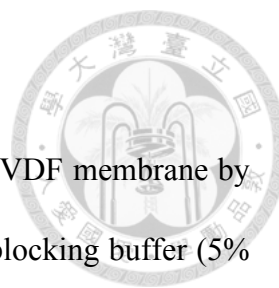


and the column was washed. The proteins were eluted with a linear gradient of 0 M to 0.8 M NaCl in buffer B. The protein fractions were concentrated with an Amicon-15 filter, 10,000 MWCO (Millipore) and loaded into the same SEC column again to alter target protein from buffer B containing high concentration of salt to Buffer C (20 mM K/ PO<sub>4</sub>, 300 mM NaCl, 5 mM MgCl<sub>2</sub>, and 10 mM β -Me, pH 7.4) for crystallization. The N-DFRSc270 fractions were pooled and concentrated with an Amicon-15 filter, 10,000 MWCO (Millipore). Finally, the purified protein was analyzed by 12% SDS-PAGE, distributed into eppendorf (each one 100 μL) and stored in -80°C.<sup>43</sup>

## **2.5 Gel analysis**

### **2.5.1 SDS-PAGE analysis**

For SDS-PAGE, the amount of protein samples was determined by Bradford Protein Assay (Bio-Rad). 5 μg protein samples were mixed in 5 μL 4x sample buffer (200 mM Tris-HCl, 400 mM DTT, 8% SDS, 0.04% bromophenol blue, and 40% glycerol, pH 6.8) to final volume 20 μL and boiled at 95°C heating block module for 10 min. Loading 15 μL prepared samples into well slowly as well as 2.5 μL protein marker (Bio-Rad) in 12% SDS-PAGE which placed in electrode assembly, put into electrophoresis tank and filled with TGS buffer (25 mM Tris-HCl, 192 mM glycine, 0.1% SDS, pH 8.5). The electrophoresis was run under 80 V for 20 min and 120 V for 60 min. After the total running time, the gel was stained by instant blue (Marvelgent Biosciences) more than 15min and destained by H<sub>2</sub>O. The bands of protein could be visualized.



### 2.5.2 Western blot analysis

After electrophoresis, the proteins were transfer from gel to PVDF membrane by western blotting transfer kit (Bio-Rad). Blocking the membrane in blocking buffer (5% non-fat dried milk dissolved in PBST for 1 hr at room temperature. And membrane was washed by PBST for 10 min three times. The primary antibody (mouse anti-hexahistidine) was diluted in PBST to final concentration 1  $\mu\text{g}/\text{mL}$  and incubated with membrane for 1 hr at room temperature or at 4°C overnight. The incubated membrane was washed with PBST for 10 min three times. Subsequently, the membrane was incubated with secondary antibody (peroxidase-conjugated rabbit anti-mouse IgG light chain Ab) which was diluted 2000 times in PBST for 1 hr at room temperature and was washed with PBST for 10 min three times. The signals were detected with ECL Western blot detection reagents (Thermo Fisher Scientific).

### 2.5.3 Native gel analysis

For native gel, the amount of protein samples was determined by Bradford Protein Assay (Bio-Rad). 5-10  $\mu\text{g}$  protein samples were mixed with 10  $\mu\text{L}$  2x Native sample buffer (100 mM Tris-HCl, 0.02% bromophenol blue, and 20% glycerol, pH 6.8) to final volume 20  $\mu\text{L}$ . Loading 20  $\mu\text{L}$  prepared samples into well slowly as well as 4  $\mu\text{L}$  native protein marker (Thermo Fisher Scientific) in 4–15% Mini-PROTEAN<sup>®</sup> TGX<sup>™</sup> Precast Protein Gel (Bio-Rad) which placed in electrode assembly, put into electrophoresis tank and filled with TG buffer (25 mM Tris-HCl, 192 mM glycine, pH 8.5). The electrophoresis was run under 80 V for 6 hrs at 4°C. After the total running time, the gel

was stain by instant blue (Marvelgent Biosciences) more than 15 min and destained by H<sub>2</sub>O. The bands of protein could be visualized.



## 2.6 Protein Biophysical characterizations

### 2.6.1 X-crystal structural analysis

#### *Crystallization of PylRSc270 and soaking*

The crystals of N-DFRSc270 (12 mg/mL) with 5 mM ANP was obtained at 293K by the hanging drop vapor-diffusion method in 100 mM HEPES buffer at pH 7.5, 10% PEG 8000, 10% Ethylene Glycerol. For soaking, 5 mM NCAA was added into crystallization drops 1 hr before collecting crystals. Additional 20% glycerol was added to the crystallization buffer as a cryoprotectant before removing the crystals from crystallization drops. The crystal of sfGFP-3TAG-**3** was obtained at 293K by sitting drop vapor-diffusion method in 100 mM HEPES buffer at pH 7.5, 30% PEG 400, 0.2 M MgCl. Diffraction intensities were integrated and scaled by HKL2000 package<sup>46</sup>. Molecular replacements was carried out using previous reported PylRS (PDBID: 2ZCE) as search model through PHENIX-Phaser-MR<sup>47</sup>. For sfGFP-3TAG-**3** the search model is sfGFP (PDBID: 2B3P). Final models were built after several runs of refinement in phenix.refine<sup>48</sup> and manual re-building in Coot<sup>49</sup>. The final models were analyzed with PROCHECK<sup>50</sup>. Structural presentations were generated using the PyMOL Molecular Graphics System (Version 1.7.4 Schrödinger, LLC).

## 2.6.2 Protein ESI-MS and MALDI-TOF-MS/MS analysis

### *LC-ESI-MS/MS Analysis*

Peptide samples were reconstituted in buffer A (0.1% FA in H<sub>2</sub>O) and analyzed on Waters Synapt G2 HDMS (Waters, Milford, MA). Samples were injected onto a 2 cm × 180 μm capillary trap column and separated on a 75 μm × 25 cm nanoACQUITY 1.7 μm BEH C18 column using a nanoACQUITY Ultra-Performance LC system (Waters, Milford, MA). The column was maintained at 35°C and bound peptides were eluted with a gradient of 5-35% buffer B (buffer A, 0.1% FA in H<sub>2</sub>O; buffer B, 0.1% FA in ACN) for 30 min. The MS was operated in ESI positive V mode with a resolving power of 10000 and calibrated with a synthetic human [Glu1]-Fibrinopeptide B solution (0.2 μmol/mL, Sigma-Aldrich) delivered through the NanoLockSpray source. Data acquisition was performed using data directed analysis (DDA) method. The DDA method included one full MS scan (m/z 350-1600, 1 s/scan) and three MS/MS scans (m/z 100-2000, 1 s/scan) performed sequentially on the three most intense ions present in the full scan mass spectrum.

The MS and MS/MS raw data were converted into peak lists using ProteinLynx Global Server 2.5 (Waters, Milford, MA). Analysis of all MS/MS samples was performed using Mascot (Matrix Science; version 2.5.0). Mascot was set up to search the GFP for IBC database (1 sequences), assuming trypsin as the digestion enzyme. Database search against Mascot was performed with a fragment ion mass tolerance of 0.05 Da and a parent ion tolerance of 25 ppm. Two missed cleavages were allowed during trypsin digestion. Oxidation (Met) and carbamidomethyl (Cys) were specified as variable modification.



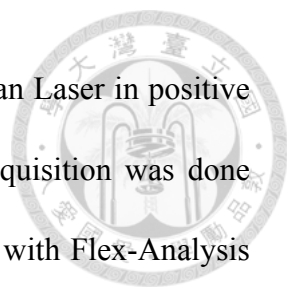
### *ESI-MS Analysis*

The pure protein was diluted with 50% acetonitrile and 1% formic acid. An aliquot corresponding to one pmol of the pure protein was injected via a ESI source (Waters LockSpray Exact Mass Ionization Source) with a syringe pump (Harvard Apparatus, MA) and held a flow rate of 5  $\mu$ l/min throughout the analysis. The mass of intact proteins was determined using Waters Synapt G2 HDMS mass spectrometer (Waters, Milford, MA). The acquired spectra were deconvoluted to single-charge state using MaxEnt1 algorithm of the MassLynx 4.1 software (Waters).

### *MALDI-TOF-MS/MS Analysis*

**In-gel digestion:** after the staining procedure, gel bands were exercised, cut into small pieces. A modified in-gel digestion protocol was applied.<sup>51</sup> Briefly, after sequentially washing the gel pieces with 25 mM  $\text{NH}_4\text{HCO}_3$ , 40% methanol solution, and 100% acetonitrile. The reduction with DTT and alkylation with iodoacetamide of proteins in gel pieces were performed and the gel pieces were washed and dried in a vacuum centrifuge before trypsin digestion. A trypsin solution in 25 mM  $\text{NH}_4\text{HCO}_3$ , 10% acetonitrile containing 65 to 100 ng of sequencing grade modified trypsin (Promega) in 25-30  $\mu$ L was added and incubated with gel pieces for 12-16 hrs at 37°C. Stop reaction by adding 1-2  $\mu$ L of 5% formic acid.

**MS and data analysis:** the digested samples (0.5  $\mu$ L) were carefully mixed with matrix solution (0.5  $\mu$ L of 5 mg/ml DHB in 0.1%TFA/30% acetonitrile) and deposited 0.5  $\mu$ L of mixture onto the MTP 600/384 AnchorChip (Bruker Daltonics,). All mass spectrometry experiments were done using a Bruker Autoflex III MALDI TOF/TOF mass



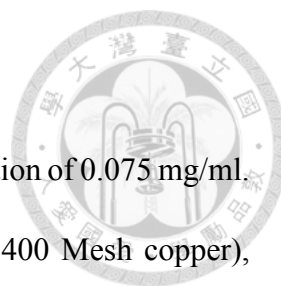
spectrometer (Bremen, Germany) equipped with a 200 Hz SmartBean Laser in positive ion mode with delayed extraction in the reflectron mode. Data acquisition was done manually with FlexControl 3.4, and data processing was performed with Flex-Analysis 3.4 (both Bruker Daltonik). Protein database searches, through Mascot, using combined PMF and MS/MS datasets were performed via Biotoools 3.2 (Bruker).

### **2.6.3 $T_m$ analyzed by CD**

The sfGFP-Y66TAG with D/L ncAA and wt-sfGFP were fresh prepared and diluted to concentration 0.5 mg/mL in PBS at pH 8.0. All CD spectra were collected on J-815 Circular Dichroism Spectropolarimeter (Jasco). And the samples were loaded in a 1 mm path-length cuvette. Denaturation midpoint experiments were carried out by monitoring the changes in CD spectra at 210 nm to 250 nm from 25°C to 95°C with a 2 °C interval and equilibrating to keep within +/- 0.1°C for 10 seconds. Refolding experiments were returned temperature from 95°C to 25°C and also measured CD spectra every 2°C. The melting curves were analyzed and fitted using Origin 8.

### **2.6.4 DLS analysis**

Ferritin and Ferritin variants were diluted in PBS to final concentration of 0.3-0.5 mg/mL and filtered to remove aggregation or dust in solution. Experiment were conducted with a Malvern Zetasizer Nano ZS at 25°C in 1 mL disposable cuvettes and repeated at least three measurements. Measurements were performed a primary size distribution by intensity value.



### **2.6.5 TEM analysis**

The samples were dissolved in PBS at pH 8.0 to final concentration of 0.075 mg/ml. And loading 5  $\mu$ l sample on a support films grid (formvar/carbon 400 Mesh copper), discharged with 25 mA for 30 seconds by EMITECH K100X Glow, for 90 seconds then the excess solution was wicked away with filter paper. The grids were rinsed with 5  $\mu$ l of distilled deionized water, immediately wicked away and repeated this step three times. 5  $\mu$ l of aqueous freshly filtered 1% uranyl acetate was added on grids for 60 seconds then wicked away. Keeping the grids in the dish and put in electronic dry cabinet until the grids fully dry. Images of Ferritin and Ferritin variants were collected on FEG-TEM, FEI Tecnai G2 TF20 Super TWIN. The microscope was used at an accelerating voltage of 120 kV.

## **2.7 Emission and Excitation analysis**

### **2.7.1 UV/Visible absorption spectrum**

The sfGFP-Y66TAG with D/LFA and wt-sfGFP were diluted to final concentration of 0.5 mg/ml in PBS at pH 8.0. Absorption spectra were acquired using a 1 cm quartz cuvette at 25°C on an V-630-Bio (Jasco). The following information was parameter setting: photometric mode\_ Abs, response\_Fast, UV/Vis bandwidth\_1.5 nm, scan speed\_200 nm/min. The Scanning wavelength was started from 600 nm to 230 nm.

### **2.7.2 Fluoresces spectrum**

The sfGFP-Y66TAG with D/LFA and wt-sfGFP were diluted to final concentration of 0.5 mg/mL in PBS at pH 8.0. Then the samples were loaded into 1 cm quartz cuvette

and measured by Fluorolog-3 (Jobin Yvon). Excitation wavelength of sfGFP-Y66TAG with D/L ncAA is 370 nm and wt-sfGFP is 488 nm. Separately, the emission spectra were recorded from 380 to 600 nm and 500 to 600 nm every 5 nm.





## Chapter 3 Results

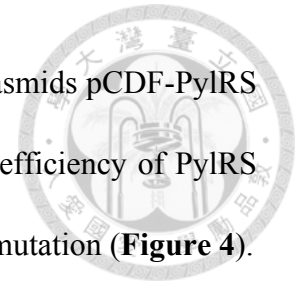


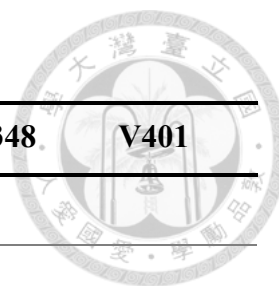
### 3.1 PylRS designing for high efficiency incorporating D-aa

#### 3.1.1 Engineering *MmPylRS* and sfGFP

In order to introduce bioorthogonal pair into protein translation system. Two type of plasmid was constructed. One is pCDF-1b plasmid encoded with wild-type *MmPylRS* or PylRS variant, which was focused on modifying N-terminus and catalytic pocket. The published lecture have revealed that N-terminus of PylRS interact with  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  and mutant at specific site on N-terminal domain can increase efficiency of incorporation.<sup>41</sup> Therefore, PylRS was modified at R61K, H63Y, and S193R in N-terminal domain, naming N-PylRS. In addition, Wenshe R. Liu and co-works have reported the ability of rationally designed PylRS mutant at N346A and C348A which were in the catalytic pocket, coupled with  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$ , to genetically incorporate a wide variety of phenylalanine derivatives into protein.<sup>52-54</sup> To expand the substrate range of PylRS, the enzyme was not only mutated at N346 and C348 with different amino acid but also mutated one more site at V401 position. The following PylRS variants was facilitated to explore more ncAAs that unknown before: LFRS containing mutation at N346V; DFRS containing mutations at N346G, C348Q, and V401G; and N-DFRS containing mutations at R61K, H63Y, S193R, N346G, C348Q, and V401G. (**Table 1**) Another one is pET-22b (+) plasmid encoding  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  and gene of target protein, sfGFP with amber codon (TAG) which was arranged at position S2, F27, and hexahistidine at C' terminus.<sup>53,54</sup> S2 and F27 is located at the random coli and second  $\beta$ -strand, respectively, of sfGFP. The side chain of F27 is toward to interior of  $\beta$ -barrel and is surrounded by several hydrophobic residues (**Figure 3**). All constructed plasmids were confirmed by DNA

sequencing. Subsequently, *E. coli* BL21(DE3) cell was harbored plasmids pCDF-PylRS variants and pET-pylT-sfGFP-S2TAG or F27TAG to evaluate the efficiency of PylRS variants•tRNA<sub>CUA</sub><sup>Pyl</sup> pair to incorporate ncAAs in response to amber mutation (**Figure 4**).

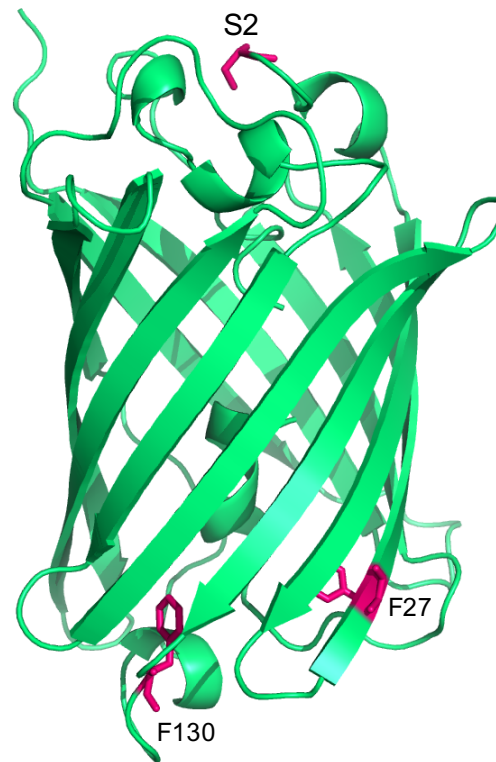




**Table 1 Mutated residues of PylRS variants in this study**

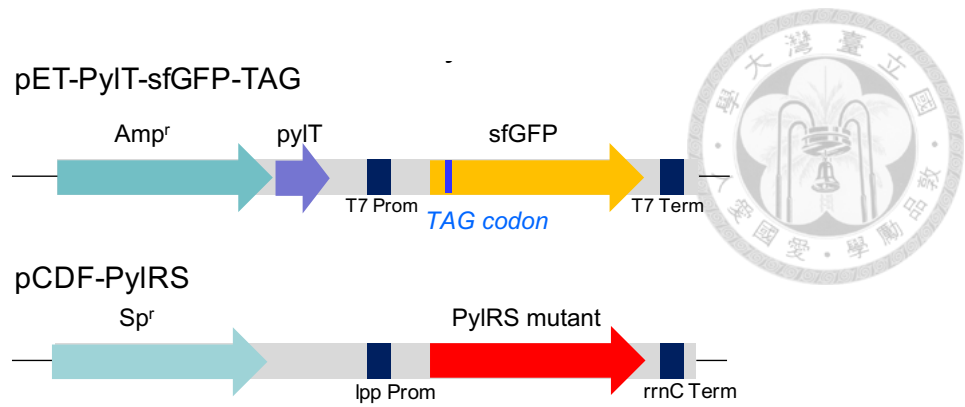
<b>wt-PylRS</b>	<b>R61</b>	<b>H63</b>	<b>S193</b>	<b>N346</b>	<b>C348</b>	<b>V401</b>
N-PylRS	K	Y	R			
LFRS				V		
DFRS				G	Q	G
N-DFRS	K	Y	R	G	Q	G

Residues at *MmPylRS* N-terminus: R61, H63, and S193; active site: N346, C348, and V401.



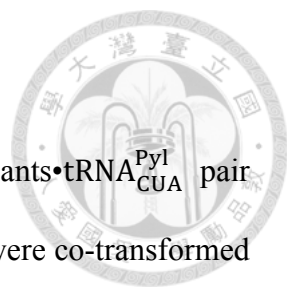
**Figure 3. The x-ray crystal structure of wild-type sfGFP.**

The  $\beta$ -barrel structure of sfGFP was illustrated by green ribbon. The amber mutation sites of sfGFP were highlighted by stick mode in red color. The cartoon figure was made by PyMOL. PDB code: 2b3p.



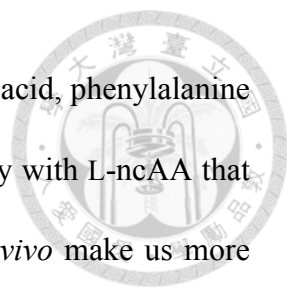
**Figure 4. Two plasmid system design for incorporating ncAA into protein in *E. coli*.** pET-22b (+) plasmid with Amp selection marker encode tRNA<sub>CUA</sub><sup>Pyl</sup> with anticodon CUA and report protein, sfGFP, with amber mutation. T7 promoter regulate target proteins expression by inducer (IPTG) and tRNA under the control of lpp promoter. Another one is pCDF-1b plasmid with Sp selection marker containing *MmPylRS* variants gene which was also controlled by lpp promoter.

### 3.1.2 High throughput screening of pylRS variants



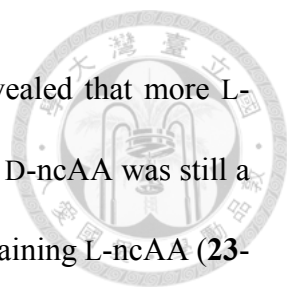
For the purpose of rapidly screening the ability of PylRS variants•tRNA<sub>CUA</sub><sup>Pyl</sup> pair to recognize the ncAAs substrates. The *E. coli* (BL21) DE3 strains were co-transformed with pCDF-PylRS variants and pET-pylT-sfGFP-S2TAG or F27TAG and grown in LB medium until OD<sub>600</sub> 0.8-1.0. After washing pellet to remove 20 canonical amino acids and re-suspended in minimal medium with 1 mM IPTG. The strains were cultured in 384-well plate with each well contained one kind of ncAA in our library at 1 mM. The library contained 359 ncAAs. Subsequently, the fluorescence of sfGFP-S2TAG or F27TAG, used as a measure of the ability of PylRS variants•tRNA<sub>CUA</sub><sup>Pyl</sup> pair to incorporate ncAAs, was detected by FilterMax F5 plate reader at 37°C for 12 hrs. If the fluorescence signal obviously appeared in the well, it represents various PylRS •tRNA<sub>CUA</sub><sup>Pyl</sup> pairs with efficient ncAA incorporation. This is because sfGFP S2TAG or F27TAG can be incorporated with ncAA at amber codon, no termination, by PylRS variants•tRNA<sub>CUA</sub><sup>Pyl</sup> pair and formed full-length protein to emit fluorescence.<sup>44,45</sup>

Initially, the strain harboring PylRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair and sfGFP-S2TAG was screened. The profile of substrate range demonstrated that the enzyme preferred to incorporate lysine derivatives **14** and **15** (**Figure 6**). A similar substrate preference was observed with N-PylRS/sfGFP-S2TAG group (**Figure 7**). In order to broaden substrate range of PylRS, the mutation site was designed at catalytic pocket N346V, naming LFRS. The N346 side chain of PylRS is adjacent to substitution on Phe derivatives. Removing the N346 side chain would potentially leave space for incorporating Phe derivatives.<sup>52</sup> After screening production of sfGFP-S2TAG, the substrate specificity profile indicated that aromatic-ring containing L-ncAAs (**21**, **38**, **50...**) were substrates with appreciable



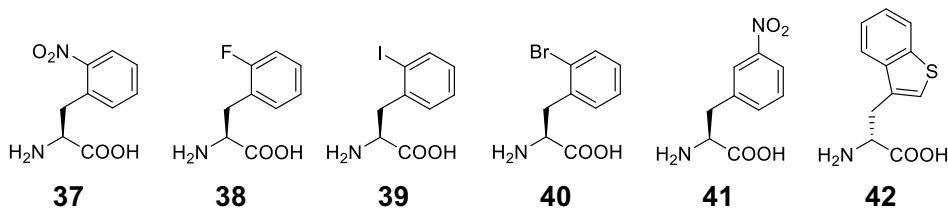
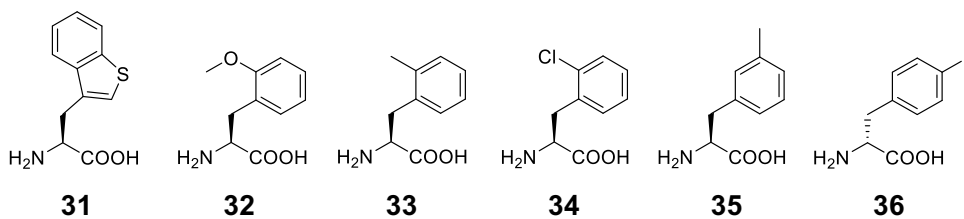
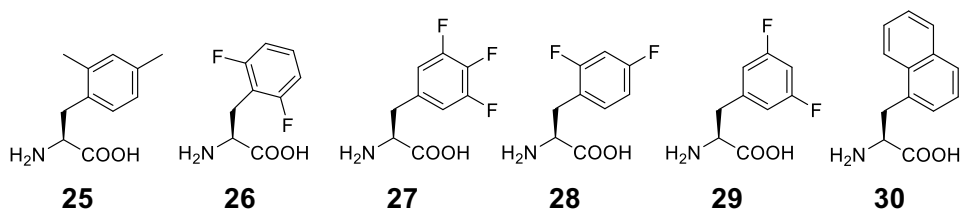
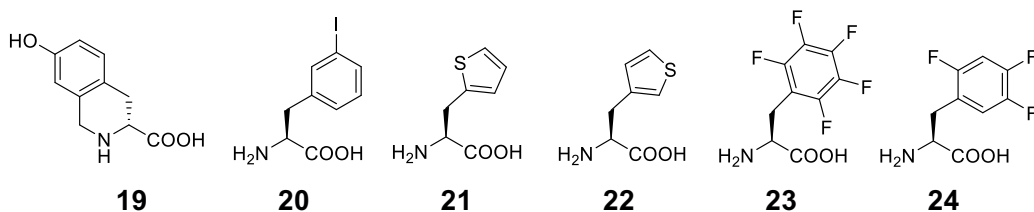
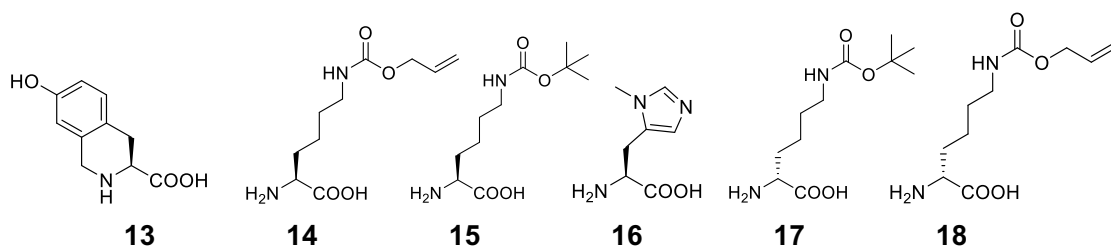
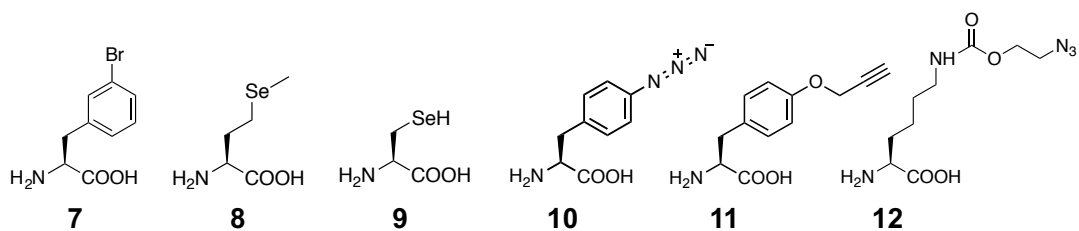
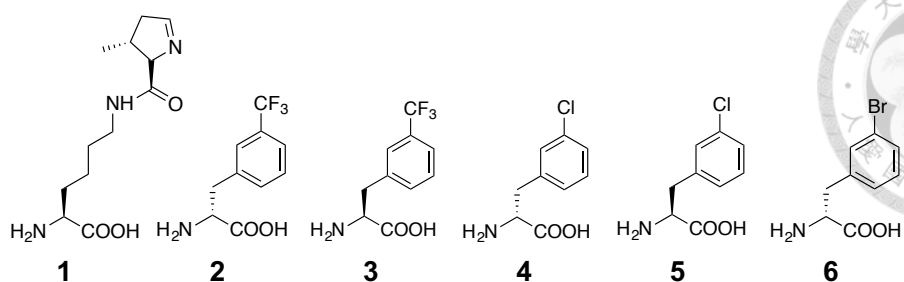
signal, in addition to this, LFRS also can recognize canonical amino acid, phenylalanine **55 (Figure 8)**. This three of PylRS variants provided good efficiency with L-ncAA that have studied previously, but incorporating D-ncAA into protein *in vivo* make us more interested in. Further, rationally evolving PylRS mutations at N346G, C348Q, and V401G enlarged the amino acid binding site, naming DFRS. With similar condition, the DFRS/sfGFP-S2TAG group was screening for 12 hrs. Interestingly, the result revealed DFRS could genetically incorporate not only L-aromatic amino acid (**3, 5, 7...e.g.**) but also D-aromatic amino acid (**2, 4, and 6 (Figure 9)**). D-amino acid is absence in normal translation system but is usually used in peptide drug or bacterial cell wall.<sup>55,56</sup> Comparing suppression efficiency of L-ncAA with D-ncAA, D-ncAA still lower than others. Continually, the residue on the NTD was modified, which interacted with tRNA<sub>CUA</sub><sup>Pyl</sup>, of DFRS to improve catalytic properties<sup>41</sup>, naming N-DFRS. As expected, the screening data revealed that the fluorescence intensity of sfGFP-S2TAG with D-ncAA (**2, 4, and 6**) were significantly increased comparing with DFRS/sfGFP-S2TAG group, although the pattern of substrate range was similar (**Figure 10**).

In order to explore whether the position of amber codon in protein influenced the suppression efficiency of PylRS variants, or not. *E. coli*. was co-transformed with PylRS variants•tRNA<sub>CUA</sub><sup>Pyl</sup> pairs separately and sfGFP-F27TAG, which's amber codon located at second  $\beta$ -strand, in *E. coli* BL21(DE3). Comparing screening result with sfGFP-S2TAG groups, the substrate ranges of PylRS variants had slight change. In addition to recognize L-lysine derivatives (**14 and 15**), PylRS apparently incorporated D-Lysine derivatives (**17 and 18**) and proline derivatives (**13 and 19**) into sfGFP-F27TAG (**Figure 11**). In N-PylRS group, the fluorescence intensities with phenylalanine derivatives (**16 and 20-22**) were



observed (**Figure 12**). And the suppression efficiency of LFRS revealed that more L-phenylalanine derivatives (**33, 34, 37**...e.g.) could be recognized but D-ncAA was still a trace (**Figure 13**). Furthermore, many members of aromatic ring containing L-ncAA (**23-40**) could be incorporated by DFRS and sfGFP-F27TAG with D-ncAA (**2, 4, and 6**) also were observed in the substrate range profile (**Figure 14**). Finally, the enzyme, N-DFRS, demonstrated the highest catalytic efficiency for D-ncAA (**2, 4, and 6**) and also broadened the substrate range, comparing with sfGFP containing amber mutation at position S2 (**Figure 15**).





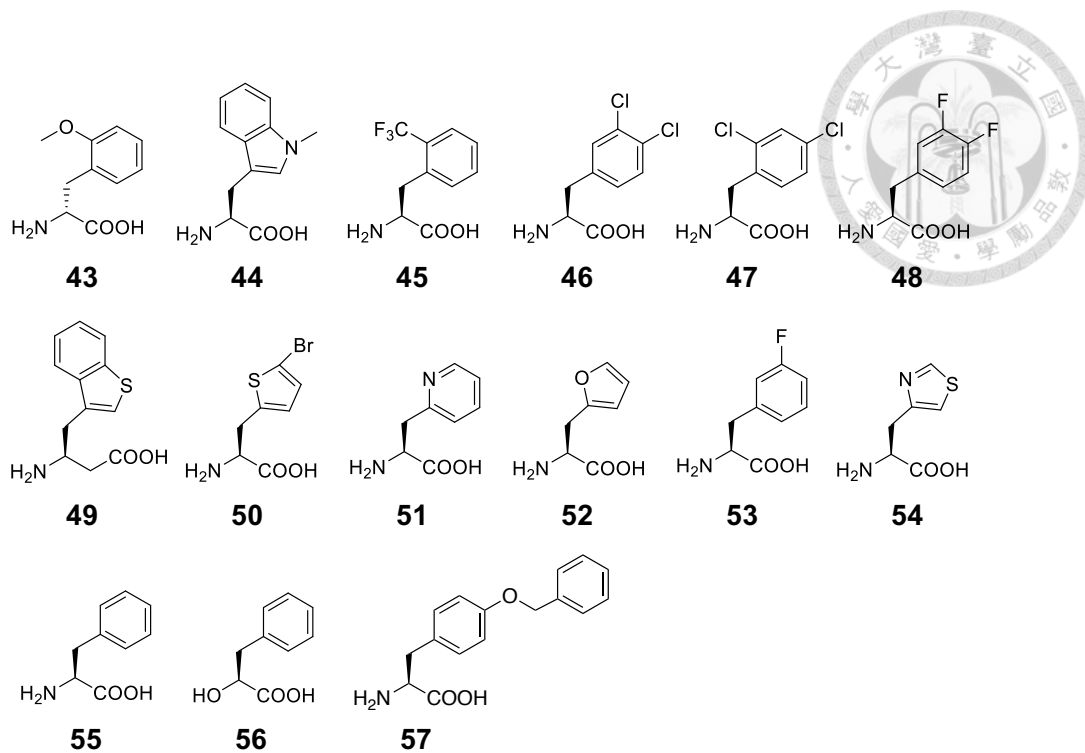
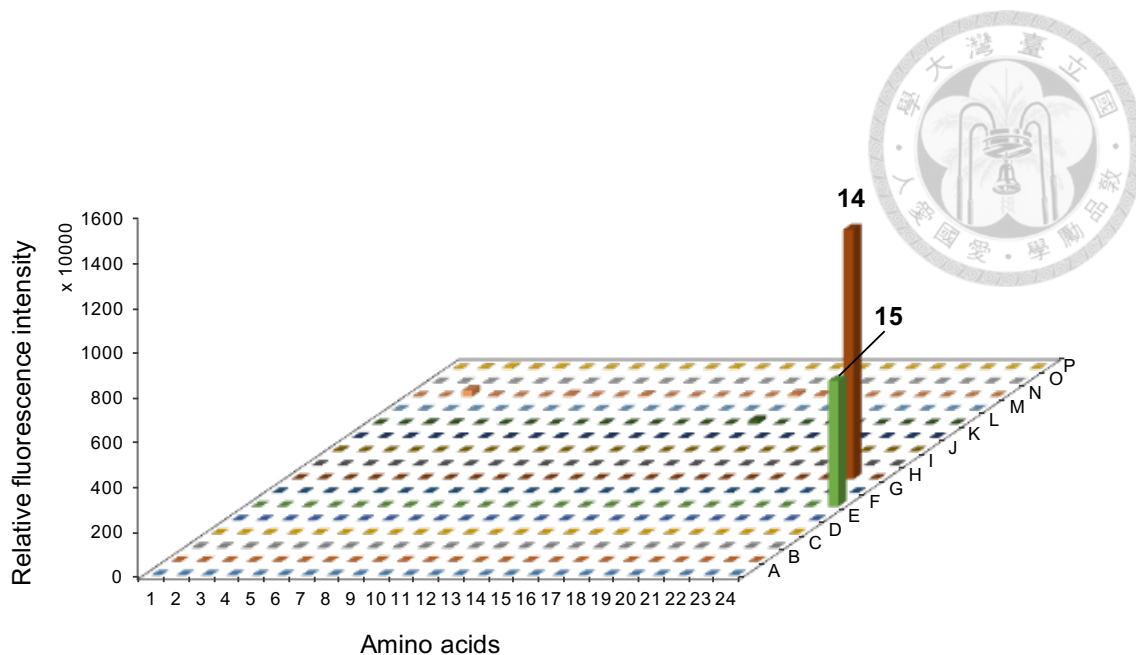
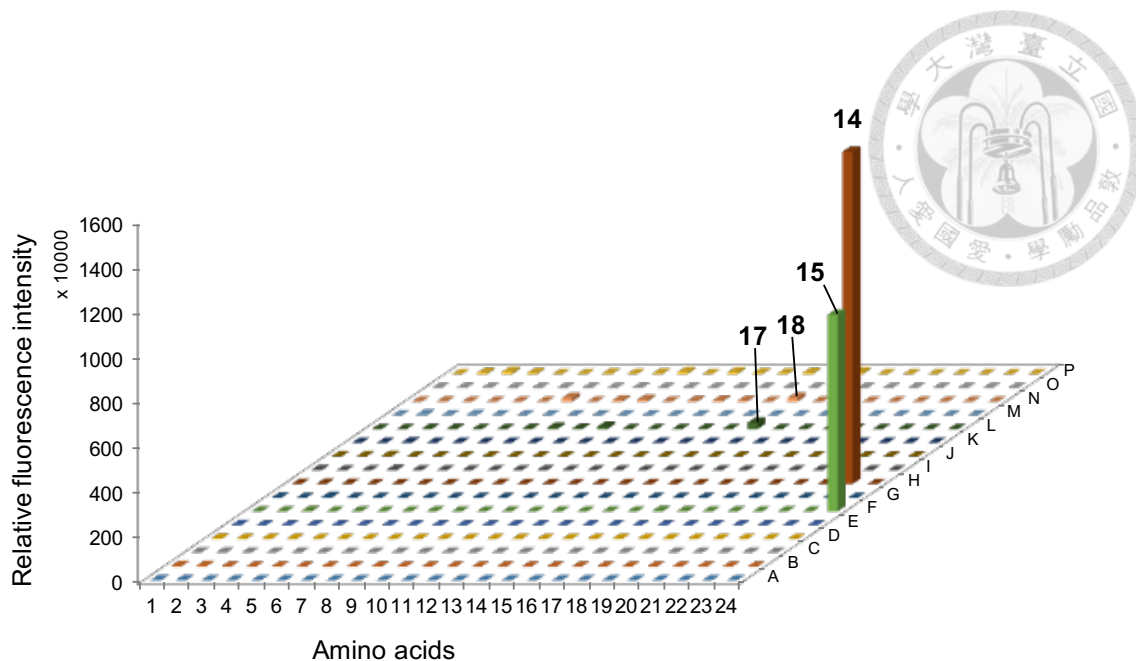


Figure 5. Chemical structures of ncAAs in this study.



**Figure 6. Range of substrate specificities of PyIRS with sfGFP-S2TAG.**

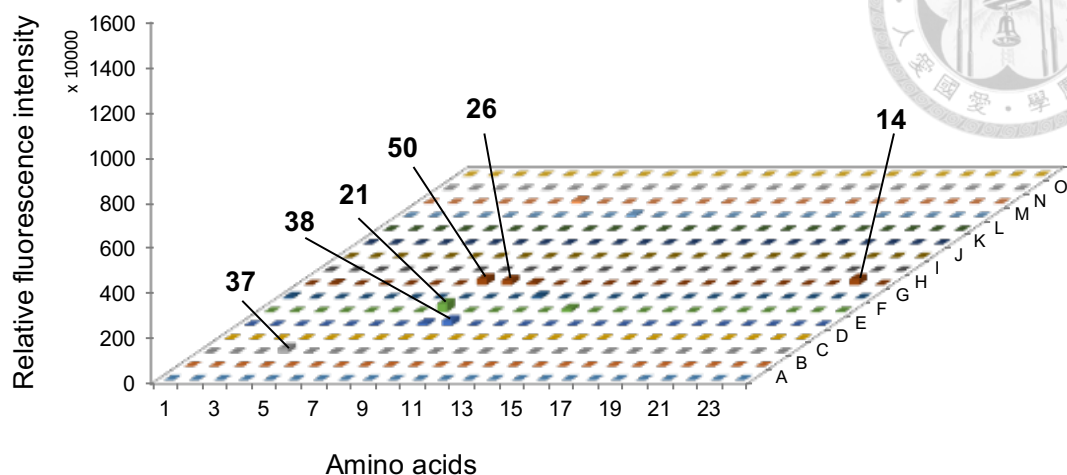
Read through of sfGFP-S2TAG with library of ncAA was measured by fluorescence intensity. Six wells (A1-2, B1-2, and C1-2) without ncAA and IPTG, the other six wells (D1-2, E1-2, and F1-2) without ncAA were controls to quantify background and thus non-production of sfGFP. The concentration of each ncAA was 1 mM with 1 mM IPTG in M9 medium at 37°C for 12 hrs. *E. coli* BL21(DE3) was used in these experiments. The library of ncAA chemical structures are given in **Figure 5**. (PyIRS: *Methanosarcina maize* PyIRS)



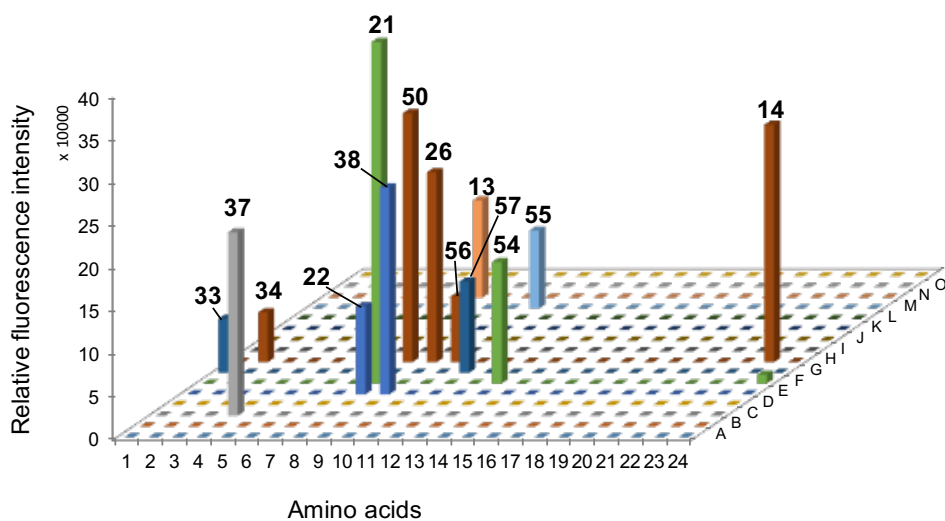
**Figure 7. Range of substrate specificities of N-PylRS with sfGFP-S2TAG.**

Read through of sfGFP-S2TAG with library of ncAA was measured by fluorescence intensity. Six wells (A1-2, B1-2, and C1-2) without ncAA and IPTG, the other six wells (D1-2, E1-2, and F1-2) without ncAA were controls to quantify background and thus non-production of sfGFP. The concentration of each ncAA was 1 mM with 1 mM IPTG in M9 medium at 37°C for 12 hrs. *E. coli* BL21(DE3) was used in these experiments. The library of ncAA chemical structures are given in **Figure 5**. (N-PylRS: *MmPylRS*-R61K/H63Y/S193R)

(A)



(B)

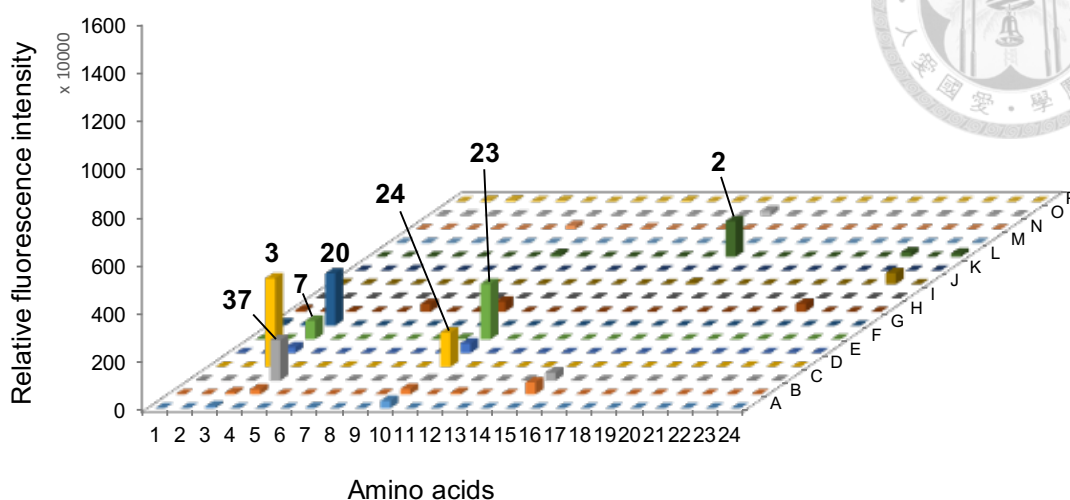


**Figure 8. Range of substrate specificities of LFRS with sfGFP-S2TAG.**

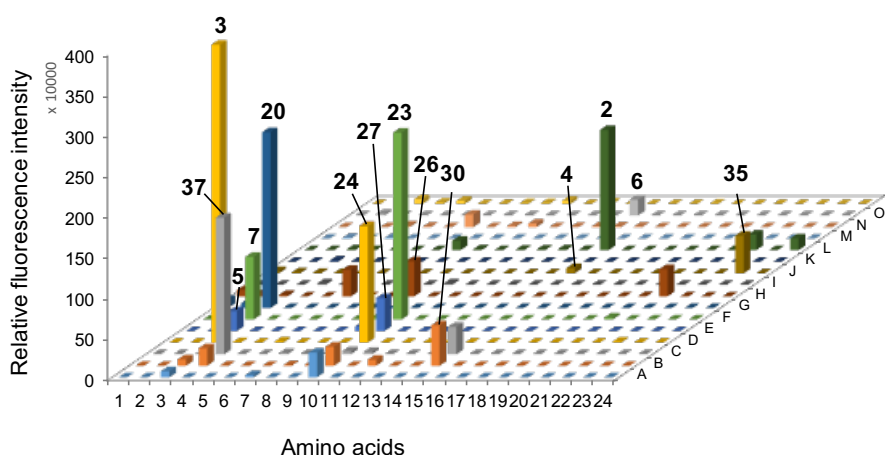
(A) Read through of sfGFP-S2TAG with library of ncAA was measured by fluorescence intensity. Six wells (A1-2, B1-2, and C1-2) without ncAA and IPTG, the other six wells (D1-2, E1-2, and F1-2) without ncAA were controls to quantify background and thus non-production of sfGFP. The concentration of each ncAA was 1 mM with 1 mM IPTG in M9 medium at 37°C for 12 hrs. *E. coli* BL21(DE3) was used in these experiments. The library of ncAA chemical structures are given in **Figure 5**. (B) Zoom-in view. (LFRS: *MmPylRS-N346V*)



(A)

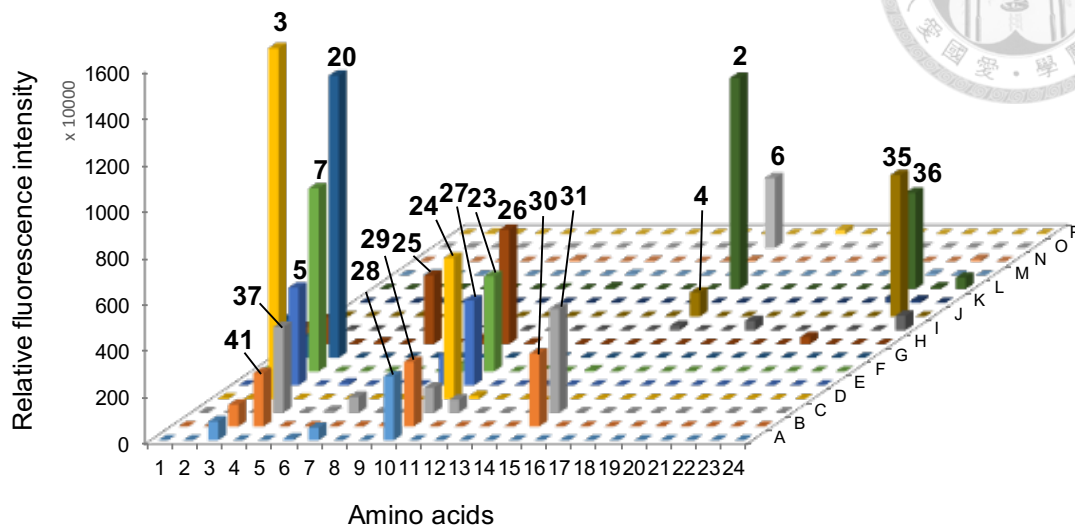


(B)



**Figure 9. Range of substrate specificities of DFRS with sfGFP-S2TAG.**

(A) Read through of sfGFP-S2TAG with library of ncAA was measured by fluorescence intensity. Six wells (A1-2, B1-2, and C1-2) without ncAA and IPTG, the other six wells (D1-2, E1-2, and F1-2) without ncAA were controls to quantify background and thus non-production of sfGFP. The concentration of each ncAA was 1 mM with 1 mM IPTG in M9 medium at 37°C for 12 hrs. *E. coli* BL21(DE3) was used in these experiments. The library of ncAA chemical structures are given in **Figure 5**. (B) Zoom-in view. (DFRS: *Mm*PyIRS-N346G/C348Q/V401G)

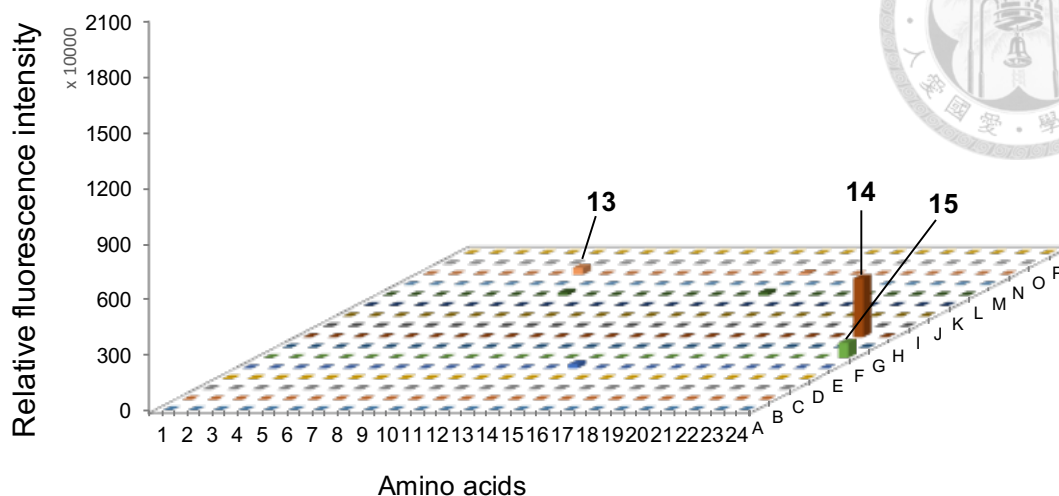


**Figure 10. Range of substrate specificities of N-DFRS with sfGFP-S2TAG.**

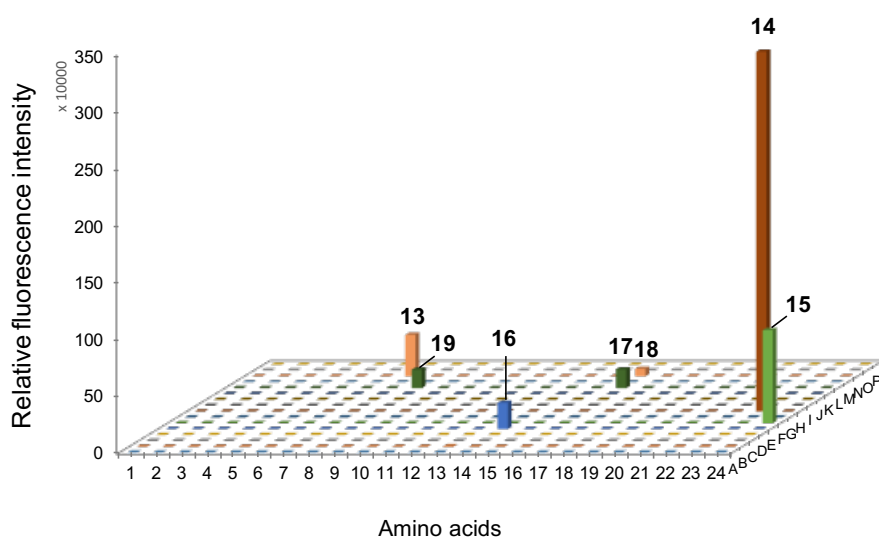
Read through of sfGFP-S2TAG with library of ncAA was measured by fluorescence intensity. Six wells (A1-2, B1-2, and C1-2) without ncAA and IPTG, the other six wells (D1-2, E1-2, and F1-2) without ncAA were controls to quantify background and thus non-production of sfGFP. The concentration of each ncAA was 1 mM with 1 mM IPTG in M9 medium at 37°C for 12 hrs. *E. coli* BL21(DE3) was used in these experiments. The library of ncAA chemical structures are given in **Figure 5**. (N-DFRS: *Mm*PyIRS-R61K/H63Y/S193R/N346G/C348Q/V401G)



(A)



(B)

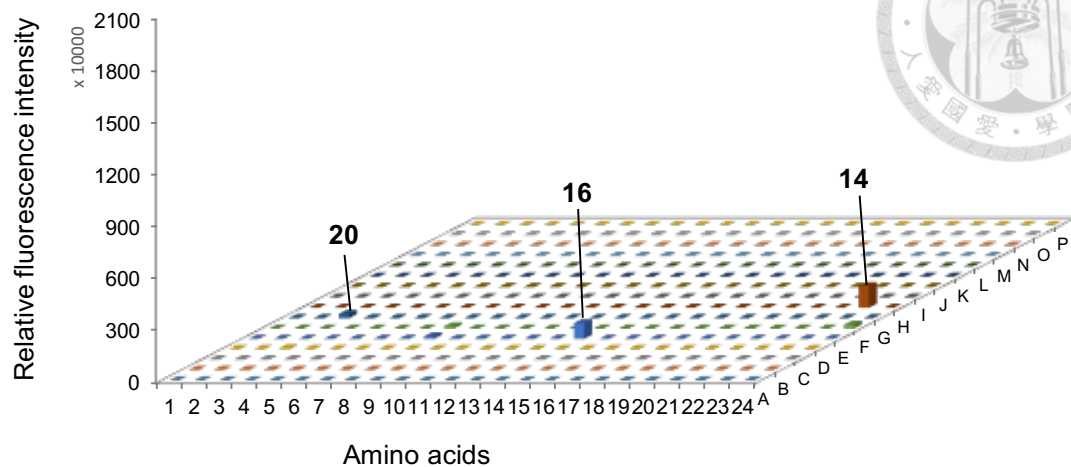


**Figure 11. Range of substrate specificities of PylRS with sfGFP-F27TAG.**

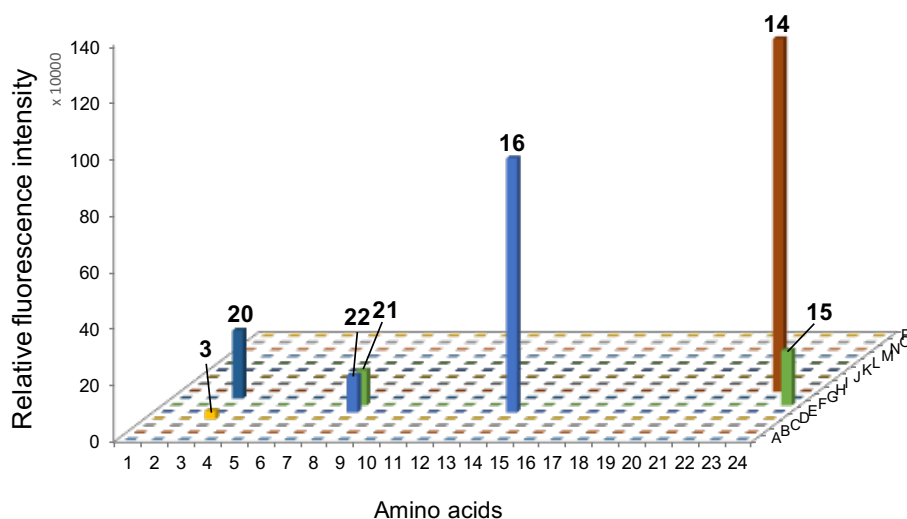
(A) Read through of sfGFP-F27TAG with library of ncAA was measured by fluorescence intensity. Six wells (A1-2, B1-2, and C1-2) without ncAA and IPTG, the other six wells (D1-2, E1-2, and F1-2) without ncAA were controls to quantify background and thus non-production of sfGFP. The concentration of each ncAA was 1 mM with 1 mM IPTG in M9 medium at 37°C for 12 hrs. *E. coli* BL21(DE3) was used in these experiments. The library of ncAA chemical structures are given in **Figure 5**. (B) Zoom-in view. (PylRS: *Methanosarcina maize* PylRS)



(A)



(B)

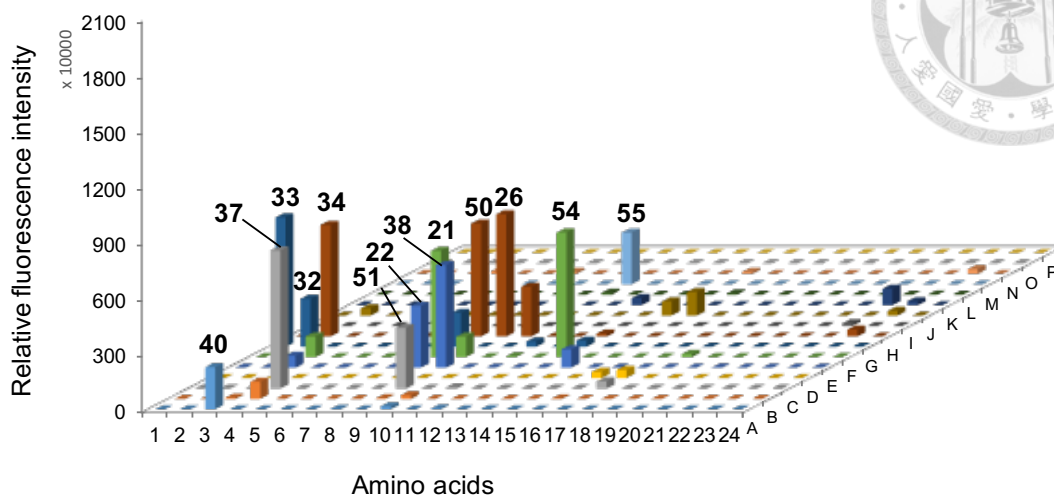


**Figure 12. Range of substrate specificities of N-PylRS with sfGFP-F27TAG.**

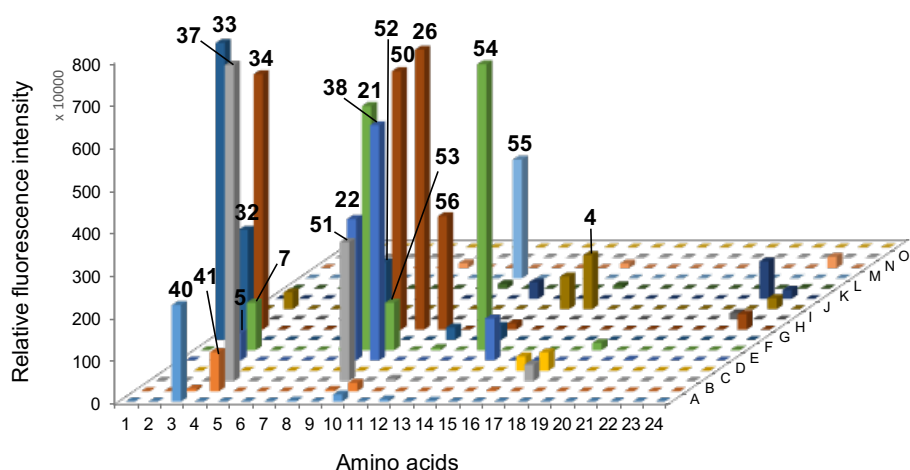
Read through of sfGFP-F27TAG with library of ncAA was measured by fluorescence intensity. Six wells (A1-2, B1-2, and C1-2) without ncAA and IPTG, the other six wells (D1-2, E1-2, and F1-2) without ncAA were controls to quantify background and thus non-production of sfGFP. The concentration of each ncAA was 1 mM with 1 mM IPTG in M9 medium at 37°C for 12 hrs. *E. coli* BL21(DE3) was used in these experiments. The library of ncAA chemical structures are given in **Figure 5**. (B) Zoom-in view. (N-PylRS: *Mm*PylRS-R61K/H63Y/S193R)



(A)



(B)

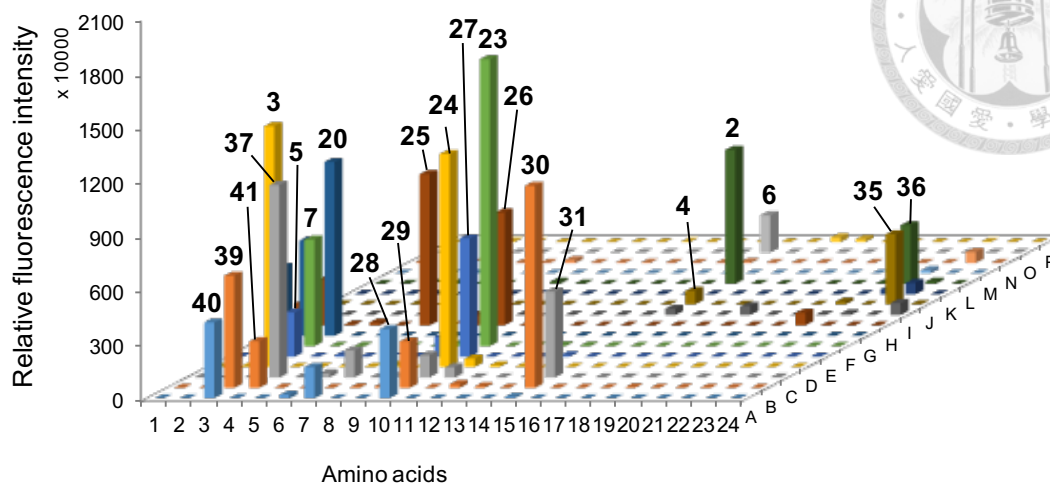


**Figure 13. Range of substrate specificities of LFRS with sfGFP-F27TAG.**

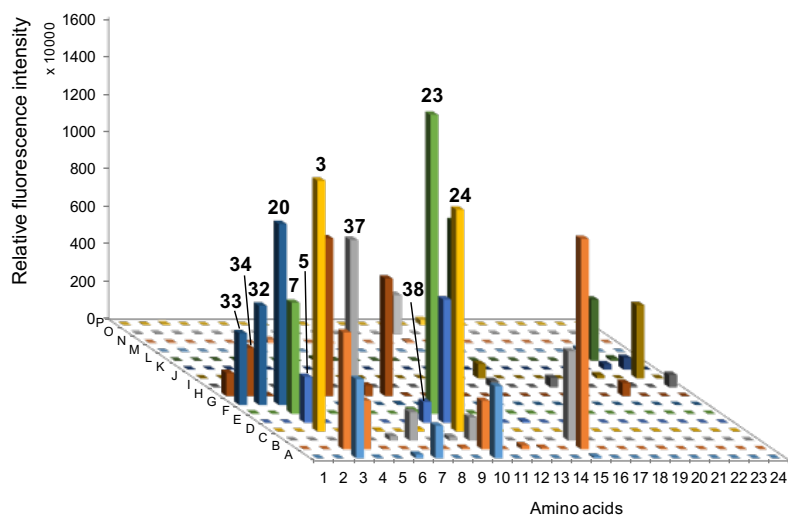
Read through of sfGFP-F27TAG with library of ncAA was measured by fluorescence intensity. Six wells (A1-2, B1-2, and C1-2) without ncAA and IPTG, the other six wells (D1-2, E1-2, and F1-2) without ncAA were controls to quantify background and thus non-production of sfGFP. The concentration of each ncAA was 1 mM with 1 mM IPTG in M9 medium at 37°C for 12 hrs. *E. coli* BL21(DE3) was used in these experiments. The library of ncAA chemical structures are given in **Figure 5**. (B) Zoom-in view. (LFRS: *MmPyIRS-N346V*)



(A)



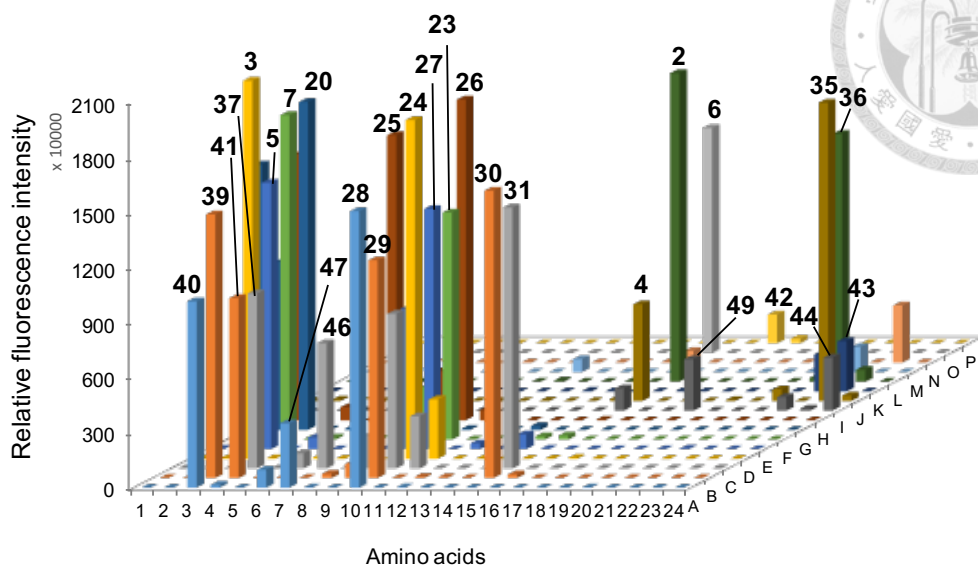
(B)



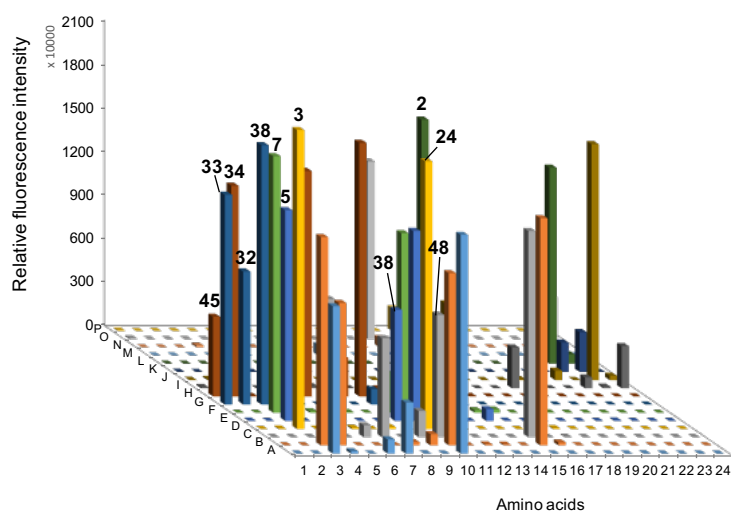
**Figure 14. Range of substrate specificities of DFERS with sfGFP-F27TAG.**

(A) Read through of sfGFP-F27TAG with library of ncAA was measured by fluorescence intensity. Six wells (A1-2, B1-2, and C1-2) without ncAA and IPTG, the other six wells (D1-2, E1-2, and F1-2) without ncAA were controls to quantify background and thus non-production of sfGFP. (B) Zoom-in view at different angle revealed the obscured area in the figure A. The concentration of each ncAA was 1 mM with 1 mM IPTG in M9 medium at 37°C for 12 hrs. *E. coli* BL21(DE3) was used in these experiments. The library of ncAA chemical structures are given in **Figure 5**. (DFERS: *Mm*PyIRS-N346G/C348Q/V401G)

(A)



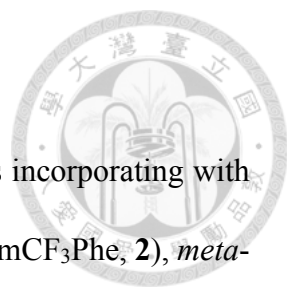
(B)



**Figure 15. Range of substrate specificities of N-DFRS with sfGFP-F27TAG.**

(A) Read through of sfGFP-F27TAG with library of ncAA was measured by fluorescence intensity. Six wells (A1-2, B1-2, and C1-2) without ncAA and IPTG, the other six wells (D1-2, E1-2, and F1-2) without ncAA were controls to quantify background and thus non-production of sfGFP. (B) Zoom-in view at different angle revealed the obscured area in the figure A. The concentration of each ncAA was 1 mM with 1 mM IPTG in M9 medium at 37°C for 12 hrs. *E. coli* BL21(DE3) was used in these experiments. The library of ncAA chemical structures are given in **Figure 5**. (N-DFRS: *Mm*PyIRS-R61K/H63Y/S193R/N346G/C348Q/V401G)

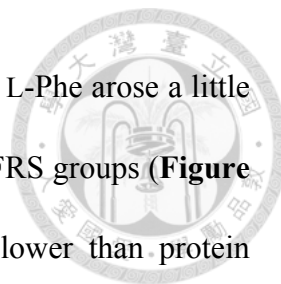
### 3.1.3 Catalytic activity of PylRS with L/DFA

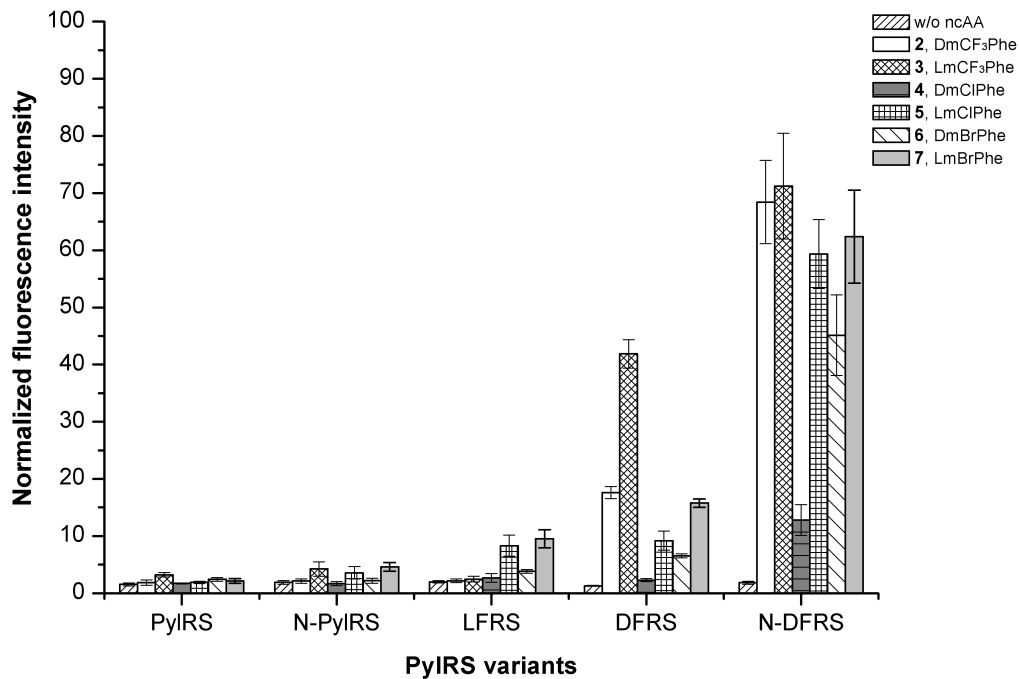


Attention was then focused on PylRS variants•tRNA<sub>CUA</sub><sup>Pyl</sup> pairs incorporating with three groups of enantiomers, *meta*-trifluoromethyl-phenylalanine (DmCF<sub>3</sub>Phe, **2**), *meta*-trifluoromethyl-L-phenylalanine (LmCF<sub>3</sub>Phe, **3**), *meta*-chloro-D-phenylalanine (DmClPhe, **4**), *meta*-chloro-L-phenylalanine (LmClPhe, **5**), *meta*-bromo-D-phenylalanine (DmBrPhe, **6**), and *meta*-bromo-L-phenylalanine (LmBrPhe, **7**) (**Figure 5**). So as to confirm the authenticity of high throughput screening, the strains harboring plasmid pCDF-PylRS variants and pET-pylT-sfGFP-F27TAG were cultured with 1 mM ncAA **2-7** or without ncAA and 1 mM IPTG, separately, and each experiment was repeated three times. The fluorescence intensities indicated the catalytic efficiency of PylRS and N-PylRS with **2-7** were unable to detection, they were similar to background (without ncAA). In LFRS group, only sfGFP-F27TAG with LFA **5** and **7** could be observed. However, after mutating more position at catalytic pocket, DFRS could recognize D/LFA, but the fluorescence signals of D-form groups were obviously low. Thus, mutations was added in NTD of DFRS, naming N-DFRS. This enzyme demonstrated high suppression efficiency not only in LFA **3**, **5**, and **7** but also in DFA **2**, **4**, and **6**, although the signal of DmClPhe, **4** was lower than others (**Figure 16**).

In order to confirm that PylRS variants•tRNA<sub>CUA</sub><sup>Pyl</sup> pairs were bioorthogonal system *in vivo* which were no interference with native translation system. The strains harboring PylRS variants•tRNA<sub>CUA</sub><sup>Pyl</sup> and sfGFP-F27TAG were grown at 37°C for 12 hrs in M9 medium and cultured in 384-well plate, with each well contain one member of 20 canonical amino acids or the enantiomer of L-aa, D-aa (1 mM). The result demonstrated PylRS (**Figure 17**), N-PylRS (**Figure 18**), and DFRS (**Figure 19**) without interference

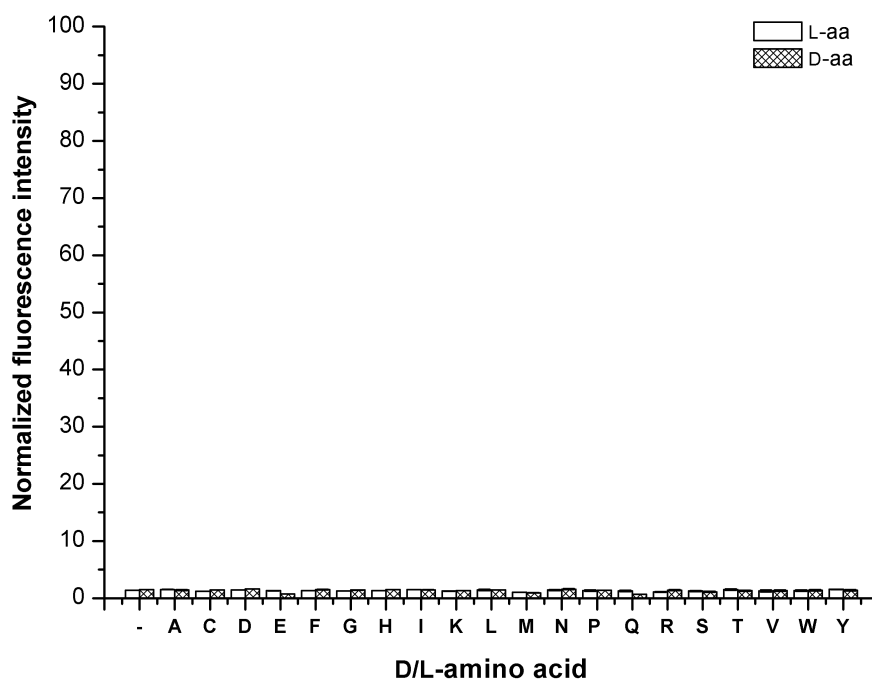
with native translation system. And the fluorescence intensities with L-Phe arose a little bit, comparing with other 38 D/L-aa, in LFRS (**Figure 20**) and N-DFRS groups (**Figure 21**). However, the signal of L-Phe incorporation was extremely lower than protein incorporated with ncAA **2-7** in N-DFRS groups.





**Figure 16. Incorporation efficiency of PylRS variants with D/LFA.**

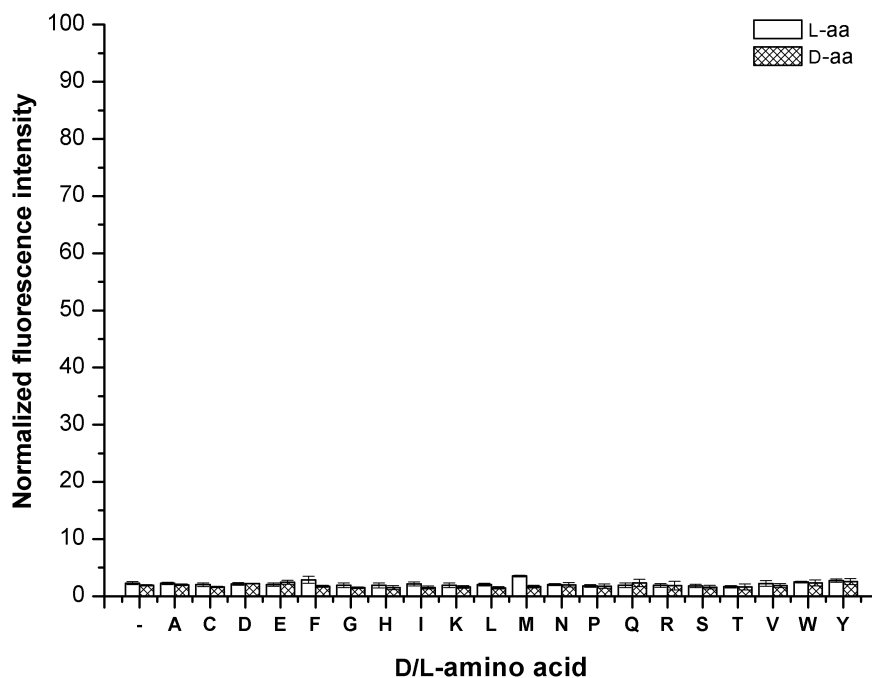
D/LFA was incorporated into sfGFP-F27TAG by the following PylRS variants : PylRS (*Methanosarcina maize* PylRS), LFRS (*MmPylRS*-N346V), N-PylRS (*MmPylRS*-R61K/H63Y/S193R), DFRS (*MmPylRS*-N346G/C348Q/V401G), and N-DFRS (*MmPylRS*-R61K/H63Y/S193R/N346G/C348Q/V401G) with tRNA<sub>CUA</sub><sup>Pyl</sup>. Protein was expressed with 1 mM amino acid and 1 mM IPTG in M9 medium at 37°C for 12 hrs. Error bars represent the standard deviation of sfGFP produced from three independent expression cultures.



**Figure 17. Incorporation efficiency of PylRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair with twenty D/L-aa.**

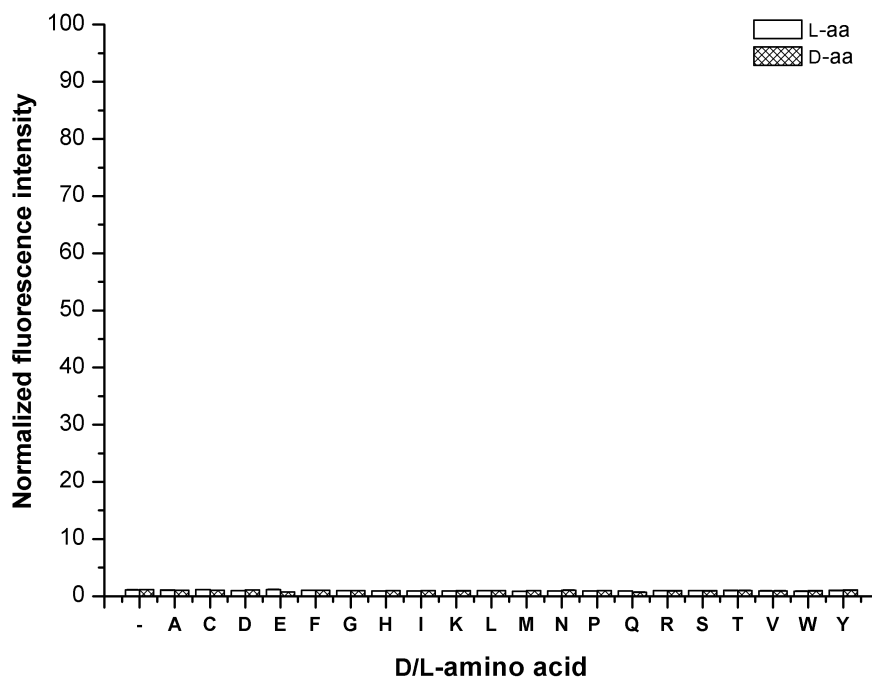
The incorporation efficiency was measured by fluorescence intensity of sfGFP protein with TAG mutation at position F27. Separately, the protein was incorporated with L-aa (white bar) and its isomer, D-aa (gray bar), but these two bars in G, Glycine without chiral center, group was the same one. First group was control (without ncAA) as the background. Protein was expressed with 1 mM amino acid and 1 mM IPTG in M9 medium at 37°C for 12 hrs. Error bars represent the standard deviation of sfGFP produced from three independent expression cultures.





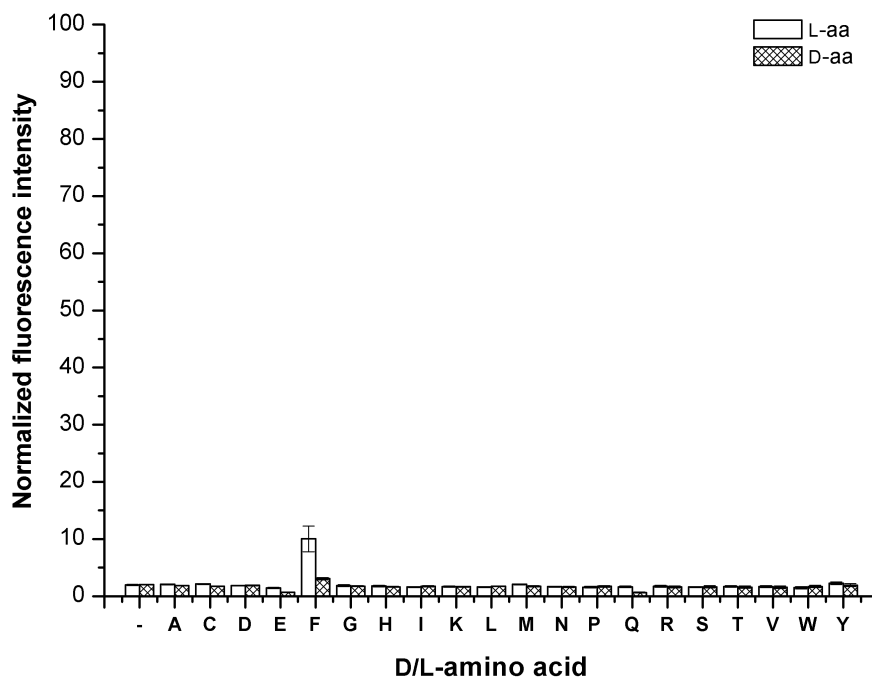
**Figure 18. Incorporation efficiency of N-PylRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair with twenty D/L-aa.**

The incorporation efficiency was measured by fluorescence intensity of sfGFP protein with TAG mutation at position F27. Separately, the protein was incorporated with L-aa (white bar) and its isomer, D-aa (gray bar), but these two bars in G, Glycine without chiral center, group was the same one. First group was control (without ncAA) as the background. Protein was expressed with 1 mM amino acid and 1 mM IPTG in M9 medium at 37°C for 12 hrs. Error bars represent the standard deviation of sfGFP produced from three independent expression cultures.



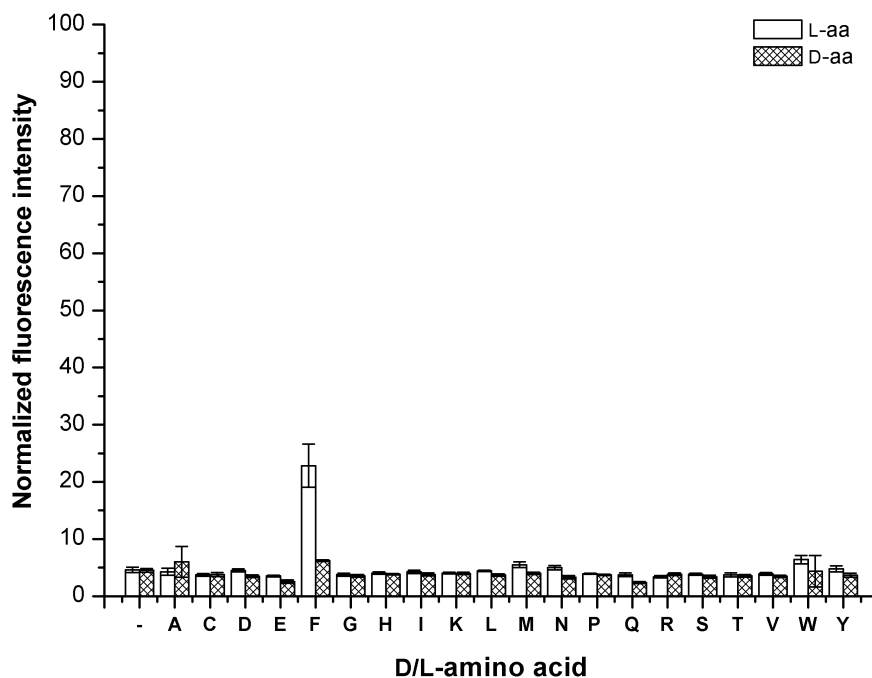
**Figure 19. Incorporation efficiency of DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair with twenty D/L-aa.**

The incorporation efficiency was measured by fluorescence intensity of sfGFP protein with TAG mutation at position F27. Separately, the protein was incorporated with L-aa (white bar) and its isomer, D-aa (gray bar), but these two bars in G, Glycine without chiral center, group was the same one. First group was control (without ncAA) as the background. Protein was expressed with 1 mM amino acid and 1 mM IPTG in M9 medium at 37°C for 12 hrs. Error bars represent the standard deviation of sfGFP produced from three independent expression cultures.



**Figure 20. Incorporation efficiency of LFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair with twenty D/L-aa.**

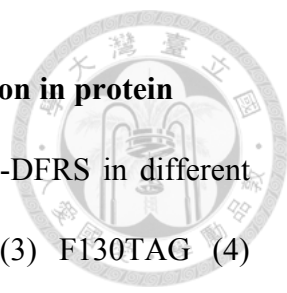
The incorporation efficiency was measured by fluorescence intensity of sfGFP protein with TAG mutation at position F27. Separately, the protein was incorporated with L-aa (white bar) and its isomer, D-aa (gray bar), but these two bars in G, Glycine without chiral center, group was the same one. First group was control (without ncAA) as the background. Protein was expressed with 1 mM amino acid and 1 mM IPTG in M9 medium at 37°C for 12 hrs. Error bars represent the standard deviation of sfGFP produced from three independent expression cultures.



**Figure 21. Incorporation efficiency of N-DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair with twenty D/L-aa.**

The incorporation efficiency was measured by fluorescence intensity of sfGFP protein with TAG mutation at position F27. Separately, the protein was incorporated with L-aa (white bar) and its isomer, D-aa (gray bar), but these two bars in G, Glycine without chiral center, group was the same one. First group was control (without ncAA) as the background. Protein was expressed with 1 mM amino acid and 1 mM IPTG in M9 medium at 37°C for 12 hrs. Error bars represent the standard deviation of sfGFP produced from three independent expression cultures.

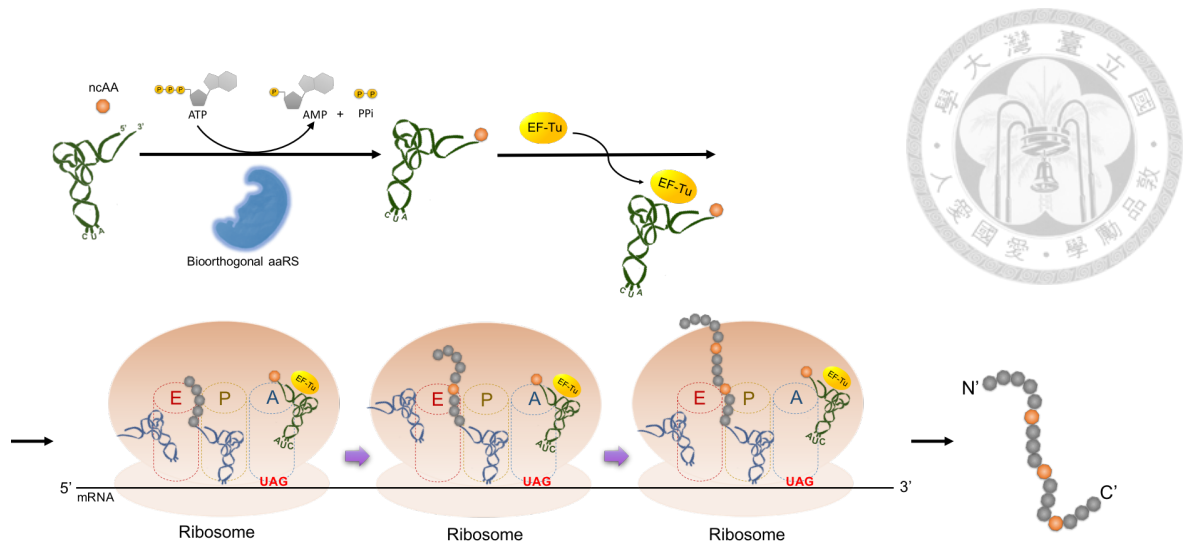
### 3.1.4 Analysis of N-DFRS catalytic activity at different position in protein



To investigate the site-specific incorporating efficiency of N-DFRS in different position, four pET-sfGFP variant: (1) S2TAG, (2) F27TAG (3) F130TAG (4) S2/F27/F130TAG (3xTAG) plasmids have been constructed and individually co-transformed with pCDF-N-DFRS plasmid into *E. coli* BL21(DE3) which grow in minimal medium and supplemented with **2-7**, separately. After over-expression for 12 hrs, the fluorescence intensities of sfGFP with amber mutation at different position were not alike (**Figure 23**). The sfGFP-F27TAG had the best incorporating efficiency with higher fluorescence intensity comparing with other group. The pattern of fluorescence intensity in each protein also exhibited some difference. The catalytic activity for *meta*-trifluoromethyl-phenylalanine (**2** and **3**) were relatively better than phenylalanine analogues containing other substitutions in sfGFP-S2TAG group. But, in sfGFP-F130TAG group, the catalytic activity for *meta*-bromo-phenylalanine (**7**) was the highest. And there was similar incorporating efficiency of all ncAAs (**2-7**) in sfGFP-F27TAG group, except DmClPhe **4** which was relatively low usage in four groups. In addition, the suppression efficiency of sfGFP with three amber codons was tested (**Figure 22**). Owing to the competition of release factor I with ncAAs, the protein production was low. But they can be detected on fluorescence and Western blot analysis.

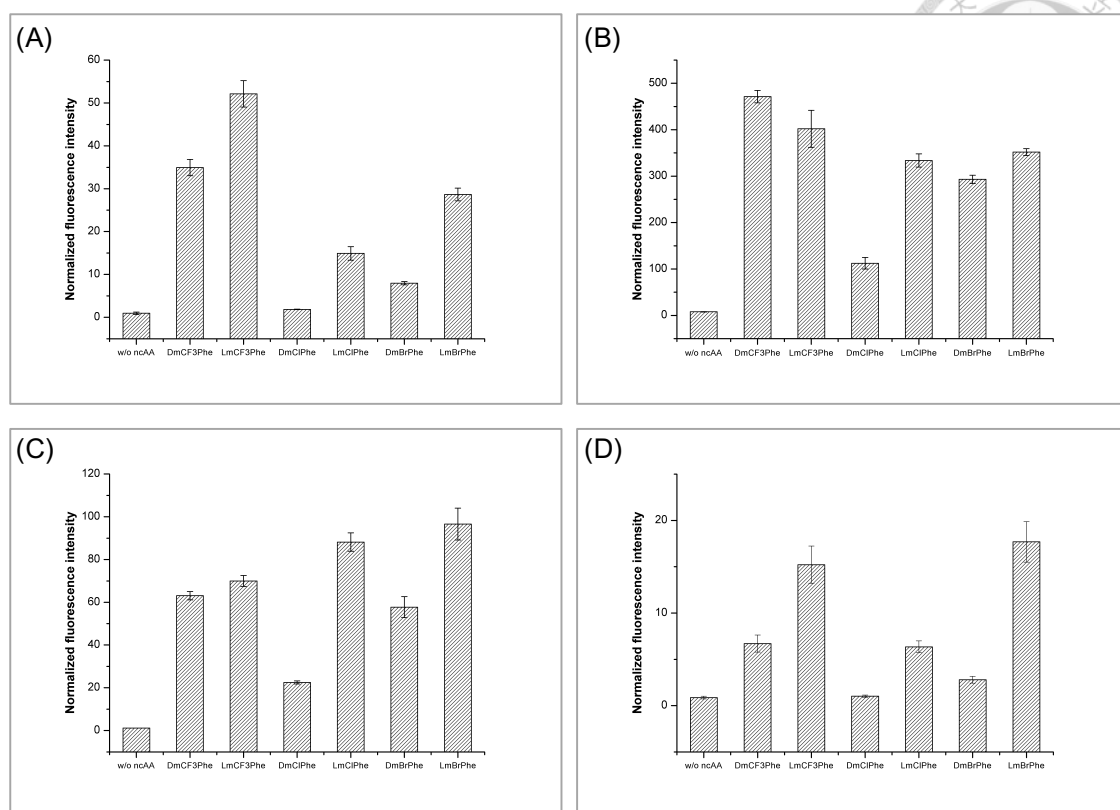
In order to ensure the suppressed yield of sfGFP-1xTAG or 3xTAG with **2-7**. The whole cell was diluted to the same OD<sub>600</sub> value and analyzed by SDS-PAGE. Subsequently, the SDS-PAGE was used for western blot analysis with an anti-hexahistidine antibody (**Figure 24, S1-4**). The sfGFP with amber mutation contained hexahistidine in C' terminus which was used to separate the full-length protein with

truncated form by Ni<sup>2+</sup>-NTA column. In western blot analysis, the antibody recognized hexahistidine in C' terminus of full-length sfGFP. The result further verified production yield of sfGFP-F27TAG with ncAAs were higher than other sfGFP variants. Interestingly, protein incorporated with **4** had the lowest production in all sfGFP variants, but with its enantiomer (**5**) had high yield. Comparing western blotting with fluorescence intensity of sfGFP-3xTAG, protein with **2** and **3** revealed the similar production yield, however, the fluorescence intensity of sfGFP with DFA (**2**) was lower than LFA (**3**) (**Figure 23**).



**Figure 22. Protein incorporated with ncAA at multiple sites.**

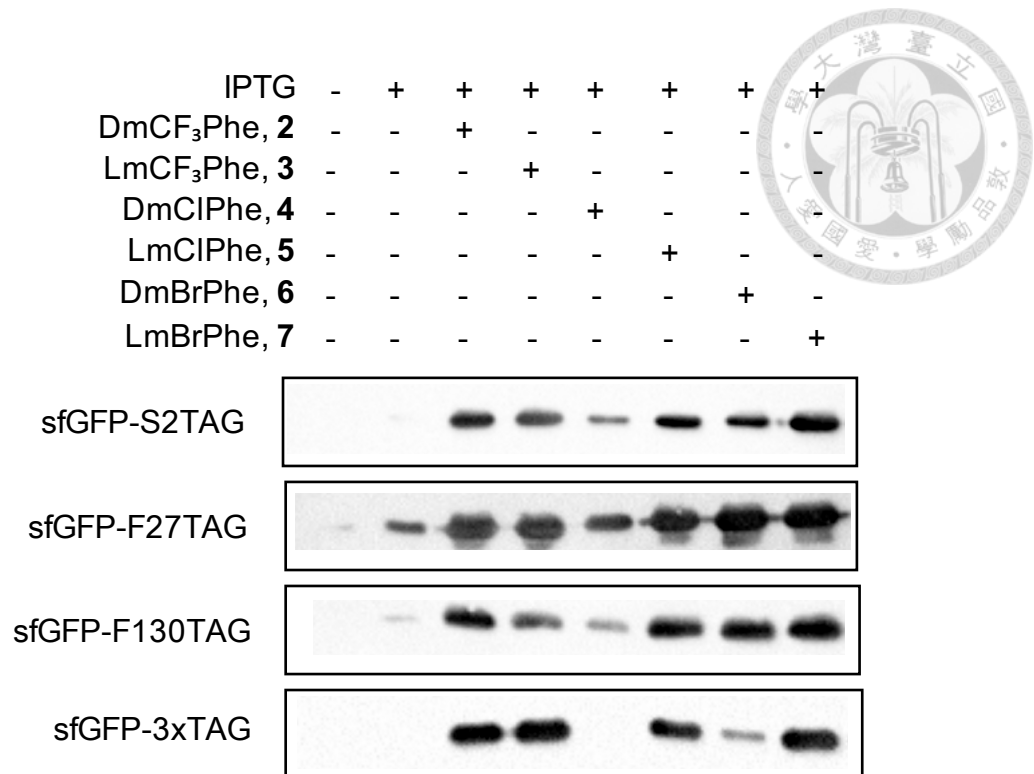
The DNA sequence of target protein was mutant with amber codon (TAG) in desired sites. After transcription, the mRNA contained multiple UAG codons which was used to encode ncAA. Subsequently, protein was produced and site-specifically incorporated with multiple ncAAs in the BTP.



**Figure 23. Site-specific incorporation of D/LFA by N-DFRS•tRNA<sup>Pyl</sup><sub>CUA</sub> pair.**

The incorporation efficiency was measured by fluorescence intensity of sfGFP protein with a TAG mutation at different position: (A) S2, (B) F27, (C) F130 or (D) with three TAG mutations at S2, F27, and F130. First bar was control (without ncAA) to quantify background. Protein was expressed with 1 mM ncAA and 1 mM IPTG in M9 medium at 37°C for 12 hrs. Error bars represent the standard deviation of sfGFP produced from three independent expression cultures. (DmCF<sub>3</sub>Phe, **2**; LmCF<sub>3</sub>Phe, **3**; DmClPhe, **4**; LmClPhe, **5**; DmBrPhe, **6**; LmBrPhe, **7**)

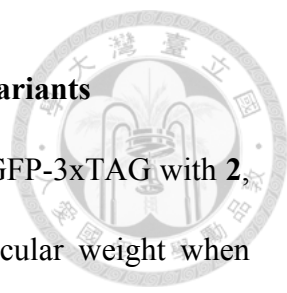




**Figure 24. Analyzed amber suppression of D/LFA in sfGFP variants by WB.**

*E. coli* BL21(DE3) cells coding DFRS• tRNA<sup>Pyl</sup><sub>CUA</sub> pair and a sfGFP gene with amber codon at position S2, F27, F130 or three of them which contained hexahistidine in C' terminus. The protein expression was controlled with an inducible T7 promoter and the cell was grown in M9 medium supplemented with 1 mM IPTG and 1 mM ncAAs. Amber suppression of sfGFP-1or 3xTAG-hexahistidine were analyzed by WB with anti-hexahistidine. And confirming that DFRS•tRNA<sup>Pyl</sup><sub>CUA</sub> pair incorporated **2-7** into protein in *E. coli* Cell.

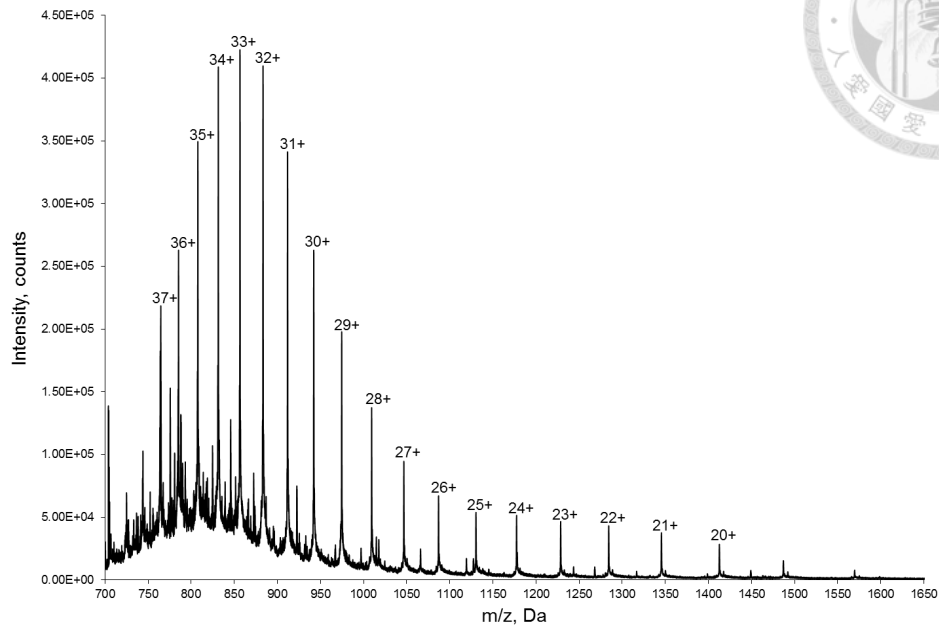
### 3.1.5 ESI-MS and MALDI-TOF-MS/MS analysis of sfGFP variants



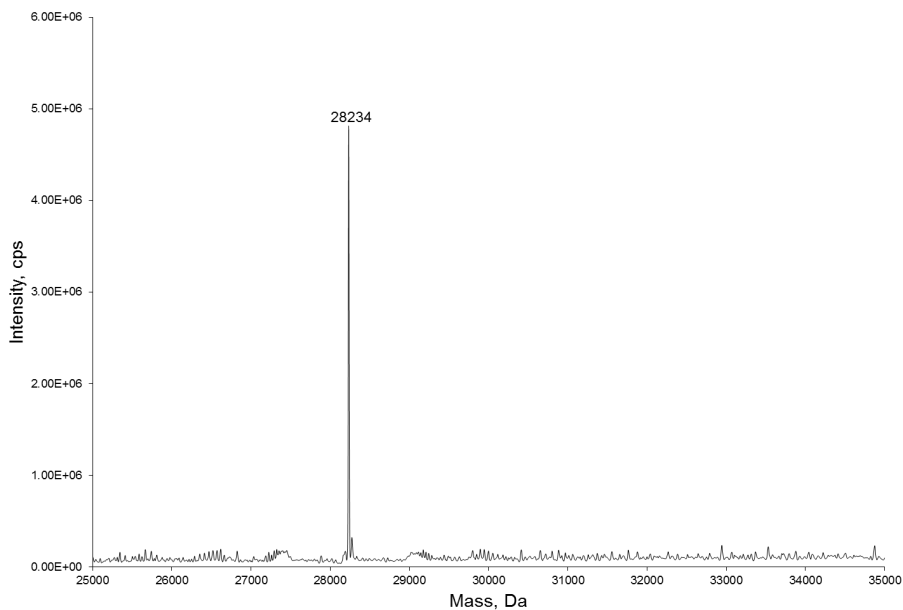
Next, the ESI-MS was performed to confirm the results of sfGFP-3xTAG with **2**, **3**, **6**, and **7**. The purified proteins all showed the expected molecular weight when analyzed by ESI-MS (**Figure 25-28**). In order to make sure that N-DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair site-specifically incorporated ncAA into protein at position S2, F27, and F130. The purified proteins were digested by trypsin or chymotrypsin and analyzed by MALDI-TOF-MS/MS. As our expectation, the fragments containing **2** and **3** at S2, F27, and F130 have been searched (**Figure 29-34**), but the fragment of sfGFP-3xTAG with **6** and **7** was only found at F27 and F130 position (**Figure 35-38**). Therefore, other enzymatic digested condition will be tested to find out the fragment containing **6** and **7** at S2.



(A)



(B)

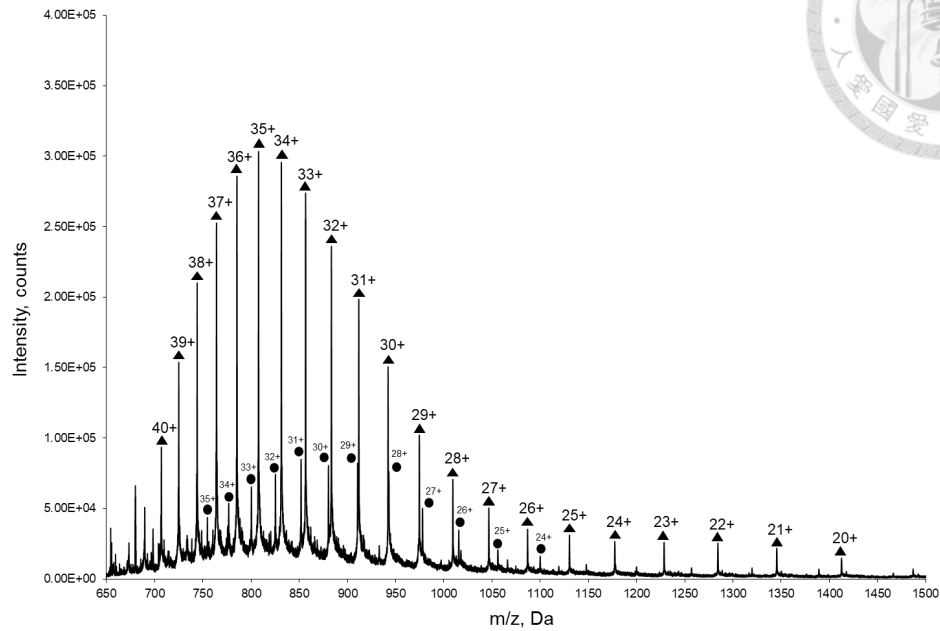


**Figure 25. Molecular mass determination of the protein sfGFP-3xTAG-2.**

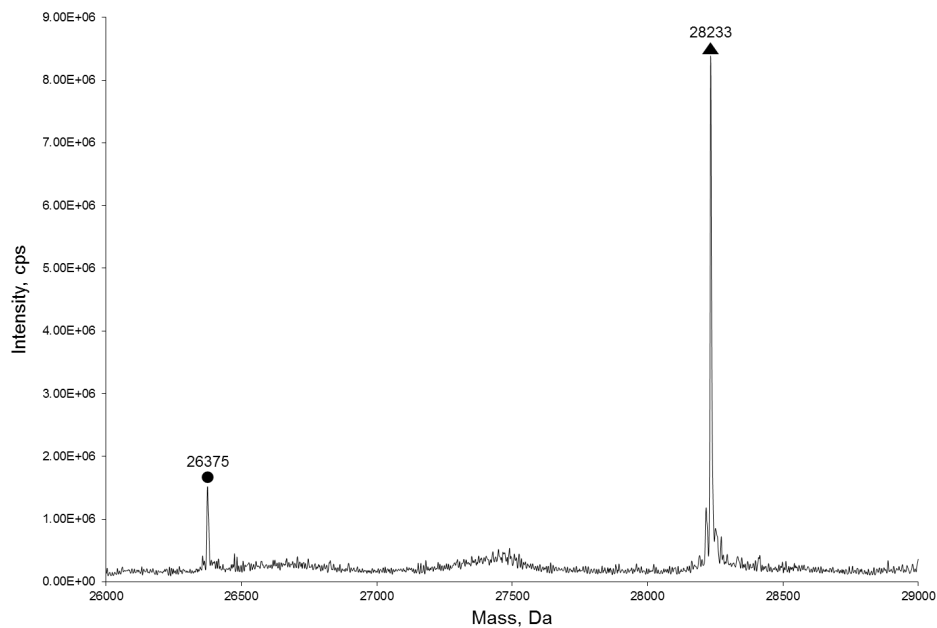
Full-length sfGFP-3xTAG-2 was produced by using PyIRS• tRNA<sup>PyI</sup><sub>CUA</sub> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum of sfGFP-3xTAG-2 (B) the deconvoluted ESI-MS spectrum of sfGFP-3xTAG-2. The calculated molecular mass is 28,234 Da; observed molecular mass is 28,234 Da.



(A)



(B)

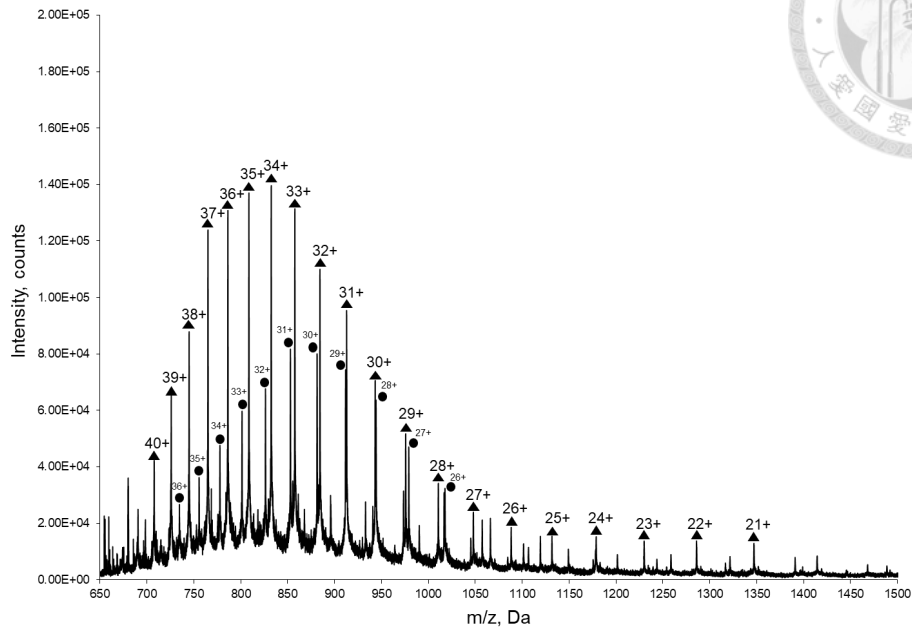


**Figure 26. Molecular mass determination of the protein sfGFP-3xTAG-3.**

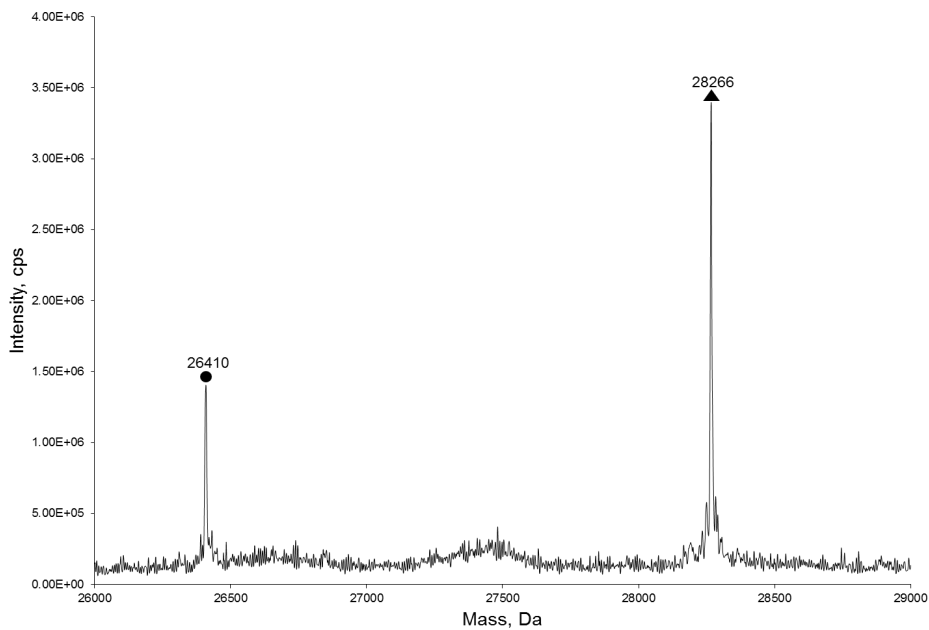
Full-length sfGFP-3xTAG-3 was produced by using PylRS• tRNA<sub>CUA</sub><sup>Pyl</sup> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 28,234 Da; observed molecular masses are 28,233 Da and 26,375 Da.



(A)



(B)

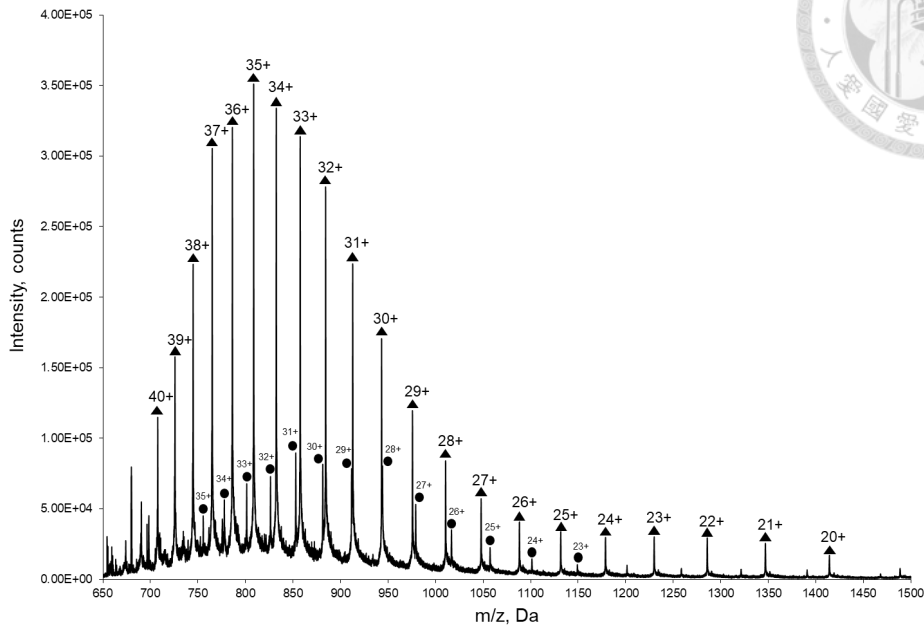


**Figure 27. Molecular mass determination of the protein sfGFP-3xTAG-6.**

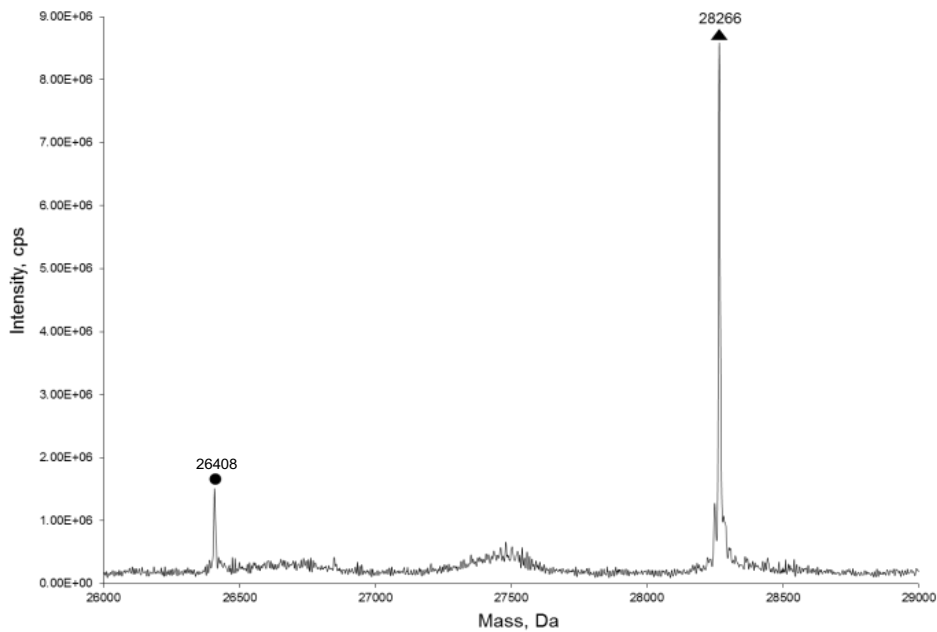
Full-length sfGFP-3xTAG-6 was produced by using PylRS• tRNA<sub>CUA</sub><sup>Pyl</sup> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 28,266 Da; observed molecular masses are 28,266 Da and 26,410 Da.



(A)

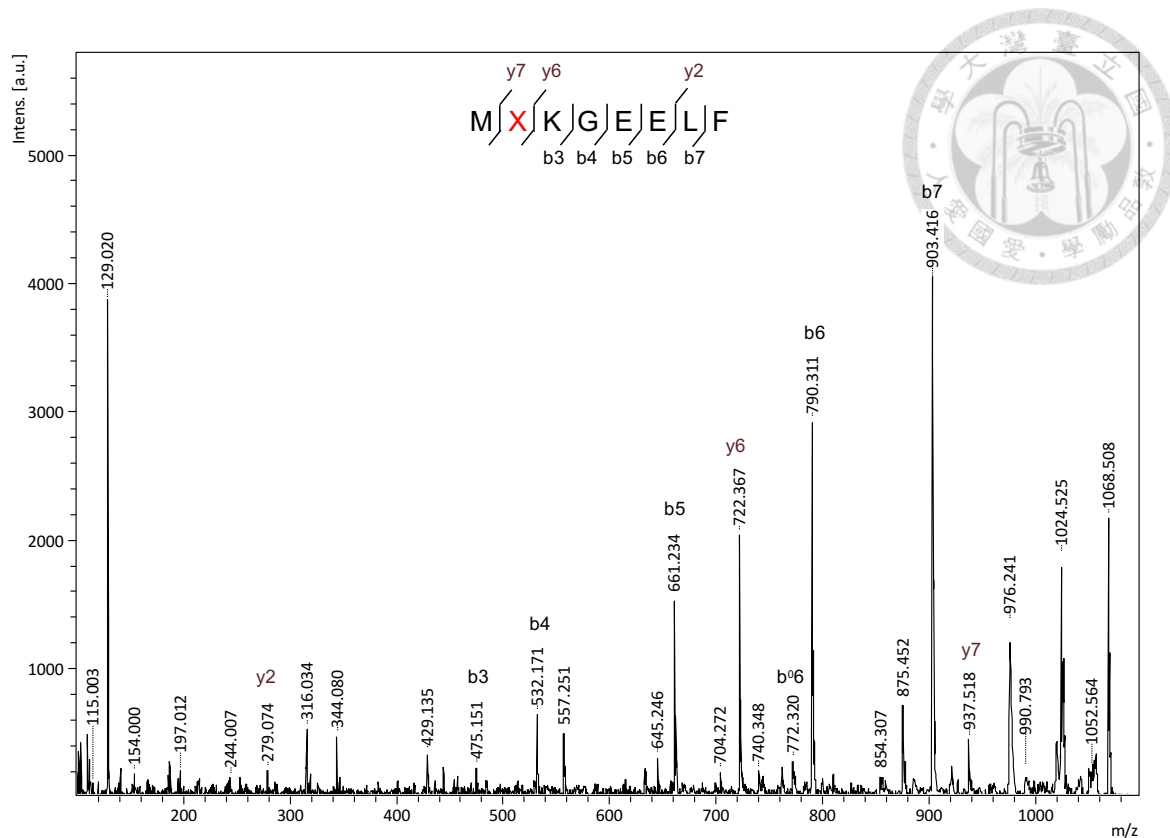


(B)



**Figure 28. Molecular mass determination of the protein sfGFP-3xTAG-7.**

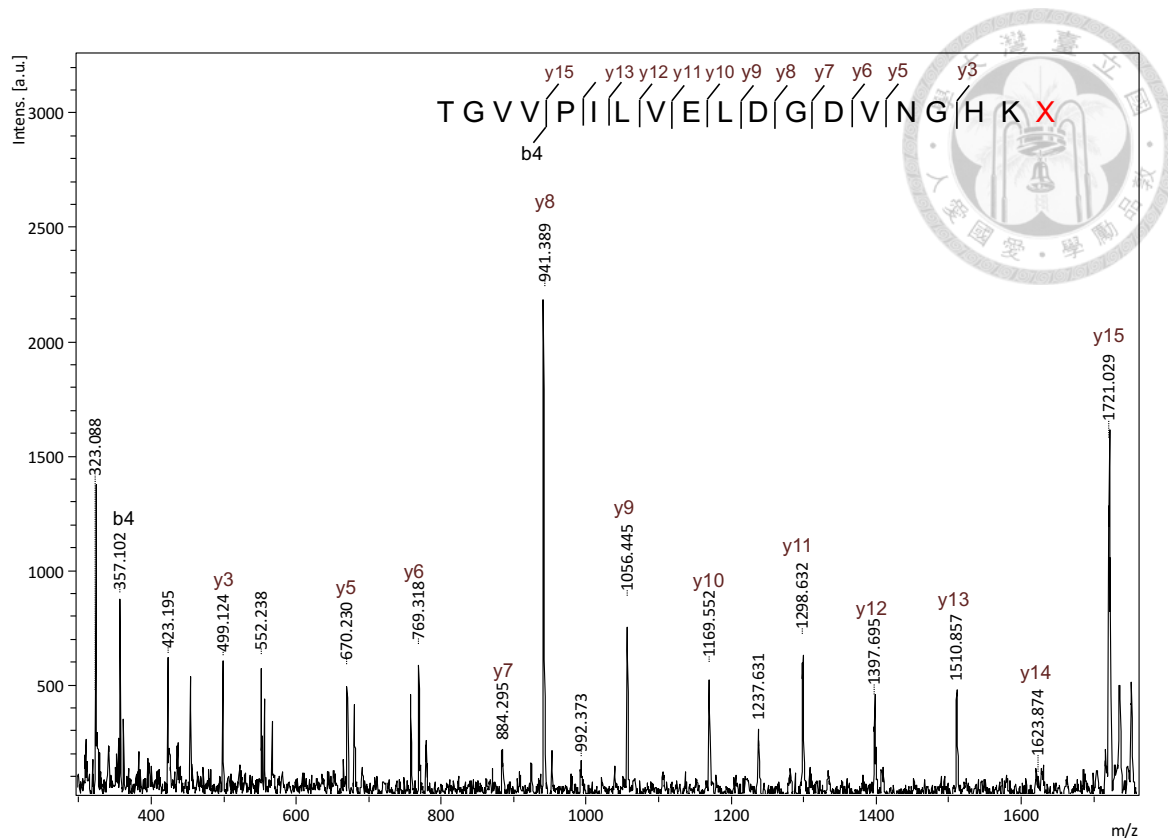
Full-length sfGFP-3xTAG-7 was produced by using PylRS• tRNA<sup>Pyl</sup><sub>CUA</sub> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 28,266 Da; observed molecular masses are 28,266 Da and 26,408 Da.



**Figure 29. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-2 at position S2.**

Fragment M<sup>1</sup>XKGEELF<sup>8</sup>, X represents **2**, was searched in tandem mass spectrometry.

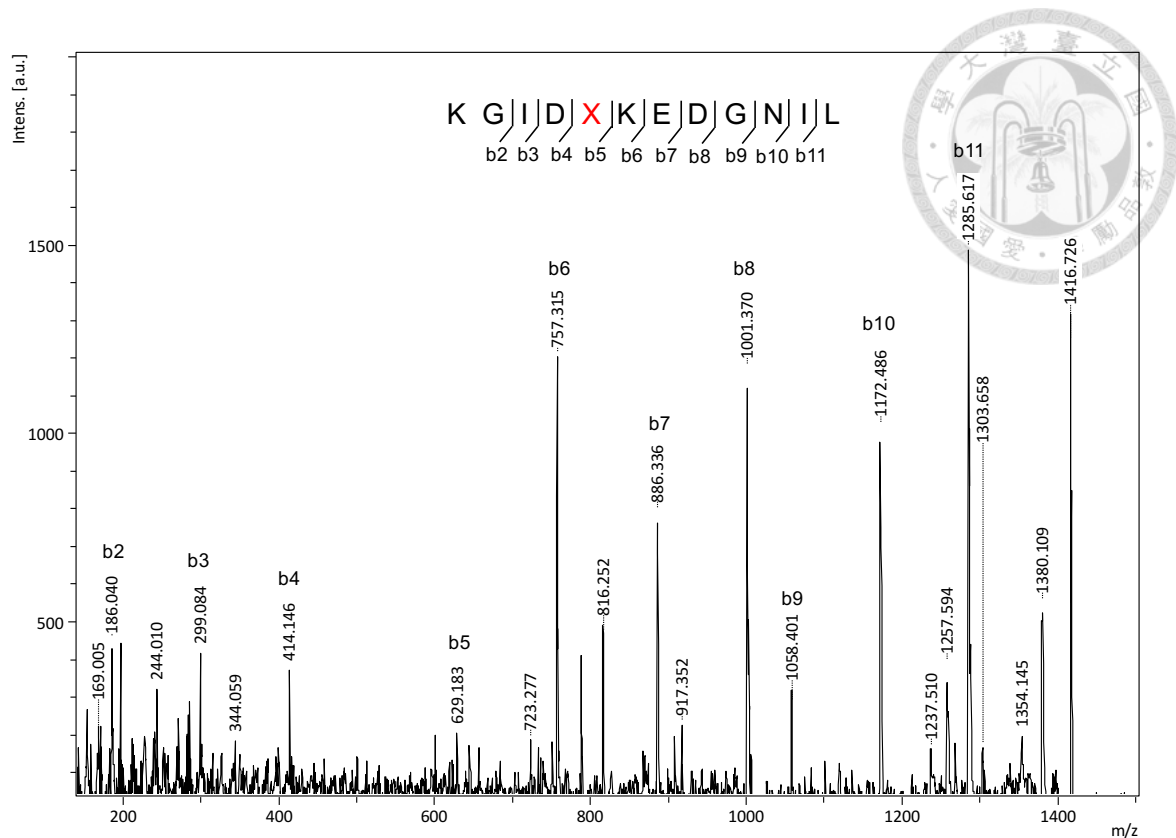
The full-length sfGFP-3xTAG-**2** protein was produced by the N-DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair in the presence of 1 mM **2**. The protein was in-gel digested by chymotrypsin.



**Figure 30. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-2 at position F27.**

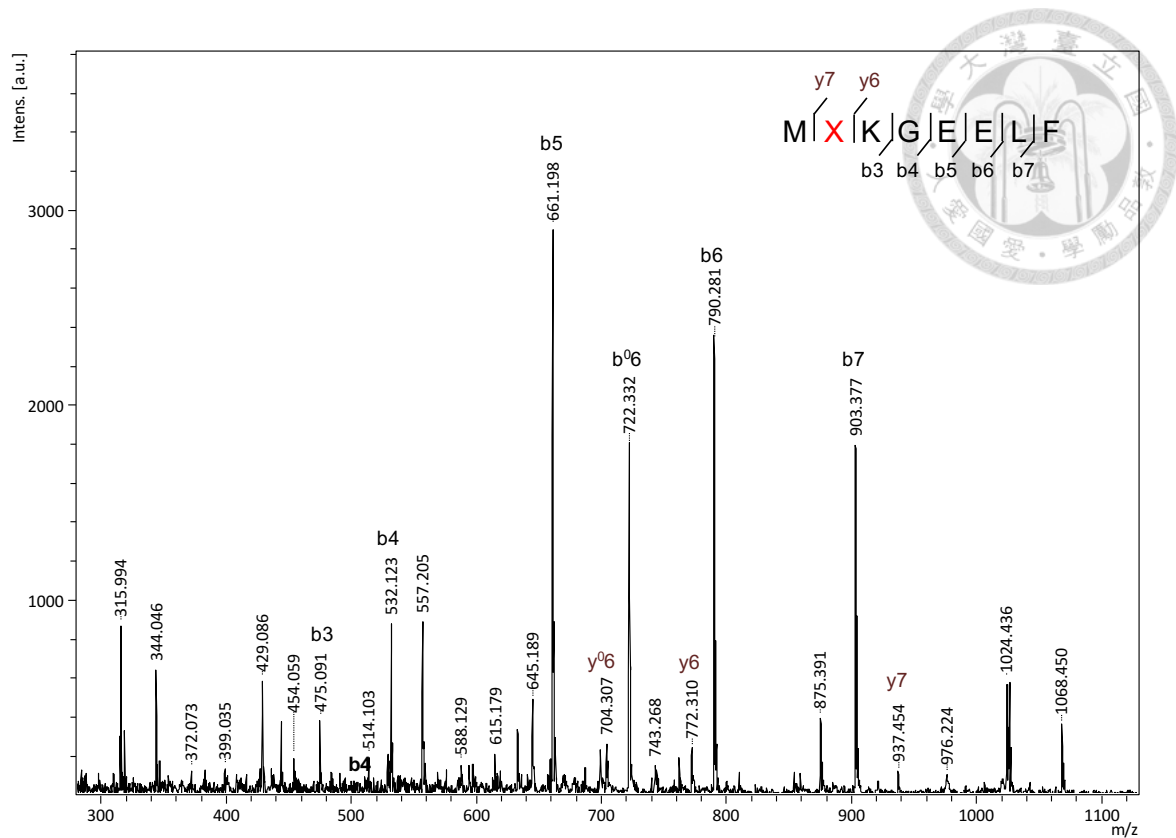
Fragment T<sup>9</sup>GVVPI<sup>1</sup>LV<sup>2</sup>EL<sup>3</sup>DL<sup>4</sup>GD<sup>5</sup>V<sup>6</sup>NG<sup>7</sup>HK<sup>8</sup>X<sup>27</sup>, X represents **2**, was searched in tandem mass spectrometry. The full-length sfGFP-3xTAG-2 protein was produced by the N-DFRS•tRNA<sub>CUA</sub><sup>Py1</sup> pair in the presence of 1 mM **2**. The protein was in-gel digested by chymotrypsin.





**Figure 31. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-2 at position F130.**

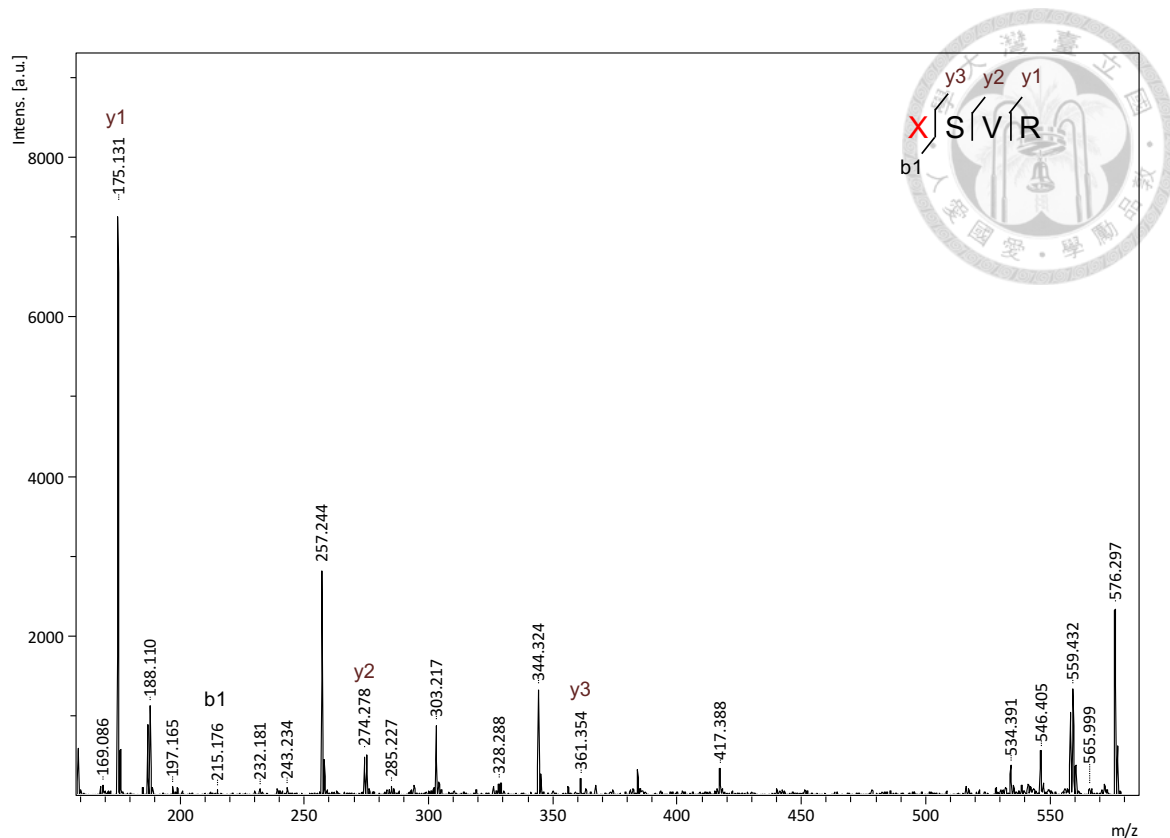
Fragment  $K^{126}GIDXKEDGNIL^{137}$ , X represents **2**, was searched in tandem mass spectrometry. The full-length sfGFP-3xTAG-2 protein was produced by the N-DFRS•tRNA<sub>CUA</sub><sup>Py1</sup> pair in the presence of 1 mM **2**. The protein was in-gel digested by chymotrypsin.



**Figure 32. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-3 at position S2.**

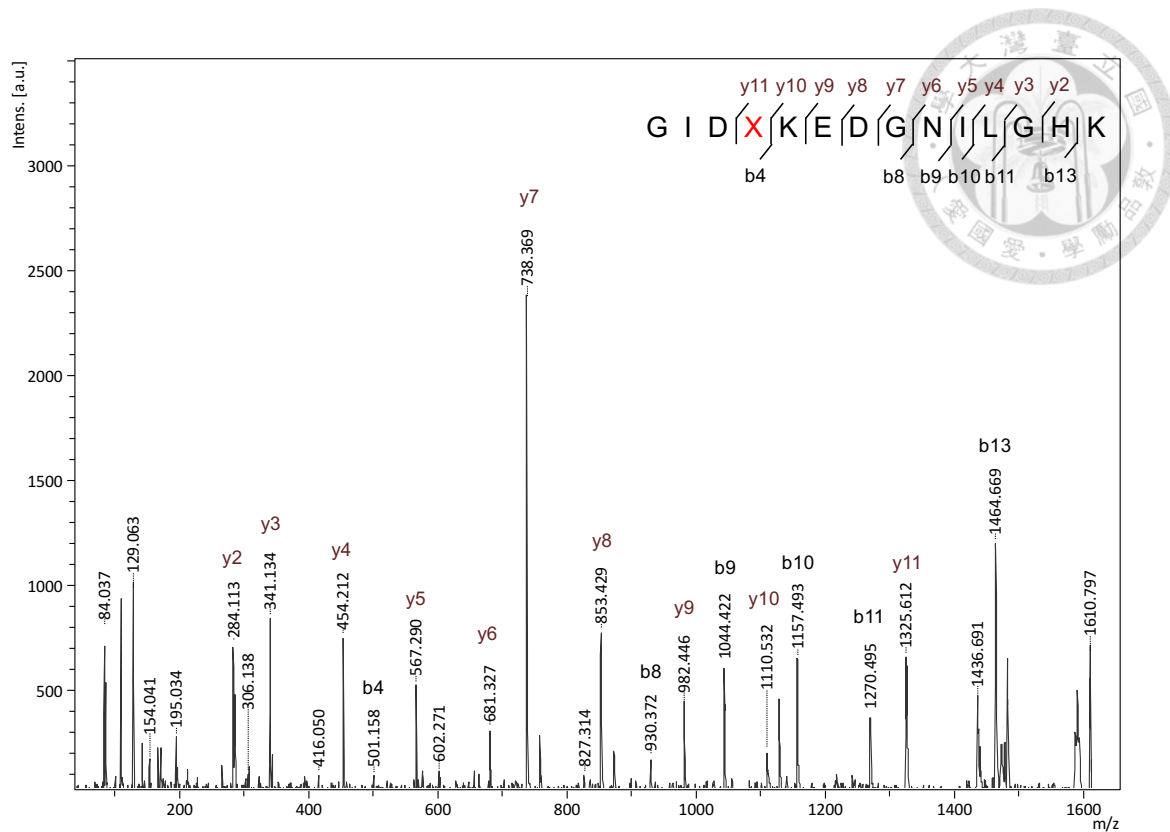
Fragment M<sup>1</sup>XKGEELF<sup>8</sup>, X represents **3**, was searched in tandem mass spectrometry.

The full-length sfGFP-3xTAG-**3** protein was produced by the N-DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair in the presence of 1 mM **3**. The protein was in-gel digested by chymotrypsin.



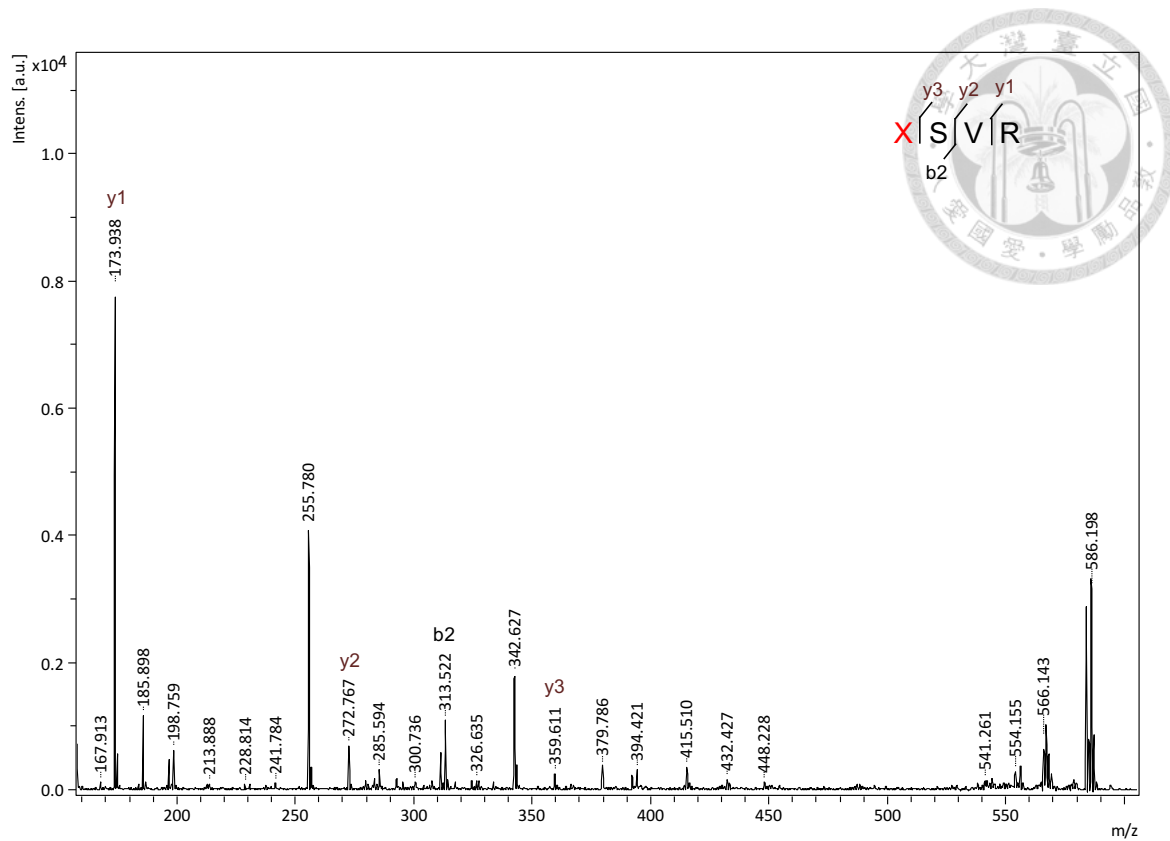
**Figure 33. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-3 at position F27.**

Fragment  $X^{27}SVR^{30}$ , X represents **3**, was searched in tandem mass spectrometry. The full-length sfGFP-3xTAG-3 protein was produced by the N-DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair in the presence of 1 mM **3**. The protein was in-gel digested by trypsin.



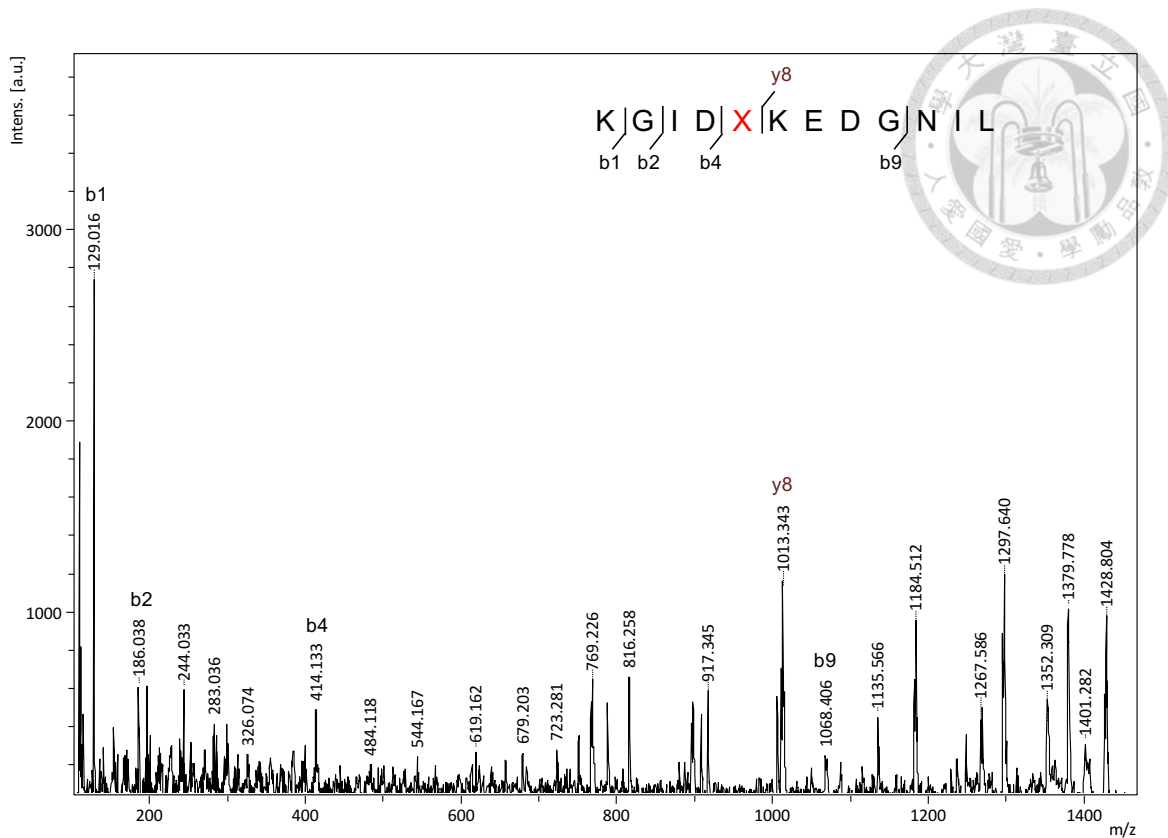
**Figure 34. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-3 at position F130.**

Fragment  $G^{127}IDXKEDGNILGHK^{140}$ , X represents **3**, was searched in tandem mass spectrometry. The full-length sfGFP-3xTAG-**3** protein was produced by the N-DFRS•tRNA<sub>CUA</sub><sup>PyI</sup> pair in the presence of 1 mM **3**. The protein was in-gel digested by trypsin.



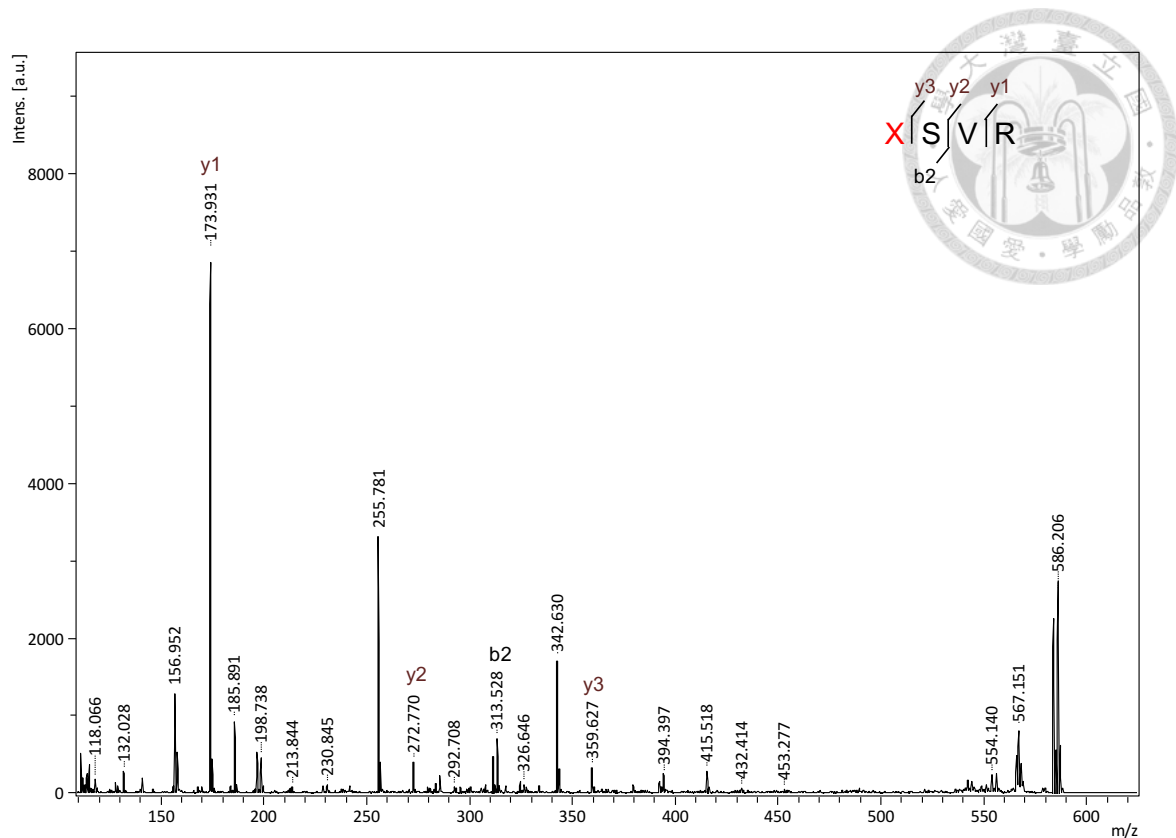
**Figure 35. MALDI-TOF-MS/MS analysis on of sfGFP-3xTAG-6 at position F27.**

Fragment  $X^{27}SVR^{30}$ , X represents **6**, was searched in tandem mass spectrometry. The full-length sfGFP-3xTAG-6 protein was produced by the N-DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair in the presence of 1 mM **6**. The protein was in-gel digested by trypsin.



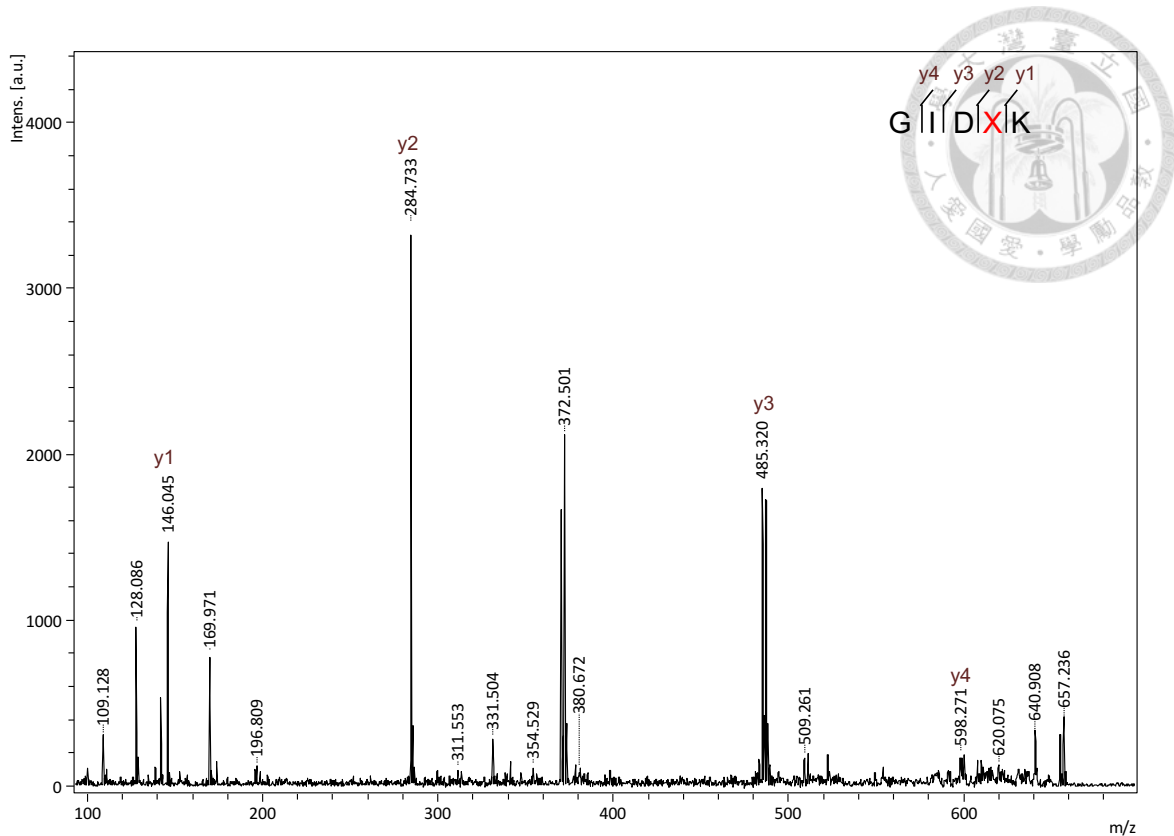
**Figure 36. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-6 at position F130.**

Fragment  $K^{126}GIDXKEDGNIL^{137}$ , X represents **6**, was searched in tandem mass spectrometry. The full-length sfGFP-3xTAG-**6** protein was produced by the N-DFRS•tRNA<sup>PyI</sup><sub>CUA</sub> pair in the presence of 1 mM **6**. The protein was in-gel digested by chymotrypsin.



**Figure 37. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-7 at position F27.**

Fragment  $X^{27}SVR^{30}$ , X represents 7, was searched in tandem mass spectrometry. The full-length sfGFP-3xTAG-7 protein was produced by the N-DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair in the presence of 1 mM 7. The protein was in-gel digested by trypsin.



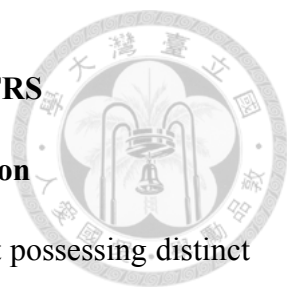
**Figure 38. MALDI-TOF-MS/MS analysis of sfGFP-3xTAG-7 at position F130.**

fragment  $G^{127}IDXK^{131}$ , X represents 7, was searched in tandem mass spectrometry. The full-length sfGFP-3xTAG-7 protein was produced by the N-DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair in the presence of 1 mM 7. The protein was in-gel digested by trypsin.

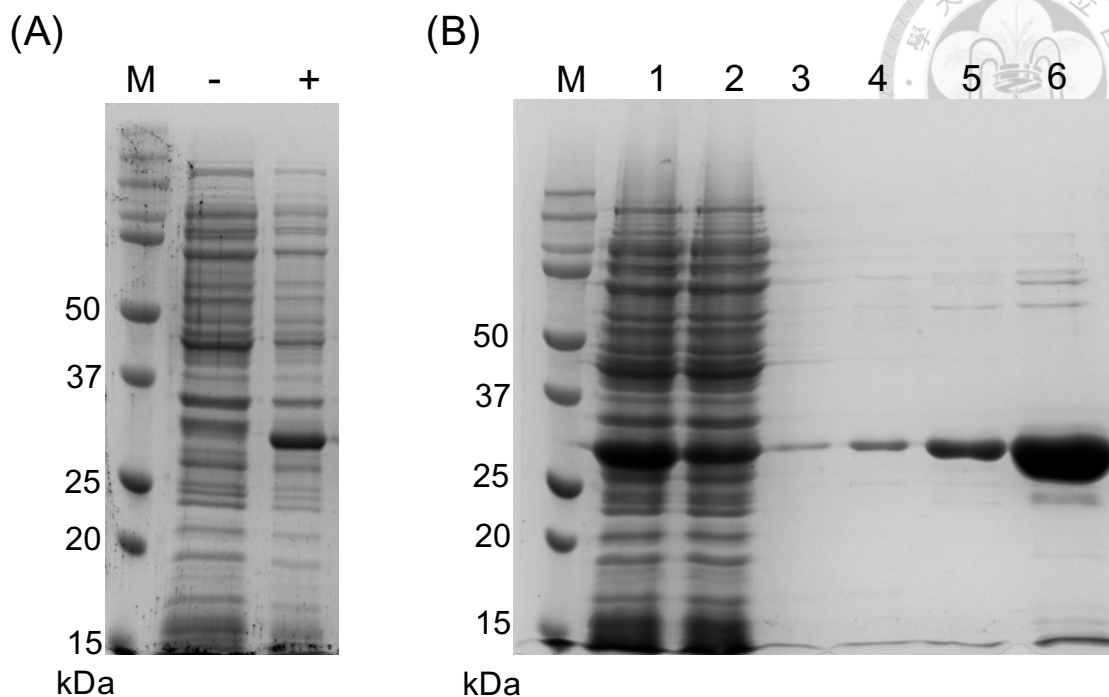


## 3.2 Crystallization revealing L/DFA binding mode in N-DFRS

### 3.2.1 Purification of N-DFRS binding pocket for crystallization

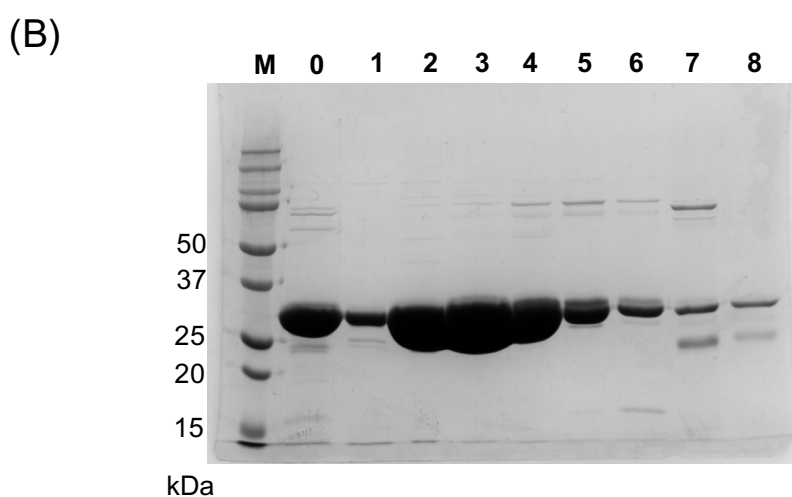
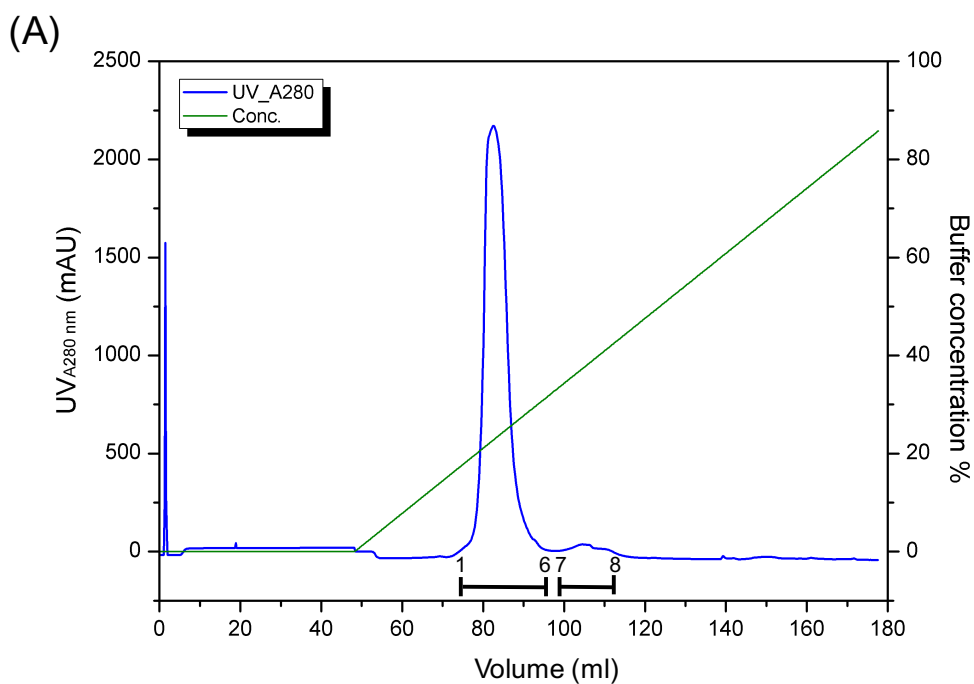


The enantiomers (**4** and **5**) are mirror images of each other but possessing distinct catalytic activity with N-DFRS. Getting to the meat of this issue, we focused on investigating binding mode of ncAA in catalytic pocket to realize the catalytic process of N-DFRS with **2-7**. Thus, pET-pyIT plasmid with the pocket site of N-DFRS (from residue 185 to 454; totally 270 amino acids) containing hexahistidine in N-terminus was constructed. Subsequently, the plasmid was transformed into *E. coli* BL21(DE3) and over-expressed with 0.5 mM IPTG. 7 L cultures were collected and purified with three different column chromatographs, the condition of purification was optimized from the published paper which focused on wt-PylRS.<sup>43</sup> Firstly, the cell lysates were purified with Ni<sup>2+</sup>-NTA open column and there were also some impurities in elution with 200 mM imidazole (**Figure 39**). Secondly, N-DFRSc270 was transferred to 25 mM potassium phosphate buffer at pH 7.4 by SEC (HiLoad 16/600 superdex 75 pg, GE Healthare). Finally, the sample was applied on Mono Q column in FPLC (**Figure 40**). The fractions containing N-DFRSc270 were pooled and changed buffer to 20 mM potassium phosphate buffer containing 300 mM NaCl, 5 mM MgCl<sub>2</sub>, and 10 mM β-Me at pH7.4 by SEC (HiLoad 16/600 superdex 75 pg, GE Healthare). Totally, 45 mg protein was obtained for crystallization.



**Figure 39. Over-expression and purification of N-DFRSc270.**


(A) SDS-PAGE shows over-expression of N-DFRSc270. BL21(DE3) cells was transformed with pET22b (+) plasmid encoding N-DFRSc270 and grown in LB with 0.5 mM IPTG at 25°C over-night. “-” before induction, “+” after induction with IPTG and “M” protein marker. Red arrow pointed target protein which’s size was 33 kDa. (B) N-DFRSc270 proteins were purified with Ni<sup>2+</sup>-NTA column and analyzed by SDS-PAGE. 1: supernatant of cell lysate; 2: flow-through; 3: washing with 25 mM imidazole; 4: washing with 40 mM imidazole; 5: washing with 50 mM imidazole, and 6: elution.



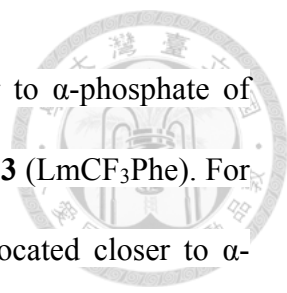
**Figure 40. Purification of N-DFRSc-270 with Mono Q column.**

Mono Q is the strong anion exchange chromatography column. (A) The elution after Ni<sup>2+</sup>-NTA column was concentrated for FPLC with Mono Q column. The major peak appeared in buffer with 160 mM to 200 mM NaCl. Blue line: the signal of UV<sub>A280</sub>; Green line: the concentration (%) of buffer with 0.8 M NaCl. (B) SDS-PAGE shows the fractions indicated in (A). 0: elution after Ni<sup>2+</sup>-NTA column; 1-6: major peak; 7-8: minor peak. Fraction 1-3 (in red box) were collected and concentrated for crystallization.

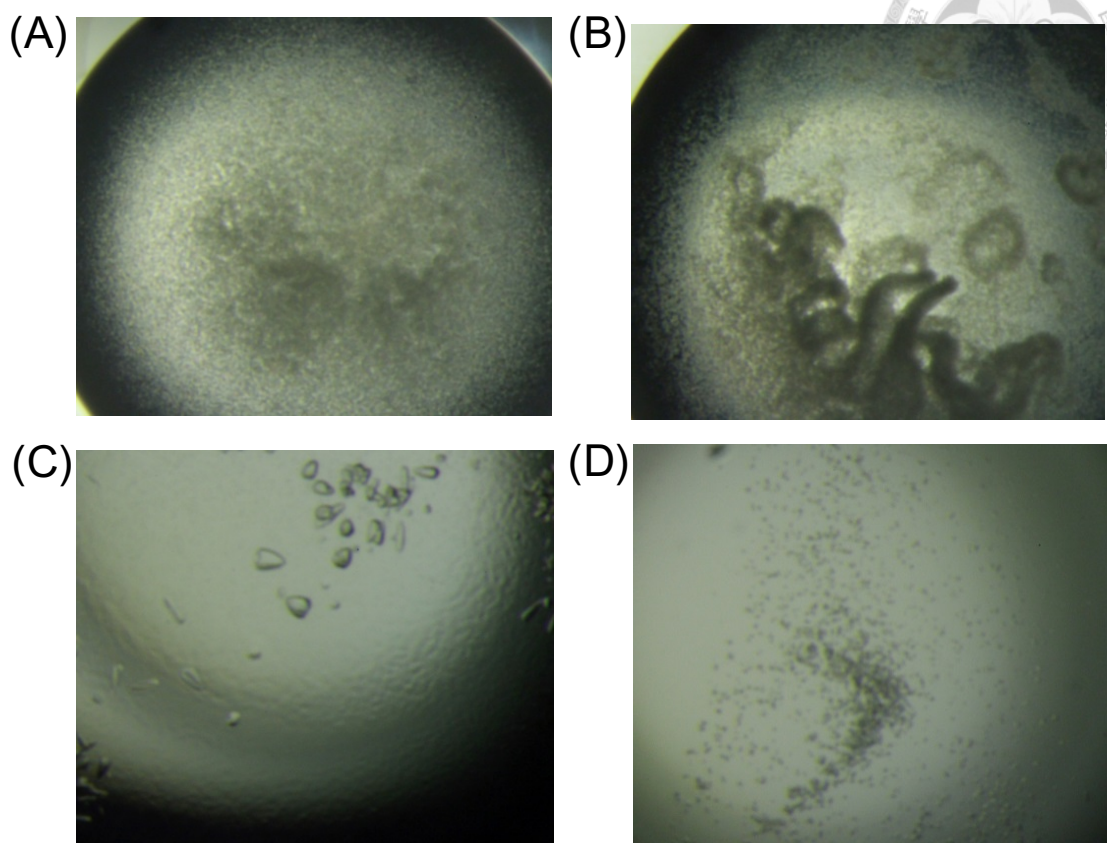
### 3.2.2 D/LFA soaking with N-DFRS binding pocketing



The condition of crystallization was screened by RIGAKU Phoenix/RE crystallization robot (**Figure 41**). then the successful condition was repeat by hanging drop vapor-diffusion method (**Figure 42**). N-DFRSc270 was soaked with AMPPNP, ATP analogs, and 5mM six ncAA (**2-7**) separately in 100 mM HEPES buffer at pH 7.5, 10% PEG 8000, 10% Ethylene Glycerol. To gain the structural insight of different D/LFA in the binding pocket of N-DFRS, X-ray crystallography was used to study the structure of N-DFRSc270 in complex with AMPPNP and different ncAAs including DmCF<sub>3</sub>Phe **2**; LmCF<sub>3</sub>Phe **3**; DmClPhe **4**; LmClPhe **5**; DmBrPhe **6**; and LmBrPhe **7**. N-DFRSc270 crystals contained one molecular in an asymmetry unit. The overall structure was similar to the previously reported wt-PylRSc270 crystal structure<sup>57-59</sup>. The mutation, N346G, C348Q, and V401G, made the active site larger than the wild type. All residues of N-DFRSc270 could be built and fitted well to the electron density except for loop  $\beta 7$ – $\beta 8$ . No discernable electron density could be identified for loop  $\beta 7$ – $\beta 8$ . In particular, the clear electron density for building the entire AMPPNP, two octahedrally coordinated Mg<sup>2+</sup> ions and the amino acid in the building pocket of the three LFA (**3**, **5**, **7**) complex structures (**Figure 43**) were observed. Both AMPPNP and the LFA were located at the conserved active site. The main chains of the three LFAs superposed very well with each other. However, the CF<sub>3</sub> substituents points at a different direction than the other Cl and Br substituents. In contrast, for three DFA complexes, even though the election density of AMPPNP and Mg<sup>2+</sup> were as clear as in the LFA complexes, only partial electron density could be identified (**Figure 44**). The sidechain of the DFAs can be identified, however, it was difficult to locate their main chains. This might suggest a more dynamic and unfixed

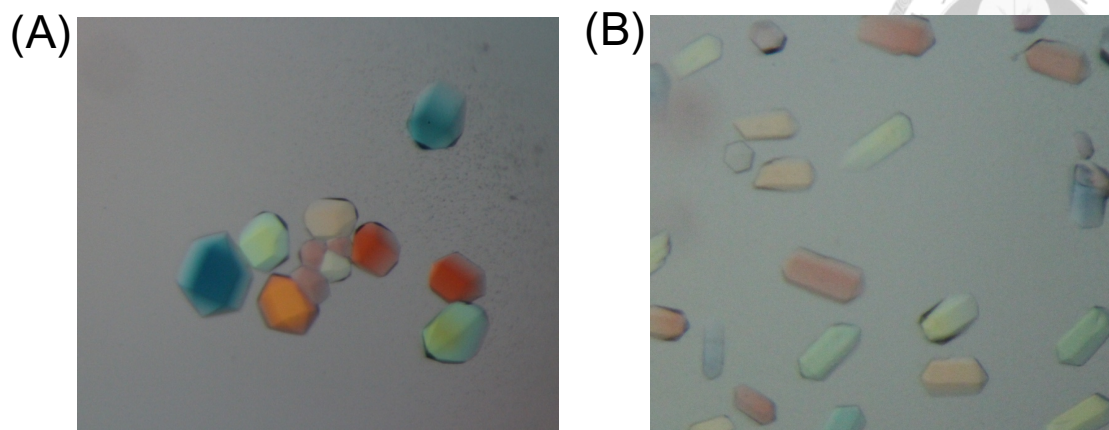


binding modes. The  $\alpha$ -carboxyl group of **2** (DmCF<sub>3</sub>Phe) is closer to  $\alpha$ -phosphate of AMPPNP, and its side chain flipped about 180 degrees compared to **3** (LmCF<sub>3</sub>Phe). For **6** (DmBrPhe) and **4** (DmClPhe), both their  $\alpha$ -carboxyl group is located closer to  $\alpha$ -phosphate of AMPPNP. The side chain of **4** is at about a 95-degree angle compared to **5** (LmClPhe) (**Figure 45**).



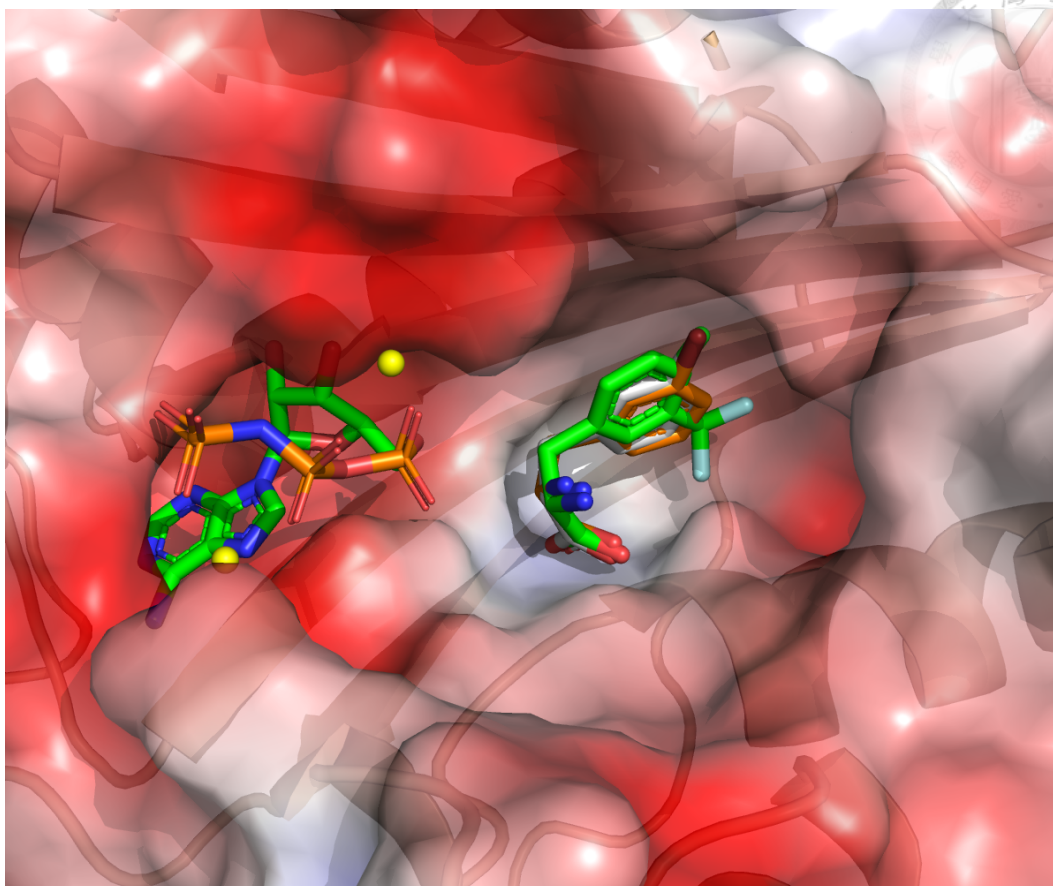
**Figure 41. Screening for crystallizing condition of N-DFRSc270.**

The condition of each picture: (A) 0.1 M Tris-HCl (pH 8.5), 8% PEG-8000, 8% Ethylene Glycol. (B) 0.1 M Tris-HCl (pH 8.5), 8% PEG-8000, 10% Ethylene Glycol. (C) 0.1 M Tris-HCl (pH 8.5), 10% PEG-8000, 8% Ethylene Glycol (D) 0.1 M Tris-HCl (pH 8.5), 10% PEG-8000, 10% Ethylene Glycol. And all condition with additional 5 mM ANP additive.



**Figure 42. Crystals of N-DFRSc270.**

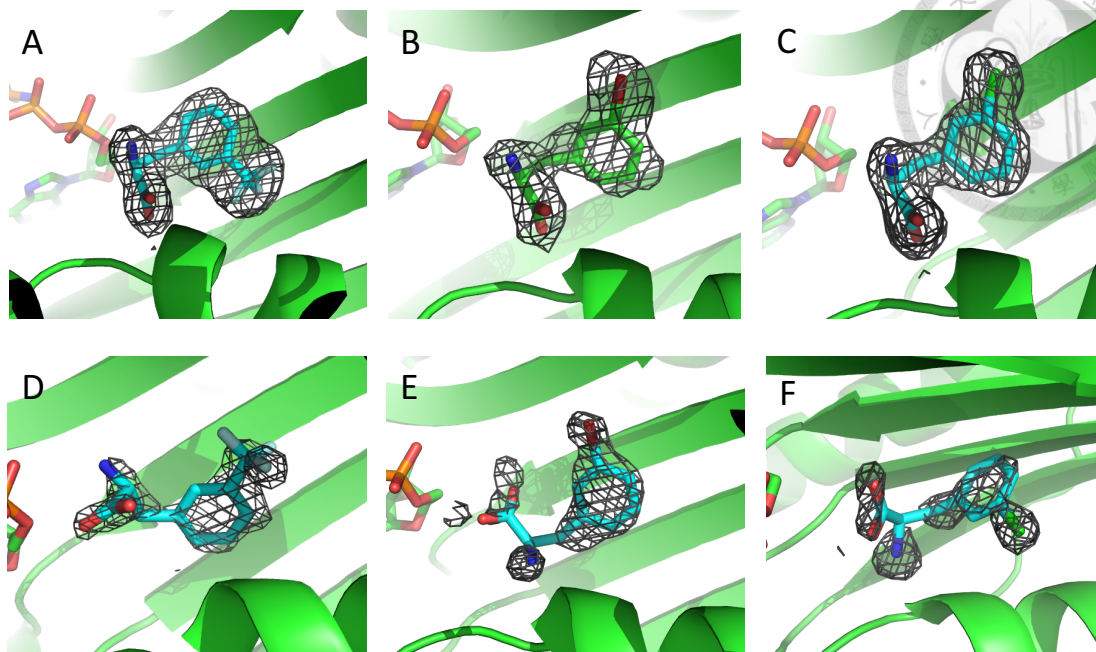
The crystals grew in different shape. The crystallizing condition after refinement was 100 mM HEPES buffer at pH 7.5, 10% PEG 8000, 10% Ethylene Glycerol with protein 10 mg/ml and 5 mM ANP. The picture was taken with polarized lenses.



**Figure 43. Crystal structure of N-DFRSc270 with ANP and LFA.**

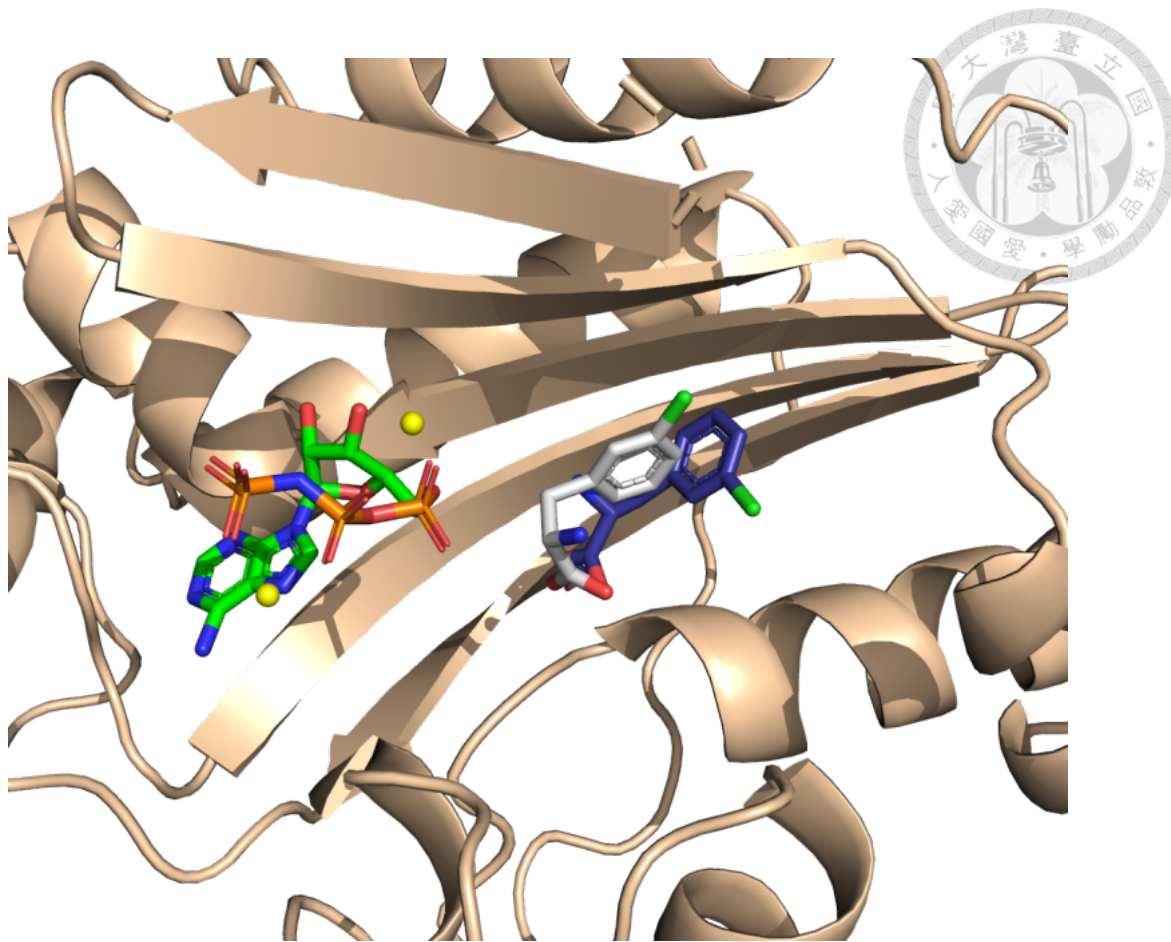
ANP and LFA were shown as stick;  $Mg^{2+}$  ions were shown as yellow spheres. The electrostatic surface potential scale was from -5 to +5 kT/e (red being negative and blue being positive). All three *L-meta*-phenylalanine derivatives, LmCF<sub>3</sub>Phe (**3**, green color), LmClPhe (**5**, white color), and LmBrPhe (**7**, red), were in the same position. The **3** with the CF<sub>3</sub> substituent had its substituents points at a different direction than the other derivatives, **5** and **7**, with Cl and Br substituents. (Structure resolution: **3**, 2.2 Å; **5**, 1.9 Å; **7**, 2.5 Å)





**Figure 44. Electron density map ( $2F_o-F_c$ ) around D/LFA in binding pocket.**

Green cartoon structure represents N-DFRSc270. (A)-(C) LmCF<sub>3</sub>Phe (**3**), LmBrPhe (**7**), and LmClPhe (**5**);  $\sigma = 1.0$ . (D)-(F) DmCF<sub>3</sub>Phe (**2**), DmBrPhe (**6**), and DmClPhe (**4**);  $\sigma = 0.9$ . All LFA (**3**, **5**, and **7**) can be built and fitted well to the electron density. In contrast, only partial electron density can be identified for the DFA (**2**, **4**, and **6**).




**Figure 45. Crystal structure of N-DFRSc270 with ANP and D/LCImPhe.**

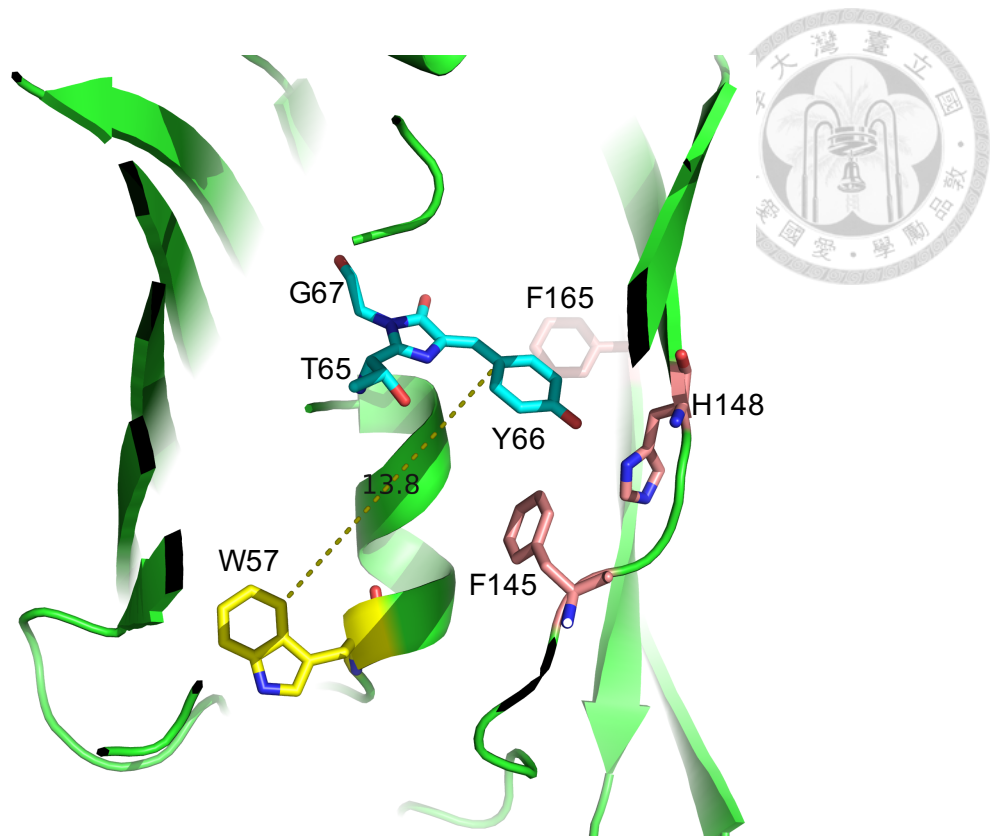
ANP and D/LFA were shown as stick;  $Mg^{2+}$  ions were shown as yellow spheres. The catalytic pocket of N-DFRS were shown as salmon cartoons. LmCIPhe (**5**, white) and DmCIPhe (**4**, blue) were in different orientations. The side chains of **5** and **4** were about 95 degrees away from each other in the active site of N-DFRS. (Structure resolution: **4**, 1.9 Å; **5**, 1.9 Å)

### 3.3 Studying biophysical property of sfGFP variant with D/LFA

#### 3.3.1 Incorporating D/LFA into chromophore of sfGFP



In nature, ribosomal translation system exclusively uses the 19 canonical L-amino acids and glycine, i.e. however, their D-form enantiomers are excluded from peptide or protein synthesis. Thus, the characteristics of protein with D-form ncAA were rarely understood. Here, the photophysical properties of sfGFP incorporating D/LFA was focused and the sfGFP was mutated with amber codon at Y66 position in chromophore, composed of T65, Y66, and G67<sup>60</sup> (**Figure 46**). And the gene of sfGFP-Y66TAG was inserted into pET-pylT plasmid and co-transformed with pCDF-N-DFRS in *E. coli* BL21(DE3). The cells were cultured with 1 mM **2-7** and 1 mM IPTG in minimal medium at 37 °C for 12 hrs. The over-expressed protein, sfGFP-Y66TAG with **2-7**, was purified and analyzed by ESI-MS. The MS spectrum confirmed that the found mass was same with the expected mass (**Figure 47-52**). However, the sfGFP-Y66TAG with **2-7** could not observed the green fluorescence at emission 485, only sfGFP-Y66TAG-**4** obtained weak fluorescence.

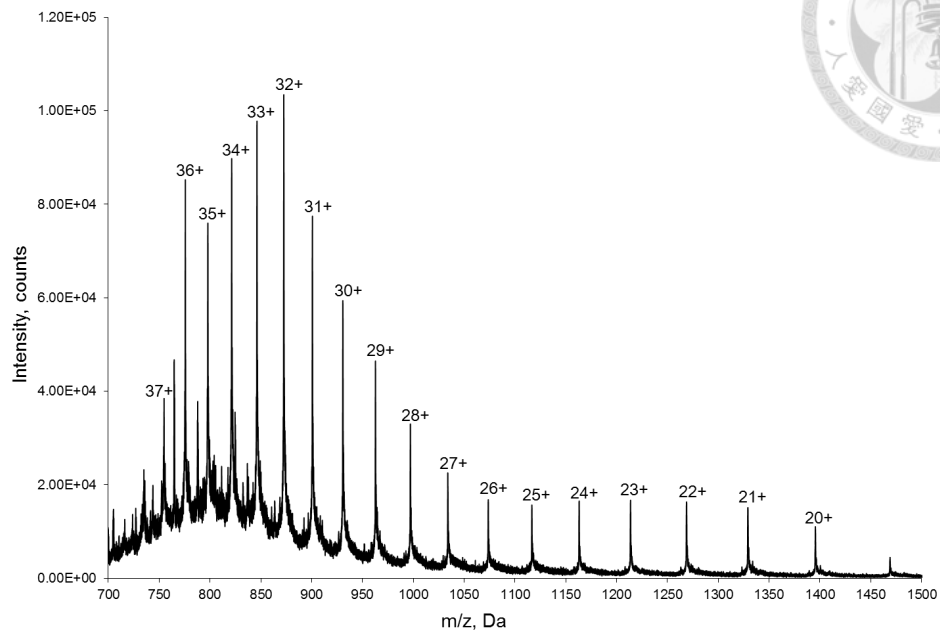


**Figure 46. The molecular insight of sfGFP in chromophore.**

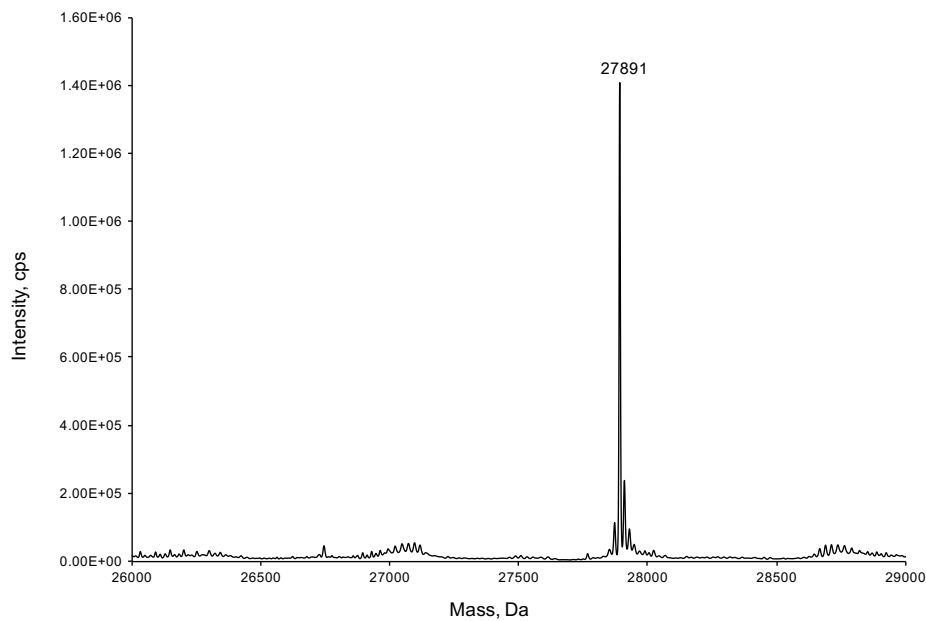
The fluorophore is composed of modified amino acid residues within the polypeptide chain at T65, Y66, and G67. And the cyclized backbone of these residues forms a  $\rho$ -hydroxybenzylidene-imidazolidone.<sup>61</sup> Moreover, the absorption spectrum has two maxima at 395 nm and at 470 nm, the fluorescence emission spectrum has a peak at 509 nm and a shoulder at 540 nm.<sup>62</sup> The cartoon figure was made by PyMOL. PDB code: 2b3p.



(A)



(B)

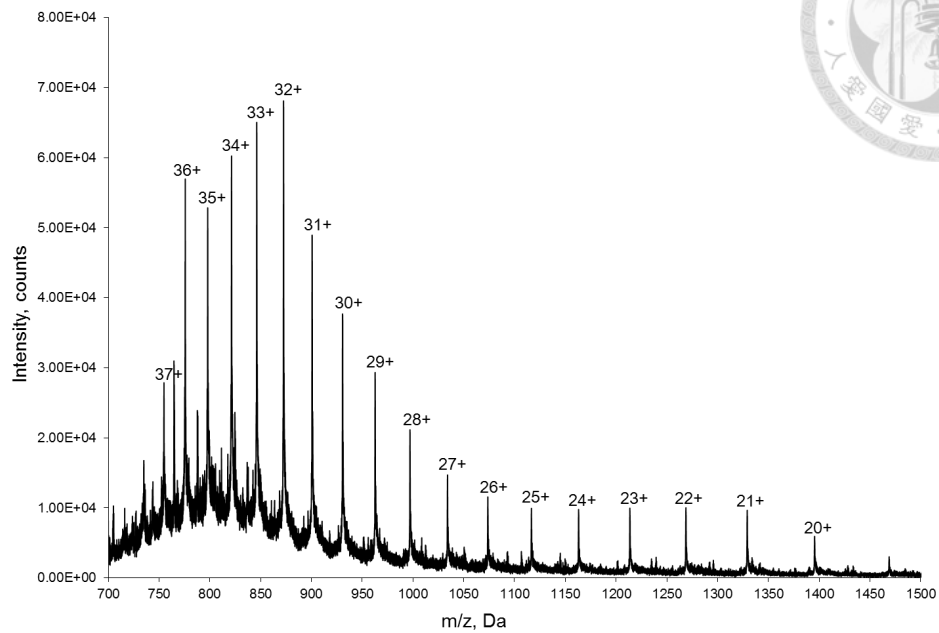


**Figure 47. Molecular mass determination of the protein sfGFP-Y66TAG-2.**

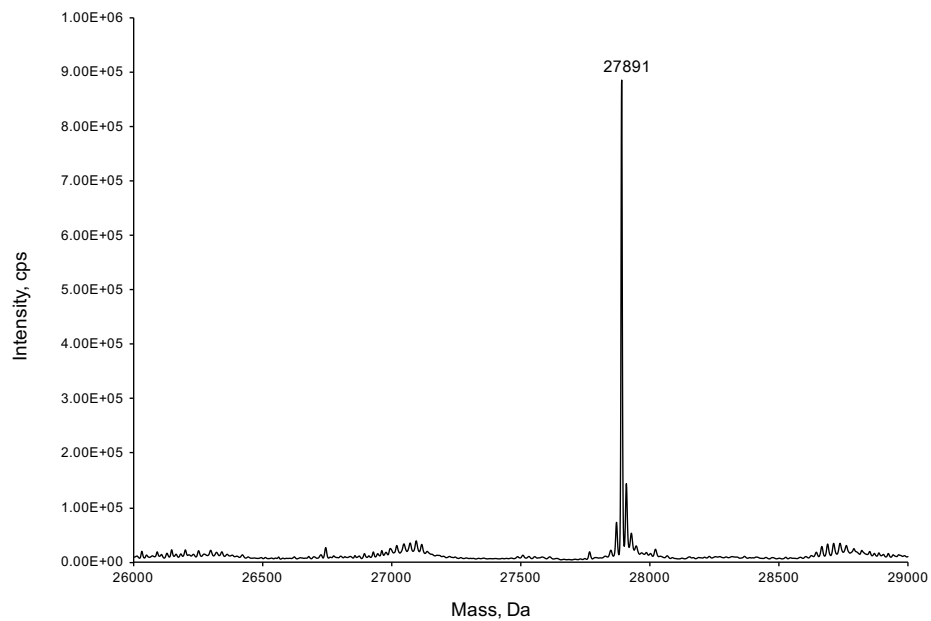
Full-length sfGFP-Y66TAG-2 was produced by using PylRS• tRNA<sup>Pyl</sup><sub>CUA</sub> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 27,891 Da (protein reducing first methionine); observed molecular mass is 27,891 Da.



(A)



(B)

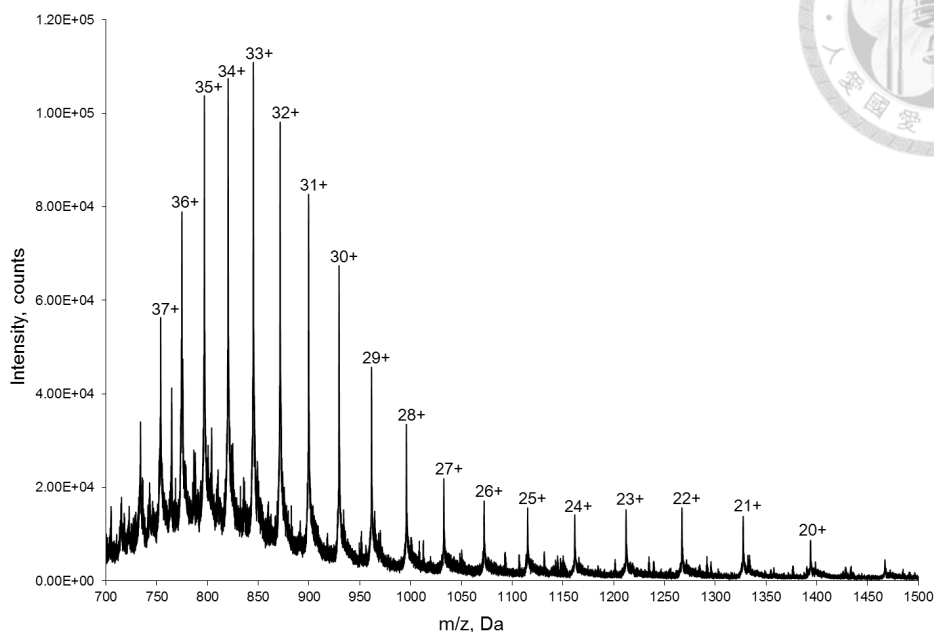


**Figure 48. Molecular mass determination of the protein sfGFP-Y66TAG-3.**

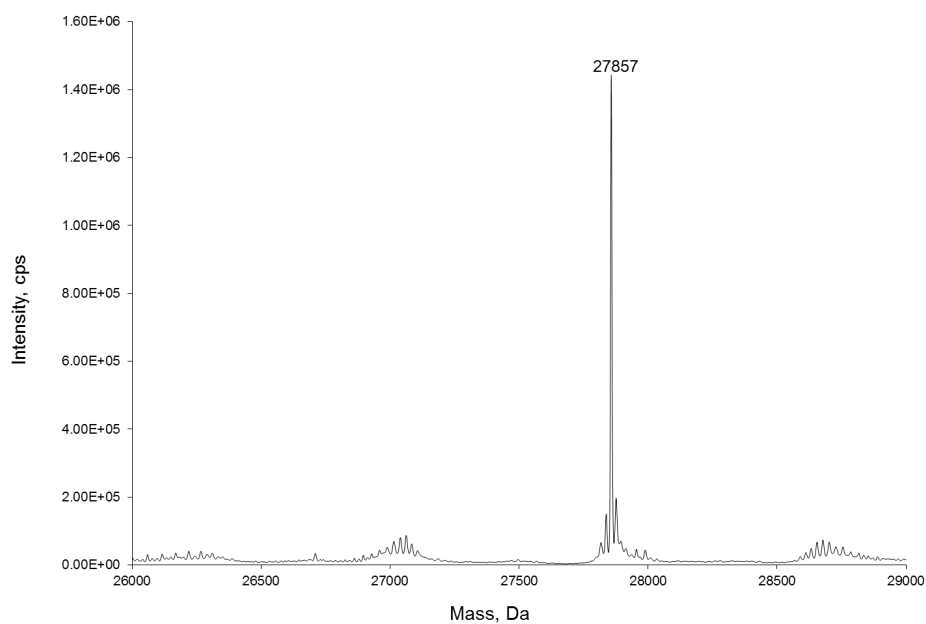
Full-length sfGFP-Y66TAG-3 was produced by using PylRS• tRNA<sup>Pyl</sup><sub>CUA</sub> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 27,891 Da (protein reducing first methionine); observed molecular mass is 27,891 Da.



(A)



(B)

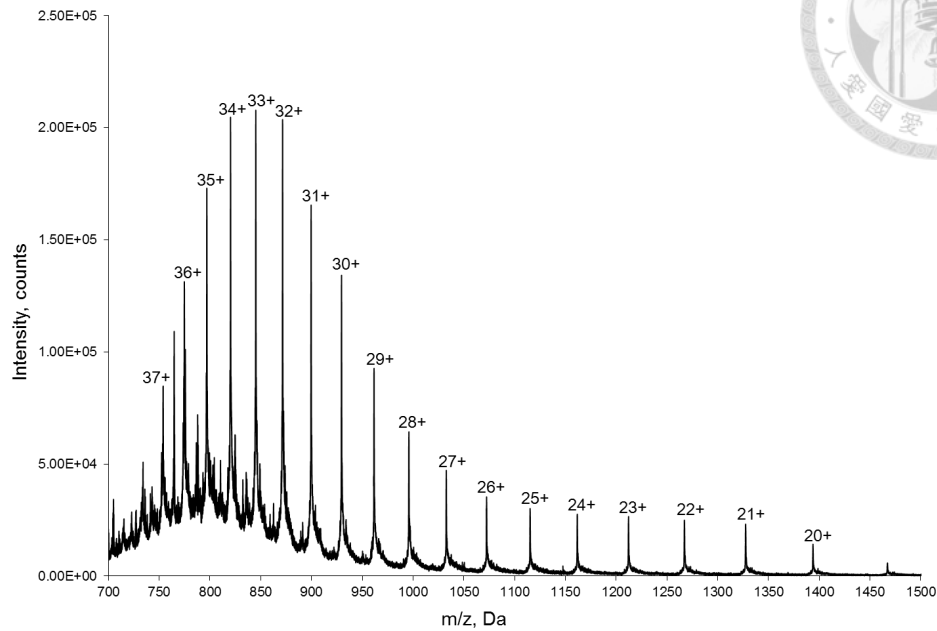


**Figure 49. Molecular mass determination of the protein sfGFP-Y66TAG-4.**

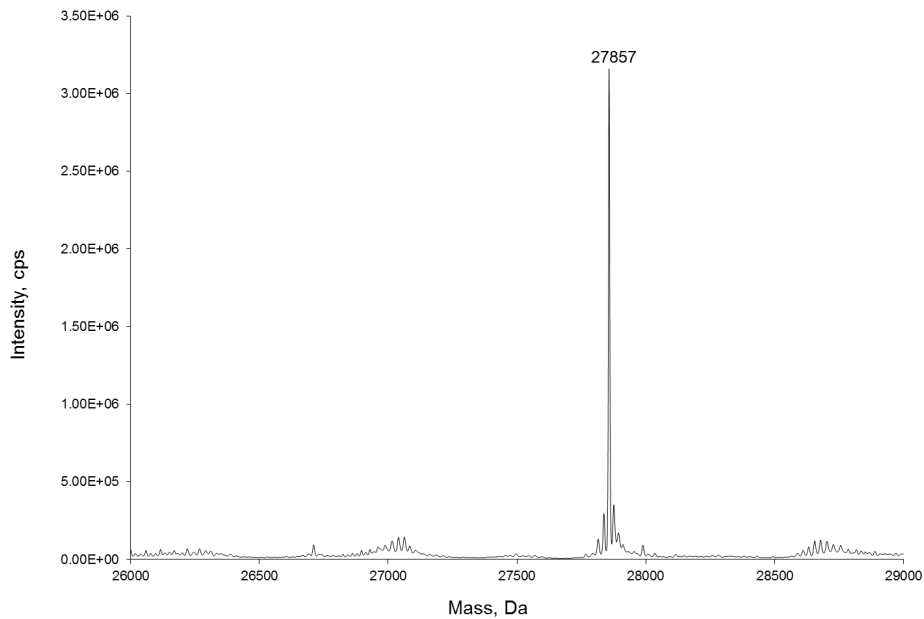
Full-length sfGFP-Y66TAG-4 was produced by using PylRS• tRNA<sup>Pyl</sup><sub>CUA</sub> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 27,857 Da (protein reducing first methionine); observed molecular mass is 27,857 Da.



(A)



(B)



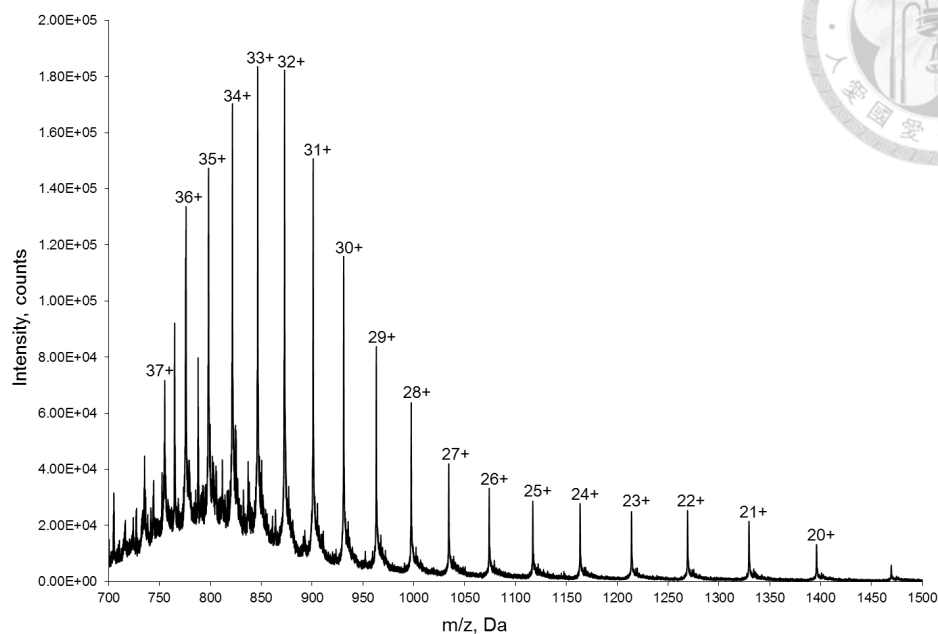
**Figure 50. Molecular mass determination of the protein sfGFP-Y66TAG-5.**

Full-length sfGFP-Y66TAG-5 was producing by using PylRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 27,857 Da (protein reducing first methionine); observed molecular mass is 27,857 Da.

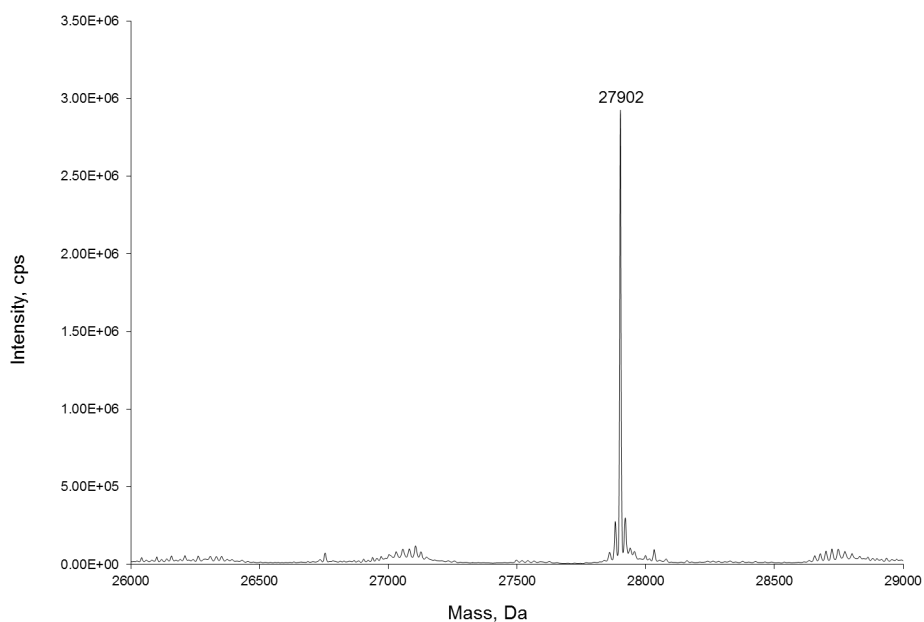




(A)



(B)

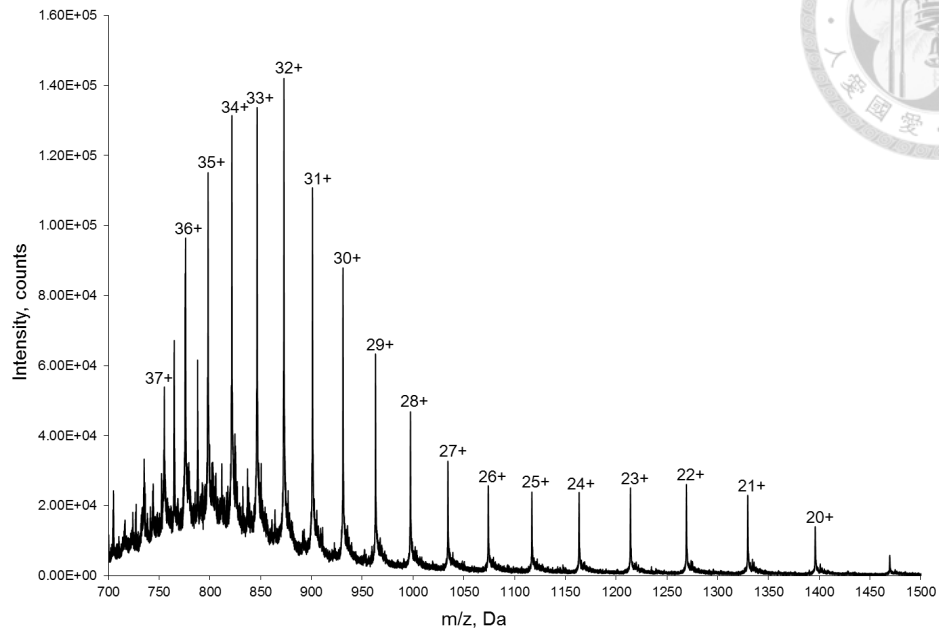


**Figure 51. Molecular mass determination of the protein sfGFP-Y66TAG-6.**

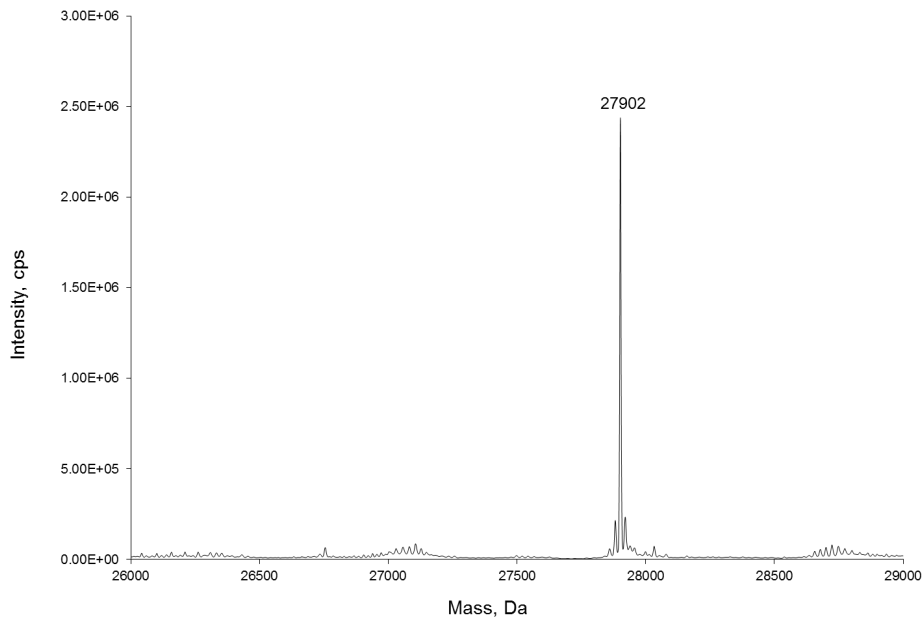
Full-length sfGFP-Y66TAG-6 was produced by using PylRS• tRNA<sup>Pyl</sup><sub>CUA</sub> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 27,902 Da (protein reducing first methionine); observed molecular mass is 27,902 Da.



(A)



(B)

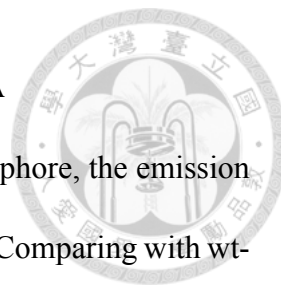


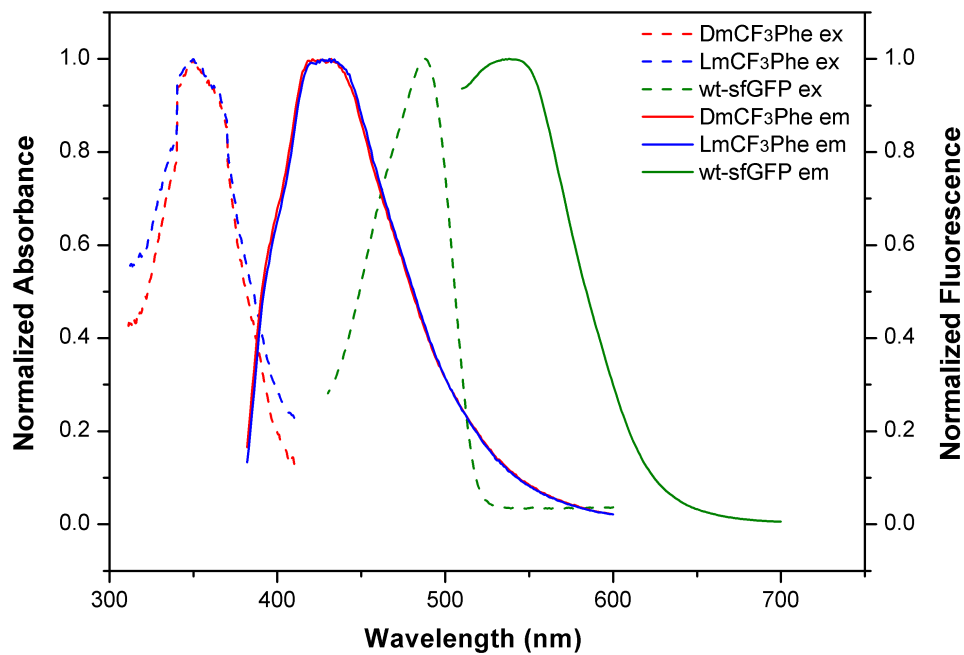
**Figure 52. Molecular mass determination of the protein sfGFP-Y66TAG-7.**

Full-length sfGFP-Y66TAG-7 was producing by using PylRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 27,092 Da (protein reducing first methionine); observed molecular mass is 27,902 Da.

### 3.3.2 Photophysical characteristics of sfGFP-Y66 with D/LFA

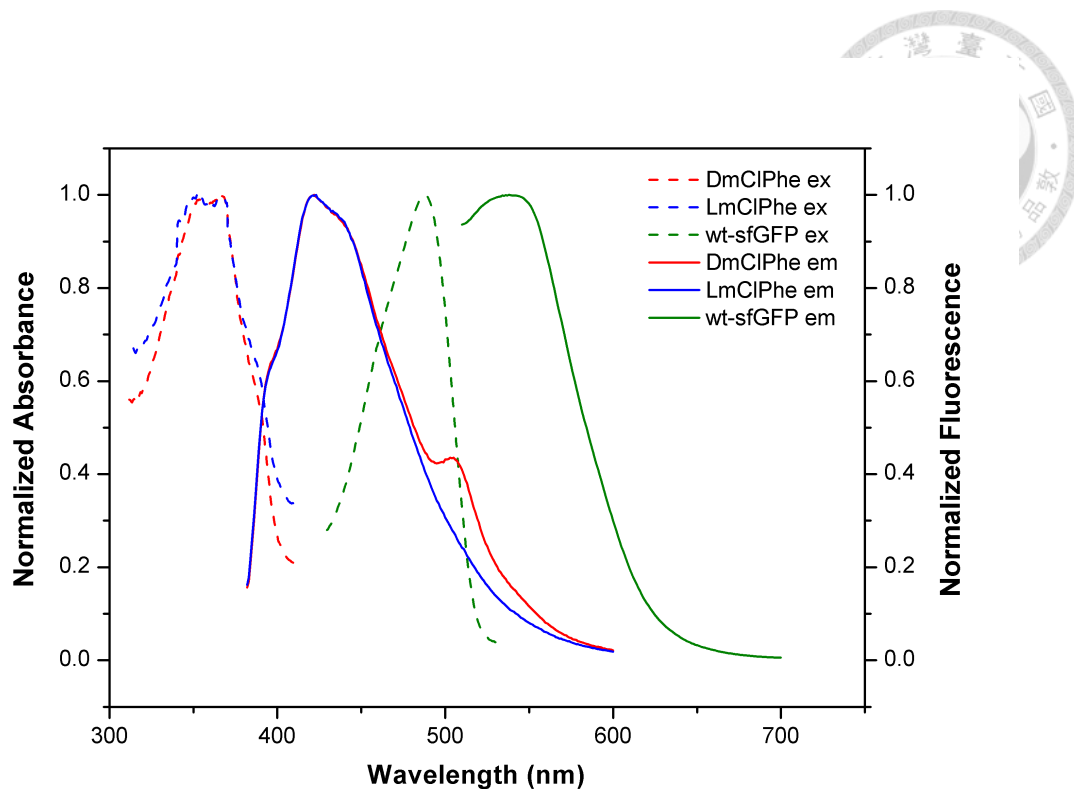
In order to realize the impact of sfGFP with D/LFA in chromophore, the emission and absorption spectrum of sfGFP-Y66TAG with **2-7** was measured. Comparing with wt-sfGFP, the wavelength of protein with **2-7** at Y66 position shifted to the shorter value in emission and excitation spectra. The emission wavelength of wt-sfGFP was 488 nm whereas sfGFP-Y66TAG with **2-7** showed shorter wavelength at about 350 nm. And the excitation wavelength of wt-sfGFP and sfGFP-Y66TAG with **2-7** was 538 nm and about 420 nm, respectively (**Figure 53-55**). In addition to this, sfGFP-Y66TAG-**4** (DmCIPhe) presented another peak in excitation spectra which could not find in spectra of sfGFP-Y66TAG-**5** (LmCIPhe) and the wavelength of peak was about 510 nm. Next, performing intrinsic fluorescence of sfGFP variants, the proteins were excited with 280 nm, is the absorption wavelength of tryptophan, and sfGFP contain a Trp at position 57. The result demonstrated the fluorescence intensity of sfGFP-Y66TAG with D/LFA was lower than wt-sfGFP, and all of sfGFP variants presented the similar emission wavelength at 330 nm but with various FRET spectra in different pair of L/DFA. Specially, comparing the intrinsic fluorescent spectrum of sfGFP-Y66TAG-**4** with sfGFP-Y66TAG-**5**, protein with **4** in chromophore appeared unique peak at 507 nm in emission spectrum without obviously FRET, but protein with its enantiomers, **5**, presented obviously FRET spectrum (**Figure 56-58**).





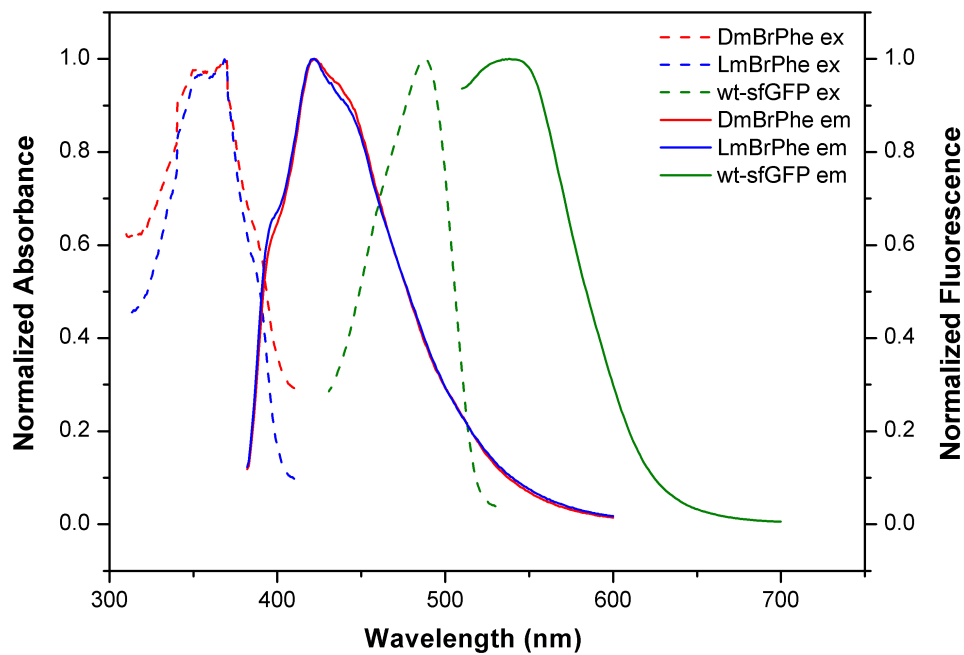
**Figure 53. Photophysical characterization of sfGFP-Y66TAG-2 and 3.**

Excitation (solid lines) and emission (dashed lines) spectrum of sfGFP-Y66-2 (red color), sfGFP-Y66-3 (blue color), and wt-sfGFP (green color) proteins were analyzed with 0.5 mg/mL concentration in 25°C. The maximum of excitation and emission were defined as 1 to normalize the value of spectra. sfGFP-Y66-2 (ex. 349.5 nm, em. 432 nm), sfGFP-Y66-3 (ex. 350 nm, em. 431 nm), and wt-sfGFP (ex. 488 nm, em. 538 nm).



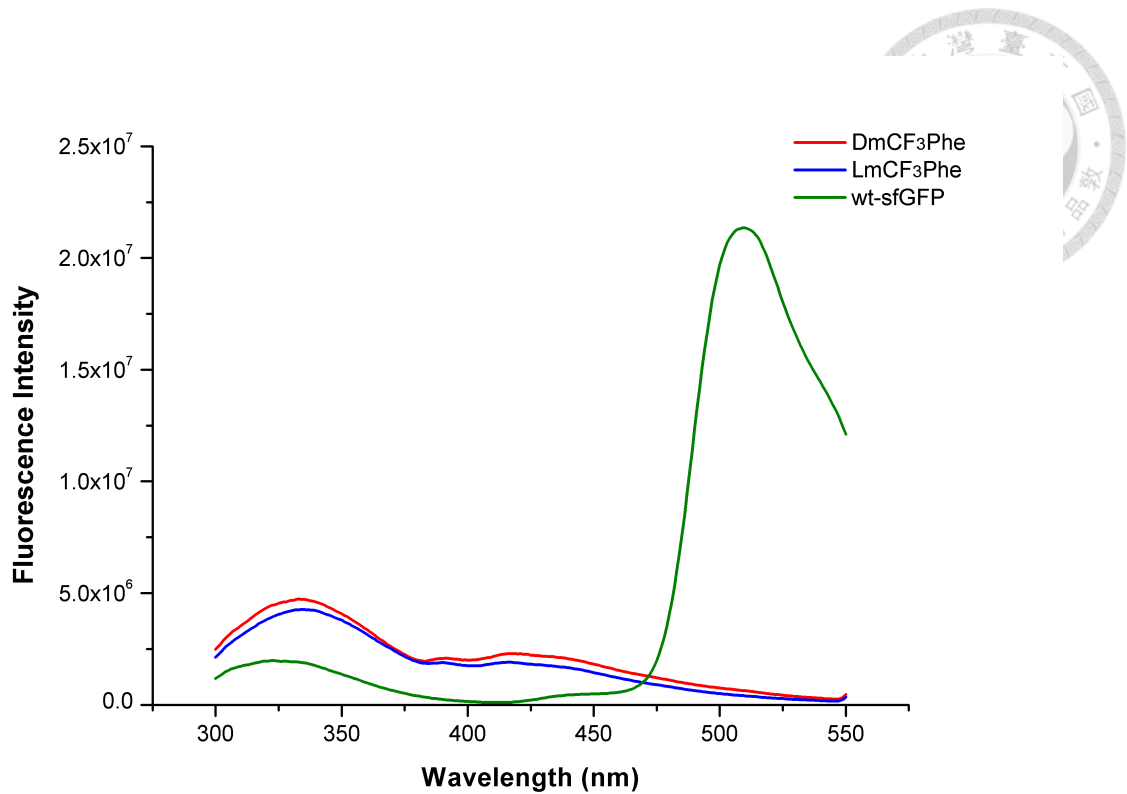
**Figure 54. Photophysical characterization of sfGFP-Y66TAG-4 and 5.**

Excitation (solid lines), and emission (dashed lines) spectrum of sfGFP-Y66-4 (red color), sfGFP-Y66-5 (blue color), and wt-sfGFP (green color) proteins were analyzed with 0.5 mg/mL concentration in 25°C. The maximum of excitation and emission were defined as 1 to normalize the value of spectra. sfGFP-Y66-4 (ex. 365.5 nm, em. 422 nm and 510 nm), sfGFP-Y66-5 (ex. 352 nm, em. 423 nm), and wt-sfGFP (ex. 488 nm, em. 538 nm).



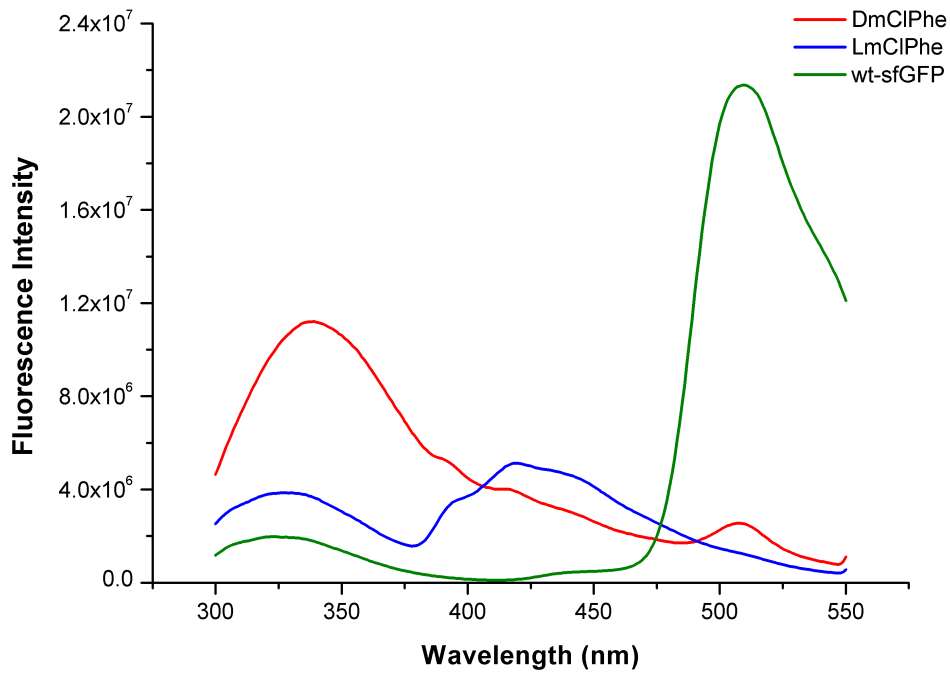
**Figure 55. Photophysical characterization of sfGFP-Y66TAG-6 and 7.**

Excitation (solid lines) and emission (dashed lines) spectrum of sfGFP-Y66-6 (red color), sfGFP-Y66-7 (blue color), and wt-sfGFP (green color) proteins were analyzed with 0.5 mg/mL concentration in 25°C. The maximum of excitation and emission were defined as 1 to normalize the value of spectra. sfGFP-Y66-6 (ex. 368.5 nm, em. 422 nm), sfGFP-Y66-7 (ex. 368.5 nm, em. 422 nm), and wt-sfGFP (ex. 488 nm, em. 538 nm).



**Figure 56. Intrinsic fluorescence spectrum of sfGFP-Y66TAG-2 and 3.**

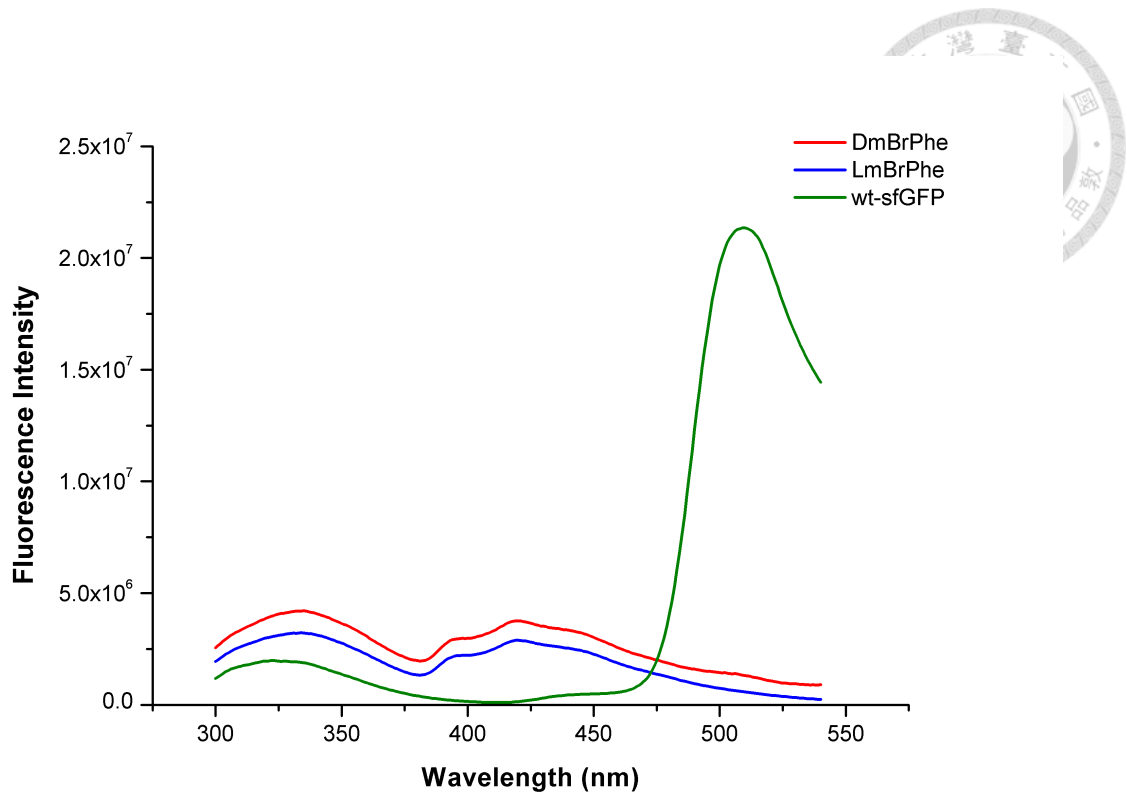
The spectra of sfGFP-Y66-2 (red color), sfGFP-Y66-3 (blue color), and wt-sfGFP (green color) proteins were excited with 280 nm in 0.5 mg/mL concentration at 25°C. The spectra of wt-sfGFP contained two obvious peaks at 322 nm and 509 nm. But sfGFP-Y66-2 spectra and sfGFP-Y66-3 spectra just contained one obvious peaks at 333 nm and 335 nm separately, in addition, the spectra raised a little bit at 415 nm.



**Figure 57. Intrinsic fluorescence spectrum of sfGFP-Y66TAG-4 and 5.**

The spectra of sfGFP-Y66-4 (red color), sfGFP-Y66-5 (blue color), and wt-sfGFP (green color) proteins were excited with 280 nm in 0.5 mg/mL concentration at 25°C. All spectra contained two obvious peaks: wt-sfGFP (322 nm and 509 nm), sfGFP-Y66-4 (339 nm and 507 nm), and sfGFP-Y66-5 (328 nm and 419 nm).






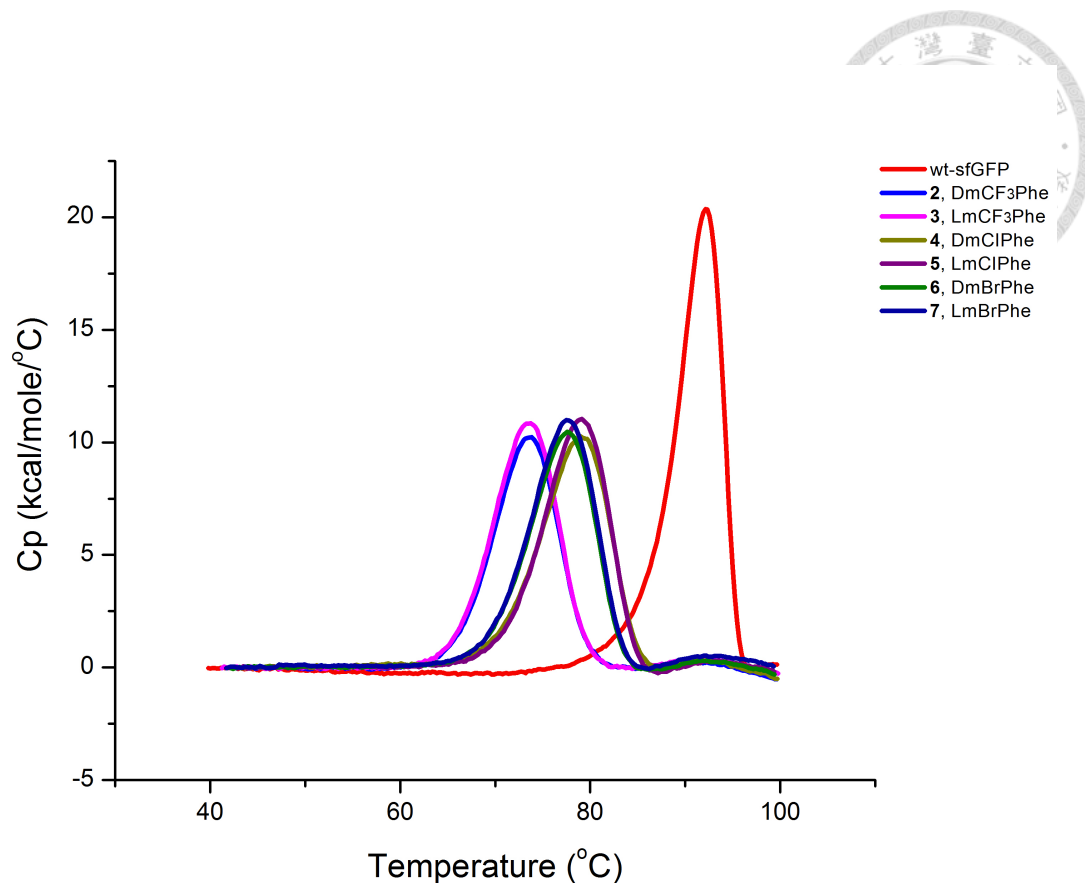
**Figure 58. Intrinsic fluorescence spectrum of sfGFP-Y66TAG-6 and 7.**

The spectra of sfGFP-Y66-6 (red color), sfGFP-Y66-7 (blue color), and wt-sfGFP (green color) proteins were excited with 280 nm in 0.5 mg/mL concentration at 25°C. All spectra contained two obvious peaks: wt-sfGFP (322 nm and 509 nm), sfGFP-Y66-6 (335 nm and 419 nm), and sfGFP-Y66-7 (334 nm and 420 nm).

### 3.3.3 Thermally induced conformational changes of sfGFP with Y66 D/LFA

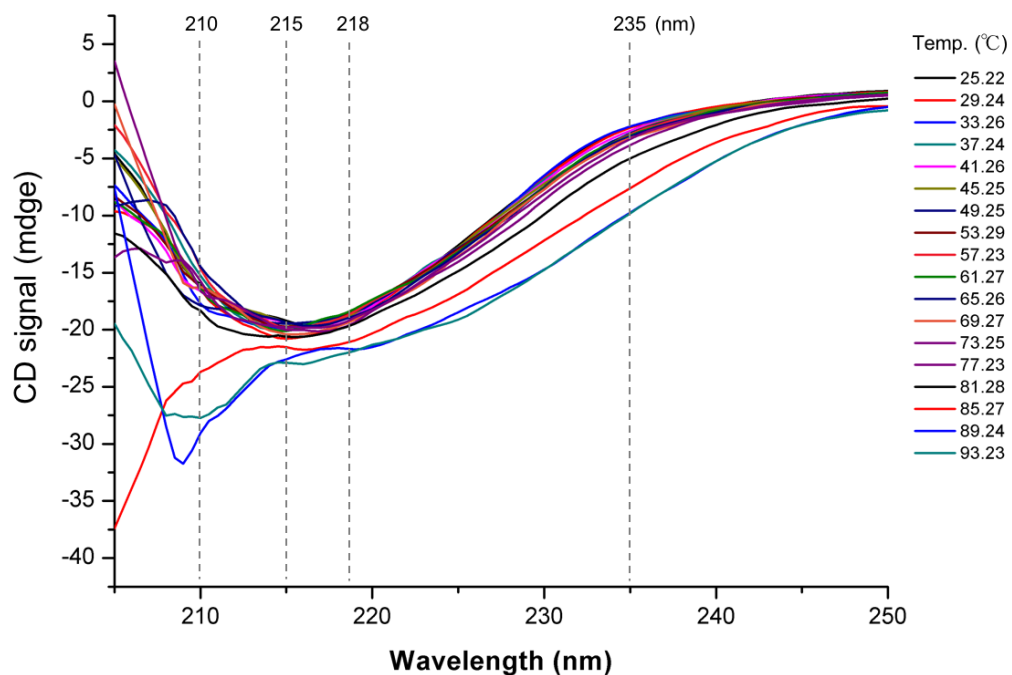


The structural information of protein with D-form amino acid was still uncommon. In our study, DSC was performed to assess the stability of sfGFP-Y66TAG with different D/LFA. Upon heating the sample from 20°C to 100°C. wt-sfGFP exhibited the highest  $T_m$  value at 92.33°C. And the sfGFP-Y66TAG with **2-7** presented slight different in  $T_m$  values, which could be separated into three group by different substitution on D/LFA. The sfGFP-Y66TAG-**2, 3** group exhibited the transition at  $73.7 \pm 0.1^\circ\text{C}$ , which lower than both sfGFP-Y66TAG-**4, 5** group ( $79.1 \pm 0.1^\circ\text{C}$ ) and fGFP-Y66TAG-**6, 7** group ( $77.5 \pm 0.1^\circ\text{C}$ ) (**Figure 59**). To determine whether protein with D/LFA at Y66 position had impact on over all secondary structure when heating, CD was performed. Because of the distinct fluorescent spectrum on sfGFP-Y66TAG-**4** and **5**, we focused on this group. Far-UV wavelength was scanned from 205 to 250 nm and collected at variable temperatures between 25°C and 95°C. The result demonstrated the CD spectra of wt-sfGFP and sfGFP with LmClPhe (**5**) became more negative following heating. However, CD spectra of sfGFP containing DCIPhe (**4**) gradually moved to positive after heating more than 65°C (**Figure 60-62**). The difference was more clear when comparing the signal of wt-sfGFP with sfGFP-Y66TAG with **4** or **5** at particular wavelengths, 210 nm, 215 nm, 218 nm used to detect secondary structure of  $\beta$ -sheet, and 235 nm. The trend of wt-sfGFP and sfGFP with LmClPhe (**5**) were toward negative but sfGFP with LmClPhe (**4**) toward the positive in CD spectra at 215 and 218 nm (**Figure 63**). And the change of sfGFP-Y66TAG-**4** was less then sfGFP-Y66TAG-**5** at 210 and 235 nm.



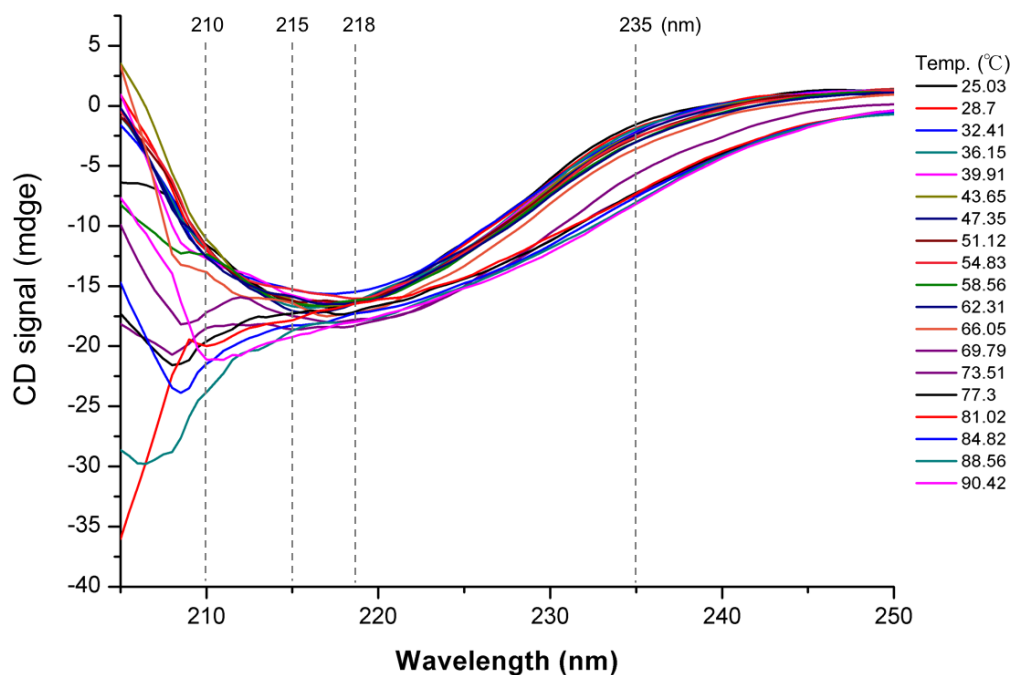
**Figure 59. DSC profile of sfGFP-Y66TAG incorporated with D/LFA.**

The protein was dissolved in phosphate buffer at pH 7.0 with 0.2 mg/ml concentration. The temperature change from 20°C to 100°C with a rate of 200°C/hr. Different color of line means sfGFP-Y66TAG incorporating with different ncAA. The midpoint of denatured temperature ( $T_m$ ) of wt-sfGFP (92.33°C), sfGFP-Y66TAG-2 (73.85°C), sfGFP-Y66TAG-3 (73.74°C), sfGFP-Y66TAG-4 (79°C), sfGFP-Y66TAG-5 (79.11°C), sfGFP-Y66TAG-6 (77.58°C), and sfGFP-Y66TAG-7 (77.51°C).



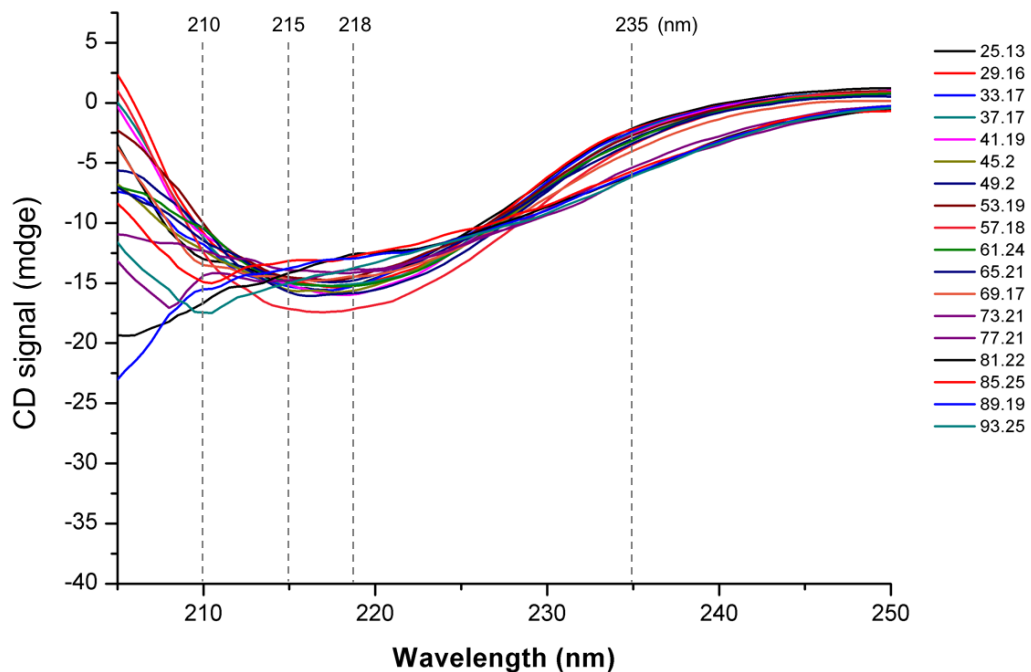
**Figure 60. Thermally induced conformational changes of wt-sfGFP.**

The structural change was monitored by CD and sfGFP was dissolved in PBS buffer at pH 7.0 with 0.5 mg/ml concentration. The Far-UV CD spectra recorded structural transformation from 25 °C to 95°C at a rate of 2°C/min and a given temperature was equilibrated to keep within +/- 0.1°C for 10 sec. Wavelength with the changeable CD signal, 210 nm, 215 nm, 218 nm, and 235 nm, were compared to the other sfGFP variants.



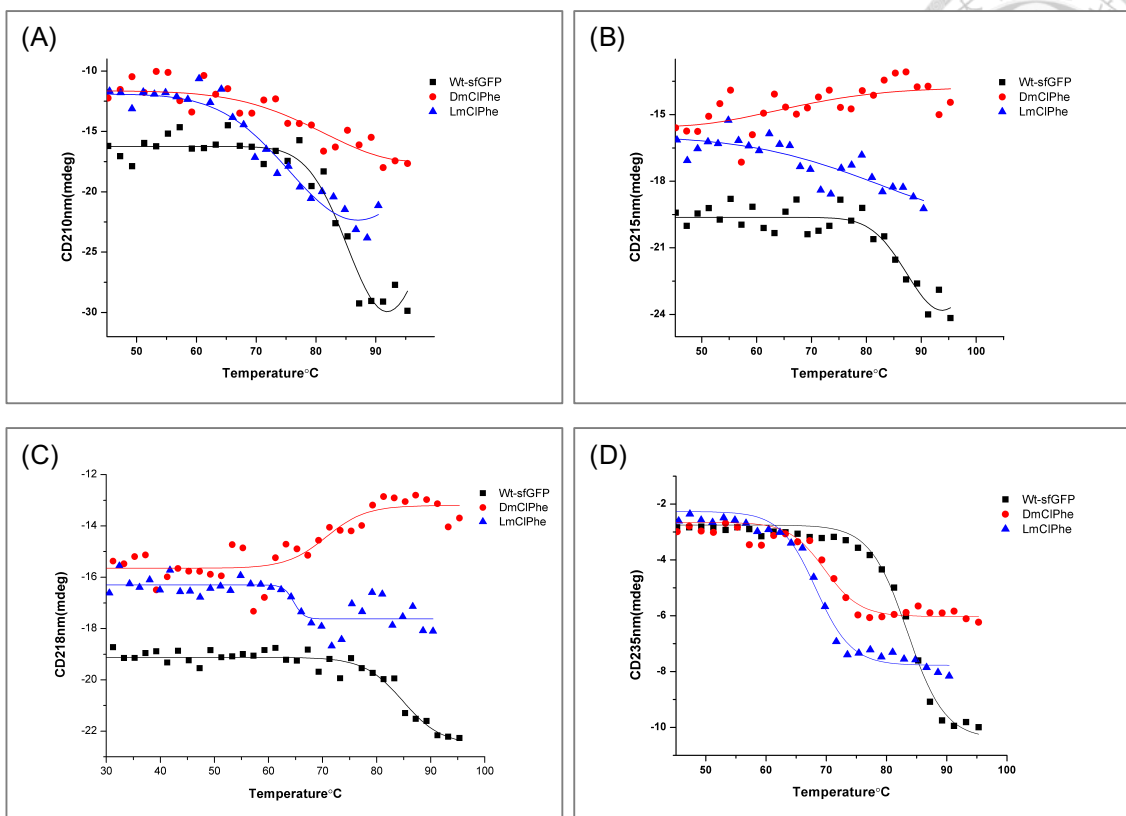
**Figure 61. Thermally induced conformational changes of sfGFP-Y66TAG-5.**

The structural change was monitored by CD and sfGFP-Y66TAG-5 was dissolved in PBS buffer at pH 7.0 with 0.5 mg/ml concentration. The Far-UV CD spectra recorded structural transformation from 25 °C to 95°C at a rate of 2°C/min and a given temperature was equilibrated to keep within +/- 0.1°C for 10 sec. Wavelength with the changeable CD signal, 210 nm, 215 nm, 218 nm, and 235 nm, were compared to the other sfGFP variants.



**Figure 62. Thermally induced conformational changes of sfGFP-Y66TAG-4.**

The structural change was monitored by CD and sfGFP-Y66TAG-4 was dissolved in PBS buffer at pH 7.0 with 0.5 mg/ml concentration. The Far-UV CD spectra recorded structural transformation from 25°C to 95°C at a rate of 2°C/min and a given temperature was equilibrated to keep within +/- 0.1°C for 10 sec. Wavelength with the changeable CD signal, 210 nm, 215 nm, 218 nm, and 235 nm, were compared to the other sfGFP variants.



**Figure 63. Thermally induced conformational changes monitored by CD.**

The concentration of protein was 0.5 mg/ml and dissolved in PBS buffer at pH 7.0. The spectra recorded signal intensity at (A) 210 nm, (B) 215 nm, (C) 218 nm, and (D) 235 nm, from 25°C to 95°C at a rate of temperature change of 2°C/min and a given temperature was equilibrated to keep within +/- 0.1°C for 10 sec. Red one: sfGFP-Y66TAG-4; blue one: sfGFP-Y66TAG-5; Black one: wt-sfGFP.

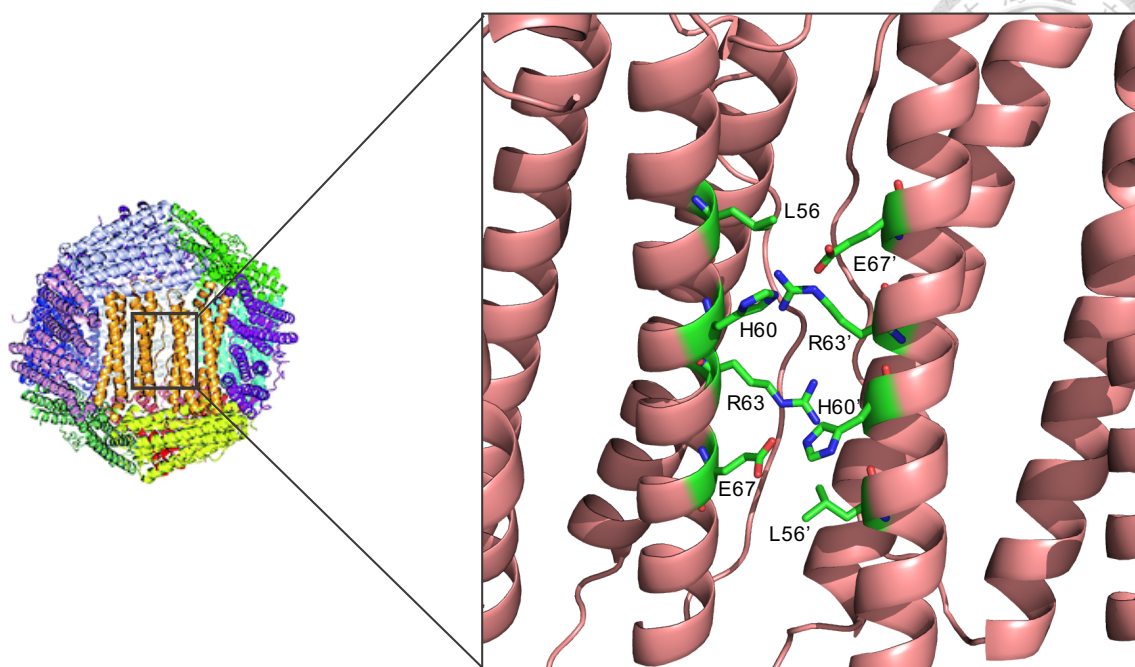


### 3.4 Application of D/LFA into protein cage design

Partial D-amino acid substitution of peptide shows high resistance against proteolytic degradation.<sup>24</sup> In order to improved enzymatic stability of protein cage which was utilized to protein/drug encapsulation. DFA incorporated ferritin was designed in the C<sub>2</sub> interface. The published literature identified sites L56, H60 and sites R63, E67' in the other ferritin monomer that lie across the interface from each other<sup>40</sup> (**Figure 64**). Thus, ferritin was constructed with amber codon at L56 position in pET-pylT plasmid and co-transformed with N-DFRS•tRNA<sup>Py1</sup><sub>CUA</sub> pair in *E. coli* BL21(DE3). The target protein was over-expressed with 1 mM **2** or **3** and 1 mM IPTG in M9 medium at 37°C for 12 hrs. And the cell was pelleted and purified by Ni<sup>2+</sup>-NTA column and SEC column (HiLoad 16/600 superdex 200 pg, GE Healthare). The FPLC profile of SEC exhibited two major peak in both of ferritin with **2** and **3**. After analyzed by SDS-PAGE, fraction in peak 2 was collected, although peak 1 containing few proteins (**Figure 65-66**). The purified protein was analyzed by LC-ESI-MS and confirmed that the found mass was same with the expected mass (**Figure 67-68**). To preliminarily understand whether ferritin-L56TAG with **2** or **3** self-assemble into 24-mer as protein cage (about 530 kDa) in PBS buffer. the native gel analysis was performed (**Figure 65-66**). As our expectation, the gel demonstrated the size of ferritin collected from peak 2 was bigger than 480 kDa in native form, and peak 1 was appeared some soluble aggregation and di-ferritin. In order to optimized the buffer condition for homogeneous dispersion of ferritin, the buffer was change to 50 mM Tris-HCl and 100 mM NaCl at pH 8.0 in purification. After SEC column, the profile of FPLC exhibited one peak (**Figure S5-S6**) and native gel demonstrated one band near 480kDa marker at pH 8.0 in ferritin-L56 with **2** and **3** (**Figure**

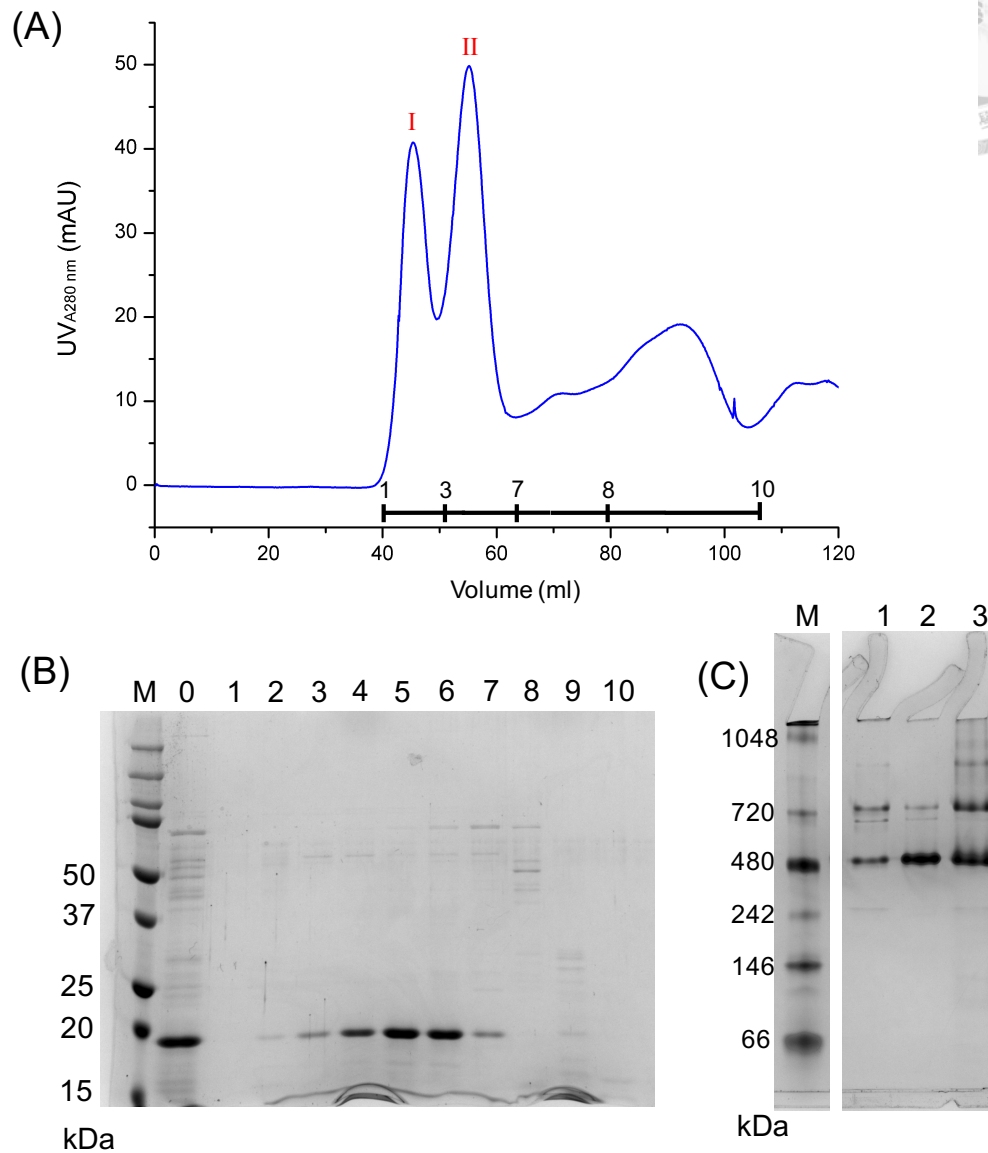


**S11**). However, the wt-ferritin still appeared two overlapping peak in FPLC profile (**Figure S7**) and three band in native gel at pH 8.0 (**Figure S11**). In addition, the wt-ferritin was apparently observed various size particles in PBS buffer, but it cannot be separated by SEC column (HiLoad 16/600 superdex 200 pg, GE Healthare). Thus the sample was later analyzed by Field-Flow Fractionation- Multi-Angle Light Scattering (FFF-MALS). The result demonstrated three peaks and molar weight was 554.7 kDa, 1006 kDa, and 1459 kDa, individually. These molar weights were near to the size of mono-ferritin, di-ferritin and tri-ferritin. (**Figure S8**).



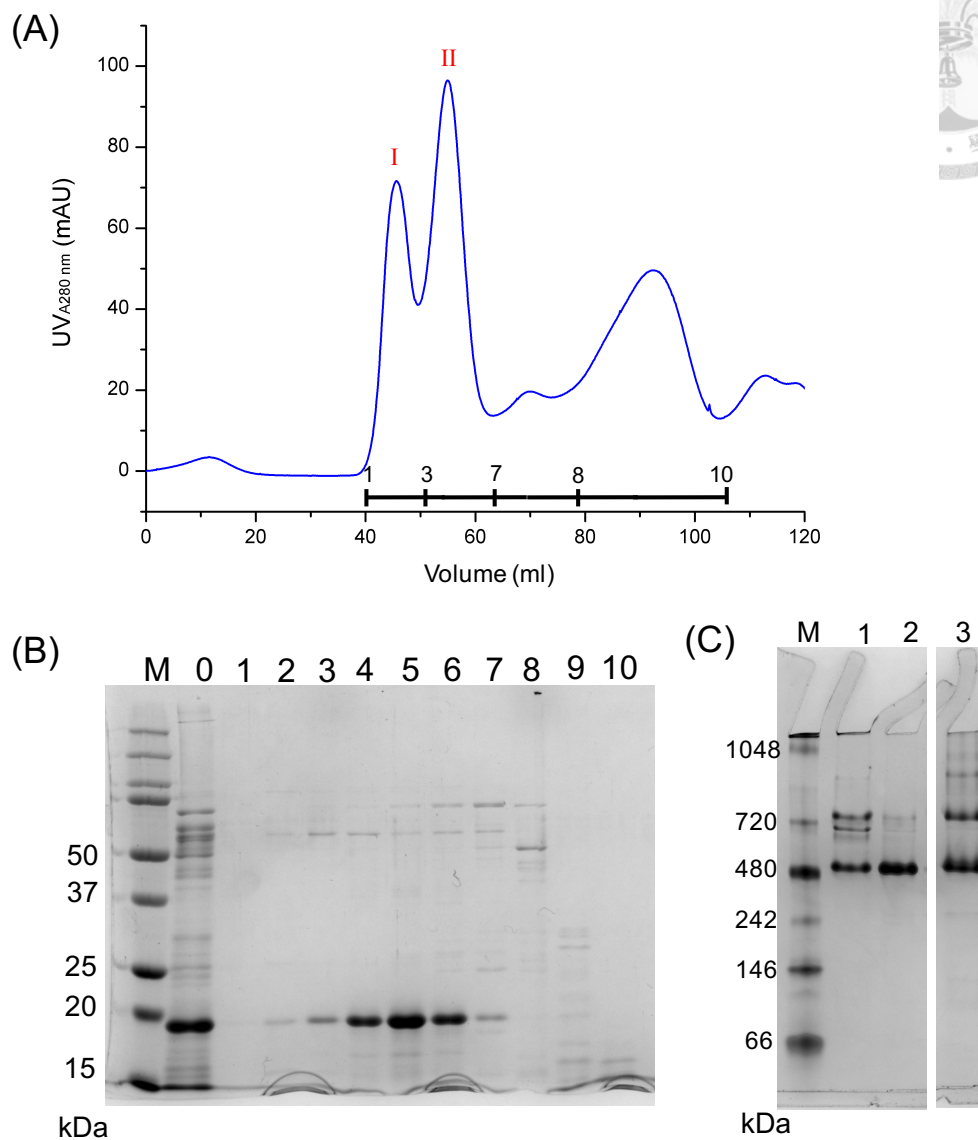
**Figure 64. The structure of  $C_2$  interface of ferritin.**

The green sticks represent residues lying across the interface from each other and play an essential role in self-assembly.<sup>40</sup> In this thesis, L56 position was focused. The cartoon figure was made by PyMOL. PDB code: 2fha.



**Figure 65. Purification of Ferritin-L56TAG-2 with SEC.**

(A) The elution in PBS buffer after Ni<sup>2+</sup>-NTA column was concentrated for FPLC with SEC column (HiLoad 16/600 superdex 200 pg, GE Healthare). Two major peak appeared in the profile. Blue line: the signal of UV<sub>A280</sub>; (B) SDS-PAGE shows the fractions indicated in (A). 0: elution after Ni<sup>2+</sup>-NTA column; 1-10: relative fractions in FPLC profile. Fraction 4-7 were collected and concentrated for TEM and biophysical experiments. (C) 5  $\mu$ g protein from peak I and peak II in (A) was invidiously mixed with native sample dye and running native gel with 80 V at 4°C for 6 hrs. 1: peak I; 2: peak II; 3: wt-ferritin (as the control).

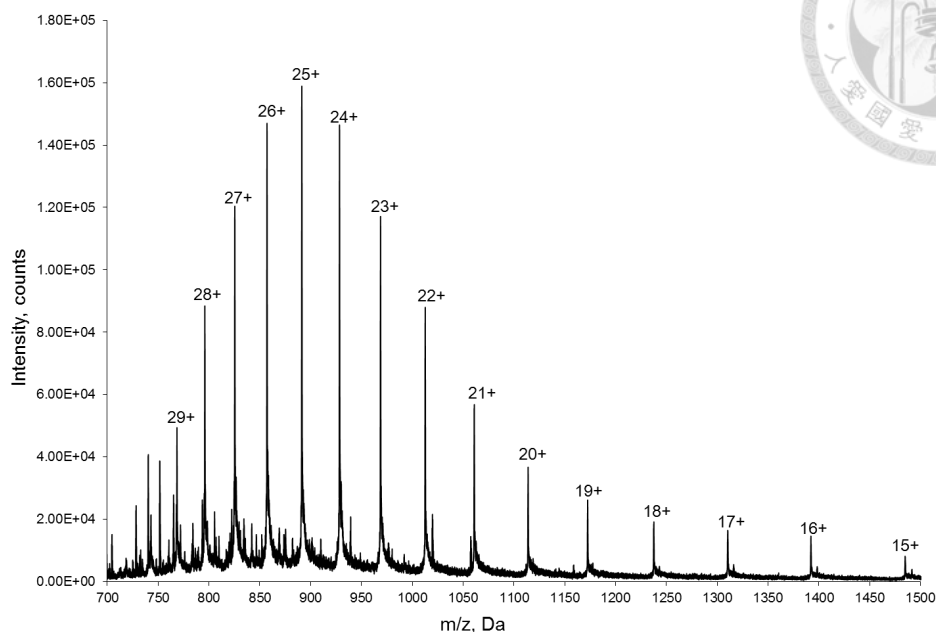


**Figure 66. Purification of Ferritin-L56TAG-3 with SEC.**

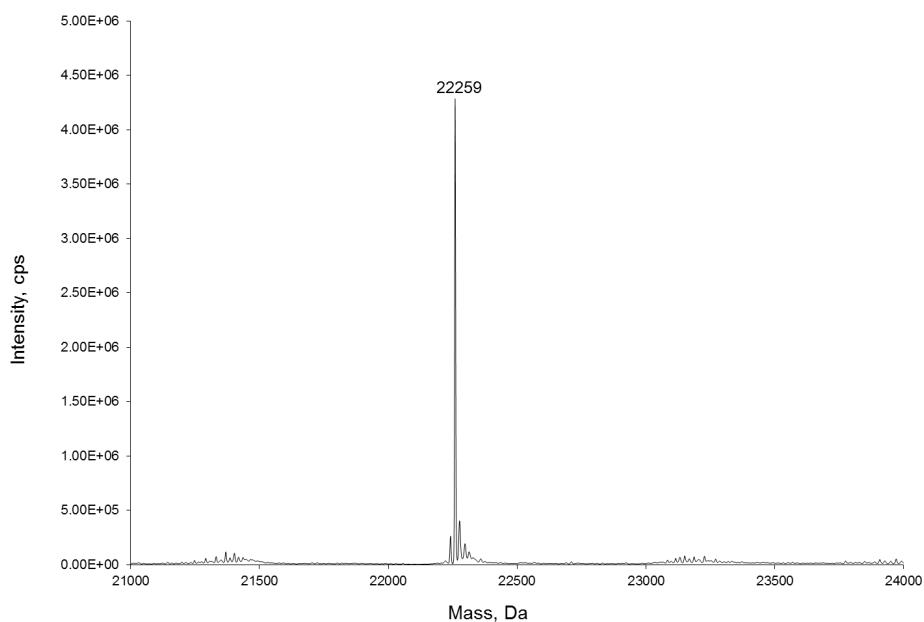
(A) The elution in PBS buffer after  $\text{Ni}^{2+}$ -NTA column was concentrated for FPLC with SEC column (HiLoad 16/600 superdex 200 pg, GE Healthcare). Two major peak appeared in the profile. Blue line: the signal of  $\text{UV}_{\lambda 280}$ ; (B) SDS-PAGE shows the fractions indicated in (A). 0: elution after  $\text{Ni}^{2+}$ -NTA column; 1-10: relative fractions in FPLC profile. Fraction 4-7 were collected and concentrated for TEM and biophysical experiments. (C) 5  $\mu\text{g}$  protein from peak I and peak II in (A) was individually mixed with native sample dye and running native gel with 80 V at 4°C for 6 hrs. 1: peak I; 2: peak II; 3: wt-ferritin (as the control).



(A)



(B)

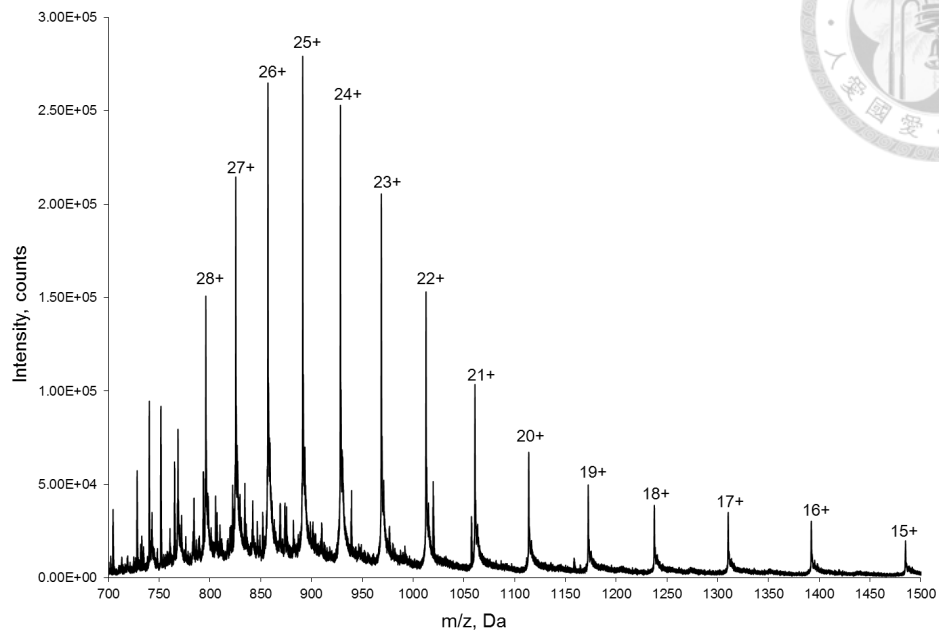


**Figure 67. Molecular mass determination of the protein Ferritin-L56TAG-2.**

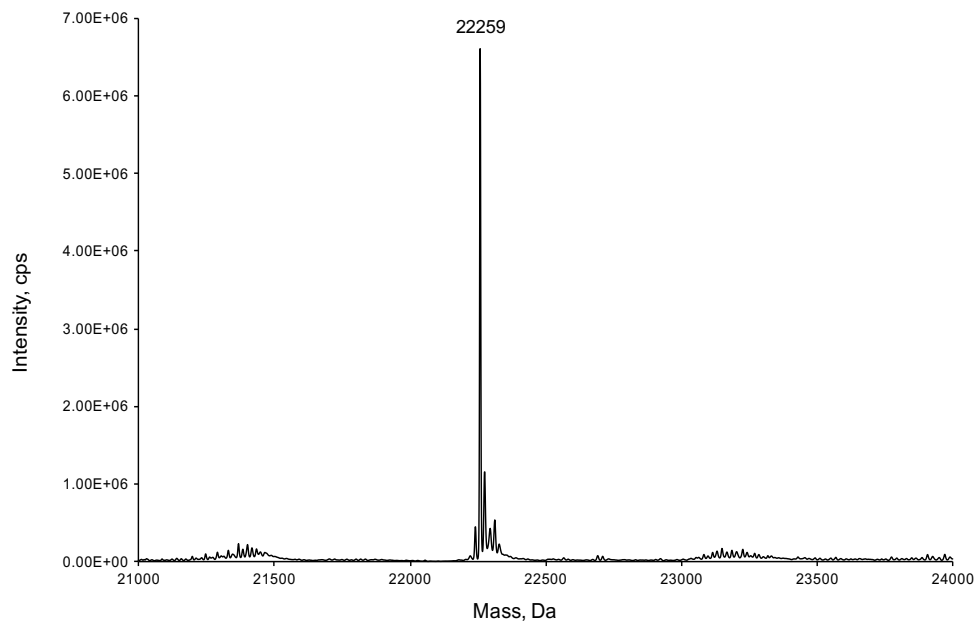
Full-length Ferritin-L56TAG-2 was produced by using PylRS•tRNA<sup>Pyl</sup><sub>CUA</sub> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 22,260 Da (protein reducing first methionine); observed molecular mass is 22,259 Da.



(A)



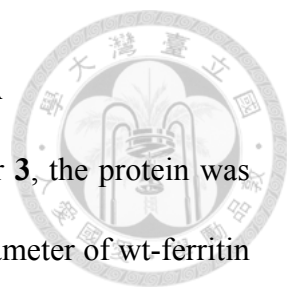
(B)



**Figure 68. Molecular mass determination of the protein Ferritin-L56TAG-3**

Full-length Ferritin-L56TAG-3 was produced by using PylRS• tRNA<sup>Pyl</sup><sub>CUA</sub> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum of (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 22,260 Da (protein reducing first methionine); observed molecular mass is 22,259 Da.

### 3.4.1 Shape and biophysical property of Ferritin with D/LFA



To correct the shape of self-assembling ferritin-L56TAG-2 or **3**, the protein was analyzed by TEM (**Figure 69**). The picture exhibited the exterior diameter of wt-ferritin and ferritin variants were similar: wt-ferritin was  $14 \pm 2$  nm, ferritin-L56TAG-2 and **3** were  $13.7 \pm 1.6$  nm. The appearances were also alike. In our study, the protein cage was decided to encapsulate protein or drug, however wt-ferritin is very stable and disassemble until the condition lower than pH 2.0,<sup>31</sup> but rare protein can tolerate. Thus, the ferritin variants were transferred into different pH buffer (pH 4.0, 8.0, and 10.0) and measured the particles size by DLS. The profile demonstrated ferritin-L56TAG-2 and **3** appeared obviously size change in pH 4.0, which was not observed in wt-ferritin (**Figure 70**). Moreover, the particle was analyzed by TEM, and the picture exhibited the same trend with the result of DLS. The self-assembly particle can be observed in all variants, but ferritin-L56TAG-2 and **3** appeared some disassembly and aggregation (**Figure 71**).

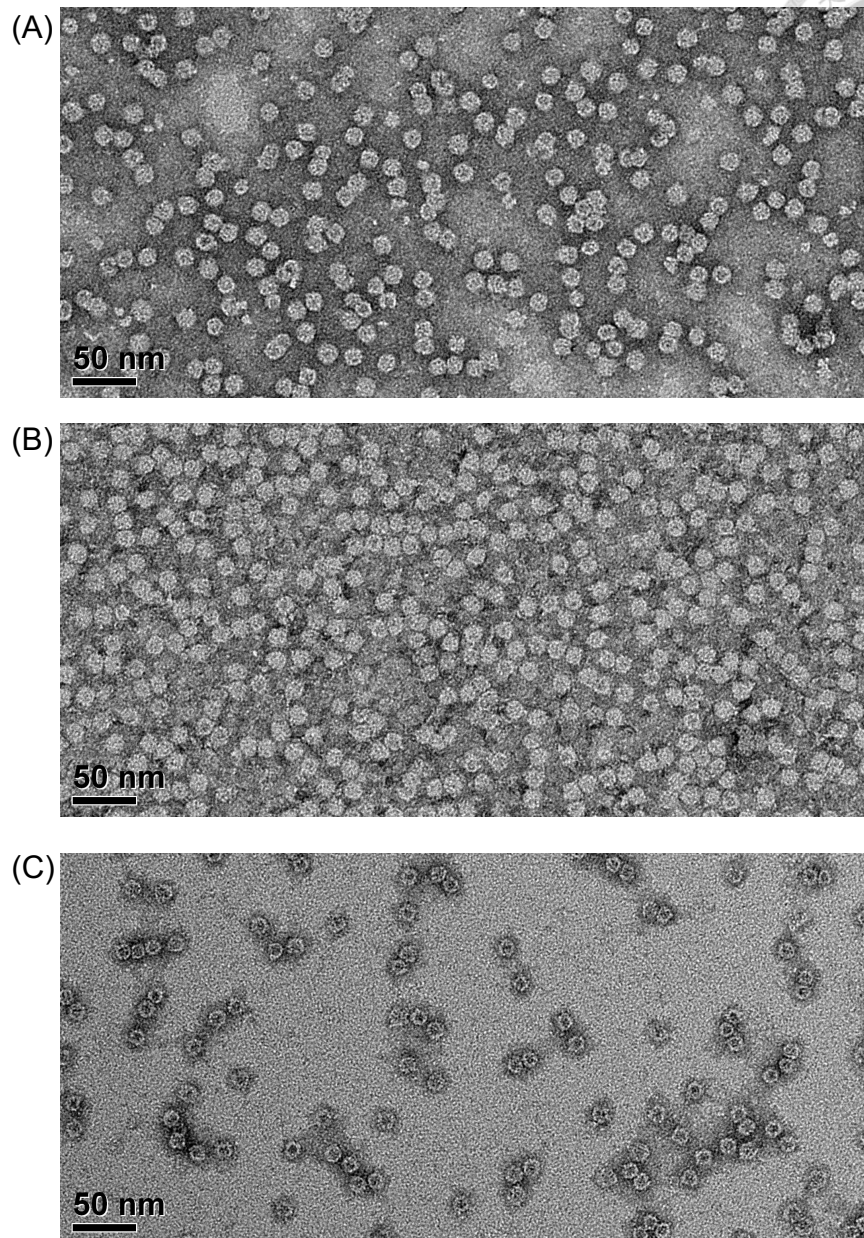
To look deeper of influences of pH values, the ferritin variants was dissolved in pH 2 to 12 buffer. The result demonstrated that ferritin-L56-2, **3**, and wt-ferritin were disassembly at pH 3.0 and pH 2.0 (**Figure S9-S10**). Contentiously, the ferritin extreme pH was analyzed by native gel, the gel revealed that ferritin variants at pH 3.0 exhibited partial disassembly and the molar weight higher than 66 kDa. In addition, the protein at pH 2.0 appeared aggregation, it can be found on the top of well. (**Figure S11**)

In addition, the protein cages were also design to incorporate with different ncAAs (**10** and **11**) containing functional group, azido group and alkyne group, for click chemistry. And the protein was purified and analyzed by ESI-MS. (**Figure S12-S13**). On

the other hand, the ncAA **12** was synthesized for decorating protein cage (**Figure S14**), and the product (**12**) was analyzed by NMR (**Figure S15**).

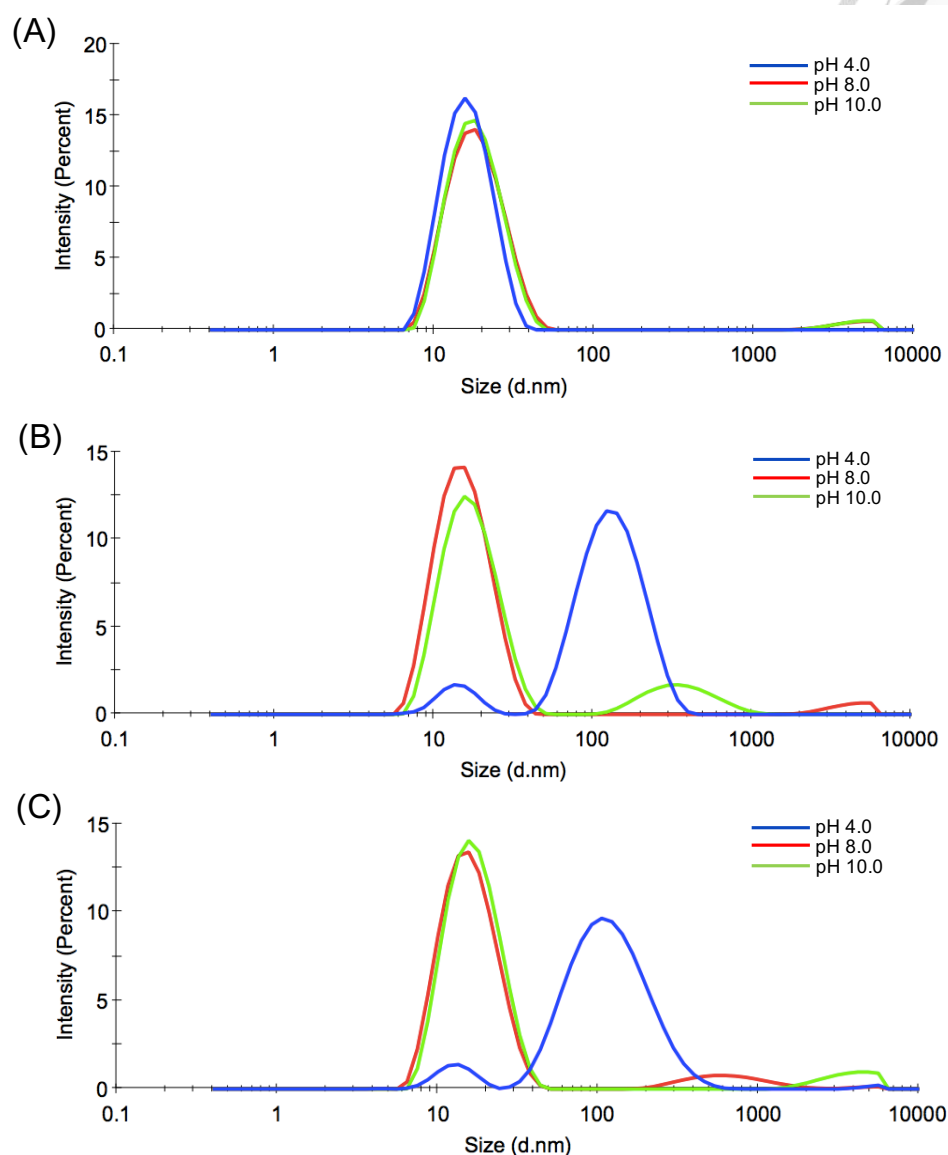






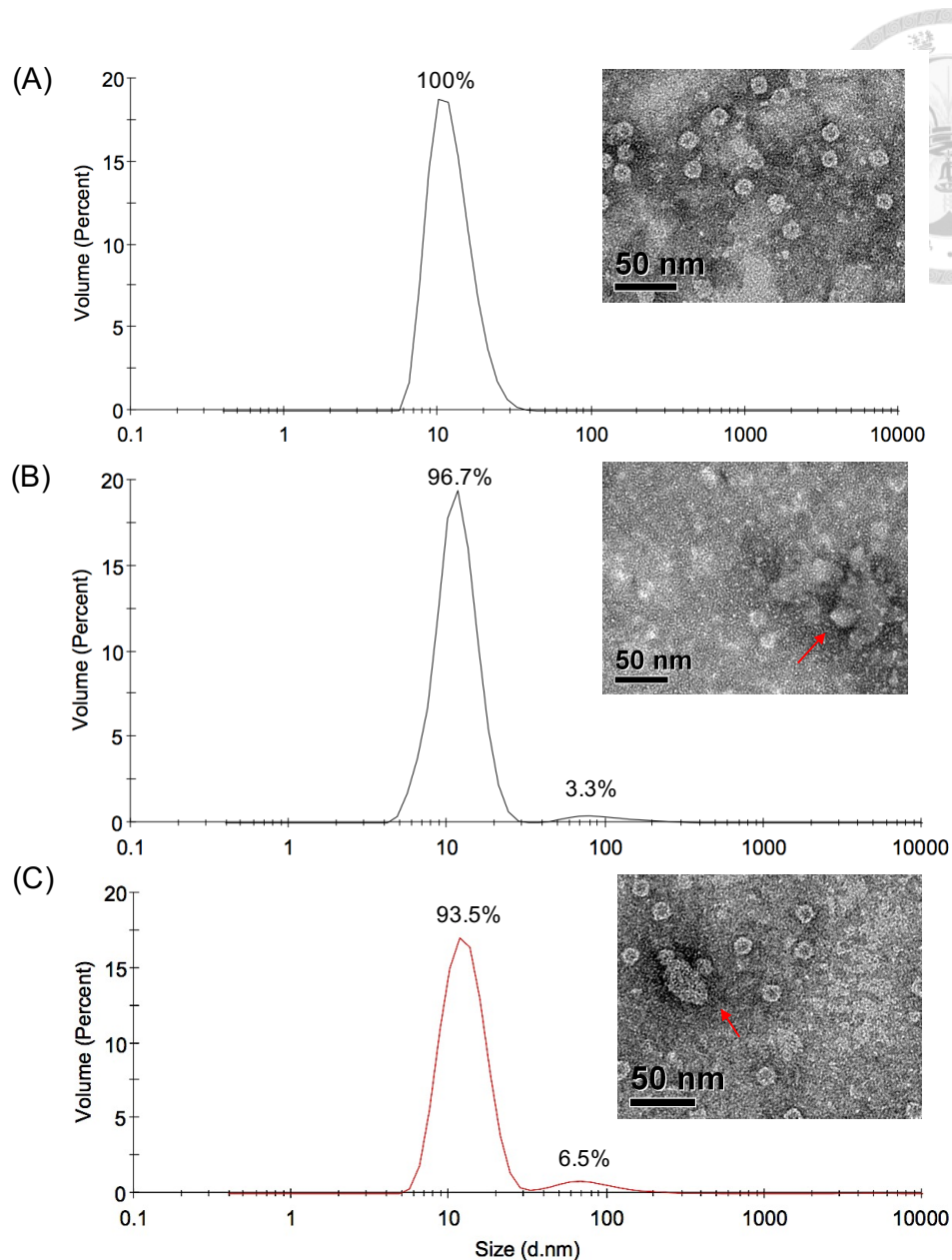
**Figure 69. TEM of Ferritin variants.**

The concentration of (A) Ferritin-L56TAG-2, (B) Ferritin-L56TAG-3, and (C) wt-Ferritin were 75 mg/ml, 100 mg/ml, and 50 mg/ml, respectively. The protein dissolved in 50 mM Tris-HCl at pH 8.0. The exterior appearance of all ferritin variants are like the circular. The size of (A) and (B) were  $13.7 \pm 1.6$  nm, (C) was  $14 \pm 2$  nm. (n = 50)



**Figure 70. DLS analysis of Ferritin variants at pH 4.0, 8.0, and 10.0.**

The ferritin variants were analyzed by DLS and the concentration of proteins were 0.5 mg/ml in buffer at pH 4, 8 or 10. The spectra was measure by scattering light intensity of the scattering particles versus their diameter (nm). The diameter of (A) wild type ferritin was about  $12 \pm 1.5$  nm at pH 4.0, 8.0, and 10.0. However, the diameter of (B) Ferritin-L56TAG-2 and (C) Ferritin-L56TAG-3 were  $12 \pm 1.5$  nm at pH 8.0 and 10.0 but significantly alter to  $100 \pm 10$  nm at pH 4.0. (blue line: at pH 4.0; red line: at pH 8.0; green line: at pH 10.0)



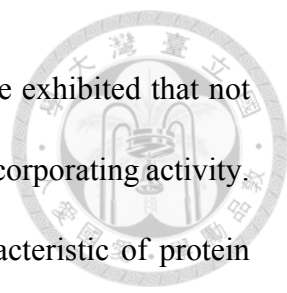
**Figure 71. Analyzed size distribution of ferritin variants at pH 4.0 by volume.**

The ferritin variants were analyzed by DLS and the concentration of proteins were 0.5 mg/ml in buffer at pH 4.0. The spectra were measure by scattering light intensity of the scattering particles versus their diameter (nm) and transferred to volume value by Zetasizer software. The size of (A) wt-ferritin maintaining 12 nm, (B) Ferritin-L56TAG-2 including 12 nm (96.7%) and 109 nm (3.3%), (C) Ferritin-L56TAG-3 including 12 nm (93.5%) and 88 nm (6.5%). The TEM picture were relative to DLS analysis in each group at pH 4.0.

## Chapter 4 Discussion

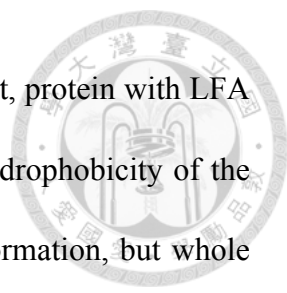


Expanding genetic codon is an essential technology in protein decoration, not only increasing the function but also providing structural insight of protein. The bioorthogonal AARS•tRNA pair was introduced into *E. coli* and responsible for genetically incorporating ncAAs in protein. In this study, *MmPylRS* was engineered to extend the substrate range from L-ncAA to D-ncAA. Firstly, the catalytic pocket of PylRS was modified at N346 position which interact with **1** (Pyl) by nitrogen on side chain and sterically clash aromatic amino acid in active side.<sup>53</sup> After mutation Asn with Val at position 346, naming LFRS, the screening result demonstrated enzyme not only decrease the binding with **1** but also increase the activity for aromatic amino acid, but it only recognizes L-ncAA. In order to release chiral selectivity of PylRS variant. More mutations at catalytic pocket was designed. Residue C348 have revealed that toward catalytic pocket of PylRS and mutated to a larger amino acid, Gln in our study, reducing the size of the active site and more adapting to aromatic amino acid.<sup>63</sup> In addition, we suggested that the residue V401 was mutated to a small one, Gly, to enlarge empty space at the active site around the *meta* position of phenylalanine. Thus, PylRS was logically designed with N346G/C348Q/V401G mutation, naming DFRS, and screening with ncAA library, result demonstrated DFRS enhance the efficiency of incorporation to *meta*-phenylalanine variants which was both LFA (**3**) and DFA (**2**). However, the incorporation efficiency of DFA was still poor, so more mutations were designed at NTD. Because the published paper illustrated that NTD of PylRS interact with  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  and mutant at relative residue can increase the activity of enzyme.<sup>41</sup> Therefore, N-DFRS, *MmPylRS* with mutations at NTD (R61K/H63Y/S193R) and in the catalytic pocket



(N346G/C348Q/V401G) was screened, the profile of substrate range exhibited that not only LFA **3**, **5**, and **7** but also their isomer DFA **2**, **4**, and **6** had high incorporating activity. Thus, we focused on exploring the structural and biophysical characteristic of protein incorporated with these three pairs enantiomers (**Figure 16**).

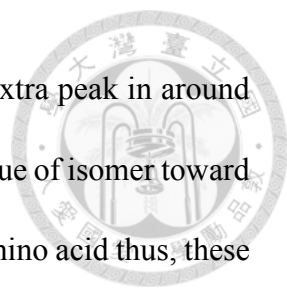
Apart from incorporating ncAAs into sfGFP-S2TAG, PylRS variants with sfGFP-F27TAG was screened with ncAAs library. Interestingly, the profile of substrate range demonstrated Phe variants more preferred F27 position, more fluorescent signals of ncAAs incorporating were detected in PylRS variants/sfGFP-F27TAG group which cannot be observed in sfGFP-S2TAG group. Thus, we suggested that steric effects influence the incorporating efficiency of ncAAs. To confirm our hypothesis, sfGFP with amber mutation at different position S2, F27 or F130 was designed and over-expressed with N-DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair and **2-7**. The result revealed that all of them were with different pattern in fluorescence intensity, meaning the preference of **2-7** in different position are variable (**Figure 23**), and sfGFP-F27TAG had higher incorporating efficiency. We suggest the position F27 of sfGFP in the  $\beta$ -barrel is more suitable for aromatic amino acid, because the residue around position 27 naturally interacted with Phe to form secondary structure. In addition, sfGFP was also mutated with three TAG to confirm that N-DFRS•tRNA<sub>CUA</sub><sup>Pyl</sup> activity and conformation effect. The fluorescence intensity and western blot primarily revealed that sfGFP-3xTAG was successfully incorporated with D/LFA and without obviously conformational disruption, because proteins were still detected fluorescence. Comparing western blot (**Figure 24**) with fluorescence intensity (**Figure 23**), particularly, sfGFP-3xTAG-**2** and sfGFP-3xTAG-**3**



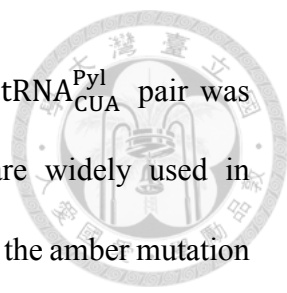
had similar suppression yield but fluorescence intensity was different, protein with LFA higher than with DFA. We suggest the fluorescence is based on hydrophobicity of the core, however, protein with DFA has a little bit influence on conformation, but whole structure still maintain, so causing decrease of the fluorescence intensity.

Besides, some differences in catalytic activity of **4** and its isomer **5** was noticed. In profile of fluorescence intensity, the all sfGFP variants with **4** is lower than **5**. And sfGFP-3xTAG-**4** cannot be detected in both western blot and fluorescence intensity. All results strongly suggested N-DFRS present diverse activity in **4** and **5**. In order to realize the detail of catalytic process, the catalytic pocket of N-DFRS was crystallized and then soaking with **2-7**. Attractively, after X-ray analysis, all LFA in catalytic pocket presented clear electron density and very similar binding mode to each other, however, DFA without clear electron density, which suggest they do not have very stable binding mode (**Figure 43**). In contrast, **4** have relatively stronger electron density in the perpendicular binding mode comparing with **5**, about 95 degrees, and studying with fluorescent intensity, **4** is lower than others (**Figure 45**). It suggests perpendicular binding mode actually does not contribute to the activity and which also means in the active site have more space in other binding mode. But the activity come from the unstable binding of DFA in the active site.

To realize the properties of protein containing DFA. the amber mutation was designed at Y66 position of sfGFP in chromophore and express with **2-7**. Especially, the sfGFP-Y66TAG-**4** and **5** with same molecular weight exhibited various emission



spectrum (**Figure 54**). the spectra of sfGFP-Y66TAG-**4** present an extra peak in around 510 nm, but it cannot find in sfGFP-Y66TAG-**5**. We suggest the residue of isomer toward different direction, causing the variously interaction with surround amino acid thus, these two protein possess distinct spectra. In addition, in the intrinsic fluorescence spectrum, different appearance was observed in sfGFP-Y66TAG-**4** and -**5** (**Figure 57**). The sfGFP-Y66TAG-**5** exhibited obviously FRET from wavelength 330 nm to 420 nm, however, it cannot be observed in sfGFP-Y66TAG-**4**. It suggests that residue 66 of sfGFP-Y66TAG-**5** is more close to the Trp57 than sfGFP-Y66TAG-**4**, thus, the energy of intrinsic fluorescence can transfer to **5** but cannot arrive **4** in 66 position. In order to more study the structural property of protein with DFA, the thermally induced conformation change was measured by DSC and CD for 25°C to 95 °C. In the DSC data, protein with different kind of enantiomers present different  $T_m$  (**Figure 59**). The sfGFP-Y66TAG-D/LFA with bigger substitution (**2** and **3**) exhibited the lower  $T_m$  which suggest the steric hindrance of **2** and **3** in protein are stronger than others, therefore, the thermal stability of these protein is slightly lower than others. On the order hand, The CD data revealed the details of thermally induced conformation change in sfGFP-Y66TAG with **4** and **5**. The negative signal at wavelength 218 nm is a distinguishing characteristic of  $\beta$ -sheet. Comparing the signal change of sfGFP variants at 218 nm, the All L-protein exhibited more negative signal, in contrast, the D and L mixture protein presented more positive signal after heating over 65°C (**Figure 63**). We suggest the direction of residue (**4** and **5**) cause the probability of interaction with different side in protein while heating and appear the different signals in CD spectra at 218 nm.



To explore the application of D/LFA into protein. N-PylRS•tRNA<sup>Pyl</sup><sub>CUA</sub> pair was used to incorporate **2** and **3** into cage protein, ferritin which are widely used in bionanotechnology applications. To make ferritin easier to dissociate, the amber mutation was designed at L56 position which is in  $C_2$  interface of ferritin. The position L56 was demonstrated that is the essential residue for protein assembly.<sup>40</sup> After TEM analysis, the ferritin variants was confirmed that they still maintain the ability of self-assembly without apparently shaped change. It means ferritin variants still can as the container for encapsulating the cargo and without side effect of ferritin with D/LFA on cage building. Moreover, the DLS data revealed that Ferritin-L56TAG-**2** and **3** appear conformational change in pH 4.0, although ferritin with LFA or DFA without obviously difference. It suggests that the ncAA at L56 position plays the important role in protein cage assembly. This phenomenon is essential for encapsulating weak acid tolerant protein and releasing drug in cancer cell, through sensing pH decreases in the acidic endocytic compartment such as endosomes (pH5-6) and lysosomes (pH4-5).<sup>37</sup>



## Chapter 5 Conclusion



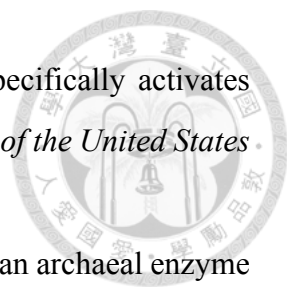
In our study, *MmPylRS* variants N-DFRS with N-terminus and active site mutations generates activity in charging both D and L form *meta*-substituted phenylalanine with high efficiency. After more exploring in catalytic process of N-DFRS, the crystal structures of N-DFRSc270 soaking with DFA and LFA reveal different binding modes. Dynamic movement of DFA in the N-DFRS engineered binding pocket provides higher opportunity in reacting with ATP cofactor and result in great activity. All N-FRSc270/LFA co-crystal structures present similar and stable and classical binding mode. Moreover, photophysical property of sfGFP with DFA or LFA mutation at chromophore Y66 position revealed different fluorescent spectrum. sfGFP protein with D/LFA possessed same molecule weight but sfGFP containing DFA appeared another peak in intrinsic fluorescence spectrum. It suggests D-aa may display curial role in studying protein photochemical property, protein stability, and protein design. The CD spectrum measured the thermally induced structural change in protein with D/LFA, comparing all L-protein with D and L mixture protein exhibited different CD signal at heating. It may suggest the contrasting direction of isomer cause interaction with different side of protein.

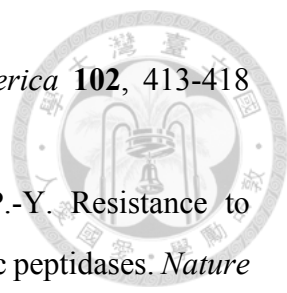
In utilizing, protein was incorporate with D/LFA at L56 position, and all of them reveal acid-induced conformation change in pH4.0, and protein contained DFA still can self-assembly without disruption. This finding displays an essential onset for protein cage design. In the future, more site at  $C_2$  interface will be modified with D-aa or other ncAAs to explore the protein (dis)assembly and enzymatic stability.

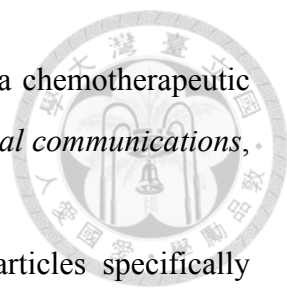
## Reference

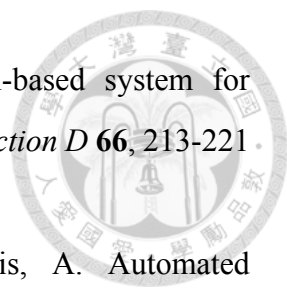


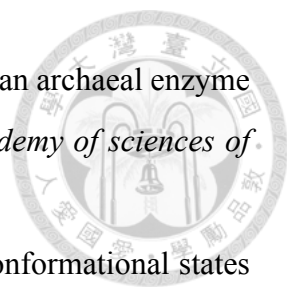
- 1 Crick, F. H. The origin of the genetic code. *Journal of molecular biology* **38**, 367-379 (1968).
- 2 Sonenberg, N. & Dever, T. E. Eukaryotic translation initiation factors and regulators. *Current opinion in structural biology* **13**, 56-63 (2003).
- 3 Andersen, G. R., Nissen, P. & Nyborg, J. Elongation factors in protein biosynthesis. *Trends in biochemical sciences* **28**, 434-441 (2003).
- 4 Petry, S. *et al.* Crystal structures of the ribosome in complex with release factors RF1 and RF2 bound to a cognate stop codon. *Cell* **123**, 1255-1266 (2005).
- 5 Grentzmann, G. *et al.* Release factor RF-3 GTPase activity acts in disassembly of the ribosome termination complex. *Rna* **4**, 973-983 (1998).
- 6 Cohen, G. & Munier, R. Incorporation of structural analogues of amino acids in bacterial proteins. *Biochimica et biophysica acta* **21**, 592 (1956).
- 7 Johnson, J. A., Lu, Y. Y., Van Deventer, J. A. & Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Current opinion in chemical biology* **14**, 774-780 (2010).
- 8 Singh-Blom, A., Hughes, R. A. & Ellington, A. D. Residue-specific incorporation of unnatural amino acids into proteins in vitro and in vivo. *Methods in molecular biology* **978**, 93-114 (2013).
- 9 Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182-188 (1989).
- 10 Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annual review of biophysics and biomolecular structure* **35**, 225-249 (2006).
- 11 James, C. M., Ferguson, T. K., Leykam, J. F. & Krzycki, J. A. The amber codon in the gene encoding the monomethylamine methyltransferase isolated from *Methanosarcina barkeri* is translated as a sense codon. *Journal of biological chemistry* **276**, 34252-34258 (2001).

- 
- 12 Polycarpo, C. *et al.* An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proceedings of the national academy of sciences of the United States of America* **101**, 12450-12454 (2004).
- 13 Kavran, J. M. *et al.* Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proceedings of the national academy of sciences of the United States of America* **104**, 11268-11273 (2007).
- 14 Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding Nε-acetyllysine in recombinant proteins. *Nature chemical biology* **4**, 232-234 (2008).
- 15 Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annual review of biochemistry* **79**, 413-444 (2010).
- 16 O'donoghue, P., Ling, J., Wang, Y.-S. & Söll, D. Upgrading protein synthesis for synthetic biology. *Nature chemical biology* **9**, 594-598 (2013).
- 17 Lipmann, F., Hotchkiss, R. D. & Dubos, R. J. The occurrence of D-amino acids in gramicidin and tyrocidine. *Journal of biological chemistry* **141**, 163-169 (1941).
- 18 Montecucchi, P. C., Castiglione, R. & Erspamer, V. Identification of dermorphin and hyp6-dermorphin in skin extracts of the brazilian frog phyllomedusa rhodei. *Chemical biology & Drug design* **17**, 316-321 (1981).
- 19 Heck, S. D. *et al.* Functional consequences of posttranslational isomerization of Ser46 in a calcium channel toxin. *Science* **266**, 1065-1068 (1994).
- 20 Kreil, G. D-amino acids in animal peptides. *Annual review of biochemistry* **66**, 337-345 (1997).
- 21 Achenbach, J. *et al.* Outwitting EF-Tu and the ribosome: translation with d-amino acids. *Nucleic acids research* **43**, 5687-5698 (2015).
- 22 Katoh, T., Iwane, Y. & Suga, H. Logical engineering of D-arm and T-stem of tRNA that enhances d-amino acid incorporation. *Nucleic acids research* **45**, 12601-12610 (2017).
- 23 Chen, S. *et al.* Improving Binding Affinity and Stability of Peptide Ligands by Substituting Glycines with D-Amino Acids. *Chembiochem* **14**, 1316-1322 (2013).
- 24 Tugyi, R. *et al.* Partial D-amino acid substitution: Improved enzymatic stability and preserved Ab recognition of a MUC2 epitope peptide. *Proceedings of the*

- 
- National academy of sciences of the United States of America* **102**, 413-418 (2005).
- 25 Li, Y.-X., Zhong, Z., Hou, P., Zhang, W.-P. & Qian, P.-Y. Resistance to nonribosomal peptide antibiotics mediated by D-stereospecific peptidases. *Nature chemical biology* **14**, 381 (2018).
- 26 Bang, D. *et al.* Dissecting the energetics of protein alpha-helix C-cap termination through chemical protein synthesis. *Nature chemical biology* **2**, 139-143 (2006).
- 27 Granick, S. & Michaelis, L. Ferritin and apoferritin. *Science* **95**, 439-440 (1942).
- 28 Harrison, P. M. & Arosio, P. The ferritins: molecular properties, iron storage function and cellular regulation. *Biochimica et biophysica acta (BBA)-Bioenergetics* **1275**, 161-203 (1996).
- 29 Theil, E. C. Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. *Annual review of biochemistry* **56**, 289-315 (1987).
- 30 Bernacchioni, C., Ghini, V., Theil, E. & Turano, P. Modulating the permeability of ferritin channels. *RSC Advances* **6**, 21219-21227 (2016).
- 31 Kim, M. *et al.* pH-dependent structures of ferritin and apoferritin in solution: disassembly and reassembly. *Biomacromolecules* **12**, 1629-1640 (2011).
- 32 Ma-Ham, A. *et al.* Apoferritin-based nanomedicine platform for drug delivery: equilibrium binding study of daunomycin with DNA. *Journal of materials chemistry* **21**, 8700-8708 (2011).
- 33 Maity, B., Abe, S. & Ueno, T. Observation of gold sub-nanocluster nucleation within a crystalline protein cage. *Nature communications* **8**, 14820 (2017).
- 34 Domínguez-Vera, J. M. & Colacio, E. Nanoparticles of Prussian blue ferritin: a new route for obtaining nanomaterials. *Inorganic chemistry* **42**, 6983-6985 (2003).
- 35 Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102 (2013).
- 36 Bhaskar, S. & Lim, S. Engineering protein nanocages as carriers for biomedical applications. *NPG Asia materials* **9**, e371 (2017).

- 
- 37 Flenniken, M. L. *et al.* Selective attachment and release of a chemotherapeutic agent from the interior of a protein cage architecture. *Chemical communications*, 447-449 (2005).
- 38 Liang, M. *et al.* H-ferritin–nanocaged doxorubicin nanoparticles specifically target and kill tumors with a single-dose injection. *Proceedings of the national academy of sciences of the United States of America* **111**, 14900-14905 (2014).
- 39 Choi, S.-H., Choi, K., Kwon, I. C. & Ahn, H. J. The incorporation of GALA peptide into a protein cage for an acid-inducible molecular switch. *Biomaterials* **31**, 5191-5198 (2010).
- 40 Huard, D. J., Kane, K. M. & Tezcan, F. A. Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. *Nature chemical biology* **9**, 169 (2013).
- 41 Suzuki, T. *et al.* Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nature chemical biology* **13**, 1261-1266 (2017).
- 42 Englert, M. *et al.* Probing the active site tryptophan of *Staphylococcus aureus* thioredoxin with an analog. *Nucleic acids research* **43**, 11061-11067 (2015).
- 43 Yanagisawa, T., Ishii, R., Fukunaga, R., Nureki, O. & Yokoyama, S. Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon *Methanosarcina mazei*. *Acta crystallographica section F: structural biology communications* **62**, 1031-1033 (2006).
- 44 Ko, J.-h. *et al.* Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function. *FEBS letters* **587**, 3243-3248 (2013).
- 45 Fan, C., Ho, J. M., Chirathivat, N., Söll, D. & Wang, Y. S. Exploring the Substrate Range of Wild-Type Aminoacyl-tRNA Synthetases. *ChemBioChem* **15**, 1805-1809 (2014).
- 46 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in enzymology* **276**, 307-326 (1997).

- 
- 47 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica section D* **66**, 213-221 (2010).
- 48 Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature protocols* **3**, 1171-1179 (2008).
- 49 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica section D* **66**, 486-501 (2010).
- 50 Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* **26**, 283-291 (1993).
- 51 Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols* **1**, 2856 (2007).
- 52 Tharp, J. M., Wang, Y.-S., Lee, Y.-J., Yang, Y. & Liu, W. R. Genetic incorporation of seven ortho-substituted phenylalanine derivatives. *ACS chemical biology* **9**, 884-890 (2014).
- 53 Wang, Y.-S., Fang, X., Wallace, A. L., Wu, B. & Liu, W. R. A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *Journal of the American chemical society* **134**, 2950-2953 (2012).
- 54 Wang, Y.-S. *et al.* Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. *ACS chemical biology* **8**, 405-415 (2012).
- 55 Katoh, T., Tajima, K. & Suga, H. Consecutive Elongation of D-Amino Acids in Translation. *Cell chemical biology* **24**, 46-54 (2017).
- 56 Lam, H. *et al.* D-amino acids govern stationary phase cell wall remodeling in bacteria. *Science* **325**, 1552-1555 (2009).
- 57 Guo, L. T. *et al.* Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proceedings of the national academy of sciences of the United States of America* **111**, 16724-16729 (2014).

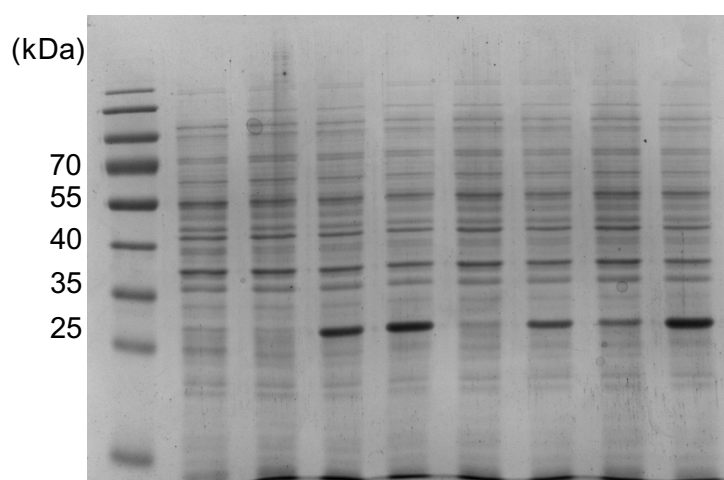
- 
- 58 Kavran, J. M. *et al.* Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation *Proceedings of the national academy of sciences of the United States of America* **104**, 11268-11273 (2007).
- 59 Yanagisawa, T. *et al.* Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *Journal of molecular biology* **378**, 634-652 (2008).
- 60 Ma, H., Liu, N., Shi, S., Wang, S. & Chen, Y. Genetic incorporation of D-amino acids into green fluorescent protein based on polysubstrate specificity. *RSC Advances* **5**, 39580-39586 (2015).
- 61 Heim, R., Prasher, D. C. & Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the national academy of sciences of the United States of America* **91**, 12501-12504 (1994).
- 62 Chalfie, M. Green fluorescent protein. *Photochemistry and photobiology* **62**, 651-656 (1995).
- 63 Wang, Y.-S. *et al.* The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of L-phenylalanine and its derivatives. *Molecular bioSystems* **7**, 714-717 (2011).
- 64 Hounsou, C. *et al.* Time-resolved FRET binding assay to investigate hetero-oligomer binding properties: proof of concept with dopamine D1/D3 heterodimer. *ACS Chemical biology* **10**, 466-474 (2014).
- 65 Nguyen, D. P. *et al.* Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNACUA pair and click chemistry. *Journal of the American chemical society* **131**, 8720-8721 (2009).

## Appendix



(A)

	IPTG	-	+	+	+	+	+	+
DmCF <sub>3</sub> Phe, <b>2</b>	-	-	+	-	-	-	-	-
LmCF <sub>3</sub> Phe, <b>3</b>	-	-	-	+	-	-	-	-
DmClPhe, <b>4</b>	-	-	-	-	+	-	-	-
LmClPhe, <b>5</b>	-	-	-	-	-	+	-	-
DmBrPhe, <b>6</b>	-	-	-	-	-	-	+	-
LmBrPhe, <b>7</b>	-	-	-	-	-	-	-	+

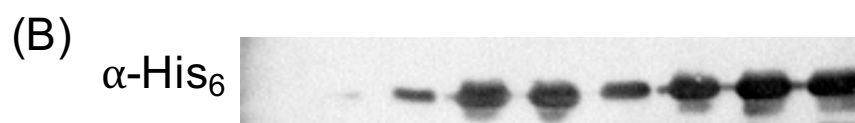
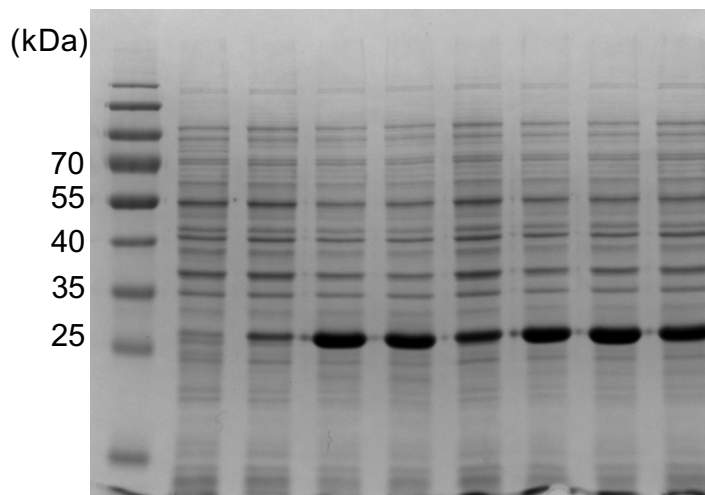


**Figure S1. Amber suppression of D/LFA in sfGFP-S2TAG.**

*E. coli* BL21(DE3) cells coding DFRS, tRNA<sup>Pyl</sup><sub>CUA</sub> and a sfGFP gene with an amber codon (TAG) at position S2 containing His-tag in 3' end and under control of an inducible T7 promoter were grown in M9 medium supplemented with 1mM IPTG and 1mM ncAAs. (A) InstantBlue-Stained gel of whole cell over-expressing sfGFP-S2TAG in the absence and in the presence **2-7**. First lane was negative control without ncAA and IPTG. (B) Western blot of lysate from whole cell over-expressing sfGFP-S2TAG-**2** to **7** with an anti-His tag antibody confirms that the DFRS, tRNA<sup>Pyl</sup><sub>CUA</sub> pair incorporates **2-7** into protein with TAG codon in *E. coli* Cell.



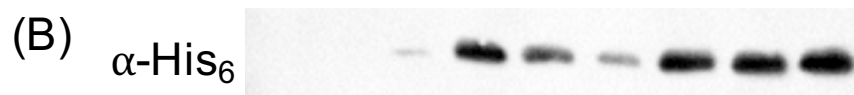
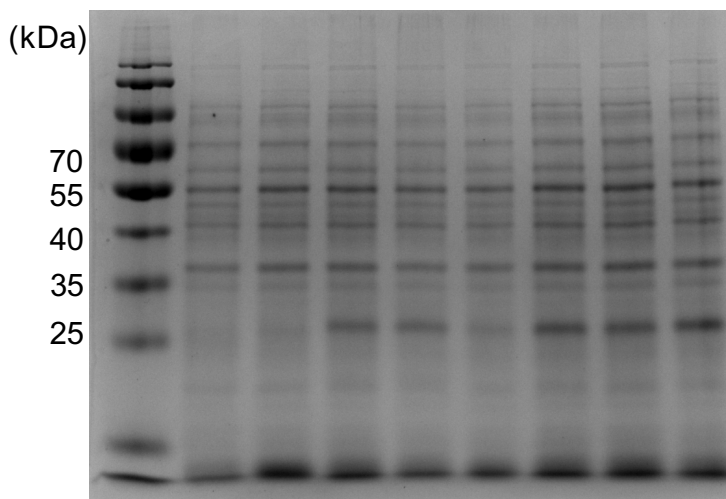
(A)	IPTG	-	+	+	+	+	+	+
	DmCF <sub>3</sub> Phe, <b>2</b>	-	-	+	-	-	-	-
	LmCF <sub>3</sub> Phe, <b>3</b>	-	-	-	+	-	-	-
	DmClPhe, <b>4</b>	-	-	-	-	+	-	-
	LmClPhe, <b>5</b>	-	-	-	-	-	+	-
	DmBrPhe, <b>6</b>	-	-	-	-	-	-	+
	LmBrPhe, <b>7</b>	-	-	-	-	-	-	+



**Figure S2. Amber suppression of D/LFA in sfGFP-F27TAG.**

*E. coli* BL21(DE3) cells coding DFRS, tRNA<sup>Pyl</sup><sub>CUA</sub> and a sfGFP gene with an amber codon (TAG) at position F27 containing His-tag in 3' end and under control of an inducible T7 promoter were grown in M9 medium supplemented with 1mM IPTG and 1mM ncAAs. (A) InstantBlue-Stained gel of whole cell over-expressing sfGFP-F27TAG in the absence and in the presence **2-7**. First lane was negative control without ncAA and IPTG. (B) Western blot of lysate from whole cell over-expressing sfGFP-F27TAG-**2** to **7** with an anti-His tag antibody confirms that the DFRS, tRNA<sup>Pyl</sup><sub>CUA</sub> pair incorporates **2-7** into protein with TAG codon in *E. coli* Cell.

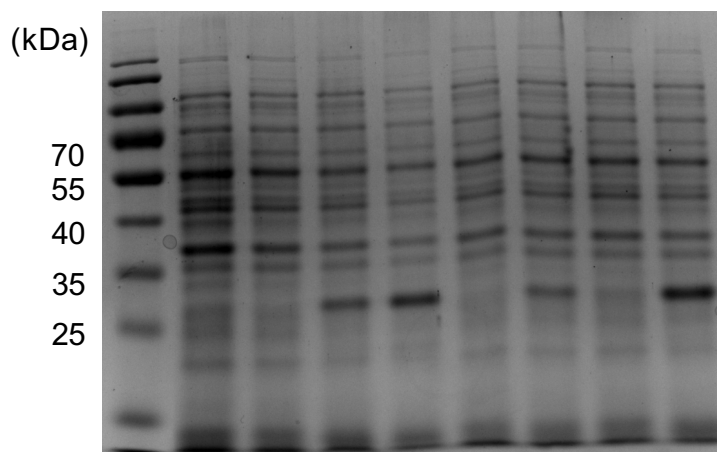
(A)	IPTG	-	+	+	+	+	+	+
<i>DmCF<sub>3</sub>Phe</i> , <b>2</b>	-	-	+	-	-	-	-	-
<i>LmCF<sub>3</sub>Phe</i> , <b>3</b>	-	-	-	+	-	-	-	-
<i>DmClPhe</i> , <b>4</b>	-	-	-	-	+	-	-	-
<i>LmClPhe</i> , <b>5</b>	-	-	-	-	-	+	-	-
<i>DmBrPhe</i> , <b>6</b>	-	-	-	-	-	-	+	-
<i>LmBrPhe</i> , <b>7</b>	-	-	-	-	-	-	-	+



**Figure S3. Amber suppression of D/LFA in sfGFP-F130TAG.**

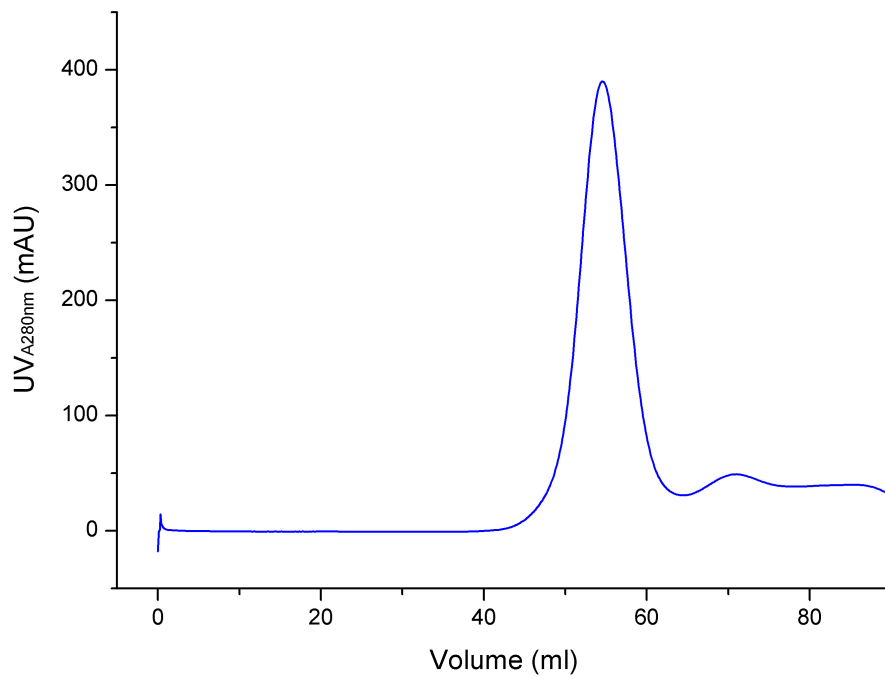
*E. coli* BL21(DE3) cells coding DFRS,  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  and a sfGFP gene with an amber codon (TAG) at position F130 containing His-tag in 3' end and under control of an inducible T7 promoter were grown in M9 medium supplemented with 1mM IPTG and 1mM ncAAs. (A) InstantBlue-Stained gel of whole cell over-expressing sfGFP-F130TAG in the absence and in the presence **2-7**. First lane was negative control without ncAA and IPTG. (B) Western blot of lysate from whole cell over-expressing sfGFP-F130TAG-**2 to 7** with an anti-His tag antibody confirms that the DFRS,  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  pair incorporates **2-7** into protein with TAG codon in *E. coli* Cell.

(A)	IPTG	-	+	+	+	+	+	+	+
DmCF <sub>3</sub> Phe, <b>2</b>	-	-	+	-	-	-	-	-	-
LmCF <sub>3</sub> Phe, <b>3</b>	-	-	-	+	-	-	-	-	-
DmClPhe, <b>4</b>	-	-	-	-	+	-	-	-	-
LmClPhe, <b>5</b>	-	-	-	-	-	+	-	-	-
DmBrPhe, <b>6</b>	-	-	-	-	-	-	+	-	-
LmBrPhe, <b>7</b>	-	-	-	-	-	-	-	+	-



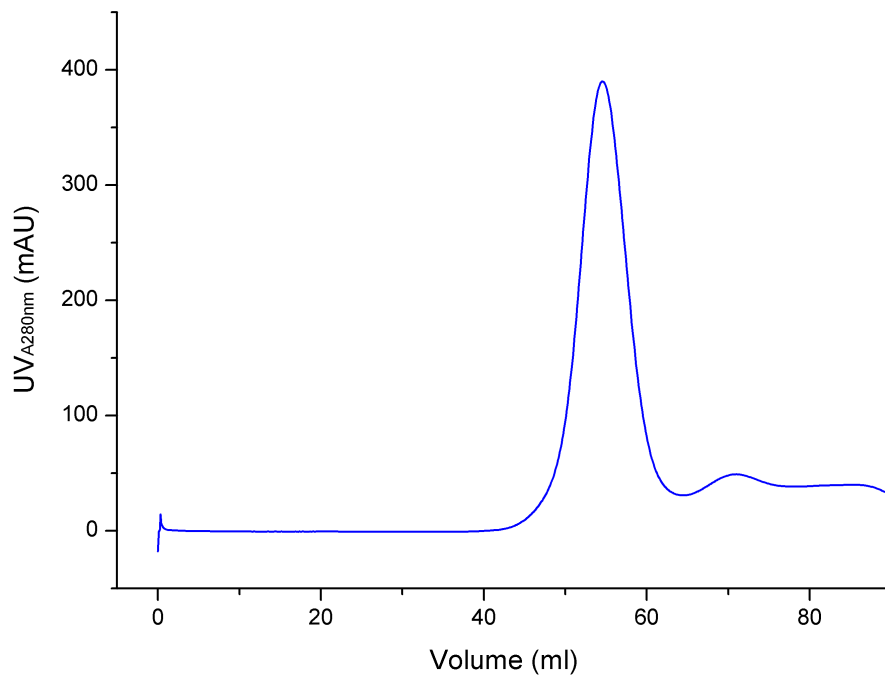
**Figure S4. Amber suppression of D/LFA in sfGFP-3xTAG.**

*E. coli* BL21(DE3) cells coding DFRS, tRNA<sup>Py1</sup><sub>CUA</sub> and a sfGFP gene with three amber codons (TAG) at position S2, F27 and F130 containing His-tag in 3' end and under control of an inducible T7 promoter were grown in M9 medium supplemented with 1mM IPTG and 1mM ncAAs. (A) InstantBlue-Stained gel of whole cell over-expressing sfGFP-3xTAG in the absence and in the presence **2-7**. First lane was negative control without ncAA and IPTG. (B) Western blot of lysate from whole cell over-expressing sfGFP-3xTAG-**2** to **7** with an anti-His tag antibody confirms that the DFRS, tRNA<sup>Py1</sup><sub>CUA</sub> pair incorporates **2-7** into protein with TAG codon in *E. coli* Cell.



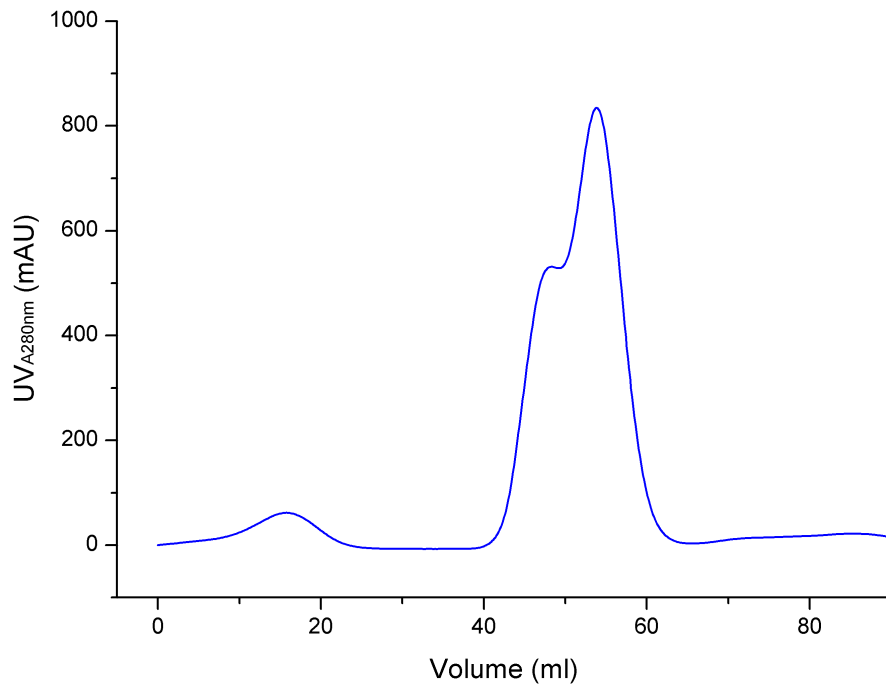
**Figure S5. Purification of Ferritin-L56TAG-2 with SEC.**

The elution after Ni<sup>2+</sup>-NTA column in 50 mM Tris, 100mM NaCl, and 150 mM imidazole was concentrated for FPLC with SEC column (HiLoad 16/600 superdex 200 pg, GE Healthare). One major peak appeared in the profile. Blue line: the signal of UV<sub>A280</sub>.



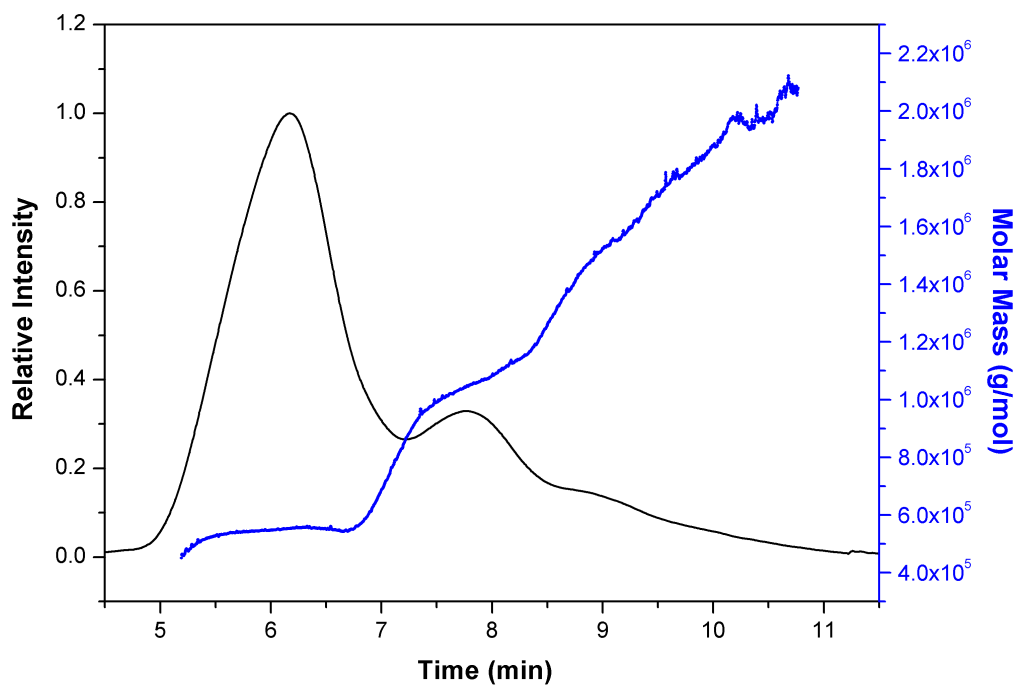
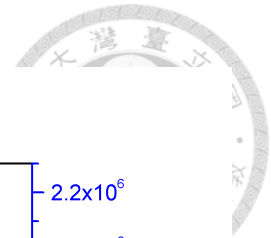
**Figure S6. Purification of Ferritin-L56TAG-3 with SEC.**

The elution after Ni<sup>2+</sup>-NTA column in 50 mM Tris, 100mM NaCl, and 150 mM imidazole was concentrated for FPLC with SEC column (HiLoad 16/600 superdex 200 pg, GE Healthare). One major peak appeared in the profile. Blue line: the signal of UV<sub>A280</sub>.



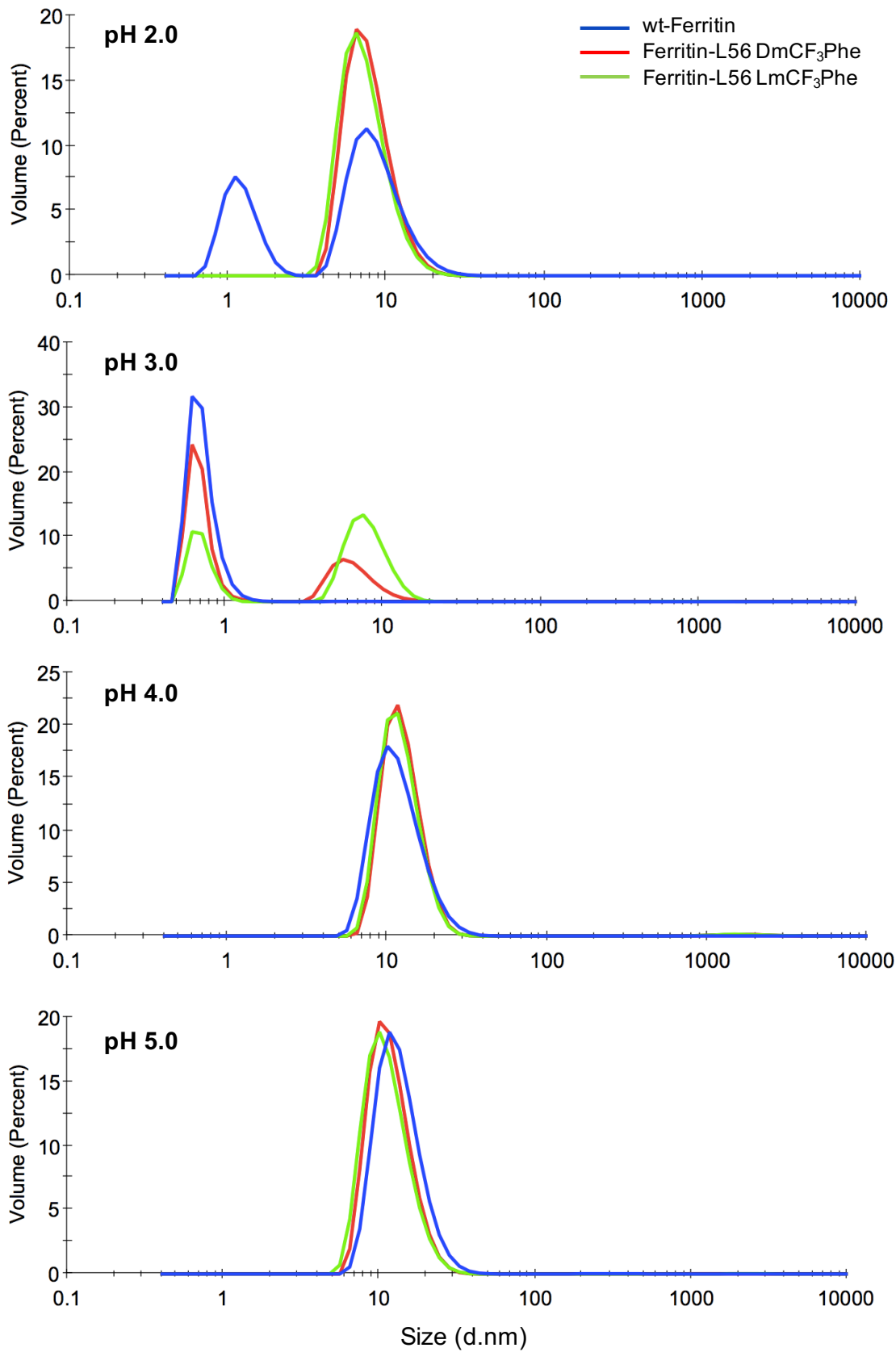
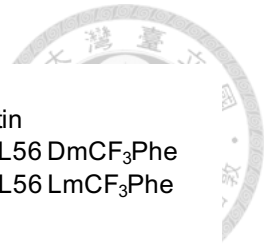
**Figure S7. Purification of wt-Ferritin with SEC.**

The elution after Ni<sup>2+</sup>-NTA column in 50 mM Tris, 100mM NaCl, and 150 mM imidazole was concentrated for FPLC with SEC column (HiLoad 16/600 superdex 200 pg, GE Healthare). Two peaks overlapped in the profile. Blue line: the signal of UV<sub>A280</sub>.

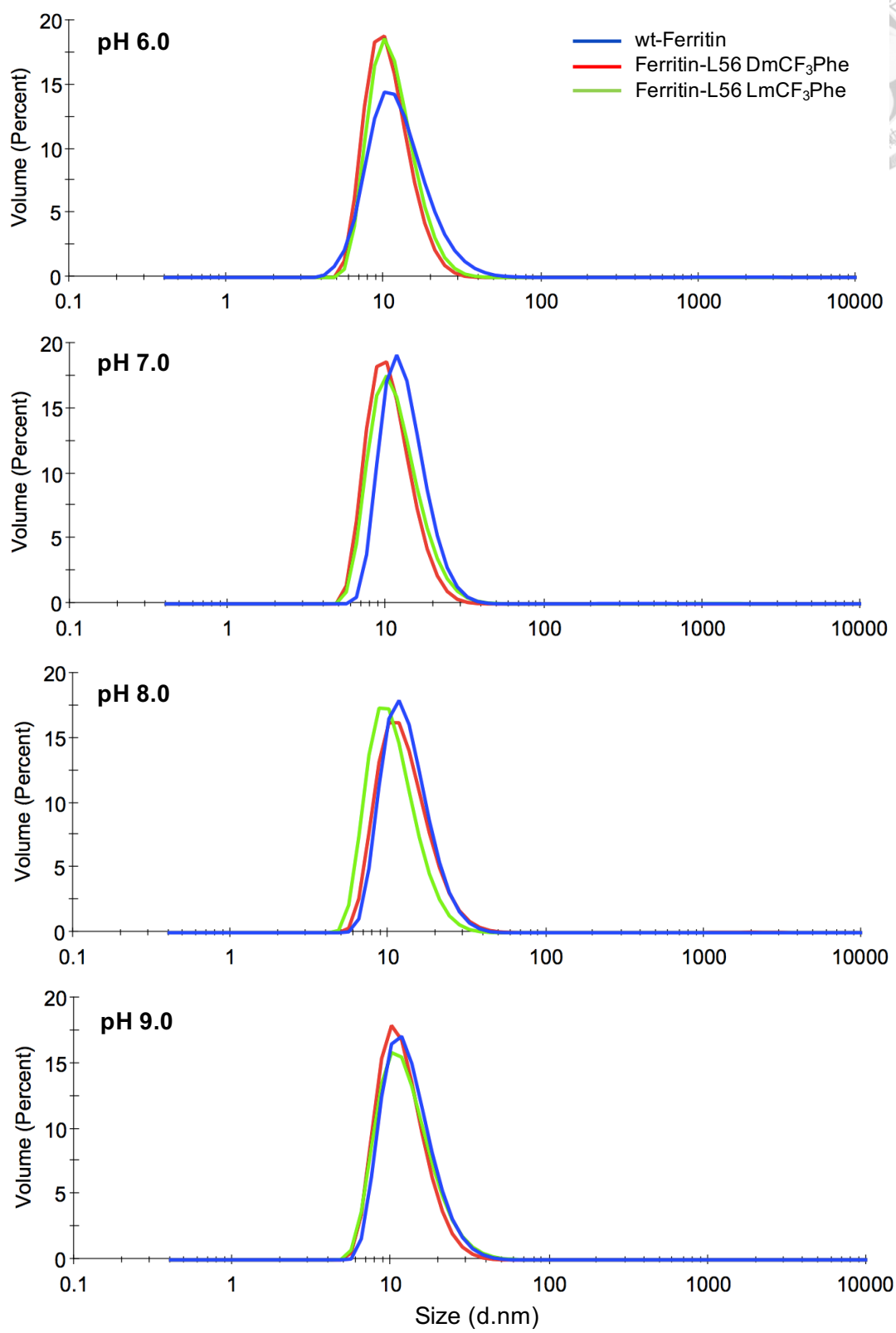
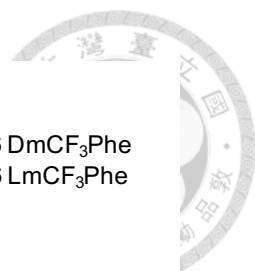


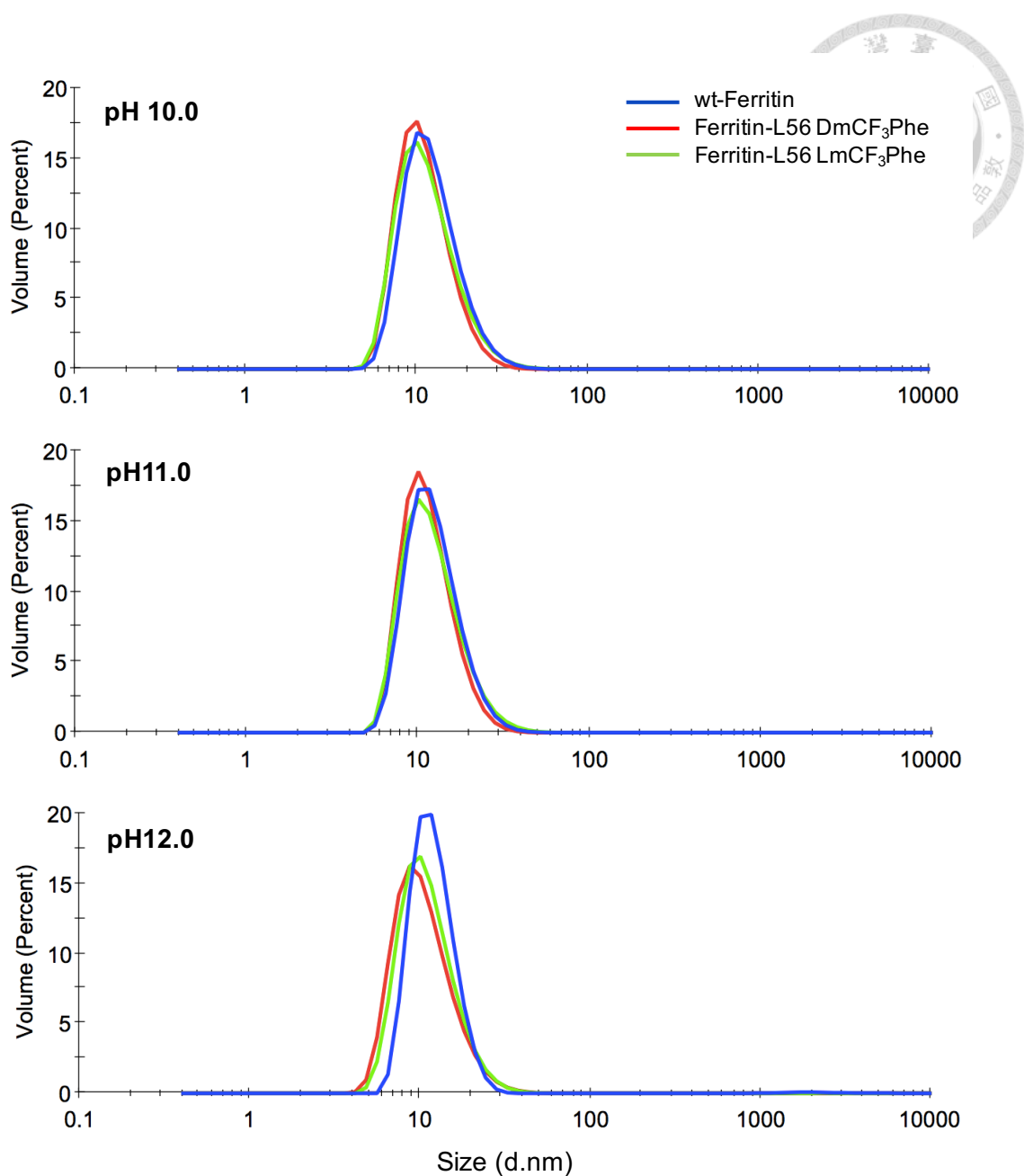
**Figure S8. FFF-MALS analysis of wt-Ferritin**

The sample after SEC column in PBS buffer was pooled for FFF-MALS analysis. The flow rate of FFF was 0.35 mL/min with light scattering instrument (TREOS). The cell type was fused silica, wavelength was 659.0 nm, and calibration constant was  $5.1573 \times 10^{-5}$  1/(V cm).



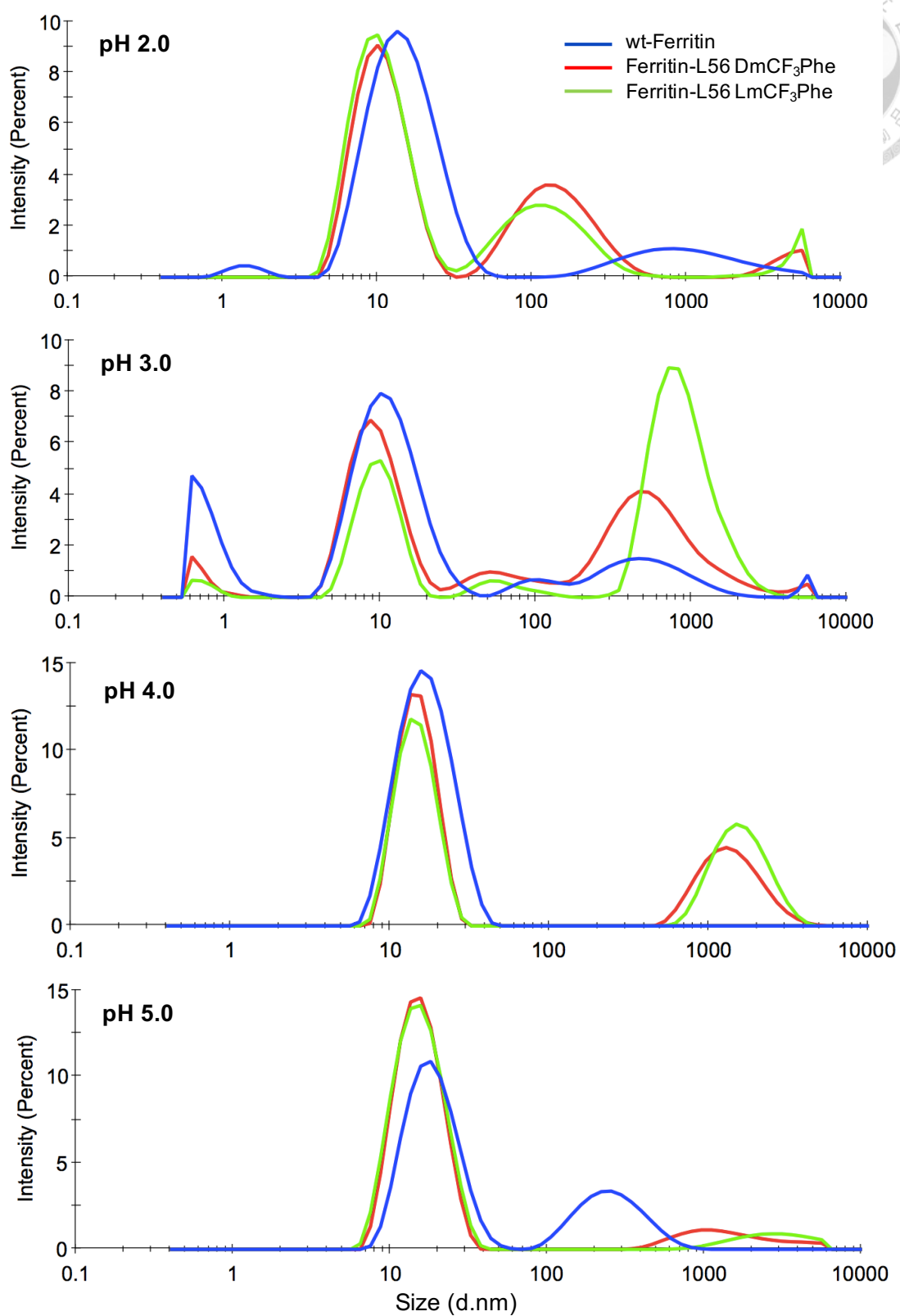
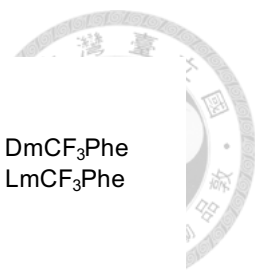


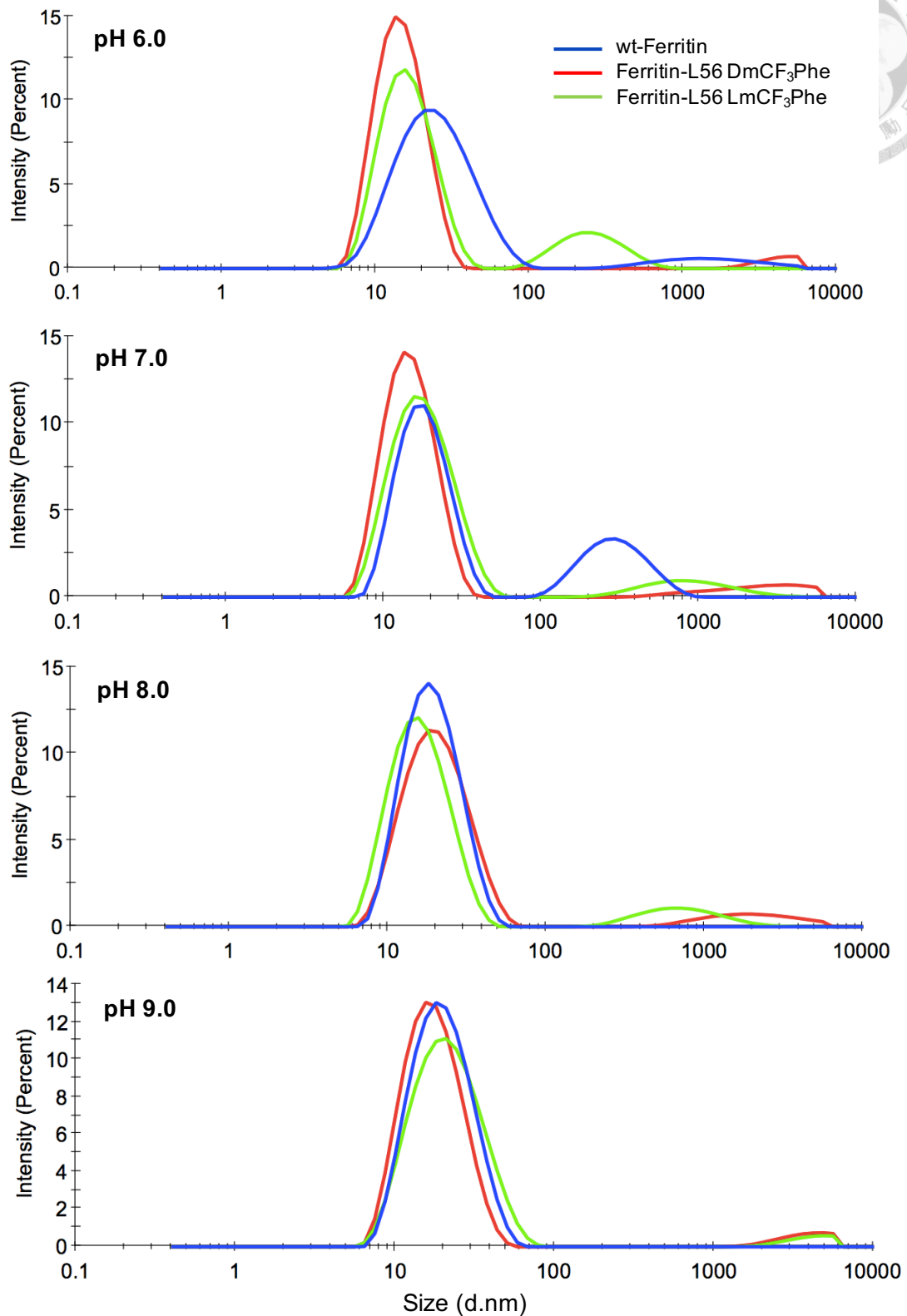
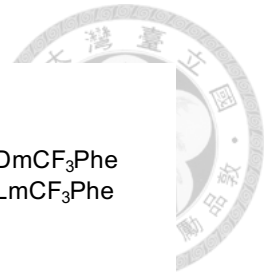


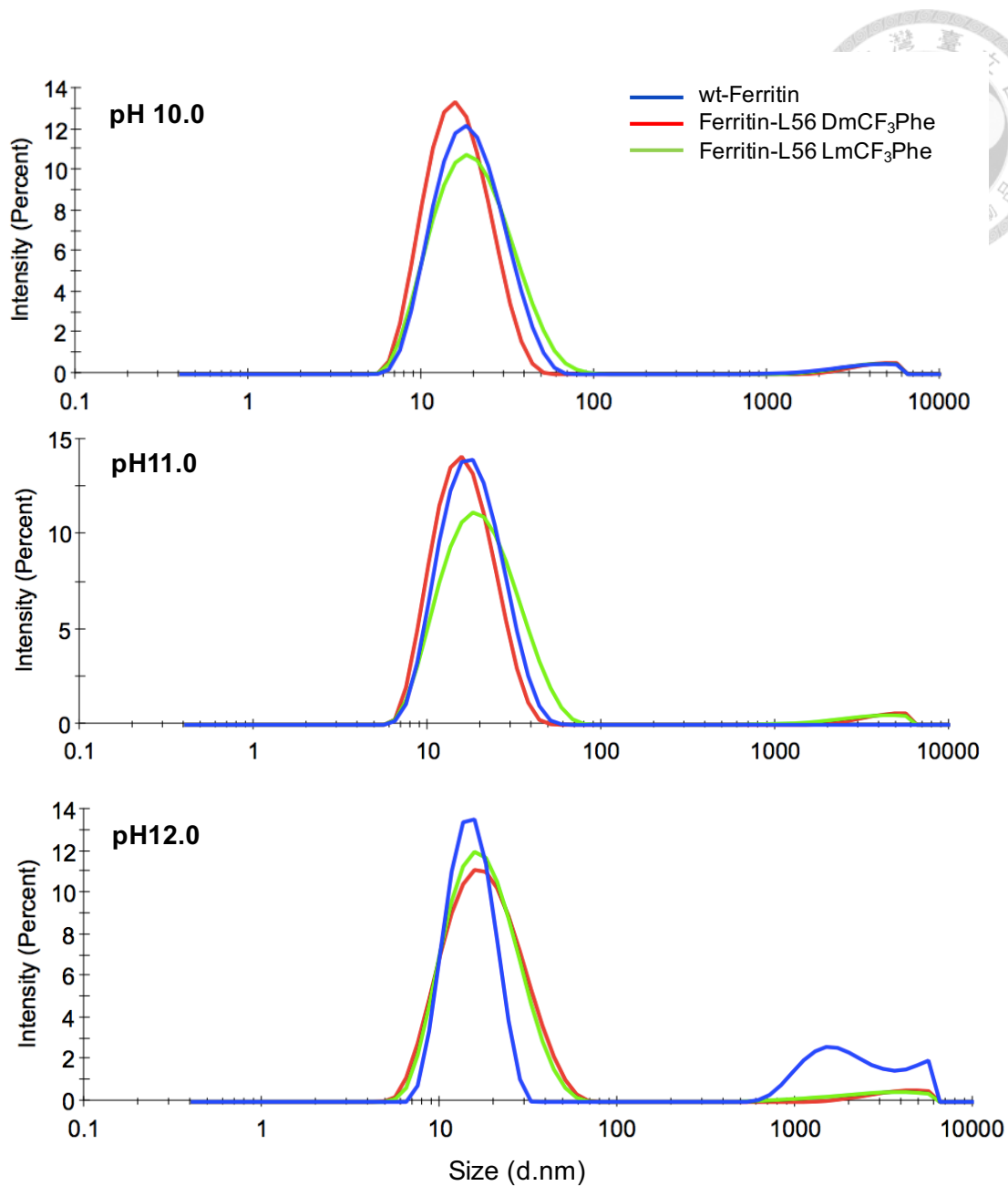


**Figure S9. DLS analysis of Ferritin variants at pH 2-12 displayed by volume.**

The ferritin variants were analyzed by DLS and the concentration of proteins were 0.5 mg/ml in buffer at pH 2 to 12. The spectra were measure by scattering light intensity of the scattering particles versus their diameter (nm) and calculated the percentage of volume. Blue line: wt-Ferritin; red line: Ferritin-L56-2; green line: Ferritin-L56-3.

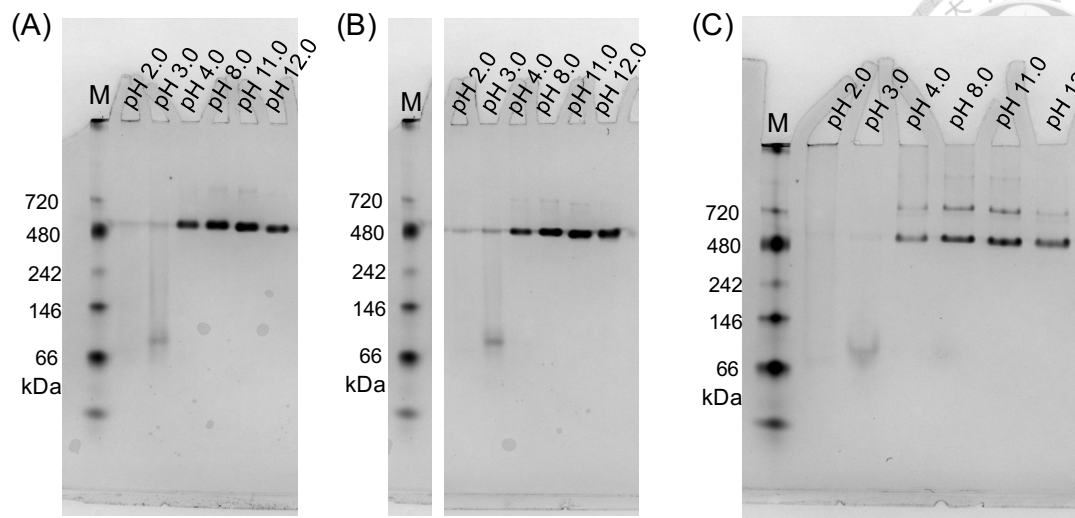






**Figure S10. DLS analysis of Ferritin variants at pH 2-12 displayed by intensity.**

The ferritin variants were analyzed by DLS and the concentration of proteins were 0.5 mg/ml in buffer at pH 2 to 12. The spectra were measure by scattering light intensity of the scattering particles versus their diameter (nm). Blue line: wt-Ferritin; red line: Ferritin-L56-2; green line: Ferritin-L56-3.

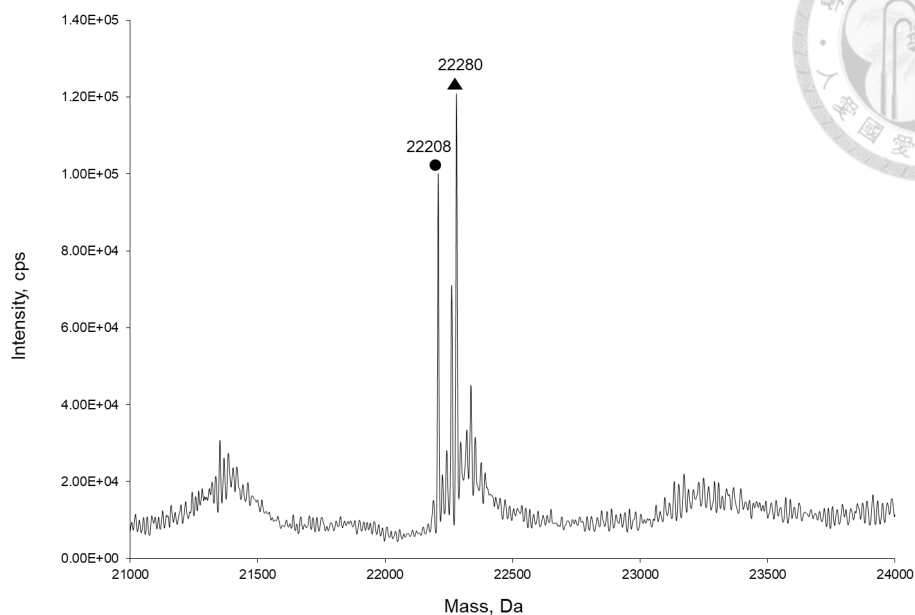


**Figure S11. Native gel analysis of Ferritin variants at pH 2, 3, 4, 8, 11, and 12.**

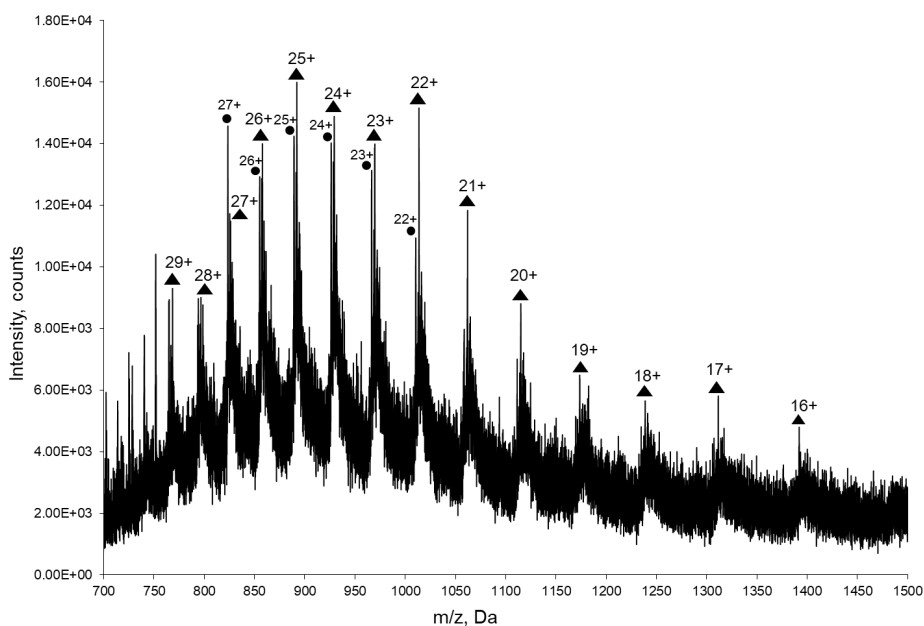
Ferritin variants was dissolved in pH 2, 3, 4, 8, 11, and 12 buffer. 5  $\mu\text{g}$  protein was invidiously mixed with native sample dye and running native gel with 80 V at 4°C for 3 hrs. (A) Ferritin-L56-2, (B) Ferritin-L56-3, and (C) wt-Ferritin. While the buffer condition was lower than pH 3.0, the ferritin variants were disassembly.



(A)



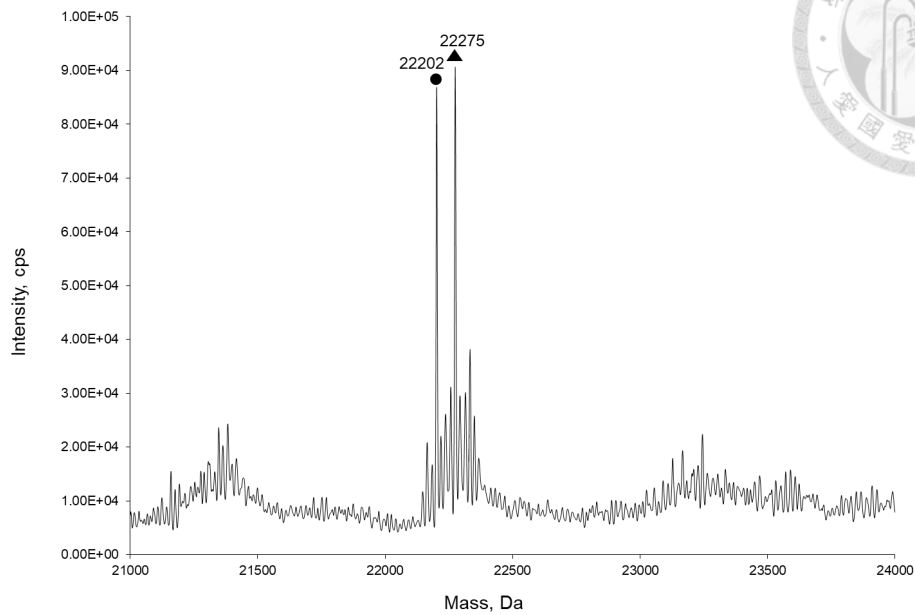
(B)



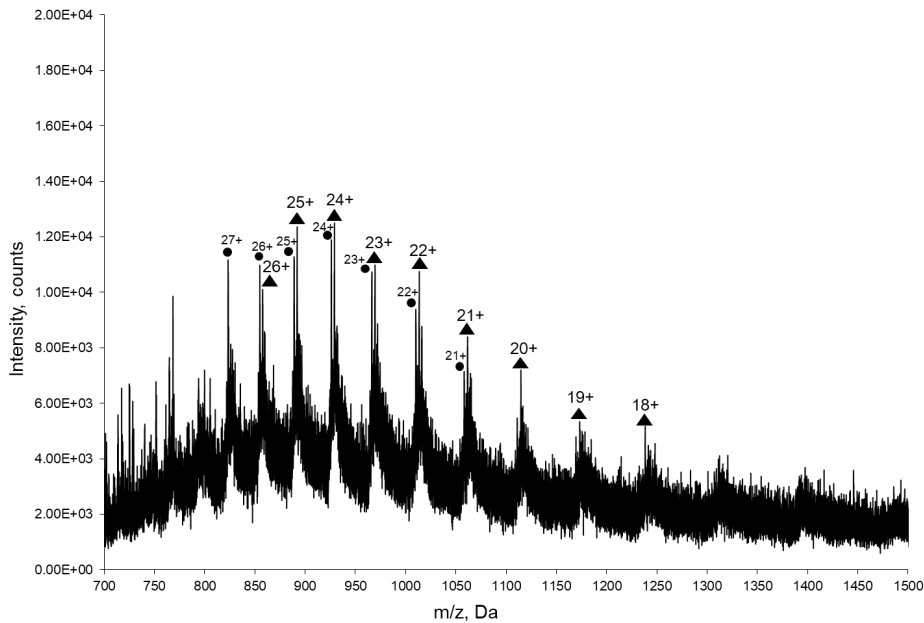
**Figure S12. Molecular mass determination of the protein Ferritin-H60TAG-10.**

Full-length Ferritin-H60TAG-10 was produced by using PylRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 22,209 Da (protein reducing first methionine); observed molecular mass was 22,208 Da and 22280 Da.

(A)



(B)



**Figure S13. Molecular mass determination of the protein Ferritin-R63TAG-11.**

Full-length Ferritin-R63TAG-11 was produced by using PylRS•tRNA<sub>CUA</sub><sup>Pyl</sup> pair in BL21(DE3) strain with 1 mM ncAA in M9 medium. (A) ESI-MS spectrum (B) the deconvoluted ESI-MS spectrum. The calculated molecular mass is 22,203 Da (protein reducing first methionine); the finding molecular mass was 22,202 Da and 22275 Da.



### Organic synthesis of *N*<sup>e</sup>-((2-azidoethoxy) carbonyl)-L-lysine

NaN<sub>3</sub> (780 mg, 10 mmol) was added to a solution of 2-bromoethanol (500 mg, 4 mmol) in water 10 mL. To ensure that there is no precipitate in solution. The mixture was stirred at 80°C in reflux apparatus overnight and cooled to room temperature (RT). The aqueous layer was extracted with ether (C<sub>2</sub>H<sub>5</sub>OC<sub>2</sub>H<sub>5</sub>). The combined organic extracts were dried over magnesium sulfate, filtered, and solvent removed *in vacuo*. The resulting oil was 2-azidoethanol in 60% yield (204 mg, 2.3 mmol).<sup>64</sup>

In the next reaction, 2-azidoethanol (500mg, 5.74 mmol) was added to a solution of 4-nitrophenyl carbonochloridate (1.153g, 5.74 mmol) in dried DCM (25 mL). And the mixture was added DIPEA (5 mL) on ice bath. The reaction was stirred overnight, and DCM was evaporated under vacuum. The resulting compound was extracted with ethyl acetate (30 mL), and washed with brine (3 x 15 mL), subsequently washed with ddH<sub>2</sub>O (3 x 15 mL). The organic layer was then dried over magnesium sulfate, filtered and evaporated, affording clean 2-azidoethyl (4-nitrophenyl) carbonate (**58**) (1276 mg, 5.06 mmol). All of **58** was mixed with Boc-Lys-OH (1039 mg, 4.2 mmol) and slowly added TFA (10 mL)/ NaOH (15mL) at 0°C. The reaction mixture was stirred for 12hrs and slowly warmed to RT. Subsequently, the solution was cooled to 0°C acidified to pH 2-3 with 1 or 2 M HCl solution. The aqueous layer was extracted with ethyl acetate (100 mL) and then the organic layer was washed with brine (2 x 100 mL), subsequently washed with ddH<sub>2</sub>O (100 mL). The organic layer was then dried over magnesium sulfate, filtered and evaporated, affording *N*<sup>e</sup>-((2-azidoethoxy)carbonyl)-*N*<sup>t</sup>-(*tert*-butoxycarbonyl)-L-lysine (**59**) in 90% yield (1357 mg, 3.78 mmol).<sup>65</sup>

Finally, **59** (1357 mg, 3.78 mmol) was dissolved in DCM (5 mL) and slowly added TFA (5 mL) solution. The reaction was stirred at RT for 1 hr, added ether about 100 mL until appearing white precipitate. The precipitate was collected and dried under vacuum at least 12 hrs, affording pure *N*<sup>ε</sup>-((2-azidoethoxy) carbonyl)-L-lysine (**12**) in 84% yield (822 mg, 3.17 mmol). (**Figure S7**) And the product was analyzed by NMR. (**Figure S8**)

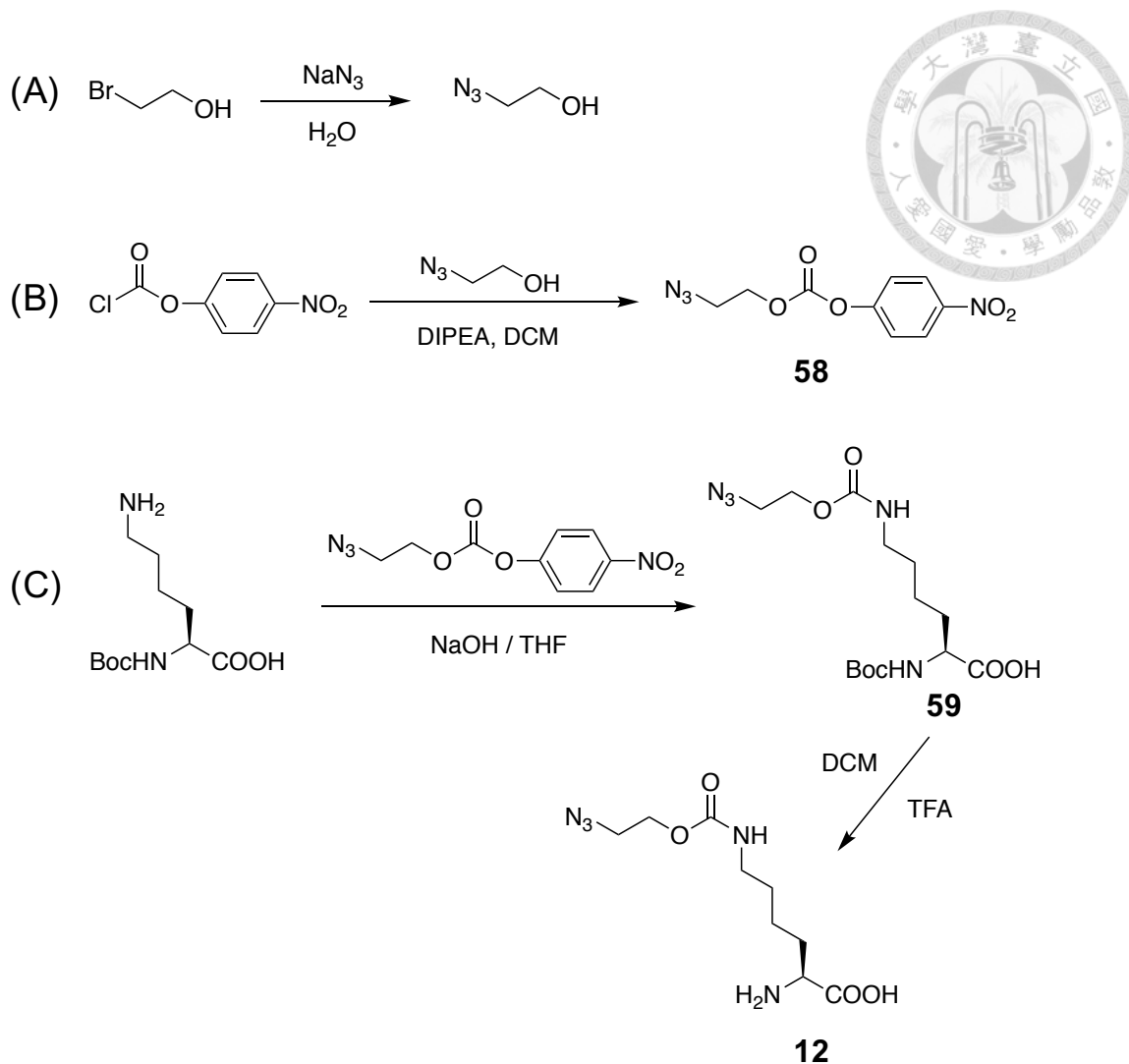
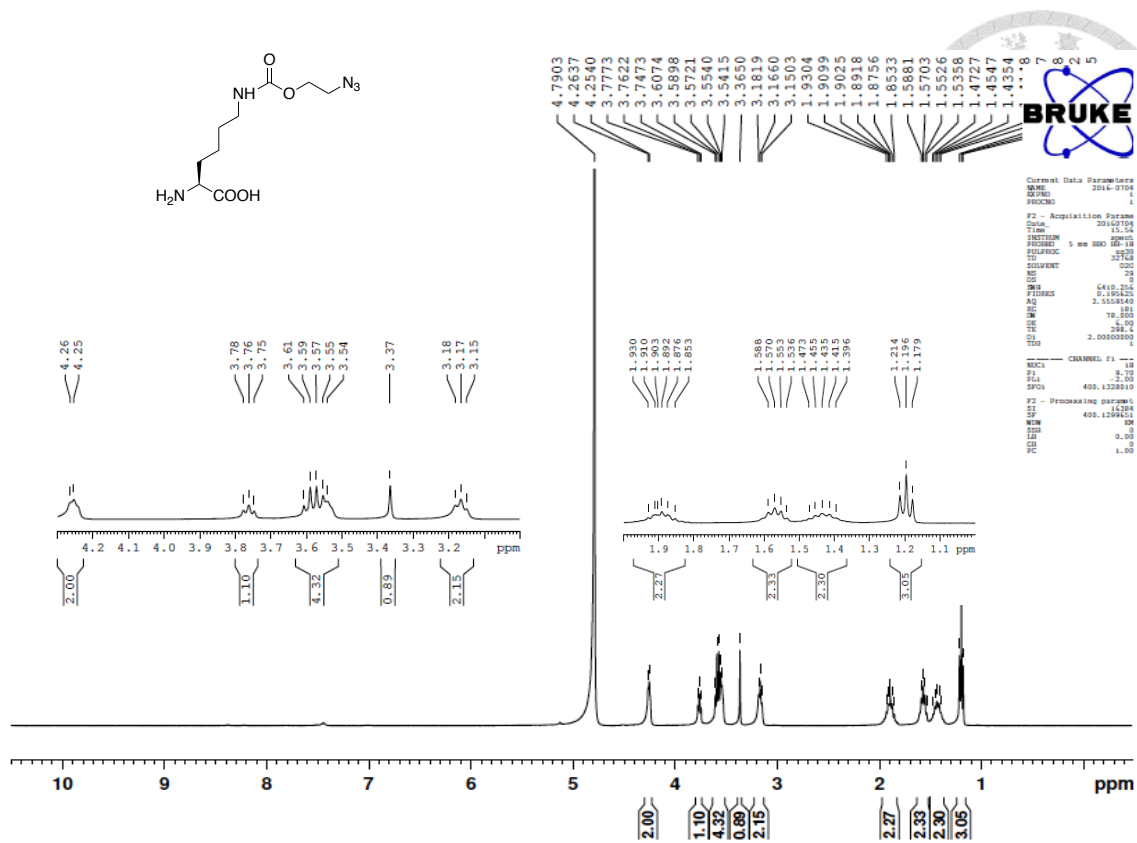


Figure S14. Synthesis of *N*<sup>ε</sup>-((2-azidoethoxy) carbonyl)-L-lysine.



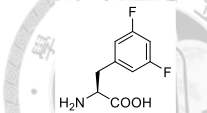
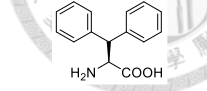
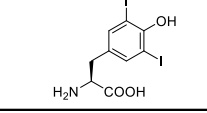
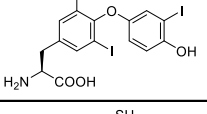
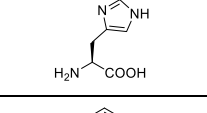
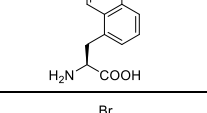
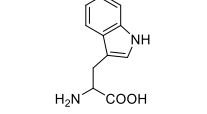
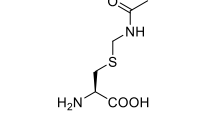
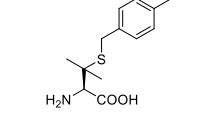
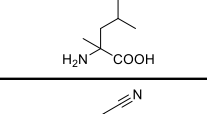
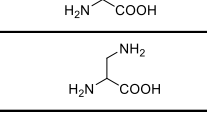
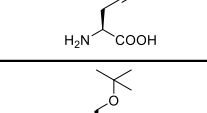
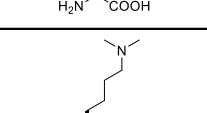
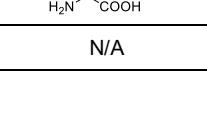
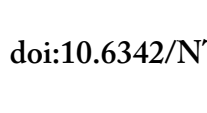
**Figure S15.** The NMR spectrum of *N*<sup>ε</sup>-((2-azidoethoxy) carbonyl)-L-lysine.

<sup>1</sup>H NMR (D<sub>2</sub>O, 400MHz), δ1.18~1.21 (t, 3H), δ1.40~1.47 (m, 2H), δ1.54~1.59 (m, 2H), δ1.85~1.93 (m, 2H), δ3.15~3.18 (t, 2H), δ3.37 (s, 1H), δ3.54~3.61 (m, 4H), δ3.75~3.78 (t, 1H), δ4.25~4.26 (m, 2H).

**Table S1: Chemical Names and Structures in ncAA Library**

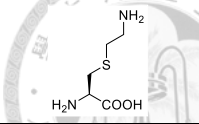
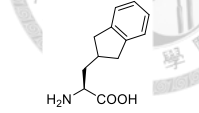
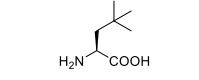
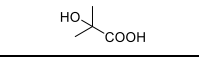
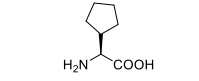
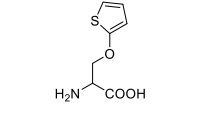
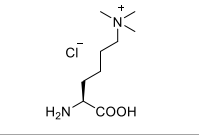
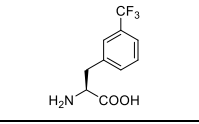
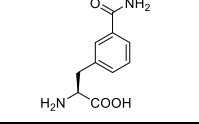
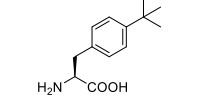
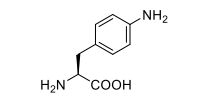
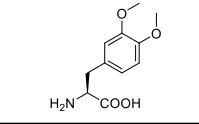
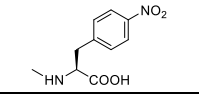
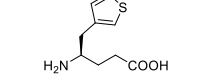
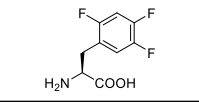
Position	No.	CAS No.	Name	Chemical Structure
A01	N/A	N/A	N/A	N/A
A02	N/A	N/A	N/A	N/A
A03	5	42538-40-9	2-Bromo-L-phenylalanine	
A04	130	57213-48-6	3-Cyano-L-phenylalanine	
A05	11	126109-42-0	4-Carboxy-L-phenylalanine	
A06	253	24250-85-9	4-Iodo-L-phenylalanine	
A07	69	111119-36-9	L-2,4-Dichlorophenylalanine	
A08	163	2566-30-5	<i>N</i> <sup>α</sup> -Methyl-L-phenylalanine hydrochloride	
A09	81	37535-49-2	3-(4'-Pyridyl)-L-alanine	
A10	51	31105-93-8	2,4-Difluoro-L-phenylalanine	
A11	235	515-30-0	α-Phenyllactic acid	
A12	35	6636-22-2	<i>O</i> -Acetyl-L-tyrosine	
A13	197	1596-67-4	L-Thyronine	
A14	156	32381-18-3	DL-α-Methylhistidine dihydrochloride	
A15	50	154-08-5	5-Fluoro-DL-tryptophan	
A16	152	1954-45-6	4-Methyl-DL-tryptophan	

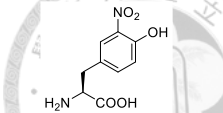
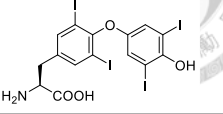
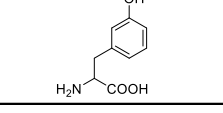
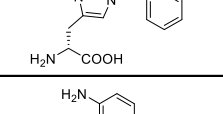
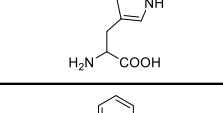
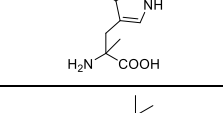
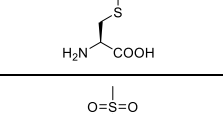
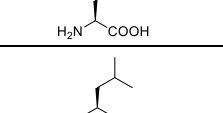
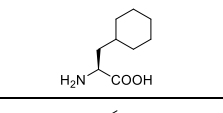
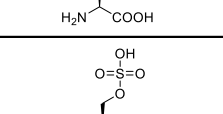
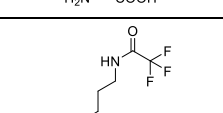
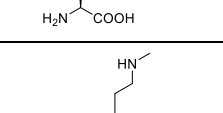
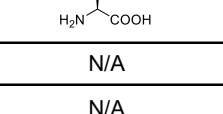

A17	117	638-23-3	<i>S</i> -Carboxymethyl-L-cysteine	
A18	268	387868-34-0	<i>S</i> -4-Methoxybenzyl-L-penicillamine	
A19	111	16338-48-0	H-Allyl-L-glycine	
A20	296	6600-40-4	L-Norvaline	
A21	103	62-57-7	$\alpha$ -Aminoisobutyric acid	
A22	302	25528-71-6	-Cyclohexyl-L-alanine hydrochloride	
A23	282	66638-22-0	<i>O</i> -Acetyl-L-serine hydrochloride	
A24	160	692-04-6	<i>N</i> <sup>ε</sup> -Acetyl-L-lysine	
B01	N/A	N/A	N/A	N/A
B02	N/A	N/A	N/A	N/A
B03	6	167817-55-2	2-Iodo-L-phenylalanine	
B04	131	19883-74-0	3-Nitro-L-phenylalanine	
B05	36	223593-04-2	L-4-Carbamoylphenylalanine	
B06	101	155760-02-4	H- <i>p</i> -Phenyl-L-Phenylalanine	
B07	99	49607-21-8	L-2,4-Dinitrophenylalanine	
B08	241	62777-25-7	L-2-Amino-5-phenylpentanoic acid	
B09	82	64090-98-8	3-(3'-Pyridyl)-L-alanine dihydrochloride	

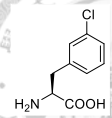
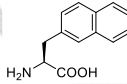
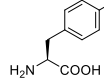
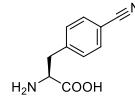


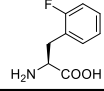
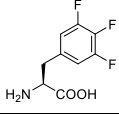
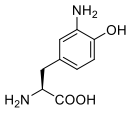
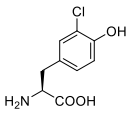
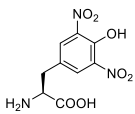

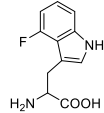
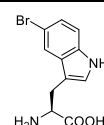
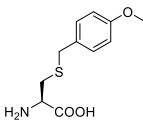
B10	59	31105-91-6	L-3,5-Difluorophenylalanine	
B11	53	149597-92-2	3,3-Diphenyl-L-alanine	
B12	52	18835-59-1	3,5-Diiodo-L-tyrosine dihydrate	
B13	211	55-06-1	3,3',5-Triiodo-L-thyronine sodium salt	
B14	157	2002-22-4	2-Mercapto-L-histidine	
B15	56	55516-54-6	3-(1-Naphthyl)-L-alanine	
B16	153	33599-61-0	6-Bromo-DL-tryptophan	
B17	118	28798-28-9	S-Acetamidomethyl-L-cysteine hydrochloride	
B18	269	1039102-11-8	S-4-Methylbenzyl-L-penicillamine	
B19	138	144-24-1	DL-α-Methylleucine	
B20	45	6232-19-5	β-Cyano-L-alanine	
B21	189	54897-59-5	DL-2,3-Diaminopropionic acid monohydrochloride	
B22	47	23235-01-0	L-Propargylglycine	
B23	283	18822-58-7	O-tert-Butyl-L-serine	
B24	228	2259-86-1	N <sup>ε</sup> -Dimethyl-L-lysine hydrochloride	
C01	N/A	N/A	N/A	N/A

C02	N/A	N/A	N/A	N/A
C03	55	263396-42-5	2-Cyano-L-phenylalanine	
C04	190	19883-75-1	2-Nitro-L-phenylalanine	
C05	37	122555-04-8	L-4-Acetylphenylalanine	
C06	126	949-99-5	4-Nitro-L-phenylalanine	
C07	134	52794-99-7	L-3,4-Dichlorophenylalanine	
C08	243	267650-37-3	3-Styryl-L-alanine	
C09	83	37535-51-6	3-(2'-Pyridyl)-L-alanine	
C10	133	31105-90-5	L-3,4-Difluorophenylalanine	
C11	191	57213-47-5	L-3-Aminomethylphenylalanine	
C12	64	6230-11-1	O-Methyl-L-tyrosine	
C13	262	21820-51-9	O-Phospho-L-tyrosine	
C14	289	332-80-9	1-Methyl-L-Histidine	
C15	58	72120-71-9	3-Benzothieryl-L-alanine	
C16	154	852391-45-8	7-Bromo-DL-tryptophan	
C17	119	2481-09-6	S-tert-Butyl-L-cysteine hydrochloride	

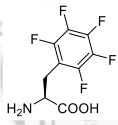
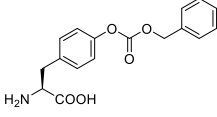
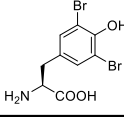

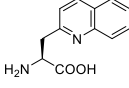
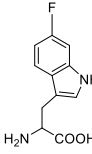
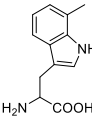
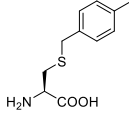




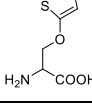
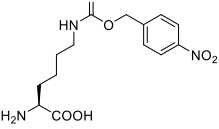


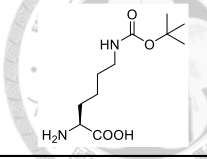
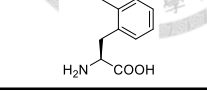
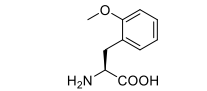
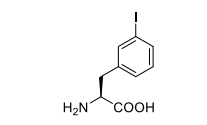
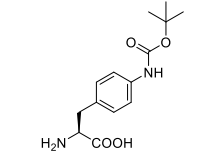
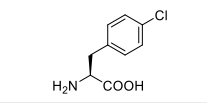
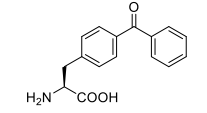
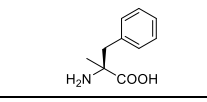
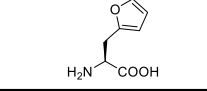
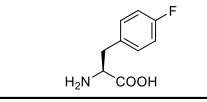
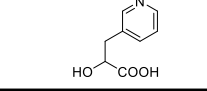
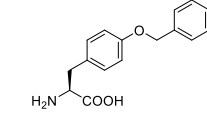
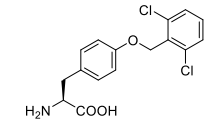
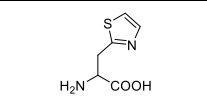
C18	303	4099-35-8	<i>S</i> -(2)-Aminoethyl-L-cysteine hydrochloride	
C19	139	155239-51-3	L-β-indanylglycine	
C20	46	57224-50-7	β- <i>tert</i> -Butyl-L-alanine	
C21	234	594-61-6	α-Hydroxyisobutyric acid	
C22	68	2521-84-8	L-Cyclopentylglycine	
C23	284	32595-59-8	DL-β-(2-Thienyl)-serine	
C24	229	55528-53-5	<i>N</i> <sup>ε</sup> -(trimethyl)-L-lysine chloride	
D01	N/A	N/A	N/A	N/A
D02	N/A	N/A	N/A	N/A
D03	259	14464-68-7	3-(Trifluoromethyl)-L-phenylalanine	
D04	300	1217651-22-3	L-3-Carbamoylphenylalanine	
D05	54	82372-74-5	4- <i>tert</i> -Butyl-L-phenylalanine	
D06	127	943-80-6	4-Amino-L-phenylalanine	
D07	135	32161-30-1	L-3,4-Dimethoxyphenylalanine	
D08	165	70663-55-7	<i>N</i> -Methyl-4-nitro-L-phenylalanine	
D09	200	270262-99-2	(3-Thienyl)-L-β-homoalanine	
D10	98	749847-57-2	L-2,4,5-Trifluorophenylalanine	

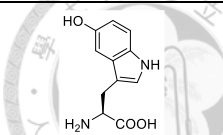
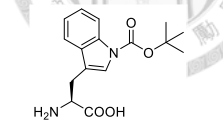
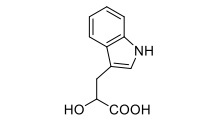
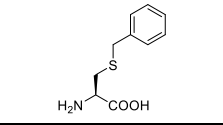
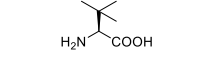
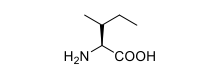
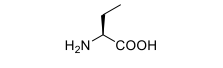
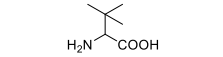
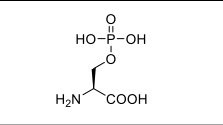
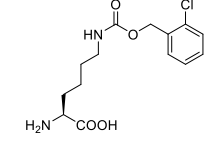
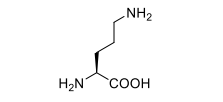
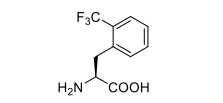
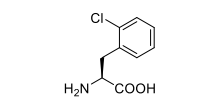
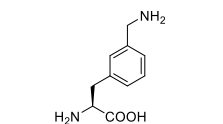
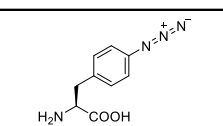
D11	25	621-44-3	3-Nitro-L-tyrosine	
D12	123	51-48-9	L-Thyroxine	
D13	287	775-06-4	DL- <i>m</i> -Tyrosine	
D14	290	2022956-35-8	<i>N</i> <sup>im</sup> -Benzyl-D-histidine	
D15	90	6383-69-3	5-Amino-DL-tryptophan	
D16	205	153-91-3	DL- $\alpha$ -Methyl-tryptophan	
D17	140	30044-51-0	<i>S</i> -tert-Butylthio-L-cysteine	
D18	121	7314-32-1	L-Methionine sulfone	
D19	187	13748-90-8	L- $\alpha$ -Hydroxyisocaproic acid	
D20	78	27527-05-5	$\beta$ -Cyclohexyl-L-alanine	
D21	238	1492-24-6	L- $\alpha$ -Aminobutyric acid	
D22	110	626-69-7	<i>O</i> -Sulfo-L-serine	
D23	28	10009-20-8	<i>N</i> <sup>ε</sup> -Trifluoroacetyl-L-lysine	
D24	231	7622-29-9	<i>N</i> <sup>ε</sup> -Methyl-L-lysine hydrochloride	
E01	N/A	N/A	N/A	N/A
E02	N/A	N/A	N/A	N/A

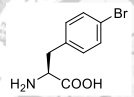
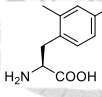
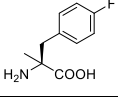
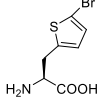
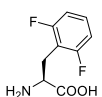

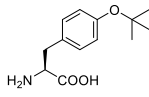
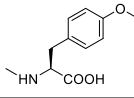
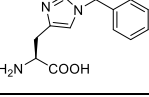
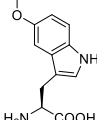
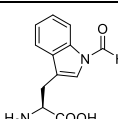
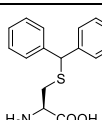




E03	258	80126-51-8	3-Chloro-L-phenylalanine	
E04	57	58438-03-2	3-(2-Naphthyl)-L-alanine	
E05	63	1991-87-3	4-Methyl-L-phenylalanine	
E06	128	104531-20-6	4-Cyano-L-phenylalanine	
E07	61	2935-35-5	L-Phenylglycine	
E08	74	3685-51-6	$\beta$ -(3-Thienyl)-L-alanine	
E09	95	19883-78-4	2-Fluoro-L-phenylalanine	
E10	132	646066-73-1	3,4,5-Trifluoro-L-phenylalanine	
E11	26	23279-22-3	3-Amino-L-tyrosine	
E12	129	7423-93-0	3-Chloro-L-tyrosine	
E13	288	17360-11-1	3,5-Dinitro-L-tyrosine monohydrate	
E14	291	368-16-1	H-N-3-Methyl-L-histidine	
E15	92	25631-05-4	4-Fluoro-DL-tryptophan	
E16	215	25197-99-3	5-Bromo-L-tryptophan	
E17	141	2544-31-2	S-4-Methoxybenzyl-L-cysteine	

E18	44	87392-13-0	4,5-Dehydro-L-leucine	
E19	260	2566-29-2	Diethylglycine	
E20	79	99295-82-6	$\beta$ -Cyclopentyl-L-alanine	
E21	292	7685-44-1	H-Allyl-DL-glycine	
E22	247	4726-96-9	O-Benzyl-L-serine	
E23	122	10009-97-9	$N^{\delta}$ -Phthaloyl-L-ornithine hydrochloride	
E24	244	1155-64-2	$N^{\epsilon}$ -Z-L-lysine	
F01	N/A	N/A	N/A	N/A
F02	N/A	N/A	N/A	N/A
F03	257	82311-69-1	3-Bromo-L-phenylalanine	
F04	8	150338-20-8	4-(Aminomethyl)-L-phenylalanine	
F05	67	114926-38-4	$\rho$ -Trifluoromethyl-L-phenylalanine	
F06	298	1080496-42-9	4-Propargyloxy-L-phenylalanine	
F07	60	943-73-7	L-Homophenylalanine	
F08	75	22951-96-8	$\beta$ -(2-Thienyl)-L-alanine	
F09	96	19883-77-3	3-Fluoro-L-phenylalanine	

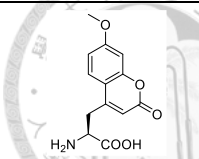
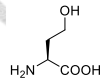
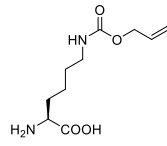
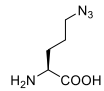
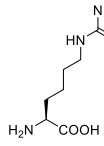
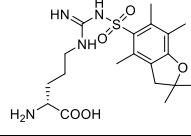
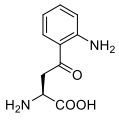
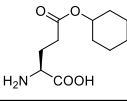
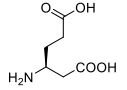
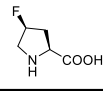
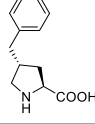
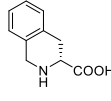
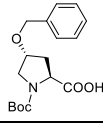
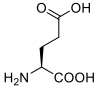
F10	149	138109-65-6	Pentafluoro-L-phenylalanine	
F11	27	21106-04-7	<i>O</i> -Z-L-tyrosine	
F12	136	300-38-9	L-3,5-Dibromotyrosine	
F13	72	119433-80-6	3-(4-Thiazolyl)-L-alanine	
F14	7	161513-46-8	3-(2'-Quinoly)-L-alanine	
F15	93	7730-20-3	6-Fluoro-DL-tryptophan	
F16	13	17332-70-6	7-Methyl-DL-tryptophan	
F17	142	42294-52-0	<i>S</i> -4-Methylbenzyl-L-cysteine	
F18	106	96386-92-4	L-Homoleucine Hydrochloride	
F19	270	1883-09-6	L-2,4-Diaminobutyric acid dihydrochloride	
F20	80	102735-53-5	H-β-Cyclopropyl-L-Alanine	
F21	293	14328-51-9	L-2-Cyclohexylglycine	
F22	248	32595-59-8	DL-β-(2-Thienyl)-serine	
F23	145	3557-90-2	<i>N</i> <sup>ε</sup> -4-Nitro-Z-L-lysine	

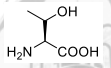
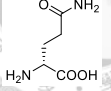
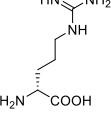
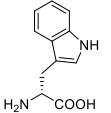
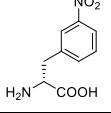
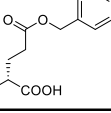
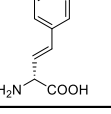
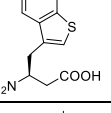
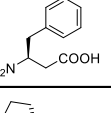
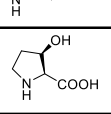
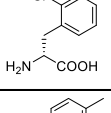
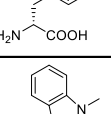
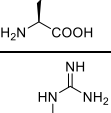
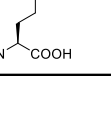
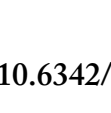
F24	249	2418-95-3	<i>N</i> <sup>ε</sup> -Boc-L-Lysine	
G01	3	80126-53-0	2-Methyl-L-phenylalanine	
G02	255	193546-31-5	L-2-Methoxyphenylalanine	
G03	124	20846-39-3	3-Iodo-L-phenylalanine	
G04	9	74578-48-6	4-(Boc-amino)-L-phenylalanine	
G05	65	14173-39-8	4-Chloro-L-phenylalanine	
G06	301	104504-45-2	4-Benzoyl-L-phenylalanine	
G07	71	23239-35-2	$\alpha$ -Methyl-L-phenylalanine	
G08	76	127682-08-0	H- $\beta$ -(2-Furyl)-L-alanine	
G09	97	1132-68-9	4-Fluoro-L-phenylalanine	
G10	184	889957-22-6	( <i>R,S</i> )2-Hydroxy-3-(3-pyridyl)propionic acid	
G11	32	16652-64-5	<i>O</i> -Benzyl-L-tyrosine	
G12	137	40298-69-9	<i>O</i> -2,6-Dichlorobenzyl-L-tyrosine	
G13	73	1596-65-2	$\beta$ -(2-Thiazolyl)-DL-alanine	

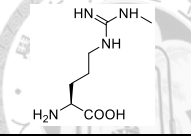
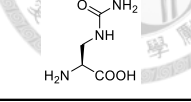
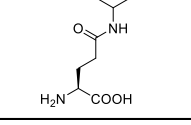
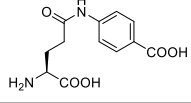
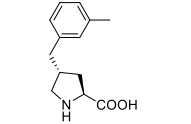
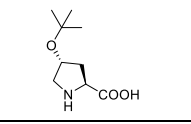
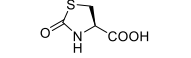
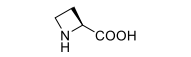
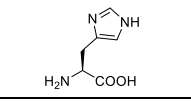
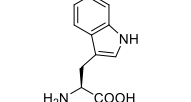
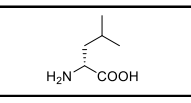
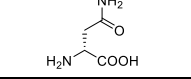
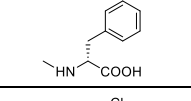
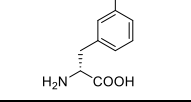
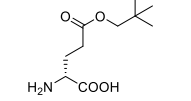
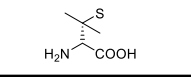
G14	22	4350-09-8	5-Hydroxy-L-tryptophan	
G15	150	146645-63-8	<i>N</i> <sup>in</sup> -Boc-L-tryptophan	
G16	233	1821-52-9	DL-Indole-3-lactic acid	
G17	246	3054-01-1	<i>S</i> -Benzyl-L-cysteine	
G18	107	20859-02-3	L- $\alpha$ - <i>tert</i> -Butyl-Gly-OH	
G19	108	1509-34-8	L-allo-Isoleucine	
G20	100	1492-24-6	L- $\alpha$ -Aminobutyric acid	
G21	294	33105-81-6	DL- $\alpha$ - <i>tert</i> -Butylglycine	
G22	261	407-41-0	<i>O</i> -Phospho-L-serine	
G23	146	42390-97-6	<i>N</i> <sup>ε</sup> -2-Chloro-Z-L-lysine	
G24	263	3184-13-2	L-Ornithine hydrochloride	
H01	62	119009-47-1	H-L-Phe(2-trifluoromethyl)-OH	
H02	4	103616-89-3	2-Chloro-L-phenylalanine	
H03	105	57213-47-5	L-3-Aminomethylphenylalanine	
H04	10	33173-53-4	4-Azido-L-phenylalanine	

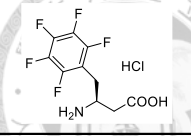
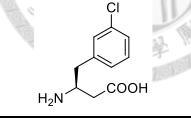
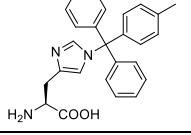
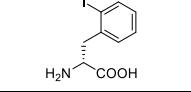
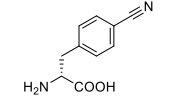
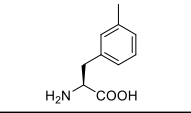
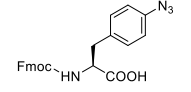
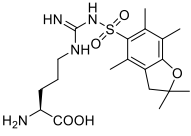
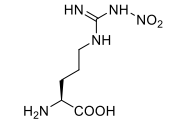
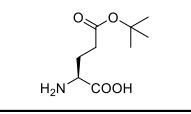
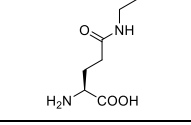
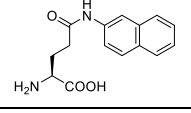
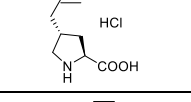
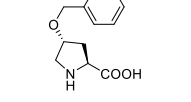
H05	66	24250-84-8	4-Bromo-L-phenylalanine	
H06	12	259726-56-2	L-2,4-Dimethylphenylalanine	
H07	70	130855-57-1	$\alpha$ -Methyl-L-4-Fluorophenylalanine	
H08	77	154593-58-5	L-2-(5-Bromothieryl)alanine	
H09	24	33787-05-2	2,6-Difluoro-L-phenylalanine	
H10	232	20312-36-1	L- $\beta$ -Phenyllactic acid	
H11	34	18822-59-8	<i>O</i> - <i>tert</i> -Butyl-L-tyrosine	
H12	162	52939-33-0	<i>N</i> -Methyl- <i>O</i> -methyl-L-tyrosine hydrochloride	
H13	125	16832-24-9	<i>N</i> <sup>im</sup> -Benzyl-L-histidine	
H14	23	25197-96-0	5-Methoxy-L-tryptophan	
H15	151	38023-86-8	<i>N</i> <sup>in</sup> -Formyl-L-tryptophan hydrochloride	
H16	20	5191-80-0	<i>S</i> -Diphenylmethyl-L-cysteine	
H17	266	1113-41-3	L-Penicillamine	
H18	109	2566-31-6	Di- <i>n</i> -propylglycine	
H19	188	17407-55-5	L- $\alpha$ -Hydroxyisovaleric acid	
H20	102	2731-73-9	$\beta$ -Chloro-L-alanine	

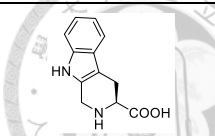
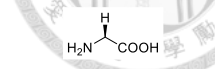
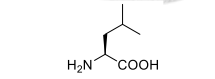
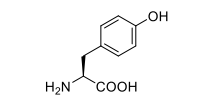
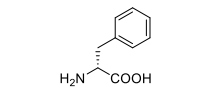
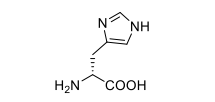
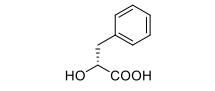
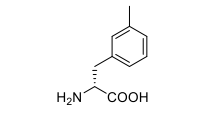
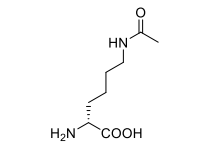
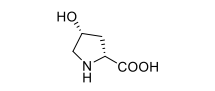
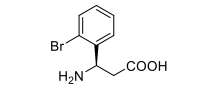
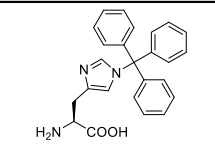
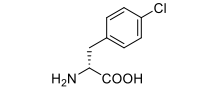
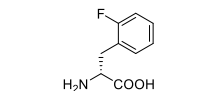
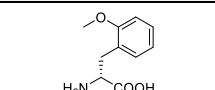


H21	367	208660-68-8	H-β-(7-Methoxycoumarin-4-yl)-Ala-OH	
H22	281	672-15-1	L-Homoserine	
H23	147	6298-03-9	N <sup>ε</sup> -Allyloxycarbonyl-L-lysine	
H24	297	156463-09-1	N <sup>δ</sup> -Azido-L-Ornithine hydrochloride	
I01	33	1190-49-4	L-Homocitrulline	
I02	225	200116-81-0	N <sup>ω</sup> -(2,2,4,6,7-Pentamethylidihydrobenzofuran)-5-sulfonyl-D-arginine	
I03	217	2922-83-0	L-Kynurenine	
I04	144	112471-82-6	L-Glutamic acid-γ-cyclohexyl ester	
I05	212	61884-74-0	L-β-Homoglutamic acid hydrochloride	
I06	39	2438-57-5	cis-4-Fluoro-L-proline	
I07	89	1279049-67-0	(R)-γ-Benzyl-L-proline·HCl	
I08	204	103733-65-9	(3R)-1,2,3,4-Tetrahydroisoquinoline-3-carboxylic acid	
I09	279	54631-81-1	Boc-O-benzyl-L-trans-4-hydroxyproline	
I10	312	56-86-0	L-Glutamic acid	

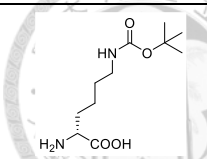
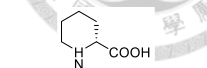
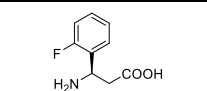
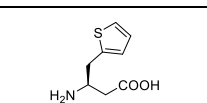
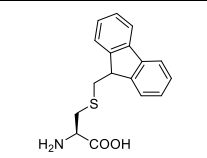
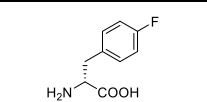
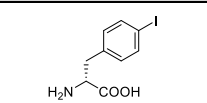
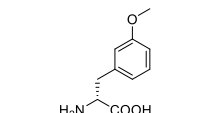
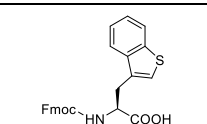
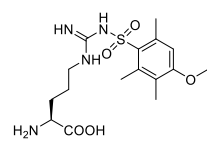
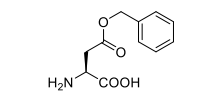
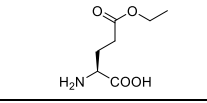
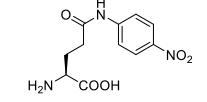
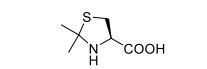
I11	324	72-19-5	L-Threonine	
I12	310	5959-95-5	D-Glutamine	
I13	333	157-06-2	D-Arginine	
I14	326	153-94-6	D-tryptophan	
I15	172	169530-97-6	3-Nitro-D-phenylalanine	
I16	236	2578-33-8	D-Glutamic acid-γ-benzyl ester	
I17	265	264903-53-9	3-Styryl-D-alanine	
I18	42	270063-44-0	L-β-HomoAla(3-benzothieryl)-OH.HCl	
I19	182	270062-92-5	3-Methyl-L-β-homophenylalanine hydrochloride	
I20	277	439918-59-9	D-β-Homoproline.HCl	
I21	112	567-35-1	cis-L-3-Hydroxyproline	
I22	350	80126-50-7	2-Chloro-D-phenylalanine	
I23	358	49759-61-7	4-Methyl-D-phenylalanine	
I24	366	21339-55-9	1-Methyl-L-tryptophan	
J01	148	156-86-5	L-Homoarginine	

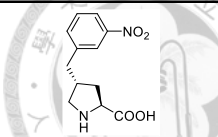
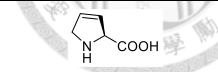
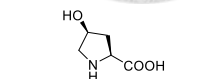
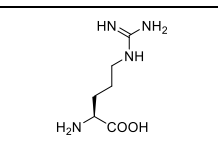
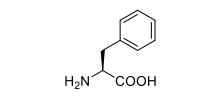
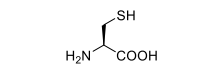
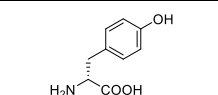
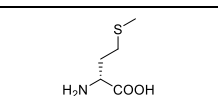
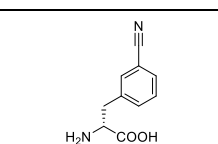
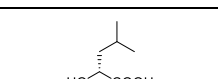
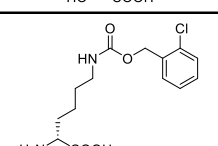
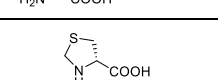
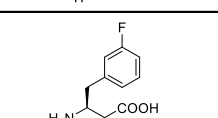
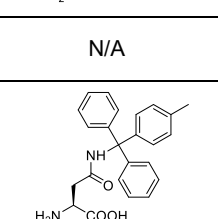
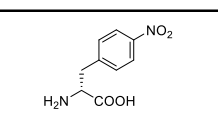
J02	230	156706-47-7	<i>N</i> <sup>ω</sup> -Methyl-L-arginine hydrochloride	
J03	242	1483-07-4	3-[(Aminocarbony)amino]-L-alanine	
J04	155	4311-12-0	<i>N</i> <sup>δ</sup> -Isopropyl-L-glutamine	
J05	214	2643-70-1	$\gamma$ -Glu-4-Abz-OH	
J06	40	1049734-52-2	( <i>R</i> )- $\gamma$ -(3-Methylbenzyl)-L-proline	
J07	113	79775-07-8	<i>O</i> - <i>tert</i> -Butyl-L- <i>trans</i> -4-hydroxyproline	
J08	209	19771-63-2	L-Thiazolidin-2-one-4-carboxylic acid	
J09	280	2133-34-8	L-Azetidine-2-carboxylic acid	
J10	314	71-00-1	L-Histidine	
J11	325	73-22-3	L-tryptophan	
J12	315	328-38-1	D-Leucine	
J13	335	2058-58-4	D-Asparagine hydrate	
J14	161	56564-52-4	<i>N</i> <sup>α</sup> -Methyl-D-phenylalanine hydrochloride	
J15	176	80126-52-9	3-Chloro-D-phenylalanine	
J16	237	45125-00-6	D-Glutamic acid- $\gamma$ - <i>tert</i> -butyl ester	
J17	267	52-67-5	D-Penicillamine	

J18	43	270063-41-7	Pentafluoro-L-β-homophenylalanine hydrochloride	
J19	183	270596-38-8	3-Chloro-L-β-homophenylalanine hydrochloride	
J20	1	133367-32-5	<i>N</i> <sup>im</sup> -4-Methyltrityl-L-histidine	
J21	344	736184-44-4	2-Iodo-D-phenylalanine	
J22	351	263396-44-7	4-Cyano-D-phenylalanine	
J23	359	114926-37-3	3-Methyl-L-phenylalanine	
J24	368	163217-43-4	Fmoc-4-azido-L-phenylalanine	
K01	158	200115-86-2	L-Arg(Pbf)-OH	
K02	239	2149-70-4	<i>N</i> <sup>ω</sup> -Nitro-L-arginine	
K03	29	2419-56-9	L-Glutamic acid-γ-tert-butyl ester	
K04	192	3081-61-6	<i>N</i> <sup>γ</sup> -Ethyl-L-glutamine	
K05	220	14525-44-1	L-Glutamic acid-γ-(β-naphthylamide)	
K06	41	1049755-32-9	( <i>R</i> )-γ-Propynyl-L-proline-HCl	
K07	114	66831-16-1	<i>O</i> -Benzyl-L- <i>trans</i> -L-4-hydroxyproline hydrochloride	

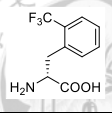
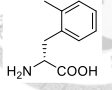
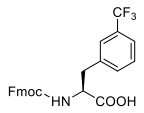
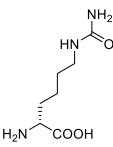
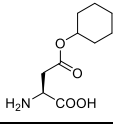
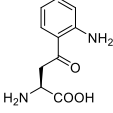
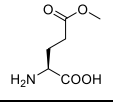
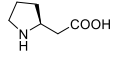
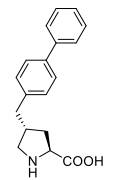
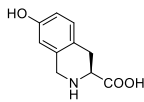
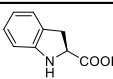
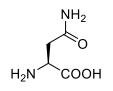
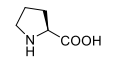
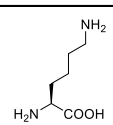
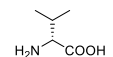
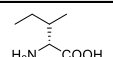
K08	218	42438-90-4	L-1,2,3,4-Tetrahydronorharman-3-carboxylic acid	
K09	313	56-40-6	Glycine	
K10	316	61-90-5	L-Leucine	
K11	327	60-18-4	L-Tyrosine	
K12	319	673-06-3	D-Phenylalanine	
K13	336	351-50-8	D-Histidine	
K14	164	7326-19-4	D-β-Phenyllactic acid	
K15	177	114926-39-5	3-Methyl-D-phenylalanine	
K16	245	51621-57-9	N <sup>ε</sup> -Acetyl-D-lysine	
K17	271	2584-71-6	cis-D-4-Hydroxyproline	
K18	48	275826-34-1	(S)-3-Amino-3-(2-bromophenyl)propionic acid	
K19	N/A	N/A	N/A	N/A
K20	2	35146-32-8	N <sup>im</sup> -Trityl-histidine	
K21	345	14091-08-8	4-Chloro-D-phenylalanine	
K22	352	122839-51-4	2-Fluoro-D-phenylalanine	
K23	360	170642-31-6	D-2-Methoxyphenylalanine	

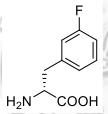

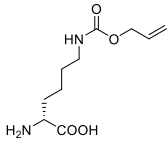
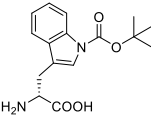
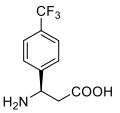
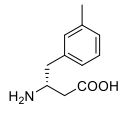
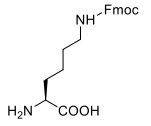
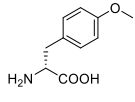
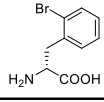
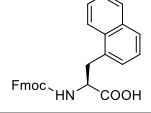
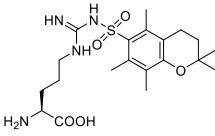
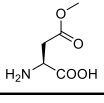
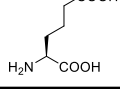
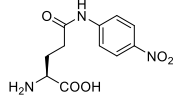
K24	369	205526-32-5	Fmoc-pentafluoro-L-phenylalanine	
L01	159	4125-79-5	$N^{\omega}, N^{\omega}$ -Di-Z-L-arginine	
L02	240	4353-32-6	$N^{\omega}$ -(4-Toluenesulfonyl)-L-arginine	
L03	30	5963-60-0	L-Glutamic acid- $\gamma$ -anilide	
L04	196	5963-60-0	L-Glutamic acid- $\gamma$ -anilide	
L05	16	336182-11-7	L- $\beta$ -Homohydroxyproline hydrochloride	
L06	84	1049743-68-1	( <i>R</i> )- $\gamma$ -(4-Trifluoromethylbenzyl)-L-proline	
L07	115	51-35-4	<i>trans</i> -L-4-Hydroxyproline	
L08	219	152286-30-1	7-hydroxy-D-Tic-OH	
L09	305	56-41-7	L-Alanine	
L10	317	63-68-3	L-Methionine	
L11	330	72-18-4	L-Valine	
L12	321	312-84-5	D-Serine	
L13	337	632-20-2	D-Threonine	
L14	166	331763-65-6	3-Fluoro-D- $\beta$ -homophenylalanine hydrochloride	
L15	180	14464-67-6	3-Trifluoromethyl-D-phenylalanine	

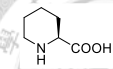
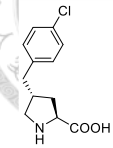
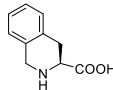
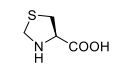
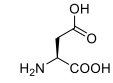
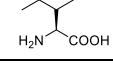
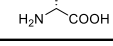
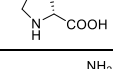
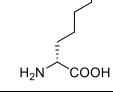
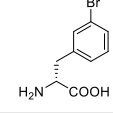
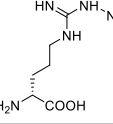
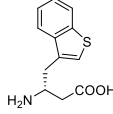
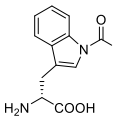
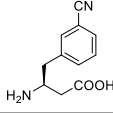
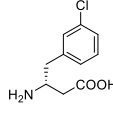
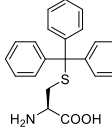
L16	250	31202-69-4	<i>N</i> <sup>ε</sup> -Boc-D-lysine	
L17	274	1723-00-8	D-Homoproline	
L18	49	151911-32-9	( <i>S</i> )-3-Amino-3-(2-fluorophenyl)propionic acid	
L19	210	270065-91-3	(2-Thienyl)-L-β-homoalanine hydrochloride	
L20	14	84888-34-6	<i>S</i> -9-Fluorenylmethyl-L-cysteine hydrochloride	
L21	346	18125-46-7	4-Fluoro-D-phenylalanine	
L22	353	62561-75-5	4-Iodo-D-phenylalanine	
L23	361	145306-65-6	D-3-Methoxyphenylalanine	
L24	370	177966-60-8	Fmoc-3-benzothienyl-L-alanine	
M01	221	80745-10-4	<i>N</i> <sup>ω</sup> -(4-Methoxy-2,3,6-trimethylbenzenesulfonyl)-L-arginine	
M02	15	2177-63-1	L-Aspartic acid β-benzyl ester	
M03	31	1119-33-1	L-Glutamic acid-γ-ethyl ester	
M04	198	67953-08-6	L-Glutamic acid-γ-(p-nitroanilide) hydrochloride	
M05	17	213475-47-9	L-2,2-Dimethyl-thiaproline hydrochloride	

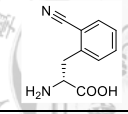
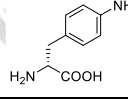
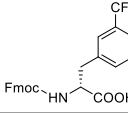
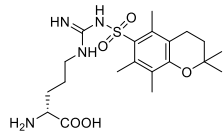
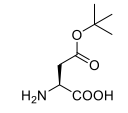
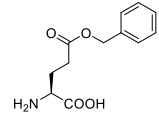
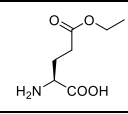
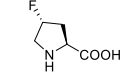
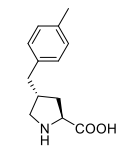
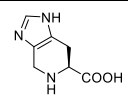
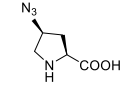
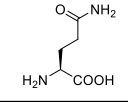
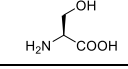
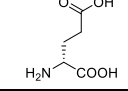
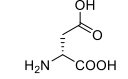
M06	85	1049740-11-5	( <i>R</i> )- $\gamma$ -(3-Nitrobenzyl)-L-proline-HCl	
M07	116	4043-88-3	3,4-Dehydro-L-proline	
M08	264	618-27-9	<i>cis</i> -L-4-Hydroxyproline	
M09	306	74-79-3	L-Arginine	
M10	318	63-91-2	L-Phenylalanine	
M11	334	52-90-4	L-Cysteine	
M12	328	556-02-5	D-Tyrosine	
M13	338	348-67-4	D-Methionine	
M14	167	263396-43-6	3-Cyano-D-phenylalanine	
M15	185	20312-37-2	D- $\alpha$ -Hydroxyisocaproic acid	
M16	251	201014-19-9	<i>N</i> <sup>ε</sup> -2-Chloro-Z-D-lysine	
M17	275	45521-09-3	D-Thiaproline	
M18	171	270596-50-4	3-Fluoro-L- $\beta$ -homophenylalanine hydrochloride	
M19	N/A	N/A	N/A	N/A
M20	21	144317-20-4	<i>N</i> <sup>γ</sup> -4-Methyltrityl-L-asparagine	
M21	347	56613-61-7	4-Nitro-D-phenylalanine monohydrate	

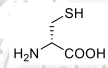
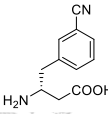
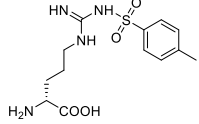
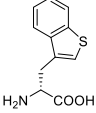
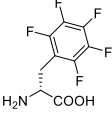
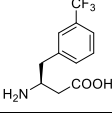
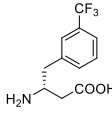
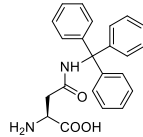
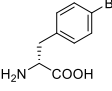
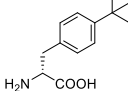
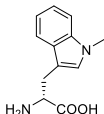
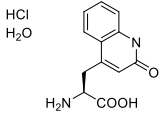


M22	354	130930-49-3	H-D-Phe(2-trifluoromethyl)-OH	
M23	362	80126-54-1	2-Methyl-D-phenylalanine	
M24	371	205526-27-8	Fmoc-3-trifluoromethyl-L-phenylalanine	
N01	222	121080-96-4	D-Homocitrulline	
N02	195	112259-66-2	L-Aspartic acid-β-cyclohexyl ester	
N03	91	2922-83-0	L-Kynurenine	
N04	199	1499-55-4	L-Glutamic acid-γ-methyl ester	
N05	18	53912-85-9	L-β-Homoproline hydrochloride	
N06	86	1049745-26-7	(R)-γ-(4-Biphenylmethyl)-L-proline-HCl	
N07	193	128502-56-7	7-hydroxy-L-Tic-OH	
N08	272	79815-20-6	L-Indoline-2-carboxylic acid	
N09	307	5794-13-8	L-Asparagine monohydrate	
N10	320	147-85-3	L-Proline	
N11	339	56-87-1	L-Lysine	
N12	329	640-68-6	D-Valine	
N13	340	319-78-8	D-Isoleucine	

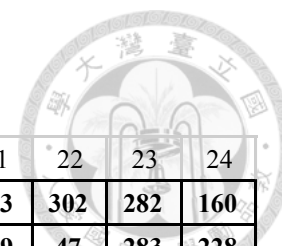
N14	168	110117-84-5	3-Fluoro-D-phenylalanine	
N15	186	17407-56-6	D-α-Hydroxyisovaleric acid	
N16	252	274260-42-3	N <sup>ε</sup> -Allyloxycarbonyl-D-lysine	
N17	285	201290-11-1	N <sup>in</sup> -Boc-D-tryptophan	
N18	173	719995-40-1	(S)-3-Amino-3-(3-trifluoromethylphenyl)propionic acid	
N19	175	269398-82-5	3-Methyl-D-β-homophenylalanine hydrochloride	
N20	94	84624-28-2	N <sup>ε</sup> -Fmoc-L-lysine	
M21	N/A	N/A	N/A	N/A
N22	355	39878-65-4	O-Methyl-D-tyrosine	
N23	363	267225-27-4	2-Bromo-D-phenylalanine	
N24	372	96402-49-2	Fmoc-3-(1-naphthyl)-L-alanine	
O01	223	112160-37-9	N <sup>ω</sup> -(2,2,5,7,8-Pentamethylchroman-6-sulfonyl)-L-arginine	
O02	201	16856-13-6	L-Aspartic acid-β-methyl ester hydrochloride	
O03	104	1118-90-7	L-α-Amino adipic acid	
O04	207	122864-94-2	L-Glutamic acid 5-(4-nitroanilide) monohydrate	

O05	19	3105-95-1	L-Homoproline	
O06	87	1049733-88-1	( <i>R</i> )- $\gamma$ -(4-Chlorobenzyl)-L-proline	
O07	202	74163-81-8	(3 <i>S</i> )-1,2,3,4-Tetrahydroisoquinoline-3-carboxylic acid	
O08	273	34592-47-7	L-Thiaproline	
O09	308	56-84-8	L-Aspartic acid	
O10	322	73-32-5	L-Isoleucine	
O11	304	338-69-2	D-Alanine	
O12	331	344-25-2	D-Proline	
O13	341	923-27-3	D-Lysine	
O14	169	99295-78-0	3-Bromo-D-phenylalanine	
O15	226	66036-77-9	<i>N</i> <sup>ω</sup> -Nitro-D-arginine	
O16	254	269398-95-0	D- $\beta$ -HomoAla(3-benzothiényl)-OH	
O17	286	367453-01-8	<i>N</i> <sup>in</sup> -Formyl-D-tryptophan	
O18	174	270065-85-5	3-Cyano-L-beta-homophenylalanine	
O19	178	331763-55-4	3-Chloro-D-beta-homophenylalanine	
O20	120	2799-07-7	<i>S</i> -Trityl-L-cysteine	

O21	348	263396-41-4	2-Cyano-D-phenylalanine	
O22	356	126257-07-6	4-Amino-D-phenylalanine	
O23	N/A	N/A	N/A	N/A
O24	373	205526-28-9	Fmoc-3-trifluoromethyl-D-phenylalanine	
P01	224	191869-60-0	<i>N</i> <sup>ω</sup> -(2,2,5,7,8-Pentamethylchroman-6-sulfonyl)-D-arginine	
P02	206	3057-74-7	L-Aspartic acid-β-tert-butyl ester	
P03	143	1676-73-9	L-Glutamic acid-γ-benzyl ester	
P04	208	1119-33-1	L-Glutamic acid-γ-ethyl ester	
P05	38	2507-61-1	<i>trans</i> -4-Fluoro-L-proline	
P06	88	1049734-62-4	( <i>R</i> )-γ-(4-Methylbenzyl)-L-proline	
P07	203	59981-63-4	L-4,5,6,7-Tetrahydro-1H-imidazo[4,5-c]pyridine-6-carboxylic acid	
P08	278	892128-58-4	(4 <i>S</i> )-Azido-L-Proline	
P09	311	56-85-9	L-Glutamine	
P10	323	56-45-1	L-Serine	
P11	309	6893-26-1	D-Glutamic acid	
P12	332	1783-96-6	D-Aspartic acid	

P13	342	921-01-7	D-Cysteine	
P14	170	269726-82-1	3-Cyano-D-β-homophenylalanine	
P15	227	97233-92-6	<i>N</i> <sup>ω</sup> -(4-Toluenesulfonyl)-D-arginine	
P16	256	111139-55-0	3-Benzothieryl-D-alanine	
P17	299	40332-58-9	Pentafluoro-D-phenylalanine	
P18	179	270065-76-4	3-Trifluoromethyl-L-β-homophenylalanine	
P19	181	269726-73-0	3-Trifluoromethyl-D-β-homophenylalanine	
P20	213	132388-58-0	<i>N</i> <sup>γ</sup> -Trityl-L-asparagine	
P21	349	62561-74-4	4-Bromo-D-phenylalanine	
P22	357	274262-82-7	4-tert-Butyl-D-phenylalanine	
P23	365	110117-83-4	1-Methyl-D-tryptophan	
P24	374	5162-90-3	3-(2-Oxo-1,2-dihydro-4-quinoliny)alanine hydrochloride monohydrate	

\*N/A designates an empty well



**Table S2. Arrangement of ncAA Library.**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	O	O	5	130	11	253	69	163	81	51	235	35	197	156	50	152	117	268	111	296	103	302	282	160
B	O	O	6	131	36	101	99	241	82	59	53	52	211	157	56	153	118	269	138	45	189	47	283	228
C	O	O	55	190	37	126	134	243	83	133	191	64	262	289	58	154	119	303	139	46	234	68	284	229
D	IP	IP	259	300	54	127	135	165	200	98	25	123	287	290	90	205	140	121	187	78	238	110	28	231
E	IP	IP	258	57	63	128	61	74	95	132	26	129	288	291	92	215	141	44	260	79	292	247	122	244
F	IP	IP	257	8	67	298	60	75	96	149	27	136	72	7	93	13	142	106	270	80	293	248	145	249
G	3	255	124	9	65	301	71	76	97	184	32	137	73	22	150	233	246	107	108	100	294	261	146	263
H	62	4	105	10	66	12	70	77	24	232	34	162	125	23	151	20	266	109	188	102	367	281	147	297
I	33	225	217	144	212	39	89	204	279	312	324	310	333	326	172	236	265	42	182	277	112	350	358	366
J	148	230	242	155	214	40	113	209	280	314	325	315	335	161	176	237	267	43	183	1	344	351	359	368
K	158	239	29	192	220	41	114	218	313	316	327	319	336	164	177	245	271	48	194	2	345	352	360	0
L	159	240	30	196	16	84	115	219	305	317	330	321	337	166	180	250	274	49	210	14	346	353	361	370
M	221	15	31	198	17	85	116	264	306	318	334	328	338	167	185	251	275	171	0	21	347	354	362	371
N	222	195	91	199	18	86	193	272	307	320	339	329	340	168	186	252	285	173	175	94	343	355	363	372
O	223	201	104	207	19	87	202	273	308	322	304	331	341	169	226	254	286	174	178	120	348	356	0	373
P	224	206	143	208	38	88	203	278	311	323	309	332	342	170	227	256	299	179	181	213	349	357	365	374

\*O and IP set as control experiments. O: without ncAA and IPTG; IP: without ncAA.