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

設計新穎含金屬酵素催化胰島素轉譯後修飾

De novo design of a ferritin metalloenzyme for insulin modifications

鄒淳如

Chun-Ju Tsou

指導教授：王彥士 博士

Advisor: Yane-Shih Wang, Ph.D.

中華民國 108 年 7 月

July 2019

doi:10.6342/NTU201903019

誌謝

一年半的碩班生涯就這樣倏地飛過，快得令人措手不及，儘管時間總是公平地一天天推進，真正面臨畢業時卻感覺不太真實。當初進入台大生化所的過程有點意外，但絕非只有運氣，首先我要感謝指導教授王彥士老師，若沒有他的指引和協助，我不會有機會申請提早畢業，也無法順利完成這本論文，回想這過去一年半的訓練，不論在邏輯思辯還是實驗設計方面我都著實學了不少。我還要感謝蔡明道老師、陳長謙老師和俞聖法老師在百忙之中撥空參加我的口試。

在繁忙的實驗中，互相扶持的 WYS lab 絕對是我隱居南港兩年最強大的後盾，謝謝翰鍇解答我許多學業和實驗相關的疑難雜症，Prashant 擔任我的有機化學顧問以及英文會話家教，曼妮打點了實驗室的生活起居還有協助助理時期的我熟悉環境，去年畢業的伊慧和健隆則提供了許多過來人的經驗和建議，讓我在面對各個階段的考驗時感覺踏實許多，同舟共濟的戰友睦龍，沉穩中帶點幽默，是做實驗時的最佳夥伴，一起準備德檢 A2 的冠文，還有帶給實驗室歡樂的神奇寶貝大師若主和嘉呈。謝謝大家兩年來的包容和幫忙，我無法想像沒有你們，碩班生活會是怎樣的沉悶。我想我也會永遠記得寫論文期間，打球跑步游泳重訓什麼都來的日子，除了瘋狂我想不到別的形容詞。

最大的感謝我想要獻給我的家人，因為有他們無條件的支持與鼓勵，我才能安心地投入研究，家一直是我溫暖的避風港，謝謝你們一直都在。大學時期的朋友們是我暫時抽離苦悶研究生活的精神依靠，謝謝念縳，妳始終是我的心靈支柱，還有智堯，你曾是我每天向前的最大力量。謝謝芸霈、傘池和 B612 群組聽我的各種碎念，還有麒文以及校友團的大家，音樂真的是最撫慰人心的存在。

還有很多感謝不知道該如何表達，謹以本論文獻給所有曾幫助過我的人們，謝謝你們。

2019 年 7 月於中研院生化所

doi:10.6342/NTU201903019

摘要

隨著現代飲食精緻化，糖尿病的盛行率逐年提升。世界衛生組織統計全世界約超過 4.2 億患有糖尿病，臺灣在擁有 200 多萬糖尿病患者成因下，此疾病也成為 2018 年國人第五大死因。自 1922 年首次將胰島素作為藥物使用以來，胰島素一直是糖尿病患者最主要的治療方式，而為了因應不同的血糖調控需求，各式胺基酸序列改變及化學修飾的胰島素類似物也相繼研究開發。然而，由於胰島素的修飾受到了自然胺基酸側鏈官能基的蛋白質化學限制，此研究藉由設計新穎的金屬酶以增加胰島素類似物多樣性，並放眼未來用於活體內進行蛋白質轉譯後修飾的催化。

文獻研究已指出銅離子可作為路易斯酸用以催化傅-克烷基化，為了催化 α,β-不飽和羰基化合物與胰島素的芳香族胺基酸之間的專一反應，此研究改造人類重鏈攜鐵蛋白設計了新穎的含銅酵素。攜鐵蛋白是一種普遍存在於生物體內用於儲存和釋放鐵離子的蛋白質，屬於 24 蛋白單體自組裝的籠蛋白群組。在這此研究中，我們將胰島素受體羧基末端 α 螺旋的 αCT 片段連接於攜鐵蛋白末端，並利用吡咯-轉核糖核酸合成酶•$tRNA^{Pyl}$ 突變株配對將非典型胺基酸嵌入籠狀攜鐵蛋白的 C_2 界面，建構新的金屬螯合位。經由蛋白質譜分析，新型攜鐵蛋白球在螯合銅離子下，專一地對胰島素的 B 鏈的 His^{B5} 位點進行位置特異的烷基化修飾。總結本研究，新型含銅攜鐵蛋白球酵素具有胰島素特定位置修飾功能，可以催化胰島素 B 鏈上 His^{B5} 的傅-克烷基化反應，球體外露 αCT 胜肽鏈和胰島素的親合鍵結設計也能增進傅-克烷基化的效率。

關鍵字: 鐵蛋白、生物催化、含金屬酵素、胰島素類似物、非典型胺基酸

doi:10.6342/NTU201903019

Abstract

According to WHO, more than 422 million people are diagnosed with diabetes around the world, and in Taiwan, 2 million people are suffering from diabetes. Since insulin is the major treatment to those patients, development of efficient insulin drugs is of great significance. However, modifications of insulin analogues are limited by the deficiency of protein chemistry, therefore our goal is to construct a post-translational modification (PTM) enzyme aimed to explore new protein chemistry and enrich the chemical diversity of insulin.

Previous literature has performed Friedel-Crafts alkylation using copper ion, however, without specificity. To catalyze reactions between small molecules with α,β-unsaturated carbonyl and aromatic amino acids on insulin, a metalloprotein based biocatalysis platform using human heavy chain ferritin was designed. Ferritin is a family of self-assembled 24-subunits protein cages who act as antioxidants by storing and releasing irons. In this study, a ferritin-αCT fusion protein containing the insulin receptor carboxy-terminal α-chain (αCT) segment on its C-terminus was constructed. Non-canonical amino acids (ncAAs), L-2-(5-Bromothienyl)alanine, H-*N*-3-Methyl-L-histidine, and 3-(4-Thiazolyl)-L-alanine, are then incorporated at the C_2 interface by evoluted pyrrolysyl-tRNA synthetase (PylRS)•$tRNA^{Pyl}$ pair, followed by the introduction of Cu(II) after iron ion removal. With evidences of protein mass spectrometry analysis, engineered ferritin-αCT variants have demonstrated specificity toward insulin modification by catalyzing Friedel-Crafts alkylation between the first histidine on B chain and diethyl ethylidenemalonate (DEEM).

Keywords: Ferritin, Biocatalysis, Metalloenzyme, Insulin analogs synthesis, non-canonical amino acids

doi:10.6342/NTU201903019

Table of contents

doi:10.6342/NTU201903019

doi:10.6342/NTU201903019

doi:10.6342/NTU201903019

List of schemes

doi:10.6342/NTU201903019

List of tables

doi:10.6342/NTU201903019

List of figures

doi:10.6342/NTU201903019

doi:10.6342/NTU201903019

Abbreviations

α-F	αCT-ferritin
Ala, A	L-alanine
Amp	ampicillin
Arg, R	L-arginine
Asn, N	L-asparagine
Asp, D	L-aspartic acid
BA	benzalacetone
BrThA	L-2-(5-Bromothienyl)alanine
Cys, C	L-cysteine
CV	column volume
Da	dalton
DEBM	diethyl benzylidenemalonate
DEEM	diethyl ethylidenemalonate
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DTT	1,4-dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EF-Tu	elongation factor thermo unstable
ER	endoplasmic reticulum
ESI-MS	electrospray ionization mass spectrometry
Ftn	human heavy chain ferritin
F-α	ferritin- αCT

doi:10.6342/NTU201903019

Glu, E	L-glutamic acid
His, H	L-histidine
ICP-MS	inductively coupled plasma mass spectrometry
IGF-1R	insulin-like growth factor-1 receptors
IPTG	isopropyl β-D-1-thiogalactopyranoside
IRS	insulin receptor substrate
LB	Luria-Bertani broth
Leu, L	L-leucine
Lys, K	L-lysine
MALDI-TOF-MS	matrix-assisted laser desorption ionization-time of flight mass spectrometry
MeH	H-His(3-Me)-OH
*Mm*PylRS	*Methanosarcina mazei* pyrrolysyl-tRNA synthetase
mRNA	messenger ribonucleic acid
ncAA	non-canonical amino acid
NPH	neutral protamine Hagedorn
NTA	nitrilotriacetic acid
PBA	phenylboronic acid
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
Phe, F	L-phenylalanine
PLGA	poly(lactic-co-glycolic acid)
Pro, P	L-proline
Ser, S	L-serine

doi:10.6342/NTU201903019

SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel
Sp	spectinomycin
TEM	transmission electron microscope
TEMED	tetramethylethylenediamine
Thr, T	L-threonine
ThzA	3-(4-Thiazolyl)-L-alanine
TLC	thin layer chromatography
TON	turnover number
tRNA	transfer ribonucleic acid
Tropone	2-chloro-2,4,6-cycloheptatrien-1-one
Trp, W	L-tryptophan
Tyr, Y	L-tyrosine

doi:10.6342/NTU201903019

Chapter 1 Introduction

1-1 Diabetes and insulin

The global prevalence of diabetes is on the rise. Globally, more than 4 million of people are diagnosed with diabetes, comprising 8.5% of the adult population (**Figure 1**). In Taiwan, the increasing prevalence is also inevitable, there are more than 2 million diabetes patients as reported by the Ministry of Health and Welfare[1]. Diabetes causes 1.6 million deaths annually, and various complications such as cardiovascular disease, blindness, and kidney failure, and imposes economic burdens on patients due to high medical expenses. Moreover, it was estimated by World Health Organization (WHO) that diabetes is the seventh leading cause of death in 2016[2].

There are two main types of diabetes, type 1 and type 2. While the exact cause of type 1 diabetes remains unknown, type 2 diabetes is largely preventable, which is mostly resulted from unhealthy lifestyles and accounts for major population of diabetes patients.

1-1-1 Type 1 and type 2 diabetes

Type 1 diabetes, also known as juvenile diabetes, is caused by autoimmune reaction and is not preventable nowadays. The pathogenesis involves both genetic and environmental factors and mainly occurs in children and adolescents. Patients with type 1 diabetes require daily insulin administration for blood glucose management due to their insufficient production of insulin.

Type 2 diabetes is caused by the ineffective use of insulin, making up 90% of diabetes patients. The risk factors including aging, obesity, and unhealthy lifestyles, however, it is often preventable through reducing overweight and healthy eating. The key feature of type 2 diabetes, insulin resistance, leads to lower glucose uptake and higher

doi:10.6342/NTU201903019

blood glucose level. Common oral antidiabetic drugs (**Table 1**) either increase insulin sensitivity or secretion, however, if non-insulin monotherapy couldn't manage blood glucose level, insulin administration will be necessary.

1-1-2 Types of insulin analogues

Insulin is a peptide hormone comprising 51 amino acids with A chain and B chain constrained by 1 intrachain and 2 interchain disulfide bonds (**Scheme 1**), it is usually secreted under high blood glucose levels. The production of insulin is involved in posttranslational modifications, it is first translated into preproinsulin to speed up the transfer and translocation within endoplasmic reticulum (ER), upon cleavage of signal peptide, proinsulin is formed as a single polypeptide. The A and B chains are connected via C-peptide who will then undergo efficient folding and synthesis of disulphide bonds[3] to generate insulin.

Insulin has been a life-saving treatment for nearly nine decades, with the progress of biotechnology, the development of insulin analogues in recent years provides diabetes patients with different choices. According to action profile, insulin analogues can be divided into 4 types; rapid-acting insulin, short-acting insulin, intermediate-acting insulin, and long-acting insulin (**Table 2**).

Rapid-acting insulin, also called mealtime insulin, is designed to disrupt the self-association of insulin to speed up absorption. The sequences of three common rapid acting insulin analogues are slightly different from native insulin. Insulin lispro has an inversion of $Pro^{B28}Lys^{B29}$ native sequence. Pro^{B28} is replaced by Asp in insulin aspart, Asn^{B3} and Lys^{B29} are substituted by Lys and Glu respectively in insulin glulisine. All of them perform earlier time action profiles compared to human insulin[4].

Short-acting and intermediate-acting insulin are both based on human recombinant

doi:10.6342/NTU201903019

insulins. Regular insulin, known as neutral insulin and soluble insulin, is a type of short-acting insulin produced in *E. coli* by recombinant DNA technology. NPH, abbreviation for neutral protamine Hagedorn, has a long history of insulin development since it was first created in 1946. NPH insulin is a cloudy solution containing the mixture of regular insulin and protamine in exact proportions with zinc and phenol addition, and it is able to maintain neutral-pH and sustain basal insulin action.

Long-acting insulin, or basal insulin, is intended to mimic the steady secretion profile of health pancrease[5], therefore, it has rather flat pharmacodynamics profile and achieves a better level of glycemic control. Current basal insulin either rely on decreased solubility or prolonged absorption to perform sustained actions. Insulin glargine reduce its solubility at physiological pH by shifting isoelectric point, which will generate precipitates at the injection site to make a slower release. However, insulin detemir and insulin degludec use chemical modifications to attain longer glucose-lowering effect.

1-1-3 Chemical modifications of insulin analogues

As mentioned above, both insulin detemir and insulin degludec are chemically modified. Insulin detemir (**Scheme 2A**) has a fatty acid acylated on its B chain terminal lysine designed to facilitate the dihexamer formation upon injection. It will then non-covalently bind to albumin and join plasma circulation, leading to delayed absorption within injection site depot[6]. Likewise, insulin degludec (**Scheme 2B**) has fatty acid acylation on the same amino acid residue and performs self-association feature. However, the glutamic acid spacer stabilizes the insulin degludec dihexamer form and arrange itself into stable multi-hexamer chain as phenol slowly releases. As zinc ions inside multihexamers start to diffuse, insulin degludec slowly dissociates into dimers and readily-absorbed monomers[7].

doi:10.6342/NTU201903019

From the design of insulin detemir and degludec, it can be summarized that fatty acid chain can prolong the glucose-lowering effect. Beside abovementioned commercial insulin analogues, several insulin therapies under study are also using chemical modification approach to achieve various properties. For example, PBA insulin (**Scheme 3A**) possesses a glucose-responsive phenylboronic acid functional group and an aliphatic domain to promote hydrophobic interactions. The glucose-mediated activation allows PBA insulin to demonstrate the self-administration feature[8]. The other insulin in development, AB101 (**Scheme 3B**), is a refined insulin designed for once a week subcutaneous injection. It is PEGylated at its N-terminus of B chain, the 5 kDa polyethylene glycol (PEG) together with poly(lactic-co-glycolic acid) (PLGA) will encapsulate insulin and emulsify slow release microspheres[9].

1-1-4 How insulin binds to its receptor

Insulin receptor is a disulfide-linked $(\alpha\beta)_2$ homodimer highly homologous to insulin-like growth factor-1 receptors (IGF-1R), and all of them are members of transmembrane tyrosine kinase receptors. Upon insulin binding, the tyrosine kinase and insulin receptor substrate (IRS) proteins will be phosphorylated and PI3K-Akt pathway is thereby activated. The Akt regulates most of insulin metabolic effects, glucose transport, lipid synthesis, gluconeogenesis, and glycogen synthesis (**Figure 2**).

However, if insulin resistance happens, cells fail to respond to insulin properly, which will then cause hyperglycemia and reduce glucose uptake in insulin-sensitive tissues. The molecular mechanisms include numerous reasons, yet it can be briefly concluded that insulin resistance, the key feature of type 2 diabetes, is often triggered by cellular perturbation such as lipid toxicity and inflammation. Those causes affect the phosphorylation of insulin receptor or insulin-signaling molecules by activating Ser/Thr

doi:10.6342/NTU201903019

kinases and finally lead to impaired insulin action[10].

Structure of insulin receptor is basically composed of ectodomain, transmembrane domain and tyrosine kinase domain (**Figure 3**); the ectodomain is made up of six subdomains (L1, CR, L2, FnIII-1, FnIII-2 and FnIII-3), additionally, an insert domain resides within FnIII-2. The terminal segment of insert domain αCT is previously reported as the primary binding site of insulin[11-12]. Insulin has two binding sites due to its interaction with receptor on both sides, the first binding site, which is involved in insulin-dimerizing residues, comprised of αCT segment and part of L1 domain of the other α-chain. The second binding site is located within FnIII domain and it primarily interacts with insulin-hexamerizing residues. The αCT segment from primary binding site is 16 amino acid residues in length and directly contacts both insulin A and B chains, moreover, it is engaged in the conformational change upon ligand binding especially insulin residues B26-30.

To measure the binding activity of insulin in lab, commercial assay is also available. For example, a cell based assay for quantification of insulin activity, product of Svar Life Science, is based on avian B-cell line DT-40 transfected with human insulin receptor genes, the luciferase reporter gene controlled by constitutive promoter is serving as the normalization of cell counts and serum matrix effects[13].

1-2 Metalloprotein and biocatalysis

Biocatalysis has appeared to be one of most powerful tools for optimizing chemical reactions. Beside academic researchers, pharmaceutical industries are also working on developing economic, sustainable, and efficient enzymes. As a main character of biocatalysis, enzyme demonstrates high specificity, enantioselectivity, and catalytic efficiency, those features have become the driven forces of protein engineering.

doi:10.6342/NTU201903019

When it comes to biocatalysis, the development of artificial metalloproteins should be mentioned, the prevalence can attribute to their redox catalysis function in nature. The inherent features provide an amenable platform for metal ion chelation engineering and directed evolution, however, recent advances in biocatalysis are not limited to metalloproteins, much progress have been made in the functional design of non-catalytic proteins.

1-2-1 Engineering of metalloproteins

Metalloproteins constitute nearly half of proteins in nature and take part in various biochemical reactions including photosynthesis and respiration. The ultimate goal of protein engineering is to create an artificial metalloprotein based on those structural features.

Generally, the design can be divided into 3 groups[14]. In the first group, metal-binding sites are introduced to protein scaffold to bind metal ions. One of de novo catalytic proteins containing a di-iron cofactor is reported to catalyze the two-electron oxidation of 4-aminophenol[15].

The method of second group is based on the re-design of existed metal-binding site combined with site directed mutagenesis, heme proteins are therefore widely used because of their diverse functions. A series of heme proteins including cytochrome c, cytochrome P450, and myoglobin have been successfully engineered into powerful metalloenzymes capable of catalyze various chemical reactions, for example, enantioselective C-H amination[16], C-Si bond formation[17], and synthesis of highly strained carbocycles[18].

The third group is to design unnatural metalloproteins by means of ncAAs incorporation or the introduction of unnatural metal-containing cofactors. For example,

doi:10.6342/NTU201903019

incorporating metal-binding amino acid (2,2'-bipyridin-5yl)alanine on the dimer interface of transcription factor Lactoccocal multidrug resistance Regulator, modifying the protein into a novel Cu(II)-binding metalloenzyme, furthermore, the artificial enzyme is able to catalyze enantioselective Friedel-Crafts reactions[19].

1-2-2 Protein modifications of selective biocatalysis

Beside small-molecule catalysis, protein modifications based on transition metals are commonly seen in protein modifications[20]. For example, π-allylpalladium, a type of electrophilic organometallic complexes, are developed to modify tyrosine on its hydroxyl group[21]. Another research uses a SH3-binding peptide coupled with rhodium(II) to generate an artificial metallopeptide, the novel enzyme is able to site-specifically modify different residues on kinase SH3 domain under different pH values[22].

In this study, the concept of metalloprotein design and site-specific protein modifications are combined to carry out Friedel-Crafts alkylation using ferritin as protein scaffold. Similar to traditional catalysts such as $AlCl_3$, $FeCl_3$, our newly designed metalloenzyme is also acting as a Lewis acid with ions coordinated to metal-binding sites to catalyze Friedel-Crafts alkylations[23] between insulin and α,β-unsaturated carbonyl group.

1-3 Properties of ferritin

1-3-1 Biological functions of ferritin

Ferritin is a self-assembled protein cage with an outer diameter of 12 nm and an inner diameter of 8 nm, it plays important roles in iron storage and release. The ferritin superfamily comprises three subfamilies: classical ferritin, bacterioferritin, and ferritin-like DNA-binding protein (Dps), The former two are maxiferritin while the later belongs

doi:10.6342/NTU201903019

to miniferritin[24]. Since classical ferritin is the most common type and widely exists in living kingdom, the following research will be focused on classical ferritin. Ferritin is also widely used as a carrier for drug delivery[25] due to its natural existence in human body, which makes it an optimal choice for reducing endogenous immunological responses.

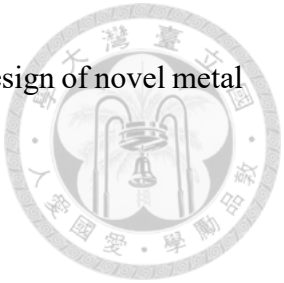
In mammals, ferritin contains two distinct subunits, heavy chain (H) and light chain (L), heavy chain ferritin possesses catalytic domain while light chain is deficient in oxidoreductase activity. The oxidoreductase activity site on each ferritin subunit is capable of catalyzing oxidation of iron(II) for the storage of iron mineral $Fe_2O_3 \bullet H_2O$ in its cavity to prevent Fenton reactions[26], moreover, it can accommodate up to 4500 iron atoms. In vertebrates, ferritin exists as heteropolymer and the ratio of H:L is tissue specific. Ferritin rich in heavy chain is mostly found in muscle and heart tissues since oxidoreductase is required for iron metabolism, while tissues responsible of iron storage such as liver and spleen favor light chain ferritin. In this research, human heavy chain ferritin will be studied in the following experiments.

1-3-2 Structural and symmetrical features of ferritin

Ferritin consists of 24 subunits and is characterized by having 432 (octahedral) point symmetry with 12 C_2 axes, 8 C_3 axes, and 6 C_4 axes (**Figure 4**). The structure of ferritin subunit consists of 5 α-helices (A-E) and a long loop between B and C helix. The N-terminus of the subunit is located near the C_3 axis pore and the C-terminus is pointing inward the C_4 channel (**Figure 5**). C_4 channel is arranged by 4 helix E and the external domain is highly hydrophobic; C_3 channel, however, is hydrophilic and defined as the entry of iron(II)[27]; C_2 interface covers the largest area of ferritin ball among all symmetries, a research therefore took the advantage of this feature and redesign the interface by mutated L56, R63, and E67 to histidine, created a copper-induced ferritin

doi:10.6342/NTU201903019

assembly[28]. In my research, those mutations will be brought into the design of novel metal chelating sites.



1-4 Incorporation of non-canonical amino acids

1-4-1 Protein translation

Proteins are synthesized through a highly conserved process called translation, it is carried out by which the sequence of mRNA is read in the 5' to 3' direction to generate polypeptide chains who will then fold into functional proteins. Translation is taken place in ribosome where mRNA codons are decoded into amino acids with the help of complementary aminoacyl-tRNA. Overall, the process of translation can be divided into three phases, initiation, elongation, and termination.

The initiation in prokaryotes requires two ribosomal subunits (50S and 30S), translation initiation regions of the mRNA, initiator $tRNA_{fmet}$, and initiation factors IF1, IF2, and IF3. The 70S ribosome, composed of 50S and 30S, has three tRNA binding sites so called the aminoacyl (A), peptidyl (P), and exit (E) sites. The initiation begins with the binding of three initiation factors to the 30S subunit which will then bind to the Shine-Dalgarno sequence upstream of the AUG initiation codon. Initiator $tRNA_{fmet}$ later joins the complex accompanied by the release of IF3. After then, a 50S ribosomal subunit associates with the complex, leading to the hydrolysis of GTP and the release of IF1 and IF2, and finally, the 70S initiation complex is formed.

The next step, elongation, starts from the binding of initiator $tRNA_{fmet}$ to P site, the subsequent conformational change facilitates the pairing between next aminoacyl tRNA and the second codon of the mRNA. The aminoacyl tRNA is brought to ribosome by elongation factor EF-Tu and GTP. As aminoacyl tRNA enters the A site, GTP is hydrolyzed and EF-Tu is released. Once EF-Tu left, peptide bonds are formed between

doi:10.6342/NTU201903019

amino acids in A and P sites. During translocation, the peptidyl-tRNA is transferred from the A site to the P site, and the uncharged tRNA will finally release from the E site as the new aminoacyl tRNA binds.

The translation of polypeptide continues until ribosome reaches a stop codon (UAA, UGA, or UAG). The stop codons are recognized by release factors RF1 and RF2 and their binding in the A site triggers the release of polypeptide chain in the P site, followed by the dissociation of tRNA and mRNA from the ribosome[29].

1-4-2 Expanding genetic code

In nature, proteins are built from 20 amino acids specified by 64 degenerate codons including 3 stop codons (UAG, UAA, and UGA). The codons were thought to be static in the past, however, they were found flexible along with the discovery of several exceptions. For example, the 21st amino acid selenocysteine who recognizes UGA codon[30], and pyrrolysine, the 22nd amino acid found in *Methanosarcina barkeri* genetically encoded by UAG codon[31].

Lots of effort have been made in the exploration of genetic codes, in 1989, ncAA was specifically incorporated into proteins by using chemically acylated suppressor tRNA coupled with a nonsense codon TAG, however, with low yield[32]. To further optimize the system, an in vitro biosynthetic method had been developed. A tyrosyl-tRNA synthetase/ $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ pair derived from *Methanococcus jannaschii* was imported to *E. coli*, and since the cross-species reaction is not efficient enough, an orthogonal aminoacyl-tRNA synthetase•tRNA pair (MjTyrRS•$Mj\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$), in *E. coli* was thereby generated[33]. This unique pair is able to incorporate *O*-methyl-L-tyrosine in response to an amber codon TAG. Although the MjTyrRS• $Mj\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ works well in *E. coli*, non-specific aminoacylation caused by endogeneous synthetase was observed because of imperfect

doi:10.6342/NTU201903019

binding pocket[34].

In 2002, pyrrolysyl-tRNA synthetase (PylRs)•$tRNA_{CUA}^{Pyl}$was identified; this pair was derived from *Methanosarcina* species that encodes pyrrolysine genetically. PylRS and its evolved variants are suitable for the incorporation of lysine, phenylalanine, tyrosine, histidine, and tryptophan derivatives while *Mj*TylRS•*Mj*$tRNA_{CUA}^{Tyr}$ pair installs tyrosine, phenylalanine, and cysteine analogues[35].

Expanding genetic code (**Figure 6**) has been a powerful tool since its discovery in 1980s. It is now widely applied to the study of protein chemistry[36], for example, post-translational modifications[37-40], protein-protein interactions[41], and fluorescently labeled proteins[42-43]. In this study, *Methanosarcina mazei* PylRS (*Mm*PylRS)•$tRNA_{CUA}^{Pyl}$ pair was engineered to incorporate three histidine analogues for Cu(II) chelating[44-46].

1-5 Specific aim

Among type 2 diabetes treatments, refined basal insulins have been reported to be one of future prospects. Though chemical modifications mentioned above indeed increase insulin varieties, however, most of them are limited to similar amino acid residues, for example, lysine. In order to enrich the diversity of insulin analogues, an enzyme that specifically catalyzes Friedel-Crafts alkylation on aromatic amino acids on insulin is proposed and it will also be applied to different protein modifications in the future (**Figure 7**).

The design of catalysis platform is based on a protein scaffold called ferritin, which is consisted of 24 self-assembled subunits. The inherent iron ions storage feature makes it an ideal subject for metalloenzyme construction, a Cu(II) binding ferritin with catalytic power can be generated by using expanding genetic code to incorporate metal-chelating

doi:10.6342/NTU201903019

ncAAs. The biocompatibility of ferritin is also a great concern for applying the platform in human body to extend the lifetime of insulin. Ferritin C_2 interface was engineered into novel metal binding sites and since single C_2 interface is capable of holding two Cu(II), 12 C_2 interfaces would supposedly chelate 24 Cu(II) and catalyze reactions. To further enhance the specificity toward insulin, a αCT peptide derived from insulin receptor primary binding site was also fused to ferritin C-terminus, and the interactions will be measured by biophysical characterizations.

In this study, the novel metalloenzyme is serving as a biocatalysis platform for the study of Friedel-Crafts alkyation on insulin. Friedel-Crafts alkylation is an electrophilic aromatic substitution reaction commonly based on Lewis acid catalysts in which copper complex is widely used, the Cu(II) binding ferritin was also a Lewis acid catalyst. Small molecules with α,β-unsaturated carbonyl group **2-5** were proposed to react with aromatic amino acids on insulin (**Scheme 4**), especially histidine. According to previous studies[47], four aromatic amino acids were potential candidates that will react with α,β-unsaturated carbonyl group under Cu(II) catalysis, however, since insulin doesn't have tryptophan, the study will be focused on the phenylalanine (B1,B24, and B25), tyrosine (A14, A19, B16, and B26), and histidine (B5 and B10).

doi:10.6342/NTU201903019

Chapter 2 Materials and methods

2-1 DNA and protein sequences

2-1-1 DNA sequences

$\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Pyl}}$:

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

*Mm*PylRS:

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaaataa

aacaccacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagc

agcaggactgcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctc

aataagttcctcacaaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaa

aaggcaatgccaaaatccgttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggat

ctaaattttcacctgcgataccggtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttct

acaggagcaactgcatccgcactggtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgc

ccccgcacttacgaagagccagactgacaggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggca

agcctttcagggagcttgagtccgaattgctctctcgcagaaaaaaagacctgcagcagatctacgcggaagaaaggga

gaattatctggggaaactcgagcgtgaaattaccaggttctttgtggacaggggttttctggaaataaaatccccgatcctg

atccctcttgagtatatcgaaaggatgggcattgataatgataccgaactttcaaaacagatcttcagggttgacaagaact

tctgcctgagacccatgcttgctccaaacctttacaactacctgcgcaagcttgacagggccctgcctgatccaataaaaat

ttttgaaataggcccatgctacagaaaagagtccgacggcaaagaacacctcgaagagtttaccatgctgaacttctgcc

agatgggatcgggatgcacacgggaaaatcttgaaagcataattacggacttcctgaaccacctgggaattgatttcaag

atcgtaggcgattcctgcatggtctatggggataccttgatgtaatgcacggagacctggaactttcctctgcagtagtcgg

acccataccgcttgaccgggaatggggtattgataaaccctggataggggcaggtttcgggctcgaacgccttctaaagg

ttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactataacgggatttctaccaacctgtaa

doi:10.6342/NTU201903019

Ferritin heavy chain (*Homo sapiens*):

atgaccaccgcctctacctcacaggtgcgtcagaattatcatcaggatagtgaagcagcaattaatcgccagattaatctg

gaactgtatgcaagctatgtgtatctgtctatgagctattattttgatcgcgatgatgttgccctgaaaaattttgccaaatatttt

ctgcatcagtctcatgaagaacgcgaacatgccgaaaaactgatgaaattacagaatcagcgtggtggtcgtatttttcttc

aagatattaaaaaaccggattgtgatgattgggaaagcggcctgaatgcgatggaatgtgccttacatcttgagaaaaat

gttaatcagtcactgctggaactgcataaactggcaaccgataaaaatgatccgcatctgtgtgattttattgaaacccatta

tctgaatgagcaggttaaagccattaaagaactgggcgatcatgttaccaatctccgcaaaatgggcgccccggaaagt

ggcttagccgaatatctgtttgataaacataccttaggcgatagcgataacgaaagtgagctc

3XG4S linker:

agtggtggtggtggtagcggaggggggaggaagcggaggaggaggcagc

αCT:

acgtttgaggattacctgcacaacgtggttttcgtccccaggccatct

2-1-2 Protein sequences

*Mm*PylRS:

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSS

RTARALRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKA

MPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGAT

ASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFREL

ESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERM

GIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKE

SDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVYGD

TLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRA

doi:10.6342/NTU201903019

ARSESYYNGISTNL

Ferritin heavy chain (*Homo sapiens*):

MTTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKNFA

KYFLHQSHEEREHAEKLMKLQNQRGGRIFLQDIKKPDCDDWESGLNAMECAL

HLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVKAIKELGDHVTNLRK

MGAPESGLAEYLFDKHTLGDSDNES

3XG4S linker:

SGGGGSGGGGSGGGGS

αCT:

TFEDYLHNVVFVPRPS

2-2 Plasmid design

2-2-1 List of primers

Entry	*Name*	*Sequence (5' to 3')*
1	pET-h-Ferritin-NdeI-F	gatatacatatgaccaccgcctctacctcacaggtgc
2	pET-h-Ferritin-SacI-R	ggtgatggagctcactttcgttatcgctatcgcctaag
3	pET-alphaCT-hFerritin-NdeI-F1	gagatatacatatgacgtttgaggattacctgcacaacgtggttttcg
4	alphaCT-3XG4S-R1	accaccaccactagatggcctggggacgaaaaccacgttgtg
5	3XG4S-F2	agtggtggtggtggtagcggaggggggaggaagcggaggaggaggcagc
6	3XG4S-hFerritin-R2	ctgtgaggtagaggcggtggtgctgcctcctcctcc
7	hFerritin-F3	accaccgcctctacctcacaggtgcgtcagaattat

doi:10.6342/NTU201903019

8	hFerritin-R1	actttcgttatcgctatcgcctaaggtatg
9	hFerritin-3XG4S-F1	agcgataacgaaagtagtggtggtggtggtagcggaggggggagga
10	3XG4S-alphaCT-R2	gtaatcctcaaacgtgctgcctcctcctccgcttcctccccctccgct
11	alphaCT-F2	acgtttgaggattacctgcacaacgtggttttcgtccccagg
12	pET-hFerritin-alphaCT-SacI-R3	ggtgatggagctcagatggcctggggacgaaaac
13	hFerritin-H60TAG /R63TAG-R	cggcatgttcctattcttcctaagactgatgcag
14	hFerritin-H60TAG/ R63TAG-F	cagtcttaggaagaataggaacatgccgaaaaactg
15	hFerritin-H60TAG/ E67TAG-F	gaagaacgcgaacatgcctagaaactgatg
16	h-Ferritin-H60TAG-R	catgttcgcgttcttcctaagactgatgcag
17	hFerritin-R63TAG/ E67TAG-R	cagtttctaggcatgttcctattcttcatgagac
18	hFerritin-R63TAG/ E67TAG-F	gaagaataggaacatgcctagaaactgatgaaatta
19	hFerritin-linker-F	gatagcgataacgaaagtagtggtggtggtggtagc
20	hFerritin-2XG4S-R	atcctcaaacgtgcttcctccccctccgctaccaccacc
21	hFerritin-1XG4S-R	gtgcaggtaatcctcaaacgtgctaccaccaccaccac
22	alphaCT-2XG4S-R	tccccctccgctagatggcctggggacgaaaaccacgttg
23	alphaCT-1XG4S-R	gctagatggcctggggacgaaaaccacgttgtg
24	alphaCT-1XG4S-F	cgtccccaggccatctagcggaggaggaggcagc
25	2XG4S-F	agcggaggggggaggaagcggaggaggaggcagc
26	h-Ferritin-R63TAG-F	cagtctcatgaagaataggaacatgccgaaaaac
27	h-Ferritin-R63TAG-R	gtttttcggcatgttcctattcttcatgagactgatg

2-2-2 Plasmids

Plasmid abbreviation	*Plasmid full name*
pCDF-AA	pCDF-PylRS-N346A/C348A

doi:10.6342/NTU201903019

pET-α-3XG4S-f	pET-αCT-3XG4S-ferritin
pET-f-linker-α variants	
pET-f-3XG4S-α	pET-pylT-ferritin-3XG4S-αCT
pET-f-2XG4S-α	pET-ferritin-2XG4S-αCT
pET-f-1XG4S-α	pET-ferritin-1XG4S-αCT
pET-ftn-1TAG variants	
pET-ftn-56	pET-ferritin-L56TAG
pET-ftn-60	pET-ferritin-H60TAG
pET-ftn-63	pET-ferritin-R63TAG
pET-ftn-67	pET-ferritin-E67TAG
pET-ftn-2TAG variants	
pET-ftn-60/63	pET-ferritin-H60/R63TAG
pET-ftn-63/67	pET-ferritin-R63/E67TAG
pET-ftn-60/67	pET-ferritin-H60/E67TAG
pET-f-3XG4S-α-1TAG	
pET-f-3XG4S-α-63	pET-ferritin-3XG4S-αCT-R63TAG
pET-f-3XG4S-α-2TAG	
pET-f-3XG4S-α-63/67	pET-ferritin-3XG4S-αCT-R63/E67TAG

pCDF-AA

pCDF-AA plasmid that encodes PylRS mutant N346A/C348A was derived from pBK-N346A/C348A plasmid, which is based on published paper[44].

Construction of pET-α-3XG4S-f

The pET-α-3XG4S-f plasmid is derived from pET-Ferritin. Genes of αCT are based

doi:10.6342/NTU201903019

on protein structure of insulin receptor ectodomain and connected to ferritin genes via 3XG4S linker. The following primers are designed to amplify DNA fragments of (1), (2), and (3). (1) entry 3 and 4; (2) entry 5 and 6; (3) entry 7 and 2. Overlap extension PCR was applied to fuse DNA fragments (1) and (2). After purification, overlap extension PCR was again conducted to join (3) together with (1)(2). The recipient plasmid pET-pylT and the PCR products were digested with restriction enzymes NdeI and SacI, purified by agarose gel extraction and ligated together to create pET-α-3XG4S-f by T4 DNA ligase. The target genes have Histag at the C- terminus.

Construction of pET-f-linker-α variants

pET-f-linker-α variants include pET-f-1XG4S-α, pET-f-2XG4S-α, and pET-f-3XG4S-α. The source of pET-Ferritin and genes of αCT are similar to above.

To construct those variants, following primers are designed to amplify DNA fragments. pET-f-1XG4S-α: (1) entry 1 and 8; (2a) entry 19 and 21; (3) entry 11 and 12. pET-f-2XG4S-α: (1) entry 1 and 8; (2b) entry 19 and 20; (3) entry 11 and 12. pET-f-3XG4S-α: (1) entry 1 and 8; (2c) entry 9 and 10; (3) entry 11 and 12. Target genes of each variant are consisted of three DNA fragments, the second fragment differs by the length of linker. To combine those fragments, overlap extension PCR was applied to fuse (2) and (3) in advance, and finally, (1) was put together with purified (2)(3). Recipient plasmid pET-pylT and each of the three PCR products were digested with restriction enzymes NdeI and SacI and ligated by T4 DNA ligase to create pET-f-linker-α variants. All the target genes contain Histag at the C- terminus.

pET-ftn-1TAG variants

pET-ftn-1TAG variants include pET-ftn-56, pET-ftn-60, pET-ftn-63, and pET-ftn-67.

doi:10.6342/NTU201903019

Those plasmids were constructed by introducing an amber mutation separately at L56, H60, R63, and E67 by overlapping extension PCR.

Construction of pET-ftn-2TAG variants

pET-ftn-2TAG variants include pET-ftn-60/63, pET-ftn-63/67, and pET-ftn-60/67.

To introduce amber stop codon to pET-Ferritin, following primers are designed to amplify DNA fragments. pET-ftn-60/63: (1a) entry 1 and 13; (2a) entry 14 and 2. pET-ftn-63/67: (1b) entry 1 and 17; (2b) entry 18 and 2. pET-ftn-60/67: (1c) entry 1 and 16; (2c) entry 15 and 2. Two DNA fragments of each variant, (1) and (2), were fused by overlap extension PCR and digested with restriction enzymes NdeI and SacI. Digested recipient plasmid pET-pylT and the PCR products were purified by agarose gel extraction and ligated together to generate pET-ftn-2TAG variants with Histag at the C-terminus of target genes.

Construction of pET-f-3XG4S-α-63

pET-f-3XG4S-α-63 is derived from pET-f-3XG4S-α. To introduce amber stop codon to pET-f-3XG4S-α, the two DNA fragments were amplified by following primers. (1) entry 1 and 27; (2) entry 26 and 12. Overlap extension PCR was applied to fuse DNA fragments of (1) and (2) to generate the target gene. After the digestion of recipient plasmid pET-pylT and PCR products with NdeI and SacI, DNA ligation was performed generate pET-f-3XG4S-α-63 with Histag at the C-terminus of target genes.

Construction of pET-f-3XG4S-α-63/67

pET-f-3XG4S-α-63/67 is derived from pET-f-3XG4S-α. To introduce amber stop codons to pET-f-3XG4S-α, the two DNA fragments were amplified by following primers. (1) entry 1 and 17; (2) entry 18 and 12. Overlap extension PCR was applied to fuse DNA

doi:10.6342/NTU201903019

fragments of (1) and (2) to generate the target gene. After the digestion of recipient plasmid pET-pylT and PCR products with NdeI and SacI, DNA ligation was performed generate pET-f-3XG4S-α-63/67 with Histag at the C-terminus of target genes.

2-3 Protein overexpression and purification

Overexpression and purification of recombinant proteins

To overexpress αCT-ferritin (α-F), ferritin-1XG4S-αCT (F-1XG4S-α), ferritin-2XG4S-αCT (F-2XG4S-α), and ferritin-3XG4S-αCT (F-3XG4S-α), pET-pylT plasmids containing target genes were transformed to *E. coli* BL21(DE3) competent cell. Competent cells were thawed on ice and mixed with 0.5 μL plasmid (80 ng/μL), after incubation on ice for 10 minutes, heat shock was performed in 42ºC water bath for 1 minute and the cells are put back on ice for 2 minutes. To recover the cells, 1 mL LB is added to the mixture followed by 37ºC for 1 hour. The cells were then spread over on LB agar plate with Ampicillin (Amp) (100 μg/mL). After overnight incubation, a single colony was selected and inoculated in 5 mL LB overnight. For large-scale overexpression, the 5mL overnight broth was cultured in 500 mL LB supplemented with Amp and grown at 37ºC, 200 rpm upon OD_{600} reaching 0.6. The cells were then induced by adding 1 mM isopropyl β-D-thiogalactoside (IPTG) and incubated for 3 hours at 37ºC. Cells were harvested by centrifugation at 6,000 rpm., re-suspended in the lysis buffer (20 mM Tris, 100 mM NaCl, pH 7.5) and disrupted by sonication.

The cell lysate was clarified by centrifugation at 9,500 rpm for 50 min at 4ºC and the supernatant was collected and transferred to column containing 2 mL Ni^{2+}-NTA column resin (EBL Inc.). The resin was washed with 10 column volume (CV) lysis buffer and 3 CV wash buffer (10 mM imidazole, 20 mM Tris, 100 mM NaCl, pH 7.5) for

doi:10.6342/NTU201903019

removing proteins binding non-specifically. Finally, target protein was eluted by elution buffer (300 mM imidazole, 20 mM Tris, 100 mM NaCl, pH 7.5). All the fractions were analyzed by 15% SDS-PAGE and fractions containing target protein were centrifuged by using Amicon Ultra-15 Centrifuge Filter Devices (10,000 MWCO cut) (Merck Ltd.) to reduce the concentration of imidazole. The concentrated samples were filtered by 0.22 μm Minisart Syringe Filter (Sartorius Inc.) and injected into Superdex 200 Increase 10/300 GL (GE healthcare Ltd.) for size exclusion chromatography. The column was equilibrated with buffer (20 mM Tris, 100 mM NaCl, pH 7.5) in advance. The fractions were analyzed by 15% SDS-PAGE and concentrated before being distributed into Eppendorf and stored in -20ºC.

Overexpression and purification of proteins with ncAA incorporation

To overexpress human heavy chain ferritin (Ftn) and F-α containing ncAA, pET-pylT plasmids containing target genes and pCDF-AA are co-transformed to *E. coli* BL21(DE3) competent cell. After heat shock, cells were recovered in 1 mL LB medium at 37ºC for 1hour and spread on LB agar plate with Spectinomycin (Sp) (100 μg/mL) and Amp (100 μg/mL). After overnight incubation, single colony on LB agar plate was selected and inoculated in 5ml culture. On the next day, 2 mL of culture broth was cultured in 200 mL LB media with Amp and Sp supplementation. The cell culture was incubated at 37ºC, 200 rpm until it reach an OD_{600} of 0.8-1.0 and centrifuged at 4,500 rpm for 10 min to remove the supernatant. The cells were washed by M9 minimal medium twice and transferred to M9 minimal media (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl) supplemented with 1% glycerol, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2% NaCl, 100 μg/mL Sp, 100 μg/mL Amp, 1 mM IPTG, and 1 mM ncAA. ncAA **1-3**: ThzA, BrThA, and MeH were dissolved in water to make 50 mM stock solutions. After 12-hour

doi:10.6342/NTU201903019

induction at 37ºC, cells were harvested and the target protein with ncAA incorporation was purified and analyzed by similar protocol mentioned above.

2-4 Protein gel electrophoresis

2-4-1 SDS-PAGE analysis

To conduct SDS-PAGE analysis, the concentration of proteins was measured by Bradford Protein Assay (Bio-Rad Laboratories, Inc.). Protein samples were prepared by mixing 16 μL protein with 4 μL 5x sample buffer (200 mM Tris-HCl ,400 mM DTT, 8% SDS, 0.04% bromophenol blue, and 40% glycerol, pH6.8) to make the final volume 20 μL. The samples were then heated at 95°C for 5 minutes and spun down. 10 μL of denatured sample and 2 μL protein marker (Bio-Rad Laboratories, Inc.) were loaded into each well of the 15% SDS-PAGE. Gels were held vertically between the electrode chambers and the tank was filled with 1X TGS buffer (25 mM Tris-HCl, 192 mM glycine, 0.1% SDS, pH8.5). The electrophoresis was performed at 120 V for 85 min and the bands were visualized by staining the gel with instant blue (Marvelgent Biosciences Inc.) for 30 minutes.

2-4-2 Native gel analysis

To prepare samples for native gel analysis, the concentration of proteins was determined by Bradford Protein Assay. 10-20 μg of proteins were mixed with 4x Native sample buffer (100 mM Tris-HCl, 0.02% bromophenol blue, and 20% glycerol, pH6.8), the final volume should be less than 20 μL. The samples and 4 μL native protein marker (Thermo Fisher Scientific Inc.) were slowly loaded into 4–15% Mini-PROTEAN® TGX™ Precast Protein Gel (Bio-Rad Laboratories, Inc.). The gel was then held vertically between the electrode chambers and the tank was filled with TG buffer (25 mM Tris-HCl,

doi:10.6342/NTU201903019

192 mM glycine, pH8.5). The electrophoresis was performed at 60 V for 6 hours at 4°C and the bands were visualized by staining the gel with instant blue for 30 minutes.

2-5 Biophysical characterization

2-5-1 Dynamic light scattering (DLS) analysis

Proteins were prepared in buffer (20 mM Tris, 100 mM NaCl, pH 7.5) and the concentration was adjusted to 0.5 mg/mL, followed by 0.22 μm filtration to remove aggregation or impurities in protein samples. Experiments were conducted with a Malvern Zetasizer Nano ZS at 25°C in 1 mL disposable cuvettes and repeated at least three times. Measurements were presented in size distribution, which are derived from intensity distribution.

2-5-2 Transmission electron microscope (TEM) analysis

The protein samples were prepared in buffer (20 mM Tris, 100 mM NaCl, pH 7.5) with the final concentration of 0.075 mg/ml. The glow discharge of carbon-coated copper grid was performed by using EMITECH K100X Glow with 25 mA for 30 seconds. 5 μL of sample was then loaded on a discharged support films grid (formvar/carbon 400 Mesh copper). After staying for 90 seconds, the excess solution was wicked away with filter paper, and the grids were washed by rinsing with 5 μL of distilled deionized water and wicked away immediately for three times. 5 μL of aqueous freshly filtered 1% uranyl acetate was dropped to the grids and stayed for one minute before been wicked away. The grids were placed in a dish and put in the desiccator until they fully dry. Images of Ftn and F-α were collected on FEG-TEM, FEI Tecnai G2 TF20 Super TWIN. The microscope was used at an accelerating voltage of 120 kV.

doi:10.6342/NTU201903019

2-6 Friedel-Crafts alkylation of aromatic amino acids

Amino acids (Chem-Impex International, Inc.) and $CuCl_2$ were dissolved in water to make 100 mM stock solutions. The reaction solution containing 10 mM amino acid, 20 mM diethyl ethylidenemalonate (DEEM) (Sigma-Aldrich, Inc.), and 1 mM $CuCl_2$ is incubated (37°C, 600 rpm) for 4 hours and monitored by thin layer chromatography (TLC). TLC was performed using silica gel (TLC Silica gel 60 F_{254}) with either 90%(v/v) or 70%(v/v) isopropanol as mobile phase and stained with ninhydrin.

2-8 Protein chemistry

2-8-1 Cu(II) binding Ftn and F-α variants preparation

The concentration of Ftn and F-α variants were determined by Bradford Protein Assay. To charge Ftn variants with Cu(II), protein solution was dialyzed against buffer (6 μM $CuCl_2$, 20 mM Tris, 100 mM NaCl, pH 7.5) at 4°C for 12 hours. Finally, protein solution was concentrated by using Amicon Ultra-0.5, 10 kDa after removal of excess $CuCl_2$ in dialysis buffer.

2-8-2 Cu(II)-binding Ftn and F-α variants-catalyzed insulin modifications

Chemicals with α,β-unsaturated carbonyl groups used in reactions include DEEM, 2-Chloro-2,4,6-cycloheptatrien-1-one (tropone) (Sigma-Aldrich, Inc.), benzalacetone (BA) (Honeywell Fluka™), and diethyl benzylidenemalonate (DEBM) (Thermo Fisher Scientific Inc.).

To conduct the experiment, Ftn and F-α variants with $CuCl_2$ treatment were concentrated via Amicon Ultra-0.5, 10 kDa to reach protein concentration at 2 mg/ml. 0.58 mg of human recombinant insulin (Sigma-Aldrich, Inc.) was dissolved in buffer (20 mM Tris, 100 mM NaCl, pH 7.5) containing Ftn or F-α variants to make a 1 mM solution.

doi:10.6342/NTU201903019

2 mM small molecules are either dissolved in buffer (20 mM Tris, 100 mM NaCl, pH 7.5) in advance or directly added to the solution. The 50 μL reaction mixture was adjusted to pH 7.5 before 37ºC incubation for 12 hours. The control experiments were catalyzed by 0.1 mM $CuCl_2$ and the rest of reaction conditions are the same.

To reduce the disulfide bonds of insulin, 30 equivalents of DTT was added to modified insulin. The mixture was reacted under 37ºC for 30 minutes and centrifuged to separate supernatant and aggregate. The supernatant and aggregate were dissolved separately in 20% (v/v) acetonitrile and 0.1% (v/v) formic acid and sent to MALDI-TOF for analysis.

2-9 Mass spectrometry

2-9-1 ESI-MS

The proteins were diluted with 20% (v/v) acetonitrile and 0.1% (v/v) formic acid and concentrated by using Amicon Ultra-0.5, 10 kDa. An aliquot corresponding to one ρmol of the pure protein was injected via a ESI source (Waters LockSpray Exact Mass Ionization Source) with a syringe pump (Harvard Apparatus, MA) and held a flow rate of 5 μL/min throughout the analysis. The mass of intact proteins was determined by Waters Synapt G2 HDMS mass spectrometer (Waters, Milford, MA), and the deconvolution of acquired spectra was calculated by using MaxEnt1 algorithm of the MassLynx 4.1 software (Waters) to generate the single-charge state of target protein.

2-9-2 MALDI-TOF-MS/MS

In-gel digestion: A modified in-gel digestion protocol was applied in this experiment[48]. The SDS-PAGE gel bands were exercised and cut into small pieces. After sequential washing of the gel pieces with 25 mM NH_4HCO_3, 40% methanol solution, and

doi:10.6342/NTU201903019

100% acetonitrile. The reduction and alkylation of proteins in gel pieces were performed by using DTT and iodoacetamide, respectively. The gel pieces were then washed again and dried in a vacuum centrifuge before trypsin digestion. 25-30 μL trypsin solution containing 65 to 100 ng of sequencing grade modified trypsin (Promega Corp.) was prepared in 25 mM NH_4HCO_3, 10% acetonitrile, it was mixed with gel pieces and incubated for 12-16 hours at 37°C. The reaction was quenched by adding 1-2 μL of 5% formic acid.

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

2-9-3 ICP-MS

The concentration of F-α and Ftn variants were measured by Bradford Protein assay and noted down for further calculation, 200 μL of Ftn variants were transferred to 1.5 mL Eppendorf and then sent to Biotech Incubation Center.

Sample preparation：100 uL samples were added to 3 mL HNO_3 (ultrex II, J.T.Baker) in the microwave tubes. The mixtures were digested with microwave oven (MARS, CEM). Every sample has 30 W. The first digestion step was that spend 15min to reach the temperature 130°C. Hold with 20 min and then cool down to room temperature. The digested samples were diluted with ddH_2O to 25 mL.

Instrument：Analyses were carried out using an inductively coupled plasma mass spectrometer (ICP-MS) X series II (Thermo Fisher, Hemel Hempstead, UK). The instrument was used with direct nebulization in normal mode using optimized conditions

doi:10.6342/NTU201903019

consisting of an Xs skimmer cone with a Platinum sampler cone . The extraction voltage was typically -35 V, the Rf Power was 510 W, focus voltage was 4.5 V, and the nebulizer gas flow rate was 1.04 L/min. Dwell times were 10 ms, 100 sweeps were required per replicate and 5 replicates per sample.

doi:10.6342/NTU201903019

Chapter 3 Results and Discussion

3-1 Design of Friedel-Crafts metalloenzyme for insulin modification

3-1-1 Construction of αCT peptide fusion ferritin

In order to design a metalloenzyme that specifically modify insulin, an αCT peptide derived from insulin receptor was fused to either N or C-terminus of ferritin with Histag at the C-terminus of target genes. Since αCT is the crucial segment involved in the primary binding site of insulin as reported[11], it should be able to equip the fusion protein with binding affinity toward insulin, moreover, a flexible linker 3XG4S was applied to connect αCT and ferritin[49]. Based on those ideas mentioned above together with 24mer self-assembled feature of ferritin, the fusion protein could exhibit better specificity on insulin modifications using its 24 fusion peptides.

3-1-2 Protein sequence design of αCT peptide fusion ferritin

From the structural view of ferritin subunit, the N-terminus is exposed to solvent while the C-terminus is buried below surface. In order to know whether αCT peptide is pointing outward ferritin cage, ferritin with αCT on either N or C-terminus are constructed and analyzed after protein overexpression and Ni^{2+}-NTA column purification.

SDS-PAGE (**Figure 8**) had shown that both of them are overexpressed in *E. coli* since induced samples have thicker bands. However, after disrupting the cells by sonication and Ni^{2+}-NTA column purification, only ferritin with αCT at C-terminus (F-α) is soluble, ferritin with αCT at N-terminus (α-F) remained mostly in inclusion bodies. The elution bands of F-α matched the expected molecular weight 25 kDa. The protein sequence of α-F (**Figure S1**) and F-α (**Figure S2**) were determined by protein identification. In native page, Ftn and F-α are folded and self-assembled into a 24 subunits

doi:10.6342/NTU201903019

ball, the molecular weights are 531 kDa and 602 kDa respectively.

From these results, it can be summarized that αCT on N-terminus would impair the assembling of ferritin, making the protein insoluble in buffer, and that some of the αCT are pointing outward since the protein can be purified by Ni^{2+}-NTA column via the C-terminus Histag. F-α was then chosen for following experiments because of the soluble and assembled structure.

3-1-3 Analysis of F-α binding affinity with different linkers

As mentioned above, some αCT peptides of F-α are exposed outside the protein cage. So the next question will be, whether the linker length between ferritin and αCT affect the binding of Histag during purification? Three constructs with different length linkers are tested in this experiment: one, two, and three repeat $S(GGGGS)_n$ (n=1, 2, or 3) (**Figure 9**). During Ni^{2+}-NTA column purification, higher concentration of imidazole was required to elute F-α with longer linker; on the contrary, the construct with shorter linker performed lower affinity during Ni^{2+}-NTA column purification. Native PAGE had shown that three of them are assembled. In here, it is again proved that αCT is pointing outward and that longer linker increases the binding affinity. 3XG4S linker was then applied to following experiments.

3-2 Characterization of interactions between F-α and insulin

DLS and TEM are utilized to characterize the interaction between F-α and insulin. Ferritin is a protein cage with an external diameter of 12 nm and an internal diameter of 8 nm. DLS analysis (**Figure 10**) had shown that the calculated size distribution of Ftn is slightly larger than 10 nm, which is similar to theoretical value. However, the size of F-α is a little bigger due to the αCT fusion, and what is more, the smaller peak on the left

doi:10.6342/NTU201903019

during control experiment disappeared upon adding insulin, suggesting the possible interaction between insulin and F-α.

Under TEM images (**Figure 11**), both F-α and Ftn are assembled protein cages, however, F-α didn't seem hollow compared to Ftn and this may resulted from the C-terminus αCT peptide since some of them were pointing inward the cage. Besides, mixing insulin with F-α didn't change its pattern, indicating that insulin didn't disrupt its structure and assembly. Image J was also used to estimate the size of protein cages, the external diameters of Ftn and F-α are calculated as 12.3±0.9 nm and 14.8±1.2 nm respectively, which were matching the corresponding size distribution observed in DLS.

3-3 Friedel-Crafts alkylation of aromatic amino acids

Since Friedel-Crafts alkylation is an electrophilic aromatic substitution, the activities of aromatic amino acids were investigated in advance. Among α,β-unsaturated carbonyl compounds, DEEM was chosen due to its proposed activity. 20 mM DEEM and 10 mM aromatic amino acid (His, Phe, Trp, and Tyr) were catalyzed by $CuCl_2$ (0 mM, 0.1 mM, and 1 mM) at 37°C for 4 hours. The TLC (**Figure 12**) had shown that histidine reacted with DEEM in the absence of $CuCl_2$, and the concentration of catalyst did not affect its conversion. The R_f value of histidine is 0.03 and the product is 0.35. The other three amino acids, Phe, Trp, and Tyr, however, require 1 mM $CuCl_2$ to react with DEEM and the R_f values between reactant and product are very close.

The products of reaction between DEEM and aromatic amino acids were analyzed by ESI-MS spectrometry, and only products of alkylated histidine (**Figure S3**) and tryptophan (**Figure S4**) were found in mass spectrum, the expected mass are 341 Da and 390 Da, respectively.

doi:10.6342/NTU201903019

3-4 Insulin modifications by Cu(II) binding Ftn variants

3-4-1 Modification of ferritin C_2 interface by ncAA incorporation

A research studying ferritin C_2 interface was taken as a reference[28] in the design of novel Cu(II) binding site. In this study, L56, H60, R63, E67 at the C_2 interface (**Scheme 5**) are all mutated to histidine to perform a Cu(II)-assisted assembly feature, in other words, ferritin couldn't assemble in the absence of Cu(II). However, since the goal is to create de novo metal-binding Ftn variants that remain assembled under normal conditions, metal-binding ncAAs **1-3** (**Scheme 6**) were introduced to C_2 interface. Expanding genetic code was applied by using biorthogonal *Mm*PylRS-N346A/C348A• $tRNA_{CUA}^{Pyl}$ pairs (**Figure 13**) to engineer ferritin C_2 interface with either one or two ncAAs to generate six Ftn variants.

Whole cell lysates of Ftn with one amber suppression (Ftn-1x) of ncAAs **1-3** at R63 are analyzed by SDS-PAGE (**Figure S5**), so as Ftn with two amber suppression (Ftn-2x) at R63 and E67 (**Figure S6**), the darker target bands at 22 kDa after induction indicated the expression of Ftn variants. Ftn-1x (**Figure S7-S9**) and Ftn-2x variants (**Figure S10-12**) were purified by Ni^{2+}-NTA column. SDS-PAGE analysis had indicated that these proteins didn't perform strong binding affinity since they are eluted at low concentration of imidazole, which is also the feature of Ftn. However, all of them were assembled under native PAGE analysis (**Figure 14**), suggested that the engineering of C_2 interface didn't affect their assembly strongly.

3-4-2 ICP-MS analysis of metal binding engineering

Before charging Ftn variants with Cu(II), the content of Fe was measured by ICP-MS (**Table 3**). Commercial standards apo ferritin and equine spleen ferritin were also taken for iron content analysis. Since apo ferritin refers to ferritin without iron ions, it

doi:10.6342/NTU201903019

was serving as a control, which had only 1.2 iron ions per ball. Equine spleen ferritin is mostly consisted of ferritin light chain, therefore, it would supposedly store large amount of iron ions, the resulted showed that it contains more than 2000 iron ions.

Compared to equine spleen ferritin, Ftn had lower iron ions concentration, 15.6±8.5 iron ions per ball, therefore, it is proposed that iron ions merely bind on the surface rather than being stored inside its cavity. Besides Ftn, F-α also contained fewer amount of iron ions, 7.0±0.2 per ball, and this is possibly resulted from the interference of αCT that blocked the C_3 channels for iron entry.

The iron concentration of C_2 interface-engineered variants differed a lot, Ftn-2x-**3** had more than 50 iron ions per ball, while one Ftn-1x-**3** had only 4 iron ions. In order to reduce the interference of iron ions during catalysis, the two Ftn variants with lowest iron concentration, Ftn-1x-**3** and Ftn-2x-**2**, were selected for the following $CuCl_2$ treatment.

After $CuCl_2$ treatment, the copper ion content of Ftn (19.7±2.3/ ball), F-α (29.8±2.9/ ball) had increased as compared to the proteins without $CuCl_2$ treatment. On the other hand, the iron concentration of Ftn (29.5±19.9/ ball) and F-α (8.3±5.0/ ball), did not change a lot. The results indicated that $CuCl_2$ treatment would not affect the binding of iron ions. The two selected Ftn variants, Ftn-1x-**3** (5.2/ ball) and Ftn-2x-**2** (4.7/ ball) had also demonstrated consistent iron ions concentration after $CuCl_2$ treatment, while the copper ions concentration, Ftn-1x-**3** (29.5/ ball) and Ftn-2x-**2** (22.6±6.8/ ball), had increased. These results had suggested that the engineering of C_2 interface will impair the binding of iron ions and that the novel metal chelating sites chelated only copper ions since the concentration of iron ions remained similar.

In short, the results mentioned above can be summarized that both αCT and C_2 interface modifications disrupt either the binding or entrance of iron ions. It is also proposed that the chosen C_2 interface-engineered Ftn variants with modifications would

doi:10.6342/NTU201903019

possibly contain nearly no iron ions after αCT fusion.

3-4-3 Construction of metal binding Ftn variants with αCT fusion

To further optimize the metal binding capacity of Ftn variants, Ftn-1x-**3** and Ftn-2x-**2** with αCT on C terminus were constructed by cloning. After overexpression, F-α variants, F-α-1x-**3** (**Figure S15**) and F-α-2x-**2 (Figure S16),** were purified by Ni^{2+}-NTA column and performed similar elution profiles as F-α who was eluted under high concentration imidazole. It can be therefore proved again that 3XG4S is the only reason that affect binding during Ni^{2+}-NTA column purification.

3-4-4 ESI-MS analysis of Ftn variants

In order to determine the molecular weight of F-α and Ftn variants, ESI-MS was applied to confirm the site-specific incorporation of ncAAs into target proteins. The molecular weight of Ftn variants (Ftn-1x-**1-3** and Ftn-2x-**1-3**) were plotted in waterfall (**Figure 15**), the molecular weights were around 22 kDa. F-α and F-α variants (F-α-1x-**3** and F-α-2x-**2**) were 25 kDa (**Figure S17, S24-25**). All the observed molecular weights matched the expected molecular weights.

3-4-5 MALDI-TOF-MS analysis of insulin modification

All the Ftn and F-α variants were serving as metalloenzymes to catalyze reactions between insulin (1 mM) and DEEM (2 mM), and the concentration of single metalloenzymes was 0.04 mM (0.1 mM monomers). Before reactions, the molecular weight of human recombinant insulin was determined by MALDI-TOF (**Figure 17)** to be 5808 Da, the adduct cluster peaks around major peak came from the sodium ions in solution since insulin was dissolved in buffer (20 mM Tris, 100 mM NaCl, pH 7.5). In **Table 4**, entry 1-2 refer to reactions catalyzed by Ftn variants (Ftn-1x-**3** and Ftn-2x-**2**).

doi:10.6342/NTU201903019

The single modification of DEEM, 186 Da shift, presented in both results indicated the specificity of the newly designed metalloenzymes. Reactions were conducted under 37°C for 12 hours and the conversion of Ftn-1x-**3** (**Figure 18**) and Ftn-2x-**2** (**Figure 19**) were 20% and 33%, respectively. To further optimize the catalytic efficiency of Ftn variants, F-α variants were constructed and applied.

MALDI-TOF-MS analysis had shown that, F-α-1x-**3** (**Figure 20**) and F-α-2x-**2** (**Figure 21**), who possess αCT, displayed increased conversion as compared to Ftn-1x-**3** and Ftn-2x-**2**, at the same time, DEEM modification peaks remained single. It is proposed that the optimization came from the fusion of αCT that improve the interaction between insulin and metalloenzymes. However, since the metalloenzymes mentioned above didn't undergo iron ion removal process so the exact reason that lead to better efficiency has not been proved yet.

In order to determine the location of DEEM modification on insulin, reaction catalyzed by merely $CuCl_2$ was conducted. Insulin disulfide bonds were reduced with DTT, and separated into A chain (2384 Da) and B chain (3430 Da). However, the solution precipitated after the reduction, the pellet was dissolved by 20% (v/v) acetonitrile, 0.1% formic acid and sent for analysis together with the soluble fraction. MALDI-TOF analysis (**Figure 22**) had shown two modifications appeared on the right side of insulin B chain and the peaks were differed by 186 Da, which was calculated to be the molecular weight of DEEM. In conclusion, Ftn and F-α variants had demonstrated catalytic power.

3-4-6 MALDI-TOF-MS/MS analysis of insulin modification

In Friedel-Crafts alkylation of insulin, four aromatic amino acids were proposed to react with DEEM. In order to locate the exact amino acid residue been modified, insulin catalyzed by Cu(II) binding F-α-2x-**2** (entry 4) was analyzed by 20% SDS-PAGE in-gel

doi:10.6342/NTU201903019

digested by glutamyl endopeptidase GluV8. The MALDI-TOF-MS/MS (**Figure 23**) has shown that the first histidine on B chain (His^{B5}) was the one been modified by DEEM, however, with an extra 26 Da adduct, which was possibly resulted from the processing of in-gel digestion.

Chapter 4 Conclusion

In this study, de novo metalloenzymes that catalyzed the reaction between insulin and DEEM had been constructed. The αCT peptide derived from insulin receptor which had been fused to the C-terminus human heavy chain ferritin was proved to be pointing outward by Ni^{2+}-NTA column purification and that increasing the length of linker connecting ferritin and αCT indeed increased the binding affinity during Ni^{2+}-NTA column purification. Moreover, DLS analysis also showed a size distribution difference after mixing insulin with F-α. The assembly feature of F-α was also seen under TEM images.

On the other hand, C_2 interface of Ftn variants was successfully engineered with the assistance of expanding genetic codes. By incorporating ncAAs into specific amino acid residues, novel metal binding sites for Cu(II) chelation were designed. The endogenous iron content and copper ion concentration were measured by ICP-MS. The results have showed that Ftn had much lower iron ion concentration compared to commercial ferritin standard. What is more, both αCT fusion and C_2 interface engineering led to decreased iron ion concentration which may resulted from the disruption of iron ions binding or entrance.

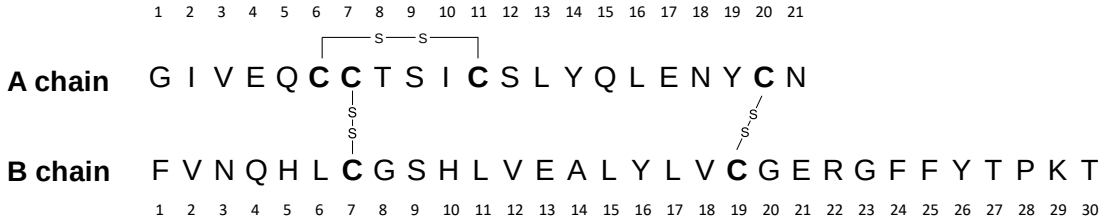
Finally, the newly designed metalloenzymes, Cu(II) binding Ftn (Ftn-1x-**3** and Ftn-2x-**2**) and F-α (F-α-1x-**3** and F-α-2x-**2**) variants were able to catalyze reactions between DEEM and insulin specifically. Moreover, F-α variants, the optimized metalloenzymes,

doi:10.6342/NTU201903019

demonstrated better efficiency because of the αCT fusion. MALDI-TOF-MS results also indicated that the modification was located on insulin B chain after reduction of endogenous disulfide bonds. To further confirm the exact amino acid residue been modified, insulin reaction catalyzed by F-α-2x-**2** was digested by GluC and analyzed by MALDI-TOF-MS/MS, the alkylation was presented on histidine as predicted. Although a 26 Da adduct was found on the fragment, it was proved that DEEM had specifically reacted with the first histidine on insulin B chain (His^{B5}).

In the future, Ftn and F-α variants will undergo iron removal before $CuCl_2$ treatment to clarify factors that had contributed to the catalysis. The detailed chemistry and modification sites between insulin and DEEM will be further studied. In conclusion, the de novo metalloenzyme platform could possibly serve as an in vivo enzyme treatment in future, the tuning of insulin action profile can be achieved by small molecules oral treatment and metalloenzyme injection.

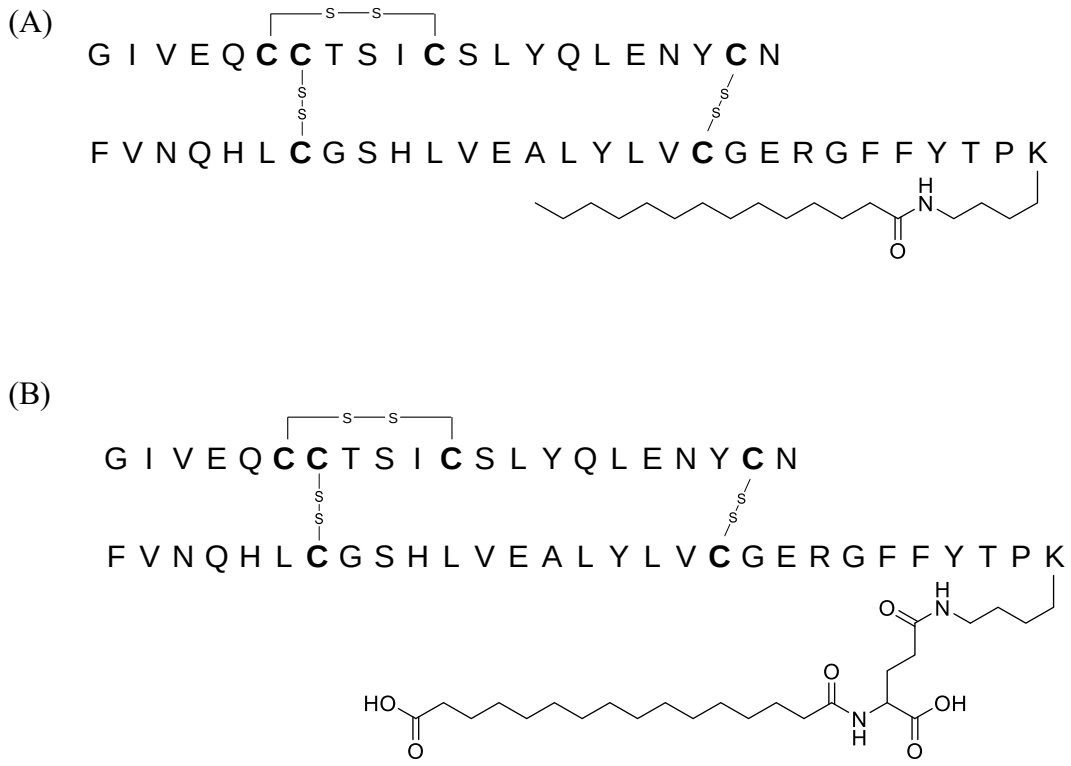
doi:10.6342/NTU201903019



Scheme 1. Amino acid sequences of insulin.

Insulin is a type of peptide hormone whose amino sequence is highly conserved within vertebrates. Human insulin has 51 amino acids, and it is consisted of A chain and B chain constrained by one intrachain and two interchain disulfide bonds.

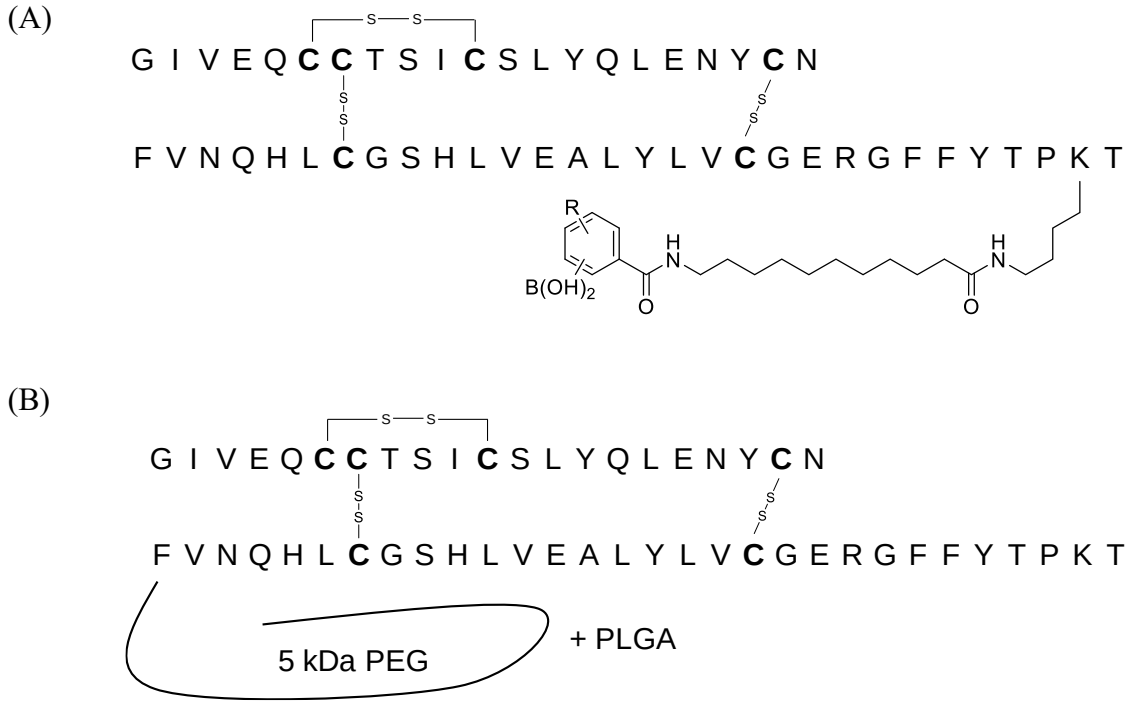
doi:10.6342/NTU201903019



Scheme 2. Structure of approved basal insulin analogues.

(A) Insulin detemir has a long fatty acid chain linked to its B chain, which can slow absorption and improve plasma circulation through non-covalent albumin binding. (B) Insulin degludec exists in a stable multi-hexameric structure that slowly releases insulin monomer.

doi:10.6342/NTU201903019

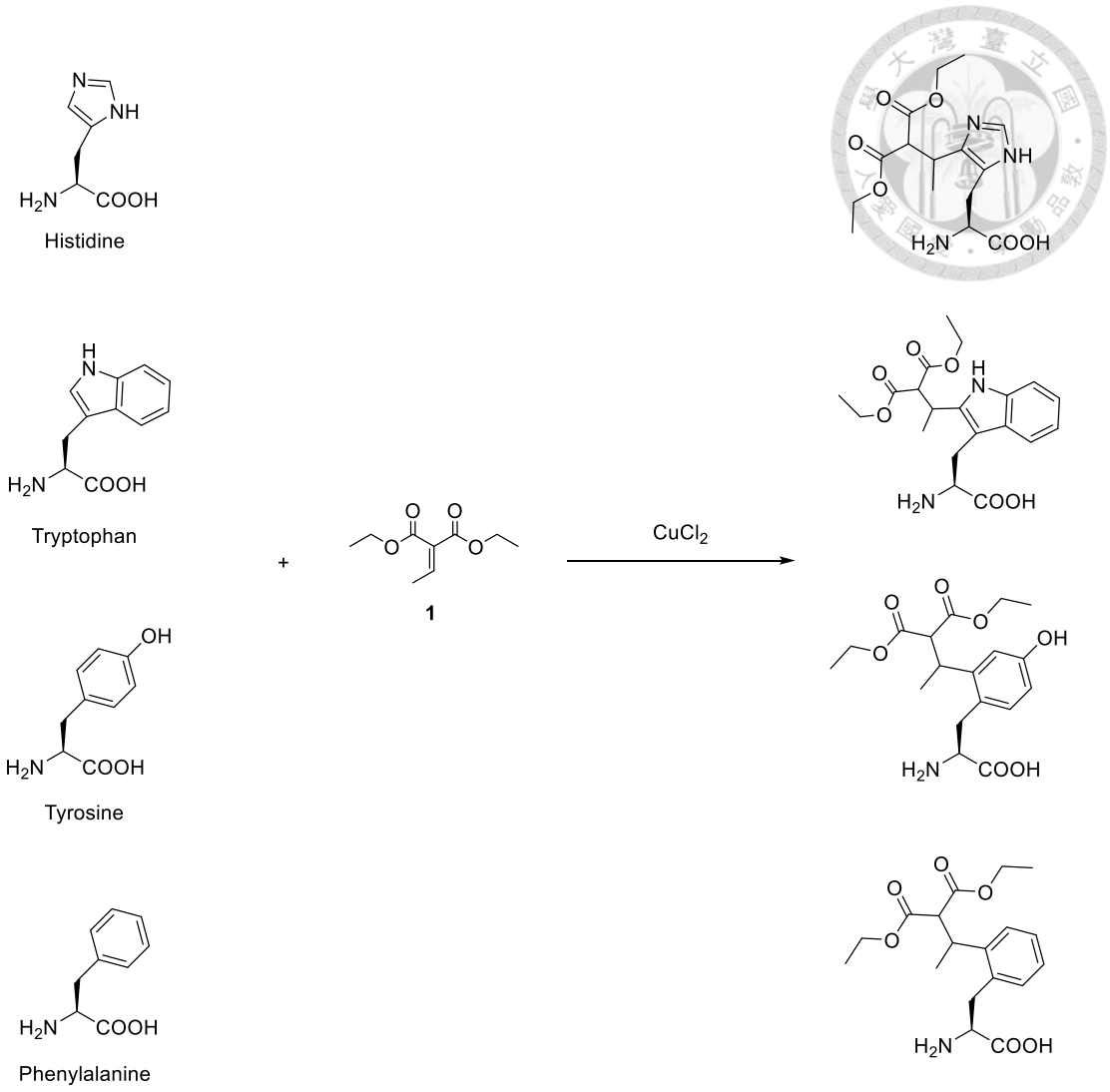


Scheme 3. Structure of basal insulin analogues under study.

(A) PBA-insulin is a glucose responsive insulin possesses a glucose sensor group who will react with glucose and release itself upon high blood glucose levels. (B) AB101 is a refined insulin with a 5kDa PEG linked to its B chain, the PEG together with PLA could encapsulate insulin and form slow release microspheres.

Source: (A) Adapted from Zaykov, A. N.; Mayer, J. P.; DiMarchi, R. D., Pursuit of a perfect insulin. *Nat. Rev. Drug. Discov.* **2016**, 15 (6), 425-39.

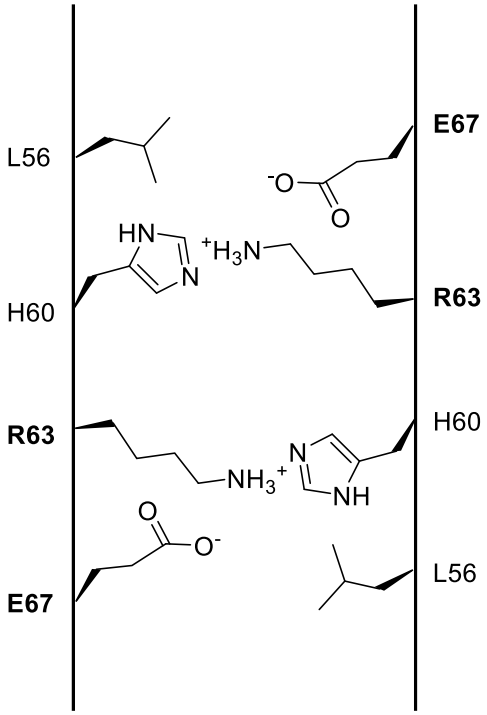
doi:10.6342/NTU201903019



Scheme 4. Proposed reaction of Friedel-Crafts reaction on aromatic amino acids

Proposed Friedel-Crafts alkylation mechanisms of reactions between four aromatic amino acids (His, Trp, Tyr, and Phe) and small molecules **1** were catalyzed by $CuCl_2$.

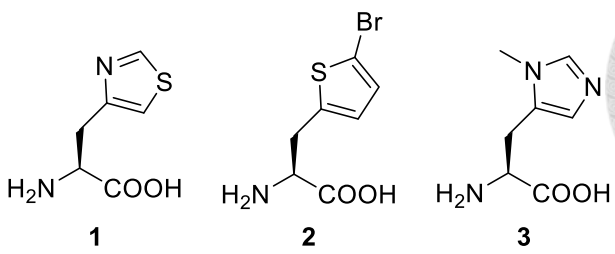
doi:10.6342/NTU201903019



Scheme 5. C_2 interface of wild type ferritin

Intersubunit interactions between ferritin viewed parallel to the C_2 symmetry axis. The C_2 interface is consisted of dimers, to engineer its Cu(II)-chelation properties, selected amino acid residues (marked in bald) will be displaced by ncAAs.

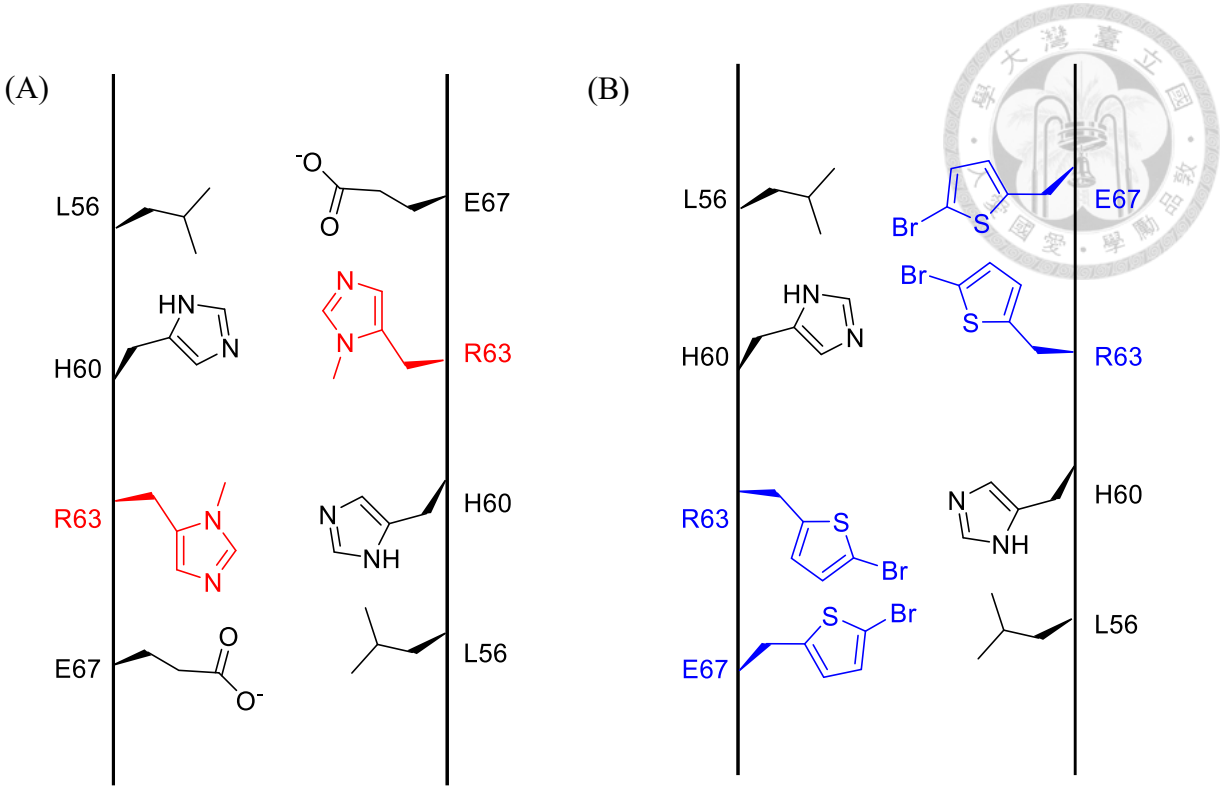
doi:10.6342/NTU201903019



Scheme 6. ncAAs used in this study.

1: 3-(4-Thiazolyl)-L-alanine. **2**: L-2-(5-Bromothienyl)alanine. **3**: H-His(3-Me)-OH

doi:10.6342/NTU201903019



Scheme 7. C_2 interface engineered by ncAAs.

(A) Engineered ferritin C_2 interface with amber suppression of ncAA **3** at R63. (B) Engineered ferritin C_2 interface with amber suppression of ncAA **2** at R63 and E67.

doi:10.6342/NTU201903019

Table 1. Class of oral antidiabetic drugs.

Class	*Mechanism*	*Representative agents*
Biguanide	Insulin sensitizer	Metformin
Incretin Mimetics		
DPP-4 inhibitors	Inhibition of degradation of GLP	Sitagliptin
GLP-1 receptor agonists	Activate GLP1 receptor	Liraglutide
SGLT2 Inhibitors	Inhibition of glucose reabsorption	Canagliflozin, Dapagliflozin
Sulfonylureas	Insulin secretion	Glimepiride, Glipizide,
Thiazolidinedione	Insulin sensitizer	Rosiglitazone, Pioglitazone

doi:10.6342/NTU201903019

Table 2. Action profile of insulin analogues.

Type of insulin	*Onset*	*Peak*	*Duration*
Rapid-acting			
Insulin aspart	10-20 mins	1-3 hrs	2-5 hrs
Insulin lispro			
Insulin glulsine			
Short-acting			
Regular insulin	30-60 mins	1-5 hrs	5-9 hrs
Intermediate-acting			
NPH insulin	60-90 mins	2-12 hrs	12-24 hrs
Long-acting			
Insulin detemir	2-4 hrs	6-14 hrs	16-20 hrs
Insulin glargine	2-4 hrs	No peak	20-24 hrs
Insulin degludec	30-90 mins	No peak	Over 42 hrs

doi:10.6342/NTU201903019

Table 3. Fe and Cu analysis of Ftn variants and F-α by ICP-MS

	Fe/ ball			*Cu/ ball*	
Standard					
Apo ferritin	1.2			-	
Equine spleen ferritin	2173.8				
Ftn variants and F-α					
Ftn	28.8	31.5	15.6	0.1	0.1
F-α	7.1	6.8		0.1	
Ftn-1x**-1**	31.6			-	
Ftn-1x**-2**	33.7				
Ftn-1x**-3**	4.1	4.1		0	
Ftn-2x**-1**	44.9			-	
Ftn-2x**-2**	4.7	7.1		0	
Ftn-2x**-3**	50.8			-	
Ftn variants and F-α after $CuCl_2$ treatment					
Ftn	43.6	15.5	21.3	18.0	
F-α	11.9	4.9	31.8	27.8	
Ftn-1x**-3**	5.2		29.5		
Ftn-2x**-2**	4.7		27.4	17.7	

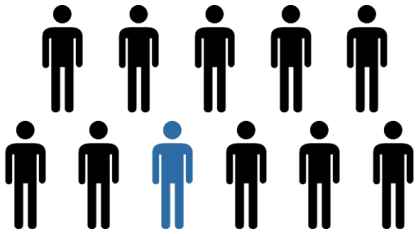
doi:10.6342/NTU201903019

Table 4. Insulin reactions catalyzed by different Cu(II) binding metalloenzymes.

Entry	*Metalloenzymes*	*Concentration*	*Conversion (%)*	*TON ($hr^{-1}ball^{-1}$)*
1	Ftn-1x-**3**	2.2 mg/ml	20	4.0
2	Ftn-2x-**2**	2.2 mg/ml	30	6.6
3	F-α-1x-**3**	2.5 mg/ml	46	9.2
4	F-α-2x-**2**	2.5 mg/ml	46	9.2

doi:10.6342/NTU201903019

(A)

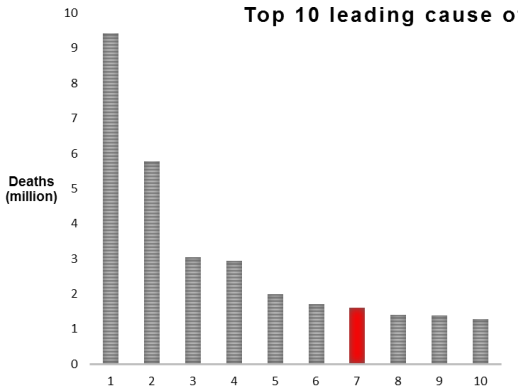


422 million diabetes patients worldwide

=

1 in 11 adult people have diabetes

(B)



Rank	Cause of deaths
1	Ischaemic heart disease
2	Stroke
3	Chronic obstructive pulmonary disease
4	Lower respiratory infections
5	Alzheimer disease and other dementias
6	Trachea, bronchus, lung cancers
7	Diabetes mellitus
8	Road injury
9	Diarrhoeal diseases
10	Tuberculosis

Source: Top 10 causes of death. *World Health Organization* (2018)

Figure 1. Illustration of diabetes prevalence around the world.

(A) According to WHO, there are 422 million diagnosed with diabetes. In other words, 1 in 11 adult people have diabetes and the prevalence is still rising. (B) In 2016, diabetes is the seventh leading cause of deaths, 1.6 million deaths are directly attributed to diabetes each year.

doi:10.6342/NTU201903019

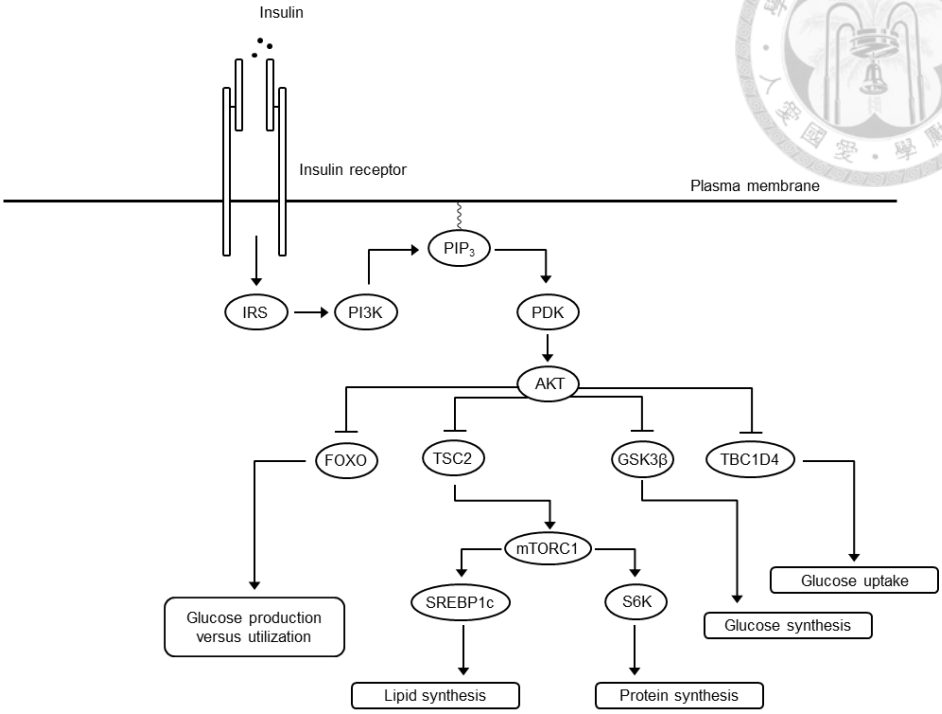


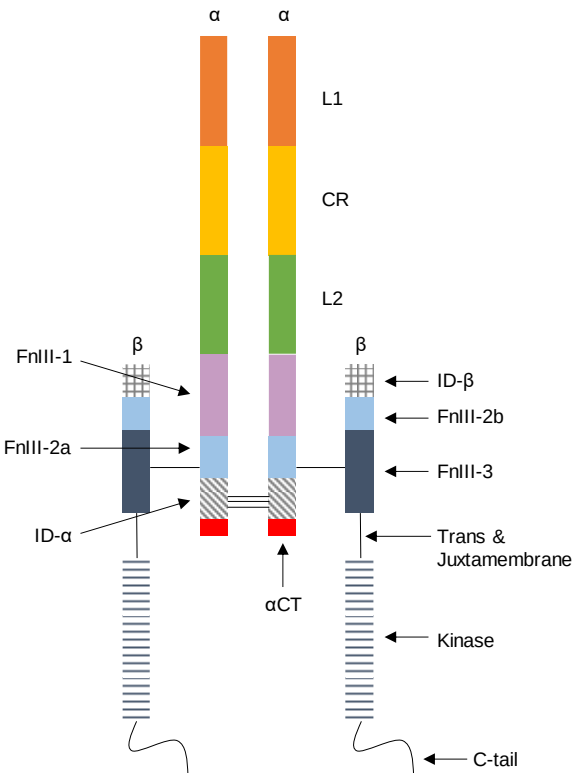
Figure 2. Activation of insulin signaling.

Insulin receptor (IR) is a type of transmembrane tyrosine kinase receptors, the tyrosine kinase and insulin receptor substrate (IRS) proteins are phosphorylated upon insulin binding and activate the PI3K-Akt pathway. Akt regulates most of insulin metabolic effects, glucose transport, lipid synthesis, gluconeogenesis, and glycogen synthesis.

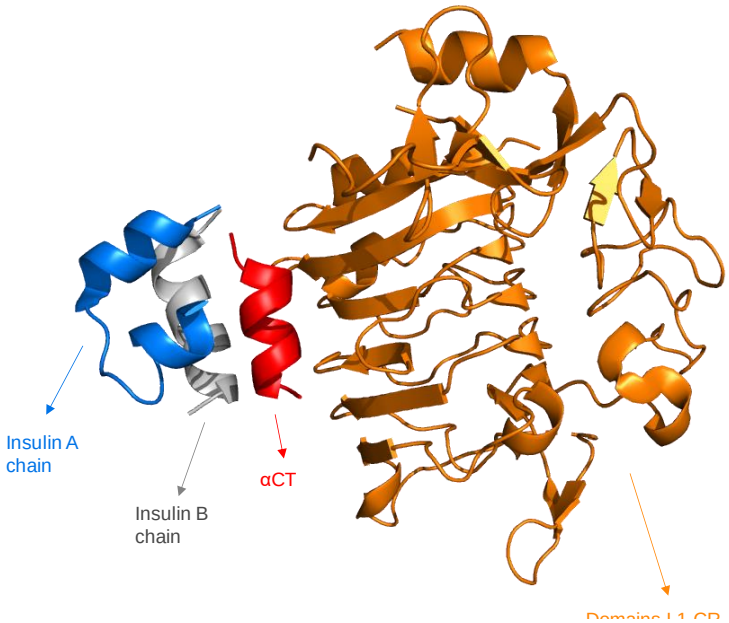
Source: Adapted from Haeusler, R. A.; McGraw, T. E.; Accili, D., Biochemical and cellular properties of insulin receptor signalling. *Nat. Rev. Mol. Cell Biol.* **2018,** *19* (1), 31-44.

doi:10.6342/NTU201903019

(A)



(B)



doi:10.6342/NTU201903019

Figure 3. Illustration of insulin primary binding site.

(A) Schematic diagram of the insulin receptor[50]: insulin receptor is a disulfide-linked $(\alpha\beta)_2$ homodimer, the α chain is ended with αCT segment (red). (B) X-Ray crystal structure of insulin and its primary binding site: insulin receptor ectodomain construct comprising domains L1-CR in complex with human insulin, αCT peptide. (PDB code: 3W11; illustrated in PyMol)

Source: (A) Adapted from McKern, N. M.; Lawrence, M. C.; Streltsov, V. A.; Lou, M. Z.; Adams, T. E.; Lovrecz, G. O.; Elleman, T. C.; Richards, K. M.; Bentley, J. D.; Pilling, P. A.; *et al. Nature* **2006**, *443* (7108), 218-21.

doi:10.6342/NTU201903019

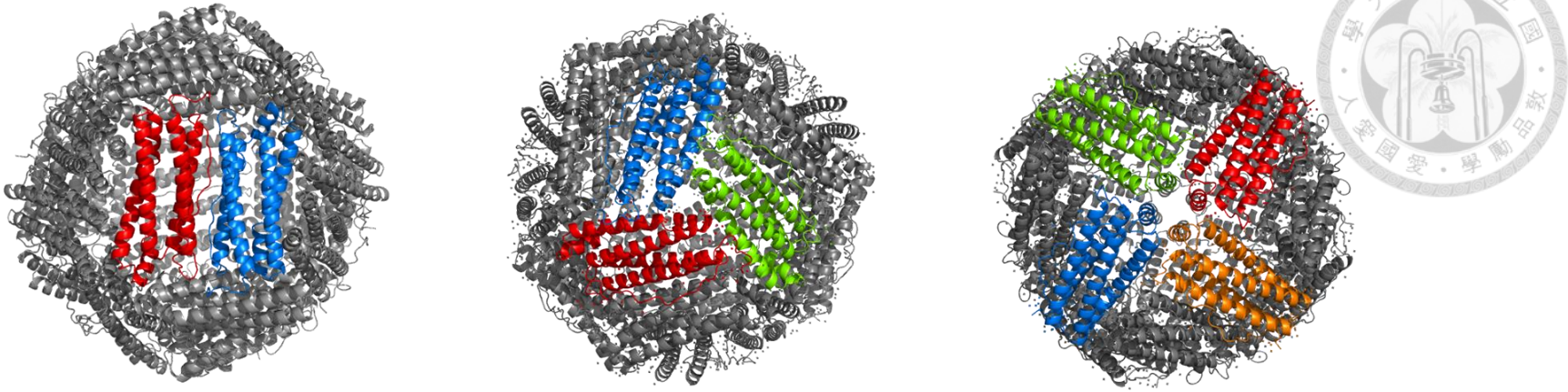


Figure 4. Illustration of ferritin exhibiting the 432 point-group symmetry.

Left graph: C_2 interface Middle graph: distribution. Middle graph: C_3 interface Right graph: C_4 interface. (PDB code: 1MFR; illustrated in PyMol)

doi:10.6342/NTU201903019

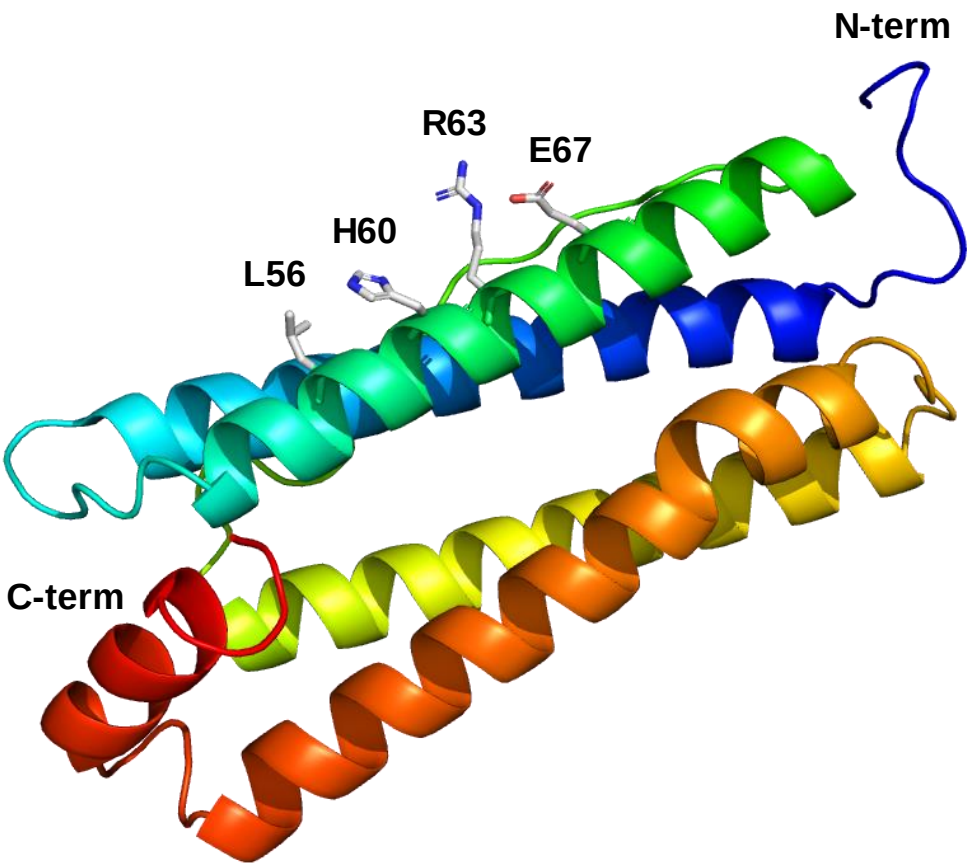


Figure 5. Structure of human heavy chain ferritin monomer.

The monomer of human heavy chain ferritin is made up of 5 helices with its N-terminus pointing toward the threefold axis pore and C-terminus located at the fourfold axis pore. The four amino acid residues on C_2 interface are L56, H60, R63, and E67. (PDB code: 1FHA; illustrated in PyMol)

doi:10.6342/NTU201903019

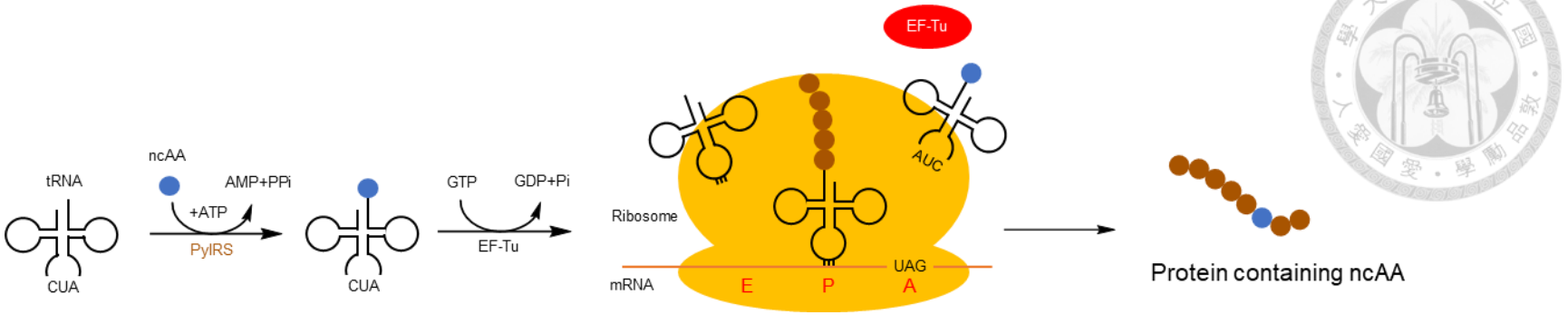


Figure 6. Illustration of expanding genetic code methodology.

In this orthogonal translation system, two components are required, a tRNA that recognizes an unused codon and a PylRS recognizes only that tRNA and the only ncAA. By using this method, more than 20 amino acids can be encoded. Take PylRS•$tRNA^{Pyl}_{CUA}$ pair for example, the PylT gene encodes tRNA with CUA anticodon to pair with TAG amber stop codon on the mRNA and then the PylRS gene produces pyrrolysyl-tRNA synthetase to charge $tRNA^{pyl}$ ncAA of interest.

doi:10.6342/NTU201903019

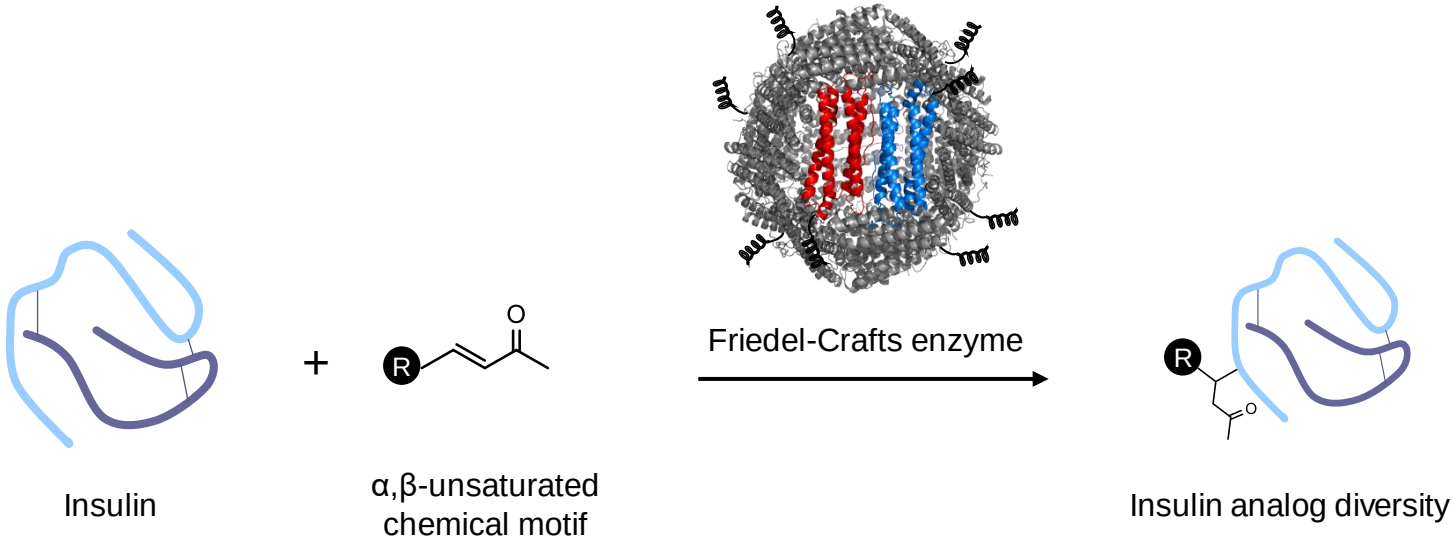


Figure 7. Specific aim.

To specifically catalyze Friedel-Crafts reaction on insulin aromatic amino acids, Cu(II) binding Ftn and F-α variants with ncAA incorporation via *Mm*PylRS-AA•tRNA$^{Pyl}_{CUA}$ pairs were designed as novel metalloenzymes. Small molecules with α,β-unsaturated carbonyl groups are then applied to modify insulin and enrich insulin with various characteristics.

doi:10.6342/NTU201903019

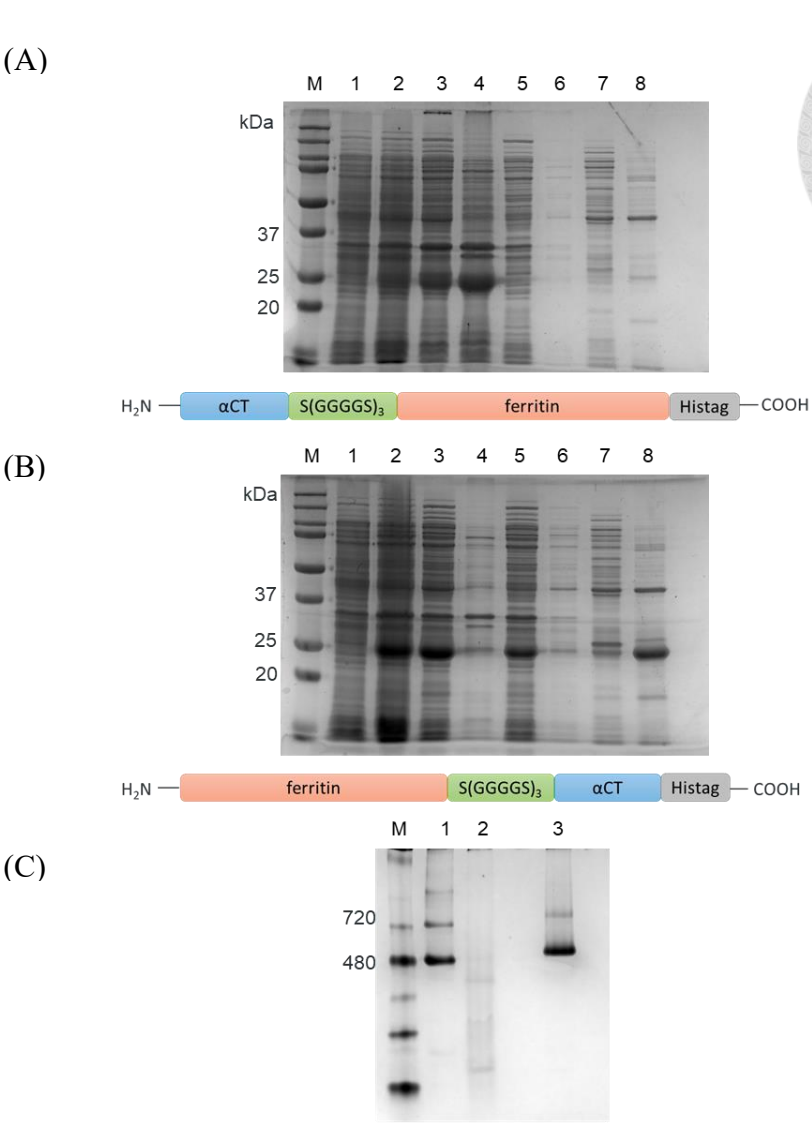


Figure 8. Purification of αCT peptide fusion ferritin by Ni^{2+}-NTA column and self-assembly study.

12% SDS-PAGE analysis of the purified His-tagged αCT peptide fusion ferritin from *E. coli* lysate under denaturing conditions, bands were visualized using Instant Blue. (A) αCT-ferritin (α-F). (B) ferritin-αCT (F-α). M: protein marker; Lane 1: uninduced cell lysate; Lane 2: whole cell sample after induction; Lane 3: supernatant of cell lysate after sonication; Lane 4: cell pellets; Lane 5: flow-through; Lane 6: wash (20 mM Tris, 100 mM NaCl, pH 7.5); Lane 7: 20 mM imidazole elution; Lane 8: 300 mM imidazole elution. (C) Native PAGE analysis of αCT peptide fusion ferritin. M: protein marker; Lane 1: wild type ferritin (531 kDa); Lane 2: soluble form of α-F; Lane 3: F-α (624 kDa).

doi:10.6342/NTU201903019

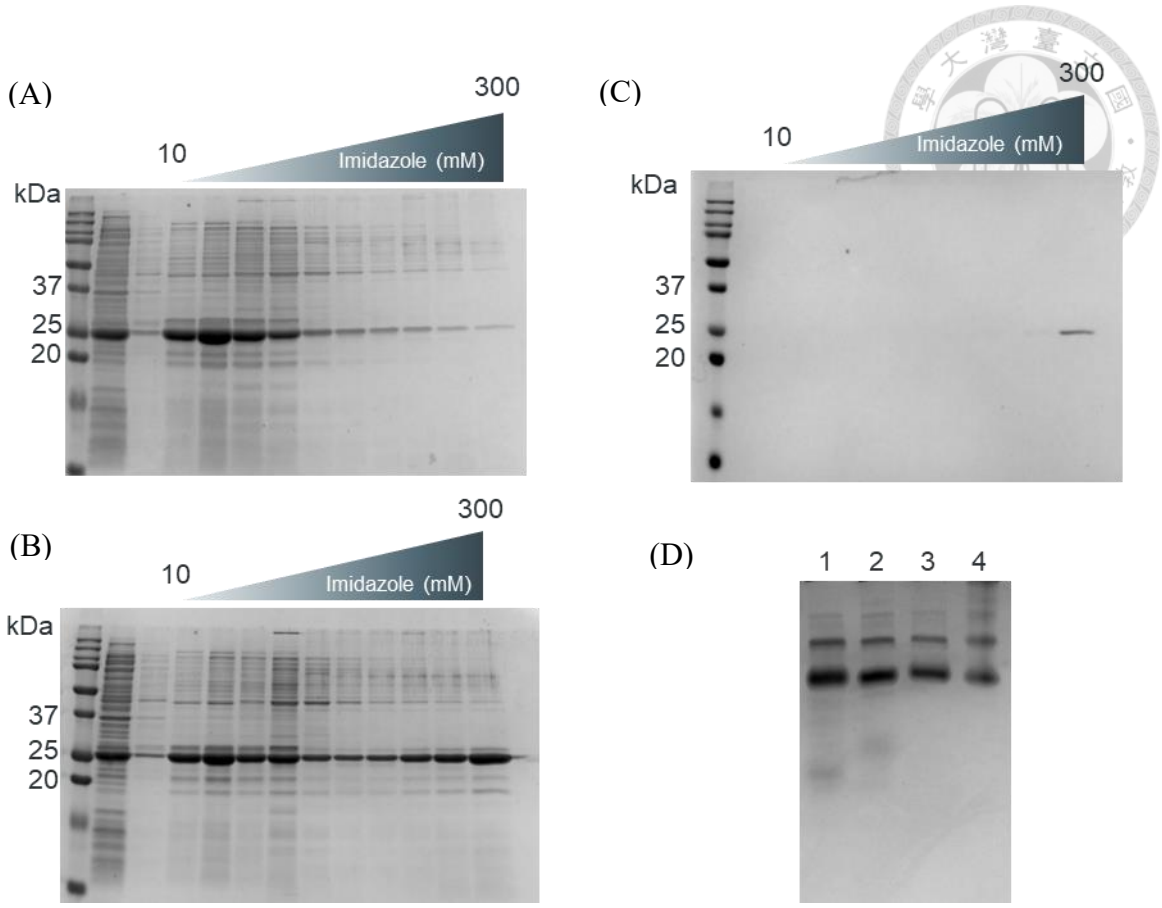


Figure 9. Purification of F-α with different linkers by Ni^{2+}-NTA column.

SDS-PAGE of F-α gradient elution starting from 10 mM to 300 mM imidazole. F-α with longer linker possesses stronger binding to Ni^{2+}-NTA column. (A) ferritin-S$(GGGGS)_1$-αCT. (B) ferritin-S$(GGGGS)_2$-αCT. (C) ferritin-3XG4S-αCT. (D) Self-assembly feature of F-α with different linkers proved by native PAGE. Lane 1: wild type ferritin (531 kDa); Lane 2: ferritin-S$(GGGGS)_1$-αCT (609 kDa); Lane 3: ferritin-S$(GGGGS)_2$-αCT (616 kDa); Lane 4: ferritin-3XG4S-αCT (624 kDa).

doi:10.6342/NTU201903019

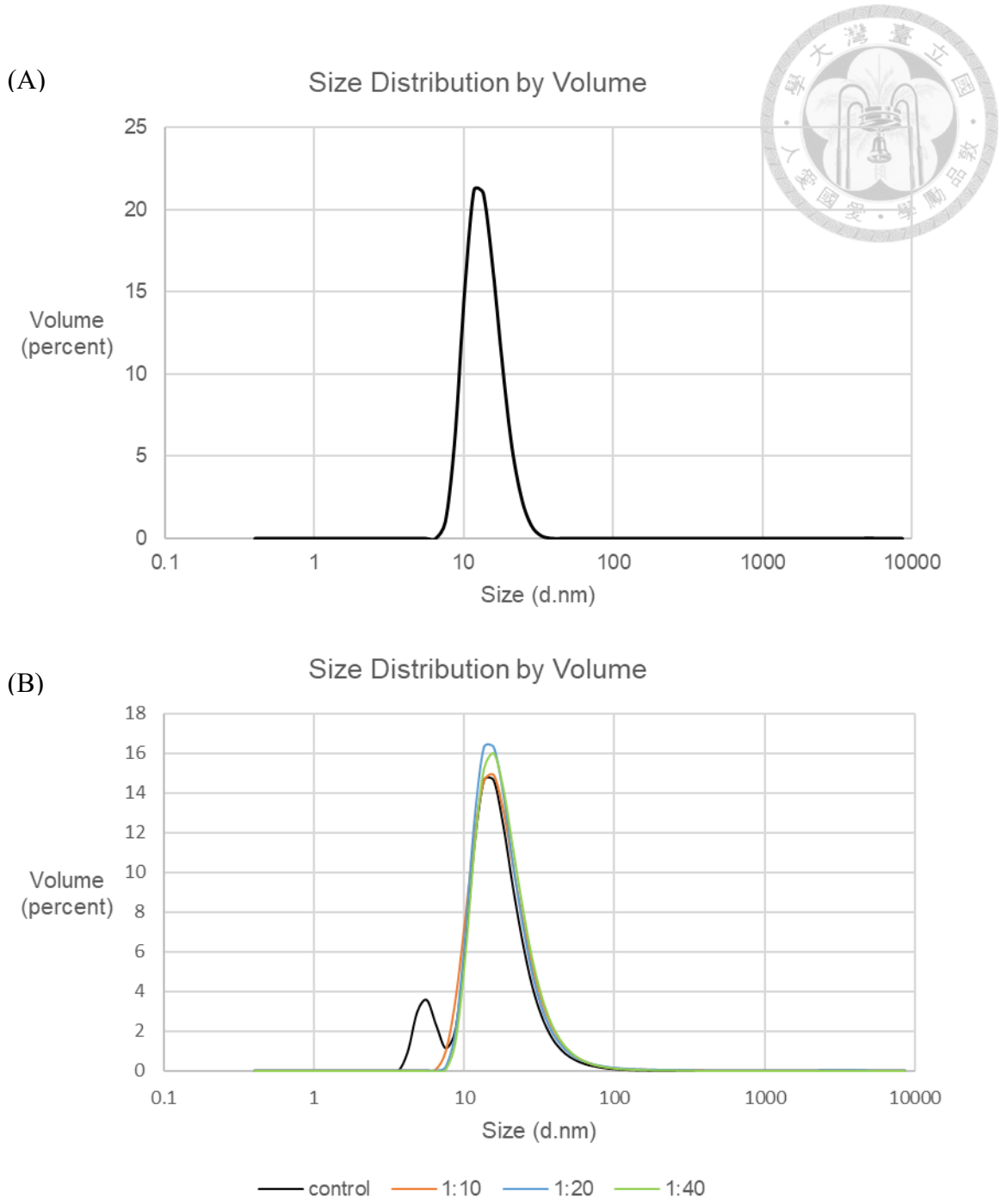


Figure 10. DLS analysis of F-α and insulin mixture.

Dynamic light scattering reveals the size distribution of wild type ferritin and F-α and the spectra are shown the volume versus diameter (d.nm). (A) Wild type ferritin (B) Mixing insulin and F-α in different molar ratio.

doi:10.6342/NTU201903019

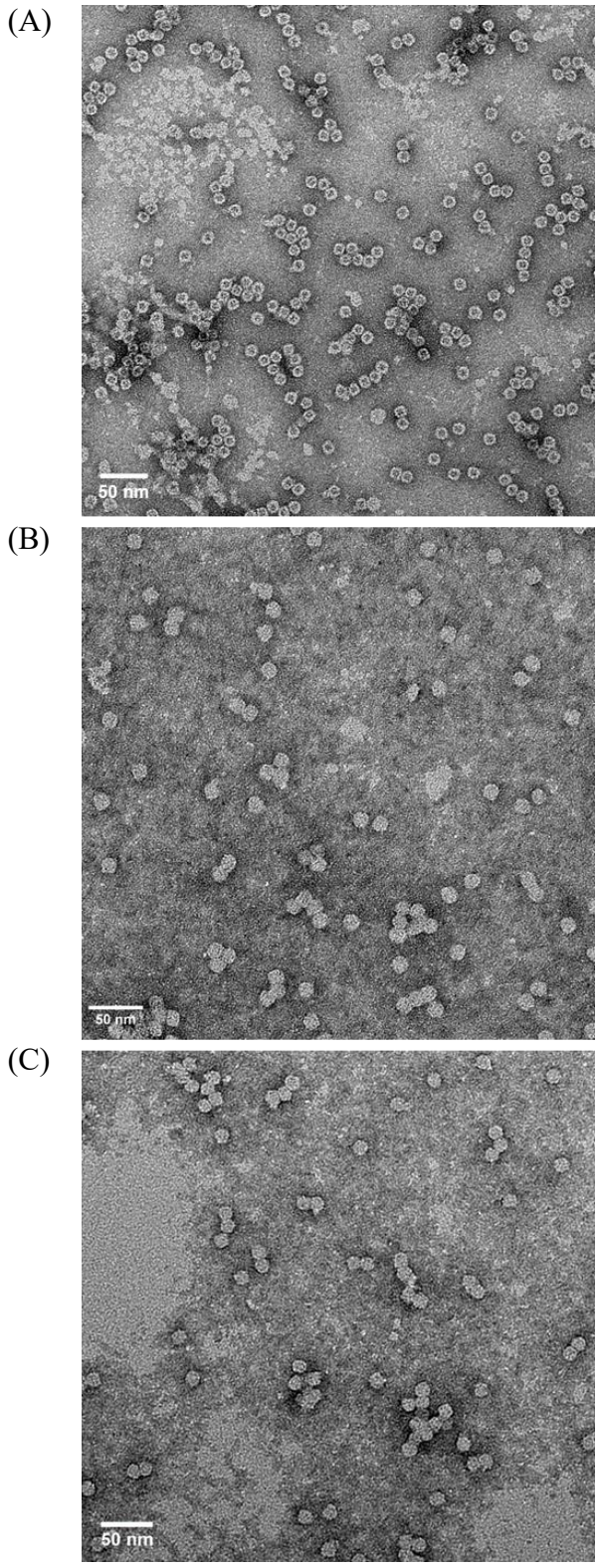


Figure 11. TEM images of F-α and insulin mixture.

Under TEM images, F-α remain assembled after mixing with insulin. (A) Wild type ferritin. (B) F-α. (C) F-α: insulin = 1:20.

doi:10.6342/NTU201903019

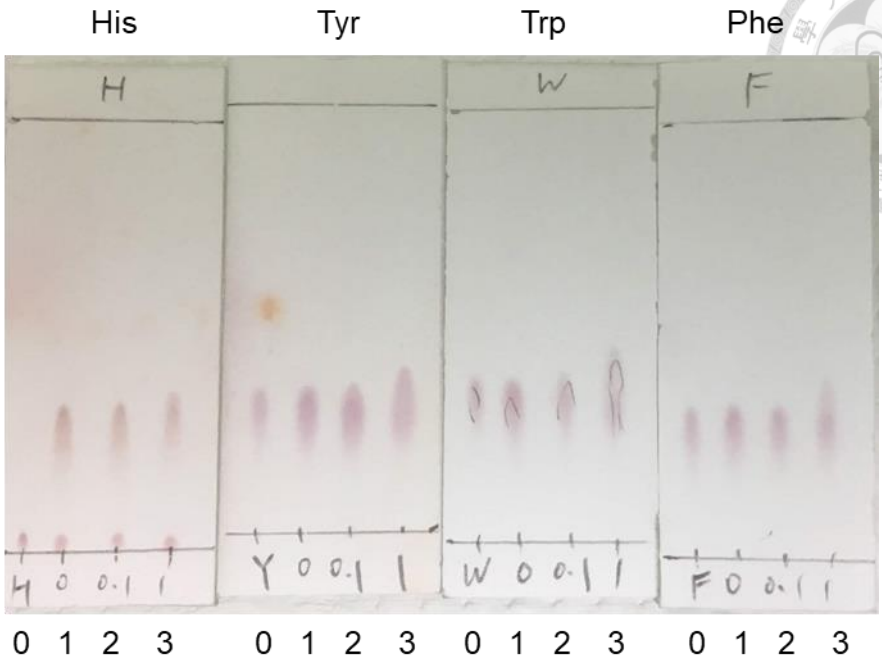


Figure 12. TLC analysis of reaction between DEEM and aromatic amino acids.

TLC had shown that histidine can react with DEEM without $CuCl_2$, the mobile phase is 70%(v/v) isopropanol. The other three amino acids require at least 1 mM $CuCl_2$ to catalyze the reaction, the mobile phase is 90%(v/v) isopropanol.

Lane 0: standard; Lane 1: reaction without $CuCl_2$; Lane 2: reaction with 0.1 mM $CuCl_2$; Lane 3: reaction with 1 mM $CuCl_2$.

doi:10.6342/NTU201903019

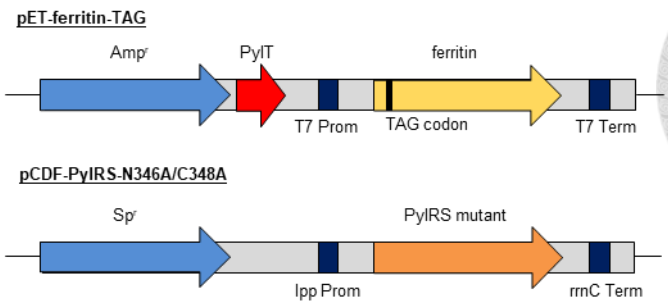


Figure 13. Two plasmids system.

pET-ferritin with Amp selection marker encodes the ferritin genes with amber mutation and $tRNA^{Pyl}_{CUA}$ with anticodon CUA. The T7 promoter regulates the transcription of target protein upon IPTG induction, and the pylT gene is flanked by the lpp promoter at 5' end and the rrnC at 3'. The other plasmid pCDF-PylRS-AA has a Sp selection marker and contains genes of *Mm*PylRS variants which is also flanked by lpp and rrnC promoters.

doi:10.6342/NTU201903019

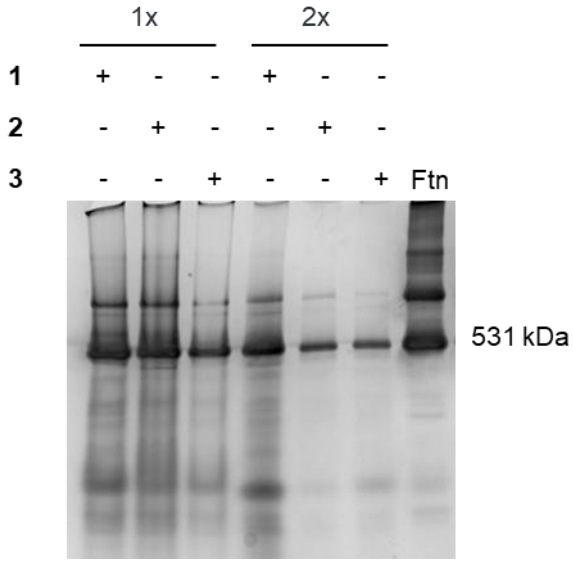


Figure 14. Native PAGE analysis of C_2 interface-engineered Ftn variants.

Ferritin containing either one or two ncAAs were analyzed by 4~15% native PAGE under non-denaturing conditions. Bands were visualized using Instant Blue. M: protein marker Ftn: wild type human heavy chain ferritin.

doi:10.6342/NTU201903019

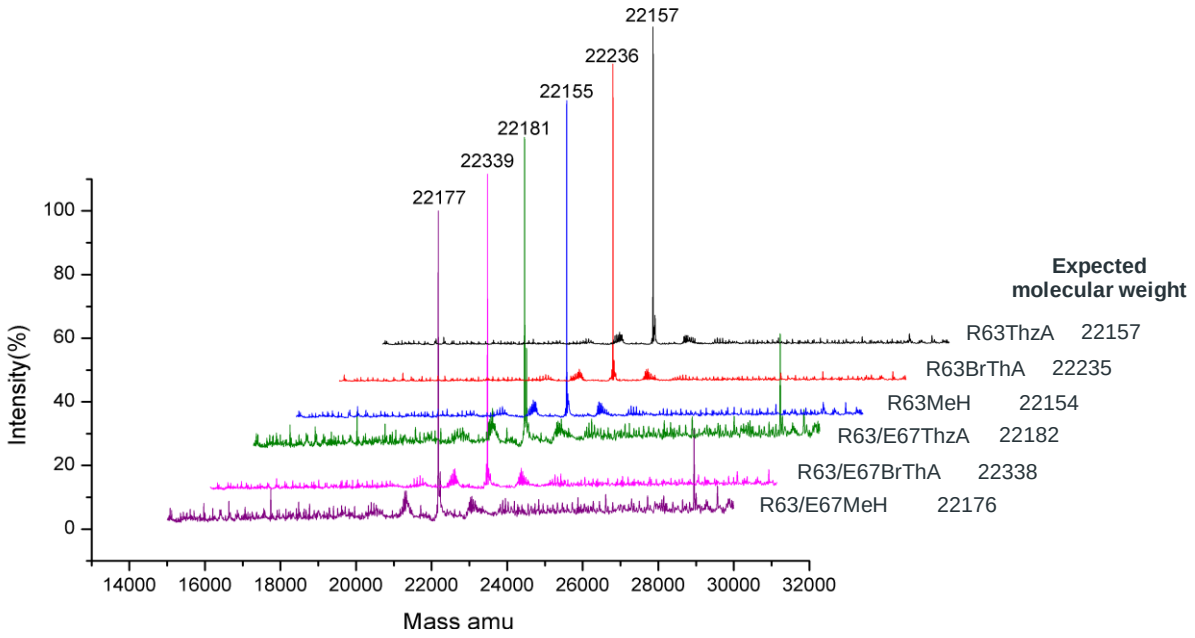


Figure 15. Molecular mass determination of C_2 interface-engineered Ftn variants.

ESI-MS spectrum of 6 C_2 interface-engineered Ftn variants. Protein were produced by *Mm*PylRS-AA•$tRNA^{Pyl}_{CUA}$ pair in *E. coli* BL21(DE3) supplemented with 1 mM IPTG, 1 mM ncAA, and M9 medium at 37°C for 16 hours. The ESI-MS spectrum before and after deconvolution are shown in **Figure S18-23.** All the found molecular weights were correct as expected.

doi:10.6342/NTU201903019

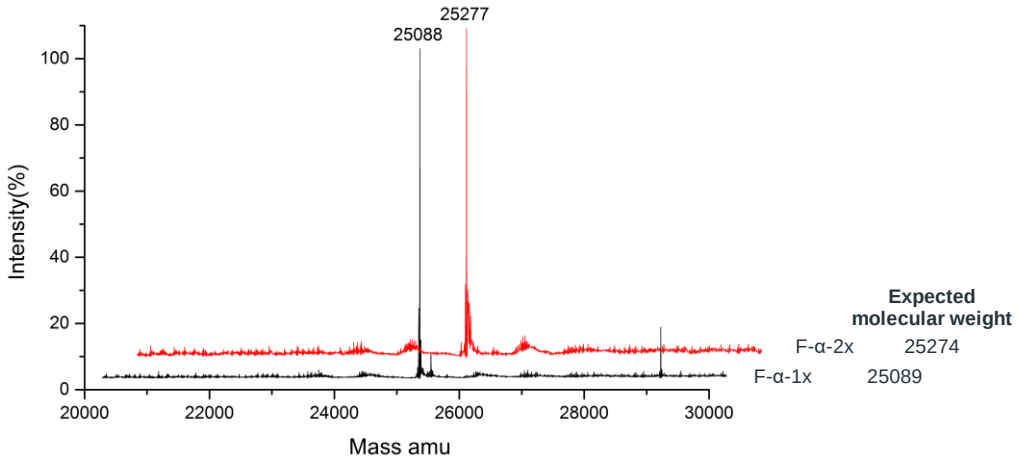


Figure 16. Molecular mass determination of C_2 interface-engineered F-α variants.

ESI-MS spectrum of 6 C_2 interface-engineered F-α variants. Protein were produced by *Mm*PylRS-AA•$tRNA^{Pyl}_{CUA}$ pair in *E. coli* BL21(DE3) supplemented with 1 mM IPTG, 1 mM ncAA, and M9 medium at 37°C for 16 hours. The ESI-MS spectrum before and after deconvolution are shown in **Figure S24-25**. All the found molecular weights were correct as expected.

doi:10.6342/NTU201903019

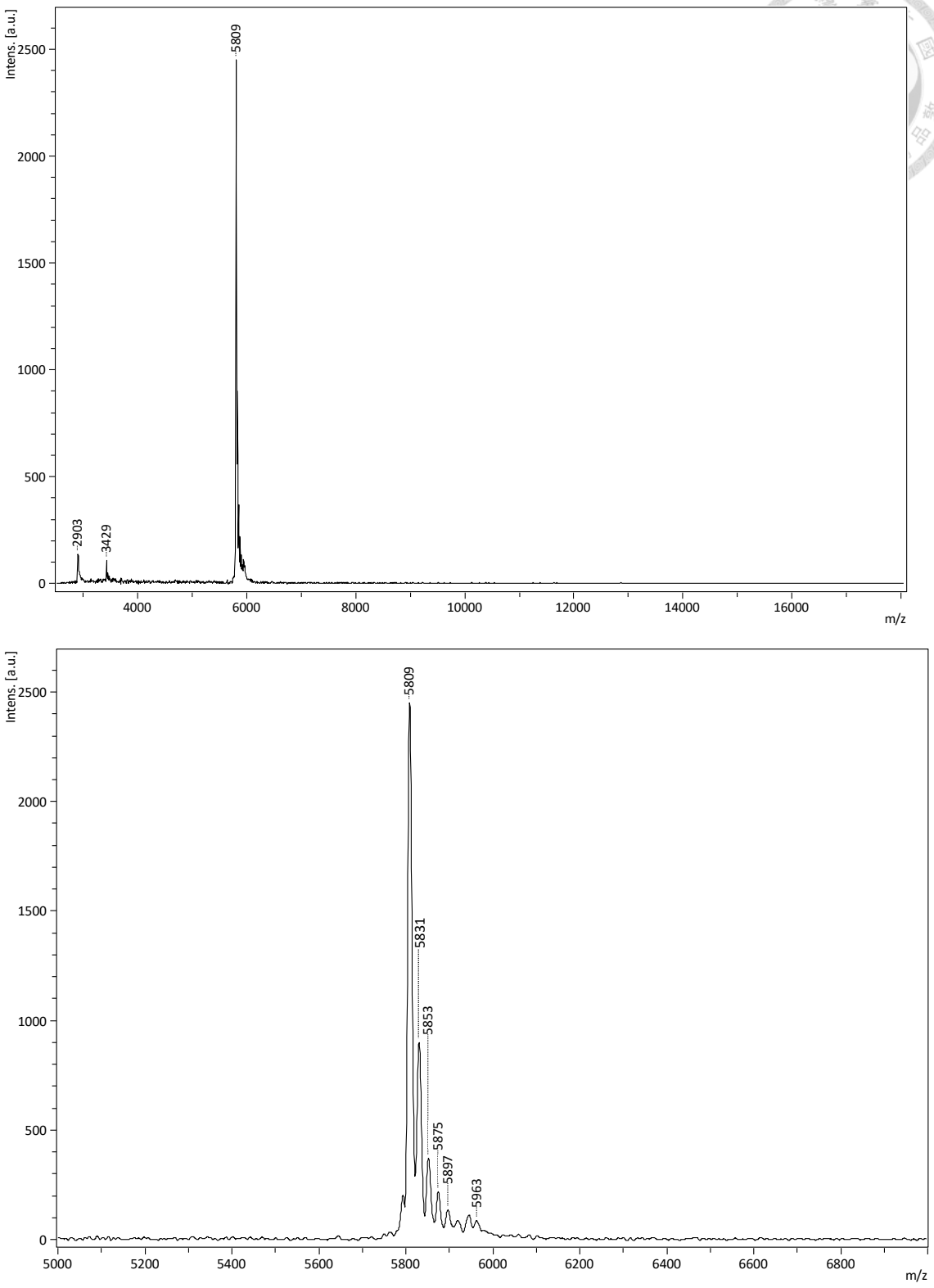


Figure 17. MALDI-TOF-MS analysis of insulin.

MALDI-TOF-MS spectrum of insulin. The calculated molecular weight is 5808 Da and the found molecular weight is 5809 Da.

doi:10.6342/NTU201903019

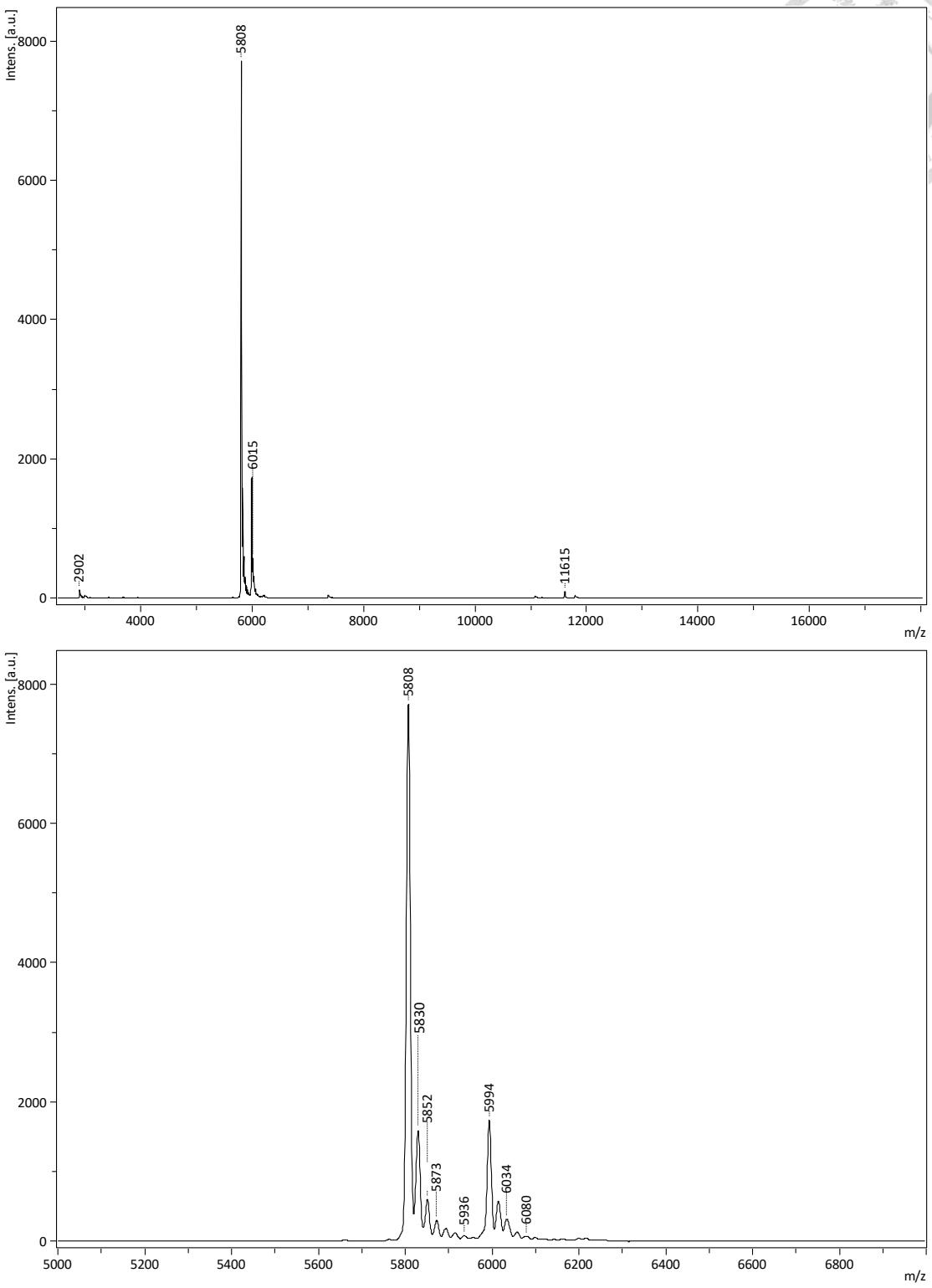


Figure 18. MALDI-TOF-MS analysis of insulin modification by Cu(II) Ftn-1x-3.

MALDI-TOF-MS spectrum of reaction mixtures with insulin and DEEM catalyzed by Ftn-1x-**3** treated with $CuCl_2$. The calculated molecular weight is 5994 Da and found molecular weights are 5808 Da and 5994 Da.

doi:10.6342/NTU201903019

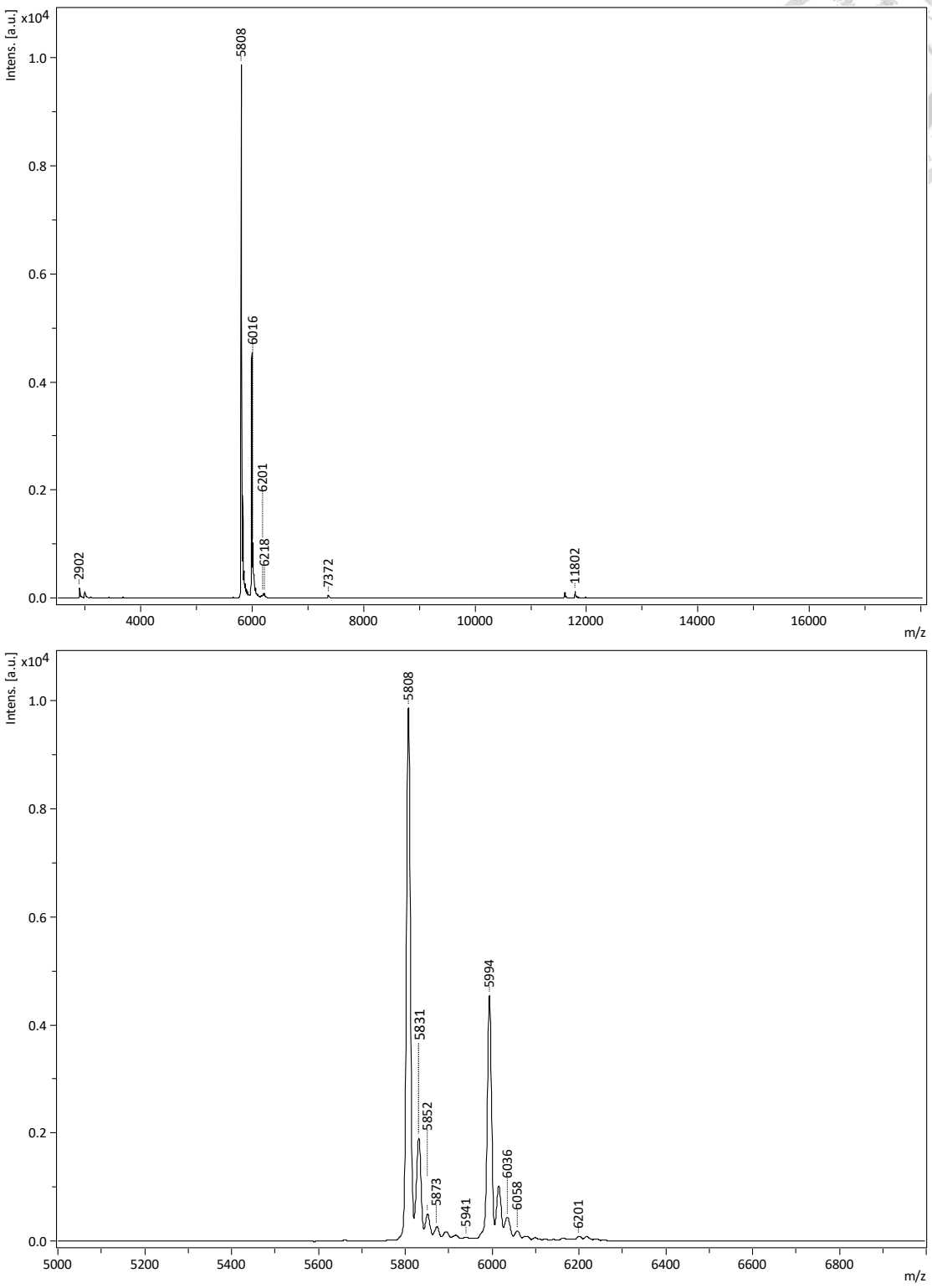


Figure 19. MALDI-TOF-MS analysis of insulin modification by Cu(II) Ftn-2x-2.

MALDI-TOF-MS spectrum of reaction mixtures with insulin and DEEM catalyzed by Ftn-2x-**2** treated with $CuCl_2$. The calculated molecular weight is 5994 Da and found molecular weights are 5808 Da and 5994 Da.

doi:10.6342/NTU201903019

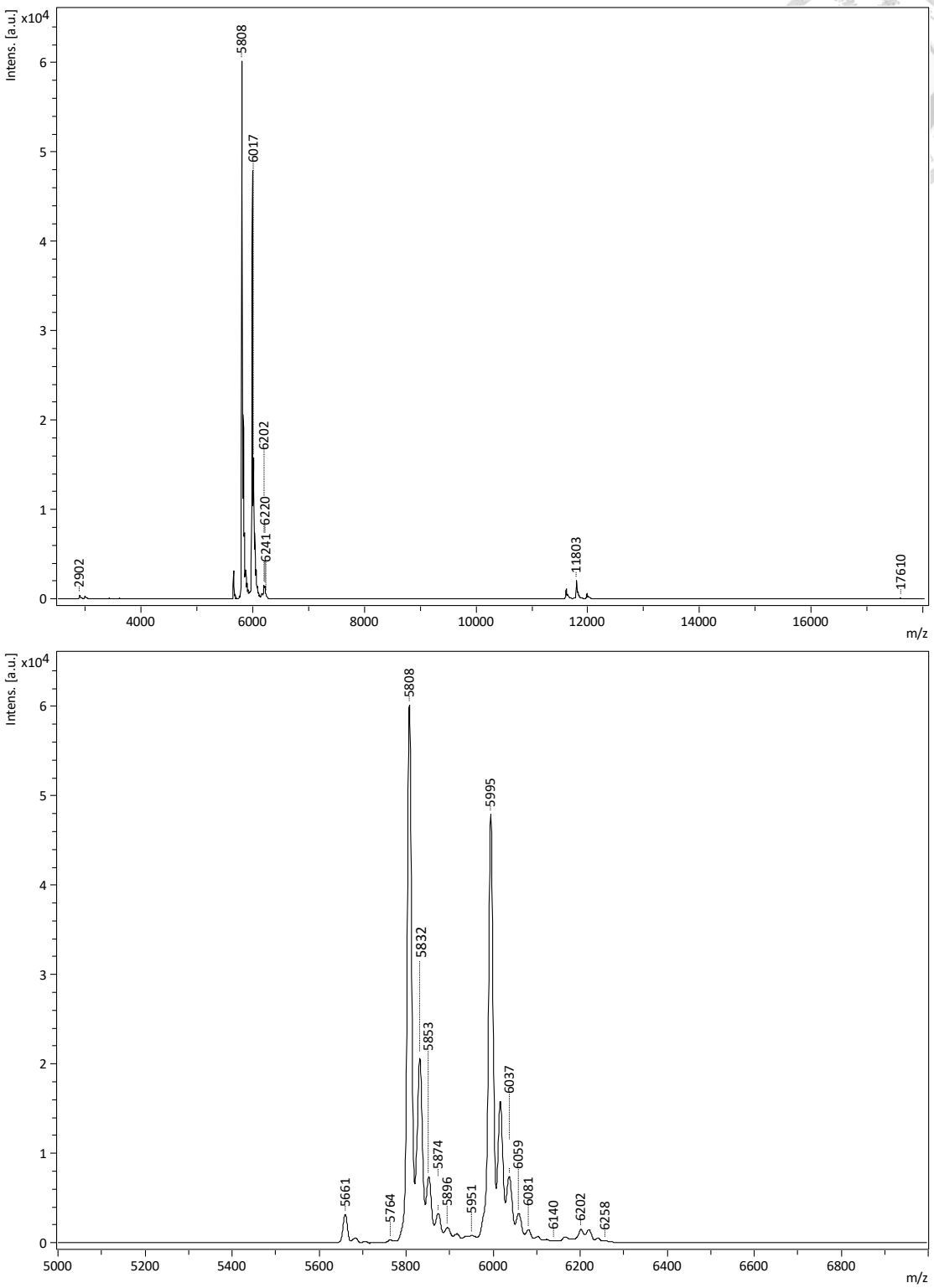


Figure 20. MALDI-TOF-MS analysis of insulin modification by Cu(II) F-α-1x-3.

MALDI-TOF-MS spectrum of reaction mixtures with insulin and DEEM catalyzed by F-α-1x-**3** treated with $CuCl_2$. The calculated molecular weight is 5994 Da and found molecular weights are 5808 Da and 5994 Da.

doi:10.6342/NTU201903019

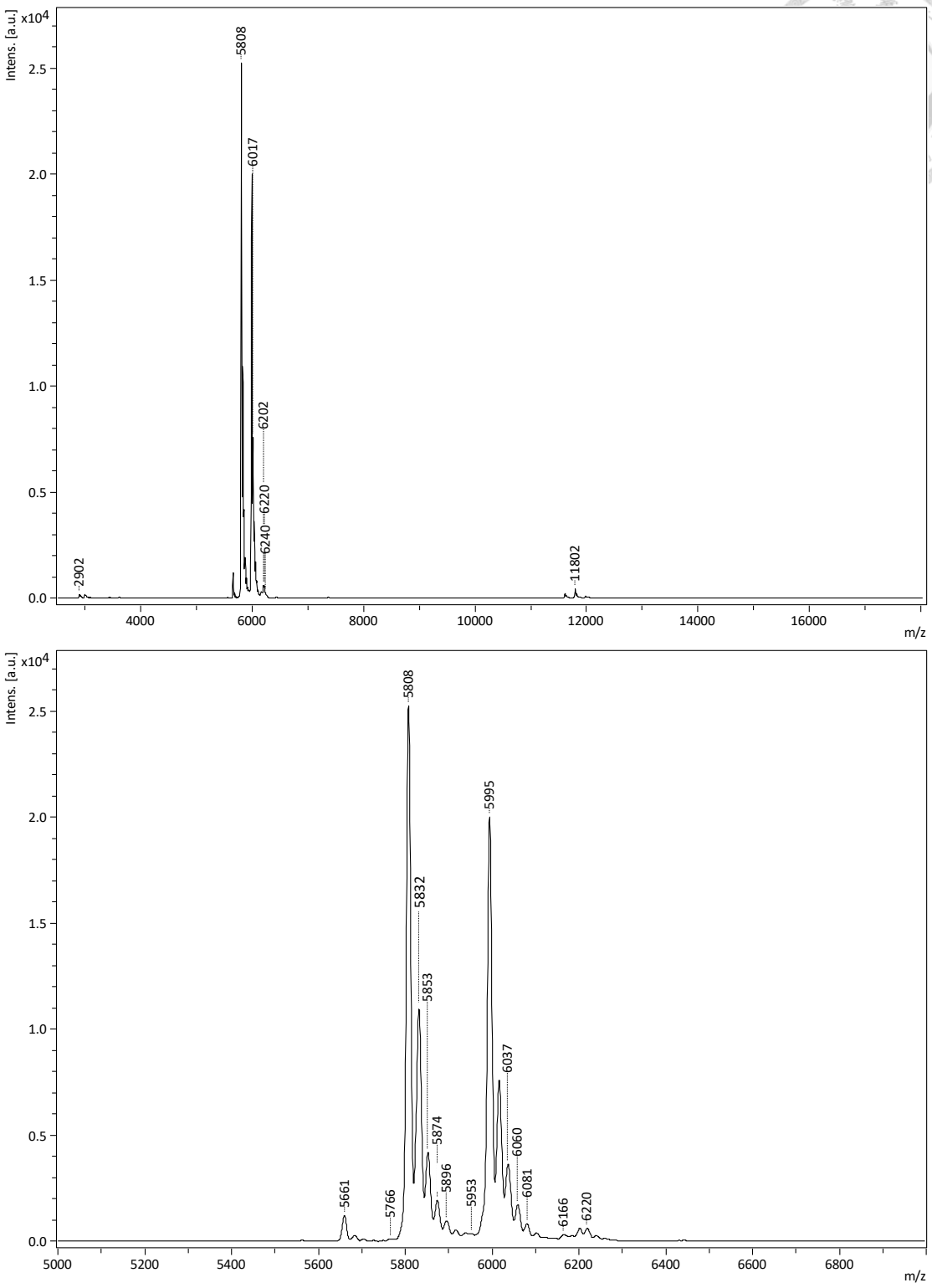
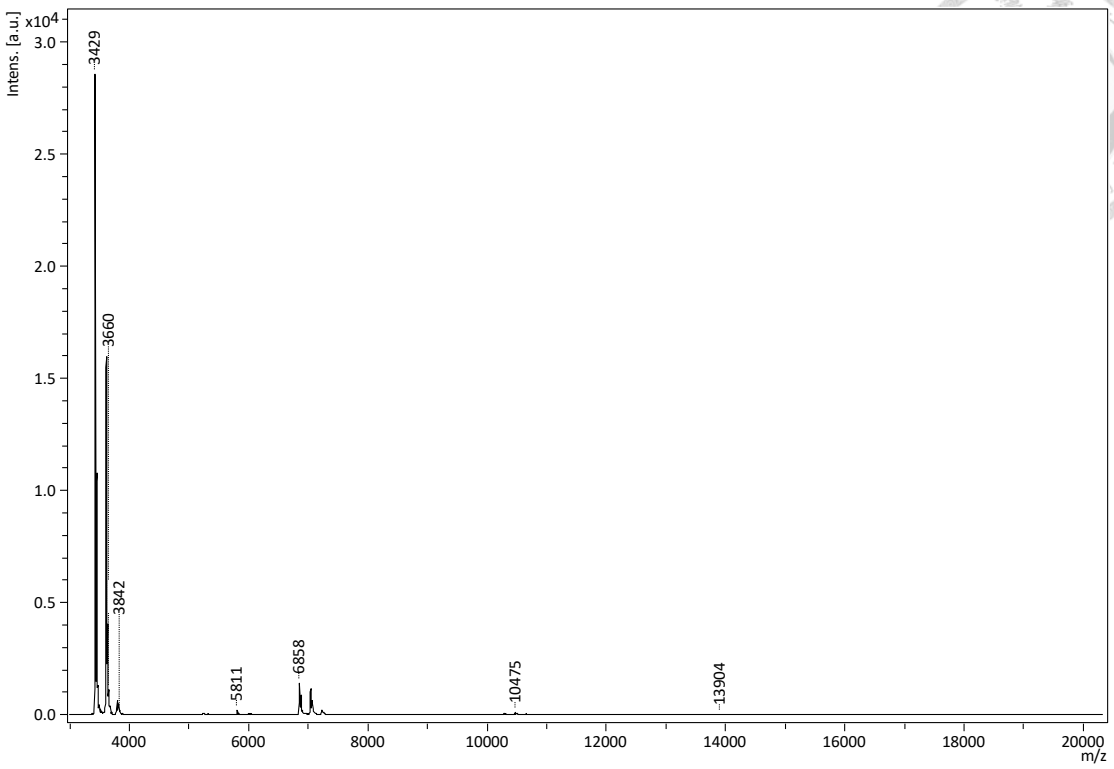
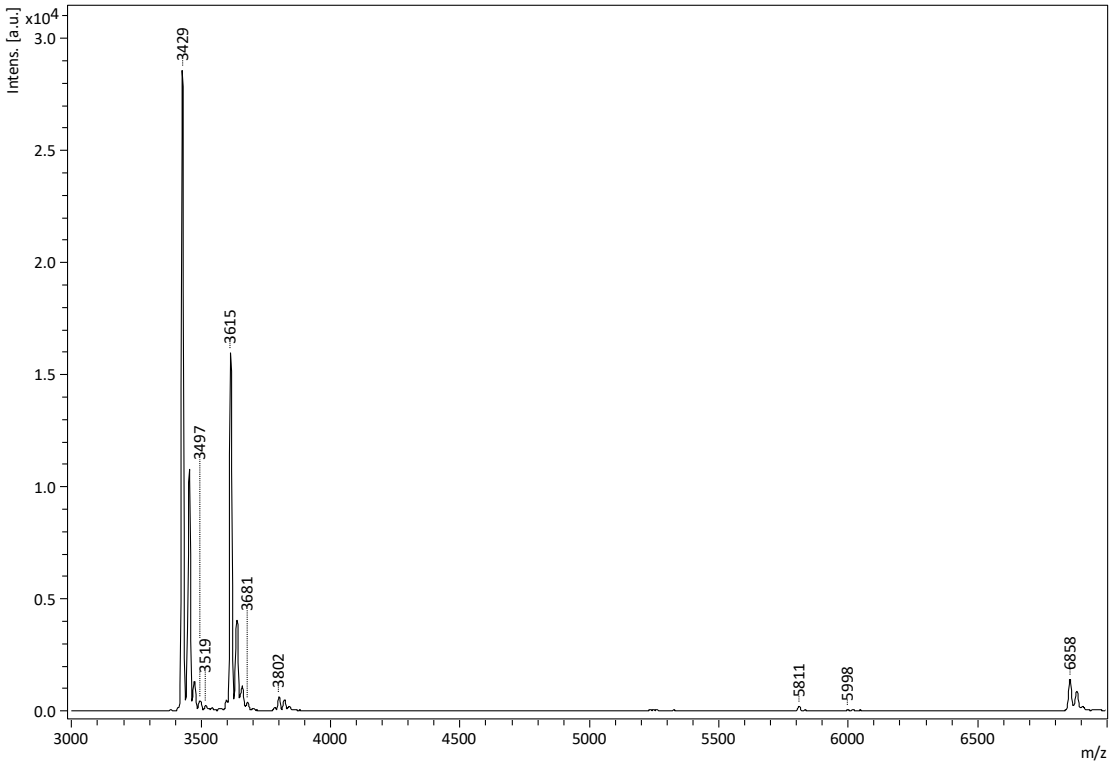


Figure 21. MALDI-TOF-MS analysis of insulin modification by Cu(II) F-α-2x-2.

MALDI-TOF-MS spectrum of reaction mixtures with insulin and DEEM catalyzed by F-α-2x-**2** treated with $CuCl_2$. The calculated molecular weight is 5994 Da and found molecular weights are 5808 Da and 5995 Da.

doi:10.6342/NTU201903019



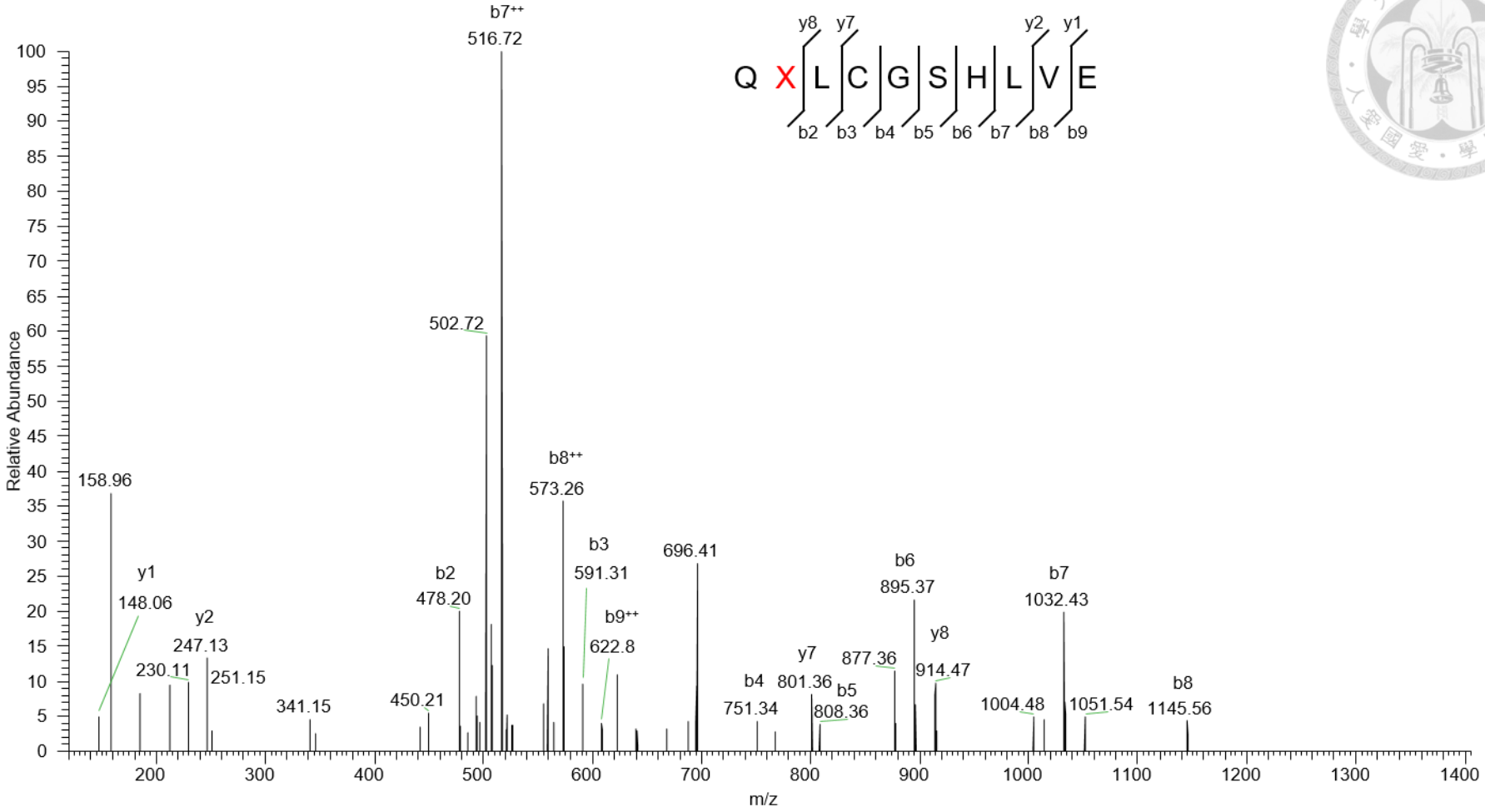


doi:10.6342/NTU201903019

Figure 22. Molecular mass determination of insulin modification after disulfide bonds reduction.

MALDI-TOF-MS spectrum of reduced insulin modified by DEEM. The reaction mixture was treated with DTT under 37ºC for 30 minutes and the precipitation is dissolved in 20% (v/v) acetonitrile, 0.1% formic acid. The calculated insulin B chain molecular weight is 3430 Da and the theoretical modified insulin B chain is 3630 Da. Found molecular weights are 3429 Da, 3615 Da, and 3802 Da.

doi:10.6342/NTU201903019



doi:10.6342/NTU201903019

Figure 23. MALDI-TOF-MS/MS analysis of insulin modified by Cu(II) binding F-α-2x-2.

Fragment $Q^{B4}XLCGSHLVE^{B13}$, X represents histidine modified by DEEM, was found in tandem mass spectrometry. The reaction of insulin and DEEM was catalyzed by F-α-2x-**2** treated with $CuCl_2$. The protein was in-gel digested by glutamyl endopeptidase GluV8.

doi:10.6342/NTU201903019

Reference

(1) 衛生福利部國民健康署 世界糖尿病日由來與國內宣導響應. https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=1090&pid=6426 (accessed Aug 3).

(2) World Health Organization The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (accessed Aug 3).

(3) Steiner, D. F.; Park, S. Y.; Stoy, J.; Philipson, L. H.; Bell, G. I., A brief perspective on insulin production. *Diabetes Obes. Metab.* **2009,** *11 Suppl 4*, 189-96.

(4) Tibaldi, J. M., Evolution of insulin: from human to analog. *Am. J. Med.* **2014,** *127* (10 Suppl), S25-38.

(5) Zaykov, A. N.; Mayer, J. P.; DiMarchi, R. D., Pursuit of a perfect insulin. *Nat. Rev. Drug. Discov.* **2016,** *15* (6), 425-39.

(6) Havelund, S. P., A.; Ribel, U.; Jonassen, I.; Vølund, A.; Markussen, J.; Kurtzhals, P. , The Mechanism of Protraction of Insulin Detemir, a Long-Acting, Acylated Analog of Human Insulin. *Pharm. Res.* **2004,** *21* (8), 1498-1504.

(7) Jonassen, I.; Havelund, S.; Hoeg-Jensen, T.; Steensgaard, D. B.; Wahlund, P. O.; Ribel, U., Design of the novel protraction mechanism of insulin degludec, an ultra-long-acting basal insulin. *Pharm. Res.* **2012,** *29* (8), 2104-14.

(8) Chou, D. H.; Webber, M. J.; Tang, B. C.; Lin, A. B.; Thapa, L. S.; Deng, D.; Truong, J. V.; Cortinas, A. B.; Langer, R.; Anderson, D. G., Glucose-responsive insulin activity by covalent modification with aliphatic phenylboronic acid conjugates. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112* (8), 2401-6.

(9) Roberts, B. K.; Wang, X.; Rosendahl, M. S. *The in Vitro and in Vivo Pharmacology of AB101, a Potential Once-Weekly Basal Subcutaneous Insulin*; AntriaBio, Inc.: 2015.

doi:10.6342/NTU201903019

(10) Boucher, J.; Kleinridders, A.; Kahn, C. R., Insulin receptor signaling in normal and insulin-resistant states. *Cold Spring Harb. Perspect Biol.* **2014,** *6* (1).

(11) Menting, J. G.; Whittaker, J.; Margetts, M. B.; Whittaker, L. J.; Kong, G. K.; Smith, B. J.; Watson, C. J.; Zakova, L.; Kletvikova, E.; Jiracek, J.; Chan, S. J.; Steiner, D. F.; Dodson, G. G.; Brzozowski, A. M.; Weiss, M. A.; Ward, C. W.; Lawrence, M. C., How insulin engages its primary binding site on the insulin receptor. *Nature* **2013,** *493* (7431), 241-5.

(12) Weis, F.; Menting, J. G.; Margetts, M. B.; Chan, S. J.; Xu, Y.; Tennagels, N.; Wohlfart, P.; Langer, T.; Muller, C. W.; Dreyer, M. K.; Lawrence, M. C., The signalling conformation of the insulin receptor ectodomain. *Nat. Commun.* **2018,** *9* (1), 4420.

(13) Lallemand, C.; Simansour, M.; Liang, F.; Ferrando-Miguel, R.; Tovey, M. G., Establishment of a Novel iLite™ Reporter Gene Assay for the Quantification of the Activity and Neutralizing Antibody response to Insulin Products. SAS, B., Ed. 2016.

(14) Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M., Design of functional metalloproteins. *Nature* **2009,** *460* (7257), 855-62.

(15) Kaplan, J.; Degrado, W. F., De novo design of catalytic proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (32), 11566–11570.

(16) Prier, C. K.; Zhang, R. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H., Enantioselective, intermolecular benzylic C-H amination catalysed by an engineered iron-haem enzyme. *Nat. Chem.* **2017,** *9* (7), 629-634.

(17) Kan, S. B. J.; Lewis, R. D. C., K.; Arnold, F. H., Directed Evolution of Cytochrome c for Carbon–Silicon Bond Formation: Bringing Silicon to Life. *Science* **2016,** *354* (6315), 1048–1051.

(18) Chen, K.; Huang, X.; Kan, S. B. J.; Zhang, R. K.; Arnold, F. H., Enzymatic

doi:10.6342/NTU201903019

construction of highly strained carbocycles. *Science* **2018,** *360* (6384), 71-75.

(19) Drienovska, I.; Rioz-Martinez, A.; Draksharapu, A.; Roelfes, G., Novel artificial metalloenzymes by in vivo incorporation of metal-binding unnatural amino acids. *Chem. Sci.* **2015,** *6* (1), 770-776.

(20) Antos, J. M.; Francis, M. B., Transition metal catalyzed methods for site-selective protein modification. *Curr Opin Chem Biol.* **2006,** *10* (3), 253-62.

(21) Tilley, S. D.; Francis, M. B., Tyrosine-selective protein alkylation using pi-allylpalladium complexes. *J. Am. Chem. Soc.* **2006,** *128* (4), 1080-1.

(22) Vohidov, F.; Coughlin, J. M.; Ball, Z. T., Rhodium(II) metallopeptide catalyst design enables fine control in selective functionalization of natural SH3 domains. *Angew. Chem. Int. Ed.* **2015,** *54* (15), 4587-91.

(23) Poulsen, T. B.; Jorgensen, K. A., Catalytic asymmetric Friedel-Crafts alkylation reactions--copper showed the way. *Chem. Rev.* **2008,** *108* (8), 2903-15.

(24) Zhang, Y.; Orner, B. P., Self-assembly in the ferritin nano-cage protein superfamily. *Int. J. Mol. Sci.* **2011,** *12* (8), 5406-21.

(25) Truffi, M.; Fiandra, L.; Sorrentino, L.; Monieri, M.; Corsi, F.; Mazzucchelli, S., Ferritin nanocages: A biological platform for drug delivery, imaging and theranostics in cancer. *Pharmacol. Res.* **2016,** *107*, 57-65.

(26) Theil, E. C.; Behera, R. K.; Tosha, T., Ferritins for Chemistry and for Life. *Coord. Chem. Rev.* **2013,** *257* (2), 579-586.

(27) Pozzi, C.; Di Pisa, F.; Bernacchioni, C.; Ciambellotti, S.; Turano, P.; Mangani, S., Iron binding to human heavy-chain ferritin. *Acta Cryst.* **2015,** *71* (Pt 9), 1909-20.

(28) Huard, D. J.; Kane, K. M.; Tezcan, F. A., Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. *Nat. Chem. Biol.* **2013,** *9* (3), 169-76.

(29) Cooper, G. M.; Hausman, R. E., *The cell: a molecular approach*. 2 ed.; ASM press:

doi:10.6342/NTU201903019

Washington, DC, 2000.

(30) Böck, A.; Stadtman, T., Selenocysteine, a highly specific component of certain enzymes, is incorporated by a UGA-directed co-translational mechanism. *BioFactors* **1988,** *1* (3), 245-250.

(31) Gaston, M. A.; Jiang, R.; Krzycki, J. A., Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* **2011,** *14* (3), 342-349.

(32) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G., A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **1989,** *244* (4901), 182-188.

(33) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001,** *292* (5516), 498-500.

(34) Wang, L.; Xie, J.; Schultz, P. G., Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **2006,** *35*, 225-49.

(35) O'Donoghue, P.; Ling, J.; Wang, Y. S.; Soll, D., Upgrading protein synthesis for synthetic biology. *Nat. Chem. Biol.* **2013,** *9* (10), 594-8.

(36) Wang, Q.; Parrish, A. R.; Wang, L., Expanding the genetic code for biological studies. *Chem Biol.* **2009,** *16* (3), 323-36.

(37) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W., Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **2008,** *4* (4), 232-4.

(38) Xie, J.; Supekova, L.; Schultz, P. G., A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli. *ACS Chem. Biol.* **2007,** *2* (7), 474-478.

(39) Liu, H.; Wang, L.; Brock, A.; Wong, C.-H.; Schultz, P. G., A method for the generation of glycoprotein mimetics. *J. Am. Chem. Soc.* **2003,** *125* (7), 1702-1703.

(40) Liu, W. R.; Wang, Y. S.; Wan, W., Synthesis of proteins with defined posttranslational

doi:10.6342/NTU201903019

modifications using the genetic noncanonical amino acid incorporation approach. *Mol. Biosyst.* **2011,** *7* (1), 38-47.

(41) Yang, B.; Tang, S.; Ma, C.; Li, S. T.; Shao, G. C.; Dang, B.; DeGrado, W. F.; Dong, M. Q.; Wang, P. G.; Ding, S.; Wang, L., Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions. *Nat. Commun.* **2017,** *8* (1), 2240.

(42) Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G., Addition of the keto functional group to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100* (1), 56-61.

(43) Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G., Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J. Am. Chem. Soc.* **2003,** *125* (39), 11782-11783.

(44) Wang, Y. S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R., A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J. Am. Chem. Soc.* **2012,** *134* (6), 2950-3.

(45) Takimoto, J. K.; Dellas, N.; Noel, J. P.; Wang, L., Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids. *ACS Chem. Biol.* **2011,** *6* (7), 733-43.

(46) Xiao, H.; Peters, F. B.; Yang, P. Y.; Reed, S.; Chittuluru, J. R.; Schultz, P. G., Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase. *ACS Chem. Biol.* **2014,** *9* (5), 1092-6.

(47) Poppe, L.; Retey, J., Friedel-Crafts-type mechanism for the enzymatic elimination of ammonia from histidine and phenylalanine. *Angew. Chem. Int. Ed.* **2005,** *44* (24), 3668-88.

doi:10.6342/NTU201903019

(48) Shevchenko, A.; Tomas, H.; Havli, J.; Olsen, J. V.; Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006,** *1* (6), 2856.

(49) Trinh, R.; Gurbaxani, B.; Morrison, S. L.; Seyfzadeh, M., Optimization of codon pair use within the $(GGGGS)_3$ linker sequence results in enhanced protein expression. *Mol. immunol.* **2004,** *40* (10), 717-722.

(50) McKern, N. M.; Lawrence, M. C.; Streltsov, V. A.; Lou, M. Z.; Adams, T. E.; Lovrecz, G. O.; Elleman, T. C.; Richards, K. M.; Bentley, J. D.; Pilling, P. A.; Hoyne, P. A.; Cartledge, K. A.; Pham, T. M.; Lewis, J. L.; Sankovich, S. E.; Stoichevska, V.; Da Silva, E.; Robinson, C. P.; Frenkel, M. J.; Sparrow, L. G.; Fernley, R. T.; Epa, V. C.; Ward, C. W., Structure of the insulin receptor ectodomain reveals a folded-over conformation. *Nature* **2006,** *443* (7108), 218-21.

doi:10.6342/NTU201903019

Appendix

MATRIX SCIENCE Mascot Search Results

Protein View

Match to: **gi|2018052401** Score: **90** Expect: **1.8e-008**
□\CT□VhFerritin

Nominal mass (M_r): **25381**; Calculated pI value: **5.65**
NCBI BLAST search of gi|2018052401 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)
Cleavage by TrypChymo: cuts C-term side of FYWLKR unless next residue is P
Number of mass values searched: **49**
Number of mass values matched: **15**
Sequence Coverage: **49%**

Matched peptides shown in **Bold Red**

```
     1 MTFEDYLHNV VFVPRPSSGG GGSGGGGSGG GGSTTASTSQ VRQNYHQDSE
    51 AAINRQINLE LYASYVYLSM SYYFDRDDVA LKNFAKYFLH QSHEEREHAE
   101 KLMKLQNQRG GRIFLQDIKK PDCDDWESGL NAMECALHLE KNVNQSLLEL
   151 HKLATDKNDP HLCDFIETHY LNEQVKAIKE LGDHVTNLRK MGAPESGLAE
   201 YLFDKHTLGD SDNESELHHH HHH
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

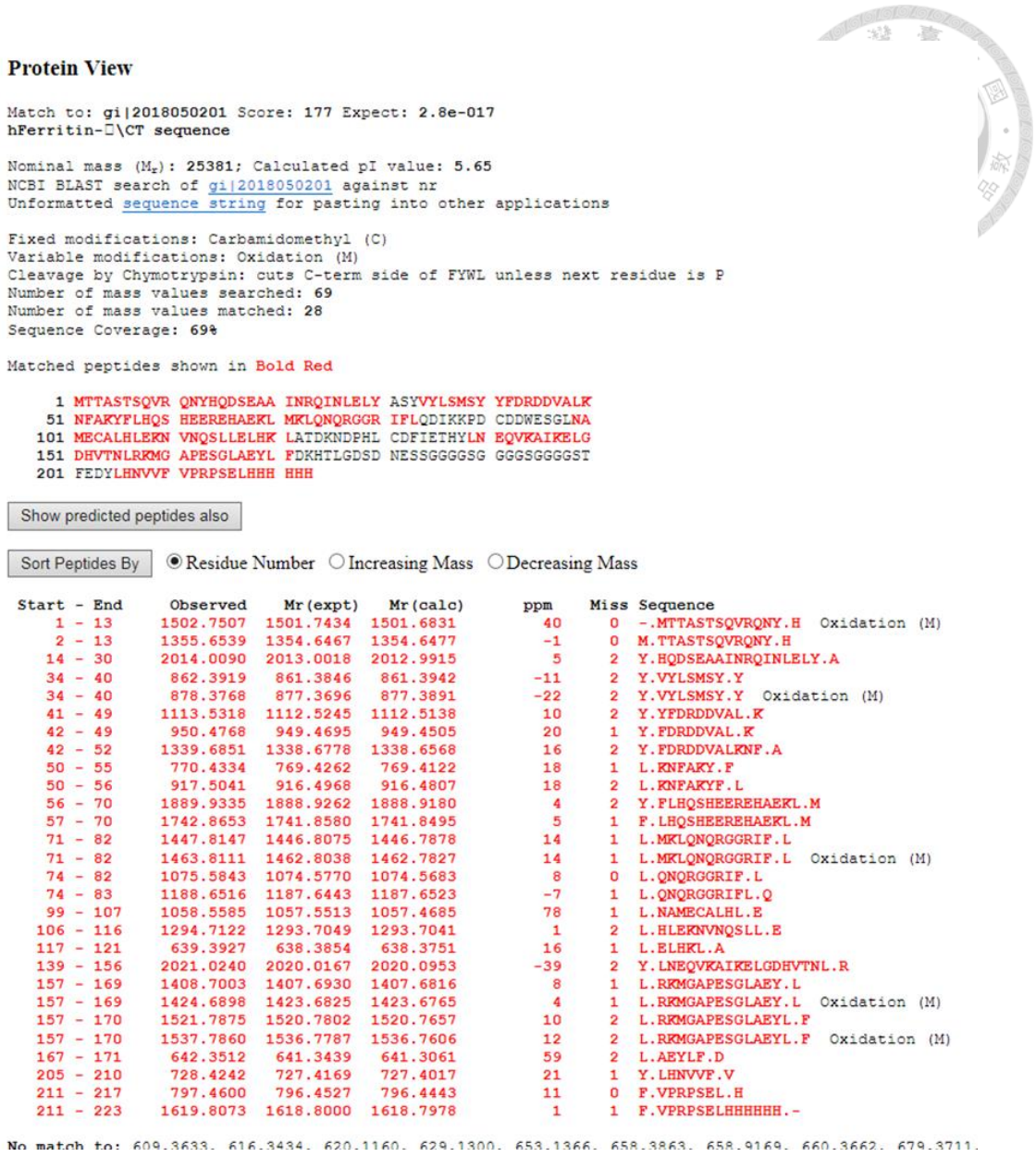
Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
13 - 42	2519.1364	2518.1292	2518.1757	-18	0	F.VPRPSSGGGGSGGGGSGGGGSTTASTSQVR.Q
43 - 55	1545.7065	1544.6993	1544.6967	2	1	R.QNYHQDSEAAINR.Q
46 - 55	1140.5394	1139.5321	1139.5319	0	0	Y.HQDSEAAINR.Q
82 - 86	607.3708	606.3636	606.3489	24	2	L.KNFAK.Y
83 - 87	642.3348	641.3276	641.3173	16	2	K.NFAKY.F
88 - 96	1182.5690	1181.5618	1181.5578	3	2	Y.FLHQSHEER.E
97 - 101	613.3108	612.3035	612.2867	27	0	R.EHAEK.L
103 - 109	917.5193	916.5120	916.4912	23	2	L.MKLQNQR.G
103 - 109	933.5029	932.4957	932.4862	10	2	L.MKLQNQR.G Oxidation (M)
105 - 109	658.3752	657.3679	657.3558	18	1	K.LQNQR.G
113 - 119	876.5243	875.5170	875.5116	6	2	R.IFLQDIK.K
115 - 119	616.3818	615.3746	615.3592	25	1	F.LQDIK.K
171 - 176	730.4203	729.4130	729.4021	15	1	Y.LNEQVK.A
180 - 189	1153.6079	1152.6006	1152.5887	10	2	K.ELGDHVTNLR.K
206 - 223	2138.8902	2137.8829	2137.9063	-11	2	K.HTLGDSDNESELHHHHHH.-

No match to: 603.3045, 609.3541, 643.3684, 654.3367, 670.3509, 677.3803, 680.3572, 685.3797, 688.3949

Figure S1. Protein identification of α-F.

The αCT-ferritin protein was in-gel digested by chymotrypsin. MALDI-TOF-MS/MS was then used to produce tandem mass spectra, which were identified to query in a known protein database. The sequence coverage is 49% and the matched peptides are shown in red.

doi:10.6342/NTU201903019



Protein View

Match to: gi|2018050201 Score: 177 Expect: 2.8e-017
hFerritin-□\CT sequence

Nominal mass (M_r): 25381; Calculated pI value: 5.65
NCBI BLAST search of gi|2018050201 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Chymotrypsin: cuts C-term side of FYWL unless next residue is P
Number of mass values searched: 69
Number of mass values matched: 28
Sequence Coverage: 69%

Matched peptides shown in Bold Red

```
  1 MTTASTSQVR QNYHQDSEAA INRQINLELY ASYVYLSMSY YFDRDDVALK
 51 NFAKYFLHQS HEEREHAEKL MKLQNQRGGR IFLQDIKKPD CDDWESGLNA
101 MECALHLEKN VNQSLLELHK LATDKNDPHL CDFIETHYLN EQVKAIKELG
151 DHVTNLRKMG APESGLAEYL FDKHTLGDSD NESSGGGGSG GGGSGGGGST
201 FEDYLHNVVF VPRPSELHHH HHH
```

Show predicted peptides also

Sort Peptides By ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
1 - 13	1502.7507	1501.7434	1501.6831	40	0	-.MTTASTSQVRQNY.H Oxidation (M)
2 - 13	1355.6539	1354.6467	1354.6477	-1	0	M.TTASTSQVRQNY.H
14 - 30	2014.0090	2013.0018	2012.9915	5	2	Y.HQDSEAAINRQINLELY.A
34 - 40	862.3919	861.3846	861.3942	-11	2	Y.VYLSMSY.Y
34 - 40	878.3768	877.3696	877.3891	-22	2	Y.VYLSMSY.Y Oxidation (M)
41 - 49	1113.5318	1112.5245	1112.5138	10	2	Y.YFDRDDVAL.K
42 - 49	950.4768	949.4695	949.4505	20	1	Y.FDRDDVAL.K
42 - 52	1339.6851	1338.6778	1338.6568	16	2	Y.FDRDDVALKNF.A
50 - 55	770.4334	769.4262	769.4122	18	1	L.KNFAKY.F
50 - 56	917.5041	916.4968	916.4807	18	2	L.KNFAKYF.L
56 - 70	1889.9335	1888.9262	1888.9180	4	2	Y.FLHQSHEEREHAEKL.M
57 - 70	1742.8653	1741.8580	1741.8495	5	1	F.LHQSHEEREHAEKL.M
71 - 82	1447.8147	1446.8075	1446.7878	14	1	L.MKLQNQRGGRIF.L
71 - 82	1463.8111	1462.8038	1462.7827	14	1	L.MKLQNQRGGRIF.L Oxidation (M)
74 - 82	1075.5843	1074.5770	1074.5683	8	0	L.QNQRGGRIF.L
74 - 83	1188.6516	1187.6443	1187.6523	-7	1	L.QNQRGGRIFL.Q
99 - 107	1058.5585	1057.5513	1057.4685	78	1	L.NAMECALHL.E
106 - 116	1294.7122	1293.7049	1293.7041	1	2	L.HLEKNVNQSLL.E
117 - 121	639.3927	638.3854	638.3751	16	1	L.ELHKL.A
139 - 156	2021.0240	2020.0167	2020.0953	-39	2	Y.LNEQVKAIKELGDHVTNL.R
157 - 169	1408.7003	1407.6930	1407.6816	8	1	L.RKMGAPESGLAEY.L
157 - 169	1424.6898	1423.6825	1423.6765	4	1	L.RKMGAPESGLAEY.L Oxidation (M)
157 - 170	1521.7875	1520.7802	1520.7657	10	2	L.RKMGAPESGLAEYL.F
157 - 170	1537.7860	1536.7787	1536.7606	12	2	L.RKMGAPESGLAEYL.F Oxidation (M)
167 - 171	642.3512	641.3439	641.3061	59	2	L.AEYLF.D
205 - 210	728.4242	727.4169	727.4017	21	1	Y.LHNVVF.V
211 - 217	797.4600	796.4527	796.4443	11	0	F.VPRPSEL.H
211 - 223	1619.8073	1618.8000	1618.7978	1	1	F.VPRPSELHHHHHH.-

No match to: 609.3633, 616.3434, 620.1160, 629.1300, 653.1366, 658.3863, 658.9169, 660.3662, 679.3711,

Figure S2. Protein identification of F-α.

The ferritin-αCT protein was in-gel digested by chymotrypsin. MALDI-TOF-MS/MS was then used to produce tandem mass spectra, which were identified to query in a known protein database. The sequence coverage is 69% and the matched peptides are shown in red.

doi:10.6342/NTU201903019

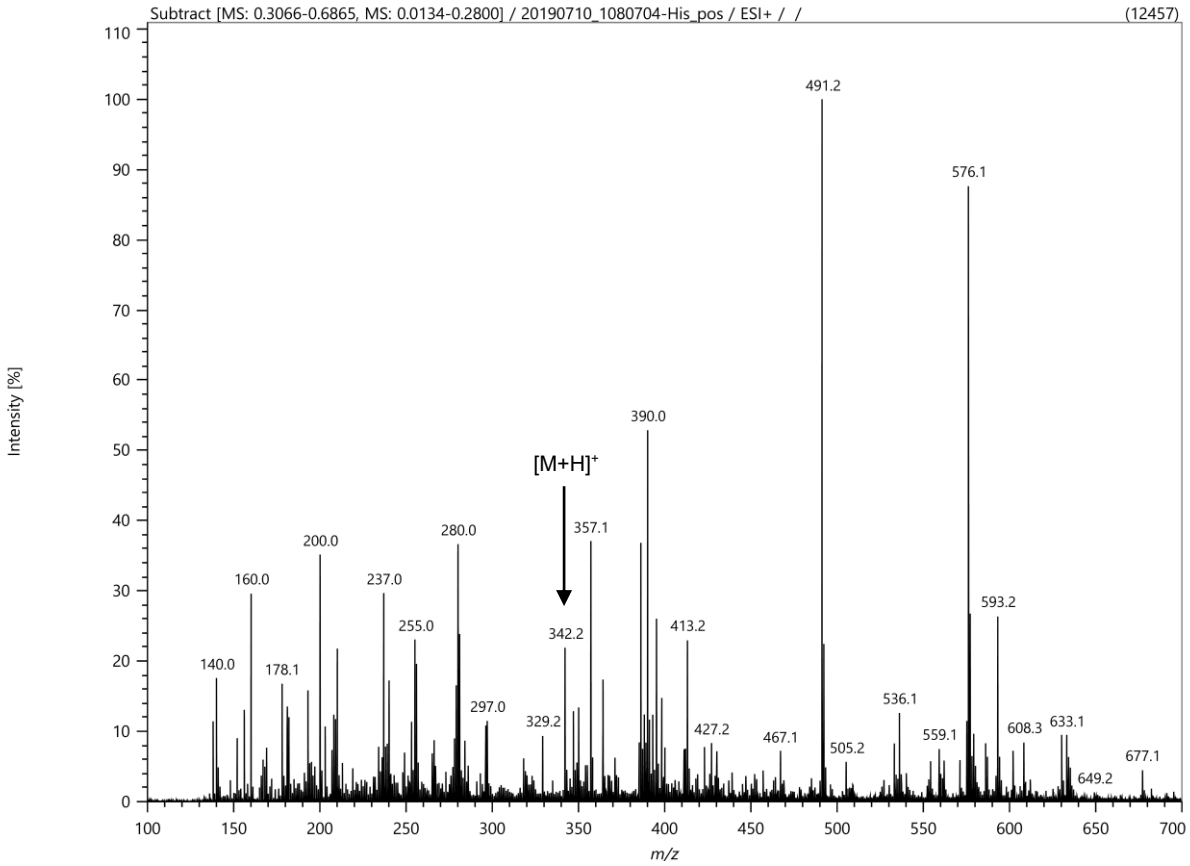


Figure S3. Molecular determination of reaction between DEEM and histidine.

The reaction mixture was analyzed by ESI-MS spectrometry and the expected molecular weight of product is 321. In the positive ion mode, product was found at m/z 342.2 ($[M+H]^+$), and histidine was found at m/z 178.1 ($[M_H+Na]^+$).

doi:10.6342/NTU201903019

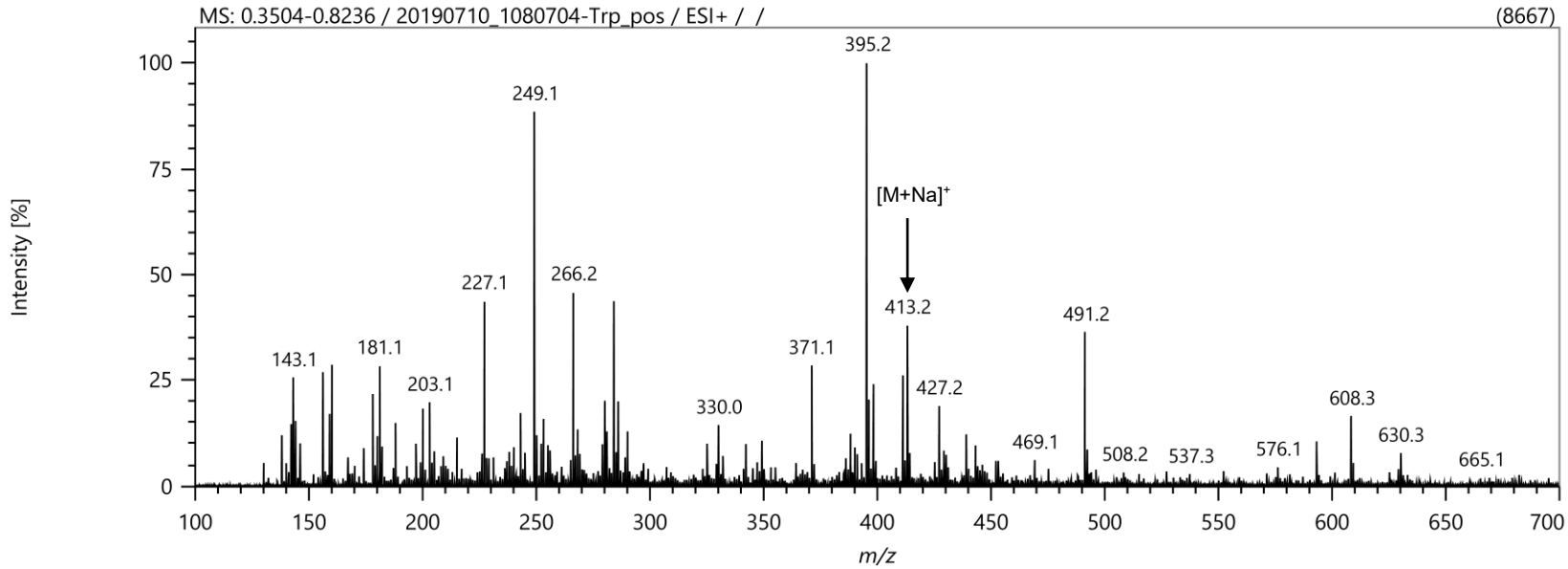


Figure S4. Molecular determination of reaction between DEEM and tryptophan.

The reaction mixture was analyzed by ESI-MS spectrometry and the expected molecular weight of product is 390. In the positive ion mode, the product was found at m/z 413.2 ($[M+Na]^+$), and tryptophan was found at m/z 227.1 ($[M_W+Na]^+$).

doi:10.6342/NTU201903019

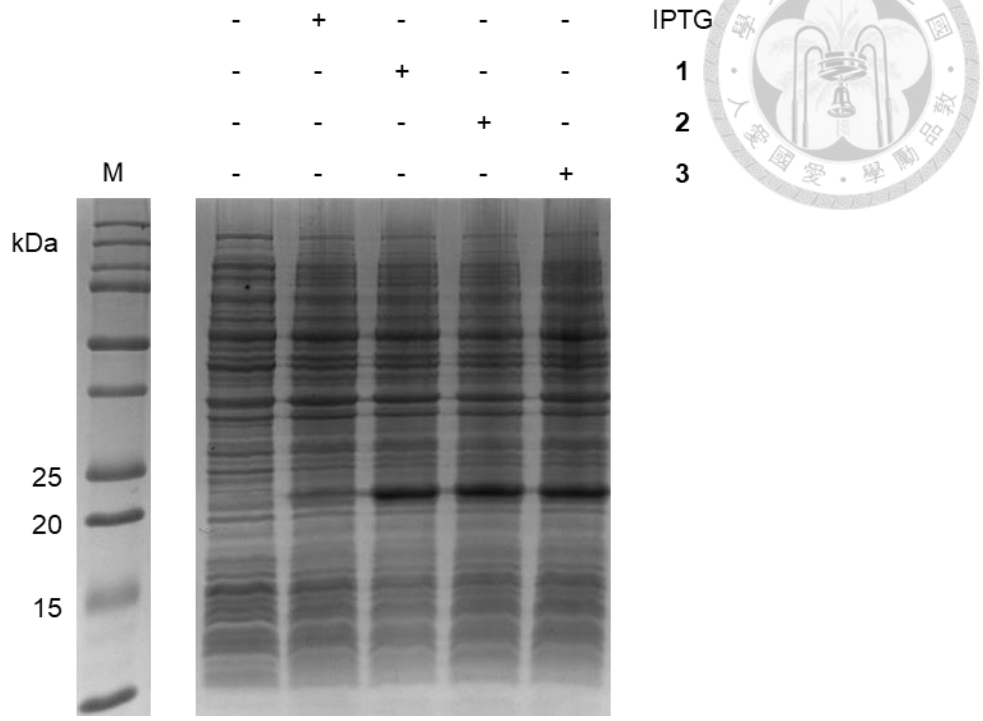


Figure S5. Amber suppression of ncAAs 1-3 at Ftn-1x.

E. coli BL21 (DE3) encoding PylRS•$tRNA^{Pyl}_{CUA}$ and ferritin genes with an amber codon at R63 were cultured in M9 minimal medium containing 1 mM IPTG and ncAA. The whole cell lysates were analyzed by SDS-PAGE. Bands were visualized using Instant Blue. M: protein marker; Lane 1: uninduced cell lysate; Lane 2: induced cell absent of ncAA; Lane 3-5: cell induced in the presence of three different ncAAs (1 mM) under 37°C for 12 hours.

doi:10.6342/NTU201903019

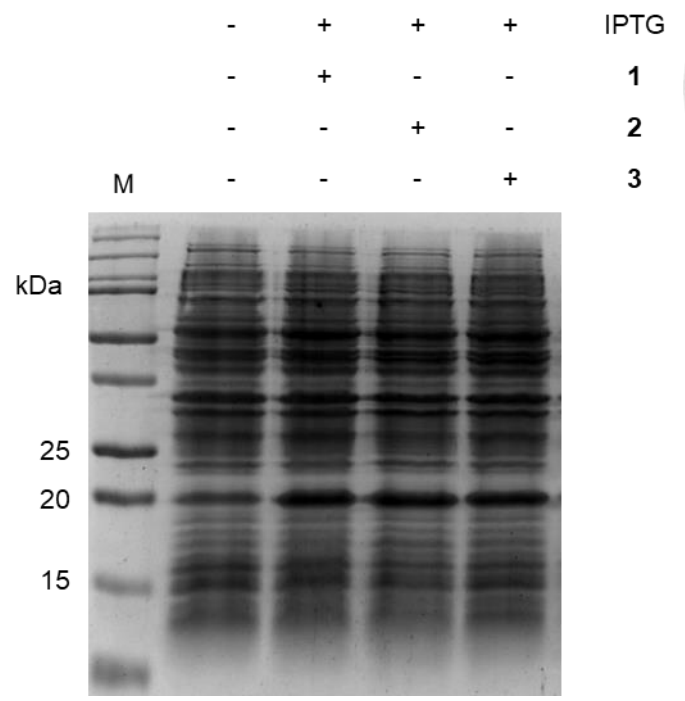


Figure S6. Amber suppression of ncAAs 1-3 in Ftn-2x.

E. coli BL21 (DE3) encoding PylRS•$tRNA^{Pyl}_{CUA}$ and ferritin genes with 2 amber codons at R63 and E67 were cultured in M9 minimal medium containing 1 mM IPTG and ncAA. The whole cell lysates were analyzed by SDS-PAGE. Bands were visualized using Instant Blue. M: protein marker; Lane 1: uninduced cell lysate; Lane 2-4: cell induced in the presence of three different ncAAs (1 mM) under 37ºC for 12 hours.

doi:10.6342/NTU201903019

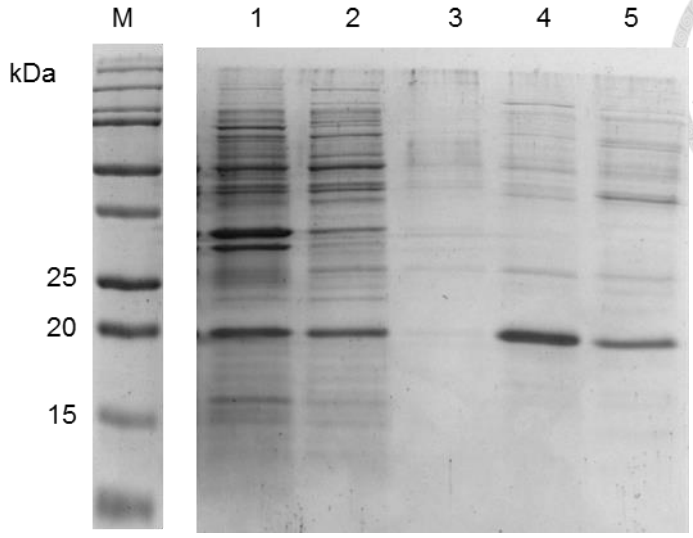


Figure S7. Purification of Ftn-1x-1 by Ni^{2+}-NTA column.

SDS-PAGE analysis of the purified His-tagged Ftn-1x-**1** from *E. coli* lysate under denaturing conditions. Bands were visualized using Instant Blue. M: protein marker; Lane 1: cell pellets; Lane 2: flow-through; Lane 3: wash (20 mM Tris, 100 mM NaCl, pH 7.5); Lane 4: 20 mM imidazole elution; Lane 5: 300 mM imidazole elution.

doi:10.6342/NTU201903019

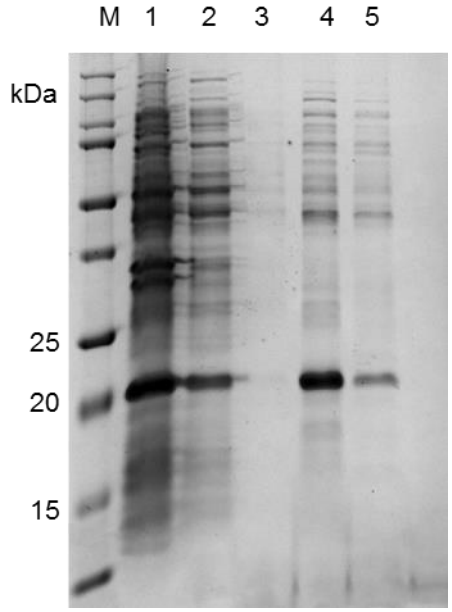


Figure S8. Purification of Ftn-1x-2 by Ni^{2+}-NTA column.

SDS-PAGE analysis of the purified His-tagged Ftn-1x-**2** from *E. coli* lysate under denaturing conditions. Bands were visualized using Instant Blue. M: protein marker; Lane 1: cell pellets; Lane 2: flow-through; Lane 3: wash (20 mM Tris, 100 mM NaCl, pH 7.5); Lane 4: 20 mM imidazole elution; Lane 5: 300 mM imidazole elution.

doi:10.6342/NTU201903019

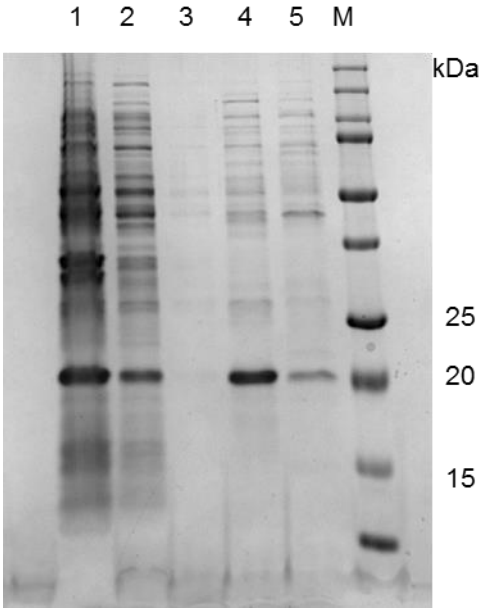


Figure S9. Purification of Ftn-1x-3 by Ni^{2+}-NTA column.

SDS-PAGE analysis of the purified His-tagged Ftn-1x-**3** from *E. coli* lysate under denaturing conditions. Bands were visualized using Instant Blue. M: protein marker; Lane 1: cell pellets; Lane 2: flow-through; Lane 3: wash (20 mM Tris, 100 mM NaCl, pH 7.5); Lane 4: 20 mM imidazole elution; Lane 5: 300 mM imidazole elution.

doi:10.6342/NTU201903019

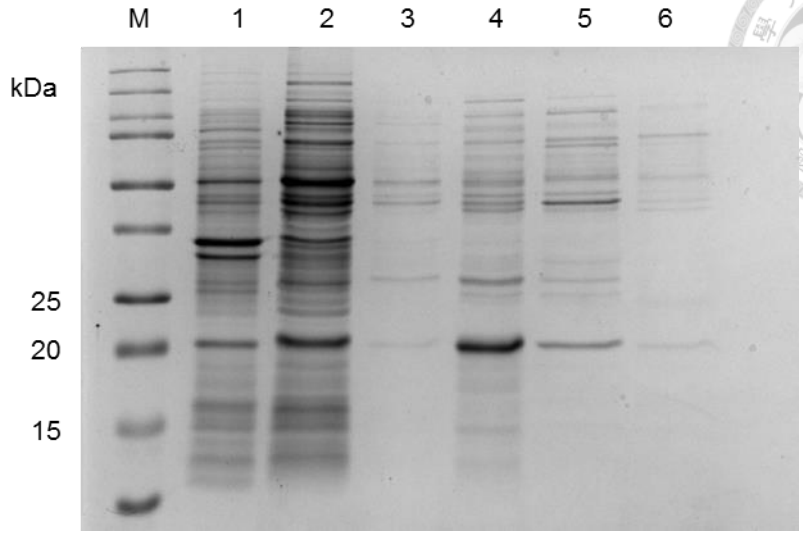


Figure S10. Purification of Ftn-2x-1 by Ni^{2+}-NTA column.

SDS-PAGE analysis of the purified His-tagged Ftn-2x-**1** from *E. coli* lysate under denaturing conditions. Bands were visualized using Instant Blue. M: protein marker; Lane 1: cell pellets; Lane 2: flow-through; Lane 3: wash (20 mM Tris, 100 mM NaCl, pH 7.5); Lane 4: 10 mM imidazole wash; Lane 5: 20 mM imidazole elution; Lane 5: 300 mM imidazole elution.

doi:10.6342/NTU201903019

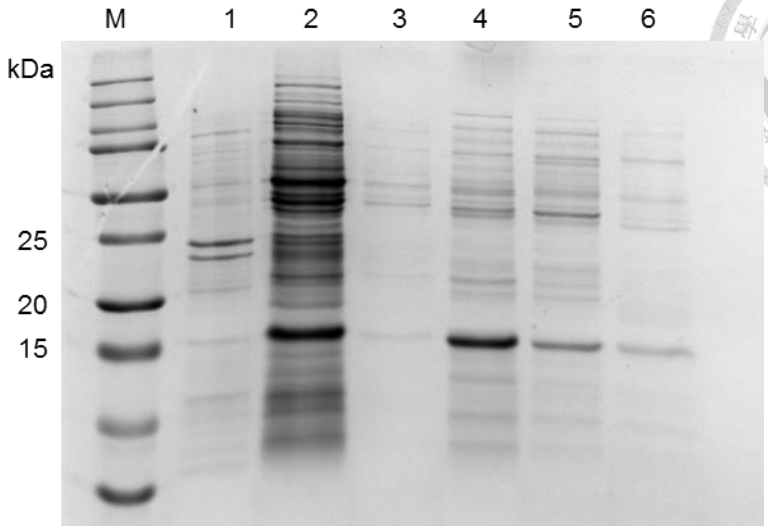


Figure S11. Purification of Ftn-2x-2 by Ni^{2+}-NTA column.

SDS-PAGE analysis of the purified His-tagged Ftn-2x-**2** from *E. coli* lysate under denaturing conditions. Bands were visualized using Instant Blue. M: protein marker; Lane 1: cell pellets; Lane 2: flow-through; Lane 3: wash (20 mM Tris, 100 mM NaCl, pH 7.5); Lane 4: 10 mM imidazole wash; Lane 5: 20 mM imidazole elution; Lane 5: 300 mM imidazole elution.

doi:10.6342/NTU201903019

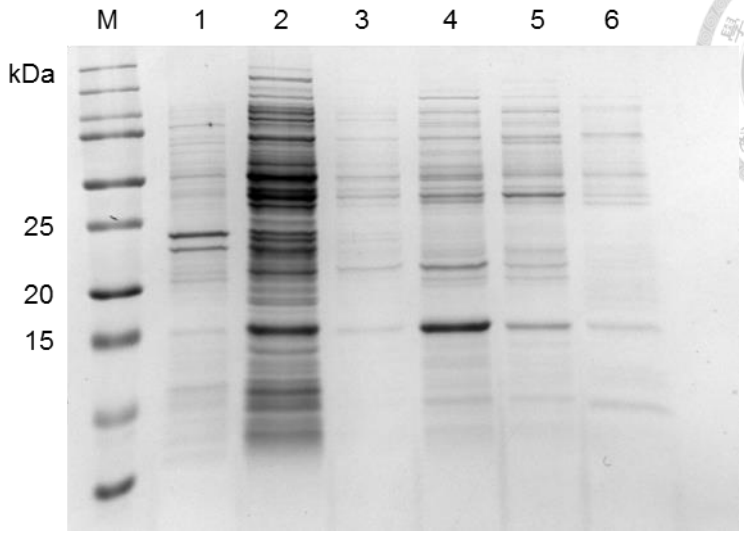


Figure S12. Purification of Ftn-2x-3 by Ni^{2+}-NTA column.

SDS-PAGE analysis of the purified His-tagged Ftn-2x-**3** from *E. coli* lysate under denaturing conditions. Bands were visualized using Instant Blue. M: protein marker; Lane 1: cell pellets; Lane 2: flow-through; Lane 3: wash (20 mM Tris, 100 mM NaCl, pH 7.5); Lane 4: 10 mM imidazole wash; Lane 5: 20 mM imidazole elution; Lane 5: 300 mM imidazole elution.

doi:10.6342/NTU201903019

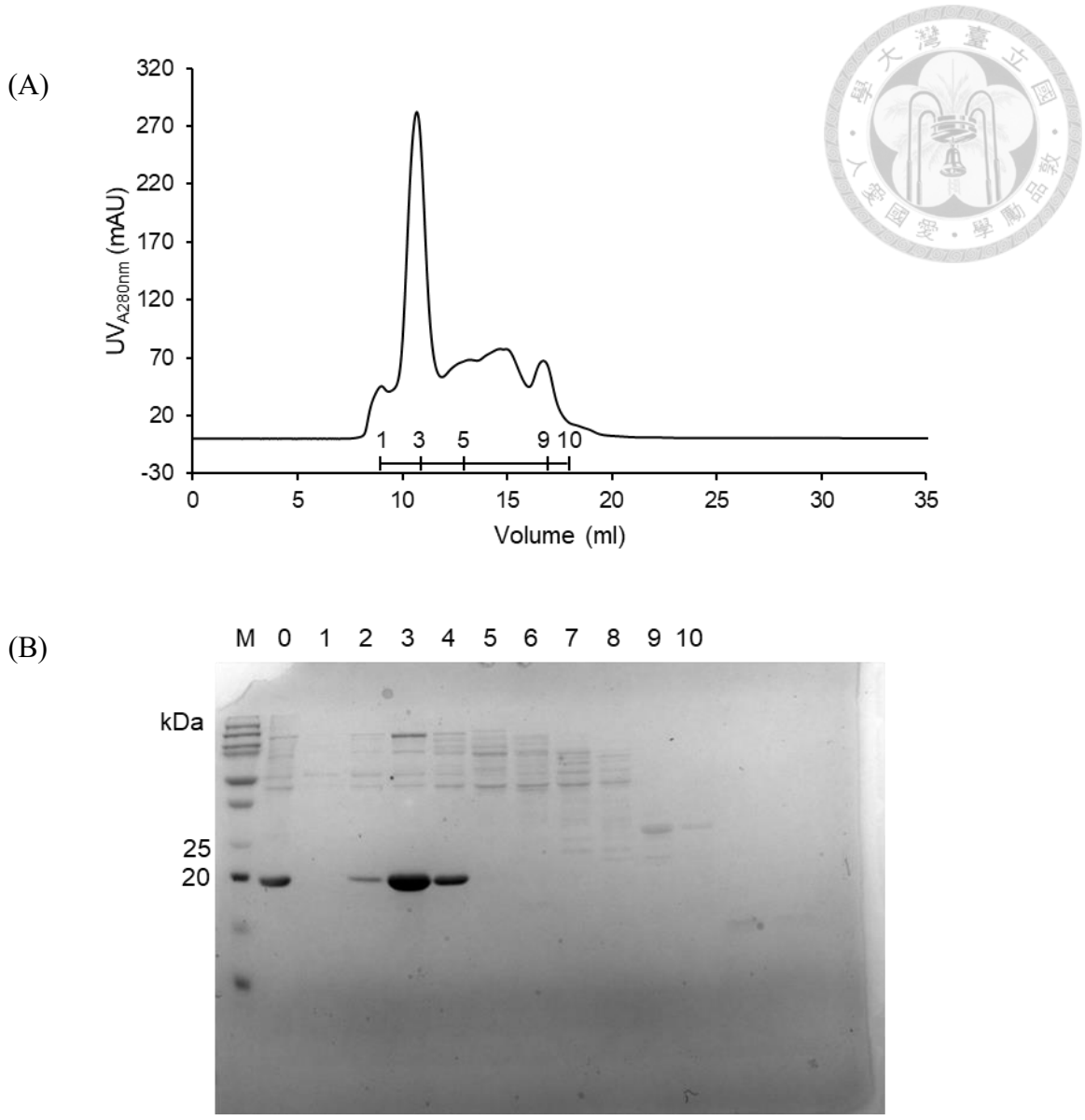


Figure S13. Purification of Ftn-1x-3 by SEC.

(A) Ftn-1x-**3** was purified by Superdex 200 Increase 10/300 GL. (B) SDS-PAGE analysis of purified Ftn-1x-**3** fractions. M: protein marker; Lane 0: Ftn-1x-**3** after Ni^{2+}-NTA column purification; Lane 2-4: major peak. Fraction 2-4 were collected.

doi:10.6342/NTU201903019

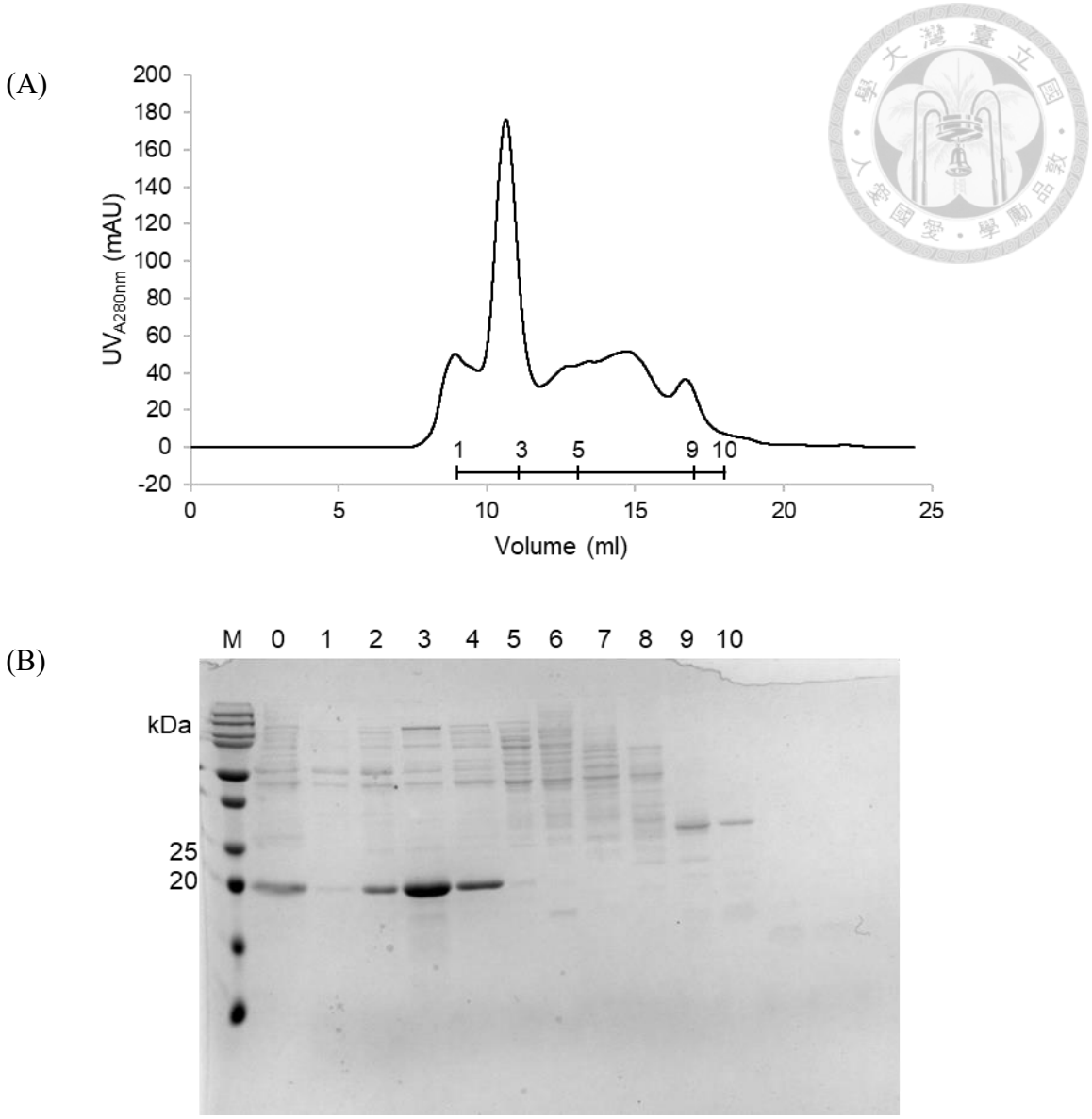


Figure S14. Purification of Ftn-2x-2 by SEC.

(A) Ftn-2x-**2** was purified by Superdex 200 Increase 10/300 GL. (B) SDS-PAGE analysis of purified Ftn-2x-**2** fractions. M: protein marker; Lane 0: Ftn-2x-**2** after Ni^{2+}-NTA column purification; Lane 2-4: major peak. Fraction 2-4 were collected.

doi:10.6342/NTU201903019

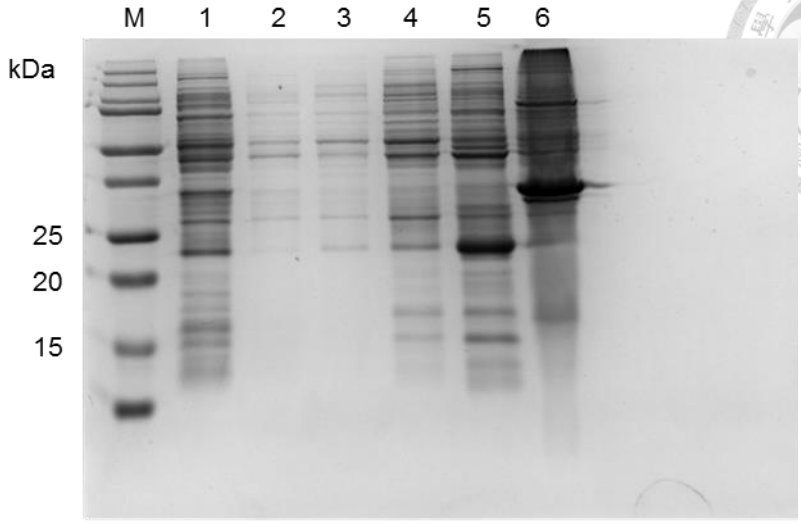


Figure S15. Purification of F-α-1x-3 by Ni^{2+}-NTA column.

SDS-PAGE analysis of the purified His-tagged F-α-1x-**3** from *E. coli* lysate under denaturing conditions. Bands were visualized using Instant Blue. M: protein marker; Lane 1: flow-through; Lane 2: wash (20 mM Tris, 100 mM NaCl, pH 7.5); Lane 3: 10 mM imidazole wash; Lane 4: 20 mM imidazole wash; Lane 5: 300 mM imidazole elution; Lane 6: cell pellets.

doi:10.6342/NTU201903019

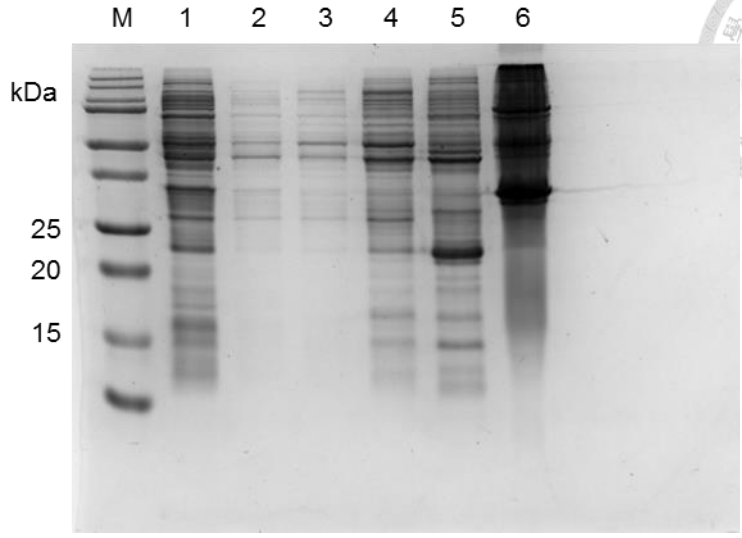
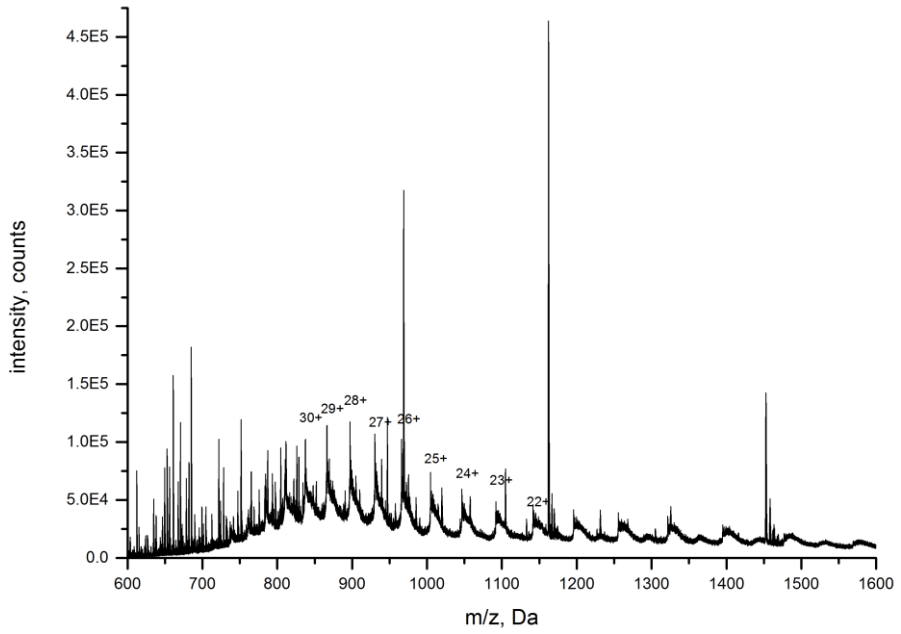


Figure S16. Purification of F-α-2x-2 by Ni^{2+}-NTA column

SDS-PAGE analysis of the purified His-tagged F-α-2x-**2** from *E. coli* lysate under denaturing conditions. Bands were visualized using Instant Blue. M: protein marker; Lane 1: flow-through; Lane 2: wash (20 mM Tris, 100 mM NaCl, pH 7.5); Lane 3: 10 mM imidazole wash; Lane 4: 20 mM imidazole wash; Lane 5: 300 mM imidazole elution; Lane 6: cell pellets.

doi:10.6342/NTU201903019

(A)



(B)

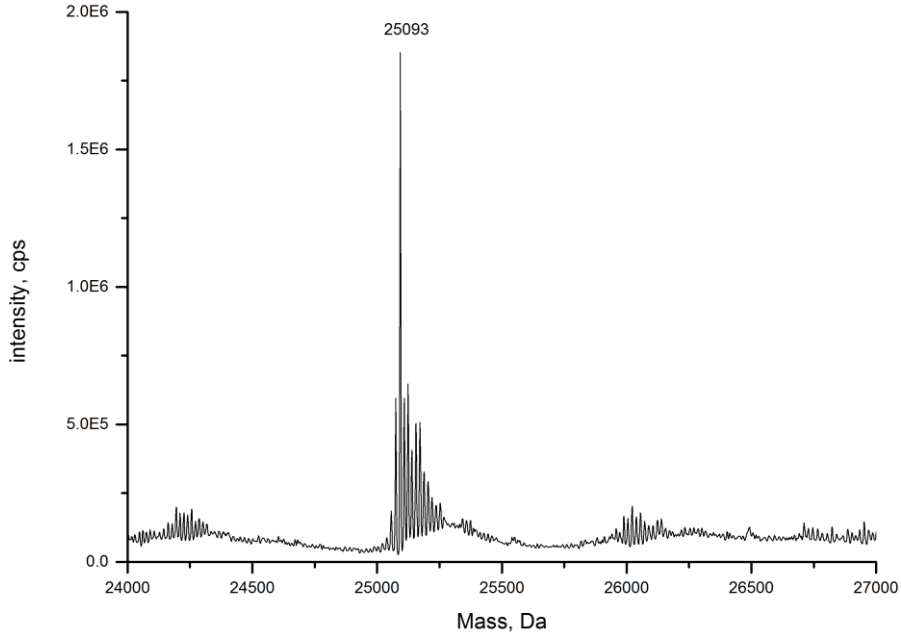
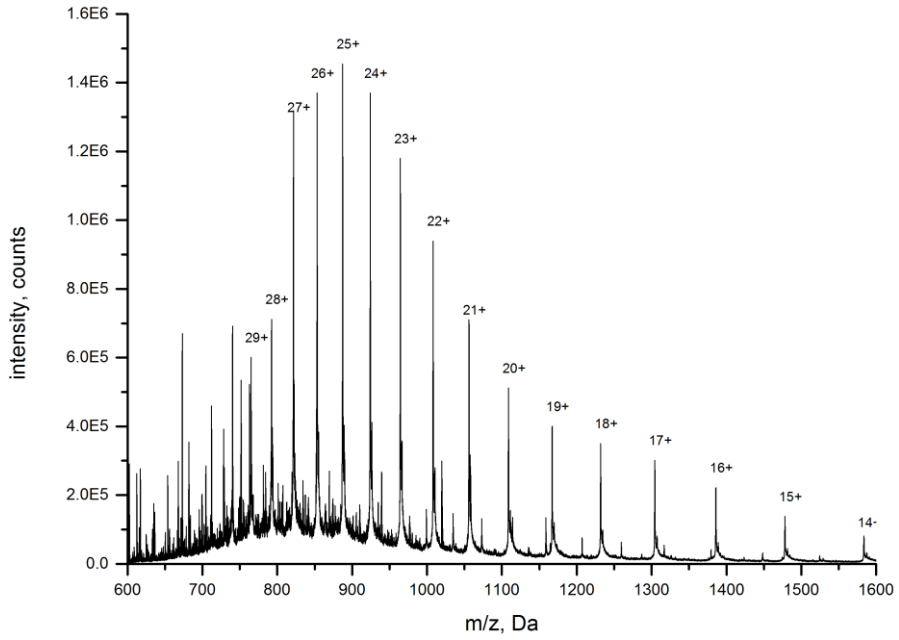


Figure S17. Molecular mass determination of F-α.

(A) ESI-MS spectrum of F-α in 20% (v/v) acetonitrille, 0.1% (v/v) formic acid. (B) Deconvoluted ESI-MS spectrum. The calculated molecular weight is 25,094 Da and the found molecular weight is 25,093 Da.

doi:10.6342/NTU201903019

(A)



(B)

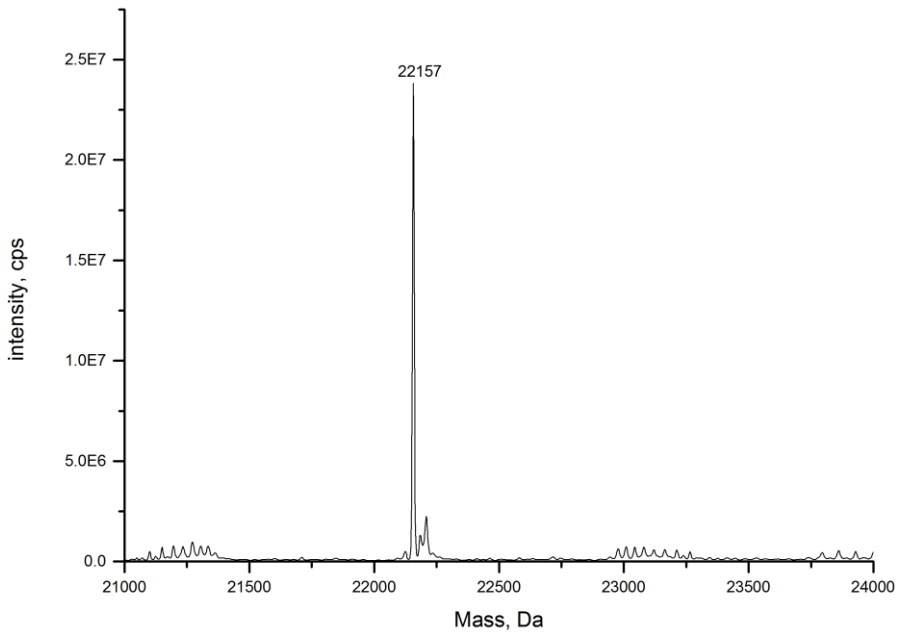
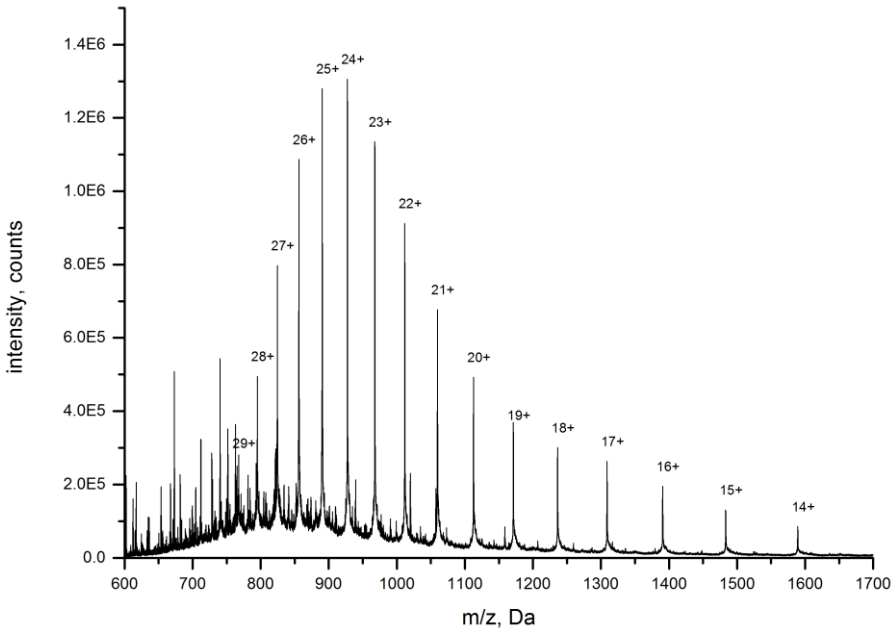


Figure S18. Molecular mass determination of Ftn-1x-1.

(A) ESI-MS spectrum of Ftn-1x-**1** in 20% (v/v) acetonitrille, 0.1% (v/v) formic acid. (B) Deconvoluted ESI-MS spectrum. The calculated molecular weight is 22,157 Da and the found molecular weight is 22,157 Da.

doi:10.6342/NTU201903019

(A)



(B)

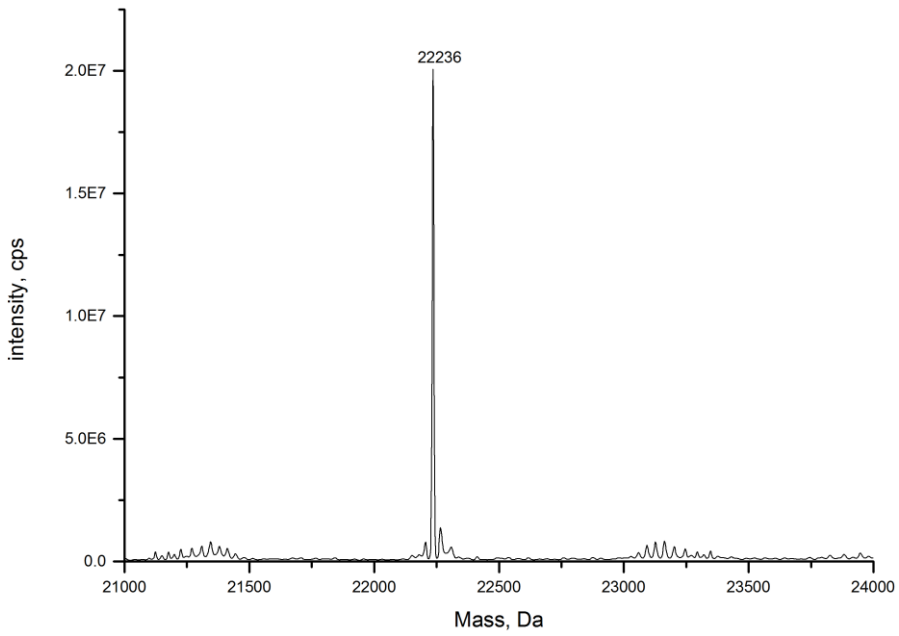
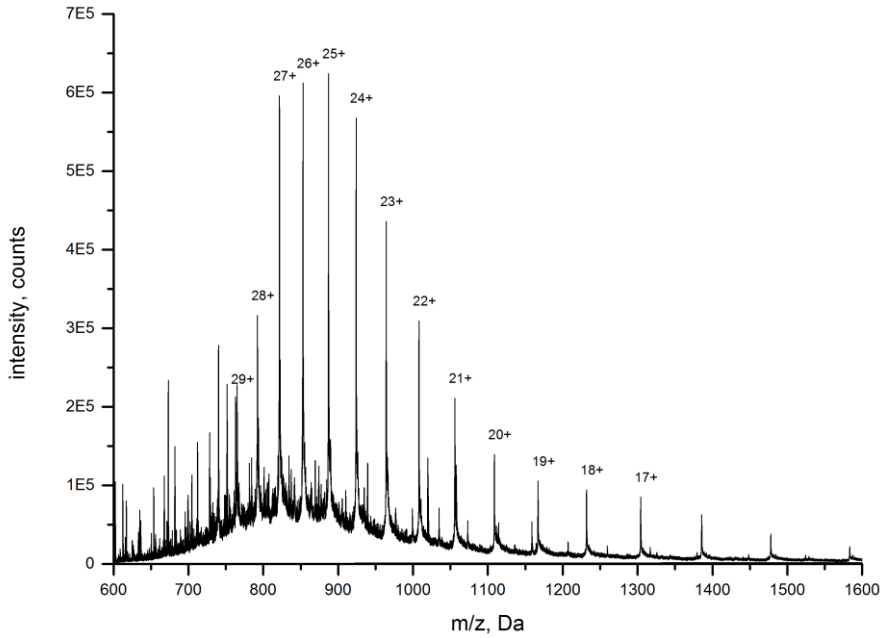


Figure S19. Molecular mass determination of Ftn-1x-2.

(A) ESI-MS spectrum of Ftn-1x-**2** in 20% (v/v) acetonitrille, 0.1% (v/v) formic acid. (B) Deconvoluted ESI-MS spectrum. The calculated molecular weight is 22,235 Da and the found molecular weight is 22,236 Da.

doi:10.6342/NTU201903019

(A)



(B)

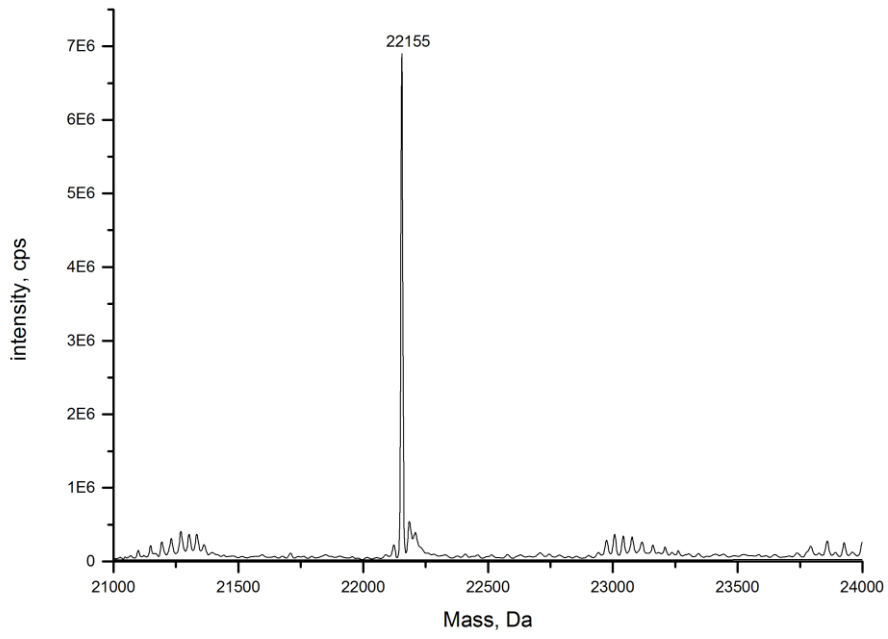
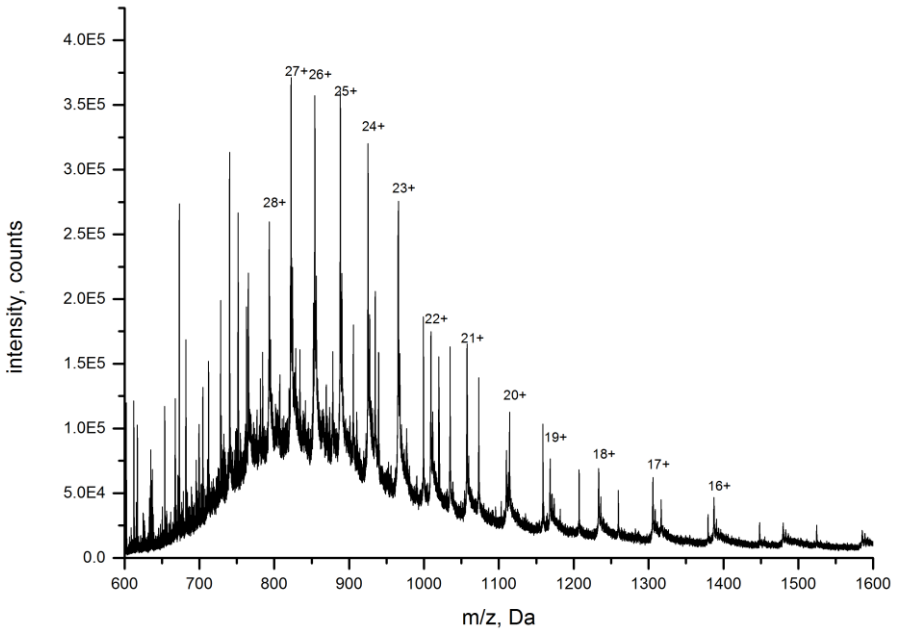


Figure S20. Molecular mass determination of Ftn-1x-3.

(A) ESI-MS spectrum of Ftn-1x-**3** in 20% (v/v) acetonitrille, 0.1% (v/v) formic acid. (B) Deconvoluted ESI-MS spectrum. The calculated molecular weight is 22,154 Da and the found molecular weight is 22,155 Da.

doi:10.6342/NTU201903019

(A)



(B)

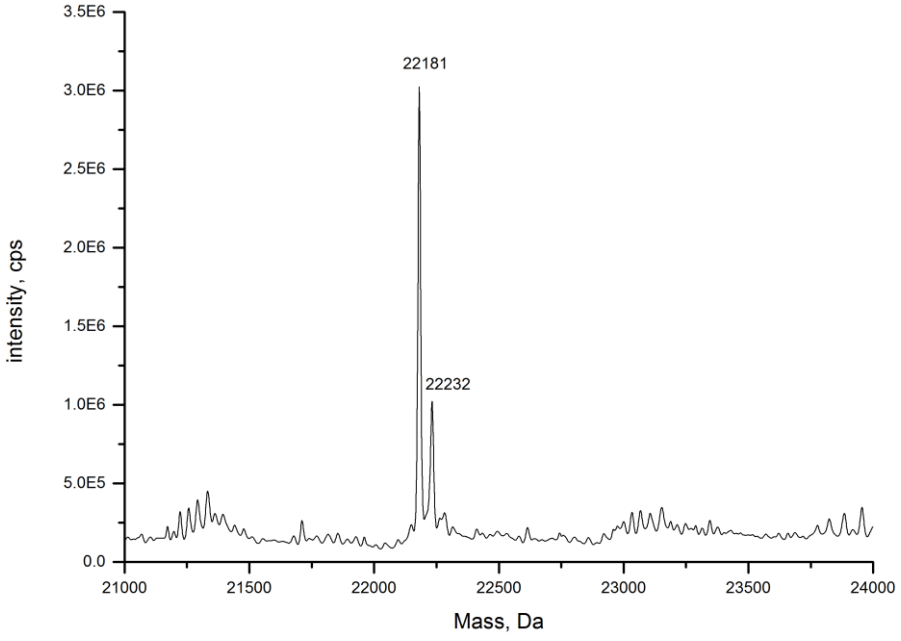
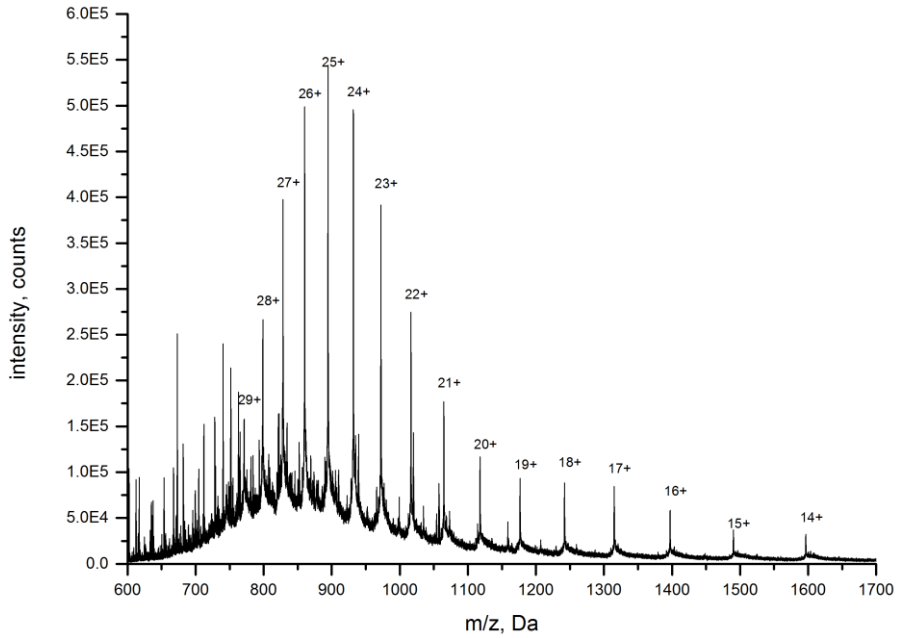


Figure S21. Molecular mass determination of Ftn-2x-1.

(A) ESI-MS spectrum of Ftn-2x-**1** in 20% (v/v) acetonitrille, 0.1% (v/v) formic acid. (B) Deconvoluted ESI-MS spectrum. The calculated molecular weight is 22,182 Da and the found molecular weight are 22,181 and 22,232 Da.

doi:10.6342/NTU201903019

(A)



(B)

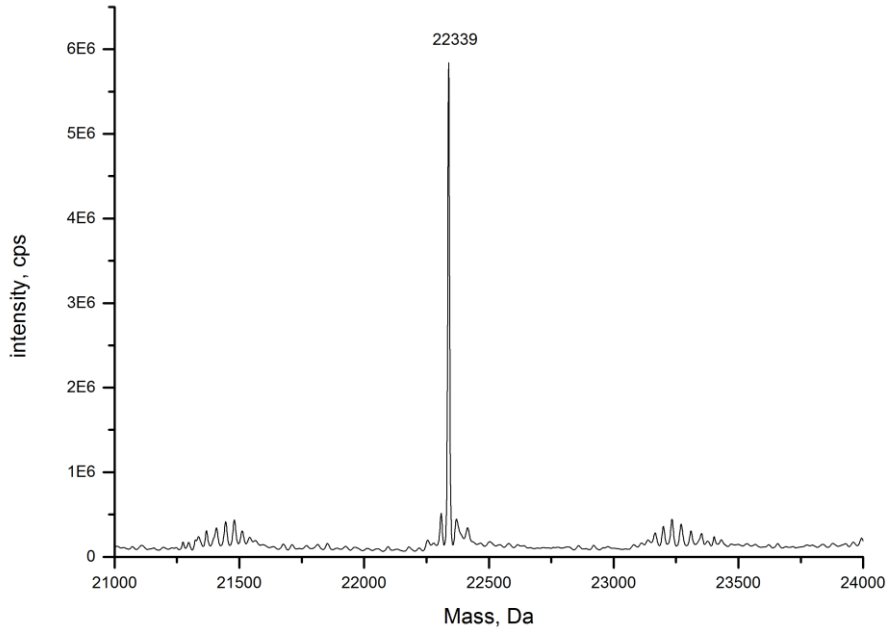


Figure S22. Molecular mass determination of Ftn-2x-2.

(A) ESI-MS spectrum of Ftn-2x-**2** in 20% (v/v) acetonitrille, 0.1% (v/v) formic acid.

(B) Deconvoluted ESI-MS spectrum. The calculated molecular weight is 22,338 Da and the found molecular weight is 22,339 Da.

doi:10.6342/NTU201903019

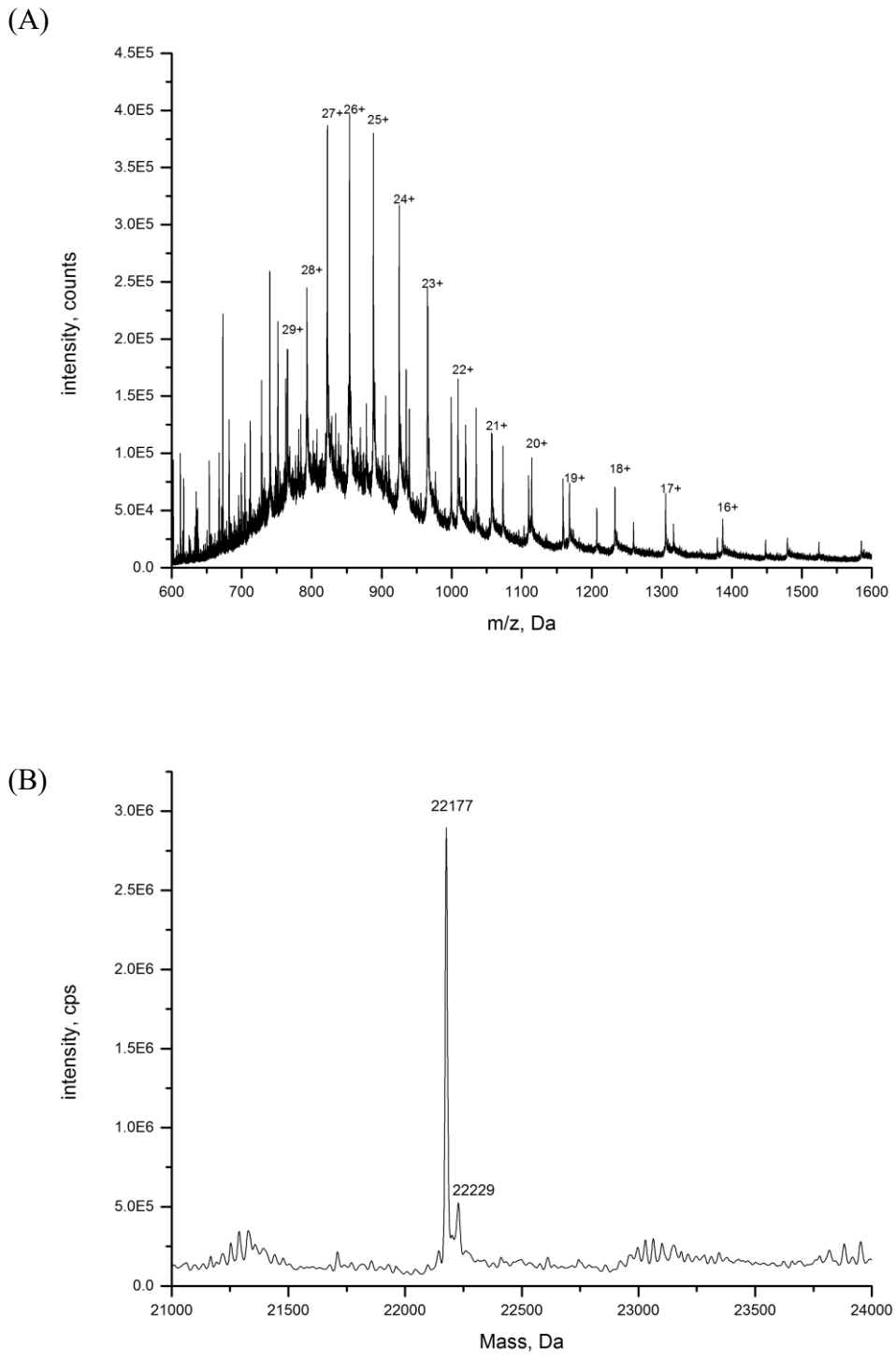


Figure S23. Molecular mass determination of Ftn-2x-3.

(A) ESI-MS spectrum of Ftn-2x-**3** in 20% (v/v) acetonitrille, 0.1% (v/v) formic acid. (B) Deconvoluted ESI-MS spectrum. The calculated molecular weight is 22,176 Da and the found molecular weight is 22,177 Da.

doi:10.6342/NTU201903019

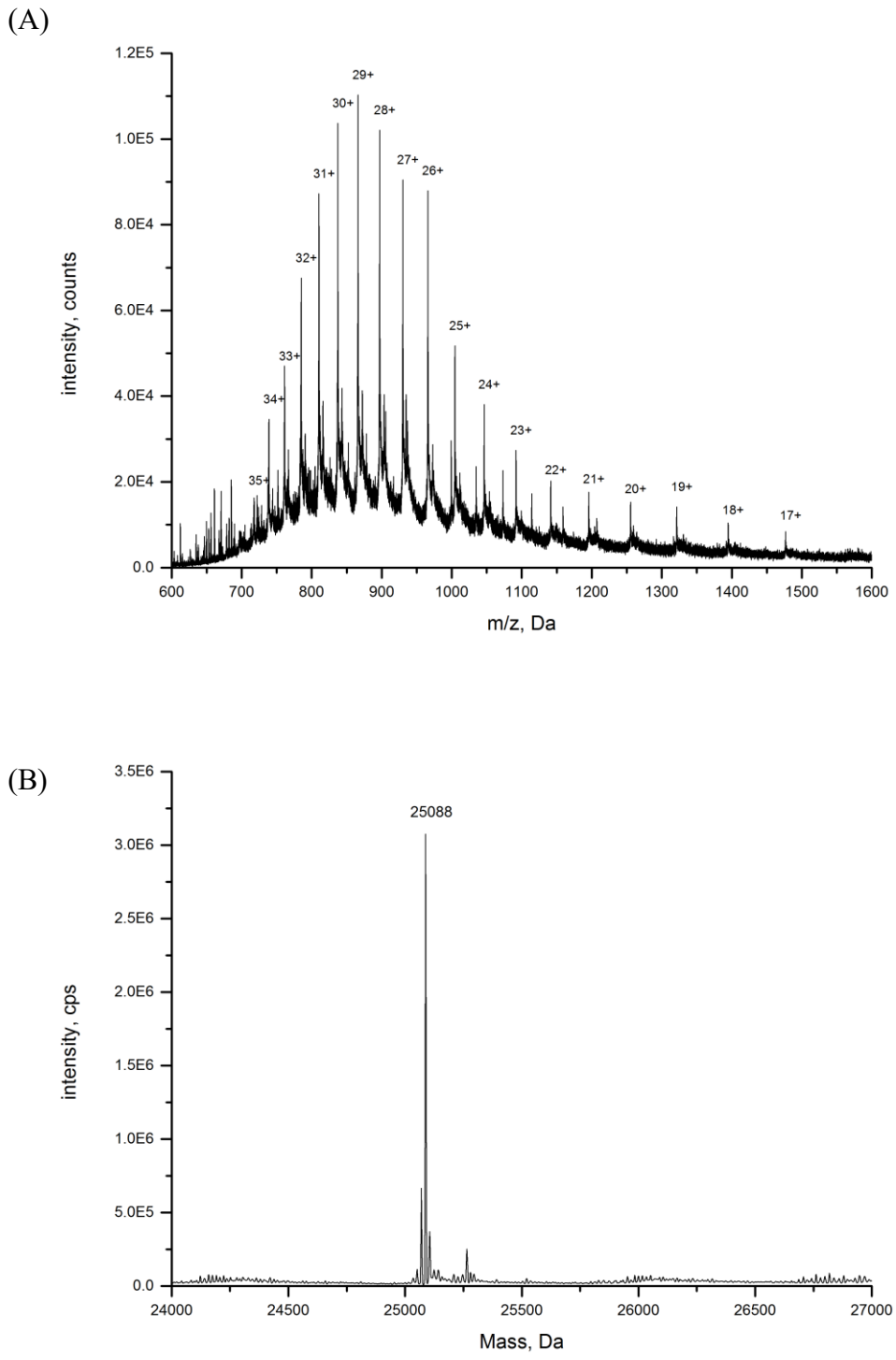


Figure S24. Molecular mass determination of F-α-1x-3.

(A) ESI-MS spectrum of F-α-1x-**3** in 20% (v/v) acetonitrille, 0.1% (v/v) formic acid. (B) Deconvoluted ESI-MS spectrum. The calculated molecular weight is 25,089 Da and the found molecular weight is 25,088 Da.

doi:10.6342/NTU201903019

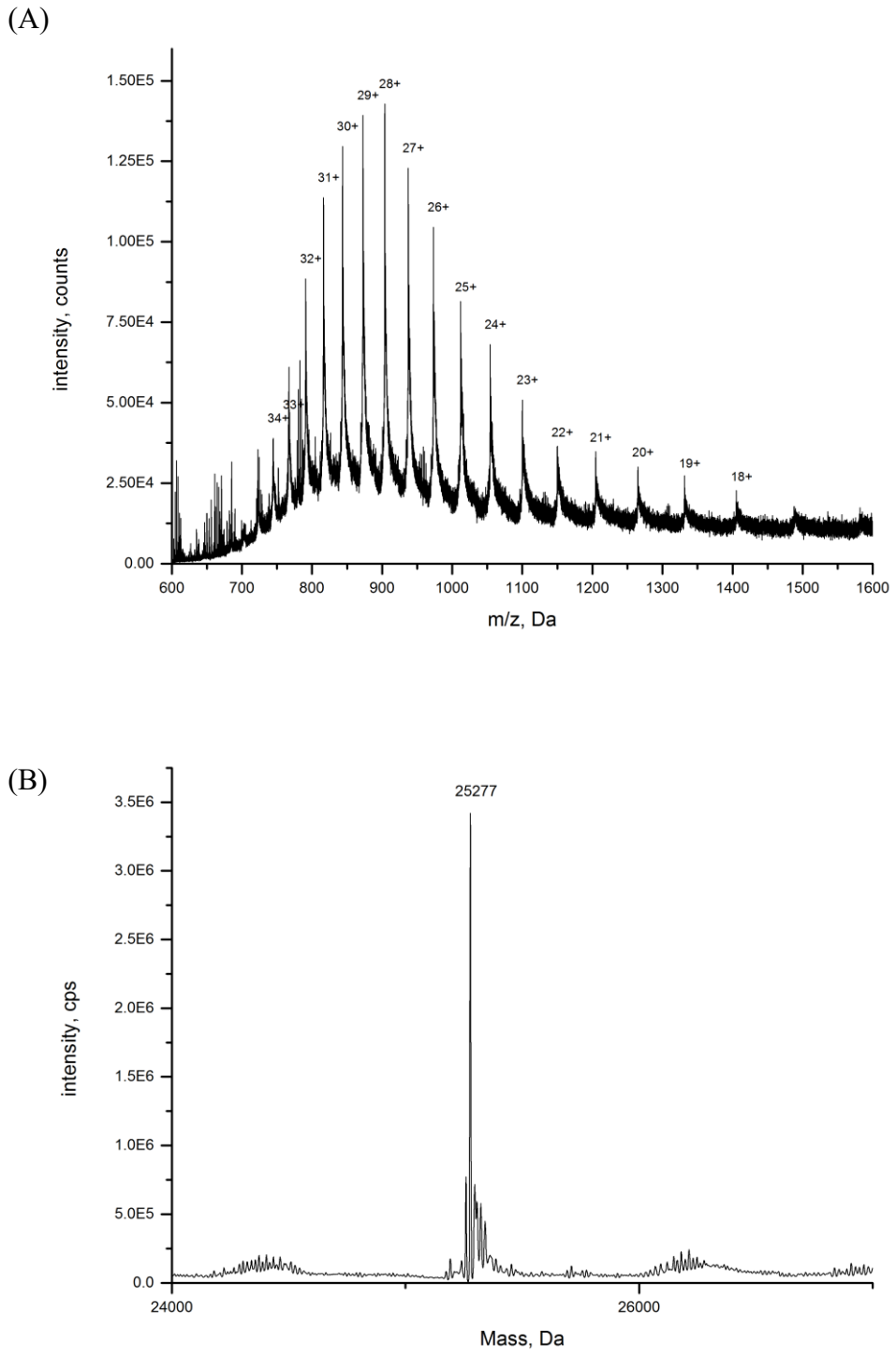


Figure S25. Molecular mass determination of F-α-2x-2.

(A) ESI-MS spectrum of F-α-2x-**2** in 20% (v/v) acetonitrille, 0.1% (v/v) formic acid. (B) Deconvoluted ESI-MS spectrum. The calculated molecular weight is 25,274 Da and the found molecular weight is 25,277 Da.

doi:10.6342/NTU201903019