論文 / 著書情報
Article / Book Information

題目(和文)	ミスアシル化tRNAによるPEG修飾および光応答性ペプチドアプタマーのin vitroセレクション
Title(English)	Misacylated tRNA for PEGylation and in vitro selection of photo-responsive peptide aptamer
著者(和文)	ZangQingmin
Author(English)	Qingmin Zang
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第10007号, 授与年月日:2015年9月25日, 学位の種別:課程博士, 審査員:山村　雅幸，小長谷　明彦，伊藤　嘉浩，木賀　大介，瀧ノ上　正浩
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第10007号, Conferred date:2015/9/25, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Misacylated tRNA for PEGylation and *In Vitro* Selection of Photo-responsive Peptide Aptamer

QINGMIN ZANG

Department of Computational Intelligence and Systems Science
Tokyo Institute of Technology

2015

Table of Contents

ABSTRACT

Genetic incorporation of non-canonical amino acids enables us to prepare precisely designed macromolecular bioconjugates. The approach attracts a great interest in protein engineering and evolutionary engineering to generate new medical drugs and novel molecular devices.

In this thesis polyethylene glycol (PEG) of differing lengths was genetically incorporated into the backbone of a polypeptide using stop-anticodon and frameshift anticodon-containing tRNAs which were acylated with PEG-containing amino acids. Although the incorporation yields decreased as the length of PEG increased, and the codon and the incorporation sites also affected the yield, the incorporation was successfully performed. It was also indicated the yields of truncated peptide increased as the lengths of PEG increased in the ribosomal synthesis.

As an application for evolutionary engineering, photo-responsive peptide aptamer to glutathione-immobilized microbeads was selected using ribosome display incorporated with tRNA carrying an amino acid coupled with an azobenzene. The selected peptide aptamer showed high affinity (micromolar) to glutathione-immobilized microbeads. Under UV irradiation, the peptide aptamer detached from the glutathione-immobilized microbeads.

In conclusion, by using non-canonical amino acid carrying tRNAs, incorporation of PEGs of different lengths into one peptide backbone and in *in vitro* selection of photo-responsive peptide aptamer was successfully achieved. The precise synthesis of bioconjugates contributes future biotechnology fused with chemistry.

List of Abbreviations

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
tRNA	Transfer ribonucleic acid
mRNA	Messenger ribonucleic acid
cDNA	Complementary deoxyribonucleic acid
PCR	Polymerase chain reaction
RT-PCR	Reverse transcription polymerase chain reaction
PRM	Peptide-ribosome-mRNA
EDTA	2,2’,2’’,2’’’-(Ethane-1,2-diyldinitrilo)tetraacetic acid
UV	Ultraviolet
T7	T7 promoter
SD	Shine-Dalgarno
NMR	Nuclear magnetic resonance
MS	Mass spectra
SPPS	Solid phase peptide synthesis
RF	Release factor
PEG	Polyethylene glycol
ADCs	Antibody-drug conjugates
Azo	Azobenzene
GSH	Glutathione

CHAPTER 1

GENERAL INTRODUCTION

1.1 Introduction

Since the discovery of the double helical structure of DNA molecules in 1953,[1] a great development has been taking place in molecular biology. One of the milestones was the famous Central Dogma[2] which implied the relationship and the flow of information among DNA, RNA and proteins. Many serious confirmation experiments[3] showed that the genetic code was based on non-overlapping triplets of bases, which are called codons. The codons of the standard genetic code were deciphered.[4-7] Each of 61 codons is assigned to a specific amino acid; the other three (UAA, UAG, UGA) are stop codons which terminate the elongation of peptide sequence. Crick proposed the frozen-accident hypothesis[8] that the genetic code was universal and had no further evolution with few exceptions. While in 1981 it was discovered that in mammalian mitochondria AUA was not assigned to isoleucine, but to methionine, and UGA for tryptophan instead of the stop codon.[9] Then it was realized that the genetic code could evolve.

In 1986, a great surprise was the discovery of the 21st amino acid: selenocysteine, which is found in every life on Earth.[10] Another surprising discovery was the 22nd amino acid: pyrrolysine, which is used in methanogenic archaea.[11] Both of them are genetically encoded in the same manner to the standard 20 amino acids. They are regarded as a natural expansion of the genetic code.

The artificial expansion of the genetic code was established by Schultz, who developed a method of site-specifically introducing non-natural amino acids into proteins using the amber suppression.[12] This method opened the way to expand proteins functions.

1.2 Non-natural amino acids

The natural amino acids contain an amino group, a carboxylate group, a hydrogen atom and a side chain (R group), all of which are attached to a chiral carbon atom (Cα) in the L form (except glycine). Similar to the natural amino acids, the compounds with an amino group, a carboxylate group and especially the altered R group are generally considered as non-natural amino acids. The changes of R group are diverse, ranging from the structural analogs of the natural amino acids to some compounds with novel chemical functional groups. Moreover, D-amino acids have also been selectively incorporated into proteins.[13]

The non-natural amino acids can circumvent the constraint of using the 20 natural amino acids on investigating various structure and function of proteins. And the non-natural amino acids can also enable the introduction of new functions that are difficult or impossible to create with the natural 20 amino acids, such as biophysical probes,[14] photo-active groups[15] and various other functional groups.

1.3 Approaches for expansion of genetic code

Both chemical and biochemical approaches have been developed to expand the genetic code. The chemical approaches are simple and straightforward, but the selectivity is quite low by the modification of the proteins or peptides or the size of the proteins is limited using the solid phase peptide synthesis. The biochemical approaches are a general *in vitro* method for the incorporation of non-natural amino acids into peptides or

proteins using the nonsense or frameshift suppressor tRNAs that are misacylated with a non-natural amino acid. The advantages of these approaches are high selectivity and high fidelity while the yield is low.

1.3.1 Chemical approaches

1.3.1.1 Chemical modification

The introduction of non-natural amino acids into peptides or proteins can be done *in vitro* through the chemical modification of the active groups on the surface of the peptides or proteins.[16] The selectivity of chemical modification depends on the reactivity of functional groups on different amino acids. The typical chemical modification is on the SH group of cysteine, NH_2 group of lysine, carboxylate group of aspartic acid and glutamic acid and the OH group of serine, threonine and tyrosine. The initial application of this method was to detect the relationship between functional groups and the biological activity. The applications were expanded to introduce biophysical probes, tags, or conjugate with kinds of functional compounds and so on. However, to achieve site-specific incorporation of non-natural amino acids into peptides using this approach is very difficult because the modification can occur to any exposed active sites of the peptides or proteins if there is more than one active group. This random modification can also lead to the loss of bioactivity.

1.3.1.2 Chemical synthesis

The non-natural amino acids can be introduced into proteins or peptides through pure chemical synthesis.[17] One of the typical chemical methods is solid phase peptide synthesis (SPPS)[18] (Figure 1). In this method, usually the C-terminus of the first amino acid is attached to an aqueous-compatible solid support through a cleavable linker, and the following N-protected amino acids can be reacted with the support-bound segment one by one after repeating deprotection/condesation. With SPPS, all kinds of non-natural amino acids can be introduced into peptides or proteins easily. One clear advantage of this method is that the non-natural amino acids can be precisely introduced into any sites of the proteins and peptides. But the size of the peptides or proteins is limited, to approximately 100 amino acids.[13]

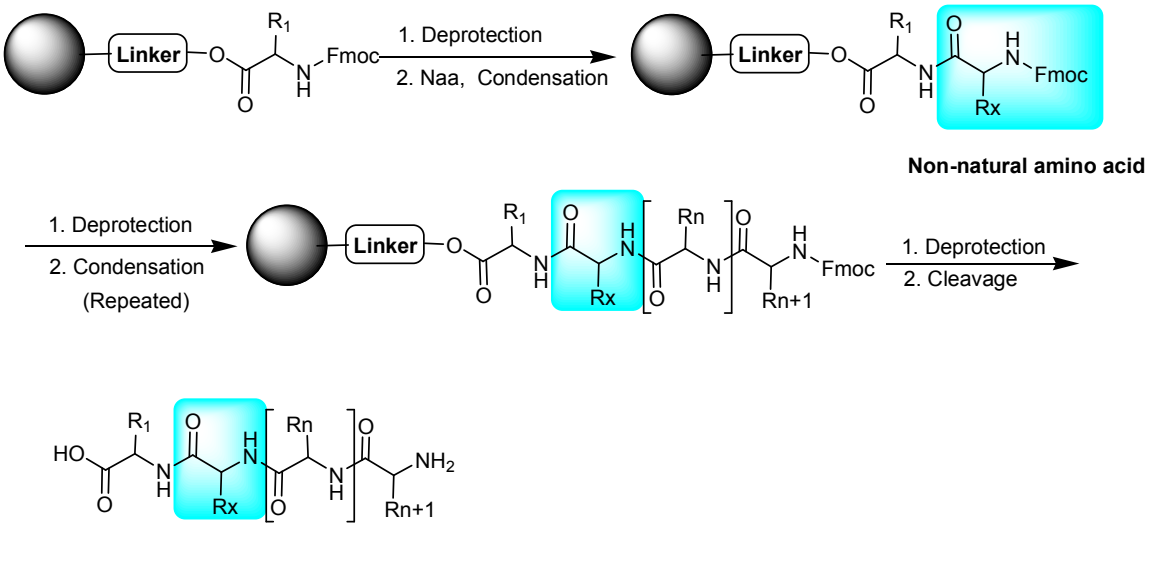


Fig. 1. Solid-phase peptide synthesis (SPPS)

Several efficient semi-synthetic methods have been developed to circumvent the size limitation of SPPS[19] (Figure 2). The principle strategy is to link two peptide fragments together, one or two of which containing the

non-natural amino acids is synthesized by using SPPS method. The other fragment is produced by recombinant techniques. These fragments are modified at their termini with some kind acitve groups such as bromide, thiol and amino group. Finally, these fragments are combined to make a larger peptide or protein contaning the non-natural amino acids. Theoretically, there is no limitation of the number of peptide fragments in semisynthetic methods. However, as the fragments increase, this approach becomes impractical and costly due to the yield decreases in each step.

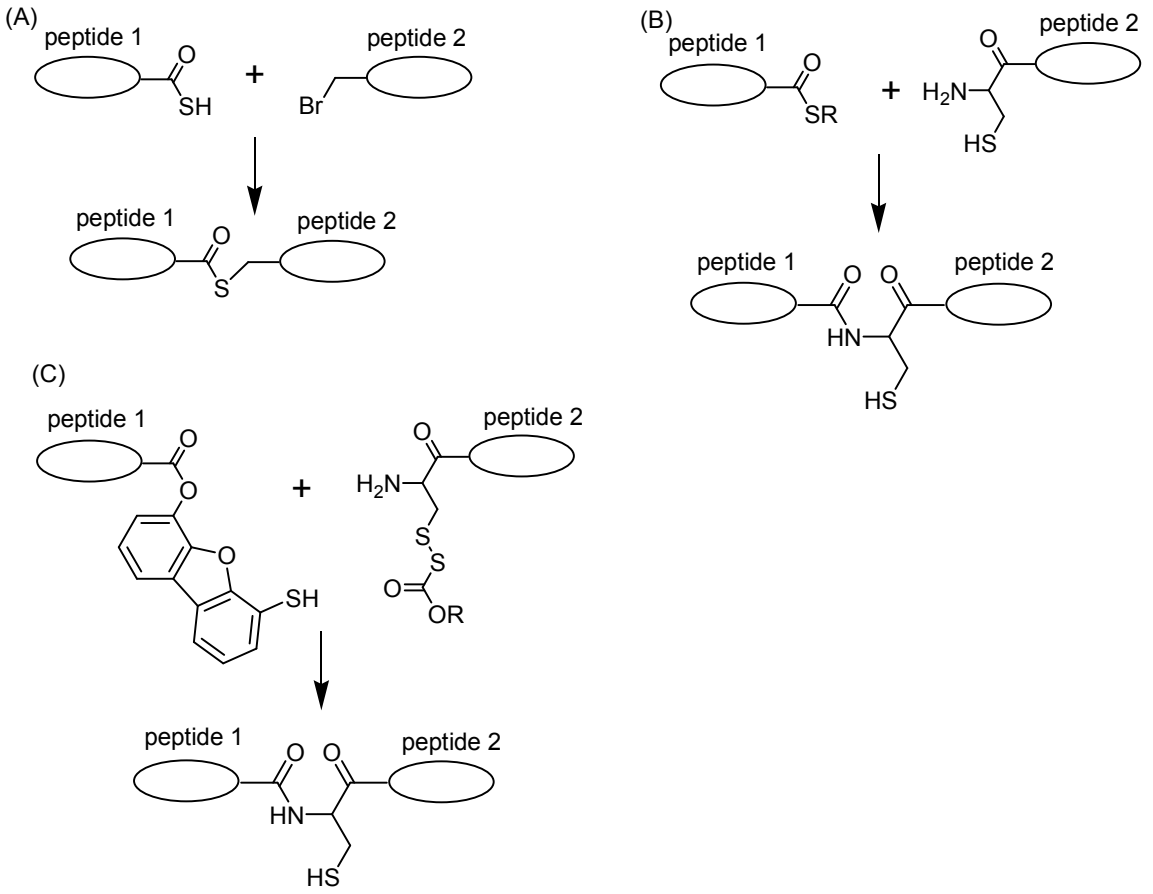


Fig. 2. Examples of chemoselective reactions for peptide ligation. (A) thioester formation; (B) native chemical ligation; (C) thiol capture.

1.3.2 Biochemical approaches

It is quite clear that the translation process of natural peptides or proteins *in vivo*. By mimicking the natural translation process, the non-natural amino acids can be introduced into peptides or proteins utilizing the existing cellular machinery of the cell.

Several *in vitro* methods have been developed to incorporate non-natural amino acids into proteins, which are based on the specific anticodon-codon recognition between mRNA and tRNA that is aminoacylated with the appropriate non-natural amino acid. The key intermediate is the aminoacyl tRNA conjugating with the non-natural amino acid. So how to form the key intermediate is the research hot pot.

1.3.2.1 Enzymatic aminoacylation

Quite similar to the natural amino acids, the incorporation of the non-natural amino acids into proteins requires a unique tRNA-codon pair and a corresponding aminoacyl-tRNA synthetase (aaRS), which are termed the orthogonal set. The orthogonal set must not cross-react with any of the endogenous tRNAs, aminoacyl tRNA synthetase, amino acids or codons and must be functionally compatible with other components of the translation system. The orthogonal synthetase should charge specifically a non-natural amino acid onto the orthogonal tRNA only (Figure 3).

The first orthogonal tRNA-amino acid tRNA synthetase was derived from tyrosyl-tRNA synthetase (TyrRS)-$tRNA^{Tyr}$ pair from the archaea *Methanococcus jannaschii* (Mj) because archaeal tRNAs have little cross-reacting with other translational machinery in *E. coli*.[20] And the TyrRS was

further mutated that not to accept tyrosine or any other amino acids, but only to accept O-methyl-tyrosine. A mutated E.coli TyrRS that recognizes m-iodotyrosine more efficiently than tyrosine was reported by Hirao and co-workers. Three positions were mutated and one mutant, V37C195, recognized iodotyrosine 10-fold than to recognize tyrosine.

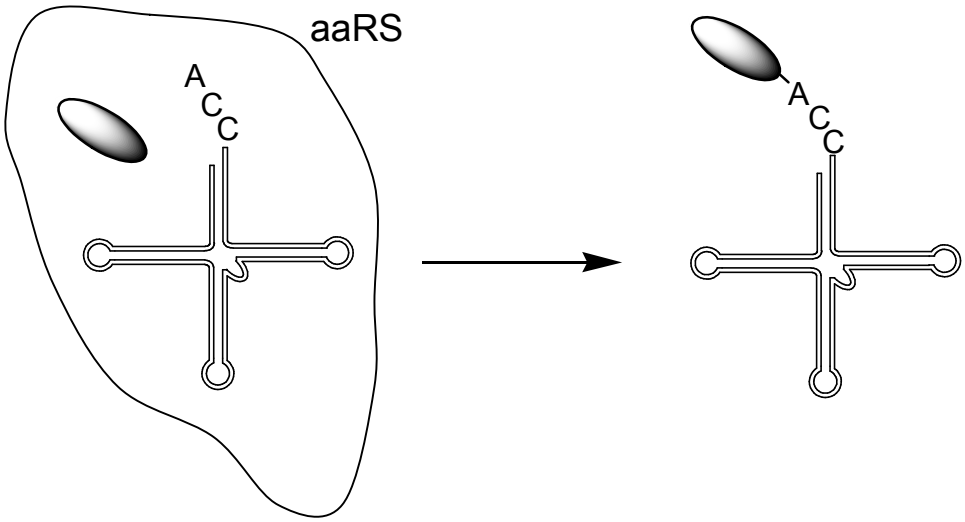


Fig. 3. Enzymatical aminoacylation. Aminoacylation by aaRS generated by directed evolution to recognize both a non-natural amino acid and an anticodon of tRNA.

Almost all the methods for genetic incorporation non-natural amino acids use four aminoacyla-tRNA synthetases. The MjTyrRS is used in E.coli, not in eukaryotic cells. The E.coli Tyrosyl-tRNA synthetase and E.coli Leucyl-tRNA synthetase[21-23] are used in yeast and mammalian cells, but not in bacteria. And the pyrrolysyl-tRNA syntheatase[24] can be used in E.coli, yeast, mammalian cells and C.elegans.

Up to now, about 70 non-natural amino acids have been incorporated into proteins using aaRS with high fidelity (Figure 4). Some of them have unique reactive groups such as azido, amino, iodo, bromo, cyano, and so on,

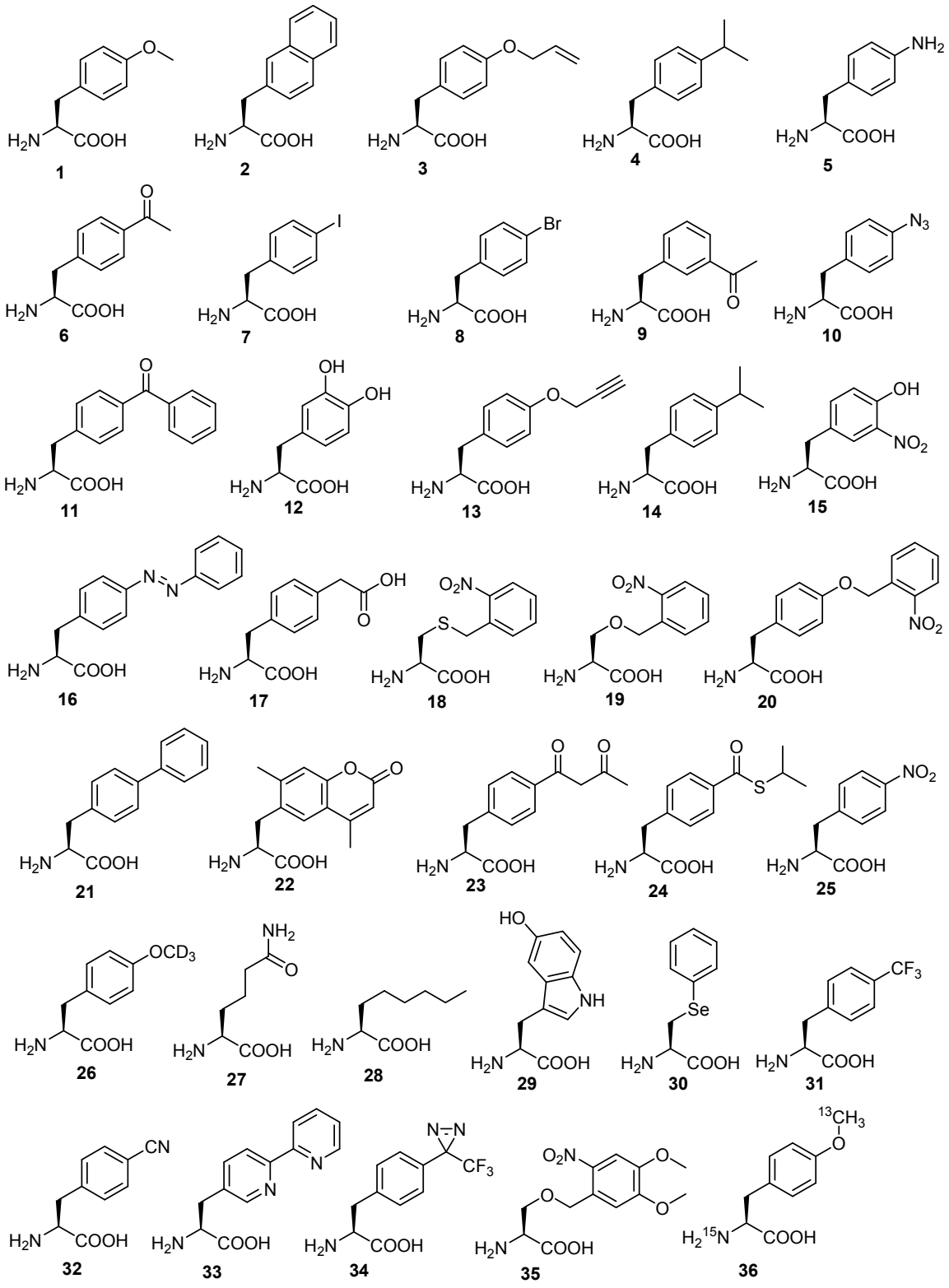


Fig. 4. Some non-natural amino acids that have been added to the genetic code with aaRS.

and these groups can be reacted with other molecules. And some non-natural amino acids with special functions are also added to the genetic code in this method, for example, azobenezene, biphenyl and coumarine. So the protein functions are widely expanded.[25]

1.3.2.2 Semi-chemical acylation

Direct chemical acylation of tRNAs is not practical due to the high degree of reactive sites in the tRNA. A semi-chemical acylation of tRNAs has been developed according to the fact that the common feature of all tRNAs has the same sequence CCA at the 3' terminus (Figure 5). The strategy is that the truncated tRNAs (without A or CA at the 3' terminus) are enzymatically ligated with the mono- (A) or the dinucleotides (CA) which is conjugated with the non-natural amino acids by chemical method.

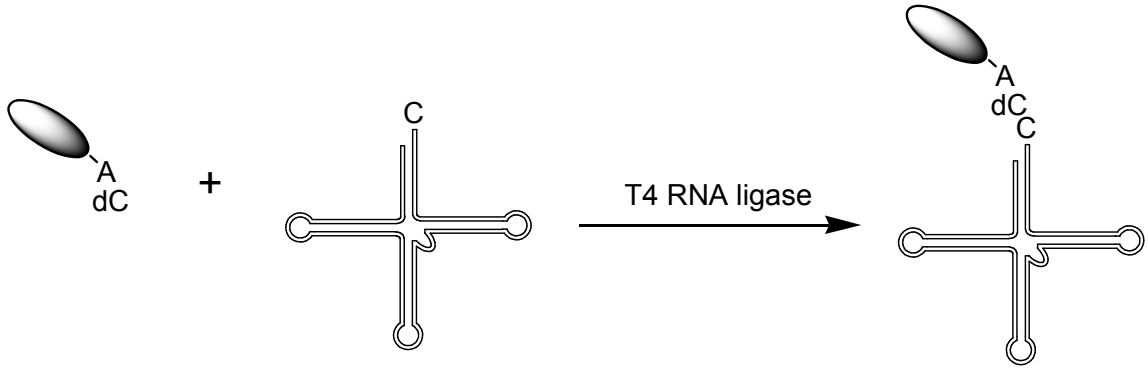


Fig. 5. Semi-chemical aminoacylation of tRNAs. A chemically synthesized aminoacyl-dinucleotide is enzymatically ligated with a truncated tRNA by T4 RNA ligase.

Pioneering work in this field was undertaken by Hecht and co-workers.[26] The mononucleotide pA was chemically reacted with the N^{α}-protected amino acid and then it was ligated with the truncated tRNA with RNA ligase.

Finally the protective group was removed. But the yield was quite low. Later, this method was improved by Hecht[27-28] and Brunner[29] separately. The dinucleotides pCpA was used instead of pA to conjugate with the N^{α}-protected amino acid. And then the pCpA conjugating with the N^{α}-protected amino acid was ligated with the truncated tRNA by T4 RNA ligase. A lot of non-natural amino acids were incorporated into the peptides by this method. But the method also suffered some limitations: the protective groups could not be removed efficiently and the yields of the dinucleotides conjugating with the non-natural amino acids were quite low. In 1989, Schultz and co-workers developed a general method using the dinucleotides pdCpA to conjugate with non-natural amino acids.[30] And then the pdCpA conjugating with non-natural amino acid was ligated with the truncated tRNA with T4 RNA ligase. The advantage of pdCpA is that it involves a few protective groups and the yield of the dinucleotides conjugating with the non-natural amino acids is high. Therefore, pdCpA is usually used for conjugating with the non-natural amino acids.

Comparing to the enzymatic aminoacyaltion method, semi-chemical aminoacylation is a more general method for the preparation of tRNA carrying any non-natural amino acid while this method is time-consuming and laborious.

1.3.2.3 Flexizymes

To overcome the specificity of the aminoacyl-tRNA synthetase (aaRS), one kind of general enzymes is developed, called flexizyme. Flexizymes are

highly flexible *de novo* tRNA acylation ribozymes that enable the preparation of a variety of non-natural amino acids onto tRNAs.[31-32]

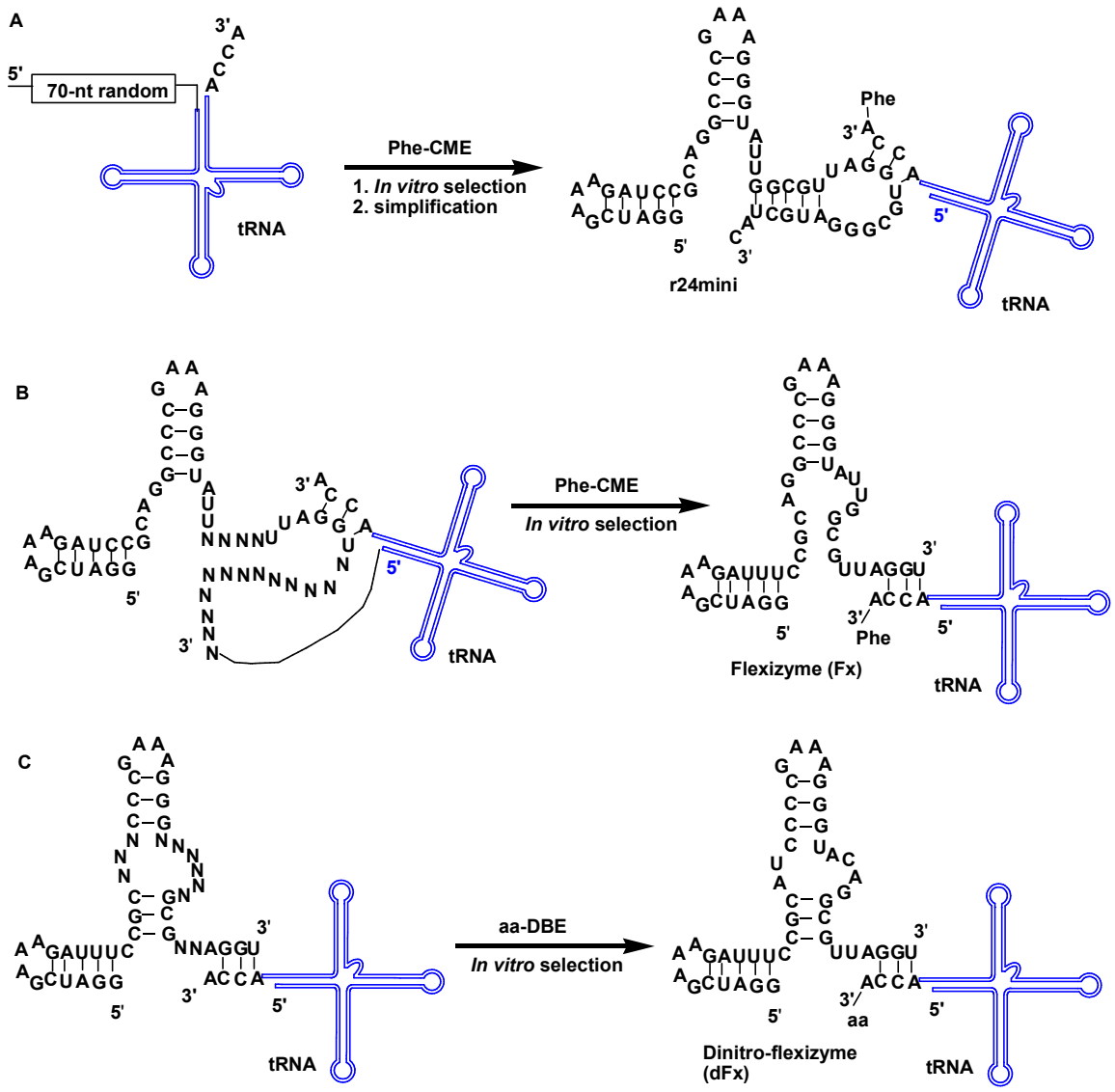


Fig. 6. *In vitro* selection of flexizymes. A: The first generation of flexizymes; B: the second generation of flexizymes; C: *In vitro* selection of dinitro-flexizyme (dFx).

Up to now, it has been developing for two generations. The first generation of flexizyme was obtained by *in vitro* selection from a random RNA sequence pool which contained a 70-necletiode random region at the

5’-terminus of tRNA.[33] After the pool was reacted with N-biotinylated phenylalanine cyanomethyl ester (N-biotin-Phe-CME), one single active sequence was selected. After the mapping and simplification of the secondary structure, a minimal ribozyme was generated, called r24mini, which functioned as both *cis*-acting and *trans*-acting ribozyme (Figure 6A). This opened up the opportunity to develop a practical tool for tRNA aminoacylation. However, r24mini could not be used as a versatile catalyst. Then the second generation of flexizyme was developed to gain more versatile and effective flexizymes toward kinds of tRNAs based on the structure-activity relationship of r24mini.[31] One doped pool where 19nt random mutations were introduced was constructed and performed using N-biotin-Phe-CME, giving a lot of active sequences (Figure 6B). After the alignment of these sequences, a 45-nt long sequence was obtained, called Fx, which is able to aminoacylate various of tRNAs.

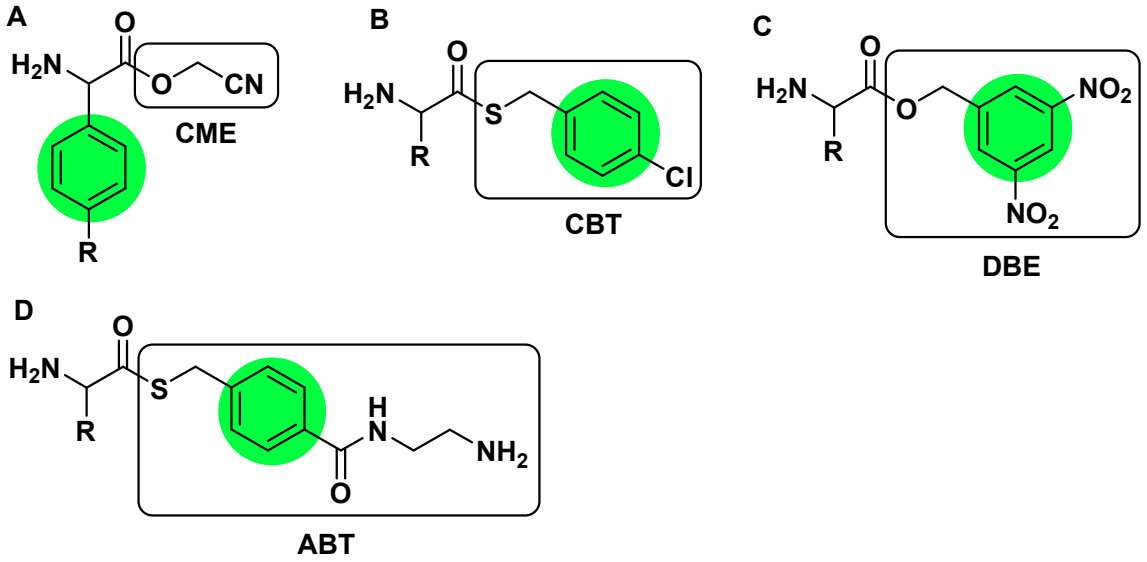


Fig. 7. Structures of amino acids substrates. The aromatic moiety recognized by flexizyme is shown in green and the leaving group is indicted by the boxed region. A: amino acid-CME; B: amino acid-CBT; C: amino acid-DBE; D: amino acid-ABT.

Although the second generation Fx was versatile toward tRNAs, it was not so effective to recognize the amino acids. Flexizymes are active to the aromatic moieties of aminoacyl-CMEs, such as Phe, Tyr, Trp and their derivatives. But flexizymes are not active for those nonaromatic side chains because Fx strictly recognizes the aromatic side chain. Based on this, the structure of the leaving group is not critical to the Fx. Chlorobenzyl thioester (CBT) is used in place of CME, the Fx recognizes the aromatic moiety of CBT more effectively than the amino acid side chain. Similar to CBT, 3, 5-Dinitrobenzyl ester (DBE) and 2-(aminoethyl)amidocarboxybenzyl thioester (ABT) are also recognized effectively by Fx (Figure 6C and Figure 7).

Using the flexizymes, many kinds of compounds can be introduced into peptides and proteins. The compounds are not limited to the α-L-amino acids and their derivatives with diverse of side chains, D-amino acids,[34-35] β-amino acids, the small polypeptides and other non-natural amino acids.

1.3.2.4 The encoded codons

One of the requirements for incorporating non-natural amino acids into proteins or peptides in these two biochemical approaches is that the tRNA must carry the anticodon which can recognize the specific codon on mRNA. There are mainly two kinds of codons for the incorporation of non-natural amino acids. One is the stop codons (UAA, UAG and UGA) which do not encode any amino acids. Since only one stop codon is required during the translation and the other two can be used to specify the non-natural amino acid. But usually the amber codon (UAG) is used to reassign the non-natural amino acid. Because that the amber stop codon (UAG) is the least used of

these three stop codons in E.coli (～9%) and yeast (～23%) and it rarely terminates the translation. In 1989, Schultz and co-workers[12] developed a general method which took advantage of the stop codon UAG (the amber codon) to incorporate the non-natural amino acid. And this method also allows the site-specific incorporation of a large number of non-natural amino acids into proteins with excellent translational fidelity. However, the amber suppression codon has a disadvantage that the incorporation of non-natural amino acid must compete with a release factor (RF).

Another one is the frame-shift codon (CGGG and AGGU) which is also the popular method for site-specific incorporation of the non-natural amino acids into proteins, which was developed by Sisido and co-workers.[36] The four-base codon also must compete with a three-base decoding by endogenous tRNAs. But the competition becomes in favour by the used of four-base codons, which derived from rarely used codons such as CGG and GGG. Not only the four-base codon, but also the five-base codon (CGGUA)[37] which was also developed by Sisido was used for the site-specific incorporation of non-natural amino acids.

A lot of non-natural amino acids have been incorporated into peptides or proteins through the amber codon or the frame-shift codon with excellent translational fidelity.

1.4 The applications

The incorporation of non-natural amino acids with biophysical labels, spectroscopic probes, metal chelators and unique functional groups will provide new tools for investigating the structure and function, and

generating proteins or peptides with novel physical, chemical and biological properties.

1.4.1 Altering the properties

1.4.1.1 Improving the enzyme activities

It is found that the modified enzyme exhibited very high catalysis and novel bioactivity after the incorporation of non-natural amino acids.[38] One example is the bacterial phosphotriesterases (PTE) which catalyze hydrolysis of the pesticide paraoxon with very fast turnover rates and are thought to be near their evolutionary limit for this activity. But the turnover rates were increased to 8-11 folds when the non-natural amino acids L-(7-hydroxycoumarin-4-yl)ethylglycine and L-(7-methylcoumarin-4-yl)ethylglycine were site-specifically incorporated into PTE in place of Tyr309.[39] These results provide that one easy access for improving the existing enzyme functions.

1.4.1.2 Enhancing the thermo-stability

The thermo-stability of the proteins can be enhanced by the incorporation of non-natural amino acids into proteins. One non-natural amino acid, *p*-fluorophenylalanine (pFF), was site-specifically incorporated into phosphotriesterase (PTE) by replacing the phenylalanine residues.[40] This modified protein retained the refoldability and no loss of the activity was observed at 60℃, but the wild-type PTE deactivated significantly at 60℃. The thermo-stability of Horseradish peroxidase was increased about 10 and 9 folds respectively after it was modified by phthalic anhydride and

glucosamine hydrochloride.[41] Another example is the bovine pancreatic α-chymotrypsin. Its thermo-stability was enhanced by about 4-6℃ after conjugating with mono-6-formyl-β-cyclodextrin and mono-6-succinyl-6-deoxy-β-cyclodextrin.[42]

1.4.1.3 Enhancing the bioactivity

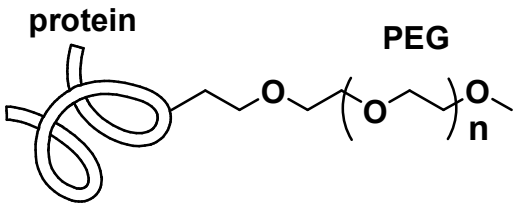


Fig. 8. PEGylated protein.

The bioactivity can be enhanced by the incorporation of non-natural amino acids. One of the successful strategies is the incorporation of polyethylene glycol (PEG),[43] which is non-toxic, non-immunogenic, hydrophilic, and approved by the U. S. Food and Drug Administration (FDA). And the process is termed as PEGylation. PEGylation of proteins, peptides or drugs can overcome their drawbacks: the low stability, the bad biocompatibility and fast degradation *in vivo* (Figure 8). Until to now, the PEGyaltion has been developed for two generations. The first generation of PEGylation is the attachment of PEG chains employing the natural amino acids. The selectivity of PEGylation is low and the product is a mixture. The second generation is on the non-natural amino acids with chemoselective reactions. One non-natural amino acid with a functional group is incorporated into proteins by biochemical approach. And then the PEGylation reagent is addressed to react with the functional group, for

example, the click reaction and the thiol-ene reaction. The selectivity of PEGylation in this method is very high.

Up to now, a lot of PEGylated proteins and peptides have been applied in biological and medical fields.[44] Some of them have been approved by FDA and used in clinical, such as PEG-ADA,[45] PEG-L-asparaginase,[46] peginterferon alfa-2b, peginterferon alfa-2a, and epgfilgrastim and so on.

1.4.2 Biophysical probes

Fluorescent groups, spin labels and IR probes can be site-specifically introduced into proteins as biophysical probes to investigate the relationship between the protein structure and function with great accuracy.

A large number of fluorescent non-natural amino acids have been site-specifically incorporated into proteins to image the protein expression, dynamics and function. The fluorescent amino acid L-(7-hydroxycoumarin-4-yl)ethylglycine (HceG), which is sensitive to solvent polarity and pH, has been site-specifically into sperm whale myoglobin to monitor biomolecular interactions or conformational changes in proteins.[47] Another environmentally sensitive fluorophore, dansyl amino acid, was also site-specifically incorporated into human superoxide dismutase and used as an environmentally sensitive reporter of protein unfolding.[48] And the fluorophore, 6-propionyl-2-(N, N-dimethyl)aminonaphthalene,[49] was used to study ligand-induced local conformational changes in proteins and biomolecular interactions *in vitro*. Most interestingly, fluorescence resonance energy transfer (FRET) between the fluorescent donor and the acceptor could be used to investigate the interaction or conformational changes between proteins. Two fluorescent amino acids, including 4-biphenyl-L-pehnylalanine and L-(7-hydroxycoumarin-4-yl)ethylglycine,

were site-specifically incorporated into dihydrofolate reductase (DHFR) to detect the conformational change associated with inhibitor binding by FRET.[50]

Some spin labels are introduced into a protein at a defined position, which can greatly simplify signal assignment of NMR spectra of large proteins or complexes. The site-specific incorporations of ^{19}F-[51], ^{13}C-[52], ^{15}N-labeled[53] non-natural amino acids into proteins can identify the conformational changes occurred in ligand binding by comparing the NMR spectra of different mutants.

Some non-natural amino acids with unique IR spectroscopy have also been incorporated into proteins to investigate the local environment and the conformational changes. The azido group is sensitive to its electrostatic environment and its antisymmetric stretch vibration is at $\sim$ 2100 cm^{-1} in a clear spectral window. One IR- active amino acid, *p*-azido-L-phenylalanine has been incorporated into the G protein-coupled receptor (GPCR) rhodopsin.[54] Through Fourier transform infrared (FTIR) difference spectroscopy, it can reveal the electrostatic environment changes during conformational transitions of the GPCR rhodopsin. Another IR probe is *p*-cyano-L-phenylalanine, the cyano group of which absorbs in a clear spectral window at $\sim$2200cm^{-1}. It was incorporated into myoglobin at His64 to probe metal ion and ligand binding because it is sensitive to subtle changes in local environment, as well as the changes of nitrogen with metal ions.[55]

1.4.3 Photo-control of structure and function

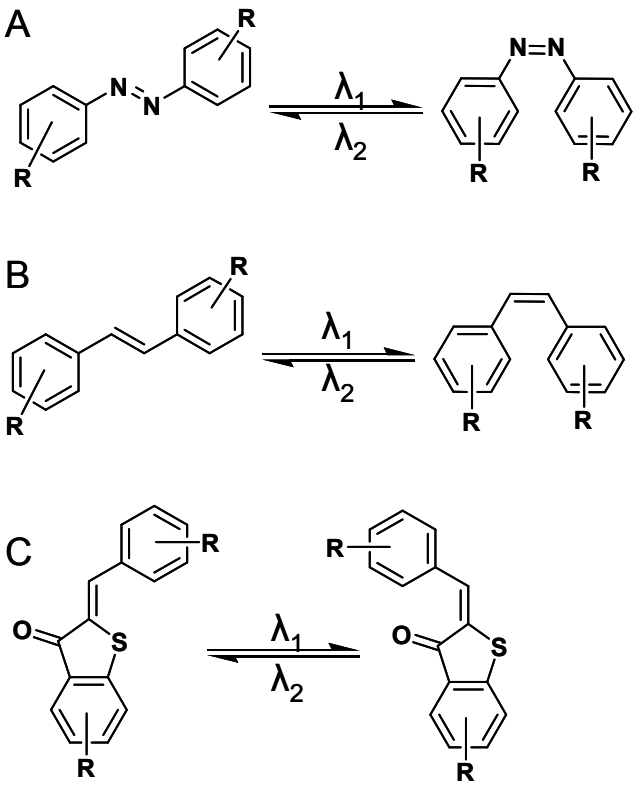


Fig. 9. Some photo-responsive molecules. A: azobenzenes; B: stilbenes; C: Hemithioindigos.

The incorporation of photo-responsive molecules into peptides or proteins can control the structure and function through the photo-induced transformation. Light seems to be the ideal external control element without the contamination of the sample. Various photo-responsive molecules can undergo a reversible change in structure under light irradiation (Figure 9). Among these photo-responsive molecules, azobenzenes are one of the largest and most popular molecules. The photo-responsive peptides were reported for the first time by Goodman,[56-60] who incorporated azobenzenes into peptides and investigated the effects of photo-responsiveness and the structure changes of polypeptides. Since then, a lot of polypeptides containing azobenzenes have been developed. It showed that the photo-

responsiveness was low when the azobenzene molecule was on the side chain of the polypeptide[61] (Figure 10A). The reason is that the small structure changes of azobenzene can not change the structure of the large peptides effectively. It was realized that the position of the azobenzene molecules was quite important. Then the azobenzene molecules are incorporated into the linear peptide backbone.[62] Through the structure changes of the azobenzene molecules (Figure 10B), the structure of the peptide can change much. Also, the cyclic peptides containing the azobenzene molecules have been developed[63] (Figure 10C). It showed very high photo-responsiveness.

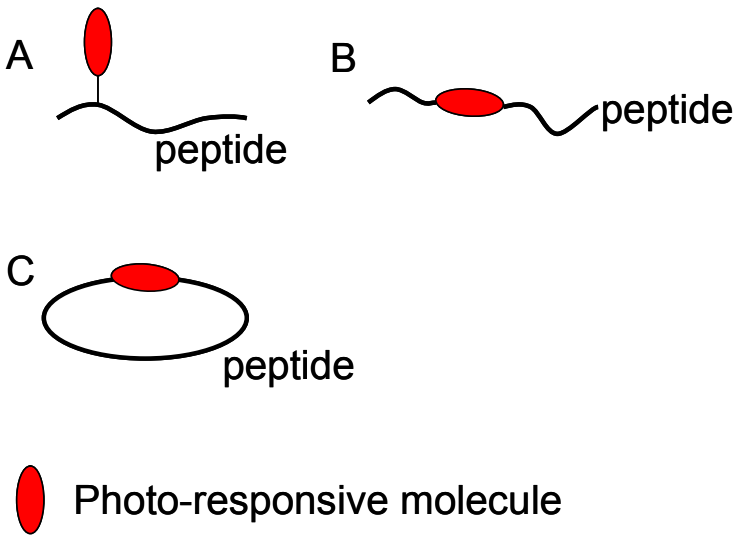
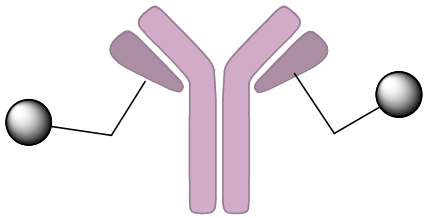


Fig. 10. The different positions of photo-responsive molecules. A: on the side chain; B: in the linear backbone; C: in the cyclic backbone.

The photo-responsive non-natural amino acids not only can control the structure and function of peptides, but also can provide tools for purify proteins, regulation of drug release and other special functions.

1.4.4 Therapeutics

The non-natural amino acids can be genetically introduced for synthesizing antibody-drug conjugates (ADCs) (Figure 11), which are a new class of highly potent biopharmaceutical drugs for the treatment of cancer.[64] ADCs are complex molecules composed of an antibody linked to a drug with labile bonds. These ADCs are highly-targeted because ADCs possess sensitive discrimination between healthy cells and the tumour cells.



Antibody-drug conjugates

Fig. 11. Antibody-drug conjugates (ADC)

The first generation of ADCs conjugate drugs non-selectively to cysteine or lysine residues in the antibody, resulting in a mixture of ADCs.[65] This kind of ADCs has the challenge of suboptimal safety and efficacy properties. The site-specific incorporation of non-natural amino acids provides the control of the attachment of drugs. So it allows ADCs can be used for market.

To date, more than 30 ADCs are in clinical trials. But only three ADCs have received market approval. The first ADC received market approval was gumtuzumab ozogamicin,[66] an anti-CD33 antibody conjugated with the DNA cleaving agent calicheamicin, which was withdrawn by the U.S. Food and Drug Administration (FDA) in 2010 because of toxicity and lack of

efficacy in larger trials.[67] Brentuximab vedotin,[68] an anti-CD30 antibody conjugating with the microtubule-disrupting agent monoethyl auristatin E, has been approved for Hodgkin's lymphoma by the U. S. FDA. Another one is Trastuzumab emtansine,[69] which is used for the treatment of HER2-positive metastatic breast cancer.

This site-specific incorporation method can also be applied to the synthesis of antibodies conjugated to kinase and phosphatise inhibitors, nuclear hormone receptor agonists and antagonists, and other drug classes for treatment of autoimmune, cardiovascular, and metabolic disease.

1.4.5 New materials

The non-natural amino acids can be used to develop the new materials that can satisfy the requirements such as biocompatibility, biodegradability, surface properties and mechanical properties. For example, sodium montmorillonite (Na-MMT) clay was modified with three different non-natural amino acids: (±)-2-aminopimelic acid, 5-aminovaleric acid and DL-2-aminocaprylic acid. Cell culture experiments showed that all the clays modified with these three different amino acids were biocompatible. It showed that these modified Na-MMT clays can be used for bone biomaterials applications.[70] The thin films were prepared by photo-cross-linking artificial extracellular matrix proteins containing the photosensitive amino acid *p*-azidophenylalalnine. Under ultraviolet irradiation, elastomers with moduli were yielded and these materials were used in conjunction with photolithographilic techniques to obtain patterned surfaces and with microfluidic technology to generate protein surface gradients.[71]

The incorporation of non-natural amino acids into proteins has generated new materials with useful properties and has begun to meet the important biological requirements.

1.5 Objects & Aims of this thesis

The objects & aims of this thesis are the following:

1. To study the *in vitro* incorporation of two PEGs with different lengths into one polypeptide backbone.
2. To get a photo-responsive peptide aptamer by *in vitro* selection with incorporating an azobenzene molecule;

Both of the objects are the applications of the expanded genetic code.

1.6 Structure of this thesis

Chapter 2 is to do the *in vitro* incorporation of two PEGs with different length into one peptide using both the amber codon and the four-base codon.

Chapter 3 is to get a photo-responsive peptide aptamer containing an azobenzene residue using ribosome display by *in vitro* selection. This aptamer can be a new tool for the purification of proteins because of the conformation changes of azobenene under UV irradiation / visible light. And meanwhile, the synthesis of tRNA carrying an azobenzene residue will be also studied.

Reference:

1. R J. D. Watson and F. H. C. Crick, A structure for deoxyribose nucleic acid. *Nature*, 171, 737-738, 1953.
2. F. Crick, Central dogma of molecular biology. *Nature*, 227, 561-563, 1970.
3. F. H. C. Crick, F. R. S. Leslie Barnett, S. Brenner and R. J. Watts-Tobin, General nature o the genetic code for proteins. *Nature*, 4809, 1227-1232, 1961.
4. M. W. Nirenberg and J. H. Matthaei, The dependence of cell-free ptrotein synthesis in E.coli upon naturally occurring or synthetic polyribobycleotides. *Proc. Natl. Acad. Sci. USA*. 47, 1588-1602,1961.
5. R. S. Gardner, A. J. Wahba, C. Basilio, R. S. Miller, P. Lengyel and J. F. Speyer, Synthetic polynucleotides and the amino acid code, VII-. *Proc. Natl. Acad. Sci. USA,* 48, 2087-2094, 1962.
6. A. J. Wahba, R. S. Gardner, C. Basilio, R. S. Miller, J. F. Speyer and P. Lengyel, Synthetic polynucleotides and the amino acid code, VIII-. *Proc. Natl. Acad. Sci. USA*, 49, 116-122, 1963.
7. M. Nirenberg, P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman and CO'Neal, RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc. Natl. Acad. Sci. USA,* 53, 1161-1168, 1965.
8. F.H. C. Crick, The origin of the genetic code. *J. Mol. Biol.,* 38, 367-379, 1968.
9. S. Anderson, A.T. Bankier, B. G. Barrell, M. H. L. De Bruijin, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden and I. G. Young, Sequence and organization of the human mitochondrial genome. *Nature*, 290, 457-465, 1981.
10. F. Zinoni, A. Birkmann, T. C. Stadtman and A. Böck, Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from Escherichia coli. *Proc Natl Acad Sci USA*. 83, 4650-4654, 1986.
11. G. Srinivasan, C. M. James and J. A. Krzycki, Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science*, 1459-1462, 2002.

12. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins. *Science,* 244, 182-188, 1989.
13. T. N. M. Schumacher, L. M. Mayr, D. L. Minor Jr., M. A. Milhollen, M .W. Burgess and P. S. Kim, Indentification of D-peptide ligands through mirror-image phage display. *Science*, 29, 1854-1857, 1996.
14. D. A. Dougherty, Unnatural amino acids as probes of proteins structure and function. *Curr. Opin. Cell. Biol.,* 4, 645-652, 2000.
15. Q. Shao and B. G. Xing, Photoactive molecules for applications in molecular imaging and cell biology. *Chem. Sci. Rev.,* 39, 2835-2846, 2010.
16. G. E. Means and R. E. Feeney, Chemical modification of proteins: history and applications. *Bioconjug. Chem.,* 1, 2-12, 1990.
17. S. B. Kent, Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem*., 57, 957-989, 1988.
18. R. B. Merrifield, Solid phase peptide synthesis. I. The synthesis of a tetrapepetide. *J. Am. Chem. Soc*., 85, 2149-2154, 1963.
19. E. T. Kaiser, Synthetic approaches to biologically active peptides and proteins including enzymes. *Acc. Chem. Res*., 22, 47, 1989
20. J. M. Xie and P. G. Schultz, An expanding genetic code. *Methods,* 36, 227-238, 2005.
21. J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. W. Zhang and P. G. Schultz, An expanded eukaryotic genetic code. *Science*, 301, 964-967, 2003.
22. J. W. Chin, T. A. Cropp, S. Chu, E. Meggers and P. G. Schultz, Progress toward an expanded eukaryotic genetic. *Chem. Biol*., 10, 511-519, 2003.
23. N. Wu, A. Deiters, T. S. Cropp, D. King and P. G. Schultz, A genetically encoded photocaged amino acid. *J. Am. Chem. Soc*., 126, 14306-14307, 2004.
24. H. Neumann, S. Y. Peak-Chew and J. W. Chin, Genetically encoding N^{ε}-acetyllysine in recombinant proteins. *Nat. Chem. Biol*., 4, 232-234, 2008.
25. P. G. Schultz and L. Wang, Expanding the genetic code. *Angew. Chem. Int. Ed*., 44, 34-66, 2005
26. S. M. Hecht, B. L. Alford, Y. Kuroda and S. Kitano, Chemical aminoacylation of tRNA's. *J. Biol. Chem*., 253, 4517-4520, 1978.

27. T. G. Heckler, L. H. Chang, Y. Zama, T. Naka, M. S. Chorghade and S. M. Hecht, T4 RNA ligase mediated preparation of novel chemically misacylated tRNAPheS. *Biochemistry*, 23, 1468-1473, 1984.
28. J. R. Roesser, C. Xu, R. C. Payne, C. K. Surratt and S. M. Hecht, Prepareation of misacylated aminoacyl-tRNAPhes useful as probes of the robisomal acceptor site. *Biochemistry*, 28, 5185-5195, 1989.
29. G. Baldini, B. Martoglio, A. Schachenmaann, C. Zugliani and J. Brunner, Mischarging Escherichia coli tRNAPhe with L-4'-[3-(trifluoromethyl)-3H-diazirin-3-yl]phenylalanine, a photoactivatable analogue of phenylalanine. *Biochemistry*, 27, 7952-7959, 1988.
30. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith and P. G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins. *Science,* 244, 182-188, 1989.
31. H. Murakami, H. Saito and H. Suga, A versatile tRNA aminoacylation catalyst based on RNA. *Chem. Biol*., 10, 655-662, 2003.
32. H. Murakami, A. Ohata, H. Ashigai and H. Suga, A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods*, 3, 357-359, 2006.
33. H. Saito, D. Kourouklis and H. Suga, An in vitro evolved precursor tRNA with aminoacylation activity. *EMBO. J.,* 20, 1797-1806, 2001.
34. L. M. Dedkova, N. E. Fahmi, S. Y. Golovine and S. M. Hecht, Construction of modified ribosomes for incorporation of D-amino acids into proteins. *Biochemistry*, 45, 15541-15551, 2006.
35. L. M. Dedkova, N. E. Fahmi, S. Y. Golovine and S. M. Hecht, Enhanced D-amino acid incorporation into protein by modified ribosomes. *J. Am. Chem. Soc*., 125, 6616-6617, 2003.
36. T. Hohsaka, Y. Ashizuka, H. Taira, H. Murakami and M. Sisido, Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli in vitro translation system. *Biochemistry,* 40, 11060-11064, 2001.
37. T. Hohsaka, Y. Ashizuka, H. Murakami and M. Sisido, Five-base codons for incorporation of nonnatural amino acids into proteins. *Nucl. Acids Res*., 29, 3646-3651, 2001.

38. D. G. Benjamin, Chemical modification of biocatalysts. *Curr. Opin. Cell. Biol.,* 14, 379-386, 2003.
39. I. N. Ugwamba, K. Ozawa, Z. Q. Xu, F. Ely, J. L. Foo, A. J. Herlt, C. Coppin, S. Brown, M. C. Taylor, D. L. Ollis, L. N. Mander, G. Schenk, N. E. Dixon, G. Otting, J. G. Oakeshott and C. J. Jackson, Improving a natural enzyme activity through incorporation of unnatural amino acids. *J. Am. Chem. Soc*., 133, 326-333, 2011.
40. P. J. Baker and J. K. Montclare, Enhaanced refoldability and thermoactivity of fluorinated phosphotriesterase. *Chembiochem*., 12, 1845-1848, 2011.
41. J. Z. Liu, H. Y. Song, L. P. Weng and L. N. Ji, Increased thermostability and phenol removal efficiency by chemical modified horseradish peroxidise. *J. Mol. Catal. B Enzyme,* 18, 225-232, 2002.
42. M. Fernandez, M. De L. Villalonga, A. Fragoso, R. Cao and R. Villalonga, Stabilization of α -chymotrypsin by modification with β-cyclodextrin derivatives. *Biotechnol. Appl. Biochem.,* 36, 235-239, 2002.
43. N. Nischan and C. P. R. Hackenberger, Site-specific PEGylation of proteins: recent development. *J. Org. Chem*., 79, 10727-10733, 2014.
44. S. M. Ryan, G. Mantovani, X. X. Wang, D. M. Haddleton and D. J. Brayden, Advances in PEGylation of important biotech molecues: delivery aspects. *Expert Opin. Drug Delicery*, 5, 371-383, 2008.
45. M. S. Hershfield, PEG-ADA: an alternative to haploidentical bone marrow transplantation and an adjunct to gene therapy for adenosine deaminase deficiency. *Hum. Mutat*., 2, 107-112, 1995.
46. C. H. Fu and K. M. Sakamoto, PEG-asparaginase. *Expert Opin. Pharmacother.*, 8, 1977-1984, 2007.
47. J. Y. Wang, J. M. Xie and P. G. Schultz, A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc*., 128, 8738-8739, 2006.
48. D. Summerer, S. Chen, N. Wu, A. Deiters, J. W. Chin and P. G. Schultz, A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. USA*, 103, 9785-9789, 2006.

49. H. S. Lee, J. T. Guo, E. A. Lemke, R. D. Dimla and P. G. Schultz, Genetic incorporation of small, environmentally sensitive, fluorescent probe into proteins in *Saccharomyces cerevisiae*. *J. Am. Chem. Soc*., 131, 12921-12923, 2009.
50. S. X. Chen, N. E. Fahmi, L. Wang, C. Bhattacharya, S. J. Benkovic and S. M. Hecht, Detection of dihydrofolate reductase conformational change by FRET using two fluorescent amino acids. *J. Am. Chem. Soc*., 135, 12924-12927, 2013.
51. M. Neerathilingam, L. H. Greene, S. A. Colebrooke, I. D. Campbell and D. Staunton, Quantitation of protein expression in a cell-free system: efficient detection of yields and 19F NMR to identify folded protein. *J. Biomol*., 31, 11-19, 2005.
52. J. N. Lampe, R. Brandman, S. Sivaramakrishnan and P. R. Ortiz de Montellano, Two-dimensional NMR and all-atom molecular dynamics of cytochrome P450 CYP19 reveal hidden conformational substates. *J. Bio. Chem.,* 285, 9594-9603, 2010.
53. S. E. Cellitti, D. H. Jones, L. Lagpacan, X. S. Hao, Q. Zhang, H. Y. Hu, S. M. Brittain, A. Brinker, J. Caldwell, B. Bursulaya, G. Spraggon, A. Brock, Y. H. Ryu, T. Uno, P. G. Schultz and B. H. Geierstanger, In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc*., 130, 9268-9281, 2008.
54. S. X. Ye, T. Huber, R. Vogel and T. P. Sakmar, FITR analysis of GPCR activation using azido probes. *Nat. Chem. Biol*., 5, 397-399, 2009.
55. K. C. Schultz, L. Supekova, Y. Ryu, J. M. Xie, R. Perera and P. G. Schultz, A generally encoded infrared probe. *J. Am. Chem. Soc.*, 128, 13984-13985, 2006.
56. M. Goodman and A. Kossoy, Conformational aspects of polypeptide structure. XIX. Azoaromatic side-chain effects. *J. Am. Chem. Soc.,* 88, 5010-5015, 1966.
57. M. Goodman and M. L. Falxa, Conformational aspects of polypeptide structure. XXIII. Photoisomerization of azoaromatic polypeptides. *J. Am. Chem. Soc*., 89, 3863-3867, 1967.
58. M. Goodman and E. Benedetti, Conformational aspects of polypeptide structure. XXVI. Azoaromatic side-chain effect from poly-L-p-(p'-hydroxyphenylazo)phenylalanine. *Biochemistry*, 7, 4226-4234, 1968.

59. E. Benedetti, A. Kossoy, M. L. Falxa and M. Goodman, Conformational aspects of polypeptide structure. XXVII. Solvent effects on azoaromatic polypeptides. *Biochemistry*, 7, 4234-4242, 1968.
60. E. Benedetti and M. Goodman, Conformational aspects of polypeptide structure. XXVIII. Side-chain cotton effect from poly-L-p-(2'-hydroxy-5'-methylphenylazo)phenylalanine. *Biochemistry,* 7, 4242-4247, 1968.
61. M. Z. Liu, S. Tada, M. Ito, H. Abe and Y. Ito, In vitro selection of a photo-responsive peptide aptamer using ribosome display. *Chem. Commun*., 48, 11871-11873, 2012.
62. T. M. Doran, E. A. Anderson, S. E. Latchney, L. A. Opanashuk and B. I. Nisson, An azobenzene photoswitch sheds light on turn nucleation in amyloid- β-self-assembly. *ACS Chem. Neurosic*., 3, 211-220, 2012.
63. R. M. Abaskharon, R. M. Culik, G. A. Woolley and F. Gai, Tuning the attempt frequency of protein folding dynamics via transition-state rigidification: application to Trp-Cage. *J. Phys. Chem. Lett*., 3, 521-526,
64. I. Sassoon and V. Blanc, Antibody-drug conjugate (ADC) clinical pipeline: a review. *Methods Mol. Biol*., 1045, 1-27, 2013.
65. H. Bouchard, C. Viskov and C. Garcia-Echeverria, Antibody-drug conjugates- a new wave of cancer drugs. *Bioorg. Med. Chem. Lett*., 24, 5357-5363, 2014.
66. F. J. Giles, H. M. Kantarjian, S. M. Kornblau, D. A. Thomas, G. Garcia-Manero, T. A. Waddelow, C. L. David, A. T. Phan, D. E. Colburn, A. Rashid and E. H. Estey, Mylotarg TM (gemtuzumab ozogamicin) therapy is associated with hepatic venoocclusive disease in patients who have not received stem cell transplantation. *Cancer*, 92, 406-413, 2001.
67. A. Beck, J. F. Haeuw, T. Wurch, L. Goetsch, C. Bailly and N. Corvaia, The next generation of antibody-drug conjugates comes of age. *Discov. Med*., 53, 329-339, 2010.
68. J. Katz, J. E. Janik and A. Younes, Brentuximab vedotin (SGN-35). *Clin. Cancer. Res*., 17, 6428-6436, 2011.
69. N. Duvaz, Trastuzumab emtansine, an antibody-drug conjugate for the treatment of HER2+ metastatic breast cancer. *Curr. Opin. Mol. Ther*., 12, 350-360, 2010.

70. K. S. Katti, A. H. Amber, N. Peterka and D. R. Katti, Use of unnatural amino acids for design of novel organomodified clays as components of nanocomposite biomaterials. *Phil. Trans. R. Soc. A,* 368, 1963-1980, 2010.
71. P. J. Nowataki, C. Franck, S. A. Maskarinec, G. Ravichandran and D. A. Tirrell, Mechanically tunable thin films of photosensitive artificial proteins: preparation and characterization by nanoindentation. *Macromolecules*, 41, 1839-1845, 2008.

CHAPTER 2

IN VITRO TRANSLATION OF TWO PEGS WITH DIFFERENT LENGTHS INTO ONE POLYPEPTIDE BACKBONE

2.1 Introduction

Proteins and peptides have great promise as therapeutic agents.[1] However, their applications *in vivo* are limited due to their low stability, low biocompatibility, and fast clearance.[2] These problems can be overcome by attaching the macromolecules.[3,4] One of the most successful methods is the covalent coupling of polyethylene glycol (PEG), termed as PEGylation, which is chemically inert, non-toxic, non-immunogenic, hydrophilic, and approved by the U. S. Food and Drug Administration (FDA) (Figure 1). The first PEGylated protein was approved to enter the market by the U. S. FDA in 1990, which is used to treat severe combined immunodeficiency disease (SCID) associated with a deficiency of adenosine deaminase (ADA).[5] Since the first introduction of PEGylated protein, a large number of PEGylated proteins have been taking the clinical trial or approved by the U. S. FDA.[6]

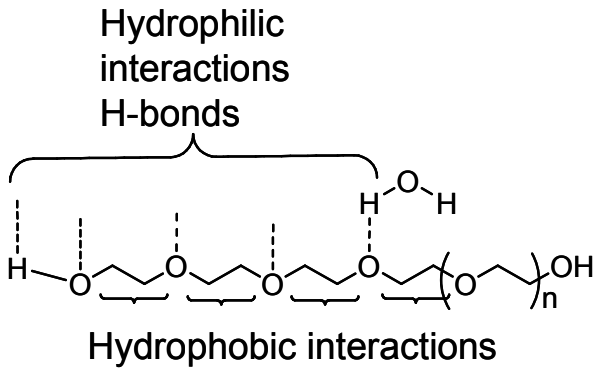


Fig. 1. Structure of polyethylene glycol.

2.2 The development of PEGylation

The PEGylation has attracted great attention since the first report of PEGylation of proteins by Abuchowsky and Davies in 1970s.[7,8] After that, many efforts have been done to modify peptides and proteins with PEG. Up to now, the PEGylation has been developed for two generations.

2.2.1 The first generation of PEGylation

The first generation of PEGylation is the attachment of PEG mainly on the nucleophiles in the side chains of natural amino acids in the proteins (Figure 2). These nucleophiles include amino group, thiol group, and hydroxyl group and so on. Then the electrophiles carrying PEG are reacted with the active groups on the proteins. Because of the numerous of nucleophiles in one protein, usually an excess of PEGylation reagents is needed to achieve reasonable conversions. However, this kind of modification with PEG leads to a mixture of PEGylation product, whereas the analysis and purification are very difficult.[9] Moreover, the activity of the proteins decreases by the unwanted PEGylation near or at the active site of the proteins.

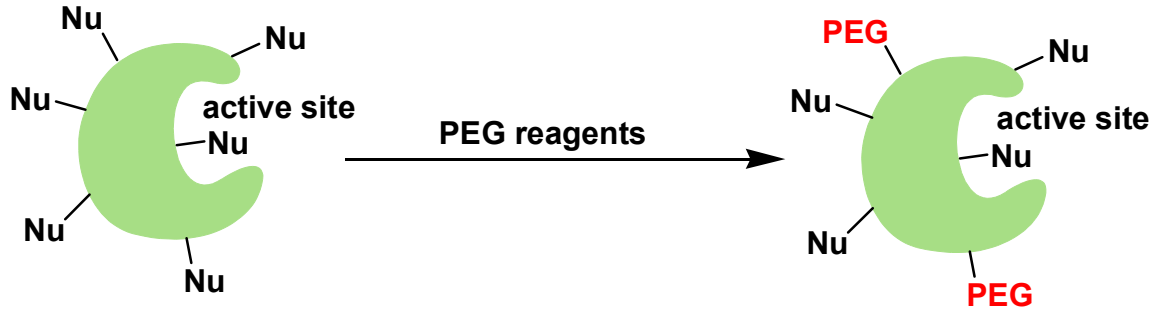


Fig. 2. The first generation of PEGylation. Random PEGylation of proteins.

In order to enhance the selectivity of the PEGylation, some special reactive groups are placed on the surface of the proteins. The thiol group is often used to react with the PEGylation reagent.[10] Also, tyrosine can also react with diazo acetophenone, which can conjugate with PEGylation reagent.[11] Another approach is that the PEGylation of the N- or C- terminus of a protein. The N-terminus can be selectively reacted with a PEGylation reagent such as thioester by native chemical ligation.[12] The C-terminus can be modified with oxyamines and then functionalized with a pyruvoyl-PEG reagent.[13]

The main problem of the first generation PEGyaltion is that the selectivity is limited or the reactive sites are confined to a few groups.

2.2.2 The second generation of PEGylation

The second generation of PEGylation is developed to achieve complete control of the PEGylation through the site-specific introduction of non-natural amino acids (Figure 3). The PEGylated amino acids can be directly incorporated into proteins using the amber codon or the frameshift codon. However, the length of PEG is limited to very short PEGs.[14]

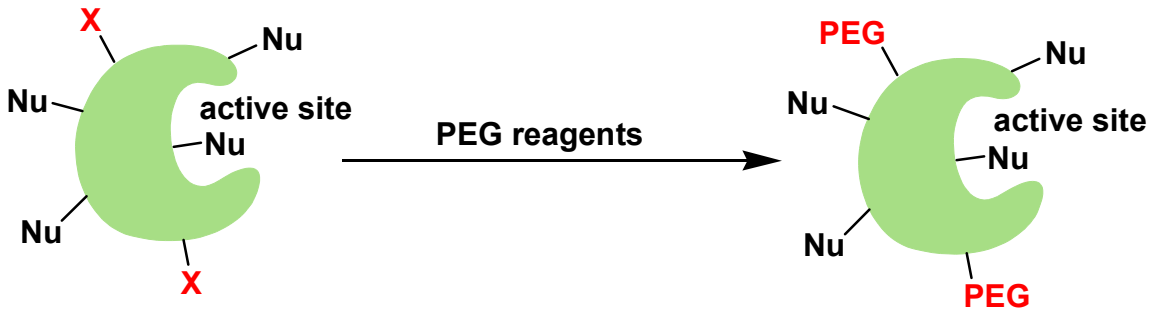
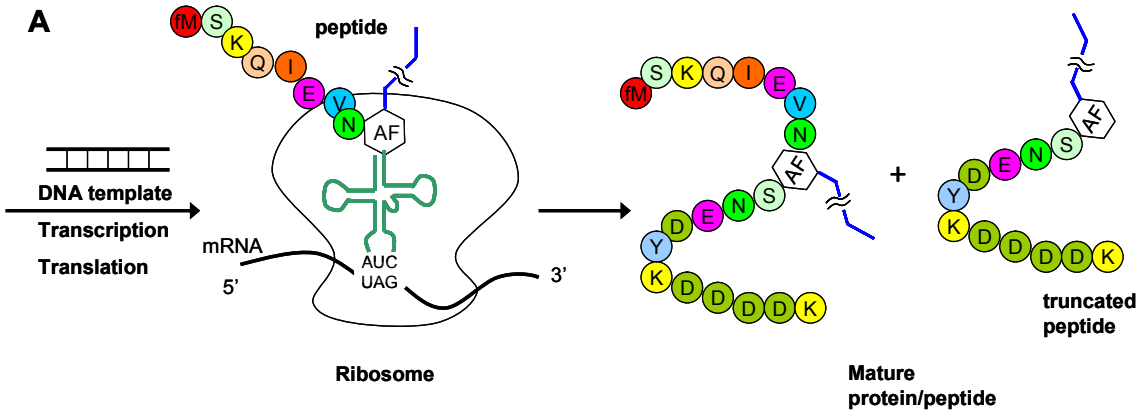


Fig. 3. The second generation of PEGylation.

Another approach is that biochemical incorporation of a non-natural amino acid carrying an orthogonal reaction group which can react with the PEGylation reagents with high selectivity.[15] A variety of orthogonal groups have been introduced into proteins. For example, azides are one very popular group, which can be reacted in copper-catalyzed azide alkyne cycloadditions (CuAAC) to generate a triazole linker.[16] Although catalytic amount of Cu (I) are sufficient for the mild chemoselective [3+2] cylcoaddition in aqueous media, there are some limitations including the high cytotoxicty.[17] A metal-free chemoselective method for azides is the strain-promoted azide alkyne cycloaddition (SPAAC) using strained cyclic alkynes.[18] Azides can also be reacted with triaryphosphine-PEG reagent using Staudinger reaction.[19] Besides the azides, some keto/aldehyde functionalities,[20,21] palladium-mediated conjugation methods[22] are also developed for the PEGylation.

Various efforts are being made to achieve the site PEGylation and maintain the full activity of the protein. The site-specific PEGylation is achieved by exploring new chemoselective reactions, but the full activity is difficult to maintain because all the introduced PEGs have the same lengths until to now. The incorporation of two PEGs with different lengths into one peptide or protein is not reported.



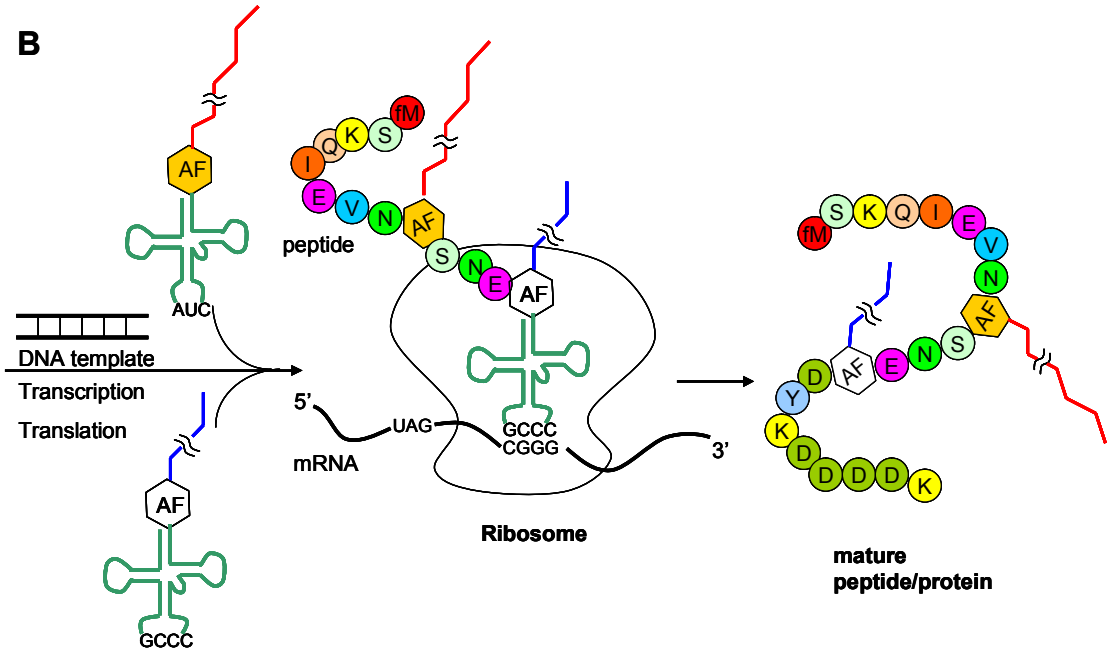


Fig. 4. The incorporation of PEG into a polypeptide by *in vitro* translation. A: *in vitro* translation of single PEG; B: *in vitro* translation of double PEGs with different lengths.

I prepared two DNA sequences containing both the amber and the frame-shift codon. Then PEGs with different lengths were genetically incorporated into one polypeptide backbone with tRNAs which were acylated with PEG-containing amino acids (Figure 4). The advantage of using misacylated tRNA is multiple specific incorporation of non-natural amino acids into proteins.[23,24] By using the genetic method, it is possible to precisely insert more than two PEG chains of different lengths into each desirable position in one protein molecule. Therefore, in this study, we attempted to site-specifically incorporate one and two PEGs by adding tRNAs carrying PEG of various lengths that recognize a stop codon and a frameshift codon via a translation system, as shown in Figure 4. And we also observed the interesting truncated peptides occurred in the single PEGylation, which have

been reported by other researchers,[25] although the reason is unknown. These data are expected to be as the raw materials for studying the reason.

2.3 Materials and methods

2.3.1 Materials

Primer fp-(FL92-T7prom): 5'- CGCGAAATTAATACGAGTCAC-3'

Primer rp-(Prox)FLAG-TGA: 5'-
TCACTTGTCATCGTCATCCTTGTAGTCCTCATTAGACTAGTTTACT
TCGATTTG-3'

Primer rp(ProX)4bFLAG-TGA:
5'TCACTTGTCATCGTCATCCTTGTAGTCCCCGCTCATTAGACTAGT
TTACTTCGATTTG-3'

Primer rp-(Prox)FLAG-CGGG:
5'-
TCACTTGTCATCGTCATCCTTGTAATCCTCATTAGACCCGGTTTAC
TTCGATTTG-3'

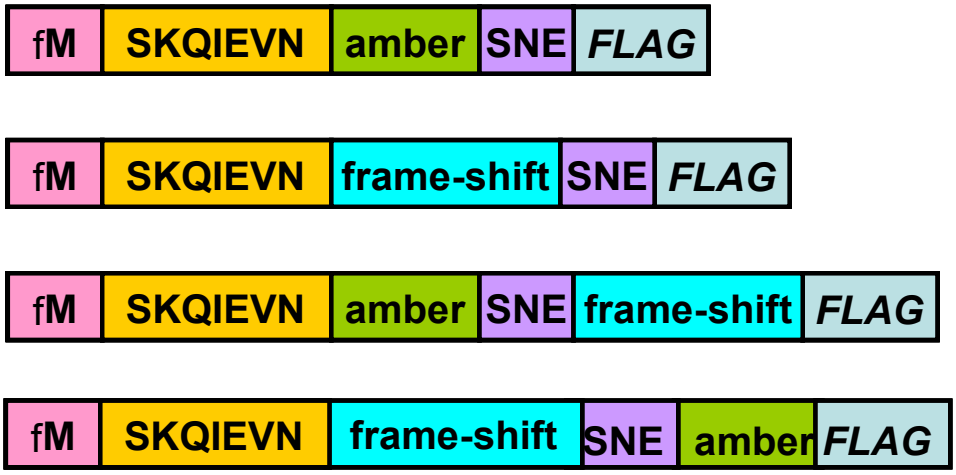
Primer rp(Prox-4b)-amb-FLAG-TGA:
5'-
TCACTTGTCATCGTCATCCTTGTAGTCCTACTCATTAGACCCGGTT
TACTTCGATTTG-3'

Remarkable Yield Translation System Kit (RYTS Kit) (ProteinExpress, Chiba, Japan)

Washing buffer: 50 mM Tris-HCl, 150 mM NaCl, pH 7.6.

2.3.2 Preparation of amioacyl tRNAs

First, PEG-AF-pdCpA was synthesized as following. To a solution of PEG-succinimide ester in DMSO (25mM, 40μl) was added a solution of aminophenylalanyl-pdCpA (AF-pdCpA) in DMSO (5mM, 40μl) and aqueous pyridine-HCl (2.5M, pH=5.0, 80μl). After incubation at 37℃ for 3h, PEG-AF-pdCpA was purified by reverse-phase HPLC (Waters XTerra C18: 2.5μm, 4.6mm x 20mm) at a flow rate of 1.5 ml/min, with a linear gradient of 0-100% acetonitrile in 0.1% trifluoroacetic acid for 10 min. The product was confirmed by MALDI-TOF-MS.



fM: formylated methionine

SKQIEVN: Ser-Lys-Gln-Ile-Glu-Val-Asn

SNE: Ser-Asn-Glu

FLAG: Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys

amber: incorporation site of functionalized amino acid-tRNA^{CUA}

frame-shift: incorporation site of functionalized amino acid-tRNA^{CCCG}

Fig. 5. Prepared DNA templates.

The resulting PEG-AF-pdCpA was ligated to the amber or frame-shift suppressor tRNA derived from Mycoplasma capricolum Trp1 tRNA without the 3'dinucleotide by chemical ligation method as described. The PEG-AF-tRNA can be obtained as commercially available (CoverDirect tRNA reagents for site-directed protein labelling, ProteinExpress, Chiba, Japan).

2.3.3 Prepared DNA templates

The DNA templates for the cell-free translation encode the peptides were prepared (Figure 5). The plasmid DNA was prepared by pROX-FL92.1amber as a template, and primers were shown below. The PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) and used as the templates for cell-free translation reactions.

2.3.4 *In vitro* translation

The *in vitro* translation was performed using RYTS kit (ProteinsExpress, Chiba, Japan) according to the manufacture's protocol. The translation reaction was performed in the presence of 200pmol of each PEG-AF-tRNA, unless otherwise stated. The reaction mixture was incubated at 30℃ for 2h.

2.3.5 Mass spectra measurements

The samples were prepared for mass spectrometry as previously reported.[26] The translated peptides were purified from 25 μL reactions using

prewashed Anti-DDDDK-tag mAb-Magnetci Agarose (MBL). After two washes with buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.6), the peptides were eluted from the matrix with 0.2% TFA. Before freeze drying, 3xFlag Peptide DYKDHDGDYKDHDIDYKDDDDK (Sigma) (1 nmol or 10 fmol) was added to each sample as the internal standard. For the mass analysis, the peptides were desalted using ZipTip μ-C18 (Millipore) and mixed with 2,5-dihydroxybenzoic acid or 3-hydroxy-2-pyridinecarboxylic acid as the matrix. The samples were subjected to MALDI–TOF–MS analysis on an Ultraflex spectrometer (Bruker Daltonics) in linear or reflector mode. Each sample was analyzed for 3 times and the yield efficiencies of PEG-incorporated peptides were estimated by their relative amounts comparing with 3 x Flag peptide.

2.4 Results and discussion

All the polypeptide containing one PEG and two PEGs with different lengths were measured by MS. And all the MS data are shown in Table 1. From Table 1, we can find the MS peak which is assigned to the full peptide sequence and the peak corresponding to the peptide sequence without formylmethionine (fM) was also observed. This is because of the methionylaminopeptidase (MAP) in the E. coli translation system, which can catalyze the N-terminal methionine excision when the second amino acid is glycine (Gly), alanine (Ala), serine (Ser) or threonine (Thr).[27] For the incorporation of single PEG, we also observed the unexpected peptide without the peptide sequence before the non-natural amino acid (lack of fMSKQIEVN or SKQIEVN), which we termed as truncated peptide. Although the truncated peptides have been reported by other researchers,[25]

the reason is unknown. These MS data for the truncated peptides can be the raw materials to study the reason.

position	sequence	M.W.Calculated $[M+H]^+$	MW. found
PEG4-amber	fMSQKIEVNXSNEDYKDDDDK	2681.178	2681.437
	SQKIEVNXSNEDYKDDDDK	2522.142	2522.389
	XSNEDYKDDDDK	1723.719	1723.021
PEG8-amber	fMSQKIEVNBSNEDYKDDDDK	2857.282	2857.416
	SQKIEVNBSNEDYKDDDDK	2698.247	2698.390
	BSNEDYKDDDDK	1899.824	1899.948
PEG4-CGGG	fMSQKIEVNXSNEDYKDDDDK	2681.178	2861.451
	SQKIEVNXSNEDYKDDDDK	2522.142	2522.428
	XSNEDYKDDDDK	1723.719	1723.977
PEG8-CGGG	fMSQKIEVNBSNEDYKDDDDK	2857.282	2857.623
	SQKIEVNBSNEDYKDDDDK	2698.247	2698.562
	BSNEDYKDDDDK	1899.824	1900.075
PEG12-amber	fMSQKIEVNXSNEDYKDDDDK	3033.387	3033.970
	SQKIEVNXSNEDYKDDDDK	2874.352	2874.890
	XSNEDYKDDDDK	2075.928	2076.311
PEG24-amber	fMSQKIEVNXSNEDYKDDDDK	3561.702	3562.222
	SQKIEVNXSNEDYKDDDDK	3402.667	3403.168
	XSNEDYKDDDDK	2604.243	2604.622
PEG4(amber)--PEG4(CGGG)	fMSQKIEVNXSNEXDYKDDDDK	3061.372	n.d.
	SQKIEVNXSNEXDYKDDDDK	2902.337	2903.113
PEG4(amber)--PEG8(CGGG)	fMSQKIEVNXSNEBDYKDDDDK	3237.477	n.d.
	SQKIEVNXSNEBDYKDDDDK	3078.442	n.d.
PEG8(amber)--PEG4(CGGG)	fMSQKIEVNBSNEXDYKDDDDK	3237.477	3239.239
	SQKIEVNBSNEXDYKDDDDK	3078.442	3079.390
PEG8(amber)--PEG8(CGGG)	fMSQKIEVNBSNEBDYKDDDDK	3413.582	n.d.
	SQKIEVNBSNEBDYKDDDDK	3254.547	n.d.
PEG4(CGGG)--PEG4(amber)	fMSQKIEVNXSNEXDYKDDDDK	3061.372	3062.304
	SQKIEVNXSNEXDYKDDDDK	2902.337	2903.271
PEG4(CGGG)--PEG8(amber)	fMSQKIEVNXSNEBDYKDDDDK	3237.477	n.d.
	SQKIEVNXSNEBDYKDDDDK	3078.442	3079.491
PEG8(CGGG)--PEG4(amber)	fMSQKIEVNBSNEXDYKDDDDK	3237.477	3239.195
	SQKIEVNBSNEXDYKDDDDK	3078.442	3080.503
PEG8(CGGG)--PEG8(amber)	fMSQKIEVNBSNEBDYKDDDDK	3413.582	n.d.
	SQKIEVNBSNEBDYKDDDDK	3254.547	n.d.

Table 1. MS data of PEGylation product. X and B indicate AF-PEG (4, 8, 12, 24).

2.4.1 PEGylation with single codon

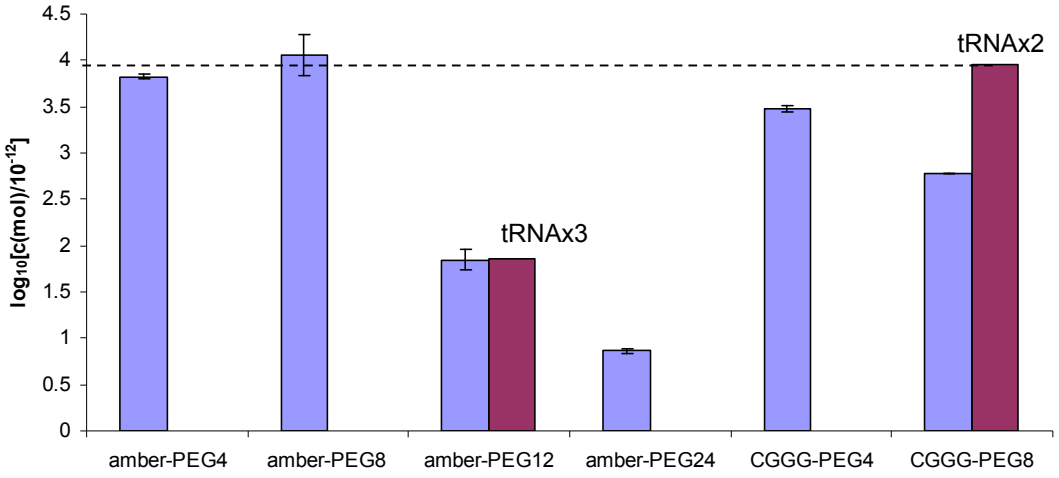


Fig. 6. Yield efficiency of single PEGylation of different lengths (PEG4, PEG8, PEG12, and PEG24) using amber or frame-shift (CGGG) codon. tRNAx2 and tRNAx3 mean two and three fold concentrations of tRNA in translation system.

PEG was successfully incorporated by the cell-free translation system. The incorporation of longer PEG chain resulted in the less amount of the translation product (Figure 6). For example in the cases of the amber tRNA system, the PEG4 and PEG8 exhibit almost the same amount of the translation products, although the amount for PEG12 and 24 are 1/100 and 1/1000 compared with that for the PEG4 and PEG8. The PEG-length dependence of the translation product from PEG8 to PEG24 may be explained by the previously reported steric hindrances between PEG and ribosome,[14] and between PEG and EF-Tu.[28] In this study we have found that the similar amount of translation products for PEG4 and PEG8, implying that the molecular weight from 170 to 340 (corresponding PEG4 and PEG8) may not sterically block the translation in a ribosome and/or binding of

tRNA to EF-Tu. It is noteworthy that we have confirmed that the amounts of PEGylation products are not limited by the amount of tRNAs; effectively the same amount of the translation product was obtained even in the case of triple amount of tRNA (tRNAx3) to incorporate PEG12.

The translation efficiency of PEG8 with the frame-shift codon is over 10-fold lower than that with the amber codon, whereas the translation efficiency of PEG4 at the frame-shift codon is just 2-fold lower than that with the amber codon (Figure 6). The translation efficiency of PEG8 with the frame-shift codon using $tRNA_{CCCG}$ is considerably lower compared with the other set of PEG4 and PEG8 presumably due to the competitive translation of $tRNA_{CCG}$(Arg) in the translation system. This consideration is supported by the experiment that the double amount of the $tRNA_{CCCG}$ (tRNAx2) increased the yield of the translation product to the similar amount as that for the amber-PEG8 (Figure 6).

2.4.2 PEGylation with double codon

In response to the above success of the incorporation of the different PEG lengths at either amber or frame-shift codon, we site-specifically incorporated two PEGs via the two codons in a peptide. First, two PEG4s were site-specifically incorporated into a peptide; the first PEG4 was at the position 9 via the amber codon and the second PEG4 was at the position 13 via the frame-shift codon (entry 1 Figure 7). Although the yield of the translation product with two PEG4s was about 2000-fold lower than single PEG4 incorporation, we were able to successfully observe the peptide with two incorporated PEG4s chains. In contrast, we could not incorporate two PEG8s by using the same mRNA (entry 4 Figure 7).

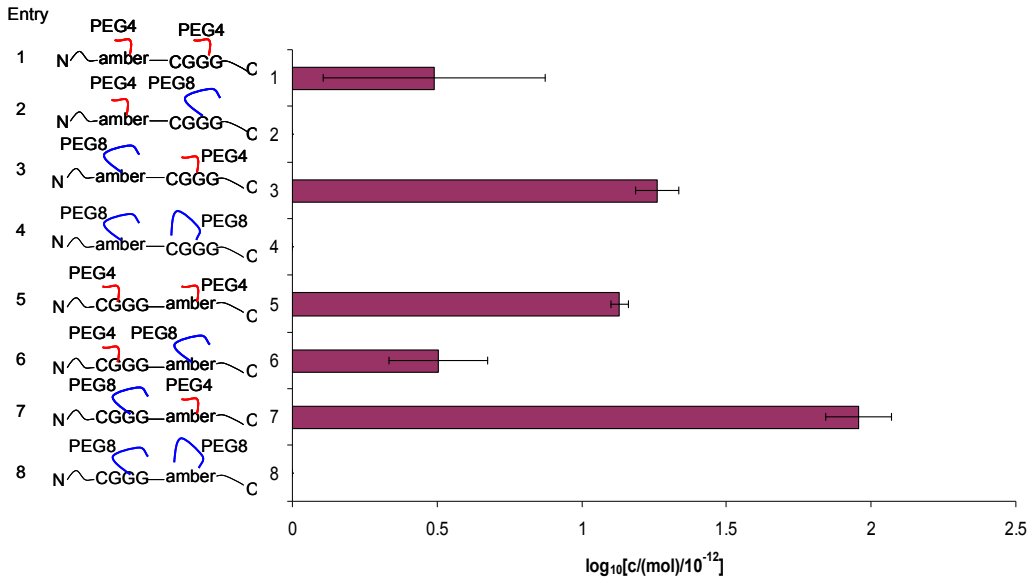


Fig. 7. PEGylation of different lengths in one peptide using two different codons. PEG4 and PEG8 were incorporated using amber or frame-shift codons. N and C indicate N- and C-terminal of peptide, respectively.

Through the usage of two different codons, it is quite possible to synthesize the peptides containing two PEGs of different sizes, each incorporated onto a different specific site using the amber or frameshift codon. We also found that a position-dependent size preference of PEGs used in the peptide synthesis by the ribosome. As shown in entry 1 of Figure 7, the addition of two PEG4 insertions using the same mRNA allows for the site-specific incorporation of the PEG8-amber and PEGs-frameshift codons are position 9 and 13 (entry 3), respectively, depending on the difference in anticodon of tRNAs carrying PEG of different lengths. Using the same mRNA in entry 2, we could not obtain a peptide with a PEG4 at position 9 and a PEG8 at position 13. The synthesis efficiency depends on PEG size (entry 3 > entry 1 > entry 2) and is also seen in the other mRNA (entry 7 > entry 5 > entry 6), which has the frameshift codon at position 9 and the

amber codon at position 13. Similarly, both mRNAs could not incorporate two PEG8s (entries 4 and 8 in Figure 7).

A preference for PEG8 over PEG4 at position 9 in these double incorporation experiments is also found in the production of single site incorporation of PEG (Figure 6) with enough amount of tRNA for frameshift codon although the preference is less different than cases of two site incorporation (Figure 7) The difference in the translation efficiency of the two mRNAs is a function of their secondary structures. Even with the above PEG8 preference, two PEG8s within 4 residues could not be incorporated simultaneously. Considering the size of the ribosome tunnel from the peptidyl transferase center, the mechanism which decreased efficiency of PEG12- and PEG24-containing peptide synthesis (Figure 6) would inhibit this two PEG8s incorporation.

2.4.3 The truncated PEGylation with single codon

When we observed the incorporation of the PEGs, we also found a mass corresponding to the truncated peptides (lack of fMSKQIEVN or SKQIEVN) of which the translation is apparently terminated just before the incorporation of the PEG. The attempted incorporation of longer PEGs results in a higher yield of the truncated peptides (Figure 8). We consider that the longer PEGs might block the peptidyl transferase reaction in the ribosome. A similar phenomenon was also reported by Abe et al.,[25] when one large fluorescent non-natural amino acid was incorporated into proteins.

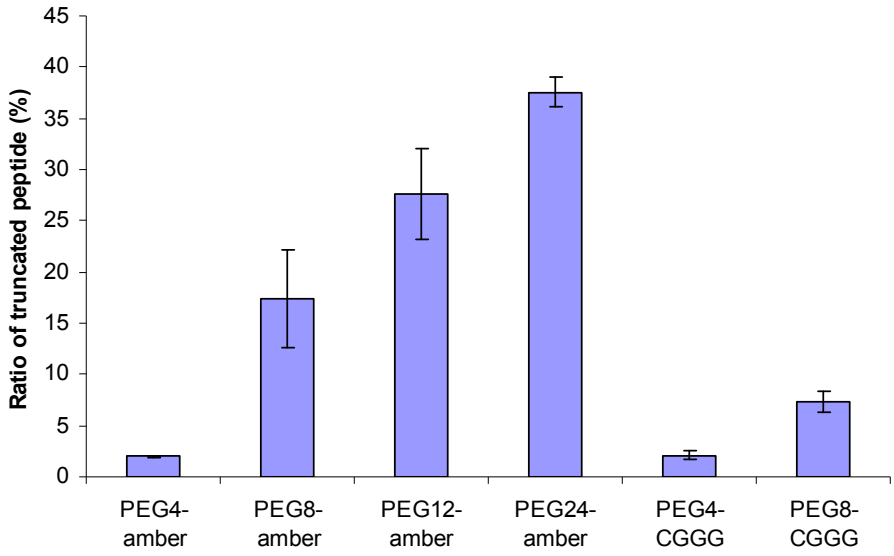


Fig. 8. The truncated PEGylation with single codon. The truncated peptides are XSNEDYKDDDDK, where X is PEG (4, 8, 12, and 24). The ratio of truncated peptides is the percentage in all peptides containing PEG.

2.5 Conclusions

In conclusion, we have demonstrated the site-specific incorporation of one and two different PEGs into one peptide. For the single incorporation, the PEGylation product decreased as the molecular weight increased. And the content of the truncated PEGylation peptide increased as the molecular weight increased. For the incorporation of two PEGs with different lengths, the PEGylation efficiency depends on the length of the PEG, the codon, and the incorporation sites. This method could have the application in the precise synthesis of bioconjugate drugs.

Reference:

1. A. Sood and R. Panchagnula, Peroral route: an opportunity for protein and peptide drug delivery. *Chem. Rev*. 101, 3275-3304, 2001.
2. M. Werle and A. Bernkop-Schnuerch, Strategies to improve plasma half life time of peptide and protein drugs. *Amino acids*, 30, 351-367, 2006.
3. F. M. Veronese and G. Pasut, PEGylation, successful approach to drug delivery. *Drug Discov. Today*, 10, 1451-1458, 2005.
4. J. M. Harris and R. B. Chess, Effect of PEGylation on pharmaceuticals. *Nat. Rev. Drug Discovery*, 2, 214-221, 2003.
5. R. Hirschhorn, Adenosine deaminase deficiency. *Immunodefic. Rev.*, 2, 175-198, 1990.
6. W. J. Li, P. Zhan, E. De Clercq, H. X. Lou and X. Y. Liu, Current drug reserach on PEGylation with small molecular agents. *Prog. Polym. Sci.,* 38, 421-444, 2013.
7. A. Abuchowski, T. van Es, N. C. Palczuk and F. F. Davis, Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. *J. Biol. Chem*., 252, 3578-3581, 1977.
8. A. Abuchowski, J. R. McCoy, N. C. Palczuk, T. van Es and F. F. Davis, Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase. *J. Biol. Chem*., 252, 3582-3586, 1977.
9. F. M. Veronese, Peptide and protein PEGylation: a review of problems and solutions. *Biomaterials*, 22, 405-417, 2001.
10. F. Lecolley, L. Tao, G. Mantovani, I. Durkin, S. Lautru and D. M. Haddleton, A new approach to bioconjugates for proteins and peptides (pegylation) utilizing living radical polymerization. *Chem. Commun*., 2026-2027.
11. T. L. Schlick, Z. B. Ding, E. W. Kovacs and M. B. Francis, Dual-surface modification of the tobacco mosaic virus. *J. Am. Chem. Soc*., 127, 3718-3723, 2005.
12. Y. Marsac, J. Crammer, D. Olschewski, K. Alexandrov and C. F. W. Becker, Site-specific attachment of polyethylene glycol-like oligomers to proteins and peptides. *Bioconjugate Chem*., 17, 1492-1498, 2006.

13. J. Thom, D. Anderson, J. McGregor and G. Cotton, Recombinant protein hydrazides: application to site-specific protein PEGylation. *Bioconjugate Chem*., 22, 1017-1020, 2011.
14. S. Tada, T. Andou, T. Suzuki, N. Dohmae, E. Kobatake and Y. Ito, Genetic PEGylation. *PLoS ONE*, 7, e49235, 2012.
15. K. Lang and J. W. Chin, Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev*., 114, 4764-4806, 2014.
16. X. Y. Li, T. H. Li, J. S. Guo, Y. Wei, X. B. Jing, X. S. Chen and Y. B. Huang, PEGylation of bovine serum albumin using click chemistry for the application as drug carriers. *Biotechnol Prog*. 28, 856-861, 2012.
17. A. J. Linkand D. A. Tirrell, Cell surface labeling of Escherichia coli via copper (I)-catalyzed [3+2] cycloaddition. *J. Am. Chem. Soc*., 125, 11164-11165, 2003.
18. N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo and C. R. Bertozzi, A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol*., 1, 644-648, 2006.
19. R. Serwa, P. Majkut, B. Horstmann, J. M. Swiecicki, M. Gerrits, E. Krause and C. P. R. Hackenberger, Site-specific PEGylation of proteins by a Staudinger-phosphite reaction. *Chem. Sci*., 1, 596-602, 2010.
20. M .J. Roberts, M. D. Bentley and J. M. Harris, Chemistry for peptide and protein PEGylation. *Adv. Drug Deliver Rev.*, 64, 116-127, 2012.
21. O. Melnyk, J. A. Fehrentz, J. Martinez and H. Gras-Masse, Fictionalization of peptides and proteins by aldehyde or keto groups. *Biopolymers*, 55, 165, 186, 2000.
22. A. Dumas, C. D. Specer, Z. H. Gao, T. Takehane, Y. A. Lin, T. Yasukohchi and B. G. Davis, Self-liganded Suzuki-Miyaura coupling for site-selective protein PEGylation. *Angew. Chem. Int. Ed*., 52, 3916-3921, 2013.
23. D. Kajihara, R. Abe, I. Iijima, C. Komiyama, M. Sisido and T. Honsaka, FRET analysis of protein conformational change through position-specific incorporation of fluorescent amino acids. *Nat. Methods,* 2006, 3, 923-929.
24. S. X. Chen, N. E. Fahmi, L. Wang, C. Bhattacharya, S. J. Benkovic and S. M. Hecht, Detection of dihydrofolate reductase conformational change by FRET using two fluorescent amino acids. *J. Am. Chem. Soc*., 2013, 135, 12924-12927.

25. R. Abe, K. Shiraga, S. Ebisu, H. Takagi and T. Hohsaka, Incorporation of fluorescent non-natural amino acids into N-terminal tag of proteins in cell-free translation and its dependence on position and neighboring codons. *J. Biosci. Bioeng*., 2010, 110, 32-38.
26. K. Josephson, M. C. T. Hartman and J. W. Szostak, Ribosomal synthesis of unnatural peptides. *J. Am. Chem. Soc.,* 2005, 127, 11727-11735.
27. P. Hirel, J. M. Schmitter, P. Dessen, G. Fayat and S. Blanquet, Extent of N-terminal methione excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. USA*, 1989, 86, 8247-8251.
28. J. Mittelstaet, A. L. Konevega and M. V. Rodnina, A kinetic safety gate controlling the delivery of unnatural amino acids to the ribosome. *J. Am. Chem. Soc*., 2013, 135, 17031-17038.

CHAPTER 3

IN VITRO SELECTION OF A PHOTO-RESPONSIVE PEPTIDE APTAMER TO GLUTATHIONE-IMMOBILIZED MICROBEADS

3.1 Introduction

The term “aptamer” is from the Latin word “aptus”, means “to fit” and the Greek word “meros”, means “part”. Aptamers are a special class of oligonucleotide acids or peptides isolated from a large random sequence library, which can bind to their targets with high specificity and high binding affinity like antibodies. However, unlike antibodies, aptamers are usually created *in vitro* and are with a number of different formats: nucleic acid (DNA/RNA/XNA) or peptide. Additionally, aptamers are artificial recognition molecules, but the natural aptamers also exist in riboswitches.[1]

The selection procedure was described for the first time independently by Tuerk and Gold,[2] Ellington and Szostak [3] in 1990 for identification of RNA based aptamers. Following the discovery of the RNA based aptamers, the DNA based aptamers were soon developed in 1994.[4] Soon after the functional RNA and DNA aptamers, the peptide aptamers were also developed.[5] The peptide aptamers are more like the antibodies and the dissociation constants are similar to, or sometimes better than antibodies.

A lot of aptamers have been selected against to numerous targets, including the small metal ions,[6] small molecules,[7-8] peptides,[9] proteins,[10] the whole cells[11] and the tissues.[12]

Systematic evolution of ligands by exponential enrichment (SELEX) or *in vitro* selection is one powerful tool for isolating the aptamers from a large random sequence pool. In general, the principal of SELEX is quite similar to the famous Darwinian Theory (Figure 1). First, one random sequence pool containing single strand oligonucleotide is constructed by solid phase synthesis (library generation). Then it is used for affinity selection against to the target of interest (binding and separation). The target-bound library

components are separated in each round of selection. Finally, target-bound library components are applied for the next round of selection (amplification).

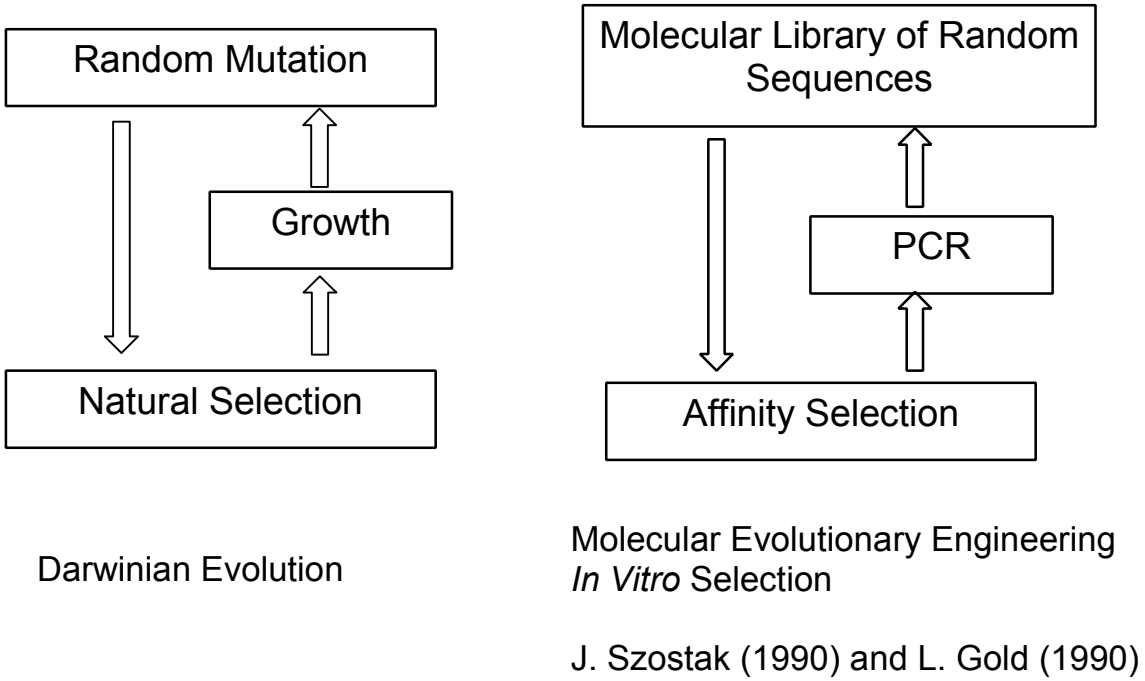


Fig. 1. Comparison of Darwinian evolution theory and *in vitro* selection.

In order to select the peptide aptamers from a random pool, several other methods have been coupled. Some classical methods are phage display,[13] bacterial surface display,[14] yeast display,[15,16] mRNA display,[17] cDNA display,[18] CIS display,[19] and ribosome display.[20] The usage of method depends on the type of aptamers. All the methods have the advantages and disadvantages. In case of the phage display, yeast display and bacterial surface display, the main problem is that the limitation of the library size in living cells. There is no such limitation for cell-free method such as mRNA display, DNA display and ribosome display. The common feature for any peptide-selection methods is the linkage of functional peptide sequence (phenotype) with its coding nucleic acid sequence (genotype). The

phenotype-genotype conjugation plays a key role for reading the genetic information of the corresponding peptide aptamer selected from the library. Among the *in vitro* methods, ribosome display is one very simple and effective method. Therefore, ribosome display was employed to select functional peptide aptamers in this study.

Peptide aptamers are most commonly used as discruptors of protein-protein interactions *in vivo*.[21] There are many fields in which aptamers can be used, can be as biomaterial,[22] a diagnostic[23,24] and therapeutic tool,[25] biosensing probe,[26] to develop new drugs[27] and delivery systems.[28] While in this research, one photoresponsive peptide aptamer (Figure 2) was selected from a large random sequence pool and it can be used to purify the other proteins through the structure changes of azobenzene under UV/visible light (Figure 3).

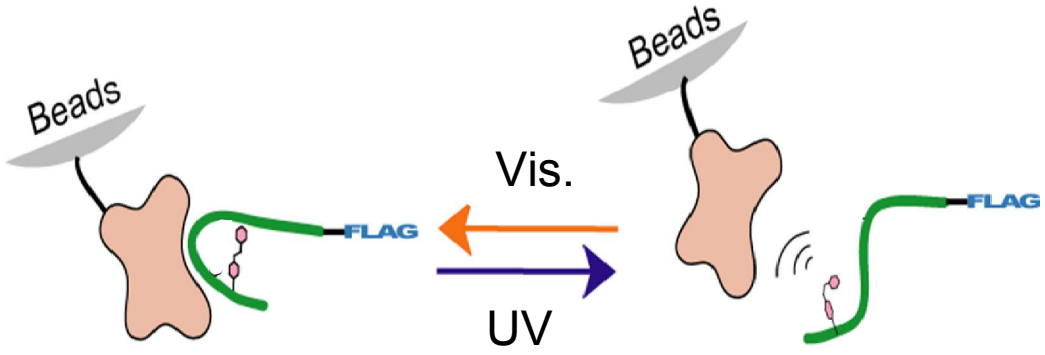


Fig. 2. The peptide aptamer can detach from the target under UV irradiation.

I chose glutathione (GSH) as the target for the selection of the photo-responsive peptide aptamer (Figure 4). GSH is a tripeptide with a gamma amide group and thiol group, which are comprised of gamma-glutamyl, cystein and glycine. GSH can be used for the purification of proteins[29] because GSH can bind to glutathione-S-transferase (GST) with high affinity.

However, this method has some limitations, such as low product purity, retention of GSH in the product, relatively high cost, as well as the risk of denaturation of the proteins.[30] To overcome these limitations, I selected one photo-responsive peptide aptamer instead of GST, which was expected to be an alternative method for proteins purification.

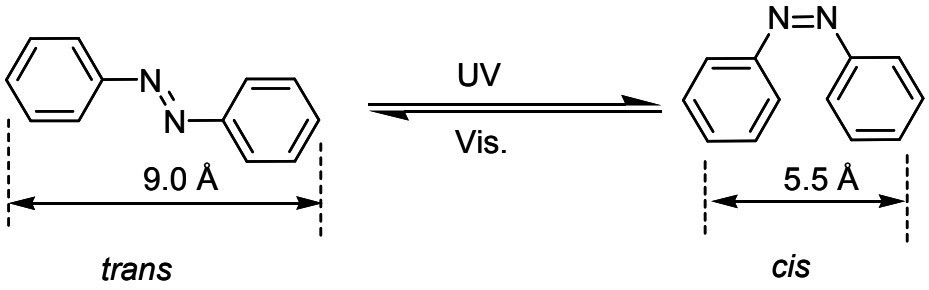


Fig. 3. The conformation changes of azobenzene under UV/Visible light.

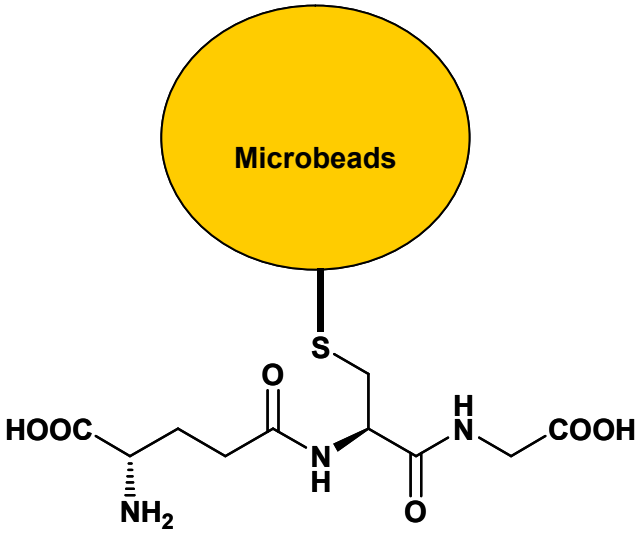


Fig. 4. Glutathione (GSH)-immobilized microbeads.

In order to select a photo-respnosive peptide aptamer, a new system was developed to introduce a non-natural photo-responsive amino acid into the library. One photo-responsive molecule, azobenzene (Azo), was employed,

which can transform under UV irradiation/visible light. Azo-linked aminophenylalanine was prepared and it was ligated with a tRNA carrying an amber codon. And one cell-free method, ribosome display, was used for the selection. In ribosome display, the translated peptide and its encoding mRNA remained linked to the ribosome because of the lack of stop codons in mRNA. This ternary complex of translated peptide, ribosome and mRNA can be used for the isolation of a functional peptide (Figure 5).

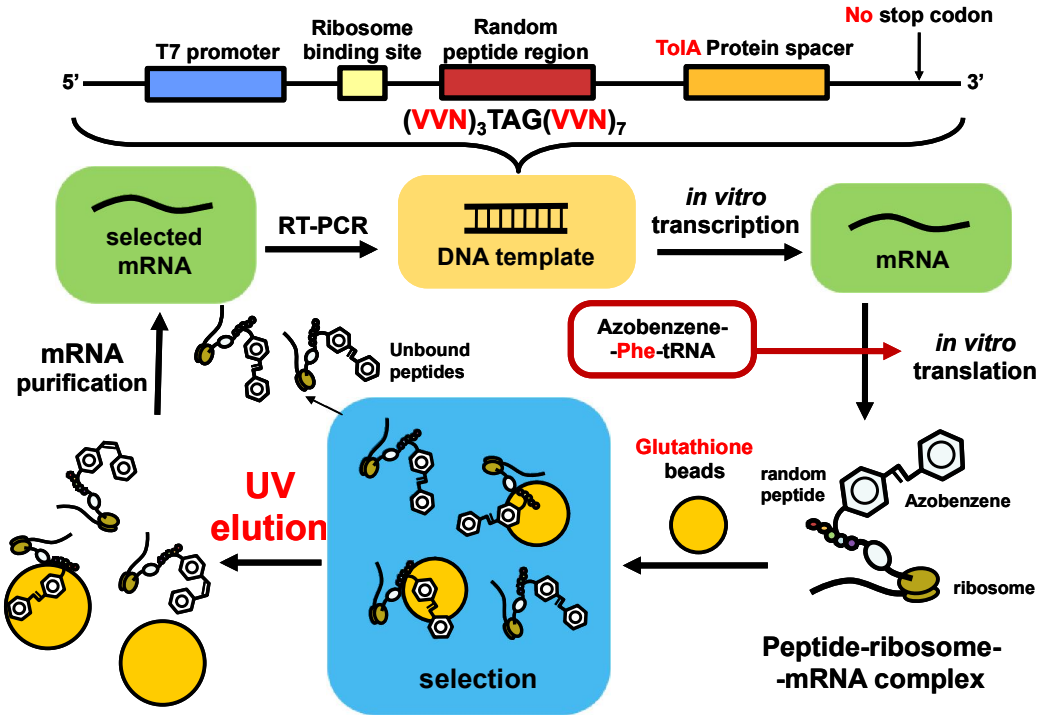


Fig. 5. The principle of *in vitro* selection using ribosome display and misacylated tRNA. A DNA library was transcribed into an mRNA library. The mRNA library and azo-aminoPhe-tRNA (tRNA carrying azobenzene-coupled aminophenylalanine) were then introduced into a cell-free translation system, resulting in the production of an azobenzene-modified peptide displayed on the surface of the ribosome. After affinity selection, the bound peptides were eluted by UV irradiation. The eluted complexes were collected and subsequently dissociated. The mRNA that encodes the peptide sequence was recovered and used for the next round of selection after RT-PCR.

Here one peptide aptamer carrying an azobenze reisude was selected with ribosome display method. The selected peptide aptamer exhibited high binding affinity to GSH-immobilized microbeads and it can detach from the microbeads under UV irradiation.

3.2 Materials and methods

3.2.1 Materials

Single strand DNA library (ssDNA) and primers were synthesized by Operon Co. Ltd (Tokyo, Japan).

Single strand DNA library (ssDNA):

5'-

ATCAGGCCAGCATGGCC$(VVN)_3$TAG$(VVN)_7$GGCCTGAGTGGCCAG AA-3', where V=G, C, or A, and N=G, C, T, or A.

FwSfiIUpstrm: 5'-GTTTAACTTTAAGAAGGAGATATCAGG-3'

ReSfiIDwstrm: 5'-CGCTGCCGCCTCTTCAGCTTGCTTCTG-3'

P3FwdT7USb: 5'-GAGTCAGTGAGCGAGGAAGC-3'

RevT7tmntDS: 5'-CCTCAAGACCCGTTTAGAGG-3'

FwT7-RTPCR2:

5'GAGGCCGGTAATACGACTCACTATAGGGAAATAATTTTGTTTAA CTTTAAGAAGGAGATATCAG-3'

PrimeSTAR®GXL DNA polymerase (Takara, Japan)

SfiI(New England Biolabs, USA)

DNA ligation Kit (Takara, Japan)

T4 RNA Ligase (Takara, Japan)

RNeasy kit (Qiagen, USA)

PURE SYSTEM Classic II kit (Wako, Japan)
Glutathione-immobilized magnetic beads (Thermo Fisher, USA)
UV-D36B filter (Asahi Technoglass, Japan)
RT-PCR (Toyobo, Japan)
QIAquick PCR purification kit (Qiagen, Germany)
Selection buffer, 0.1% (v/v) Tween 20, 60 Mm Tris-acetate, 180 mM NaCl, 60mM magnesium acetate
Washing buffer, 60 Mm Tris-acetate, 180 mM NaCl, 60mM magnesium acetate
TBS-T buffer, 20 mM Tris-acetate, 137 mM NaCl, 0.1% Tween 20, pH 7.6.

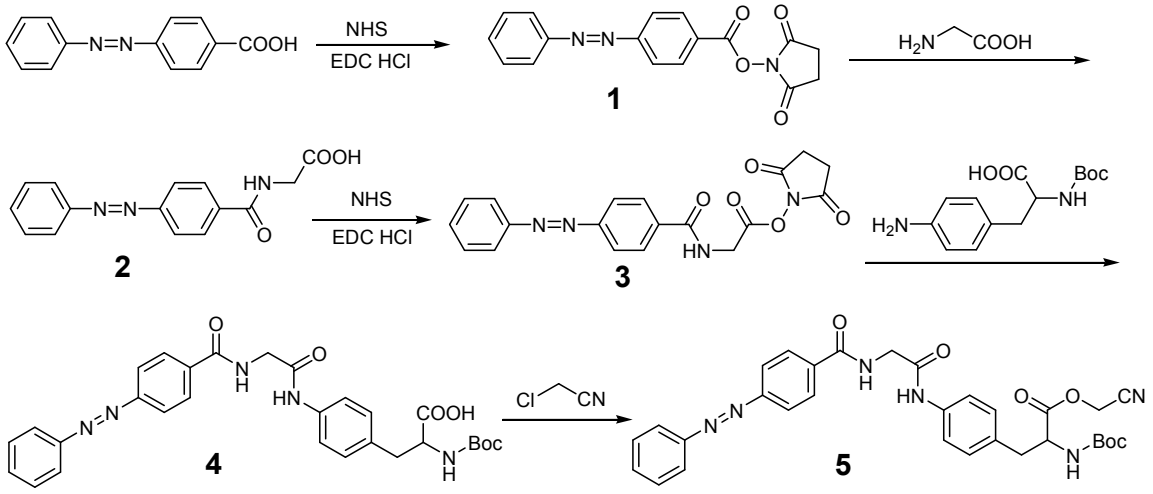
3.2.2 Preparation of aminoacyl tRNAs

tRNA carrying a non-natural amino acid coupled with an azobenzene residue was prepared from three starting materials: the truncated tRNA, pdCpA and N-Boc amino cyanomethyl ester. The truncated tRNA was prepared by transcription and the other two starting materials were chemically synthesized. And at last, all these three starting materials are combined to give the aminoacyl tRNAs.

3.2.2.1 Preparation of N-Boc amino cyanomethyl ester

2, 5-dioxopyrrolidin-1-yl 4-(phenyldiazenyl)benzoate (1) was prepared as following. 4-(phenyldiazenyl)benzoic acid (1.95g, 8.62mmol), NHS (1.19g, 10.34mmol) and EDC •HCl (1.98g, 1.34mmol) were stirred in DMF (30ml) at RT overnight. The solvent was removed under reduced pressure. The residue was purified by flash column chromatography with

neat dichloromethane to give Compound **1**. ^{1}H NMR (300 MHz, DMSO): 8.33 (d, *J*=8.4Hz, 2H), 8.10 (d, *J*=8.4Hz, 2H), 8.0-7.8 (m, 2H), 7.8-7.4 (m, 3H), 2.93(s, 4H).



Scheme 1. Synthesis of N-Boc amino cyanomethyl ester.

2-(4-(phenyldiazenyl)benzamido)acetic acid (2) Compound **1** (1.5g, 4.64mmol), glycine (0.42g, 5.57mmol) and Et3N (1.17g, 1.63ml) were stirred in a solution of DMF (15ml) and water (5ml). The reaction was carried out at RT overnight. The solvent was removed under reduced pressure. CH2Cl2 (10ml) was added to the residue and stirred at RT. Some precipitation occurred. Filtered and washed with CH2Cl2, red solid was obtained. ^{1}H NMR (300 MHz, DMSO): 8.80 (t, *J*=5.1Hz, 1H), 8.10 (d, *J*=8.1Hz, 2H), 8.04-7.88 (m, 4H), 7.72-7.58 (m, 3H), 3.91 (d, *J*= 5.4Hz, 2H).

2,5-dioxopyrrolidin-1-yl 2-(4-(phenyldiazenyl)benzamido)acetate (3) To a solution of **2** (283.0 mg, 1.0 mmol) in DMF 50 mL, NHS (126.5 mg, 1.1 mmol) and EDC •HCl (230.0 mg, 1.2 mmol) were added. The mixture

was stirred at RT for 4h and the solvent was evaporated under reduced pressure. The crude residue was washed with CH_2Cl_2 to give compound **3** as orange solid (311.6 mg, 0.81 mmol, yield 81.9%). ^{1}H NMR (400 MHz, DMSO) δ(ppm): 9.373 (t, *J*=5.2 Hz, 1H), 8.094 (d, *J*=7.2 Hz, 2H) 7.986(d, *J*=7.2 Hz, 2H) 7.941-7.923 (m, 2H), 7.627 (s, 1H), 7.613 (s, 2H), 4.471 (d, *J*=4.8 Hz, 2H). MALDI-MS: m/z, calculated 403.101 found $[M+Na]^+$ 403.110, calculated 419.210, found $[M+K]^+$ 419.086. ^{1}H NMR (300 MHz, DMSO): 9.38 (t, J=6.0Hz, 1H), 8.11 (d, *J*=8.4Hz, 2H), 8.00 (d, *J*=8.4Hz, 2H), 7.98-7.90 (m, 2H), 7.68-7.59 (m, 2H), 4.48 (d, *J*=5.4Hz, 2H), 2.83 (s, 4H).

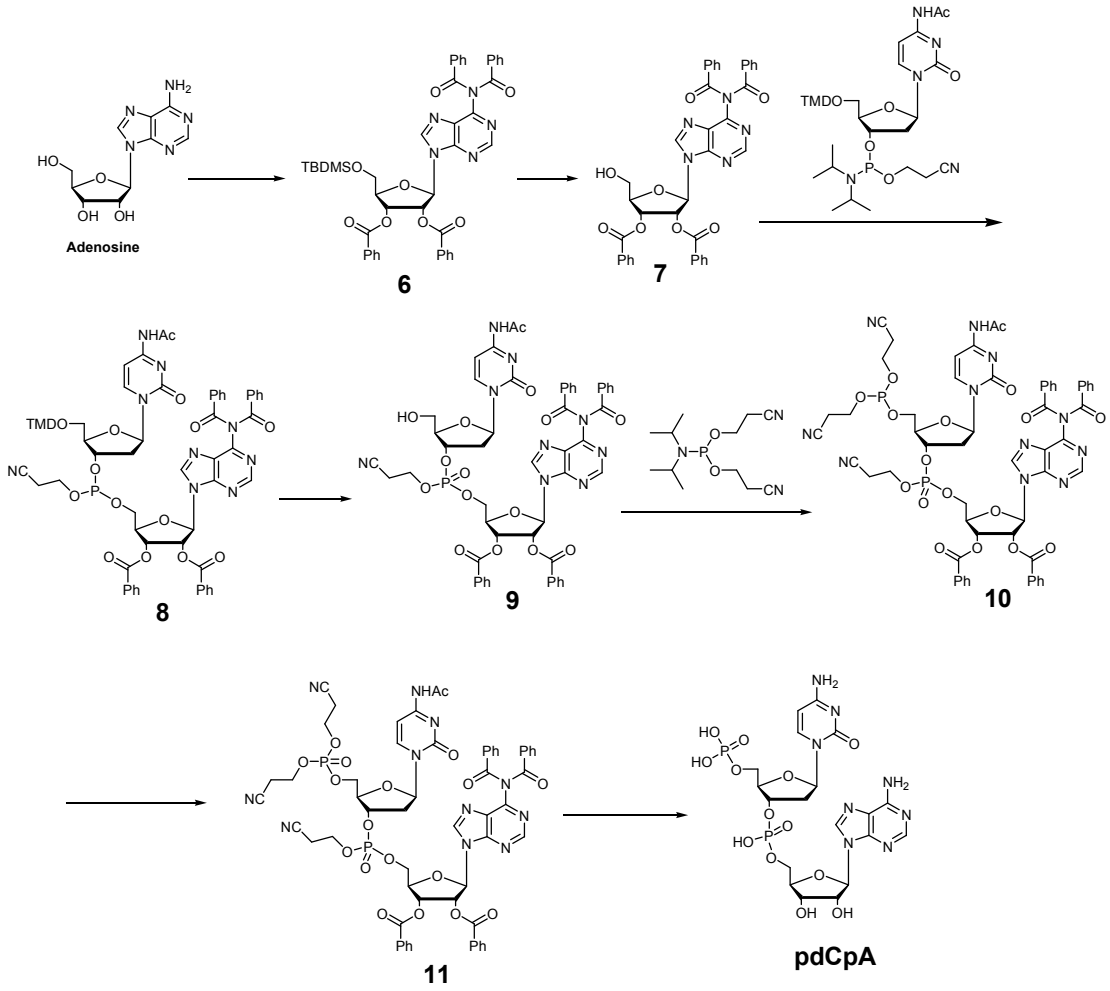
2-(tert-butoxycarbonylamino)-3-(4-(2-(4-(phenyldiazenyl)benzamido) acetamido)phenyl)propanoic acid (4) 2.5 M aqueous pyridine-HCl pH 5.0 (5mL) was added to a solution of **3** (330mg, 0.87mmol,) and Boc-4-amino-L-phenylal-anine (230mg, 0.59mmol,) in DMSO (15mL). The mixture was stirred at 37 ℃ in dark overnight. The solvent was evaporated under reduced pressure and the crude residue was purified by flash chromatography on silica gel (CH_2Cl_2:CH_3OH, 100:1 to 50:1, v/v) to afford compound **4** (248mg, 0.37 mmol, yield 65.0%) as orange solid. ^{1}H NMR (300 MHz, DMSO) δ(ppm): 10.05 (s, 1H), 9.03 (t, *J*=6.0Hz, 1H), 8.13 (d, *J*=9.0Hz, 2H), 8.03-7.88 (m, 2H), 7.99 (d, *J*=8.7Hz, 2H), 7.71-7.58 (m, 3H), 7.51 (d, *J*=8.7Hz, 2H), 7.18 (d, *J*=8.4Hz, 2H), 7.07 (d, *J*=8.1Hz, 1H), 4.16-3.95 (m, 1H), 4.09 (d, *J*=5.7Hz, 2H), 2.96 (dd, *J*=13.5Hz, 4.2Hz, 1H), 2.78 (t, *J*=10.8Hz, 1H), 1.33 (s, 9H).

N-Boc amino cyanomethyl ester (5) Compound 3 (120mg, 0.18mmol), Et3N (0.13ml, 0.90mmol) and chloroacetonitrile (115.5mg, 1.53mmol) were stirred in anhydrous DMF (8ml) at RT in dark overnight. The solvent was

evaporated under reduced pressure and the crude residue was purified by flash column chromatography (CH_2Cl_2:CH_3OH, 100:1 to 100:5, v/v) to give Compound 4 (120mg, yield: 93%). ^{1}H NMR (400 MHz, DMSO): δ =10.06 (s, 1H), 9.04 (s, 1H), 8.12 (d, *J*=8.8Hz, 2H), 7.99 (d, *J*=8.8Hz, 2H), 7.97-7.91 (m,2H), 7.68-7.50 (m, 3H), 7.53 (d, *J*=8.4Hz, 2H),7.46 (d, *J*=8.0Hz, 1H) 7.19 (d, *J*=8.4Hz, 2H), 5.00 (s,2H), 4.28-4.17 (m,1H), 4.10 (d, *J*=6.0Hz, 2H), 2.97 (dd, *J*=14Hz, 5.6Hz), 2.86 (dd, *J*=13.2Hz, 9.6Hz), 1.34 (m,9H).

3.2.2.2 Synthesis of pdCpA

(2R,3S,4S,5R)-2-(6-(N-benzoylbenzamido)-9H-purin-9-yl)-5-((tert-butyldimethylsilyloxy)methyl)-tetrahydrofuran-3,4-diyl dibenzoate (6) To a solution of **6** (100mg, 285mmol) in THF at RT was added pyridine (2mL) and TBDMSCl (52mg, 0.342mmol). After stirring for 2h at RT under N2 atmosphere, benzyl chloride (200mg, 1.425mmol) was added. The resulting mixture was stirred at RT for 3h. Then the solvent was removed under vacuum and the residue was diluted with ethyl acetate (50ml), washed with diluted HCl (0.1M), saturated $NaHCO_3$ and brine, dried, filtered and concentrated. The residue was purified by flash column chromatography using ethyl acetate/ hexane (10:1 to 4:1) to yield compound **7** as a white solid. ^{1}H NMR (300 MHz, DMSO): 8.79 (s, 1H), 8.70(s, 1H), 7.94 (d, *J*=8.1Hz, 2H), 7.88-7.73 (m, 6H), 7.71-7.54 (m, 4H), 7.54-7.36 (m, 8H), 6.65 (d, *J*=5.1Hz, 1H), 6.37 (t, *J*=5.4Hz, 1H), 6.02 (t, *J*=5.4Hz, 1H), 4.68-4.52 (m, 1H), 4.10-3.90 (m, 2H), 0.80 (s, 9H), 0.00 (s, 3H), -0.03 (s, 3H).



Scheme 2. Synthesis of pdCpA.

(2R,3S,4S,5R)-2-(6-(N-benzoylbenzamido)-9H-purin-9-yl)-5-(hydroxymethyl)-tetrahydrofuran-3,4-diyl dibenzoate (7) To a solution of **7** in THF was added a solution of TBAF (1.0M in THF, 2.85ml). After stirring at RT overnight, the solvent was removed under vacuum and the residue was washed with diluted HCl (0.1M) and brine, dried, filtered and concentrated. The residue was purified by flash column chromatography using ethyl acetate/hexane (10:1 to 4:1) to give a white solid.^{1}H NMR (300 MHz, DMSO): 8.94 (s, 1H), 8.74 (s, 1H), 8.04-7.95 (m, 2H), 7.88-7.76 (m, 6H), 7.74-7.54 (m, 4H), 7.54-7.36 (m, 8H), 6.67 (d, *J*=5.7Hz, 1H), 6.32 (t,

J=5.7Hz, 1H), 6.04-5.92 (m, 2H), 5.46 (t, *J*=6.0Hz, 1H), 4.64-4.53 (m, 1H), 3.96-3.70 (m, 2H).

(2R,3S,4S,5R)-2-((((2R,3R,5R)-5-(4-acetamido-2-oxopyrimidin-1(2*H*)-yl)-2-(hydroxymethyl)-tetrahydrofuran-3-yloxy)(2-cyano ethoxy)-phosphinooxy)methyl)-5-(6-(N-benzoylbenzamido)-9H-purin-9-yl)-tetrahydrofuran-3,4-diyl dibenzoate (9)To a solution of **8** (150mg, 0.236mmol) in dichloromethane (3ml) at RT was added 1*H*-tetrazole (19.8mg, 0.283mmol) and (2R,3R,5R)-5-(4-acetamido-2-oxopyrimidin-1(2H)-yl)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-tetrahydrofuran-3-yl 2-cyanoethyl diisopropylphosphoramidite (182mg, 0.236mmol). The mixture was stirred at RT for 4h, then a solution of iodine in THF-H_2O-pyridine (66:33:1) was added. After 15min, the mixture was poured into ethyl acetate (30ml), washed with 0.2M sodium bisulfate (10mlx2) and brine, dried, filtered and concentrated. To the residue was added a solution of 0.1% TFA in dichloromethane (25ml) at RT for 10min. The reaction mixture was quenched by adding saturated $NaHCO_3$. After separation, the organic phase was washed with brine, dried, filtered and evaporated. The residue purified by flash column chromatography using dichloromethane/methanol (20: 1 to 40:1). A white solid was obtained. ^{1}H NMR (300 MHz, DMSO): 8.94 (s, 1H), 8.74 (s, 1H), 7.99 (d, *J*=5.4Hz, 2H), 7.80 (d, *J*=5.4Hz, 6H), 7.73-7.54 (m, 4H), 7.54-7.38 (m, 8H), 6.67 (d, *J*=5.7Hz, 1H), 6.32 (t, *J*=5.7Hz, 1H), 6.04-5.93 (m, 1H), 5.46 (t, *J*=6.0Hz, 1H), 4.64-4.52 (m, 1H), 3.94-3.72 (m, 2H).

(2R, 3S, 4S, 5R)-2-((((2R,3R,5R)-5-(4-acetamido-2-oxopyrimidin-1(2H)-yl)-2-((bis(2-cyanoethoxy)phosphoryloxy)methyl)-tetra-hydrofuran-3-yloxy)(2-cyanoethoxy)phosphoryloxy)methyl)-5-(6-(N-

benzoylbenzamido)-9H-purin-9-yl)-tetrahydrofuran-3,4-diyl dibenzoate (11) To a solution of **10** (330mg, 0.309mmol) in dichloromethane (2ml) at RT was added bis(2-cyanoethyl) diisopropylphosphoramidite (167.7mg, 0.618mmol) and 1H-tetrazole (43.3mg, 0.618mmol). After stirring at RT for 3h, a solution of iodine in THF-H_2O-pyridine was added to the reaction mixture. After stirring for 10min, the mixture was poured into ethyl acetate (30ml), washed with 0.2 M sodium bisulfate (10mlx2) and brine, dried, filtered and concentrated. The residue was purified by flash column chromatography using dichloromethane/methanol (100:1 to 100:5) to yield a white solid. ^{1}H NMR (300 MHz, DMSO): 10.89 (s, 1H), 8.83 (d, *J*=2.7Hz, 1H), 8.72 (d, *J*=3.6Hz, 1H), 8.12 (d, *J*=6.3Hz, 1H), 7.92 (d, *J*=7.2Hz, 2H), 7.87-7.76 (m, 6H), 7.74-7.54 (m, 4H), 7.44-7.38 (m, 8H), 7.36-7.14 (m, 9H), 7.06 (d, *J*=7.2Hz, 1H), 6.93-6.78 (m, 4H), 6.67 (d, *J*=4.5Hz, 1H), 6.45-6.36 (m, 1H), 6.16-6.07 (m, 1H), 6.07-5.97 (m, 1H), 4.95-4.76 (m, 1H), 4.70-4.59 (m, 1H), 4.32-4.12 (m, 1H), 4.12-4.04 (m, 1H), 3.96-3.78 (m, 1H), 3.70 (s, 3H), 3.69 (s, 3H), 2.75-2.64 (m, 2H), 2.38-2.09 (m, 1H).

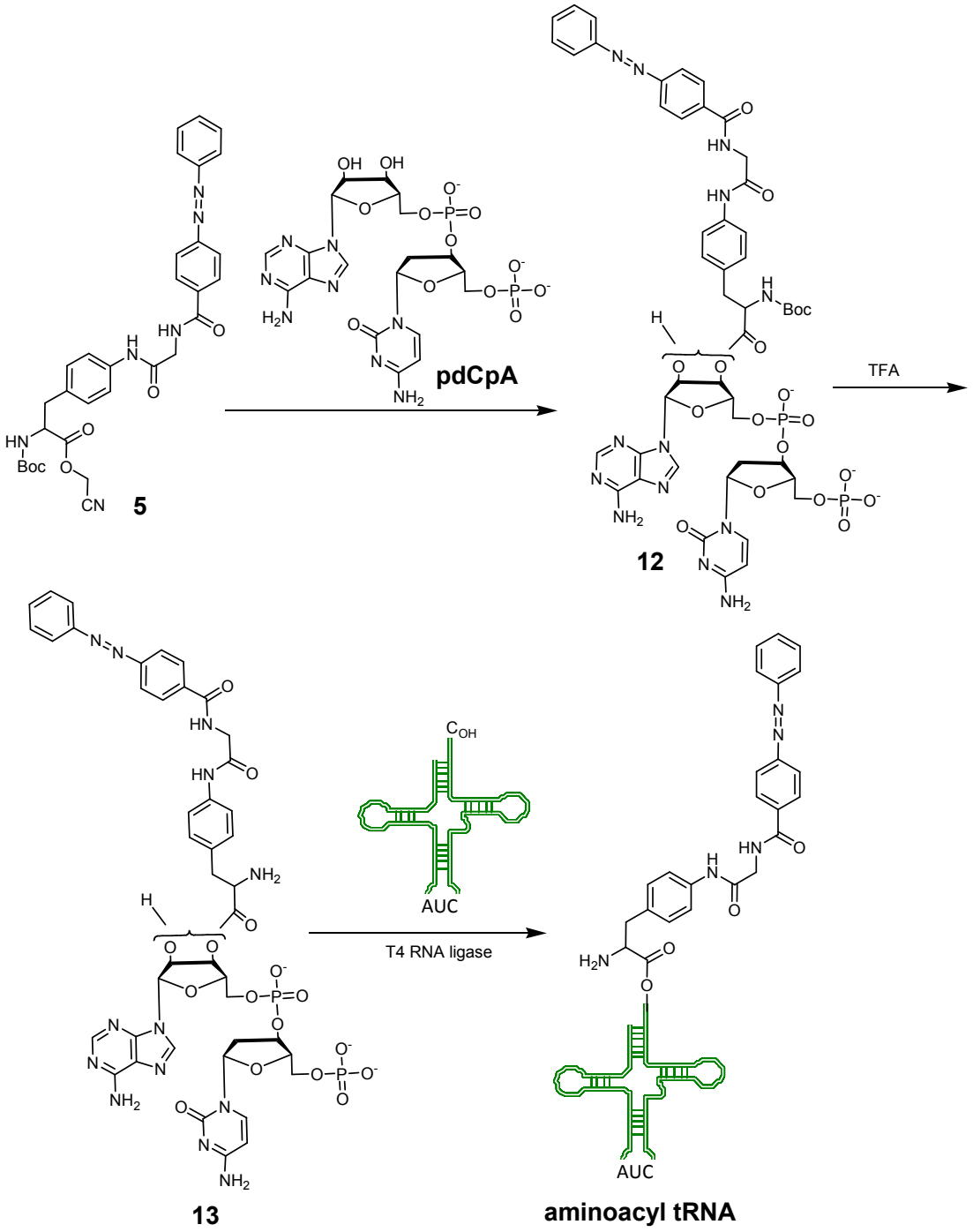
((2R,3R,5R)-5-(4-amino-2-oxopyrimidin-1(2H)-yl)-3-((((2R,3R,4S,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxy-tetrahydrofuran-2-yl) methoxy)(hydroxy)phosphoryloxy)-tetrahydrofuran-2-yl)methyl dihydrogen phosphate (pdCpA) To a solution of 10 (300mg, 0.239mmol) in dioxane (0.45ml) and methanol (3ml) was added concentrated ammonium hydroxide (3.45ml). The flask was sealed and stirred at RT for 24h. The solution was concentrated under reduced pressure and further lyophilized. The crude product was purified by reverse silica gel column chromatography using H_2O/methanol (100:1 to 100:10). After lyophilized, pdCpA (85mg, yield: 66.3%) was obtained as a white solid. ^{1}H NMR (300 MHz, DMSO): 8.45 (s, 1H), 8.29 (s, 1H), 7.96 (d, *J*=8.1Hz, 1H), 6.09 (d, *J*=8.1Hz, 1H),

6.10-5.93 (m, 2H), 4.40 (s, 1H), 4.34-4.12 (m, 2H), 410-3.92 (m, 2H), 3.89 (s, 2H), 2.56-2.17 (m, 1H), 2.10-1.96 (m, 1H).

3.2.2.3 Preparation of aminoacyl tRNAs

pdCpA conjugated with (Boc)-aminophenylalanine-azobenzene (12) The TBA salt of pdCpA was prepared using Dowex 50Wx8 20-50 mesh in its TBA form. The ion exchange beads (about 5g) were washed with milliQ, and stirred in 20% aqueous TBAOH (25ml) for 2h. Then the beads were packed into a column, and washed with water until the pH value of the wash was neutral. An aqueous pdCpA-ammonium solution was passed over the column with water as the eluant. Fractions containing the TBA-pdCpA salt were determined by TLC and UV. Lyophilization yielded the TBA form slat of pdCpA as a white solid. The product was dissolved in anhydrous DMF and stored in a sealed vial under argon at -20℃.

The tetrabutylammonium salt of pdCpA (10mg) was added to a solution of N-Boc amino cyanomethyl ester (5) in anhydrous DMF in the presence of a catalytic amount of TBA-acetate. After 2h of sonication at RT, the reaction mixture was purified by C18 reversed phase column (250mm x 10mm) using a gradient of 0 → 65% acetonitrile in 50mM NH4OAc, pH 4.5, over a period of 40min at a flow rate of 1.0 mL/min (monitoring at 260 nm). The fractions eluting at 22.742 min and 23.269 min were collected (Figure 6), combined and lyophilized to give a colorless solid. Mass spectrum (MALDI): [M+H]+: Calculated: 1164.334, Found: 1164.271.



Scheme 3. The ligation of truncated tRNA with pdCpA carrying an azobenzene residue.

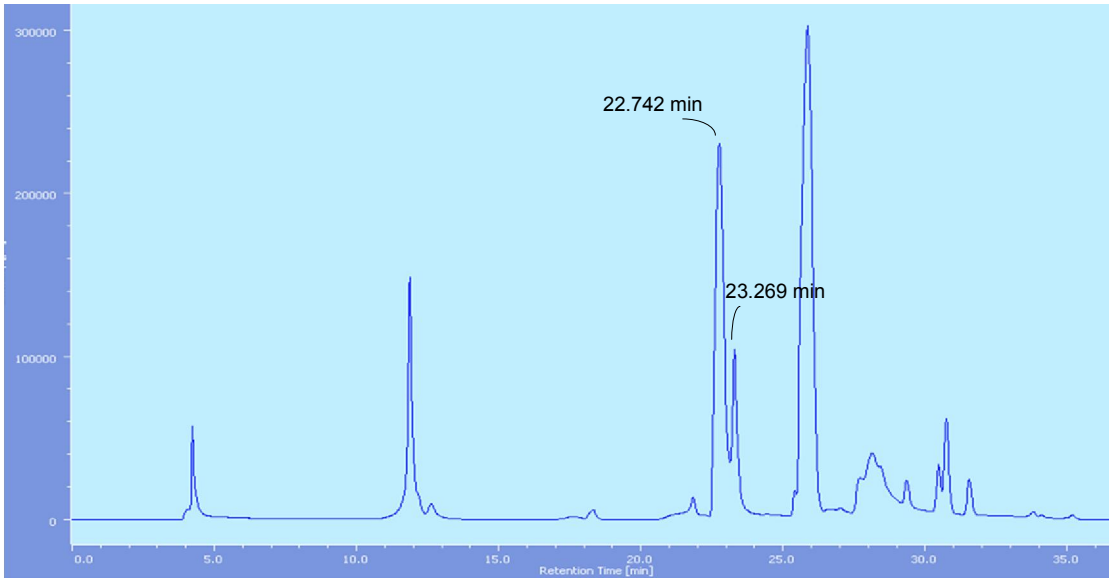


Fig. 6. HPLC of Compound **12**.

pdCpA conjugated with aminophenylalanine-azobenzene (pdCpA-AF-Azo) (13) The compound **12** was dissolved in 30μl of prechilled trifluoroacetic acid on ice. The mixture was placed on ice for 15min, and the trifluoroacetic acid was removed by reduced pressure. Mass spectrum (MALDI): $[M+H]^+$: Calculated: 1064.281, Found: 1064.140.

Aminoacyl tRNA carrying an azobenzene residue The suppressor $tRNA_{CUA}$ aminoacylation was carried out in 10 μl (total volume) of tRNA (3.2 μg), T4 RNA ligase (1 μl, 40 unit), 15% DMSO, 0.1% BSA (0.6 μl), pdCPA-AF-Azo and 10x T4 RNA ligase buffer. After incubation at 10℃ for 16h, the reaction mixture was quenched by the addition of 1μl of 3M NaOAc, pH 5.2, followed by 30 μl of ethanol. The reaction miture was incubated at -20℃ for 30min, and then centrifuged at 15,000 x g at 4℃ for 30 min. The supernatant was carefully decanted and the tRNA pellet was washed with 10 μl of 70% ethanol, and dissolved in 20 μl of RNase free H_2O. The product was analyzed by HPLC on a reversed phase column (Applied

Biosystems Poros R2/10: 4.6mm X 100mm, 10 μm) over a period of 30min at a flow rate of 1.0 mL/min (monitoring at 260 nm), with a linear gradient of 0%→40% acetonitrile in 0.1M TEAA buffer, pH=7.0. Then it was used for the in vitro selection after lyophilisation.

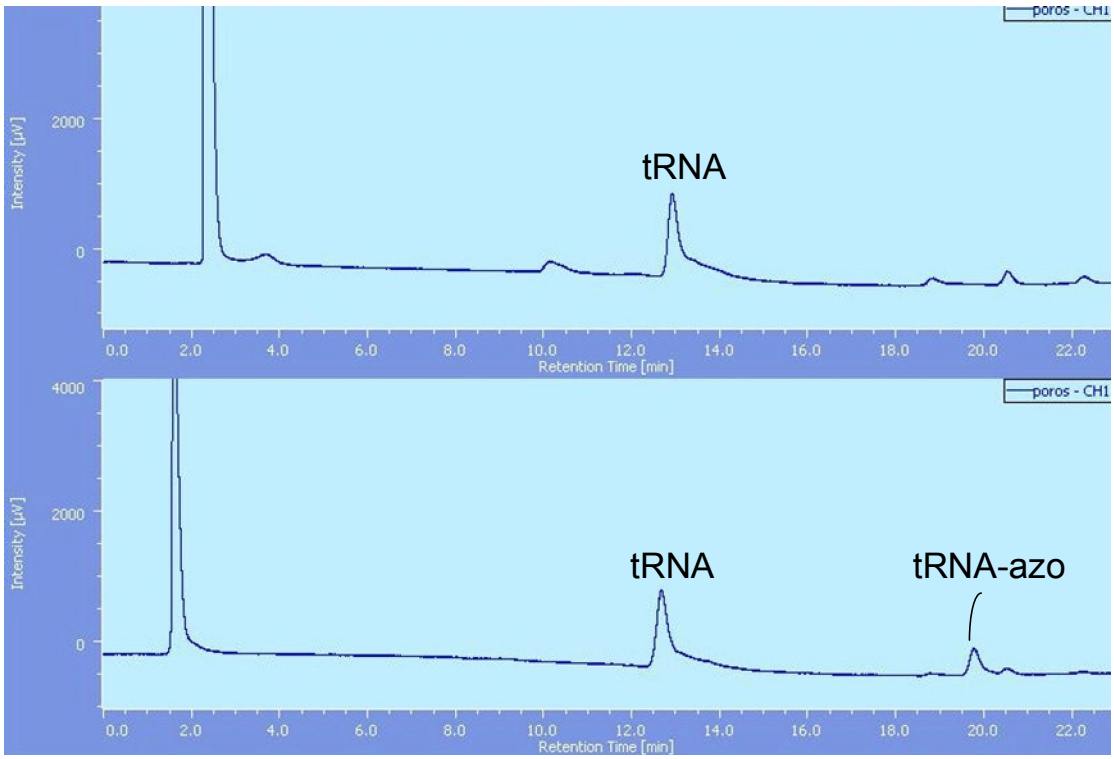


Fig. 7. HPLC of aminoacyl tRNA.

3.2.3 Preparation of random sequence library

Random double-stranded DNA (dsDNA) as a peptide library region was prepared. The sequence comprised T7 promoter; SD, a Shine-Dalgarno E.coli ribosome-binding site sequence; random sequence region and the TolA protein spacer.

For increasing the solubility of the peptides in aqueous buffer, we used VVN library sequences obtained from Operon Co.Ltd. The sequence was:5'-

ATCAGGCCAGCATGGCC(VVN)3TAG(VVN)7GGCCTGAGTGGCCAG AA-3', where V represents G, C or A and N represents G, C, T or A. all of them were synthesized by using solid-phase method and the codon usage for V and N were used in equivalent amounts. Because VVN codons do not code for most hydrophobic amino acids, such as leucine, valine, and tryptophan, this library provides water-soluble peptide sequences. We prepared dsDNA by one cycle of the polymerase chain reaction (PCR) using GXL PCR kit and primers FwSfiIUpstrm: 5'-GTTTAACTTTAAGAAGGAGATATCAGG-3' ReSfiIDwstrm: 5'-CGCTGCCGCCTCTTCAGCTTGCTTCTG-3'. The resulting dsDNA and pTolA3 plasmid were digested with a restriction enzyme and fused. Finally, a double-stranded DNA library was prepared by PCR with new primers (P3FwdT7USb: 5'-GAGTCAGTGAGCGAGGAAGC-3' and RevT7tmntDS: 5'-CCTCAAGACCCGTTTAGAGG-3'.

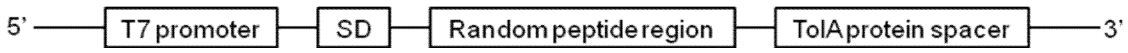


Fig. 8. Structure of DNA template.

3.2.4 *In vitro* transcription and translation

The prepared DNA was used for *in vitro* transcription and translation. The transcription was performed at 37℃ for 3h using RiboMAX large scale RNA production system. The mixture was then treated with DNase at 37℃ for 30 min. The mRNA was purified using the RNeasy kit. And *In vitro* translation was performed using a PURESYSTEM Classic II kit. The mixture was incubated 37℃ for 3h in the presence of tRNA carrying an

azobenzene residue and then the reaction was stopped by placing the mixture on ice for 10 min.

3.2.5 *In vitro* selection and recovery of mRNA

The translated peptide solution was incubated with glutathione-immoblized magnetic beads in the selection buffer (0.1%(v/v) Tween 20, 60 mM Tris-acetate, 180 mM NaCl, 60mM magnesium acetate) at 4℃ for 1h. The beads were collected by magnetic sedimentation and washed free of the unbound complex using the washing buffer (60 mM Tris-acetate, 180 mM NaCl, 60 mM magnesium acetate) at 4℃. The bound complexes to the microbeads were eluted by adding a 20mM glutathione solution at the first to third rounds of selection, or irradiating with UV light (6.0 mW/cm^2) for 2min using a UV spot light instrument through a UV-D36B filter at the fourth to eighth rounds of selection. In the first three rounds of selection, we aimed to just amplifying high-affinity peptide sequences, where in the latter five rounds of selection we collected the photo-responsive peptide sequences. After complex elution, we recovered the mRNA from the eluted complexes by incubation with EDTA at room temperature for 15min. The isolated mRNA was purified using the RNeasy kit.

3.2.6 RT-PCR of isolated mRNA

The purified mRNA was amplified by reverse transcription PCR (RT-PCR) using a one-step RT-PCR kit and primers FwT7-RTPCR2 and RevT7tmntDS, according to the manufacturer's instructions. The DNA product was purified with QIAquick PCR purification kits (Qiagen). The

concentration and quality of the DNA was verified by UV absorbance and electrophoresis using 2% agarose gel. The isolated DNA was used as the template for the next round of selection.

3.2.7 Cloning and sequencing

After eight rounds of selection, the library sequences of the selected DNA were analyzed. The selected library sequences were amplified again by PCR using the GXL PCR kit with primers FwT7-RTPCR2 and ReSfiIDwstrm, and purified with the QIAquick PCR purification kit. The obtained dsDNA product was inserted into the PCR-Blunt II-TOPO vector using the Zero Blunt TOPO Cloning Kit (Thermo Fisher Scientific Inc., Rockford, IL, USA), according to the manufacturers protocol. The ligated product was transformed into E.coli, and the transformed individual clones were sequenced at RIKEN BSI Research Source Center.

3.2.8 Binding assay

The binding assay of the selected peptides to glutathione-immobilized beads was analyzed as follows. The selected peptides labelled with fluorescein were incubated with glutathione-immobilized beads at 20 ℃ for 1h in TBS-T buffer (20 mM Tris-acetate, 137 mM NaCl, 0.1% Tween 20, pH 7.6). The peptide-binding microbeads were washed twice with washing buffer (60 mM Tris-acetate, 180 mM NaCl, 60 mM magnesium acetate) and the fluorescence intensity of the beads was quantified with a microplate reader (Mithras LB940; Berthold Technologies Ltd., Bad Wildbad, Germany). Estimation of the photo-responsiveness of the selected peptides

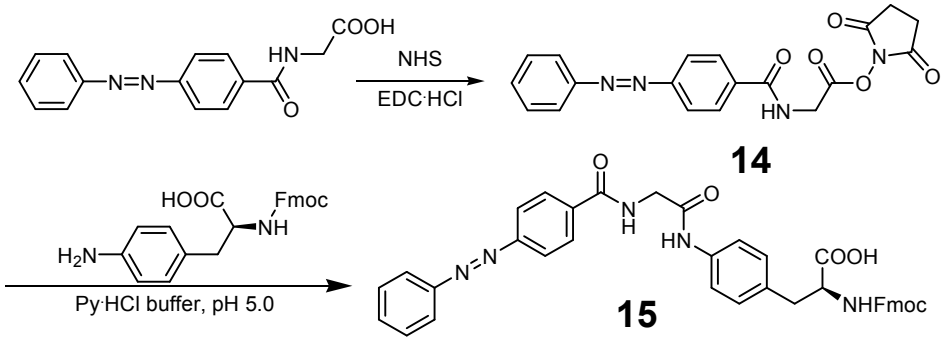
were conducted by irradiating at 6.0 mW/cm^2 UV light for 2min after the incubation of the peptides with glutathione-immobilized beads using the UV spot light instrument (Hamamatsu Photonics).

3.2.9 Spectropics measurements

UV-visible spectra of B09 and B69 peptide solutions were measured with a V-550 spectrophotometer (Jasco, Tokyo, Japan). The peptide solution was irradiated with UV light at 334 nm with the light source REX-250 (Asahi Spectra Co. Ltd., Tokyo, Japan).

3.2.10 Chemical synthesis of the selected peptides

Fmoc-protected azobenzene-aminophenylalanine was synthesized as shown in Scheme 4 and used for the solid-phase peptide synthesis.



Scheme 4. Synthesis of Fmoc-protected azobenzene-aminophenylalanine for the solid phase peptide synthesis.

2, 5-dioxopyrrolidin-1-yl 2-(4-(phenyldiazenyl)benzamido)acetate (14) The starting material 2-(4-(phenyldiazenyl)benzamido)acetic acid was synthesized according the reported method. 2-(4-(phenyldiazenyl)-

benzamido)acetic acid (0.5g, 1.76mmol) was dissolved in DMF (5ml). NHS and EDC·HClwere added to the mixture. The reaction was carried out at RT in dark for about 2hrs. The solvent was removed under reduced pressure. CH_2Cl_2 (5ml) was added to the residue and stirred at RT. The product was filtered and washed with CH_2Cl_2. Compound **14** was obtained as yellow solid (0.4g, yield 59.7%). The product was directly used for the next step without any purification.

Fmoc-phenlyalanine-glycine-azobenzene (**15**) To a solution of **14** (330mg, 0.87mmol), Fmoc-4-amino-L-phenylalanine (230mg, 0.59mmol,) in DMSO (15mL), 2.5 M aqueous pyridine-HCl pH=5.0 (5mL) was added slowly. The mixture was stirred at 37 ℃ in dark overnight. The solvent was evaporated under reduce pressure and the crude residues was purified by flash chromatography on silica gel (CH_2Cl_2: CH_3OH, 100:1 to 50:1, v/v) to afford compound 2 (248mg, 0.37 mmol, yield 65.0%) as orange solid. 1H NMR (300 MHz, DMSO) δ(ppm): 10.06 (1H, s), 9.05 (t, *J*= 5.4 Hz, 1H), 8.13 (d, *J*=7.8 Hz, 2H), 8.04-7.93 (m, 4H), 7.88 (d, *J*= 4.2 Hz, 2H), 7.78-7.58 (m, 5H), 7.54 (d, *J*=7.8 Hz, 2H), 7.41 (t, *J*=7.2 Hz, 2H), 7.30 (q, *J*=7.2 Hz, 2H), 7.22 (d, *J*=7.8 Hz, 2H), 4.33-4.04 (m, 6H), 3.04 (dd, J_1=7.5 Hz, J_2=14.1 Hz, 1H), 2.83 (dd, J_1=3.3 Hz, J_2=14.4Hz, 1H). ^{13}C NMR (75MHz, DMSO) δ(ppm): 173.40, 167.55, 165.86, 155.95, 153.39, 151.93, 143.77, 143.75, 140.68, 137.38, 136.13, 132.76, 132.16, 129.60, 129.42, 128.72, 127.65, 127.08, 125.31, 122.78, 122.45, 120.12, 119.00, 65.63, 55.58, 46.55, 43.34, 35.91. MALDI-MS: m/z, calculated for $C_{39}H_{33}N_5O_6$ $[M+Na]^+$ 690.232, found 690.261.

Selected peptides, B09 and B69, were synthesized by a standard Fmoc solid-phase method at RIKEN Brain Science Institute. Fmoc-aminophenylalanyl azobenzene was synthesized for the incorporation of the azobenzene into the selected peptide sequences. The fluorescein-N-succinimidyl ester (Thermo Fisher Scientific) was used for fluorescent labeling at the N-terminus of the peptides. For the cleavage of these peptides from peptide synthesis resin, a △TIS cleavage cocktail consisting of trifluoroacetic acid: phenol: thioanisole: ddH_2O: ethane dithiol in the ratio 85:5:4:4:2 (v/v) was used to avoid reduction of the double bond in the azobenzene moiety. Synthesized peptides were purified by HPLC and their formulation was analyzed by MALDI-TOF-MS. Peptide B09: calculated 3380.3 for $[M+H]^+$, found 3381.0, peptide B69: calculated 3274.2 for $[M+H]^+$, found 3275.0).

3.3 Results

3.3.1 Construction of the DNA library for in vitro selection

In this study, I employed the ribosome display method to get the peptides binding to GSH-immobilized microbeads. A new random sequence library was constructed. Each sequence contained the stop codon (the amber codon) for increasing the diversity of the peptide sequences. And there is no stop codon at the terminus of each sequence, so the polypeptide, the ribosome and mRNA can form a complex.

3.3.2 Cloning and sequencing

After eight rounds of the selection process, the selected sequences were cloned and the isolated sequences were analyzed. The selected library sequences were inserted into the pCR-Blunt II-Topo vector using the Zero Blunt TOPO Cloning Kit according to the manufacturer's protocol. The ligated product was transformed into Escherichia coli, and the transformed individual clones were sequenced at the RIKEN BSI Research Source Center. The sequences are listed in Table S1. 215 types of sequences were found from 269 clones. 12 sequences showed multiple repetitions in the analyzed sequences. Two peptides derived from plural clones were chosen, RNGXSSGRHGD (B09) and KDGXGGEEGET (B69), where X corresponds to azobenezene-aminophenylalanine, and these peptides were chemically synthesized by a solid-phase method. They appeared more times than the other selected peptides in the sequence analysis. Further study focuses on these two peptides (B09 and B69).

Peptide #	Sequence	Peptide #	Sequence	Peptide #	Sequence
01	RRPXPPSGAGR	A49	ANPXAARRGTG	B95	SAAXTTKSTAG
06	GPEXLGSSSAR	A51	GHNXPDDQRAQ	C02	RHEXNKGPPGR
09	RNGXSSGRHGD	A52	GNXXERGHAAR	C04	QPRXRACTAHD
11	DGEXHGPSHAE	A54	GARXKDDRKGG	C06	RGSXTRGAADG
13	GRGXQQKGSTG	A57	ANPXAARRGTG	C07	NDRXKPAAGRH
16	RARXRAGRGKR	A58	TVAXQXGSRGH	C10	KKSXDRRAGEG
17	GEQXKDAAHKG	A62	STSXRRTEPCR	C12	GHHXSKTKKRP
18	GGTXPNRGDDR	A64	QRRXGRGPNSN	C15	TRSXTGEDGRG
21	GGRXQKADRPA	A65	TTEXEDLRDGP	C20	ARNXRDDKGQD
22	GKTXASGQKRA	A66	RQAXTGPNRER	C22	PSAXTRNTGRE
23	AGRXREGTDDE	A67	GARXGSARRPG	C25	SGTXARDGRGG
27	RGSXTRGAADG	A68	ASGXGTDGSAQ	C27	PTSXSAEKGAE
28	ATEXTGTHGRR	A69	QRGXHGGTQPA	C29	GGRXGDRESST
34	GHGXGPGAGRR	A71	HTDXNGTSGPT	C32	GRGXIGGDSGE
37	EAAXRHHRDSR	A73	GRTXGGLPHER	C33	TAGXGLARRGG
39	GDEXGACTVEW	A74	GAGXPRAHRRP	C34	SDQXGAQHGEA
42	AGHXRNSEPTT	A75	GGGXGANAARN	C37	GDGXGRKGAGS
47	GNPXGGNRDRD	A78	NGGXRRTAGAG	C38	AAEXGGKARGG
48	RGSXDRTLRRR	A79	GGGXGANAARN	C42	GRPXRQQNGRR
53	DPGXAQGKRAG	A80	RWRXKRAGRSK	C45	GRHXARADVSE
54	RRGXGGGGEPR	A81	GDSXRTPTERR	C46	EQEXDRRKRTG
55	YGRXARGGQTR	A83	QAKXAGARTPG	C47	GKTXETGDPRR
56	AGGXTVRGGGR	A84	RRTXGGWAAKG	C48	QGRXRGADEAV
59	EGGXKAAARKR	A86	GPWXEAGREDA	C50	SANXYKSSRQG
60	GKRXGGAQPXQ	A87	HGRXGQRRRKR	C51	WRRXTENQQSQ
61	RDRXTGNPREE	A89	NRGXRRNNSDV	C52	SGGXGHGQHAT
62	AGEXNDNRSNR	A90	PRDXCRRGGGR	C53	SSRXARRDHRA
63	GSTXRPGNNRR	A92	DQEXGEMGGAG	C54	NRRXGTRAEVA
64	SAVXGRHGAAT	A95	GGSXTTTTSGR	C57	HAQXRADNQEG
65	TRSXTGEDGRG	B01	RRGXDRATERG	C64	GESXRGPGSGT
69	ATEXTKGSNTG	B02	GHNXSAHDGAE	C66	GYDXAPREPSG
70	KDGXGGEEGET	B03	DRGXTAEEERR	C67	SQHXRHTRKTA
71	QTDXGKAARDA	B05	ERQXKIRAPGR	C70	DRAXPGGRRHR
72	ARSXRHGPAGR	B06	RVSXSDTCEDR	C71	GRRXHPCGSGE
73	KGRXGRRQGTG	B07	SARXESQPGAH	C72	RNGXSSGRHGD
74	EREXETHGRGC	B08	HHRXGQRAGRG	C74	DRAXGGGKLSD
78	GALXTNKREGG	B09	RNGXSSGRHGD	C76	DYPXRRRRHEM
79	DGAXPDSSLVE	B13	GRSXRRAPAGC	C77	ASEXREQKKGR
80	RAKXTGESGAT	B14	DRRXGGRNTRG	C82	SNNXSETGSRG
82	DGEXHGPSHAE	B15	RTEXHRPDPPR	C83	QRGXGHGPRRT
84	RNAXGSSGKGG	B16	GTKXATRRPGT	C85	GLNXGRPKNRP
87	GERXNHSRAGR	B19	GEPXEPGAAGG	C87	EHTXADAGKGR
88	GADXKKAAGRR	B20	RPGXGGARRNR	C89	DGLXPSPHKDG
89	NGGXSRGLDRE	B24	TNQXTQDADRS	C90	SGWXGTDRAGE
93	GGAXDGGEQTG	B25	EDEXTGTRRRK	C92	AERXGQQRRNR
96	ADRXPRQGGSA	B26	KDGXGDEEGET	C93	QQQXRDTRAGP
A01	GGAXGRRGAGE	B27	EKSXHLPDRSG	C94	RDRXTGNPREE
A02	GTGXRGAQPAR	B29	GRAXTDDRVER	C95	KDGXGGEEGET
A05	RNAXERRQKRG	B31	PGTXGSPTREQ	D01	GHEXRGRRETG
A06	EDDXERRPGAX	B33	TRSXGPEKTRP	D02	TNQXTQDADRS
A07	GKTXNHDRRDK	B36	AAAXAAGGGER	D04	HPGXHGAKRGD
A08	ASSXGWTRGGQ	B37	PDAXRGRRGRA	D07	TAEXRDGRRGK
A10	GAHXETTGATK	B38	GDTXRKGRTGS	D09	HHRXGQRASRR
A13	AEKXRDRRRKR	B41	GSRXGGGREPE	D10	GGGXDDGGNEG
A15	NKGXQGRRSAG	B42	TRTXRSDPRRT	D11	GDEXAKDASHR
A17	GKTXNHDRRDK	B44	GAKXTERAAEP	D16	DRGXAHHANEE
A20	RSAXQRKTRGN	B46	ESAXRSNDTGP	D17	AGGXRGKNTRG
A21	KHGXGDAQGGG	B47	RKSXPTRAGRG	D22	MNDXGGKDTGG
A23	RSRXKDRETDQ	B48	HHGXAGQRTGT	D23	AGNXRKARHPV
A25	GTGXNERSRPR	B53	RKRXGLKRSPR	D24	GLRXPSIGHDH
A26	GRQXAHRGKGT	B54	AAKXQRPHQRT	D28	RKRXPTKSGRG
A28	PGAXGXAGHAA	B59	GELXRHGPKEC	D33	RKRXPTKSGRG
A29	EETXREHGRRR	B61	GTQXAPDRNPA	D35	HHRXSQRAGRG
A30	RDGXRGRTTQP	B69	KDGXGGEEGET	D36	DRGXAQHASEE
A32	DREXGDGQADG	B70	XDQXPGAAKGE	D39	GTGXTESGSGG
A35	QSRXAGGREGG	B73	QHNXREGRDGG	D40	DEGXEGSTPEG
A36	EANXGGRGTES	B78	DRGXRDARGDA	D41	LNAXGRRSVSW
A39	GPPXKEDDRGG	B80	DDSXPGTGQGH	D42	ASSXSDHNTGS
A41	PRCXDTARACR	B86	DRGXAQRASEE	D44	GTGXRCQAGDT
A42	PQGXDQTAKEE	B89	RTAXEAVDSRT	D45	RGSXYRALRRR
A46	PARXRTGRRNE	B92	EAKXPDPRAAR	D46	SAVXGRHGAAT
A47	TSSXRDGATEG	B94	RNGXSSGRHGD		

Table 1. Sequences found by *in vitro* selection.

Peptide #	Sequence	Frequency
B09	RNGXSSGRHGD	4
B69	KDGXGGEEGET	3
11	DGEXHGPSHAE	2
61	RDRXTGNPREE	2
A07	GKTXNHDRRDK	2
A49	ANPXAARRGTG	2
A75	GGGXGANAARN	2
C06	RGSXTRGAADG	2
C15	TRSXTGEDGRG	2
D02	TNQXTQDADRS	2
D28	RKRXPTKSGRG	2
D46	SAVXGRHGAAT	2
D48	GGRESTQSSEV	2

Table 2. Sequence list of duplicated peptides. X indicates aminophenylalanine coupled with azobenzene.

Abbreviation	Sequence[a]
B09	MA**RNGXSSGRHGD**GPSDYKDDDDK
B69	MA**KDGXGGEEGET**GLSGDYKDDDDK

Table 3. Peptides synthesized by the solid phase method.

[a] Bold letters indicate peptide sequence selected from library. Underlined letters indicate FLAG sequence. X indicates aminophenylalanine coupled with azobenzene.

3.3.3 Binding activity

The interaction of the selected peptides with the GSH-immobilized microbeads was investigated. Peptides containing the sequence of B09 and B69 with fluorescein at the N-terminus for fluorescence labelling, and the FLAG sequence (DYKDDDDK) through a spacer sequence (GPS and GLSG) at the C-terminus for improving the solubility, were synthesized.

The binding behaviour onto the GSH-immobilized microbeads was assayed by measuring the fluorescent intensity and the result is shown in Figure 8. Higher binding affinity for B09 over B69 was observed. The dissociation constants (Kd) of the B09 and B69 peptides bound to the microbeads were calculated as 5.21 and 1.19 μM, respectively, by curve fitting to a Langmuir isotherm.

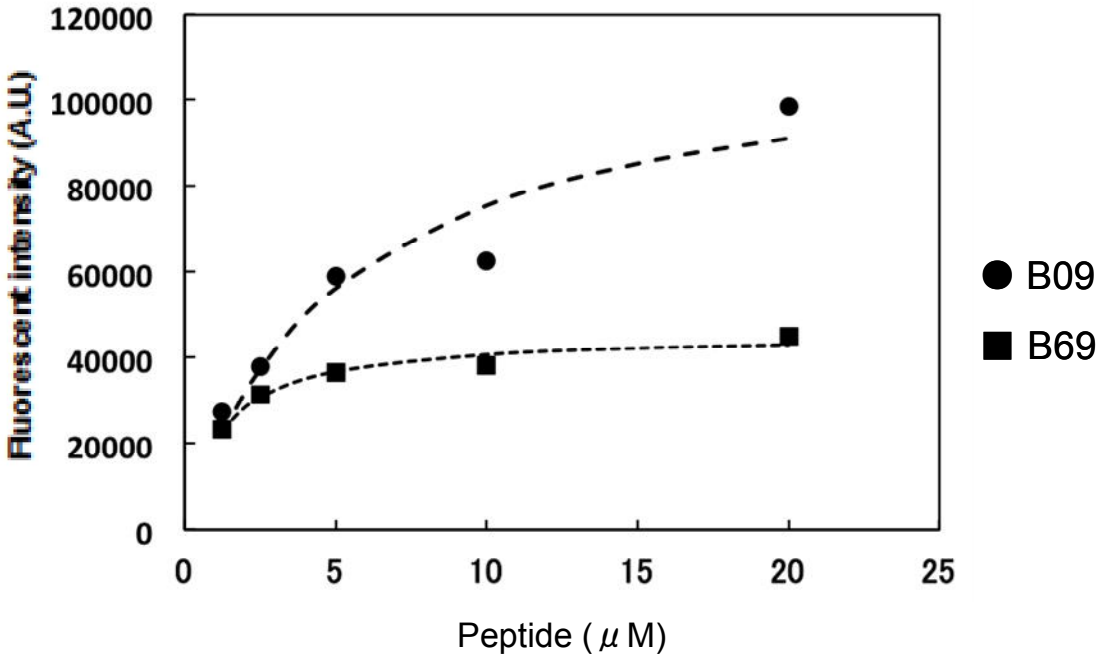


Fig. 9. Detection of dissociation constant of peptide B09 and B69 against GSH.

One of the peptides, B09, adsorbed onto the microbeads under visible light irradiation but this adsorption was significantly reduced by UV irradiation, although the B69 peptide did not show significant photo-responsive differences in adsorption behaviour.

In contrast to this difference between the B09 and B69 peptides, azobenzene on both the B09 and B69 peptides showed conformational changes in response to photo-irradiation, as judged by changes in the UV

spectra. These spectral changes for peptides B09 and B69 were similar. This result also indicates that the *cis-trans* conformational change of the azobenezene occurred in the B09 and B69 peptides. Thus, the less photo-responsiveness of the B69 peptide is likely to be from a lack of some components in the ribosome display. A longer peptide sequence length, close to the full length displayed, may show the photo-responsiveness.

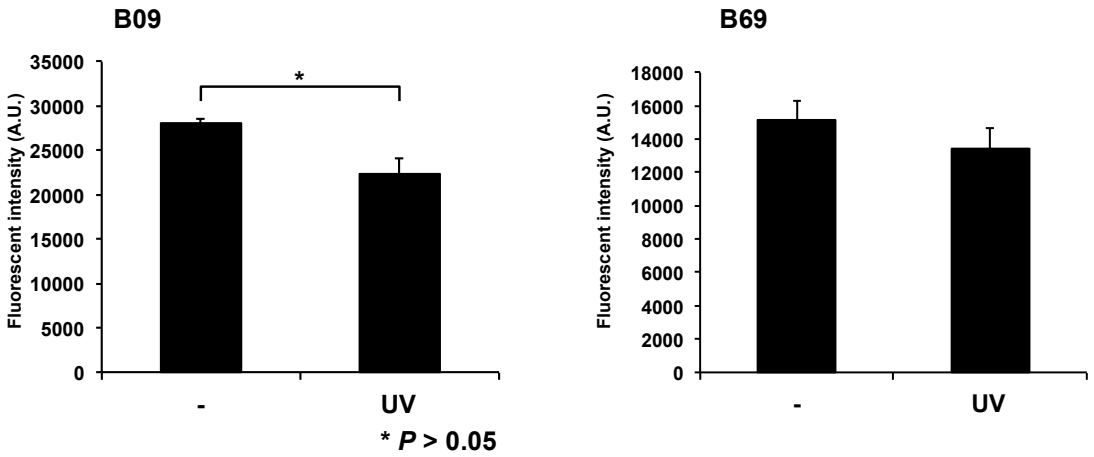


Fig. 10. Photo-responsive binding of peptides containing either the B09 or B69 sequence.

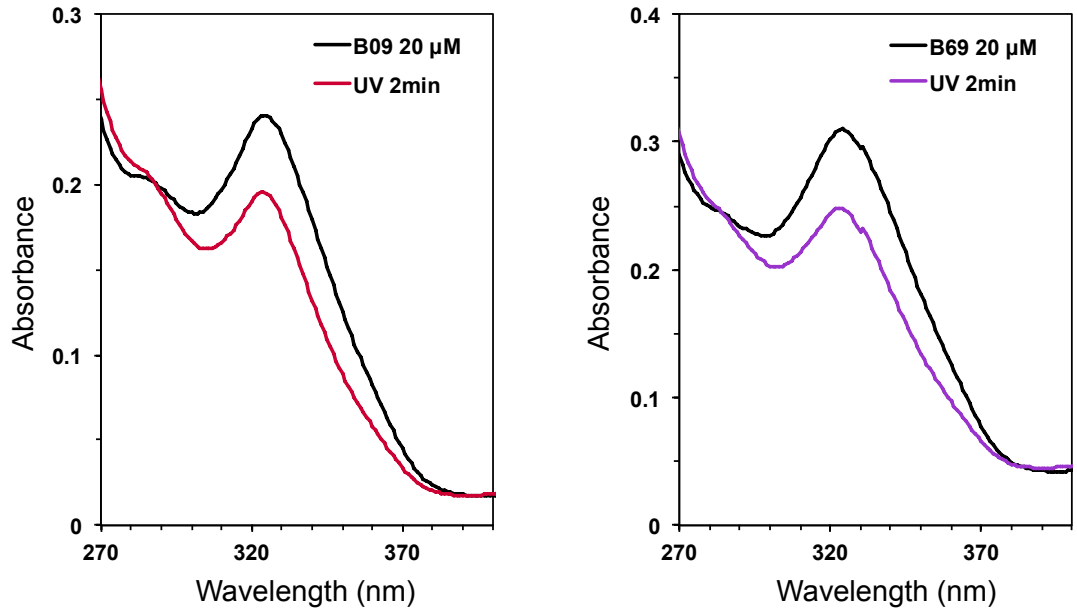


Fig. 11. UV spectra of B09 and B69 under irradiation of visible and ultraviolet light.

3.4 Conclusions

In conclusion, a photo-responsive peptide aptamer against glutathione-immobilized microbeads was isolated using *in vitro* selection combined with photo-manipulation. The application of the strategy proposed in this study renders in vitro selection of photo-responsive host molecules against various guests as promising.

Reference:

1. W. C. Winkler and R. R. Breaker, Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol*., 59, 487-517, 2005.
2. C. Tuerk and L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249, 505-510, 1990.
3. A. D. Ellington and J. W.Szostak, In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346, 818-822, 1990.
4. W. Mok and Y. F. Li, Recent progress in nucleic acid aptamer-based biosensors and bioassays. *Sensors,* 8, 7050-7084, 2008.
5. M. Crawford, R. Woodman and P. Ko Ferrigno, Peptide aptamers: tool for biology and drug discovery. *Brief Funct. Genomic Proteomic*, 2, 72-79, 2003.
6. A. Wada, S. Y. Sawata and Y. Ito, Ribosome display selection of a metal-binding motif from an artificial peptide library. *Biotechnol. Bioeng*., 101, 1102-1107, 2008.
7. J. Ciesiolka and M. Yarus, Small RNA-divalent domains. *RNA*, 2, 785-793, 1996.
8. E. Vianini, M. Palumbo and B. Gatto, In vitro selection of DNA aptamers that bind L-tyrosinamide. *Bioorg. Med. Chem*., 9, 2543-2548, 2001.

9. S. D. Mendonsa and M. T. Bowser, In vitro selection of aptamers with affinity for neuropeptide Y using capillary electrophoresis. *J. Am. Chem. Soc*., 127, 9382-9383, 2005.
10. S. E. Lupold, B. J. Hicke, Y. Lin and D. S. Coffey, Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen. *Cancer Res.*, 62, 4029-4033, 2002.
11. M. S. Raddatz, A. Dolf, P. Knolle, M. Famulok and G. Mayer, Enrichment of cell-targeting and population-specific aptamers by fluorescent-activated cell sorting. *Angew. Chem. Int. Ed. Engl*., 47, 5190-5193, 2008.
12. X. L. Li, W. Y. Zhang, L. Liu, Z. Zhu, G. L. Ouyang, Y. An, C. Y. Zhao and C. Y. J. Yang, In vitro selection of DNA aptamers for metastatic breast cancer cell recognition and tissue imaging. *Anal. Chem*., 86, 6596-6603, 2014.
13. G. P. Smith and V. A. Petrenko, Phage display. *Chem. Rev*., 97, 391-410, 1997.
14. S. Stahl and M. Uhlen, Bacterial surface display: trends and progress. *Trends Biotechnol.*, 15, 185-192, 1997.
15. J. M. Wwaver-Feldhaus, K. D. Miller, M. J. Feldhaus and R. W. Siegel, Directed evolution for the development of conformation-specific affinity reagents using yeast display. *Protein Eng. Des. Sel*., 18, 527-536, 2005.
16. N. Fukuda, H. Ishii, S. Shibasaki, M. Ueda, H. Fukuda and A. Kondo, High-efficiency recovery of target cells using improved yeast display system for detection of protein-protein interactions. *Aool. Microbiol. Biotechnol*., 76, 151-158, 2007.
17. P. A. Barendt, D. T. W. Ng, C. N. McQuade and C. A. Sarkar, Streamlined protocol for mRNA display. *ACS Comb. Sci.,* 15, 77-81, 2013.
18. J. Yamaguchi, M. Naimuddin, M. Biyani, T. Sasaki, M. Machida, T. Kubo, T. Funatsu, Y. Husimi and N. Nemoto, cDNA display: a novel screening method for functional disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions. *Nucleic Acids Res*., 37, e108, 2009.
19. R. Odegrip, D. Coomber, B. Eldridge, R. Hederer, P. A. Kuhlman, C. Ullman, K. FitzGerald and D. McGregor, CIS display: in vitro selection of peptides from libraries of protein-DNA complexes. *Proc. Natl. Acad. Sci. USA*, 101, 2806-2810, 2004.

20. J. Hanes and A. Plueckthun, In vitro selection and evolution of funcational proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA*, 94, 4937-4942, 1997.
21. A. L. Hopkins and C. R. Groom, The druggable genome. *Nat. Rev. Drug Discov*., 1, 727-730, 2002.
22. E. N. Brody and L. Gold, Aptamer as therapeutic and diagnostic agents. *J. Biotechnol.*, 74, 5-13, 2000.
23. M. S. Weinberg, Therapeutic aptamers march on. *Mol. Ther. Nucl. Acids*, 3, e194, 2014.
24. A. D. Keefe. S. Pai and A. Ellington, Aptamers as therapeutics. *Nat. Rev. Drug Discov.,* 9, 537-550, 2010.
25. S. P. Song, L. H. Wang, J. Li, C. H. Fan and J. L. Zhao, Aptamer-based biosensors. *TRAC-Trend. Anal. Chem.,* 27, 108-117, 2008.
26. E. S. Gragoudas, A. P. Adamis, E. T. Cunningham, M. Feinsod and D. R. Guyer, Pegaptanib for neovascular age-related macular degeneration. *N. Engl. J. Med*., 351, 2805-2816, 2004.
27. E. W. Nq, D. T. Shima, P. Calias, E. T. Jr. Cunninqham, E. R. Guyer and A. P. Adamis, Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat. Rev. Drug Discov*., 5, 123-132, 2006.
28. K. Min, H, Jo, K. Song, M. Cho, Y. S. Chun, S. Y. Jon. W. J. Kim and C. Ban, Dual-aptamer-based delivery vehicle of docorubicin to both PSMA (+) and PSMA (-) prostate cancers. *Biomaterials,* 32, 2124-2132, 2011.
29. S. Harper and D. W. Speicher, Purification of proteins fused to glutathione S-tranferase. *Methods Mol. Biol*., 681, 259-280, 2011.

CHAPTER 4

CONCLUSIONS

This thesis illustrated the design and the development of the incorporation of non-natural amino acids into peptides using the cell-free translation systems. The incorporation of one and two PEGs with different lengths and the incorporation of an azobenzene molecule for the *in vitro* selection of photo-responsive peptide aptamer were discussed in detail.

In chapter 1, the general background and the development of the expansion of genetic code were presented. The expansion of genetic code has attracted great attention, because that a great number of non-natural amino acids are incorporated into peptides and proteins. And the incorporation of non-natural amino acids carrying special functional groups can provide powerful tools for generating peptides and proteins with novel physical, chemical, biological and pharmaceutical properties. Both chemical and biochemical approaches have been developed to incorporate the non-natural amino acids into peptides and proteins. One of the site-specific incorporation methods is utilizing misacylated tRNAs conjugating with non-natural amino acids which recognize a stop amber codon or a frameshift codon with high fidelity. Up to now, the peptides and proteins with non-natural amino acids have been applied to probe the structure and the functions, alter the properties and generate new materials and new drugs.

In chapter 2, one and two PEGs with different lengths were successfully incorporated into one polypeptide backbone. The incorporation of polyethylene glycol (PEG) can overcome the drawbacks of peptides and proteins: the low stability, the bad biocompatibility and fast degradation *in vivo*. Many efforts have been done to incorporate PEGs. The site-specific PEGylation has been achieved; however, it is highly challenging to maintain the full activity of the proteins. The main reasons are that the selectivity is low and the PEGs of the same length result in the unwanted PEGyaltion near

or at the active site of proteins. And the incorporation of PEGs with different lengths can achieve to maintain the full activity. Herein, I prepared two DNA sequences containing both the amber and the frameshift codon. Then PEGs with different lengths were genetically incorporated into one polypeptides backbone with tRNAs which were acylated with PEG-containing non-natural amino acids. All the *in vitro* translations were performed using RYTS kit according to the manufacture's protocol. The translation reaction was carried out at 30℃ for 2hrs in the presence of 200 pmol of each PEG-AF-tRNA. Then the translated peptides were purified using prewashed Anti-DDDDK-tag mAb-Magnetic Agarose. The purified peptides were analyzed by MALDI-TOF-MS and the amount of each peptide was estimated by the area intensity comparing to 3 x FLAG peptide, which was used as the internal standard. First, I studied the incorporation of one PEG with various lengths into one polypeptide via the amber codon or the frameshift codon. The incorporation of longer PEG chain resulted in the less amount of the translation product. The PEG-length-dependence of the translation product from PEG8 to PEG24 may be explained by the steric hindrance between PEG and ribosome, and between PEG and EF-Tu. I also successfully incorporated two PEGs via the amber and the frameshift codon in one peptide. It was found that the translation efficiency depended on the length of the PEG, the codon and the incorporation sites. And I also found a mass peak corresponding to truncated peptides (lack of fMSKQIEVN or SKQIEVN) for single PEGylation. The amount of the truncated peptides increased as the molecular weight of PEG increased. It appears that the longer PEGs might block the peptidyl transferase reaction in the ribosome. And this observation could be offered as the materials for studying the translation process of non-natural amino acids. This site-specific PEGylation

method could maintain the full activity of the proteins and peptides and have the application in the precise synthesis of bioconjugate drugs.

Although one and two PEGs with different lengths were successfully incorporated into peptides with high selectivity, the translation efficiency was high only for the short PEGs (PEG4 and PEG8) and was rather low for the long PEGs (PEG12 and PEG24). So this method was greatly limited for PEGylation. The reason might be the steric hindrances between PEG and ribosome, and between PEG and EF-Tu. In this research, all the translations were carried out by using E. coli system, so the translation efficiency could be high if human or other mammal translation systems were used. And for the truncated peptides, which were reported by other researchers, the mechanism is still unclear. It could be studied systematically using different peptides sequences, different translation systems, and different non-natural amino acids. All the problems could be solved when the mechanism was clear.

In chapter 3, I incorporated one non-natural amino acid carrying an azobenzene residue that is a photo-responsive molecule into the peptides for the *in vitro* selection of the peptide aptamer. Aptamers are a special class of oligonucleothide acids or peptides isolated from a large random sequence library, which can bind to their targets with high specificity and high binding affinity. The classic selection methods include phage display, yeast display, mRNA display, and ribosome display and so on. Among the *in vitro* methods, ribosome display is one very simple and effective method. Therefore, I employed ribosome display to select the peptide aptamer. And the selected aptamers can be used as biomaterials, the biosensing probes, and the diagnostic and therapeutic tools. The targets range from the small ions to the big tissues. In this study, I chose glutathione (GSH) as the target for the

selection of the photo-responsive peptide aptamer. First, I prepared a random sequence library, which contained the amber codon. Through the amber codon, the non-natural amino acid carrying an azobenzene residue can be incorporated into the peptides. And all the sequences did not contain the stop codon at the 3' terminus so that the translated polypeptide, the ribosome and the corresponding mRNA can form a complex. The *in vitro* transcription and translation was performed using PURESYSTEM at 37℃ for 3h. Then the translated peptide solution was incubated with glutathione-immobilized microbeads. The unbound complexes were washed away and the bound complexes were eluted by adding a glutathione solution or irradiating with UV light for the next round of selection. After eight rounds of selection, the selected sequences were cloning and analyzed. 215 types of sequences were found from 269 clones. 12 sequences showed multiple repetitions in the analyzed sequences. Two peptides derived from plural clones were chosen: B09 and B69. The dissociation constants (Kd) of B09 and B69 peptides bound to the GSH-immobilized microbeads were calculated as 5.21 and 1.19μm, respectively. And for peptide B09, it detached from the GSH-immobilized microbeads under UV irradiation. Utilizing the photo-responsive peptide aptamer, it could be an alternative method to purify the proteins.

Although one photo-responsive peptide aptamer was successfully selected, the photo-responsiveness showed not so much difference after UV irradiation. The reason could be the photo-responsive molecule, azobenzene, was on the side chain of the peptide or only one azobenzene was on the peptide. So it can not change the structure of peptide greatly. To get high photo-responsive peptide aptamer, the photo-responsive moiety can be incorporated into the backbone of the peptides. Or two or more azobenzenes

can be incorporated into one peoptide. Besides the linear peptide aptamer, the cyclic photo-responsive peptide aptamer can also be developed, which would show high binding affinity.

In conclusions, I succeeded in incorporation of one and two PEGs of different lengths into one polypeptide backbone and in *in vitro* selection of a photo-responsive peptide aptamer with the incorporation of an azobenzene molecule.

ACKNOWLEDGEMENTS

My immeasurable gratitude goes first and foremost to Professor Yoshihiro Ito, my promoter as well as supervisor, for giving me the opportunity to pursue my Ph.D. abroad. I am also extremely grateful for his mentorship, his professional and inspiring supervision, his continuous support and encouragement during the three years. Of course, my Ph.D. thesis would not be finished without him. Also, it is very kind of him. I still remember that he asked me whether my parents worried about me or not in the subway.

My sincere gratitude goes to Professor Masayuki Yamamura, my promoter as well as supervisor, for giving me the opportunity to study for Ph.D. at Tokyo Institute of Technology. Thanks a lot for his great support and his guidance during the last three years.

My special gratefulness goes to Professor Daisuke Kiga, my promoter as well as supervisor, for his trust in me and for his continuous support. He is also available for discussion. And he is very constructive and I learned a lot from him about writing in a scientific manner and I still have a lot to learn.

My special thanks to Dr. Wei Wang, my best friend, who picked me up from the station, for helping me so much in daily life and in scientific research. We have known each other for a long time.

I would like to express my sincere thanks to Dr. Takanori Uzawa, whose research suggestions and insights were quite useful. I did enjoy the discussions which were organized by him.

I greatly appreciate Dr. Seiichi Tada, who helped me a lot on my research work. He taught me how to carry out the experiments.

I am also thankful to the scientific staff of our group. Ms. Noriko Minagawa and Ms. Michiru Iwashita helped me a lot carry out the biological experiments. They are very nice and enthusiastic about my research.

I would like to express my gratitude to Ms. Akiko Nose and Ms. Yuko Sato, who helped me so much on daily life.

My special thanks go to Ms. Fumie Nishi for helping me a lot on kinds of documents.

In the autumn of 2012, after I had worked for 4 years, I decided to come to Japan for Ph.D. study at Tokyo Institute of Technology. I still clearly remember that how nervous and excited I was when the plane landed in Japan. How time flies! Now the time comes to the end of my Ph.D. period. In this period, I met a lot of friends from many countries, from whom I received great help. Now, I would like to take the opportunity to express my heartfelt gratitude to all the people who helped me with the research work and my life in Japan.

I can not go without mentioning the support from an International Program Associate (IPA) program for graduate students from the research personnel support section, Global Relation Office, RIKEN.

Last, but not least, my deepest thanks go to my beloved parents for their understanding, patience and support over the past years.

Qingmin Zang

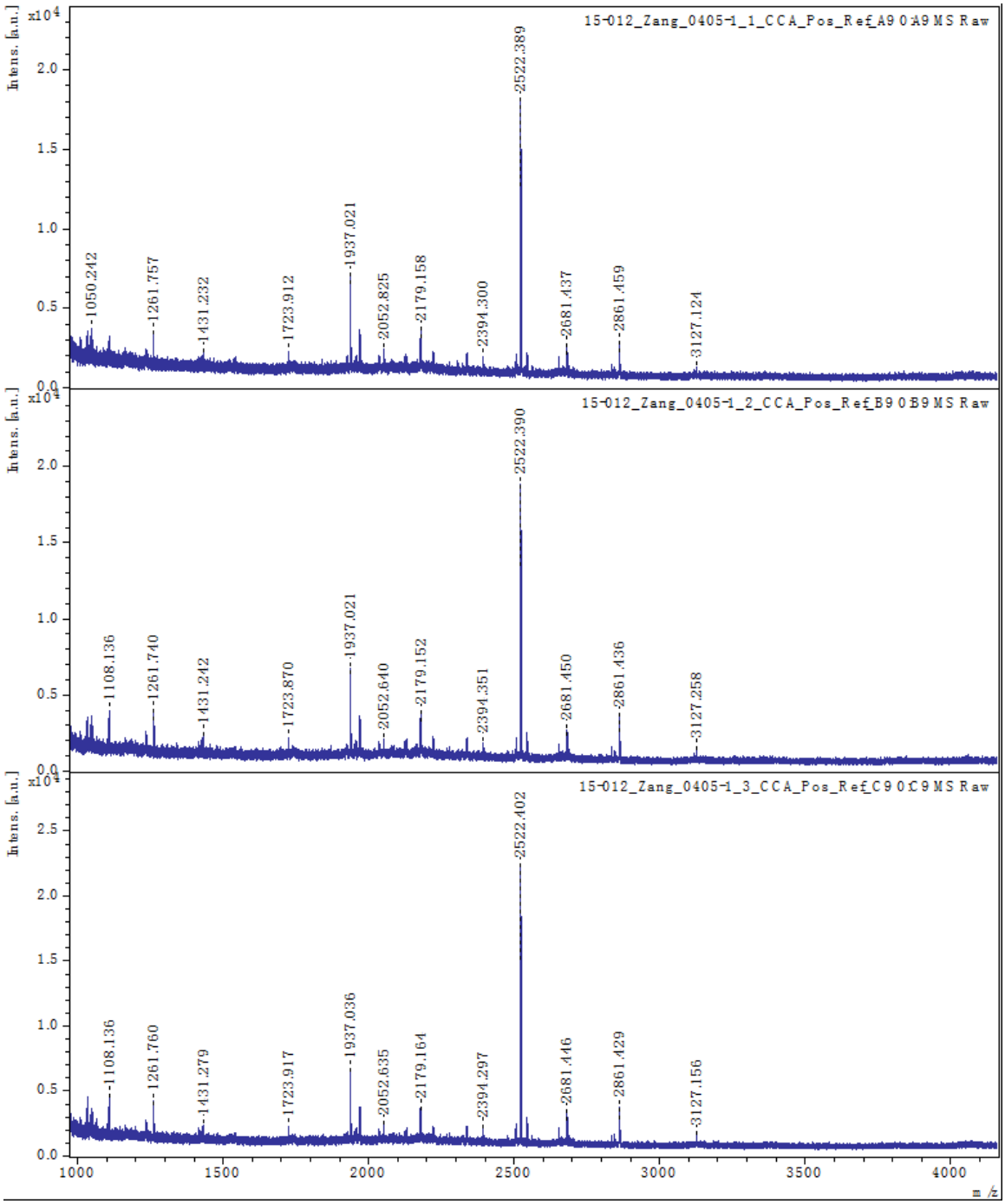
August, 2015

Wakoshi, Japan

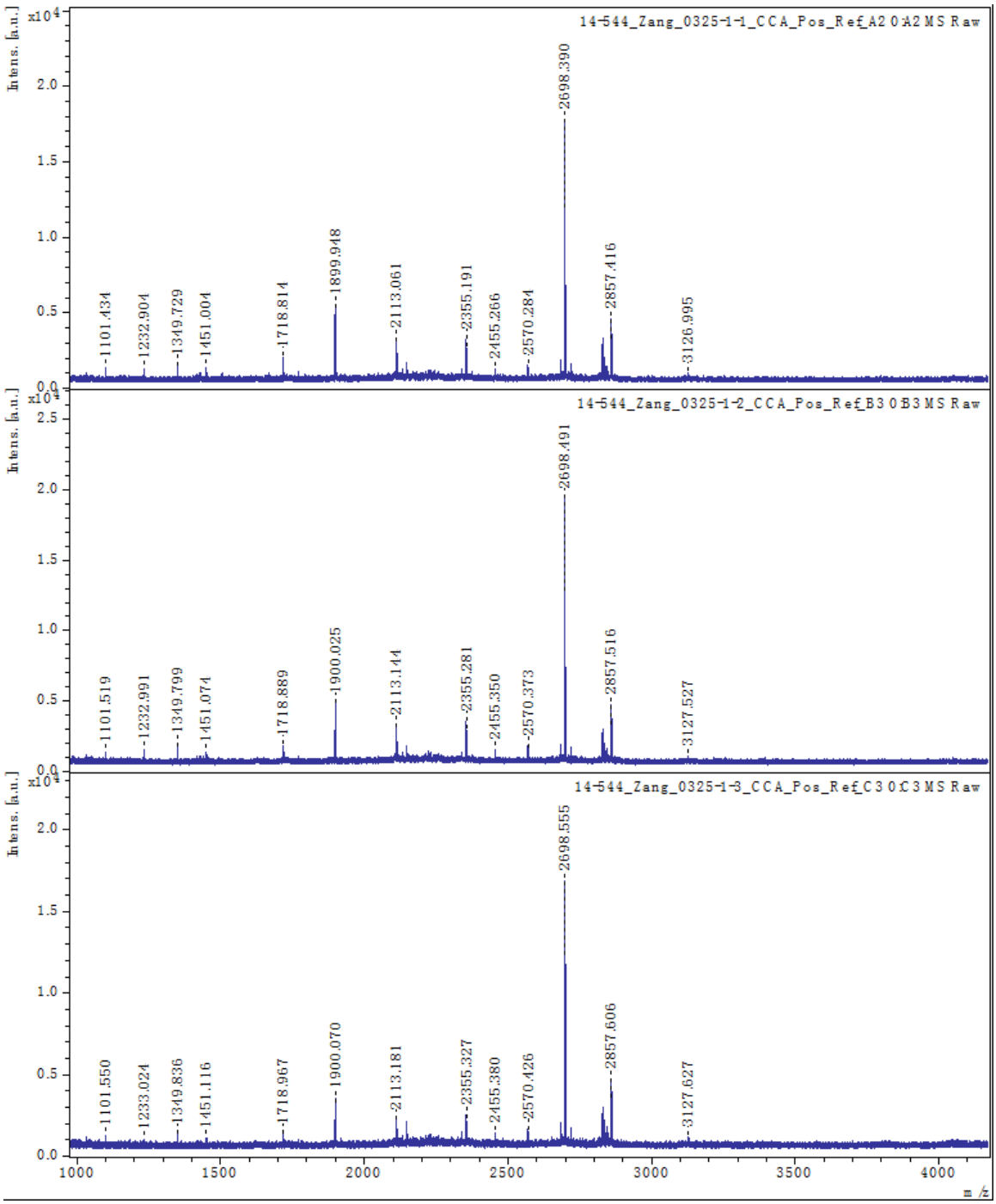
APPENDIX

MS and NMR spectra:

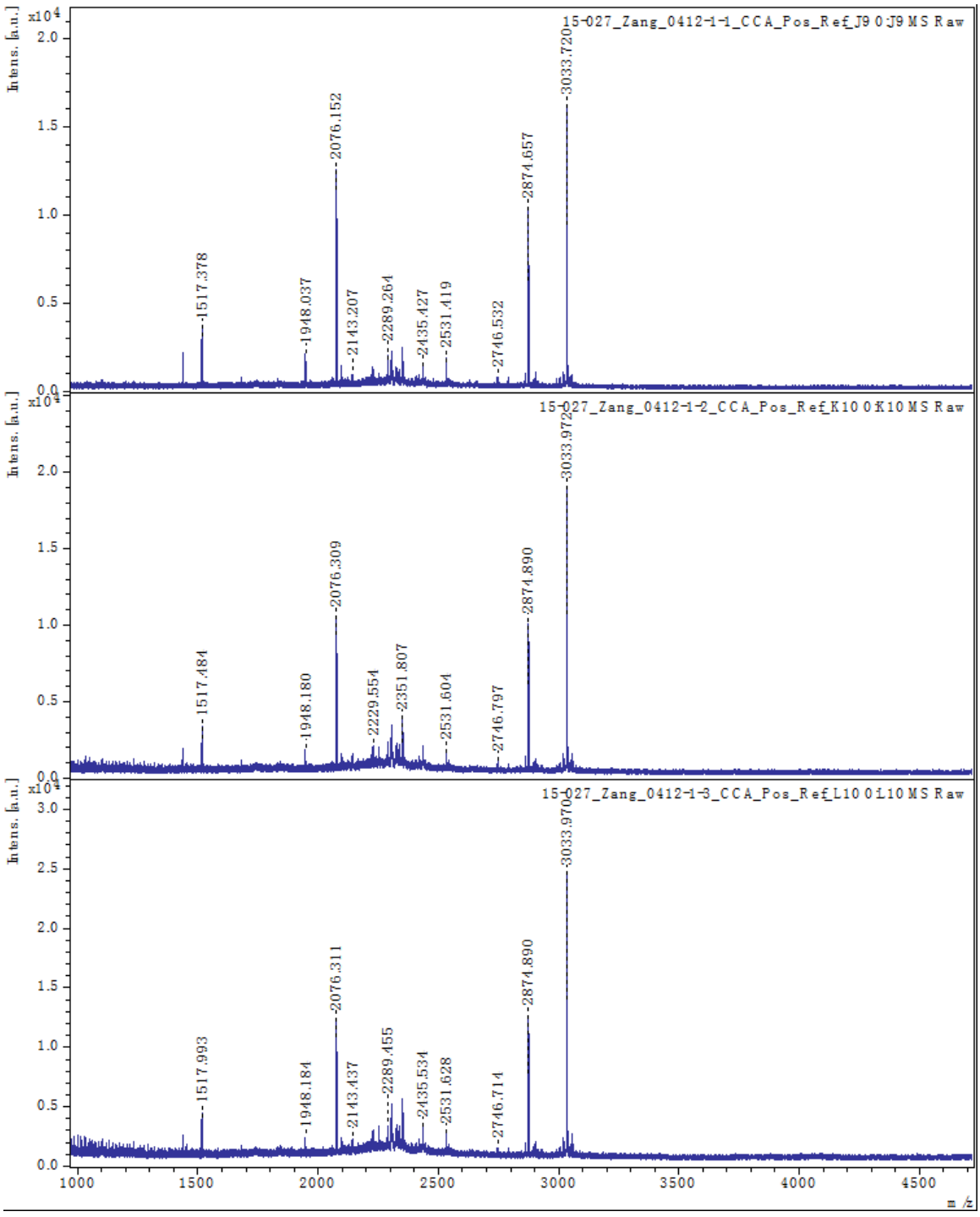
Appendix 1. MS spectra of amber-PEG4



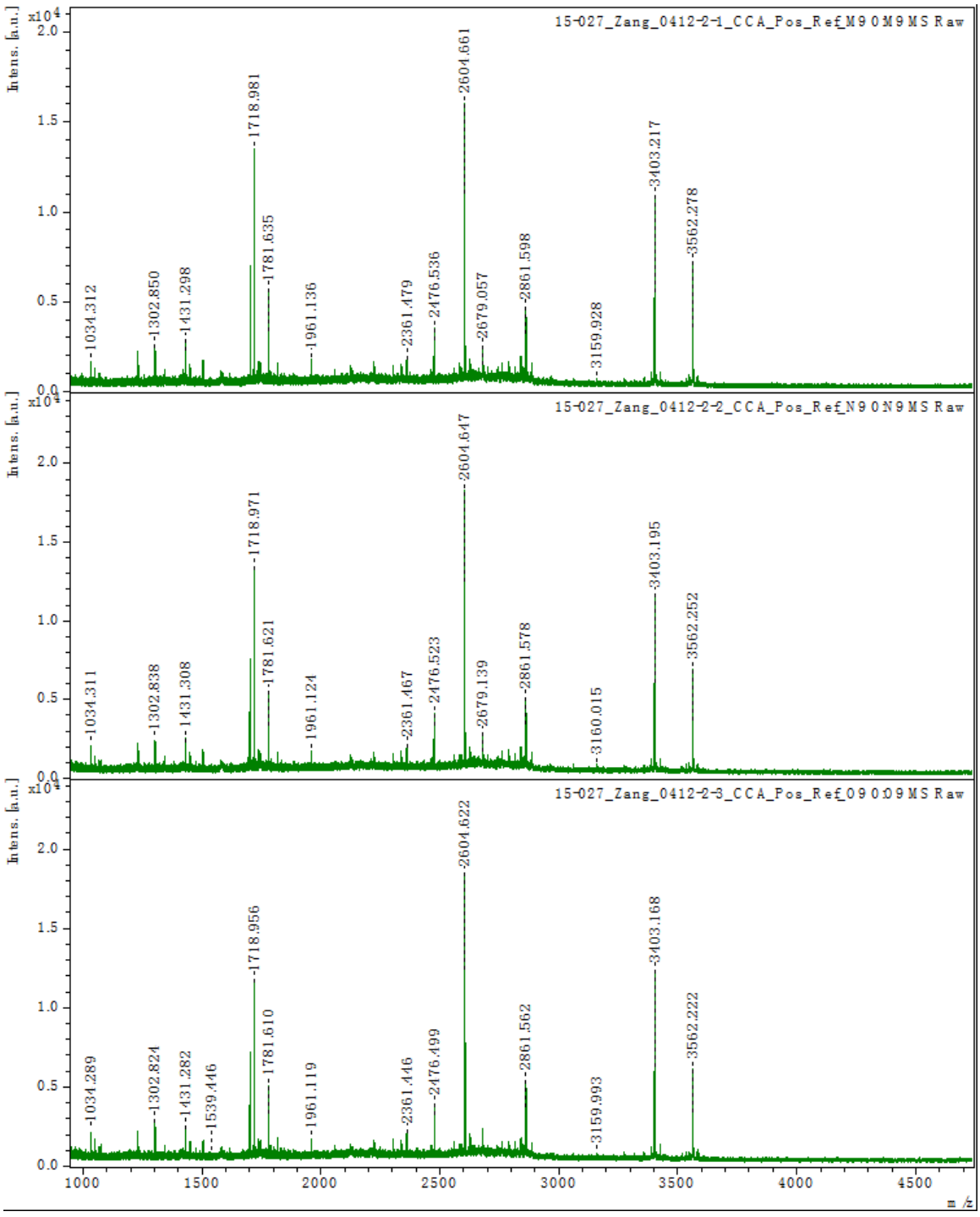
Appendix 2. MS spectra of amber-PEG8



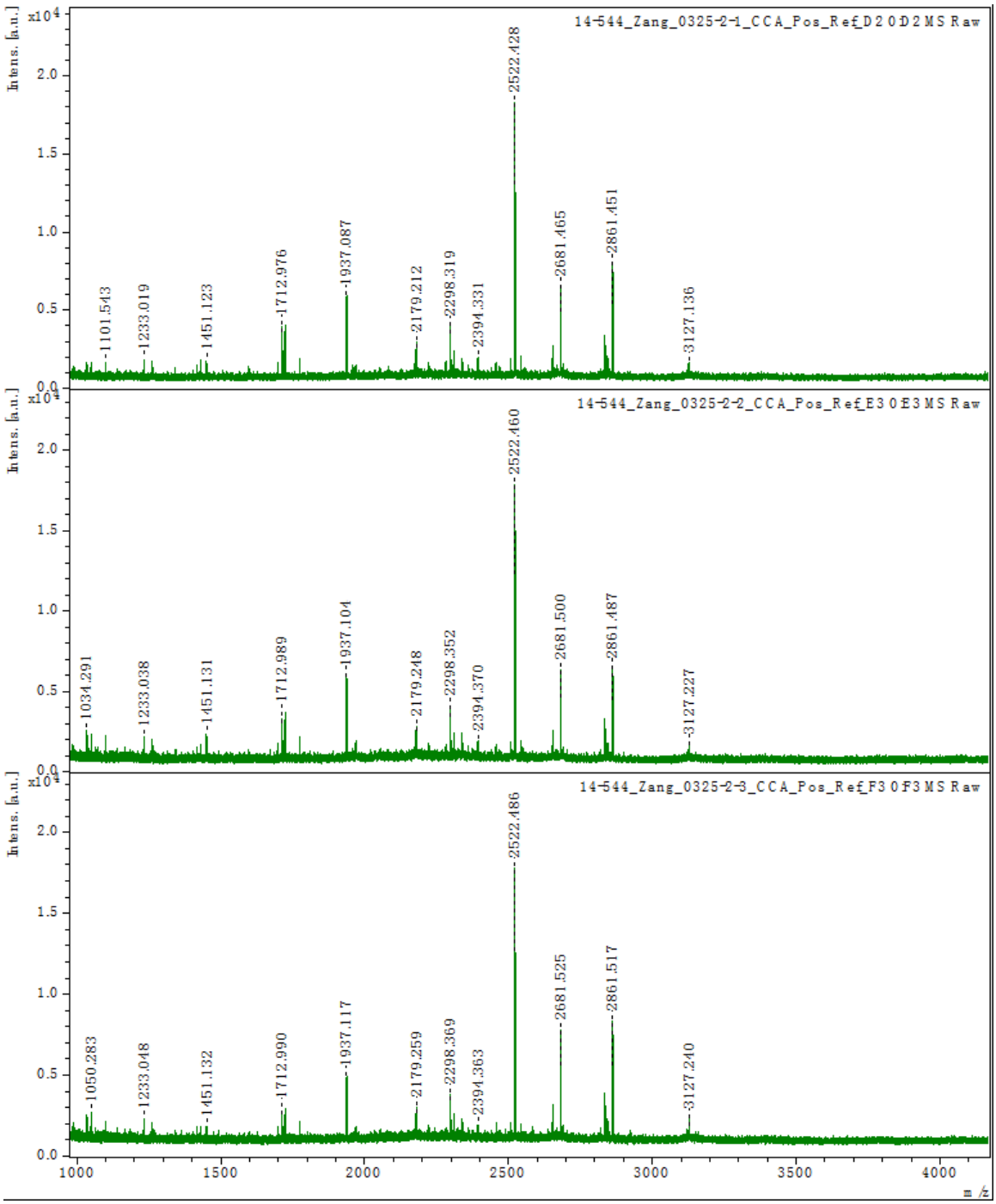
Appendix 3. MS spectra of amber-PEG12



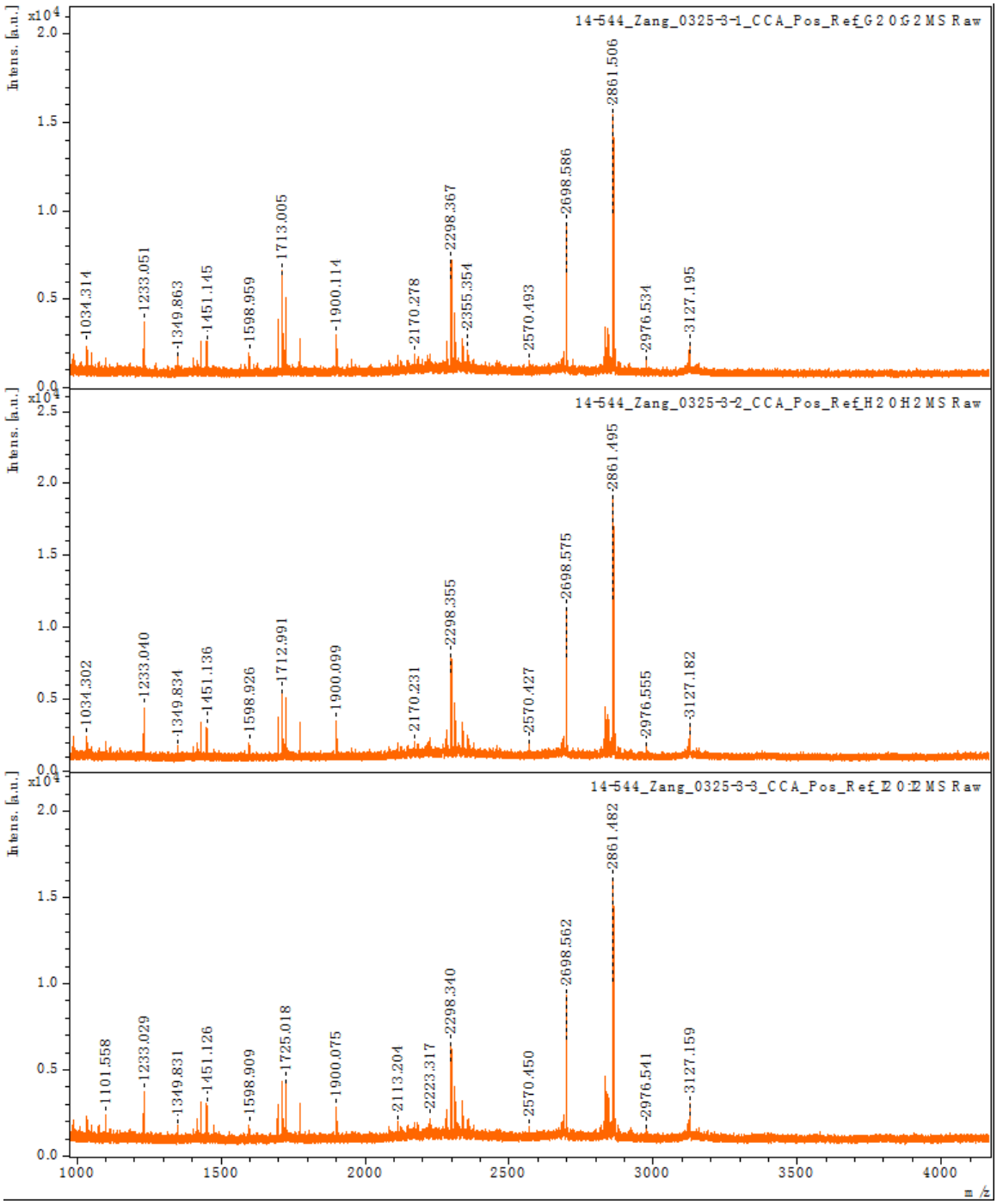
Appendix 4. MS spectra of amber-PEG24



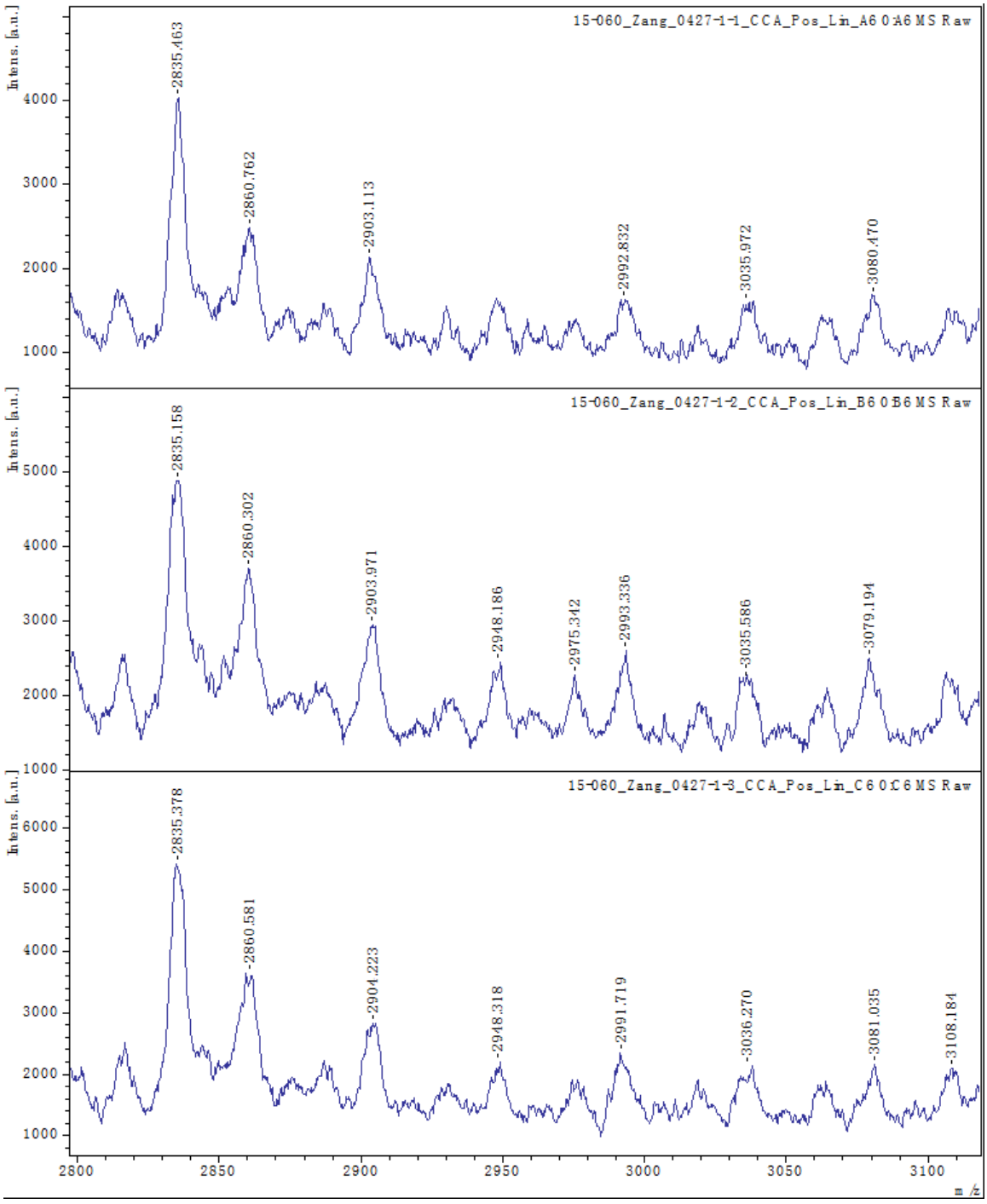
Appendix 5. MS spectra of CGGG-PEG4



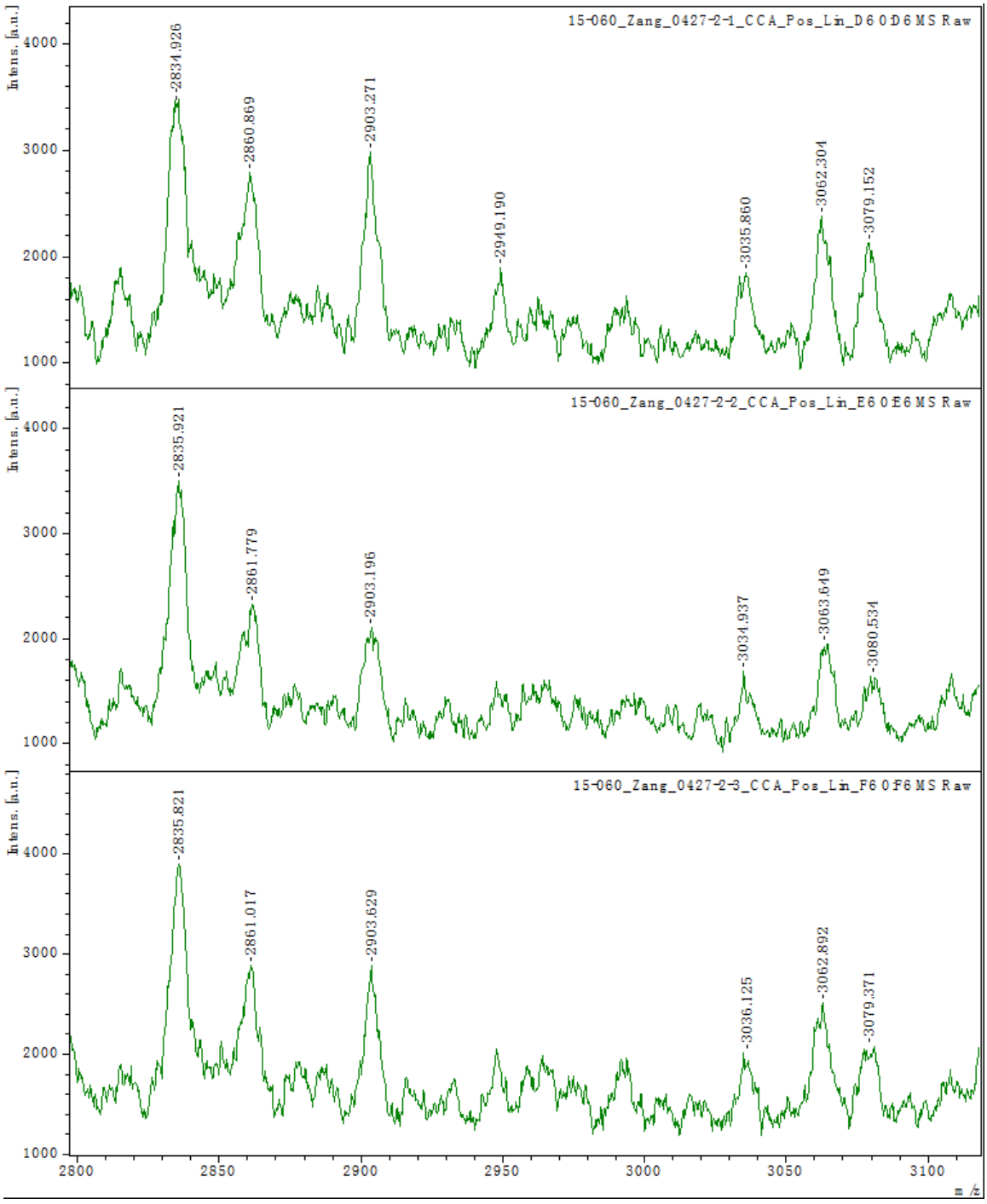
Appendix 6. MS spectra of CGGG-PEG8



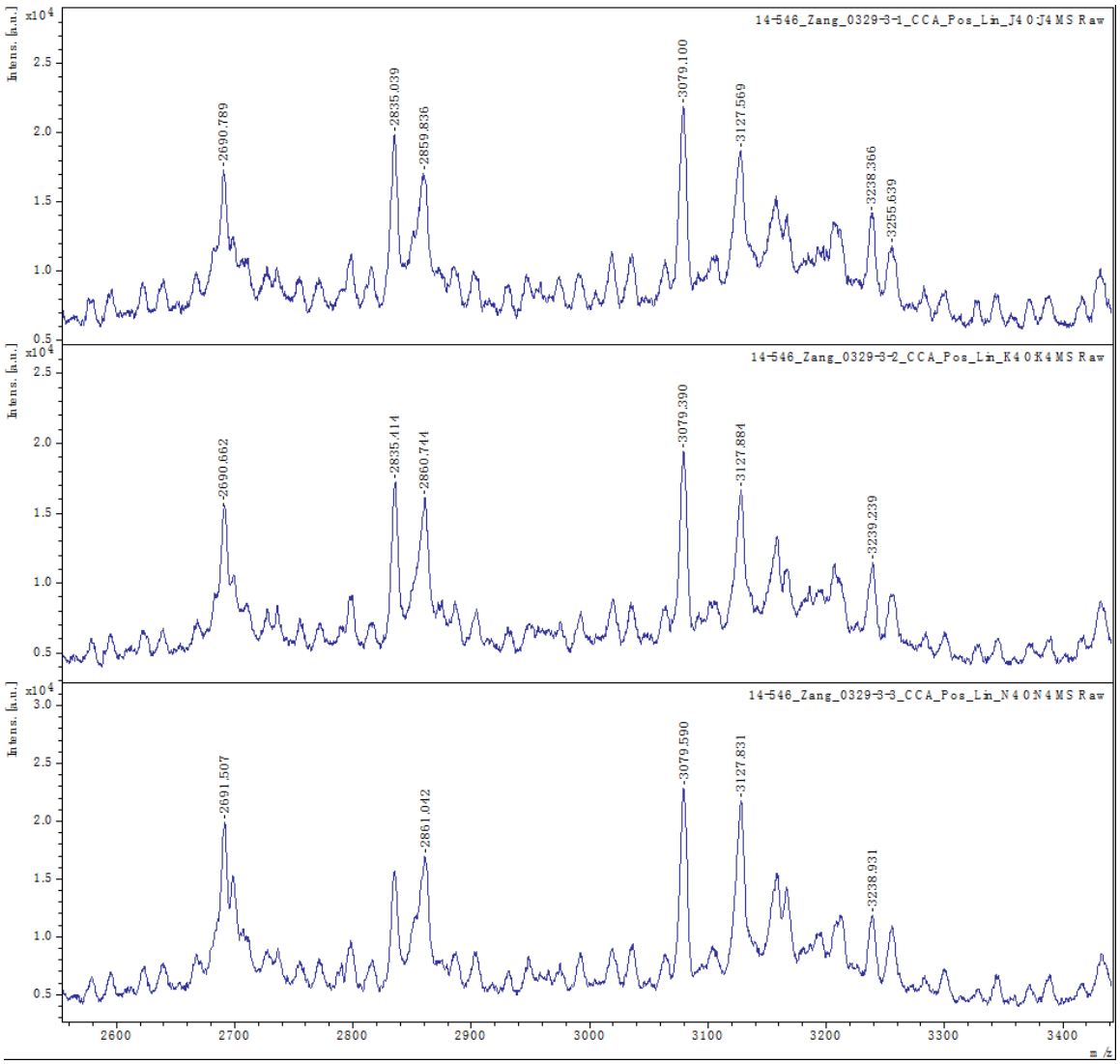
Appendix 7. MS spectra of amber (PEG4) ---CGGG (PEG4)



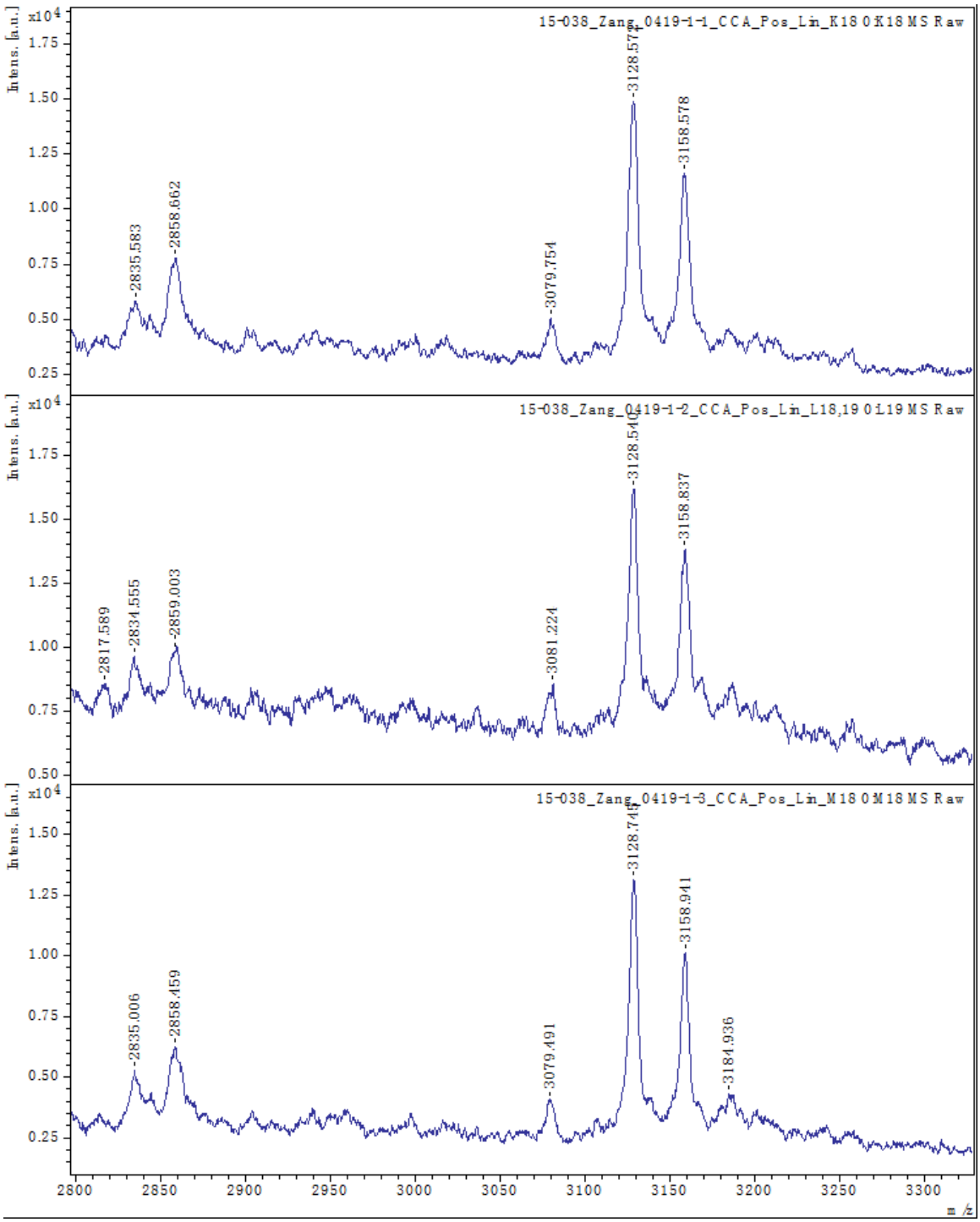
Appendix 8. MS spectra of CGGG (PEG4) ---amber (PEG4)



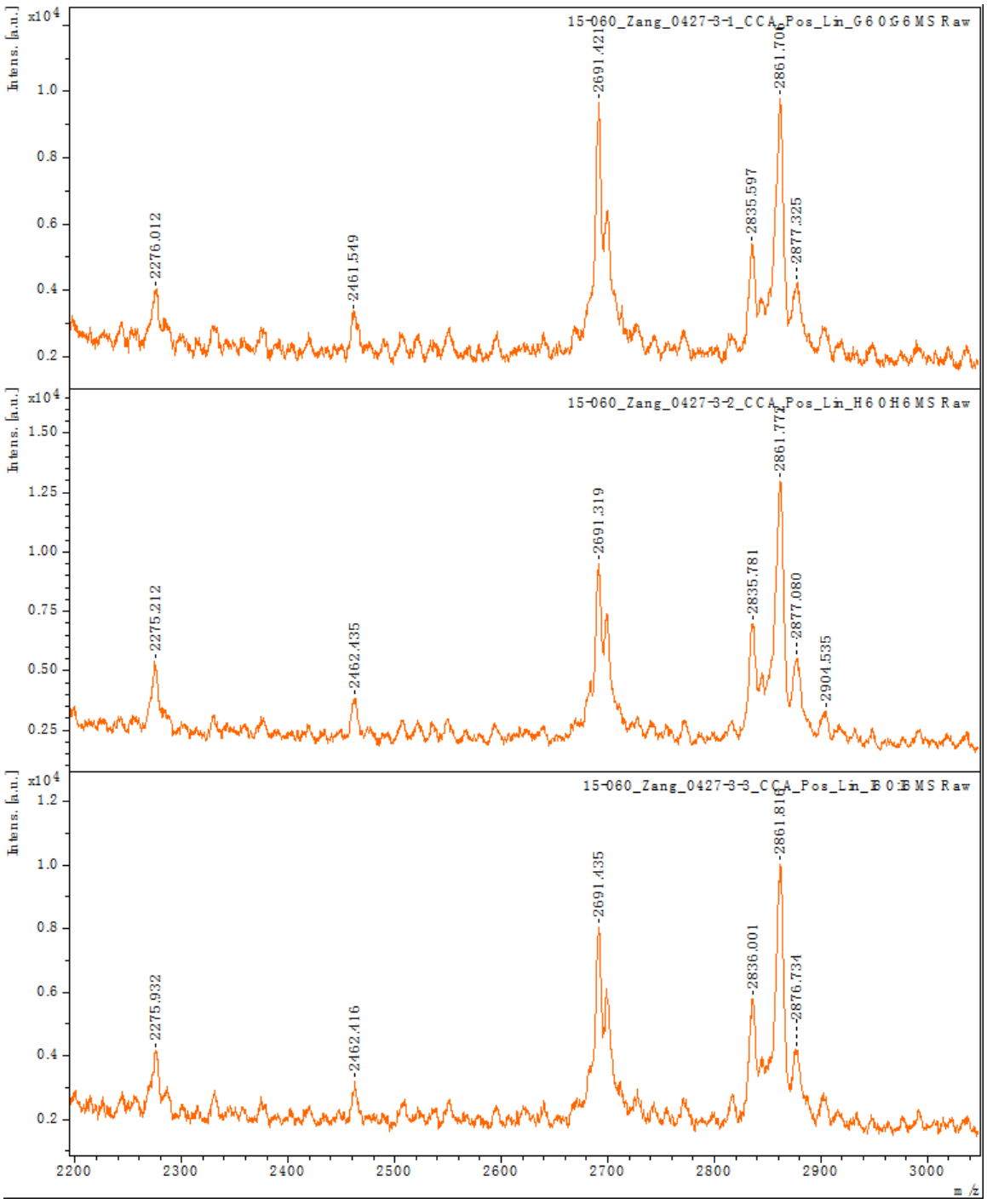
Appendix 9. MS spectra of amber (PEG8) ---CGGG (PEG4)



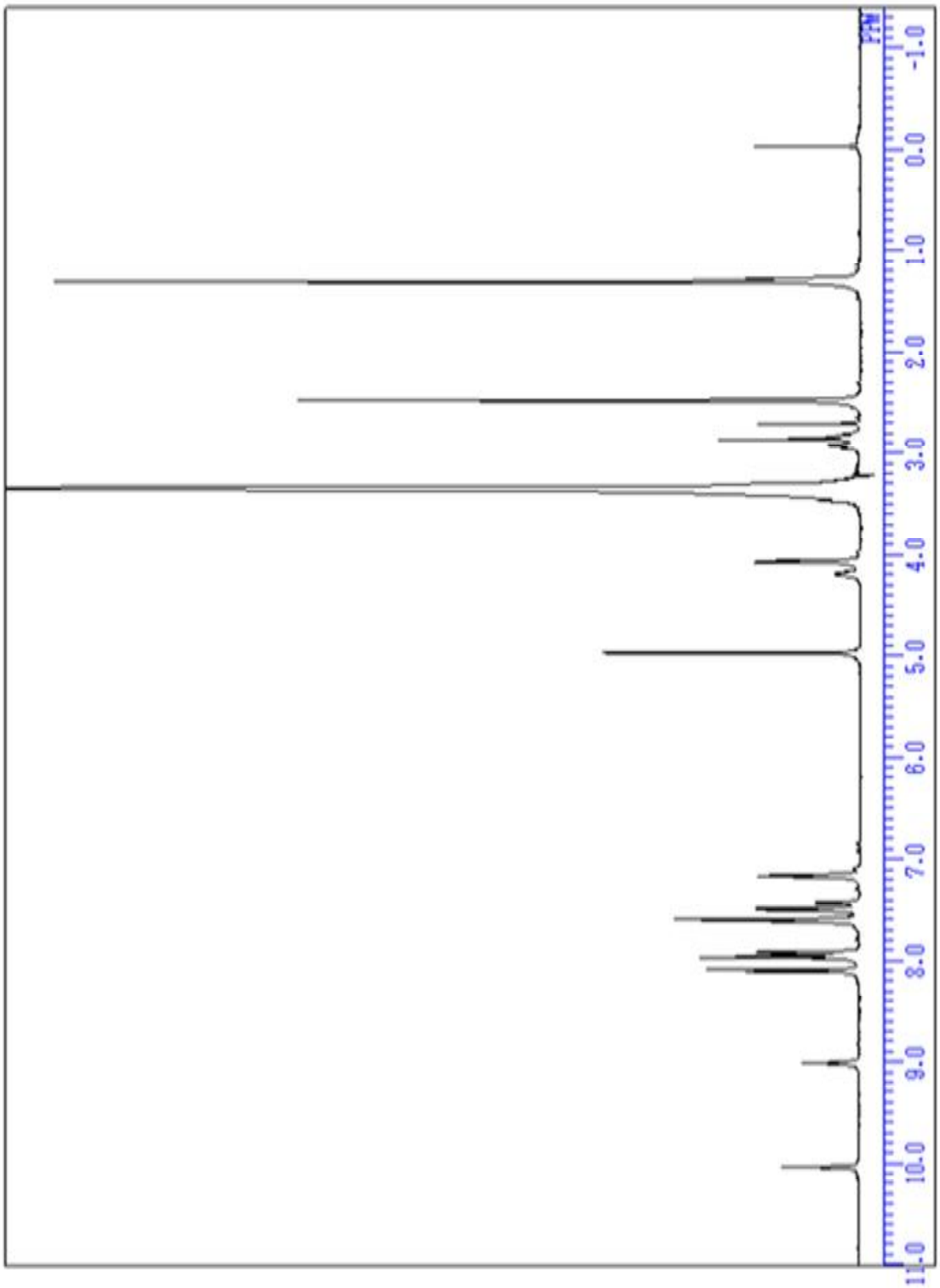
Appendix 10. MS spectra of CGGG (PEG4) ---amber (PEG8)



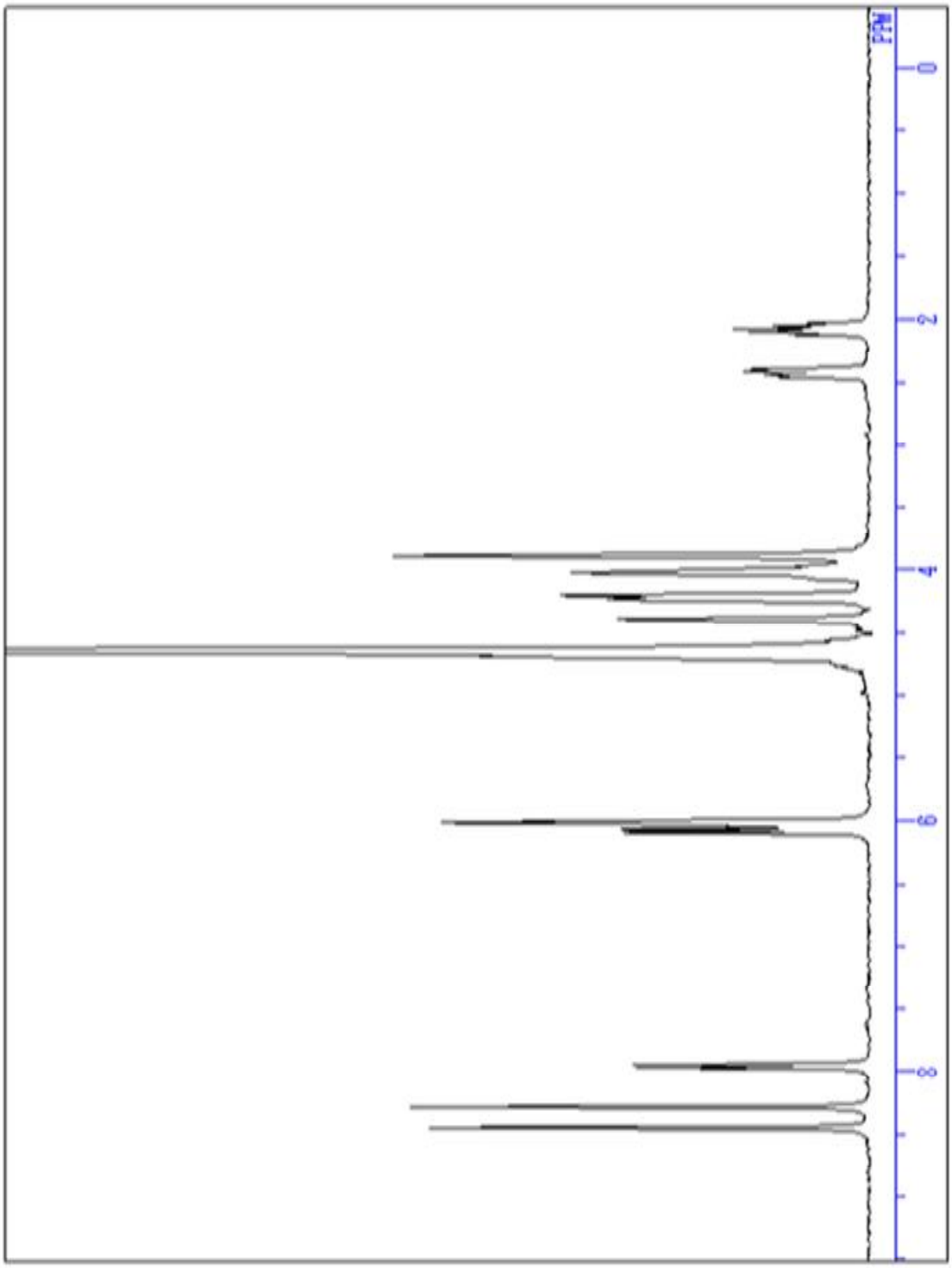
Appendix 11. MS spectra of CGGG (PEG8) ---amber (PEG4)



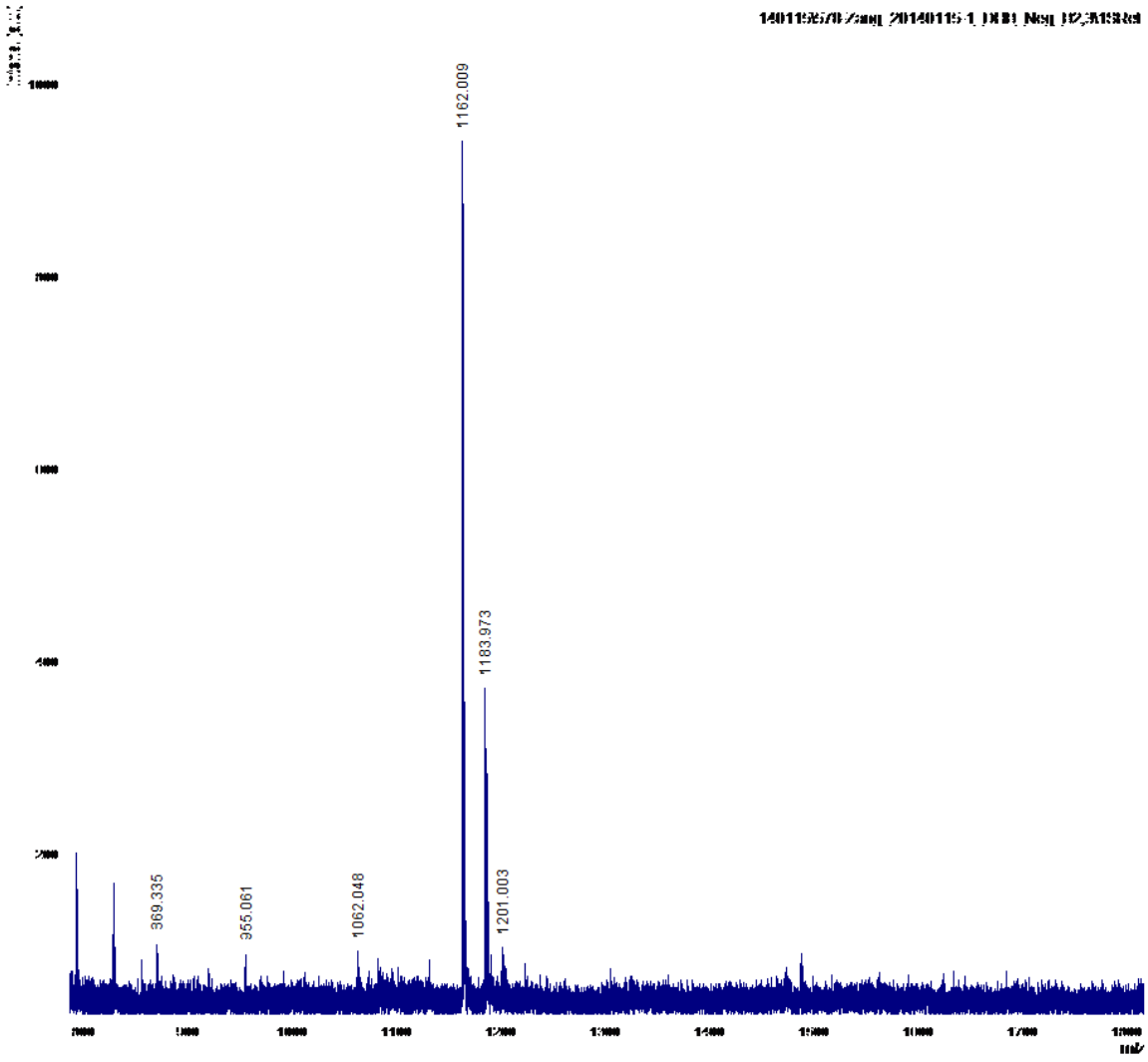
Appendix 12. NMR of N-Boc amino cyanomethyl ester (**5**)



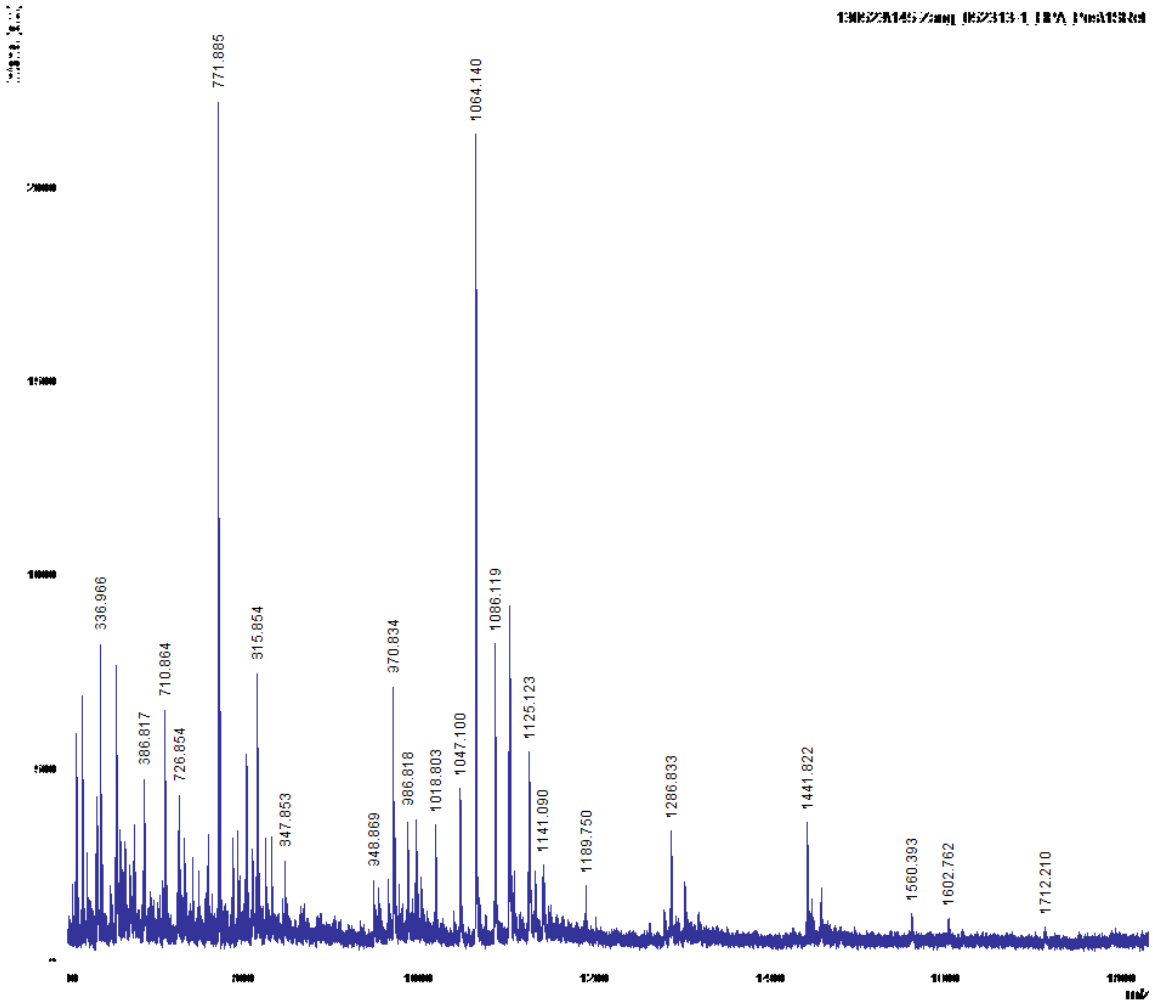
Appendix 13. NMR of pdCpA



Appendix 14. MS spectra of pdCpA conjugated with (Boc)-aminophenylalanine-azobenzene (**12**)



Appendix 15. MS spectra of pdCpA conjugated with aminophenylalanine-azobenzene (**13**)



List of Publications

Original papers:

1. Qingmin Zang, Seiichi Tada, Takanori Uzawa, Daisuke Kiga, Masayuki Yamamura and Yoshihiro Ito, Two site genetic incorporation of varying length polyethylene glycol into the backbone of one peptide. *Chem. Commun*. (accepted)
2. Seiichi Tada, Qingmin Zang, Wei Wang, Masuki Kawamoto, Mingzhe Liu, Michiru Iwashita, Takanori Uzawa, Daisuke Kiga, Masayuki Yamamura and Yoshihiro Ito, *In vitro* selection of a photoresponsive peptide aptamer to glutathione-immobilized microbeads. *J. Bicosci. Bioeng*., 119, 137-139, 2015.

Poster presentation:

1. Qingmin Zang, Seiichi Tada, Wei Wang, Mingzhe Liu, Daisuke Kiga, Masayuki Yamamura and Yoshihiro Ito, *In vitro* selection of a photo-responsive peptide aptamer containing an azobenzene residue. 62nd symposium on macromolecules, May, 2013, Kyoto, Japan.
2. Qingmin Zang, Seiichi Tada and Yoshihiro Ito, *In vitro* selection of a photo-responsive peptide aptamer for purifying proteins. Noyori summer school, Sept, 2013, Kobe, Japan.