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

利用擴充基因密碼方法合成泛素-SUMO2 異元二聚體與其功能分析之研究

Expanding genetic code for synthesis and functional studies of Ubiquitin-SUMO2 heterodimers

李健隆

Chien-Lung Li

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

doi:10.6342/NTU201803256

Abstract

Protein Ubiquitination and SUMOylation are unique protein post-translational modifications involved in a myriad biological processes. The discovery of USP7 as exclusive SUMO deubiquitinase has manifested the cross-talk between Ubiquitin (Ub) and SUMO. However, the corresponding effectors and biological language of such hetero-linkage are still indecipherable. Here, we report an unbiased novel chemical method in synthesizing Ub-SUMO2 dimers. Se-alkylselenocysteine was incorporated into SUMO2 protein at eight different lysine positions individually and further converted to SUMO2-dehydroalanine (Dha) through selenoxide-elimination chemistry. Subsequently, SUMO2-Dha was covalently tethered with Ub-ethylthio by thiol-Michael addition yielding Ub-SUMO2 heterodimer.

This study also centers on the development of sensitive heterocyclic fluorescent probe which can sense Di-Ub or Ub-SUMO linkage in response to environmental changes. With the evolved pyrrolysyl-tRNA synthetase • pyrrolysyl-tRNA (PylRS•$tRNA_{CUA}^{Pyl}$) pairs, the tyrosine 66 and phenylalanine 27 positions of superfolder green fluorescent protein were efficiently substituted with seven heterocyclic noncanonical amino acids. The photophysical characterization of these proteins have unraveled the significant differences at the emission wavelengths, stoke shift, intramolecular FRET, and additional chromophore. These findings have demonstrated that ncAA-based tool was employed to establish molecular probe which can specifically recognize Ubiquitin dimers.

Keywords：Expanding genetic code、Ubiquitin-SUMO2 dimer、thiol-Michael addition reaction、heterocyclic noncanonical amino acid incorporation、PylRS•$tRNA_{CUA}^{Pyl}$ pair

doi:10.6342/NTU201803256

摘要

泛素化和類泛素化為細胞內重要之蛋白質轉譯後修飾作用，廣泛參與並影響生物體作用。USP7 近期發現亦為泛素-類泛素降解酶，能將鍵結在類泛素 2 上的泛素移除，暗示泛素-類泛素的動態鍵結之間有著調控受質蛋白功能的生物機制。然而，調控類泛素-泛素異元聚合物轉譯後修飾的其餘調控因子，尚未進一步發現及探討。本研究利用化學生物學方法及非典型胺基酸體內嵌入的技術，合成特定鍵結的泛素-類泛素異元二聚體，做為深入探討相關蛋白間作用的基體。本研究中，類泛素 2 中的八個賴胺酸基因位點，分別以 TAG 基因密碼突變並以生物正交配對 PylRS•$tRNA_{CUA}^{Pyl}$體內嵌入非典型胺基酸 SeCbzK；類泛素 2 進一步經過硒亞碸脫去反應轉換成類泛素 2-Dha，進而衍生化的泛素-ethylthio 進行硫醇麥可加成反應，得到泛素-類泛素 2 異元二聚體產物。

此論文研究亦針對以雜環類非典型胺基酸中對環境敏感而有螢光強弱變化的特性，發展辨認特定鍵結泛素二聚體的分子探針。利用演化後的 PylRS•$tRNA_{CUA}^{Pyl}$正交配對，在超折疊綠螢光蛋白的發色團中酪胺酸 66 及苯丙胺酸 27 分別取代為 7 種不同雜環胺基酸。在螢光光譜的測定實驗中，顯示放射光譜和控制組實驗相比，有著顯著發射光譜位移、斯托克斯位移、分子內多重 FRET 光學現象、增加螢光發光團，相關結果可做為非典型胺基酸分子探針分辨特定泛素二聚體的工具。

關鍵字: 非典型胺基酸體內嵌入、泛素-類泛素異元二聚體、硫醇麥可加成反應、雜環非典型胺基酸、PylRS•$tRNA_{CUA}^{Pyl}$配對

doi:10.6342/NTU201803256

Table of contents

Pages

doi:10.6342/NTU201803256

doi:10.6342/NTU201803256

doi:10.6342/NTU201803256

doi:10.6342/NTU201803256

List of figures

doi:10.6342/NTU201803256

doi:10.6342/NTU201803256

doi:10.6342/NTU201803256

doi:10.6342/NTU201803256

List of tables

doi:10.6342/NTU201803256

Abbreviations

Ala, A	L-alanine
AARS	aminoacyl-tRNA synthetase
APS	ammonium persulfate
bp	base pair
CV	column volume
Cys, C	L-cysteine
DNA	deoxyribonucleic acid
DTT	1,4-dithiothreitol
E. coli.	*Escherichia coli*
EF-G	elongation Factor-G
EF-Ts	elongation factor-thermo stable
EF-Tu	elongation factor-thermo unstable
ESI-MS	electrospray ionization mass spectrometry
EtBr	ethidium bromide
Gln Q	L-glutamine
Glu, E	L-glutamate
Gly, G	L-glycine
His, H	L-histidine
Histag	hexahistidine tag
IPTG	isopropyl β-D-1-thiogalactopyranoside
kDa	kilo Dalton
LB	lysogeny broth
Leu, L	L-leucine

doi:10.6342/NTU201803256

Lys, K	L-lysine
MALDI-TOF-MS/MS	Matrix-assisted laser desorption ionization-time of flight - tandem mass spectrometry
Met, M	L-methionine
*Mm*PylRS	*Methanosarcina mazei* pyrrolysyl-tRNA synthetase
mRNA	messenger RNA
ncAA	noncanonical amino acid
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBST	phosphate buffered saline with Tween-20
PBS	phosphate buffered saline
PCR	polymerase chain reaction
Phe, F	L-phenylalanine
PVDF	polyvinylidene difluoride
Pyl	pyrrolysine
RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
Sec	L-seleocysteine
SEC	size exclusion chromatography
Ser, S	L-serine
sfGFP	superfolder green fluorescent protein
Sp	spectinomycin
SUMO	small ubiquitin-like modifier
TCEP	tris(2-carboxyethyl)phosphine

doi:10.6342/NTU201803256

TEMED	tetramethylethylenediamine
Thr, T	L-threonine
tRNA	transfer RNA
Tyr, Y	L-tyrosine
UV	ultraviolet
Val, V	L-valine
wt	wild-type

doi:10.6342/NTU201803256

Chapter 1 Introduction

1.1 Protein translation

In 1960s, research in genetic code coding and decoding have initialized fascinated molecular genetics and molecular biology studies[1]. Current genetic code decoding system was composed with 4 basic nucleotide, A, T, G, and C; three nucleotide decoding rule; and 64 codon which code 20 canonical amino acids with codon degeneracy fashion. Three stop codons, TAG, TGA, and TAA, were evolved as protein translation termination signals.

DNA sequence from genome is the blueprint in producing proteins through interface of messenger RNA (mRNA), so calls DNA transcription and protein translation steps in maintaining dynamic living systems. The protein translation mechanism of prokaryotes is divided into three stages, peptide chain initiation, peptide chain elongation, and peptide chain termination. The mRNA, 21 different aminoacyl-tRNAs are required message and building blocks before protein synthesis in the ribosome.

1.1.1 Initiation

In all species, the polypeptide chains start at the N-terminus and form amide bond between two amino acid and release one molecule of water. In prokaryotes, the first amino acid to be installed at ATG codon is N-formyl-L-methionine (fMet)[2]. Formyl group of

doi:10.6342/NTU201803256

fMet, however, will be removed after finishing polyamide bond synthesis by peptide deformylase[3,4] at stage of protein posttranslational modifications. Polypeptide chain synthesis was initiated from formation of initiation complex, including mRNA, fMet-tRNA^{fMet}, 30S ribosomal subunit, GTP, and three initiation factors, IF-1, IF-2, IF-3. IF-3 factor promotes mRNA binding with 30S ribosomal subunit. Then, IF-1 binds with the 30S ribosome subunit at the A site and avoid other aminoacyl-tRNA entering. On the other hand, IF1 modulates IF2 binding to the ribosome by increasing affinity and may block the 50S subunit binding, preventing the formation of the 70S subunit. Later, IF-2 binds GTP and help to bring fMet-tRNA^{fMet} from all acylated tRNAs specifically. With GTP hydrolysis to GDP, IF2 facilitates 50S ribosomal subunit binds to 30S ribosomal subunit. Subsequently, IF3 departs the 30S and the 50S subunit will bind with 30S, when the 30S initiation complex has been assembled. At the same time, the 70S initiation complex generate to lead IF1 and IF2 released.

1.1.2 Elongation

In protein elongation stage, three aminoacyl-tRNAs binding sites, aminoacyl (A) site, peptidyl (P) site, and exit (E) site, are formed in the space in 50S/30S complex, 70S ribosome. Localized fMet-tRNA^{fMet} meets followed aminoacyl-tRNA at A site, which was companied with elongation factor-thermo unstable (EF-Tu) and then release by GTP

doi:10.6342/NTU201803256

hydrolysis[5] with correct codon/anticodon decoding of mRNA/tRNA pairing process. EF-Tu•GTP is further regenerated by elongation factor- thermo stable (EF-Ts). The first and more amide bonds are formed at P site and generate a water molecule. Uncharged tRNA at P site then exit to plasma from E site and generated nascent peptide chains come out from extended channel of E site. Sequential corresponding aminoacyl-tRNA will be accommodated at empty A site. Elongation factor G (EF-G), GTPase, catalyzes the aminoacyl-tRNA/mRNA translocation in A to P site movement. Polypeptide chain will be extended until meet one of three stop codons, UAG (Amber), UAA (Ochre), and UGA (Opal).

1.1.3 Termination

A stop signal is essential to the termination of protein synthesis. The codons UAG (Amber), UAA (Ochre), and UGA (Opal) are the stop signal. These stop codons are not recognized by tRNA, but the stop codons are recognized by proteins known as release factor (RF). RF1 recognizes UAA or UAG and RF2 recognizes UAA or UGA[6]. The completed polypeptide chain is released from the last tRNA by peptidyl transferase. When the stop codons binds with RF1 or RF2, RF1 or RF2 activates the peptidyl transferase which hydrolyzes the ester bond between the completed polypeptide chain and the tRNA

doi:10.6342/NTU201803256

in the P site. Then, RF3 releasesRF1 and RF2 from the ribosome utilizing GTP as an energy source.

1.2 Protein post-translational modifications

Protein post-translational modifications (PTMs) on modifying an existing functional group of amino acid play a significant role in eukaryotic systems. Moreover, the diversity connections between them also control several biological processes. Protein PTMs are generally various, there are a lot of chemical modifications, like protein acetylation, methylation, hydroxylation, phosphorylation, nitrosylation or glycosylation while small protein modifications: for example, protein Ubiquitination or SUMOylation[7].

1.2.1 Protein SUMOylation

The small ubiquitin-like modifier (SUMO) plays a crucial role in protein post-translational modifications and modulates protein functions in eukaryotic systems such as altered localization and activity in the cell, or regulate protein-protein interactions in signaling pathways. SUMO proteins are highly conserved in the eukaryotes (**Figure 1**). The mammalian SUMO family is classified into three different isoforms, SUMO-1, SUMO-2, and, SUMO-3. To be more definite, SUMO-1 protein contains 97 amino acids;

doi:10.6342/NTU201803256

SUMO2 protein is composed of 92 amino acids and SUMO3 protein consists of 93 amino acids[8]. Moreover, abundant research evidence has implied that SUMO1 prefer to form monomeric chain types whereas SUMO2/3 prefer to form polymeric chain types. SUMOylation is carried out by 3 enzymes termed, E1 SUMO-activating enzyme (SAE1/SAE2 heterodimer), E2 SUMO-conjugating enzyme (Ubc9) and SUMO E3 ligase cascade in transferring SUMO molecules tethering to target proteins by forming isopeptide linkage between lysine residues of target protein via its C-terminal Glycine residue[9] (**Figure 2**). Like Ubiquitination, SUMOylation is a reverisible process carried out by SUMO specific proteases[10]. Polymeric SUMO chains have been observed in some organism like *S. cerevisiae* but their functions remain in the veil[11]. Initially it was common notion that SUMOylation and Ubiquitination are not complementary and assist each other functions as their activating and conjugating enzymes are different. But now days the cross-talk between two systems has been enormously studied. In future this studies will unveil the more insight into their intricate biological functions[12,13].

doi:10.6342/NTU201803256

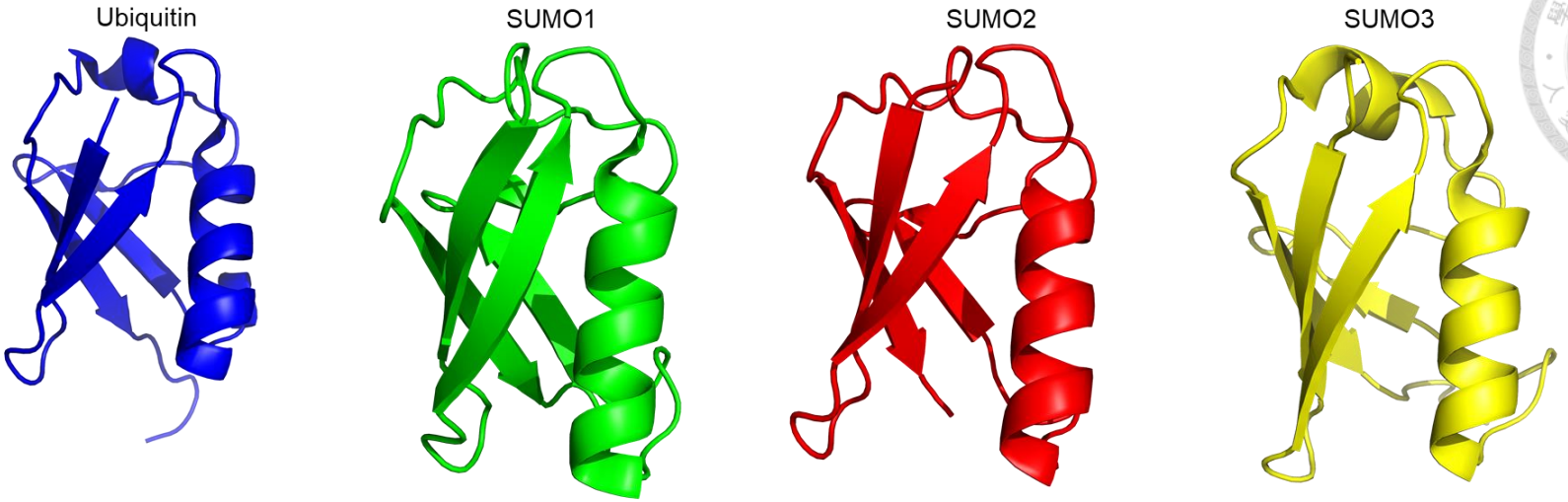


Figure 1. The structure of Ubiquitin and SUMO from *Homo sapiens*.
Drawn using PyMOL, and the PDB code is 1D3Z, 2N1V, 1WM3, and 1U4A.

doi:10.6342/NTU201803256

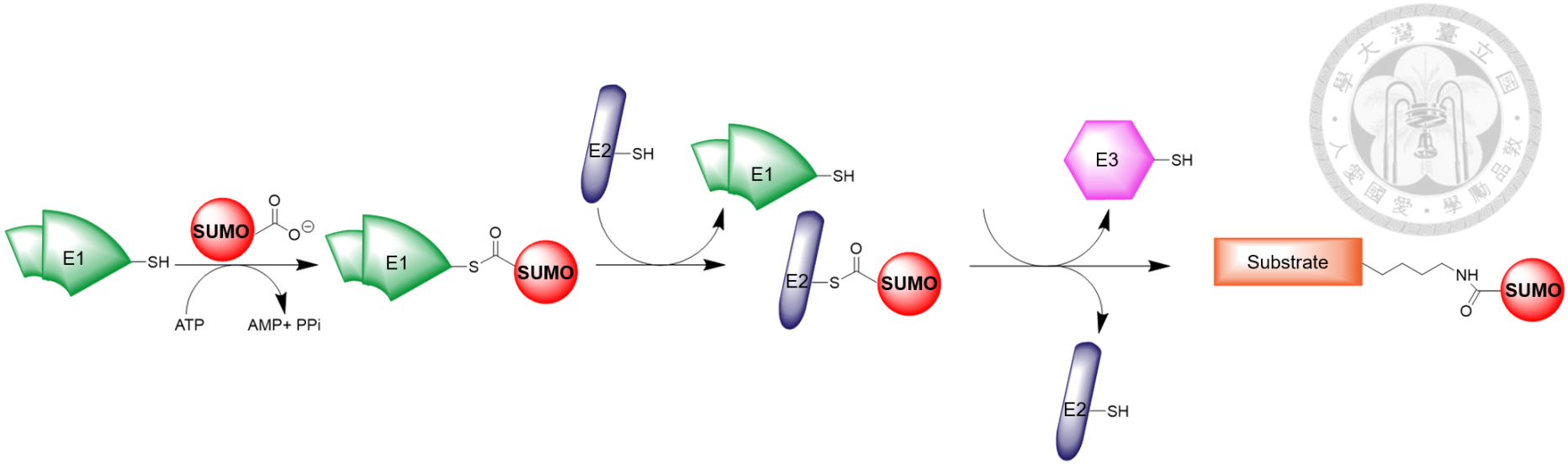


Figure 2. The SUMO is conjugated to substrates via enzymatic cascade.

Protein SUMOylation through enzymatic method. Protein SUMOylation go through E1, E2 and E3 enzymatic cascades to attach SUMO on substrate protein through lysine linkage. By using enzymatic method, products probably cannot control the SUMOylation site.

doi:10.6342/NTU201803256

1.2.2 Protein Ubiquitination

Ubiquitination, a 76 amino acid protein[14], plays a crucial role in regulating myriad of biological processes other than its well-known function of tagging proteins for degradation by 26S proteasome[15-17] for example, in response to DNA damage[18], endocytosis[19], protein trafficking to the lysosome/vacuole and, signal transduction[20]. β−grasp fold and a flexible C-terminal tail characteristics promotes stability of protein. Especially, β1/β2 loop contains Leu8 residues which indicated flexibility is essential for identifying binding proteins. Then, Ubiquitin is constantly recognized by hydrophobic region Ile44, Val70, and His68. The Ile44 binds with most UBDs and the proteasome. Another hydrophobic surface Ile36 dominates Ubiquitin chain interactions and is recognized by DUBs, UBDs, and HECT E3[21]. Basically, Ubiquitination utilizes E1, E2, and E3 enzymatic cascades to form isopeptide bond between C-terminal of Ub and lysine residues of the substrates (**Figure 3**). The process is catalyzed by a Ub-activating enzyme (E1), Ub-conjugating enzyme (E2), and Ub ligase (E3). E1 catalyzes two steps reaction to active Ub, Firstly, Ub and ATP form Ub-AMP, afterwards, this intermediate reacts with the E1 active-site cysteine to form thioester bond. E2 active-site cysteine connects with Ub from the E1-Ub complex and transfers Ub to the substrate, sometimes the final step is catalyzed by Ubiquitin ligase[22]. Further, the 7 lysine residues and N-ter of Ubiquitin also get ubiquitaneted to form polymeric long chain. This chains are divided in Homotopic chains and Heterotopic chains. Homotopic chains consist of same lysine isopeptide bond while Heteropic chains are conjugated by different lysine[23] residues. It is also possible that diiferent lysines of substrates and even Ubiqutin can be further polymerised to form different topological stuctures. Ubiquitin chain type triggers various

doi:10.6342/NTU201803256

consequences in the biological process. This intricacies in Ubiquitin chains are referred to as “Ubiquitin Code”[24] (**Figure 4**).

doi:10.6342/NTU201803256

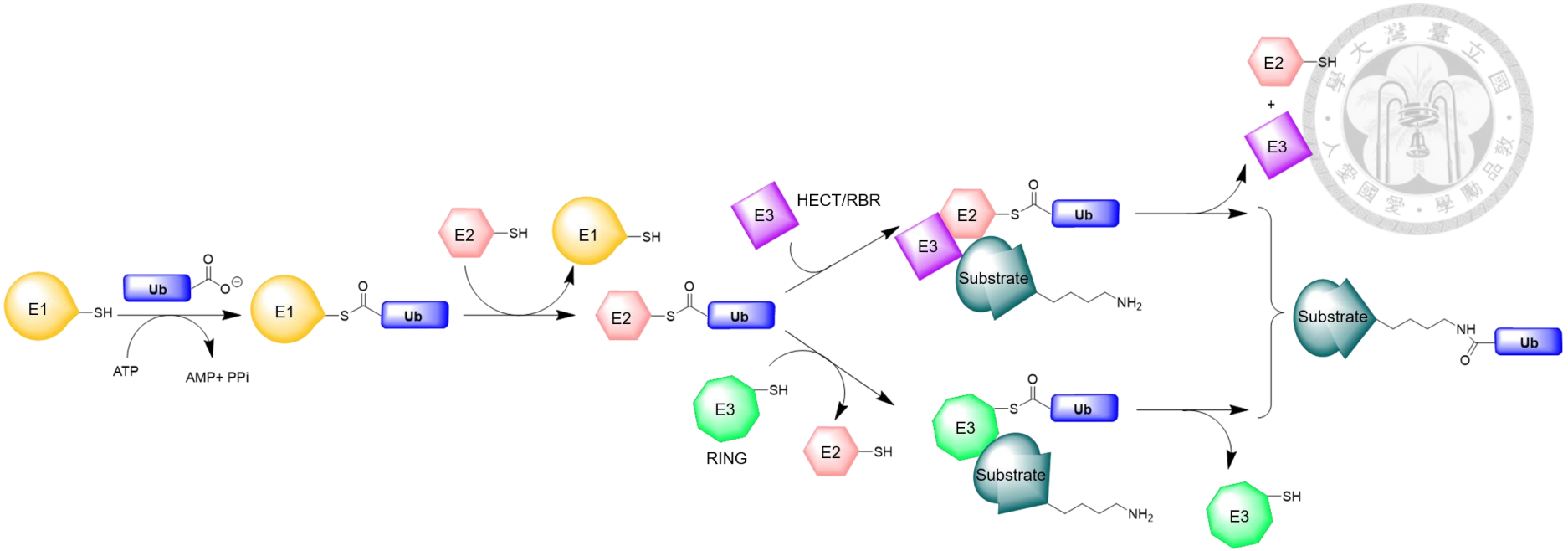


Figure 3. The Ubiquitin reaction cascade utilizes enzymes to modify substrates.

Protein Ubiquitination through enzymatic method. Protein Ubiquitination is carried out by 3 cascading multi-enzyme termed E1, E2 and E3 which has been used *in vitro* synthesis but suffered by non-specific lysine modifications. This approach generates multiple and mixed protein Ubiquitination product.

doi:10.6342/NTU201803256

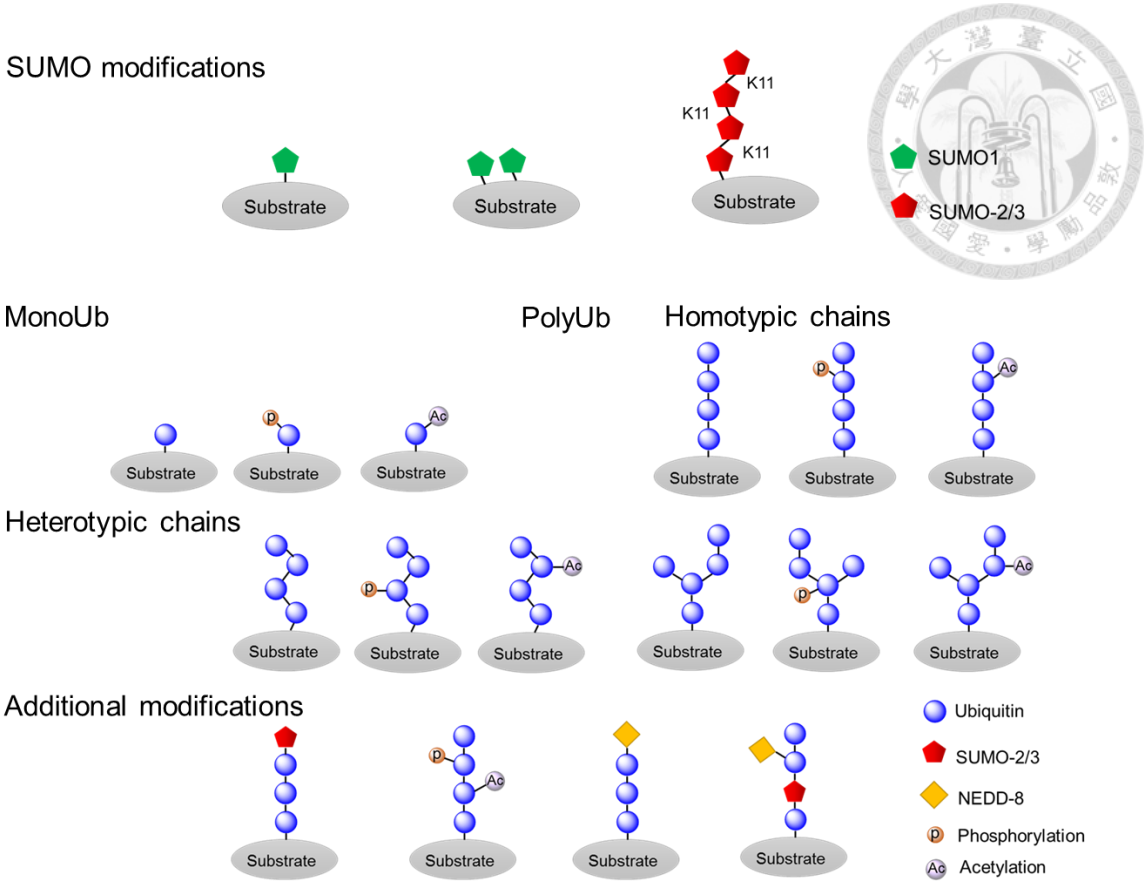


Figure 4. The complexity of Ubiquitin and SUMO modifications [8,24].

doi:10.6342/NTU201803256

1.3 Expanding genetic codes

1.3.1 Genetic code reassignment

The genetic codes are correlated with the same specific amino acids in all organisms. With the natural exception of Sec [25] and Pyl [26] (**Figure 5**), among the 20 amino acids are encoded through corresponded codes. These 20 amino acids set up the sophisticated balance on structure and function of protein under biological condition. However, the ability of canonical amino acids restrict the development of organisms and properties of proteins. Thus, to address this problem, the strategy of development of a method to incorporate ncAA into protein is crucial. Sixty one of the 64 unique codons encode the 20 canonical amino acids, the other three function as stop codon, TAA, TAG, and TGA. Among the three stop codons are deficient in a corresponding tRNA and instead bind to termination factors. Especially, the amber codon (TAG) is the least used in *Escherichia coli* and yeast. Therefore, use of the amber codon is the most appropriate candidate for codon reassignment to ncAA. To utilize the amber codon to encode ncAA, it is necessary for a suppressor tRNA and a cognate aminoacyl-tRNA synthetase. Importantly, the tRNA synthetase pairs must maintain orthogonal in the host. Firstly, the orthogonal pairs build for application in *E. coli*, the most useful was the *Methanocaldococcus jannaschii* TyrRS•$tRNA^{Tyr}$ (*Mj*TyrRS•*Mj*$tRNA^{Tyr}$) pair that was derived from archaea[27,28]. Although the *Mj*TyrRS•*Mj*$tRNA^{Tyr}$ pair enable to encode common amino acids in *E. coli,* this pair is not orthogonal in eukaryotes. Furthermore, Current research has made effort to develop tRNA synthetase pairs can incorporate ncAA and are also orthogonal in eukaryotes.

doi:10.6342/NTU201803256

1.3.2 Reside-specific incorporation of ncAAs

In the 1950s, Cohen and coworkers indicated that methionyl-tRNA synthetase could recognize the ncAA, seleomethionine, and also incorporated into protein instead of methionine[29], called the residue-specific incorporation. Especially, the residue-specific methodology means canonical amino acids tolerate partial to replace by their noncanonical analogues[30]. Since the diverse ncAAs are incorporate into protein at various site, the following protein may unveil considerably distinct chemical and physical properties. For instance, labeling radioactive compounds can detect the protein synthesis under the post-translational modification or localization, and performing reactive functional group such as alkyne, azide. Both of which can be subjected to selective modification via click chemistry in protein. Besides, expanding the scope of residue-specific incorporation was also critical topics. Generally, AARS were mutated at the amino acid binding pocket or editing domain to recognize whether the new ncAA can be incorporated protein by wild-type AARS. Consequently, residue-specific incorporation of ncAAs into protein is an efficient approach to decipher the mystery in combination with chemistry and biology[31,32].

1.3.3 Site-specific incorporation of ncAAs

Substitution of a specific site represents the site-specific incorporation. This method is an optimal for analyzing point mutations into protein with minimal disorder of structure. The advantage of this approach for explaining details of protein function and stability, structure, and protein-protein interaction is incomparable[33,34]. Thus, to further enhance the ability to modulate the structure and properties of proteins, the researcher investigate to directly design organisms that genetically encode more ncAAs. In the 1989s, Chamberlin and Peter G. Schultz with their co-workers established new method to site-specifically incorporate ncAAs into proteins[35]. Especially, the strategy of synthesis

doi:10.6342/NTU201803256

protein with ncAA was a suppressor tRNA chemically acylated with ncAA, and its following recognized through the amber codon at the specific site. This chemical method restrains on aminoacyl-tRNA cause the yields of protein. Thus, biosynthetic methodology has been set up for overcoming the hurdles. In 2001, Lei Wang discovered an orthogonal AARS•tRNA pair in *E. coli*. In other words, this orthogonal AARS does not recognize the endogenous *E. coli* tRNAs and 20 amino acids under the existing translation system in *E. coli*. (**Figure 6**) The candidate is *Mj*TyrRS•*Mj*tRNA^{Tyr} pair from archaea[36]. In addition, this orthogonal pair has demonstrated it could efficiently incorporate *O*-methyl-L-tyrosine into protein at the amber codon (TAG). However, the ability of evolving AARSs to specifically incorporate amino acid determines on the structural limitations in the catalytic pocket sites. Hence, in order to incease the diversity of ncAAs, it was necessary to explore more orthogonal AARS•tRNA pairs. In 2002, the scientists discovered the 22nd natural amino acid, Pyl, and genetically encoded by UAG in *Methanosarcina barkeri*[37]. The PylRS• $tRNA^{Pyl}_{CUA}$ pair was utilized to specifically incorporate **1** into protein[38,39]. Afterwards, Jason W Chin and co-workers introduced PylRS• $tRNA^{Pyl}_{CUA}$ into expanding genetic codon system and genetically encoded N^{ε}-acetyllysine in recombinant proteins[40]. Since the PylRS• $tRNA^{Pyl}_{CUA}$ pair maintains exceptional orthogonality in the eukaryotes, the *Methanosarcina barkeri* (*Mb*PylRS)• $tRNA^{Pyl}_{CUA}$, *Methanosarcina mazei* PylRS (*Mm*PylRS)• $tRNA^{Pyl}_{CUA}$ and *Desulfitobacterium hafniense* PylRS (*Dh*PylRS)• $tRNA^{Pyl}_{CUA}$ pairs are commonly utilized in site-specific incorporating ncAA into protein[41]. However, the techniques were limited by low production, and were restricted the number of ncAAs incorporated into target protein. Owing to the suppressor tRNA compete with the release factors. In this study, we focus on *Mm*PylRS• $tRNA^{Pyl}_{CUA}$ pair engineering to incorporate ncAA[41].

doi:10.6342/NTU201803256

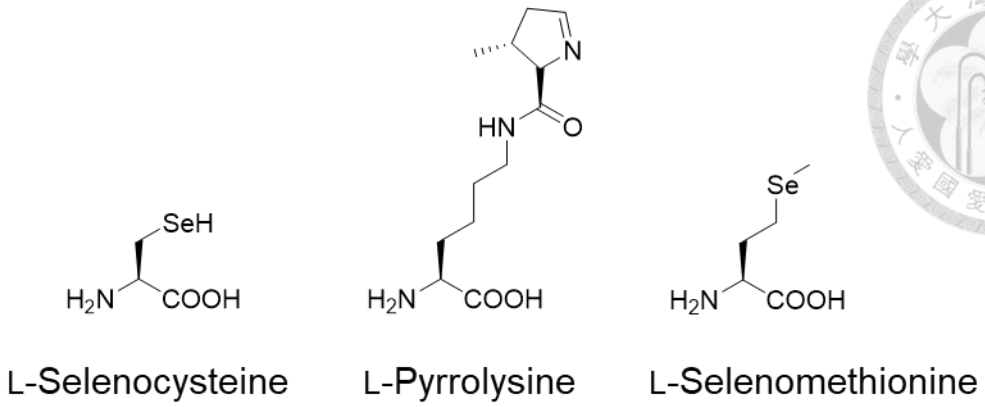


Figure 5. Chemical structures of natural amino acids.
The chemical structures of natural amino acids are L-Selenocysteine (Sec), L-Pyrrolysine (Pyl), and L-Selenomethionine (SeMet).

doi:10.6342/NTU201803256

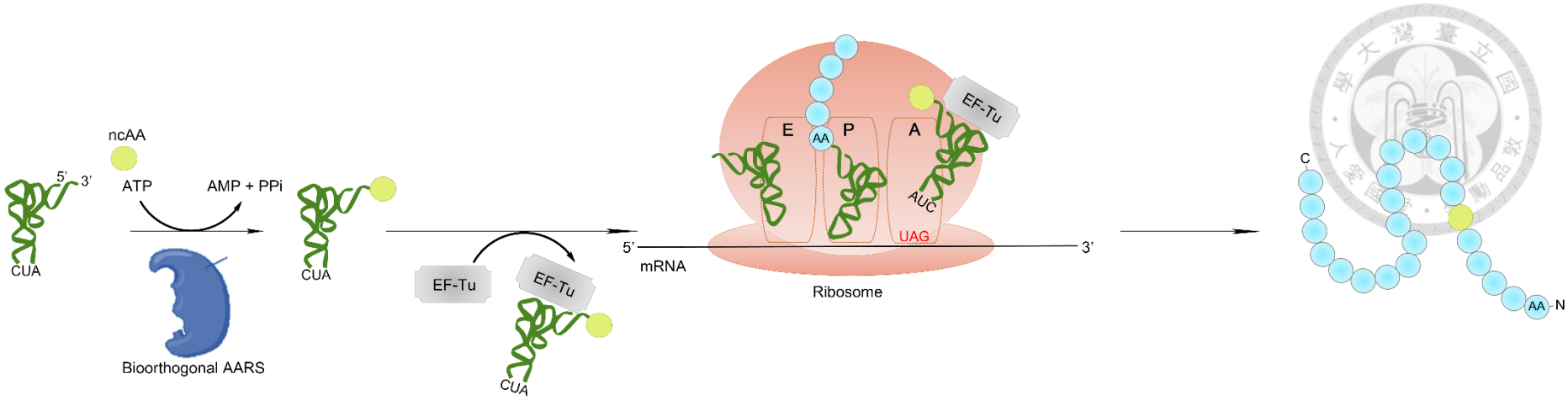


Figure 6. Illustration of protein synthesis with ncAA.

Protein translation system utilizes bioorthogonal AARS•$tRNA^{Pyl}_{CUA}$ pair to incorporate ncAA in response to the amber stop codon. The evolved bioorthogonal AARS•$tRNA_{CUA}$ pair (TyrRS•$tRNA^{Tyr}_{CUA}$ or PylRS•$tRNA^{Pyl}_{CUA}$) are required to apply into the translation system. PylRS is an aminoacyl-tRNA synthetase which derived from archaea and orthogonal in *E. coli*. Evolved PylRS catalyzes the esterification between suppressor tRNA and appropriate ncAA. The aminoacyl-tRNA is bought to ribosome with the assist of EF-TU and recognized by the corresponding amber codon on mRNA, generating the protein contains ncAA.

doi:10.6342/NTU201803256

1.4 Protein ligation

1.4.1 Native chemical ligation

Native chemical ligation is one of the powerful method used for chemical synthesis of protein and peptides. This reaction involves two chemical steps. One is the transthiolesterification. Since its first introduction in 1990s by Stephen Kent, Native chemical ligation (NCL) has been proven important techniques in large polypeptide synthesis[42]. The mechanism involves the N-terminal thiol group attack from cysteine of the unprotected peptide to the C-terminal thioester bond of another peptide at neutral pH. The intermediate further undergoes migration of S, N acyl shift to yield native peptide bond[43]. The main shortcoming of this method is prerequisite of cysteine at N-terminal of the peptide. To solve this limitation, recently, the various variations of NCL has been developed such as C-terminal Thioacid and the N-terminal Bromoacetamide residue as an alternative to conventional NCL[42].

1.4.2 Selenoxide elimination

Selenoxide elimination is a way for using selenoxides to synthesize alkenes. Generally, this chemical reaction is used to synthesize α,β-unsaturated carbonyl compounds.[44] The origin of sulfoxide elimination is a credible technique to generate alkene double bonds.[45] Subsequently, selenoxides were discovered going through similar process. Elimination of selenoxides takes place an intramolecular *syn* elimination pathway.[46] The carbon-hydrogen and carbon-selenium bonds maintain co-planar in the transition state. The most universal oxidizing agent are hydrogen peroxide, and sodium periodate. In addition, the hydrogen peroxide is frequently operated in excess, to overcome sensitive decomposition of hydrogen peroxide. But using hydrogen peroxide as oxidizing agent may generate undesired oxidation of starting material.

doi:10.6342/NTU201803256

1.4.3 Michael addition

The Michael addition reaction is an efficient synthetic methodology[47] for the nucleophile to α,β-unsaturated carbonyl compound.[48,49] This is a thermodynamically controlled reaction and one of the most useful methods for the moderate formation of C–C bonds. However, the Michael addition reaction involves the oxa-Michael reactions aza-Michael reactions, and the thiol-Michael reactions. In the 1960s, Allen reported this reaction and applied to synthesize and block polymer conjugation. In biochemistry, the site specific labelling of proteins such as post translational modification has a critical role in biological functions and protein drug discovery. The use of enzymatic methods is limited by enzyme availability, complexity, and purification hurdles[50]. The incorporation of non-canonical amino acid to perform novel chemistry is an effective and promising way to circumvent natural limitations. The site-specifically incorporated Phenylselenocysteine (PhSeCys) protein on treatment with H_2O_2 produces dehydroalanine which further undergoes Michael addition with a nucleophile such as a thiol at pH 8.0. This semisynthetic approach is better in yield and selectivity[51-53].

1.5 Specific aim of thesis

High abundance of SUMOylation and minimum Ubiquitinatiion of chromatin proteins are found near to replisome while opposite phenomenon is observed for resting chromatin. Further, this SUMOylated proteins undergoes complex Ubiquitination and deubiquitination. The interplay between the SUMOylation and Ubiquitination of chromatin proteins maintain the replisome activity. USP7, DUB enzyme, was found to play role in deubiquitination of SUMOylated chromatin protein. (**Figure 7**). Interestingly, deubiquitinase[22,54] USP7 can recognize SUMO2-Ub linkages at 11, and 33 positions [55].

doi:10.6342/NTU201803256

However, it is enigmatic to interpret the relationship between writers, readers, and erasers and their regulations of SUMOylated Ubiquitination in cells .Thus, this study aims to understand the biological meanings of different Ub-SUMO2 heterodimers. Here, we attempt to generate Ub-SUMO2 heterodimers with unbiased chemical synthesis (**Figure 8**). The expanding genetic code was explored to incorporate various ncAAs into protein *in vivo.* The specific aim is to interpret the biological consequences of different Ub-SUMO2 linkages and specificity of deubiquitinase.

Besides, the *Mm*PylRS was engineered to site-specifically incorporated environment sensitive heterocyclic fluorescent ncAAs into Ub dimers or Ub-SOUMO2 dimers which are envision to function as fluorescent probe (**Figure 9**). Fluorescence is switched after forming specific linkage with other Ub or SUMO2 proteins will assist to investigate the biological consequences of heterotopic linkages.

Finally, combining with these protein synthesis study, the study intend to comprehend the biological language and logic of protein poly-heteroUb/SUMOylation.

doi:10.6342/NTU201803256

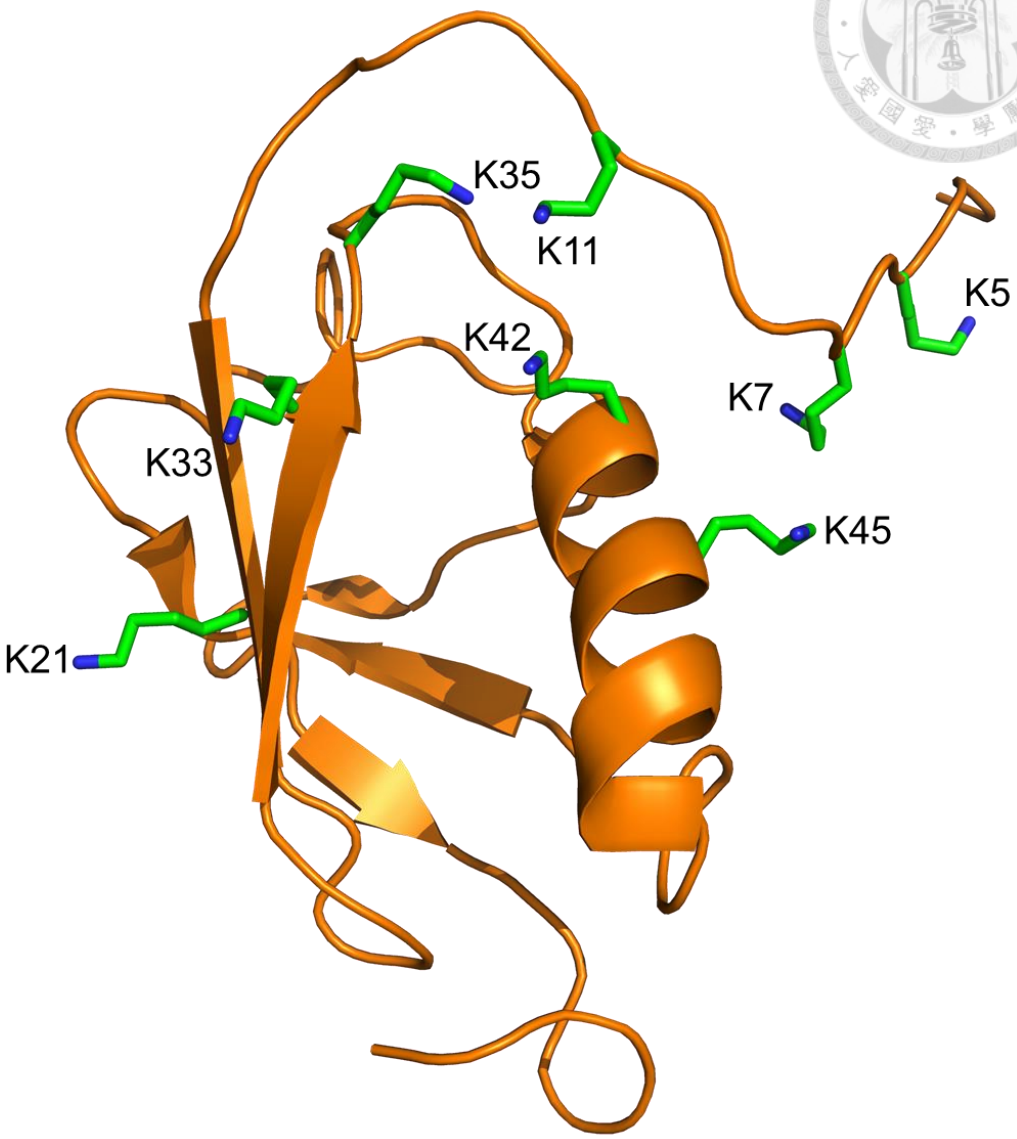


Figure 7. The structure of SUMO2 from *Homo sapiens*.
The eight lysines of SUMO2 were labeled with green stick mode at position K5, K7, K11, K21, K33, K35, K42, and K45, exhibit the residues that were mutated to amber codon for genetically incorporating SeCbzK. Drawn using PyMOL, and the PDB code is 2AWT.

doi:10.6342/NTU201803256

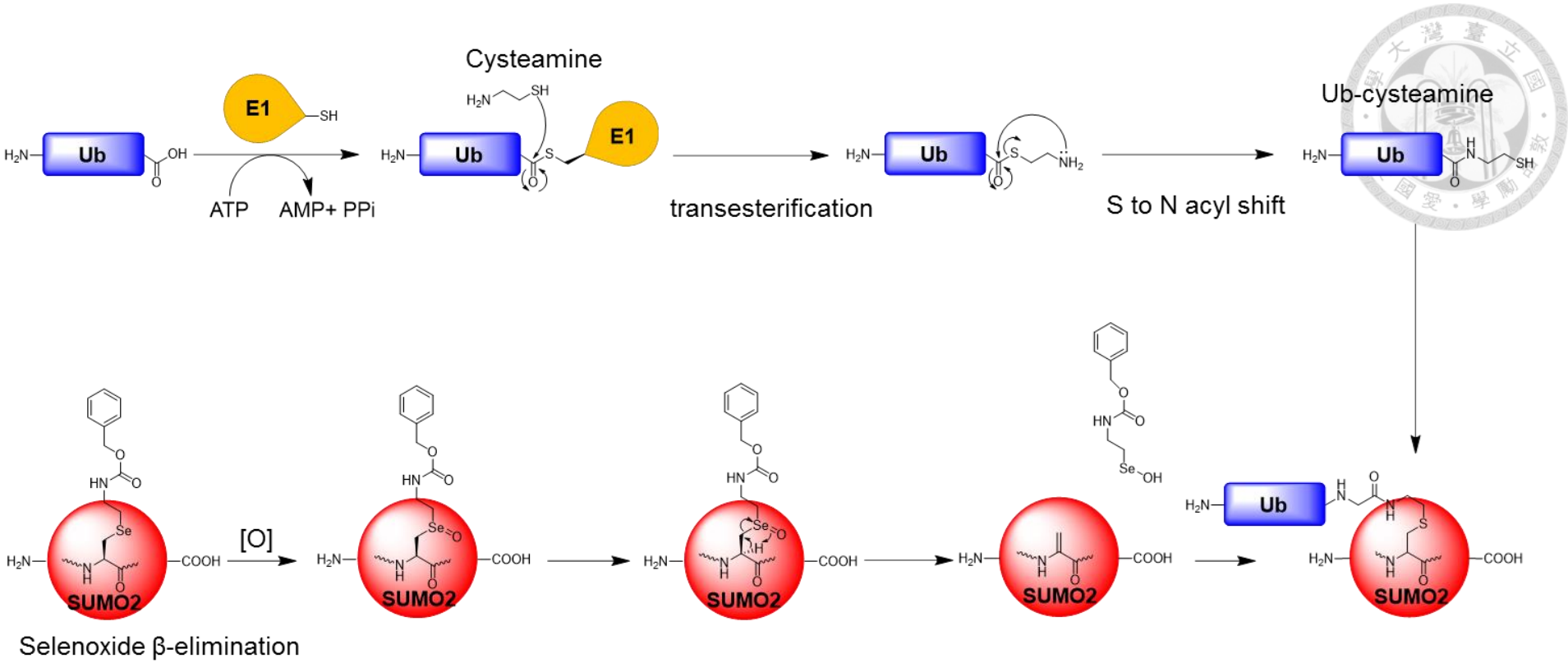


Figure 8. Ub-SUMO2 heterodimer synthesis scheme.

Ub-SUMO2 heterodimer synthesis by Michael addition reaction. ncAA SeCbzK was incorporated into proteins by evolved PylRS•tRNA$^{Pyl}_{CUA}$pair at different lysine positions with amber TAG codon mutation. Following dehydroalanine formation at SUMO2 can react with Ubiquitin with C-terminal ethylthio group, then form near-native isopeptide linkage.

doi:10.6342/NTU201803256

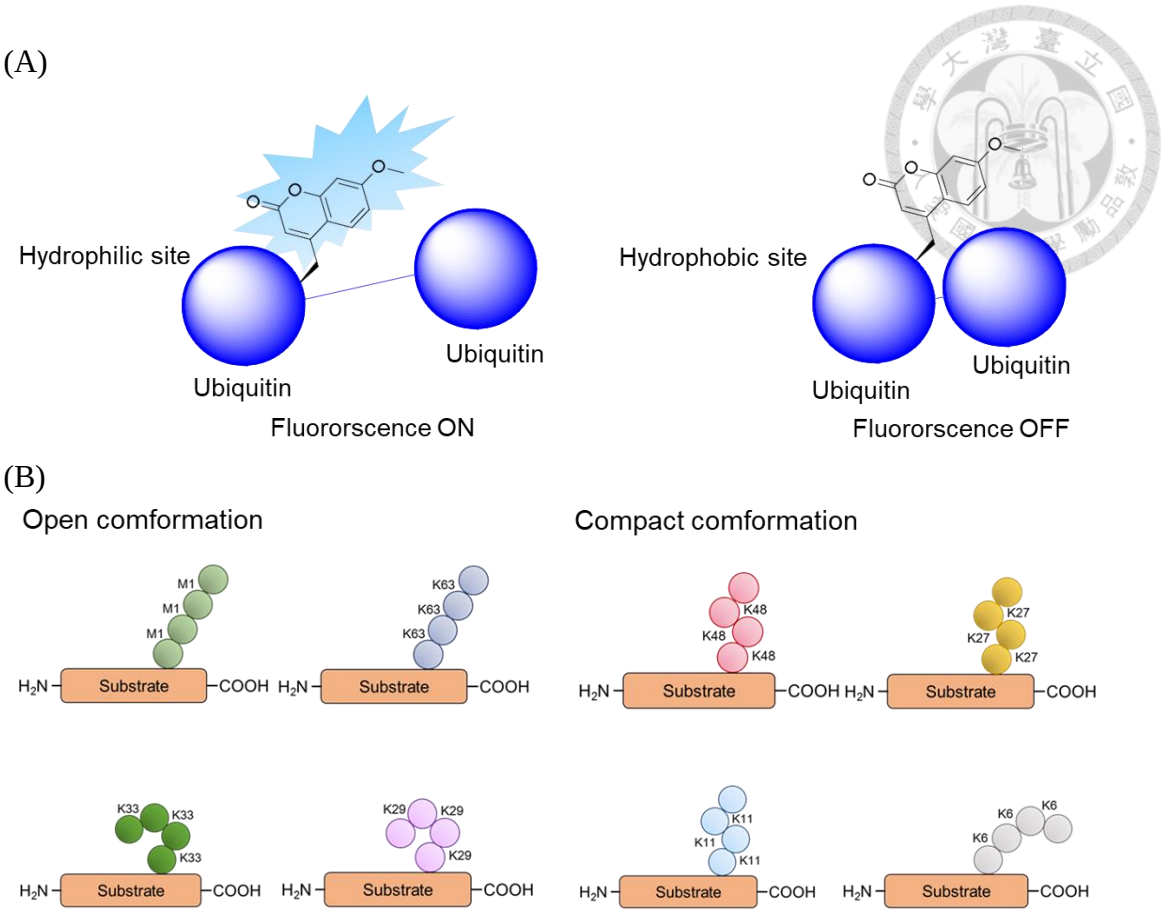


Figure 9. The probe recognizes specific linkage of Di-Ub or Ub-SUMO dimer.
(A) The heterocyclic ncAA was employed to probe specific Di-Ubiquitin. (B) Within homotypic polyubiquitination, Ubiquitin can utilize one of the seven intrinsic lysine residues or methionine at position 1 to form eight different linkages. Individual linkages generate diverse structure and distinct conformations. These characterizations indicate the regulation of various biological functions.

doi:10.6342/NTU201803256

Chapter 2 Experimental materials and methods

2.1 DNA and Protein sequences

2.1.1 DNA sequences

Methanosarcina mazei PylRS:

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacac
cacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggac
tgcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctc
acaaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgccaaaa
tccgttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggatctaaattttcacctgcgata
ccggtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgc
actggtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccaga
ctgacaggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaatt
gctctctcgcagaaaaaaagacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaatt
accaggttctttgtggacaggggttttctggaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattga
taatgataccgaactttcaaaacagatcttcagggttgacaagaacttctgcctgagacccatgcttgctccaaacctttacaact
acctgcgcaagcttgacagggccctgcctgatccaataaaaatttttgaaataggcccatgctacagaaaagagtccgacggc
aaagaacacctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgcacacgggaaaatcttgaaagcataatt
acggacttcctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctatggggataccttgatgtaatgca
cggagacctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaaccctggatagggg
caggtttcgggctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactataa
cgggatttctaccaacctgtaa

pylT:

tggcggaaaccccgggaatctaacccggctgaacggatttagagtccattcgatctacatgatcaggtttcc

doi:10.6342/NTU201803256

mkRS1F:

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacac

cacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggac

tgcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctc

acaaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgccaaaa

tccgttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggatctaaattttcacctgcgata

ccggtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgc

actggtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccaga

ctgacaggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaatt

gctctctcgcagaaaaaaagacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaatt

accaggttctttgtggacaggggttttctggaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattga

taatgataccgaactttcaaaacagatcttcagggttgacaagaacttctgcctgagacccatgcttgctccaaaccttatgaact

acgcgcgcaagcttgacagggccctgcctgatccaataaaaatttttgaaataggcccatgctacagaaaagagtccgacgg

caaagaacacctcgaagagtttaccatgctgaacttcacgcagatgggatcgggatgcacacgggaaaatcttgaaagcata

attaaggacttcctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctttggggataccttgatgtaatg

cacggagacctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaaccctggatagg

ggcaggtttcgggctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactat

aacgggatttctaccaacctgtaa

sfGFP:

atgcatcatcatcatcatcatagcaagggcgaagaactgtttacgggcgtggtgccgattctggtggaactggatggtgatgtc

aatggtcacaaattcagcgtgcgcggcgaaggtgaaggcgatgcaaccaatggtaaactgacgctgaagtttatttgcaccac

gggtaaactgccggttccgtggccgaccctggtcaccacgctgacgtatggtgttcagtgtttcagtcgttacccggatcacat

gaaacgccacgacttttttcaagtccgcgatgccggaaggttatgtccaagaacgtaccatctcatttaaagatgacggcaccta

caaaacgcgcgccgaagtgaaattcgaaggtgatacgctggttaaccgtattgaactgaaaggcatcgattttaaggaagacg

doi:10.6342/NTU201803256

gtaatattctgggccataaactggaatataacttcaattcgcacaacgtgtacatcaccgcagataagcagaagaacggtatca

aggctaacttcaagatccgccataatgtggaagatggcagcgttcaactggccgaccactatcagcaaaacaccccgattggt

gatggcccggtcctgctgccggacaatcattacctgagcacgcagtctgtgctgagtaaagatccgaacgaaaagcgtgacc

acatggtcctgctggaattcgtgaccgcggccggcatcacgcacggtatggacgaactgtataaaggctca

SUMO2 from Homo sapiens:

atggcagatgaaaaaccgaaagaaggcgttaaaaccgaaaacaatgatcatattaatctgaaagttgccggccaggatggct

cagtggtgcagtttaaaattaaacgtcataccccgctgagtaaactgatgaaagcctattgtgaacgccagggcctgtctatgcg

ccagattcgctttcgctttgatggtcagccgattaacgagaccgataccccggcacagttagagatggaagatgaagataccat

tgatgtgtttcagcagcagaccggtggctaa

Ubiquitin from Homo sapiens:

atgcatcatcatcatcatcatgaaaatctgtattttcagatgcagatttttgtgaaaaccctgaccggcaagaccattaccttagaa

gttgaaccgagtgataccattgaaaatgttaaagccaaaattcaggataaagaaggcattccgccggatcagcagcgtctgatt

tttgccggcaaacagctggaagatggtcgcaccctgagcgattataatattcagaaagaatctaccttacatctggtgctgcgct

tacgcggtggctaa

Ubiquitin activating enzyme E1 from Mus musculus:

atgtccagctcgccgctgtccaagaaacgtcgcgtgtccgggcctgatccaaagccgggttctaactgctcccctgcacagtc

tgcgctgtccgaagtgtcctcagtgccaaccaacggaatggcgaagaacggcagtgaagcagacatagacgagagccttta

ctcccggcagctgtacgttttgggccatgaggcaatgaaaatgctccagacatccagcgtccttgtctcaggcttgcggggctt

gggtgtagaaattgctaagaacatcatccttggtggggtcaaggctgtcaccctacatgaccaaggaactacccagtgggctg

atctctcctcccagttttaccttcgggaggaggacattggtaaaaatcgagcggaggtatcccagccccgacttgctgaactca

acagctatgtacctgtcactgcctacactgggcctcttgtcgaggacttccttagtagcttccaggtggtggtcctcaccaacag

ccccctggaagcccagctgcgagtgggggagttctgtcatagccgtggtatcaagctagtggtggcagatacaagaggcctg

tttgggcaacttttctgtgattttggagaggaaatggtcctcacagattccaatggggagcagccactcagtgctatggtttcaat

doi:10.6342/NTU201803256

ggtcaccaaggacaaccccggtgtggttacctgccttgatgaggcccgacatggctttgagactggtgacttcgtctcattctca

gaagtacagggcatgatccaactcaatggatgtcagcccatggagatcaaagtgctgggtccttataccttagtatctgtgaca

cttccaacttctctgactacatccgtggaggcatcgtcagtcaggtcaaagtaccgaaaaagattagttttaaatccttgccagca

tcactggtagagcctgactttgtgatgactgactttgccaagtattctcgccctgcccaactgcacattggcttccaagctctgca

ccaattctgtgctctgcacaaccaaccacctcgaccacgaaatgaggaagatgcaacagagctggtgggcctggctcaggct

gtaaacgctcggtccccaccttcagtaaaacagaacagcttggatgaagaccttattcggaagctagcttatgttgctgctggg

gacctggcacccataaatgctttcattgggggccttgctgcccaggaagtcatgaaggcctgctctggaaagtttatgcccatc

atgcagtggttgtactttgatgctcttgaatgtctcccagaggacaaagaggctctgacagaggagaagtgcctcccacgtcag

aaccgttacgatgggcaggtagctgtatttgggtcagactttcaggagaagctgagcaagcaaaagtacttcctggtgggtgc

aggggccattggctgtgaattgctcaagaactttgccatgattgggctgggttgtggagagggtggagaggtcgtggtcacag

acatggacaccattgagaaatcaaatctgaaccgacagtttctcttccggccctgggatgtcacgaagttaaagtctgacacgg

ccgctgcagctgtgcgccagatgaatccttacatccaggtgacaagccaccagaaccgtgtaggtcctgacactgagcgcat

ctatgatgatgatttcttccaaaatttggatggtgtggccaatgctctggacaacatagatgcccgcatgtacatggatcgccgat

gtgtgtactatcggaagccactgctagagtctggcacactgggcacaaagggcaacgtgcaggtggtaatccccttcctgaca

gaatcctacagctctagccaggatccaccagagaaatccatccccatttgtaccctgaaaaactttcccaatgccatcgaacac

actcttcagtgggcccgggatgaatttgaaggccttttcaagcagccagcagaaaatgttaatcagtacctcacagactccaaat

ttgtggagcggacattgcggctggctggtacccagccattggaggtgctggaggctgtgcagcgcagcctggtgttgcagcg

accacagacttggggagactgtgtgacctgggcctgccaccactggcacacccagtactgtaacaacatccggcaactgctg

cacaactttcctcctgaccagctcaccagctcaggggccccttttctggtctggacccaaacgctgtccacacccacttacttttg

atgttaacaatacattgcatctggattatgtgatggctgctgccaacctttttgcccagacctatgggttgactggatcccaagacc

gagctgctgtggcctcactcctgcagtcagtacaagtcccagagttcacccccaagtctggtgtcaagattcatgtttctgacca

ggagttgcagagtgccaatgcctctgttgatgacagccgtcttgaggagctcaaagccacattgcccagcccagacaagttac

ccggatttaagatgtaccccattgattttgagaaggatgatgacagcaatttccacatggatttcattgtggctgcatccaatcttc

gggccgaaaactatgatatttcccctgcagaccgacacaagagcaagctgattgcagggaagatcatcccagccattgccac

aaccacagctgctgtggttggccttgtgtgtctggagctctacaaggtagttcaagggcaccaacagctcgattcctacaaaaat

doi:10.6342/NTU201803256

ggtttcctgaacttggccctgcccttctttgggttttctgaacctcttgctgcacctcgtcaccagtactataatcaagagtggacat

tgtgggatcgctttgaagtacaagggctgcagcctaatggtgaggagatgaccctcaagcagttccttgattactttaagacaga

gcacaaattggagatcaccatgctgtcccagggcgtgtccatgctctattctttcttcatgccagctgctaagctcaaggaacgat

tggatcagccgatgacagagattgtgagccgagtgtcaaagagaaagctgggccgccatgtgcgggcactggtgcttgagct

gtgctgcaacgatgaaagcggcgaggacgtcgaggtcccttatgtccgatataccattcgctga

2.1.2 Protein sequences

PylRS:

M D K K P L N T L I S A T G L W M S R T G T I H K I K H H E V S R S K I Y I E M

A C G D H L V V N N S R S S R T A R A L R H H K Y R K T C K R C R V S D E D

L N K F L T K A N E D Q T S V K V K V V S A P T R T K K A M P K S V A R A P

K P L E N T E A A Q A Q P S G S K F S P A I P V S T Q E S V S V P A S V S T S I S

S I S T G A T A S A L V K G N T N P I T S M S A P V Q A S A P A L T K S Q T D R

L E V L L N P K D E I S L N S G K P F R E L E S E L L S R R K K D L Q Q I Y A E

E R E N Y L G K L E R E I T R F F V D R G F L E I K S P I L I P L E Y I E R M G I D

N D T E L S K Q I F R V D K N F C L R P M L A P N L Y N Y L R K L D R A L P D

P I K I F E I G P C Y R K E S D G K E H L E E F T M L N F C Q M G S G C T R E N

L E S I I T D F L N H L G I D F K I V G D S C M V Y G D T L D V M H G D L E L S

S A V V G P I P L D R E W G I D K P W I G A G F G L E R L L K V K H D F K N I K

R A A R S E S Y Y N G I S T N L

mkRS1F:

M D K K P L N T L I S A T G L W M S R T G T I H K I K H H E V S R S K I Y I E M

A C G D H L V V N N S R S S R T A R A L R H H K Y R K T C K R C R V S D E D

L N K F L T K A N E D Q T S V K V K V V S A P T R T K K A M P K S V A R A P

doi:10.6342/NTU201803256

K P L E N T E A A Q A Q P S G S K F S P A I P V S T Q E S V S V P A S V S T S I S

S I S T G A T A S A L V K G N T N P I T S M S A P V Q A S A P A L T K S Q T D R

L E V L L N P K D E I S L N S G K P F R E L E S E L L S R R K K D L Q Q I Y A E

E R E N Y L G K L E R E I T R F F V D R G F L E I K S P I L I P L E Y I E R M G I D

N D T E L S K Q I F R V D K N F C L R P M L A P N L M N Y A R K L D R A L P D

P I K I F E I G P C Y R K E S D G K E H L E E F T M L N F T Q M G S G C T R E N

L E S I I K D F L N H L G I D F K I V G D S C M V F G D T L D V M H G D L E L S

S A V V G P I P L D R E W G I D K P W I G A G F G L E R L L K V K H D F K N I K

R A A R S E S Y Y N G I S T N L

sfGFP:

M H H H H H H S K G E E L F T G V V P I L V E L D G D V N G H K F S V R G E G

E G D A T N G K L T L K F I C T T G K L P V P W P T L V T T L T Y G V Q C F S

R Y P D H M K R H D F F K S A M P E G Y V Q E R T I S F K D D G T Y K T R A E

V K F E G D T L V N R I E L K G I D F K E D G N I L G H K L E Y N F N S H N V Y

I T A D K Q K N G I K A N F K I R H N V E D G S V Q L A D H Y Q Q N T P I G D

G P V L L P D N H Y L S T Q S V L S K D P N E K R D H M V L L E F V T A A G I

T H G M D E L Y K

SUMO2 from Homo sapiens:

M A D E K P K E G V K T E N N D H I N L K V A G Q D G S V V Q F K I K R H T P

L S K L M K A Y C E R Q G L S M R Q I R F R F D G Q P I N E T D T P A Q L E M

E D E D T I D V F Q Q Q T G G

doi:10.6342/NTU201803256

Ubiquitin from Homo sapiens:

M H H H H H H E N L Y F Q M Q I F V K T L T G K T I T L E V E P S D T I E N V K A K I Q D K E G I P P D Q Q R L I F A G K Q L E D G R T L S D Y N I Q K E S T L H L V L R L R G G

Ubiquitin activating enzyme E1 from Mus musculus:

M S S S P L S K K R R V S G P D P K P G S N C S P A Q S A L S E V S S V P T N G M A K N G S E A D I D E S L Y S R Q L Y V L G H E A M K M L Q T S S V L V S G L R G L G V E I A K N I I L G G V K A V T L H D Q G T T Q W A D L S S Q F Y L R E E D I G K N R A E V S Q P R L A E L N S Y V P V T A Y T G P L V E D F L S S F Q V V V L T N S P L E A Q L R V G E F C H S R G I K L V V A D T R G L F G Q L F C D F G E E M V L T D S N G E Q P L S A M V S M V T K D N P G V V T C L D E A R H G F E T G D F V S F S E V Q G M I Q L N G C Q P M E I K V L G P Y T F S I C D T S N F S D Y I R G G I V S Q V K V P K K I S F K S L P A S L V E P D F V M T D F A K Y S R P A Q L H I G F Q A L H Q F C A L H N Q P P R P R N E E D A T E L V G L A Q A V N A R S P P S V K Q N S L D E D L I R K L A Y V A A G D L A P I N A F I G G L A A Q E V M K A C S G K F M P I M Q W L Y F D A L E C L P E D K E A L T E E K C L P R Q N R Y D G Q V A V F G S D F Q E K L S K Q K Y F L V G A G A I G C E L L K N F A M I G L G C G E G G E V V V T D M D T I E K S N L N R Q F L F R P W D V T K L K S D T A A A A V R Q M N P Y I Q V T S H Q N R V G P D T E R I Y D D D F F Q N L D G V A N A L D N I D A R M Y M D R R C V Y Y R K P L L E S G T L G T K G N V Q V V I P F L T E S Y S S S Q D P P E K S I P I C T L K N F P N A I E H T L Q W A R D E F E G L F K Q P A E N V N Q Y L T D S K F V E R T L R L A G T Q P L E V L E A V Q R S L V L Q R P Q T W G D C V T W A C H H W H T Q Y C N N I R Q L L H N F P P D Q L T S S G A P F W S G P K R C P H P L T F D V N N T L H L D Y V M A A A N L F A Q T Y G L T G S Q D

doi:10.6342/NTU201803256

R A A V A S L L Q S V Q V P E F T P K S G V K I H V S D Q E L Q S A N A S V D D S R L E E L K A T L P S P D K L P G F K M Y P I D F E K D D D S N F H M D F I V A A S N L R A E N Y D I S P A D R H K S K L I A G K I I P A I A T T T A A V V G L V C L E L Y K V V Q G H Q Q L D S Y K N G F L N L A L P F F G F S E P L A A P R H Q Y Y N Q E W T L W D R F E V Q G L Q P N G E E M T L K Q F L D Y F K T E H K L E I T M L S Q G V S M L Y S F F M P A A K L K E R L D Q P M T E I V S R V S K R K L G R H V R A L V L E L C C N D E S G E D V E V P Y V R Y T I R

2.2 Plasmid construction

2.2.1 Primer list

Name	Sequences(5' to 3')
pET-His6X-sfGFP-NdeI-F1	gagatatacatatgcatcatcatcatcatcatagcaaggg
SUMO2-K5-F2	cagatggcagatgaatagccgaaagaaggcgttaaaac
SUMO2-K5-R1	gccttctttcggctattcatctgccatctgaaaatacag
SUMO2-K7-F2	gatgaaaaaccgtaggaaggcgttaaaaccgaaaacaatg
SUMO2-K7-R1	gttttaacgccttcctacggtttttcatctgccatctg
SUMO2-K11-F2	ccgaaagaaggcgtttagaccgaaaacaatgatcatattaatc
SUMO2-K11-R1	cattgttttcggtctaaacgccttctttcggtttttcatc
SUMO2-K21-F2	ccgaaagaaggcgtttagaccgaaaacaatgatcatattaatc
SUMO2-K21-R1	ctggccggcaacctacagattaatatgatcattgttttcg
SUMO2-K33-F2	gtggtgcagttttagattaaacgtcataccccgctgag
SUMO2-K33-R1	gtatgacgtttaatctaaaactgcaccactgagccatcctg
SUMO2-K35-F2	gcagtttaaaatttagcgtcataccccgctgagtaaactgatg
SUMO2-K35-R1	cagcggggtatgacgctaaattttaaactgcaccactgag
SUMO2-K42-F2	ccccgctgagtaaactgatgtaggcctattgtgaacg
SUMO2-K42-R1	caataggctttcatcagctaactcagcggggtatgacg

doi:10.6342/NTU201803256

SUMO2-K45-F2	gagtaaactgatgtaggcctattgtgaacgccagggc
SUMO2-K45-R1	cgttcacaataggcctacatcagtttactcagcgggg
pET-hSUMO2-SacI-R2	ggtgatggagctcttagccaccggtctgctg
pET-His6X-TEV-hUb-NdeI-F	gagatatacatatgcatcatcatcatcatcatgaaaatctgtattttc
pET-hUb-SacI-R2	gatggagctcttagccaccgcgtaagcgcag
pET-sfGFP-NdeI-F1	gagatatacatatgagcaagggcgaag
pET-sfGFP-sacI-R2	gatggtgatggagctctgagcctttatac
pET-sfGFP-F27TAG-R	caccttcgccgcgcacgctctatttgtgacc
pET-sfGFP-F27TAG-F	ggtcacaaatagagcgtgcgcggcgaaggtg
sfGFP-Y66TAG-F2	gtcaccacgctgacgtagggtgttcag
sfGFP-Y66TAG-R1	ctgaacaccctacgtcagcgtggtgac

2.2.2 Plasmids

Construction of pET-pylT

The plasmid pET-pylT was derived from the pET-22b (+) plasmid having an ampicillin (Amp) selection marker, purchasing from Novegen. The gene of pylT with anticodon CUA flanked by lpp promoter at the 5'end and rrnC terminator at 3' end. This gene was amplified 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

The plasmid pCDF-MmPylRS was derived form pCDF-1b plasmid containing Sp selection marker, purchasing form Novegen. The wild type *Mm*PylRS gene was synthesize by MD bio, Inc company, and encoded to plasmid. The gene flanked under the control of lpp promoter at the 5'end and rrncC terminator at 3' end. The two restrictions sites, BamHI and NcoI were used to double digest the *Mm*PylRS gene. The digested gene was ligated to generate pCDF-*Mm*PylRS.

doi:10.6342/NTU201803256

Construction of pCDF-mkRS1F

The pCDF- mkRS1F plasmid with Y306M, L309A, C348T, T364K, Y384F mutation sites, based on published paper were amplified by overlapping PCR from pCDF-*Mm*PylRS plasmid. The gene of mkRS1F flanked under the control of lpp promoter at the 5'end and rrncC terminator at 3' end. The mkRS1F gene was digested with restriction enzyme BamHI and NcoI, and ligated to generate pCDF-mkRS1F.

Construction of pCDF-FOWRS2

The pCDF-FOWRS2 plasmid with R61K, H63Y, S193R, N346G, C348Q, V401G mutation sites, were amplified by overlapping PCR from pCDF-*Mm*PylRS plasmid. The gene of FOWRS2 flanked under the control of lpp promoter at the 5'end and rrncC terminator at 3' end. The FOWRS2 gene was digested with restriction enzyme BamHI and NcoI, and ligated to generate pCDF-FOWRS2.

Construction of pCDF- FOWRS6

The pCDF-FOWRS6 plasmid with R61K, H63Y, S193R, N346G, C348Q, V401G, W417T mutation sites, were amplified by overlapping PCR from pCDF-*Mm*PylRS plasmid. The gene of FOWRS6 flanked under the control of lpp promoter at the 5'end and rrncC terminator at 3' end. The FOWRS6 gene was digested with restriction enzyme BamHI and NcoI, and ligated to generate pCDF- FOWRS6.

Construction of pET-pylT- sfGFP-TEV-SUMO2 variants

The Plasmid pET-pylT- sfGFP-TEV-SUMO2 was derived from the plasmid pET-pylT$_{CUA}$. The gene of optimized superfolder GFP and SUMO2 were synthesize by MD bio, Inc. company, and encoded to plasmid. The gene of SUMO2 constructs were

doi:10.6342/NTU201803256

consisted of N-terminal hexahistidine sfGFP fusion protein, a TEV protease cleavage site between sfGFP and SUMO2. This full-length gene was amplified by overlapping PCR. The two restriction sites NdeI and SacI were used to digest the full-length gene. The digested full-length gene was ligated to form pET-pylT- sfGFP-TEV-SUMO2. The gene flanked under the control of T7 promoter.

Construction of pET-pylT- Ubiquitin

The Plasmid pET-pylT-Ubiquitin was derived from the plasmid pET-pylT$_{CUA}$. The gene of optimized Ubiquitin was synthesize by MD bio, Inc. company, and encoded to plasmid. The gene of Ubiquitin constructs were consisted of N-terminal hexahistidine, a TEV protease cleavage site between hexahistidine and Ubiquitin. This full-length gene was amplified by overlapping PCR. The two restriction sites NdeI and SacI were used to digest the full-length gene. The digested full-length gene was ligated to form pET-pylT-Ubiquitin. The gene flanked under the control of T7 promoter.

Construction of pET-pylT-sfGFP

The Plasmid pET-pylT-sfGFP was derived from the plasmid pET-pylT$_{CUA}$. The gene of optimized superfolder GFP was synthesize by MD bio, Inc. company, and encoded to plasmid. The sfGFP gene with the hexahistidine at C terminus were amplified by overlapping PCR. The two restriction sites NdeI and SacI were used to digest the sfGFP gene. The digested gene was ligated to form pET-pylT-sfGFP. The gene flanked under the control of T7 promoter.

Construction of pET-pylT-sfGFPF27TAG

The plasmid pET-pylT-sfGFPF27TAG was derived from the plasmid pET-pylT-sfGFP. The gene of sfGFP has amber mutation at F27. The sfGFP gene with the

doi:10.6342/NTU201803256

hexahistidine at C terminus was amplified by overlapping PCR from pET-pylT-sfGFP plasmid. The gene of sfGFPF27TAG flanked under the control of T7 promoter. The sfGFPF27TAG gene was digested with restriction enzyme NdeI and SacI, and ligated to generate pET-pylT-sfGFPF27TAG.

doi:10.6342/NTU201803256

2.2.3 Molecular cloning

Constructions of pET-pylT- sfGFP -SUMO2XTAG

DNA of human SUMO2 was obtained from totally synthesis with codon optimization for bacterial expression. All constructs were cloned into pET-22b (+) vector using KOD hot start polymerase kit. DNA sequence of SUMO2 constructs were consisted of N-terminal hexahistidine sfGFP fusion protein, a TEV protease cleavage site between sfGFP and SUMO2. Then, the respective codon of lysine were mutated to amber codon (KXTAG; where X represents the position within SUMO2 sequence, X = 5, 7, 11, 21, 33, 35, 42, 45). The site-directed mutagenesis of all the unique positions accomplished with overlapping PCR technique. The PCR was done to generate the gene of insert. The amplified PCR products were cleaned up and subjected to double digestion, and its following ligated to the pET vector with ampicillin resistance gene. The ligated plasmid was chemically transformed into *E. coli* DH5α strain and subsequently spread on the medium plate with corresponding antibiotic. Eight colonies were chosen to perform colony PCR. The colonies indicating positive results were set out for DNA sequencing. After results analysis, the positive colonies were picked and cultured in LB medium at 37 ^{o}C for 12-16 hrs. Bacteria medium were collected and the plasmids were extracted by mini-prep.

DNA of wild-type SUMO2 construct were containing N-terminal hexahistidine sfGFP fusion protein, a TEV protease cleavage site between sfGFP and SUMO2 constructed in pET vector in advance and was taken as template for PCR. In detail, the Primer F1 and primer R1 were used to generate the template His6X-sfGFP-TEV-SUMO2 XTAG (KXTAG; where X represents the position within SUMO2 sequence, X = 5, 7, 11, 21, 33, 35, 42, 45).The primer F2 and primer R2 were used to generate template SUMO2 XTAG (KXTAG; where X represents the position within SUMO2 sequence, X = 5, 7, 11,

doi:10.6342/NTU201803256

21, 33, 35, 42, 45). Especially, the primer R1 and F2 involved the amber mutation were used to replace the respective codon of lysine. These two fragments were used for overlapping PCR to get His6X-sfGFP-TEV-SUMO2 XTAG (KXTAG; where X represents the position within SUMO2 sequence, X = 5, 7, 11, 21, 33, 35, 42, 45). The PCR products were analyzed by 1% agarose gel with safe dye staining. Expected PCR products were collected and cleaned up for overlapping PCR. The overlapped fragment was subjected to double digestion experiment with restriction enzyme as well as the other backbone of pET vector. Restriction enzyme NdeI and SacI were used to digest the sticky end at the target gene fragment and pET vector. The fragments were recycled and purified, then used for DNA ligation.

DNA ligation used the T4 DNA ligase. The vector and insert were mixed with 1:3 molar ratio and added the T4 DNA ligase at 25^{o}C for 1 hour. The plasmids were then chemically transformed into *E. coli* DH5α respectively. The strain with ligated products added on ice bath for 20 minutes then via heat-shock process (42^{o}C for 1 minute). The bacteria were then rapidly put on ice bath for 5 minutes to increase the transformation efficiency. The recovery process was performed following 1 mL autoclaved LB medium added with the transformed bacteria at 37^{o}C for 1 hour. Plates were prepared then executed after recovery following incubation at 37^{o}C for 12 hrs. Eight colonies were picked for colony PCR.

Colony PCR was treatment with PCR mastermix kit and verified using T7 promoter, T7 terminator. The PCR products can be visualized by 1% agarose gel in TAE buffer. The colonies revealed positive consequences were chosen for DNA sequencing.

doi:10.6342/NTU201803256

Constructions of pET-pylT-Ubiquitin

DNA of human Ubiquitin was obtained from totally synthesis with codon optimization for bacterial expression. The construct was cloned into pET-22b (+) vector using KOD hot start polymerase kit. DNA sequence of Ubiquitin construct was consisted of N-terminal hexahistidine, a TEV protease cleavage site between hexahistidine and Ubiquitin. Then, the Plasmid pET-pylT- sfGFP-TEV- Ubiquitin construct was used as template for PCR. The PCR technique was used to amplify the gene of insert. In detail, the Primer F1 and primer R1 were used to generate the template His6X-TEV- Ubiquitin. Subsequently, the process of molecular cloning technique was the same as described previously.

2.3 Protein productions and purifications

Ubiquitin expression and purification

The plasmid was chemically transformed into *E. coli* BL21 (DE3) for protein expression. A single colony was cultured in 5 mL LB medium with the 100 ug/ml ampcillim and was incubated at 37°C overnight. 1 mL bacterial liquid was transferred to 100 mL fresh LB medium with appropriate antibiotics and incubated at 37°C until an optical density (OD_{600}) value of 0.4-0.6. Overexpression was induced by isopropylthio-β-D-galactopyranoside (IPTG). Final concentration of IPTG was 1 mM and recombinant protein was induced at 25°C for 12 hours. The bacteria culture were harvested by centirfugation at 10000 × *g* for 10 minutes at 4°C and discarded the supernatant. Bacteria pellets were quickly frozen by liquid nitrogen and stored at -80°C.

After freeze thawed, the bacteria was resuspended in lysis buffer (300 mM Sodium Chloride and 20 mM Monopotassium phosphate, 0.5% Triton X-100, pH 8.0) and one protease inhibitor. The suspension was incubated at 4 °C, sonicated in ice/water bath for

doi:10.6342/NTU201803256

10 minutes (5 s on/10 s off, 35W). The supernatant was gathered after centrifuged at 20000 × *g* for 50 minutes at 4 ºC. The recombinant protein was purified by Ni^{2+}-NTA chromatography. 25 column volume of lysis buffer and 10 column volume of washing buffer (300 mM Sodium Chloride, 1mM imidazole and 20 mM Monopotassium phosphate, pH 8.0) was used to wash out the non-specific binding proteins. 5 column volume of elution buffer (300 mM NaCl, 20 mM Monopotassium phosphate and 250 mM imidazole, pH4.5) was used to elute target protein. The eluted fraction in elution buffer then changed to storage buffer (20 mM HEPES, and 100 mM NaCl, pH8.0) by dialysis, and concentrated by Amicon Ultra- Centrifugal Filter -3,000 MWCO (Millipore). The concentration was determined by Braford assay (Bio-Rad). The concentrated protein flash frozen in liquid nitrogen at -80ºC for preservation.

Mouse Ubiquitin-Activating Enzyme (mE1) expression and purification

The DNA sequence of mE1 in pET28b vertor was purchased from Addgene, expressed and purified from *E. coli* BL21-CodonPlus(DE3)-RIPL cells. At first culture was grown in 3 ml LB medium with chloramphenicol and kanamycin at 37℃ overnight. And the culture was inoculated into LB medium (12 L) with the same concentration of antibiotics. The culture was then cooled to 16℃ after the OD_{600} reached 0.6~0.8 and then added 0.6 mM isopropyl 1-thio-β-D- galactopyranoside (IPTG) to grow for another 16 hrs at the same temperature. The cells were pelleted, resuspended in 150 ml of lysis buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% (w/v) Triton X-100, pH 8.0, 1 mM DTT, and protease inhibitor mixture), and sonicated for 10 minutes in an ice/water bath . The lysate was collected after centrifuged at 20000 × *g* for 40 min at 4℃. The recombinant protein was purified by Ni^{2+}-NTA column. And then washed with 10 column volume of lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM imidazole, pH 7.8), and then eluted

doi:10.6342/NTU201803256

out with 12 column volume of elution buffer (50 mM HEPES, 500 mM NaCl, 250 mM imidazole,pH 7.8) then changed to 10 mM Tris–HCl, pH 8.0, pH 8.0, 1 mM DTT by dialysis. Recombinant histidine-tagged mouse E1 (mE1) was purified by anion-exchange chromatography[56]. Finally, the recombinant protein was purified by gel filtration chromatography, and concentrated by Amicon Ultra- Centrifugal Filter -30,000 MWCO (Millipore). The concentration was determined by Braford assay (Bio-Rad). The concentrated protein flash frozen in liquid nitrogen at -80°C for preservation.

*sfGFP-SUMO2X-**2** (where X represents the position within SUMO2 sequence, X = 5, 7, 11, 21, 33, 35, 42, 45) expression and purification*

The plasmid was chemically transformed into *E. coli* BL21 (DE3) for protein expression. A single colony was cultured in 5 mL LB medium with 100 ug/ml ampcillim and 100 ug/ml streptomycin then was incubated at 37°C overnight. 1 mL bacterial liquid was transferred to 100 mL fresh LB medium with appropriate antibiotics and incubated at 37°C until an optical density (OD_{600}) value of 0.8-1.0. Overexpression was induced by isopropylthio-β-D- galactopyranoside (IPTG). Final concentration of IPTG was 1 mM and recombinant protein was induced at 37°C for 12 hours. Moreover, 1 mM **2** was added in the LB medium during induction. The bacteria culture were harvested by centirfugation at 10000 × *g* for 10 minutes and discarded the supernatant. Bacteria pellets were quickly frozen by liquid nitrogen and stored at -80°C.

After freeze thawed, the bacteria was resuspended in the lysis buffer (300 mM Sodium Chloride, 0.5% Triton X-100, and PBS buffer, pH 8.0) and one protease inhibitor. The suspension was incubated at 4 °C, sonicated in ice/water bath for 10 minutes (5 s on/10 s off, 35W). The supernatant was collected after centrifuged at 20000 × *g* for 50 minutes at 4°C. The recombinant protein was purified by Ni^{2+}-NTA chromatography. 10

doi:10.6342/NTU201803256

column volume of lysis buffer and 5 column volume of washing buffer (300 mM Sodium Chloride, 5mM imidazole, and PBS buffer, pH 8.0 was used to wash out the non-specific binding proteins. 5 column volume of elution buffer (300 mM Sodium Chloride, 250 mM imidazole, and PBS buffer, pH 8.0) was used to elute target protein. The eluted fraction in elution buffer then changed to buffer (100 mM NaCl, and 50 mM Tris-HCl pH 8.0) for TEV cleavage. To purify the non-tagged SUMO2 variants, TEV protease and N-terminal hexahistidine sfGFP fusion protein connected a TEV protease cleavage site with SUMO2XSeCbzK (where X represents the position within SUMO2 sequence, X = 5, 7, 11, 21, 33, 35, 42, 45**)** were mixed with 1:30 molar ratio (final volume 15 mL) at 30℃ for 2 hours. The His6X-sfGFP-TEV was removed using TEV protease. The next round of affinity purification enhanced the fragment of His6X-sfGFP-TEVon Ni^{2+}-NTA resin while the free SUMO2 variants was separated out by the flow through. If necessary, proteins were further purified by anion exchange chromatography (HiTrap Q HP, GE Healthcare) with 30CV linear elution into buffer (50 mM Tris, 500 mM NaCl, pH 8.0). All proteins were finally subjected to size exclusion chromatography (HiLoad 16/600 Superdex75pg, GE Healthcare) in PBS buffer. Purity of peak fractions was analyzed by 15% SDS-PAGE, and concentrated by Amicon Ultra- Centrifugal Filter -3,000 MWCO (Millipore). The concentration was determined by Braford assay (Bio-Rad). The concentrated protein flash frozen in liquid nitrogen at -80^{o}C for preservation.

sfGFPF27 and sfGFPY66 variants expression and purification

The plasmid was chemically transformed into *E. coli* BL21 (DE3) for protein expression, single colony was picked and transferred in 20 mL fresh LB medium. 1 mL bacteria solution was transferred to 100 mL fresh LB medium and incubated at 37 ^{o}C until OD_{600} value reached to 0.8-1.0. Ampicillin and Streptomycin was added and the final

doi:10.6342/NTU201803256

concentration was 100 nM. IPTG was added after OD_{600} value reached to 0.8-1.0 The bacteria then centrifuged and washed with GMML medium for three times. GMML medium was prepared for recombinant protein expression. Recombinant protein was induced by isopropylthio-β-D-galactoside (IPTG). Final concentration of IPTG was 1 mM and recombinant protein was induced at 37°C for 12 hours. 1.0 mM ncAA was added in the GMML medium during induction. Bacteria was centrifuged at 10000 × g for 10 minutes and harvested. Bacteria was disrupted by sonicator and centrifuged at 20000 × g for 50 minutes. The supernatant was collected and poured into Ni^{2+}-NTA column for protein purification. 10 column volume of binding buffer was used to wash out the unbound and nonspecific proteins. 5 column volume of washing buffer containing 5 mM imidazole was used to wash out nonspecific proteins on resin. 5 column volume of elution buffer containing 250 mM imidazole was used to elute the target protein from resin. Purity of fractions were analyzed by 12% SDS-PAGE, The eluted fraction in elution buffer then changed to storage buffer (PBS, pH 7.6) by dialysis, and concentrated by Amicon Ultra-Centrifugal Filter -10,000 MWCO (Millipore). The concentration was determined by Braford assay (Bio-Rad). The concentrated protein flash frozen in liquid nitrogen at -80°C for preservation.

2.4 Gel analysis

2.4.1 SDS-PAGE

To ensure success with the high resolution SDS-PAGE, we need to choose appropriate percentage acrylamide based on the molecular weight range of proteins. Then, the components are mixed for the separation gel to pour into the gap between the glass plates and subsequently add isobutanol over the top of the resolving gel to prevent the

doi:10.6342/NTU201803256

oxygen into the gel and inhibit the polymerization. After the separation gel polymerize in 30 minutes, isobutanol is removed from top of the separation gel. Later, the components are mixed for the stacking gel to pour directly on top of the polymerized separating gel, and insert comb to the top of the spacers immediately in 20 minutes, the stacking gel become completely polymerized. Then 16 μL sample mixed with 4 μL loading buffer (final volume 20 μL) and boiled for 5-10 minutes. Clamp the gel and fill freshly prepared 1x running buffer (500 ml) to chambers of the apparatus. Load samples and molecular mass protein markers into wells for separation by electrophoresis

2.4.2 Western blot analysis

After a complete running of SDS PAGE, the gel was then blotted with polyvinylidene fluoride (PVDF) membrane and immersed in transfer buffer. Then, 400 mA applied for 4 hours or 400 mA for 2 hours followed by 200 mA for 12 hours. After transfer process was complete, the PVDF membrane washed with TBST buffer thrice, 5 minutes each. Further, blocking non-specific binding sites was done by using 5% skim milk, and allowed blocking reaction for 1 hour. Again, the membrane washed for thrice, 5 minutes each. Primary antibody (dilution depending on experiment) was added to react with the membrane. After 1 hour of reaction, the membrane was washed with TBST buffer for three times, 5 minute each. Next, the secondary antibody conjugated with AP (alkaline phosphatase) added to the membrane and incubated for 1 hour. Another round of washing was repeated, 5 minutes for three times with TBST buffer. Finally, the BCIP/NBT substrate solution was spread onto the membrane and rinsed with sufficient tap water to stop the reaction. The membrane air-dried and blotted with kitchen roll.

doi:10.6342/NTU201803256

2.5 Protein chemistry

2.5.1 Synthesis of Ub-cysteamine

Synthesis of Ub-cysteamine

The purified mouse E1 stored at -80℃ were rapidly thawed and placed on 4℃. For the 1 μM mouse E1, 50 μM Ubiquitin, 2 mM ATP, 4 mM DTT, 5 mM $MgCl_2$ and 5 mM cysteamine were incubated in 50μL (final volume) of a buffer containing 100 mM NaCl, 20 mM HEPES, pH 8.0 at 4℃ for12 hrs. The purification of Ub-cysteamine was purified by size-exclusion chromatography[57]. The eluted fraction changed to storage buffer (100 mM NaCl, 20 mM HEPES, 4 mM DTT, pH 8.0) by dialysis, and concentrated by Amicon Ultra- Centrifugal Filter -3,000 MWCO (Millipore). The concentration was determined by Braford assay (Bio-Rad). The concentrated protein flash frozen in liquid nitrogen at -80^{o}C for preservation.

2.5.2 Selenoxide β–elimination

Selenoxide β-elimination in SUMO2K11-**2**

The purified SUMO2K11-**2** stored at -80℃ were thawed and placed on 4℃. For the 9 μM SUMO2K11-**2** incubated with 200 equivalent H_2O_2 in 50 μL (final volume) of a PBS buffer pH 8.0 at 37℃ for 2 hours. When the reaction time reached, we add $Na_2S_2O_3$ as quencher to end up all the reaction. The following changed to storage buffer (PBS buffer pH 8.0) by dialysis, and concentrated by Amicon Ultra- Centrifugal Filter -3,000 MWCO (Millipore). The concentration was determined by Braford assay (Bio-Rad). The concentrated protein flash frozen in liquid nitrogen at -80^{o}C for preservation.

doi:10.6342/NTU201803256

2.6 Mass spectrometry characterization

2.6.1 Protein MALDI-TOF-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.[58] Briefly, after sequentially washing the gel pieces with 25 mM NH_4HCO_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 NH_4HCO_3, 10% acetonitrile containing 65 to 100 ng of sequencing grade modified trypsin (Promega) in 25-30 μl was added and incubated with gel pieces for 12-16 hours at 37ºC. Stop reaction by adding 1-2μl of 5% formic acid.

MS and data analysis

The digested samples (0.5 μl) were carefully mixed with matrix solution (0.5μl of 5 mg/ml DHB in 0.1%TFA/30% acetonitrile) and deposited 0.5μl 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 Biotools 3.2 (Bruker).

doi:10.6342/NTU201803256

2.6.2 Protein ESI-MS analysis

The Intact Protein Molecular Weight Determination

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 an ESI source (Waters LockSpray Exact Mass Ionization Source) with a syringe pump (Harvard Apparatus, MA) and held a flow rate of 5 μl/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).

2.7 Protein Biophysical characterizations

2.7.1 UV/Visible absorption spectrum

The sfGFP-Y66TAG with heterocyclic amino acids and wt-sfGFP were diluted to final concentration of 0.1 mg/ml in PBS at pH 8.0. Absorption spectra were obtained using a 1 cm quartz cuvette at 25℃ on a V-630-Bio (Jasco). The following information was parameter setting: photometric mode_ Abs, response_Fast, UV/Vis bandwith_1.5 nm, scan speed_200 nm/min. The Scanning wavelength was started from 600 nm to 230 nm.

2.7.2 Fluorescence spectrum

The sfGFP-Y66TAG with heterocyclic amino acids and wt-sfGFP were diluted to final concentration of 0.1 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.

doi:10.6342/NTU201803256

Chapter 3 Results

3.1 Synthesis of Ubiquitin-cysteamine

3.1.1 Purification of Ubiquitin

In order to synthesize Ubiquitin derivative on its C- terminal modification, the purified Ubiquitin needs to be prepared. The construction of pET-pylT-Ubiquitin fused with N-terminal hexahistidine and TEV cutting site has been confirmed by the DNA sequencing results. First of all, the plasmid pET-pylT-Ubiquitin was chemically co-transformed into *E. coli* BL21 (DE3) for protein expression. Ubiquitin protein was produced in LB medium with 1 mM IPTG at 25℃ for 12 hrs. Subsequently, the Ubiquitin protein fused with N-terminal hexahistidine was purified by Ni^{2+}-NTA affinity chromatography[59] and analyzed by the 15% SDS-PAGE (**Figure 10**). To ensure the next reaction can be worked in progress, the purified Ubiquitin protein also analyzed the expected molecular weight (MW) by ESI-MS (**Figure 11**).

doi:10.6342/NTU201803256

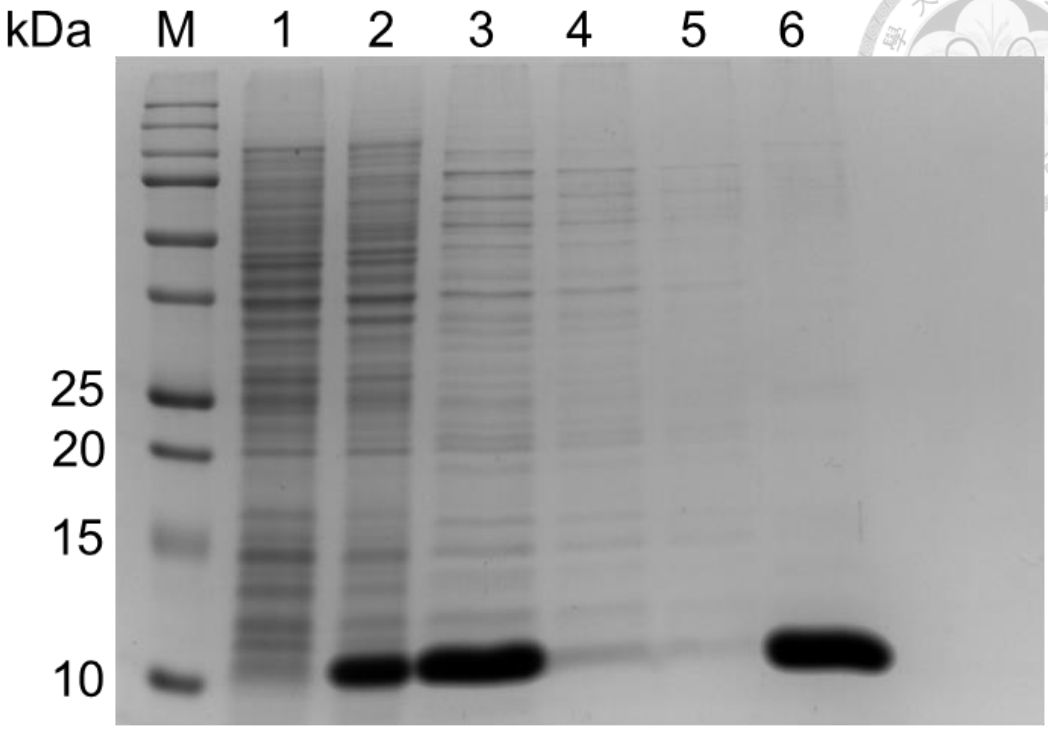
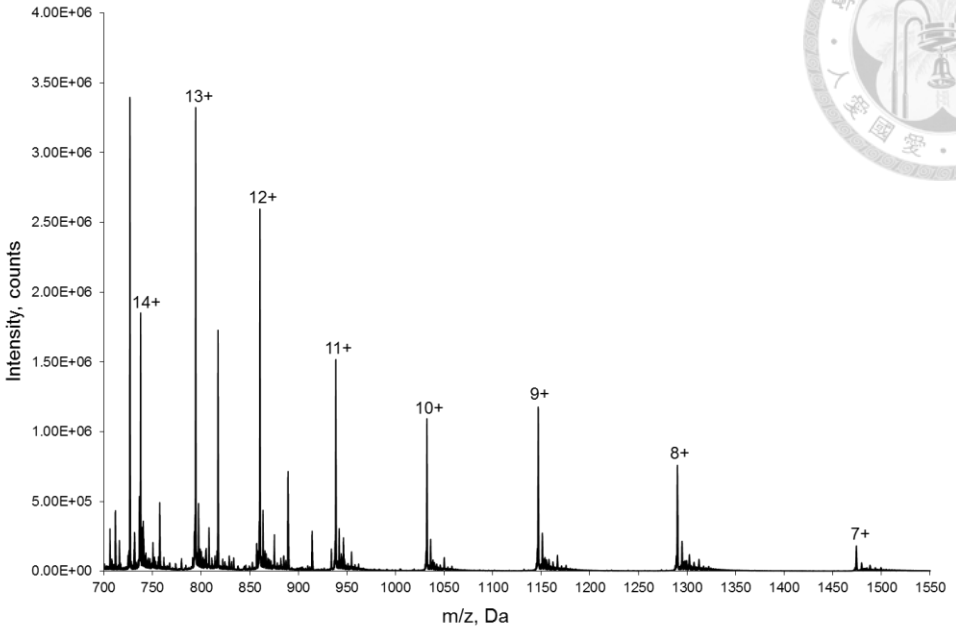


Figure 10. Purification of Ubiquitin with Ni^{2+}–NTA column.

Analysis of the purification of recombinant His6x-TEV-Ubiquitin protein on 15% SDS-PAGE under denaturing conditions with Instant Blue staining. His6x-TEV-Ubiquitin protein with the size of 10 kDa can be observed. Lane M: Protein Molecular Weight Marker; Lane 1: *E. coli* BL21(DE3) were without the addition of 1 mM IPTG; Lane2: *E. coli* BL21(DE3) were with the addition of 1 mM IPTG at 25℃ for 12 hrs; Lane3: samples prepared from soluble fraction of bacterial lysates; Lane4: Flow-through from soluble protein fractions were purified by Ni^{2+}–NTA affinity chromatography; Lane5: washing buffer (300 mM Sodium Chloride, 1mM imidazole and 20 mM Monopotassium phosphate, pH 8.0) wash non-specific binding on Ni^{2+}–NTA resin; Lane6: elution buffer (300 mM NaCl, 20 mM Monopotassium phosphate and 250 mM imidazole, pH 4.5) elute his-tag target protein.

doi:10.6342/NTU201803256

(A)



(B)

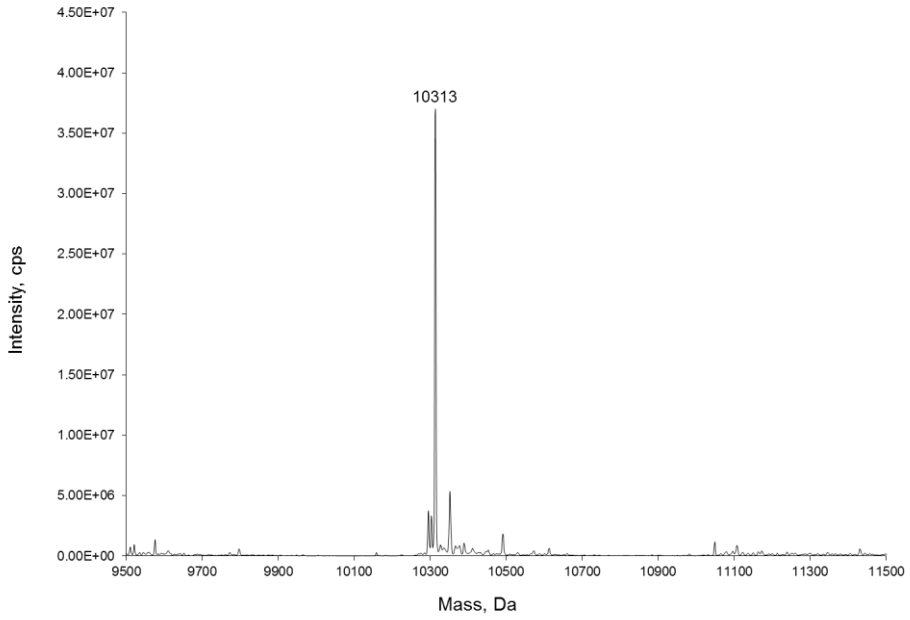


Figure 11. Molecular mass determination of the protein Ubiquitin.

(A) ESI-MS spectrum of Ubiquitin with N-terminal hexahistidine (B) The deconvoluted ESI-MS spectrum of Ubiquitin with N-terminal hexahistidine. The calculated molecular weight is 10314 Da; the found molecular weight is 10313 Da. The proteins were produced by pET-pylT-Ubiquitin with 1 mM IPTG in *E. coli* BL21 (DE3) cells in LB medium at 37 ^{o}C for 3 hours.

doi:10.6342/NTU201803256

3.1.2 Purification of Ubiquitin-activating enzyme

On the other hand, to generate Ubiquitin derivative on its C- terminal modification, the purified Ubiquitin-activating enzyme (E1) also needs to be prepared. The construction of pET-mE1 fused with N-terminal hexahistidine have been confirmed by the DNA sequencing results. First of all, the plasmid pET-mE1 was chemically co-transformed into *E. coli* BL21-CodonPlus (DE3)-RIPL cells for protein expression. The Ubiquitin-activating enzyme was produced in LB medium with 1 mM IPTG at 25℃ for 12 hrs. Next, the Ubiquitin-activating enzyme (E1) fused with N-terminal hexahistidine was purified by Ni^{2+}-NTA affinity chromatography, anion-exchange chromatography, and size-exclusion chromatography[60]. Finally, the purity of Ubiquitin-activating enzyme (E1) was analyzed by SDS-PAGE to ensure its efficient activity (**Figure 12-14**).

doi:10.6342/NTU201803256

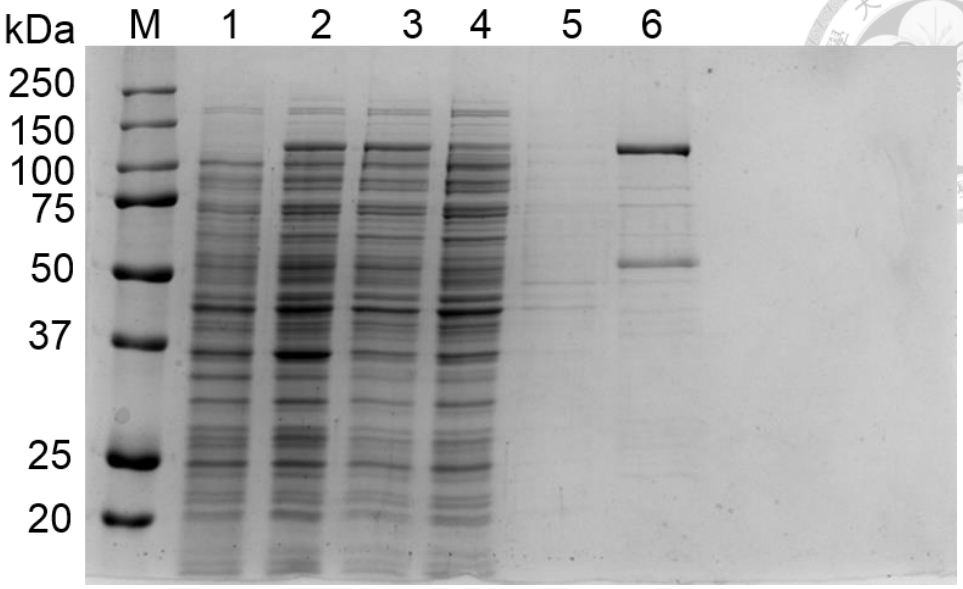
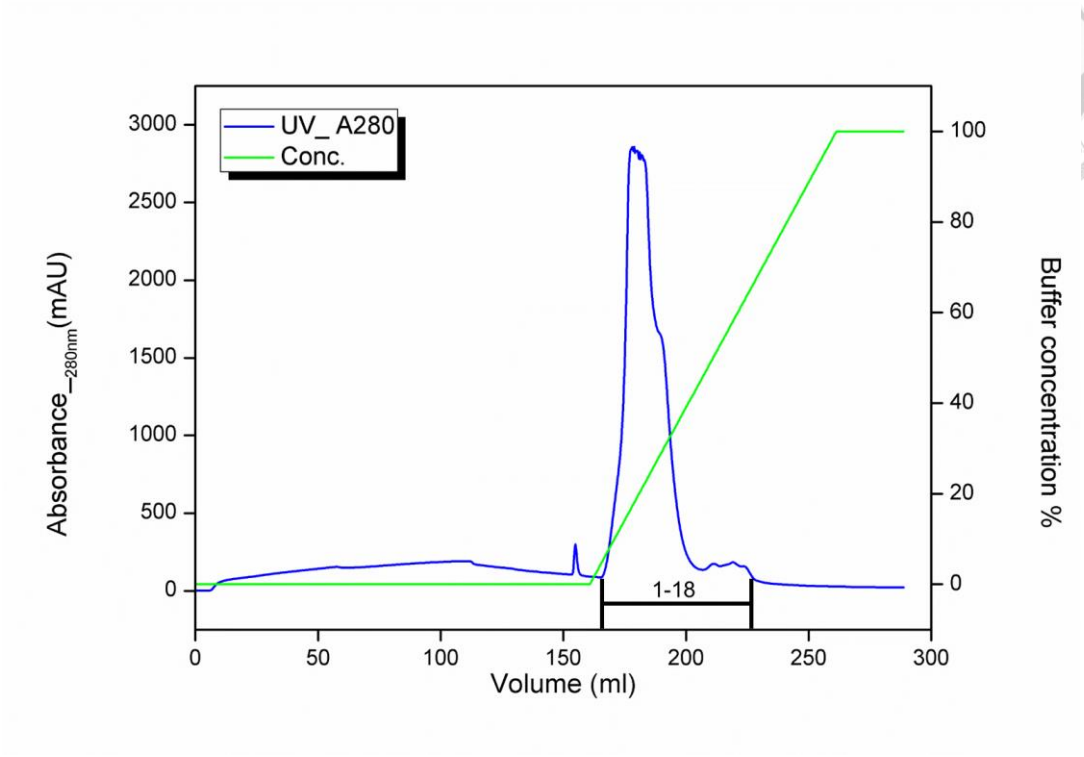


Figure 12. Purification of Ubiquitin-activating enzyme with Ni^{2+}–NTA column.
Analysis of the purification of recombinant His6x-Ubiquitin-activating enzyme protein on 10% SDS-PAGE under denaturing conditions with Instant Blue staining. His6x-Ubiquitin-activating enzyme protein with the size of 116 kDa can be observed. Lane M: Protein Molecular Weight Marker; Lane 1: *E. coli* BL21-CodonPlus(DE3)-RIPL cells were without the addition of 1 mM IPTG; Lane2: *E. coli* BL21-CodonPlus(DE3)-RIPL cells were with the addition of 1 mM IPTG at 25℃ for 12 hrs; Lane3: samples prepared from soluble fraction of bacterial lysates; Lane4: Flow-through from soluble protein fractions were purified by Ni^{2+}–NTA affinity chromatography; Lane5: washing buffer (50 mM HEPES, 500 mM NaCl, 10 mM imidazole, pH 7.8) wash non-specific binding on Ni^{2+}–NTA resin; Lane6: elution buffer (50 mM HEPES, 500 mM NaCl, 250 mM imidazole, pH 7.8) elute his-tag target protein.

doi:10.6342/NTU201803256

(A)



(B)

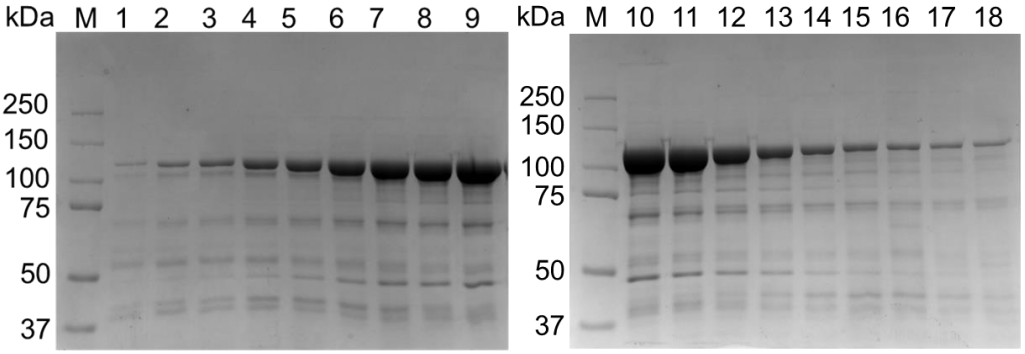
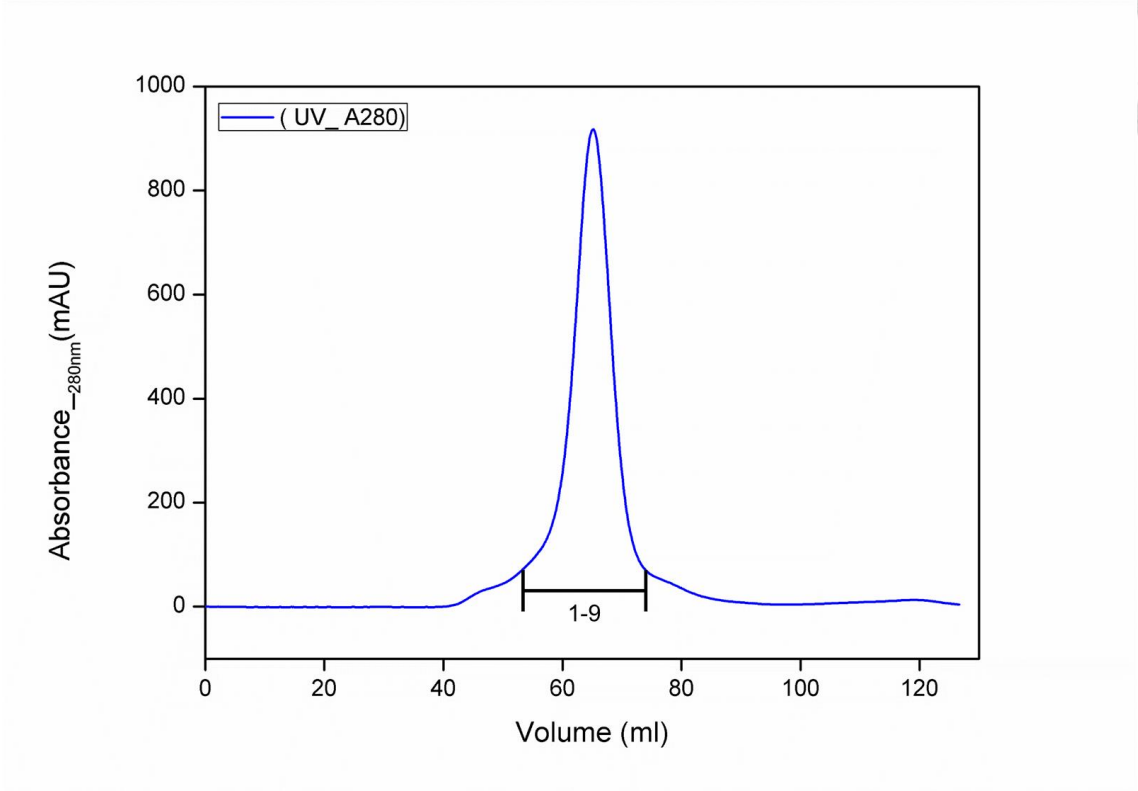


Figure 13. Purification of Ubiquitin-activating enzyme with the anion-exchange. (A) The elution after Ni^{2+}-NTA column was concentrated for FPLC with HiTrap Q HP column (GE Healthcare). The major peak appeared in buffer with 185 mM to 225 mM NaCl. Blue line: the signal of UV_{A280}; Green line: the concentration (%) of buffer with 1 M NaCl. (B) SDS-PAGE shows the fractions indicated in (A). 1-12: major peak; 13-18: minor peak. Fraction 6-12 were collected and concentrated for reaction.

doi:10.6342/NTU201803256

(A)



(B)

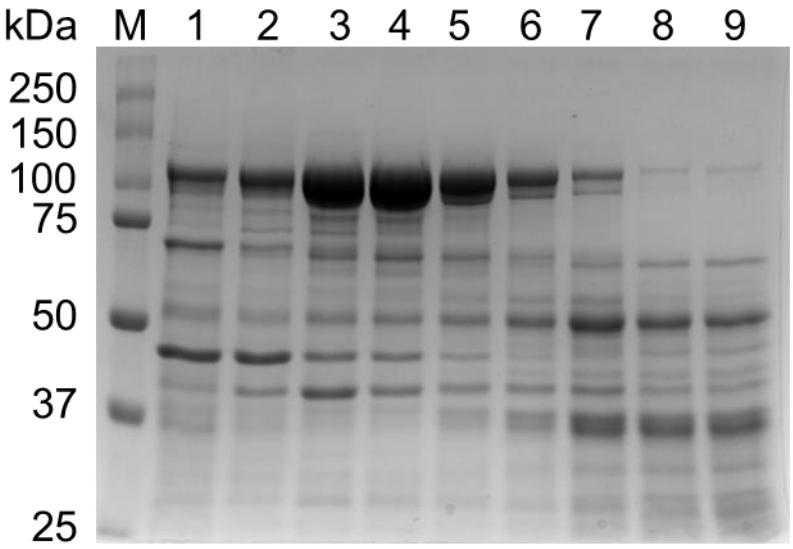


Figure 14. Purification of Ubiquitin-activating enzyme with the SEC.
(A) The elution after HiTrap Q HP column was concentrated for FPLC with S200 column. Blue line: the signal of UV_{A280} (B) SDS-PAGE shows the fractions indicated in (A). Fraction 4-5 were collected and concentrated for reaction.

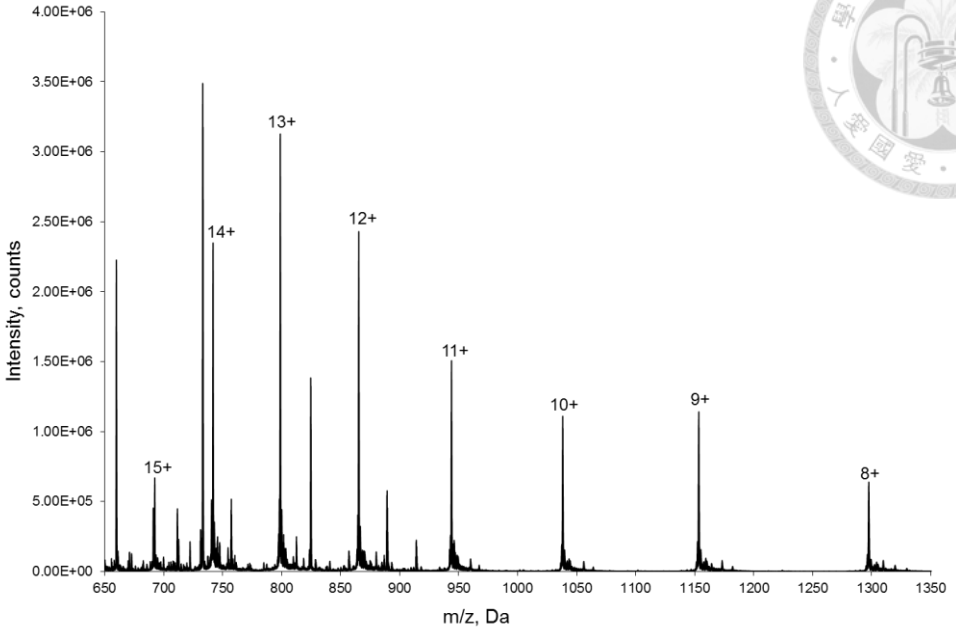
doi:10.6342/NTU201803256

3.1.3 ESI-MS analysis of Ubiquitin-cysteamine

Having Ubiquitin and Ubiquitin-activating enzyme (E1) in advance, we had made effort to achieve synthesis of Ubiquitin-cysteamine. Then E1 catalyzes two consecutive reactions to activate Ubiquitin. First of all, Ubiquitin and ATP form Ub-AMP, then E1 active-site cysteine reacts Ub-AMP to form thioester conjugate with the C-terminal Gly^{76} of Ubiquitin. It has been previously reported that E1-Ub thioester can react with 2-mercaptoethanesulfonic acid (MES) to form Ub-MES thioester and the peptides consisted of N-terminal cysteine to undergo native chemical ligation. Thus, we utilized E1 to catalyze the C-terminal modification of Ubiquitin with cysteamine. When 1 μM mouse E1, 50 μM Ubiquitin, 2 mM ATP, 4 mM DTT, 5 mM $MgCl_2$ and 5 mM cysteamine were mixed at pH 8.0 and 4℃ for 12hrs, Ubiquitin was coverted to an Ub-cysteamine under native chemical ligation in two chemical steps. The molecular weight was analyzed by ESI-MS; the calculated mass was 10374 Da and the observed mass was10372 Da (**Figure 15**). Besides, we also did a control experiment in the absence of ATP. The ESI-MS spectra showed peak at a molecular weight 10313 Da which was in agreement with calculated mass 10314 Da (**Figure 16**). Finally, the Ubiquitin was modified with cysteamine at the C-terminal (**Figure 17**).

doi:10.6342/NTU201803256

(A)



(B)

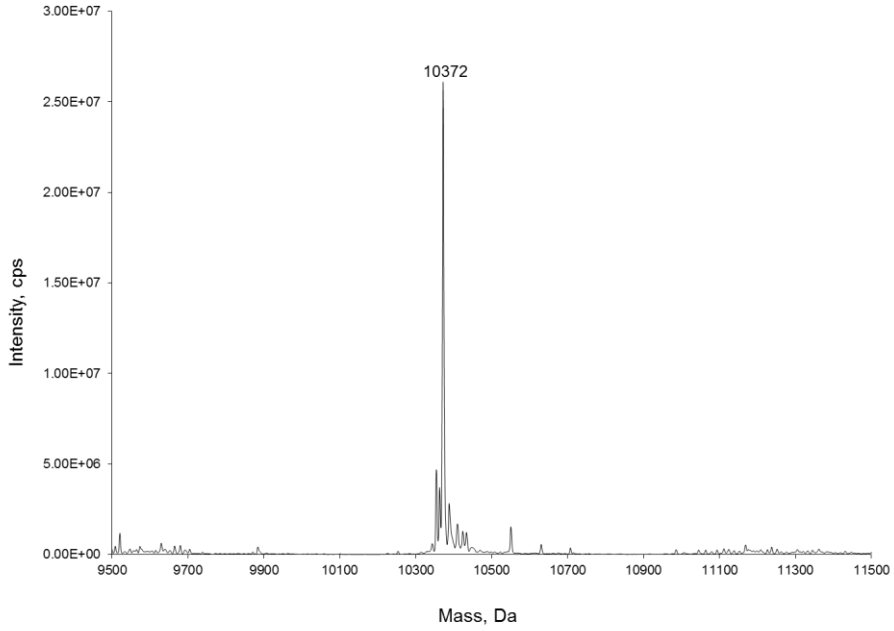
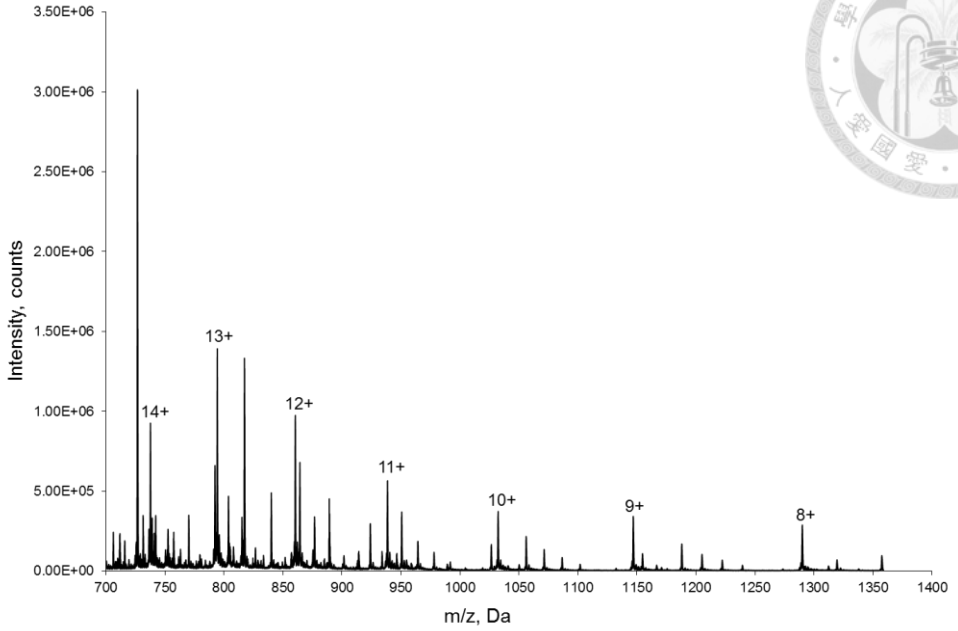


Figure 15. Molecular mass determination of Ub-Cysteamine.

(A) ESI-MS spectrum of Ub-Cysteamine (B) The deconvoluted ESI-MS spectrum of Ub-Cysteamine. The calculated molecular weight is 10374 Da; the found molecular weight is10372 Da.

doi:10.6342/NTU201803256

(A)



(B)

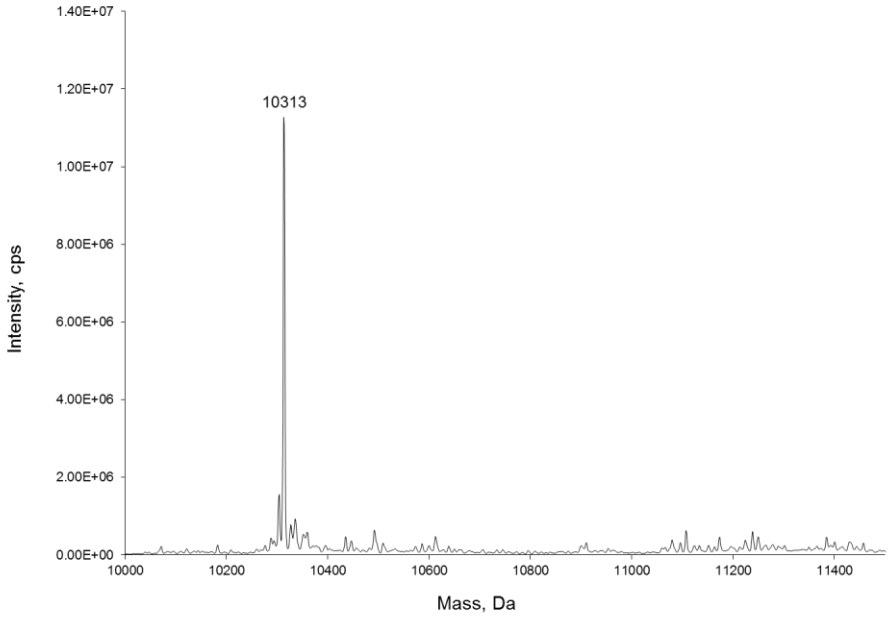


Figure 16. Molecular mass determination of Ub-Cysteamine in the absence of ATP.

(A) ESI-MS spectrum of Ub-Cysteamine in the absence of ATP (B) The deconvoluted ESI-MS spectrum of Ub-Cysteamine in the absence of ATP. The calculated molecular weight is 10314 Da; the found molecular weight is 10313 Da.

doi:10.6342/NTU201803256

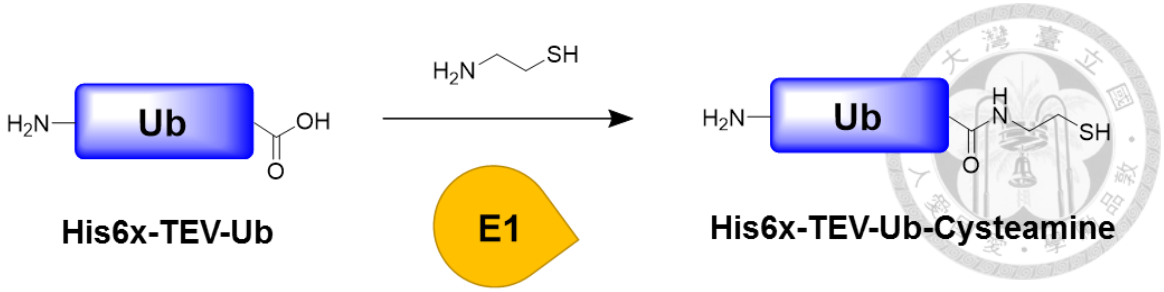


Figure 17. Protein chemistry study of Ubiquitin-cysteamine.

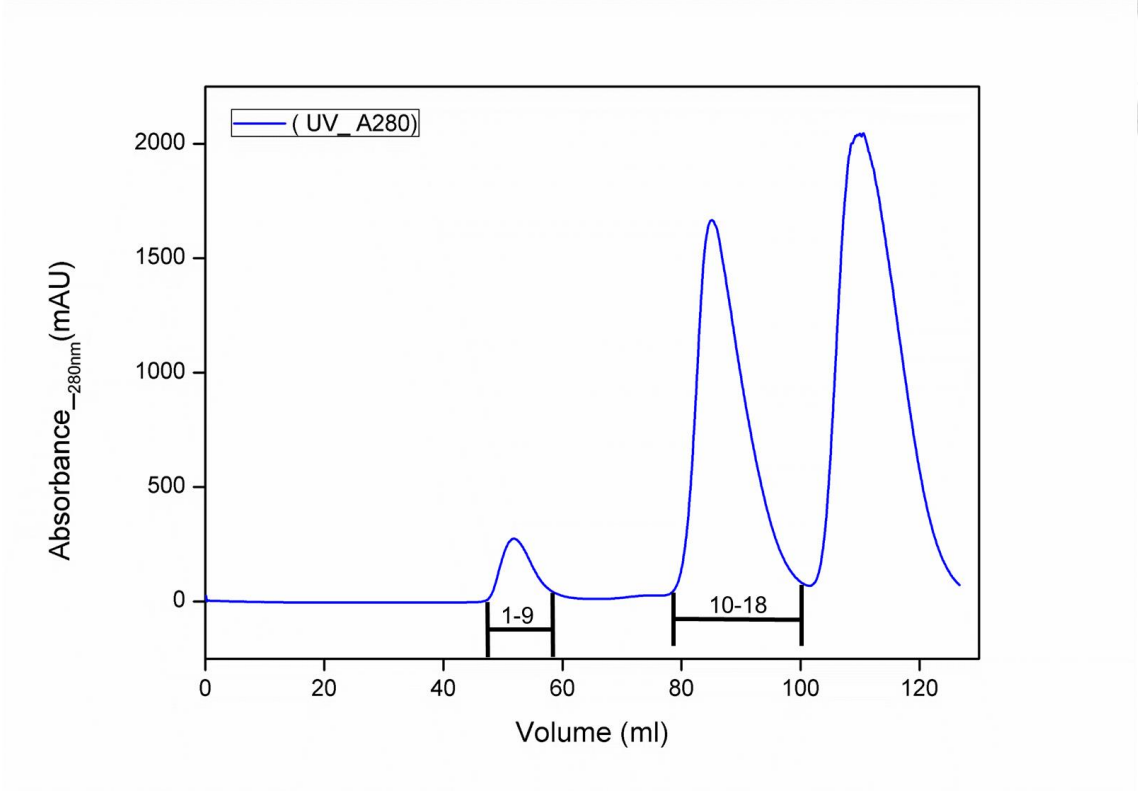
doi:10.6342/NTU201803256

3.1.4 Purification of Ubiquitin-cysteamine

Next, when the ESI-MS confirmed the results of Ubiquitin derivative on its C-terminal modification, the Ubiquitin-cysteamine was purified by size-exclusion chromatography. Finally, the SDS-PAGE analyzed the purity of the Ubiquitin-cysteamine (**Figure 18**).

doi:10.6342/NTU201803256

(A)



(B)

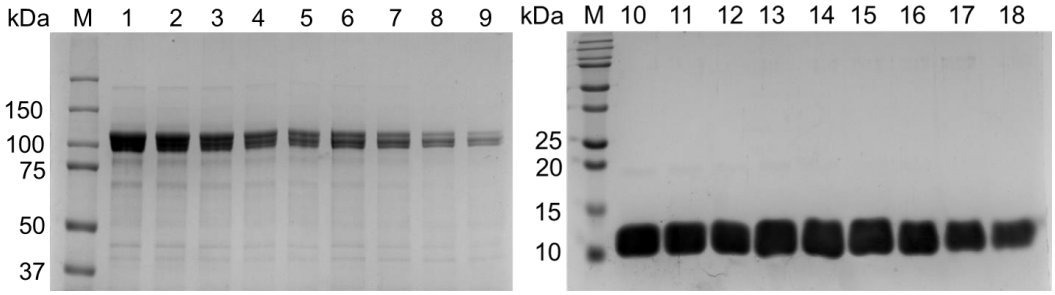


Figure 18. Purification of Ubiquitin-cysteamine with the SEC column.
(A) The ubiquitin with C terminus modification were concentrated for FPLC with S75 column (HiLoad 16/600 superdex 75 pg, GE Healthare). Blue line: the signal of UV_{A280} (B) SDS-PAGE shows the fractions indicated in (A). 1-9: minor peak; 10-18: secondary peak. Fraction 10-18 were collected and concentrated for reaction.

doi:10.6342/NTU201803256

3.2 Studying selenoxide β-elimination of SUMO2 variants

3.2.1 Purification of SUMO2 variants

In order to generate SUMO2 protein with ncAA, we utilized evolved PylRS, mkRS1F combines with $tRNA_{CUA}^{Pyl}$ to support the specific incorporation of Se-alkylselenocysteine at amber codon of SUMO2 protein in *E.coli* (**Figure 19**). The mkRS1F has the mutation site on Y306M, L309A, C348T, T364K, and Y384F from *Mm*PylRS. The constructs of pET-pylT-sfGFP-SUMO2KXTAG (X represents lysine position) (**Table 1**) and pCDF- mkRS1F have been identified by the DNA sequencing results.

Initially, the plasmid pET-pylT-sfGFP-SUMO2KXTAG (X represents lysine position) and pCDF- mkRS1F were chemically co-transformed into *E. coli* BL21 (DE3) and cultured in LB medium in *E. coli* with 1 mM IPTG and 1 mM **2** (**Figure 20**) at 37℃ for 12 hrs. Then, the eight different lysines at amber mutation site on SUMO2 protein fused with N-terminal sfGFP were purified by Ni^{2+}-NTA affinity chromatography, and run the SDS-PAGE to analyze diverse amber suppression. The SDS-PAGE displayed that amber suppression yield is about 21, 27, 69, 56, 67, 30, 49, 34% (**Figure 21**). The suppression yields for different linkages are conspicuous. Further, to gain the complete SUMO2 variants, TEV protease was used to cleave the TEV cutting site between the sfGFP and SUMO2 variant. Later, we separated SUMO2 variants from mixture by Ni^{2+}-NTA affinity chromatography, anion-exchange chromatography, and size-exclusion chromatography. Purity of SUMO2 variants were analyzed by 15% SDS-PAGE (**Figure 22**).

doi:10.6342/NTU201803256

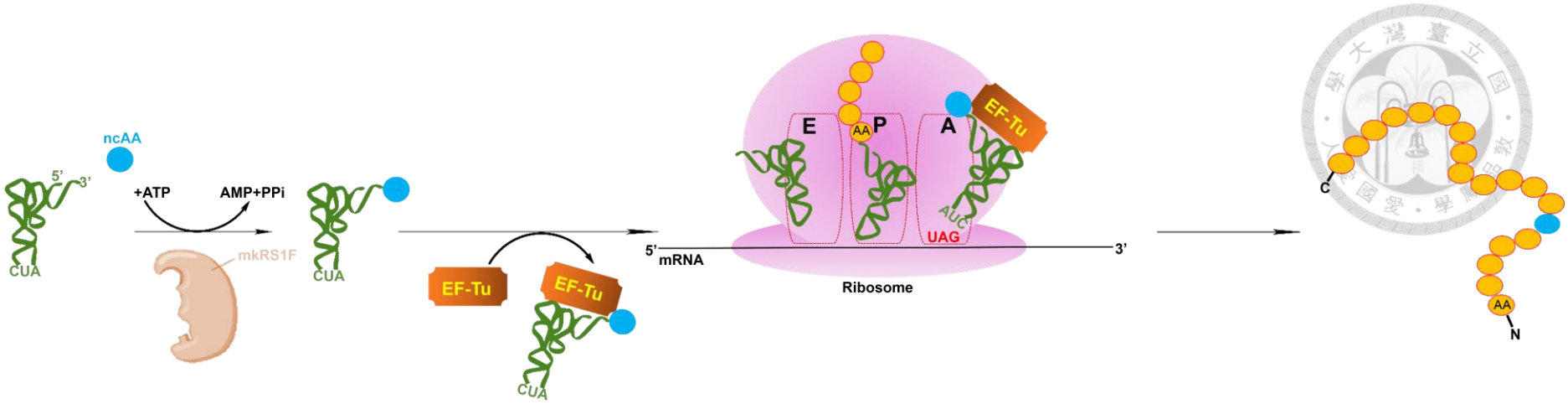


Figure 19. Protein synthesis with 2 through expanding genetic code method.

Protein translation by using pyrrolysyl-tRNA synthetase•$tRNA^{Pyl}_{CUA}$ bioorthogonal pair through amber codon suppression. PylRS is actually an aminoacyl-tRNA synthetase which derived from archaea and it is orthogonal in *E. coli*. In this thesis, engineerd PlyRS, mkRS1F would catalyze the esterification between suppressor tRNA and appropriate **2**. With the assistance of EF-TU, the aminoacyl-tRNA can recognize the corresponding amber codon on mRNA.

doi:10.6342/NTU201803256

Table 1. Mutated lysine residues of SUMO2 in this study.

Amber mutation in sfGFP-SUMO2								
Name/**Position**	**K5**	**K7**	**K11**	**K21**	**K33**	**K35**	**K42**	**K45**
GS2K5	TAG	-	-	-	-	-	-	-
GS2K7	-	TAG	-	-	-	-	-	-
GS2K11	-	-	TAG	-	-	-	-	-
GS2K21	-	-	-	TAG	-	-	-	-
GS2K33	-	-	-	-	TAG	-	-	-
GS2K35	-	-	-	-	-	TAG	-	-
GS2K42	-	-	-	-	-	-	TAG	-
GS2K45	-	-	-	-	-	-	-	TAG

doi:10.6342/NTU201803256

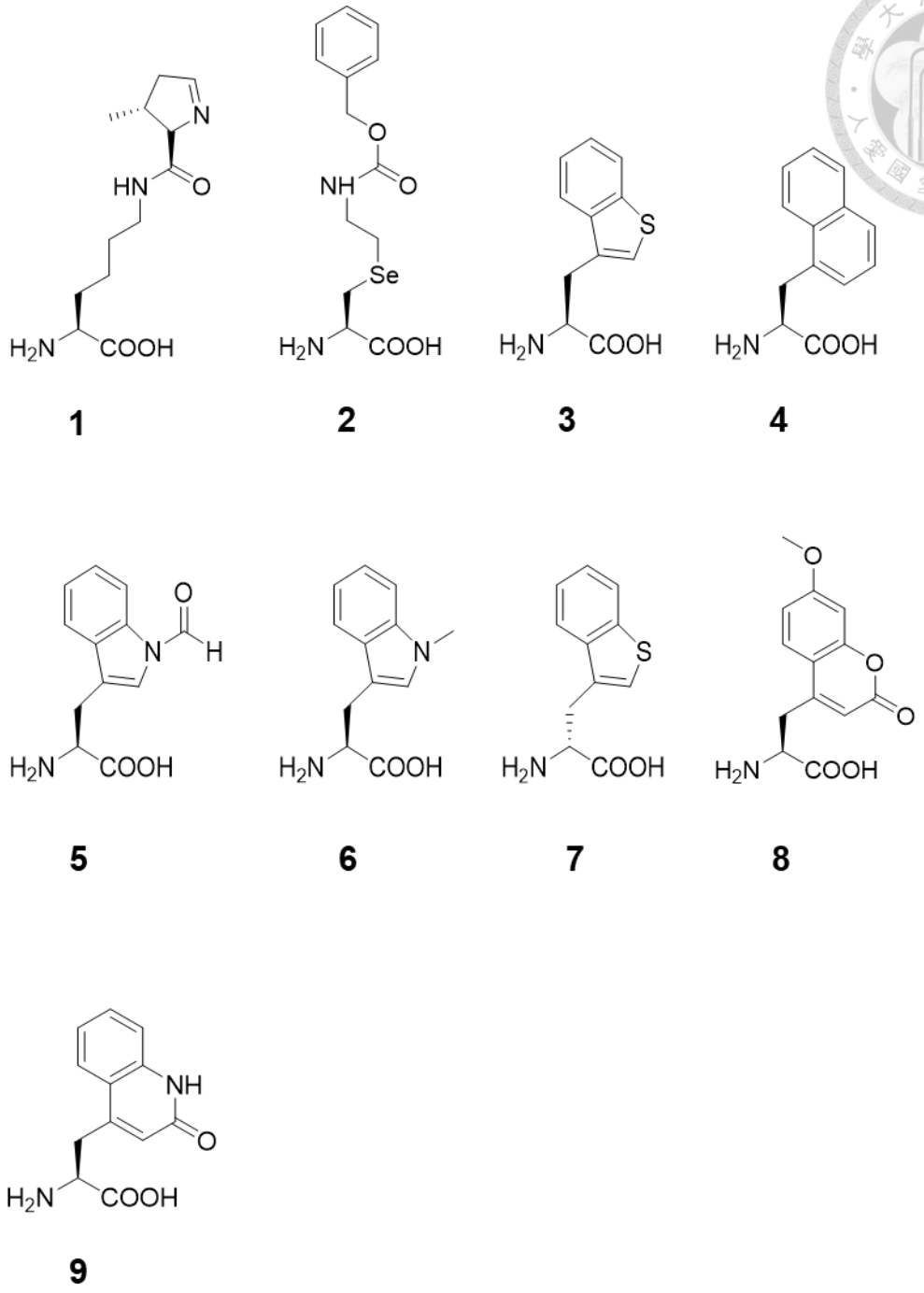
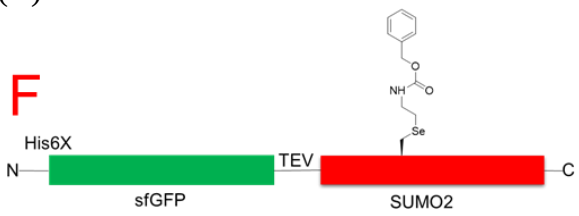


Figure 20. Chemical structures of ncAAs in this study.
1: L-Pyrrolysine (Pyl) **2**: Se-alkylselenocysteine (SeCbzK) **3**: 3-Benzothienyl-L-alanine (LBTA) **4**: 3-(1-Naphthyl)-L-alanine (1NapA) **5**: N^{in}-Formyl-L-tryptophan (1ForTrp) **6**: 1-Methyl-L-tryptophan (1MeTrp) **7**: 3-Benzothienyl-D-alanine (DBTA) **8**: H-*β*-(7-Methoxycoumarin-4-yl)-Ala-OH (7MeOCuA) **9**: 3-(2-Oxo-1,2-dihydro-4-quinolinyl)alanine (QunA).

doi:10.6342/NTU201803256

(A)



(B)

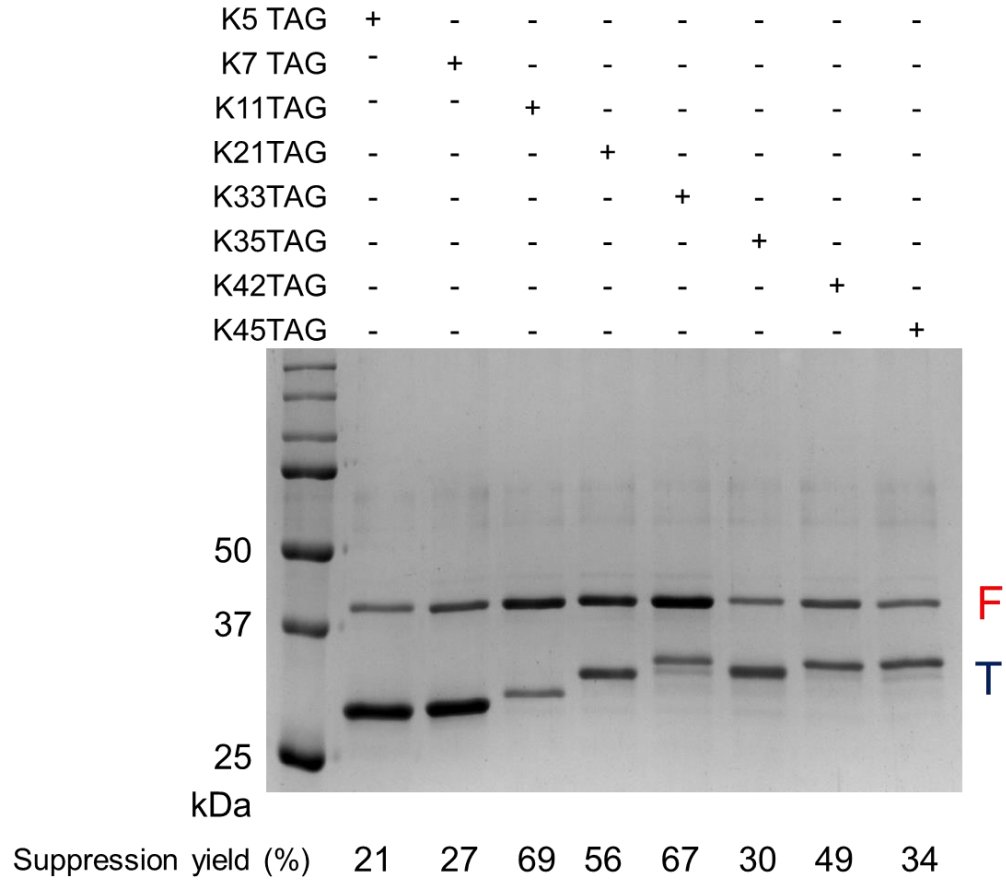


Figure 21. Purification of sfGFP-SUMO2 variants with Ni^{2+}–NTA column.
(A) Illustration of synthesis sfGFP-SUMO2 variants through amber suppression.
(B) sfGFP-SUMO2 protein production with site-specific **2** incorporation. mkRS1F•$tRNA^{Pyl}_{CUA}$ pair was utilized for these experiments in LB medium with 1mM IPTG and 1mM ncAA **2** at 37℃ for 12 hrs. Amber suppression yield is about 21 to 69% F: full length sfGFP-SUMO2, T: truncated protein

doi:10.6342/NTU201803256

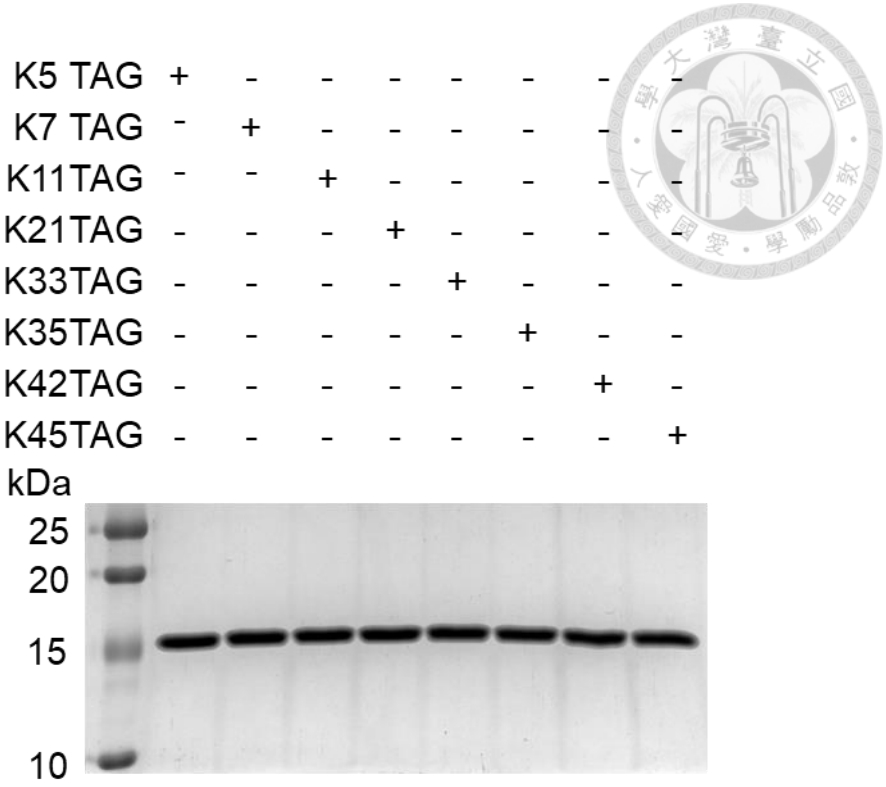


Figure 22. Purification of SUMO2 variants.

SUMO2 protein production with site-specific **2** incorporation. mkRS1F•tRNA$^{Pyl}_{CUA}$ pair was utilized for these experiments in LB medium with 1mM IPTG and 1mM ncAA **2** at 37°C for 12 hrs. Purification of SUMO2 variants with Ni^{2+}–NTA affinity chromatography, anion-exchange chromatography, and size exclusion chromatography.

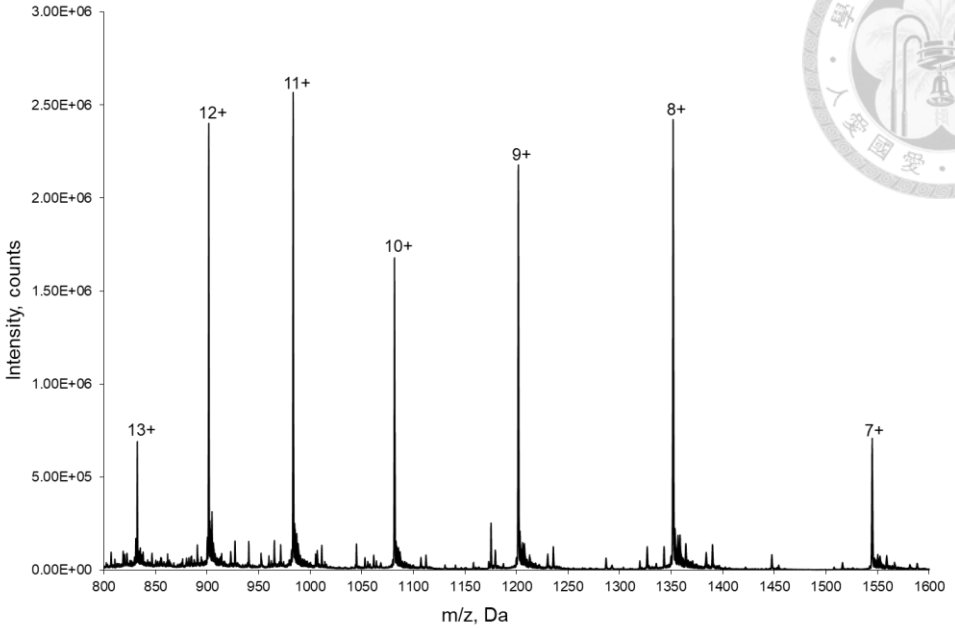
doi:10.6342/NTU201803256

3.2.2 ESI-MS analysis and MALDI-TOF-MS/MS analysis of SUMO2 variants

Finally, the purified SUMO2 protein was dissected the expected molecular weight (MW) by ESI-MS apparatus (**Figure 23-30**). In addition, to confirm the SUMO2 protein was genetically incorporated **2** at the particular position, the target proteins were digested by trypsin protease and analyzed by MALDI-TOF-MS/MS. As our expectation, the fragments containing **2** at K5, K7, and K45 have been searched, (**Figure 31-33**) but the fragment of other lysine position were not observed the consequences. Thus, other enzymatic digested condition will be tested to search the fragment containing **2** at K11, K21, K33, K35, and K42 position.

doi:10.6342/NTU201803256

(A)



(B)

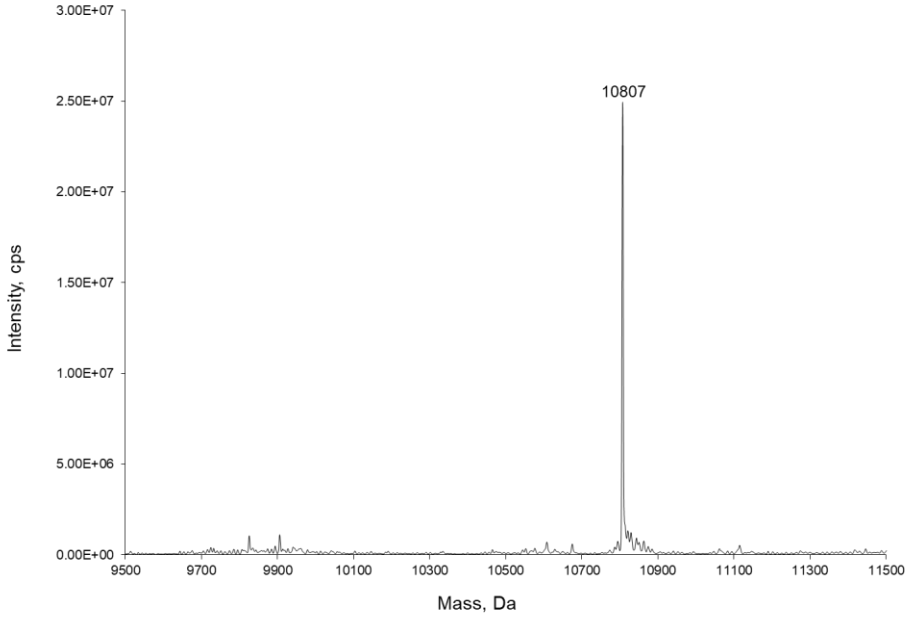
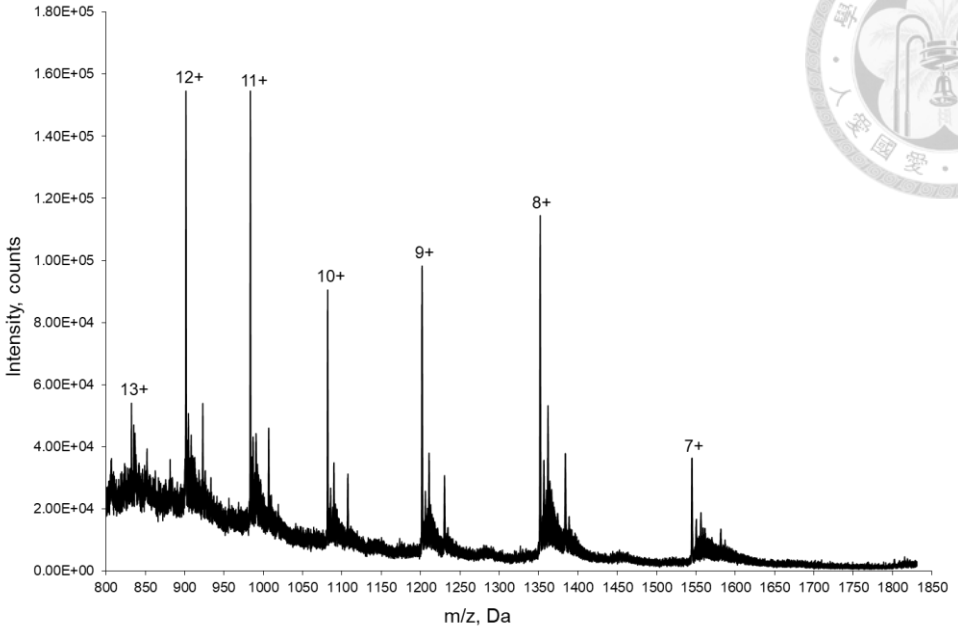


Figure 23. Molecular mass determination of the protein SUMO2K5-2.
(A) ESI-MS spectrum of SUMO2K5-**2** (B) The deconvoluted ESI-MS spectrum of SUMO2K5-**2.** The calculated molecular weight is 10808 Da; the found molecular weight is 10807 Da. The proteins were produced by mkRS1F•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **2** in *E. coli* BL21 (DE3) cells in LB medium at 37 ºC for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

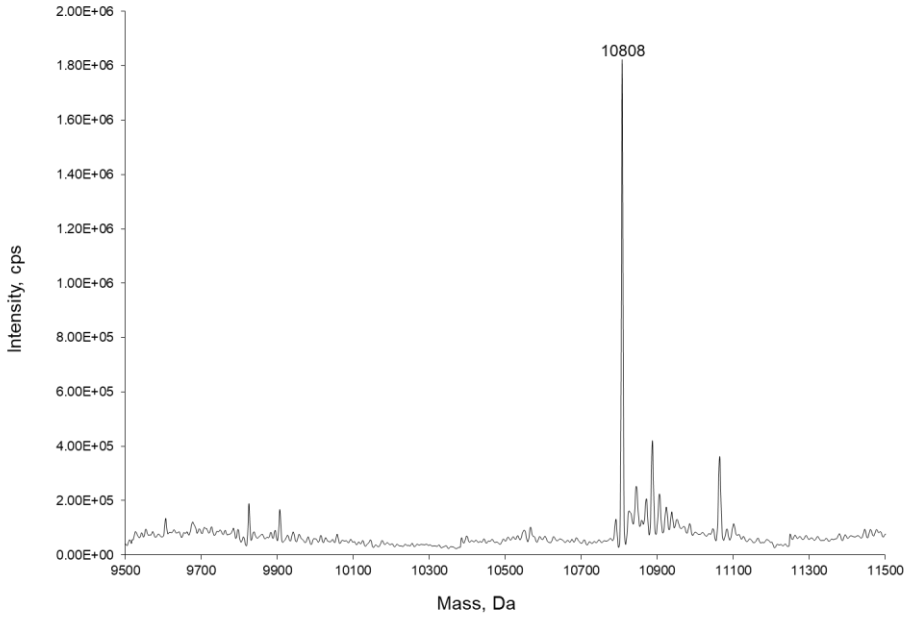
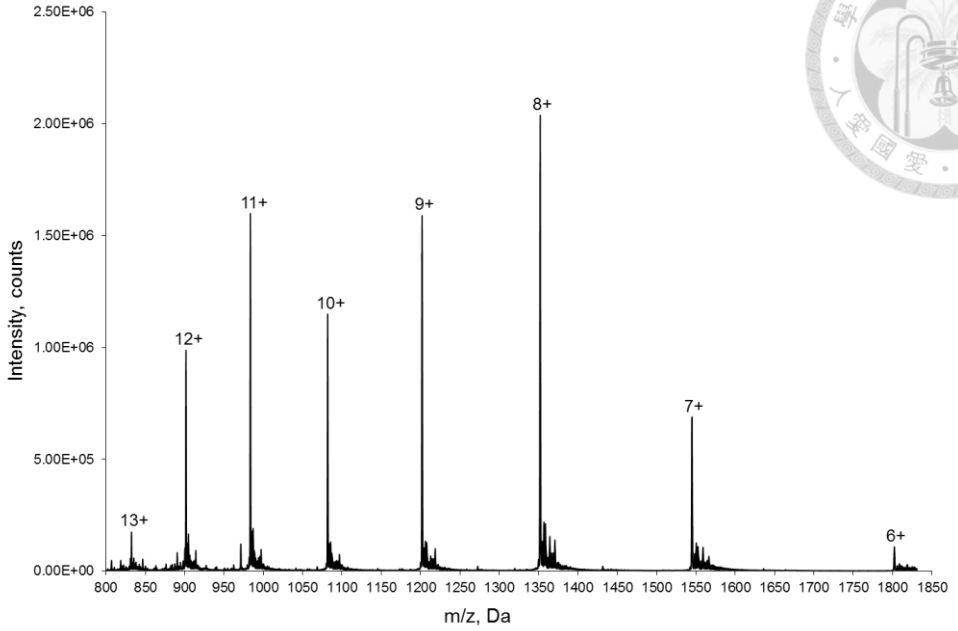


Figure 24. Molecular mass determination of the protein SUMO2K7-2.

(A) ESI-MS spectrum of SUMO2K7-**2** (B) The deconvoluted ESI-MS spectrum of SUMO2K7-**2.** The calculated molecular weight is 10808 Da; the found molecular weight is 10808 Da. The proteins were produced by mkRS1F•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **2** in *E. coli* BL21 (DE3) cells in LB medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

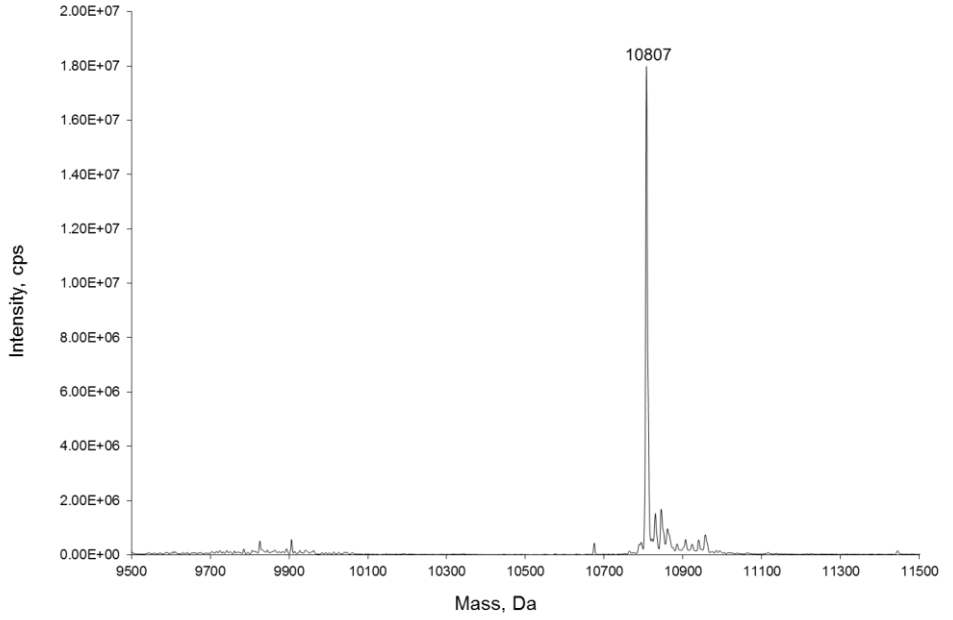
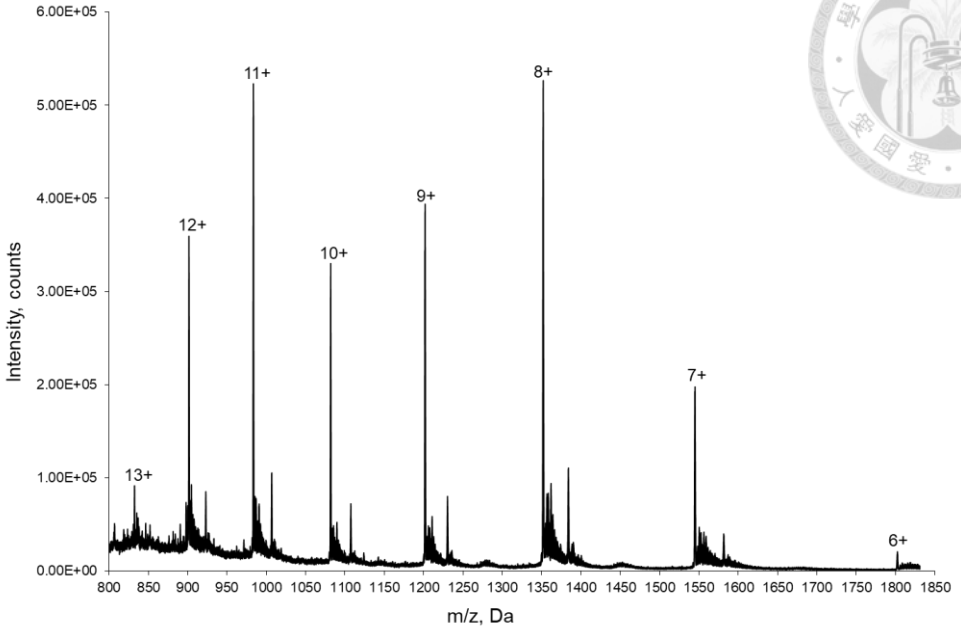


Figure 25. Molecular mass determination of the protein SUMO2K11-2.
(A) ESI-MS spectrum of SUMO2K11-**2** (B) The deconvoluted ESI-MS spectrum of SUMO2K11-**2.** The calculated molecular weight is 10808 Da; the found molecular weight is 10807 Da. The proteins were produced by mkRS1F•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **2** in *E. coli* BL21 (DE3) cells in LB medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

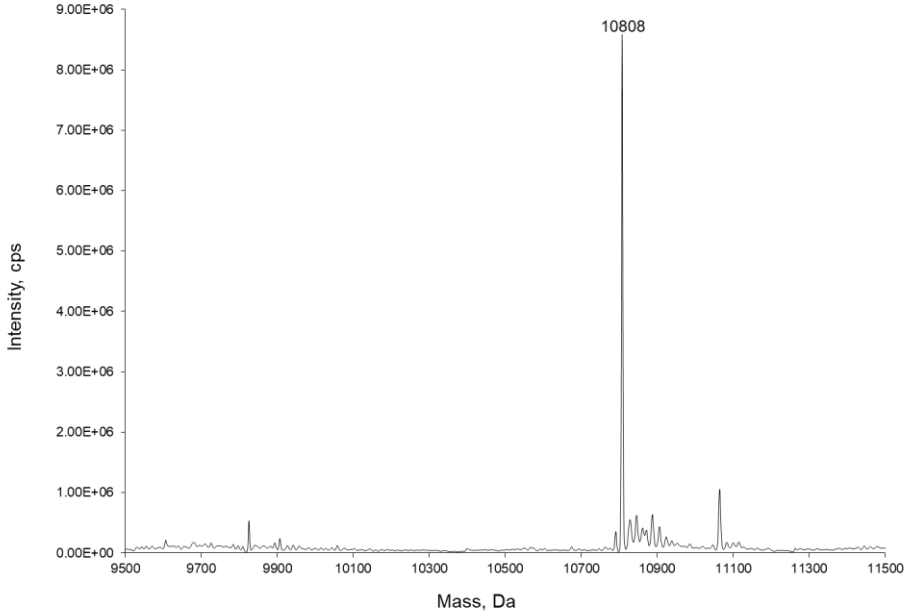
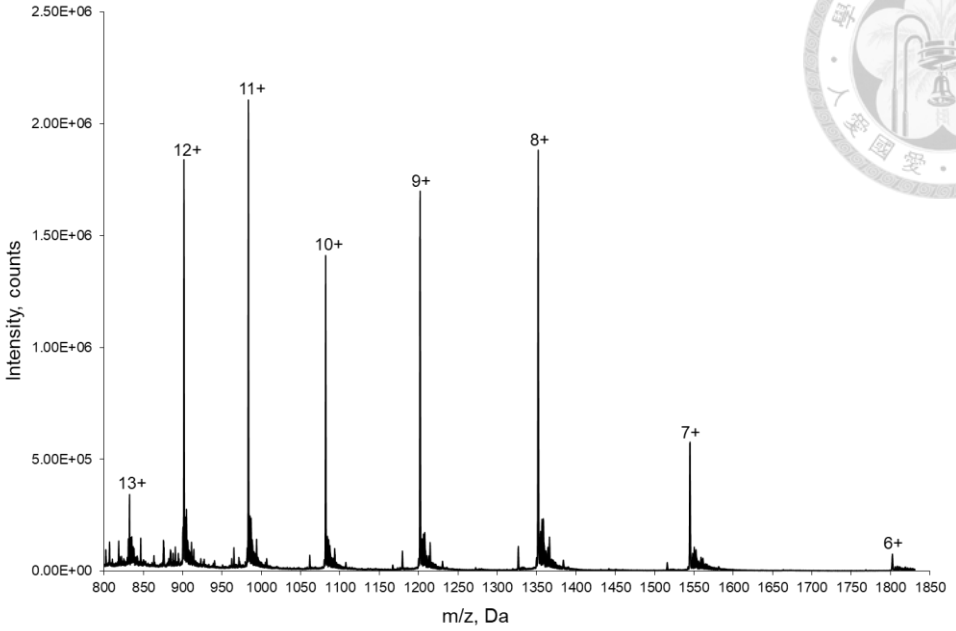


Figure 26. Molecular mass determination of the protein SUMO2K21-2.
(A) ESI-MS spectrum of SUMO2K21-**2** (B) The deconvoluted ESI-MS spectrum of SUMO2K21-**2.** The calculated molecular weight is 10808 Da; the found molecular weight is 10808 Da. The proteins were produced by mkRS1F•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **2** in *E. coli* BL21 (DE3) cells in LB medium at 37 ^{o}C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

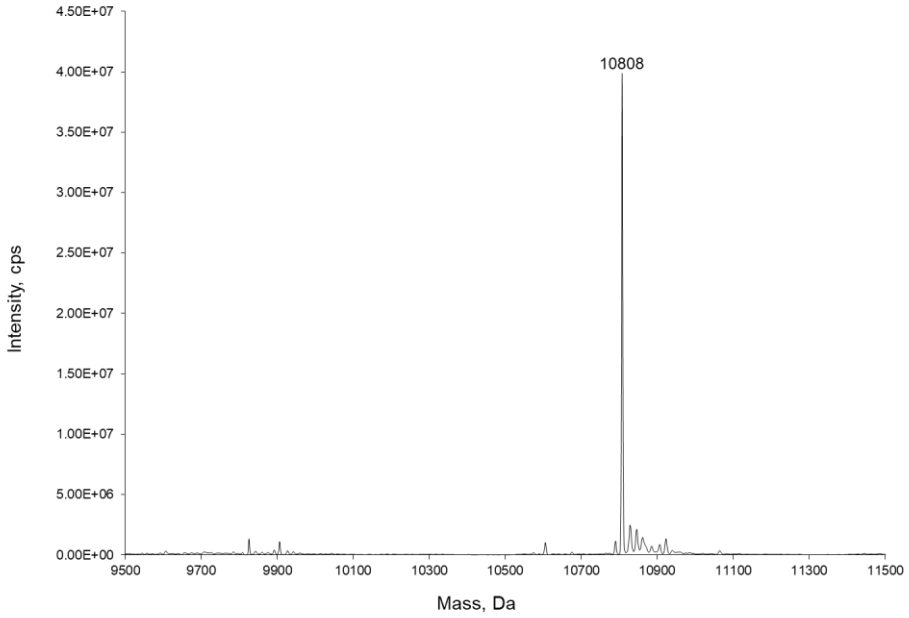
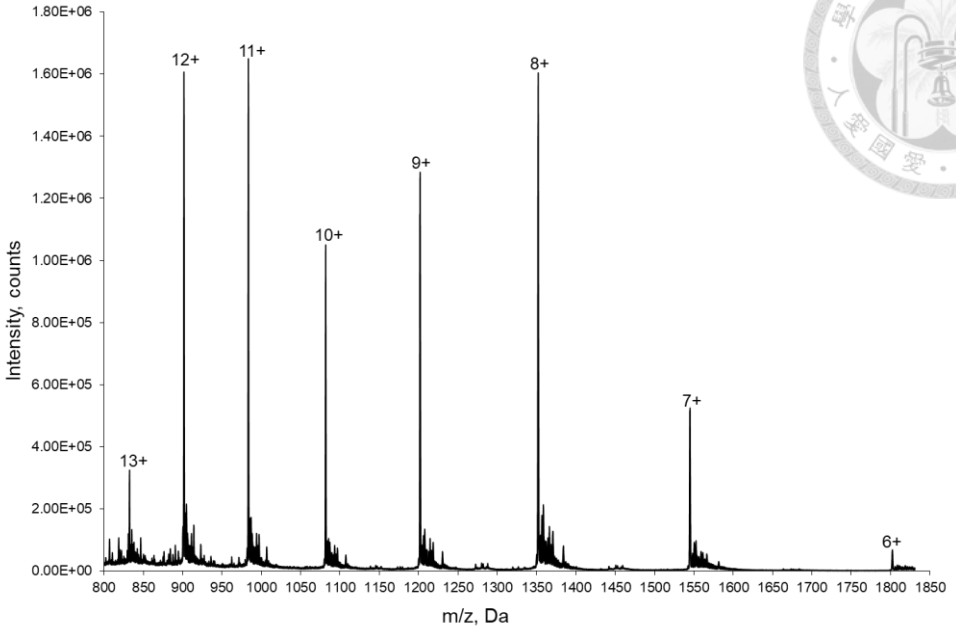


Figure 27. Molecular mass determination of the protein SUMO2K33-2.
(A) ESI-MS spectrum of SUMO2K33-**2** (B) The deconvoluted ESI-MS spectrum of SUMO2K33-**2.** The calculated molecular weight is 10808 Da; the found molecular weight is 10808 Da. The proteins were produced by mkRS1F•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **2** in *E. coli* BL21 (DE3) cells in LB medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

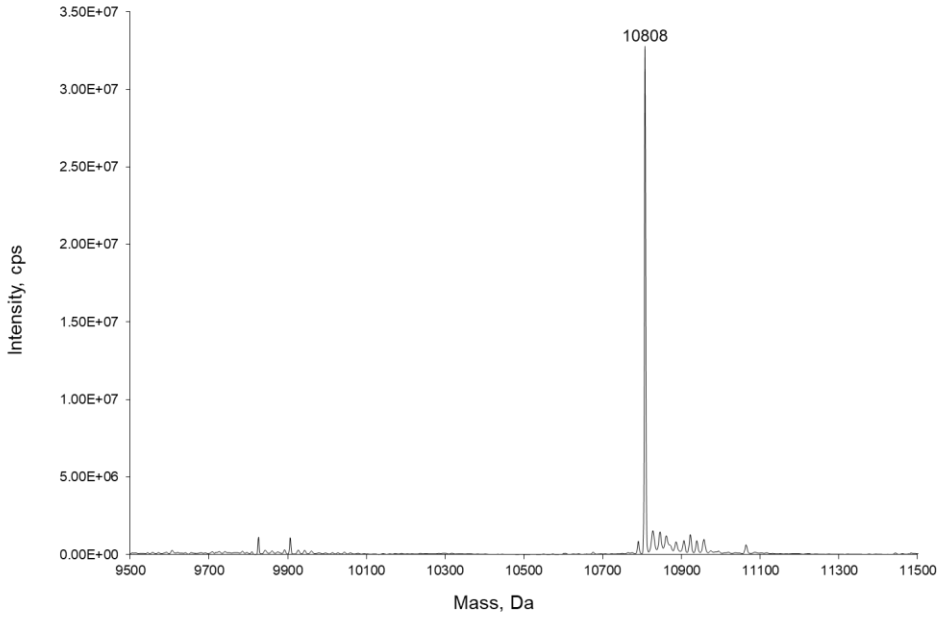
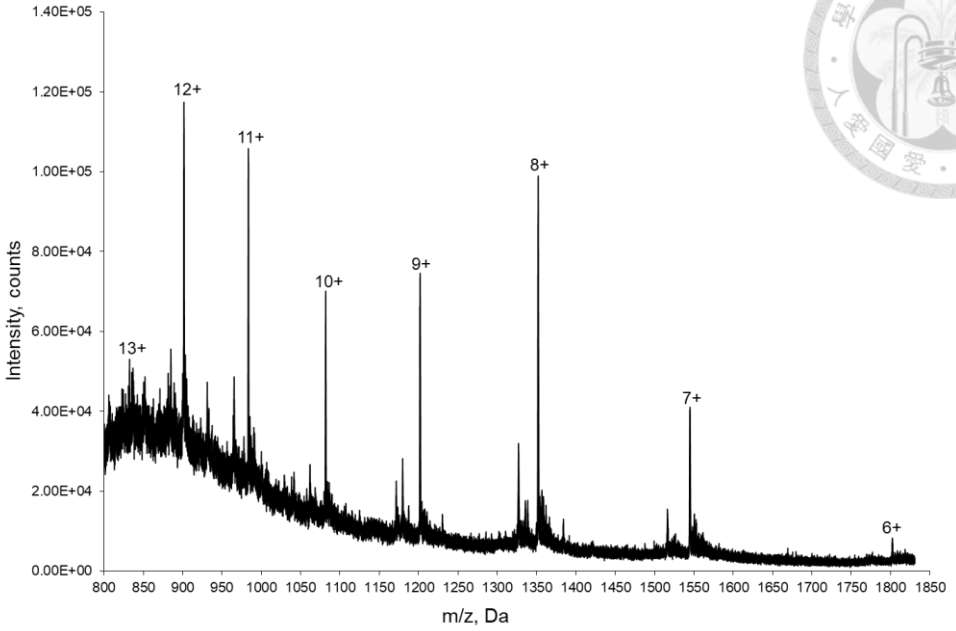


Figure 28. Molecular mass determination of the protein SUMO2K35-2.
(A) ESI-MS spectrum of SUMO2K35-**2** (B) The deconvoluted ESI-MS spectrum of SUMO2K35-**2.** The calculated molecular weight is 10808 Da; the found molecular weight is 10807 Da. The proteins were produced by mkRS1F•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **2** in *E. coli* BL21 (DE3) cells in LB medium at 37 ^{o}C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

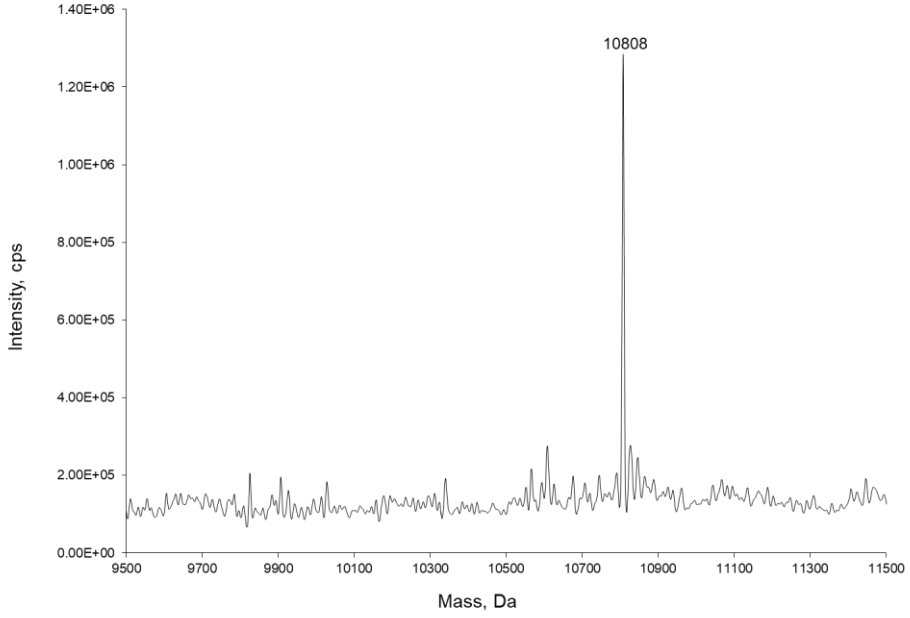
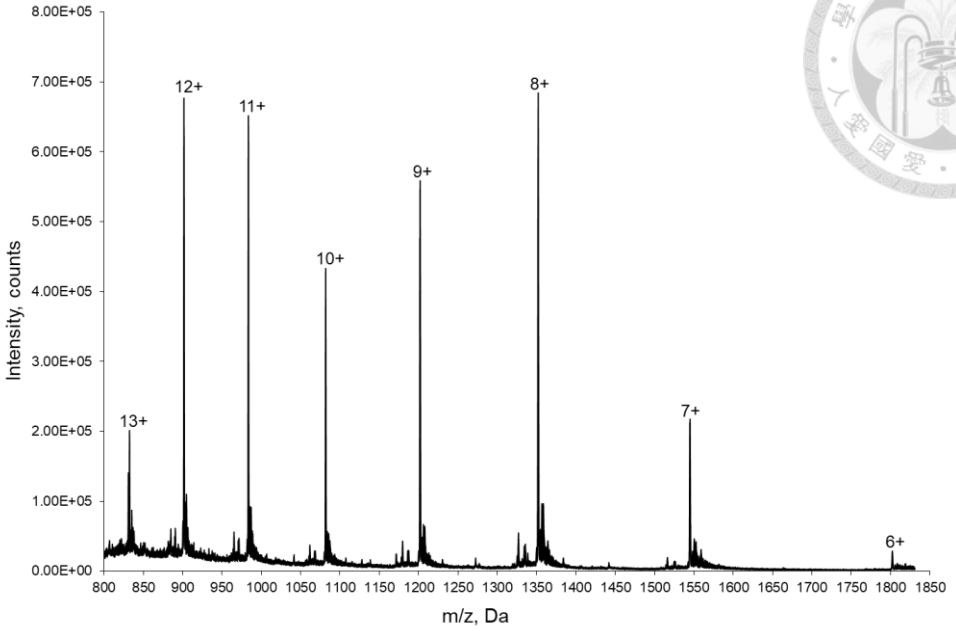


Figure 29. Molecular mass determination of the protein SUMO2K42-2.
(A) ESI-MS spectrum of SUMO2K42-**2** (B) The deconvoluted ESI-MS spectrum of SUMO2K42-**2.** The calculated molecular weight is 10808 Da; the found molecular weight is 10808 Da. The proteins were produced by mkRS1F•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **2** in *E. coli* BL21 (DE3) cells in LB medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

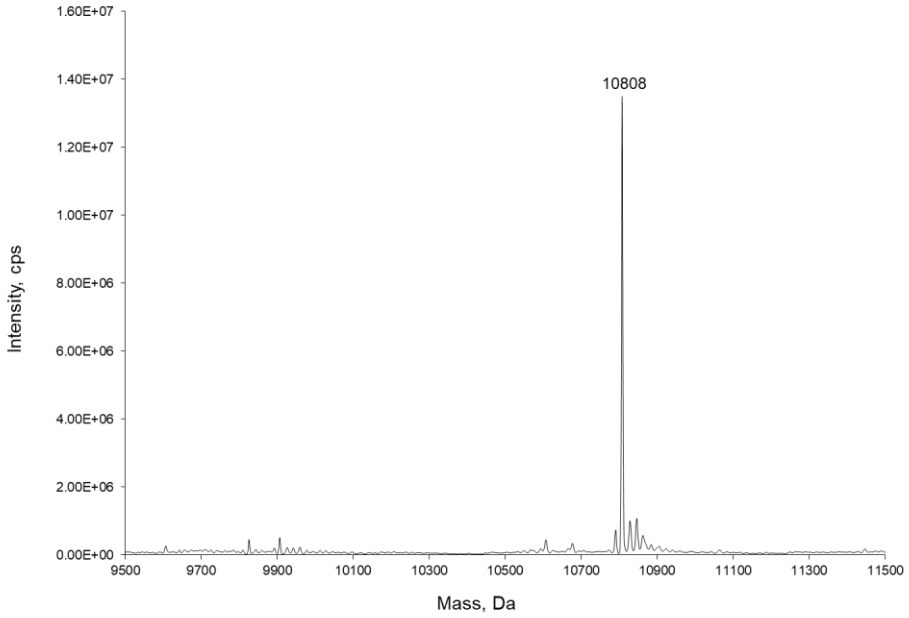


Figure 30. Molecular mass determination of the protein SUMO2K45-2.
(A) ESI-MS spectrum of SUMO2K45-**2** (B) The deconvoluted ESI-MS spectrum of SUMO2K45-**2.** The calculated molecular weight is 10808 Da; the found molecular weight is 10808 Da. The proteins were produced by mkRS1F•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **2** in *E. coli* BL21 (DE3) cells in LB medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

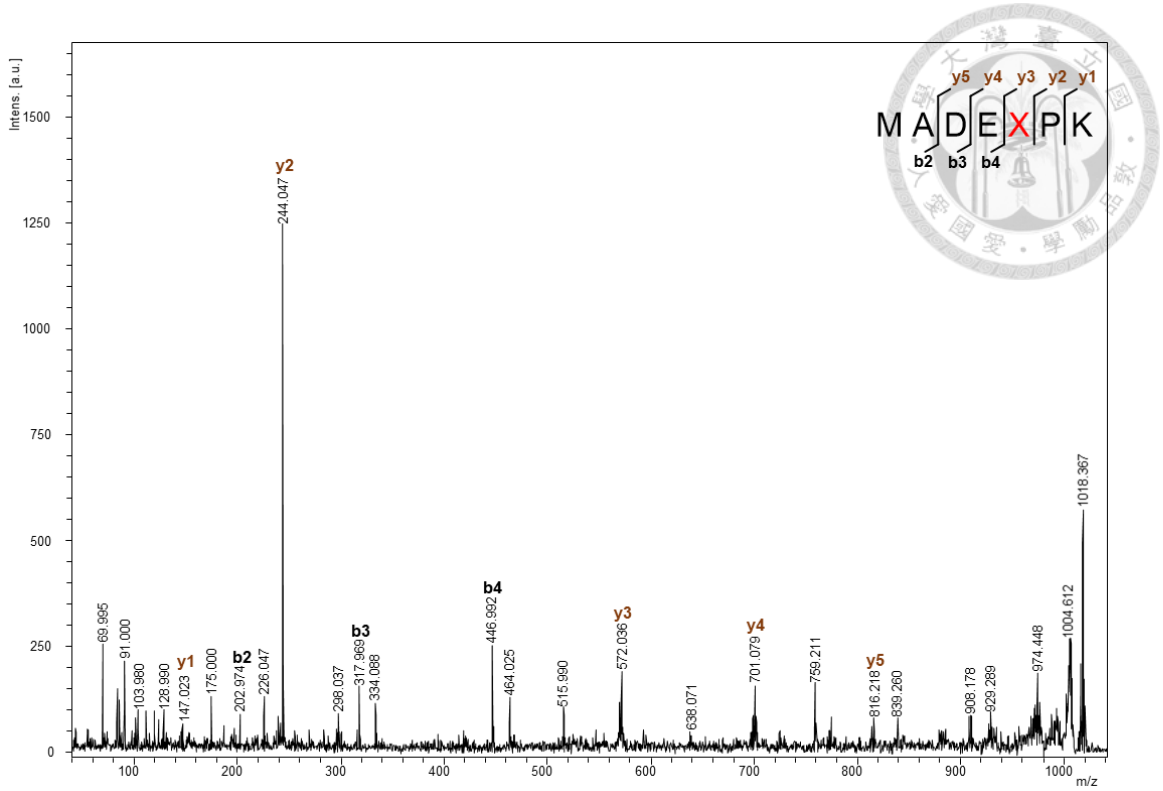


Figure 31. MALDI-TOF-MS/MS analysis of SUMO2K5-2.

Fragment $M^{1}ADEXPK^{7}$, X represents ncAA **2**. The full-length SUMO2K5-2 protein was produced by mkRS1F • $tRNA^{Pyl}_{CUA}$ pair in the presence of 1 mM ncAA **2**. The protein was in-gel digested by trypsin.

doi:10.6342/NTU201803256

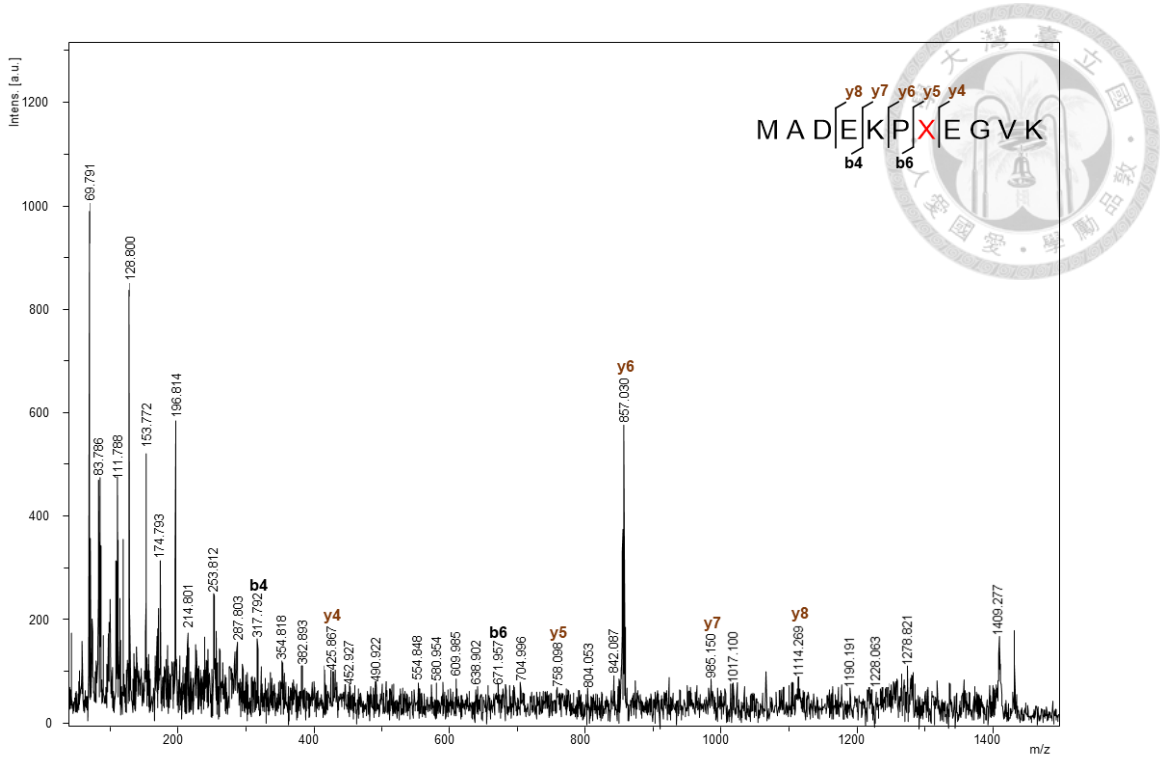


Figure 32. MALDI-TOF-MS/MS analysis of SUMO2K7-2.

Fragment $M^{1}ADEKPXEGVK^{11}$, X represents ncAA **2**. The full-length SUMO2K7-2 protein was produced by mkRS1F • $tRNA^{Pyl}_{CUA}$ pair in the presence of 1 mM ncAA **2**. The protein was in-gel digested by trypsin.

doi:10.6342/NTU201803256

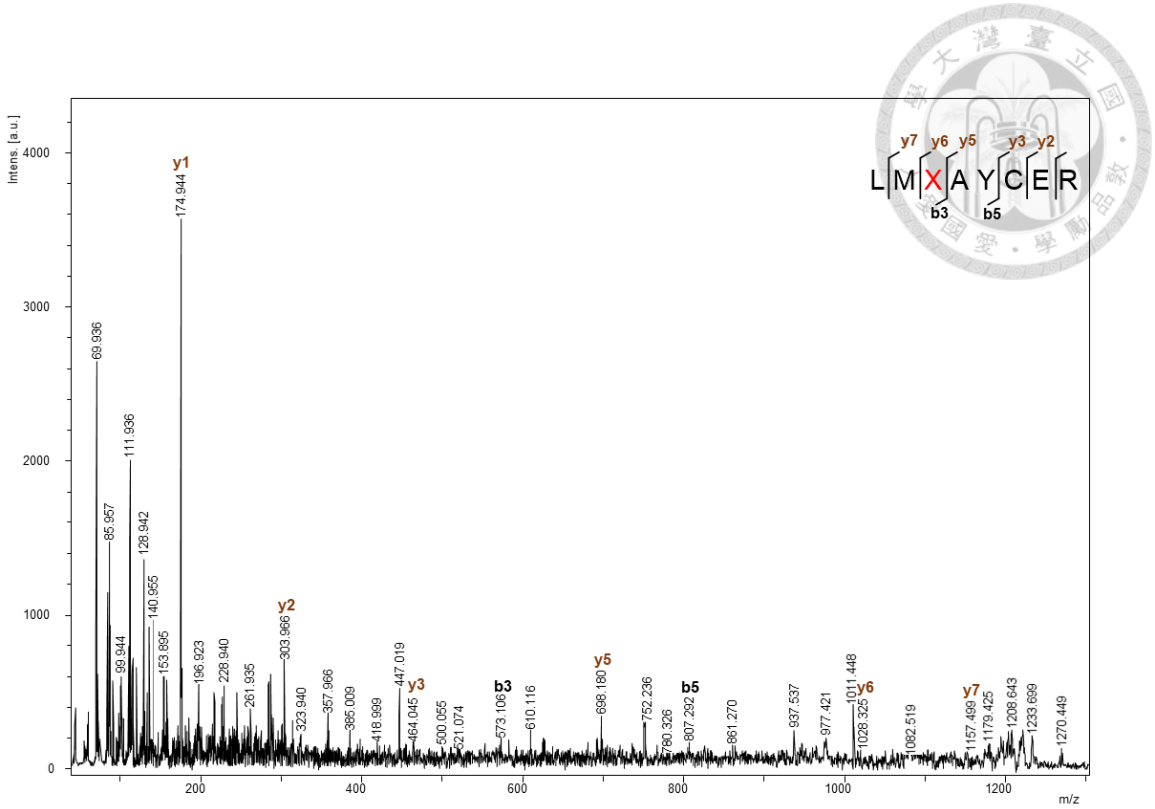


Figure 33. MALDI-TOF-MS/MS analysis of SUMO2K45-2.

Fragment $L^{43}MXAYCER^{50}$, X represents ncAA **2**. The full-length SUMO2K45-2 protein was produced by mkRS1F • $tRNA^{Pyl}_{CUA}$ pair in the presence of 1 mM ncAA **2**. The protein was in-gel digested by trypsin.

doi:10.6342/NTU201803256

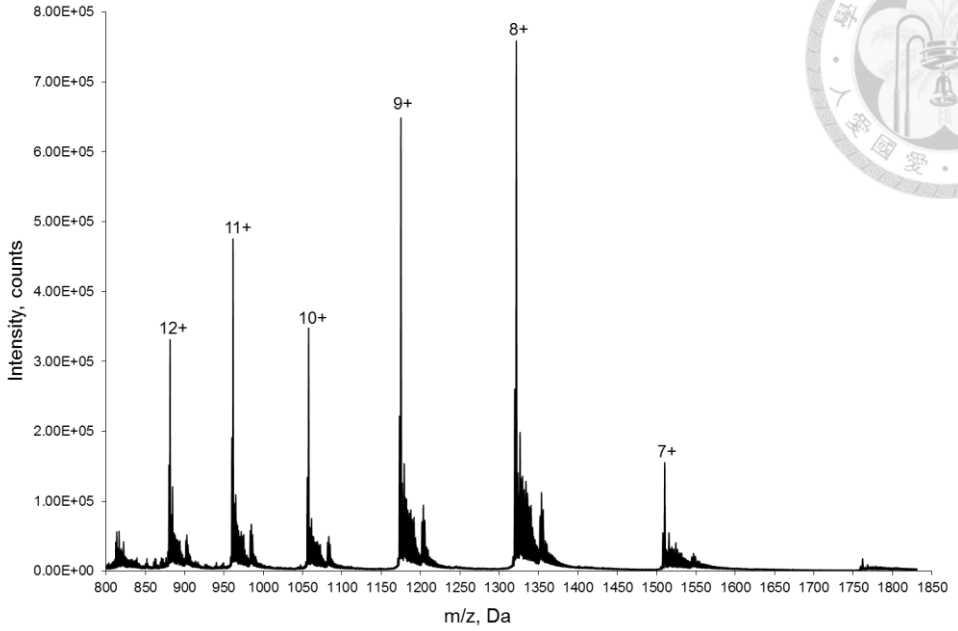
3.2.3 ESI-MS analysis of selenoxide β-elimination of SUMO2 variants

The success of incorporation of **2** on the SUMO2 proteins were demonstrated, we have focused on synthesizing SUMO2 proteins with dehydroalanine (Dha) at eight different lysines position. Previous studies have indicated that Se-alkylselenocysteine in a protein can undergo oxidative elimination to generate dehydroalanine. We chose the SUMO2K11-**2** with high suppression yield to try conditions. To convert SeCbzK in SUMO2K11-**2** to dehydroalanine, we utilized H_2O_2 to generate selenoxide β-elimination in PBS buffer. When the equivalent of H_2O_2 was increased, unexpected products were gradually escalated. Single and multiple oxygen atom(s) were added to the SUMO2 variant after ESI-MS characterization.

The expected molecular weight 10550 Da, and ESI-MS analysis of the final products indicated the found molecular weight 10549 Da, 10565 Da (**Figure 34**). Thus, we required to diminish severely oxidative reactions. Oxidation of SUMO2K11-**2** with 200 eq. of H_2O_2 for 2hrs in PBS buffer converted **2** to dehydroalanine moderately. The expected molecular weight 10550 Da was found to similar to observed molecular weight 10549 Da along with oxidative product peak at 10565 Da (**Figure 35**). Finally, the SUMO2K11-**2** was successfully converted to the SUMO2K11Dha (**Figure 36**).

doi:10.6342/NTU201803256

(A)



(B)

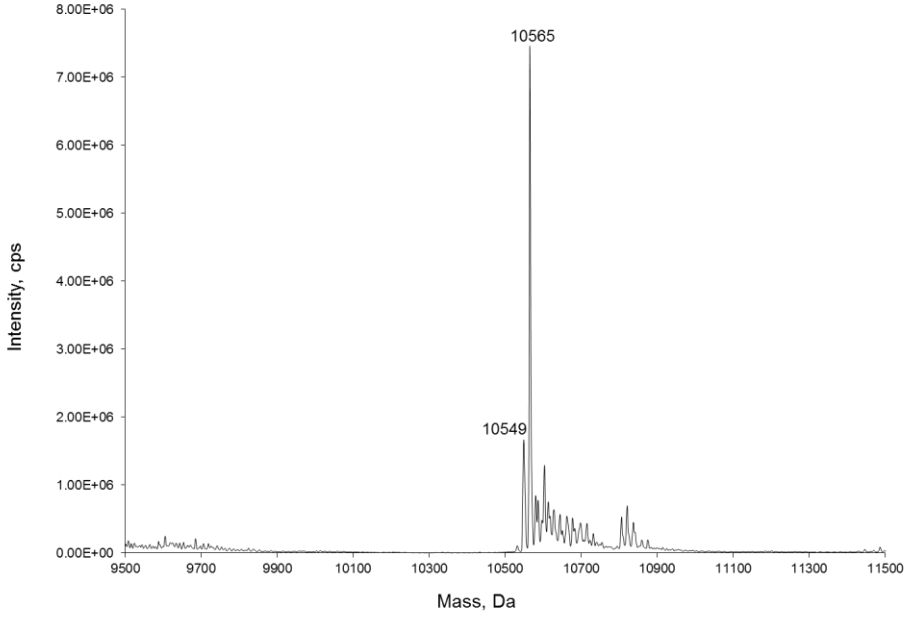
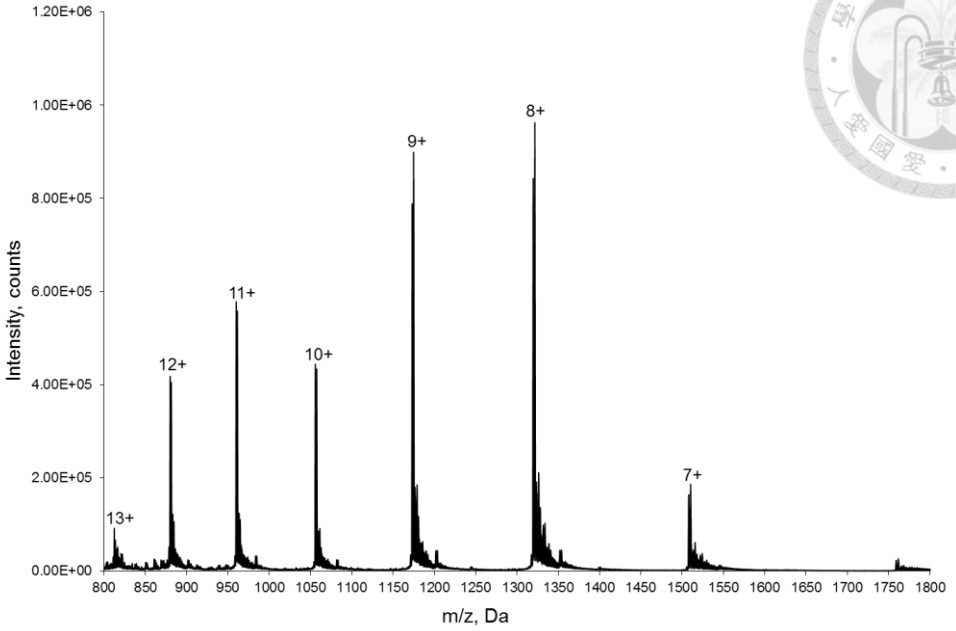


Figure 34. Molecular mass weight determination of SUMO2K11Dha with increased equivalent of H_2O_2.

(A) ESI-MS spectrum of SUMO2K11Dha (B) The deconvoluted ESI-MS spectrum of SUMO2K11Dha. The calculated molecular weight is 10550 Da; the found molecular weight is 10549 Da and 10565 Da.

doi:10.6342/NTU201803256

(A)



(B)

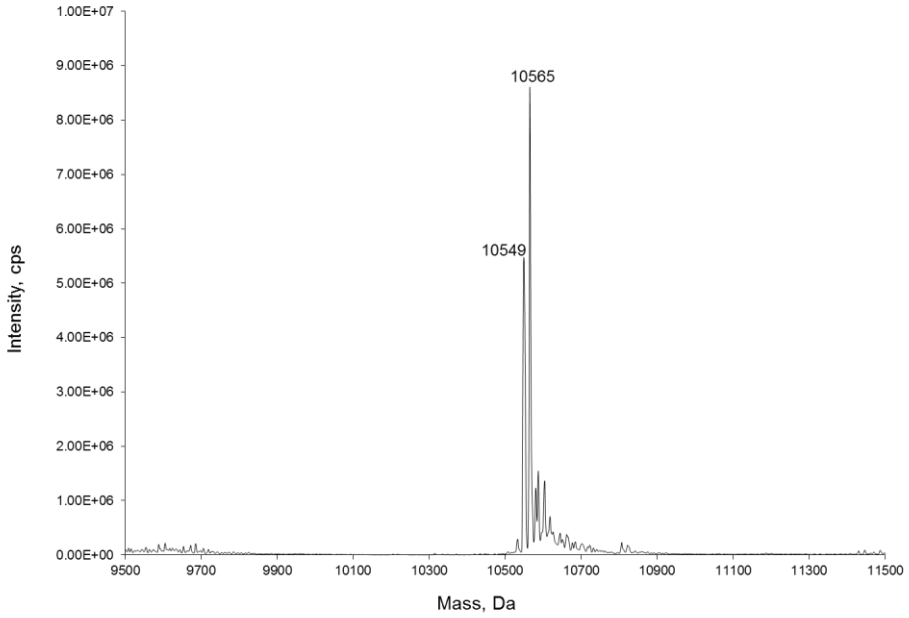


Figure 35. Molecular mass weight determination of SUMO2K11Dha.
(A) ESI-MS spectrum of SUMO2K11Dha (B) The deconvoluted ESI-MS spectrum of SUMO2K11Dha. The calculated molecular weight is 10550 Da; the found molecular weight is 10549 Da and 10565Da.

doi:10.6342/NTU201803256

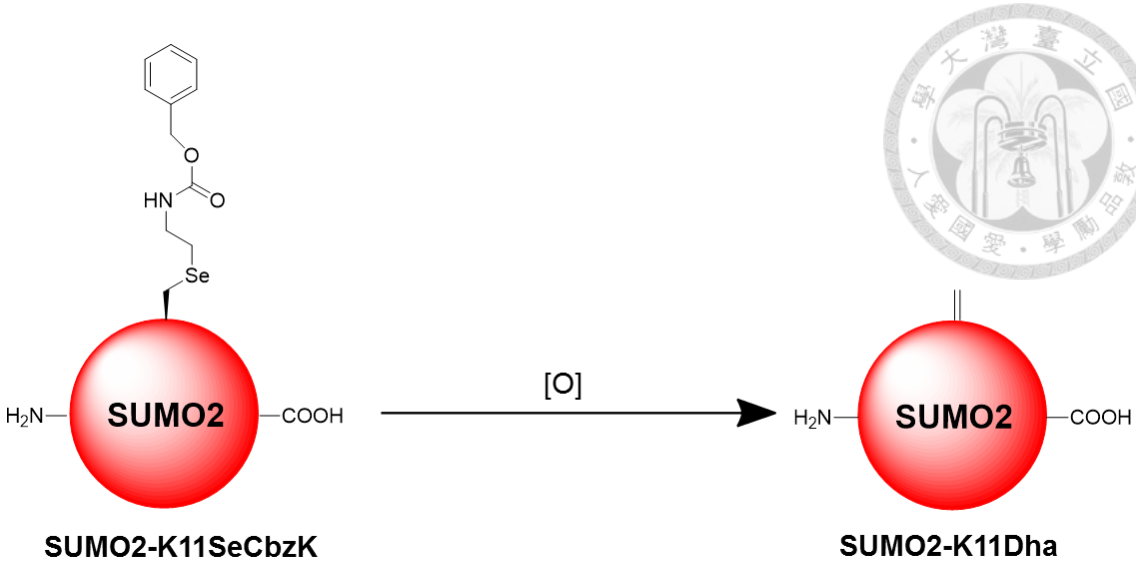


Figure 36. Protein chemistry study of SUMO2 dehydroalanine.

doi:10.6342/NTU201803256

3.3 Protein ligation through thiol-Michael addition on SUMO2Dha

To make sure thiol-Michael addition can go through in the SUMO2 variants (**Figure 37**), SUMO2K11Dha was reacted with small molecule sodium thiophosphate and the results were confirmed by ESI-MS study. Then, to ensure small molecule did not react with other functional groups of amino acids, the SUMO2K11-**2** also react with thiophosphate as control. The expected molecular weight 10808Da, and ESI-MS analysis of the final products indicated the found molecular weight 10807 Da (**Figure 38**). Thiol-Michael addition of SUMO2K11Dha with 500 eq. of thiophosphate at 37℃ for 12hrs resulted in thiophosphate addition to dehydroalanine. Moreover, the small molecule also added on the expected product by disulfide bond. The expected molecular weight 10660Da, and ESI-MS analysis of the final products indicated the found molecular weight 10799 Da, 10816 Da (**Figure 39**).

Certainly, the successful addition of thiophosphate, the Ub-cysteamine was reacted with SUMO2K11Dha. Next, the reagents were mixed with the different molar ratios to find out suitable reaction condition. Later, the results indicated that Ub-cysteamine and the SUMO2K11Dha mixed with 100:1 molar ratio and added 4mM DTT at 4^{o}C for 4 days gave lower yield as observed in the 15% SDS-PAGE (**Figure 40**). When the reaction time was increased, the yield was not improved (**Figure 41**). Thus, the temperature was increased from 37^{o}C and 50^{o}C as new condition for reaction for 1 day and 2 days respectively(**Figure 42-43**). Compared with the temperature at 37^{o}C, the reaction at 50 ^{o}C might cause unexpected products as shown in the 15% SDS-PAGE. Finally, in this thesis, the unbiased Ub-SUMO2 heterodimer was successfully synthesized at 37^{o}C for 2 days (**Figure 44**).

Finally, the results of Ub-SUMO2 heterodimer was little amount in the reactive products. We assumed that the other chemistry can be perform in the total reagents. Hence,

doi:10.6342/NTU201803256

the reactive products were analyzed by MALDI-TOF-MS apparatus (**Figure 45**). The found molecular weight 10373 Da and 10722 Da implied the high amount of Ub-cysteamine existed in the reaction and SUMO2K11Dha occur addition with DTT by thiol-Michael addition. The SUMO2K11Dha was attempted to ensure DTT as a nucleophile caused the reaction. The ESI-MS spectra showed peaks at molecular weight 10705 Da, and 10719 Da. which was in agreement with calculated mass 10704 Da, and 10720 Da (**Figure 46**). Next, the power of reducing agent, TCEP, was utilized in the reaction. But the SUMO2K11Dha also went through phospha-Michael addition with TCEP. The ESI-MS spectra indicated the found molecular weight 10798 Da, and 10816 Da. which was in accord with calculated mass 10800 Da, and 10816 Da (**Figure 47**).

doi:10.6342/NTU201803256

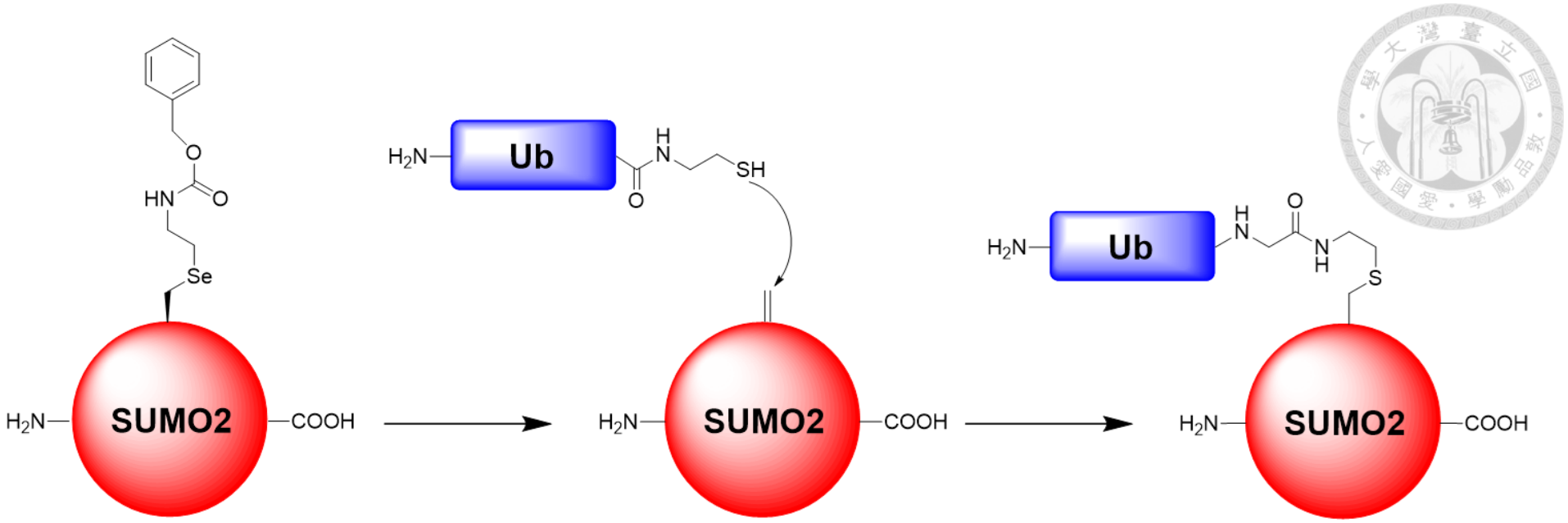
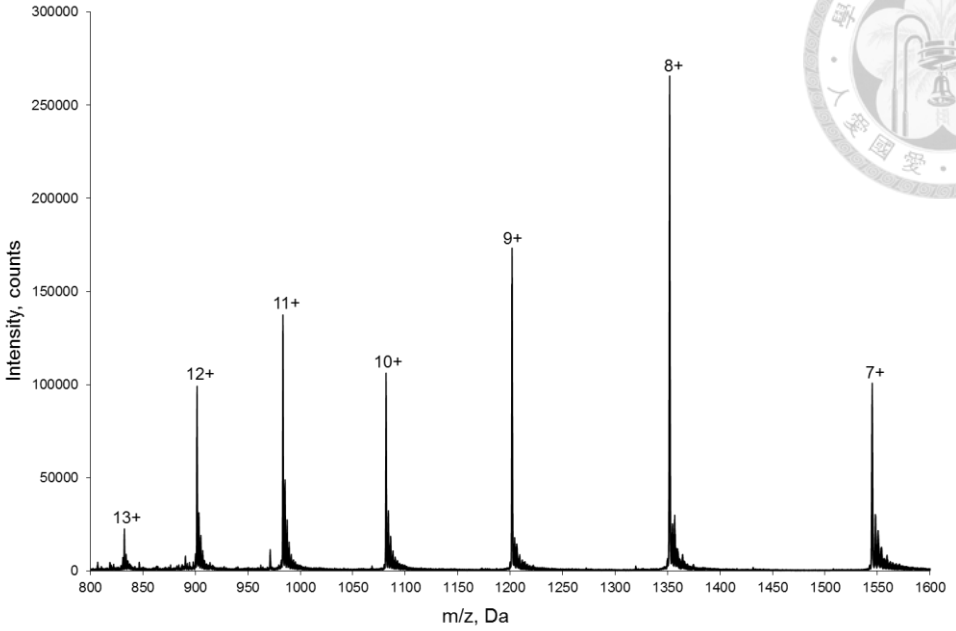


Figure 37. SUMO2-Ubiquitination through expanding genetic codes approach.
Ub-SUMO2 heterodimer synthesis using amino acylated amber suppressor tRNA *in vivo* can achieve site-specifically installation in synthesizing designed heterodimer homogenously.

doi:10.6342/NTU201803256

(A)



(B)

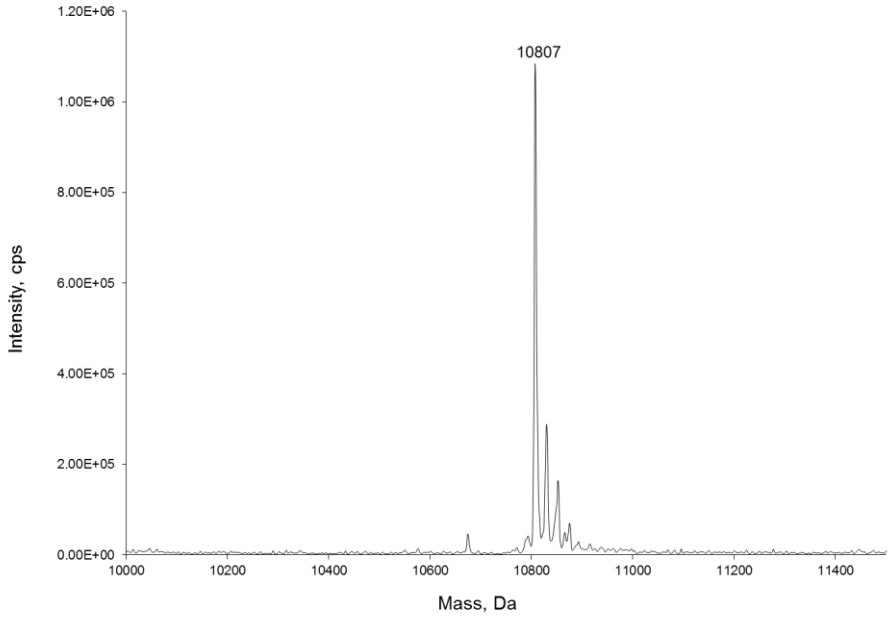
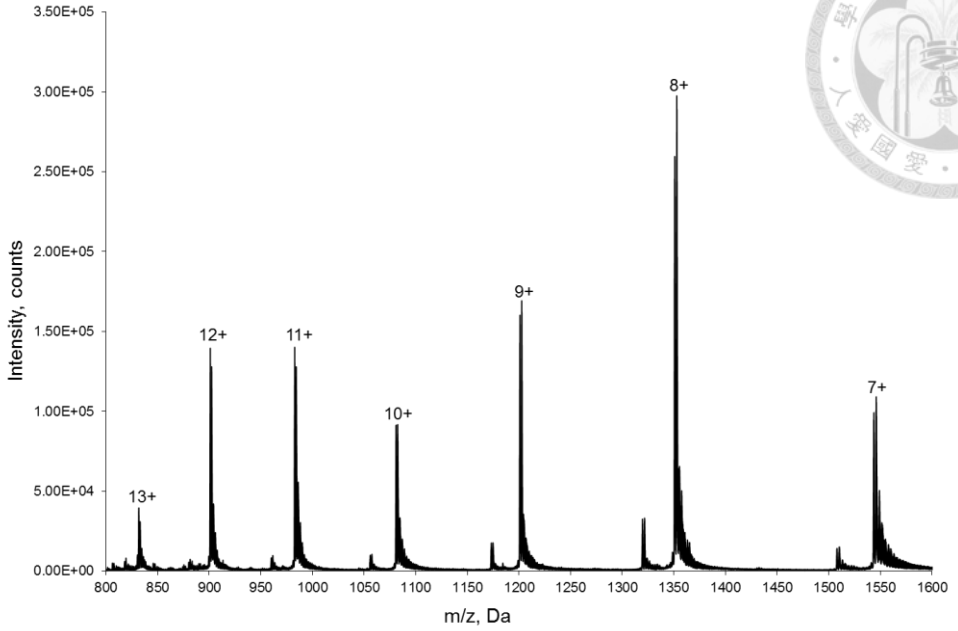


Figure 38. Molecular mass determination of the protein SUMO2K11-2 react with thiophosphate.

(A) ESI-MS spectrum of SUMO2K11-**2** react with thiophosphate (B) The deconvoluted ESI-MS spectrum of SUMO2K11-**2** react with thiophosphate**.** The calculated molecular weight is 10808 Da; the found molecular weight is 10807 Da.

doi:10.6342/NTU201803256

(A)



(B)

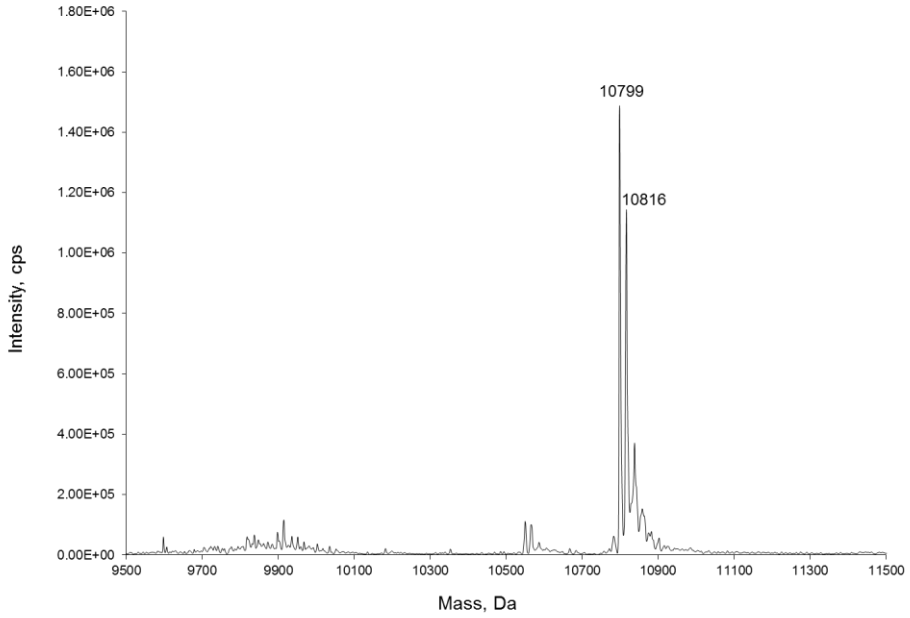


Figure 39. Molecular mass weight determination of SUMO2K11Dha react with thiophosphate.

(A) ESI-MS spectrum of SUMO2K11Dha react with thiophosphate (B) The deconvoluted ESI-MS spectrum of SUMO2K11Dha react with thiophosphate. The calculated molecular weight is 10660 Da; the found molecular weight is 10799 Da and 10816 Da.

doi:10.6342/NTU201803256

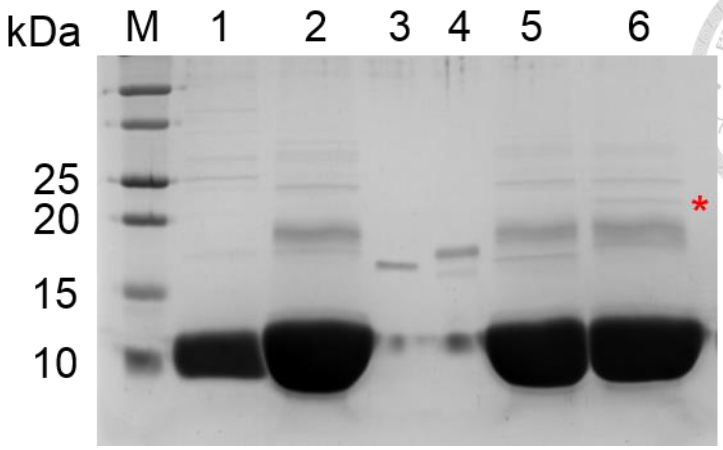


Figure 40. Analysis of Ubiqutinin-SUMO2 heterodimer synthesis.
Analysis of Ubiqutinin-SUMO2 heterodimer on 15% SDS-PAGE under denaturing conditions with Instant Blue staining. Ubiqutinin-SUMO2 heterodimer protein with the size of 20 kDa can be observed. Lane M: Protein Molecular Weight Marker; Lane 1: His6x-TEV-Ub; Lane2: His6x-TEV-Ub-cysteamine; Lane3: SUMO2K11-**2**; Lane4: SUMO2K11Dha; Lane5: 2+3 (100:1), 4^{o}C, 4 days; Lane6: 2+4 (100:1), 4^{o}C, 4 days. Red star: target protein.

doi:10.6342/NTU201803256

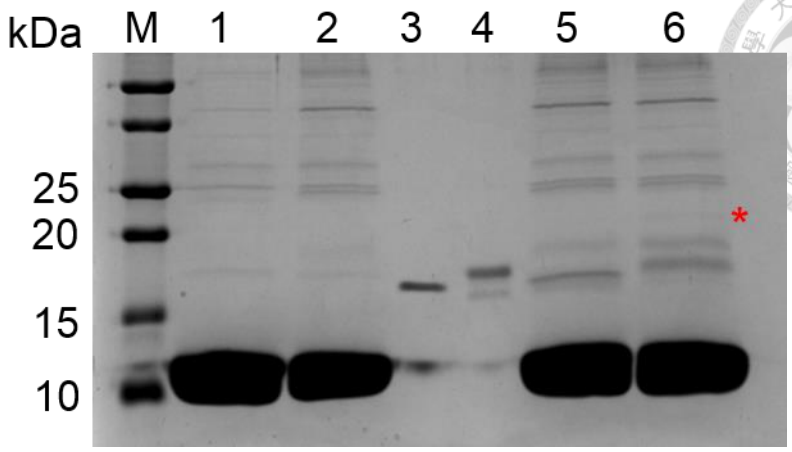


Figure 41. Analysis of Ubiqutinin-SUMO2 heterodimer synthesis.
Analysis of Ubiqutinin-SUMO2 heterodimer on 15% SDS-PAGE under denaturing conditions with Instant Blue staining. Ubiqutinin-SUMO2 heterodimer protein with the size of 20 kDa can be observed. Lane M: Protein Molecular Weight Marker; Lane 1: His6x-TEV-Ub; Lane2: His6x-TEV-Ub-cysteamine; Lane3: SUMO2K11-**2**; Lane4: SUMO2K11Dha; Lane5: 2+3 (100:1), 4^{o}C, 6 days; Lane6: 2+4 (100:1), 4^{o}C, 6 days.

doi:10.6342/NTU201803256

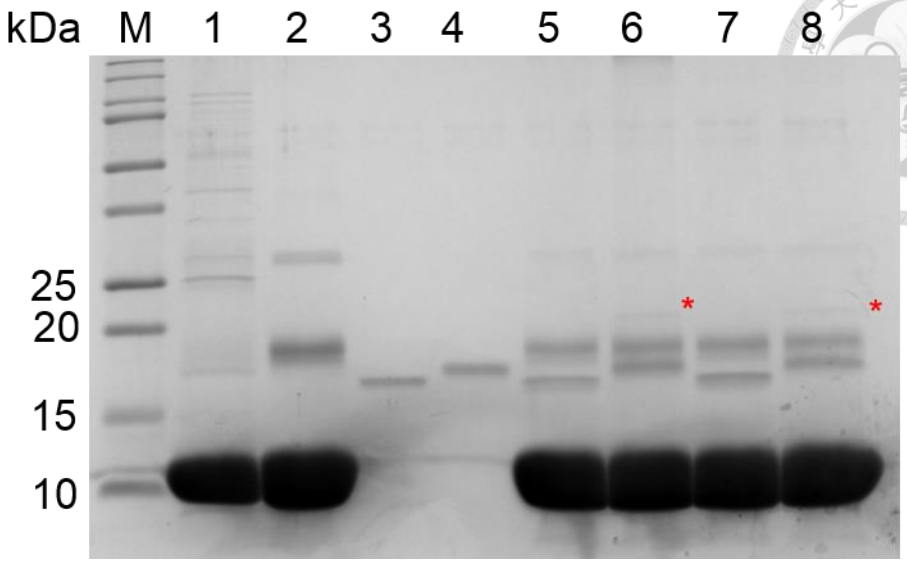


Figure 42. Analysis of Ubiqutinin-SUMO2 heterodimer synthesis.
Analysis of Ubiqutinin-SUMO2 heterodimer on 15% SDS-PAGE under denaturing conditions with Instant Blue staining. Ubiqutinin-SUMO2 heterodimer protein with the size of 20 kDa can be observed. Lane M: Protein Molecular Weight Marker; Lane 1: His6x-TEV-Ub; Lane2: His6x-TEV-Ub-cysteamine; Lane3: SUMO2K11-**2**; Lane4: SUMO2K11Dha; Lane5: 2+3 (100:1), 37°C, 1 day; Lane6: 2+4 (100:1), 37°C, 1 day; Lane7: 2+3 (100:1), 37°C, 2 days; Lane8: 2+4 (100:1), 37°C 2 days.

doi:10.6342/NTU201803256

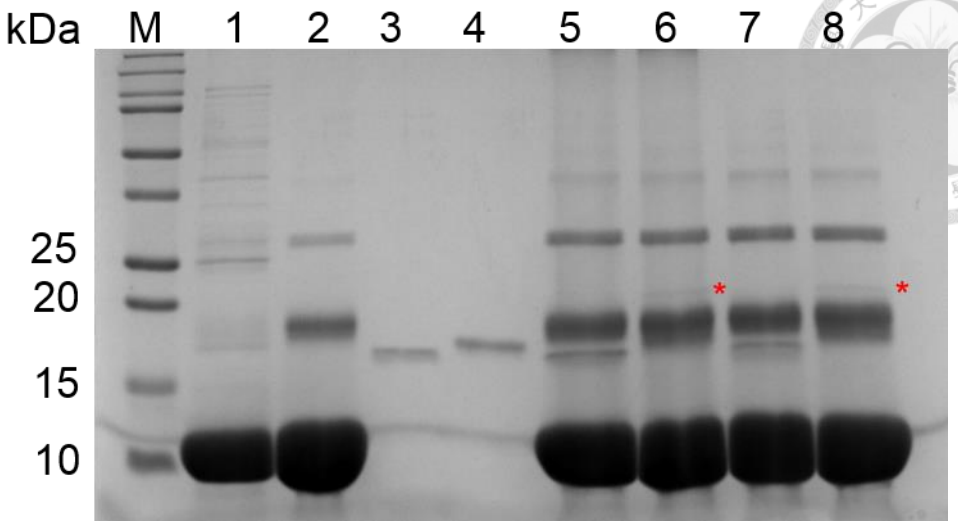


Figure 43. Analysis of Ubiqutinin-SUMO2 heterodimer synthesis.
Analysis of Ubiqutinin-SUMO2 heterodimer on 15% SDS-PAGE under denaturing conditions with Instant Blue staining. Ubiqutinin-SUMO2 heterodimer protein with the size of 20 kDa can be observed. Lane M: Protein Molecular Weight Marker; Lane 1: His6x-TEV-Ub; Lane2: His6x-TEV-Ub-cysteamine; Lane3: SUMO2K11-**2**; Lane4: SUMO2K11Dha; Lane5: 2+3 (100:1), 50^{o}C, 1 day; Lane6: 2+4 (100:1), 50^{o}C, 1 day; Lane7: 2+3 (100:1), 50^{o}C, 2 days; Lane8: 2+4 (100:1), 50^{o}C 2 days.

doi:10.6342/NTU201803256

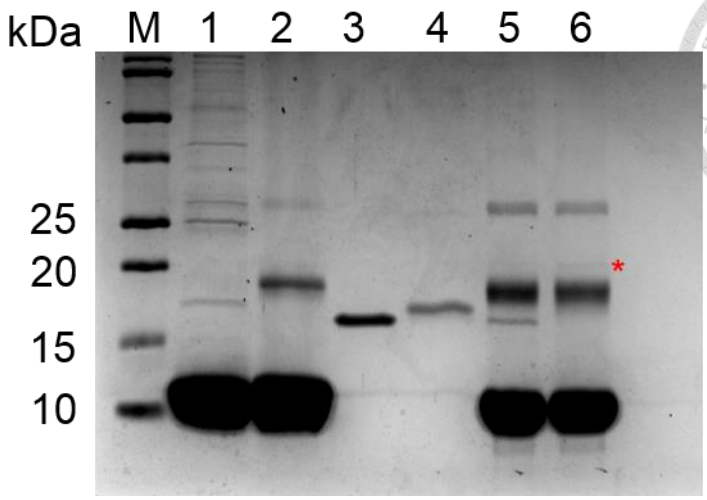


Figure 44. Analysis of Ubiqutinin-SUMO2 heterodimer synthesis.
Analysis of Ubiqutinin-SUMO2 heterodimer on 15% SDS-PAGE under denaturing conditions with Instant Blue staining. Ubiqutinin-SUMO2 heterodimer protein with the size of 20 kDa can be observed. Lane M: Protein Molecular Weight Marker; Lane 1: His6x-TEV-Ub; Lane2: His6x-TEV-Ub-cysteamine; Lane3: SUMO2K11-**2**; Lane4: SUMO2K11Dha; Lane5: 2+3 (100:1), 37^{o}C, 2 days; Lane6: 2+4 (100:1), 37^{o}C, 2 days.

doi:10.6342/NTU201803256

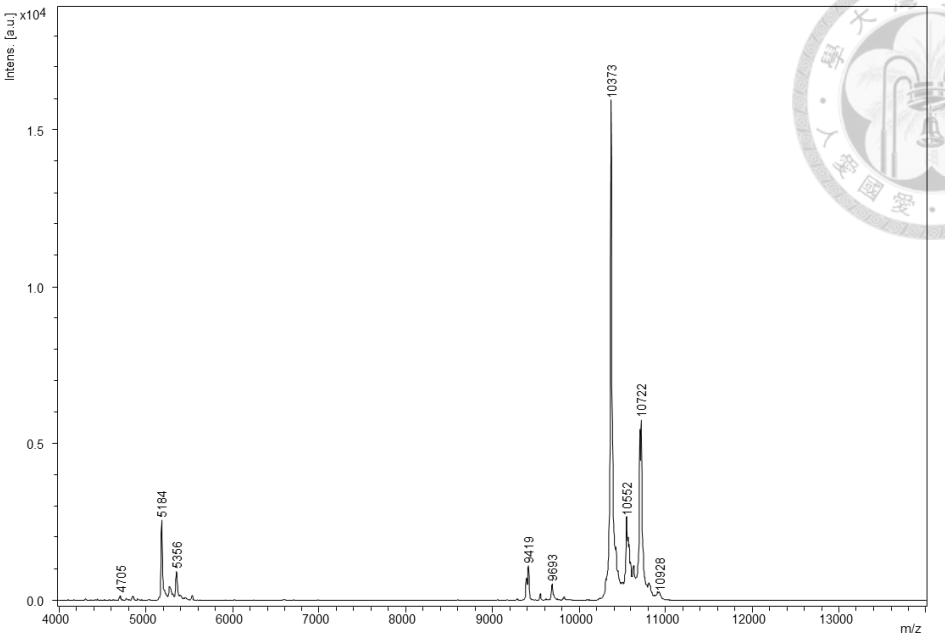
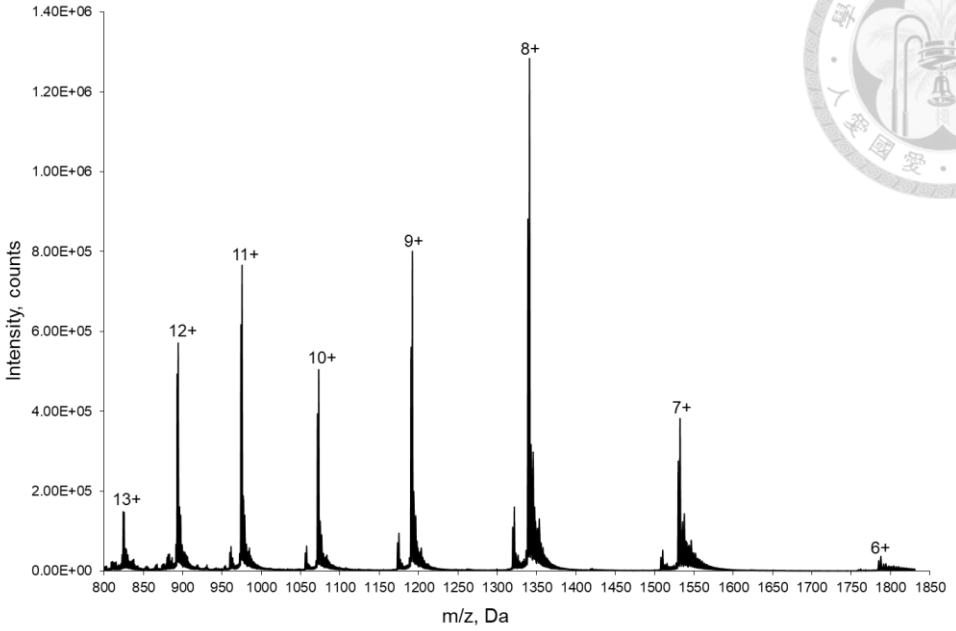


Figure 45. MALDI-TOF-MS analysis of Ub-SUMO2 heterodimer products.
MALDI-TOF-MS spectrum of Ub-SUMO2 heterodimer products. The calculated molecular weight is 10705 Da and 10719 Da; the found molecular weight is 10705 Da and 10719 Da.

doi:10.6342/NTU201803256

(A)



(B)

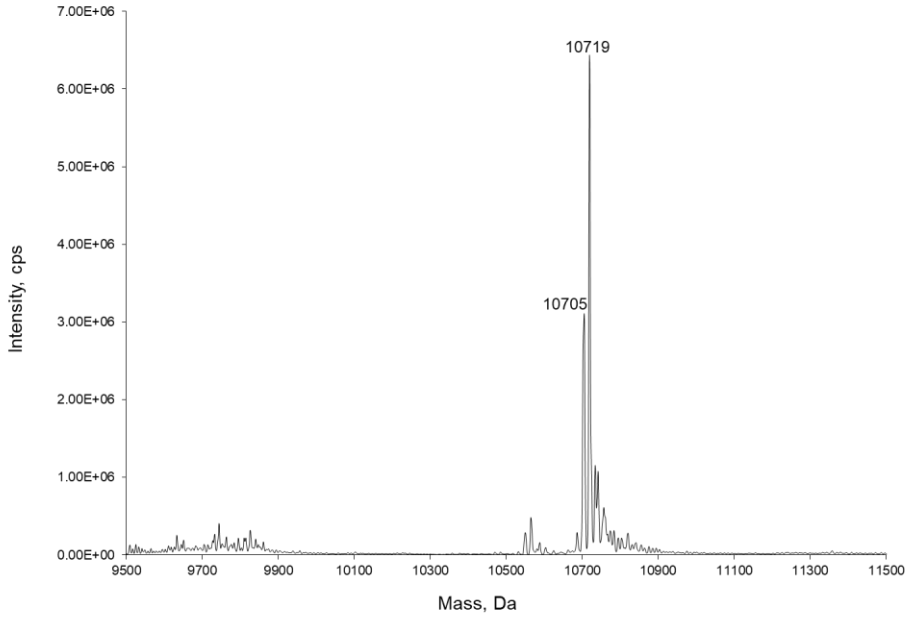
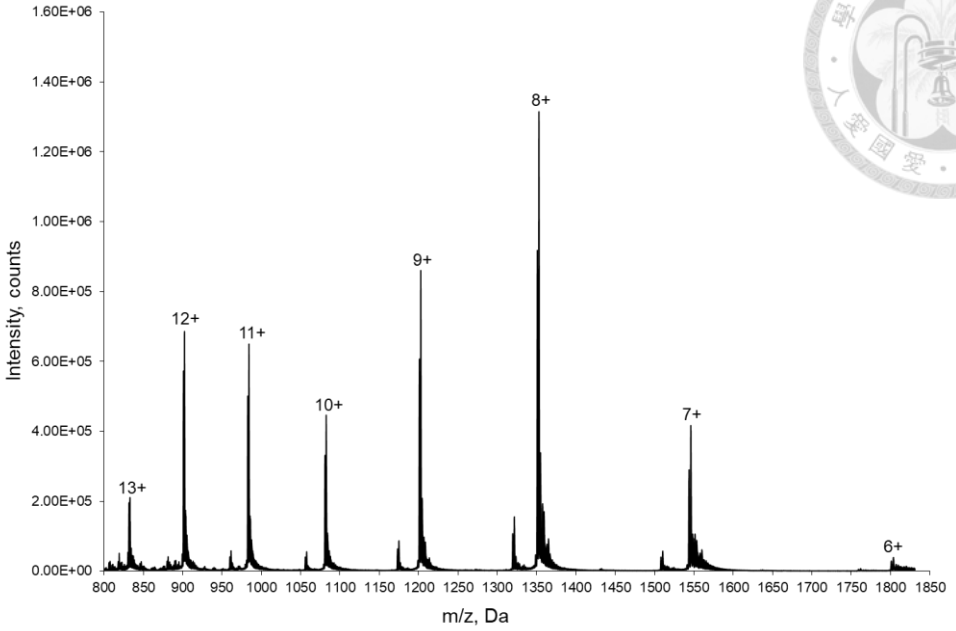


Figure 46. Molecular mass weight determination of SUMO2K11Dha react with DTT.

(A) ESI-MS spectrum of SUMO2K11Dha react with DTT (B) The deconvoluted ESI-MS spectrum of SUMO2K11Dha react with DTT. The calculated molecular weight is 10705 Da and 10719 Da; the found molecular weight is 10705 Da and 10719 Da.

doi:10.6342/NTU201803256

(A)



(B)

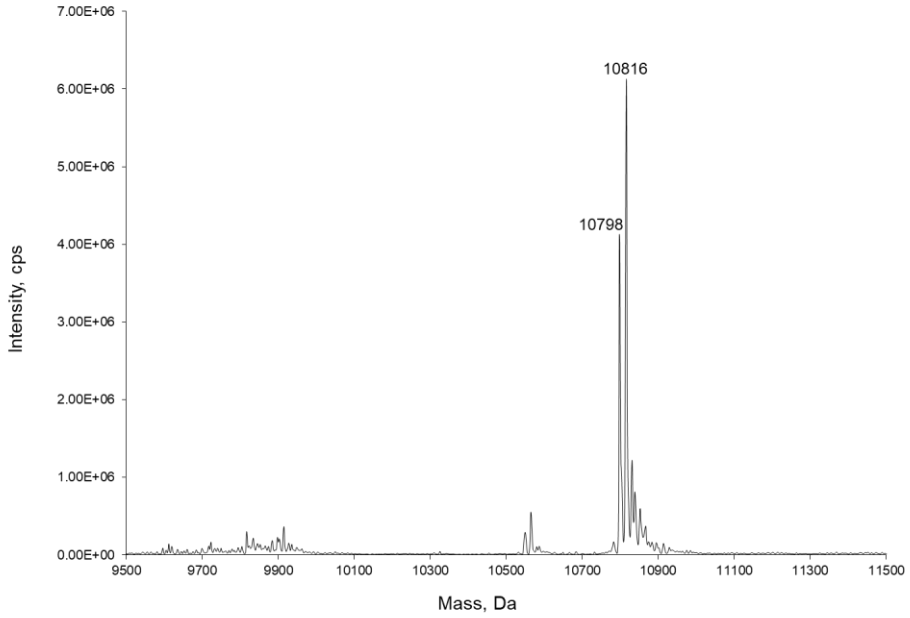


Figure 47. Molecular mass weight determination of SUMO2K11Dha react with TCEP.

(A) ESI-MS spectrum of SUMO2K11Dha react with TECP (B) The deconvoluted ESI-MS spectrum of SUMO2K11Dha react with TCEP. The calculated molecular weight is 10800 Da and 10815 Da; the found molecular weight is 10798 Da and 10816 Da.

doi:10.6342/NTU201803256

3.4 Engineering *Mm*PylRS for heterocyclic ncAAs incorporation

3.4.1 Analysis of sfGFP variants

To confirm the evolved PylRS combining with $tRNA_{CUA}^{Pyl}$ could incorporate heterocyclic ncAAs **3-7** to sfGFP variants[61] (**Table 2**), the plasmid pET-pylT-sfGFPF27TAG and pCDF-FOWRS2 were chemically co-transformed into *E. coli* BL21 (DE3) for protein expression. The sfGFP variants were produced in GMML medium with 1 mM IPTG and 1 mM **3-7** at 37℃ for 12 hrs (**Figure 48**). To ensure the suppression yield of sfGFP with **3-7**, the whole cell was diluted to the same OD_{600} value and run the 12% SDS-PAGE to analyze (**Figure 49**). Then, the five different ncAAs **3-7** with amber mutation at F27 on sfGFP protein fused with C-terminal hexahistidine were purified by Ni^{2+}-NTA affinity chromatography.

In addition, to confirm the other evolved PylRS combining with $tRNA_{CUA}^{Pyl}$ could incorporate heterocyclic ncAAs **8-9** to sfGFP variants, the plasmid pET-pylT-sfGFPF27TAG and pCDF-FOWRS6 were chemically co-transformed into *E. coli* BL21 (DE3) for protein expression. The sfGFP variants were produced in GMML medium with 1 mM IPTG and 1 mM **3-7** at 37℃ for 12 hrs. Then, the two different ncAAs **3-7** with amber mutation site at F27 on sfGFP protein fused with C-terminal hexahistidine were purified by Ni^{2+}-NTA affinity chromatography.

doi:10.6342/NTU201803256

Table 2. Mutated residues of PylRS variants in this study.

	K3	R61	H63	S193	N346	F347	C348	V401	W417	K431	D433	A441
FOWRS1					G		Q	G				
FOWRS2		K	Y	R	G		Q	G				
FOWRS3	N	K	Y	R	G		Q	G				
FOWRS4		K	Y	R	G	Y	Q	G		M	G	S
FOWRS5					G		Q	G		M	G	S
FOWRS6		K	Y	R	G		Q	G	T			

doi:10.6342/NTU201803256

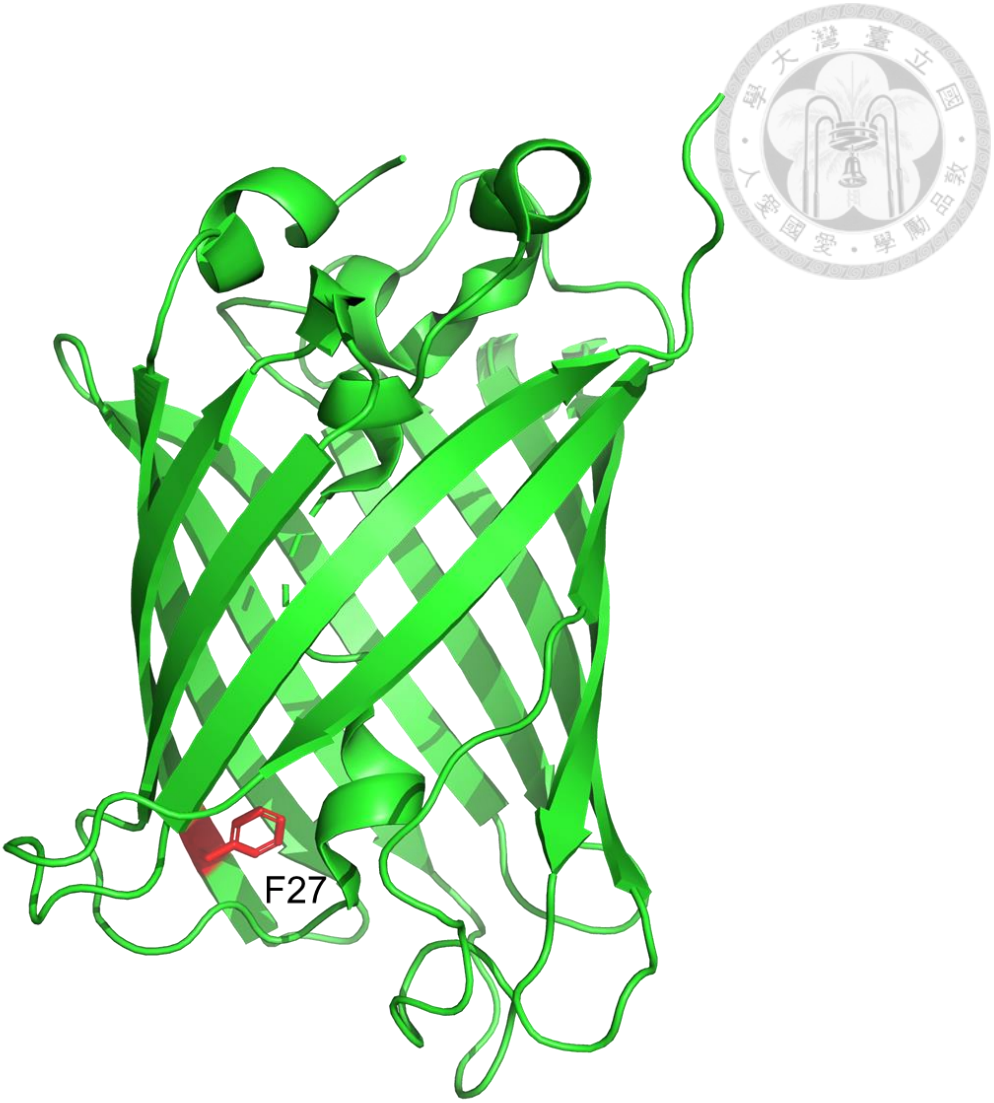


Figure 48. The structure of wild-type sfGFP.
The red labels, position F27, exhibit the residues that were mutated to ncAAs in response to amber codon. Drawn using PyMOL and the PDB code is 2B3P.

doi:10.6342/NTU201803256

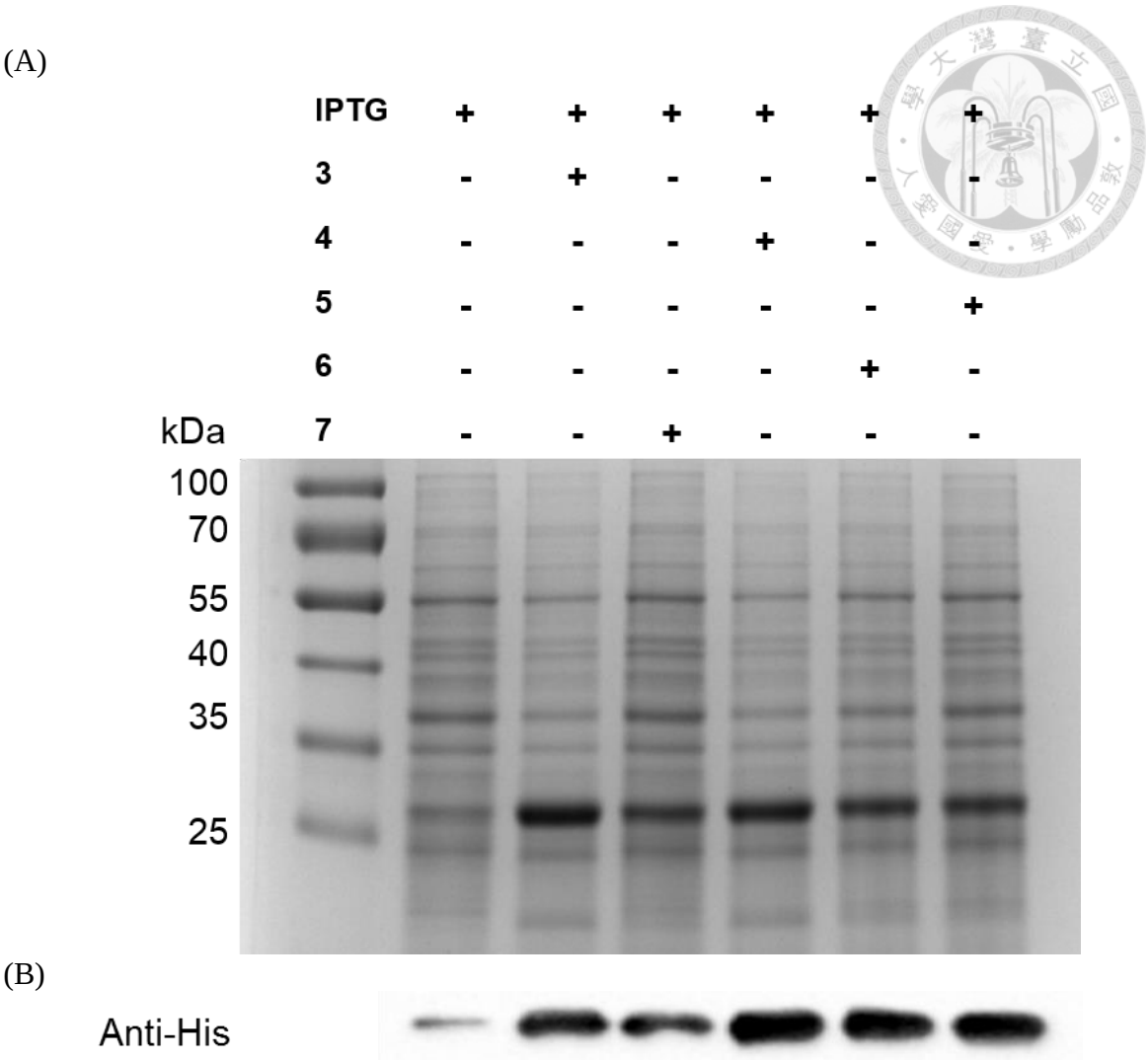


Figure 49. Amber suppression of different heterocyclic ncAAs 3-7 in sfGFP-F27TAG.

E. coli BL21(DE3) cells coding FOWRS2•$tRNA^{Pyl}_{CUA}$ 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 minimal medium supplemented with 1 mM IPTG and 1 mM ncAAs. (A) Instant Blue-Stained gel of whole cell over-expressing sfGFP-F27TAG in the absence and in the presence **3-7**. First lane was negative control without ncAA and IPTG. (B) Western blot of lysate from whole cell over-expressing sfGFP-F27TAG-**3** to **7** with an anti-His tag antibody.

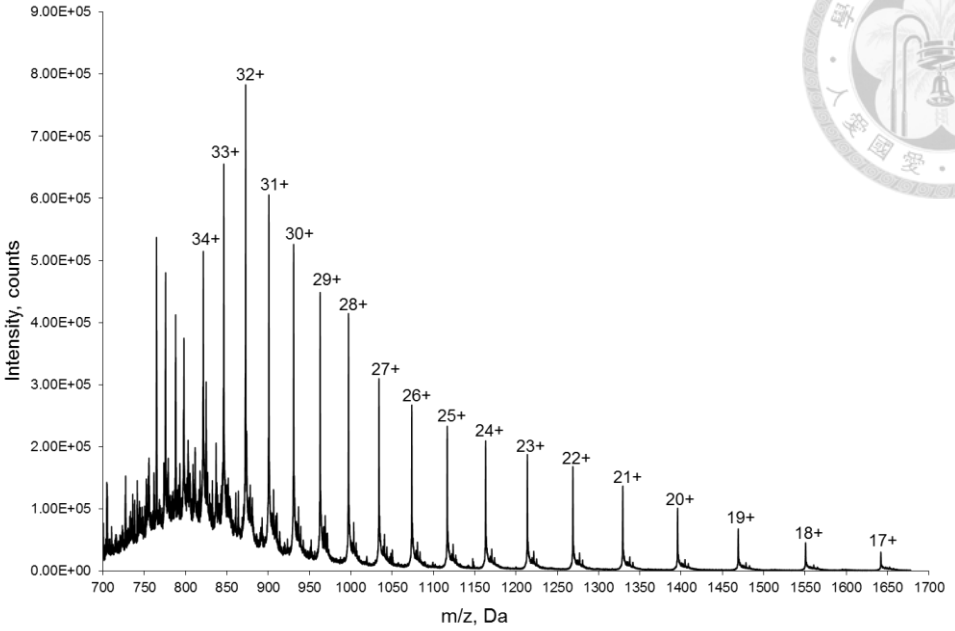
doi:10.6342/NTU201803256

3.4.2 ESI-MS analysis and MALDI-TOF-MS/MS analysis of sfGFP variants

Next, the ESI-MS was performed to demonstrate the results of sfGFPF27TAG with 3, 4, 5, 6, 7, 8, and 9 (**Figure 50-58**). The all purified proteins showed the expected molecular weight when analyzed by ESI-MS. In order to ensure whether the evolved PylRS•$tRNA^{Pyl}_{CUA}$ pair could incorporate ncAA into protein at F27. The purified proteins were digested by trypsin and analyzed by MALDI-TOF-MS/MS. As our expectation, the fragments containing **3** and **6** at F27 were observed, but the fragment of other ncAAs were not observed (**Figure 59-60**). Thus, other enzymatic digested condition will be tested to search the fragment containing 4, 5, 7, 8, and 9 at F27.

doi:10.6342/NTU201803256

(A)



(B)

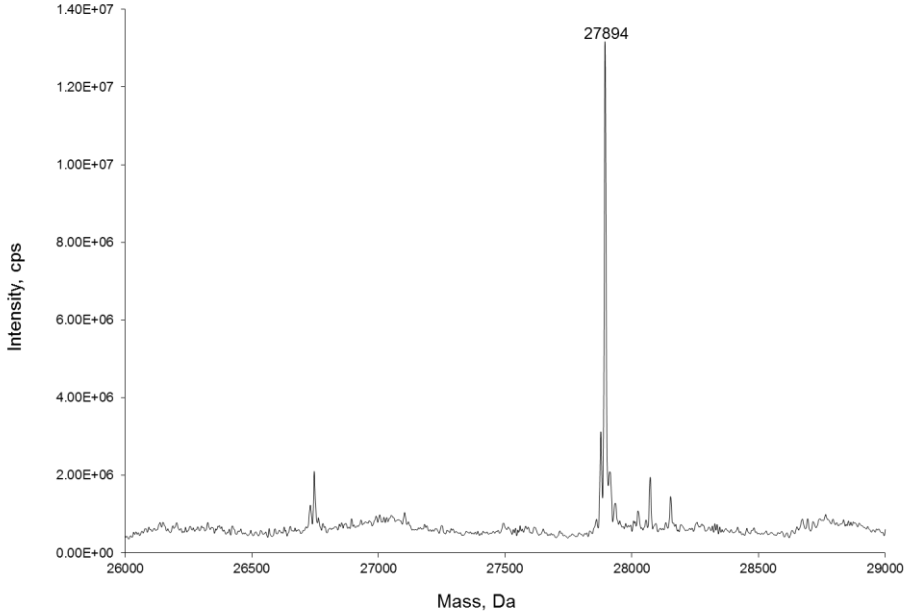
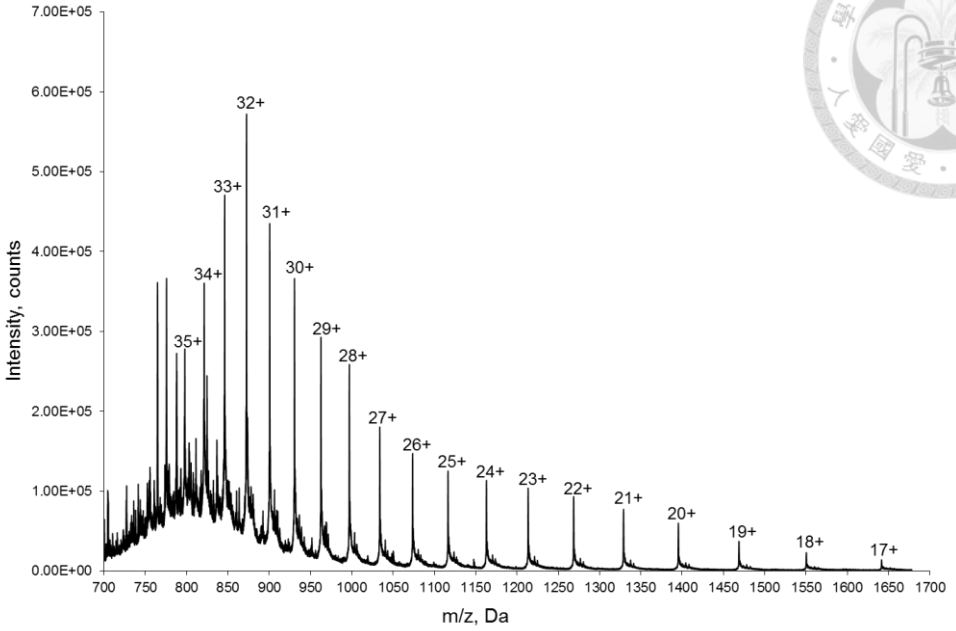


Figure 50. Molecular mass weight determination of sfGFPF27-3.

(A) ESI-MS spectrum of sfGFPF27-**3** (B) The deconvoluted ESI-MS spectrum of sfGFPF27-**3.** The calculated molecular weight is 27895 Da; the found molecular weight is 27894 Da. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **3** in *E. coli* BL21 (DE3) cells in M9 medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

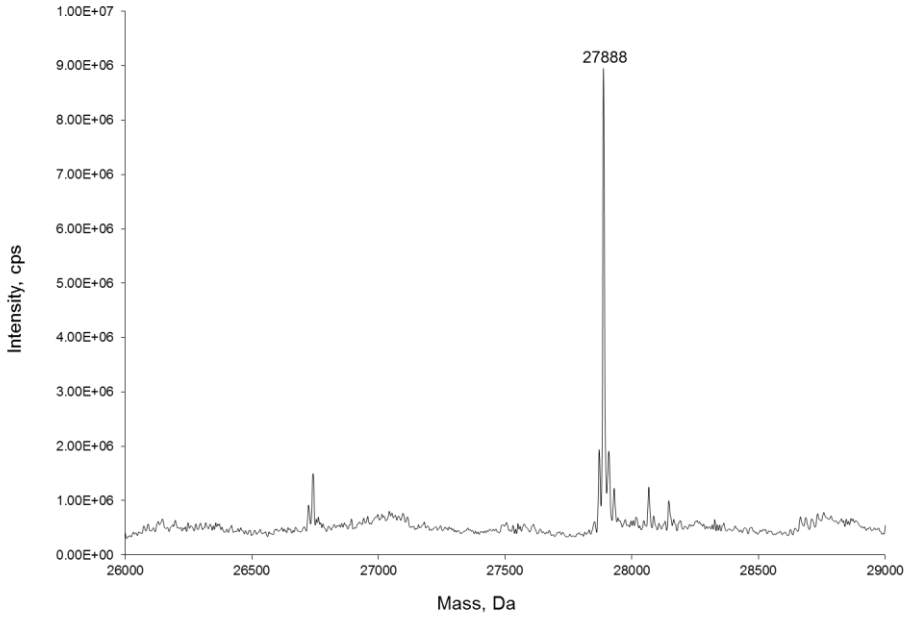
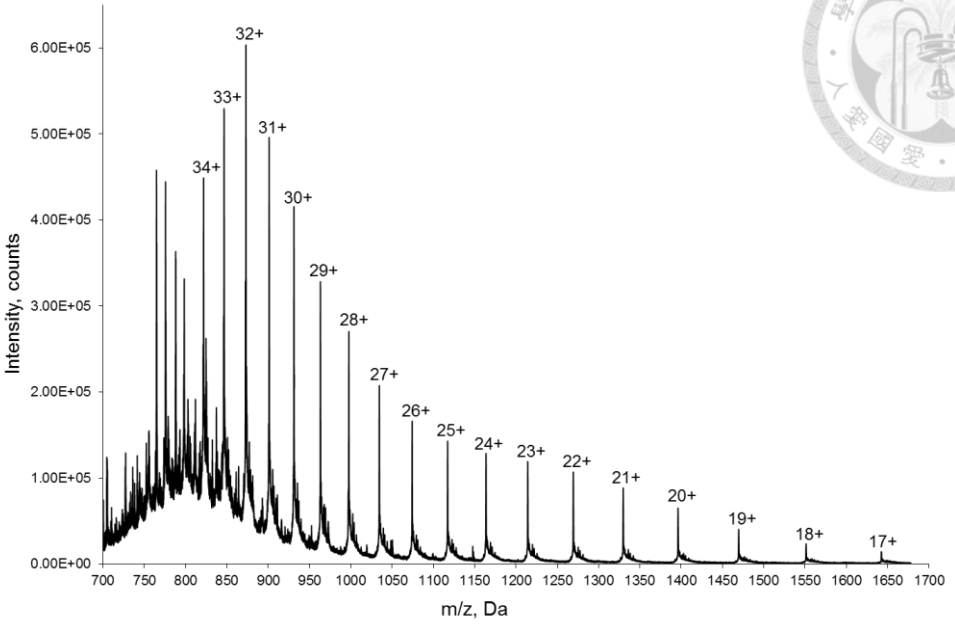


Figure 51. Molecular mass weight determination of sfGFPF27-4.

(A) ESI-MS spectrum of sfGFPF27-**4** (B) The deconvoluted ESI-MS spectrum of sfGFPF27-**4.** The calculated molecular weight is 27889Da; the found molecular weight is 27888 Da. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **4** in *E. coli* BL21 (DE3) cells in M9 medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

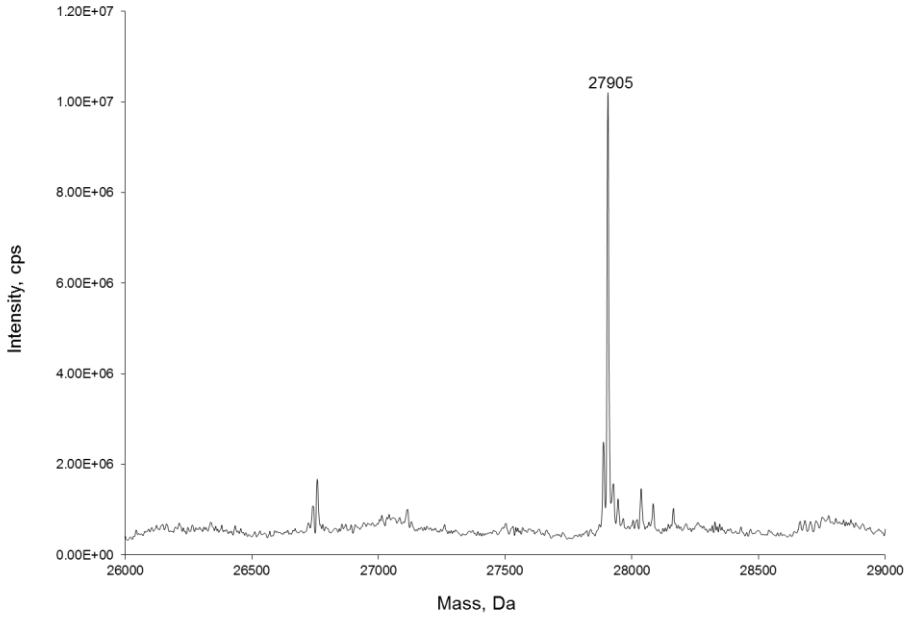
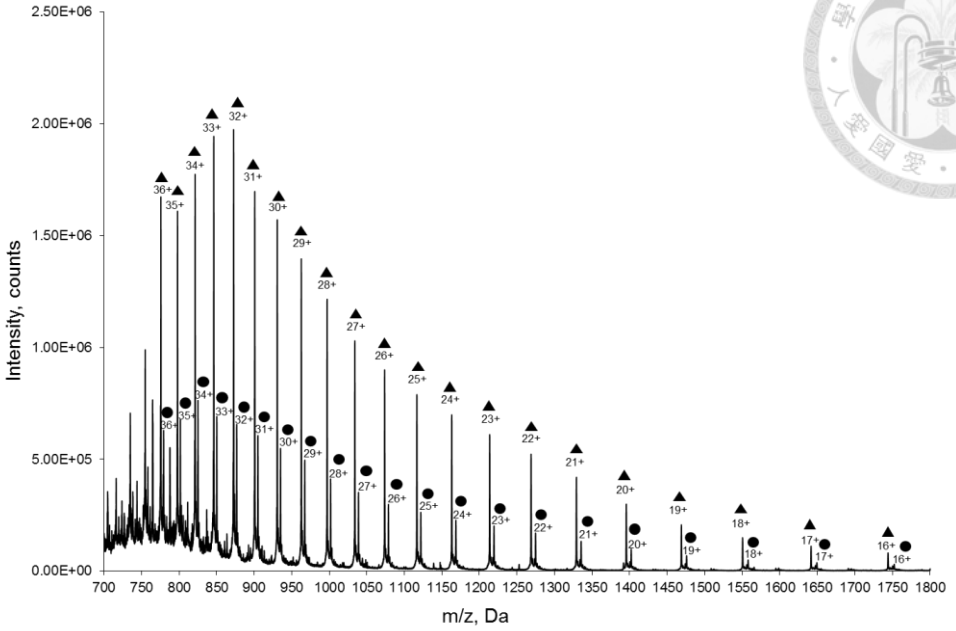


Figure 52. Molecular mass weight determination of sfGFPF27-5.

(A) ESI-MS spectrum of sfGFPF27-**5** (B) The deconvoluted ESI-MS spectrum of sfGFPF27-**5.** The calculated molecular weight is 27906 Da; the found molecular weight is 27905 Da. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **5** in *E. coli* BL21 (DE3) cells in M9 medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

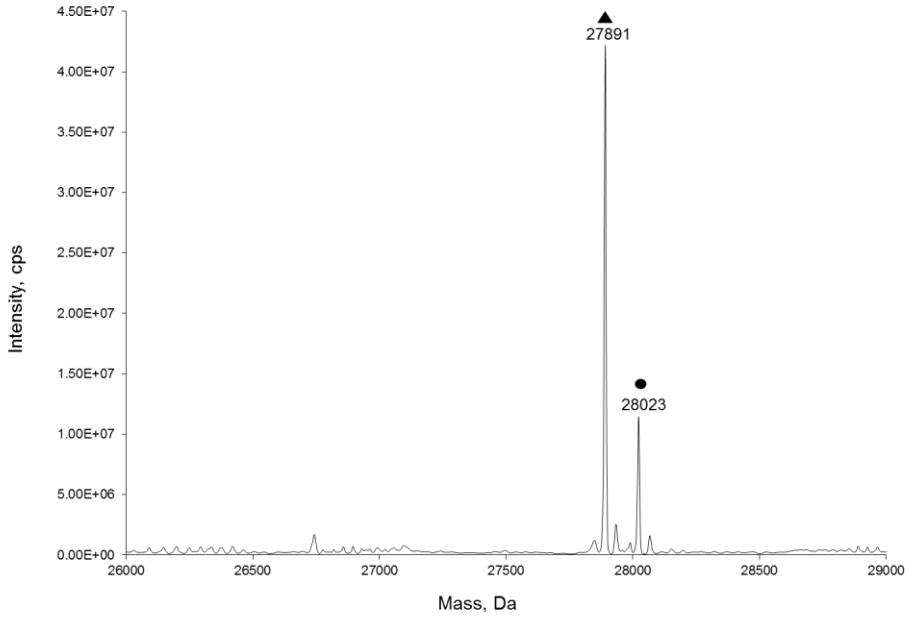
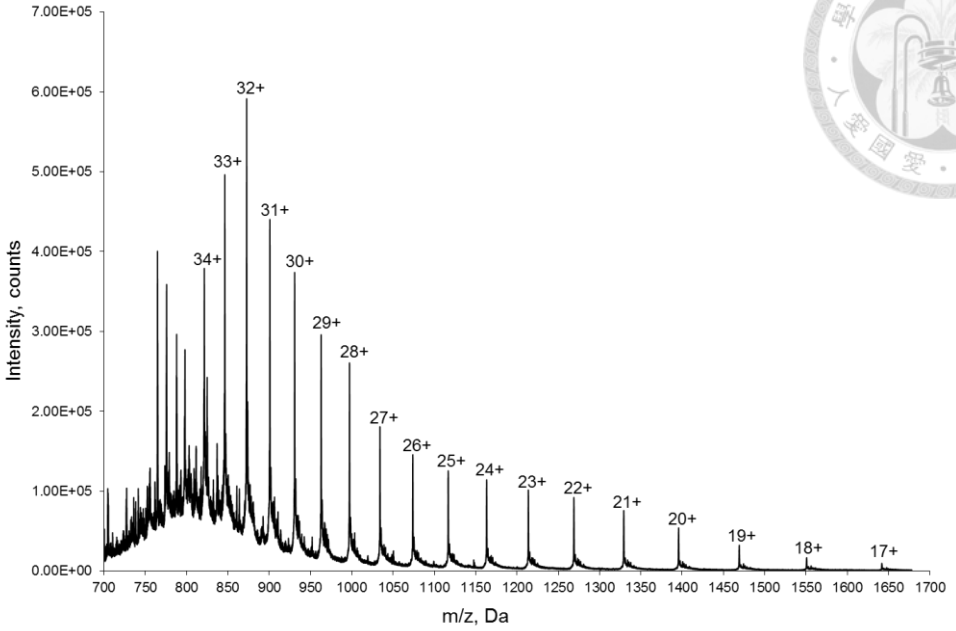


Figure 53. Molecular mass weight determination of sfGFPF27-6.

(A) ESI-MS spectrum of sfGFPF27-**6** (B) The deconvoluted ESI-MS spectrum of sfGFPF27-**6.** The calculated molecular weight is 27892Da; the found molecular weight is 27891Da. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **6** in *E. coli* BL21 (DE3) cells in M9 medium at 37 ^{o}C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

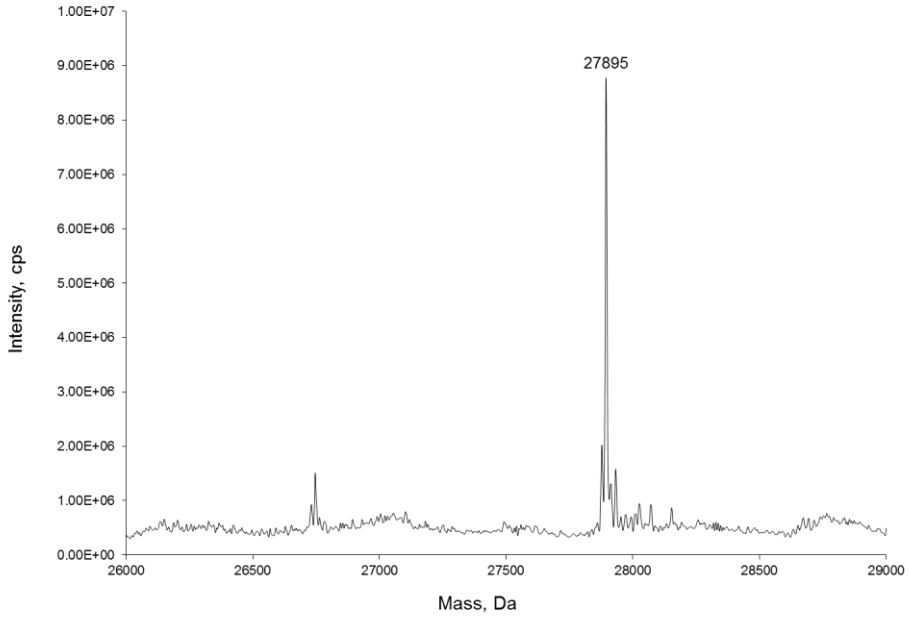
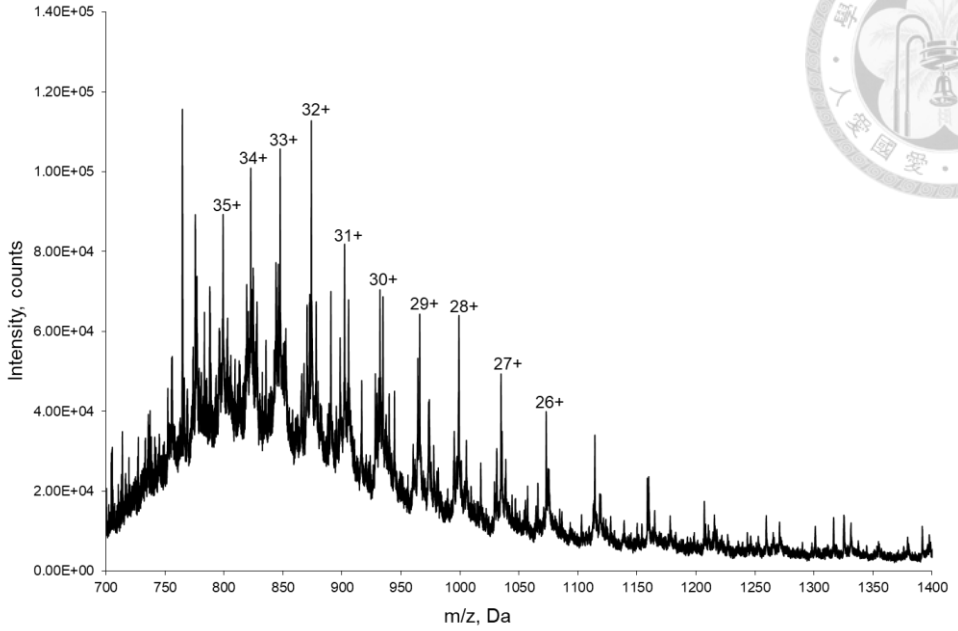


Figure 54. Molecular mass weight determination of sfGFPF27-7.

(A) ESI-MS spectrum of sfGFPF27-**7** (B) The deconvoluted ESI-MS spectrum of sfGFPF27-**7.** The calculated molecular weight is 27896 Da; the found molecular weight is 27895 Da. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **7** in *E. coli* BL21 (DE3) cells in M9 medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

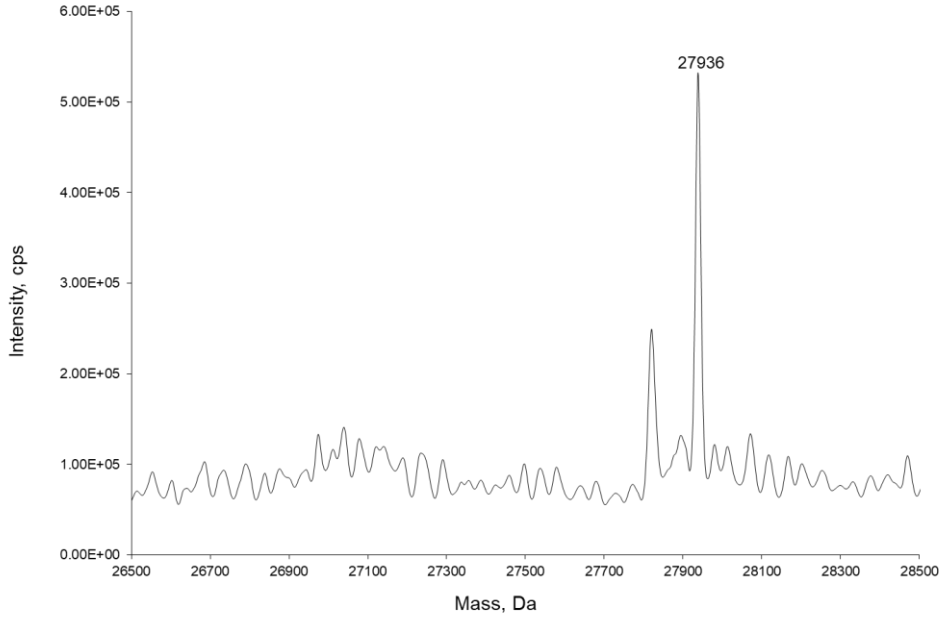
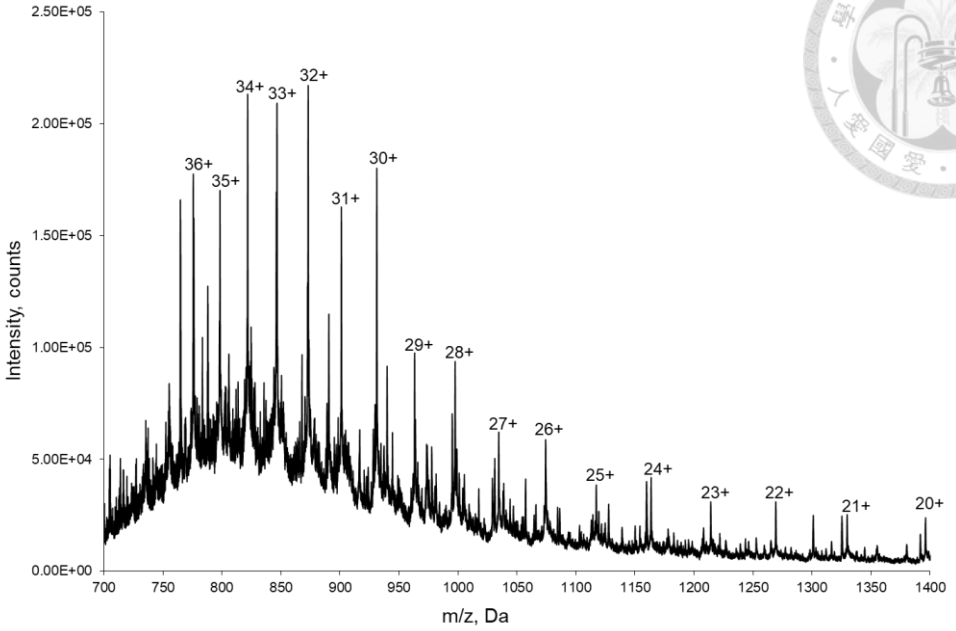


Figure 55. Molecular mass weight determination of sfGFPF27-8.

(A) ESI-MS spectrum of sfGFPF27-**8** (B) The deconvoluted ESI-MS spectrum of sfGFPF27-**8.** The calculated molecular weight is 27937 Da; the found molecular weight is 27936 Da. The proteins were produced by FOWRS6•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **8** in *E. coli* BL21 (DE3) cells in M9 medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

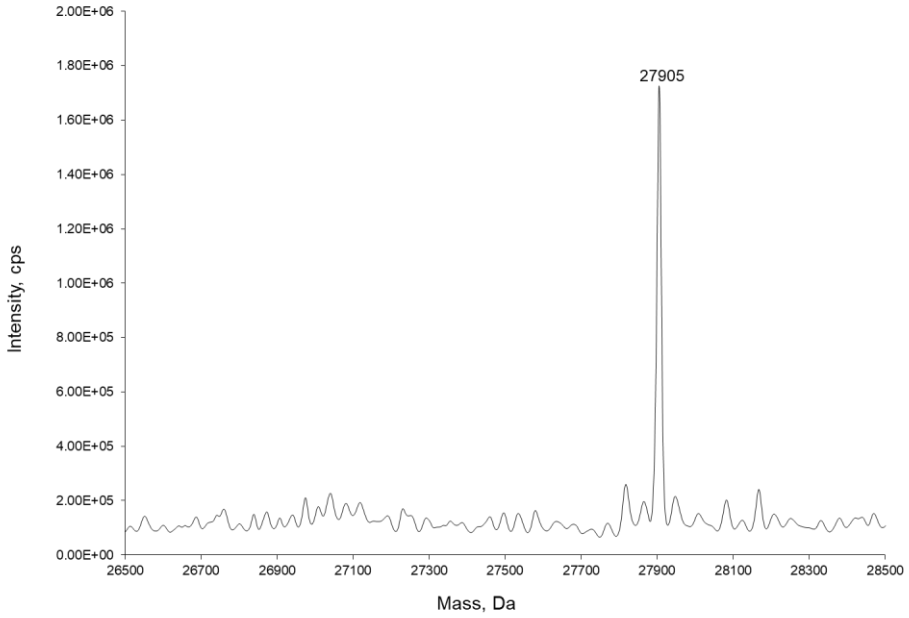


Figure 56. Molecular mass weight determination of sfGFPF27-9.

(A) ESI-MS spectrum of sfGFPF27-**9** (B) The deconvoluted ESI-MS spectrum of sfGFPF27-**9.** The calculated molecular weight is 27906 Da; the found molecular weight is 27905 Da. The proteins were produced by FOWRS6•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **9** in *E. coli* BL21 (DE3) cells in M9 medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

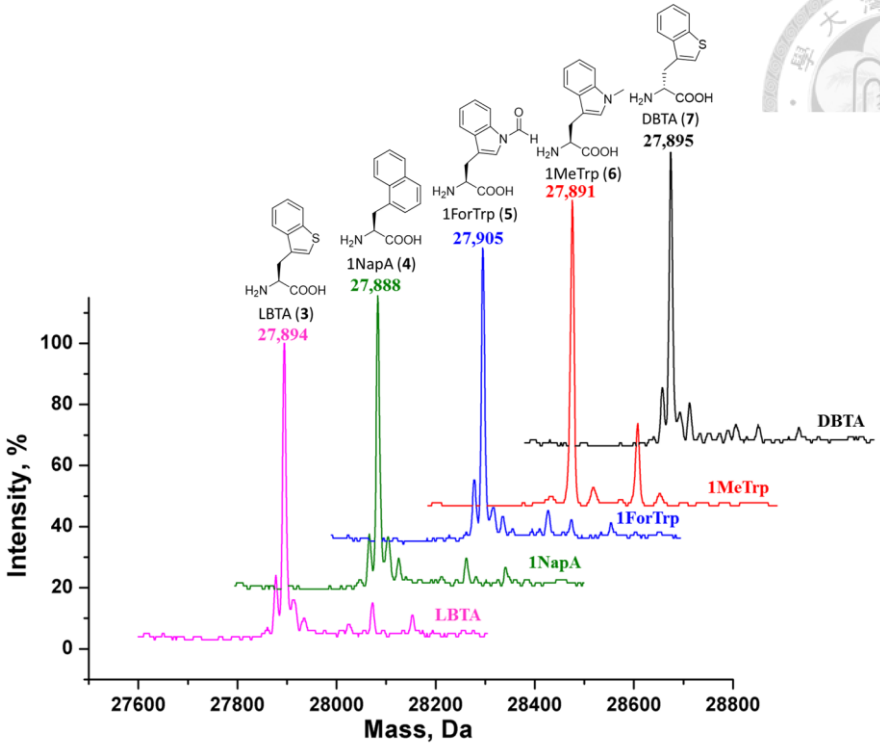


Figure 57. Molecular mass weight determination of sfGFPF27-3, 4, 5, 6, and 7. The deconvoluted ESI-MS spectrum of sfGFPF27-**3, 4, 5, 6, and 7**. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **3, 4, 5, 6, and 7** in *E. coli* BL21 (DE3) cells in M9 medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

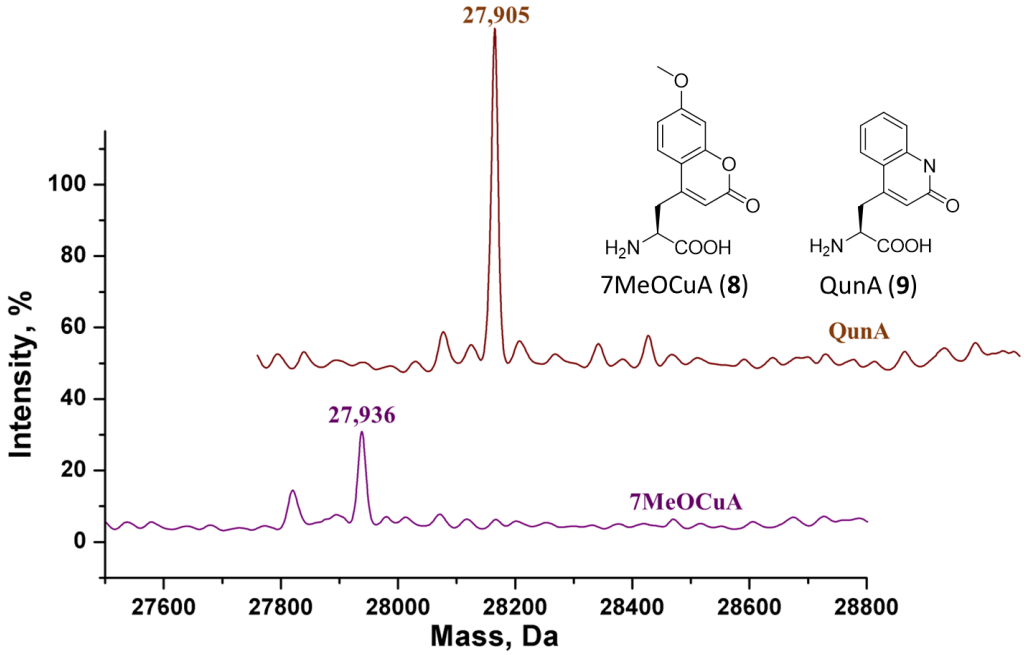


Figure 58. Molecular mass weight determination of sfGFPF27-8, and 9.
The deconvoluted ESI-MS spectrum of sfGFPF27-**8** and **9**. The proteins were produced by FOWRS6•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **8** and **9** in *E. coli* BL21 (DE3) cells in LB medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

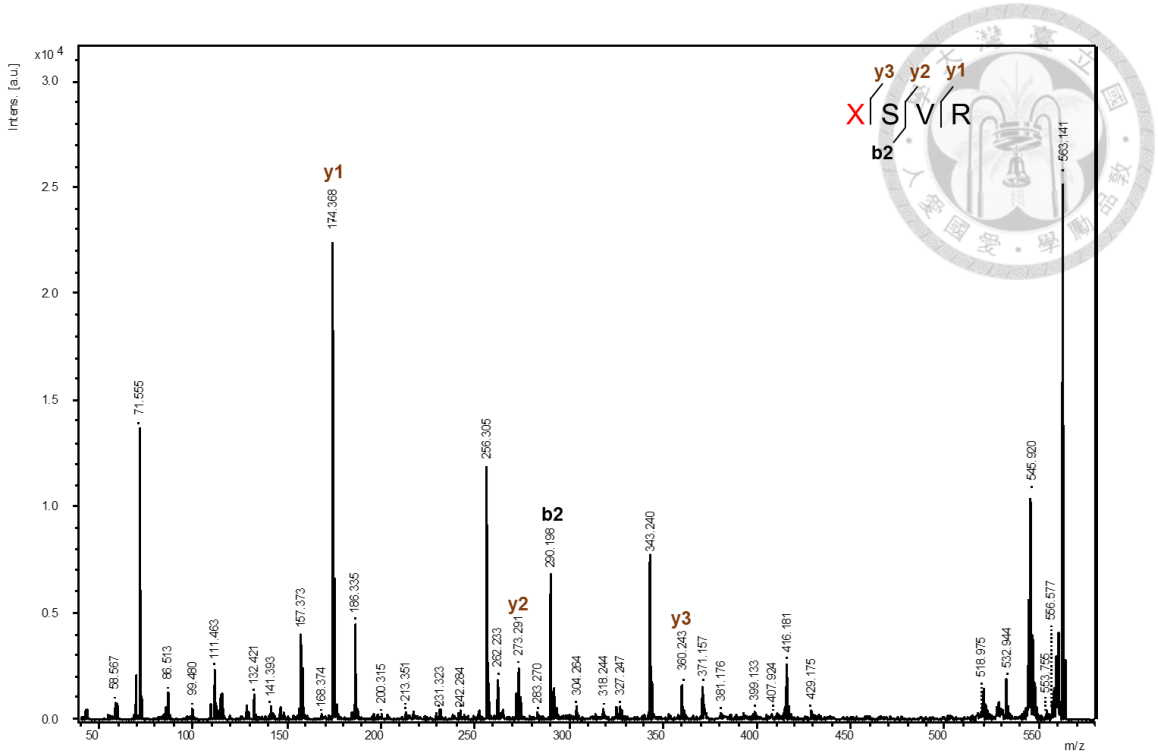


Figure 59. MALDI-TOF-MS/MS analysis of sfGFPF27-3.

Fragment $X^{27}SVR^{30}$, X represents ncAA **3**. The full-length sfGFPF27-3 protein was produced by FOWRS2・$tRNA^{Pyl}_{CUA}$ pair in the presence of 1 mM ncAA **3**. The protein was in-gel digested by trypsin.

doi:10.6342/NTU201803256

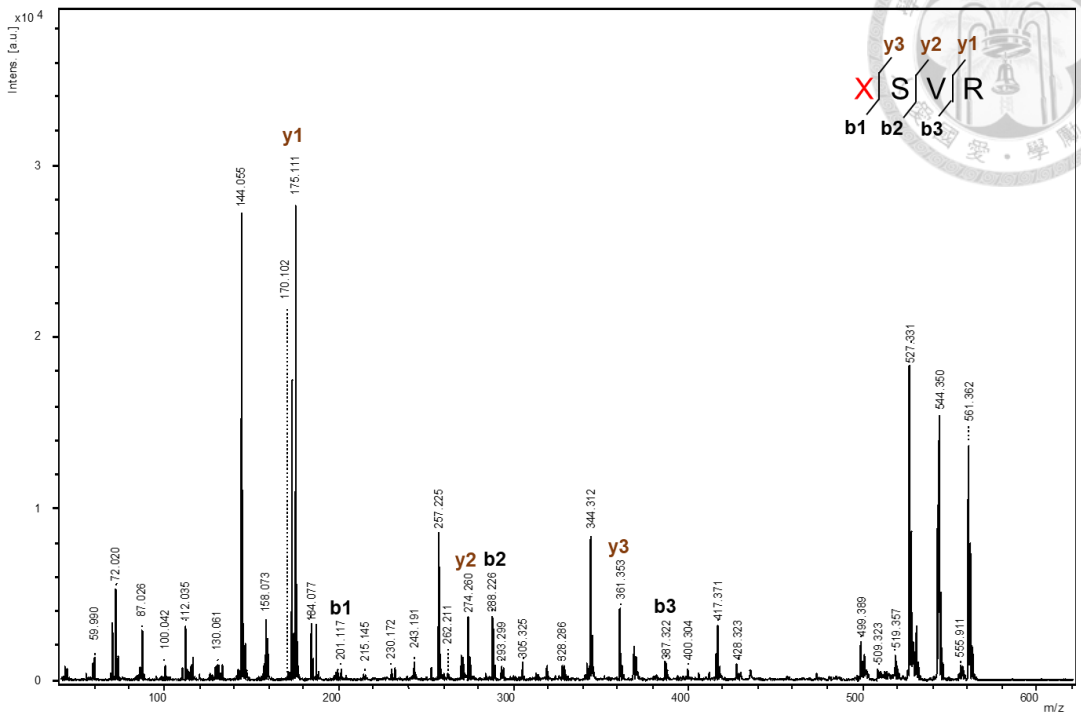


Figure 60. MALDI-TOF-MS/MS analysis of sfGFPF27-6.

Fragment $X^{27}SVR^{30}$, X represents ncAA **6**. The full-length sfGFPF27-6 protein was produced by FOWRS2 • $tRNA^{Pyl}_{CUA}$ pair in the presence of 1 mM ncAA **6**. The protein was in-gel digested by trypsin.

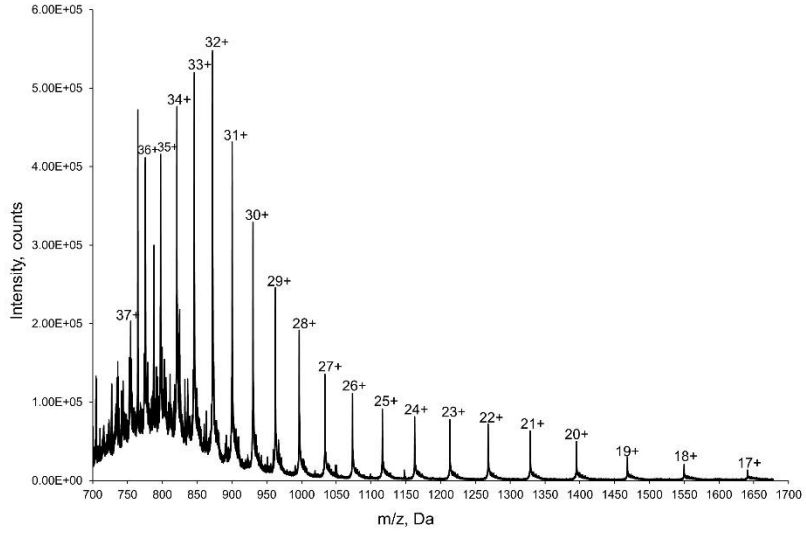
doi:10.6342/NTU201803256

3.4.3 Incorporating heterocyclic ncAAs into chromophore of sfGFP

To investigate the photophysical properties of sfGFP incorporating heterocyclic ncAA was focused and the sfGFP was mutated with amber codon at Y66 position in chromophore,which composed of T65, Y66, and G67.[62]The gene of sfGFP-Y66TAG was inserted into pET-pylT plasmid and co-transformed with pCDF-FOWRS2 in *E. coli* BL21(DE3). The cells were cultured with 1 mM **3**-**7** and 1 mM IPTG in minimal medium at 37 ℃ for 12 hrs. The over-expressed protein, sfGFP-Y66TAG with **3**-**7**, was purified and analyzed by ESI-MS. The MS spectrum confirmed that the found mass was same with the expected mass (**Figure 61**-**66**).

doi:10.6342/NTU201803256

(A)



(B)

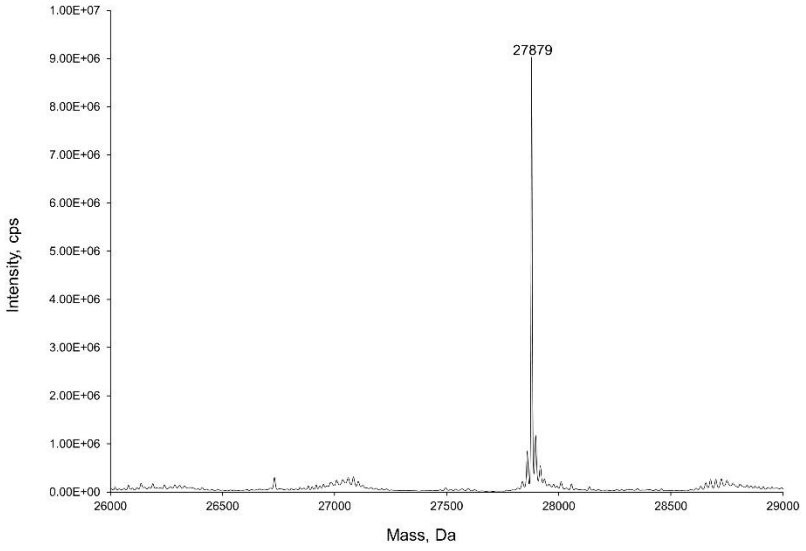
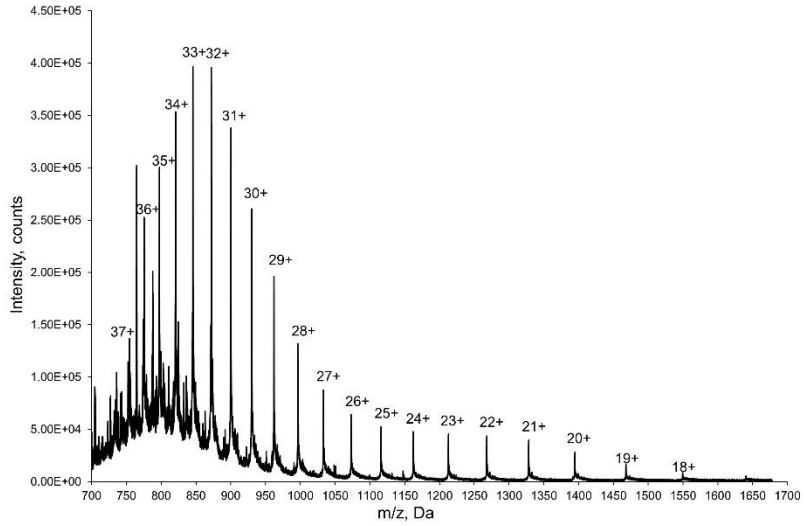


Figure 61. Molecular mass weight determination of sfGFPY66-3.

(A) ESI-MS spectrum of sfGFPY66-**3** (B) The deconvoluted ESI-MS spectrum of sfGFPY66-**3.** The calculated molecular weight is 27879 Da; the found molecular weight is 27879 Da. The proteins were produced by FOWRS2•tRNA$^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **3** in *E. coli* BL21 (DE3) cells in LB medium at 37 ºC for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

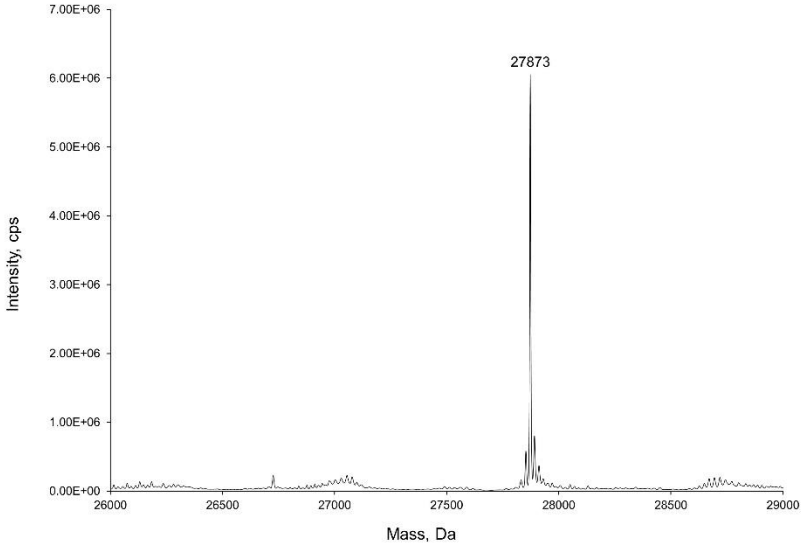
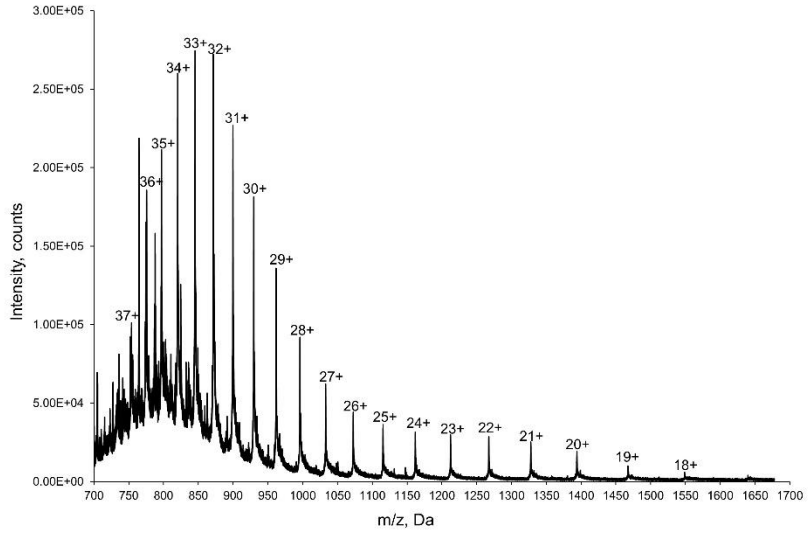


Figure 62. Molecular mass weight determination of sfGFP Y66-4.
(A) ESI-MS spectrum of sfGFPY66-**4** (B) The deconvoluted ESI-MS spectrum of sfGFPY66-**4.** The calculated molecular weight is 27872Da; the found molecular weight is 27873 Da. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **4** in *E. coli* BL21 (DE3) cells in LB medium at 37 ºC for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

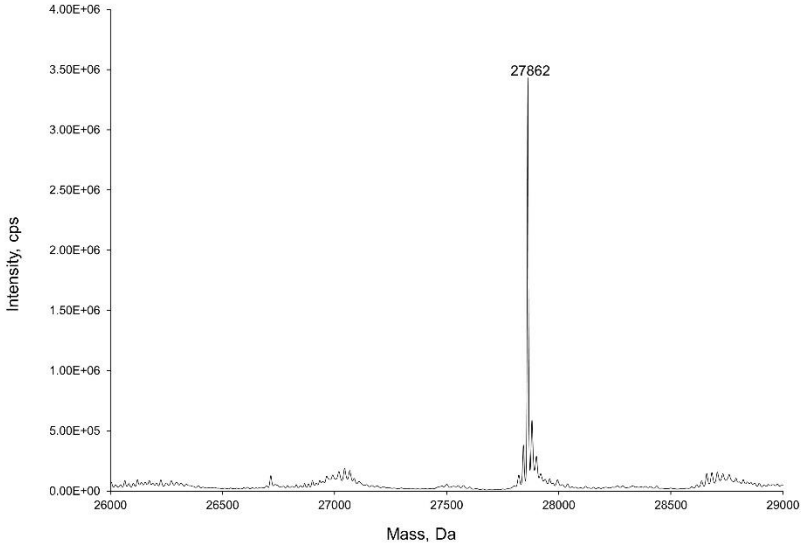
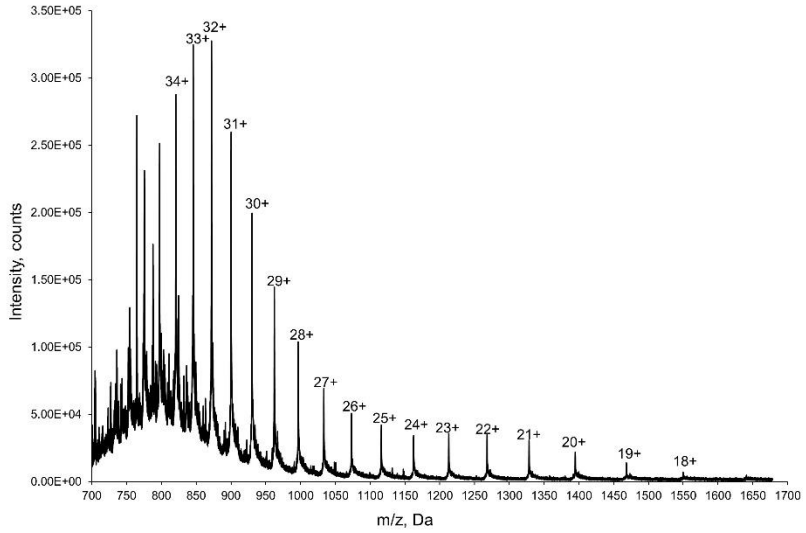


Figure 63. Molecular mass weight determination of sfGFPY66-5.

(A) ESI-MS spectrum of sfGFPY66-**5** (B) The deconvoluted ESI-MS spectrum of sfGFPY66-**5.** The calculated molecular weight is 27889 Da; the found molecular weight is 27862 Da. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **5** in *E. coli* BL21 (DE3) cells in LB medium at 37 ºC for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

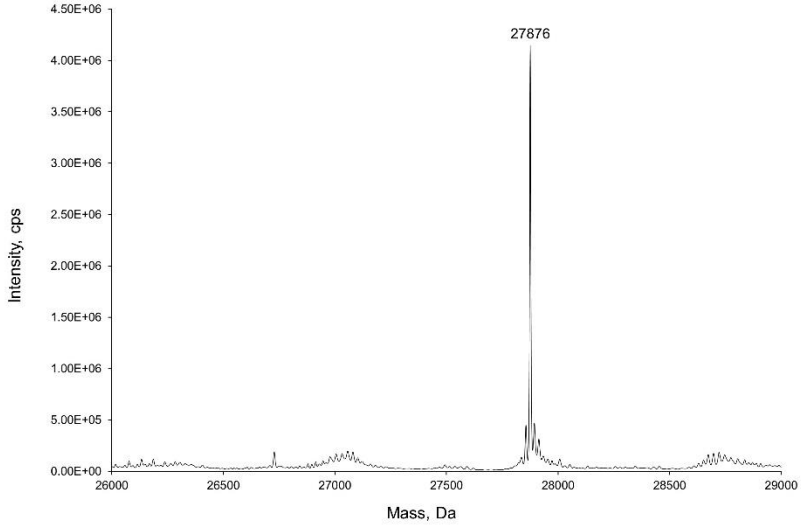
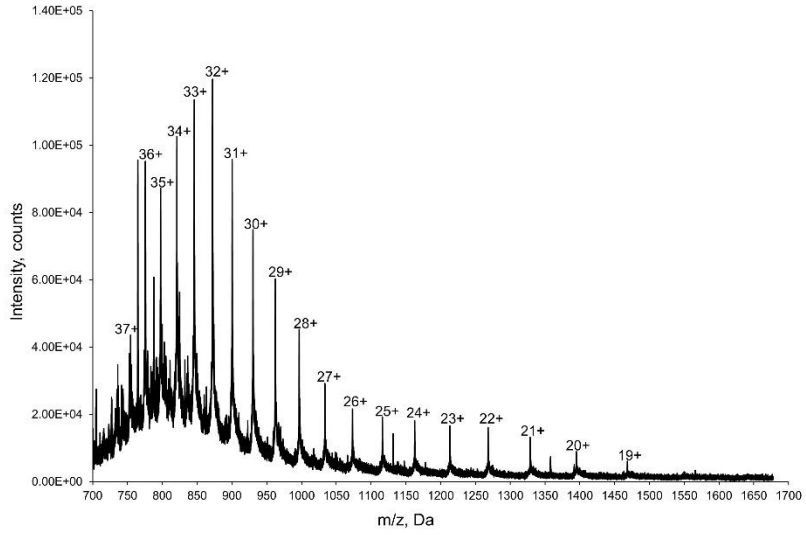


Figure 64. Molecular mass weight determination of sfGFPY66-6.
(A) ESI-MS spectrum of sfGFPY66-**6** (B) The deconvoluted ESI-MS spectrum of sfGFPY66-**6.** The calculated molecular weight is 27875Da; the found molecular weight is 27876Da. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **6** in *E. coli* BL21 (DE3) cells in LB medium at 37 ºC for 12 hours.

doi:10.6342/NTU201803256

(A)



(B)

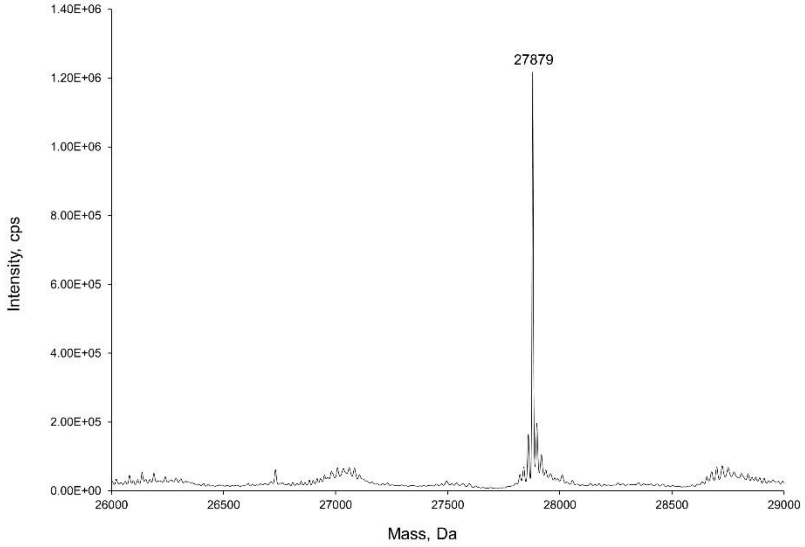


Figure 65. Molecular mass weight determination of sfGFPY66-7.

(A) ESI-MS spectrum of sfGFPY66-**7** (B) The deconvoluted ESI-MS spectrum of sfGFPY66-**7.** The calculated molecular weight is 27896 Da; the found molecular weight is 27895 Da. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **7** in *E. coli* BL21 (DE3) cells in LB medium at 37 ºC for 12 hours.

doi:10.6342/NTU201803256

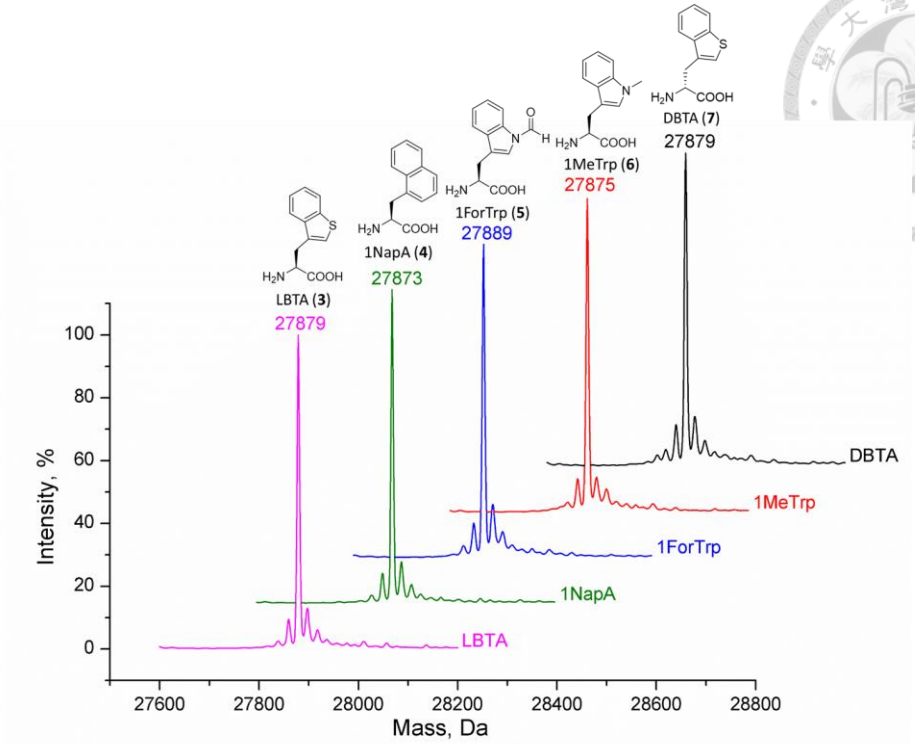


Figure 66. Molecular mass weight determination of sfGFPY66-3, 4, 5, 6, and 7. The deconvoluted ESI-MS spectrum of sfGFPY66-**3, 4, 5, 6, and 7**. The proteins were produced by FOWRS2•$tRNA^{Pyl}_{CUA}$ pair with 1 mM IPTG and 1 mM ncAA **3, 4, 5, 6, and 7** in *E. coli* BL21 (DE3) cells in M9 medium at 37 °C for 12 hours.

doi:10.6342/NTU201803256

3.4.4 Photophysical characteristics of sfGFP variants with heterocyclic ncAAs

In order to realize the impact of sfGFP with heterocyclic amino acids in chromophore (**Figure 67**), the emission and absorption spectrum of sfGFP-Y66TAG with **3-7** was measured. Comparing with wt-sfGFP, the wavelength of protein with **3-7** at Y66 position shifted to the shorter value in emission and excitation spectra. The absorption wavelength of wt-sfGFP was 488 nm whereas sfGFP-Y66TAG with **3-7** revealed shorter wavelength at about 420 nm (**Figure 68-70**). However, the emission wavelength of these variants were slightly different, the emission wavelength of wt-sfGFP was 538 nm, and sfGFP-Y66-**5** and **6** were near 485 nm with disparate waveform of spectra. In addition, sfGFP-Y66TAG-**7** (DBTA) exhibited various peak in emission spectra comparing with sfGFP-Y66TAG-**3** (LBTA) and the wavelength of peak was about 510 nm. Especially, sfGFP-Y66TAG-**7** (DBTA) has the individual stokes shift. Protein with ncAA **4** appeared board spectra of emission from 470 nm to 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 heterocyclic amino acids was lower than wt-sfGFP, and all of sfGFP variants presented the similar emission wavelength at 330 nm. (**Figure 71-73**).

On the other hand, to understand the influence of sfGFP with heterocyclic amino acids around Trp at position 57, the intrinsic fluorescence of sfGFP-F27TAG with **3-9** was measured. The proteins were excited with 280 nm, is the absorption wavelength of tryptophan, and sfGFP contain a Trp at position 57. Comparing with wt-sfGFP, the wavelength of protein with **4, 5, 6, 8,** and **9** at F27 position shifted to the longer value in absorption spectra (**Figure 74**). On the contrary, the wavelength of protein with **3,** and **7** at F27 position shifted to the shorter value in absorption spectra (**Figure 75**). The

doi:10.6342/NTU201803256

absorption wavelength of wt-sfGFP was 323 nm whereas sfGFP-F27TAG with **4, 5, 6, 8,** and **9** revealed longer wavelength at about 330 nm, sfGFP-F27TAG with **3,** and **7** revealed shorted wavelength at about 316 nm. Especially, the results demonstrated the fluorescence intensity of sfGFP-F27TAG with **8** and **9** were longer than wt-sfGFP, but with various FRET spectra of **8** and **9** (**Figure 76).** Comparing with the intrinsic fluorescence spectrum of sfGFP-F27TAG-**8** and sfGFP-F27TAG-**9**, protein with **8** and **9** in chromophore appeared unique peak at 509 nm in emission spectrum and also presented obviously FRET spectrum at 563nm and 562 nm respectively.

doi:10.6342/NTU201803256

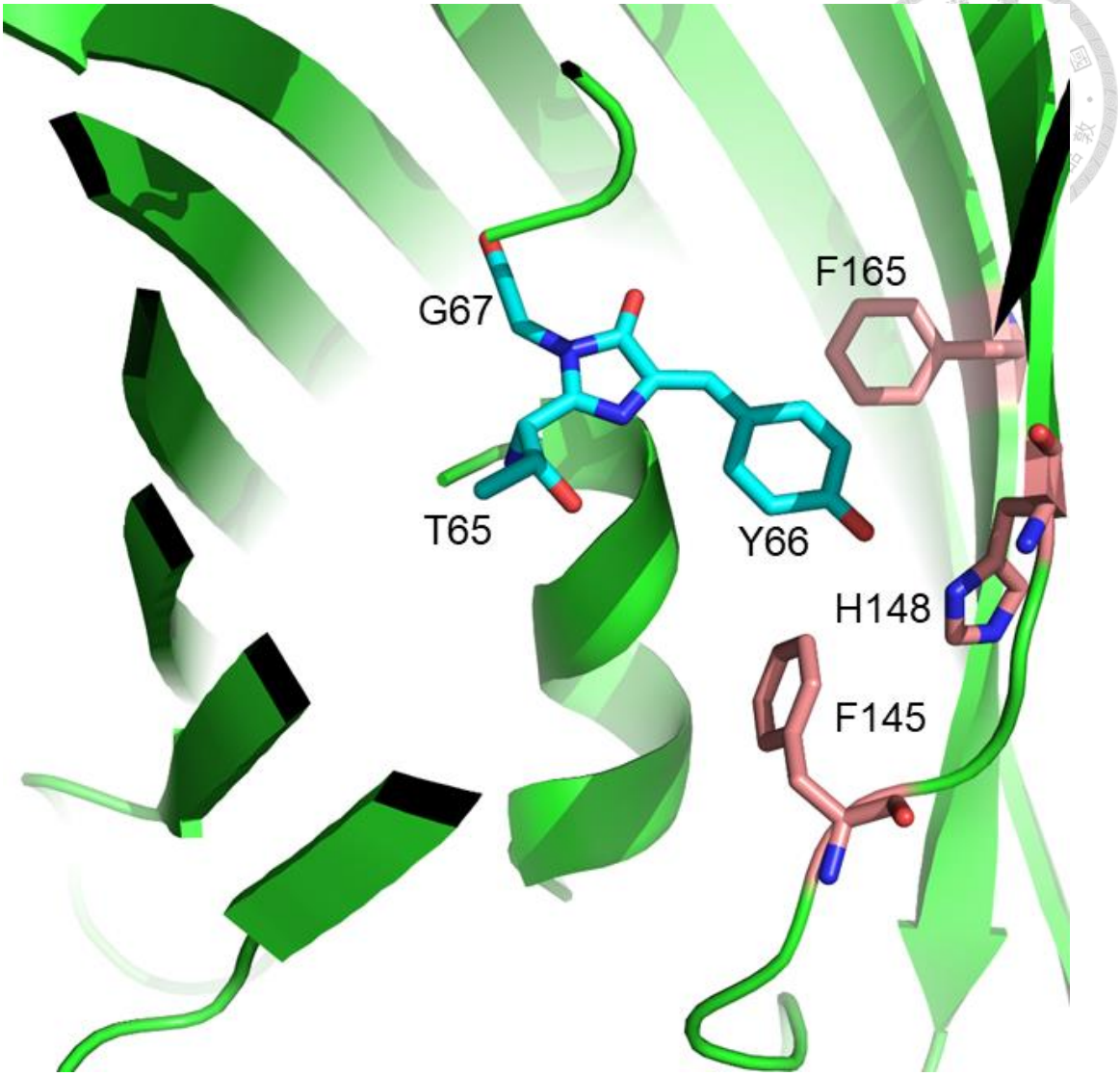


Figure 67. The molecular insight of sfGFP in chromophore.

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 ρ-hydroxybenzylidene-imidazolidone. 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. Drawn using PyMOL and the PDB code is 2B3P.

doi:10.6342/NTU201803256

(A)

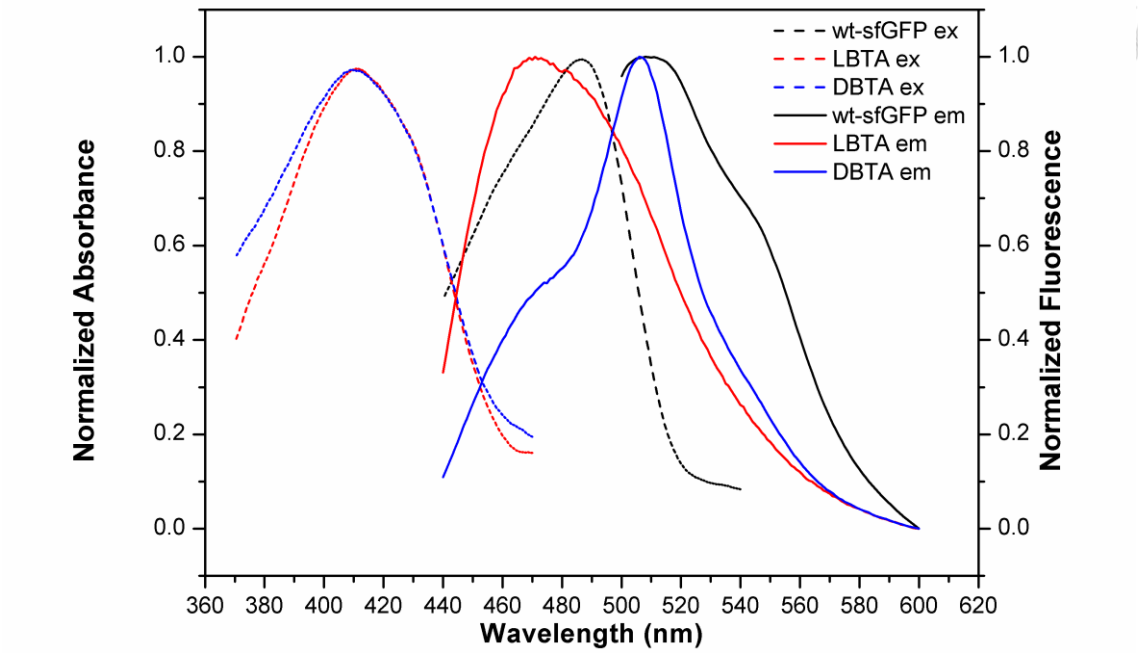


Figure 68. Photophysical characterization of sfGFP-Y66TAG-3 and 7.
Absorption (solid lines) and emission (dashed lines) spectrum of sfGFP-Y66-**3** (red color), sfGFP-Y66-**7** (blue color), and wt-sfGFP (black color) proteins were analyzed with 0.1 mg/mL concentration in 25℃. The maximum of excitation and emission were defined as 1 to normalize the value of spectra. sfGFP-Y66-**3** (ex. 420 nm, em. 471 nm, stokes shift 51), sfGFP-Y66-**7** (ex. 420 nm, em. 506 nm, stokes shift 66, 86), and wt-sfGFP (ex. 488 nm, em. 538 nm).

doi:10.6342/NTU201803256

(A)

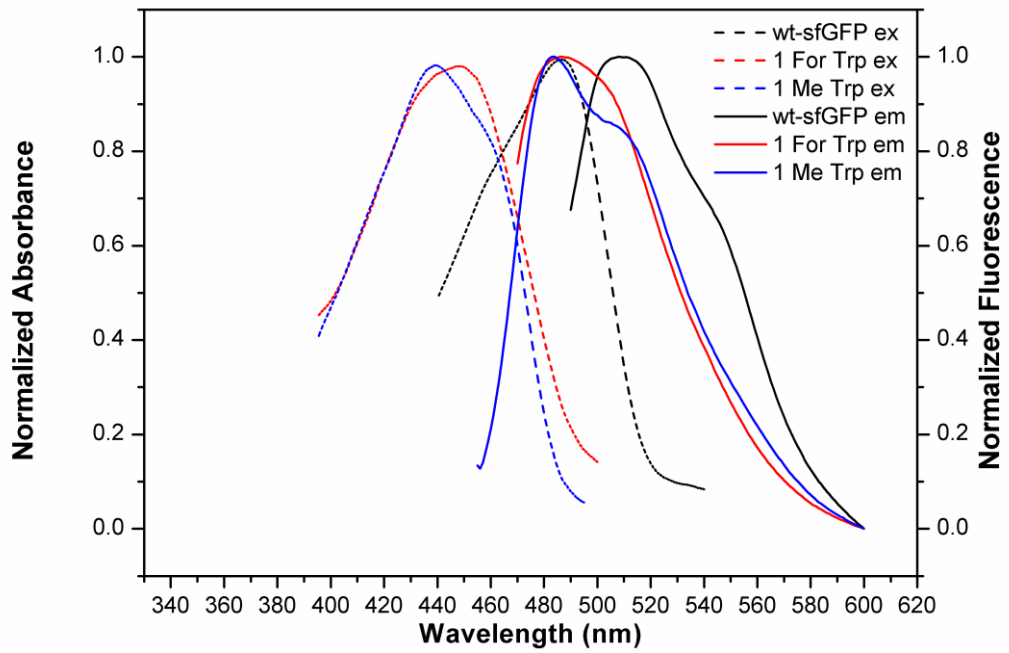


Figure 69. Photophysical characterization of sfGFP-Y66TAG-5 and 6.
Absorption (solid lines) and emission (dashed lines) spectrum of sfGFP-Y66-**5** (red color), sfGFP-Y66-**6** (blue color), and wt-sfGFP (black color) proteins were analyzed with 0.1 mg/mL concentration in 25℃. The maximum of excitation and emission were defined as 1 to normalize the value of spectra. sfGFP-Y66-**5** (ex. 450 nm, em.486 nm, stokes shift 36), sfGFP-Y66-**6** (ex. 445 nm, em. 484 nm, stokes shift 39), and wt-sfGFP (ex. 488 nm, em. 538 nm).

doi:10.6342/NTU201803256

(A)

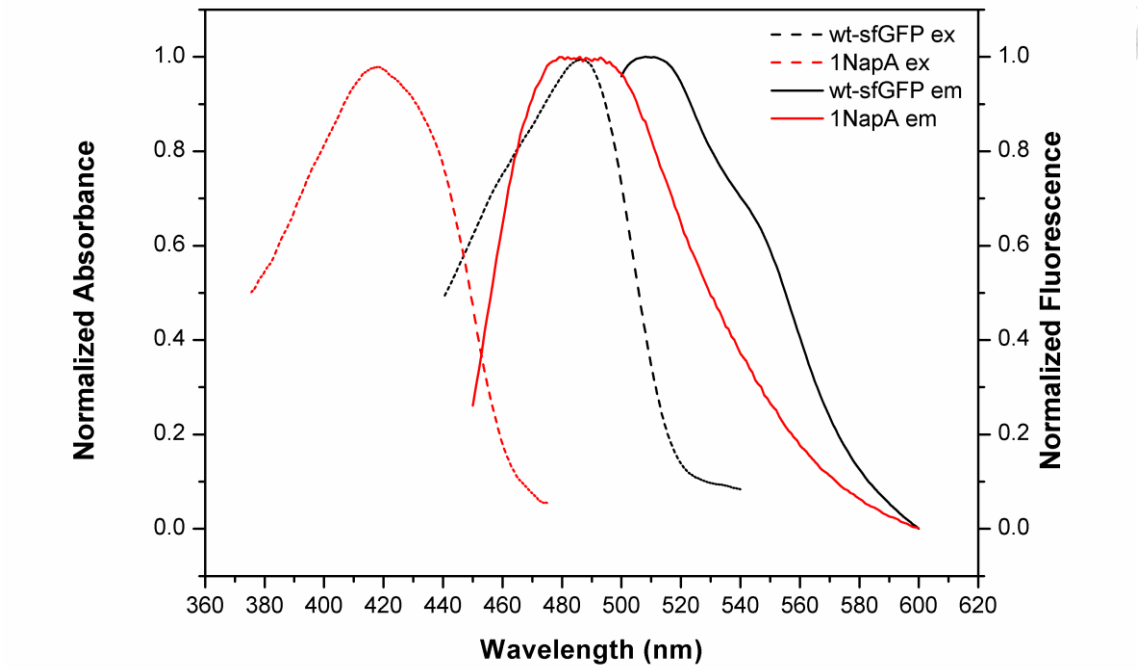


Figure 70. Photophysical characterization of sfGFP-Y66TAG-4.
Absorption (solid lines) and emission (dashed lines) spectrum of sfGFP-Y66-**4** (red color), and wt-sfGFP (black color) proteins were analyzed with 0.1 mg/mL concentration in 25℃. The maximum of excitation and emission were defined as 1 to normalize the value of spectra. sfGFP-Y66-**4** (ex. 425 nm, em.486 nm, stokes shift 61), and wt-sfGFP (ex. 488 nm, em. 538 nm).

doi:10.6342/NTU201803256

(A)

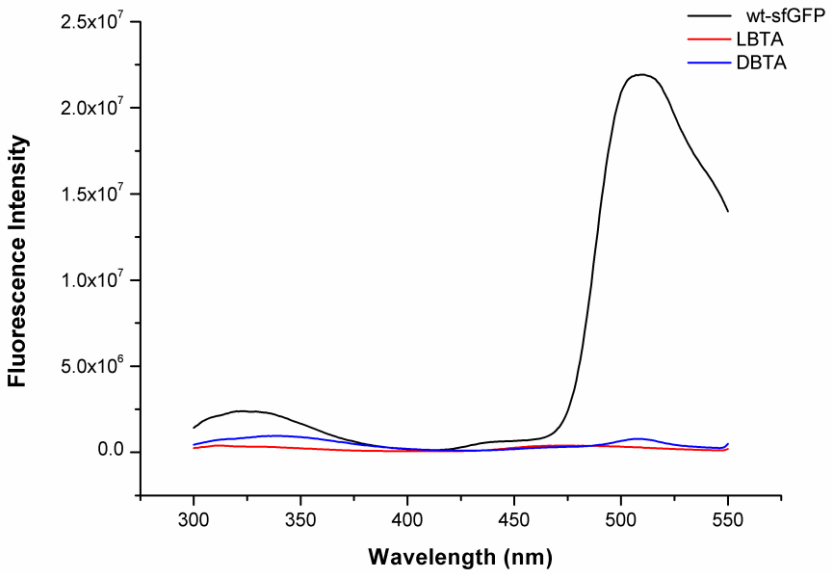


Figure 71. Intrinsic fluorescence spectrum of sfGFP-Y66TAG-3 and 7.
The spectra of sfGFP-Y66-**3** (red color), sfGFP-Y66-**7** (blue color), and wt-sfGFP (black color) proteins were excited with 280 nm in 0.1 mg/mL concentration at 25℃. The spectra of wt-sfGFP contained two obvious peaks in 338 nm and 502 nm. But sfGFP-Y66-**3** spectra and sfGFP-Y66-**7** spectra just contained one obvious peaks in 313 nm and 340nm separately, in addition, the spectra raised a little bit around 510 nm.

doi:10.6342/NTU201803256

(A)

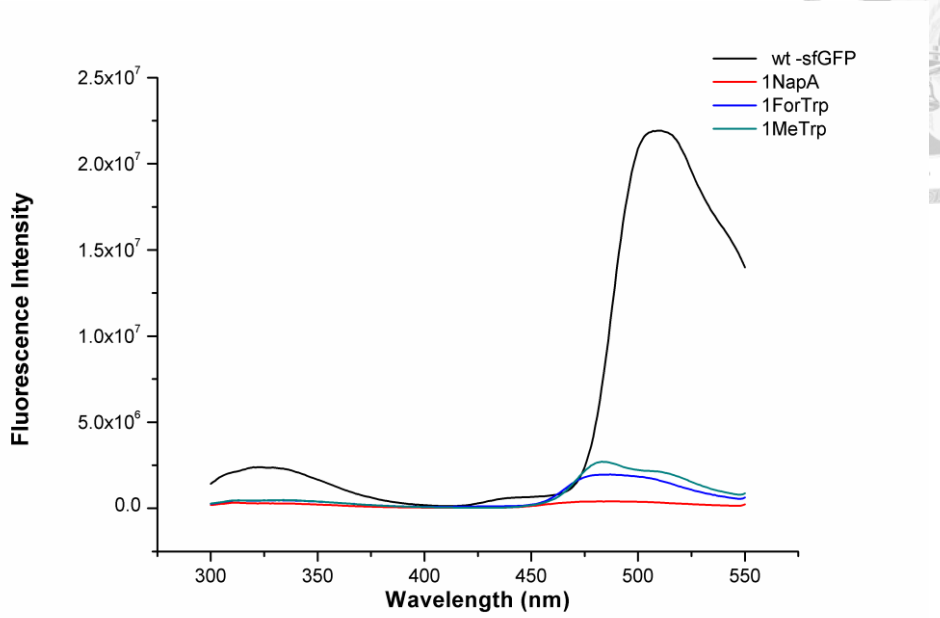


Figure 72. Intrinsic fluorescence spectrum of sfGFP-Y66TAG-4, 5 and 6.
The spectra of sfGFP-Y66-**4** (red color), sfGFP-Y66-**5** (blue color), sfGFP-Y66-**6** (green color), and wt-sfGFP (black color) proteins were excited with 280 nm in 0.1 mg/mL concentration at 25℃. The spectra of wt-sfGFP contained two obvious peaks in 338 nm and 502 nm. But sfGFP-Y66-**4** spectra, sfGFP-Y66-**5** spectra and sfGFP-Y66-**6** spectra just contained one obvious peaks in 488 nm, 486 nm and 484nm separately.

doi:10.6342/NTU201803256

(A)

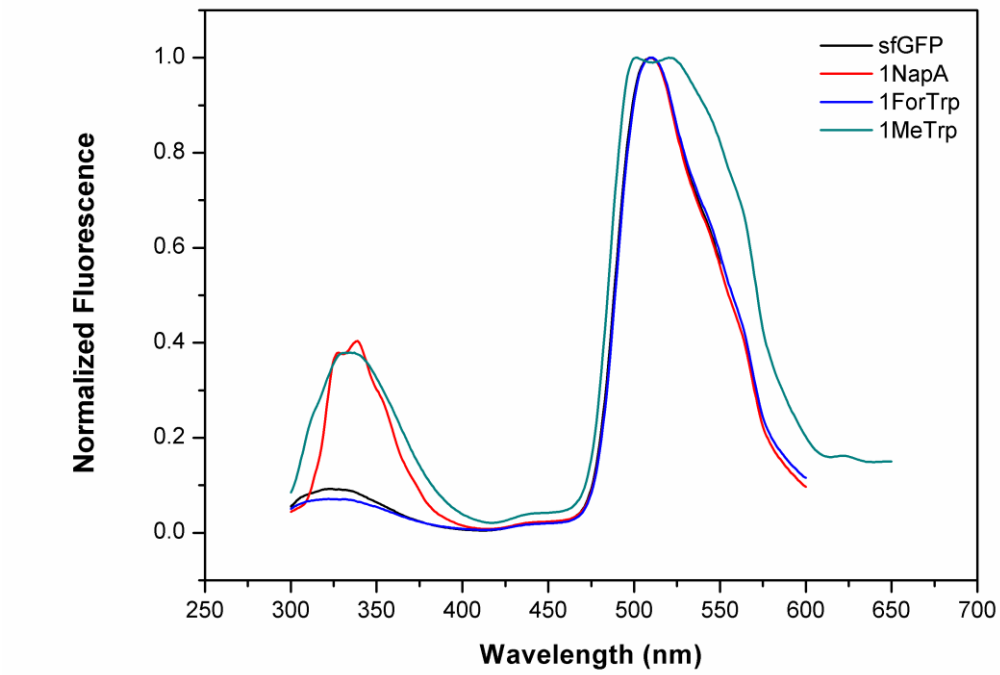


Figure 73. Intrinsic fluorescence spectrum of sfGFP-Y66TAG-4, 5 and 6.
The spectra of sfGFP-F27-**3** (red color), sfGFP-F27-**7** (blue color), and wt-sfGFP (black color) proteins were excited with 280 nm in 1 mg/mL concentration at 25℃. All spectra contained two obvious peaks, which's wavelengths were 323 nm and 509 nm in wt-sfGFP spectra, 316 nm and 511 nm in sfGFP-Y66-**3** spectra, 316 nm and 510 nm in sfGFP-Y66-**7** spectra.

doi:10.6342/NTU201803256

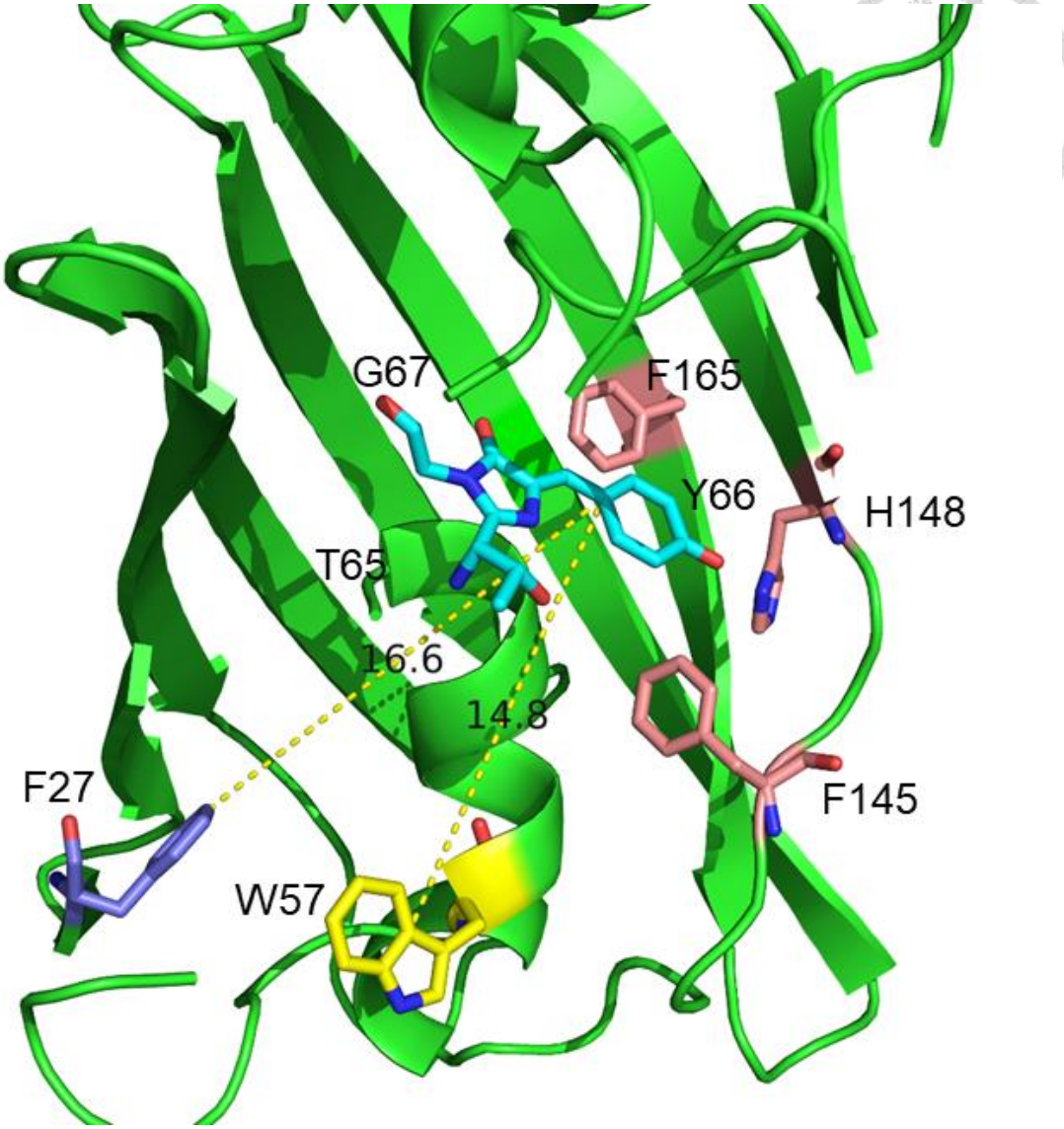


Figure 74. The molecular insight of sfGFP at 27 and 57 position.

The intrinsic fluorescence is resulted from Trp at 57 position in sfGFP. When the heterocyclic amino acids are incorporated at 27 position in sfGFP, the aromatic ring might increase their interaction. Drawn using PyMOL and the PDB code is 2B3P.

doi:10.6342/NTU201803256

(A)

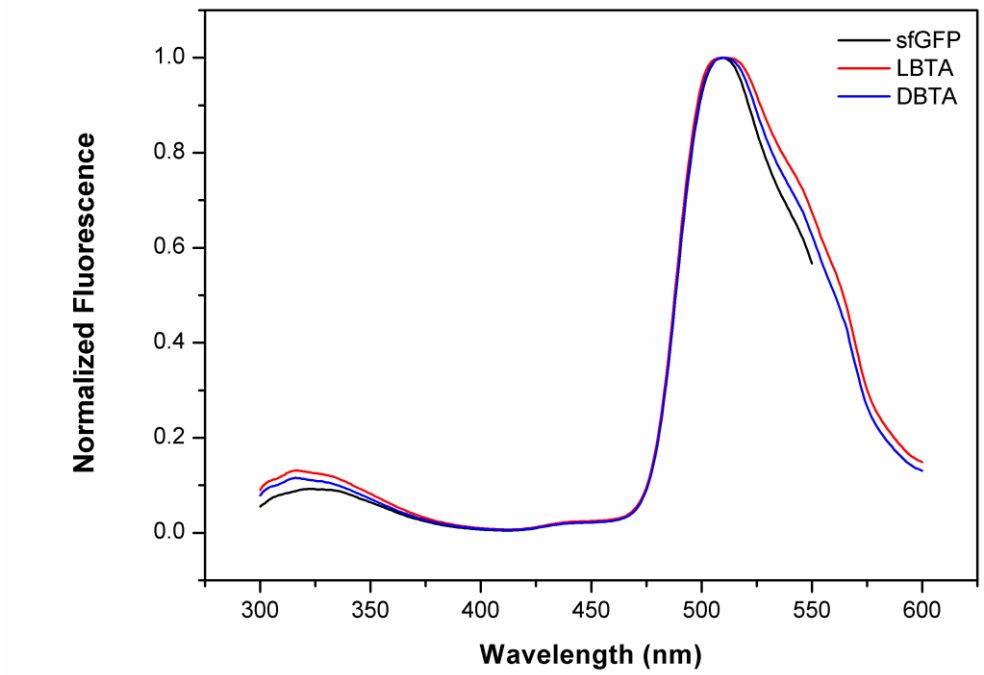


Figure 75. Intrinsic fluorescence spectrum of sfGFP-F27TAG-3 and 7.
The spectra of sfGFP-F27-**3** (red color), sfGFP-F27-**7** (blue color), and wt-sfGFP (black color) proteins were excited with 280 nm in 1 mg/mL concentration at 25℃. All spectra contained two obvious peaks, which's wavelengths were 323 nm and 509 nm in wt-sfGFP spectra, 316 nm and 511 nm in sfGFP-Y66-**3** spectra, 316 nm and 510 nm in sfGFP-Y66-**7** spectra.

doi:10.6342/NTU201803256

(A)

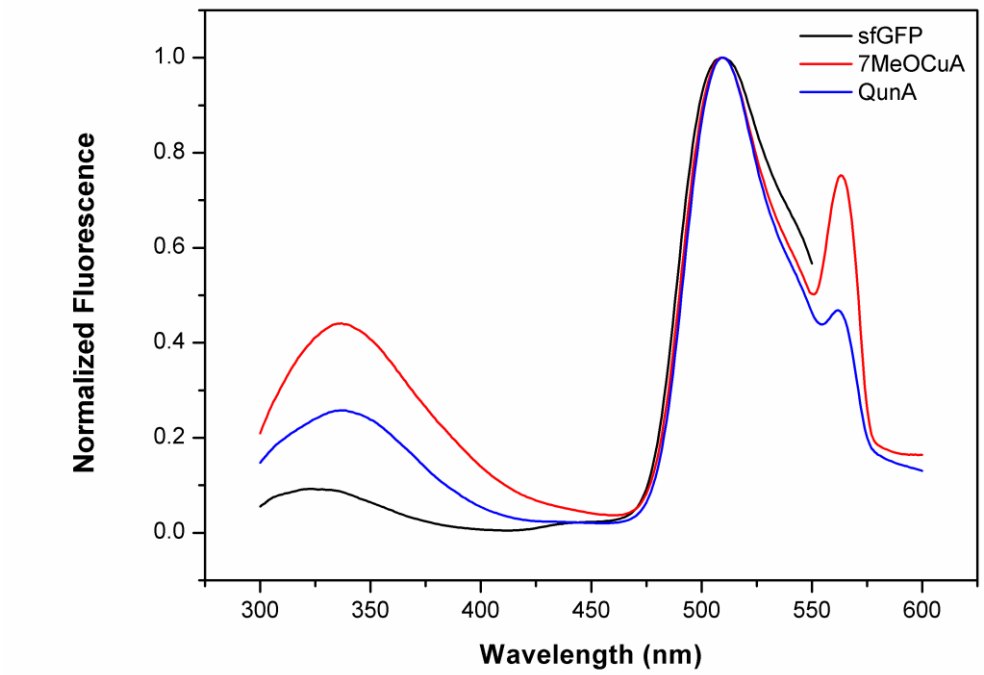


Figure 76. Intrinsic fluorescence spectrum of sfGFP-Y66TAG-8 and 9.
The spectra of sfGFP-Y66-**8** (red color), sfGFP-Y66-**9** (blue color), and wt-sfGFP (black color) proteins were excited with 280 nm in 1 mg/mL concentration at 25℃. All spectra contained two obvious peaks, which's wavelengths were 323 nm and 509 nm in wt-sfGFP spectra, 337 nm and 509 nm in sfGFP-Y66-**8** spectra, 337 nm and 509 nm in sfGFP-Y66-**9** spectra, in addition, the spectra of **8** and **9** have small additional peak at 563 nm and 562 nm respectively.

doi:10.6342/NTU201803256

Chapter 4 Discussion

Expanding genetic codes, an efficient methodology to decipher the mystery of proteins, is not only serving as a platform for protein of interest in structural biology study but also providing observation for modification of proteins. The bioorthogonal system of AARS•tRNA pair was utilized in *E. coli* and capable of genetic incorporation of ncAAs in protein. Basically, PylRS is actually an aminoacyl-tRNA synthetase which derived from archaea and catalyzed the esterification between suppressor tRNA and appropriate ncAA. The research previously showed that evolved *Mm*PylRS on Y306M, L309A, C348T, T364K, and Y384F mutation sites, mkRS1F associated with $tRNA^{Pyl}_{CUA}$ site-specifically incorporated **2** at amber mutation site of a protein in *E. coli*. Two plasmids were constructed for overexpressing the recombinant protein.By using this strategy, the SUMO2 variants containing **2** were produced. Subsequently, the purified SUMO2 variants were confirmed by ESI-MS.

After oxidation process, the reaction undergoes selenoxide β-elimination then generate dehydroalanine for further experiments of protein chemistry.

In the presence of 200 eq. of H_2O_2 for 2hrs in PBS buffer, **2** in SUMO2K11-**2** was converted to dehydroalanine moderately. When the equivalent was increased, the unexpected products were gradually produced. The results of ESI-MS showed single and multiple oxygen atom(s) were added to the SUMO2Dha. The ESI-MS analysis of the final products indicated the oxidation of the one amino acid residue in SUMO2Dha was generated primarily. The found molecular weight 10549 Da, 10565 Da were compared with the calculated molecular weight of SUMO2 with dehydroalanine at its K11 position (SUMO2K11Dha) revealed that it has single oxidation. Thus, the peak of oxidized reaction in SUMO2K11Dha could obviously observed. Compared with enhancing the

doi:10.6342/NTU201803256

equivalent, peak intensity of ESI-MS under the optimized equivalent had a major product in our expectation.

Besides, Ubiquitin derivative on its C- terminal modification synthesized under the E1 catalyzed Ub with cysteamine. When the reaction time was increased, unexpected products were gradually produced. The results of ESI-MS revealed the small chemical compoud, cysteamine formed disulfide bond with the final product. To solve this problem on final product, the equivalent of reducing agent increase to 4mM to reduce the small chemical formation.

Ultimately, protein ligation of Ub-SUMO2 heterodimer could be slightly synthesized through thiol-Michael addition. The reaction might slowly occur under 4 ^{o}C for 4 days but on the contrary it showed the unexpected products under 50 ^{o}C for 2 days. We assumed that the thiol group of Ubiquitin-cysteamine may cause oxidation, protein-protein interaction under the high temperature. In addition, when the reaction time prolonged, the mixture of proteins lead to precipitate. Finally, the results observed that SUMO2K11Dha occur addition with DTT by thiol-Michael addition. Additionally, the SUMO2K11Dha also went through phospha-Michael addition with TCEP. We assumed that the addition of DTT and TCEP are the major reason which lead the low yield of Ubiquitin-SUMO2 heterodimer. In the future, we aspire to gain more Ubiquitin-SUMO2 heterodimer and analyze by distinct apparatus to demonstrate the non-native linkage form.

In combination with development of heterocyclic fluorescent probe specifically recognizes Ub-Ub or Ub-SUMO linkage *in vivo*, we evolved the PylRS to incorporate the bulky aromatic side chain. The PylRS variants were designed with R61K, H63Y, S193R, N346G, C348Q, V401G mutation, together with $tRNA_{CUA}^{Pyl}$ could site-specifically incorporated heterocyclic ncAAs. To ensure the evolved PylRS systems maintained

doi:10.6342/NTU201803256

specific, sfGFP with amber mutation site at F27 position was overexpressed in *E. coli*. Later, the purified sfGFP variants were demonstrated by ESI-MS. Consequently, the heterocyclic ncAAs **3-7** could genetically incorporate to sfGFP with amber mutation site at F27 position.

On the contrary, the PylRS variants was designed with additional mutations at W417T, combine with $tRNA_{CUA}^{Pyl}$ were not site-specifically incorporated heterocyclic ncAAs **8-9** to sfGFP with amber mutation site at F27 position. Especially, the fluorescence intensity of sfGFP-F27TAG with **8** and **9** were longer than wt-sfGFP and revealed additional chromophore. We assumed that Trp at 57 position in sfGFP might increase their π-π interaction with the heterocyclic amino acids at 27 position in sfGFP. These resulted intramolecular FRET and additional emissions wavelength.

Besides, to realize the impact of sfGFP with heterocyclic amino acids in chromophore. FOWRS2• $tRNA_{CUA}^{Pyl}$ could site-specifically incorporated heterocyclic ncAAs at 66 position. Especially, the sfGFP-Y66TAG-**3** and **7** with same molecular weight exhibited various emission spectrum. The results of photophysical characterizations in sfGFP-Y66TAG-**7** (DBTA) exhibited various peak in emission spectra comparing with sfGFP-Y66TAG-**3** (LBTA). We suggested the residue of isomer toward different direction, causing the variously interaction with surround amino acid thus, these two protein lead distinct spectra.

doi:10.6342/NTU201803256

Chapter 5 Conclusion

In summary, using expanding genetic codes approach, we have developed a method to genetically incorporate a *Se*-alkylselenocysteine into SUMO2 variants at eight different lysines sites in *E.coli*. The suppression yield of the SUMO2 variants with N-terminal His_{6}-tagged sfGFP fusion protein were ranging from 21 to 69%. Subsequently, the purification of SUMO2 variants has been accomplished. Thus, this methodology has been demonstrated on synthesizing SUMO2 variants. Protein chemistry study of *Se*-alkylselenocysteine on SUMO2 variants was converted to dehydroalanine through selenoxide β-elimination reaction. The chemoenzymatic synthesis of Ub variants on its C-terminal modification utilizing Ubiquitin-activating enzyme and cysteamine was successfully complished. Furthermore, the small molecule of thiophosphate was reacted with SUMO2K11Dha to validate the Michael addition reaction.

On the other hand, to expand the diversity of PylRS systems to incorporate heterocyclic and bicyclic aromatic amino acid, PylRS was rationally mutated at R61K, H63Y, S193R, N346G, C348Q, V401G. Heterocyclic aromatic acid can be successfully incorporated into protein. Unexpectedly, the bulky aromatic amino acids were also partially incorporated at the desired position by additional mutations at W417T in PylRS. This suggest the wide substrate range of engineered PylRS for heterocyclic amino acid incorporation. Next it is necessary to alter specificity of the PylRS mutant to increase the chemical diversity, as well as make it possible to incorporate more heterocyclic amino acids in protein simultaneously. Finally, the photophysical characterization of **8** and **9** incorporated sfGFP at F27 have revealed different fluorescent spectra and the biophysical properties of D-form heterocyclic ncAAs exhibited additional fluorophore compare to the L-form.

doi:10.6342/NTU201803256

Reference

1 Crick, F. H. The origin of the genetic code. *Journal of molecular biology* **38**, 367-379 (1968).

2 McCarthy, J. E. *et al.* Prokaryotic translation: the interactive pathway leading to initiation. *Trends in genetics* **10**, 402-407 (1994).

3 Adams, J. M. On the release of the formyl group from nascent protein. *Journal of molecular biology* **33**, 571-589 (1968).

4 Ball, L. A. *et al.* Cleavage of the N-terminal Formymlethionine Residue from a Bacteriophage Coat Protein in vitro. *Journal of molecular biology* **79**, 531-537 (1973).

5 Andersen, G. R. *et al.* Elongation factors in protein biosynthesis. *Trends in biochemical sciences* **28**, 434-441 (2003).

6 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).

7 Liu, W. R. *et al.* Synthesis of proteins with defined posttranslational modifications using the genetic noncanonical amino acid incorporation approach. *Molecular biosystems* **7**, 38-47 (2011).

8 Saitoh, H. *et al.* Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *Journal of biological chemistry* **275**, 6252-6258 (2000).

9 Keusekotten, K. *et al.* Multivalent interactions of the SUMO-interaction motifs in RING finger protein 4 determine the specificity for chains of the SUMO. *Biochemical journal* **457**, 207-214 (2014).

10 Hay, R. T. SUMO: a history of modification. *Molecular cell* **18**, 1-12 (2005).

11 Johnson, E. S. Ubiquitin branches out. *Nature cell biology* **4**, 295-298 (2002).

12 Ulrich, H. D. Mutual interactions between the SUMO and ubiquitin systems: a plea of no contest. *Trends in cell biology* **15**, 525-532 (2005).

13 Martin, S. *et al.* Emerging extranuclear roles of protein SUMOylation in neuronal function and dysfunction. *Nature reviews neuroscience* **8**, 948-959 (2007).

14 Hershko, A. *et al.* The ubiquitin system. *Annual Reviews* 425-479 (1998)

15 Hershko, A. *et al.* The ubiquitin system for protein degradation. *Annual review of biochemistry* **61**, 761-807 (1992).

16 Peng, J. *et al.* A proteomics approach to understanding protein ubiquitination. *Nature biotechnology* **21**, 921-926 (2003).

17 Inobe, T. *et al.* Defining the geometry of the two-component proteasome degron. *Nature chemical biology* **7**, 161-167 (2011).

18 Hoege, C. *et al.* RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**, 135-141 (2002).

19 Mukhopadhyay, D. *et al.* Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* **315**, 201-205 (2007).

20 Pickart, C. M. Mechanisms underlying ubiquitination. *Annual review of biochemistry* **70**, 503-533 (2001).

21 Dou, H. *et al.* BIRC7–E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nature structural and molecular biology* **19**, 876-883 (2012).

22 Komander, D. *et al.* Breaking the chains: structure and function of the deubiquitinases. *Nature reviews molecular cell biology* **10**, 550-563 (2009).

doi:10.6342/NTU201803256

23 Carter, S. *et al.* C-terminal modifications regulate MDM2 dissociation and nuclear export of p53. *Nature cell biology* **9**, 428-435 (2007).
24 Kaiser, S. E. *et al.* Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. *Nature methods* **8**, 691-696 (2011).
25 Leinfelder, W., Zehelein, E., MandrandBerthelot, M. & Bock, A. Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* **331**, 723-725 (1988).
26 Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459-1462 (2002).
27 Kwok, Y. & Wong, J. T.-F. Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. *Canadian journal of biochemistry* **58**, 213-218 (1980).
28 Steer, B. A. & Schimmel, P. Major anticodon-binding region missing from an archaebacterial tRNA synthetase. *Journal of Biological Chemistry* **274**, 35601-35606 (1999).
29 Cohen, G. & Munier, R. Incorporation of structural analogues of amino acids in bacterial proteins. *Biochimica et biophysica acta* **21**, 592-593 (1956).
30 Wan, W. *et al.* A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angewandte Chemie International Edition* **49**, 3211-3214 (2010).
31 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).
32 Singh-Blom, A., Hughes, R. A. & Ellington, A. D. in *Enzyme Engineering* 93-114 (2013).
33 Xiu, X., Puskar, N. L., Shanata, J. A., Lester, H. A. & Dougherty, D. A. Nicotine binding to brain receptors requires a strong cation–π interaction. *Nature* **458**, 534-537 (2009).
34 Banerjee, A. *et al.* Molecular bases of cyclodextrin adapter interactions with engineered protein nanopores. *Proceedings of the national academy of sciences of the United States of America* **107**, 8165-8170 (2010).
35 Fujiki, M. *et al.* Optically active polysilanes. Ten years of progress and new polymer twist for nanoscience and nanotechnology. *Polymer journal* **35**, 297-344 (2003).
36 Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annual review of biophysics and biomolecular structure* **35**, 225-249 (2006).
37 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).
38 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).
39 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).
40 Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding N^{ε}-acetyllysine in recombinant proteins. *Nature chemical biology* **4**, 232-234 (2008).

doi:10.6342/NTU201803256

41 Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annual review of biochemistry* **79**, 413-444 (2010).
42 Raibaut, L., Ollivier, N. & Melnyk, O. Sequential native peptide ligation strategies for total chemical protein synthesis. *Chemical society reviews* **41**, 7001-7015 (2012).
43 Kulkarni, S. S., Sayers, J., Premdjee, B. & Payne, R. J. Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. *Nature reviews chemistry* **2**, 0122 (2018).
44 Reich, H. J. & Wollowitz, S. Preparation of α, β - Unsaturated Carbonyl Compounds and Nitriles by Selenoxide Elimination. *Organic reactions* (1993).
45 Emerson, D. W., Craig, A. P. & Potts Jr, I. W. Pyrolysis of unsymmetrical dialkyl sulfoxides. Rates of alkene formation and composition of the gaseous products. *The Journal of organic chemistry* **32**, 102-105 (1967).
46 Sharpless, K., Young, M. & Lauer, R. Reactions of selenoxides: Thermal syn-elimination and $H_2{}^{18}O$ exchange. *Tetrahedron letters* **14**, 1979-1982 (1973).
47 Emori, E., Arai, T., Sasai, H. & Shibasaki, M. A catalytic Michael addition of thiols to α, β-unsaturated carbonyl compounds: Asymmetric Michael additions and asymmetric protonations. *Journal of the American chemical society* **120**, 4043-4044 (1998).
48 Mather, B. D., Viswanathan, K., Miller, K. M. & Long, T. E. Michael addition reactions in macromolecular design for emerging technologies. *Progress in polymer science* **31**, 487-531 (2006).
49 Little, R. D., Masjedizadeh, M. R., Wallquist, O. & Mcloughlin, J. I. The Intramolecular M ichael Reaction. *Organic reactions* **47**, 315-552 (2004).
50 Windle, C. L. *et al.* Extending enzyme molecular recognition with an expanded amino acid alphabet. *Proceedings of the national academy of sciences of the United States of America*, 201616816 (2017).
51 Dadová, J., Galan, S. R. & Davis, B. G. Synthesis of modified proteins via functionalization of dehydroalanine. *Current opinion in chemical biology* **46**, 71-81 (2018).
52 Guo, J., Wang, J., Lee, J. S. & Schultz, P. G. Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angewandte Chemie International Edition* **47**, 6399-6401 (2008).
53 Meledin, R., Mali, S. M., Singh, S. K. & Brik, A. Protein ubiquitination via dehydroalanine: development and insights into the diastereoselective 1,4-addition step. *Org Biomol Chem* **14**, 4817-4823 (2016).
54 Mevissen, T. E. & Komander, D. Mechanisms of deubiquitinase specificity and regulation. *Annual review of biochemistry* **86**, 159-192 (2017).
55 Lecona, E. *et al.* USP7 is a SUMO deubiquitinase essential for DNA replication. *Nature structural and molecular biology* **23**, 270-277 (2016).
56 Carvalho, A. F. *et al.* High-yield expression in Escherichia coli and purification of mouse ubiquitin-activating enzyme E1. *Molecular biotechnology* **51**, 254-261 (2012).
57 Wang, X. A., Kurra, Y., Huang, Y., Lee, Y. J. & Liu, W. R. E1 - Catalyzed Ubiquitin C - Terminal Amidation for the Facile Synthesis of Deubiquitinase Substrates. *ChemBioChem* **15**, 37-41 (2014).

doi:10.6342/NTU201803256

58 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**, 2856-2860 (2006).
59 Beers, E. & Callis, J. Utility of polyhistidine-tagged ubiquitin in the purification of ubiquitin-protein conjugates and as an affinity ligand for the purification of ubiquitin-specific hydrolases. *Journal of biological chemistry* **268**, 21645-21649 (1993).
60 Krist, D. T., Park, S., Boneh, G. H., Rice, S. E. & Statsyuk, A. V. UbFluor: a mechanism-based probe for HECT E3 ligases. *Chemical science* **7**, 5587-5595 (2016).
61 Englert, M. *et al.* Probing the active site tryptophan of Staphylococcus aureus thioredoxin with an analog. *Nucleic acids research* **43**, 11061-11067 (2015).
62 Ma, H., Liu, N., Shi, S., Wang, S. & Chen, Y. Genetic incorporation of D-amino acids into green fluorescent protein based on polysubstrate specificity. *Royal society of chemistry advances* **5**, 39580-39586 (2015).
63 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).

doi:10.6342/NTU201803256

Appendix

Organic synthesis of N^{ε}-((2-azidoethoxy) carbonyl)-L-lysine

NaN_3 (4.68 g, 72 mmol) was added to a solution of 2-bromoethanol (1.17 mL, 24 mmol) in water (10 mL) and acetone (10 mL). The mixture was stirred at 80℃ in reflux apparatus overnight and cooled to room temperature. The aqueous layer was extracted with ether (3 x 40 mL). The combined organic extracts were dried over magnesium sulfate then the precipitate removed by filtration, and the solvent evaporated *in vacuo.* The resulting was 2-azidoethanol as a yellow oil in 60% yield (1.24 g, 14.22 mmol).[63]

In the second reaction, 2-azidoethanol (258 mg, 2.96 mmol) was add a solution of triphosgene (0.9 g, 3.03 mmol) in dried THF (5 mL) under ice bath. The reaction was stirred overnight, and THF was evaporated under vacuum. The resulting compound was 2-azidoethyl carbonochloridate in 87.5% yield (387 mg, 2.59 mmol).

The Boc-Lys-OH (0.85 g, 3.44 mmol) was added in NaOH (3 mL)/ DMF (3 mL) solution under ice bath then slowly added 2-azidoethyl carbonochloridate (0.57 g, 3.81 mmole) dissolved in DMF (3 mL). The reaction mixture was stirred for 24 h. The resulting compound was extracted with ethyl acetate (30 mL), and washed with brine (3 x 15 mL), subsequently washed with dd H_2O (3 x 15 mL). The organic layer was then dried over magnesium sulfate then the precipitate removed by filtration, and evaporated, affording clean as a yellow solid. Subsequently, the compound was deprotected in DCM (5 mL) and slowly added TFA (5 mL) solution. The reaction was stirred at RT for 1 hour, added ether about 100 mL until appearing white precipitate. The precipitate was centrifuged and dried under vacuum at least 12 hrs, affording pure N^{ε}-((2-azidoethoxy) carbonyl)-*L*-lysine in 84% yield (0.343 mg, 1.32 mmol).

doi:10.6342/NTU201803256

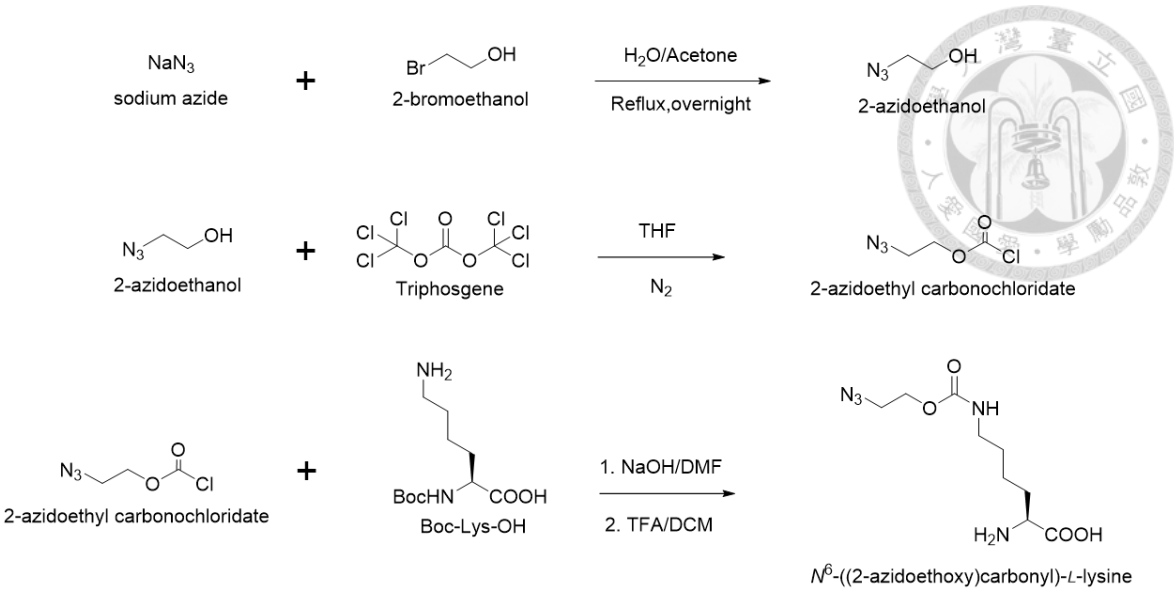


Figure S1. Synthesis of N$^{\varepsilon}$-((2-azidoethoxy) carbonyl)-L-lysine.

doi:10.6342/NTU201803256

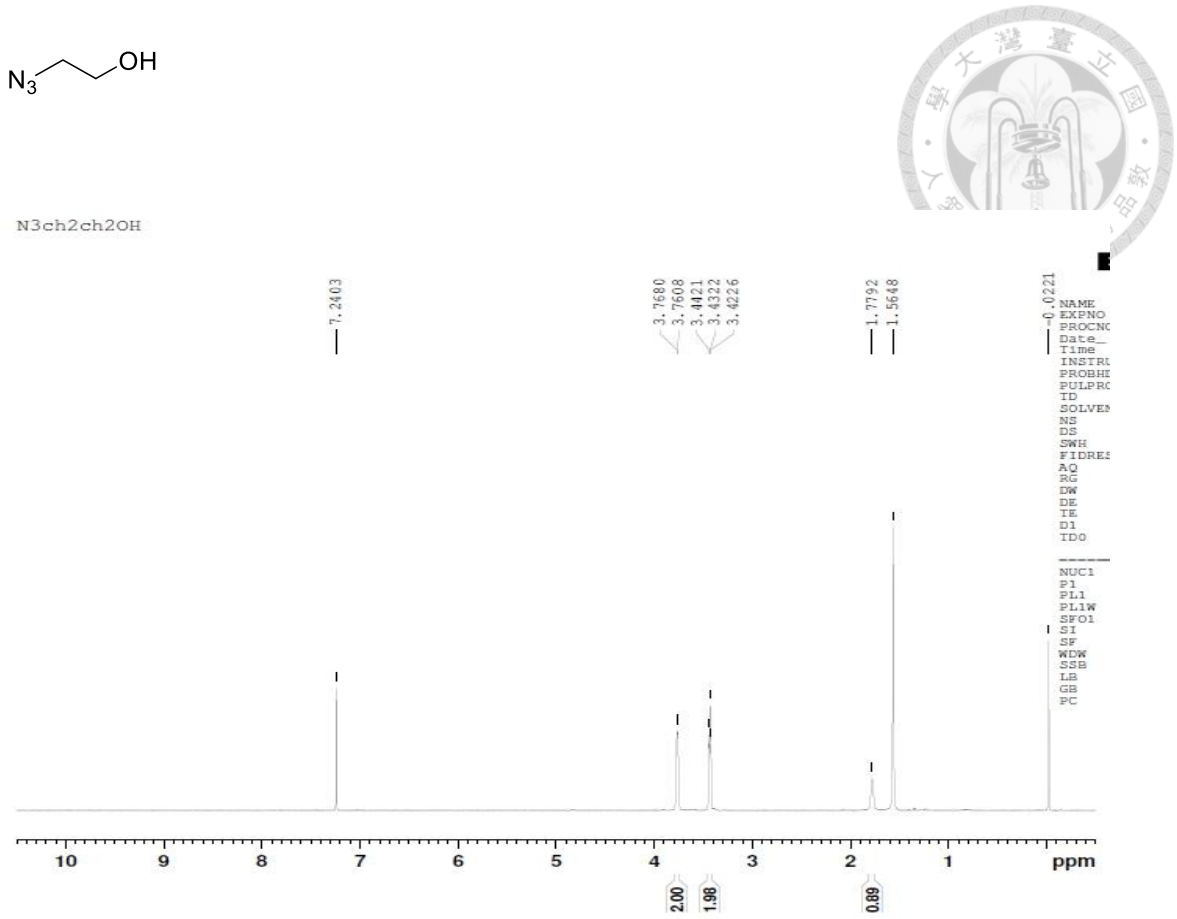


Figure S2. ^{1}H NMR spectra (500 MHz) of 2-azidoethanol in $CDCl_3$.
^{1}H NMR ($CDCl_3$, 500 MHz), δ 1.78 (s, 1H), 3.42~3.44 (t, 2H), 3.76~3.77(d, 2H)

doi:10.6342/NTU201803256

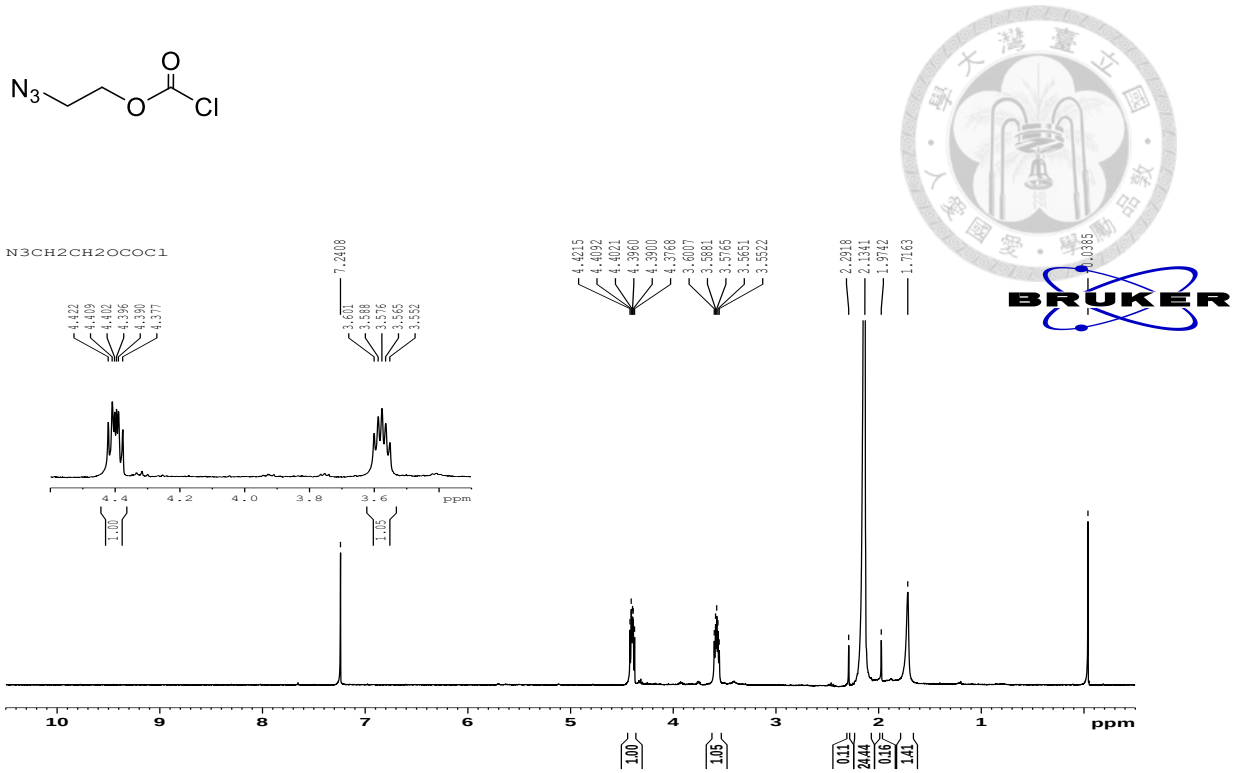
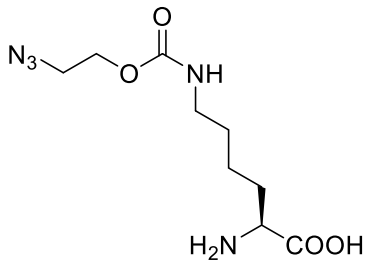


Figure S3. **^{1}H NMR spectra (500 MHz) of 2-azidoethyl carbonochloridate in $CDCl_3$.**

^{1}H NMR ($CDCl_3$, 500 MHz), δ 3.55~3.60 (m, 2H), 4.38~4.42 (m, 2H)

doi:10.6342/NTU201803256



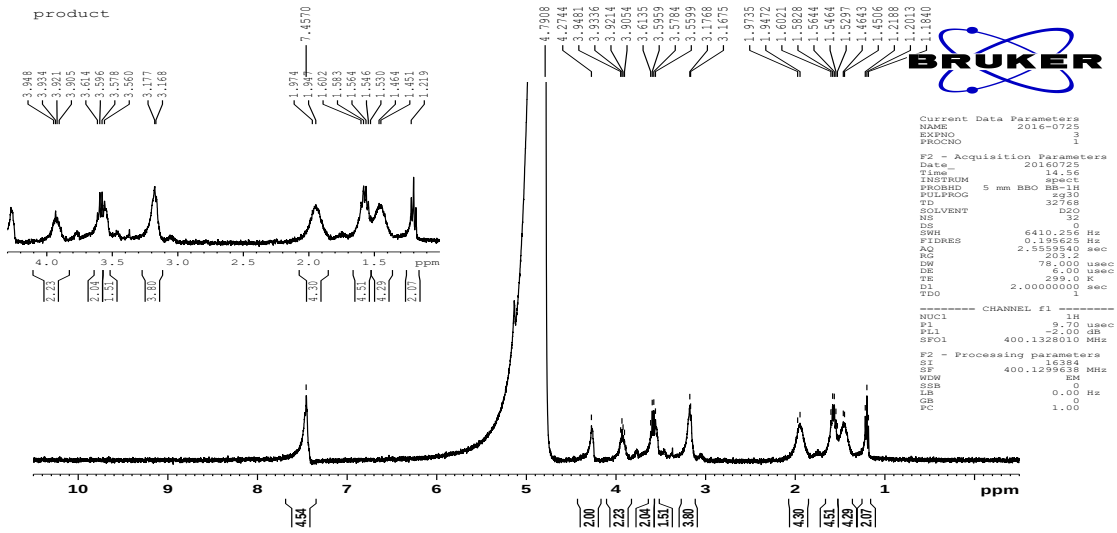


Figure S4. ^{1}H NMR spectra (400 MHz) of (*S*)-2-amino-6-(((2-azidoethoxy)carbonyl)amino)hexanoic acid in D_2O.

^{1}H NMR (D_2O, 400 MHz), δ1.21 (s, 2H), δ1.45~1.80 (m, 8H), 1.94~1.97 (d, 4H), 3.16~3.17 (d, 4H), 3.56~3.57 (d, 2H), 3.59~3.61 (d, 2H), 3.90~3.94 (q, 2H)

doi:10.6342/NTU201803256

Organic synthesis of (*S*)-2-Amino-3-((2-(((benzyloxy)carbonyl)amino)ethyl)selanyl)propanoic acid

Benzyl (2-hydroxyethyl)carbamate (1.22 g, 6.25 mmol) was dissolved in DCM (20 mL) then added methanesulfonyl chloride (0.58 mL, 7.49 mmol), and NEt_3 (1.12 mL). After the mixture was stirred for 45 mintunes, the solution was added LiBr (5.43 g, 62.5 mmole) and acetone (20 mL) respectively for 24 hrs. The aqueous layer was evaporated *in vacuo* then added dd H_2O (40 mL), and ether (40 mL) to extract compounds. The organic layer was wash with citric acid (15 mL) and brine (15 mL). Subsequently, the organic layer was dried over magnesium sulfate then the precipitate removed by filtration, and the solvent evaporated *in vacuo.* The resulting was a white solid with Benzyl (2-bromoethyl)carbamate in 78.4% yield.

L-Seleno cysteine (90 mg, 0.269 mmol) was added the dd H_2O (3 mL) under ice bath at 0℃ then slowly added $NaBH_4$ (120 mg, 3.17 mmol). Finally, Benzyl (2-bromoethyl)carbamate (300 mg, 1.16 mmol) and 1,4-dioxane (one drop) mixed with the solution and stirred for 4hrs。The solution was acidified to pH 7 with 1M HCl solution then extracted with ddH_2O and ether. The precipitate was centrifuged and dried under vacuum at least 12 hrs, affording pure (*S*)-2-Amino-3-((2-(((benzyloxy)carbonyl)amino)ethyl)selanyl)propanoic acid in 94% yield.

doi:10.6342/NTU201803256

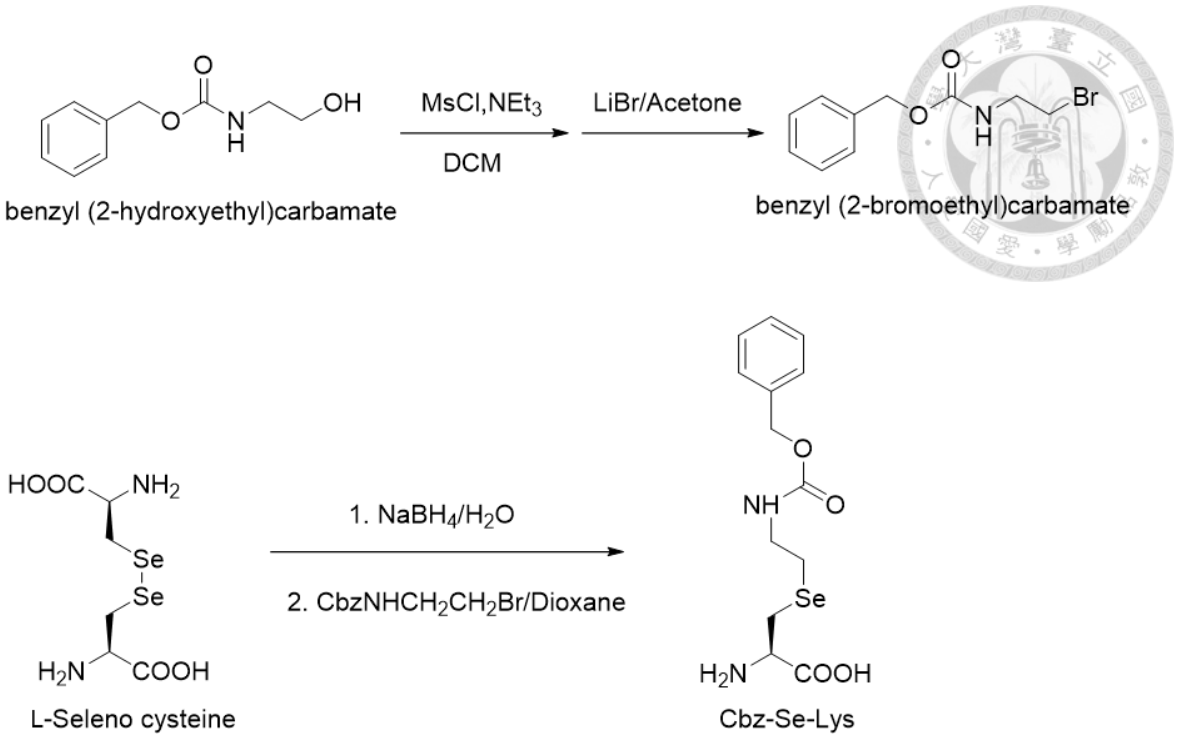


Figure S5. Synthesis of (*S*)-2-Amino-3-((2-(((benzyloxy)carbonyl)amino)ethyl)selanyl)propanoic acid.

doi:10.6342/NTU201803256

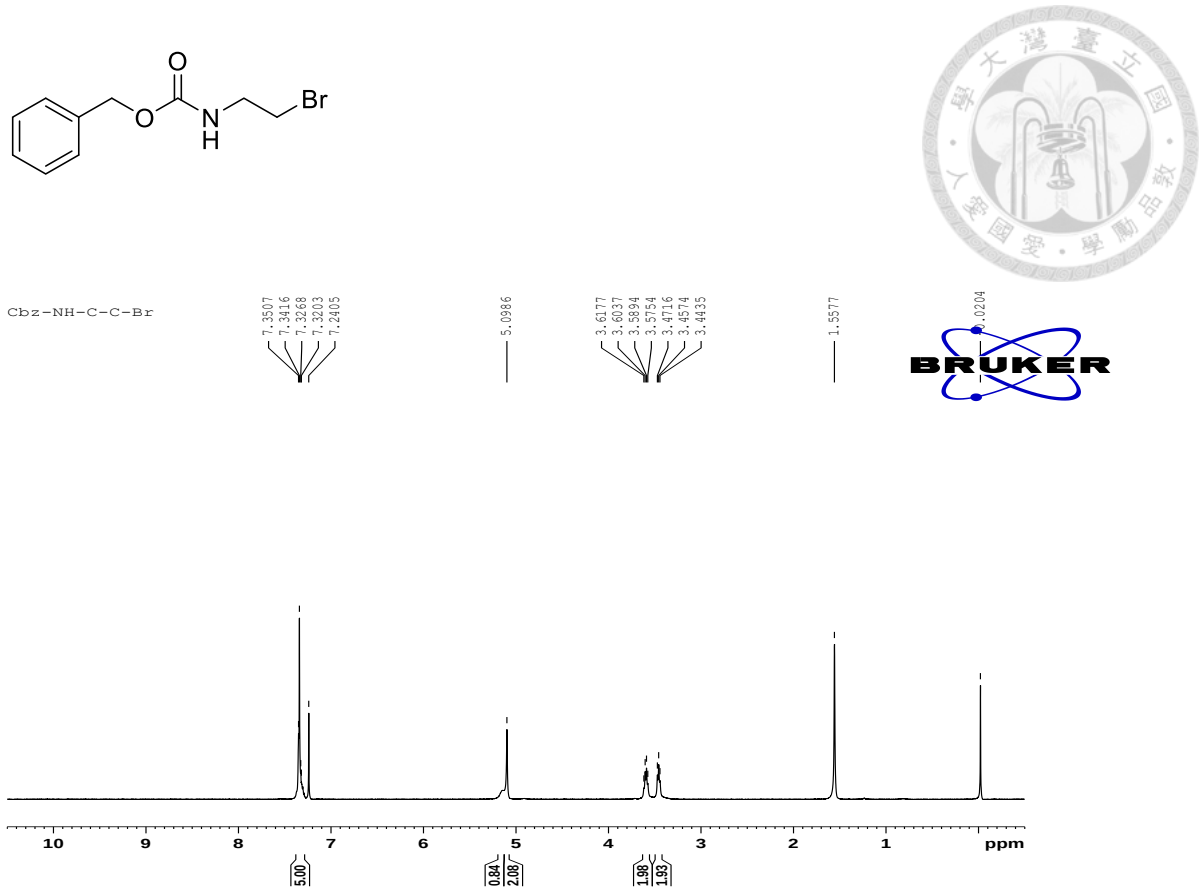
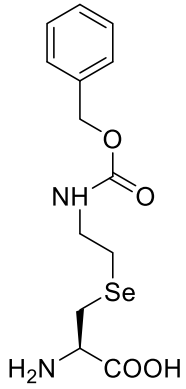


Figure S6. ^{1}H NMR spectra (500 MHz) of Benzyl (2-bromoethyl)carbamate in $CDCl_3$.

^{1}H NMR ($CDCl_3$, 500 MHz), δ 3.44~3.47 (t, 2H), 3.59~3.62 (q, 2H), 5.10 (s, 2H), 7.32~7.35 (m, 5H)

doi:10.6342/NTU201803256



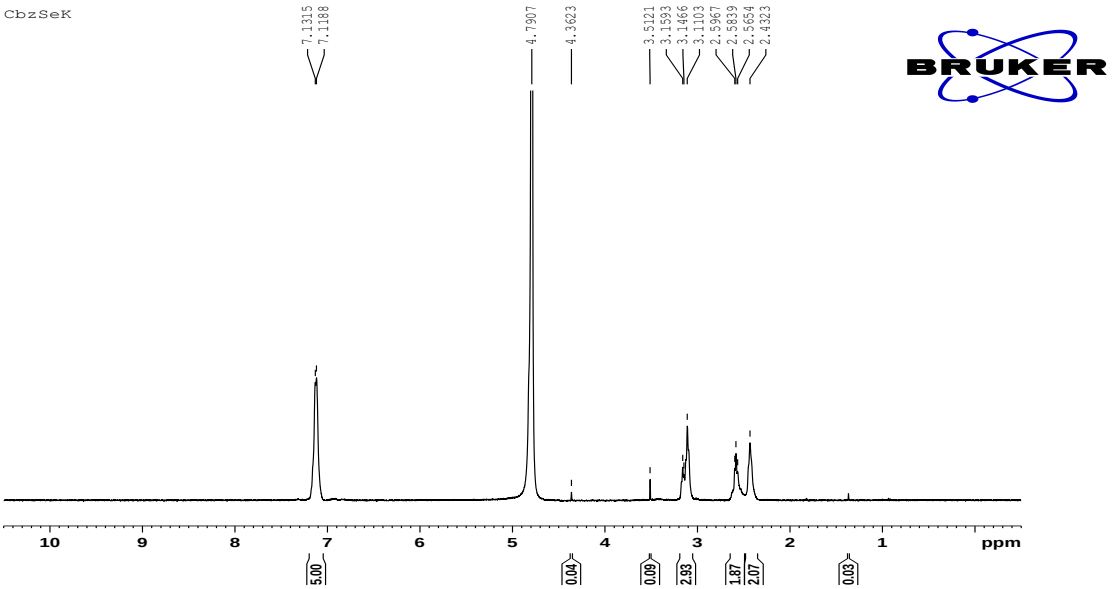


Figure S7. ^{1}H NMR spectra (500 MHz) of Cbz-Se-Lys in D_2O.

^{1}H NMR (D_2O, 500 MHz), δ2.43 (s, 2H), 2.57~2.60 (t, 2H), 3.11~3.16 (t, 3H), 7.12~7.13 (d, 5H).

doi:10.6342/NTU201803256

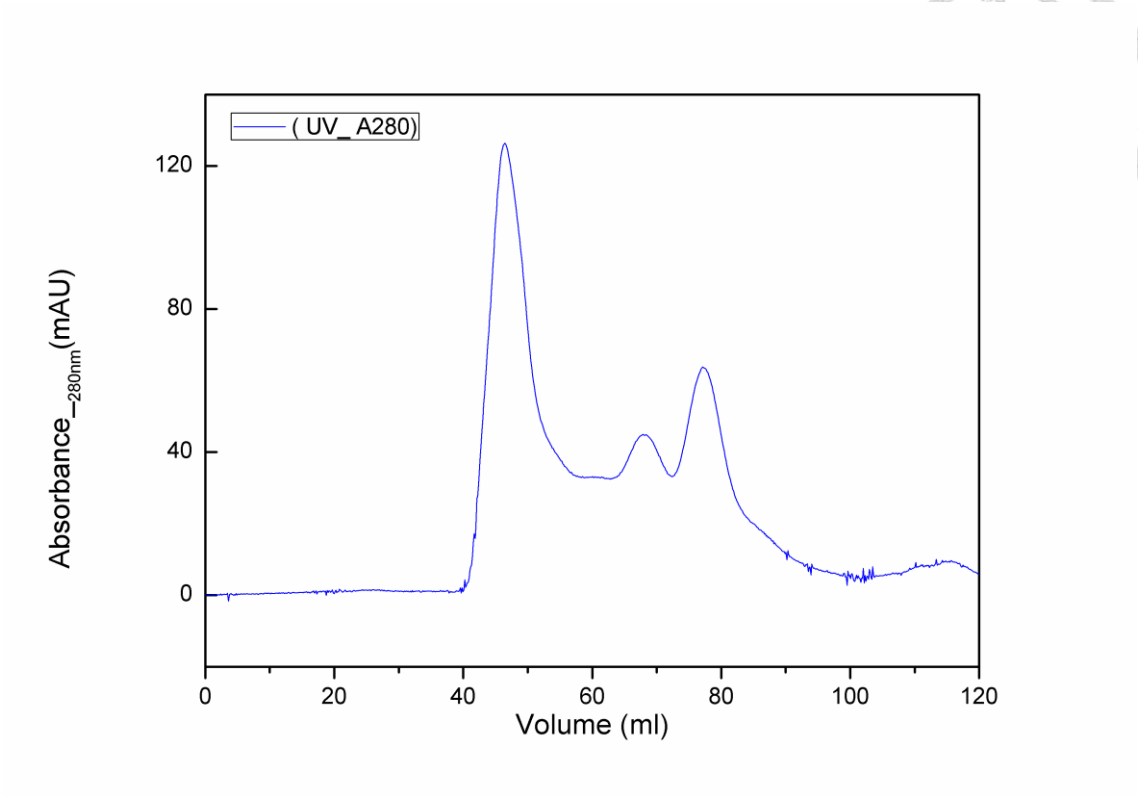


Figure S8. Purification of SUMO2K5-2 with SEC.
The flow-through after Ni^{2+}-NTA column in 50 mM Tris, and 200 mM NaCl was concentrated for FPLC with SEC column (HiLoad 16/600 superdex 75 pg, GE Healthare). Blue line: the signal of UV_{A280}.

doi:10.6342/NTU201803256

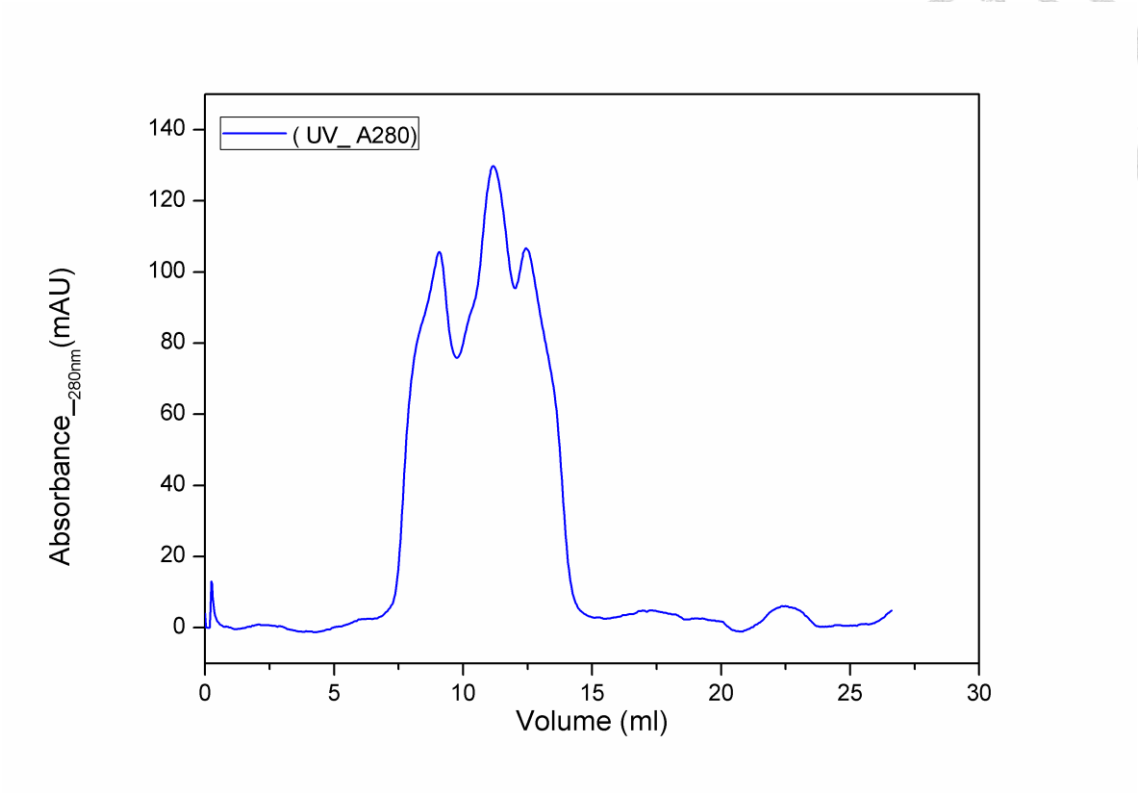


Figure S9. Purification of SUMO2K7-2 with SEC.
The flow-through after Ni^{2+}-NTA column in 50 mM Tris, and 200 mM NaCl was concentrated for FPLC with SEC column (Superdex 75 10/300 GL, GE Healthare). Blue line: the signal of UV_{A280}.

doi:10.6342/NTU201803256

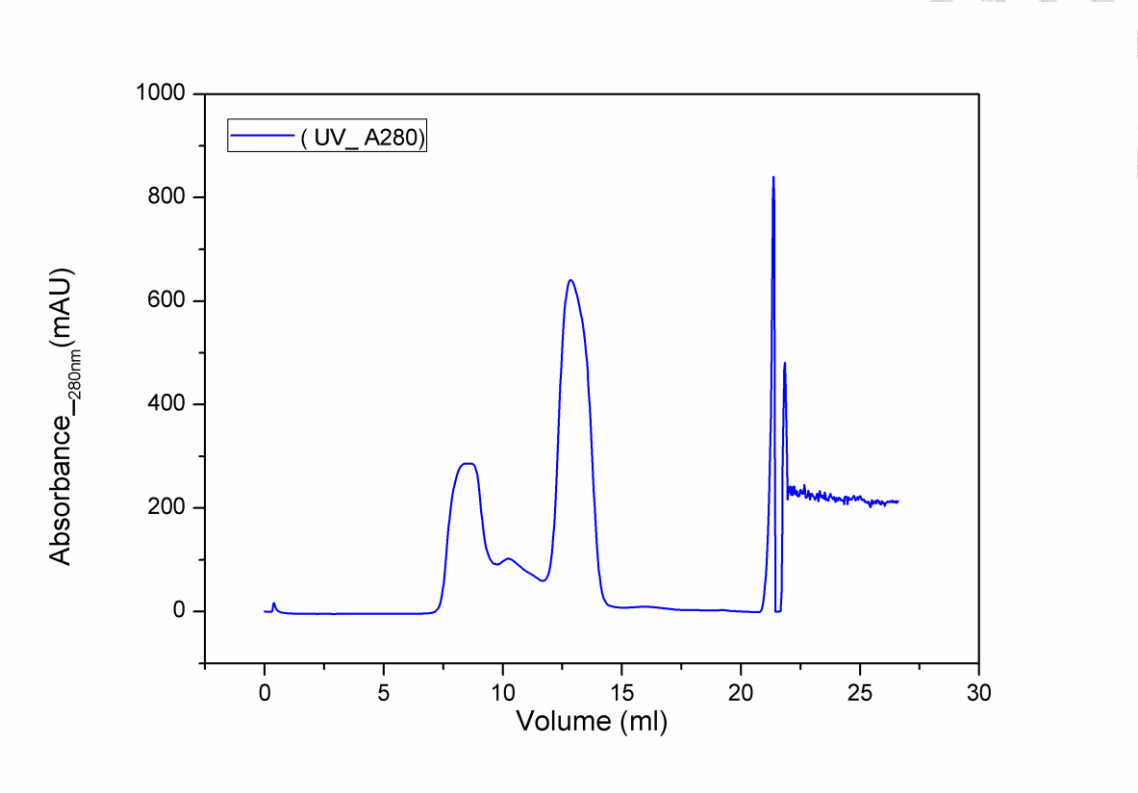


Figure S10. Purification of SUMO2K11-2 with SEC.
The flow-through after Ni^{2+}-NTA column in 50 mM Tris, and 200 mM NaCl was concentrated for FPLC with SEC column (Superdex 75 10/300 GL, GE Healthare). Blue line: the signal of UV_{A280}.

doi:10.6342/NTU201803256

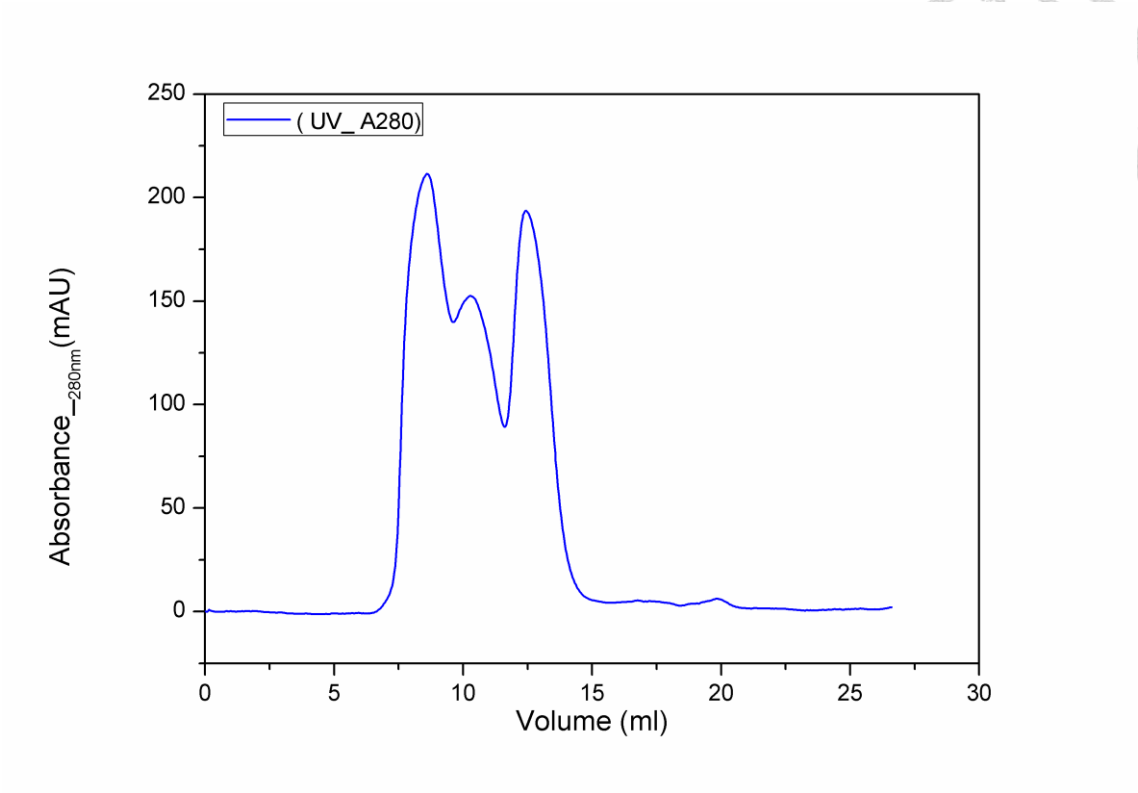


Figure S11. Purification of SUMO2K21-2 with SEC.
The flow-through after Ni^{2+}-NTA column in 50 mM Tris, and 200 mM NaCl was concentrated for FPLC with SEC column (Superdex 75 10/300 GL, GE Healthare). Blue line: the signal of UV_{A280}.

doi:10.6342/NTU201803256

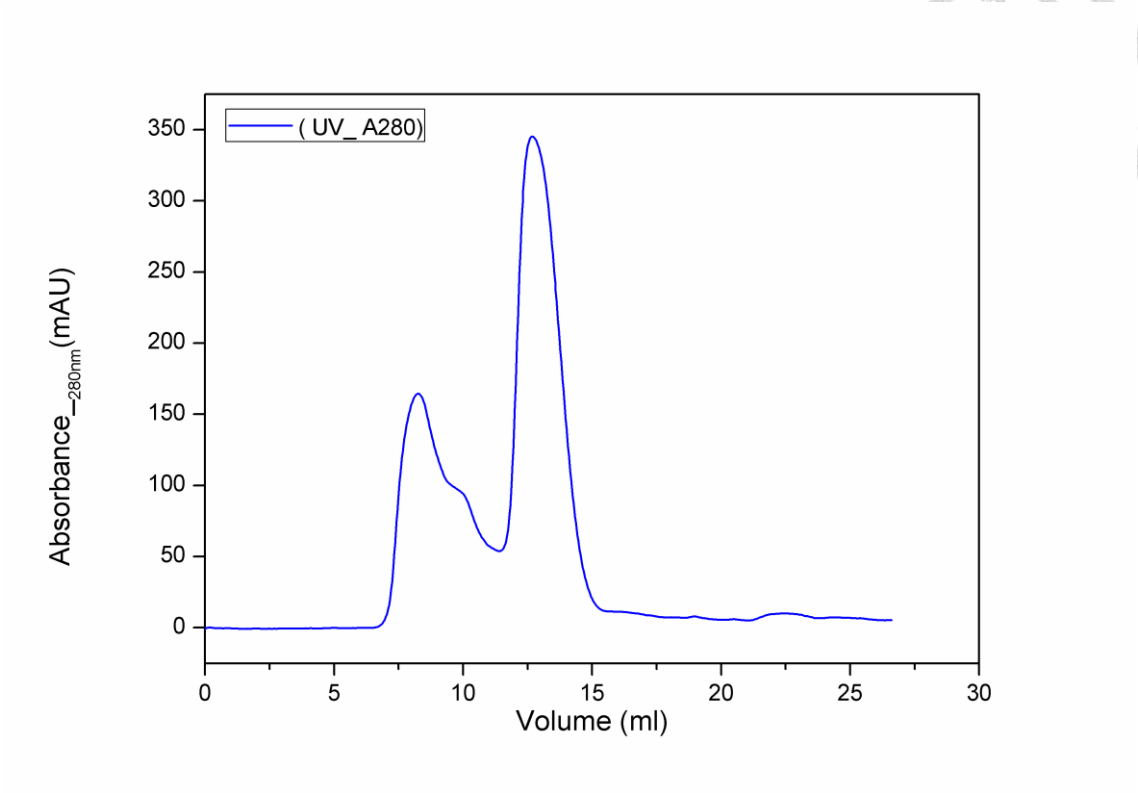


Figure S12. Purification of SUMO2K33-2 with SEC.
The flow-through after Ni^{2+}-NTA column in 50 mM Tris, and 200 mM NaCl was concentrated for FPLC with SEC column (Superdex 75 10/300 GL, GE Healthare). Blue line: the signal of UV_{A280}.

doi:10.6342/NTU201803256

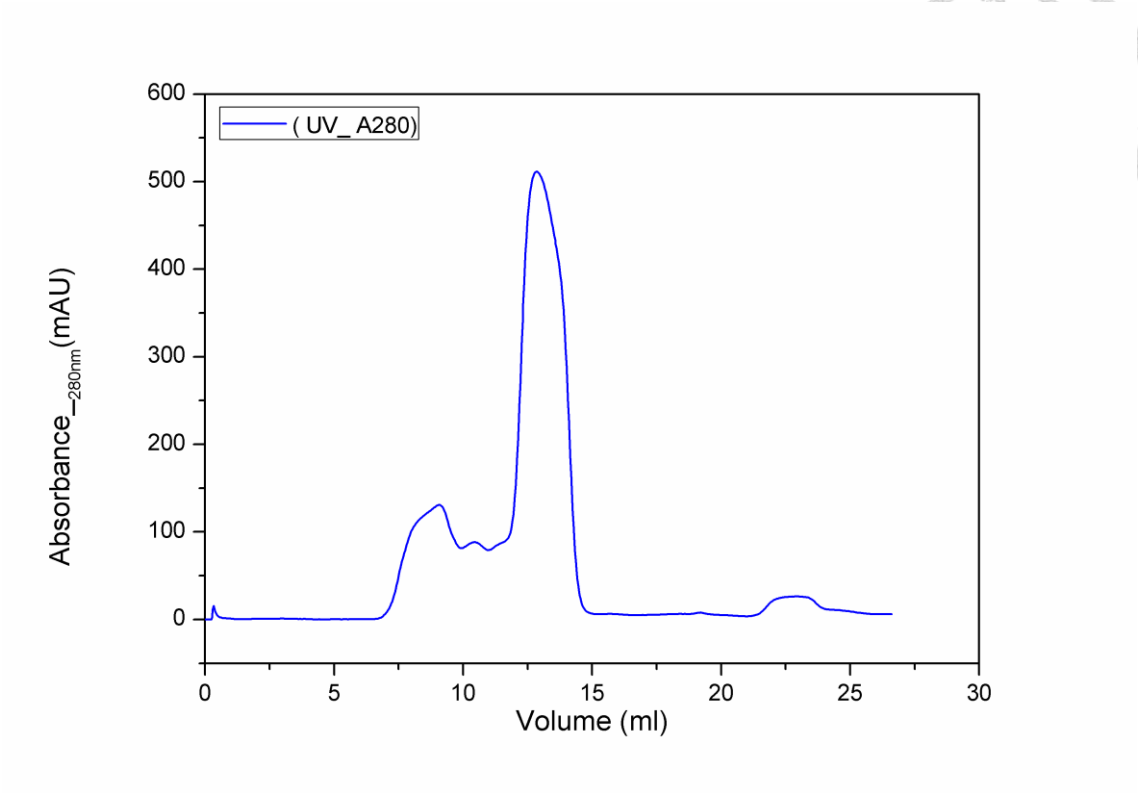


Figure S13. Purification of SUMO2K35-2 with SEC.
The flow-through after Ni^{2+}-NTA column in 50 mM Tris, and 200 mM NaCl was concentrated for FPLC with SEC column (Superdex 75 10/300 GL, GE Healthare). Blue line: the signal of UV_{A280}.

doi:10.6342/NTU201803256

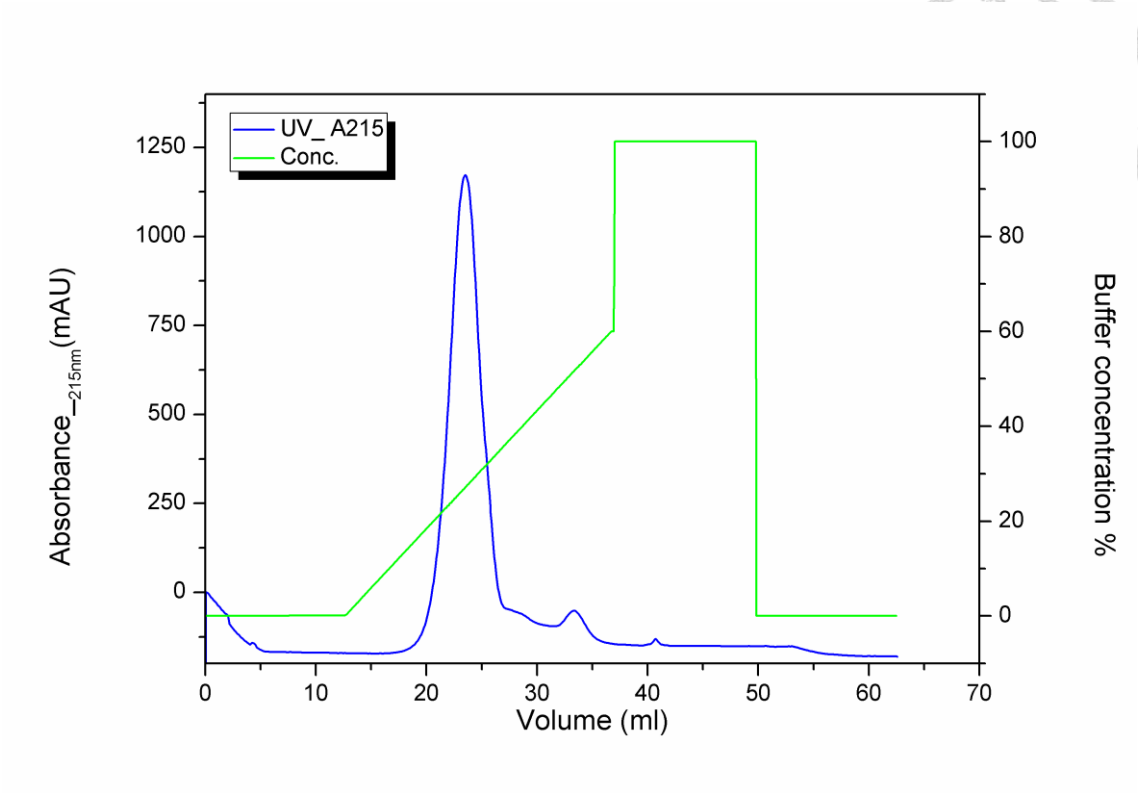


Figure S14. Purification of SUMO2K42-2 with the anion-exchange.
The flow-through after Ni^{2+}-NTA column was concentrated for FPLC with HiTrap Q HP column (GE Healthcare). Blue line: the signal of UV_{A215}; Green line: the concentration (%) of buffer with 1 M NaCl.

doi:10.6342/NTU201803256

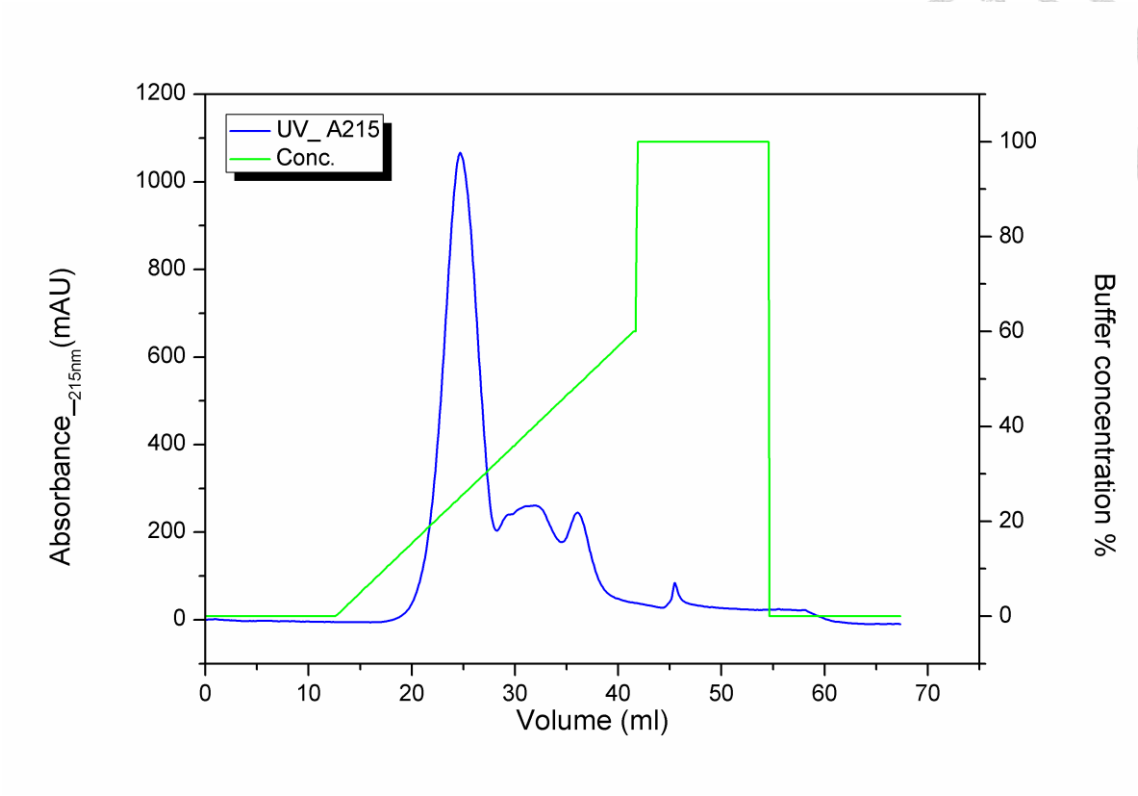


Figure S15. Purification of SUMO2K45-2 with the anion-exchange.
The flow-through after Ni^{2+}-NTA column was concentrated for FPLC with HiTrap Q HP column (GE Healthcare). Blue line: the signal of UV_{A215}; Green line: the concentration (%) of buffer with 1 M NaCl.

doi:10.6342/NTU201803256

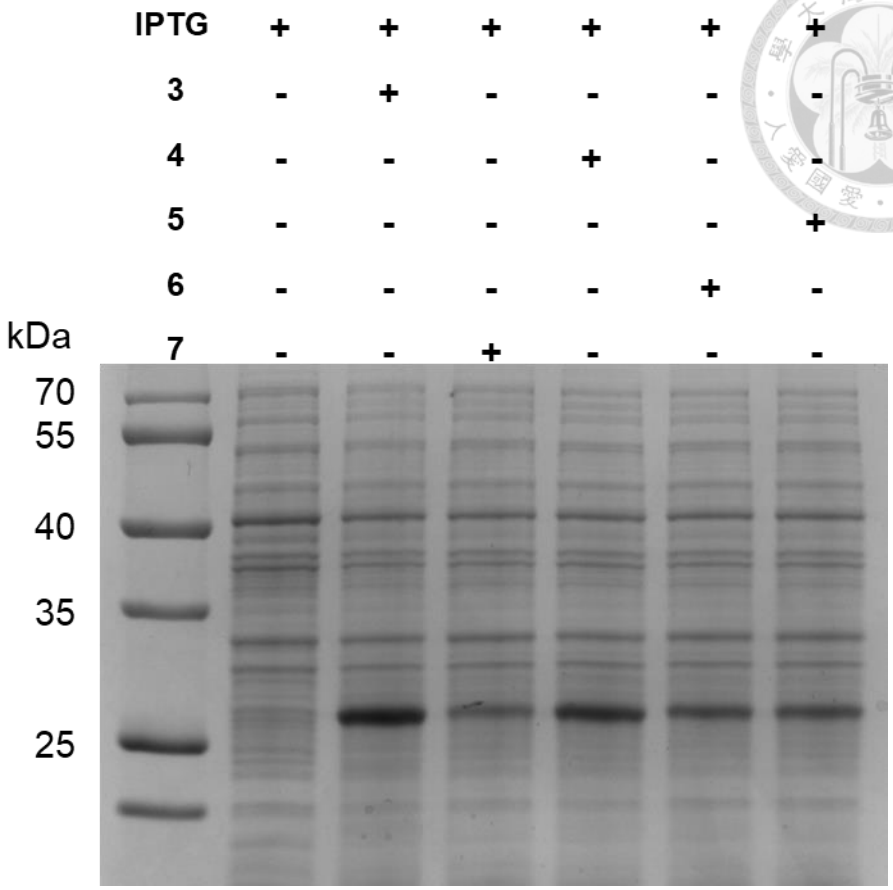


Figure S16. Amber suppression of different heterocyclic ncAAs 3-7 in sfGFP-Y66TAG.

E. coli BL21(DE3) cells coding FOWRS2•$tRNA^{Pyl}_{CUA}$ and a sfGFP gene with an amber codon (TAG) at position Y66 containing His-tag in 3' end and under control of an inducible T7 promoter were grown in M9 medium supplemented with 1 mM IPTG and 1 mM ncAAs. InstantBlue-Stained gel of whole cell over-expressing sfGFP-Y66TAG in the absence and in the presence **3-7**. First lane was negative control without ncAA and IPTG.

doi:10.6342/NTU201803256