N-TERMINAL MODIFICATION OF PEPTIDES AND PROTEINS USING 2-ETHYNYLBENZALDEHYDES *VIA* ISOQUINOLINIUM FORMATION

LAI CHUN HIM NATHANAEL

MPhil

The Hong Kong Polytechnic University

2018

The Hong Kong Polytechnic University
Department of Applied Biology and Chemical Technology

N-Terminal Modification of Peptides and Proteins using 2-Ethynylbenzaldehydes *via* Isoquinolinium Formation

Lai Chun Him Nathanael

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy

August 2017

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Lai Chun Him Nathanael

Aug 2017

Abstract

Bioconjugation is an enabling technology for studies of fundamental biological events through selective modification and labeling of biomolecules that is of great application potential in the development of protein therapeutics. To assemble structurally well-defined bioconjugates, it is of importance to develop novel site-selective bioconjugation reactions that are compatible with complex biological systems under mild reaction conditions. However, common approaches for bioconjugation, such as lysine modification, generally give heterogeneous bioconjugate mixtures due to the difficulties in controlling the number of modifications of the multiple lysine residues on the protein surface. Note that the α-amino group of *N*-terminal residues stand out as a unique reactive site rendering site-selective bioconjugation possible, as proteins usually contain one *N*-terminal α-amino residue. Thus, the development of efficient *N*-terminal α-amino group modification reaction is a promising strategy for site-selective bioconjugation.

Along with our long-term interest in selective *N*-terminal bioconjugation, a novel method for highly selective *N*-terminal α-amino group modification of peptides and proteins *via* isoquinolinium formation by using easily accessible 2-ethynylbenzaldehydes was developed in this thesis.

A series of 2-ethynylbenzaldehydes (**2a**-**2q**) were prepared from Sonogashira

coupling reaction and deprotection of trimethylsilyl group in good yields. Treatment of a model peptide YTSSSKNVVR **1a** with 20 equivalents of 2-ethynylbenzaldehyde **2a** under mild reaction conditions (50 mM PBS buffer (pH 6.3) at 37 ^{o}C for 16 h) afforded stable isoquinolinium-peptide conjugates in moderate *N*-terminal selectivity (the ratio of modified *N*-terminal α-amino group/modified lysine ε-amino group = 21:1) analyzed by LC-MS/MS. A model reaction study by treatment of (*S*)-2-amino-*N*-benzylpropanamide **1b** with 2-ethynylbenzaldehyde **2a** afforded the corresponding isoquinolinium product **1ba** in 30% isolated yield, as confirmed by NMR analysis.

Screening of 2-ethynylbenzaldehydes (**2a**-**2q**) with YTSSSKNVVR **1a** gave the corresponding isoquinolinium-peptide conjugates in low to excellent conversion (8-86%) and moderate to excellent *N*-terminal selectivity (12:1 to >99:1). Among the 2-ethynylbenzaldehydes (**2a**-**2q**) examined, reaction of peptide **1a** with 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde **2b** was found to afford good conversion (86%) and excellent *N*-terminal selectivity (>99:1) in the peptide modification. Studies on the reactions of peptide **1a** with **2b** in different reaction media with a range of pH values (pH 5.0-9.0 of 50 mM PBS) revealed that the reaction provided excellent *N*-terminal selectivity (>99:1) in slightly acidic reaction medium (pH 6.3).

Studies of the *N*-terminal selectivity of different *N*-terminal amino acid residues was conducted by treatment of 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde **2b** with a peptide library of XSKFR (X = 20 amino acids). Excellent *N*-terminal selectivity (>99:1) was achieved in 13 out of the 20 peptides with good to excellent conversions (32-93%).

Proteins including insulin, lysozyme, ribonuclease A (RNaseA), bovine serum albumin (BSA), human serum albumin (HSA), and a therapeutic protein BCArg were subjected to the newly developed *N*-terminal bioconjugation reaction at 37 ^{o}C using 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde **2b** to give good conversions (44-71%).

Acknowledgement

I would like to express my deepest gratitude to my supervisor Dr. Wong Man-Kin for offering me an opportunity to further my studies under his precious guidance. I will be forever indebted for his invaluable support. Under his guidance, I have learnt not only chemistry and chemical biology but also gained valuable experiences in other professional aspects. These fruitful experiences have become the treasure and cornerstone for my future. For all of these, I owe my heartfelt appreciation to Dr. Wong.

A special tribute is paid to my undergraduate supervisor, Dr. Chan Lai-Fung Sharon for her gentle comments and encouragement throughout my university years. Dr. Chan showed me what research is, and served as a role model of a professional chemist. Without her, I will not be able to consider graduate studies.

My special thanks go to Prof. Leung Yun-Chung and his teammates for access to the therapeutic protein (BCArg). I would also like to express my appreciation to all the staff and technicians in the Department of Applied Biology and Chemical Technology, particularly to Dr. So Pui-Kin and Dr. Wong Yee-Man Melody for training in mass spectrometry and Dr. Yan Siu-Cheung Kenneth in assistance in nuclear magnetic spectroscopy. In addition, I must thank Ms. Kwok Fung-Yee Peggy, Ms. Tsang Sau-Fung Carol, Ms. Ho Chui-Kwan Joyce, Ms. Ng Yuk-Fan, Miss Pang

Wai-Yin and Miss Wong Wing-Si for their support during my postgraduate studies.

My sincere thanks go to my groupmates in Dr. Wong Man-Kin's group not only for precious guidance in organic synthesis and bioconjugation techniques, but also contributing to the joyful and enlightening environment. Dr. Cui Jian-Fang, Dr. Kung Ka-Yan Karen, Mr. Ko Hok-Ming, Mr. Deng Jie-Ren Kenny, Mr. Bin Yang, Mr. Jia-Jun Jiang and Mr. Yip Wai-Ming are also on my thank you list.

I certainly cannot forget my PolyU friends Dr. Lam Tsz-Lung, Mr. Wong Ho-Yin Chris, Mr. Lau Yan-Fung, Miss Lee Pui-Yu, Mr. Ling Cho-Hon, Mr. Li Chun-Kit, Mr. Sin Ka-Yiu, Mr. Chan Chun-Ming, Miss Xue Min-Ying, Mr. Lai Kin-Ho, Mr. Lam Ning-Fung, Mr. Ho Ping-Luen, Miss Cheng Nga-Man, Mr. Zhang Jun-Hui, Miss Chan Shu-Ting and Miss Cheng Wai Yin; all of them have given me numerous joyful moments.

I am also thankful to have Mr. Goh Zhen-Hao, Mr. Yiu Kin-Chung, Mr Wong Man-Tsun, Mr. Chan Tsz-Wa, Miss Chan Yan-Yee and Miss Ng Miu-Yi as my final year project students. I have gained vital experience in learning to provide appropriate guidance to these dear students.

Most important of all, my gratitude goes to my parents, sisters, beloved grandmother and Jesus Christ for their unconditional love and continuous support that empower me to overcome obstacles during my studies.

In this two-year postgraduate study period, I have definitely acquired knowledge and skills in chemistry; on top of these, I went through a critical journey to remain persistent and determined in attaining my goals. I am forever grateful, proud and honoured to have gone through such fruitful, joyful and irreplaceable experiences.

Content

Chapter 1.

Introduction

1.1 Chemical modifications of peptides and proteins

1.1.1 An introduction of bioconjugation[1]

Protein modification has been one of the paramount research areas in biology. The human proteome has been discovered to be vastly more complex (over 1 million proteins) than human genome (only 20,000-25,000 genes), this disparity is presumably due to post-translational modifications (PTMs), referring to enzymatic covalent modifications after mRNA translation. There are over 200 types of PTMs discovered, including phosphorylation, glycosylation, acetylation, lipidation, and ubiquitination, targeting different residues. PTMs take place in multiple sites of a protein leading to a combinatorial expansion in the number of proteome thereby setting up the foundation for complex biological events.[1]

Inspired by nature's work, scientists have been developing a panel of bioconjugation methods for peptides and proteins that mimic natural modifications or develop novel bioconjugates. These bioconjugation methodologies have been extensively applied in biological studies and drug development, such as biochemical

assays, diagnosis of diseases, *in vitro* and *in vivo* imaging, PEGylation, biomolecule engineering and recently, to develop an emerging novel class of anticancer antibody-drug conjugates that combines the selectivity of targeted treatment with the cytotoxic potency of chemotherapy drugs.[2-6]

Different from PTMs in which multiple modifications may take place in combination within the same biomolecule, chemical protein modification/ bioconjugation aims to append biomolecules of interest with small synthetic molecules (Figure 1.1). The small molecules usually contain two parts: a tag and a reactive group; the reactive group of the small molecules should be site-specific and covalently bonded to the biomolecules, while a tag is responsible to confer desired functions (serves as an affinity probes for protein immobilization to improve the stability of biomolecules and serves as a drug delivery agent).

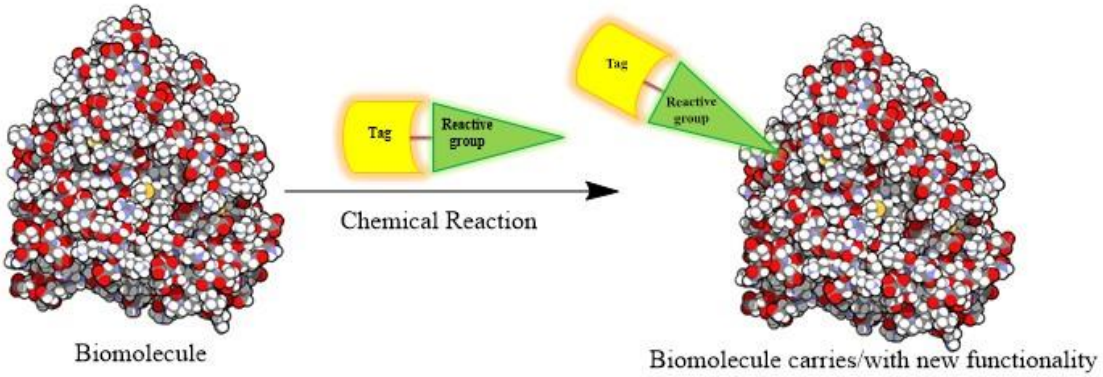


Figure 1.1 A conceptual scheme of bioconjugation.

Bioconjugation reaction development faces two major challenges: reactivity and selectivity. Unlike organic transformations that can be conducted in organic solvents,

high temperature, extreme pH conditions and with protecting groups (e.g. tert-butyloxycarbonyl group), the reaction conditions of bioconjugation are exceptionally stringent: such as good water solubility of bioconjugation reagents, biocompatibility, and reaction kinetics in physiological settings (i.e. ambient reaction temperature, aqueous medium, narrow reaction pH range 6-8, small amount of biomolecules <0.1 mM). The ideal bioconjugations are able to modify the biomolecules effectively without denaturation of bioconjugates under these stringent conditions.

The second challenge is to achieve site-selective protein modification. In view of the chemical reactivity of the 20 canonical amino acid residues, the functional group side chains of arginine, lysine, cysteine, aspartic acid, glutamic acid and tyrosine are potent nucleophiles to react with the bioconjugation agents.[7] In addition, the amino acids exist in multiple sites of a protein. The bioconjugation reagents need to be chemo-selective that can react with one particular amino acid residue effectively, but remains inert to other amino acid residues. As it is difficult to perform purification after protein modifications, development of selective reactions are a paramount issue in bioconjugation chemistry.

In the continuous endeavor of chemists, several approaches of bioconjugation reactions have been developed that target different residues specifically, including

lysine (K), serine (S), threonine (T), cysteine (C), tryptophan (W), histidine (H), tyrosine (Y), arginine (R), aspartic acid (D), glutamic acid (E), methionine (M), proline (P), *N*-terminal and C-terminal residues, as well as some sequence-selective approaches and proximity-driven modification.[8]

1.1.2 Targeting the lysine ε-amino group[8]

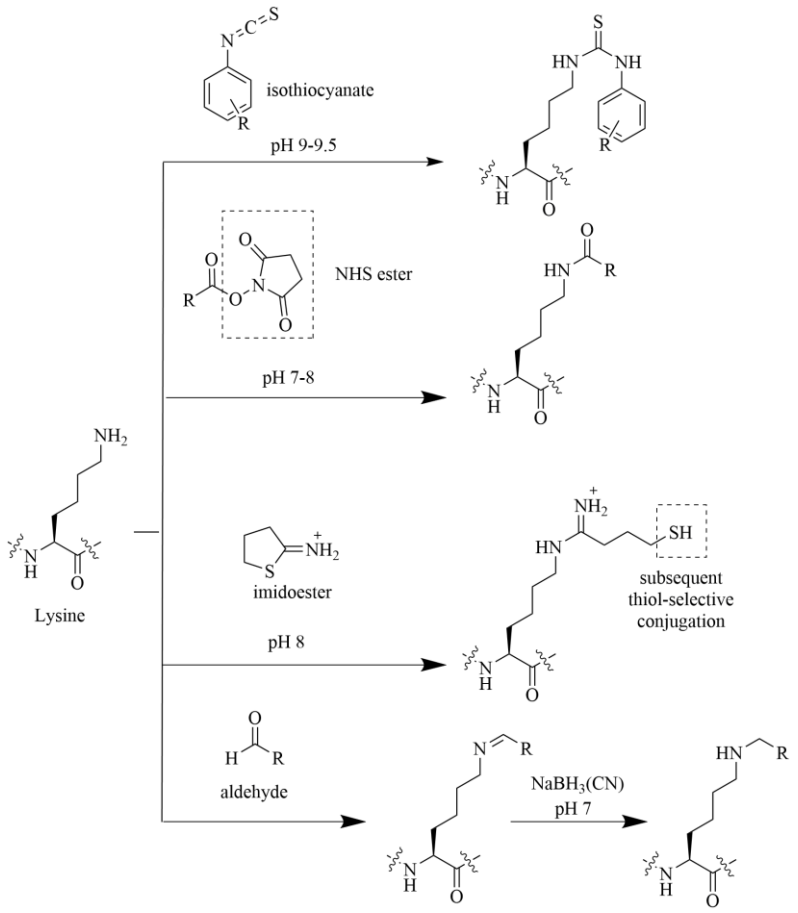
1.1.2.1 Classical methods for lysine modification

The use of chemical reagents that target the primary amines is one of the most commonly used methods for protein bioconjugation. This can be attributed to two reasons: Firstly, the p*K*a of the lysine ε-amino group is about 10 which in a vast majority of cases they are protonated at physiological pH. Consequently, the lysine residues present predominantly on the solvent-exposed surface render them an easy target for bioconjugation agents to append.[8] Secondly, lysine is of high abundance in proteins (5.2~7.2%),[9] so that lysine-target bioconjugations can be easily applied to a majority of proteins.

As lysine residues protonated at physiological pH drastically decrease their reactivity, the bioconjugations are usually performed in higher pH value (pH of 8.5-9.5) so that their nucleophilicity retains.[8] It is also worth noting that at higher pH lysine bioconjugations are more favorable than that of the *N*-terminal α-amino group

(p*K*a~8) as the lysine side chain amino groups are generally more reactive towards electrophiles.

With fine-tuning of the reaction conditions, selective modification of lysine ε-amino group can be achieved by using various chemical reagents. Classical agents used in lysine bioconjugation usually belong to one of the following classes: activated esters, isothiocyanates, isocyanates, aldehydes, epoxides and imidoesters (Scheme 1.1).[10-13] Among these reagents, N-hydroxysuccinimide esters (NHS esters) and isothiocyanates represent the most common agents that are commercially available for protein bioconjugation.



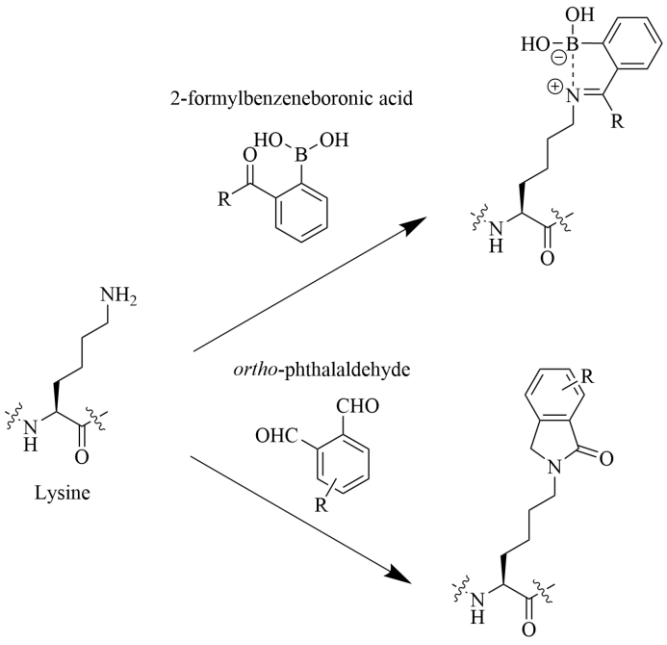
Scheme 1.1 Classical methods for lysine modification.

Unlike isocyanates and isothiocyanates that may suffer from deterioration in the presence of other nucleophiles, such as water and alcohols and form the corresponding carbamates and amines,[12] NHS esters possess faster reaction kinetics to give more stable conjugates for both peptides and proteins, and are able to proceed at physiological pH conditions (NHS esters vs isothiocyanates: pH 7-8 vs pH 9-9.5), which enable the conjugation of alkaline-sensitive proteins. NHS is considered as an "activated reagent" which forms a NHS-activated ester with the corresponding carboxylic acid. The NHS ester formed is a good leaving group that the carboxylic acid is activated and susceptible to nucleophilic attack.

Virtually any molecule containing an acid functionality can be converted into its NHS esters, in which such excellent substrate tolerance is another reason for the popularity of NHS ester-based conjugation. However, upon activation of carboxylic acids with NHS, the water solubility decreases. The utilization of sulfo-NHS to enhance the water solubility of the modified molecules by virtue of the charged sulfonate group is needed.[14] Note that the field of NHS-mediated conjugation of proteins has been continuously expanding its utility in numerous applications.

1.1.2.2 Other methods for lysine modification

In the past decades, some new methods have been developed for lysine modification. In 2012, Gois and co-workers have developed a new strategy for reversible lysine and *N*-terminal modification *via* iminoboronates formation (Scheme 1.2).[15a] The imine formed from 2-formylphenylboronic acid possesses good stability towards hydrolysis, which is presumably due to the formation of a dative N-B bond. In addition, this method has been demonstrated successfully in some well-known proteins such as lysozyme, RNaseA and myoglobin. Recently, Li and co-workers reported that *ortho*-phthalaldehyde (OPA) and its derivatives react chemoselectively with lysine residues in proteins *via* phthalimidine formation (Scheme 1.2).[15b] The authors also demonstrated an OPA-based PEGylation on asparaginase with 47% retained enzymatic activity.



Scheme 1.2 Lysine modifications developed by Gois's and Li's research groups.

1.1.3 Targeting the *N*-terminal α-amino group[16]

Despite successful cases were obtained in lysine modification, the high abundance of lysine residues on the surface of proteins may result in multi-site lysine modifications and form mixtures of modified proteins. The problematic consequences of the heterogeneous products are that the bioconjugates cannot be well-characterized and possess inconsistent properties. Therefore, it is of importance to control the site-selectivity of modification in bioconjugation.[17]

α-Amino group of *N*-terminal residues stand out as a unique reactive site which render specific selectivity in bioconjugation possible. The p*K*a of *N*-terminal α-amino group ranges from 6 to 8, whereas that of the lysine ε-amino group is 10. In addition, the *N*-terminal residues are usually exposed on the surface of proteins (over 80%).[18] Therefore, it is of interest to develop a *N*-terminal selective bioconjugation.

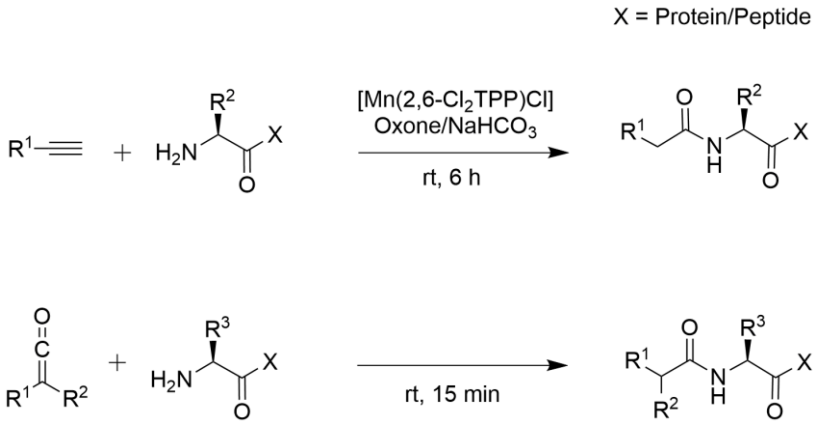
1.1.3.1 *N*-terminal modification using pH control

It is noteworthy to highlight that the bioconjugation reagents used for lysine modifications can be tuned to *N*-terminal selective by judicious pH control, usually under slightly acidic conditions (below pH 7), since the p*K*a value of the *N*-terminal α-amino group and that of the lysine ε-amino group are quite different. For example, the ratio of modified α-amino group to modified ε-amino group was found to be 3 : 1 by using the commonly employed N-hydroxysuccinimide ester (NHS ester) at pH

6.5.[19] Generally, a lower pH (down to pH 4.5) offers a higher ratio of *N*-terminal selectivity, but often at the cost of decreased conjugation yields. Thus, it is of importance to develop new methods for *N*-terminal selective bioconjugation.

1.1.3.2 N-terminal modification *via* oxidative amidation

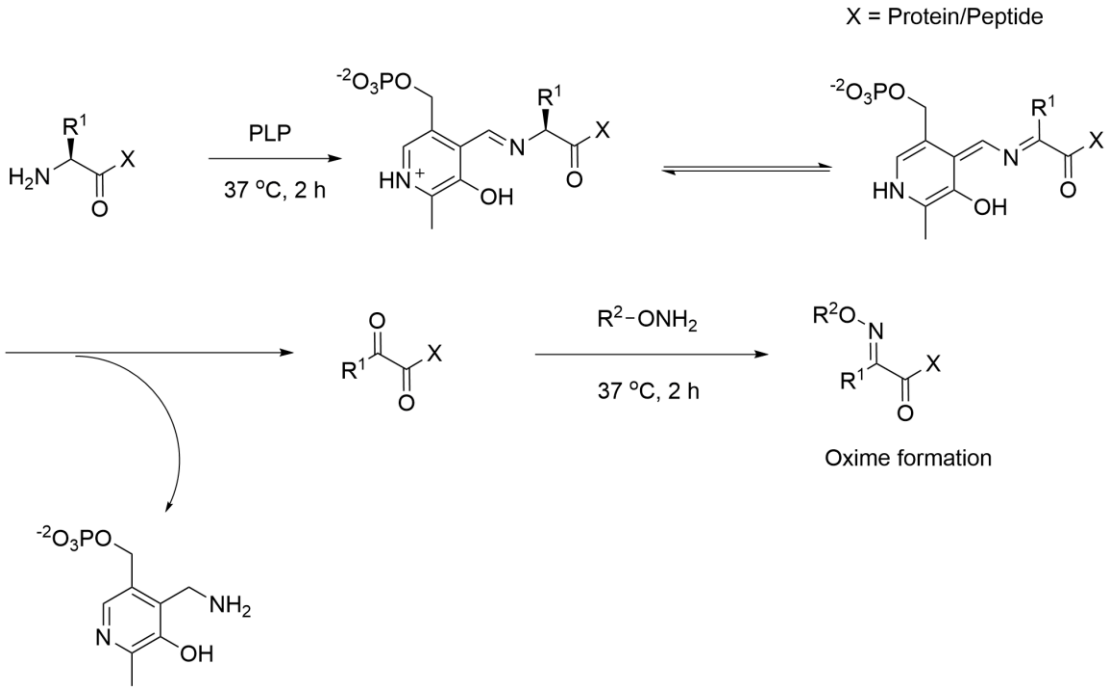
Che and our group discovered an oxidative amidation of *N*-terminal α-amino group of peptides with alkyne moieties catalyzed by [Mn(2,6-Cl_2TPPCl)] with Oxone as a terminal oxidant. This protocol was used for *N*-terminal modification of peptides with aryl, aliphatic, and internal alkynes under mild conditions.[20] The follow up studies indicated that ketenes *in situ* generated were the key intermediate for *N*-terminal modification. Accordingly, ketenes could be prepared directly and have been demonstrated successfully in protein modifications (Scheme 1.3).[20]



Scheme 1.3 (Top) *N*-terminal modification *via* oxidative amidation; (Bottom) *N*-terminal modification using ketenes.

1.1.3.3 *N*-terminal transamination and oxime ligation

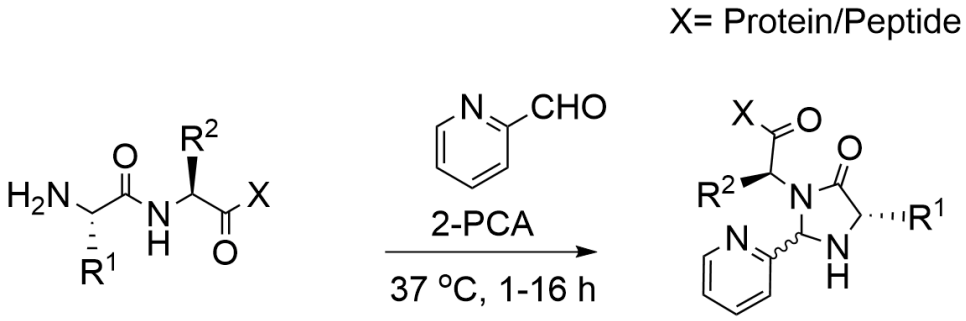
Another *N*-terminal modification approach is based on transamination. Although transamination reactions were discovered more than 30 years ago, the application on protein modifications were limited by its harsh conditions.[21] Francis and co-workers discovered a PLP-mediated transamination that allows the generation of carbonyl-containing products, that can be further functionalized *via* oxime formation (Scheme 1.4).[22] Despite the high efficiency and *N*-terminal selectivity towards a panel of peptides and proteins, this protocol was found to be ineffective in some *N*-terminus with residues such as histidine (H), isoleucine (I), lysine (K), glutamine (Q), tryptophan (W), proline (P), valine (V), and phenylalanine (F).[22]



Scheme 1.4 PLP-mediated *N*-terminal modification *via* oxime formation.

1.1.3.4 *N*-terminal modification *via* imidazolidinone formation

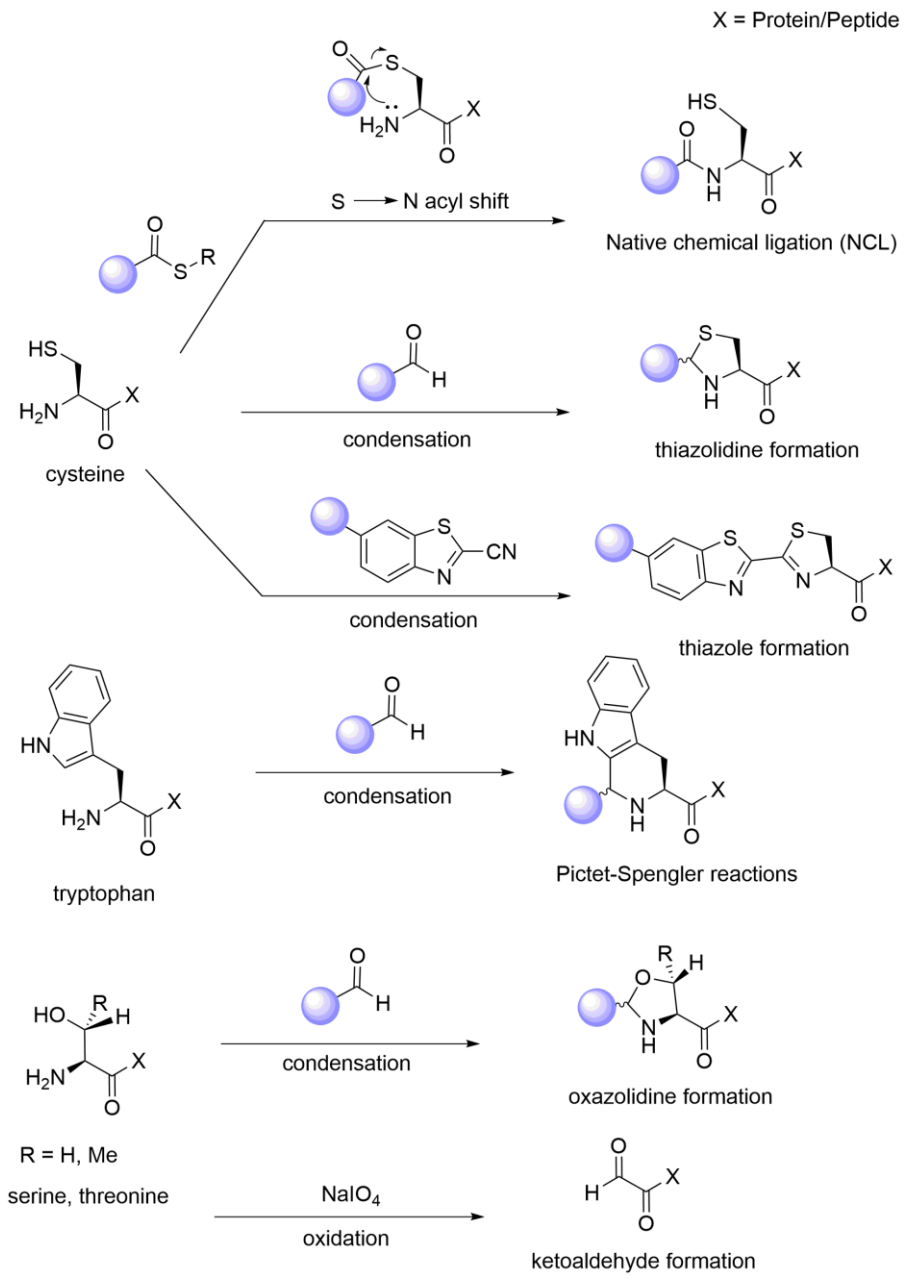
In addition, Francis and co-workers found that 2-pyridine-carboxyaldehydes (2-PCA) could modify the *N*-terminus of proteins and peptides in a one-step formation of cyclic imidazolidinone product (Scheme 1.5). This protocol generally achieves good conversion over 20 canonical amino acid residues at the *N*-terminus and notably, can also serve as a site-selective heterobifunctional cross-linker after coupling with several commercially available bioconjugation agents.[23]



Scheme 1.5 One-step *N*-terminal modification using 2-PCA.

1.1.3.5 *N*-terminal modification by side chain participation

Apart from 2-PCA that possesses a general cyclization with the amide NH group of the second amino acid residues, there are some methods involving particular side chain for cyclic product formation. Combined with the unique reactivity of *N*-terminal α-amino group, these methods are highly chemoselective and serve as a versatile tool for protein labeling purposes (Scheme 1.6).[24-25]



Scheme 1.6 Modification of specific amino acids at protein *N*-termini.

Native chemical ligation (NCL) is a renowned example of *N*-terminal modification by side chain participation, in which transthioesterification was involved, followed by a spontaneous intramolecular S → N acyl transfer. It is noteworthy to highlight that *N*-terminal cysteines are rare in nature in which they often need to be introduced by genetic engineering.[24] Other cyclization products such as thiazolidine,

oxazolidine and Pictet-Spengler reaction are formed *via* typical condensation reaction. However, these products are prone to pH-dependent hydrolysis.[25]

1.1.3.6 Challenges in *N*-terminal modification of peptides and proteins

The major challenge of *N*-terminal peptide and protein modification is to well differentiate the *N*-terminal α-amino group from the lysine ε-amino group. It is known that the p*K*a value of *N*-terminal amine (6-8) is lower than that of the lysine amine (10) (Figure 1.2). However, the high abundance of lysine residues on the surface of proteins may result in lysine modifications during the bioconjugation, making *N*-terminal specific modification a difficult task.

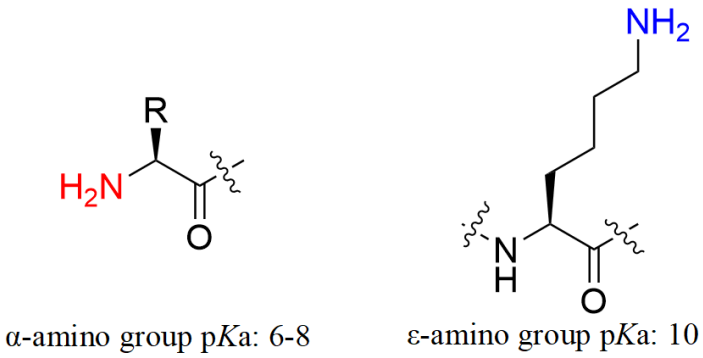


Figure 1.2 The p*K*a of *N*-terminal α-amino group (left) and lysine ε-amino group (right).

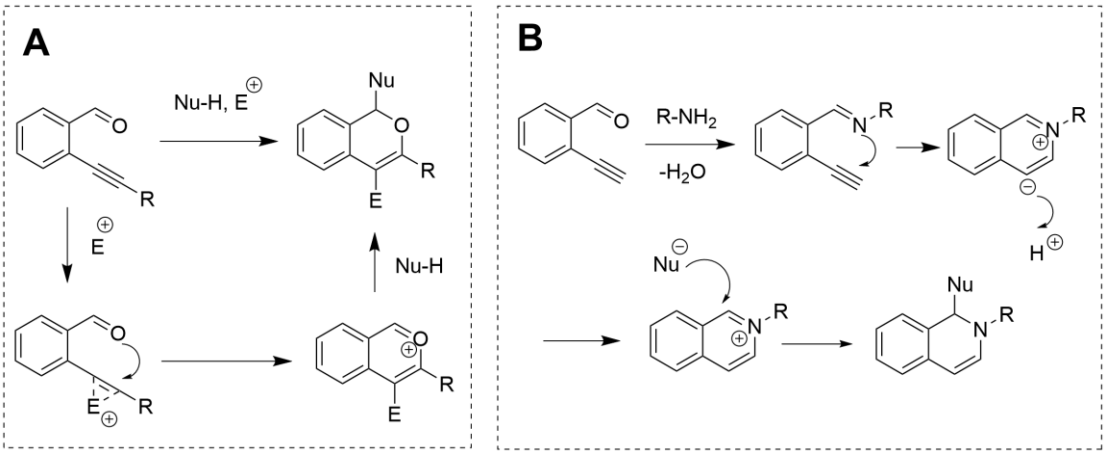
The current regiment of *N*-terminal selective modifications usually bases on several strategies: selective *N*-terminal modifications *via* pH control,[19] *N*-terminal modifications by side chain participation,[24-25] *N*-terminal oxidation[20, 24-25] and

N-terminal modification *via* heterocycles formation.[23] However, only low *N*-terminal selectivity (modified α-amino group vs ε-amino group, 3:1) was obtained using the commonly employed N-hydroxysuccinimide ester (NHS ester) at pH 6.5.[19] Other cyclization products such as thiazoline, are prone to pH-dependent hydrolysis or require certain side chains for cyclic product formation. The limited sets of easily-accessible bioorthgonal reaction underscore the demand for new contributions.

1.2 Chemistry of 2-ethynylbenzaldehydes

2-Ethynylbenzaldehyde is one of the most frequently used building blocks in the synthesis of heterocyclic compounds, including iodoisoquinoline-fused benzimidazoles [26] and isoquinolines.[27] The 2-ethynylbenzaldehydes are good electrophilic cyclization species for heterocyclic synthesis, either *via* direct cyclization,[28] or by the nucleophilic attack of the imine nitrogen to the alkyne (Scheme 1.7).[29]

These reactions are usually performed under mild conditions [28-29] (room temperature and open atmosphere) with good yields, indicating that 2-ethynylbenzaldehydes are excellent building blocks for heterocyclic synthesis.



Scheme 1.7 (A) Direct cyclization of 2-ethynylbenzaldehyde;[28] (B) Cyclization *via* nucleophilic attack of the imine nitrogen to alkyne.[29]

1.3 Synthesis and transformation of isoquinoliniums and isoquinolines

As one of the N-heterocyclic quaternary ammonium salts, isoquinolinium compounds are found in the core structures of many bioactive natural compounds, pharmaceuticals and organic compounds (Figure 1.3).[30] Their neutral form (i.e. isoquinolines) also serve similar functions, therefore, chemists have been developing efficient chemical reactions to synthesize various isoquinoliniums and isoquinolines.

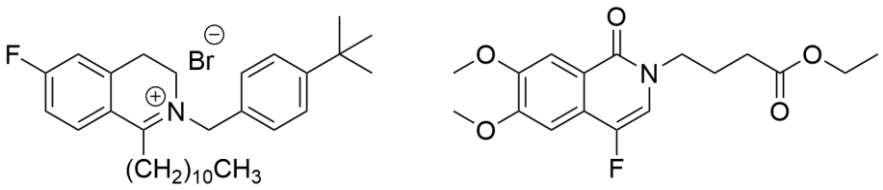
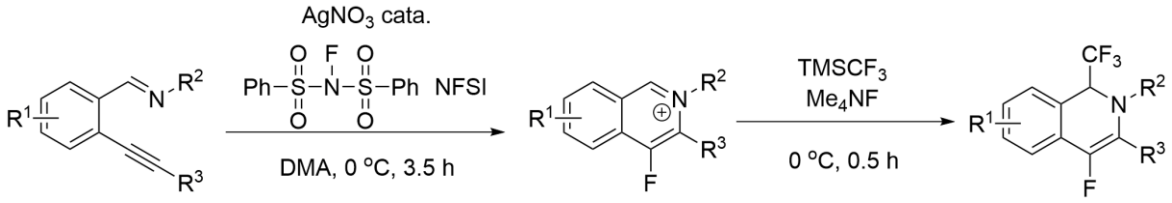


Figure 1.3 (Left) An isoquinolinium-containing inhibitor of the kinases of MAPK cascades in eukaryotes;[31] (Right) An isoquinoline-containing tumor necrosis factor.[32]

1.3.1 Silver-catalyzed synthesis of isoquinoliniums and isoquinolines

Although the synthesis of isoquinoliniums and isoquinolines were discovered several decades ago,[33] there are only few efficient synthetic methods available in literature. These methods generally involve multiple steps, and stoichiometric or catalytic amounts of noble-metal catalysts are needed.[34] In 2013, Liu reported a cascade approach for the synthesis of fluorinated isoquinolines *via* a silver-catalyzed intramolecular aminofluorination of alkynes (Scheme 1.8).[35]



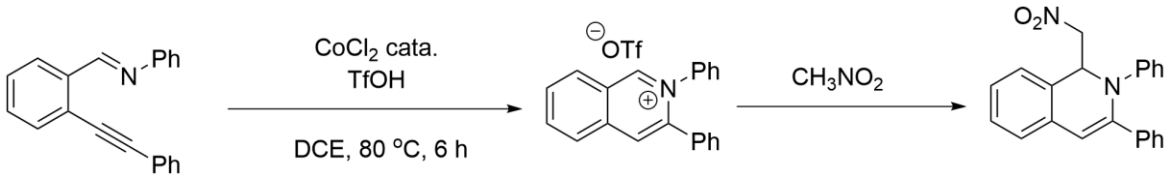
Scheme 1.8 Reaction scheme for a silver-catalyzed isoquinoliniums and isoquinolines synthesis.

The isoquinoliniums were first achieved *via* a silver-catalyzed aminofluorination, using N-fluorobenzenesulfonimide (NFSI) as a fluorine source. The intermediate can be further functionalized by subsequent nucleophilic addition to form isoquinolines in one pot.

1.3.2 Cobalt-catalyzed synthesis of isoquinoliniums and isoquinolines

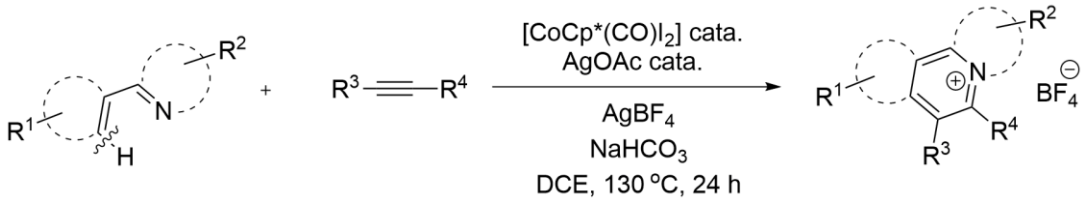
The current regiment of isoquinolinium and isoquinoline synthesis usually relies on the use of the third row transition metal catalysts in which the synthetic cost is

considerably high. Recently, there are some new methods using inexpensive metal-catalyzed synthesis of isoquinoliniums and isoquinolines reported. In 2015, Tandon reported that 6-*endo-dig* ring closure reactions of 2-(1-alkynyl) arene carboxyaldehyde imines can be achieved in the presence of $CoCl_2$ (in which cobalt has been known to activate alkynes) and form isoquinoliniums.[36] Further functionalization with nitro group can be achieved by Mannich condensation to form isoquinolines (Scheme 1.9).



Scheme 1.9 A $CoCl_2$ catalyzed isoquinolinium and isoquinoline formation.

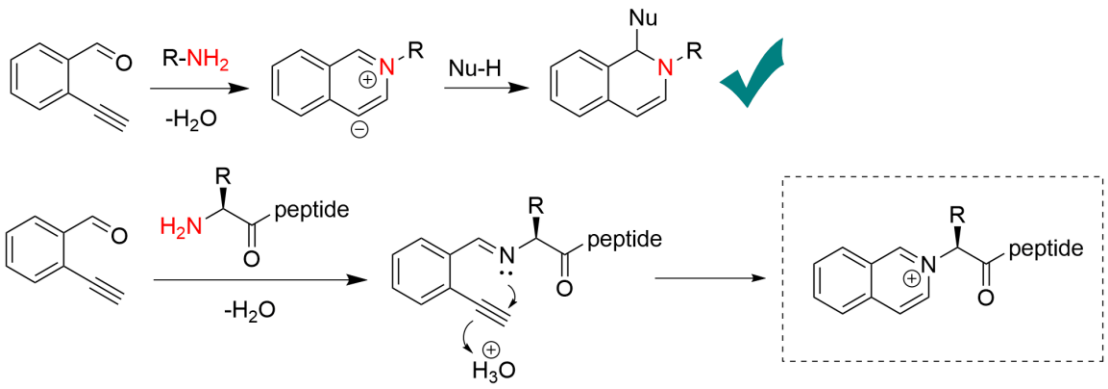
Another cobalt-catalyzed isoquinolinium formation was reported by Cheng's group. Different from the aforementioned method, the product was formed *via* oxidative annulation of nitrogen-containing arenes with alkynes, which allows a facile and more straightforward synthesis of isoquinoliniums using only a one-step C-H activation (Scheme 1.10).[37]



Scheme 1.10 Isoquinolinium formation *via* cobalt-catalyzed oxidative annulation of nitrogen-containing arenes with alkynes.

1.4 2-Ethynylbenzaldehydes as a bioconjugation reagent

2-Ethynylbenzaldehydes have been demonstrated as a good starting material for isoquinoline synthesis. The 2-ethynylbenzaldehydes were usually first reacted with amines and converted into the corresponding imines, followed by cyclization to form isoquinolines (Scheme 1.11). In this regard, we hypothesize that 2-ethynylbenzaldehydes would have the similar reactivity towards the amino group of peptides and proteins. We postulate that the imines formed between 2-ethynylbenzaldehydes and the primary amino group of peptides, will undergo nucleophilic attack to the alkyne and subsequent protonation to form a stable aromatic product isoquinolinium. The *N*-terminal selectivity (the ratio of modified *N*-terminal α-amino group/modified lysine ε-amino group) of the isoquinolinium-peptide conjugates would be fine-tuned through judicious pH control.



Scheme 1.11 (Top) Reaction between 2-ethynylbenzaldehye and amine in organic synthesis; [29] (Bottom) Reaction between 2-ethynylbenzaldehye and peptide.

Reference

1. Kalia, J.; Raines, R. T. *Curr Org Chem*., **2010**, *2*, 138.

2. (a) Bokinsky, G.; Zhuang, X. *Acc. Chem. Res*., **2005**, *38*, 566; (b) Schuler, B.; Eaton, W. A. *Curr. Opin. Struct. Biol*., **2008**, *18*, 16; (c) Smith, G. J.; Lee, K. T.; Qu, X.; Xie, Z.; Pesic, J.; Sosnick, T. R.; Pan, T.; Scherer, N. F. *J. Mol. Biol*., **2008**, *378*, 941.

3. Jares-Erijman, E. A.; Jovin, T. M. *Nat. Biotechnol*., **2003**, *21*, 1387.

4. (a) Zalipsky, S. *Bioconjugate Chem*., **1995**, *6*, 150; (b) Roberts, M. J.; Bentley, M. D.; Harris, J. M. *Adv. Drug. Deliv. Rev*., **2002**, *54*, 459.

5. Chapman, A. P. *Adv. Drug. Deliv. Rev*., **2002**, *54*, 531.

6. Diamantis, N.; Banerji, U. *Br. J. Cancer*., **2016**, *4*, 362.

7. (a) Hermanson, G. T. *Bioconjugate Techniques. 2nd ed*., Academic Press, Cambridge, 2008; (b) Bischoff, R.; Schluter, H. *J Proteomics*, **2012**, *75*, 2275.

8. Koniev, O.; Wagner, A. *Chem. Soc. Rev*., **2015**, *44*, 5495.

9. (a) King, J. L.; Jukes, T. H. *Science*, **1969**, *164*, 788; (b) Gilis, D.; Massar, S.; Cerf, N. J.; Rooman, M. *Genome Biol*., **2001**, *2*, 1; (c) Fodje, M. N.; Al-Karadaghi, S. *Protein Eng.*, **2002**, *15*, 353.

10. (a) Anderson, G. W.; Zimmerman, J. E.; Callahan, F. M. *J. Am. Chem. Soc*., **1964**, *86*, 1839; (b) Lowe, A. *Proc. R. Soc. Med*., **1970**, *63*, 367.

11. Annunziato, M. E.; Patel, U. S.; Ranade, M.; Palumbo, P. S. *Bioconjugate Chem*., **1993**, *4*, 212.

12. Coons, A. H.; Creech, H. J.; Jones, R. N.; Berliner, E. *J. Immunol*., **1942**, *45*, 159.

13. Rifai, A.; Wong, S. S. *J. Immunol. Methods*, **1986**, *94*, 25.

14. Staros, J. V.; Wright, R. W.; Swingle, D. M. *Anal. Biochem*., **1986**, *156*, 220.

15. (a) Cal, P. M. S. D.; Vicente, J. B.; Pires, E.; Coelho, A. V.; Veiros, L. F.; Cordeiro, C.; Gois, P. M. P. *J. Am. Chem. Soc*., **2012**, *134*, 10299; (b) Tung, C. L.; Wong, C. T. T.; Fung, E. Y. M.; Li, X. *Org. Lett*., **2016**, *18*, 2600.

16. Rosen, C. B.; Francis, M. B. *Nat. Chem. Bio*., **2017**, *13*, 697.

17. Kazane, S. A.; Sok, D.; Cho, E. H.; Uson, M. L.; Kuhn, P.; Schultz, P. G.; Smider, V. V. *Proc. Natl. Acad. Sci.* **2012**, *109*, 3731.

18. Jacob, E.; Unger, R. *Bioinformatics*, **2007**, *23*, e225.

19. Selo, I.; Negroni, L.; Creminon, C.; Grassi, J.; Wa, J. M. *J. Immunol. Methods*, **1996**, *199*, 127.

20. (a) Chan, W.-K.; Ho, C.-M.; Wong, M.-K.; Che, C.-M. *J. Am. Chem. Soc.*, **2006**, *128*, 14796; (b) Chan, A. O.-Y.; Ho, C.-M.; Chong, H.-C.; Leung, Y.-C.; Huang, J.-S.; Wong, M.-K.; Che, C.-M. *J. Am. Chem. Soc*., **2012**, *134*, 2589.

21. Dixon, H.B. F. *J. Protein Chem*., **1984**, *3*, 99.

22. (a) Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. *Angew. Chem., Int. Ed*., **2006**, *45*, 5307; (b) Scheck, R. A.; Dedeo, M. T.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc*., **2008**, *130*, 11762.

23. (a) MacDonalde, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nat. Chem. Bio*., **2015**, *11*, 326; (b) Lee, J. P.; Kassianidou, E.; MacDonald, J. I.; Francis, M. B.; Kumar, S. *Biomaterials*, **2016**, *102*, 268; (c) ElSohly, E. M.; MacDonald, J. I.; Hentzen, N. B.; Aanei, I. L.; EI Muslemany, K. M.; Francis, M. B. *J. Am. Chem. Soc*., **2017**, *139*, 3767.

24. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science*, **1994**, *266*, 776.

25. (a) Geoghegan, K. F.; Stroh, J. G. *Bioconjug. Chem*., **1992**, *3*, 138; (b) Chen, J. K.; Lane, W. S.; Brauer, A. W.; Tanaka, A.; Schreiber, S. L. *J. Am. Chem. Soc*., **1993**, *115*, 12592; (c) Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. *Tetrahedron Lett*., **2000**, *41*, 4069.

26. Ouyang, H. C.; Tang, R. Y.; Zhong, P.; Zhang, X. G.; Li, J. H. *J. Org. Chem*., **2011**, *76*, 223.

27. Tokimizu, Y.; Ohta, Y.; Chiba, H.; Oishi, S.; Fujii, N.; Ohno, H. *Tetrahedron*, **2011**, *29*, 5168.

28. (a) Yue, D. W.; Ca, N. D.; Larock, R. C. *Org. Lett*., **2004**, *6*, 1581; (b) Yue, D. W.;

Ca, N. D.; Larock, R. C. *J. Org. Chem*., **2006**, *71*, 3381.

29. Asao, N.; Iso, K.; Salprima, Y. S. *Org. Lett*., **2006**, *8*, 4149.

30. (a) Korivi, R. P.; Cheng, C. H. *Chem. Eur. J*., **2010**, *16*, 282; (b) Facchetti, A.; Abbotto, A.; Beverina, L.; van der Boom, M. E.; Dutta, P.; Evmenenko, G.; Marks, T .J.; Pagani, G. A. *Chem. Mater*., **2002**, *14*, 4996; (c) Seddon, K. R. *Nat. Mater*., **2003**, *2*, 363; (d) Krane, B. D.; Fagbule, M. O.; Shamma, M. *J. Nat. Prod*., **1984**, *47*, 1.

31. Kim, H. J.; Park, J. E.; Jin, S. W.; Kim, J. H.; Song, K. W. *Chem. Biol.*, **2006**, *13*, 881.

32. Chao, Q.; Deng, L.; Shih, H.; Leoni, L. M.; Genini, D.; Carson, D. A.; Cottam, H. B. *J. Med. Chem*., **1999**, *42*, 3860.

33. (a) Kornfeld, E. C. *J. Am. Chem. Soc*., **1948**, *70*, 1373; (b) Boekelheide, V.; Gall, W. G. *J. Am. Chem. Soc*., **1954**, *76*, 1832; (c) Wu, G.; Rheingold, A. L.; Geib, S. J.; Heck, R. F. *Organometallics*, **1987**, *6*, 1941; (d) Wu, G.; Geib, S. J.; Rheingold, A. L.; Heck, R. F. *J. Org. Chem*., **1988**, *53*, 3238.

34. (a) Li, L.; Brennessel, W. W.; Jones, W. D. *J. Am. Chem. Soc*., **2008**, *130*, 12414; (b) Korivi, R. P.; Wu, Y. C.; Cheng, C. H. *Chem. Eur. J*., **2009**, *15*, 10727.

35. Liu, Q. L.; Wu, Y. C.; Chen, P. H.; Liu, G. S. *Org. Lett*., **2013**, *15*, 6210.

36. Urvashi, Rastogi, G. K.; Ginotra, S. K.; Agarwal, A.; Tandon, V. *Org. Biomol.*

Chem., **2015**, *13*, 1000.

37. Prakash, S.; Muralirajan, K.; Cheng, C. H. *Angew. Chem., Int. Ed*., **2016**, *55*, 1844.

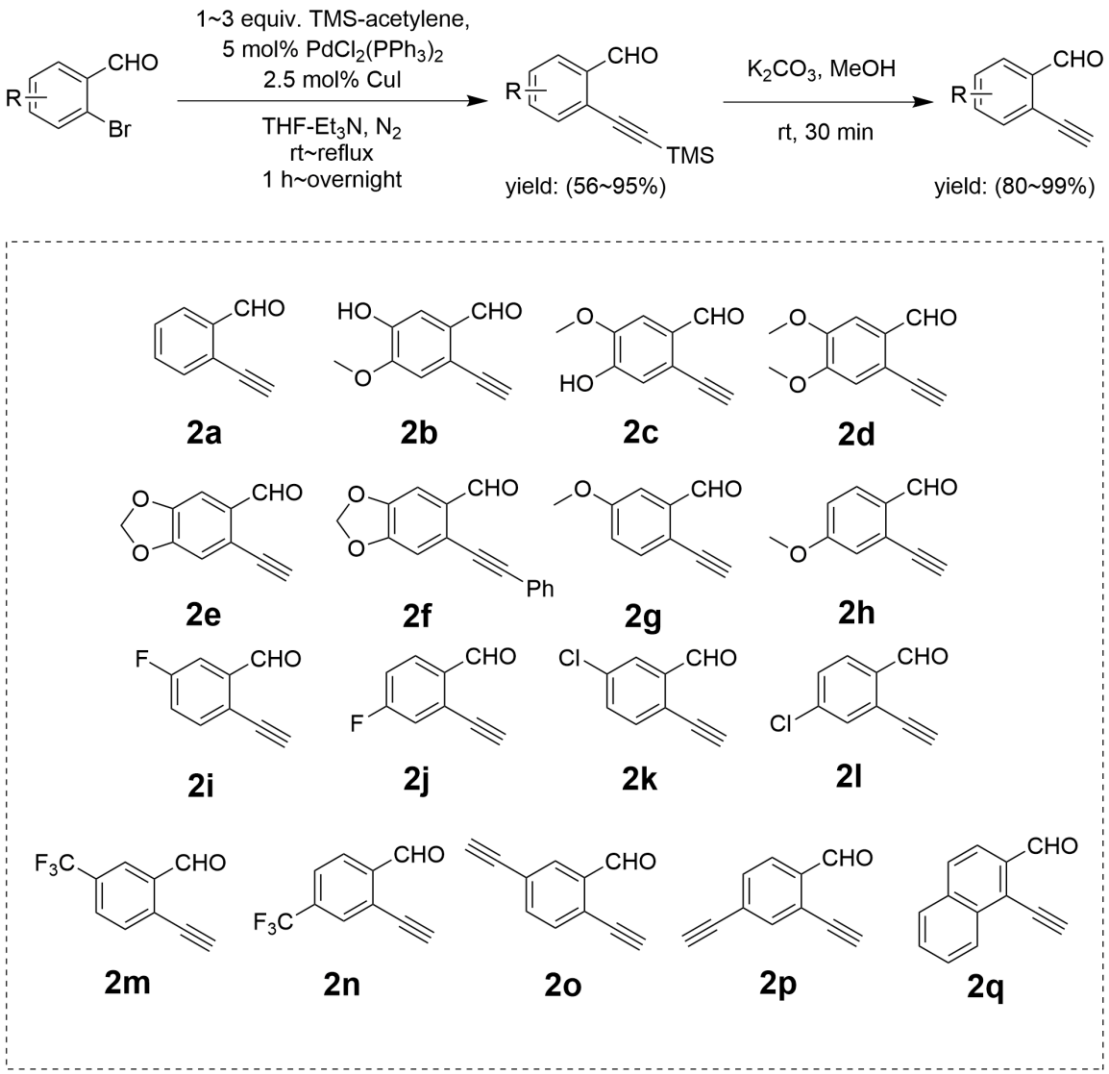
Chapter 2.

N-Terminal Modification of Peptides using 2-Ethynylbenzaldehydes

2.1 Synthesis of 2-ethynylbenzaldehydes

Easy accessibility to structurally diverse bioconjugation agents is one of the most crucial factors for practical application. For example, by using commercially available reagent N-hydroxysuccinimide, different NHS esters can be prepared from their corresponding acid functionality. The preparation of NHS ester-based reagents is facile; therefore, they are frequently used in bioconjugation.

It is noteworthy to point out that 2-ethynylbenzaldehydes are easily accessible, which can be prepared from the corresponding aromatic halides in good to excellent yields. The preparation of 2-ethynylbenzaldehydes **2a-2q** can be achieved by employing commercially available aromatic halides with Sonogashira coupling reaction, followed by deprotection of the trimethylsilane group (Scheme 2.1).[1]

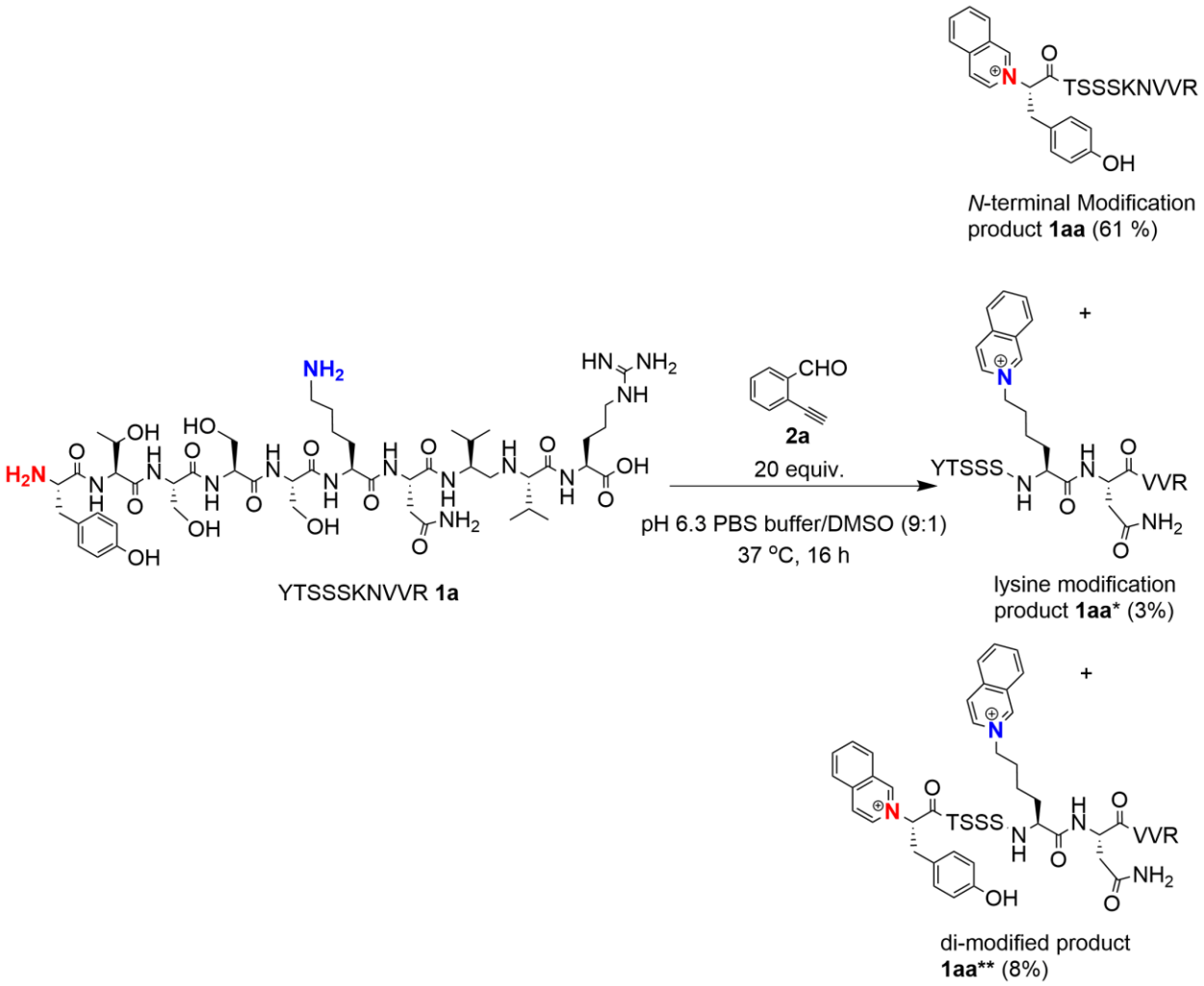


Scheme 2.1 (Top) General Procedure for preparation of 2-ethynylbenzaldehydes; (Bottom) 2-Ethynylbenaldehydes **2a**-**2q** for peptide modification.

2.2 Modification of peptides using 2-ethynylbenzaldehydes

2.2.1 Modification of peptide YTSSSKNVVR 1a using 2-ethynylbenzaldehyde 2a

To examine if 2-ethynylbenzaldehye **2a** would function as a potent *N*-terminal selective bioconjugation reagent, a study was conducted by treating a lysine-containing peptide YTSSSKNVVR **1a** (0.1 mM) with 2-ethynylbenzaldehyde **2a** (2 mM, 20 equiv.) in 100 μL of pH 6.3 50 mM PBS buffer/DMSO (9:1) solution at 37 ^{o}C for 16 h to give an isoquinolinium-modified YTSSSKNVVR conjugate (Scheme 2.2), affording moderate peptide conversion (72%), including *N*-terminal modified peptide **1aa** (61%) and moderate *N*-terminal selectivity (ratio of α-amino- to ε-amino-modified peptide = 21:1), analyzed by LC-MS/MS (Figure 2.1).



Scheme 2.2 Peptide modification using 2-ethynylbenzaldehyde **2a**.

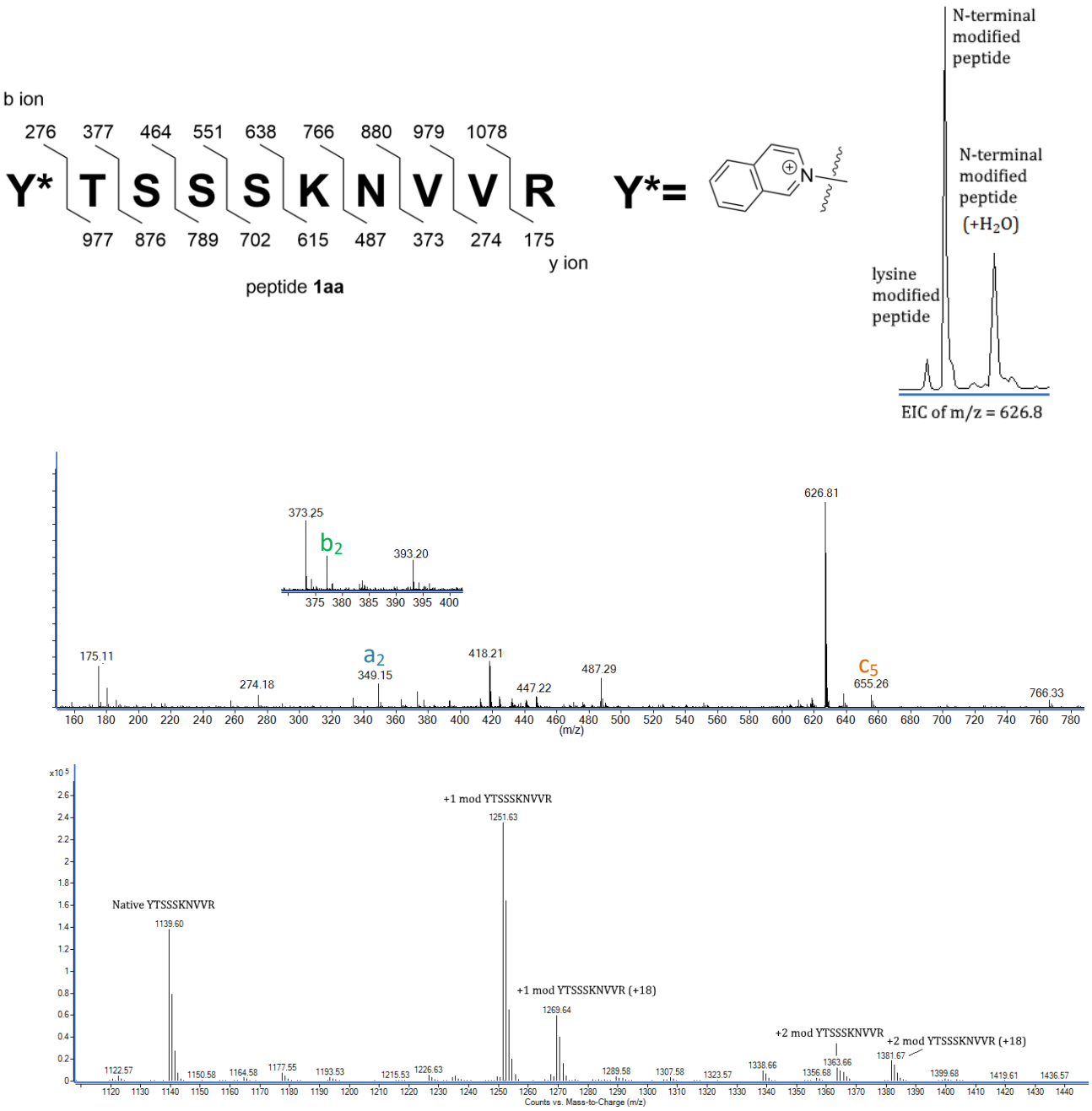
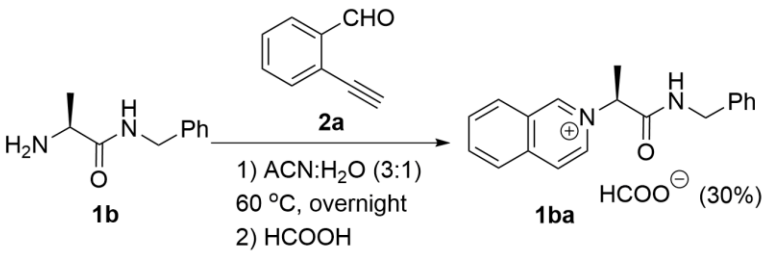


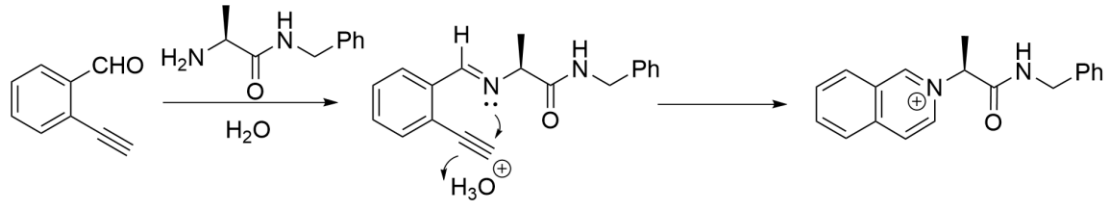
Figure 2.1 (Top) Q-TOF MS/MS spectrum of ***N*-terminal 2a-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 626.8); (Bottom) Deconvoluted spectrum of **2a-modified YTSSSKNVVR**.

2.2.2 Model reaction of isoquinolinium formation

In a model study using (*S*)-2-amino-N-benzylpropanamide **1b**, treatment with 2-ethynylbenzaldehyde **2a** showed the formation of isoquinolinium product **1ba**, as confirmed using ^{1}H NMR spectroscopy, ^{13}C NMR spectroscopy and ESI-MS (Scheme 2.3). A mechanism of the isoquinolinium formation is proposed.



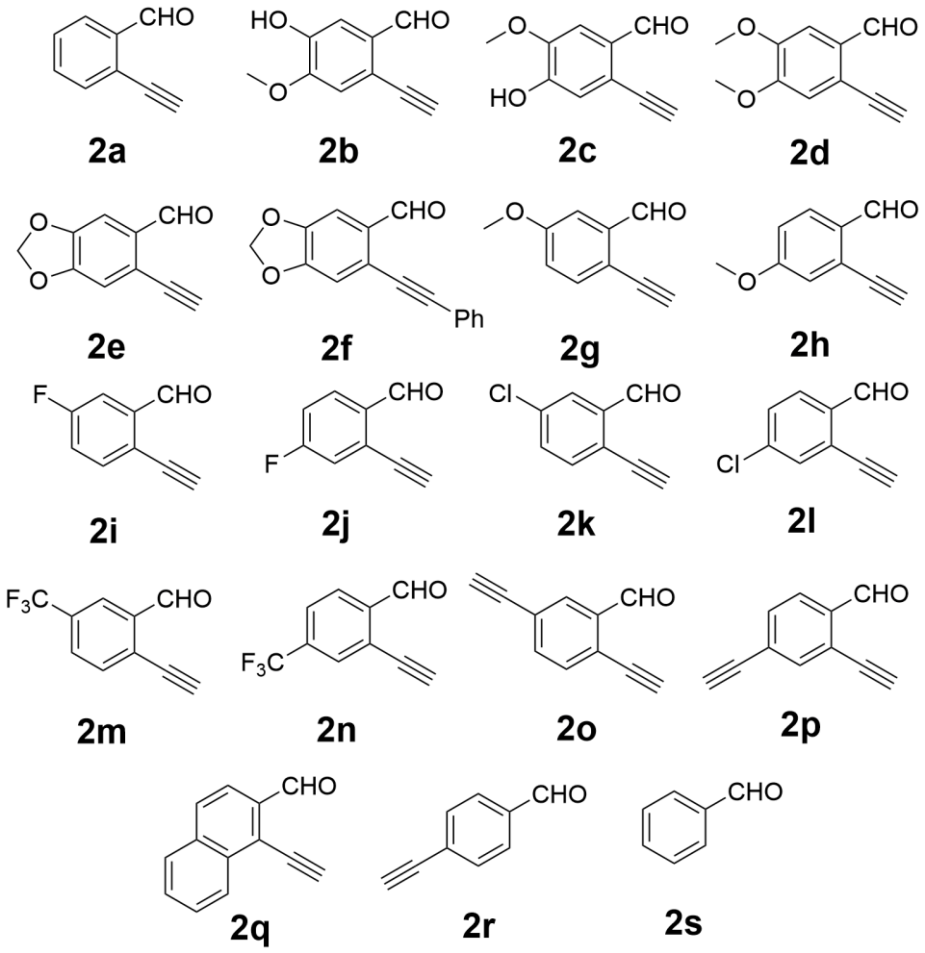
Proposed mechanism:



Scheme 2.3 (Top) Model reaction of isoquinolinium formation; (Bottom) Proposed mechanism for isoquinolinium formation.

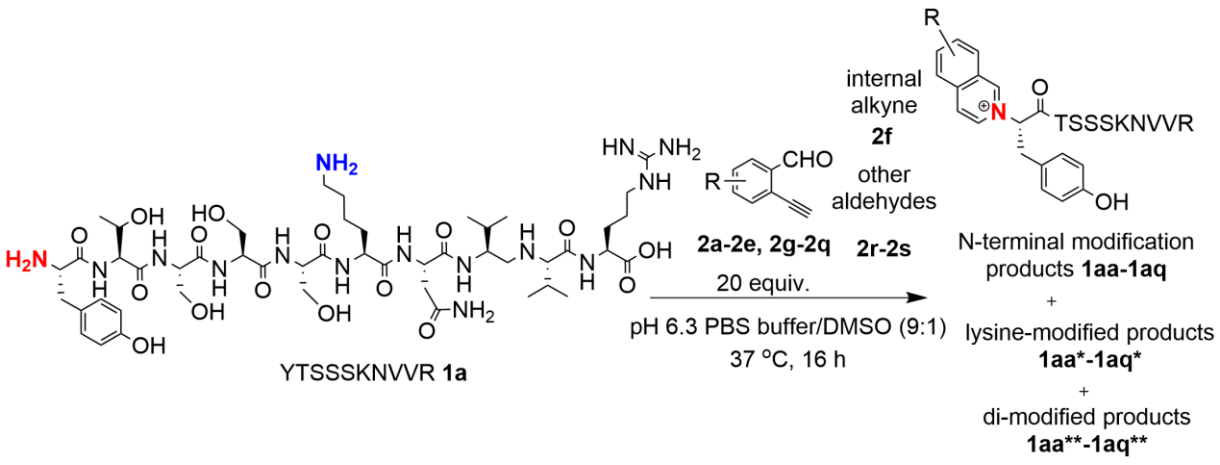
2.2.3 Modification of peptide YTSSSKNVVR 1a using 2-ethynylbenzaldehydes

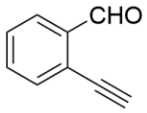
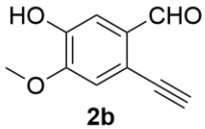
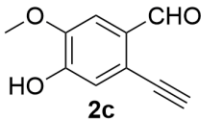
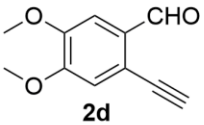
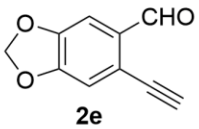
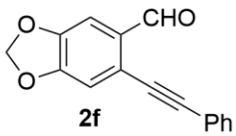
With this encouraging result of *N*-terminal modification, a variety of 2-ethynylbenzaldehydes (**2a-2q**) and other aldehydes (**2r-2s**) were screened to study the effects of substituted groups on product conversion and *N*-terminal selectivity (Scheme 2.4):

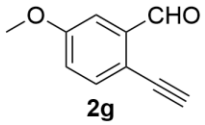
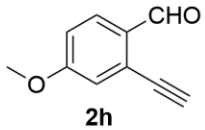
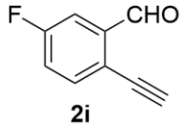
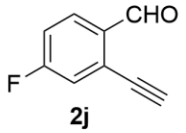
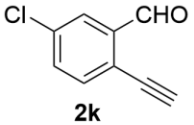
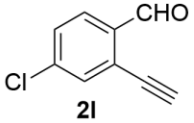
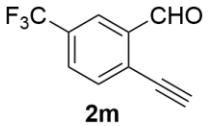
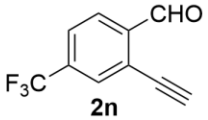
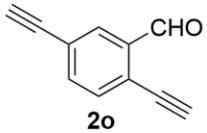
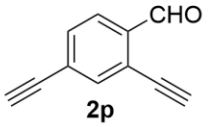


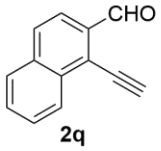
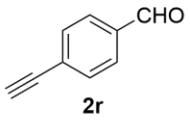
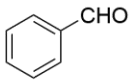
Scheme 2.4 2-Ethynylbenzaldehyde derivatives (**2a-2q**) and other aldehydes (**2r-2s**) screened for peptide modification.

Table 2.1 Modification of Peptide YTSSSKNVVR 1a by 2-Ethynylbenzaldehydes[a]



Entry	Reagent	Conversion (%)[b] +1mod	+2mod	Total	*N*-terminal selectivity of mono-modified peptide YTSSSKNVVR[c]
1	 **2a**	64	8	(72)	21:1
2	 **2b**	73	13	(86)	>99:1
3	 **2c**	28	0	(28)	20:1
4	 **2d**	60	7	(67)	24:1
5	 **2e**	40	1	(41)	50:1
6	 **2f**	0	0	(0)	-

Entry	Reagent	Conversion (%)[b]			*N*-terminal selectivity of mono-modified peptide YTSSSKNVVR[c]
		+1mod	+2mod	Total	
7	 **2g**	63	5	(68)	92:1
8	 **2h**	65	6	(71)	42:1
9	 **2i**	64	4	(68)	81:1
10	 **2j**	71	7	(78)	62:1
11	 **2k**	47	2	(49)	38:1
12	 **2l**	55	3	(58)	35:1
13	 **2m**	67	5	(72)	20:1
14	 **2n**	62	3	(65)	14:1
15	 **2o**	54	3	(57)	25:1
16	 **2p**	59	5	(64)	19:1

Entry	Reagent	Conversion (%)[b]			*N*-terminal selectivity of mono-modified peptide YTSSSKNVVR[c]
		+1mod	+2mod	Total	
17	 2q	8	0	(8)	12:1
18	 2r	0	0	(0)	-
19	 2s	0	0	(0)	-

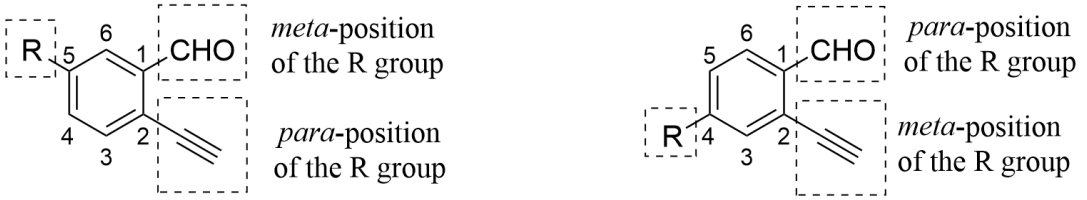
[a] Conditions: YTSSSKNVVR **1a** (0.1 mM), reagent (2 mM, 20 equiv.), in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL), 16 h, 37 ^{o}C, [b] Determined by total ion count (TIC) of LC-MS analysis. [c] Ratio of *N*-terminal α-amino group modified peptide to Lys ε-amino group modified peptide; this ratio was determined by extracted ion chromatogram (EIC) of LCMS analysis.

The substrate conversions of the peptide YTSSSKNVVR **1a** modification using different 2-ethynylbenzaldehydes were determined by LC-MS/MS analysis of the reaction mixtures. As depicted in Table 2.1, the conversions of different 2-ethynylbenzaldehydes were low to excellent (8-86%) with moderate to excellent *N*-terminal selectivity (12:1 up to >99:1) (Table 2.1 entries 1-17). No conversion was found with internal alkyne reagent **2f**, 4-ethynlbenzaldehyde **2r** and benzaldehyde **2s**, suggesting the terminal ethynyl group located at the *ortho*-position of benzaldehyde plays a key role in the isoquinolinium formation. Among the 2-ethynylbenzaldehydes (**2a-2q**) examined, the reaction of peptide **1a** with

2-ethynyl-5-hydroxy-4-methoxybenzaldehyde **2b** was found to afford good conversion (86%) and excellent *N*-terminal selectivity (>99:1) in the peptide modification.

Again, it is noteworthy that the *N*-terminal α-amino group selective peptide bioconjugation using 2-ethynylbenzaldehyde **2a** can be achieved under slightly acidic conditions (pH 6.3), as the *N*-terminal α-amino group remains neutral and possesses stronger nucleophilicity than that of the protonated lysine ε-amino group. The aldehyde group of 2-ethynylbenzaldehyde preferentially reacts with the *N*-terminal α-amino group of peptide to form imine, followed by cyclization with alkyne to form isoquinolinium.

It is of importance to understand the effects of substituents on 2-ethynylbenzaldehydes to product conversion and *N*-terminal selectivity. As the substituted groups are well-known to be affecting the electron density on the *meta*- and *para*-position of the aromatic rings, the 5-substituted and 4-substituted 2-ethynylbenzaldehydes were selected for the bioconjugation, so that their effect on the alkyne part and the aldehyde part can be studied (Figure 2.2).

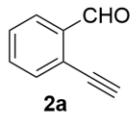


5-substituted 2-ethynylbenzaldehydes

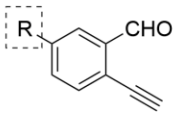
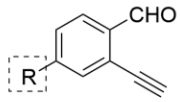
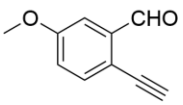
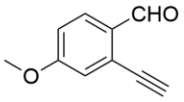
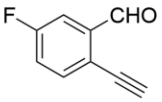
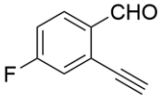
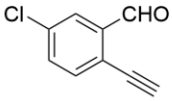
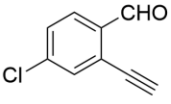
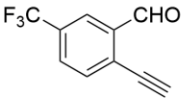
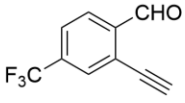
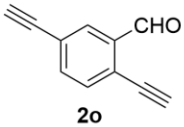
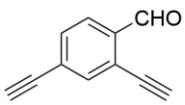
4-substituted 2-ethynylbenzaldehydes

Figure 2.2 (Left) 5-substituted 2-ethynylbenzaldehyes; (Right) 4-substituted 2-ethynylbenzaldehyes.

Table 2.2 *N*-terminal selectivity of (left) 5-substituted reagents and (right) 4-substitued reagents

Ref:  (21:1)

2a

***N*-terminal selectivity of mono-modified peptide YTSSSKNVVR:**	 5-substituted 2-ethynylbenzaldehydes **>**  4-substituted 2-ethynylbenzaldehydes
	5-substituted 2-ethynylbenzaldehydes / 4-substituted 2-ethynylbenzaldehydes
***N*-terminal selectivity of mono-modified peptide YTSSSKNVVR (N-terminal mod vs lysine mod)**	 (92:1) **2g** /  (42:1) **2h**
	 (81:1) **2i** /  (62:1) **2j**
	 (38:1) **2k** /  (35:1) **2l**
	 (20:1) **2m** /  (14:1) **2n**
	 (25:1) **2o** /  (19:1) **2p**

The effects of substituted 2-ethynylbenzaldehydes on the bioconjugation reaction, in terms of reactivity and *N*-terminal selectivity were studied. For mono-substituted 2-ethynylbenzaldehyes (**2g-2p**), the 5-substituted 2-ethynylbenzaldehydes possess better *N*-terminal selectivity than that of the 4-

substituted one (Table 2.2). This trend can be applied to both electron-donating substituted groups (**2g** vs **2h**, *N*-terminal selectivity 92:1 vs 42:1) and electron-withdrawing substituted groups (**2i** vs **2j**, **2k** vs **2l**, **2m** vs **2n** and **2o** vs **2p**). It is noteworthy that the 2-ethynylbenzaldehyde with an electron donating group on the 5-position (**2g**, 92:1) possesses much higher *N*-terminal selectivity than that of the 4-position (**2h,** 42:1).

For 2-ethynylbenzaldehydes with an electron-withdrawing group, the *N*-terminal selectivity decreased with increasing electron-withdrawing strength of the substituted group (i.e. *N*-terminal selectivity – F > -Cl > -CF_3, from 81:1 to 14:1) (Figure 2.3).

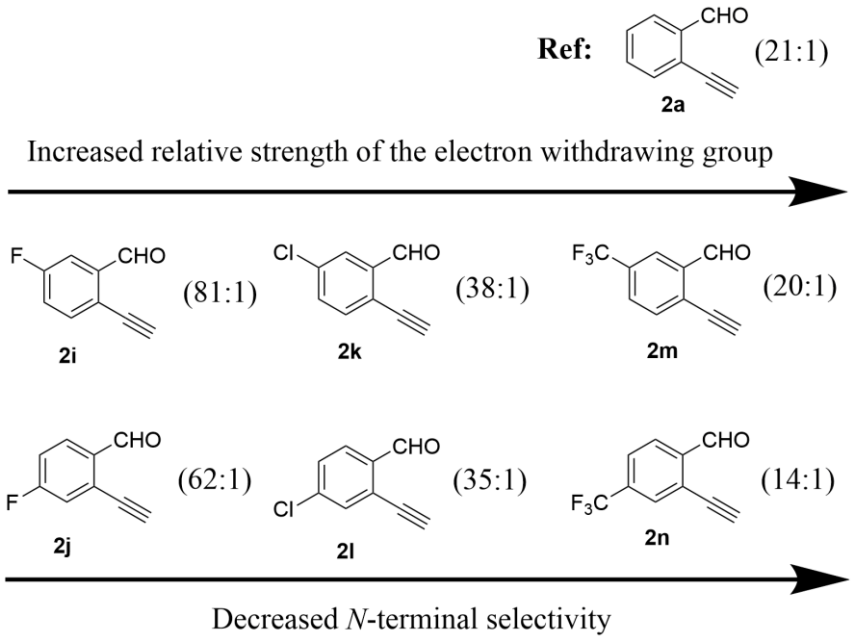
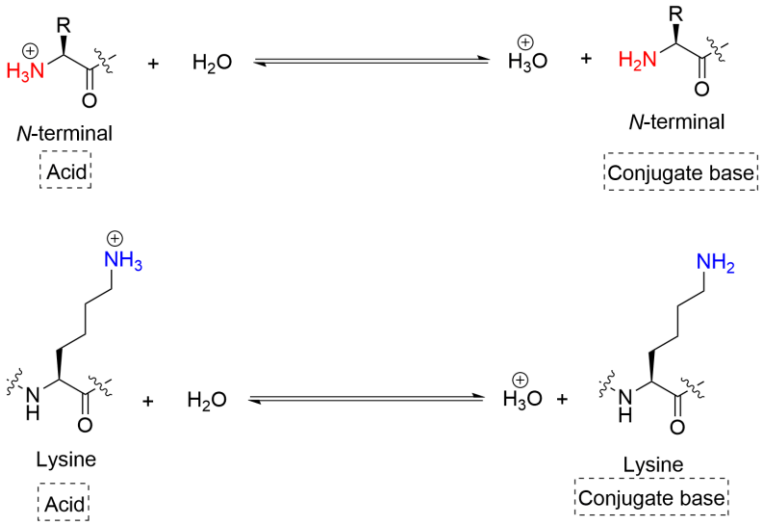


Figure 2.3 *N*-terminal selectivity of electron-withdrawing substituted 2-ethynylbenzaldehydes.

In addition, no significant difference in peptide conversion among 2-ethynylbenzaldehydes (**2g**-**2p**), suggesting the substituents on 2-ethynylbenzaldehydes did not have particular effect on the reactivity of the reagents. However, di-substituted 2-ethynylbenzaldehydes (**2b**-**2e**) showed no trend in both *N*-terminal selectivity and reactivity, the exact reason of this phenomenon is still undetermined.

It is noteworthy that the bioconjugation pH is the crucial factor in controlling the *N*-terminal selectivity of the bioconjugation reaction. Under slightly acidic pH (pH 6.3), the majority of the lysine ε-amino groups are protonated that possess weaker nucleophilicity than that of the *N*-terminal α-amino groups (less amount in protonated form). Therefore, imine formation between the aldehyde of 2-ethynylbenzaldehydes and the *N*-terminal α-amino groups is much more favorable than that of the lysine ε-amino groups, so that this bioconjugation reaction become *N*-terminal selective. In contrast, the rate-limiting step (i.e. cyclization) is independent to bioconjugation pH; the nucleophilicity of imines generated from *N*-terminal α-amino groups and lysine ε-amino groups are similar.

To illustrate the difference in population between the protonated amino group and the neutral amino group, the equation of the protonated form and the neutral form of the amino group can be expressed as an acid-base reaction (Scheme 2.5).



Scheme 2.5 Equation of the protonated form and the neutral form of (Top) *N*-terminal α-amino group; (Bottom) Lysine ε-amino group.

The equilibrium of this reaction can be expressed as acid dissociation constant *K*a (equation 1):

$$Ka = \frac{[\text{Conjugate base}][H_3O^+]}{[\text{Acid}][H_2O]} = \frac{[\text{Conjugate base}]}{[\text{Acid}]} \ldots\ldots (1)$$

p*K*a is the negative common logarithm value of *K*a, that described the ratio between the conjugate base and the acid of a molecule. When the reaction pH equals

to the p*K*a of a molecule, the ratio between the conjugate base and the acid of a molecule is the same (i.e. 1:1).

The ratio of the protonated form and the neutral form of the amino group can be estimated by the Henderson-HasselBalch equation (equation 2).[2]

$$\mathrm{pH} = \mathrm{p}K\mathrm{a} + \log\frac{[\mathrm{Conjugate\ base}]}{[\mathrm{Acid}]} \ldots\ldots\ldots (2)$$

When the bioconjugation pH is 6.3, then the ratio of the protonated form and the neutral form of the *N*-terminal α-amino group (p*K*a = 8) is:

$$6.3 = 8 + \log\frac{[\mathrm{Conjugate\ base}]}{[\mathrm{Acid}]}$$

$$\frac{[\mathrm{Conjugate\ base}]}{[\mathrm{Acid}]} = 0.01995$$

And the ratio of the protonated form and the neutral form of the lysine ε-amino group (p*K*a = 10) is:

$$6.3 = 10 + \log\frac{[\mathrm{Conjugate\ base}]}{[\mathrm{Acid}]}$$

$$\frac{[\mathrm{Conjugate\ base}]}{[\mathrm{Acid}]} = 0.0001995$$

Although the majority of the *N*-terminal α-amino groups and the lysine ε-amino groups are protonated form under pH 6.3, the *N*-terminal α-amino groups possess 100 times more in the neutral form population than that of the lysine ε-amino groups. Therefore, the bioconjugation reagents should preferentially react with the

N-terminal α-amino group that possesses more in the neutral form population (i.e. higher nucleophilicity in compared with the lysine ε-amino group).

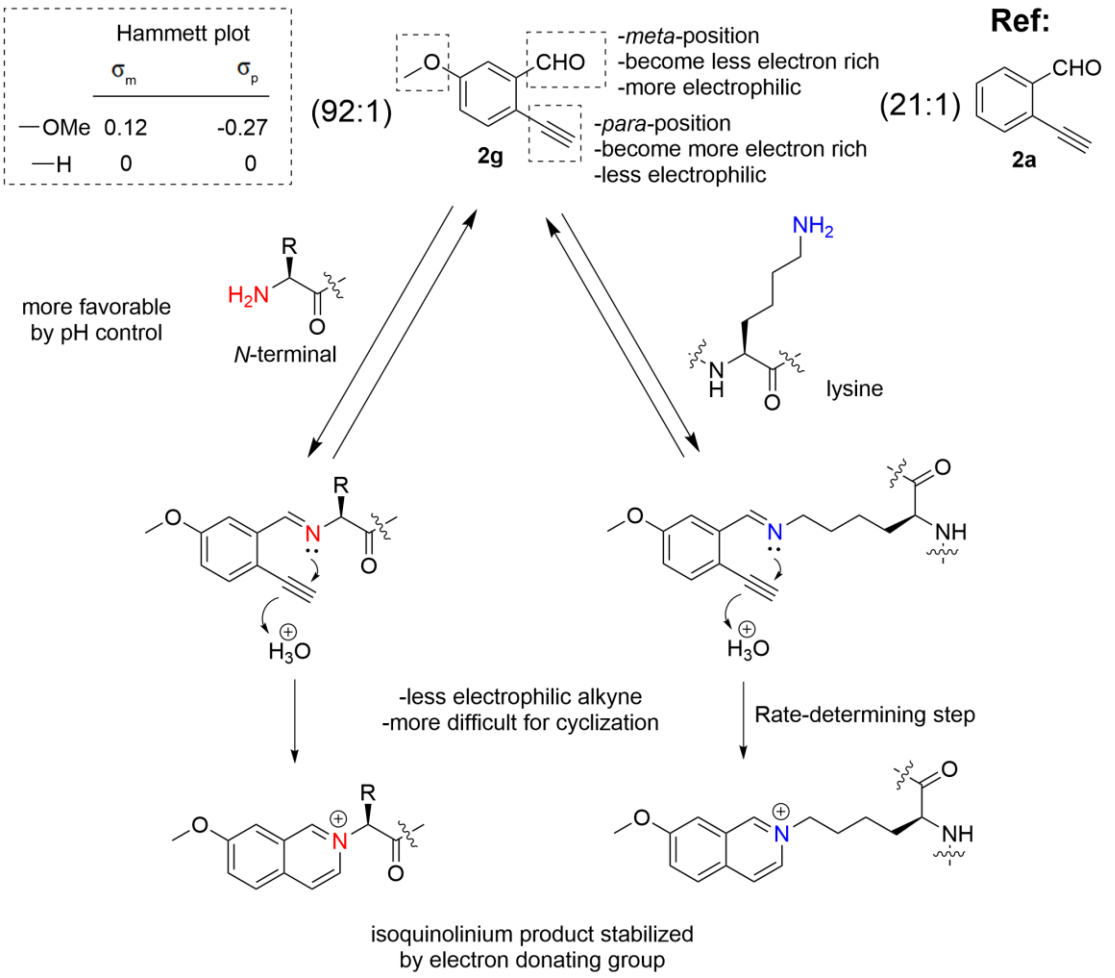
The effects of substituents on mono-substituted 2-ethynylbenzaldehydes to *N*-terminal selectivity can be explained by the Hammett equation,[3] which is an equation that describes a linear free-energy relationship relating to reaction rates and equilibrium constants, such equation applies to reactions involving benzoic acid derivatives with *meta*- and *para*-substituents. There are two parameters used in this equation: substituent constant (σ, σ_m for *meta*-position and σ_p for *para*-position) and reaction constant (ρ).[3] In this case, the reaction considered is the isoquinolinium formation between peptide and 2-ethynylbenzaldehydes, the reaction constants (ρ) among different substituents are the same in this bioconjugation reaction, therefore, only the substituent constant (σ_m and σ_p) is considered.

Table 2.3 The substituent constant σ_m and σ_p of different substituents

Hammett plot

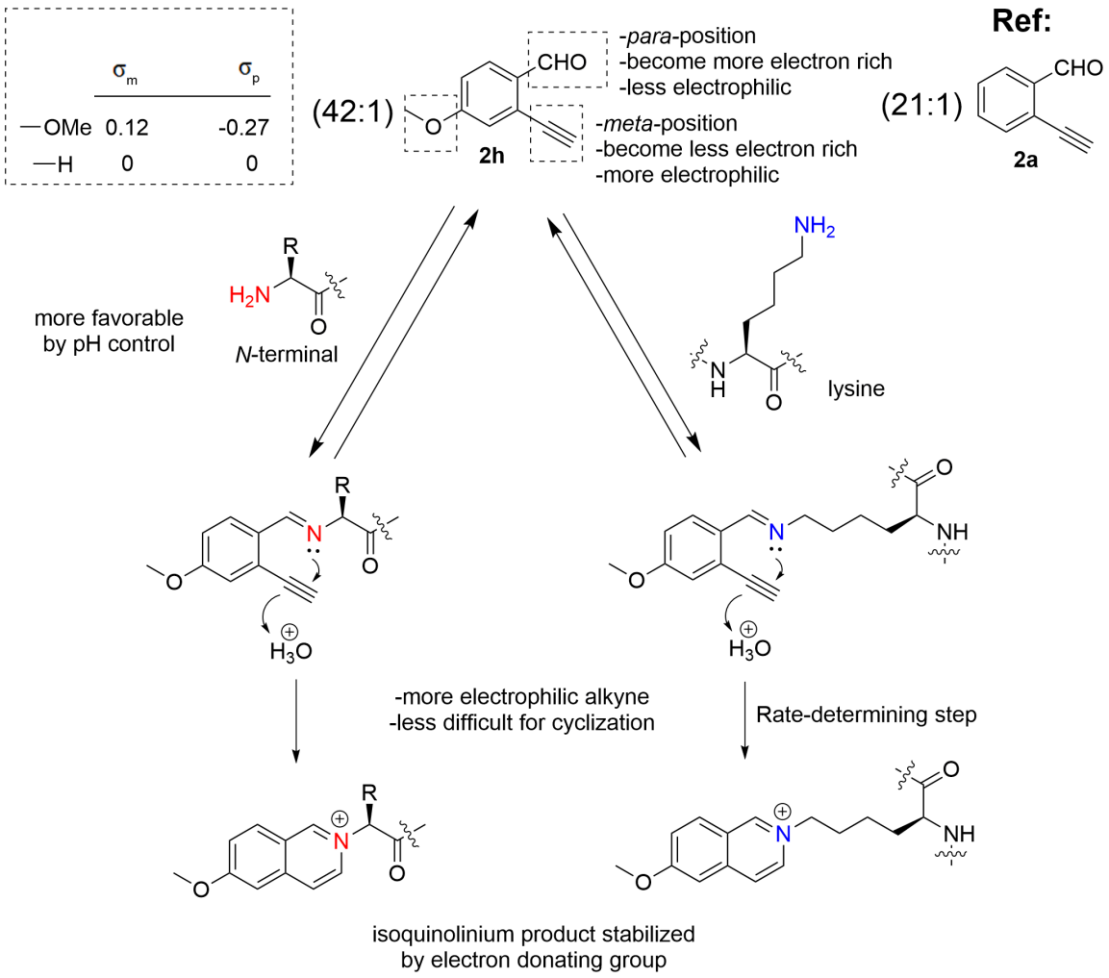
	σ_m	σ_p
—H	0	0
—OH	0.12	-0.37
—OMe	0.12	-0.27
—F	0.34	0.06
—Cl	0.37	0.23
—CF_3	0.43	0.54
—C≡CH	0.21	0.23

The hydrogen (-H, i.e. no substituent) of the benzoic acid is set as reference, in which its substituent constants (σ_m and σ_p) are set as 0. The positive value of a substituent constant σ_m/σ_p indicates the relative strength of electron withdrawing effect of the substituents, whereas the negative σ_m/σ_p value indicates the relative strength of electron donating effect of the substituents (Table 2.3). Therefore, σ_m and σ_p indicate the relative inductive effect to the *meta*-position and the *para*-position of the aromatics, respectively.[4]



Scheme 2.6 Isoquinolinium formation using 5-substituted 2-ethynylbenzaldehyde **2g**.

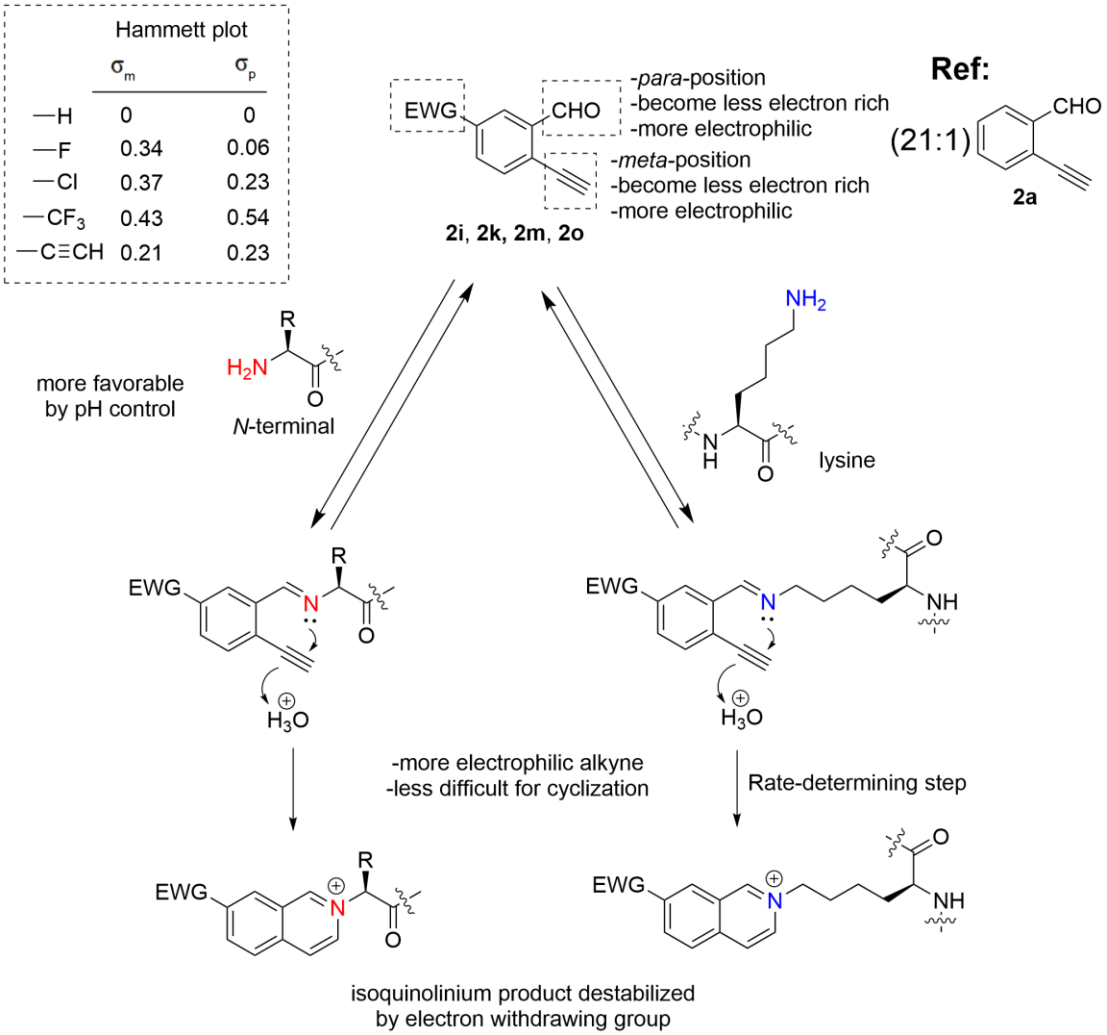
For reagent **2g** which contains an electron-donating group on the 5-position of the 2-ethynylbenzalehyde, although the aldehyde group (*meta*-position) becomes less electron rich and easier to undergo imine formation, the alkyne group (*para*-position) becomes more electron rich and less electrophilic, in which the cyclization becomes more difficult. The less reactive alkyne for cyclization favors the thermodynamic product, the isoquinolinium product that is stabilized by the electron donating group, resulting in a higher *N*-terminal selectivity (Scheme 2.6).



Scheme 2.7 Isoquinolinium formation using 4-substituted 2-ethynylbenzaldehyde **2h**.

For reagent **2h** which contains an electron-donating group on the 4-position of the 2-ethynylbenzalehyde, the aldehyde group (*para*-position) becomes more electron rich and less electrophilic, which requires a more nucleophilic amine for imine formation. As the *N*-terminal α-amino group possesses higher nucleophilicity than that of the lysine ε-amino group under slightly acidic pH (pH 6.3); the *N*-terminal α-amino group can form imine easier and subsequently form the isoquinolinium product. However, the alkyne (*meta*-position) of the 2-ethynylbenzaldehyde becomes less electron rich and more electrophilic, renders an easier cyclization to

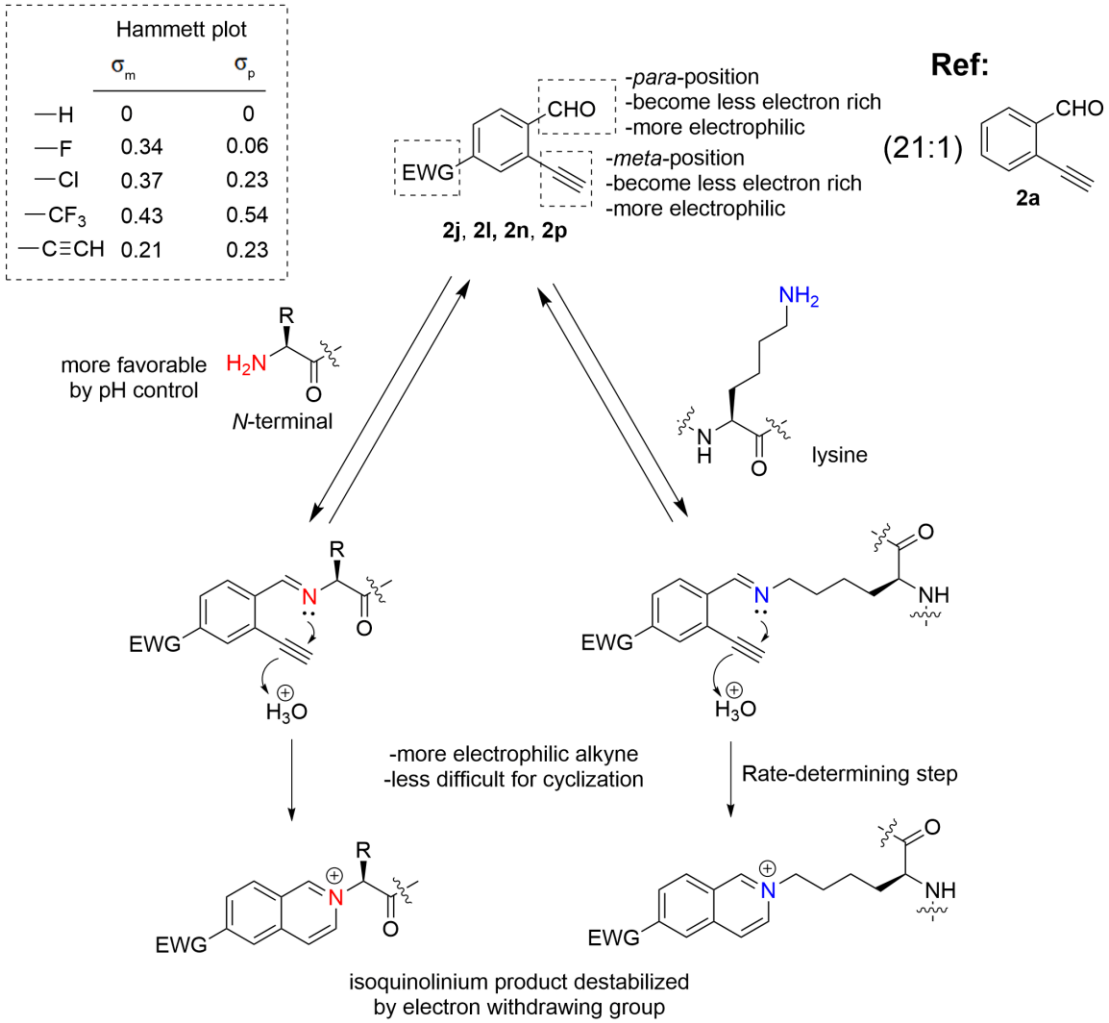
isoquinolinium, so that the lysine modified product is easier to form, resulting in a lower *N*-terminal selectivity in compared with using reagent **2g** (Scheme 2.7).



Scheme 2.8 Isoquinolinium formation using the 5-substituted 2-ethynylbenzaldehydes with different electron withdrawing groups.

As shown before, the *N*-terminal selectivity decreases with increasing electron-withdrawing strength of the substituted group (*N*-terminal selectivity –F > -Cl > -CF_3, from 81:1 to 14:1). For reagents **2i**, **2k**, **2m** and **2o** which contain an electron-withdrawing group on the 5-position of the 2-ethynylbenzalehyde, the

aldehyde group (*meta*-position) becomes less electron rich and easier to undergo imine formation. In addition, the alkyne group (*para*-position) becomes less electron rich and more electrophilic, in which a less nucleophilic imine is also reactive enough for subsequent cyclization, resulting lower *N*-terminal selectivity (Scheme 2.8).

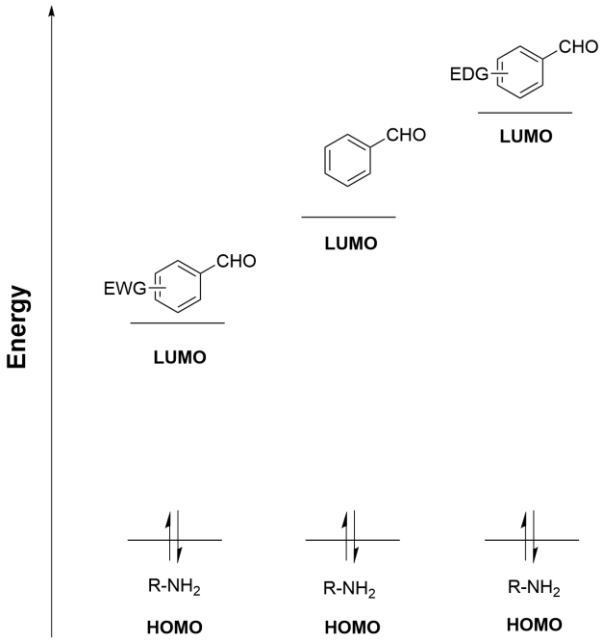


Scheme 2.9 Isoquinolinium formation using the 4-substituted 2-ethynylbenzaldehydes with different electron withdrawing groups.

Similar to the 2-ethynylbenzaldehydes with an electron-withdrawing group on the 5-position, the same trend appeared in the 2-ethynylbenzaldehydes with an

electron-withdrawing group on the 4-position (i.e. *N*-terminal selectivity: **2j**> **2l**> **2p**> **2n**). Similarly, both the aldehyde group (*para*-position) and the alkyne group (*meta*--position) become less nucleophilic, so that the imine formation and the subsequent cyclization become easier, resulting lower *N*-terminal selectivity (Scheme 2.9).

From the study on the effects of substituents on mono-substituted 2-ethynylbenzaldehydes to *N*-terminal selectivity, it is obvious that any substituents which promote imine formation and subsequent cyclization would result lower *N*-terminal selectivity and vice versa.



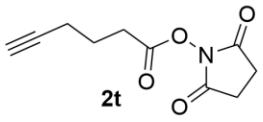
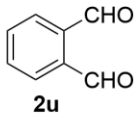
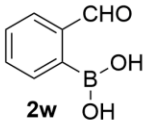
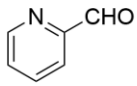
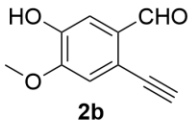
Scheme 2.10 Energy diagram of amines and aldehydes.

Although no relevant literature studied the substituent effect to isoquinoline formation, previous literature suggested that the structure-stability correlations for imine formation depend on the pK_a of amines (donor) and the lowest unoccupied

molecular orbital (LUMO) energies of aldehydes (acceptor); an electron-donating substituted aldehyde possesses higher LUMO energy in compare with benzaldehyde, thereby increase difficulty for imine formation. For an electron-withdrawing substituted aldehyde, it possesses lower LUMO energy than benzaldehyde, which promotes imine formation (Scheme 2.10).[5]

In addition, good water solubility of bioconjugation reagents also an importance issue for the bioconjugation. Most of the 2-ethynylbenzaldehyes show excellent water compatibility during addition from the DMSO stock solution except the 1-ethynyl-2-naphthaldehyde **2q**. Precipitation of reaction mixture was found upon bioconjugation using reagent **2q**, resulting low conversion (8%).

Table 2.4 Modification of Peptide YTSSSKNVVR 1a by literature reported bioconjugation reagents[a]

Entry	Reagent	Conversion (%)[b] +1mod	+2mod	Total	*N*-terminal selectivity of mono-modified peptide YTSSSKNVVR[c]
1[d]	 **2t**	38	61	(99)	19:1
2[e]	 **2u**	43	57	(100)	2:1
3[f]	 **2w**	2	0	(2)	>99:1
4	 **2x**	78	0	(78)	>99:1
5	 **2b**	73	13	(86)	>99:1

[a] Conditions: YTSSSKNVVR **1a** (0.1 mM), reagent (2 mM, 20 equiv.), in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL), 16 h, 37 °C, [b] Determined by total ion count (TIC) of LC-MS analysis. [c] Ratio of *N*-terminal α-amino group modified peptide to Lys ε-amino group modified peptide; this ratio was determined by extracted ion chromatogram (EIC) of LCMS analysis. [d] Entry 1 treated with NHS ester **2n** (2 mM, 20 equiv.), 15 min, 37 °C. [e] Entry 2 treated with OPA **2u** (0.1 mM, 1 equiv.), 4 h, 25 °C. [f] Entry 3 treated with 2-formylphenylboronic acid **2w** (2 mM, 20 equiv.), in pH 6.3 50 mM of potassium phosphate buffer/DMSO (9:1, 100 μL), 18 h, 25 °C.

For comparison, the peptide YTSSSKNVVR **1a** (0.1 mM) was modified using NHS ester **2n** (2 mM, 20 equiv.) was conducted in pH 6.3 50 mM PBS buffer for 15 min in regard of its fast reaction kinetics; with moderate *N*-terminal selectivity (19:1). However, a significant amount of doubly-modified peptide (61%) was found, indicating that the NHS ester did not have good chemoselectivity as most of the lysine residues of the peptide were modified (Table 2.4). Modification of peptide YTSSSKNVVR **1a** (0.1 mM) using OPA **2u** (reported by Li's group) (0.1 mM, 1 equiv.) was conducted in pH 6.3 50 mM PBS buffer for 4 h to give low N-terminal selectivity (2:1). In addition, modification of peptide YTSSSKNVVR **1a** (0.1 mM) using 2-formylphenylboronic acid **2w** (2 mM, 20 equiv.), was conducted in pH 6.3 50 mM KPi buffer for 18 h. Although this bioconjugation afforded excellent *N*-terminal selectivity (>99:1), the conversion was very low (2%). Among the 2-ethynylbenzaldehydes (**2a**-**2q**) examined, reaction of peptide **1a** with reagent **2b**, **2d** and **2l** were found to afford excellent *N*-terminal selectivity (81:1 to >99:1), which is comparable to 2PCA **2x** developed by Francis and co-workers.[6b]

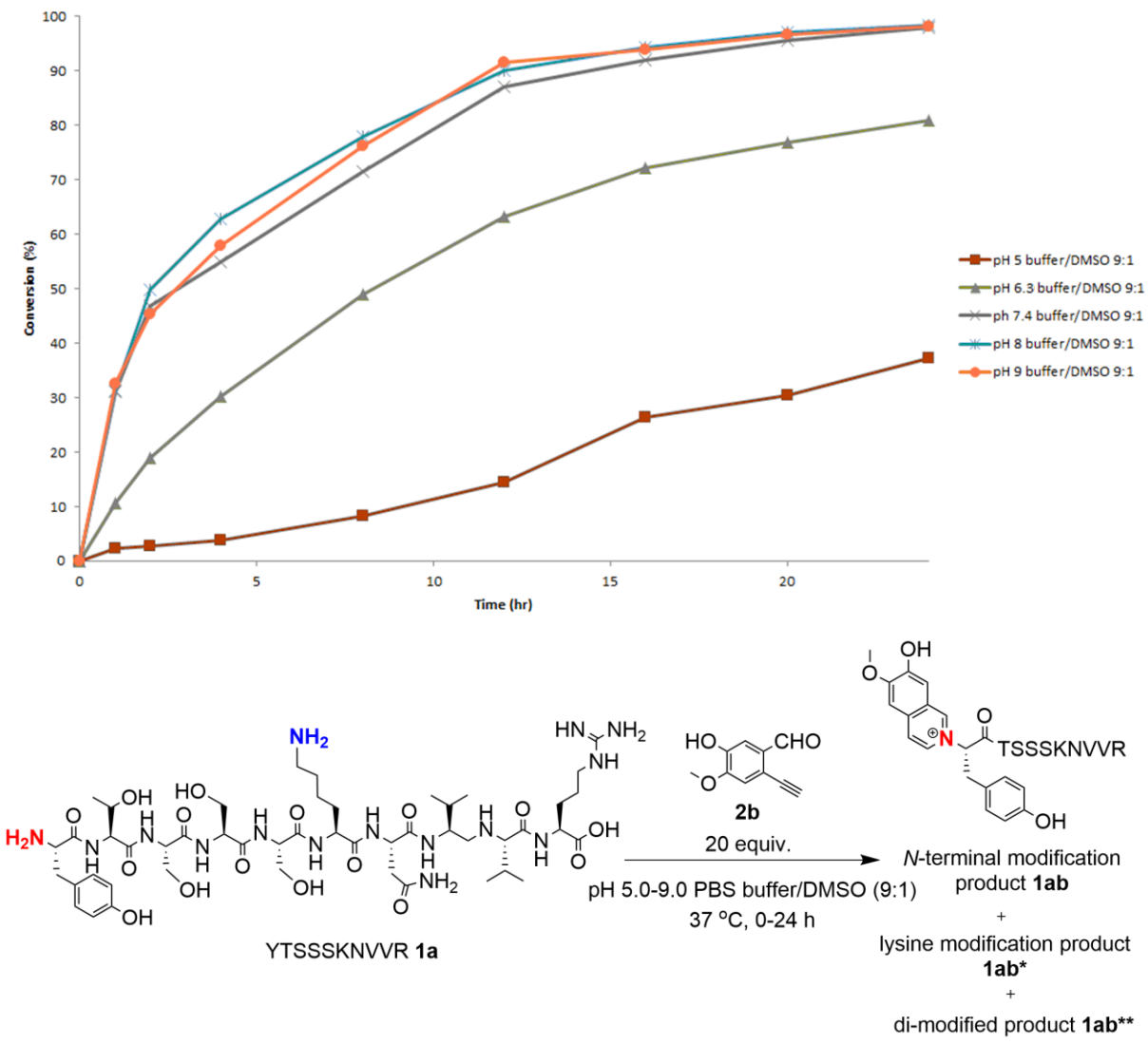
In summary, the reagent **2b** was found to be comparable to literature reported reagents,[6] in terms of reaction conditions (pH 6.3, 16 h, 37 ^{o}C), reactivity (86% conversion) and *N*-terminal selectivity (>99:1) (Table 2.4).

2.2.4 Time course study of peptide YTSSSKNVVR 1a using 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde 2b

In the previous screening of 2-ethynylbenzaldehydes, most of the 2-ethynylbenzaldehydes reacted with the peptide YTSSSKNVVR **1a** and afforded the corresponding isoquinolinium-peptide conjugates with moderate to excellent *N*-terminal selectivity. In particular, 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde **2b** formed such a product in high conversion (86%) with excellent *N*-terminal selectivity (>99:1). Therefore, reagent **2b** was selected for detailed studies.

The effect of pH on this reaction was further examined. Most of the *N*-terminal α-amino/lysine ε-amino group selective bioconjugations can be either *N*-terminal selective (near neutral pH) or lysine selective (pH 8.5~9.5) by judicious pH control. In addition, it is of importance to study the optimized conditions of this pH dependent bioconjugation, acidic pH conditions offer better *N*-terminal selectivity. However, at the cost of lower bioconjugation yield, therefore, the relationship between these two elements can be found in the time course study. Also, it is of importance to study the optimized reaction time of this bioconjugation, as the conversion generally increases overtime, but the distribution of products may be different (i.e. *N*-terminal selectivity of mono-modified peptide **1ab**/**1ab***, di-modified product **1ab****).

Reagent **2b** (2 mM, 20 equiv.) was selected to study the formation of *N*-terminal modified peptide at different pH values (from pH 5 to pH 9), treating with peptide **1a** (0.1 mM) at 37 ^{o}C for 24 h. pH time plot data were performed twice from two individual experiments to minimize the error, analyzed by LC-MS/MS (Scheme 2.11).



Scheme 2.11 Time course experiments of the formation of **1ab** at different pH values.

Table 2.5 Modification of Peptide YTSSSKNVVR 1a using reagent 2b with different reaction time and pH conditions at 12, 16, and 20 h[a]

pH	Time (h)	Conversion (%)[b]			*N*-terminal Selectivity of mono-modified YTSSSKNVVR[c]
		+1 mod	+2 mod	Total	
5.0	12	14	0	(14)	>99:1
	16	25	1	(26)	>99:1
	20	29	1	(30)	>99:1
6.3	12	59	5	(64)	>99:1
	16	64	8	(72)	>99:1
	20	66	11	(77)	>99:1
7.4	12	68	19	(87)	43:1
	16	61	30	(91)	38:1
	20	59	36	(95)	27:1
8.0	12	59	31	(90)	30:1
	16	49	46	(95)	19:1
	20	45	52	(97)	19:1
9.0	12	52	39	(91)	24:1
	16	45	49	(94)	19:1
	20	41	56	(97)	18:1

[a] Conditions: YTSSSKNVVR **1a** (0.1 mM), reagent **2b** (2 mM, 20 equiv.), in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL), (12 h, 16 h and 20 h), 37 ^{o}C, [b] Determined by total ion count (TIC) of LC-MS analysis. [c] Ratio of *N*-terminal α-amino group modified peptide to Lys ε-amino group modified peptide; this ratio was determined by extracted ion chromatogram (EIC) of LCMS analysis.

Excellent *N*-terminal selectivity (ratio of α-amino- to ε-amino-modified peptide >99:1) was observed at pH 6.3 as determined by LC-MS/MS, inevitable doubly-modified peptide was observed under basic conditions (pH 7.4-9.0) as *N*-terminal selectivity decreases upon elevating pH (*N*-terminal selectivity, α-amino- to ε-amino-modified peptide: pH 7.4 (38:1), pH 8.0 (18:1), pH 9.0 (18:1)). Although the conversion increases slowly beyond 12 h, the optimized modification time was found to be 16 h as doubly-modified product gradually formed after 16 h (Table 2.5).

The results suggest that this bioconjugation reaction is pH-dependent. *N*-terminal modification product is the sole product under acidic conditions (pH 5), however, with severely reduced reactivity of α-amino group. In contrast, this bioconjugation exerts moderate *N*-terminal selectivity and much faster reaction kinetics under physiological and basic conditions (from 18:1 to 38:1) with considerably amount of doubly modified peptide (+2 mod in 16 h: pH 7.4 (30%), pH 8.0 (46%) pH 9.0 (49%)). This result is not surprising as the portion of protonated ε-amino group decreases with increasing pH, leading to less significant difference of the nucleophilicity between *N*-terminal α-amino group and lysine ε-amino group. The

increased nucleophilicity also rendered higher bioconjugation reaction rate (Table 2.5).

Although this bioconjugation appears to be too reactive under basic conditions, it is noteworthy that when the bioconjugation time was shortened to 2 h, high *N*-terminal selectivity (pH (7.4) >99:1, pH (8.0) 83:1, pH (9.0) 72:1) and moderate conversion (45~50%) with only a trace amount of doubly modified (2~5%) can still be achieved (Table 2.6).

Table 2.6 Modification of Peptide YTSSSKNVVR 1a using reagent 2b under basic pH conditions at 2 and 4 h

pH	Time (h)	Conversion (%)[b]			*N*-terminal Selectivity of mono-modified YTSSSKNVVR[c]
		+1 mod	+2 mod	Total	
7.4	2	44	2	(46)	>99:1
	4	49	6	(55)	56:1
8.0	2	45	5	(50)	83:1
	4	52	10	(62)	44:1
9.0	2	40	5	(45)	72:1
	4	47	11	(58)	40:1

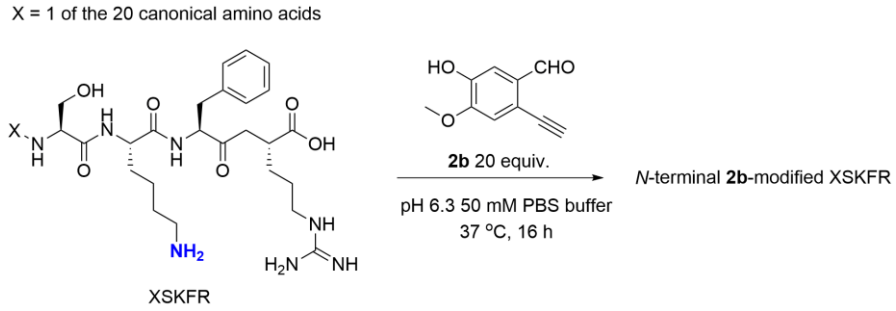
[a] Conditions: YTSSSKNVVR **1a** (0.1 mM), reagent **2b** (2 mM, 20 equiv.), in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL), (2 h and 4 h), 37 ^{o}C, [b] Determined by total ion count (TIC) of LC-MS analysis. [c] Ratio of *N*-terminal α-amino group modified peptide to Lys ε-amino group modified peptide; this ratio was determined by extracted ion chromatogram (EIC) of LCMS analysis.

Considering the balance between reactivity and *N*-terminal selectivity, this bioconjugation was optimized under slightly acidic conditions (pH 6.3) which should be satisfactory as a protocol for most bioconjugation applications. For more acidic (pH 5.0) and basic conditions (pH 7.4~9.0), the reaction time needs to be lengthened and shortened, respectively.

2.2.5 Modification of a peptide library XSKFR (X = 20 amino acid residues) using 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde 2b

The effect of different *N*-terminal amino acid residues on the *N*-terminal selectivity was systematically studied by employing a panel of peptides XSKFR with 20 different natural amino acids at the *N*-terminus. The modification reaction of peptide XSKFR (0.1 mM) with 20 equiv. of reagent **2b** (2 mM) in pH 6.3 50 mM of PBS buffer/DMSO (9:1) at 37 ^{o}C for 16 h (Table 2.7). All the *N*-terminal amino acid residues afforded moderate to excellent conversion (32-93%), with 13 out of 20 amino acids giving excellent selectivity (ratio of α-amino- to ε-amino-modified peptide, >99:1, Table 2.7, entries 1-12, Section 2.2.6, CSKFR). Moderate to high *N*-terminal selectivity (9:1 to 63:1, Table 2.7, entries 13-18). The average conversion is 76% which is comparable with the 2-PCA protocol developed by Francis and co-workers (78%).[6b]

Table 2.7 Modification of Peptide XSKFR using reagent 2b[a]



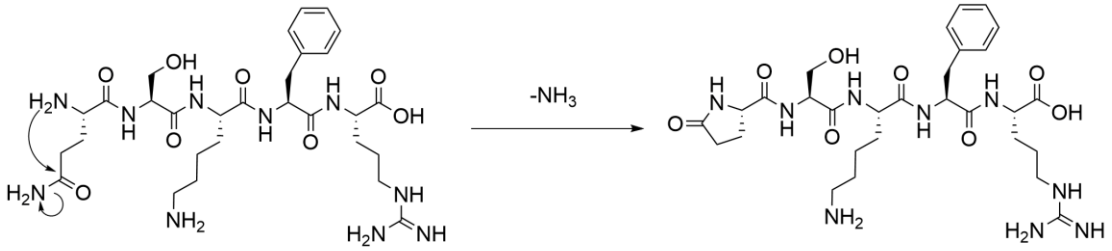
Entry	Peptide sequence	Conversion (%)[b]			*N*-terminal selectivity of mono-modified XSKFR[c]
		+1mod	+2mod	Total	
1	ASKFR	69	4	(73)	>99:1
2	GSKFR	62	8	(70)	>99:1
3	LSKFR	75	3	(78)	>99:1
4	WSKFR	66	4	(70)	>99:1
5	YSKFR	74	11	(85)	>99:1
6	DSKFR	66	13	(79)	>99:1
7	ESKFR	63	13	(76)	>99:1
8	HSKFR	61	16	(77)	>99:1
9	KSKFR	60	33	(93)	>99:1
10	SSKFR	61	22	(83)	>99:1
11	NSKFR	65	15	(80)	>99:1
12	QSKFR	18	14	(32)	>99:1
13	FSKFR	71	6	(77)	63:1
14	VSKFR	76	4	(80)	17:1
15	RSKFR	70	23	(93)	16:1
16	ISKFR	86	3	(89)	14:1
17	MSKFR	83	5	(88)	10:1
18	TSKFR	61	11	(72)	9:1
19	PSKFR	42	5	(47)	1:5
20	CSKFR	-	-	(-)	-

[a] Conditions: XSKFR (0.1 mM), reagent **2b** (2 mM, 20 equiv.), in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL), 16 h, 37 °C, [b] Determined by total ion count (TIC) of LC-MS analysis. [c] Ratio of *N*-terminal α-amino group modified peptide to Lys ε-amino group modified peptide; this ratio was determined by extracted ion chromatogram (EIC) of LCMS analysis.

This bioconjugation reaction using reagent **2b** afforded excellent reactivity and *N*-terminal selectivity in amino acid residues that contain aliphatic groups (alanine, glycine, leucine, valine), aromatic groups (phenylalanine, tryptophan, tyrosine) and acidic groups (aspartic acid, glutamic acid). Notably, the 2-ethynylbenzaldehydes have shown no cross reactivity with other potential nucleophiles such as thiol groups (cysteine), carboxylic groups (aspartic acid), and hydroxyl groups (serine), indicating that this bioconjugation is highly chemoselective.

A low conversion and poor *N*-terminal selectivity of 1:5 was observed for PSKFR having *N*-terminal proline residue, which is presumably due to iminium intermediate formed between proline and 2-ethynylbenzaldehyde that cannot undergo subsequent cyclization with the proximal alkyne group.

It is also noteworthy that the low peptide conversion of QSKFR (Table 2.7, entry 20) is presumable due to the deamidation of glutamine to deamidated cyclic product pyroglutamate under buffer solution as reported by several literatures.[7] Upon the formation of pyroglutamate, the alpha amino group is hampered and becomes unreactive (Scheme 2.12). Similar phenomenon during bioconjugation was also reported by Francis's group.[6b] Nevertheless, the mono-modified QSKFR possesses excellent *N*-terminal selectivity (>99:1).



Scheme 2.12 Deamidation of Gln in peptide QSKFR to pGln.

Moderate to high *N*-terminal selectivity of this bioconjugation reaction was found in phenylalanine, valine, arginine, isoleucine, methionine and threonine (9:1 to 63:1). The bioconjugations were performed under more acidic conditions (pH 5.0 and pH 5.5) so that a higher *N*-terminal selectivity of these *N*-terminal amino acid residues maybe achieved (Table 2.8).

Table 2.8 Modification of Peptide XSKFR using reagent 2b at pH 5.0 and pH 5.5

entry	peptide sequence	conversion (%)[a]			*N*-terminal selectivity of mono-modified XSKFR[b]
		+1mod	+2mod	Total	
1[c]	FSKFR	56	3	(59)	36:1
2[c]	VSKFR	59	1	(60)	81:1
3[c]	RSKFR	61	7	(68)	16:1
4[c]	ISKFR	69	2	(71)	16:1
5[c]	MSKFR	83	4	(87)	19:1
6[c]	TSKFR	48	0	(48)	8:1
7[d]	FSKFR	35	1	(36)	18:1
8[d]	VSKFR	30	0	(30)	33:1
9[d]	RSKFR	36	2	(38)	25:1
10[d]	ISKFR	48	1	(49)	30:1
11[d]	MSKFR	66	2	(68)	20:1
12[d]	TSKFR	32	0	(32)	4:1

[a] Determined by total ion count (TIC) of LC-MS analysis. [b] Ratio of *N*-terminal α-amino group modified peptide to Lys ε-amino group modified peptide; this ratio was determined by extracted ion chromatogram (EIC) of LCMS analysis. [c] Conditions: XSKFR (0.1 mM), reagent **2b** (2 mM, 20 equiv.), in pH 5.5 50 mM of PBS buffer/DMSO (9:1, 100 μL), 16 h, 37 ^{o}C, [d] XSKFR (0.1 mM), reagent **2b** (2 mM, 20 equiv.), in pH 5.0 50 mM of PBS buffer/DMSO (9:1, 100 μL), 16 h, 37 ^{o}C.

Peptides with the *N*-terminal amino acid residue valine, arginine, isoleucine or methionine afforded higher *N*-terminal selectivies when the bioconjugations were performed under more acidic pH conditions (entries 2, 9-11) (pH 5.5: valine (81:1), pH 5.0: arginine (25:1), isoleucine (30:1) and methionine (20:1). Surprisingly, bioconjugation of peptides FSKFR and TSKFR afforded lower *N*-terminal selectivity under more acidic pH condition (pH 5.5) in compared with the optimized pH

condition (pH 6.3), from 63:1 to 36:1 (FSKFR, entry 1), and 9:1 to 8:1 (TSKFR, entry 6), respectively. When the bioconjugations were performed under more acidic condition (pH 5.0), the *N*-terminal selectivity of **2b**-modified peptides FSKFR and TSKFR decrease further to 18:1 (entry 7) and 4:1 (entry 12), respectively. Similar trend was observed in peptide SSKFR, the *N*-terminal selectivity of **2b**-modified SSKFR decreases from 81:1 (entry 2) to 33:1 (entry 8).

The reactivity of this bioconjugation reagent decreases under more acidic conditions. However, instead of increased *N*-terminal selectivity of the modified-peptides under more acidic pH conditions (such as pH 5.5 and 5.0), their *N*-terminal selectivities vary case by case and show no significant trend.

Table 2.9 Modification of Peptide XSKFR using 10 and 15 equiv. reagent 2b

Entry	Peptide sequence	Conversion (%)[a]			*N*-terminal selectivity of mono-modified XSKFR[b]
		+1mod	+2mod	Total	
1[c]	DSKFR	45	3	(48)	>99:1
2[c]	ESKFR	47	4	(51)	>99:1
3[c]	SSKFR	47	7	(54)	>99:1
4[d]	HSKFR	51	3	(54)	>99:1
5[d]	KSKFR	47	8	(55)	>99:1
6[d]	NSKFR	64	5	(69)	>99:1
7[d]	RSKFR	56	4	(60)	7:1

[a] Determined by total ion count (TIC) of LC-MS analysis. [b] Ratio of *N*-terminal α-amino group modified peptide to Lys ε-amino group modified peptide; this ratio was determined by extracted ion chromatogram (EIC) of LCMS analysis. [c] Conditions: XSKFR (0.1 mM), reagent **2b** (1.5 mM, 15 equiv.), in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL), 16 h, 37 ^{o}C, [d] XSKFR (0.1 mM), reagent **2b** (1 mM, 10 equiv.), in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL), 16 h, 37 ^{o}C.

Although this bioconjugation reaction using reagent **2b** afforded good reactivity among 20 canonical *N*-terminal amino acid residues, considerable amounts of doubly-modified peptides (α-amino and ε-amino-modified) were found in basic *N*-terminal amino acid residues (arginine, histidine and lysine, +2mod: 16-33%), in less extent in acidic residues (aspartic acid and glutamic acid, +2mod: 13%), hydroxyl residues (serine and threonine, +2mod: 22% and 11%) and amidic residue (asparagine, +2mod: 15%). To address this problem, the amount of reagent **2b** were reduced from 20 to 15 equiv. (entries 1-3) and 10 equiv. (entries 4-7) respectively (Table 2.9). The

amounts of +2mod peptides were then reduced to acceptable amount (<10% of +2mod).

Generally, the 2-ethynylbenzaldehyde-based bioconjugation reaction gave the best *N*-terminal selectivity in *N*-terminal residues with aliphatic groups (e.g. alanine) and aromatic (e.g. phenylalanine). For other *N*-terminal residues (acidic, basic, amidic, hydroxyl), further optimizations such as more acidic conditions and reduced reagent amount were needed to achieve better results.

2.2.6 Optimization of a cysteine-containing peptide CSKFR using 2-ethynyl-5-hydroxy-4- methoxybenzaldehyde 2b

In the previous studies of the effect of *N*-terminal amino acid residues, no product formation was found in CSKFR, which may be due to the oxidation of cysteine thiol group into disulfide bond (Table 2.7, entry 20). To confirm whether the CSKFR was dimerized upon the bioconjugation or prior to the bioconjugation, a freshly prepared CSKFR stock solution was subjected to LC-MS/MS analysis. No dimerized peptide was found in the stock solution. The freshly prepared solution was added to 50 mM PBS buffer without adding any 2-ethynylbenzaldehyes and subjected to our bioconjugation protocol. Considerable amount of dimerized CSKFR was found (Figure 2.4).

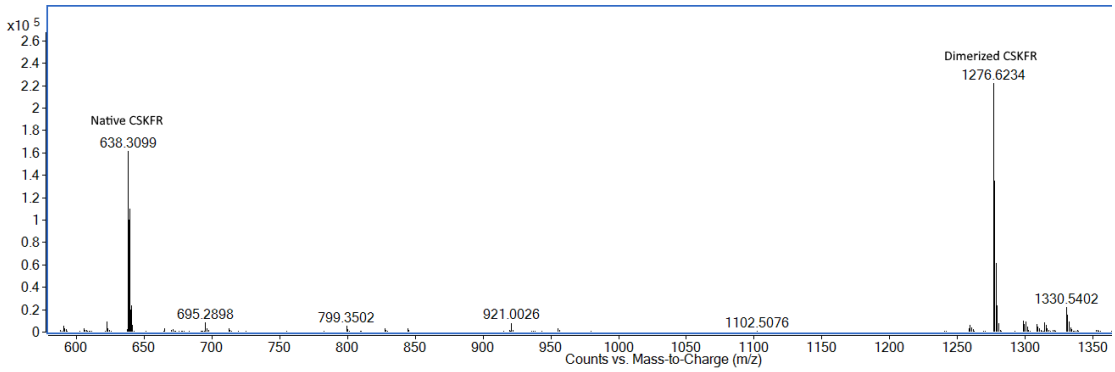
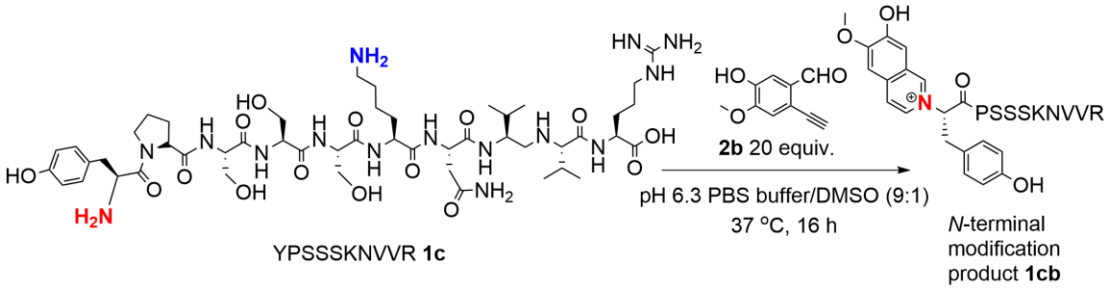


Figure 2.4 Deconvoluted spectrum of native CSKFR under 50 mM PBS buffer pH 6.3 at 37 ^{o}C for 16 h.

Therefore, a panel of reducing agents was used and the cysteine bioconjugation was performed according to a literature reported procedure.[8] The bioconjugation was found to be successful when 50 equiv. (5 mM) of tris (2-carboxyethyl) phosphine (TCEP) was incorporated to CSKFR (0.1 mM) and treated with reagent **2b** (2 mM, 20 equiv.) in pH 6.3 50 mM PBS/DMSO (9:1, 100 μL), flushed the eppendorf tube with nitrogen gas, and closed tightly to allow the reaction mixture to react for 16 h in 37 ^{o}C. With this optimized conditions, CSKFR offered good conversion (75%) and excellent *N*-terminal selectivity (>99:1).

2.2.7 Modification of peptide YPSSSKNVVR 1c using 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde 2b

In 2015, Francis's group reported a one-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes (2-PCA) with various applications (Scheme 2.13).[6b] Generally, this bioconjugation has provided satisfactory reactivity and site specificity, except for an enzyme Asparaginase (with *N*-terminal and neighbor residue Lys-Pro) that showed no conversion, which was presumably due to the proline-precluded cyclization. Therefore, YPSSSKNVVR was used to verify the reactivity of reagent **2b** towards a peptide which possesses a proline at the second residue.



Scheme 2.13 Peptide YPSSSKNVVR **1c** modification using reagent **2b.**

Table 2.10 Modification of peptide YPSSSKNVVR 1c using reagent 2b[a]

Peptide sequence	Conversion (%)[b]			*N*-terminal Selectivity of mono-modified peptide[c]
	+1 mod	+2 mod	Total	
YPSSSKNVVR	60	20	(80)	>99:1
YTSSSKNVVR	73	13	(86)	>99:1

[a] Conditions: YPSSSKNVVR **1c**/ YTSSSKNVVR **1a** (0.1 mM), reagent **2b** (2 mM, 20 equiv.), in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL), 16 h, 37 ^{o}C. [b] Determined by total ion count (TIC) of LC-MS analysis. [c] Ratio of *N*-terminal α-amino group modified peptide to Lys ε-amino group modified peptide; this ratio was determined by extracted ion chromatogram (EIC) of LCMS analysis.

The results indicated that no significant difference between YTSSSKNVVR and YPSSSKNVVR, in terms of reactivity and *N*-terminal selectivity which verified that the proline in second residue of peptide have no distinct effect to the bioconjugation reaction using reagent **2b** (Table 2.10).

2.2.8 Study of the stability of 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde 2b–modified YTSSSKNVVR 1ab

The stability of the modified-peptide conjugates is another important issue in bioconjugation chemistry. Strategies like oxidative coupling reaction rely on addition of 3,4-dihydroxy-phenylalanine (DOPA) or *o*-aminophenol to *o*-iminoquinones and *o*-quinones that are generated *in situ* using potassium ferricyanide or sodium periodate,[9-10] their conjugate products contain electron-rich aromatic rings which are unstable to be stored under ambient conditions, limiting their applications (Figure 2.5). Therefore, an ideal modified-protein/peptide conjugate should be stable under ambient conditions and towards biological oxidants and reductants.

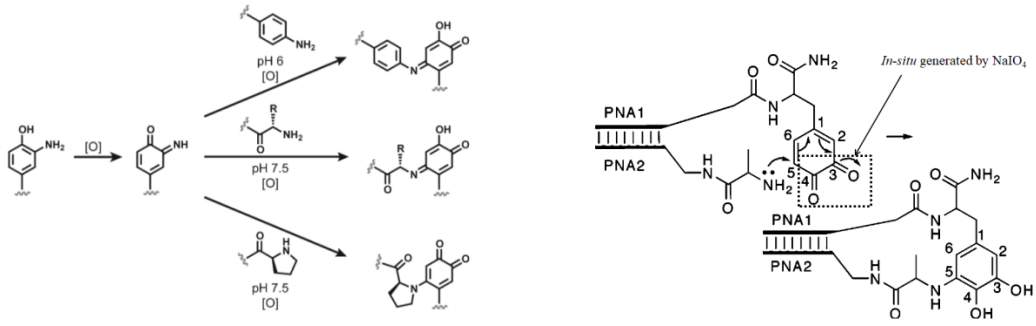
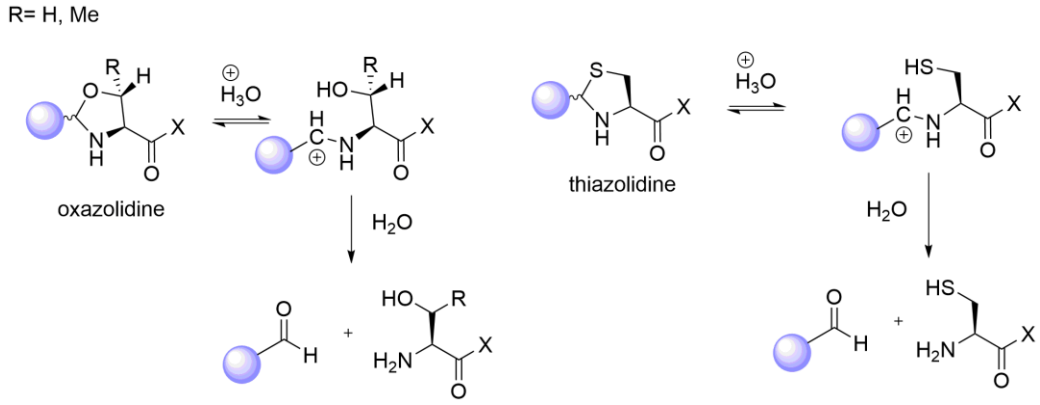


Figure 2.5 (Top) *N*-Terminal Modification of Proteins with o-Aminophenols *via in situ* oxidative coupling using potassium ferricyanide.[10a] (Bottom) Periodate-Mediated Cross-Linking of DOPA-Containing Molecules to Proteins.[10b]

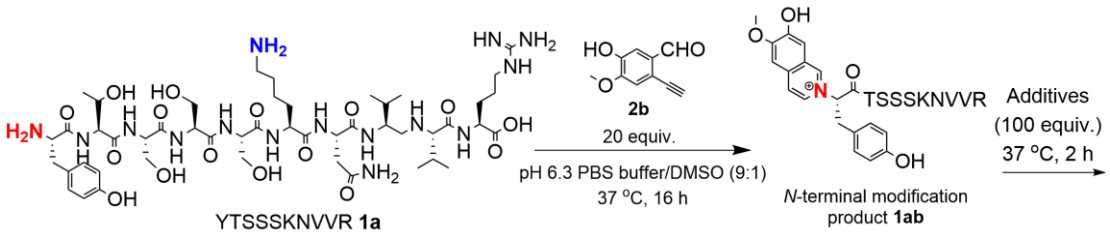
The stability of the **2b**-modified peptide **1ab** was examined with different biologically relevant thiols [9] (cysteine, homocysteine and GSH), and in the presence of oxidizing (H_2O_2) or reducing reagents (ascorbic acid, DTT, TCEP, 2-mercaptoethanol), treatment with 100 equiv. of the reagents at 37 ^oC for 2 h (Table 2.11). The isoquinolinium product was found to be stable, with no decomposition or scrambling product observed. Notably, the isoquinolinium-modified conjugate was stable in reducing environment, in which reduction of the distorted π-system was expected due to the formal positive charge nitrogen in the aromatic ring. Moreover, the isoquinolinium-modified conjugate was stable in nucleophiles such as thiols, which can be explained by hard and soft acids and bases theory (HSAB theory);[11] due to the positive charge on isoquinolinium-modified conjugate, the conjugate is regard as “hard acid” which is less susceptible towards a soft base like thiol.

The condensation product between aldehydes and *N*-terminal α-amino groups such as ox/thiazolidines, are prone to pH-dependent hydrolysis (Scheme 2.14).[12] Note that the isoquinolinium peptide conjugate was found to be stable in pH 6.3 50 mM PBS buffer in open atmosphere after 2 weeks at 4 ^{o}C, indicating isoquinolinium is a stable product.



Scheme 2.14 pH-dependent hydrolysis of (left) oxazolidine; (right) thiazolidine.[12]

Table 2.11 The stability of the isoquinolinium product



Entry	Additives[a]	% of *N*-terminal modification product 1ab
1	None	73
2	GSH	73
3	Homocysteine	74
4	Cysteine	69
5	DTT	74
6	2-mercaptoethanol	74
7	TCEP	73
8	Ascorbic acid	76
9	H_2O_2	76
10	None[b]	71

[a] Modified isoquinolinium product **1ab** treated with different biological relevant thiols, under reducing or oxidizing reagents. The data shown are based on ESI-TOF mass spectrometry analysis. [b] Product stored under 4 °C for two weeks.

2.3 Conclusion

In summary, a one-step facile *N*-terminal selective peptide and protein modification using 2-ethynylbenzaldehydes was developed. A redox, thiol stable isoquinolinium-peptide conjugate formed through imine formation between the *N*-terminal α-amino/lysine ε-amino group and aldehyde, followed by cyclization with the proximal ethynyl group. Most of the peptide modifications (except proline as *N*-terminal residue) could be achieved with high efficiency (up to 93% conversion), and excellent *N*-terminal selectivity (up to >99:1) under mild reaction conditions. In studies on the modification of peptide library XSKFR with 20 canonical *N*-terminal amino acids, 13 out of the 20 amino acids achieved excellent *N*-terminal selectivity (>99:1) with good conversion (average 76%). A model study of the isoquinolinium formation was conducted and the reaction mechanism was proposed.

2.4 Experimental Section

General Procedure

All reagents were commercially available and used without further purification. Milli-Q® water from a Milli-Q® Gradient A10 system (Millipore, Billerica, USA) was used as reaction solvent in peptide modification and LC-MS. Flash column chromatography was performed using silica gel 60 (230–400 mesh, ASTM) with *n*-hexane/EtOAc or CH_3OH/CH_2Cl_2 as eluent. ^{1}H, ^{13}C and ^{19}F NMR spectra were recorded on Bruker DPX-400 spectrometer. The chemical shifts are expressed in ppm and coupling constants are given in Hz. Data for ^{1}H NMR are recorded as follows: chemical shift (δ, ppm), multiplicity (s, singlet; br s, broad singlet; d, doublet; dd, double doublet; t, triplet; td, triplet of doublets; m, multiplet), coupling constant (Hz), integration. Data for ^{13}C NMR are reported in terms of chemical shift (δ, ppm). Low resolution mass spectra (MS) and high resolution mass spectra (HR-MS) were obtained on an Agilent 6540 UHD Accurate-Mass Q-TOF LC/MS system equipped with an ion spray source in the positive ion mode.

ESI–MS Analysis of Peptide Modification

The mass spectrometry was performed over a m/z range of 100-3000 on an Agilent 6540 UHD Accurate-Mass Q-TOF LC/MS system with positive ESI, and the raw

spectra were deconvoluted by the MassLynx 4.1 Transform Program (Waters, Manchester, UK). Desolvation and source temperatures were 150 ℃ and 80 ℃ respectively. Operating conditions optimized for the detection of reaction mixtures were the following: capillary voltage 3 kV, sample cone voltage 30 V, extraction voltage 4 V and collision cell voltage 10 eV.

LC–MS Analysis of Peptide Modification

Mass spectrometry analysis was performed using the ESI source of Agilent 6540 UHD Accurate-Mass Q-TOF LC/MS system in the positive ion mode. The CapLC® system (Waters,Manchester, UK) was equipped with a Poroshell 300SB-C18 column (1.0 mm ID ×75 mm , 5μm) with ZORBAX Poroshell guard column (1.0 mm ID ×17 mm, 5 μm) (Agilent-TechnologiesInc., Wilmington,USA). Mobile-phase A was made of 0.5% formic acid in Milli-Q® water. Mobile-phase B was made of 0.5% formic acid in CH_3CN. 2 μl of sample was injected with a flow rate of 40 μl/min at 25 ^oC. The initial conditions for separation were 3% B for 3 min, followed by a linear gradient to 70% B by 30 min, 3% B by 31 min then to 3% B by 45 min. The mass spectrometry analysis was performed over a mass range of *m/z* 100–3000, and the raw spectra were deconvoluted by the MassLynx 4.1 Transform Program (Waters, Manchester, UK). Desolvation and source temperatures were 150 ℃ and 80 ℃ respectively. Operating conditions optimized for the detection of reaction mixtures

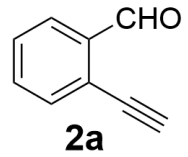
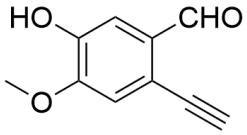
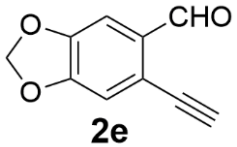
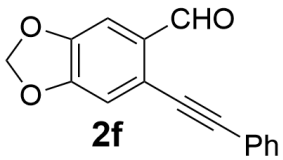
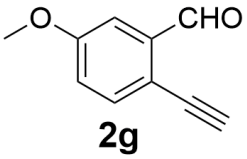
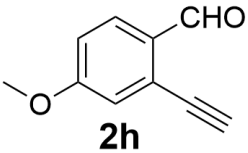
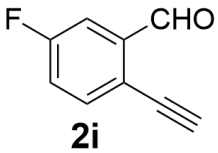
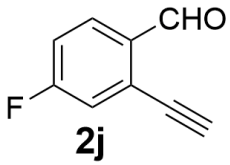
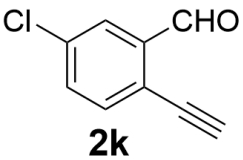
were the following: capillary voltage 3 kV, sample cone voltage 30 V, extraction voltage 4 V and collision cell voltage 10 eV.

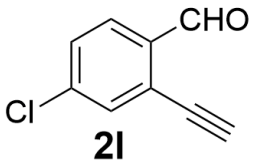
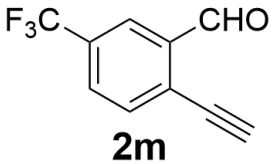
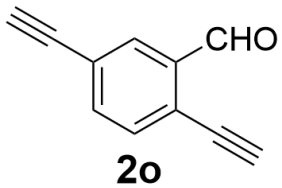
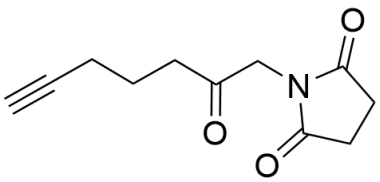
Calculation of Peptide Conversion

The crude reaction mixtures of unmodified peptides (starting) and modified peptides (product) were subjected to LC-MS or ESI-MS analysis. After data processing by MassLynx 4.1 Transform Program, peptide conversion at different time intervals was determined by measuring the relative peak intensities of peptide and product in the mass spectrum as follows:

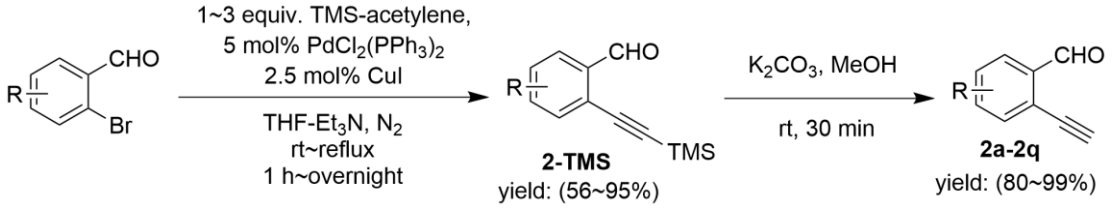
$$Peptide\ Conversion = \frac{\text{Relative Peak Intensity of Product (M)} + \text{Relative Peak Intensity of Hydrated Product (M + H2O)}}{\text{Relative Peak Intensity of Starting and Product}}\ x\ 100\%$$

Literature References

 **2a**	Ouyang, H. C.; Tang, R. Y.; Zhong, P.; Zhang, X. G.; Li, J. H. *J. Org. Chem*., **2011**, *76*, 223.
 **2b**	Cao, Z. P.; Zhang, H. Q.; Zhang, X. X.; Zhang, L. D.; Meng, X.; Chen, G.; Zhao, X. E.; Sun, X. J.; You, J. M. *RSC Adv*., **2015**, *5*, 103155.
 **2e**	Takahashi, H.; Yoshida, K.; Yanagisawa, A. *J. Org. Chem*., **2009**, *74*, 3632.
 **2f**	Huang, Q. H.; Hunter, J. A.; Larock, R. C. *J. Org. Chem*., **2002**, *67*, 3437.
 **2g**	Ohta, Y.; Kubota, Y.; Watabe, T.; Chiba, H.; Oishi, S.; Fujii, N.; Ohno, H. *J. Org. Chem*., **2009**, *16*, 6299.
 **2h**	Li, S. Y.; Luo, Y.; Wu, J. *Org. Lett*., **2011**, *13*, 3190.
 **2i**	Ohta, Y.; Kubota, Y.; Watabe, T.; Chiba, H.; Oishi, S.; Fujii, N.; Ohno, H. *J. Org. Chem*., **2009**, *16*, 6299.
 **2j**	Ohta, Y.; Kubota, Y.; Watabe, T.; Chiba, H.; Oishi, S.; Fujii, N.; Ohno, H. *J. Org. Chem*., **2009**, *16*, 6299.
 **2k**	Li, M. M.; Xing, P.; Huang, Z. G.; Jiang, B. *Chin. J. Chem*., **2013**, *31*, 49.

 **2l**	Wu, X.; Li, M. L.; Chen, D. F.; Chen, S.-S. *J. Org. Chem*., **2014**, *79*, 4743.
 **2m**	He, J.; Shi, Y. P.; Cheng, W. L.; Man, Z. M.; Yang, D. D.; Li, C. Y. *Angew. Chem., Int. Ed*., **2016**, *55*, 4557.
 **2o**	Zhou, N. Z.; Wang, L.; Thompson, D. W.; Zhao, Y. M. *Tetrahedron*, **2011**, *67*, 125.
 **2t**	Jagadish, B.; Sankaranarayanan, R.; Xu, L. P.; Richards, R.; Vagner, J.; Hruby, V. J.; Gillies, R. J.; Mash, E. A. *Bioorg. Med. Chem. Lett*., **2007**, *17*, 3310.

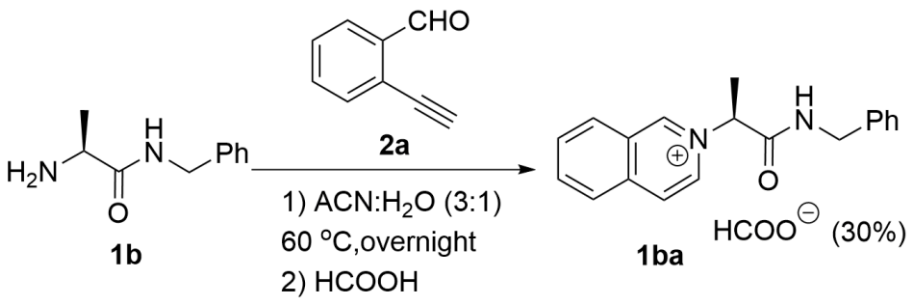
General procedure for preparation of 2-ethynylbenzaldehydes 2a-2q



To a Schlenk flask with $PdCl_2(PPh_3)_2$ (5 mol%) and CuI (2.5 mol%) under N_2 atmosphere, a solution of *o*-bromobenzaldehyde derivatives (5 mmol) in THF (50 mL) and Et_3N (0.4 M) was added at room temperature, followed by ethynyltrimethylsilane/ 1-phenyl-2-trimethylsilylacetylene (6 mmol, 1.2 equiv.). The mixture was then heated at 80 ^oC for overnight . After filtration of the mixture with celite, the filtrate was concentrated under reduced pressure and the residue was purified by column chromatography over silica gel with hexane/AcOEt (50:1) as the eluent to give **2-TMS**. **2-TMS** was reacted with K_2CO_3 (0.50 g, 3.64 mmol) in MeOH (30 mL) for 30 min at room temperature, and the solvent was removed under reduced pressure. The residue was extracted with CH_2Cl_2 and washed with saturated aqueous Na_2CO_3, and dried over $MgSO_4$. The filtrate was concentrated under reduced pressure and purified by column chromatography over silica gel with hexane/AcOEt (50:1) to give 2-ethynylbenzaldehyde derivatives **2a-2q**.

Reference: Ohta, Y.; Kubota, Y.; Watabe, T.; Chiba, H.; Oishi, S.; Fujii, N.; Ohno, H. *J. Org. Chem.*, **2009**, *16*, 6299.

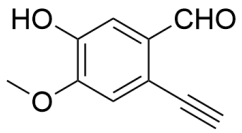
Preparation of isoquinolinium product 1ba



To a 25 mL round bottom flask with (*S*)-2-amino-*N*-benzylpropanamide **1b** (1 mmol, 178 mg) and 2-ethynylbenzaldehyde **2a** (1.1 mmol, 143 mg), a mixture of acetonitrile and distilled water (3:1, 20 mL) was added. The reaction mixture was heated at 60 ^{o}C overnight. Then, 2 equiv. of formic acid (2 mmol, 75.4 µL) was added to the mixture and stirred for 15 min. After that, the reaction mixture was concentrated under reduced pressure, and 10 mL of water was added. After filtration of the mixture with filter paper, the filtrate was concentrated under reduced pressure and the residue was purified by column chromatography over silica gel with dichloromethane/MeOH (9:1) as the eluent to give the isoquinolinium product **1ba**.

General procedure for modification of peptides using 2-ethynylbenzaldehydes

To an eppendorf tube (1.5 mL) with 80 µL of 50 mM PBS buffer pH 6.3, 10 µL of 1 mM peptide stock solution (dissolved in Milli-Q® water) was added to the buffer, followed by 10 µL of 20 mM 2-ethynylbenzaldehyde stock solution (dissolved in DMSO). The reactive mixture was allowed to react in a 37 ^{o}C water bath for 16 h. 10 µL of the mixture was drawn, diluted with 10 µL Milli-Q® and subjected to LC/MS-MS analysis.

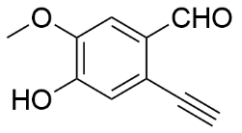


2b

2-Ethynyl-5-hydroxy-4-methoxybenzaldehyde (2b)

White solid, 50% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.37 (s, 1H), 7.45 (s, 1H), 7.03 (s, 1H), 5.90 (s, 1H), 3.98 (s, 3H), 3.35 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 190.40, 151.03, 146.96, 131.80, 119.15, 114.88, 112.67, 82.54, 79.44, 56.49. HRMS (ESI): calcd. for $C_{10}H_9O_3$ $(M + H)^+$: 177.0552, found: 177.0546.

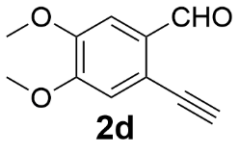


2c

2-Ethynyl-4-hydroxy-5-methoxybenzaldehyde (2c)

White solid, 45% yield

^{1}H NMR (400 MHz, $(CD_3)_2SO$): δ 10.37 (s, 1H), 7.42 (s, 1H), 7.11 (s, 1H), 6.15 (s, 1H), 3.96 (s, 3H), 3.35 (s, 1H). **^{13}C NMR** (100 MHz, $(CD_3)_2SO$): δ 189.54, 152.96, 149.43, 129.43, 119.80, 119.64, 109.61, 86.21, 79.53, 56.16. HRMS (ESI): calcd. for $C_{10}H_9O_3$ $(M + H)^+$: 177.0552, found: 177.0543.

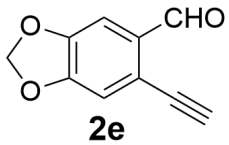


2d

2-Ethynyl-4,5-dimethoxybenzaldehyde (2d)

White solid, 70% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.39 (s, 1H), 7.40 (s, 1H), 7.02 (s, 1H), 3.96 (s, 3H), 3.95 (s, 3H), 3.38 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 190.19, 153.53, 150.12, 130.99, 120.07, 114.94, 108.23, 82.85, 79.17, 56.23, 56.15. HRMS (ESI): calcd. for $C_{11}H_{11}O_3$ $(M + H)^+$: 191.0630, found: 191.0700.

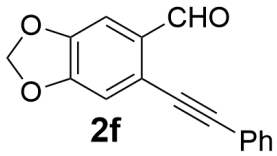


2e

6-Ethynylbenzo[d][1,3]dioxole-5-carbaldehyde (2e).

White solid, 65% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.37 (s, 1H), 7.35 (s, 1H), 6.99 (s, 1H), 6.09 (s, 2H), 3.38 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 189.92, 152.40, 149.23, 133.19, 122.25, 112.81, 106.25, 102.61, 83.23, 79.16. HRMS (ESI): calcd. for $C_{10}H_7O_3$ $(M + H)^+$: 175.0395, found: 175.0391.

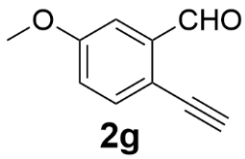


2f

6-(Phenylethynyl)benzo[d][1,3]dioxole-5-carbaldehyde (2f)

White solid, 90% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.49 (s, 1H), 7.55 – 7.52 (m, 2H), 7.39 – 7.37 (m, 4H), 7.03 (s, 1H), 6.09 (s, 2H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 190.17, 152.54, 148.85, 145.22, 132.27, 131.72, 129.12, 128.66, 123.77, 122.46, 112.13, 106.24, 102.54, 101.30, 95.28, 84.89. HRMS (ESI): calcd. for $C_{16}H_{11}O_3$ $(M + H)^+$: 251.0708, found: 251.0703.

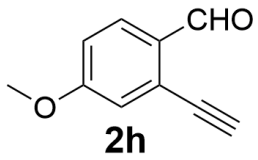


2g

2-Ethynyl-5-methoxybenzaldehyde (2g).

White solid, 80% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.49 (s, 1H), 7.53 (d, *J* = 8.5 Hz, 1H), 7.39 (s, 1H), 7.10 (d, *J* = 8.5 Hz, 1H), 3.86 (s, 3H), 3.37 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 191.48, 160.27, 138.12, 135.34, 121.62, 118.25, 110.07, 82.89, 79.30, 55.78. HRMS (ESI): calcd. for $C_{10}H_9O_2$ $(M + H)^+$: 161.0602, found: 161.0791.

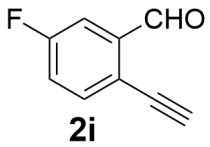


2h

2-Ethynyl-4-methoxybenzaldehyde (2h)

White solid, 83% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.38 (s, 1H), 7.90 (d, *J* = 8.8 Hz, 1H), 7.06 (s, 1H), 6.99 (d, *J* = 8.7 Hz, 1H), 3.88 (s, 3H), 3.44 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 190.24, 163.83, 130.43, 129.64, 127.67, 118.11, 116.11, 84.09, 79.31, 55.86. HRMS (ESI): calcd. for $C_{10}H_9O_2$ $(M + H)^+$: 161.0602, found: 161.0792.

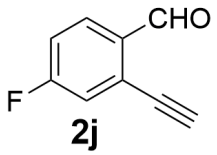


2i

2-Ethynyl-4-fluorobenzaldehyde (2i).

White solid, 90% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.49 (d, *J* = 3.1 Hz, 1H), 7.64 – 7.59 (m, 2H), 7.31 – 7.27 (m, 1H), 3.45 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 190.32, 162.86 (d, *J* = 253.3 Hz), 138.79 (d, *J* = 6.9 Hz), 136.10 (d, *J* = 7.8 Hz), 121.70 (d, *J* = 3.2 Hz), 121.40 (d, *J* = 22.8 Hz), 113.92 (d, *J* = 23.0 Hz), 84.19, 78.35. **^{19}F NMR** (376 MHz, $CDCl_3$): δ -107.89. HRMS (ESI): calcd. for C_9H_6FO $(M + H)^+$: 149.0325, found: 149.0333.

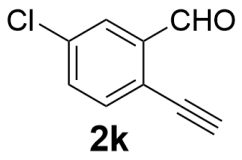


2j

2-Ethynyl-5-fluorobenzaldehyde (2j).

White solid, 82% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.44 (s, 1H), 7.97 – 7.93 (m, 1H), 7.29 – 7.27 (m, 1H), 7.19 – 7.15 (m, 1H), 3.52 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 189.72, 165.53 (d, *J* = 257.1 Hz), 133.33 (d, *J* = 3.0 Hz), 130.09 (d, *J* = 10.1 Hz), 127.92 (d, *J* = 10.9 Hz), 120.53 (d, *J* = 23.7 Hz), 117.18 (d, *J* = 22.1 Hz), 85.39, 78.01 (d, *J* = 2.9 Hz). **^{19}F NMR** (376 MHz, $CDCl_3$): δ -103.09. HRMS (ESI): calcd. for C_9H_6FO $(M + H)^+$: 149.0325, found: 149.0341.

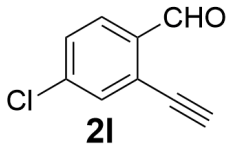


2k

5-Chloro-2-ethynylbenzaldehyde (2k)

White solid, 70% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.47 (s, 1H), 7.89 (s, 1H), 7.57 – 7.52 (m, 2H), 3.51 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 190.19, 137.80, 136.03, 135.24, 133.89, 127.41, 123.85, 85.34, 78.38. HRMS (ESI): calcd. for C_9H_6ClO $(M + H)^+$: 165.0107, found: 165.0115.

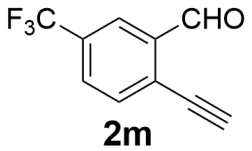


2l

4-Chloro-2-ethynylbenzaldehyde (2l)

White solid, 64% yield

^{1}H NMR (400 MHz, CD_3CN): δ 10.40 (s, 1H), 7.84 (d, *J* = 8.4 Hz, 1H), 7.69 (d, *J* = 2.0 Hz, 1H), 7.57 – 7.54 (m, 1H), 3.90 (s, 1H). **^{13}C NMR** (100 MHz, CD_3CN): δ 189.99, 139.60, 135.21, 133.50, 129.88, 128.87, 126.50, 86.22, 77.60. HRMS (ESI): calcd. for C_9H_6ClO $(M + H)^+$: 165.0107, found: 165.0111.

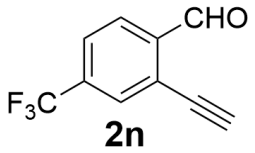


2m

2-Ethynyl-4-(trifluoromethyl)benzaldehyde (2m)

Pale yellow solid, 70% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.55 (s, 1H), 8.20 (s, 1H), 7.82 (d, *J* = 8.1 Hz, 1H), 7.75 (d, *J* = 8.1 Hz, 1H), 3.62 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 189.86, 136.87, 134.51, 131.46 (q, *J* = 33.8 Hz), 129.96 (q, *J* = 3.4 Hz), 128.65, 123.19 (q, *J* = 273.7 Hz), 124.39 (q, *J* = 3.8 Hz), 86.79, 78.02. **^{19}F NMR** (376 MHz, $CDCl_3$): δ -63.25. HRMS (ESI): calcd. for $C_{10}H_6F_3O$ $(M + H)^+$: 199.0371, found: 199.0373.

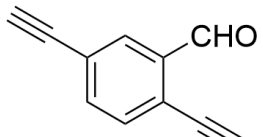


2n

2-Ethynyl-4-(trifluoromethyl)benzaldehyde (2n)

Pale yellow solid, 66% yield

^{1}H NMR (400 MHz, CD_3CN): δ 10.51 (s, 1H), 8.03 (d, *J* = 8.2 Hz, 1H), 7.97 (s, 1H), 7.83 (d, *J* = 8.1 Hz, 1H), 3.97 (s, 1H). **^{13}C NMR** (100 MHz, CD_3CN): δ 190.28, 139.00, 134.36 (q, *J* = 32.9 Hz), 130.84 (q, *J* = 3.9 Hz), 128.17, 126.03 (q, *J* = 3.6 Hz), 125.64, 123.29 (q, *J* = 272.4 Hz), 86.65, 77.59. **^{19}F NMR** (376 MHz, CD_3CN): δ -64.03. HRMS (ESI): calcd. for $C_{10}H_6F_3O$ $(M + H)^+$: 199.0371, found: 199.0359.

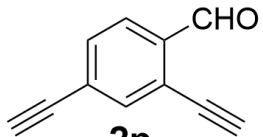


2o

2,5-Diethynylbenzaldehyde (2o)

White solid, 76% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.49 (s, 1H), 8.03 (s, 1H), 7.65 (d, *J* = 8.0 Hz, 1H), 7.57 (d, *J* = 8.0 Hz, 1H), 3.55 (s, 1H), 3.23 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 190.43, 136.64, 136.50, 133.87, 130.96, 125.32, 123.47, 85.91, 81.79, 80.38, 78.78. HRMS (ESI): calcd. for $C_{11}H_7O$ $(M + H)^+$: 155.0497, found: 155.0495.

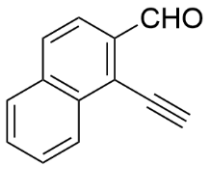


2p

2,4-Diethynylbenzaldehyde (2p)

White solid, 75% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.48 (s, 1H), 7.86 (d, *J* = 7.9 Hz, 1H), 7.69 (s, 1H), 7.53 (d, *J* = 7.7 Hz, 1H), 3.48 (s, 1H), 3.30 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 190.49, 137.27, 135.97, 132.59, 127.92, 127.23, 125.57, 84.98, 81.73, 78.31. HRMS (ESI): calcd. for $C_{11}H_7O$ $(M + H)^+$: 155.0497, found: 155.0482.

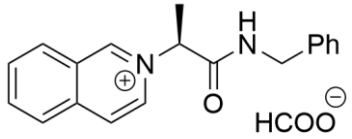


2q

1-Ethynyl-2-naphthaldehyde (2q)

Pale yellow solid, 63% yield

^{1}H NMR (400 MHz, $CDCl_3$): δ 10.79 (s, 1H), 8.55 – 8.53 (m, 1H), 7.99 – 7.96 (m, 1H), 7.92 – 7.88 (m, 2H), 7.70 – 7.65 (m, 2H), 3.92 (s, 1H). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 191.93, 135.66, 135.30, 133.42, 129.50, 129.44, 128.47, 127.89, 127.11, 126.03, 121.90, 90.10, 77.14. HRMS (ESI): calcd. for $C_{13}H_9O$ $(M + H)^+$: 181.0653, found: 181.0640.



1ba

(*S*)-2-(1-(benzylamino)-1-oxopropan-2-yl)isoquinolin-2-ium formate (1ba)

White solid, 30% yield

^{1}H NMR (400 MHz, CD_3OD): δ 10.03 (s, 1H), 8.71 (d, *J* = 6.6 Hz, 1H), 8.55 (d, *J* = 8.3 Hz, 1H), 8.49 (d, *J* = 6.6 Hz, 1H), 8.38 (s, 1H), 8.31 (m, 2H), 8.10 (t, *J* = 7.4 Hz, 1H), 7.36 – 7.24 (m, 5H), 5.79 (d, *J* = 6.7 Hz, 1H), 4.45 (q, *J* = 14.8 Hz, 2H), 2.10 (d, *J* = 6.5 Hz, 3H). **^{13}C NMR** (100 MHz, CD_3OD): δ 167.61, 149.65, 138.15, 137.77, 137.46, 134.06, 131.24, 130.62, 128.30, 128.29, 128.28, 128.27, 127.63, 127.39, 127.17, 127.08, 125.55, 68.46, 43.41, 17.60. HRMS (ESI): calcd. for $C_{19}H_{19}N_2O^+$ $(M - COOH)^+$: 291.1498, found: 291.1493.

Reference

1. Sonogashira, K.; Tohda, Y.; Hagihara, N. *Tetrahedron Lett*., **1975**, *16*, 4467.

2. (a) Henderson, L. *J. Am. J. Physiol*., **1908**, *4*, 173; (b) Hasselbalch, K. A. *Biochem. Z.*, **1917**, *78*, 112.

3. Hammett, L. P. *J. Am. Chem. Soc*., **1937**, *59*, 96.

4. Hansch, C.; Leo, A.; Taft, R. T. *Chem. Sov. Rev*., **1991**, *91*, 165.

5. (a) Godoy-Alcantar, C.; Yatsimirsky, A. K.; Lehn, J. M. *J. Phys. Org. Chem*., **2005**, *18*, 979; (b) Kulchat, S.; Chaur, M. N.; Lehn, J. M. *Chem. Eur. J*., **2017**, *23*, 11108.

6. (a) Cal, P. M. S. D.; Vicente, J. B.; Pires, E.; Coelho, A. V.; Veiros, L. F.; Cordeiro, C.; Gois, P. M. P. *J. Am. Chem. Soc*., **2012**, *134*, 10299; (b) MacDonalde, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nat. Chem. Bio*., **2015**, *11*, 326; (c) Tung, C. L.; Wong, C. T. T.; Fung, E. Y. M.; Li, X. *Org. Lett*., **2016**, *18*, 2600.

7. (a) Gowda, G. A. N.; Gowda, Y. N.; Raftery, D. *Anal. Chem*., **2015**, *7*, 3800; (b) Wright, H. T.; Urry, D. W. *Crit. Rev. Biochem. Mol. Biol*., **1991**, *1*, 1.

8. (a) Hermanson, G. T. *Bioconjugate Techniques*, Academic Press, San Diego, CA, 1996; (b) Brinkley, M. *Bioconjug. Chem*., **1992**, *3*, 2.

9. Mansoor, M. A.; Svardal, A. M.; Ueland, P. M. *Anal. Biochem*., **1992**, *200*, 218.

10. (a) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. *J. Am. Chem. Soc*., **2014**, *136*, 9572; (b) Liu, B.; Burdine, L.; Kodadek, T. *J. Am. Chem. Soc*., **2006**, *128*, 15228.

11. (a) Ralph, G. P. *J. Chem. Educ.*, **1968**, *10*, 643; (b) Ralph, G. P. *Chemical Hardness*, Wiley-VCH , New Jersey, USA. 1997.

12. (a) Fife, H. T.; Hagopian, L. *J. Am. Chem. Soc*., **1968**, *4*, 1007; (b) Fife, H. T.; Natarajan, R.; Shen, C. C.; Bembi, R. *J. Am. Chem. Soc*., **1991**, *8*, 3071; (c) Geoghegan, K. F.; Stroh, J. G. *Bioconjug. Chem*., **1992**, *3*, 138; (d) Chen, J. K.; Lane, W. S.; Brauer, A. W.; Tanaka, A.; Schreiber, S. L. *J. Am. Chem. Soc*., **1993**, *115*, 12592; (e) Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. *Tetrahedron Lett*., **2000**, *41*, 4069.

Chapter 3.

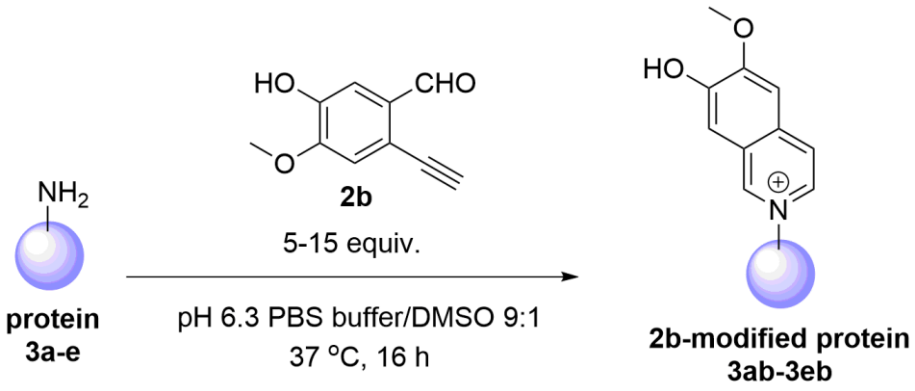
N-Terminal Modification of Proteins using 2-Ethynylbenzaldehydes

3.1 Modification of proteins using 2-ethynylbenzaldehydes

3.1.1 Modification of different proteins using 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde 2b

The reactivity and stability of 2-ethynylbenzaldehydes in peptides bioconjugation had been studied in the previous chapter. This method was then used for bioconjugation of proteins, notably, a therapeutic protein Arginase.

In order to demonstrate the applicability of this bioconjugation reaction to proteins, a panel of proteins (insulin **3a**, lysozyme **3b**, ribonuclease (RNaseA) **3c**, human serum albumin (HSA) **3d** and bovine serum albumin (BSA) **3e)** were employed. According to our previous studies on peptide ligation, the optimized reaction conditions were found to be 20 equiv. of 2-ethynylbenzaldehyde, in 50 mM PBS buffer pH 6.3 at 37 ^{o}C for 16 h (Scheme 3.1). Nevertheless, the reagents used in protein bioconjugation should be minimized to avoid multiple sites modification.



Scheme 3.1 General scheme for protein modification using 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde **2b**.

The proteins selected for bioconjugation are based on two criteria: protein size and number of lysine residues on the protein surface. Therefore, the *N*-terminal selectivity and reactivity of this bioconjugation reaction could be tested with proteins that possess different sizes and numbers of surface lysine residues (Table 3.1).[1] A panel of proteins was employed, including small proteins with various numbers of surface lysine residues (insulin, lysozyme and RNaseA) and that of large proteins (HSA and BSA). The bioconjugation of proteins (0.1 mM) with 5-15 equiv. of reagent **2b** (0.5-1.5 mM) in pH 6.3 50 mM of PBS buffer/DMSO (9:1) at 37 ^{o}C for 10-16 h. All the proteins afforded moderate to good conversion (53-71%), with average conversion of 60% (Table 3.2).

Table 3.1 Characteristics of different proteins

Protein	*N*-terminal amino acid residue	Mass (kDa)	Number of surface lysine residues
insulin	Chain A: Gly; Chain B: Phe	5.7	1
lysozyme	Lys	14.3	5
RNaseA	Lys	13.6	10
HSA	Met	66.4	~59
BSA	Met	66.4	30~35

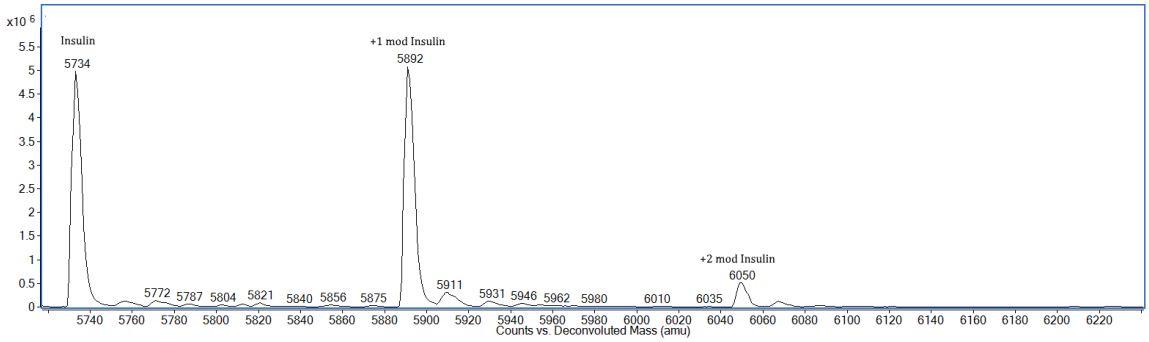
Table 3.2 Modification of different proteins using reagent 2b[a]

Protein	Equiv. of reagent 2b	Reaction time (h)	Conversion (%)[b]			
			+1 mod	+2 mod	+3 mod	Total
insulin	5	10	49	7	0	56
lysozyme	15	16	35	27	9	71
RNaseA	15	16	41	15	4	60
HSA	15	16	34	24	0	58
BSA	15	16	38	15	0	53

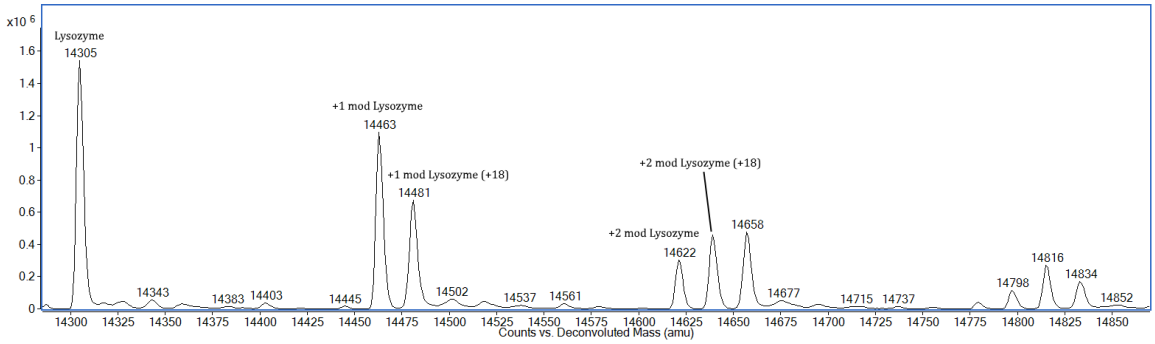
[a] Conditions: Protein (0.1 mM), reagent **2b** (5-15 mM, 5-15 equiv.), in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL), 10 or 16 h, 37 ^{o}C. [b] Determined by total ion count (TIC) of LC-MS analysis.

Despite the fairly satisfactory conversion of bioconjugation, considerably amount of +2 mod proteins are found in RNaseA and BSA (+2 mod: 15%), and particularly in lysozyme and BSA (+2 mod: 27 & 24% respectively) (Figure 3.1). The results indicated that inevitable double modification occurred in this bioconjugation, in which less amount of reagents and lower pH of buffer should be used to optimize the performance of this bioconjugation in proteins.

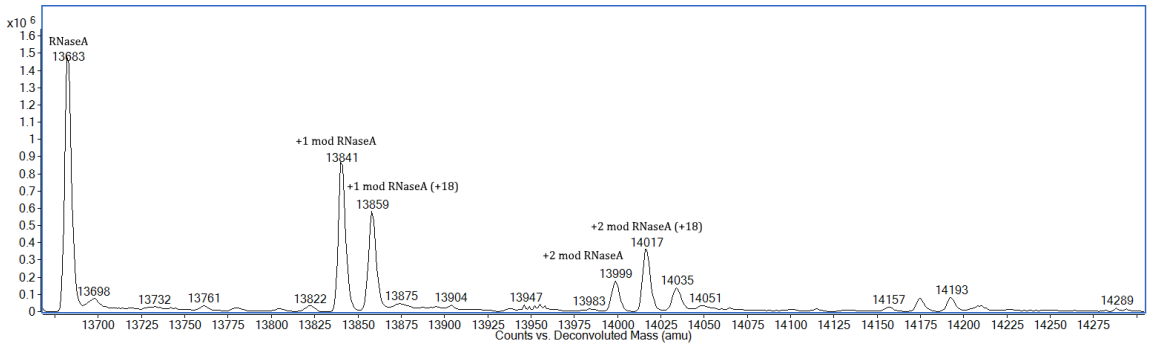
a) Mass deconvolution spectrum of modified insulin **3ab**



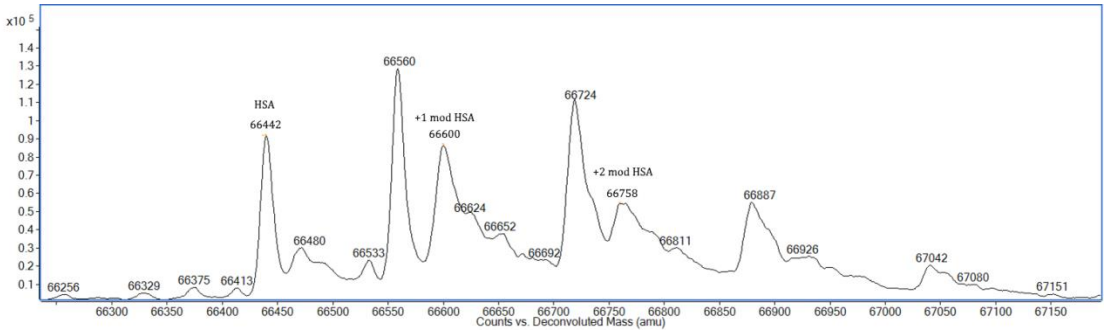
b) Mass deconvolution spectrum of modified lysozyme **3bb**



c) Mass deconvolution spectrum of modified RNaseA **3cb**



d) Mass deconvolution spectrum of modified HSA **3db**



e) Mass deconvolution spectrum of modified BSA **3eb**

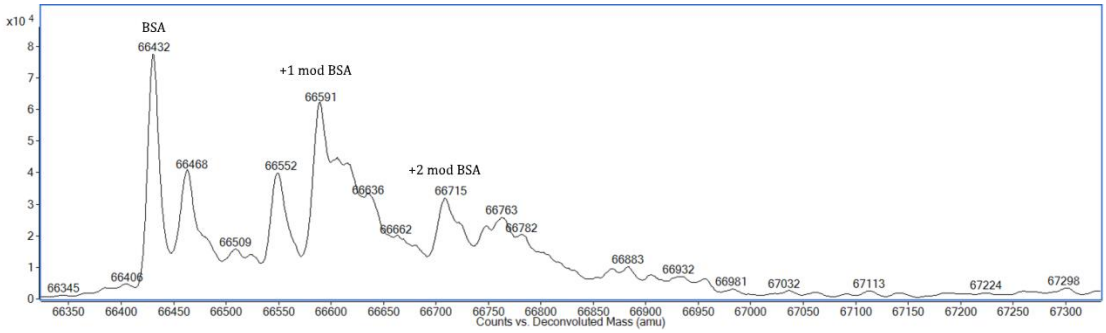


Figure 3.1 Mass deconvolution spectrum of a) modified insulin **3ab** and b) modified lysozyme **3bb** c) modified RNaseA **3cb** d) modified HSA **3db** e) modified BSA **3eb**.

3.1.2 Modification of therapeutic protein Arginase (BCArg) using 2-ethynyl-5-hydroxy- 4-methoxybenzaldehyde 2b

Therapeutic proteins are of importance to the development of novel therapeutics. From a therapeutic point of view, proteins offer the unique advantage of specific mechanisms of action and are highly potent.[2] However, therapeutic proteins often suffer from short half-lives, instability and low solubility. PEGylation is the most widely used method to improve the pharmacological profiles of proteins.[3]

Arginase (BCArg) **3f** is an anticancer therapeutic protein (*N*-terminal residue Met, possesses 12 Lys residues on the protein surface) under clinical trial developed by Leung and co-workers. Recently, Leung and his co-workers developed a pegylated analog BCArg-PEG20 which is a potential anti-cancer agent by selective inhibition of cancer cells growth.[4] The BCArg-PEG has shown anti-cancer activities on a variety of

cancers, including lung cancer, colorectal cancer, liver cancer, breast cancer, gastric and esophageal cancers.[4] Che and our group have previously prepared the alkyne-incorporated BCArg using the ketene-based protocol, the protein was then linked with the PEG moiety through click chemistry.[5]

Treatment of BCArg **3f** (0.1 mM) with reagent **2b** (5, 10, 15 and 20 equiv.) was first performed in 50 mM PBS buffer pH 6.3/DMSO (9:1) at 37 ^{o}C for 16 h (Table 3.3, entries 1-4).

Table 3.3 Modification of therapeutic protein Arginase (BCArg) using reagent 2b[a]

Entry	Equiv. of reagent 2b	Reaction time (h)	Conversion (%)[b]		
			+1 mod	+2 mod	Total
1	5	16	30	2	32
2	10	16	42	9	51
3	15	16	51	13	64
4	20	16	57	21	78
5	10	6	32	5	37
6	10	9	38	6	44

[a] Conditions: Protein (0.1 mM), reagent **2b** (5-20 mM, 5-20 equiv.), in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL), 6, 9 or 16 h, 37 ^{o}C. [b] Determined by total ion count (TIC) of LC-MS analysis.

Although the bioconjugation in entry 3 affords better conversion than that of entry 2, a larger amount of doubly-modified protein was observed. In contrast, treatment of a smaller amount of reagent **2b** (0.5 mM, 5 equiv.) reduced the conversion and no further conversion was observed beyond 16 h. Therefore, further optimization was employed by treatment of BCArg **3f** (0.1 mM) with reagent **2b** (1 mM, 10 equiv.) for a shorter reaction time (6 h and 9 h, entry 5 and 6). The optimized conditions of BCArg ligation are: treatment of reagent **2b** (10 equiv.) in 50 mM PBS buffer pH 6.3/DMSO (9:1) at 37 ^{o}C for 9 h (Figure 3.2).

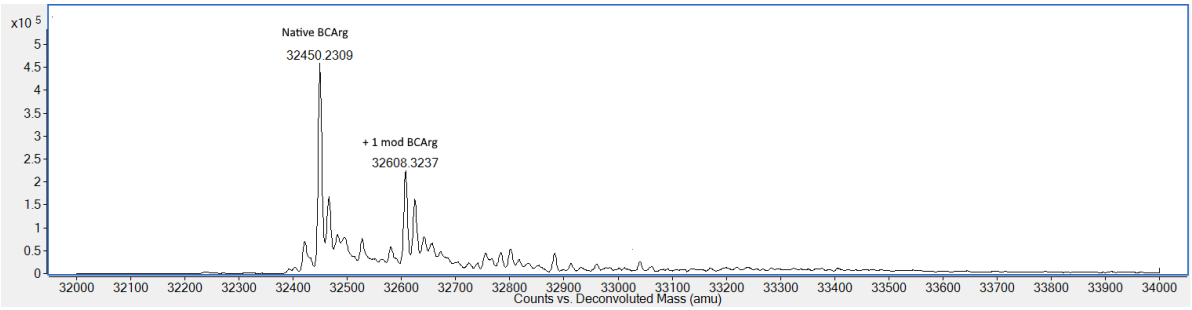


Figure 3.2 Deconvoluted spectrum of **2b**-modified BCArg.

3.3 Conclusion

In summary, a one-step facile protein modification using 2-ethynylbenzaldehyde has been demonstrated with good efficiency. Good protein conversion (average 60%) was achieved for different proteins, together with the therapeutic protein BCArg. This bioconjugation has been demonstrated to be applicable in protein modification. Our ongoing work is to analyze the site of protein modification by trypsin digestion followed by LC-MS/MS analysis, and develop fluorescent dye-linked 2-ethynylbenzaldehydes for protein bioconjugation.

3.4 Experimental Section

General procedure

High resolution mass spectra were obtained on an Agilent 6540 UHD Accurate-Mass Q-TOF LC/MS system equipped with an ion spray source in the positive ion mode. For ESI–MS/MS analysis, collision energy was set at 10, 15 or 20 eV. Milli-Q® water used as reaction solvent in protein modification and LC-MS was deionised using a Milli-Q® Gradient A10 system (Millipore, Billerica, USA).

Calculation of Protein Conversion

The crude reaction mixture of unmodified protein (starting) and modified protein (product) was subjected to LC-MS analysis with elution time of 35 min. After data processing by MassHunter, protein conversion at different time intervals was determined by measuring the relative peak intensities of starting and product in the mass spectrum as follows:

$$Protein\ Conversion = \frac{\text{Relative Peak Intensity of Product (M)} + \text{Relative Peak Intensity of Hydrated Product (M + H2O)}}{\text{Relative Peak Intensity of Starting and Product}}\ x\ 100\%$$

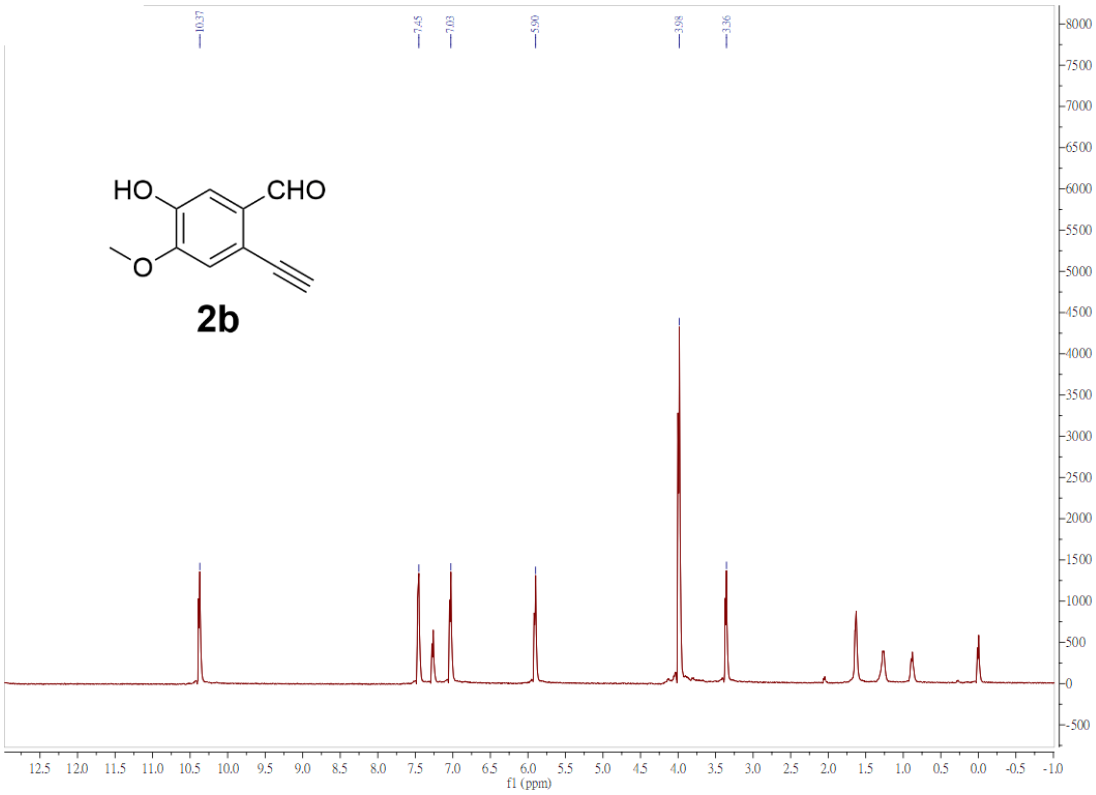
General procedure for isoquinolinium-modified protein using 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde 2b

To an eppendorf (1.5 mL) with 80 µL of 50 mM PBS buffer pH 6.3, 10 µL of 1 mM protein stock solution (dissolved in Milli-Q® water) was added to the buffer, followed by 10 µL of 5-20 mM 2-ethynyl-5-hydroxy-4-methoxybenzaldehyde **2b** stock solution (dissolved in DMSO), and allow the reaction mixture was reacted under 37 ^{o}C water bath for 16 h. 10 µL of the mixture was drawn, diluted with 10 µL Milli-Q® and subjected for LC/MS-MS analysis.

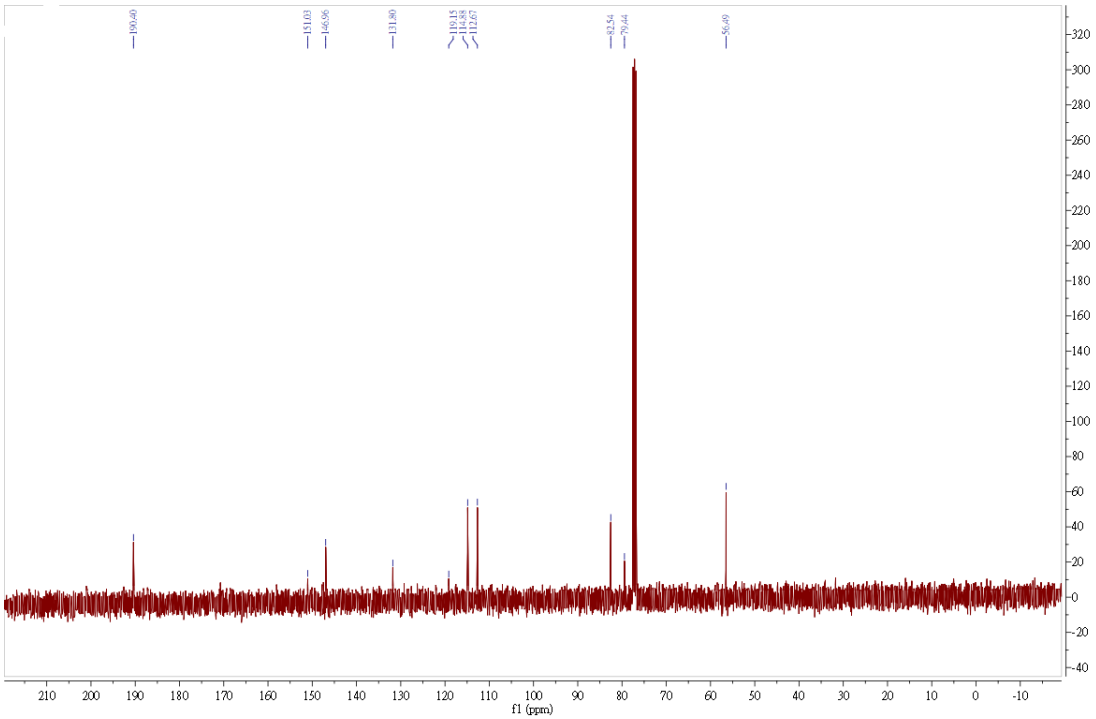
Appendix

NMR spectra of 2b

^{1}H NMR

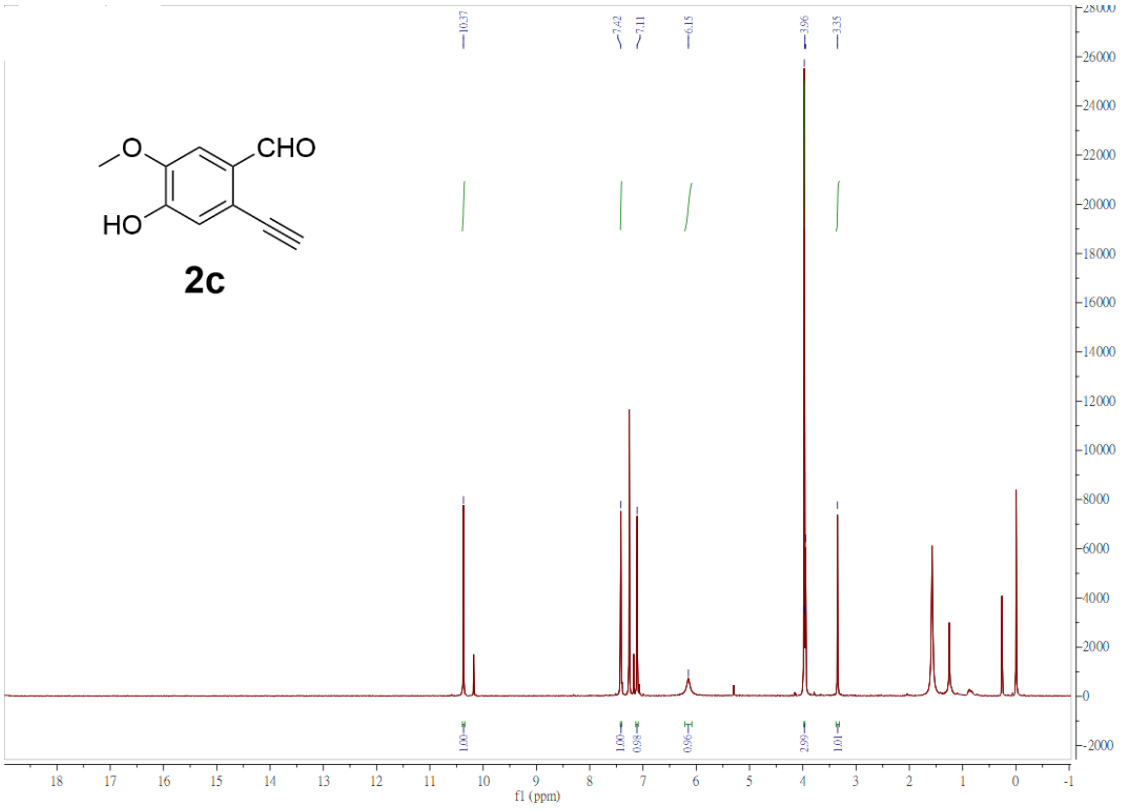


^{13}C NMR

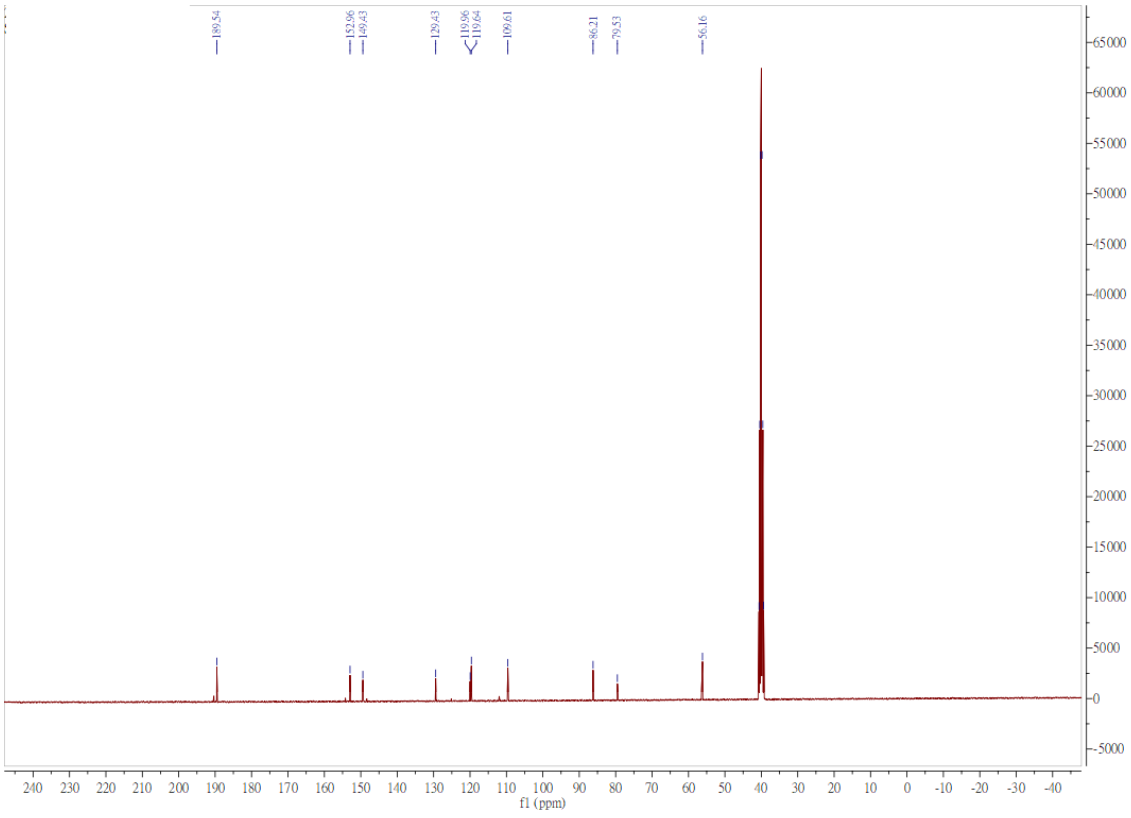


NMR spectra of 2c

^{1}H NMR

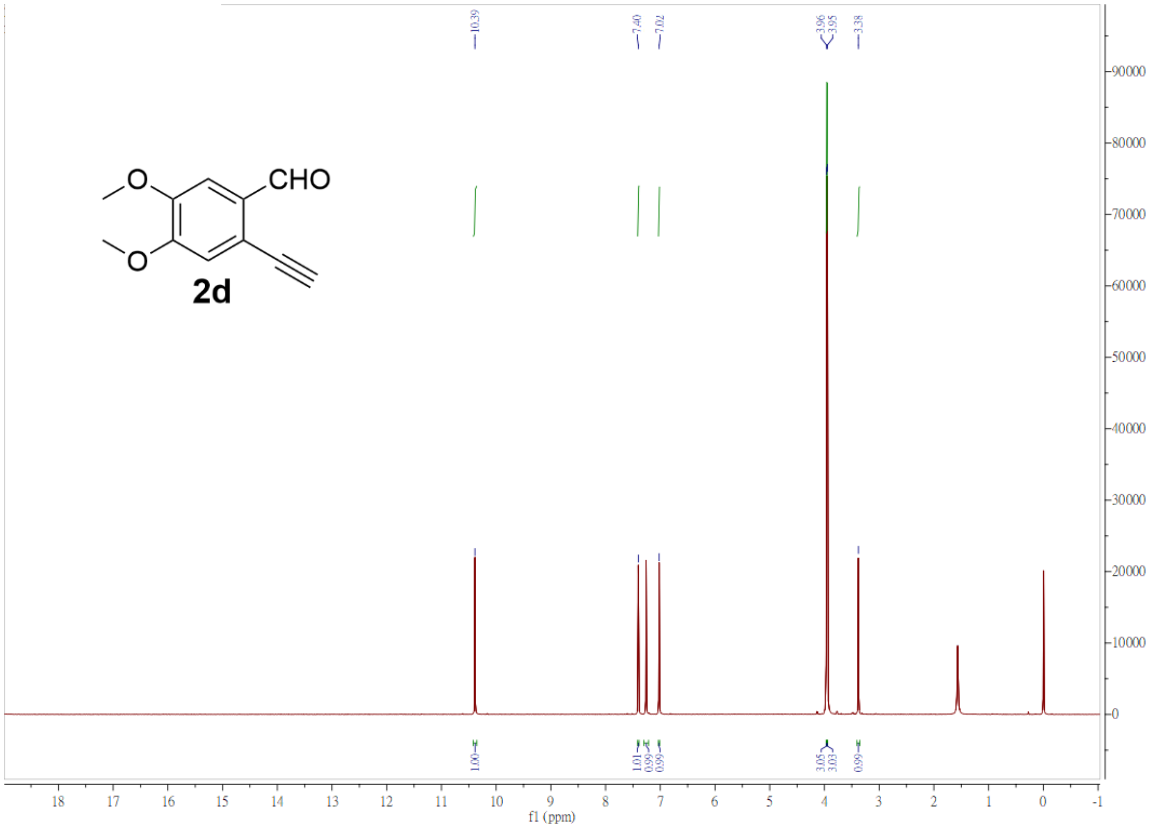


^{13}C NMR

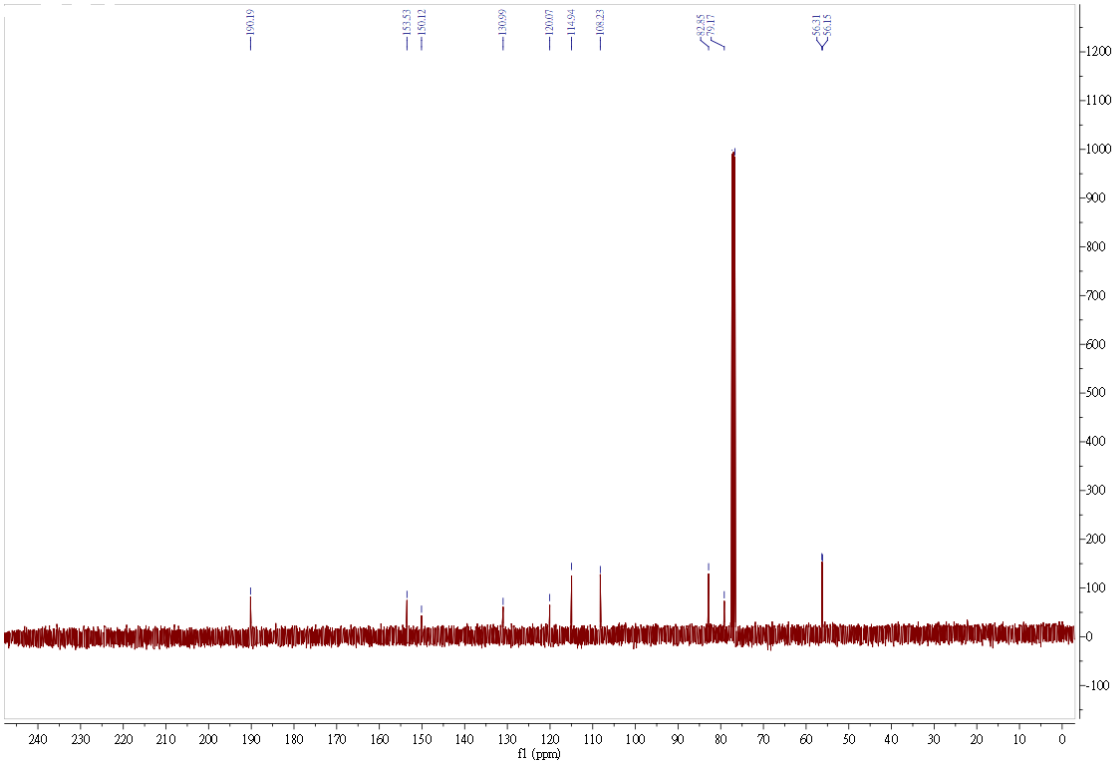


NMR spectra of 2d

^{1}H NMR

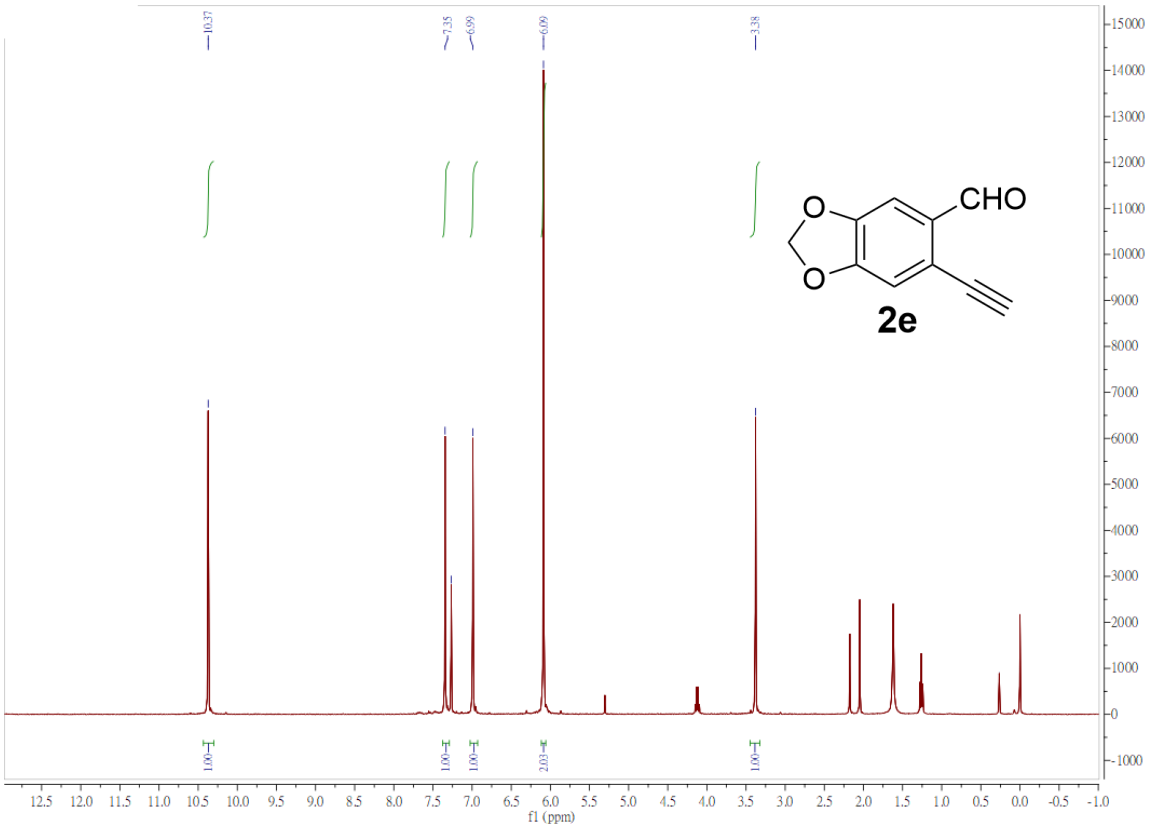


^{13}C NMR

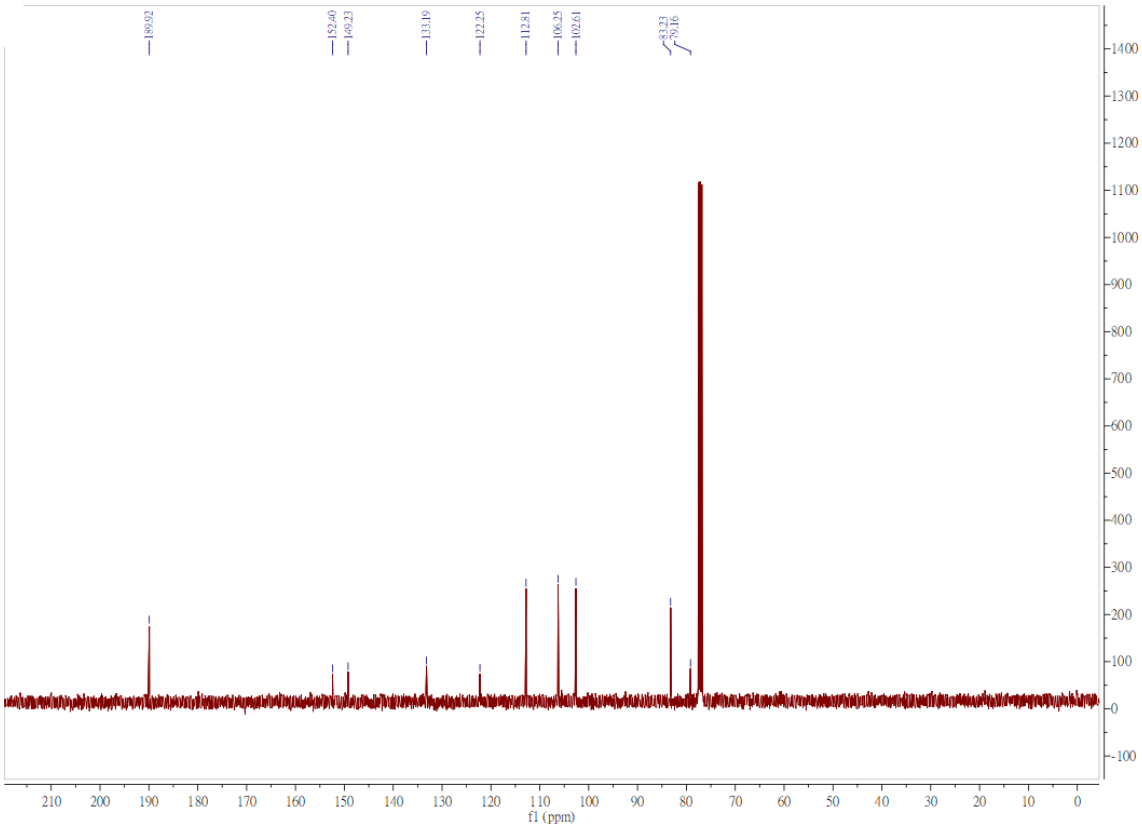


NMR spectra of 2e

^{1}H NMR

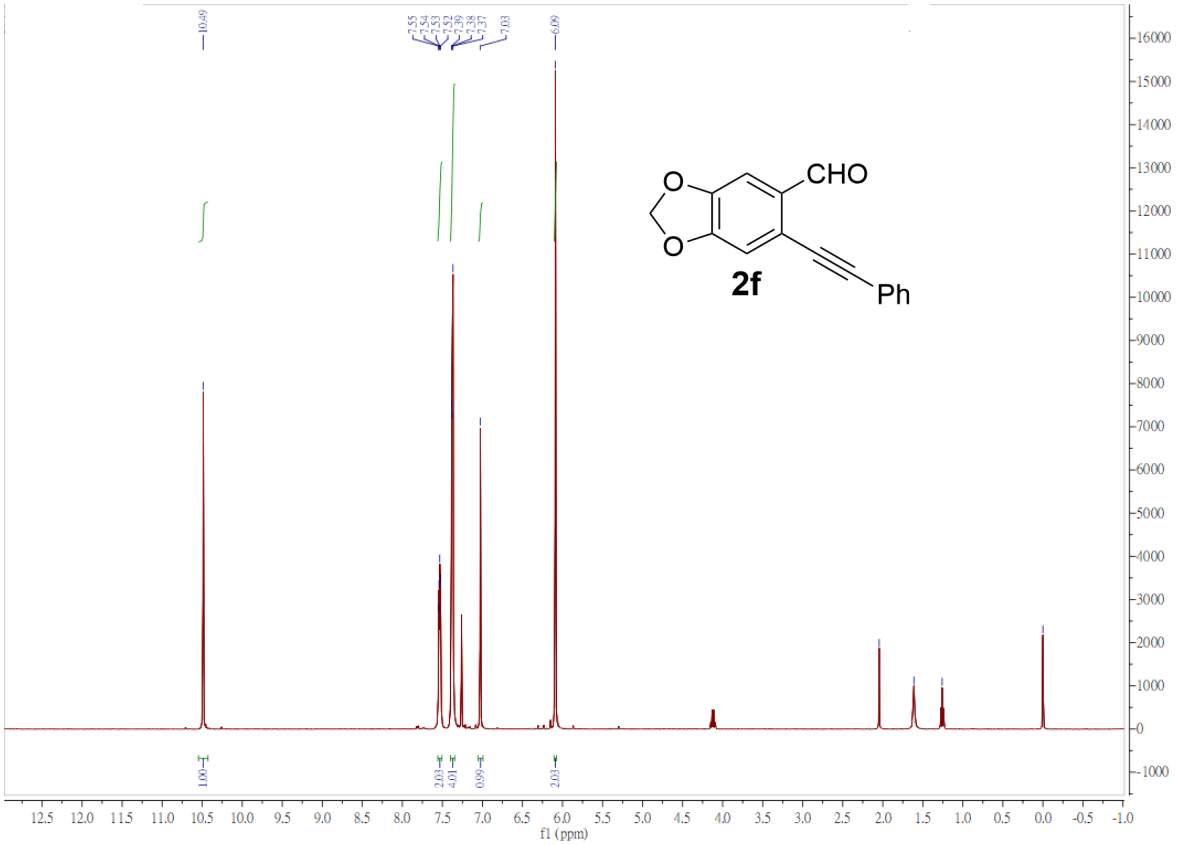


^{13}C NMR

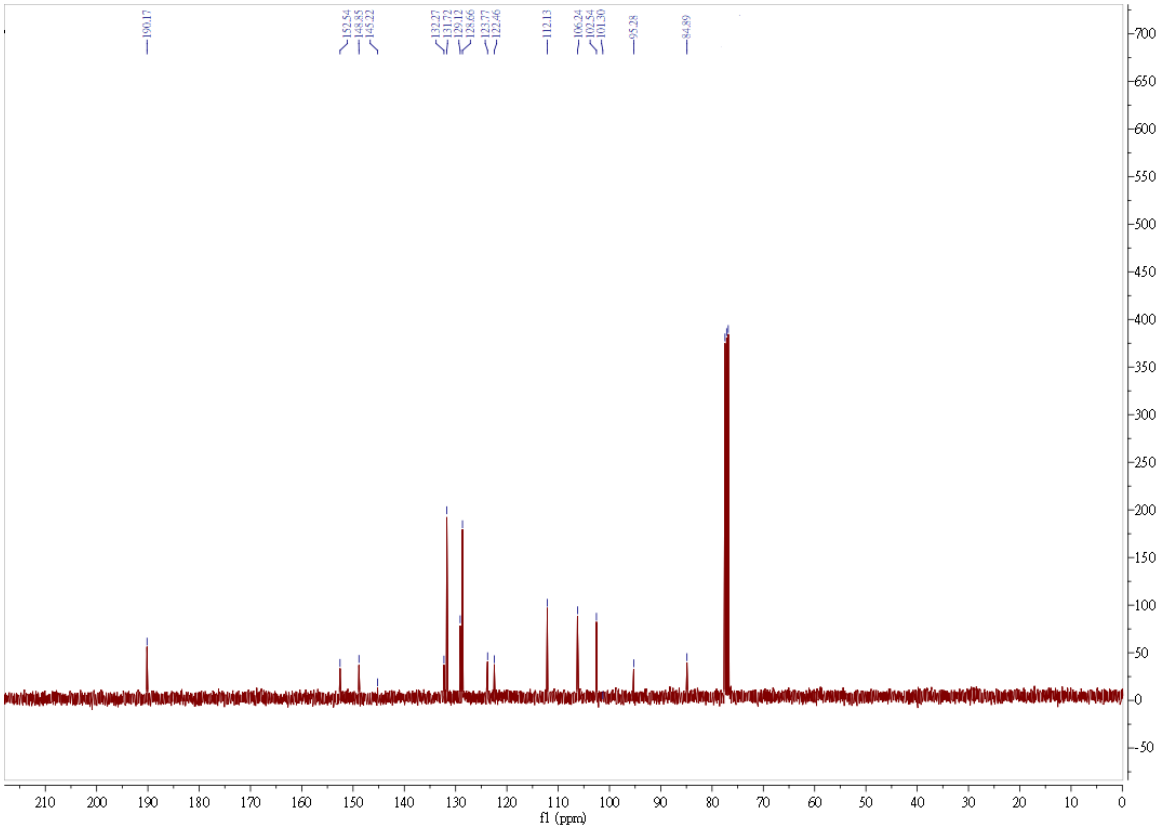


NMR spectra of 2f

^{1}H NMR

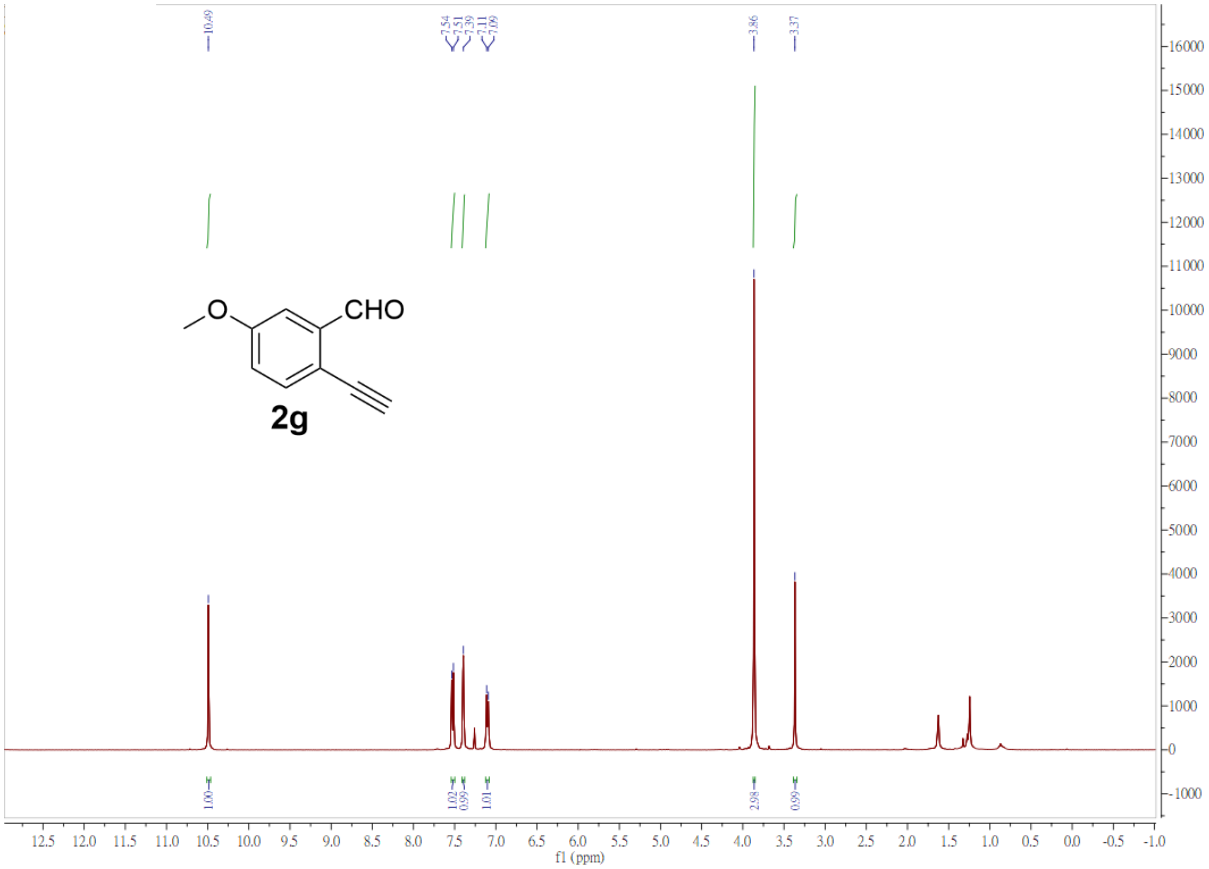


^{13}C NMR

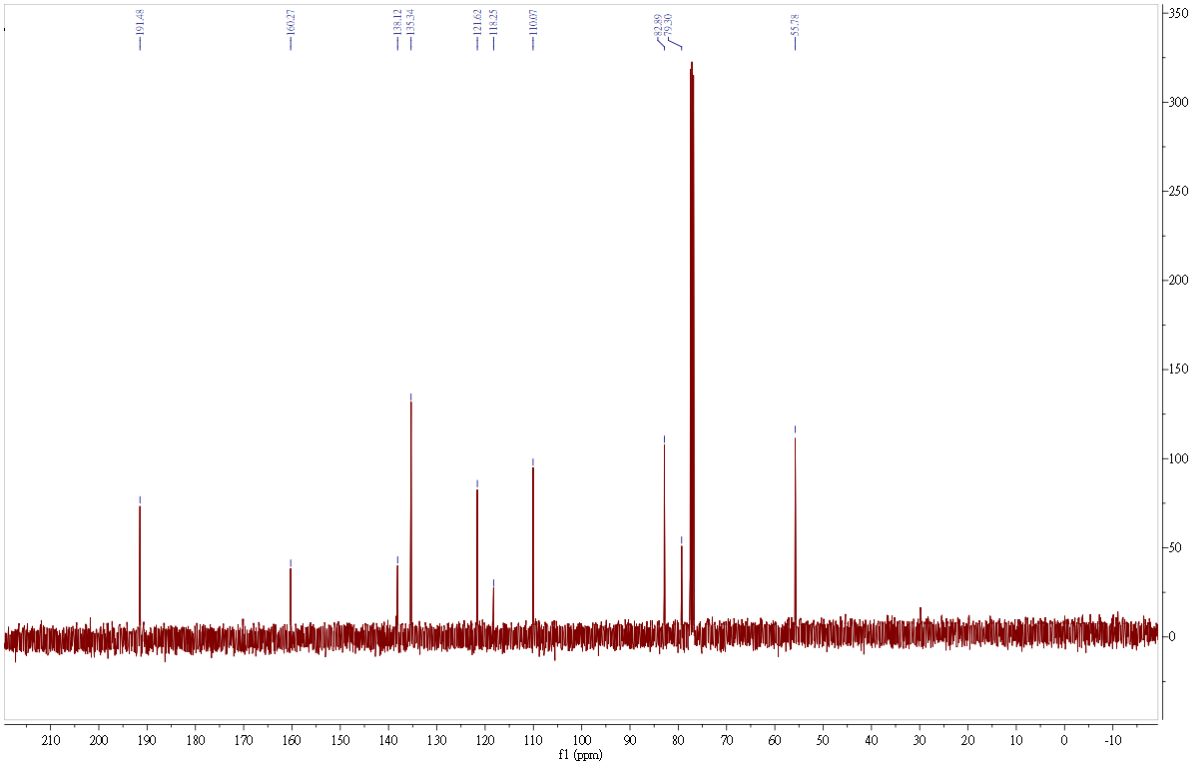


NMR spectra of 2g

^{1}H NMR

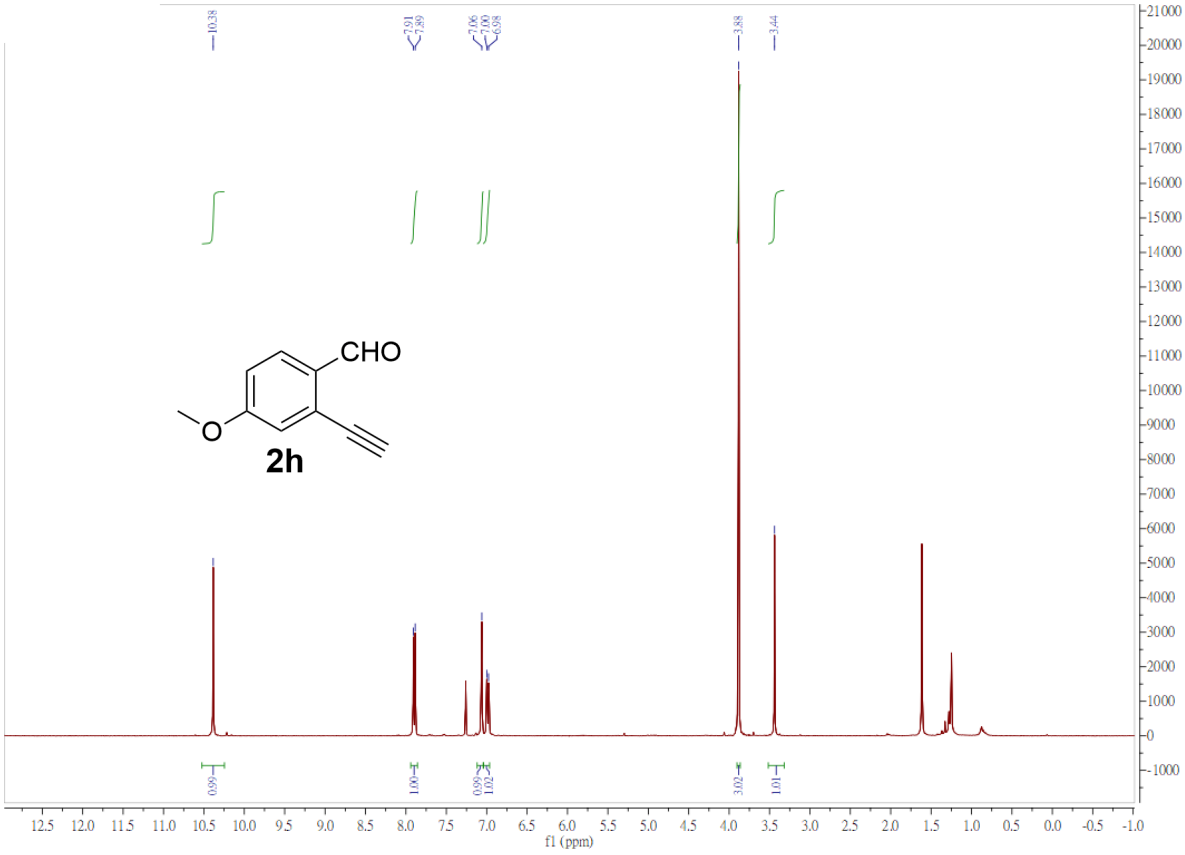


^{13}C NMR

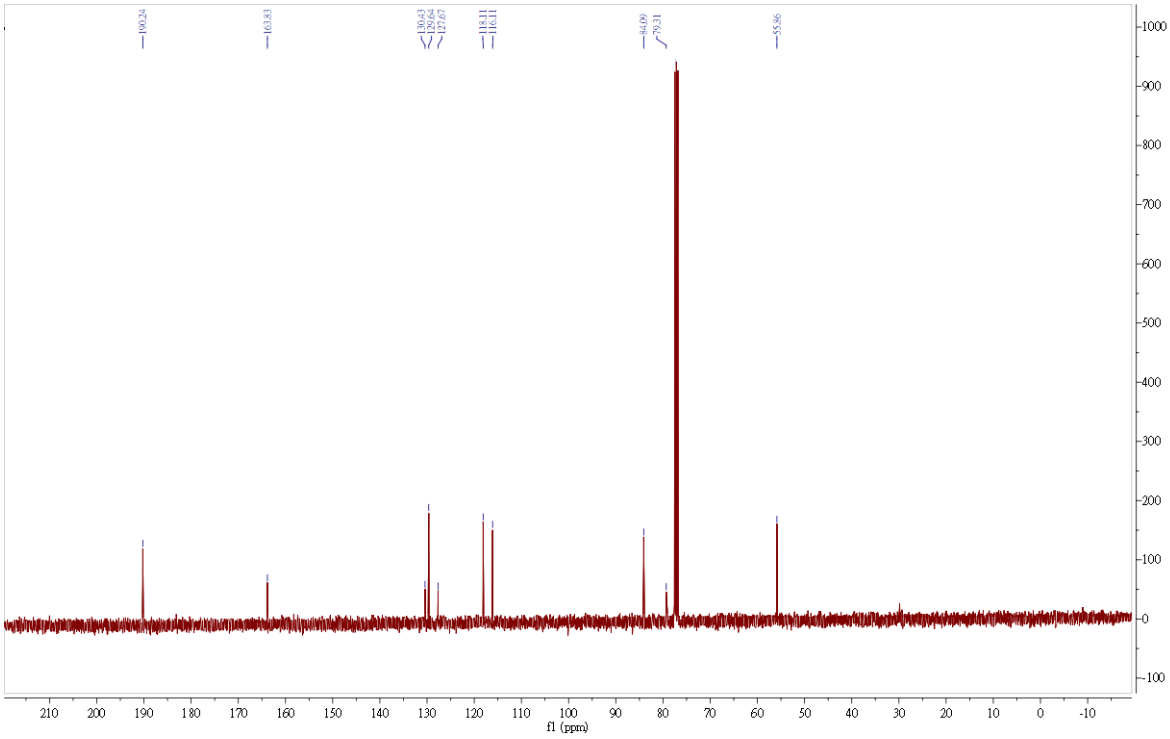


NMR spectra of 2h

^{1}H NMR

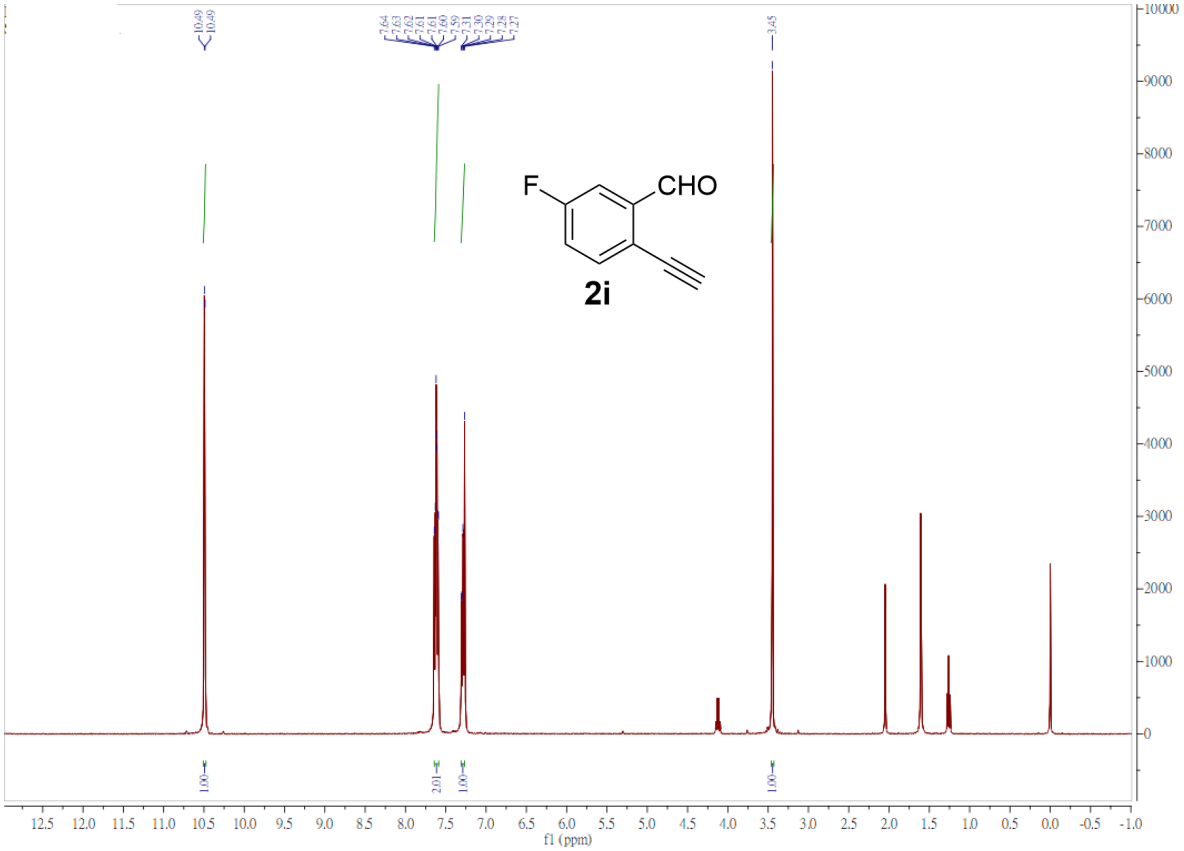


^{13}C NMR

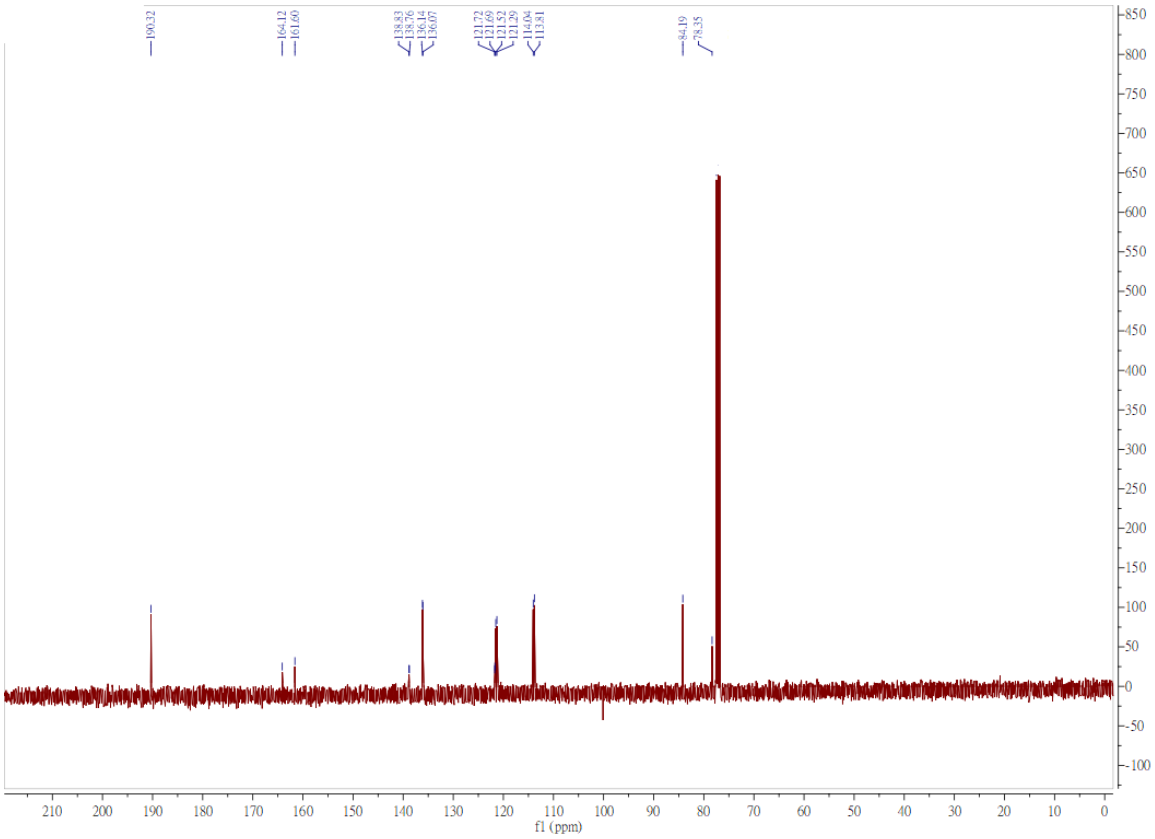


NMR spectra of 2i

^{1}H NMR

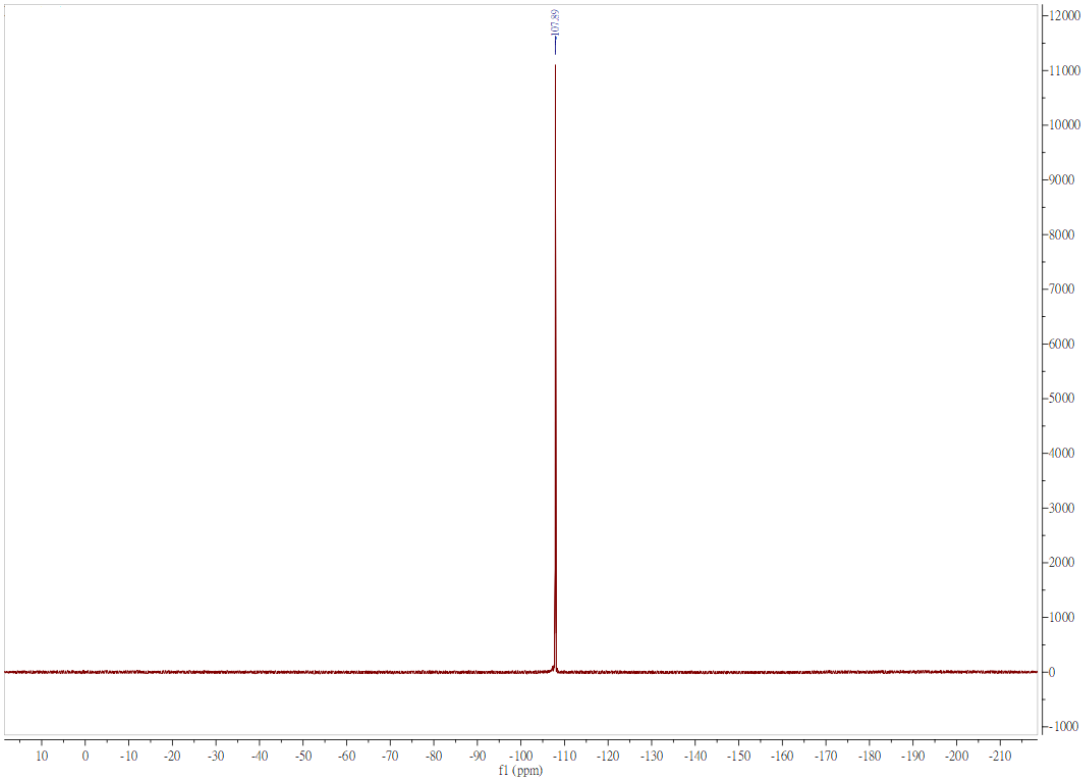


^{13}C NMR



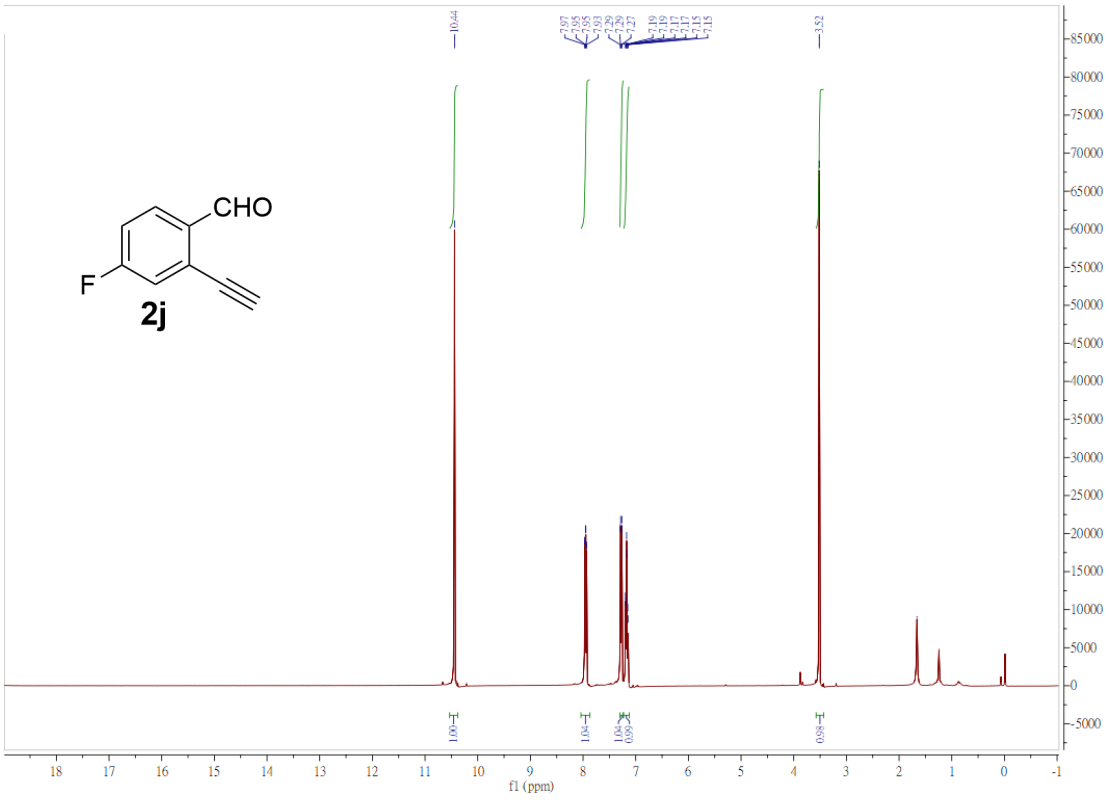
NMR spectra of 2i (cont'd)

^{19}F NMR

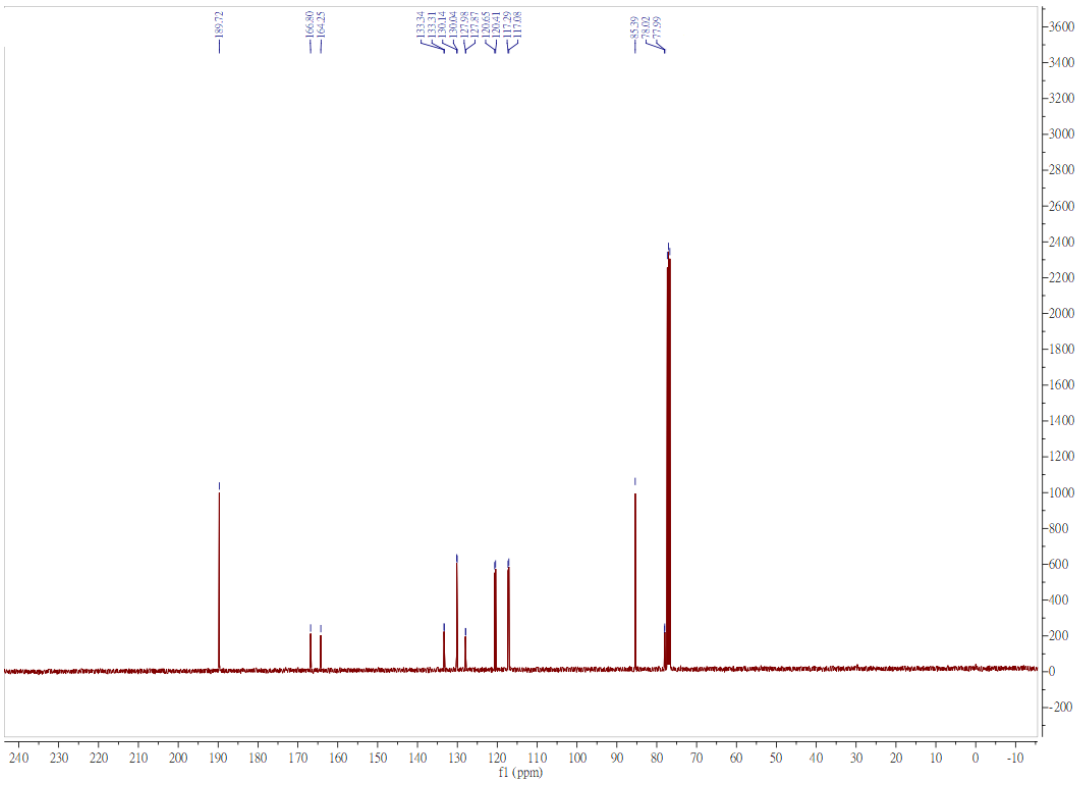


NMR spectra of 2j

^{1}H NMR

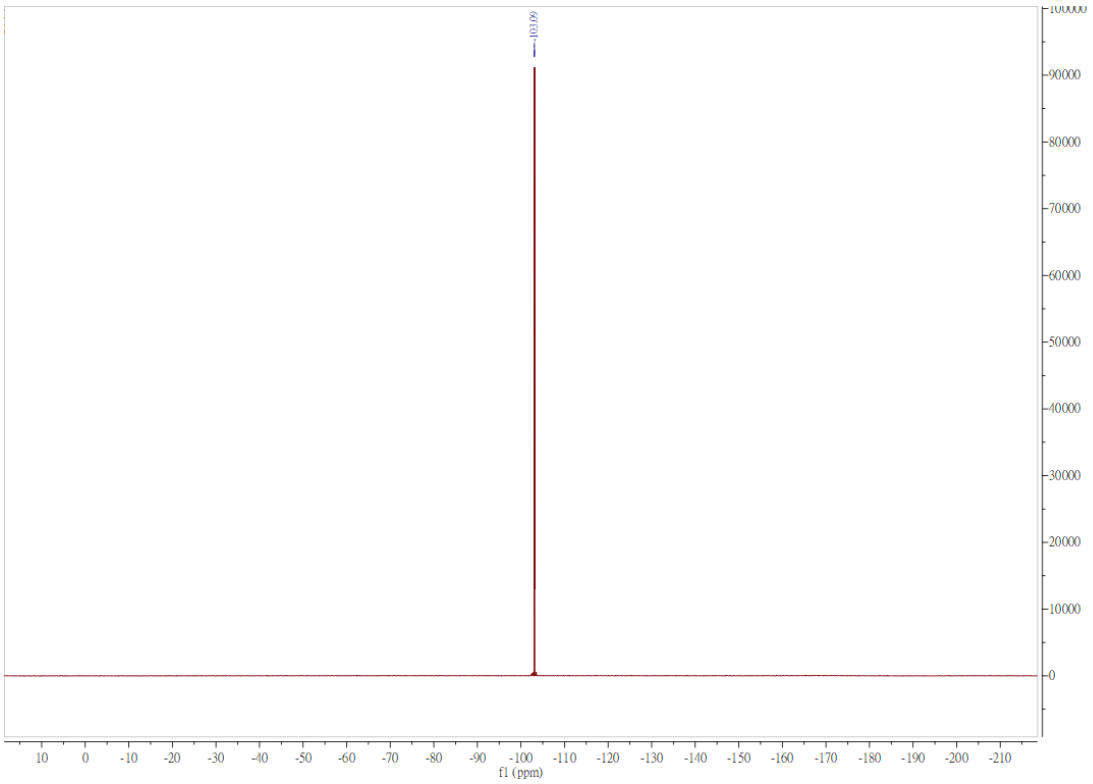


^{13}C NMR



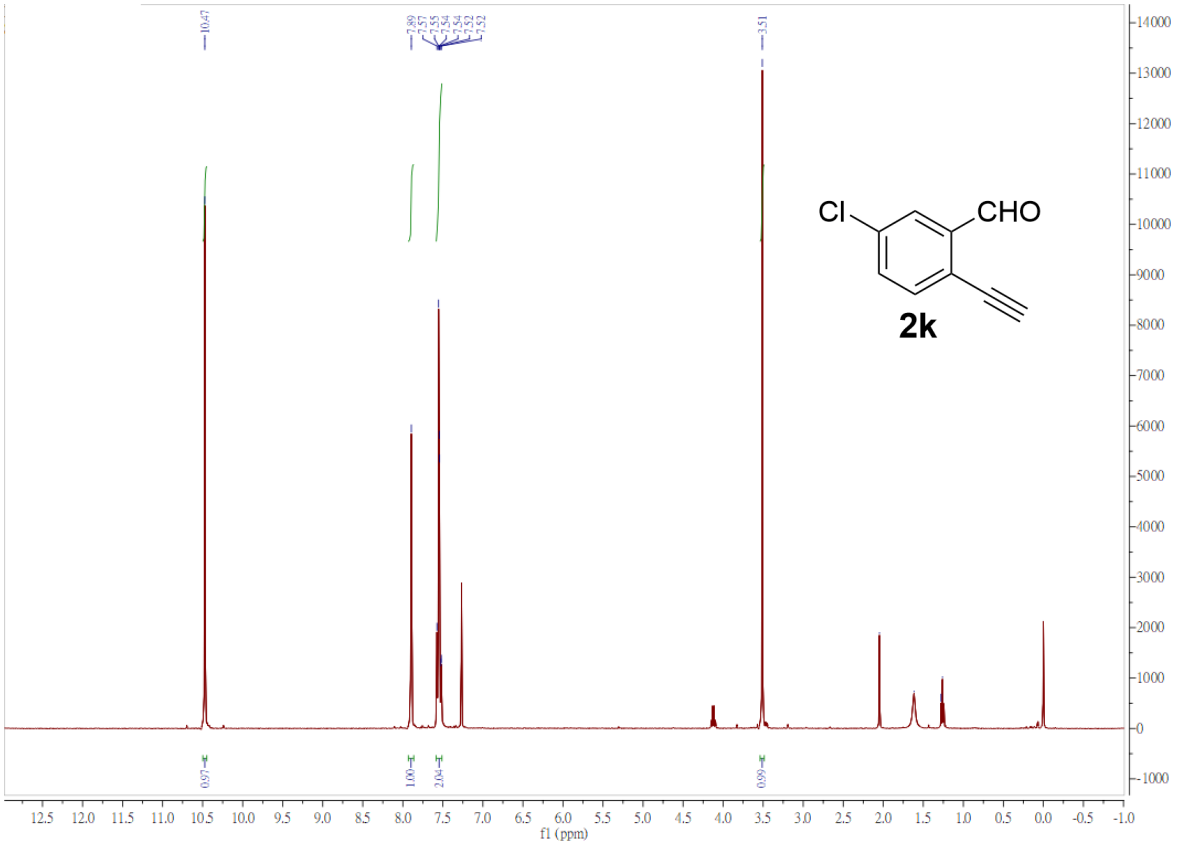
NMR spectra of 2j (cont'd)

^{19}F NMR

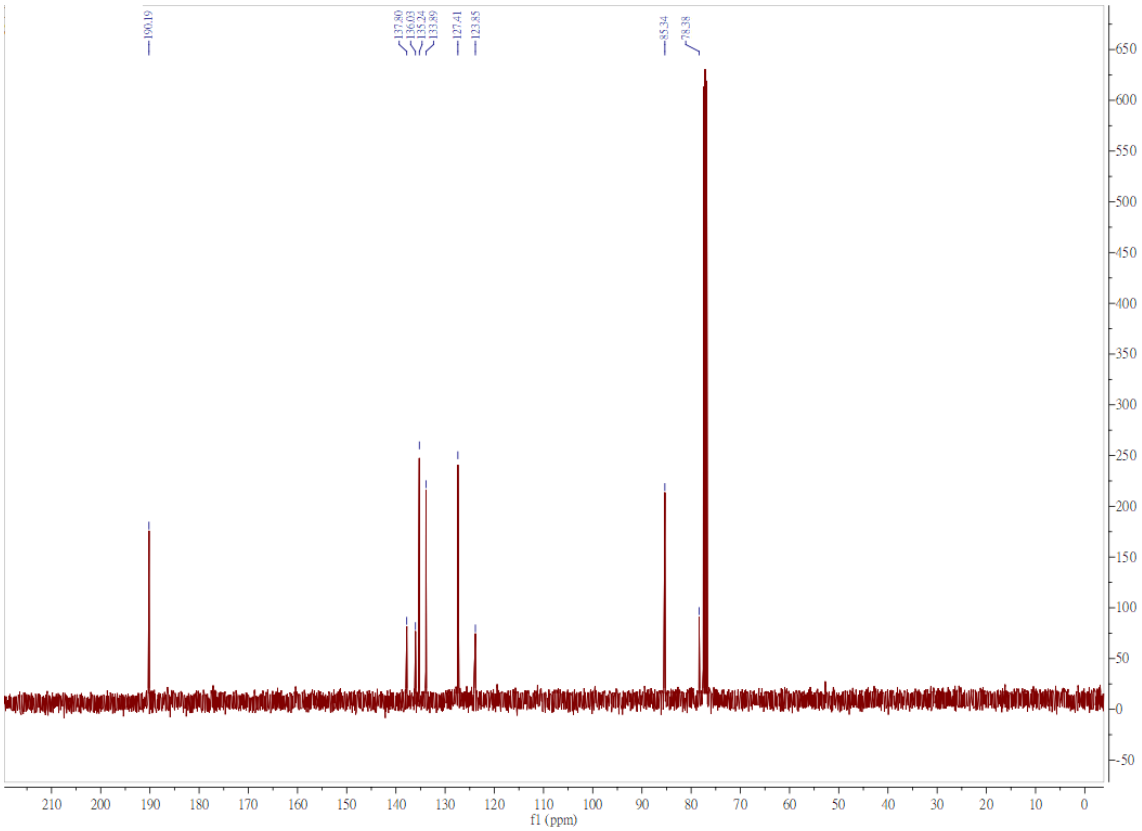


NMR spectra of 2k

^{1}H NMR

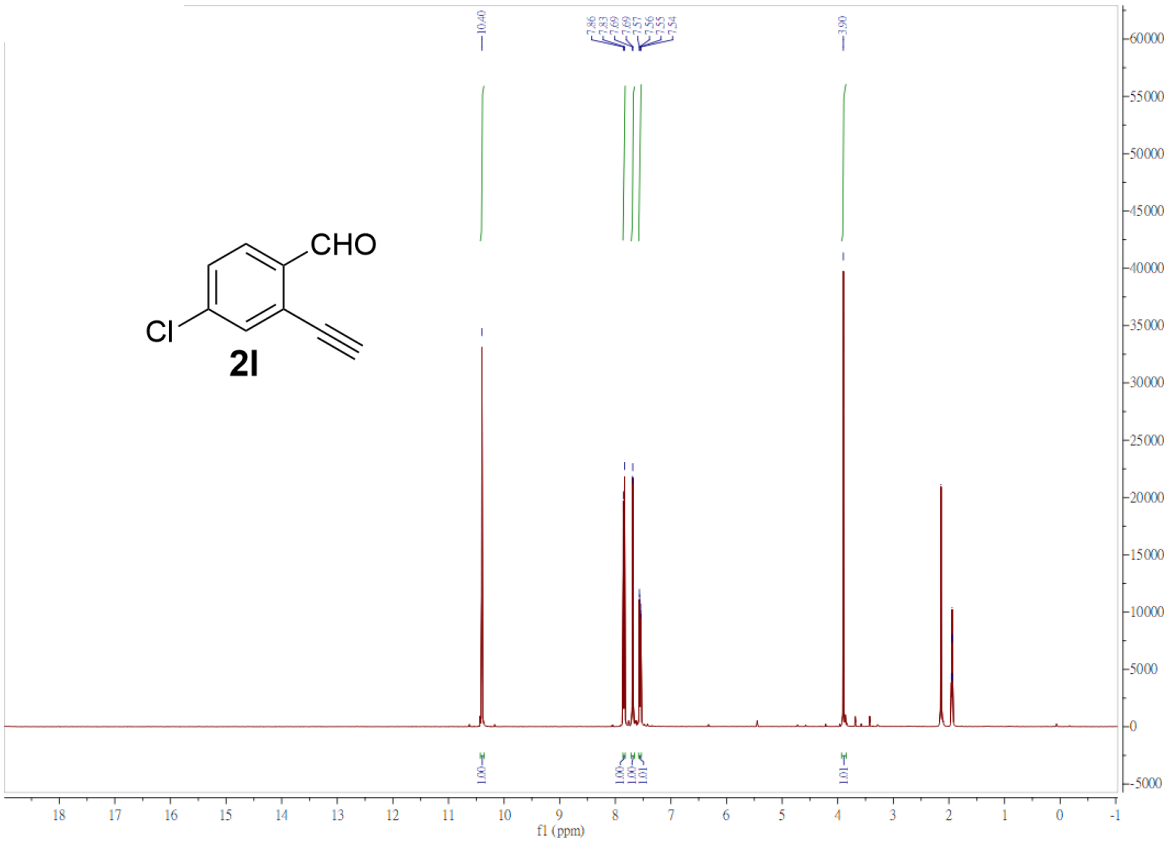


^{13}C NMR

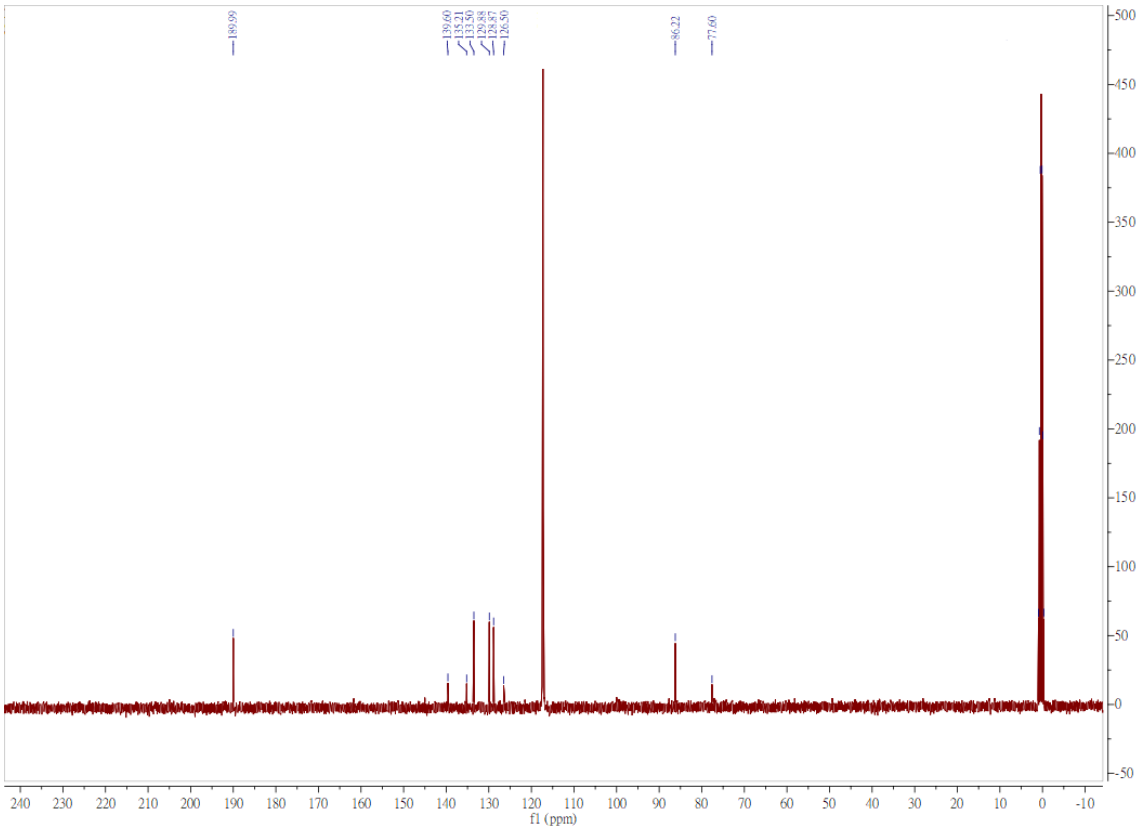


NMR spectra of 2l

^{1}H NMR

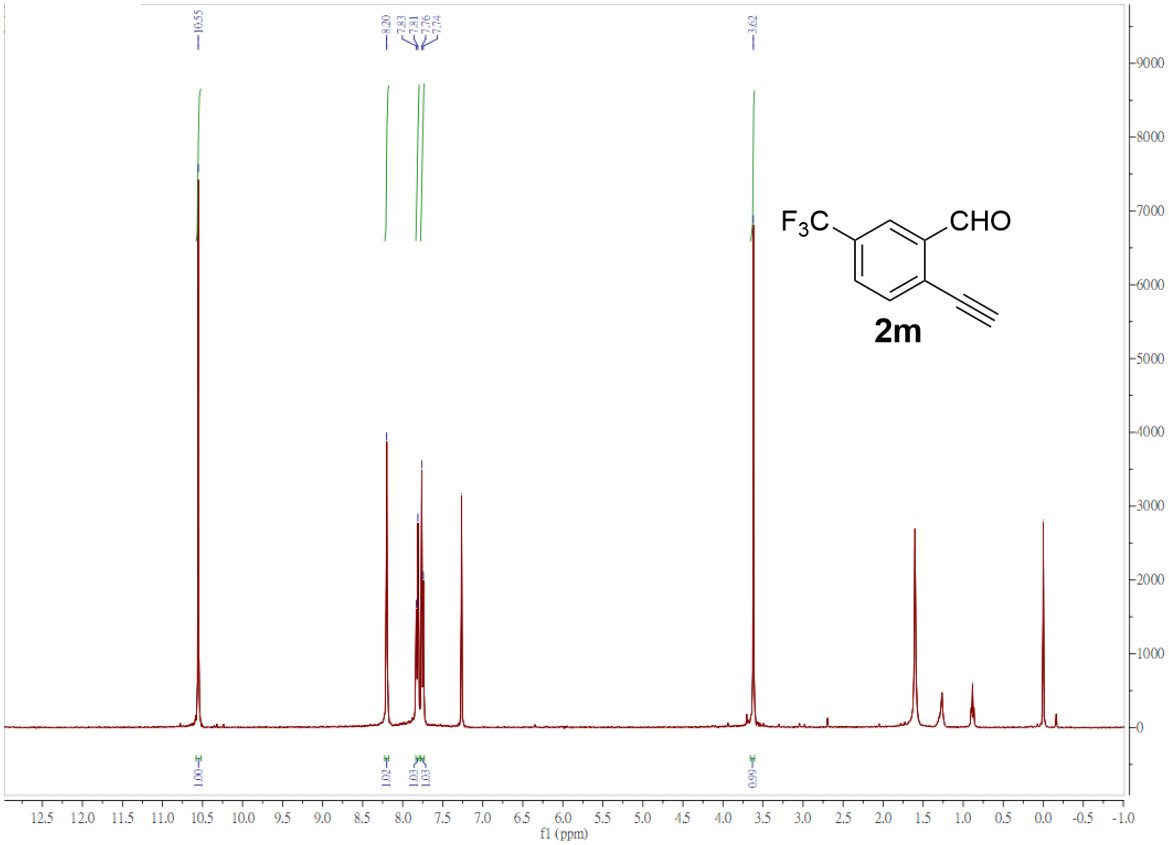


^{13}C NMR

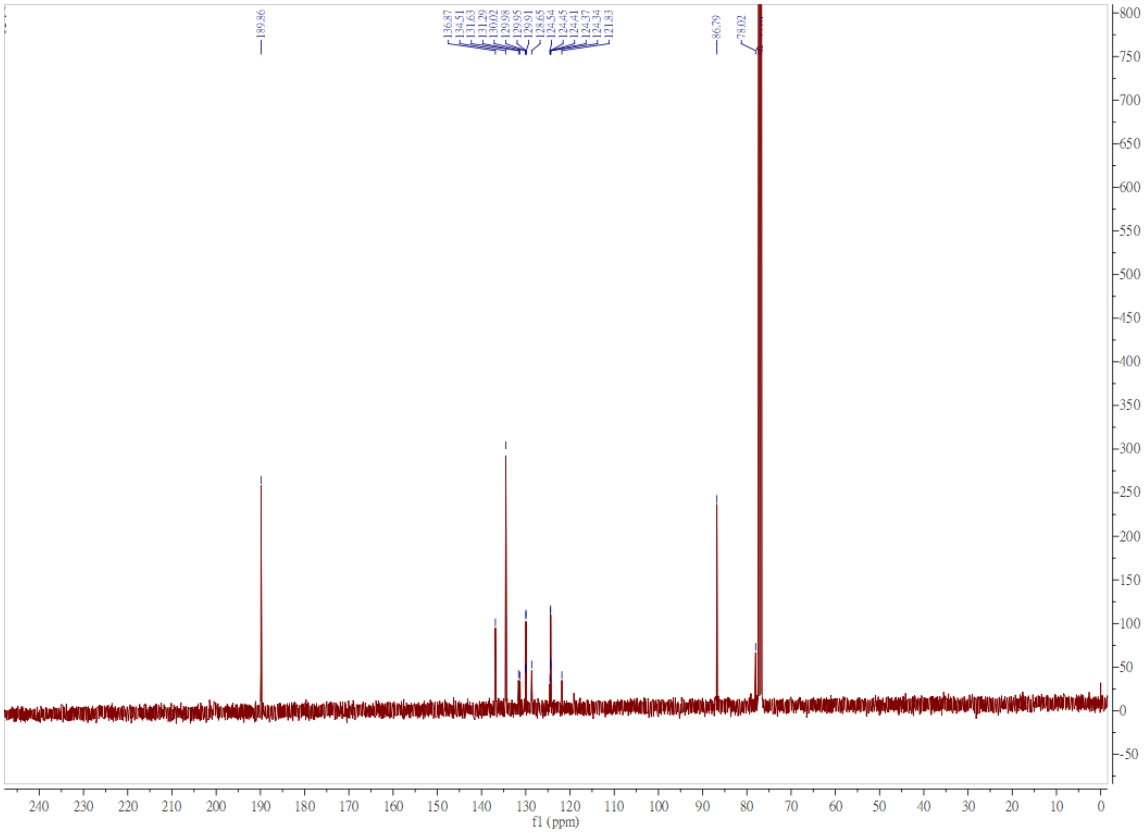


NMR spectra of 2m

^{1}H NMR

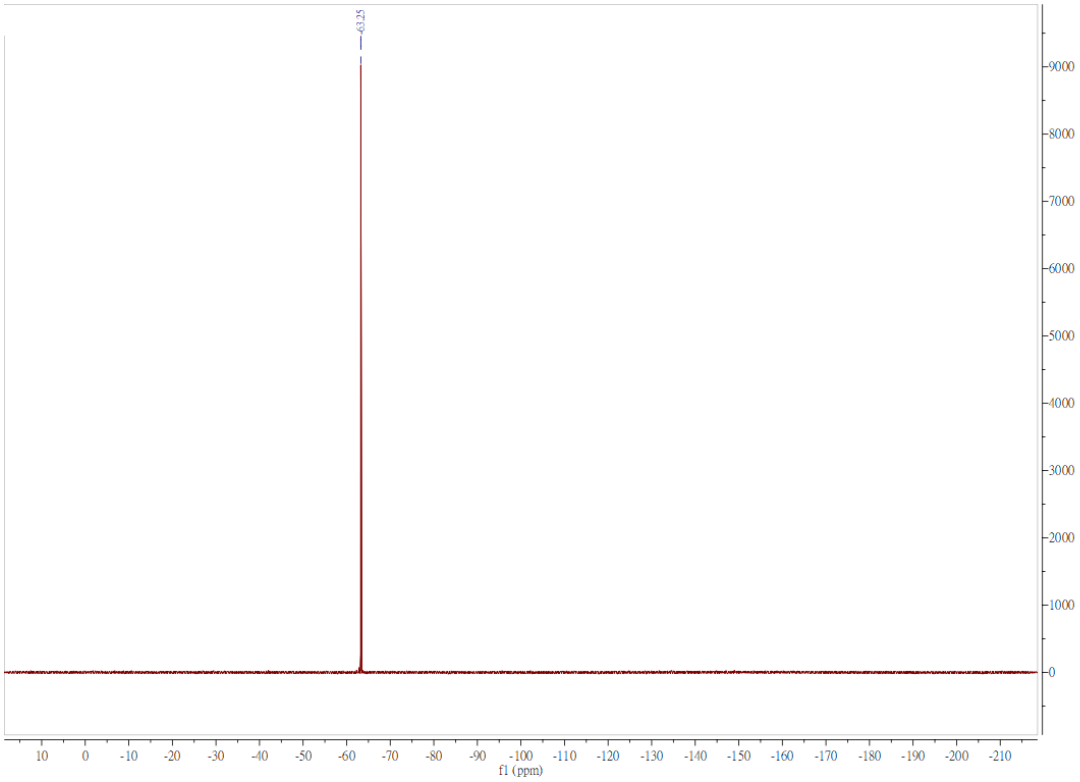


^{13}C NMR



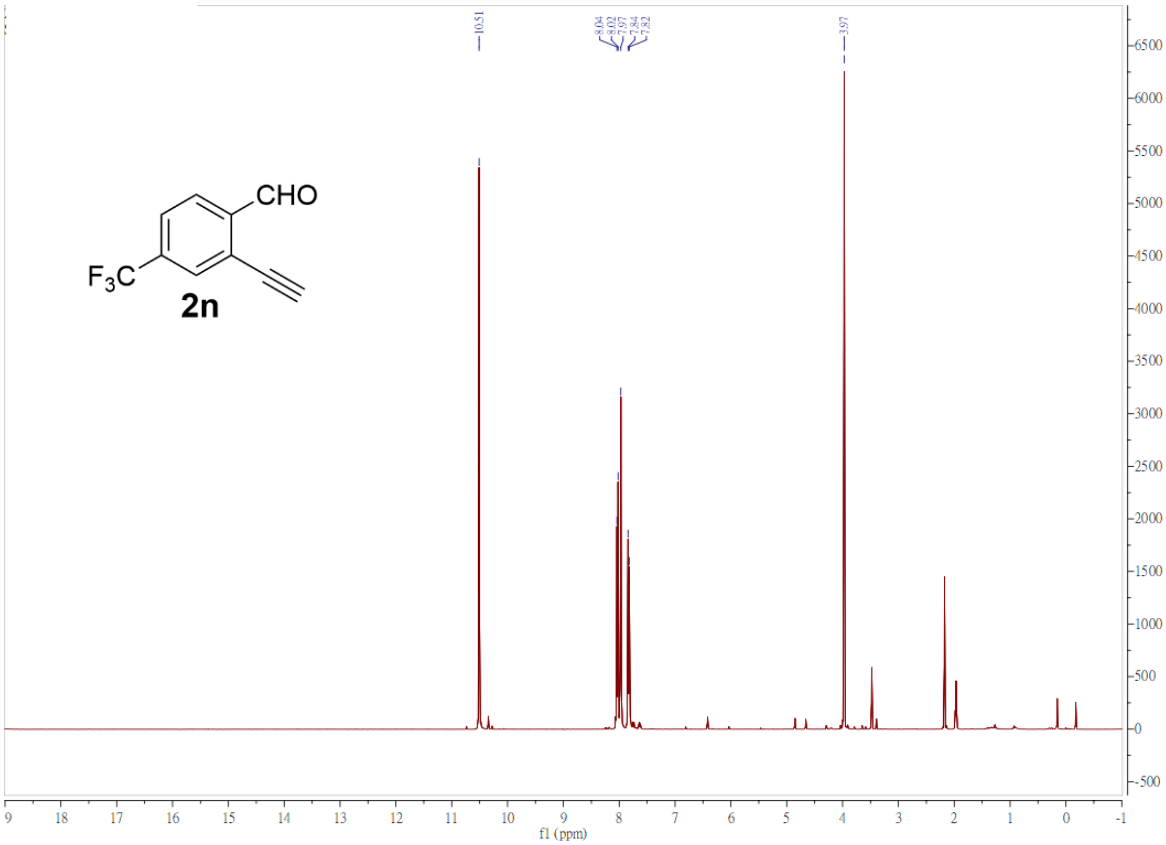
NMR spectra of 2m (cont'd)

^{19}F NMR

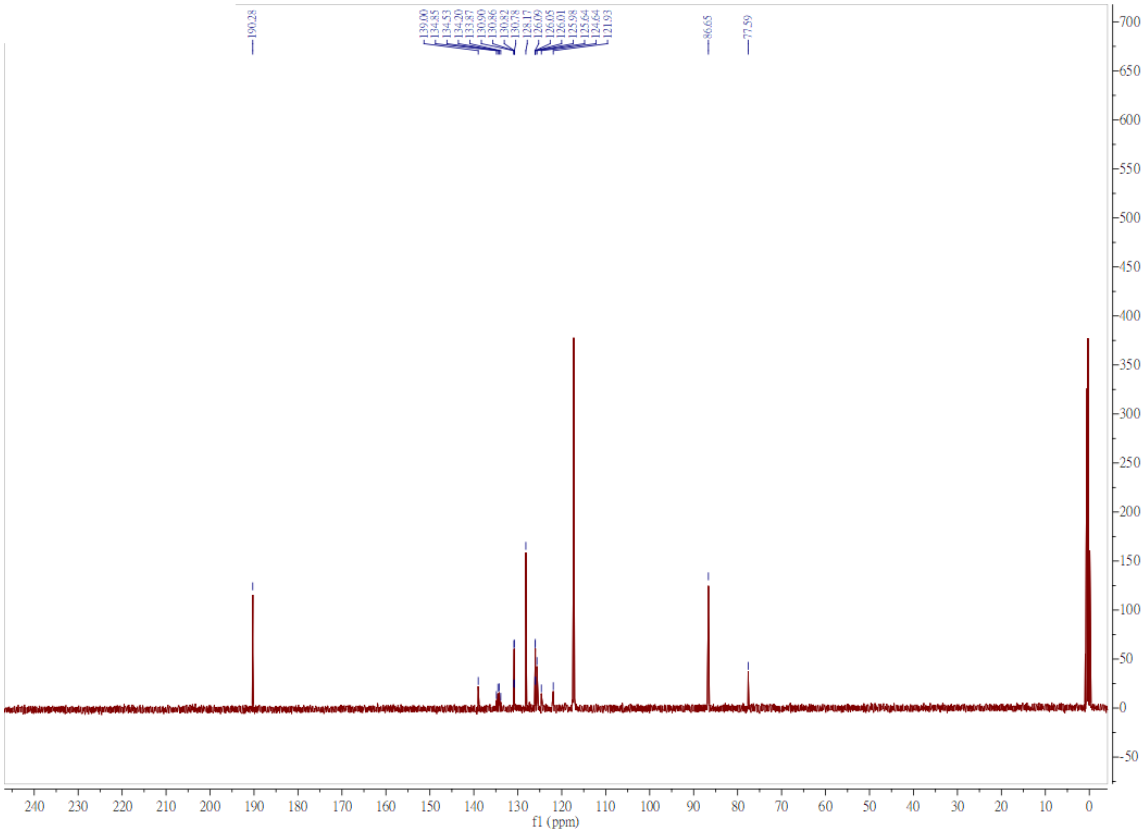


NMR spectra of 2n

^{1}H NMR

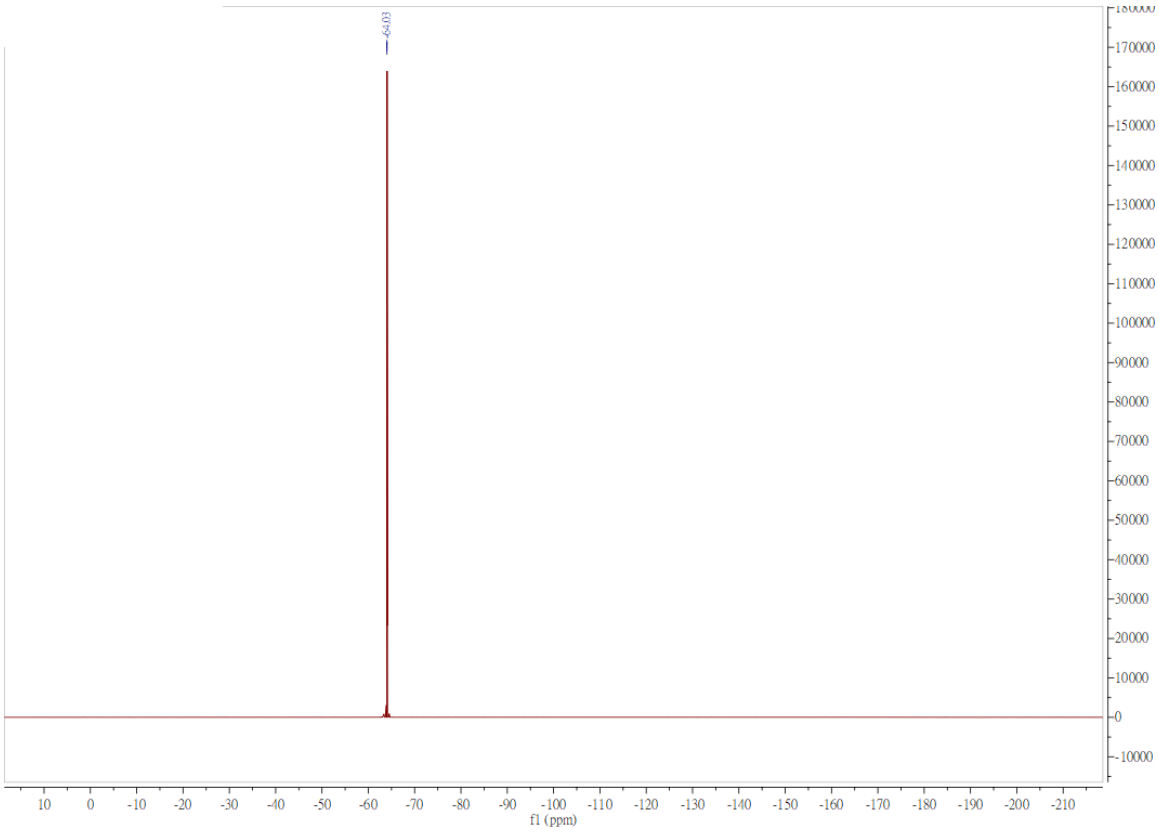


^{13}C NMR



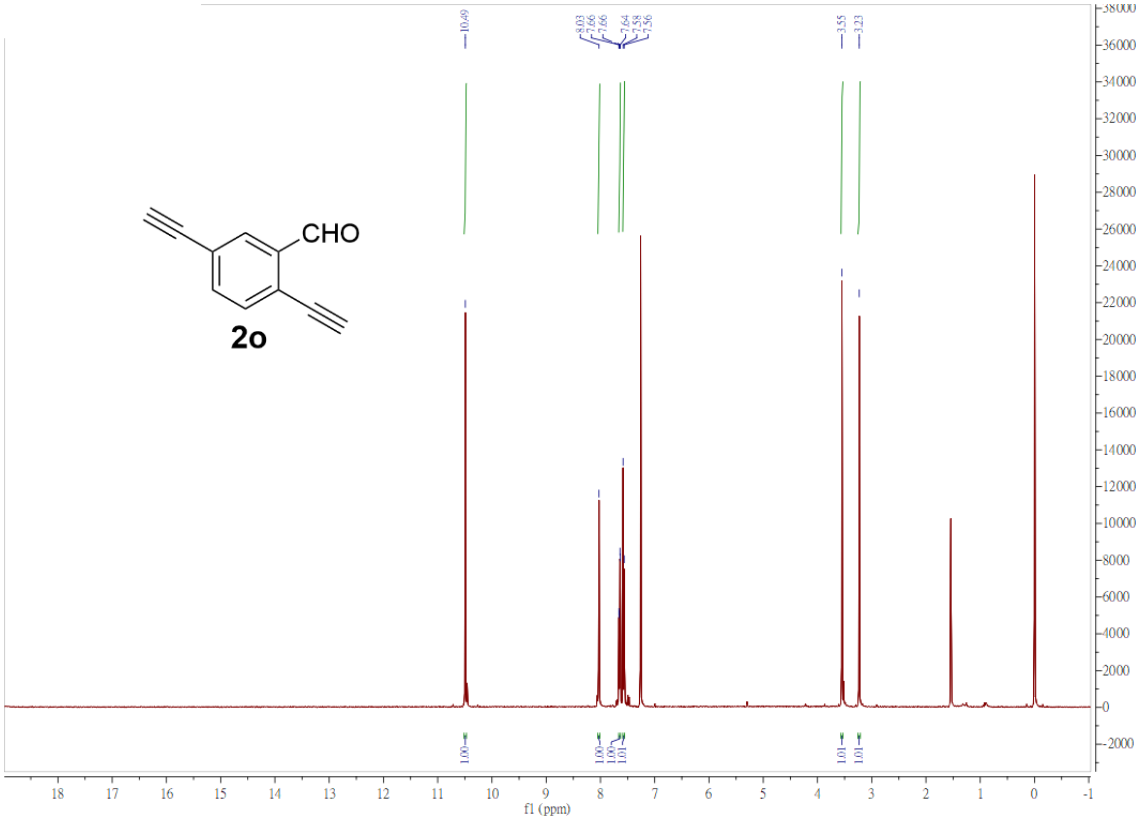
NMR spectra of 2n (cont'd)

^{19}F NMR

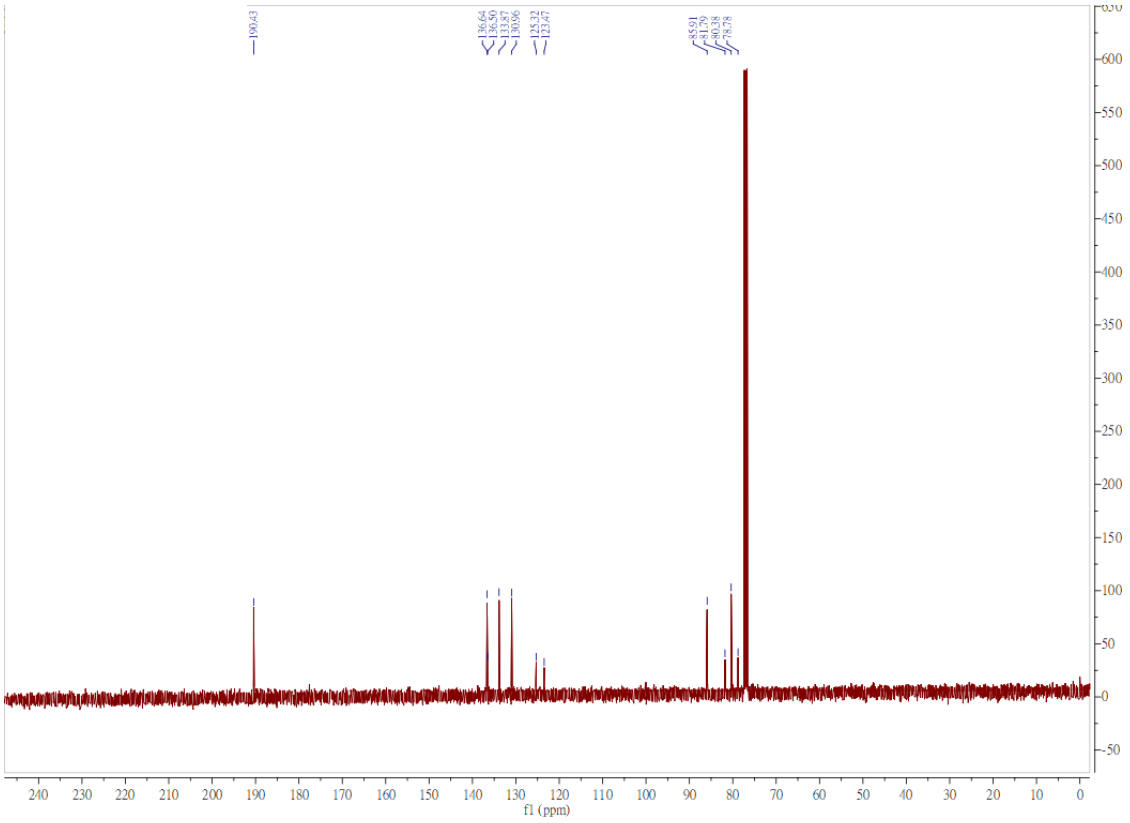


NMR spectra of 2o

^{1}H NMR

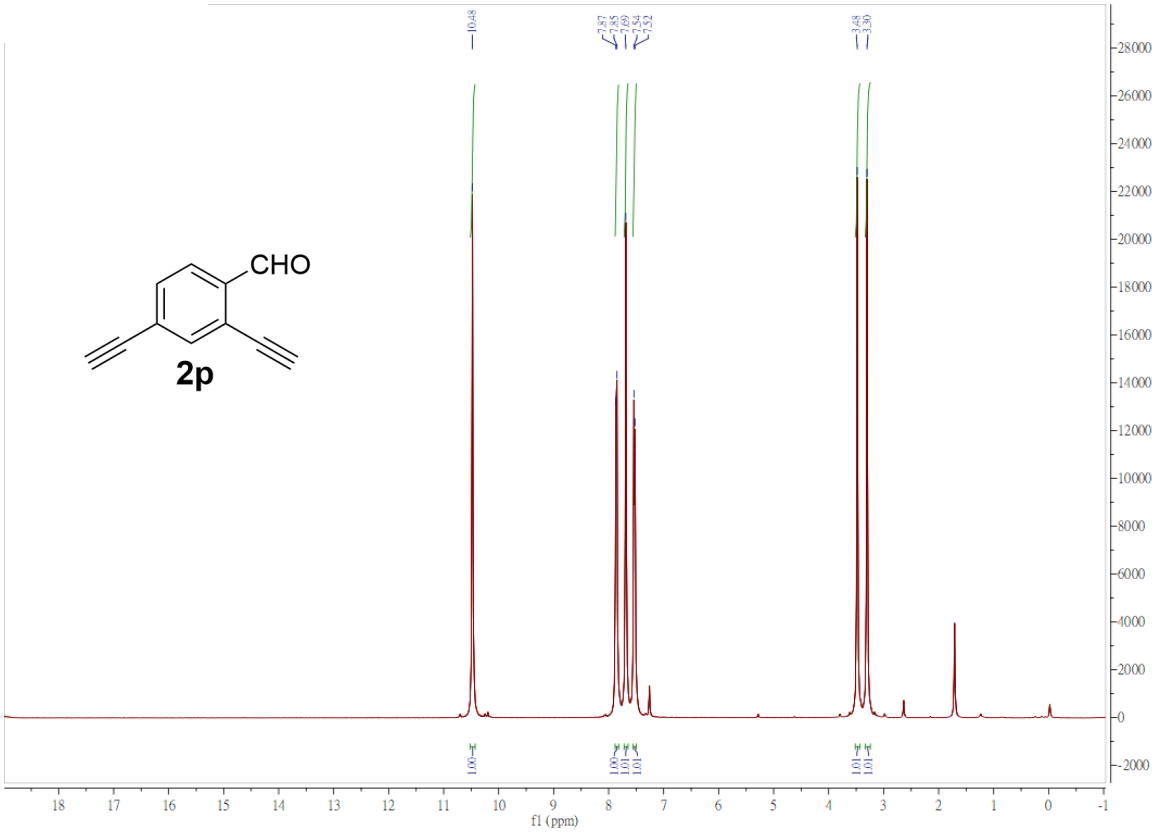


^{13}C NMR

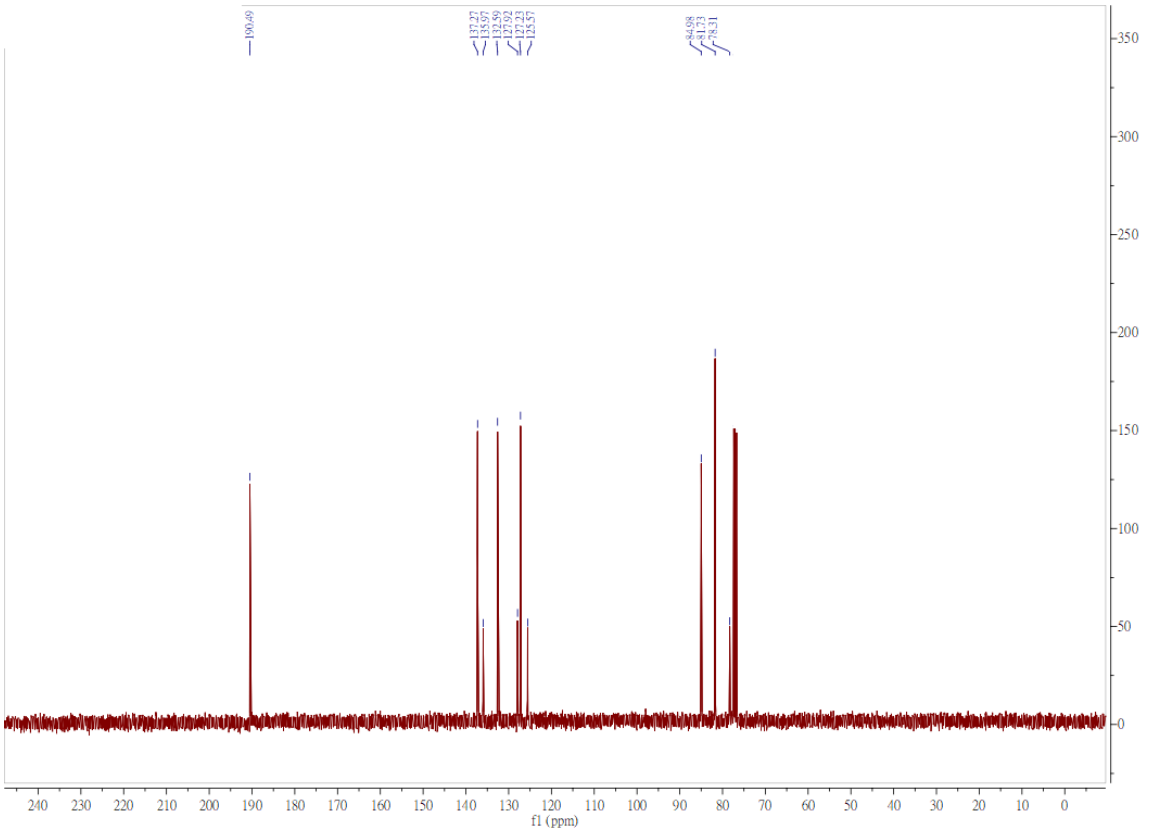


NMR spectra of 2p

^{1}H NMR

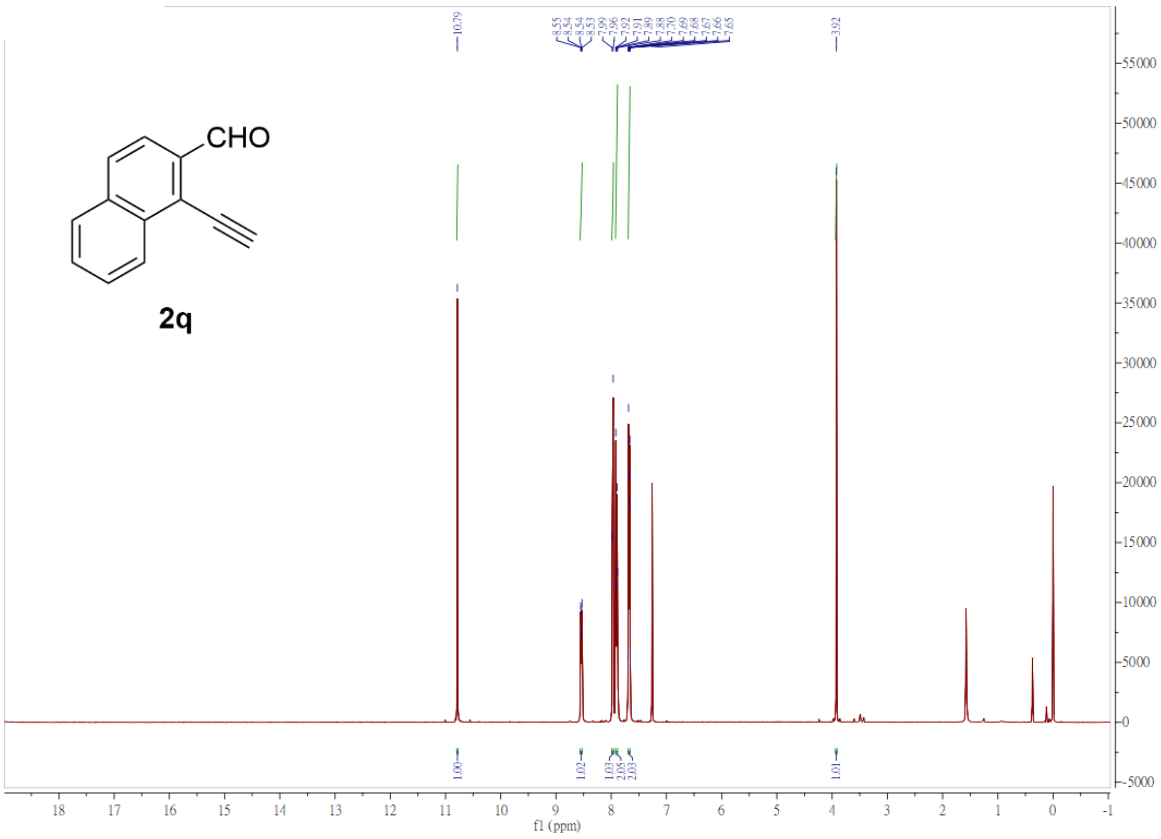


^{13}C NMR

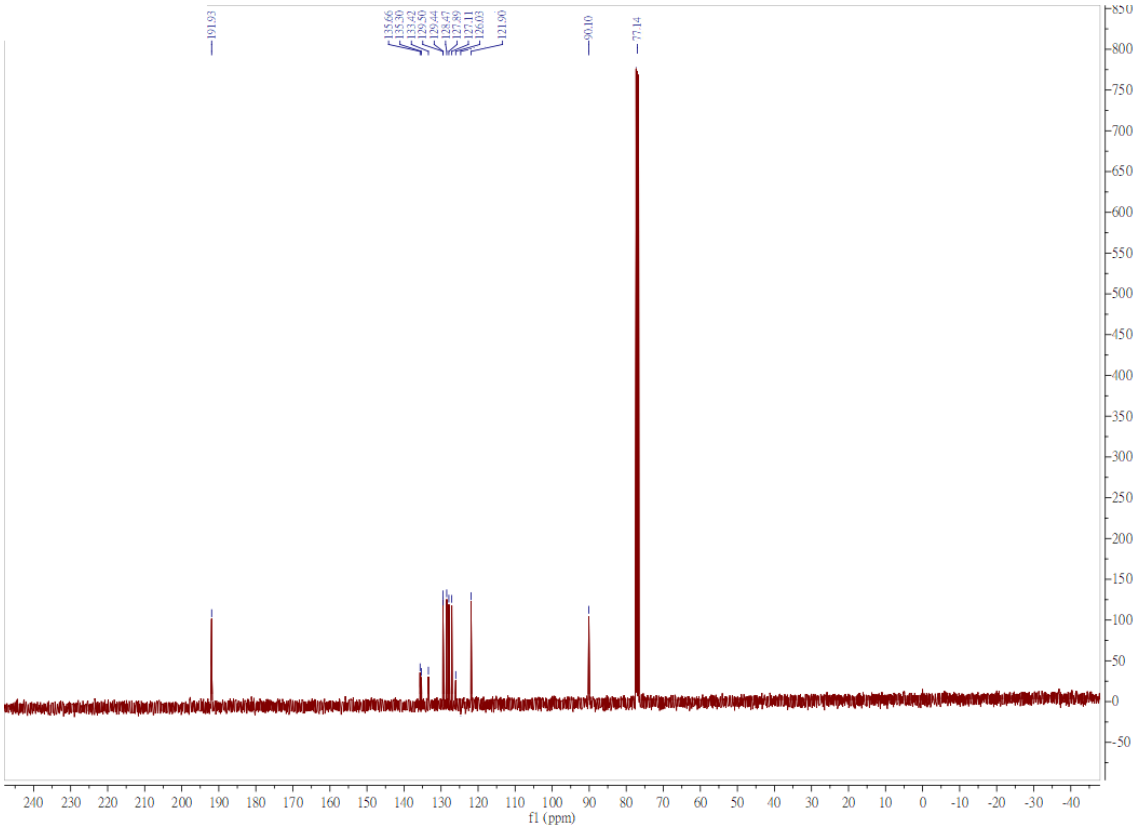


NMR spectra of 2q

^{1}H NMR

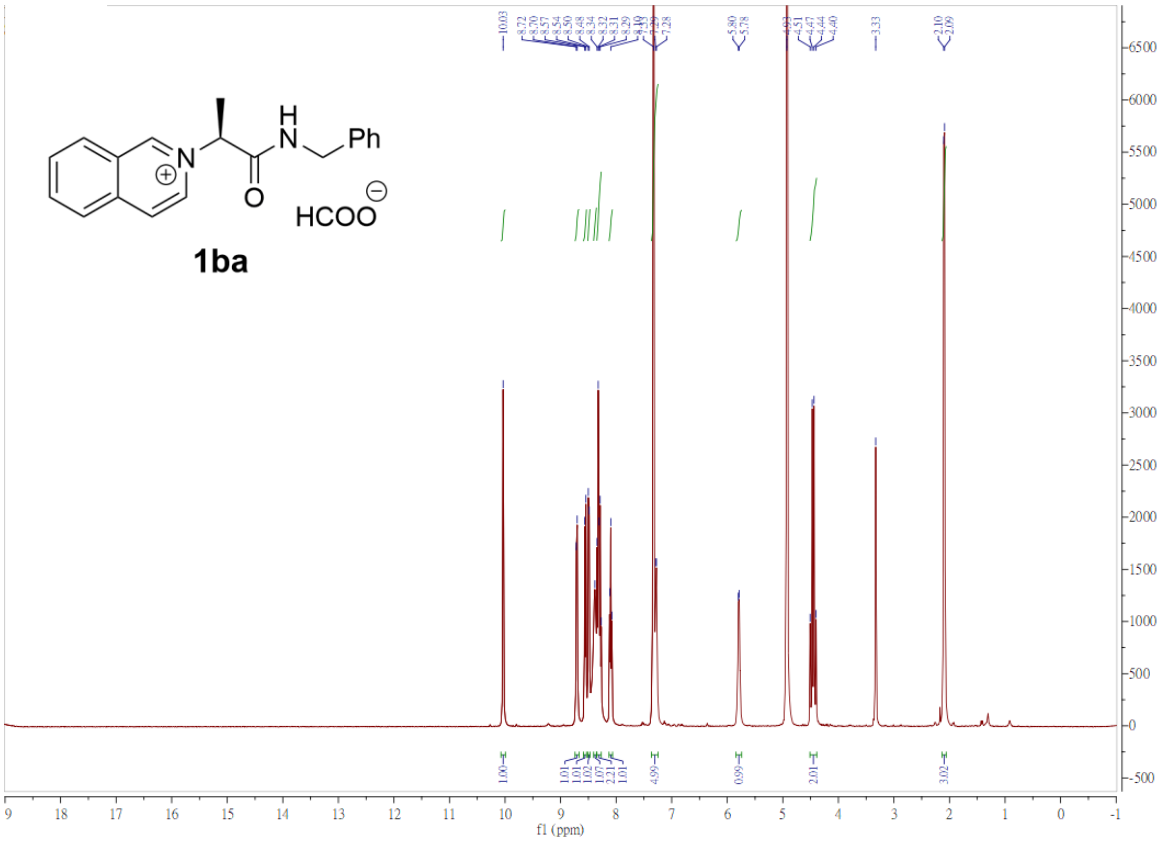


^{13}C NMR

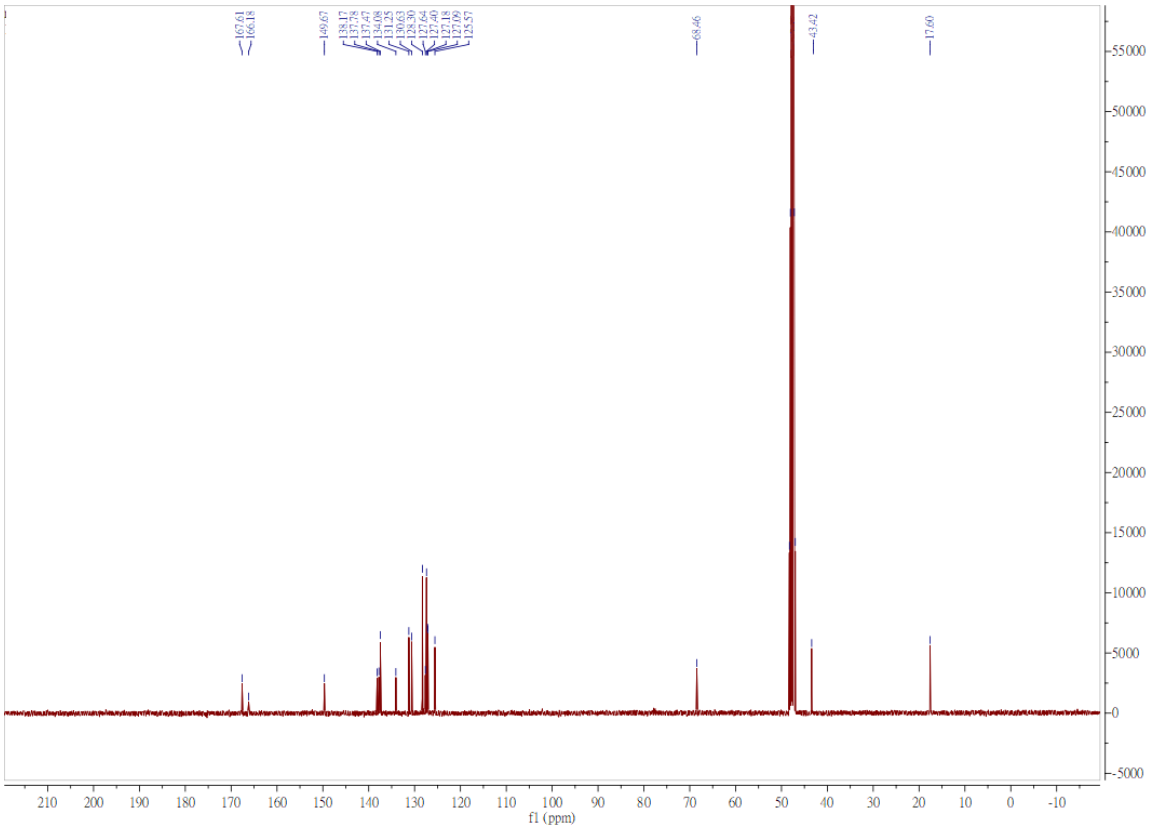


NMR spectra of 1ba

^{1}H NMR



^{13}C NMR



Modification of Peptide YTSSSKNVVR 1a using 2-Ethynylbenzaldehydes

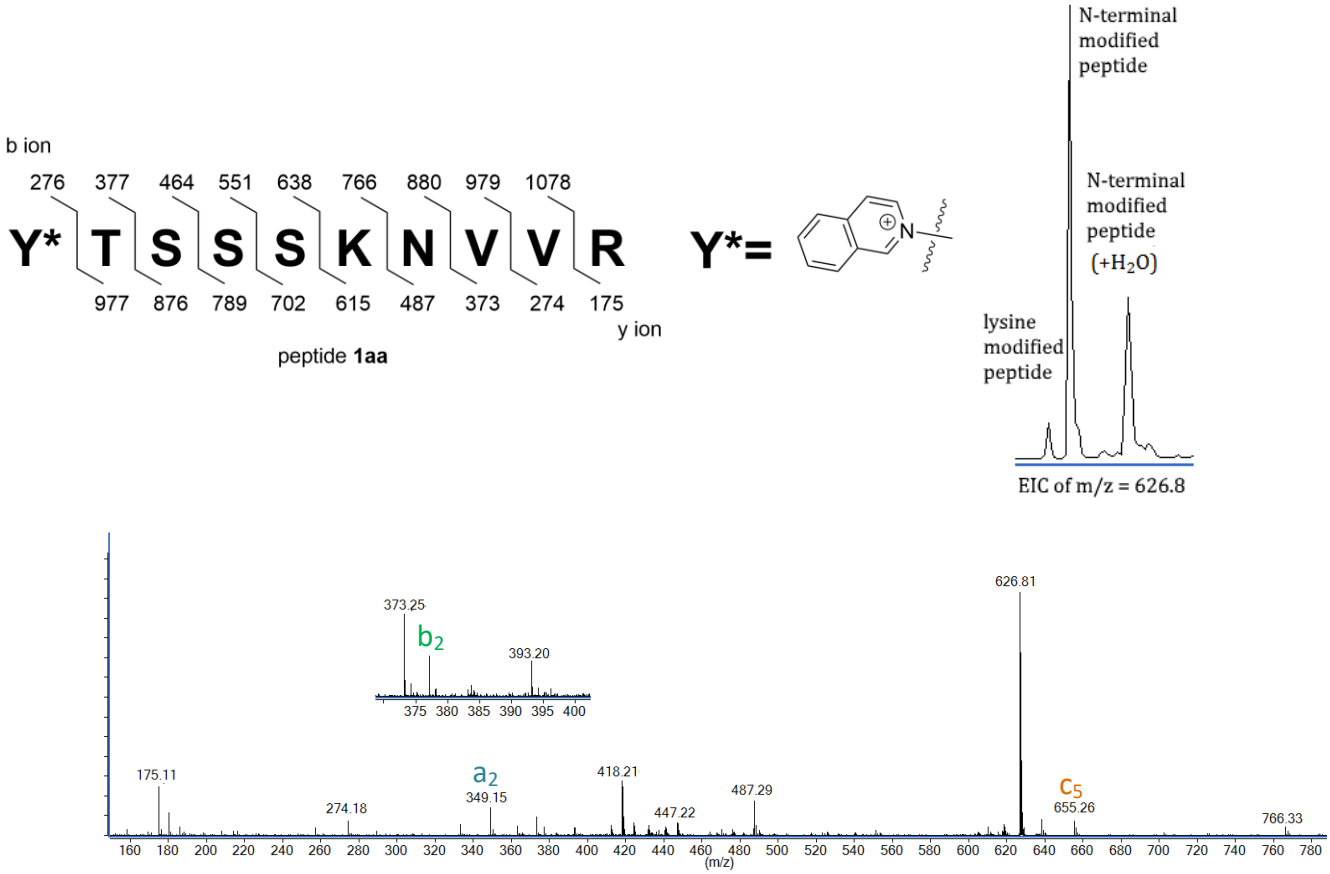


Figure S1. Q-TOF MS/MS spectrum of ***N*-terminal 2a-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 626.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 626.8.

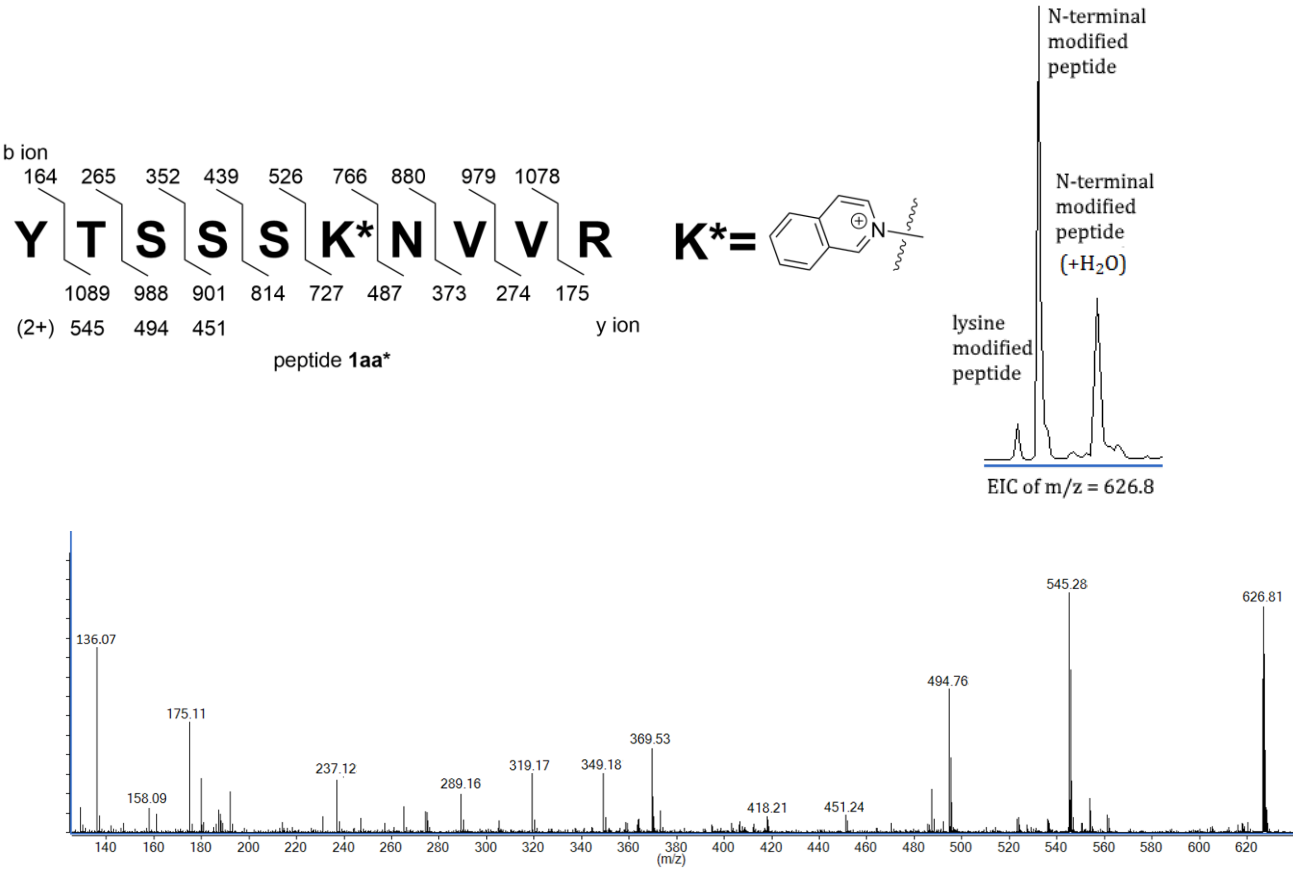


Figure S2. Q-TOF MS/MS spectrum of **Lysine 2a-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 626.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 626.8.

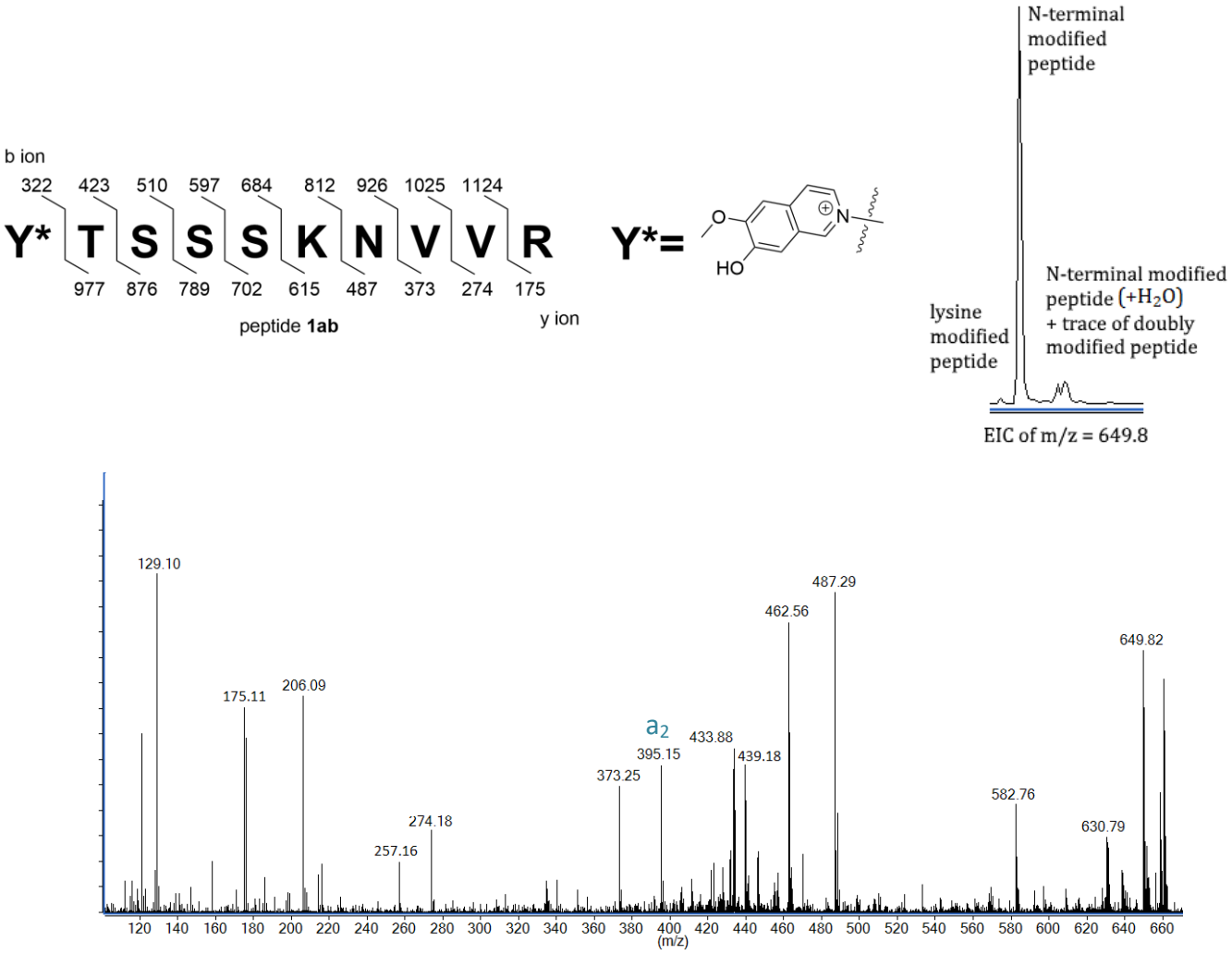


Figure S3. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 649.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 649.8.

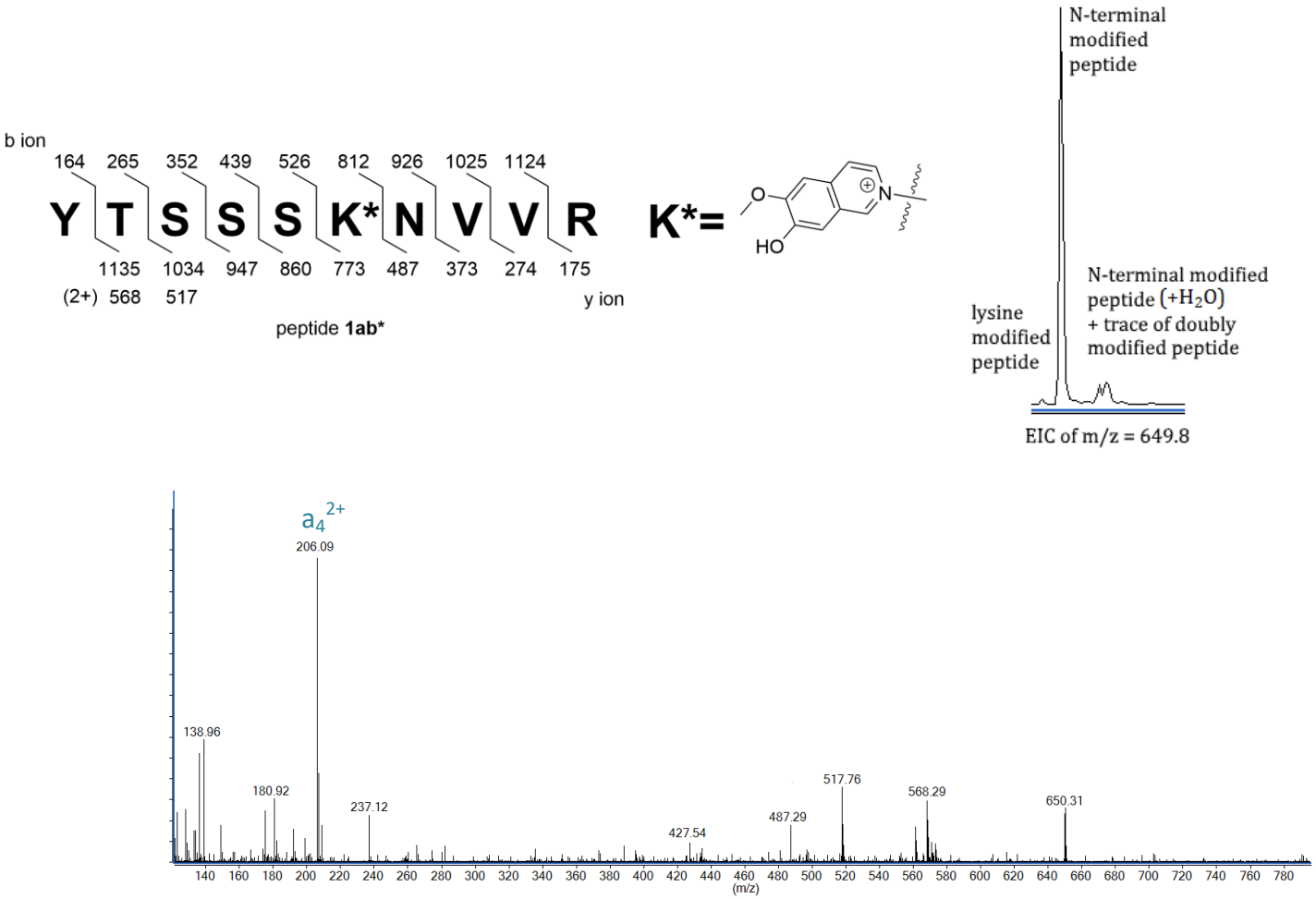


Figure S4. Q-TOF MS/MS spectrum of **Lysine 2b-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 649.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 649.8.

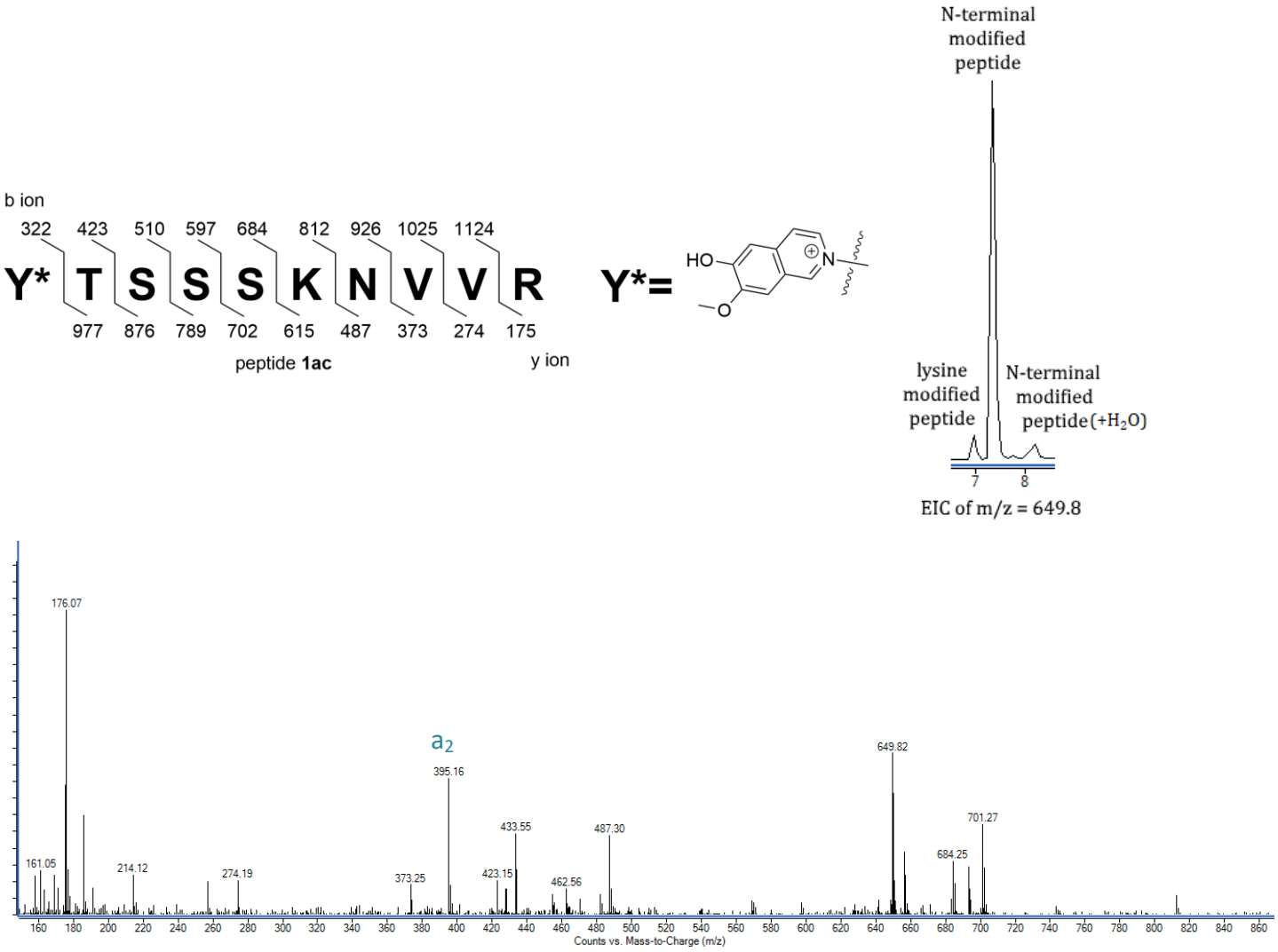


Figure S5. Q-TOF MS/MS spectrum of ***N*-terminal 2c-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 649.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 649.8.

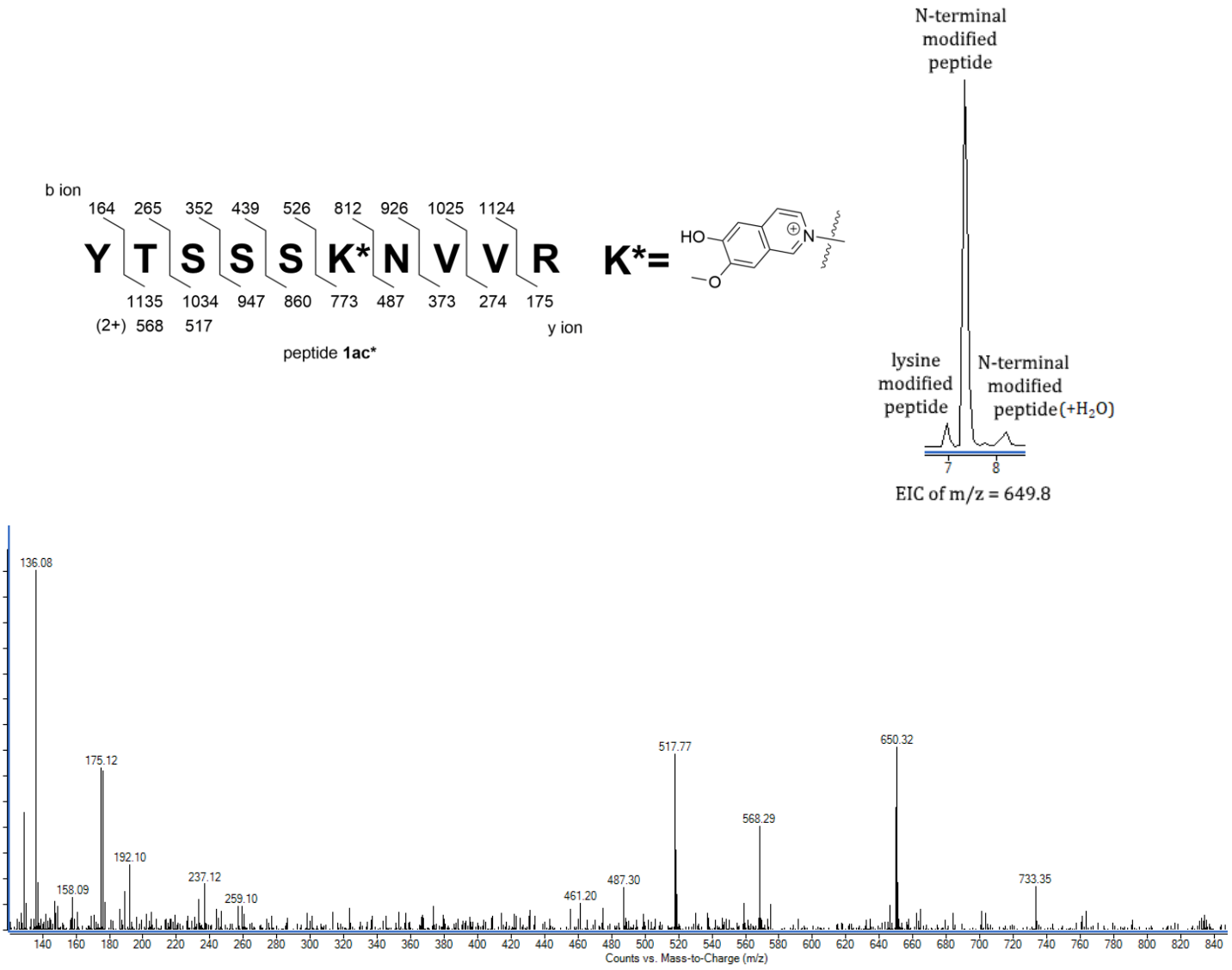


Figure S6. Q-TOF MS/MS spectrum of **Lysine 2c-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 649.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 649.8.

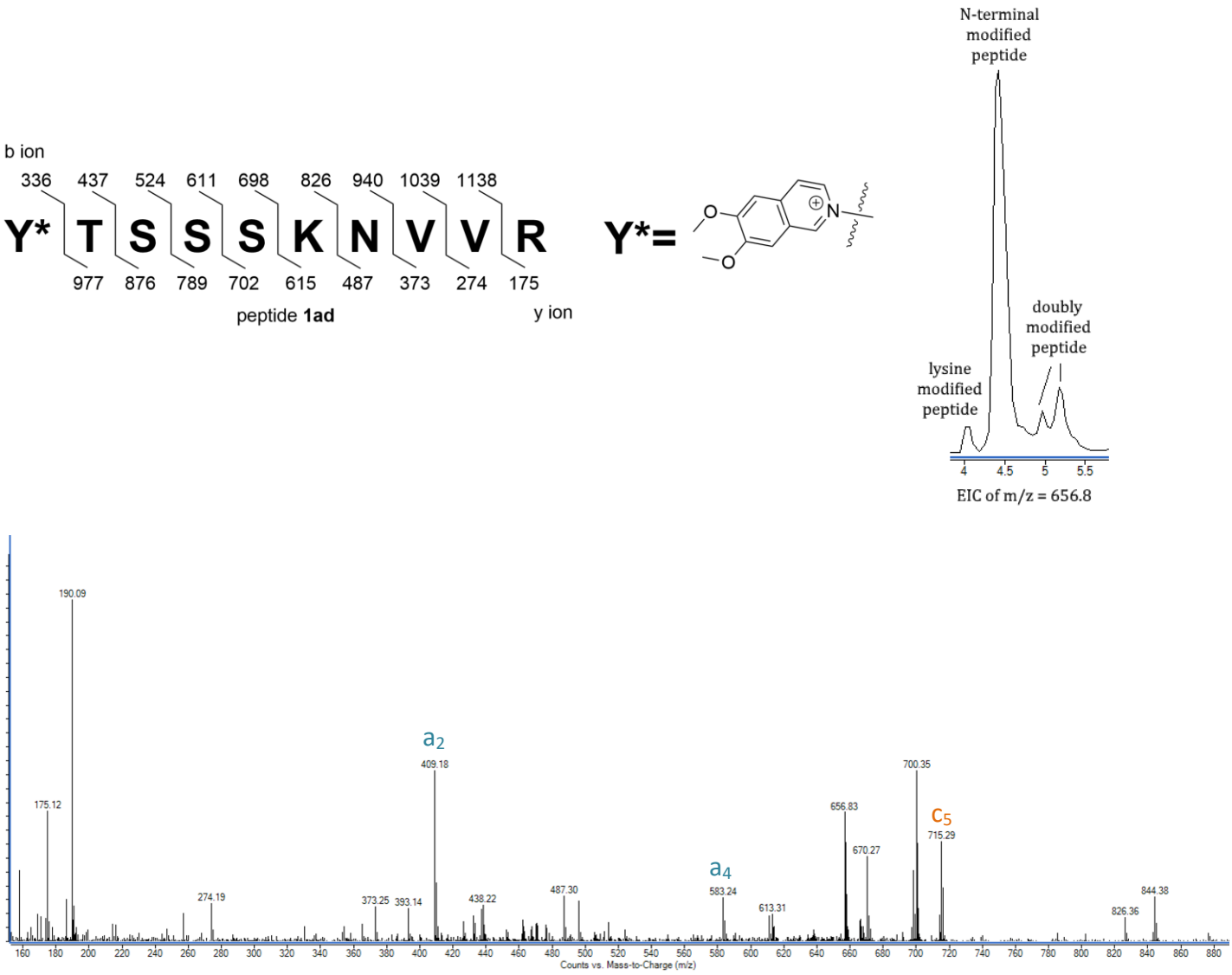


Figure S7. Q-TOF MS/MS spectrum of ***N*-terminal 2d-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 656.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 656.8.

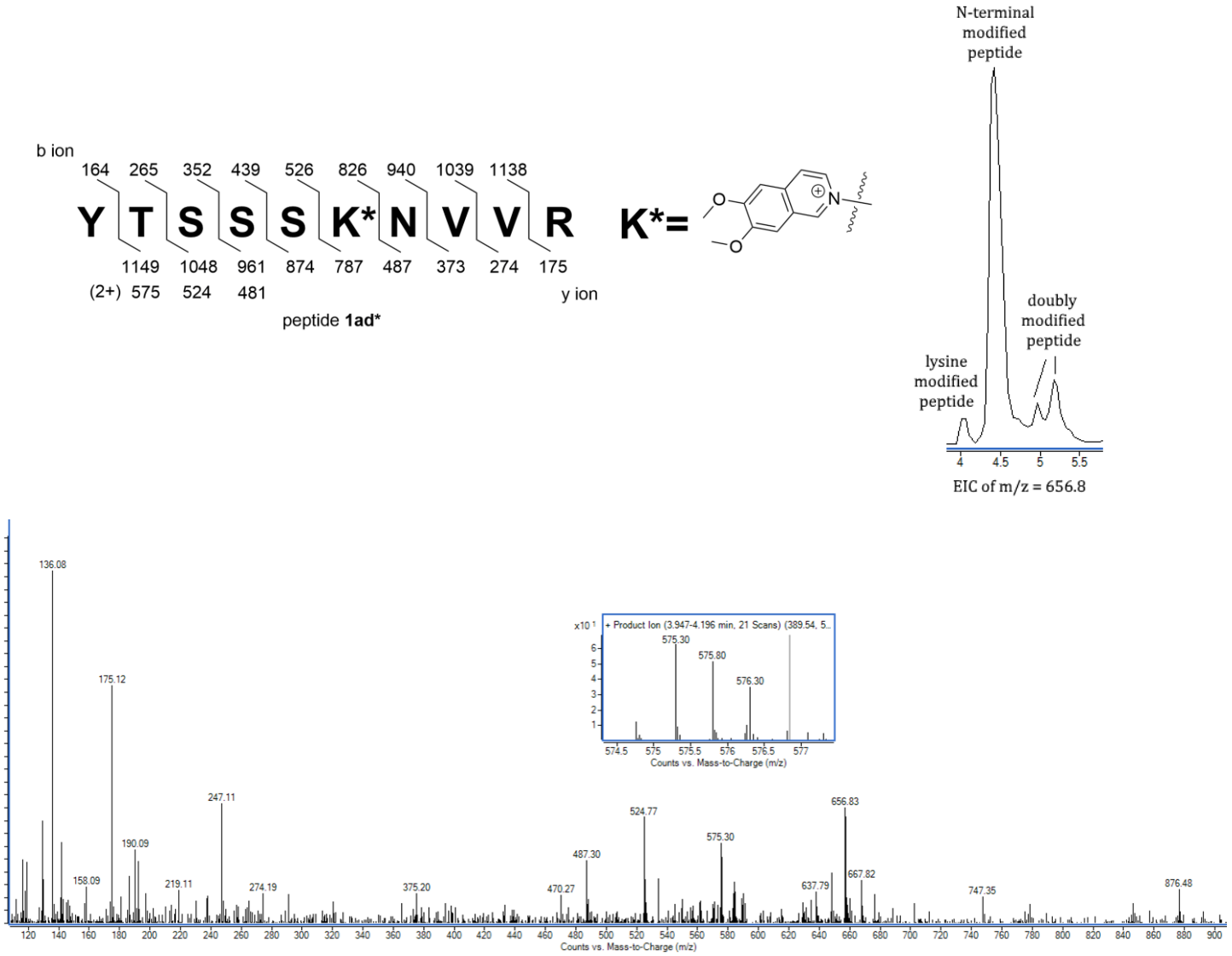


Figure S8. Q-TOF MS/MS spectrum of **Lysine 2d-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 656.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 656.8

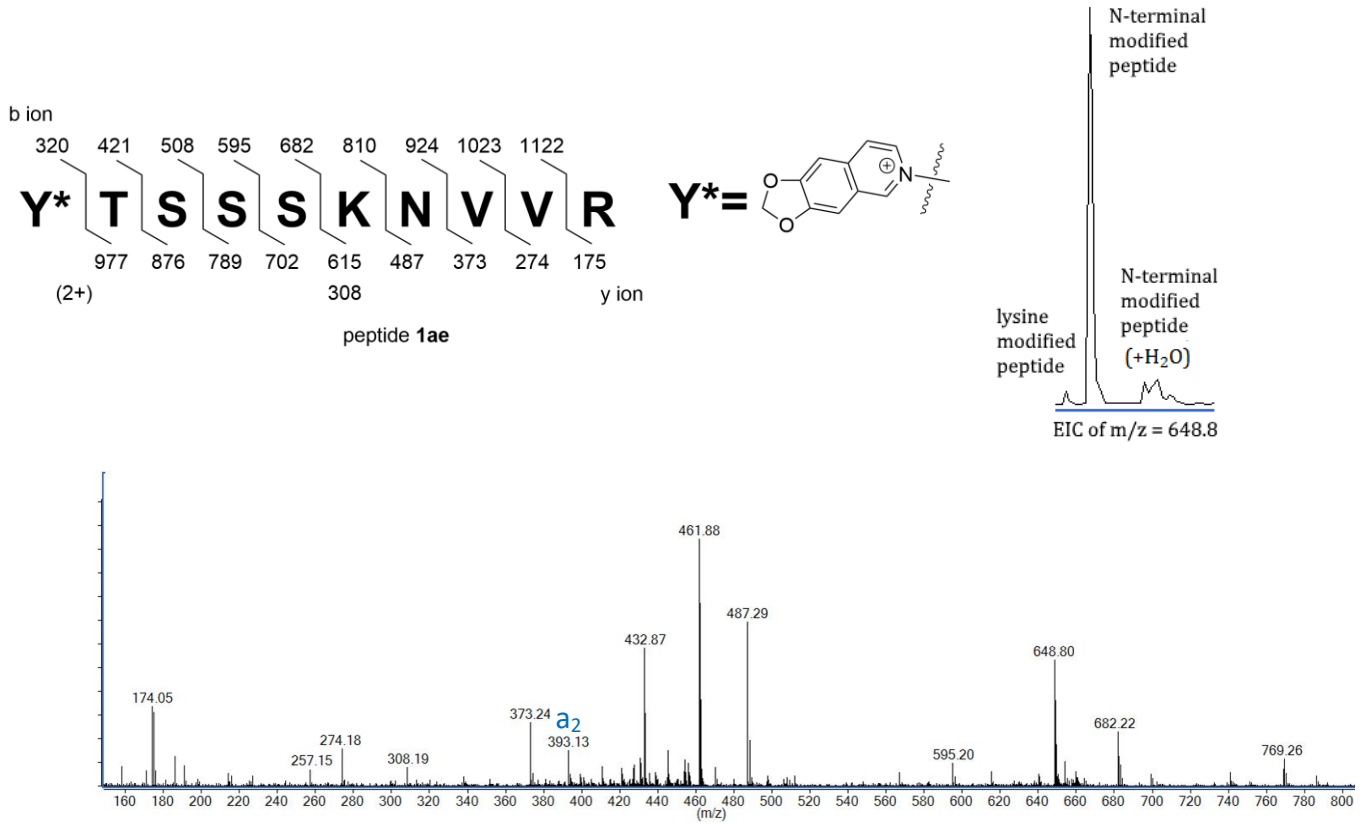


Figure S9. Q-TOF MS/MS spectrum of ***N*-terminal 2e-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 648.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 648.8.

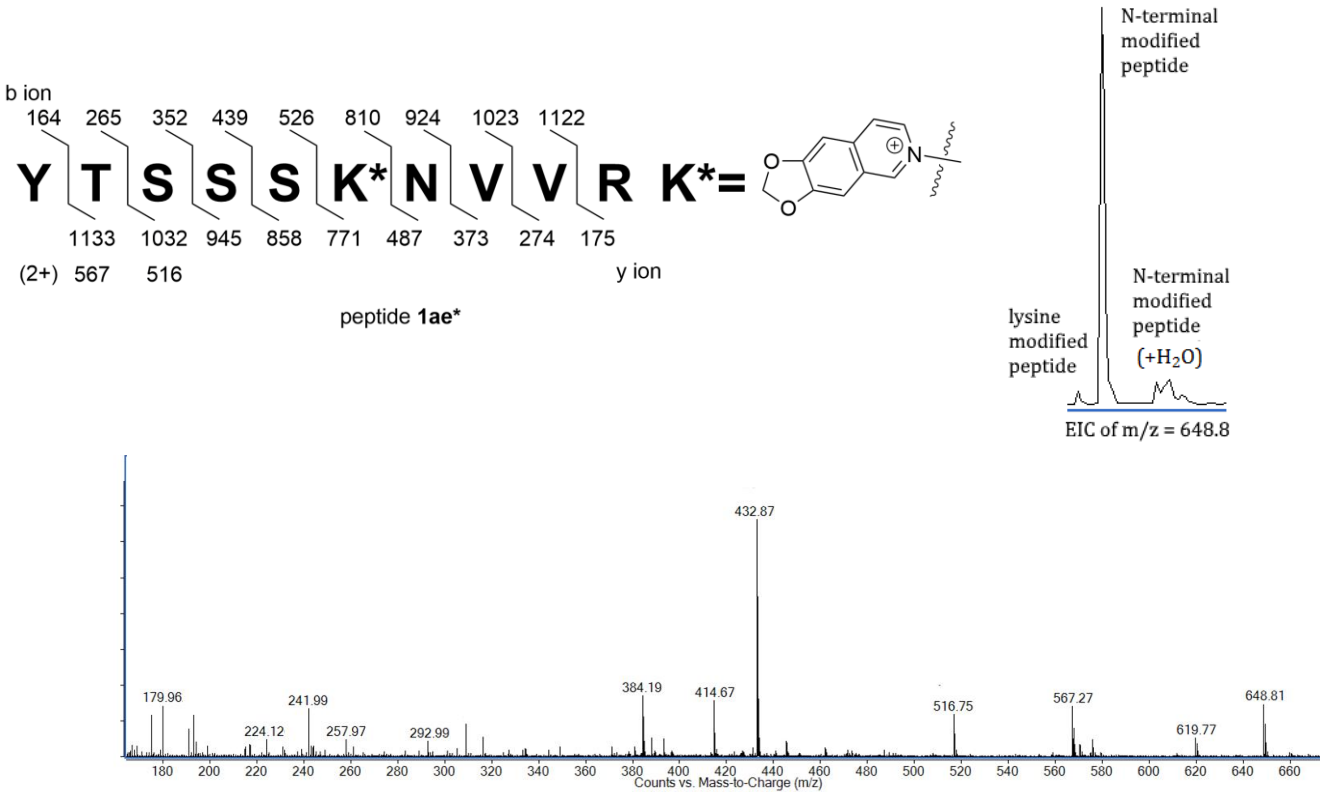


Figure S10. Q-TOF MS/MS spectrum of **Lysine 2e-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 648.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 648.8.

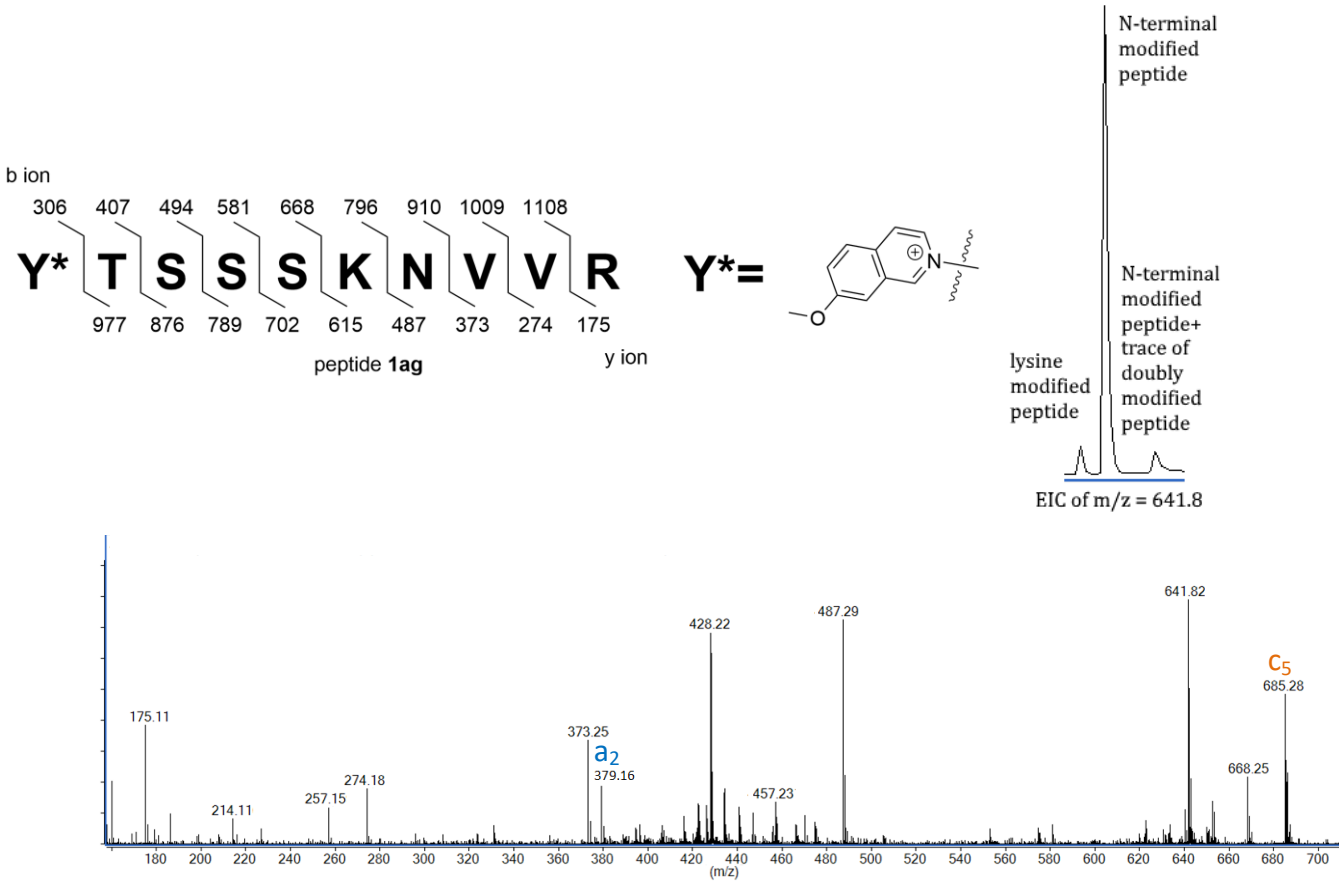


Figure S11. Q-TOF MS/MS spectrum of ***N*-terminal 2g-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 641.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 641.8.

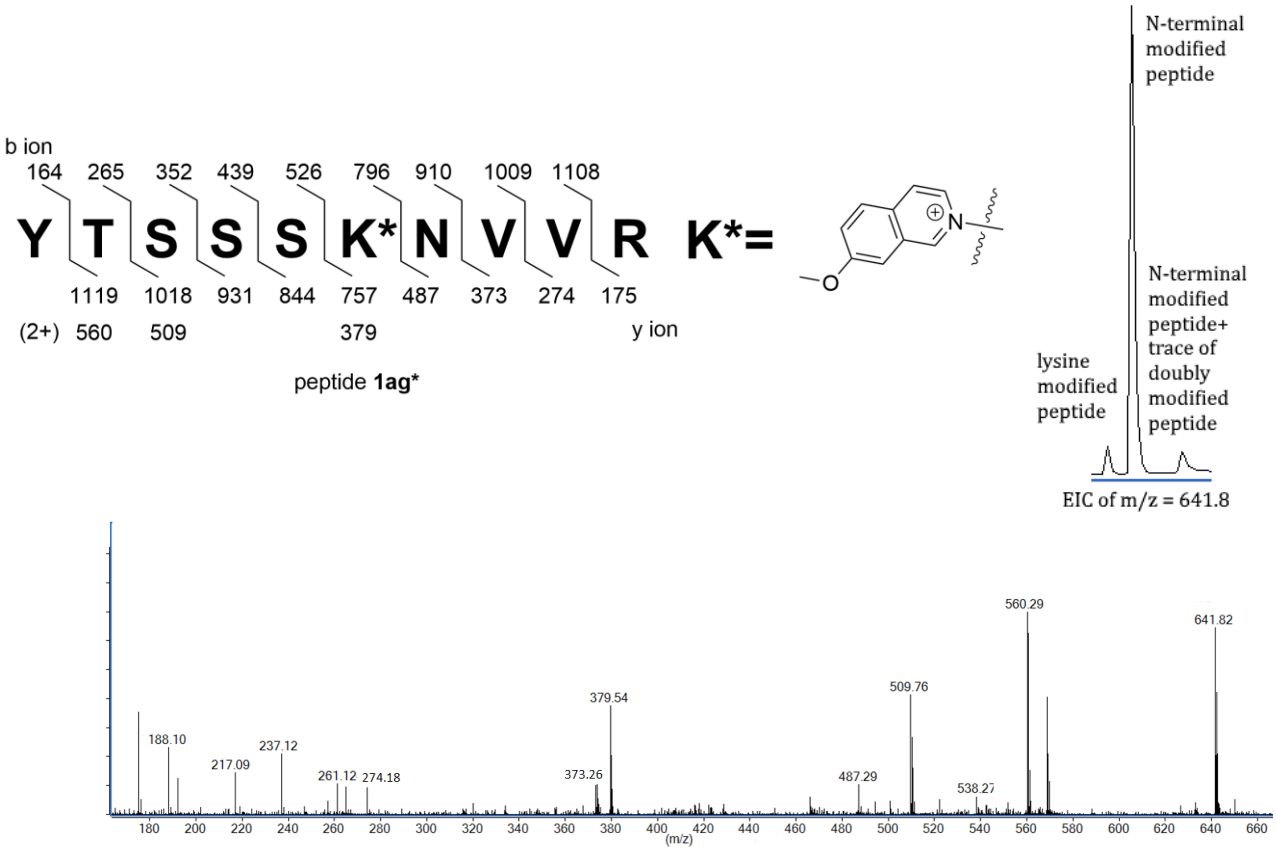


Figure S12. Q-TOF MS/MS spectrum of **Lysine 2g-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 641.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 641.8.

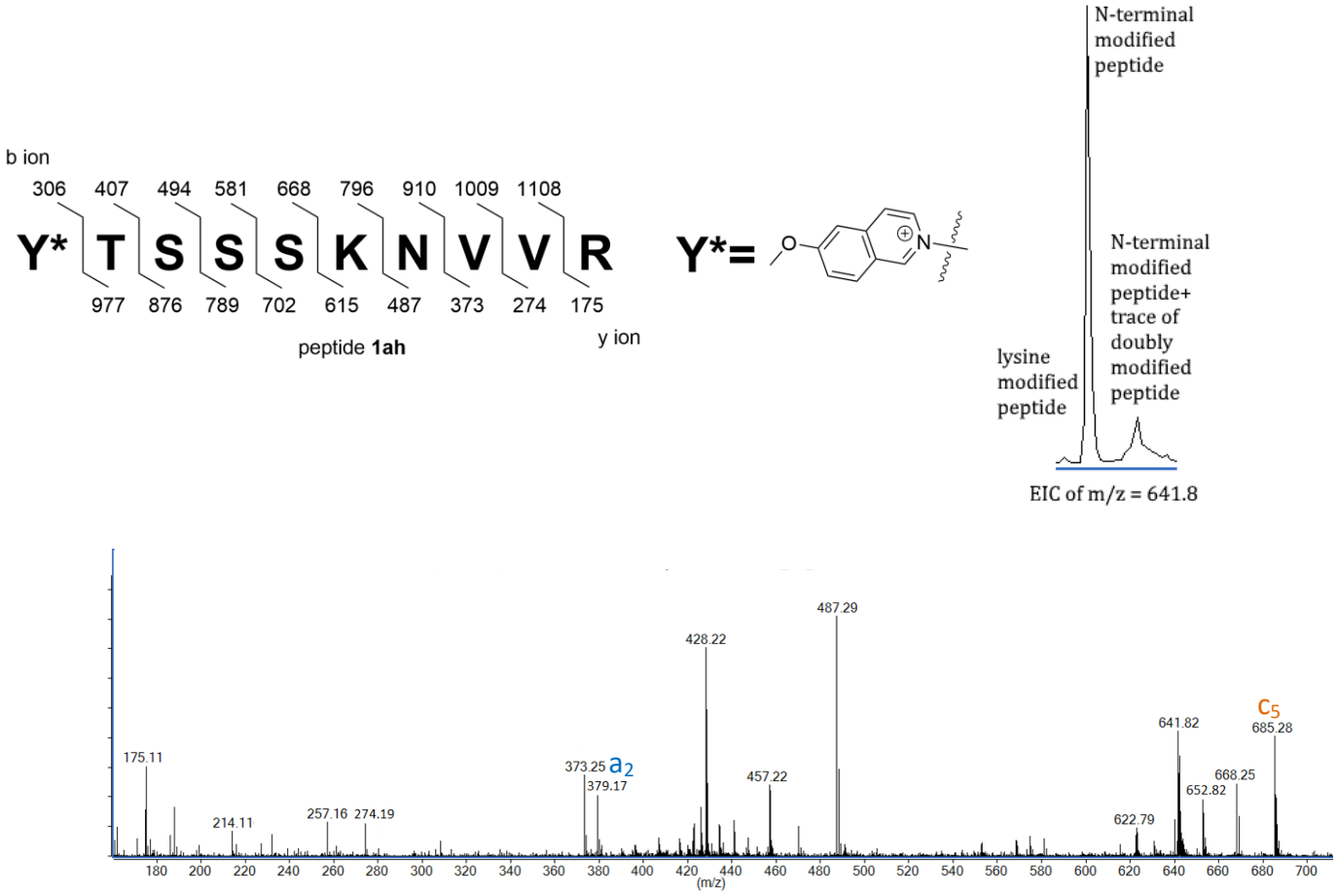


Figure S13. Q-TOF MS/MS spectrum of ***N*-terminal 2h-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 641.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 641.8.

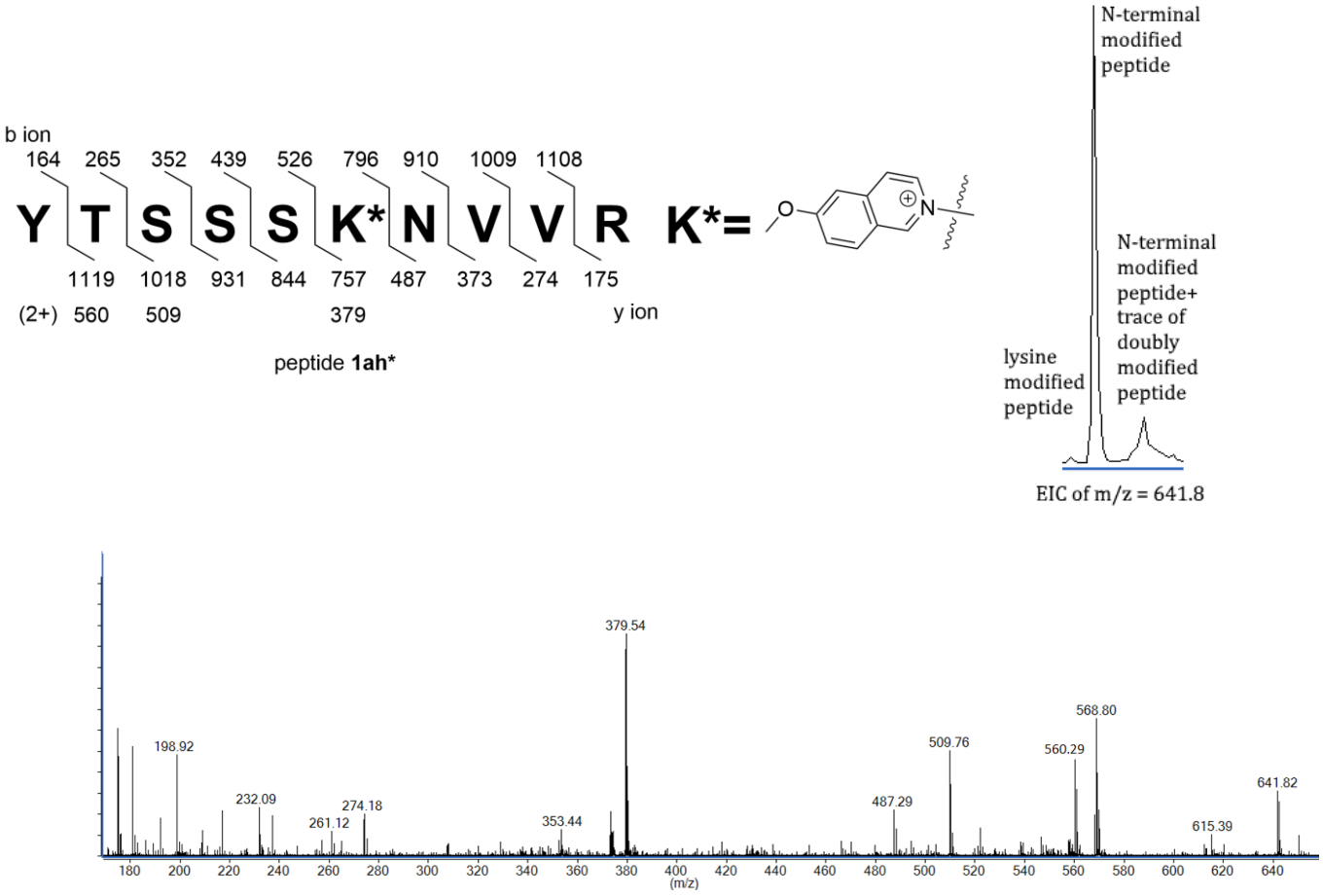


Figure S14. Q-TOF MS/MS spectrum of **Lysine 2h-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 641.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 641.8.

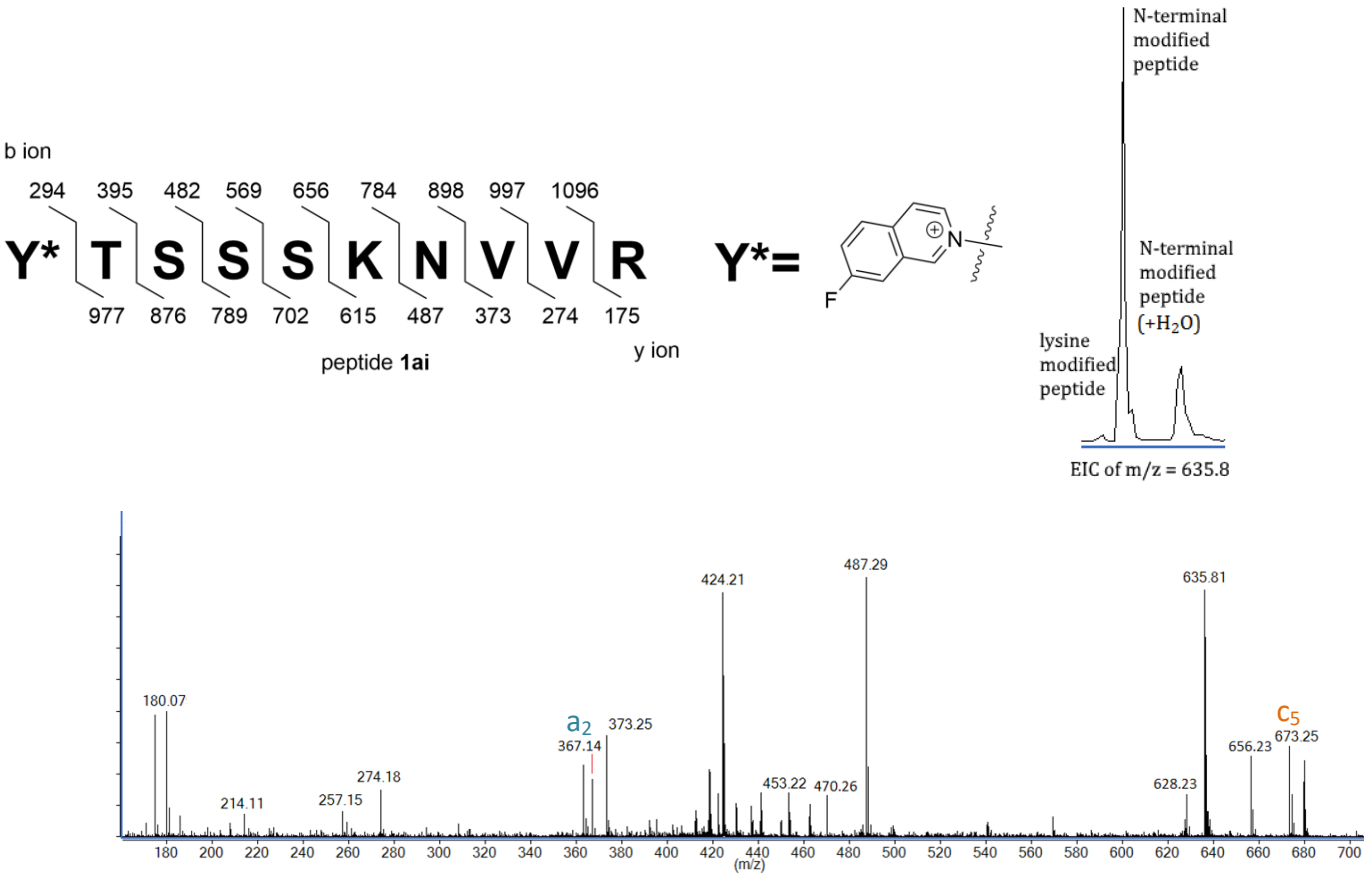


Figure S15. Q-TOF MS/MS spectrum of ***N*-terminal 2i-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 635.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 635.8.

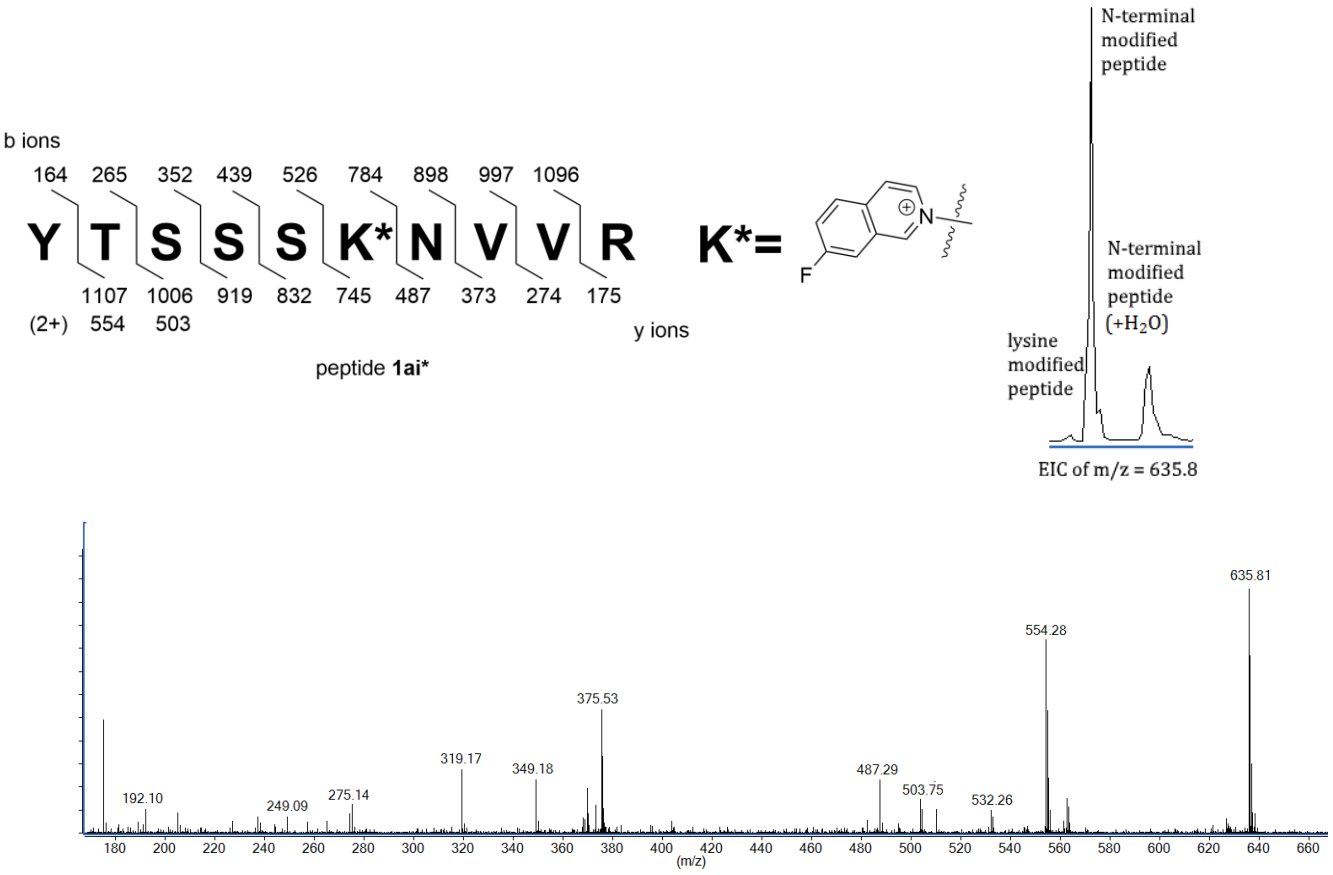


Figure S16. Q-TOF MS/MS spectrum of **Lysine 2i-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 635.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 635.8.

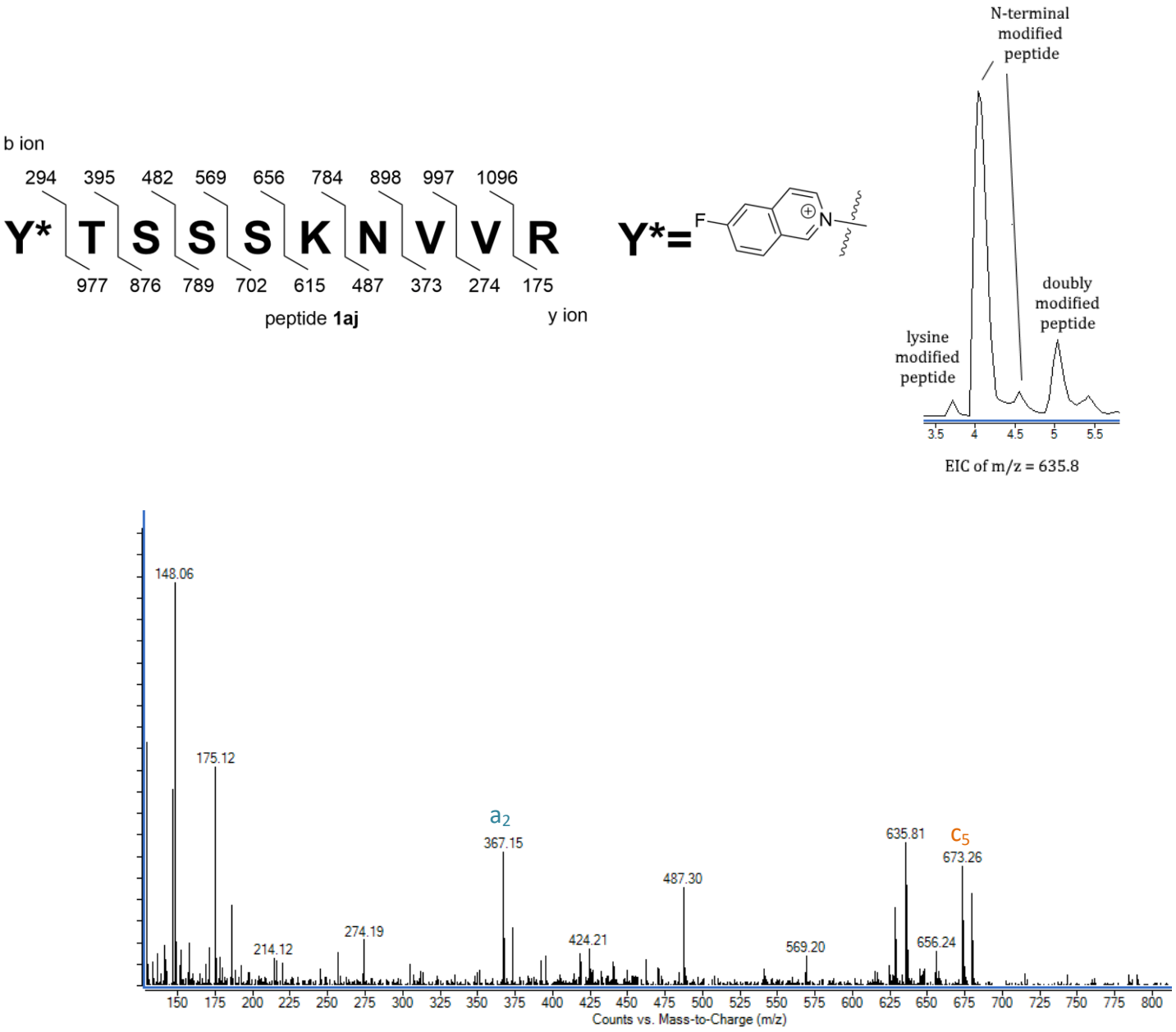


Figure S17. Q-TOF MS/MS spectrum of ***N*-terminal 2j-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 635.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 635.8.

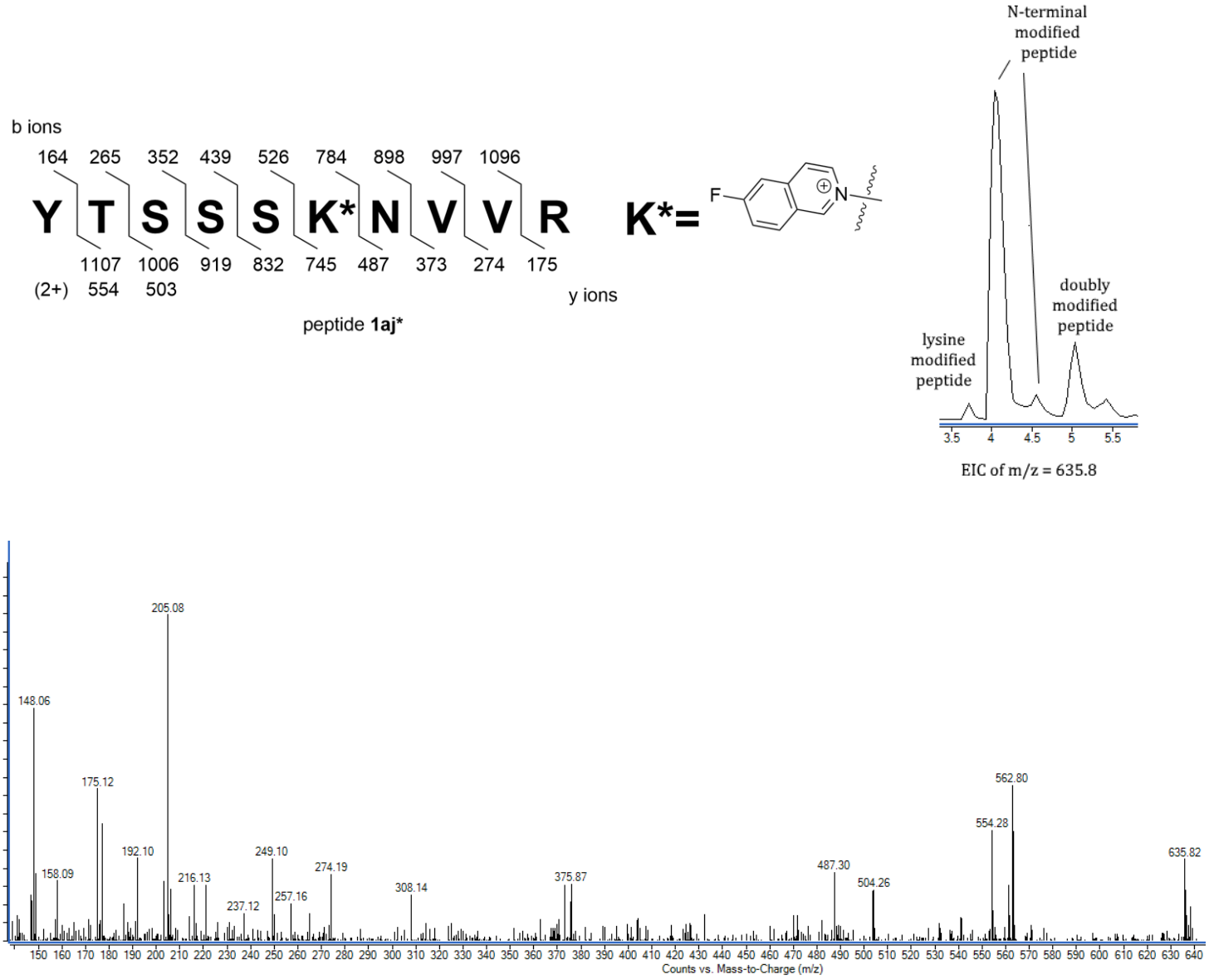


Figure S18. Q-TOF MS/MS spectrum of **Lysine 2j-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 635.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 635.8.

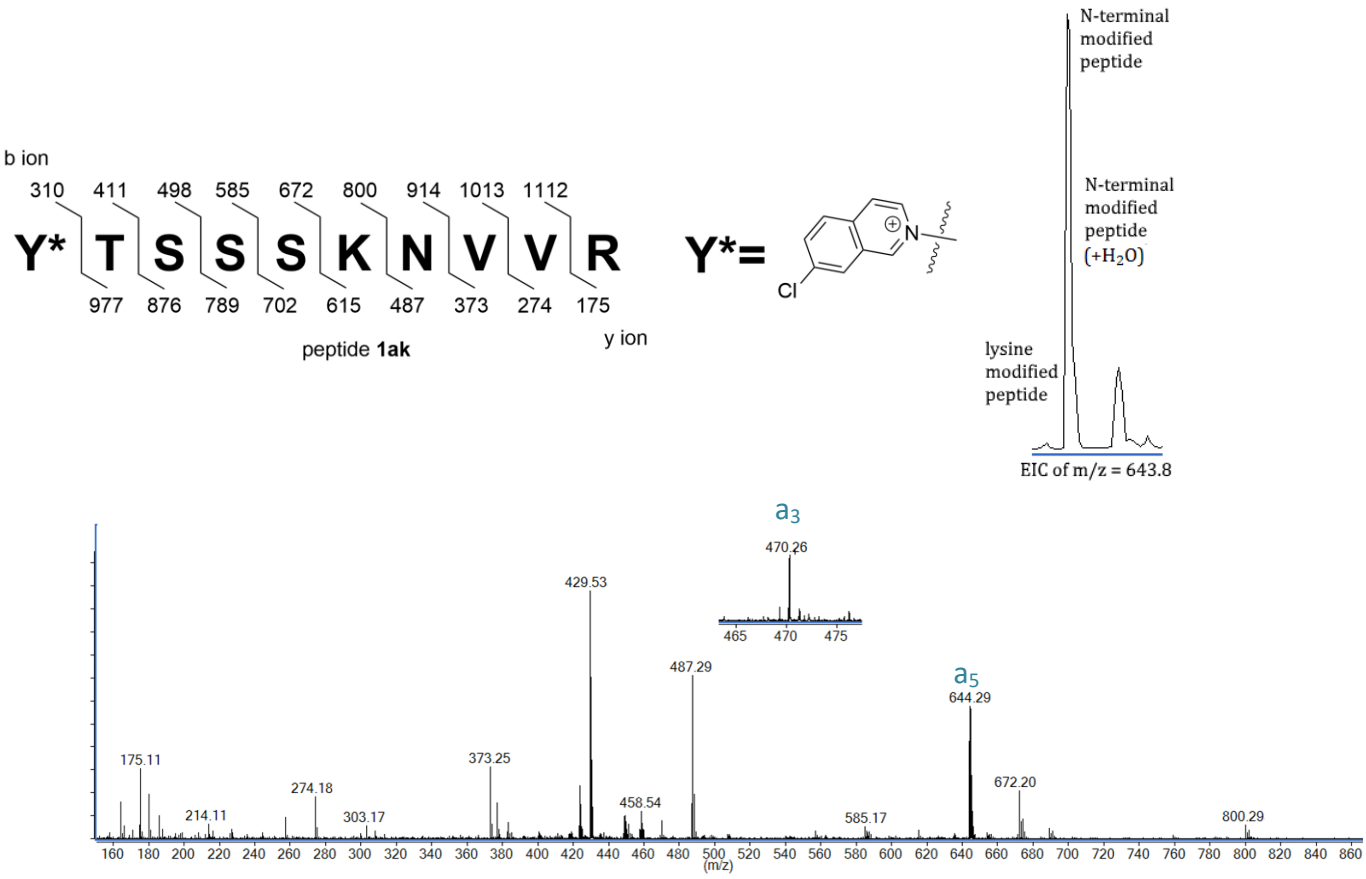


Figure S19. Q-TOF MS/MS spectrum of ***N*-terminal 2k-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 643.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 643.8.

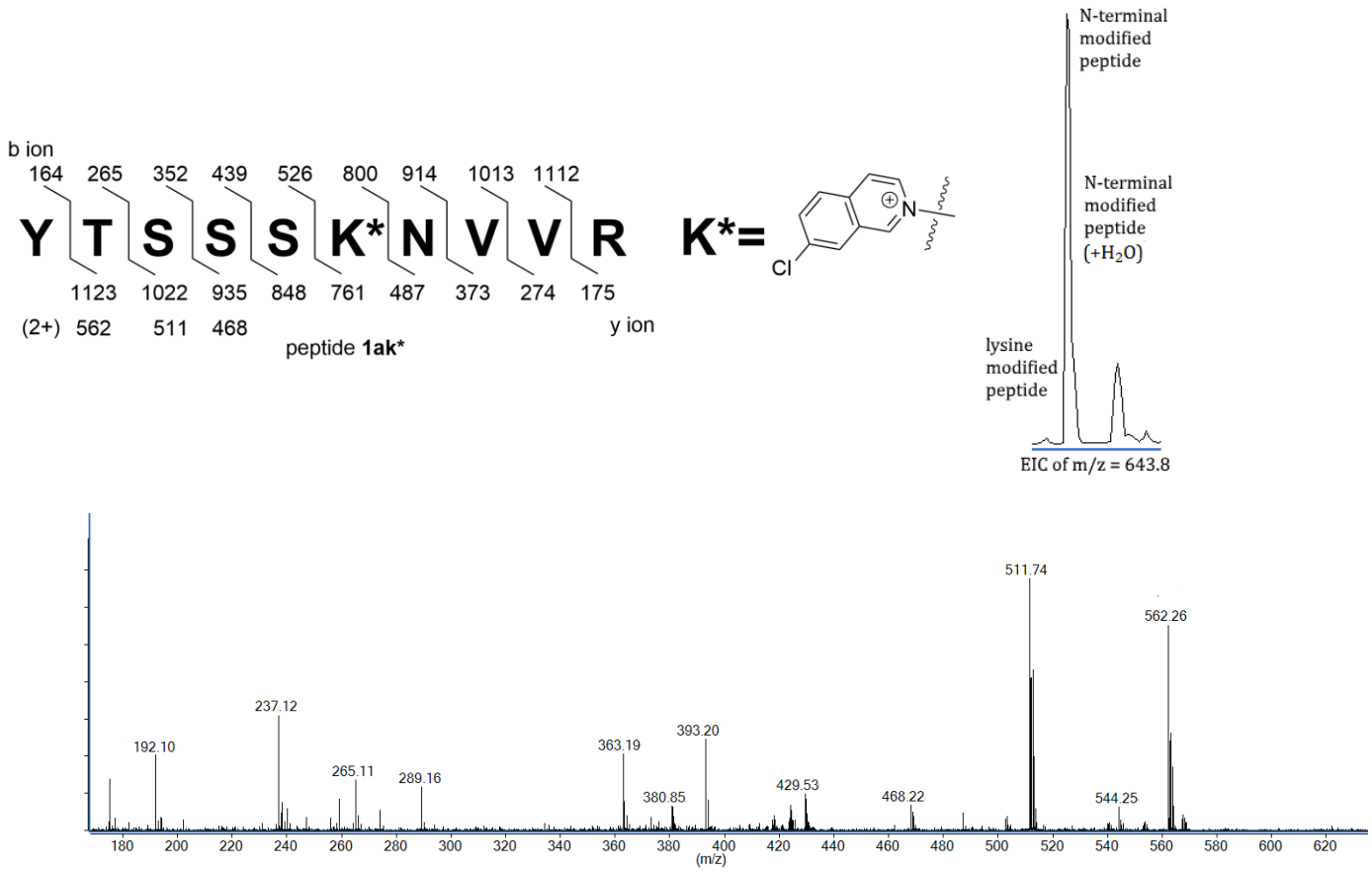


Figure S20. Q-TOF MS/MS spectrum of **Lysine 2k-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 643.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 643.8.

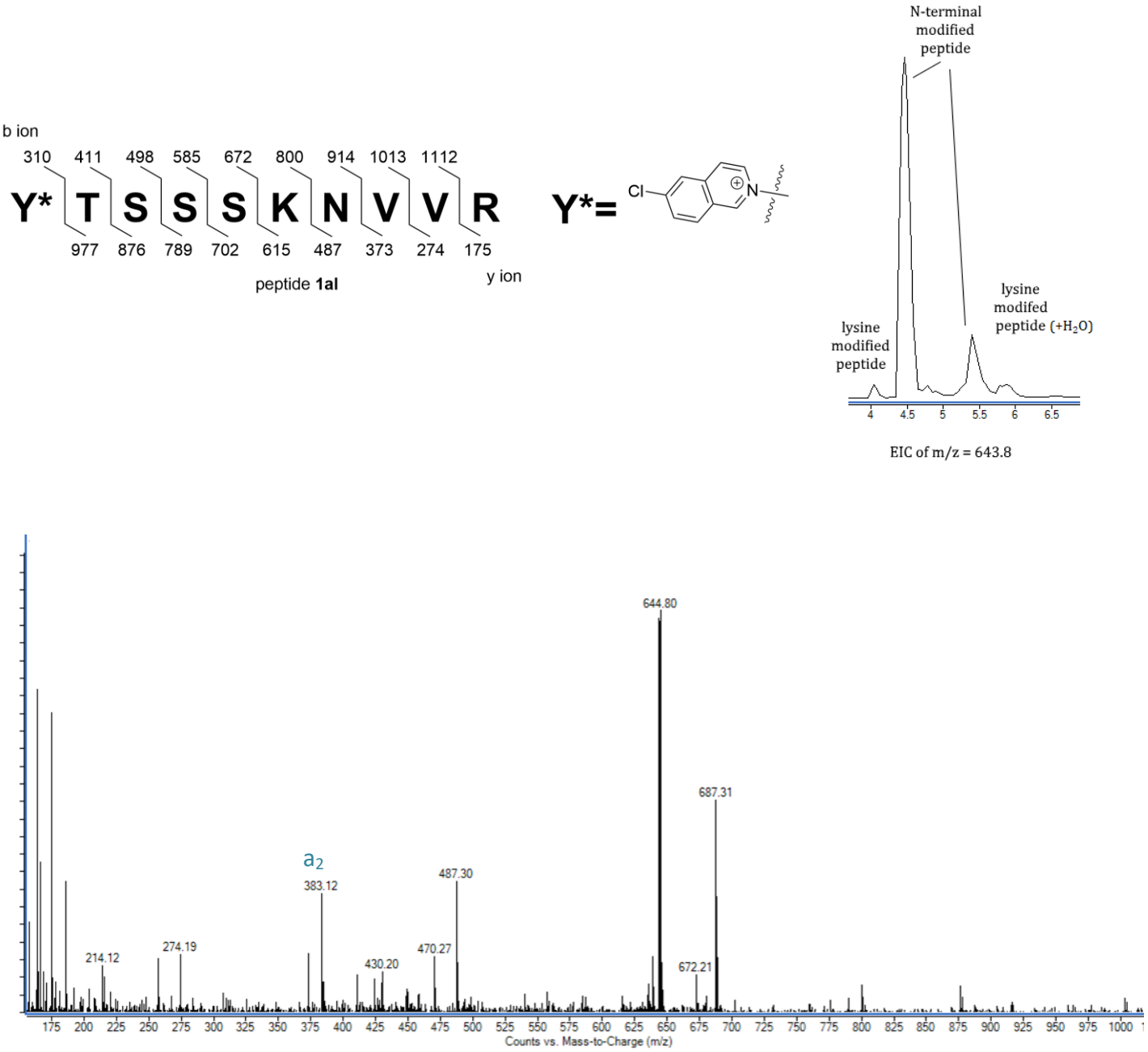
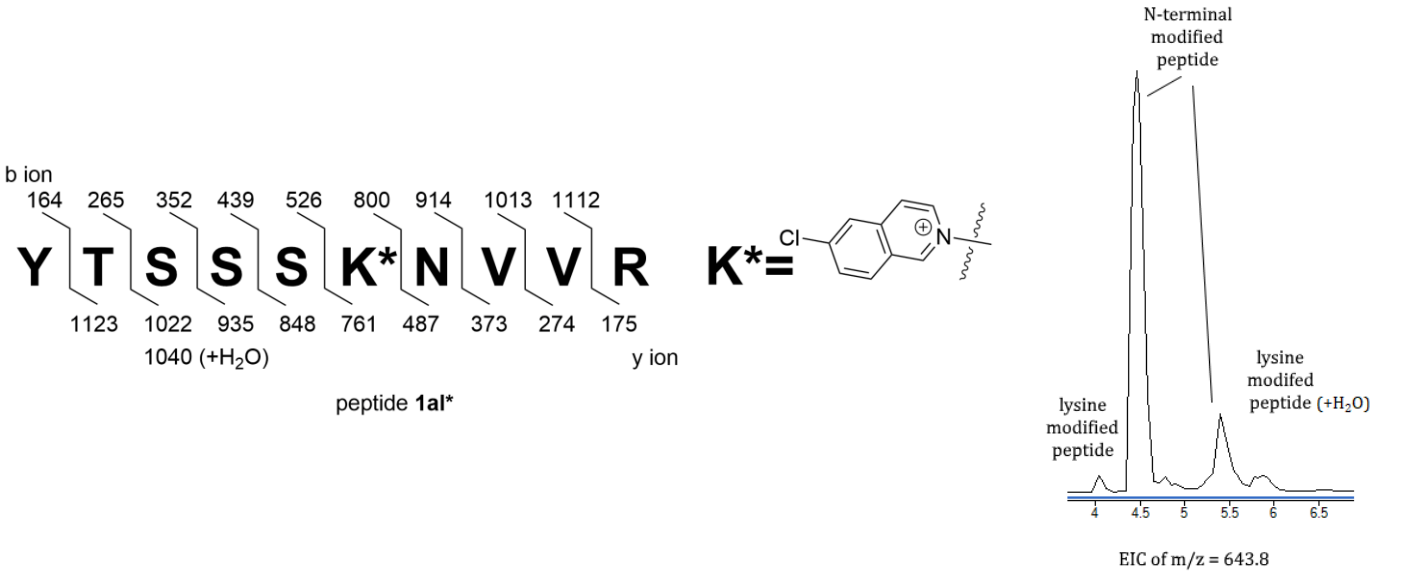


Figure S21. Q-TOF MS/MS spectrum of ***N*-terminal 2l-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 643.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 643.8.



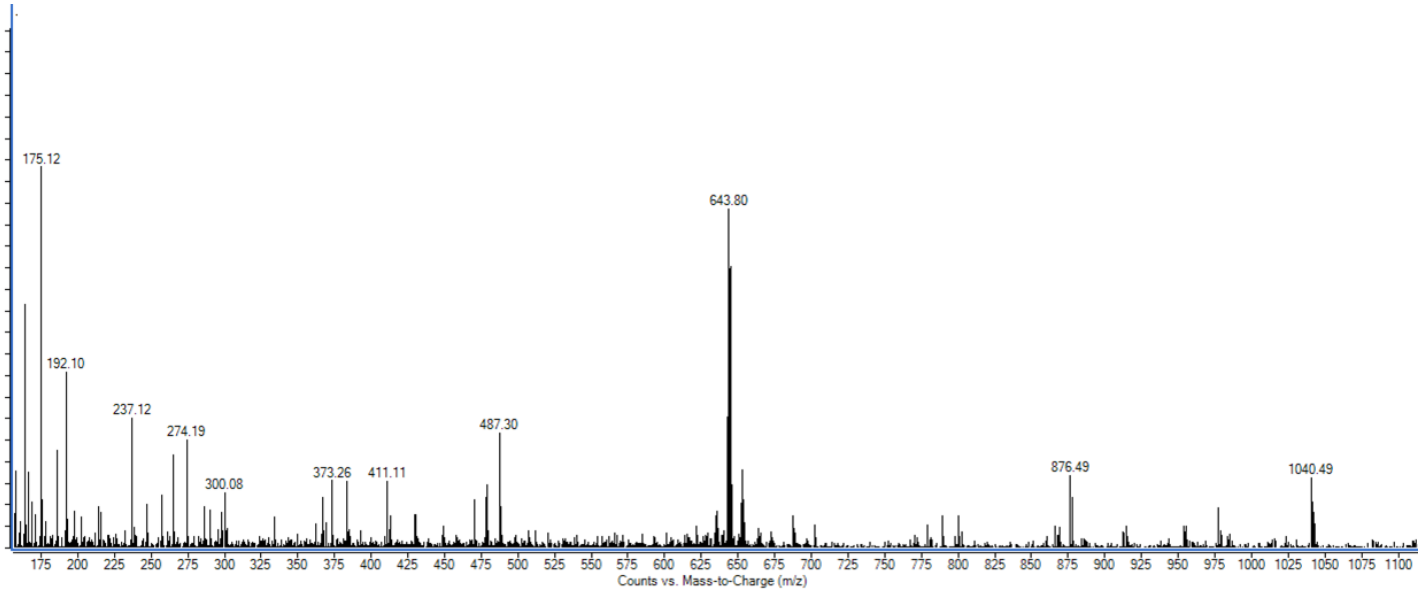


Figure S22. Q-TOF MS/MS spectrum of **Lysine 2l-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 643.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 643.8.

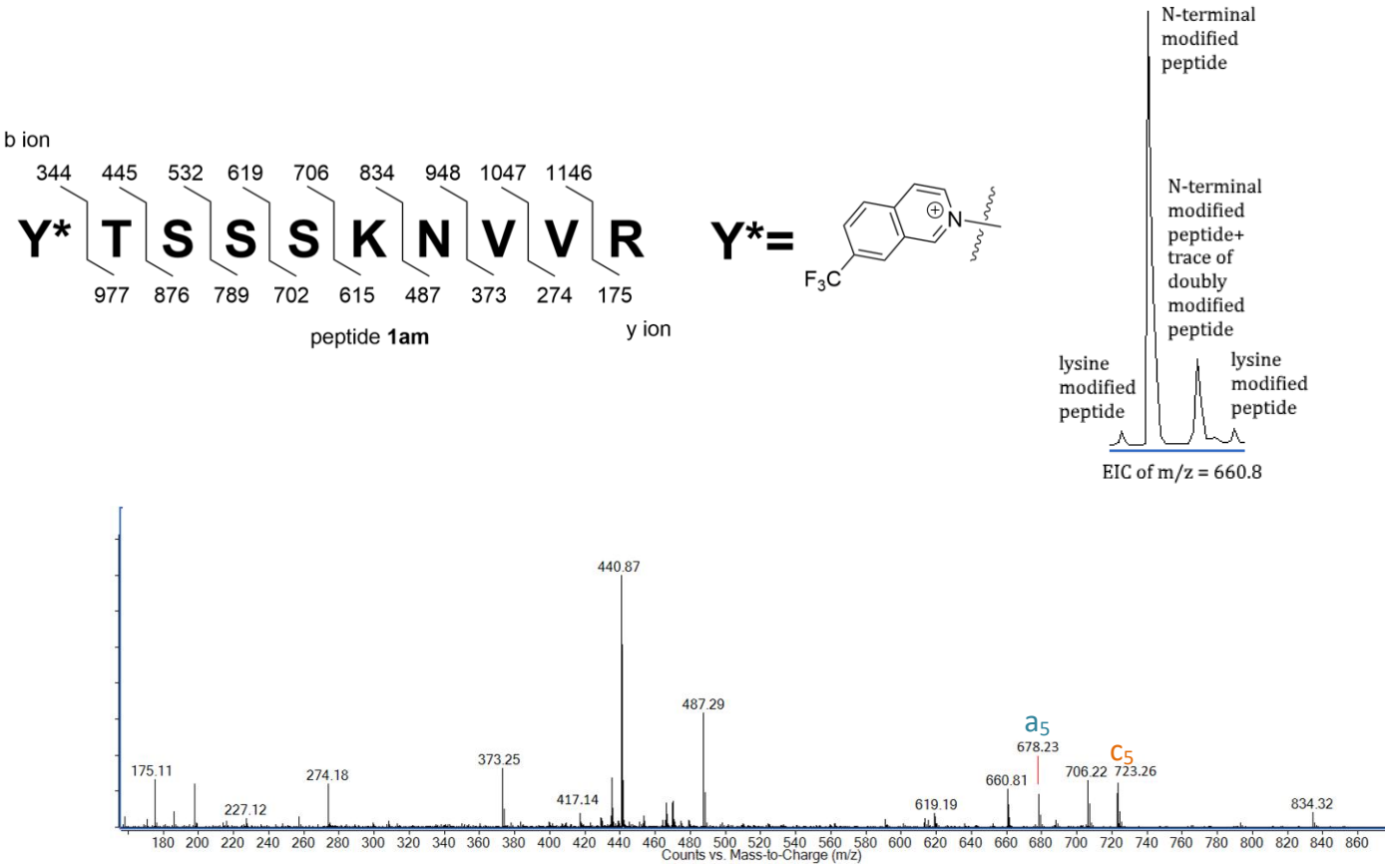


Figure S23. Q-TOF MS/MS spectrum of ***N*-terminal 2m-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 660.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 660.8.

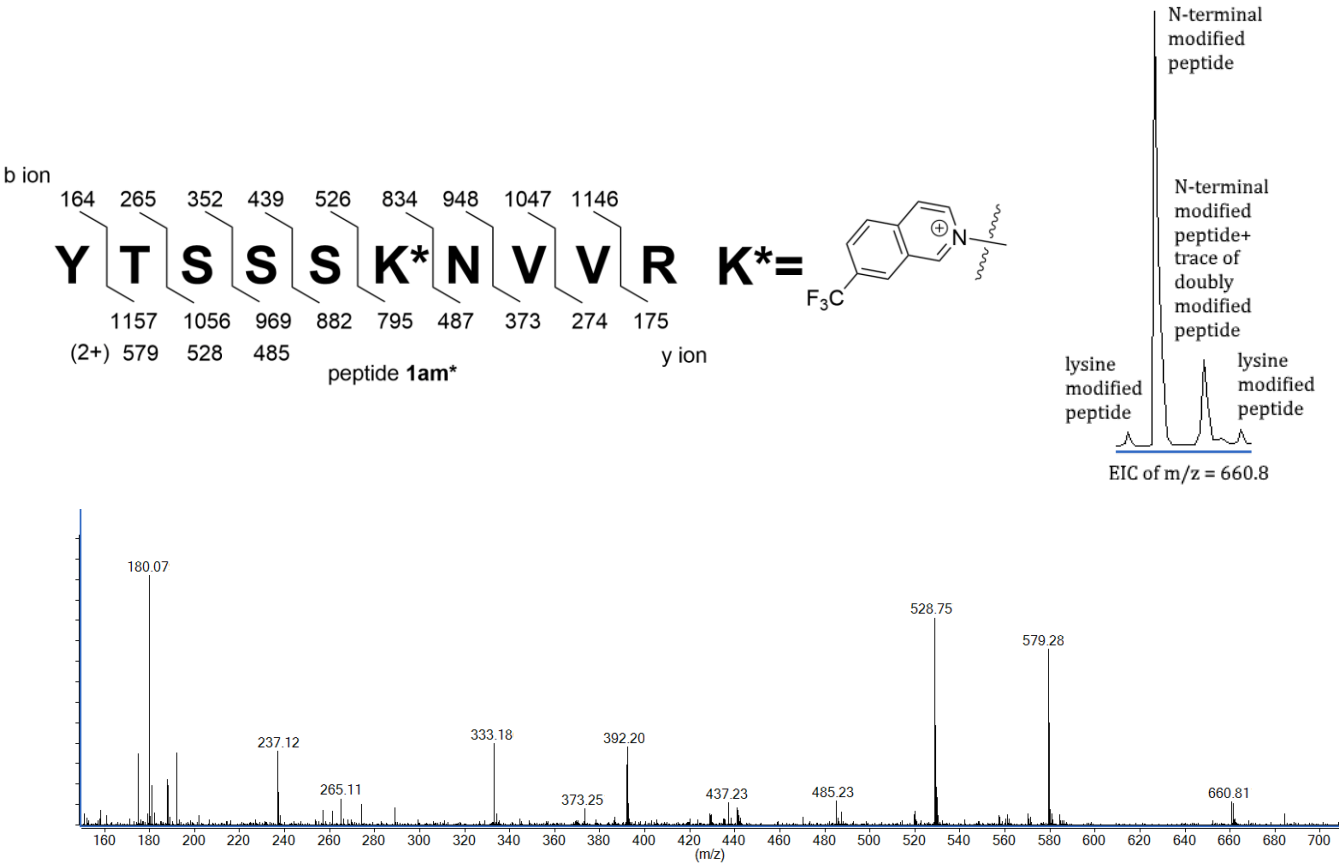


Figure S24. Q-TOF MS/MS spectrum of **Lysine 2m-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 660.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 660.8.

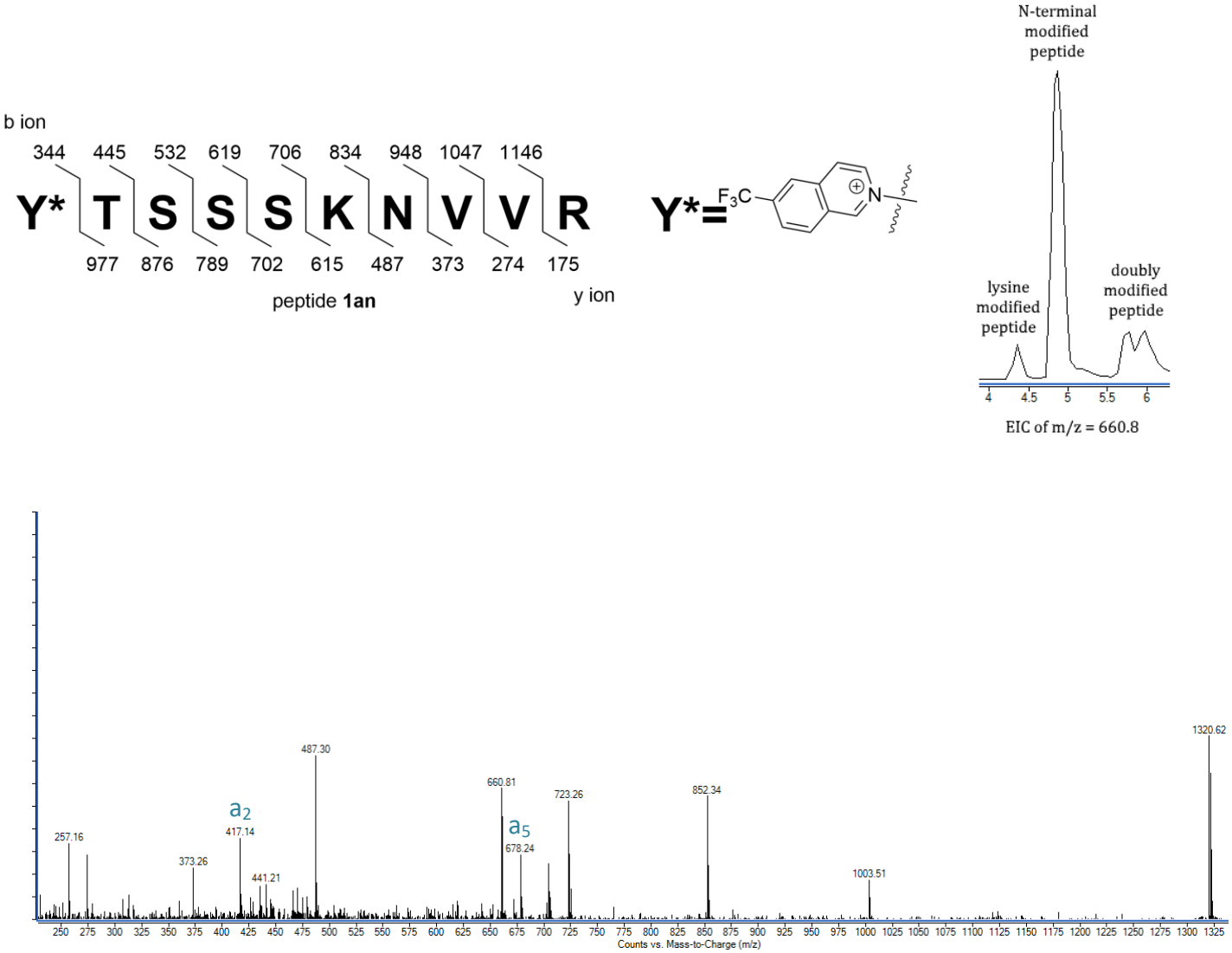


Figure S25. Q-TOF MS/MS spectrum of ***N*-terminal 2n-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 660.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 660.8.

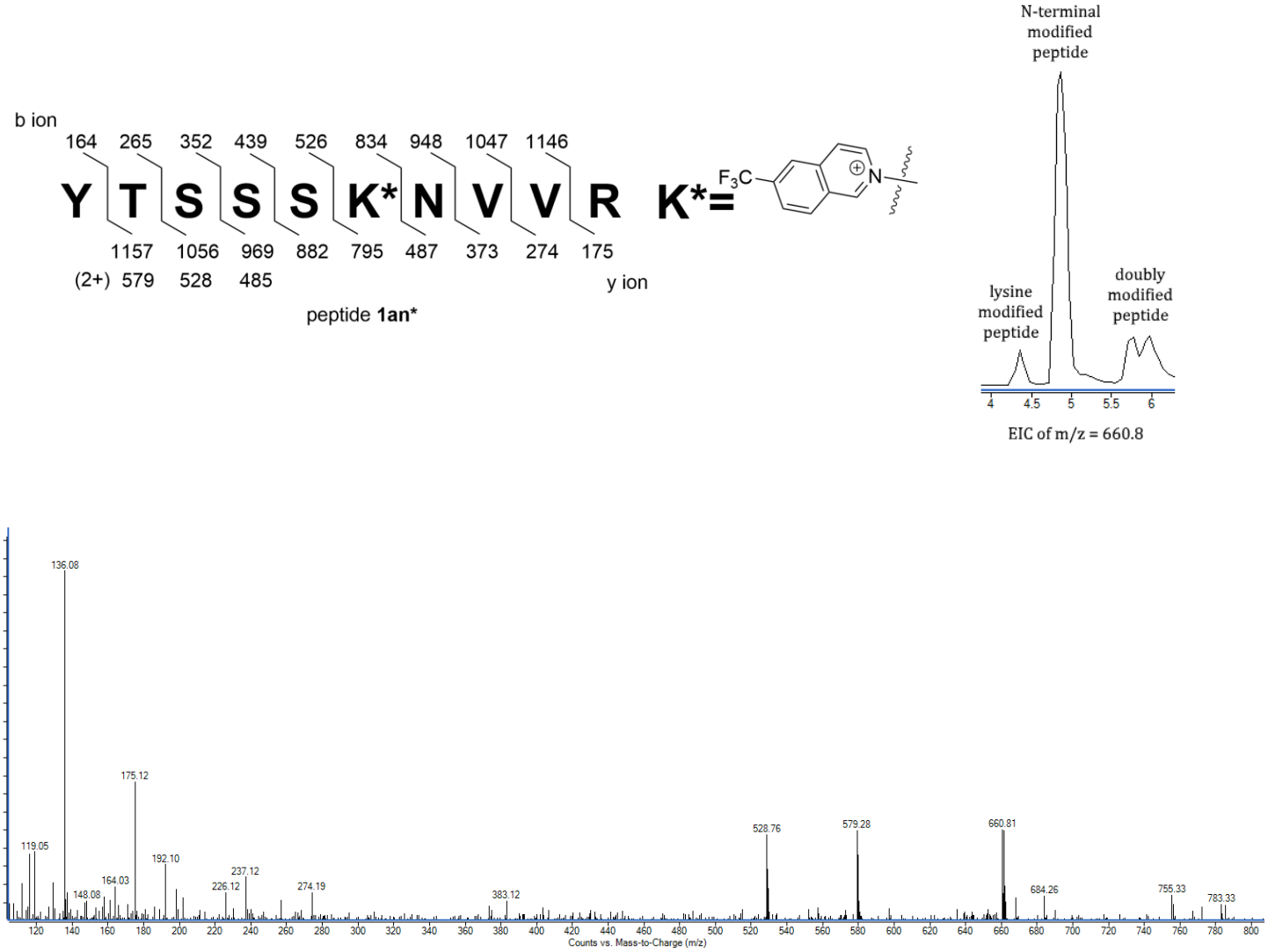


Figure S26. Q-TOF MS/MS spectrum of **Lysine 2n-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 660.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 660.8.

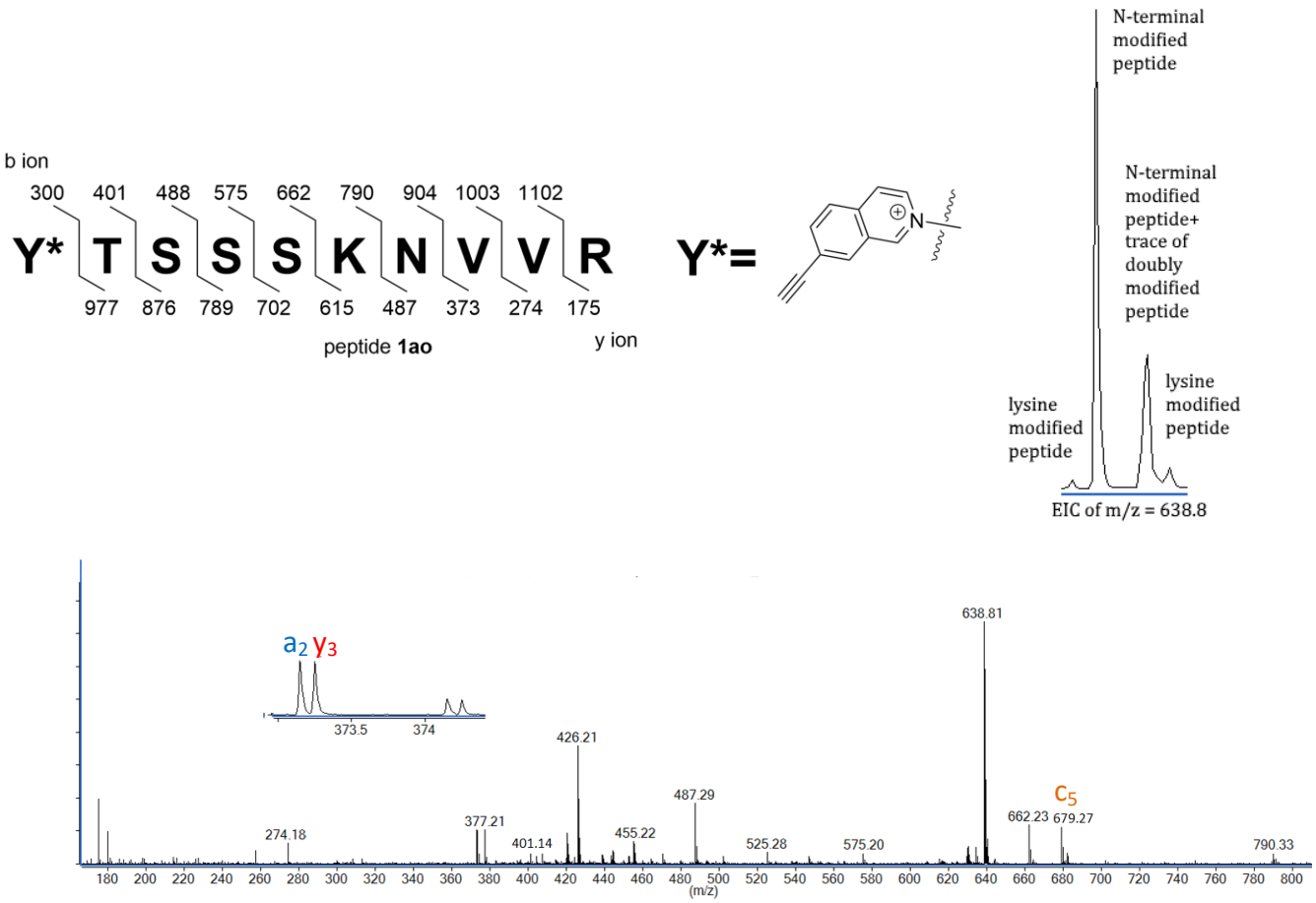


Figure S27. Q-TOF MS/MS spectrum of ***N*-terminal 2o-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 638.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 638.8.

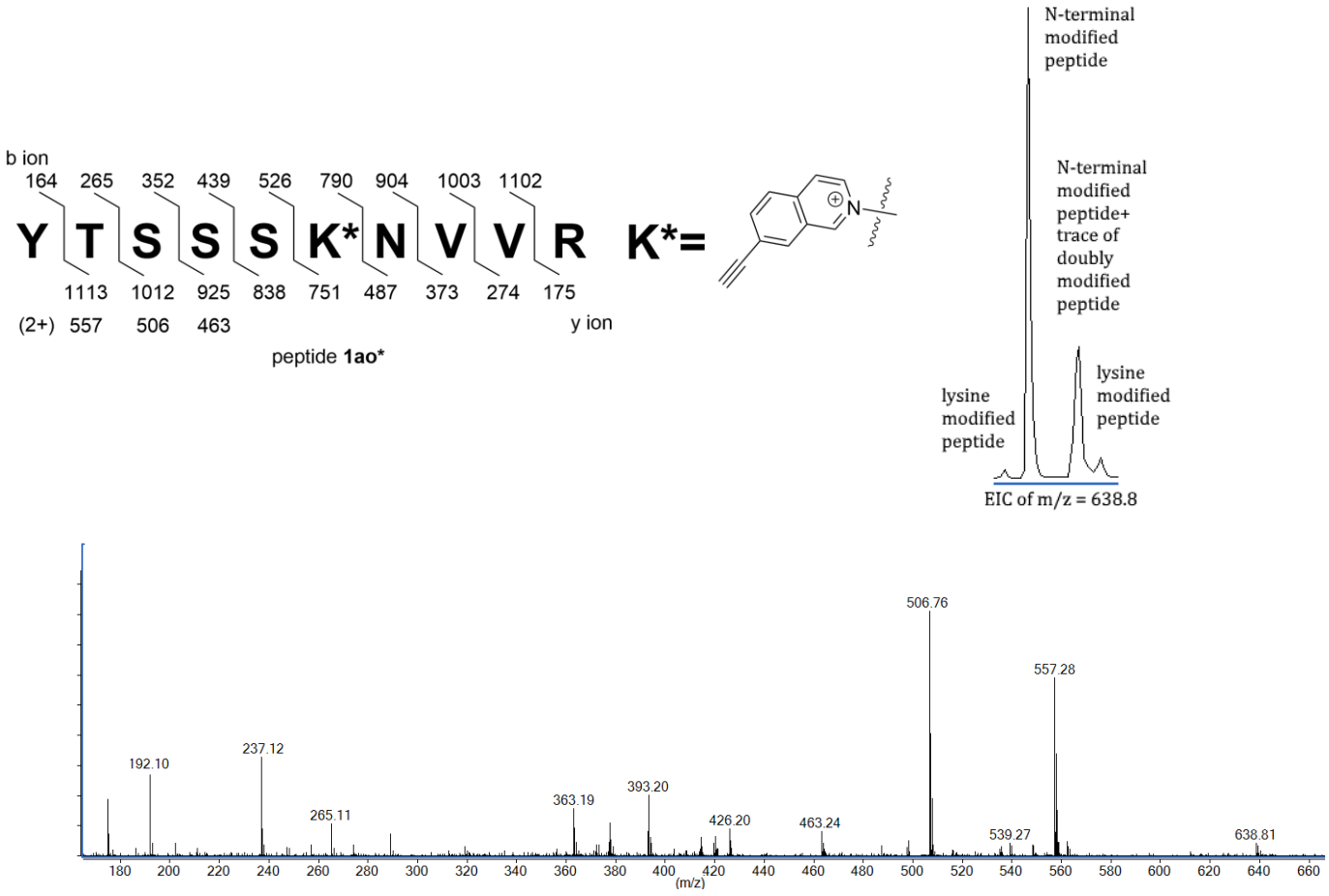


Figure S28. Q-TOF MS/MS spectrum of **Lysine 2o-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 638.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 638.8.

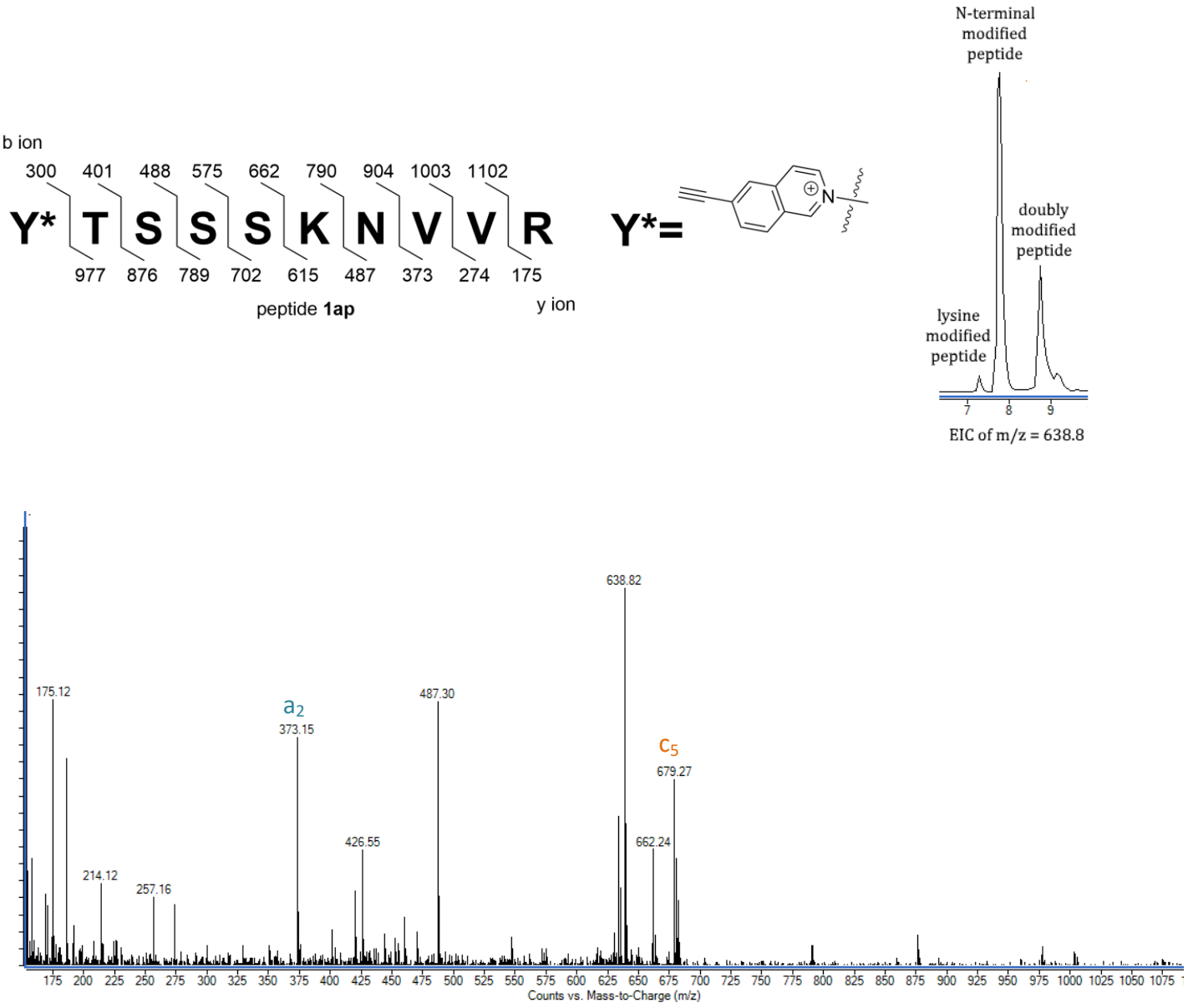


Figure S29. Q-TOF MS/MS spectrum of ***N*-terminal 2p-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 638.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 638.8.

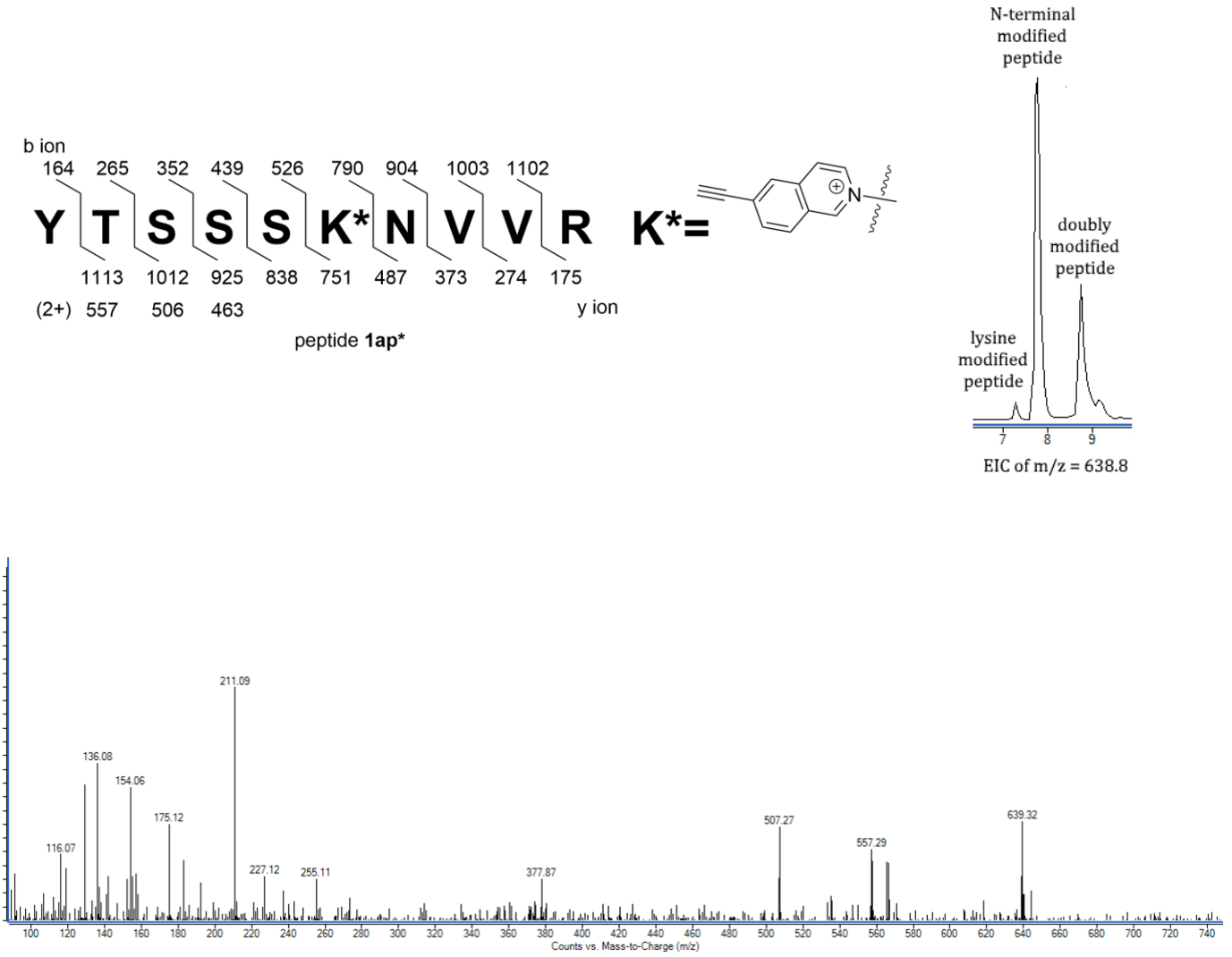


Figure S30. Q-TOF MS/MS spectrum of **Lysine 2p-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 638.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 638.8.

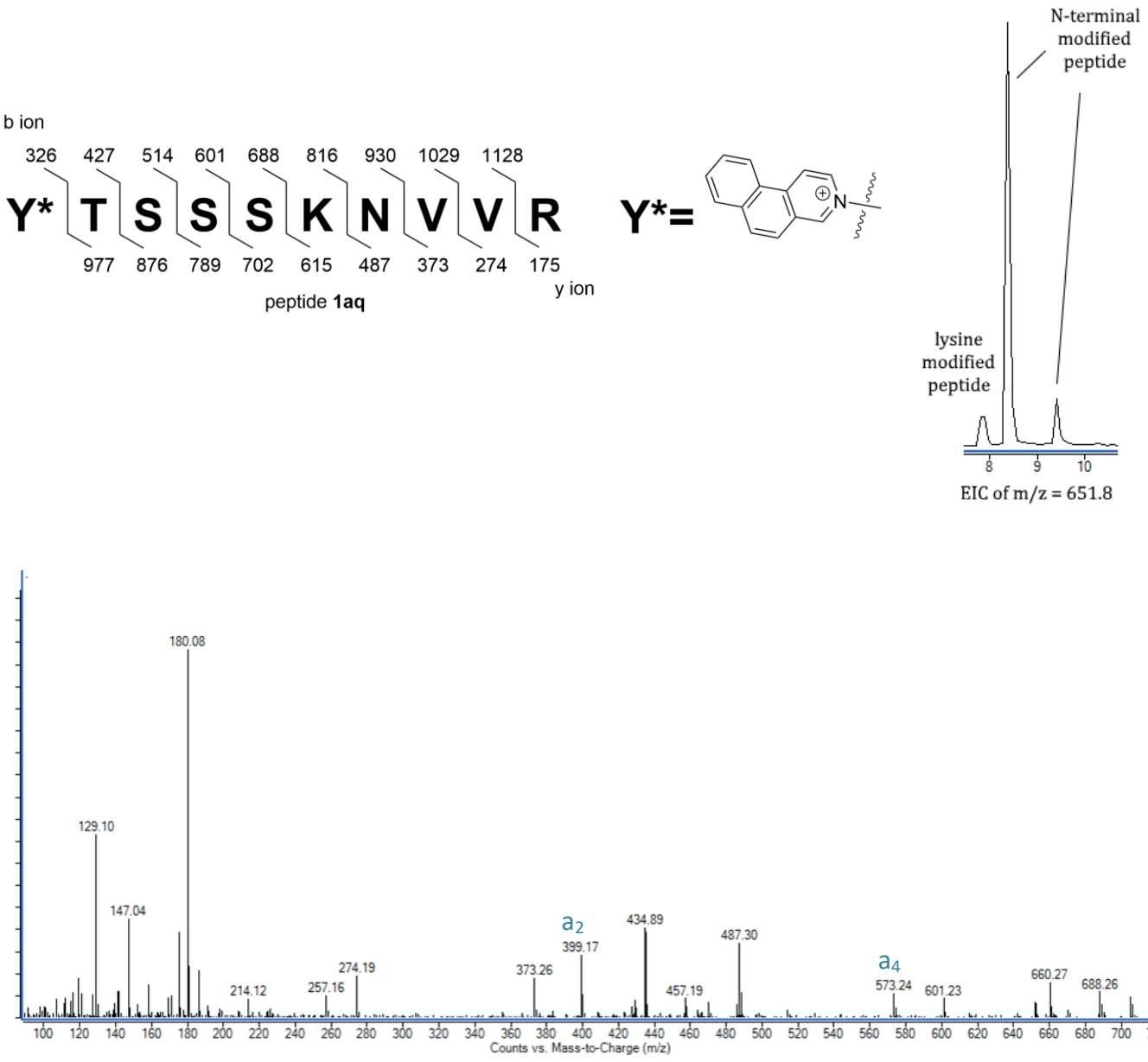


Figure S31. Q-TOF MS/MS spectrum of ***N*-terminal 2q-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 651.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 651.8.

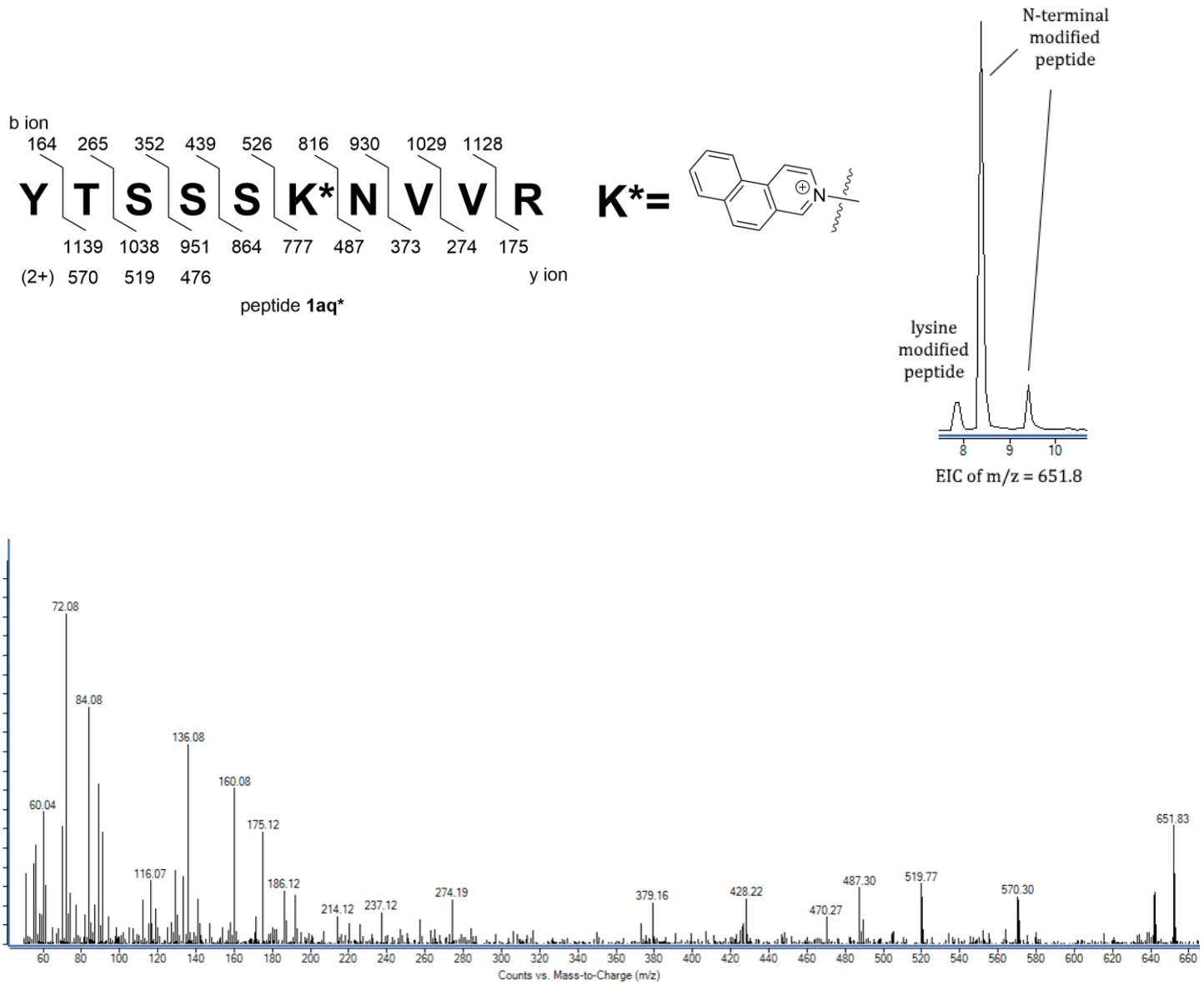


Figure S32. Q-TOF MS/MS spectrum of **Lysine 2q-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 651.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 651.8.

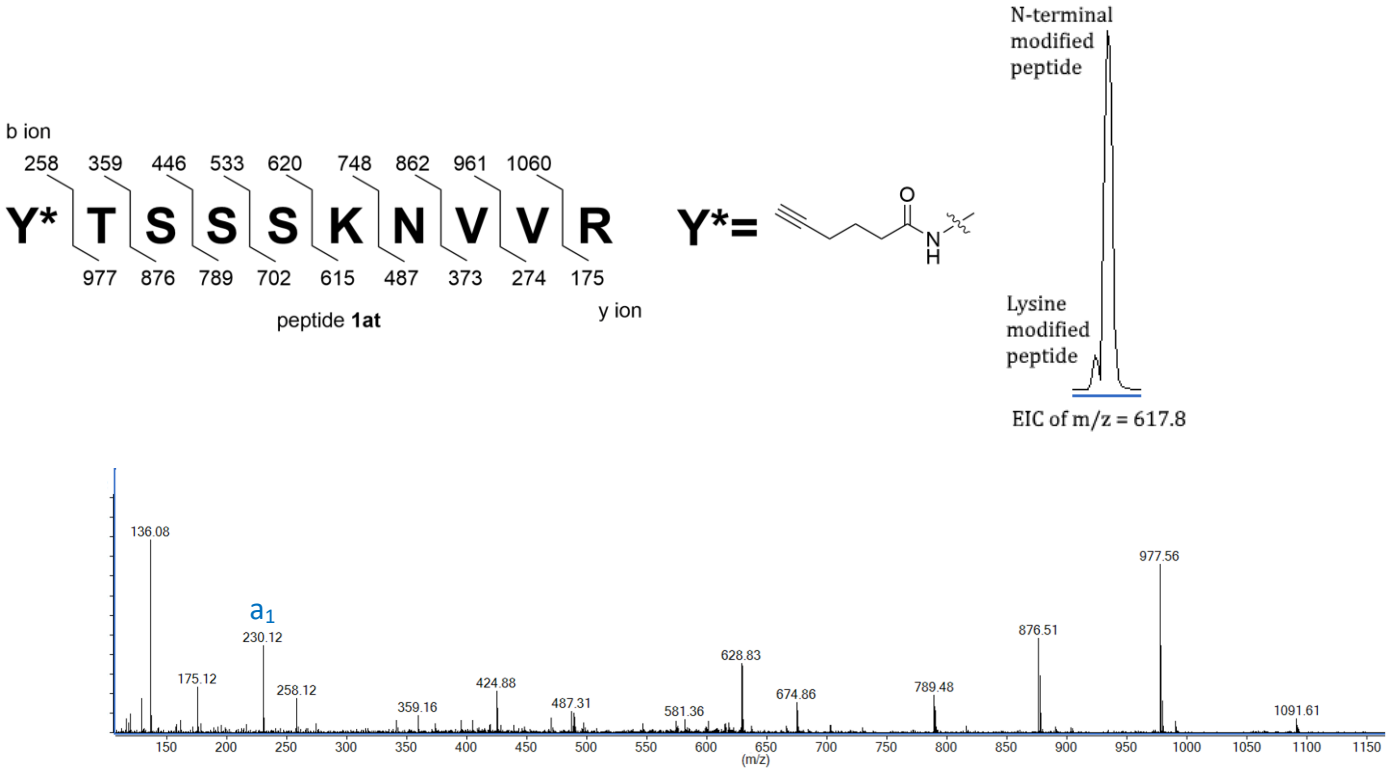


Figure S33. Q-TOF MS/MS spectrum of ***N*-terminal 2t NHS-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 617.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 617.8.

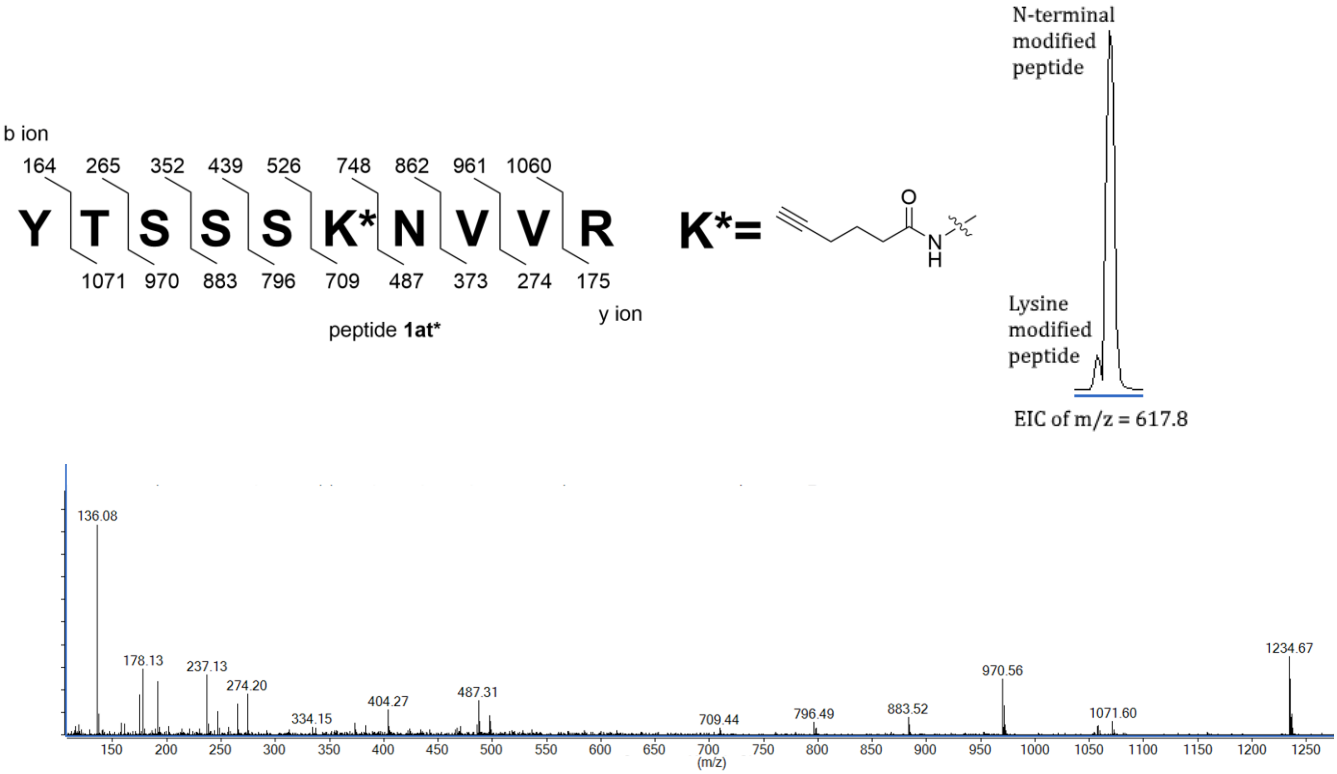


Figure S34. Q-TOF MS/MS spectrum of **Lysine 2t NHS-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 617.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 617.8.

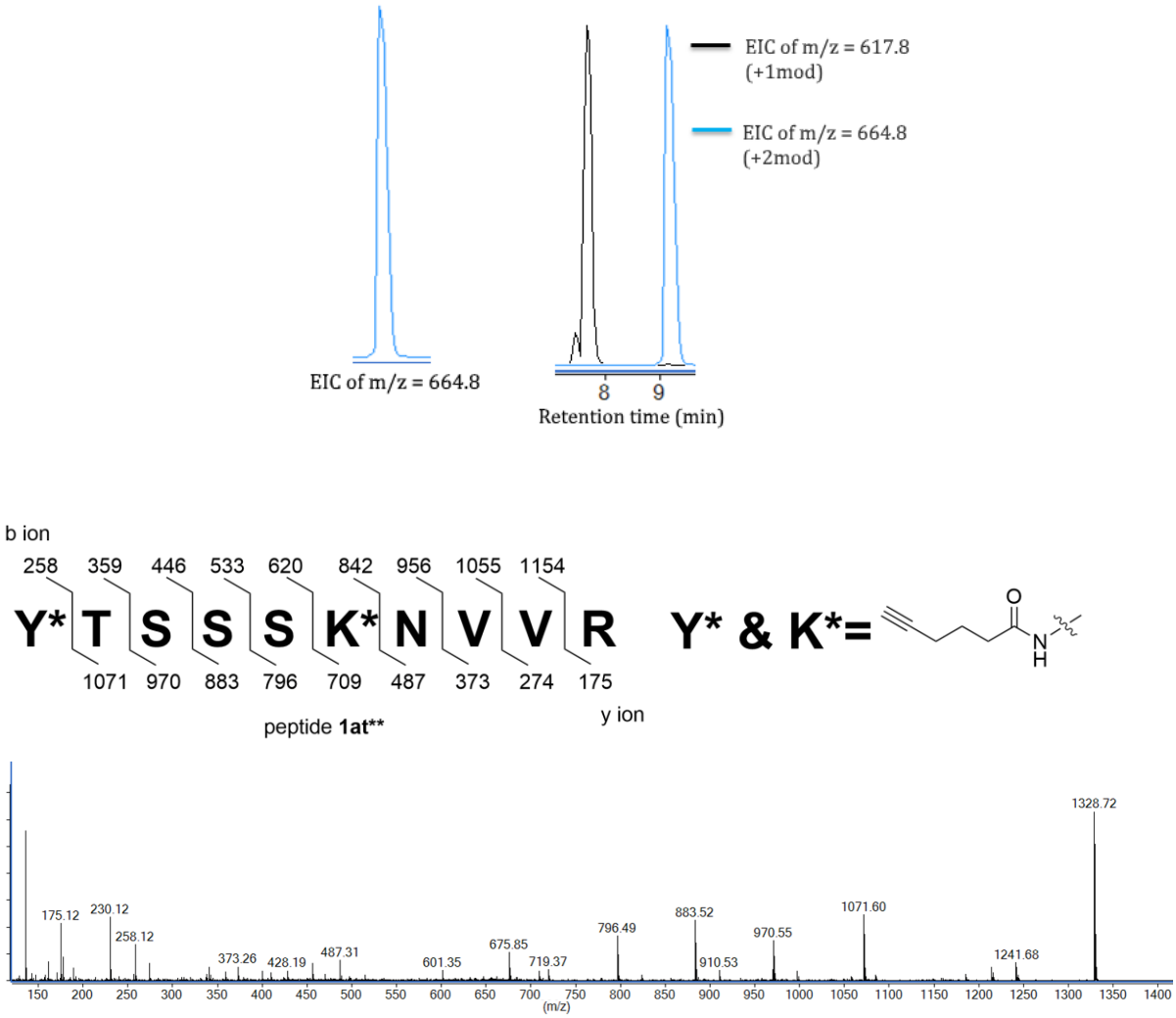
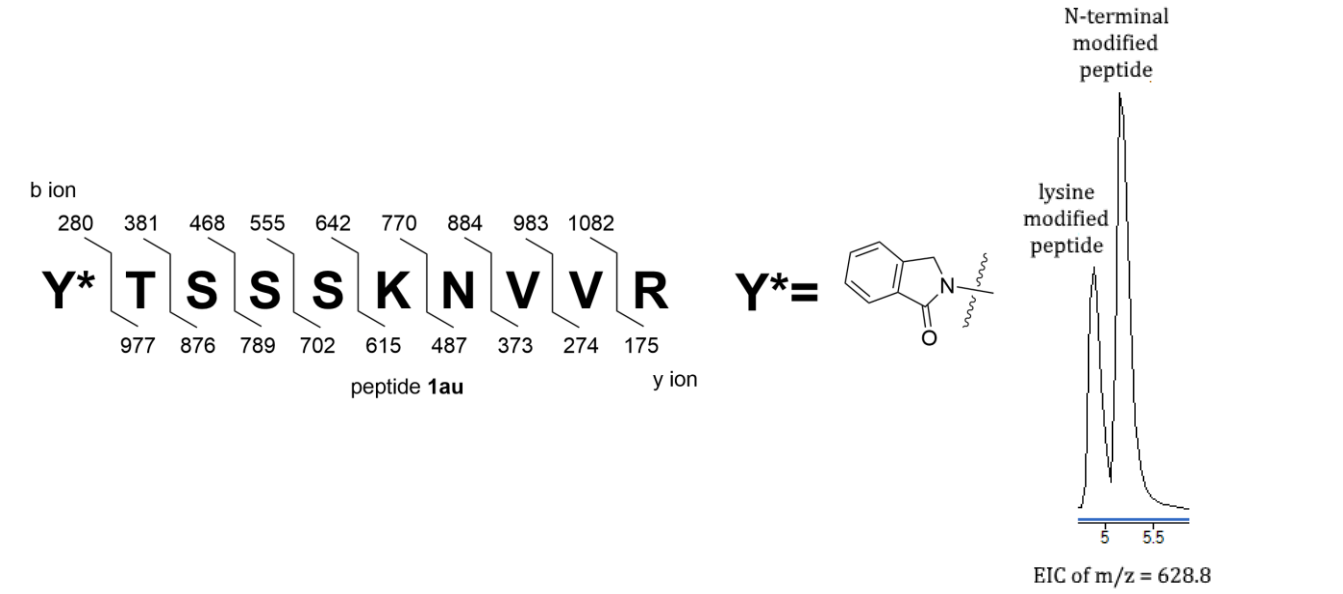


Figure S35. Q-TOF MS/MS spectrum of ***N*-terminal and Lysine 2t NHS-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 664.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 664.8.



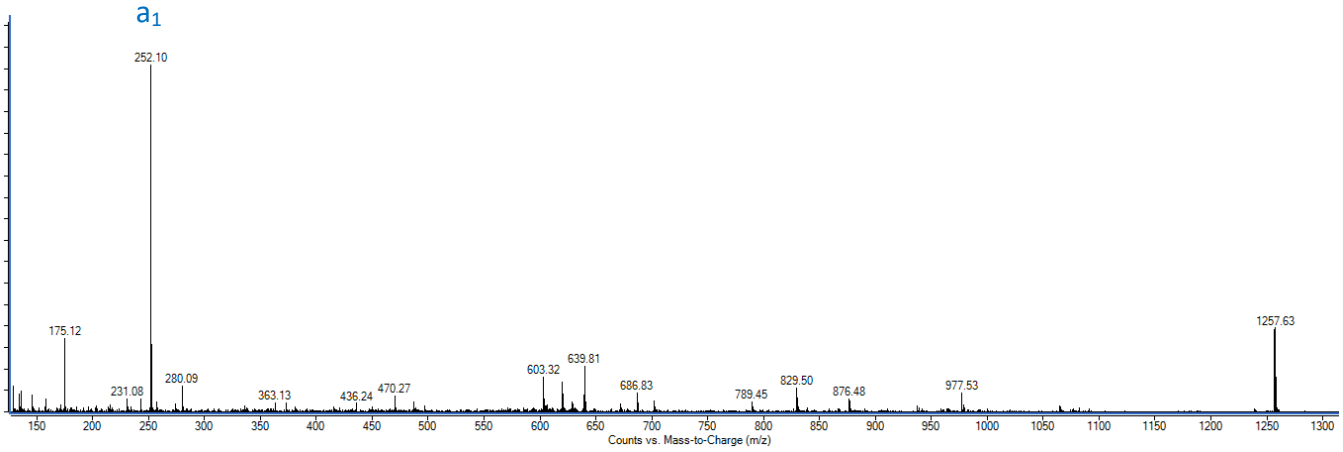


Figure S36. Q-TOF MS/MS spectrum of ***N*-terminal 2u-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 628.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 628.8.

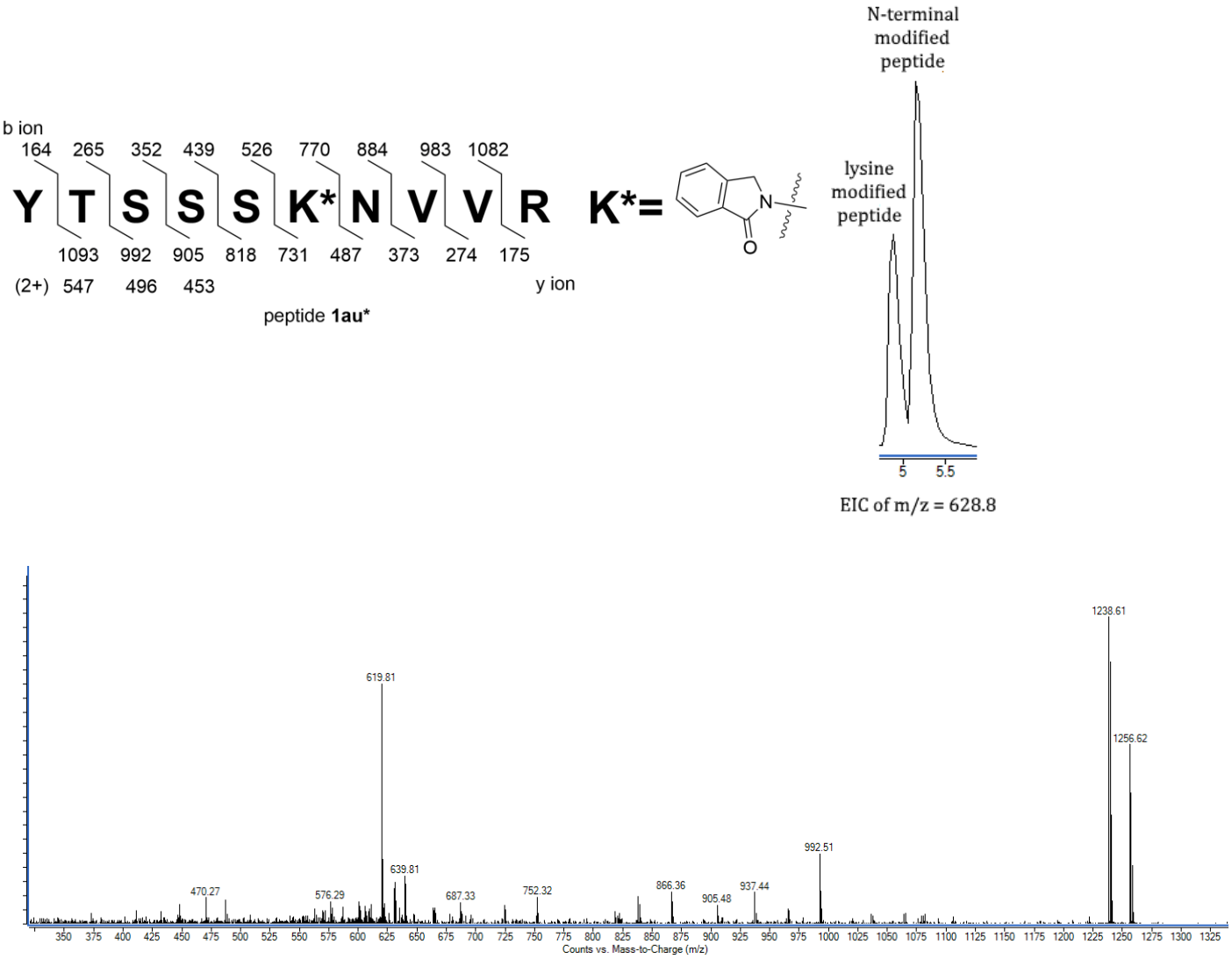


Figure S37. Q-TOF MS/MS spectrum of **Lysine 2u-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 628.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 628.8.

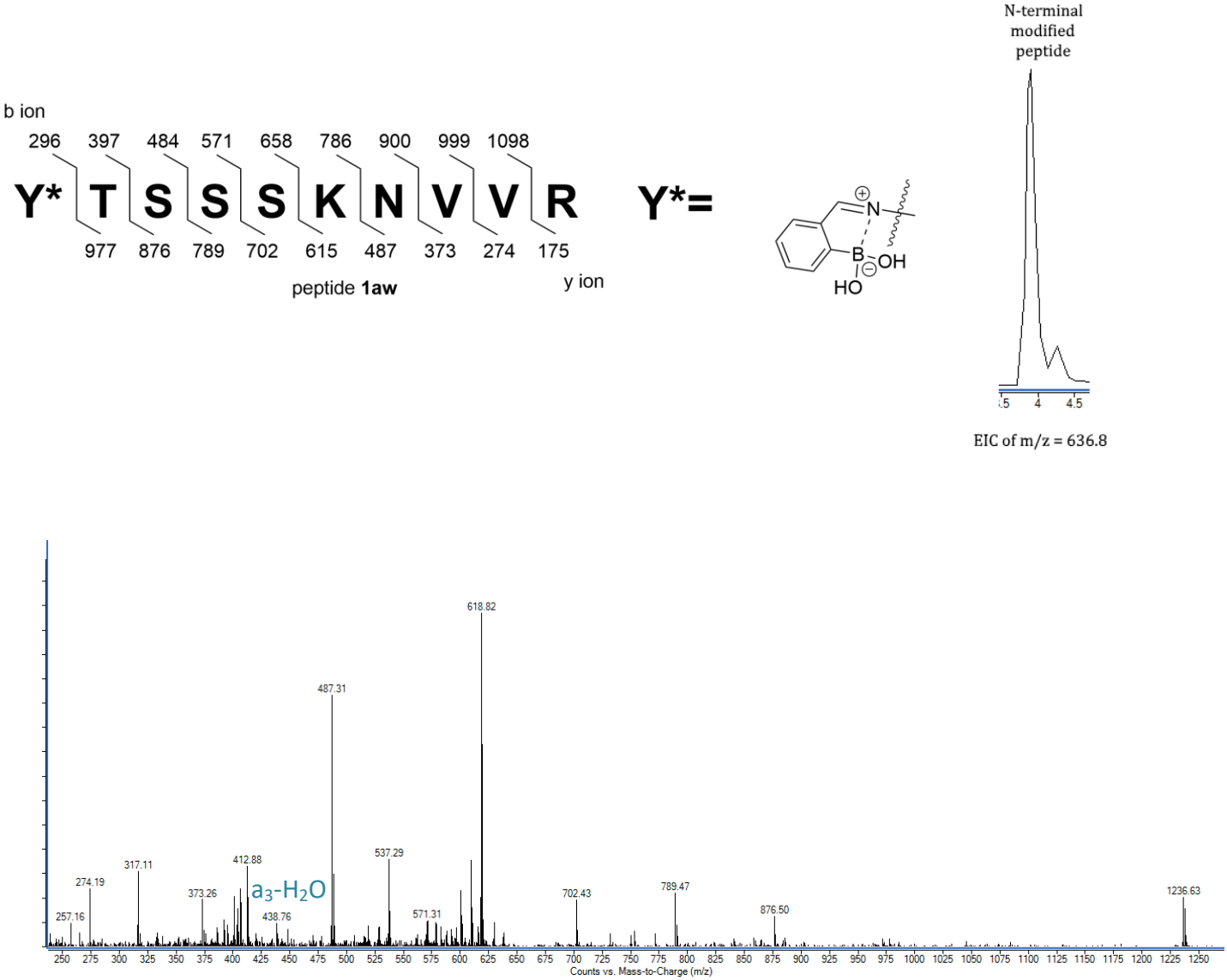


Figure S38. Q-TOF MS/MS spectrum of ***N*-terminal 2w-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 636.8). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 636.8.

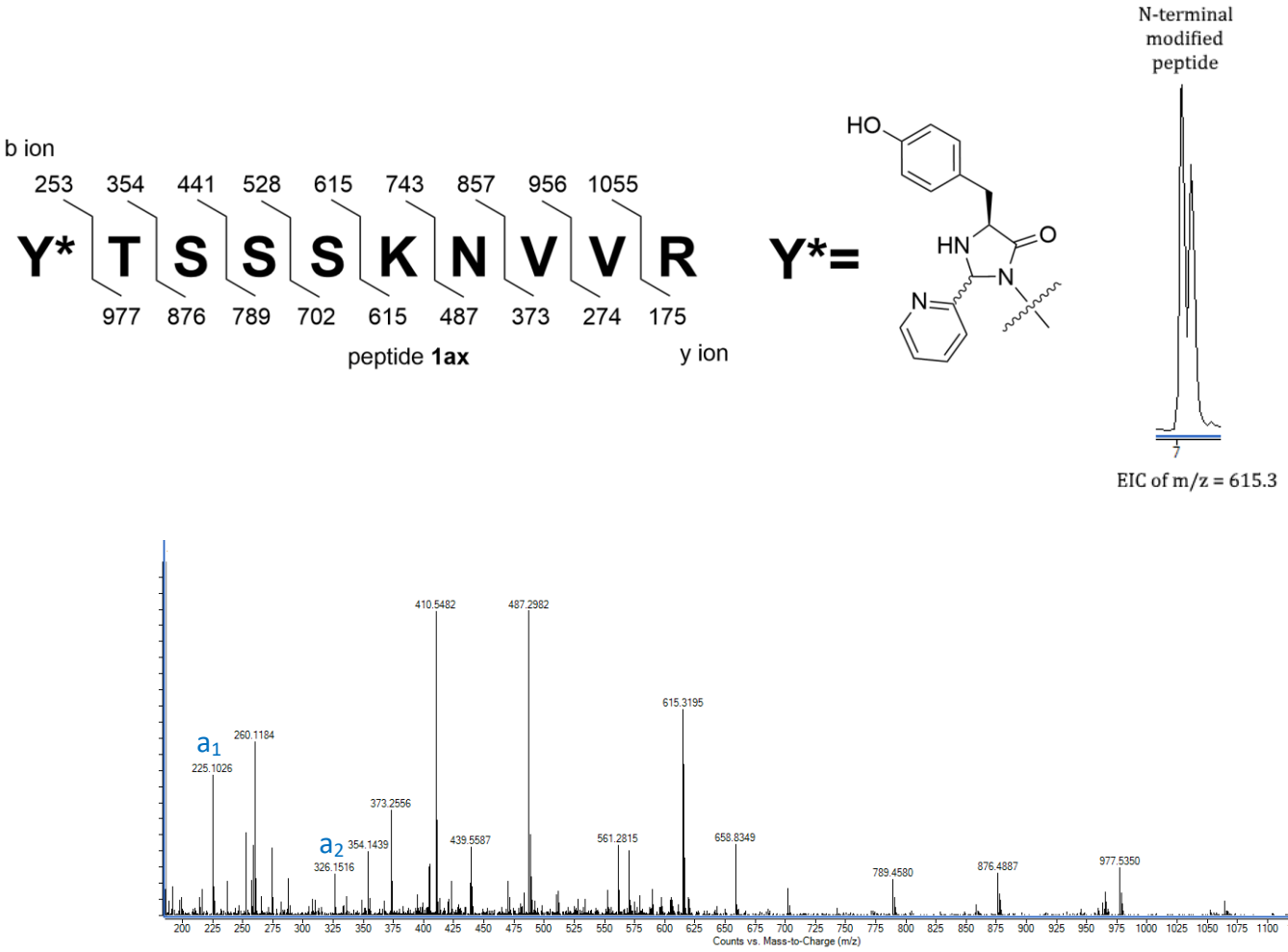


Figure S39. Q-TOF MS/MS spectrum of ***N*-terminal 2x-modified YTSSSKNVVR** (ESI source, doubly charged ion of m/z = 615.3). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 615.3.

Modification of Peptide XSKFR using 2-ethynyl- 5-hydroxy-4-methoxybenzaldehyde 2b

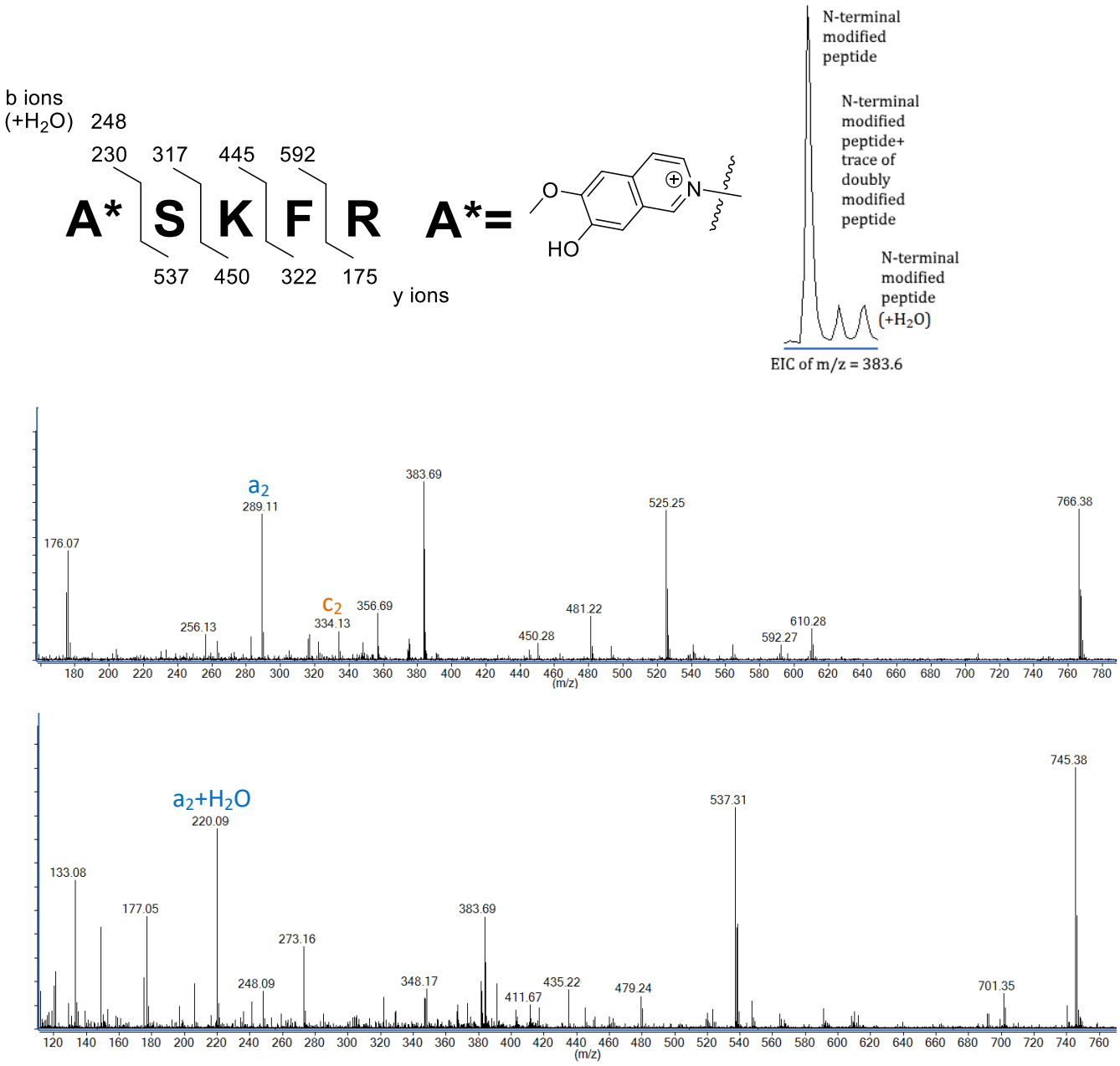


Figure S40. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified ASKFR (Top); *N*-terminal 2b-modified ASKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 383.6). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 383.6.

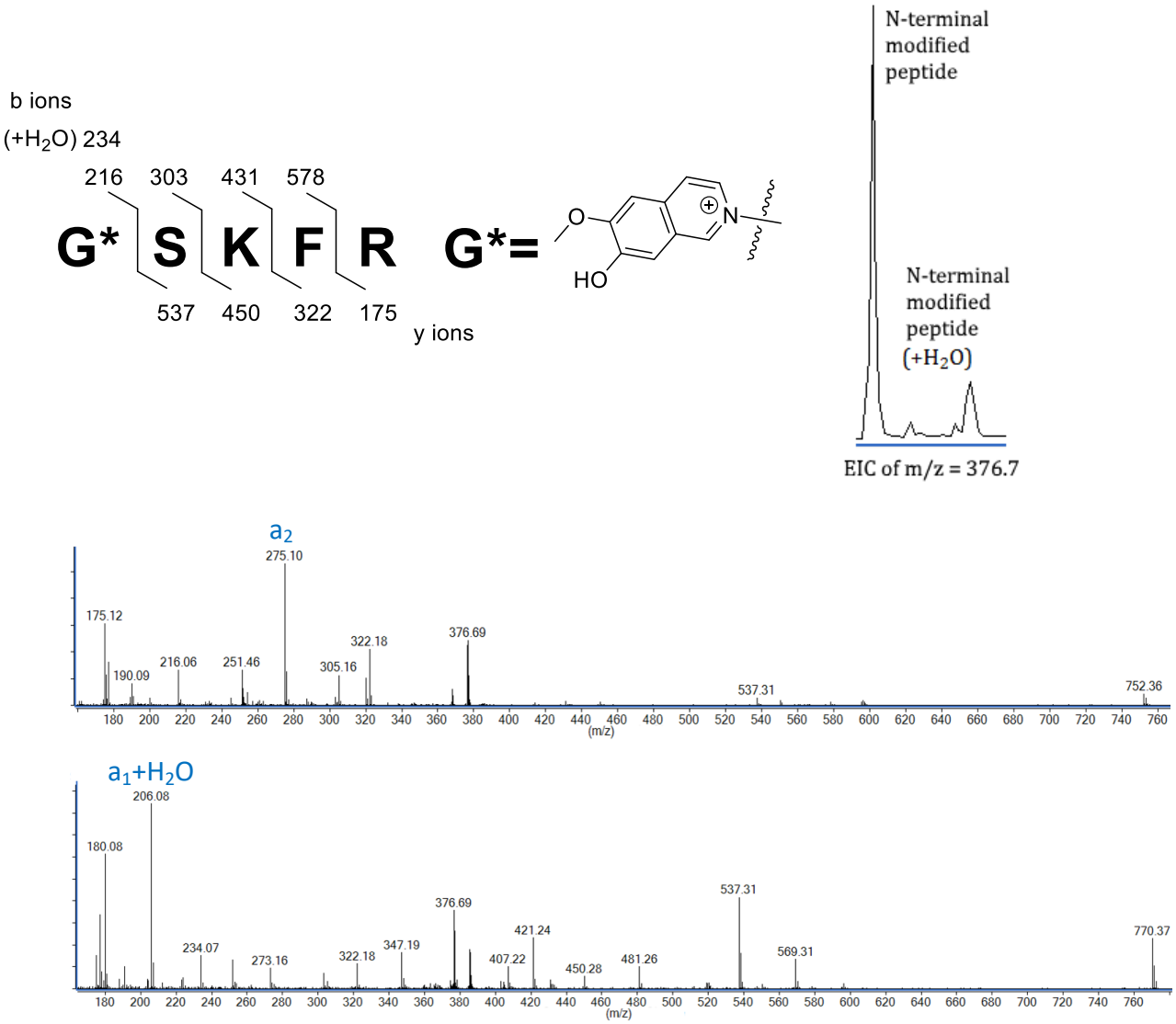


Figure S41. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified GSKFR (Top); *N*-terminal 2b-modified GSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 376.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 376.7.

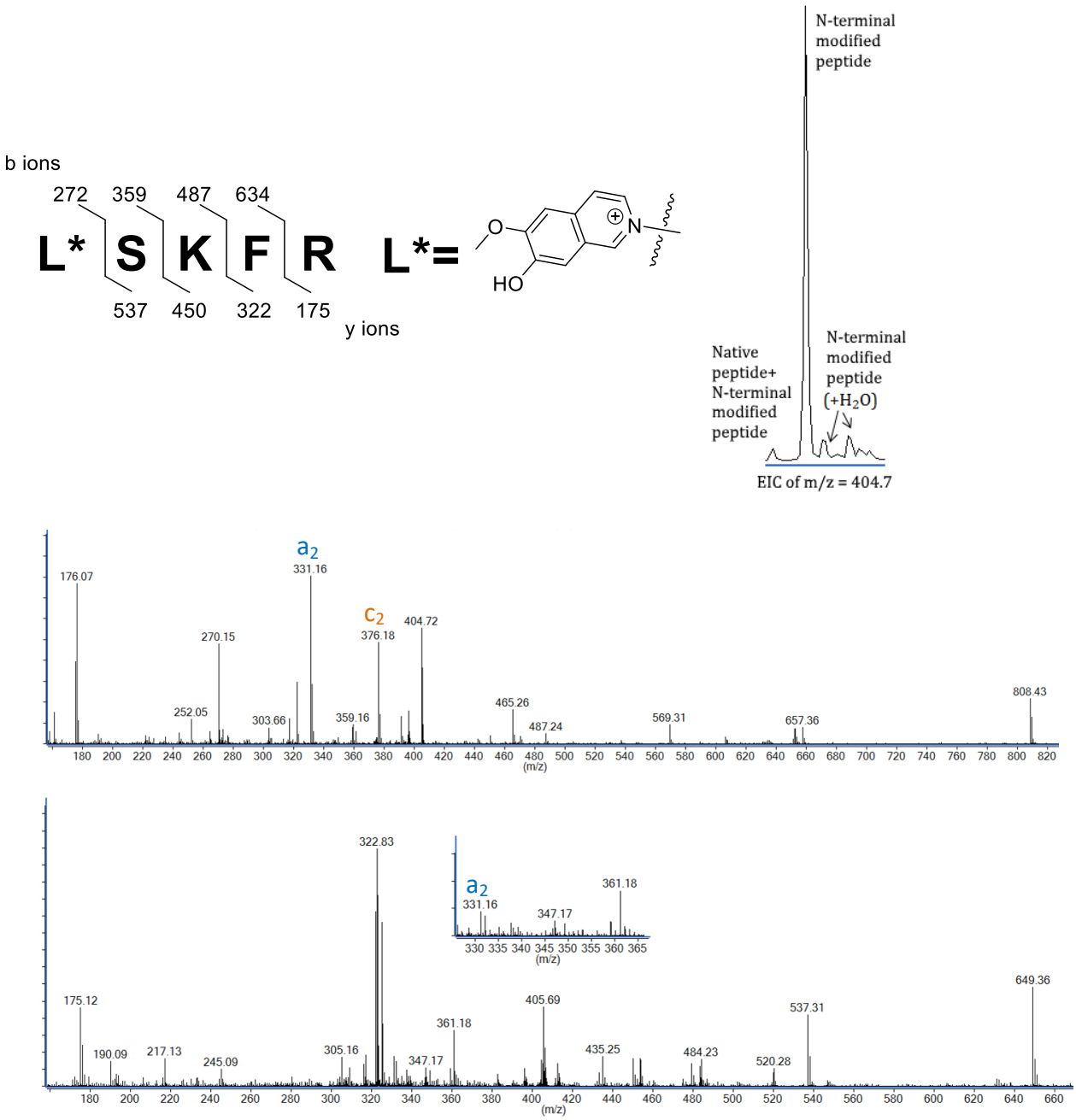


Figure S42. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified LSKFR (Top); *N*-terminal 2b-modified LSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 404.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 404.7.

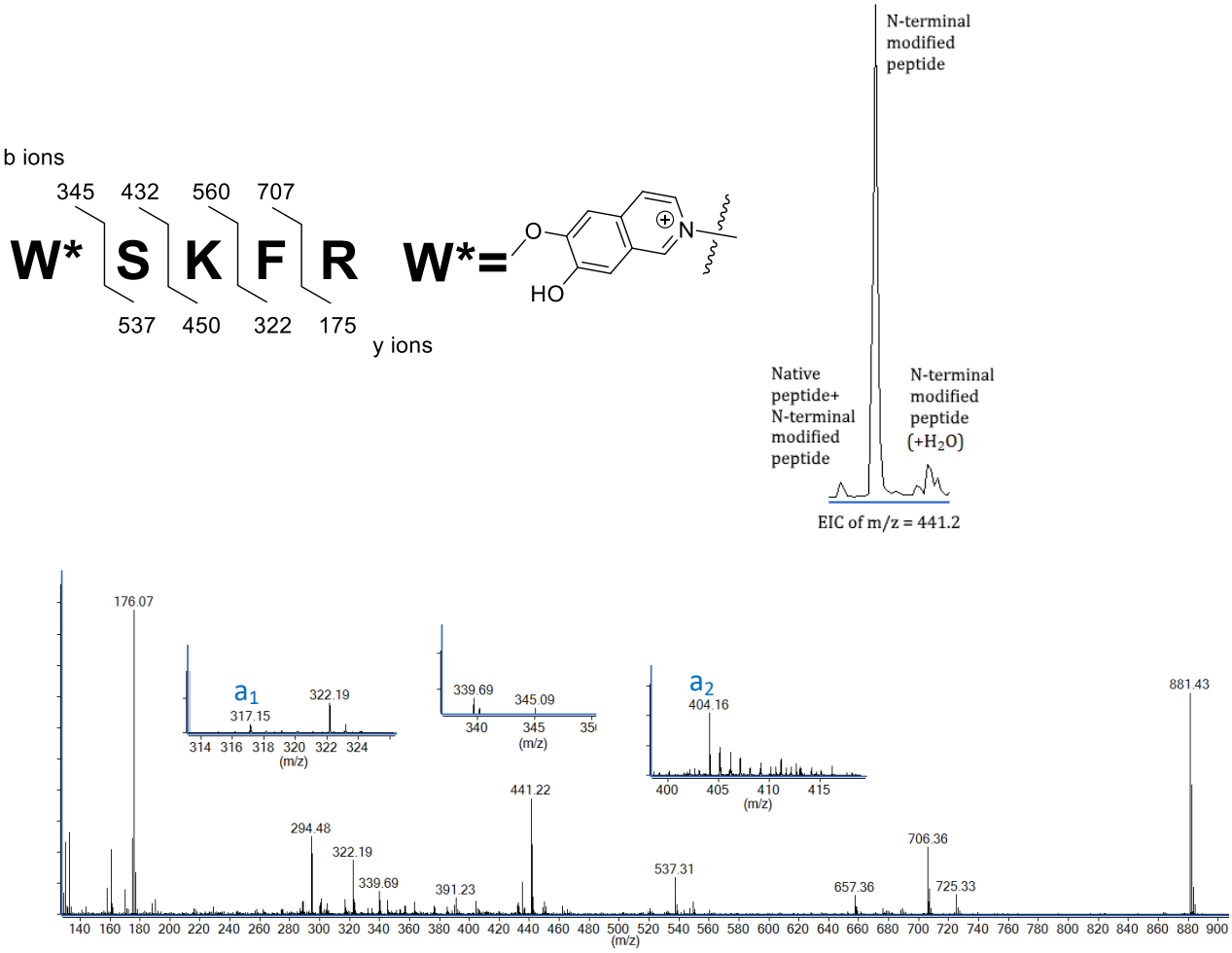


Figure S43. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified WSKFR** (ESI source, doubly charged ion of m/z = 441.2). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 441.2.

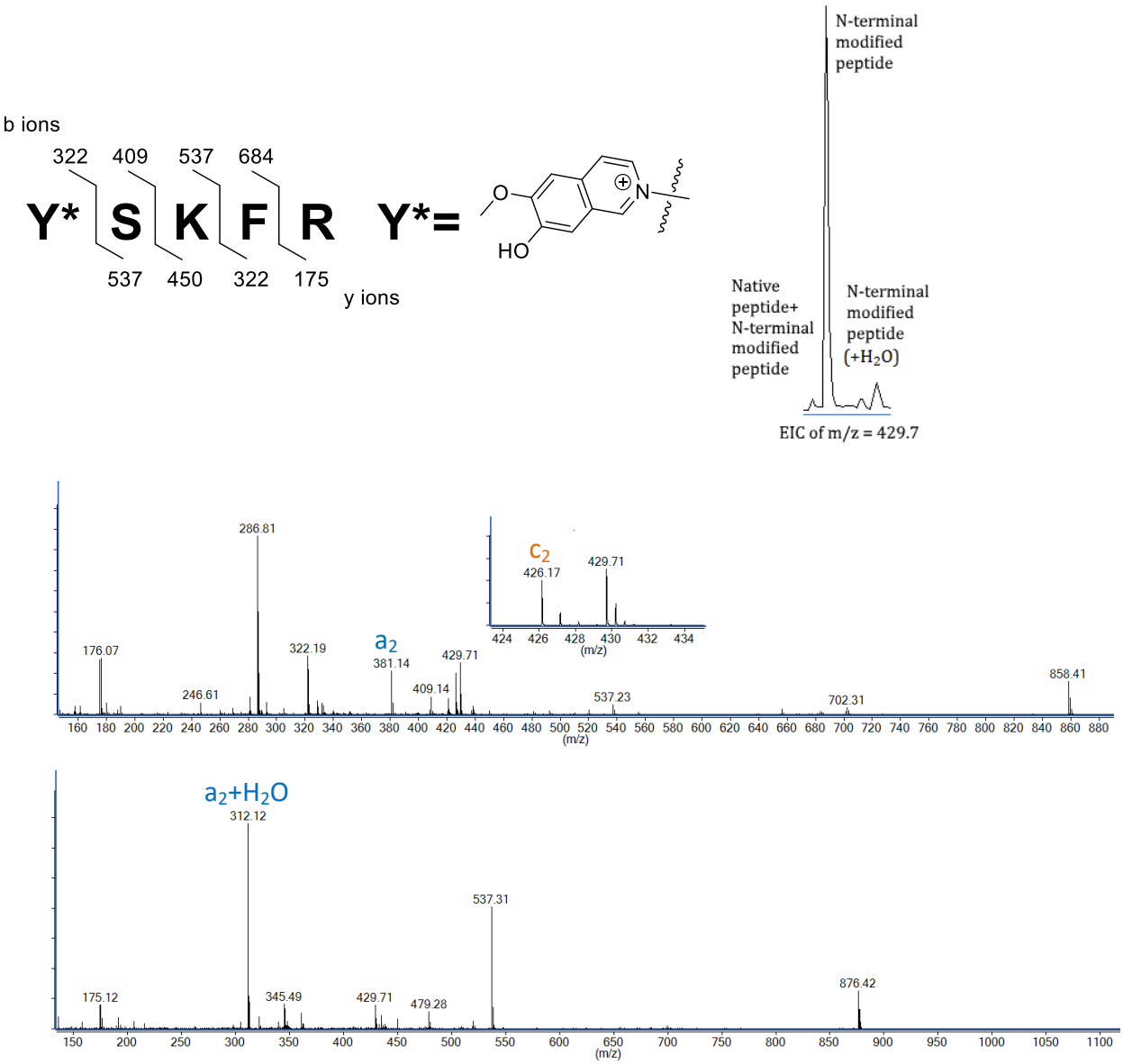


Figure S44. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified YSKFR (Top); *N*-terminal 2b-modified YSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 429.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 429.7.

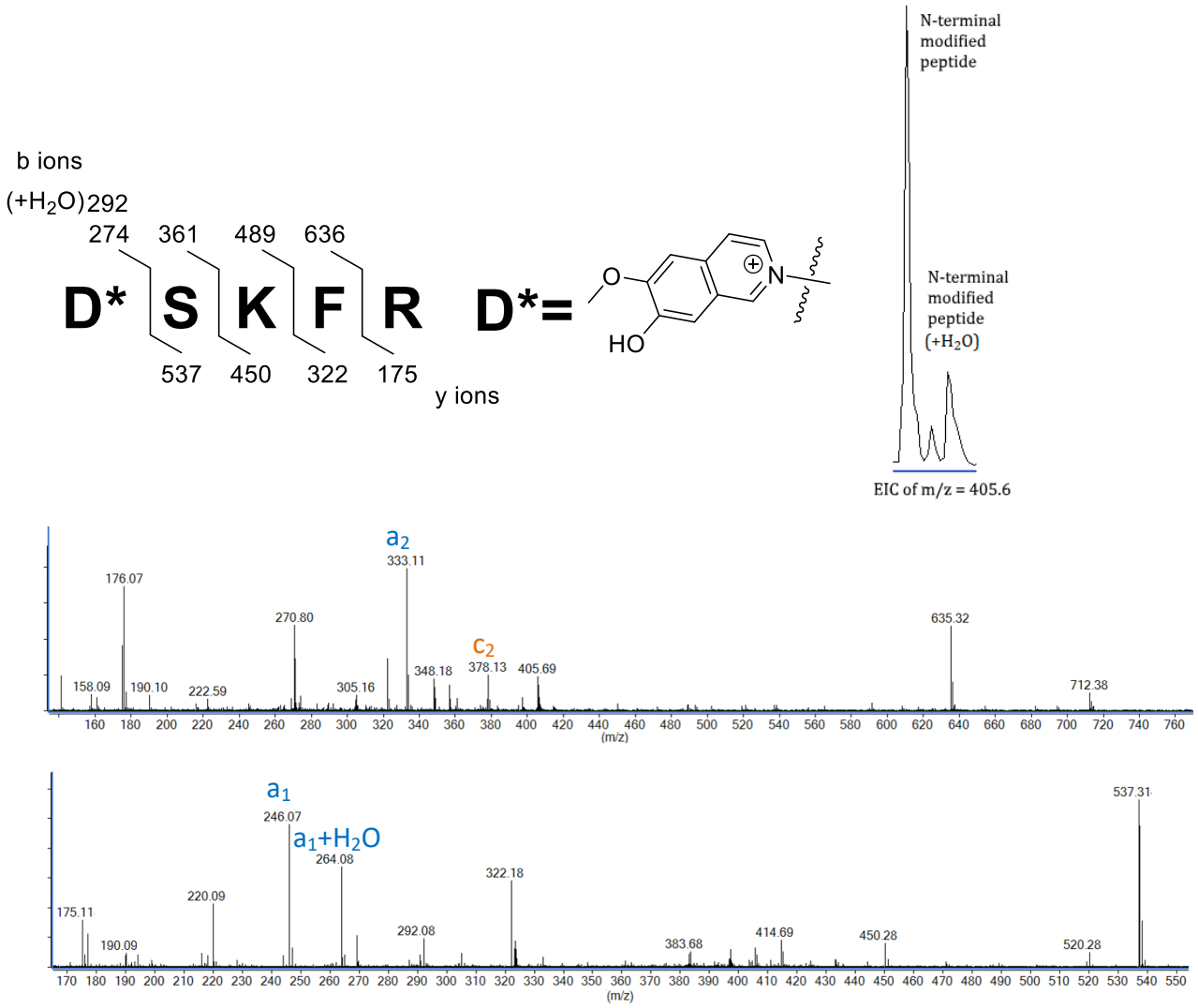


Figure S45. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified DSKFR (Top); *N*-terminal 2b-modified DSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 405.6). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 405.6.

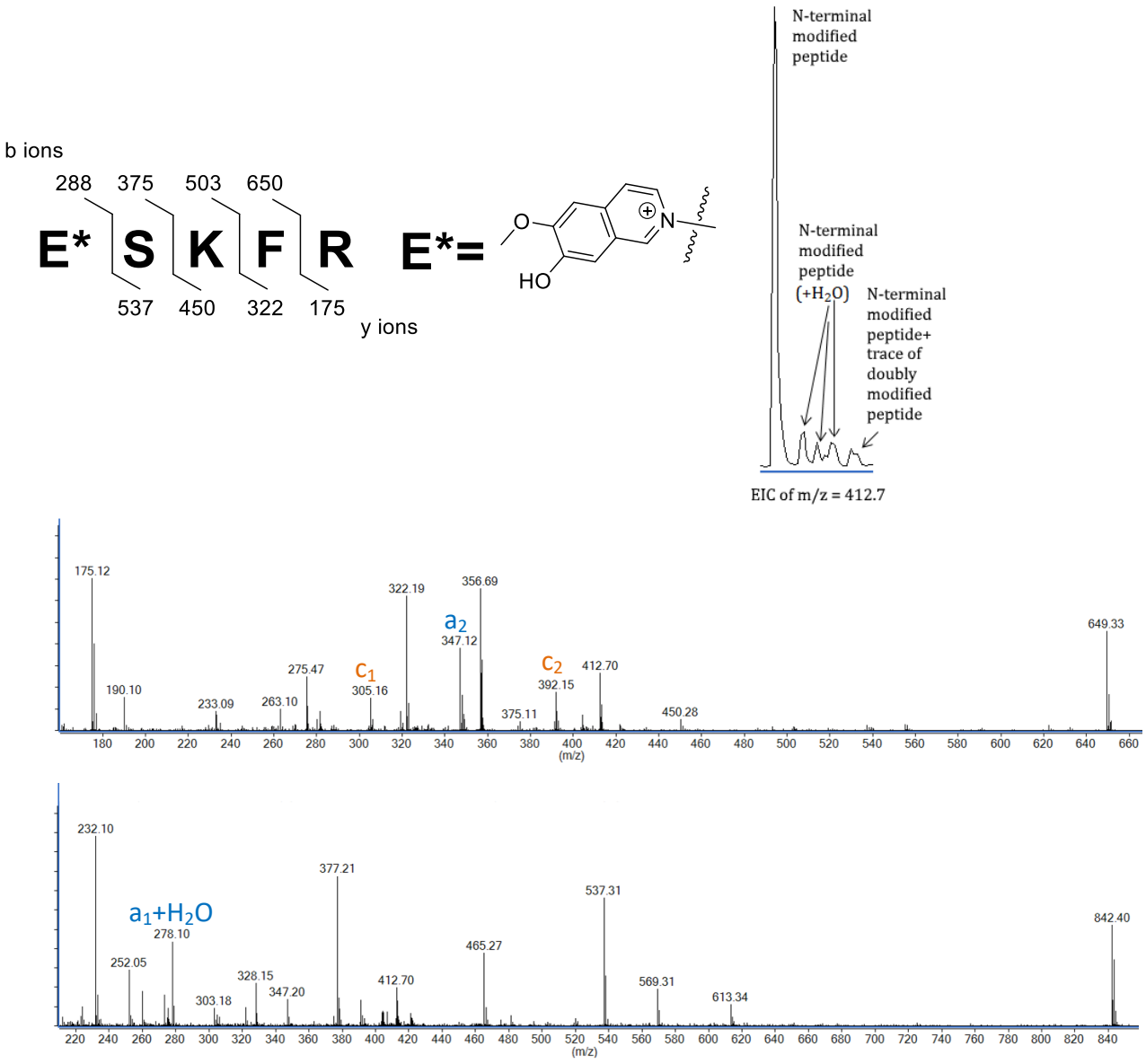


Figure S46. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified ESKFR (Top); *N*-terminal 2b-modified ESKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 412.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 412.7.

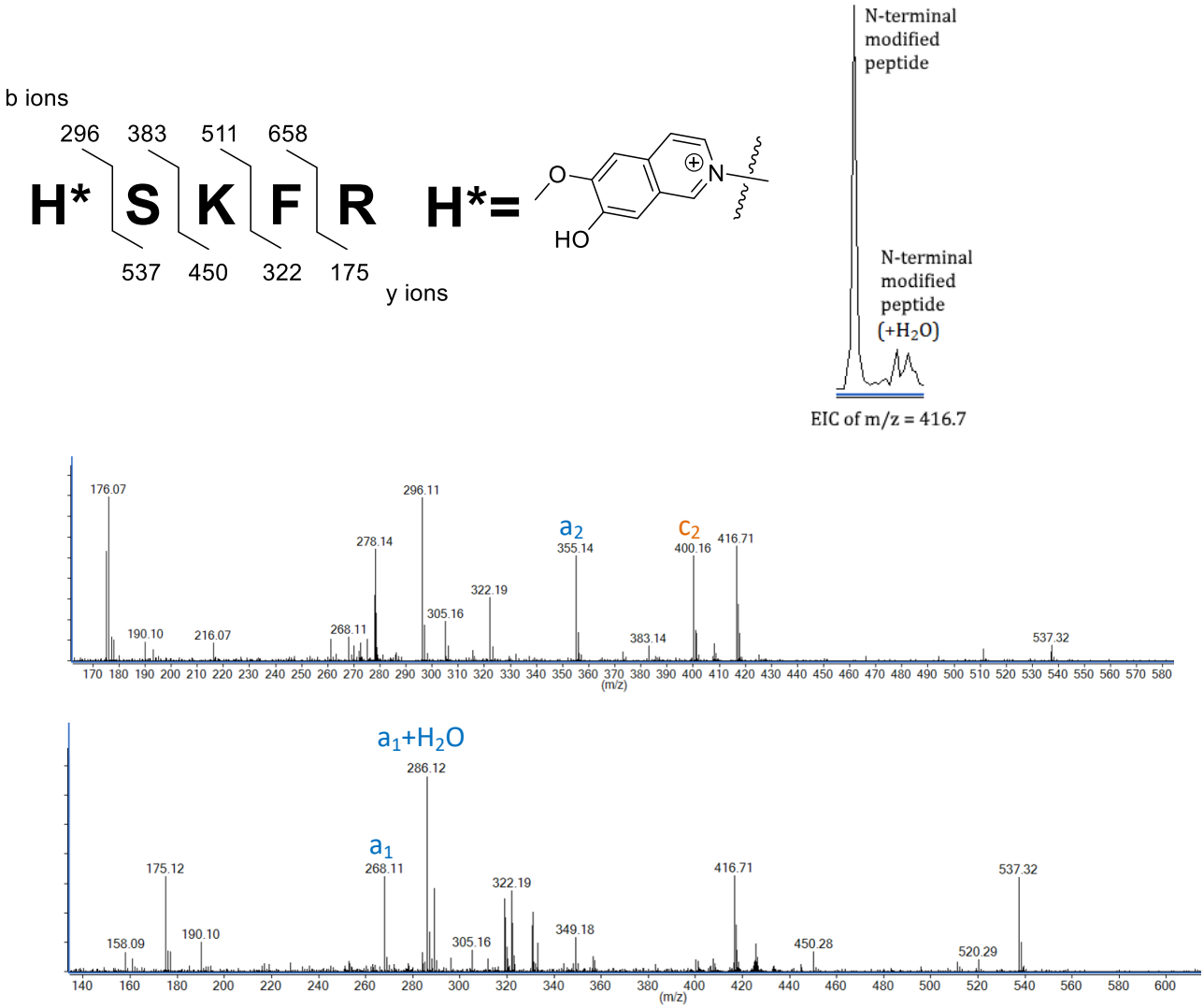


Figure S47. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified HSKFR (Top); *N*-terminal 2b-modified HSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 416.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 416.7.

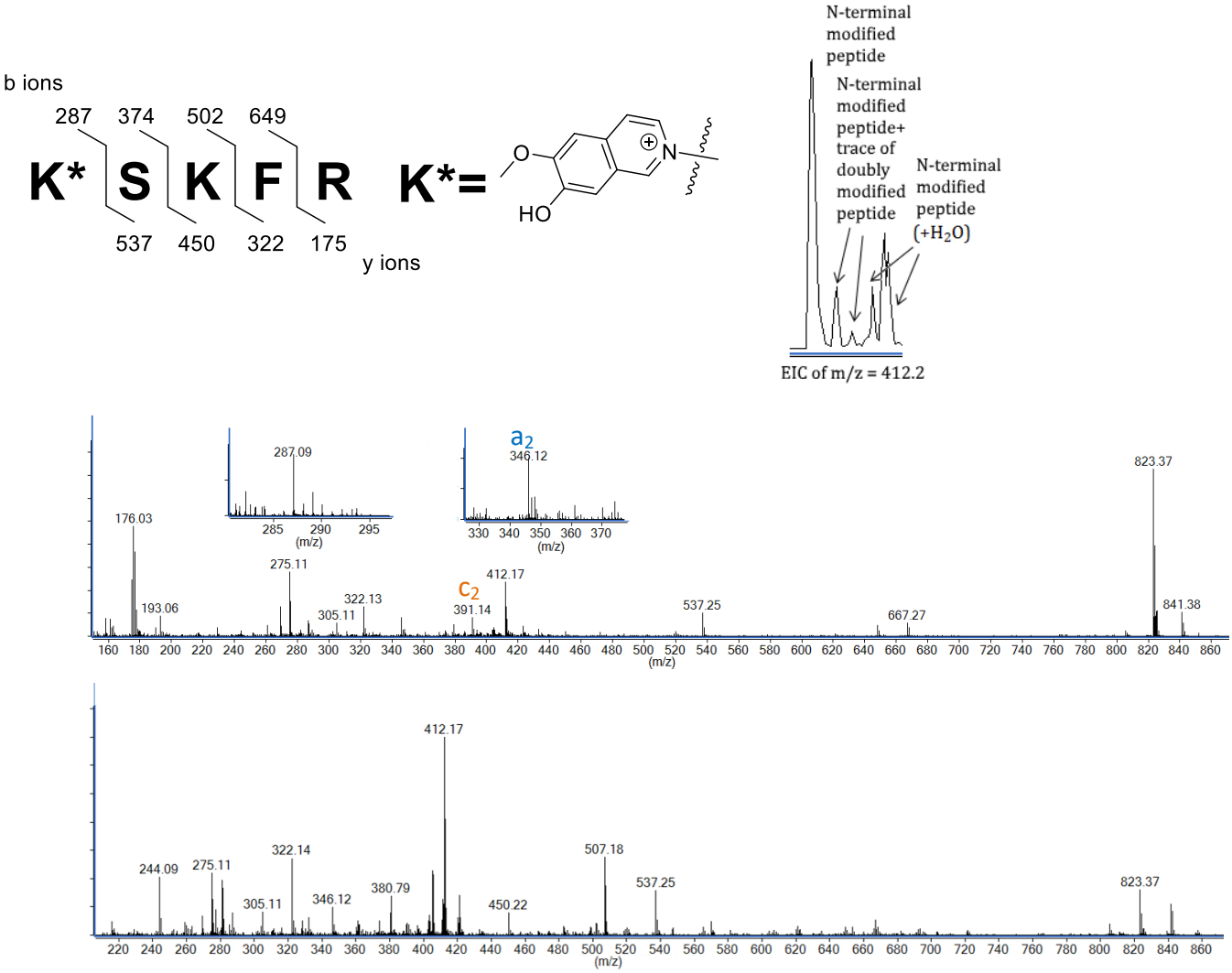


Figure S48. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified KSKFR (Top); *N*-terminal 2b-modified KSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 412.2). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 412.2.

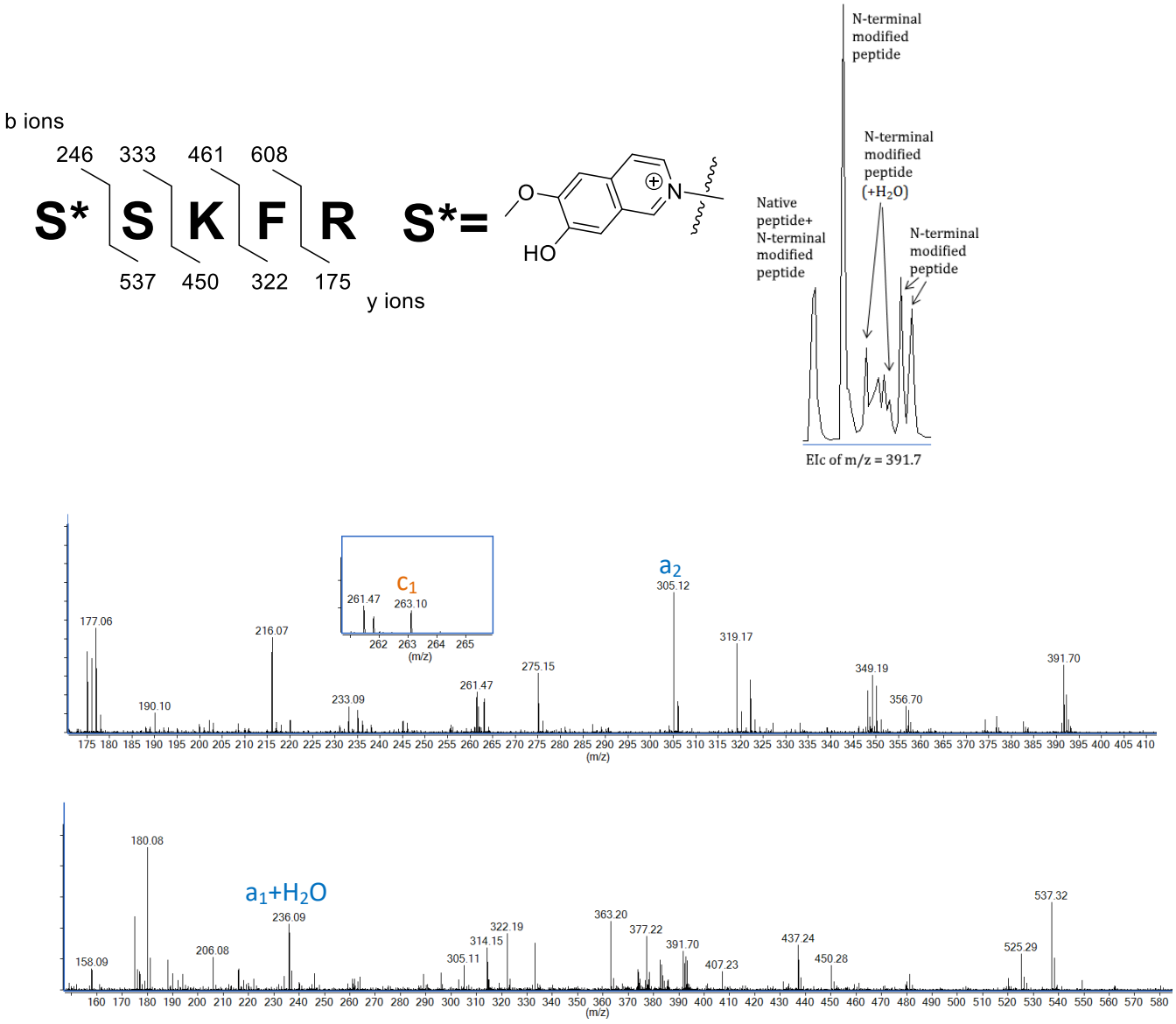


Figure S49. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified SSKFR (Top); *N*-terminal 2b-modified SSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 391.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 391.7.

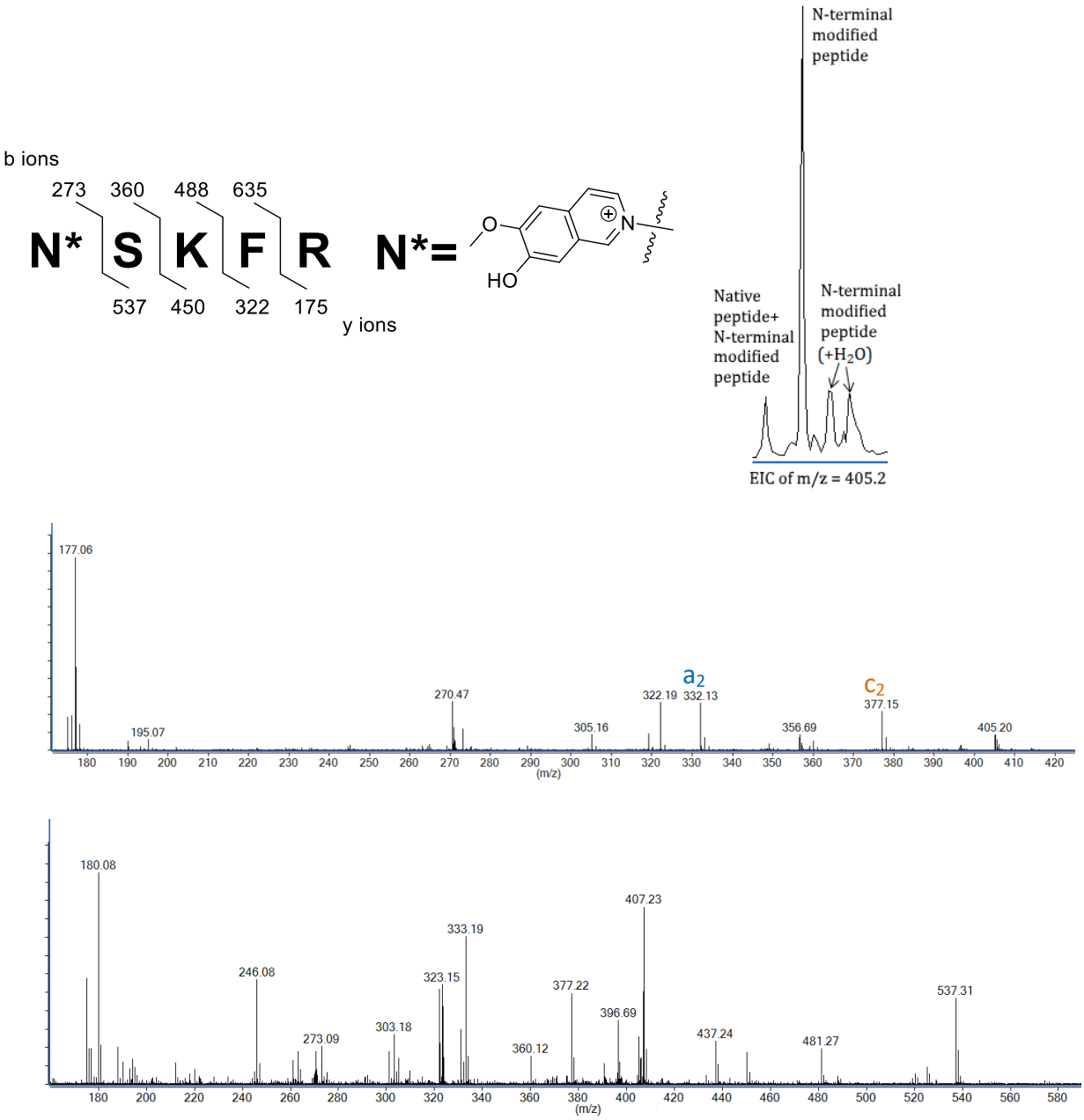


Figure S50. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified NSKFR (Top); *N*-terminal 2b-modified NSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 405.2). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 405.2.

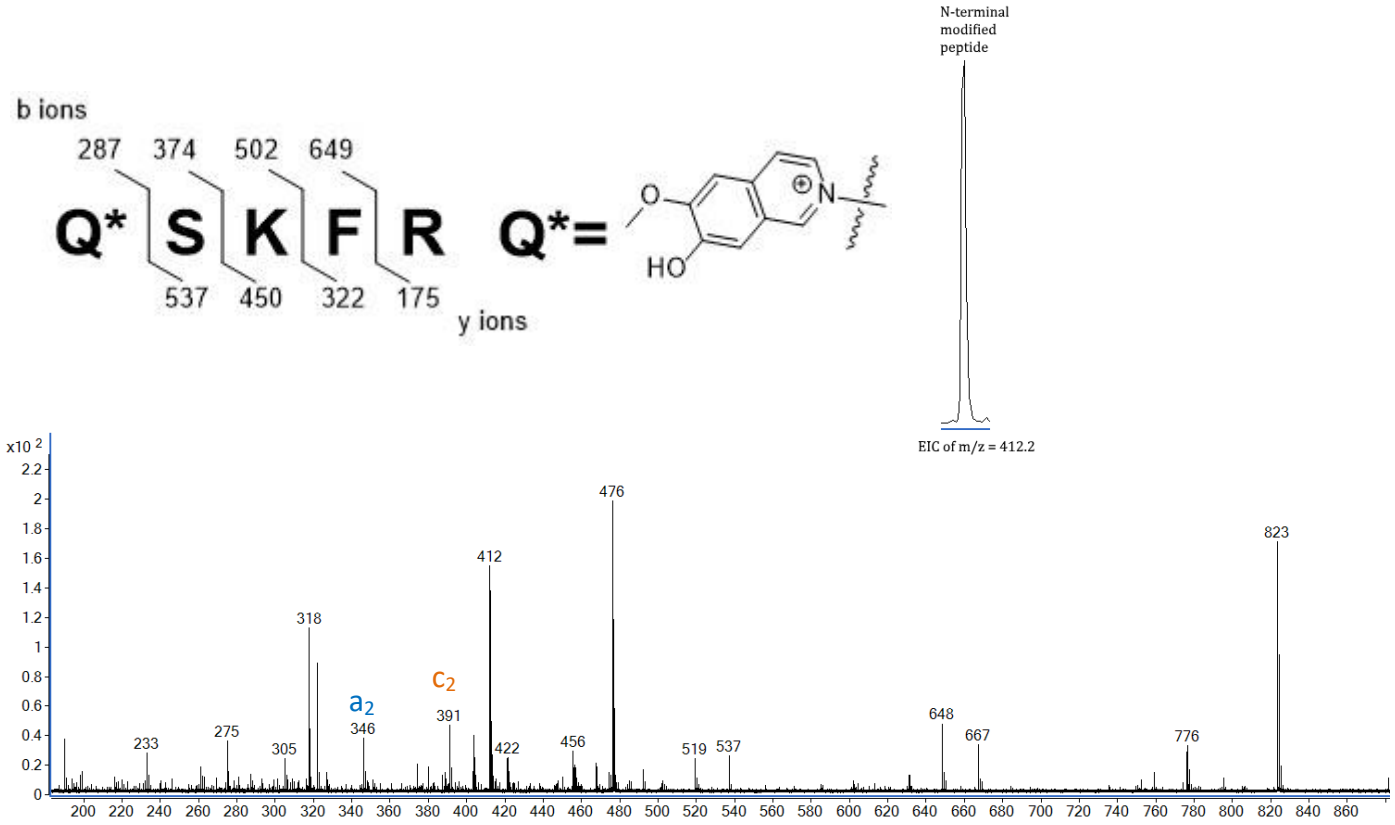


Figure S51 Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified QSKFR** (ESI source, doubly charged ion of m/z = 412.2). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 412.2.

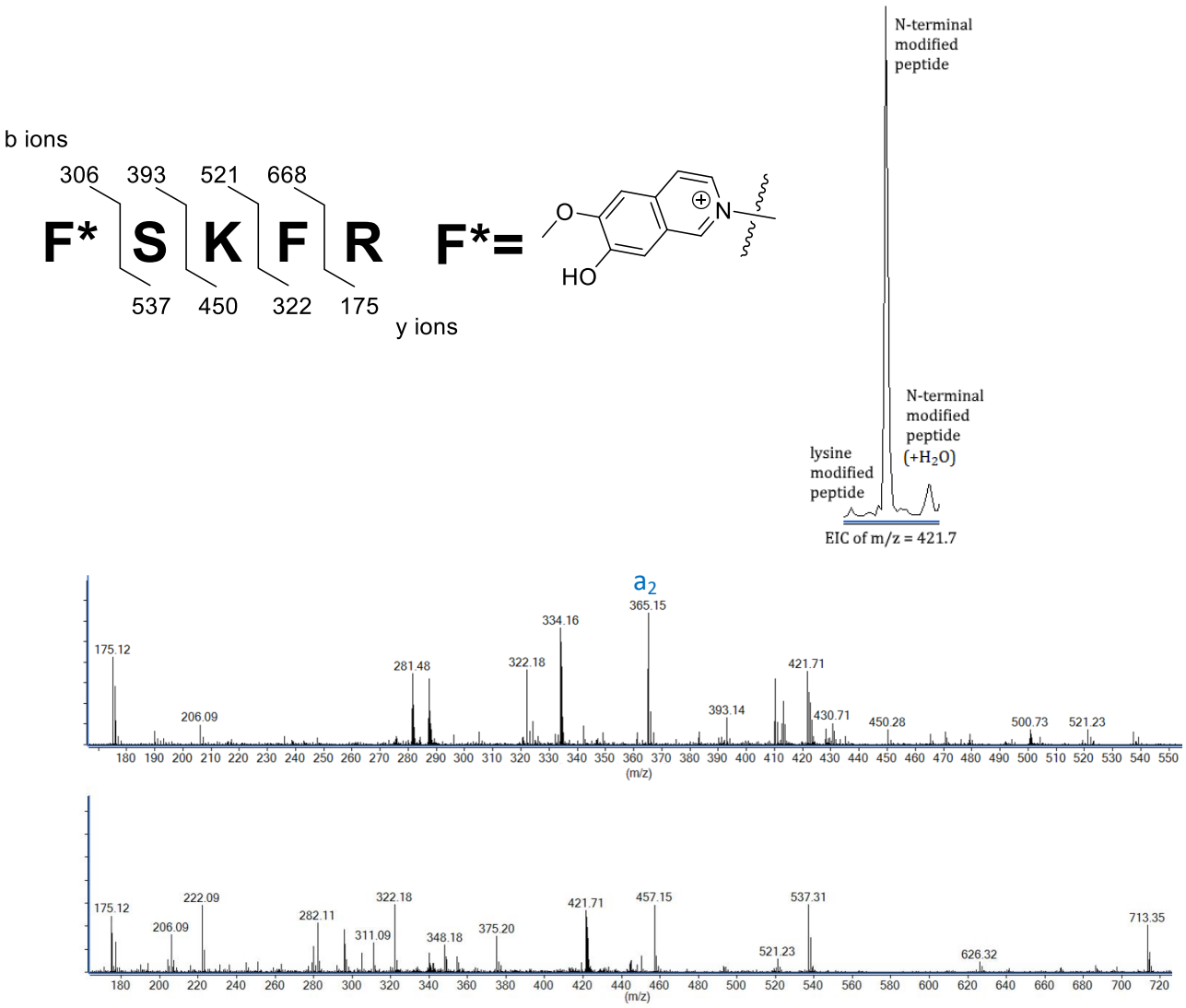


Figure S52. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified FSKFR (Top); *N*-terminal 2b-modified FSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 421.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 421.7.

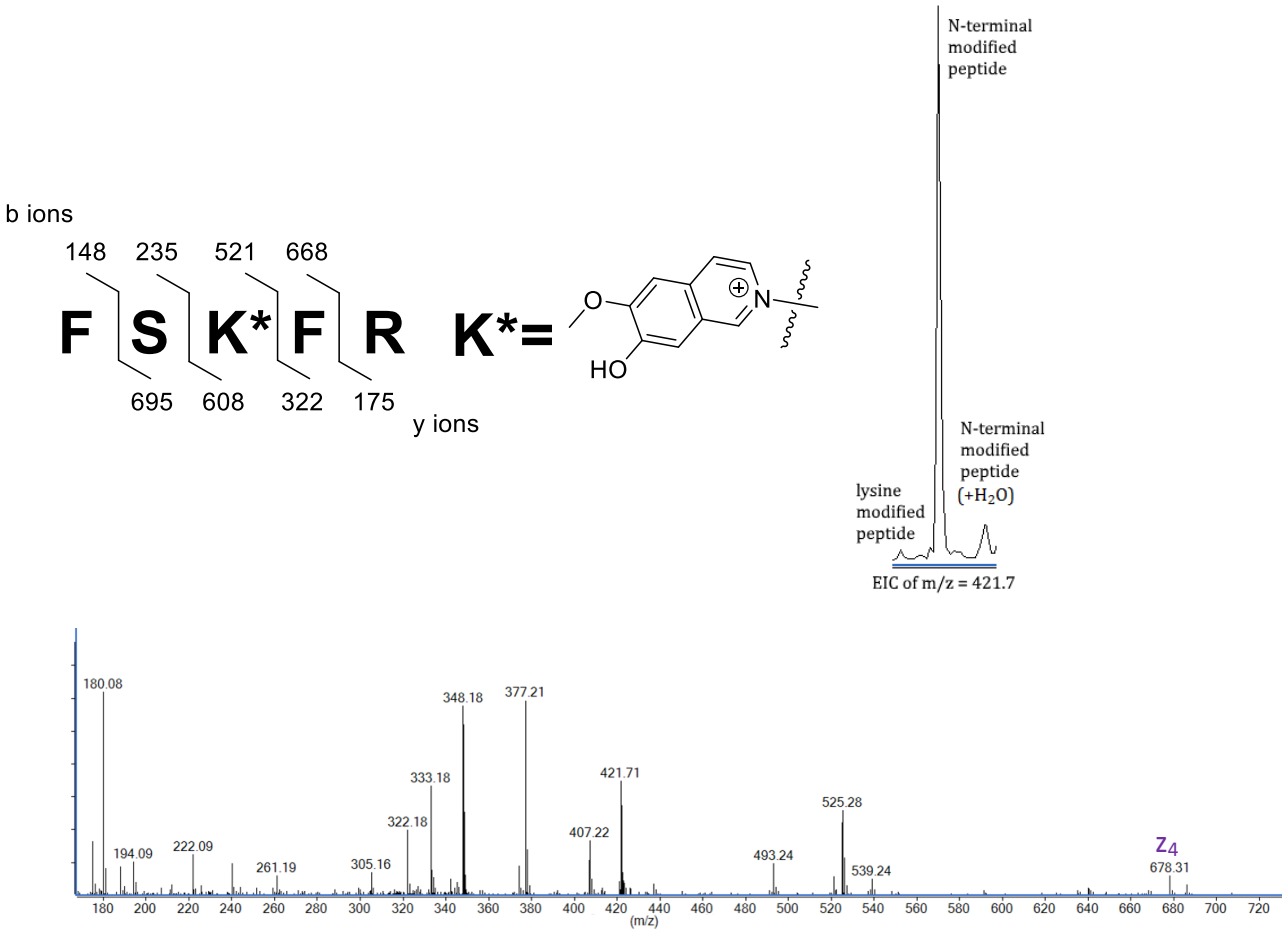


Figure S53. Q-TOF MS/MS spectrum of **Lysine 2b-modified FSKFR** (ESI source, doubly charged ion of m/z = 421.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 421.7.

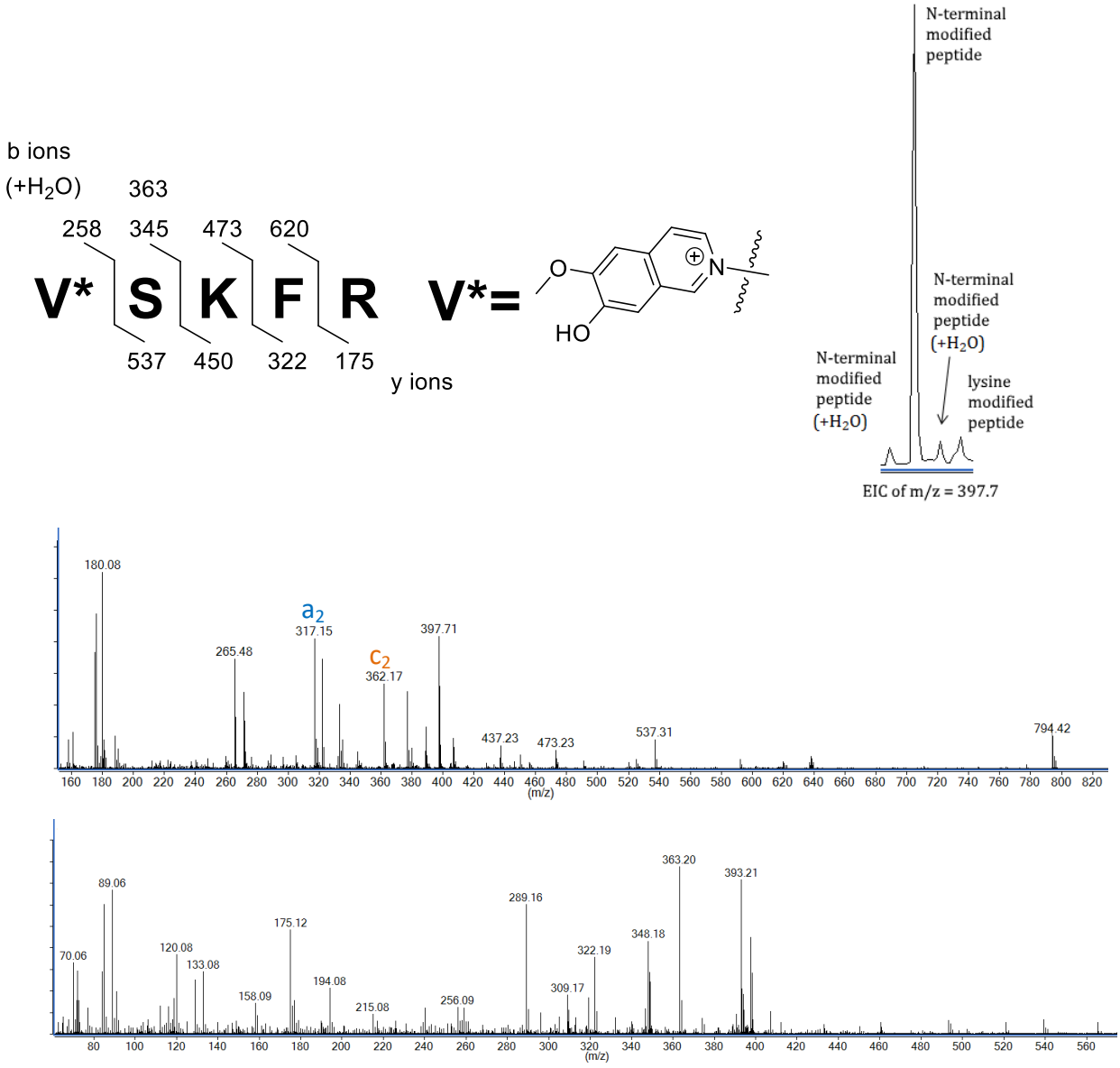


Figure S54. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified VSKFR (Top); *N*-terminal 2b-modified VSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 397.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 397.7.

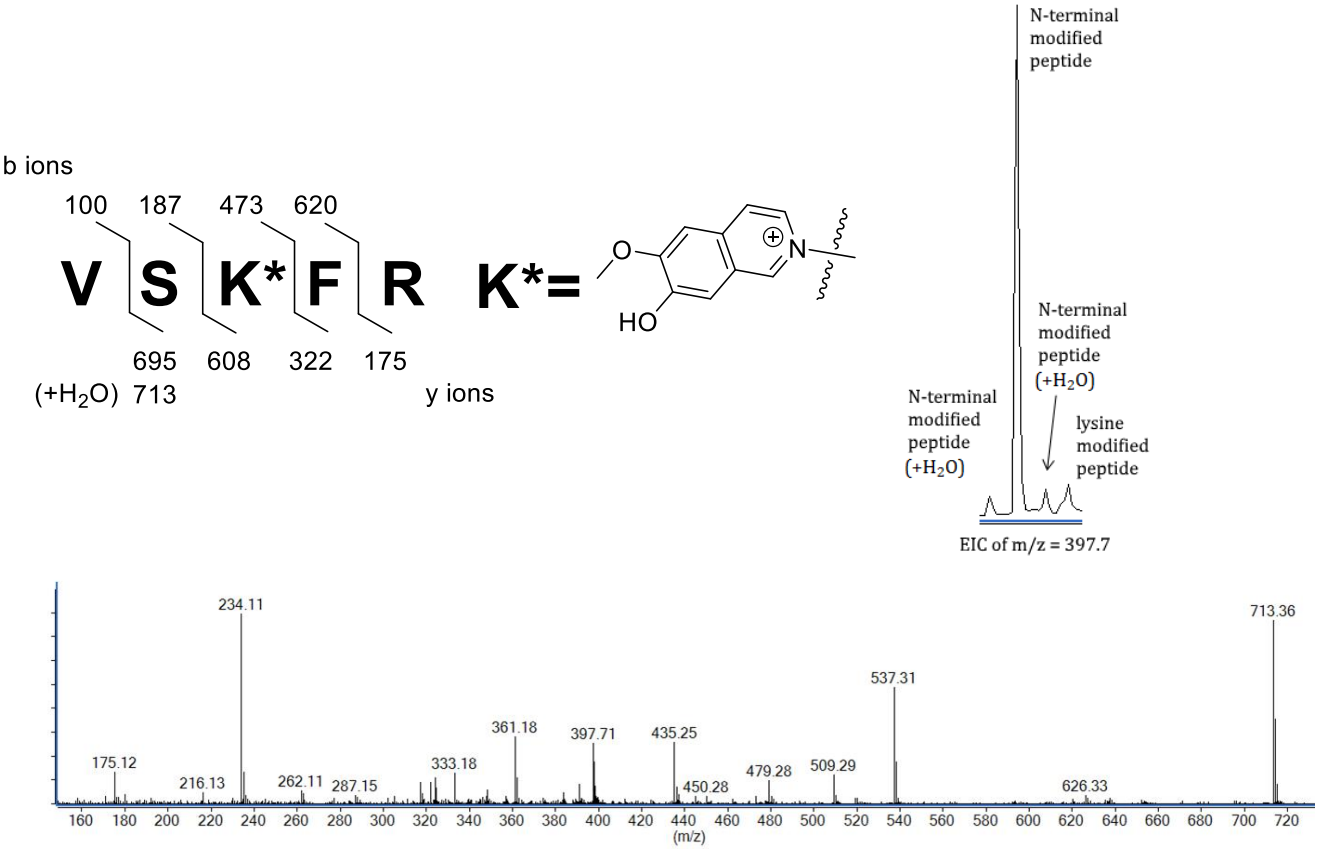


Figure S55. Q-TOF MS/MS spectrum of **Lysine 2b-modified VSKFR** (ESI source, doubly charged ion of m/z = 397.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 397.7.

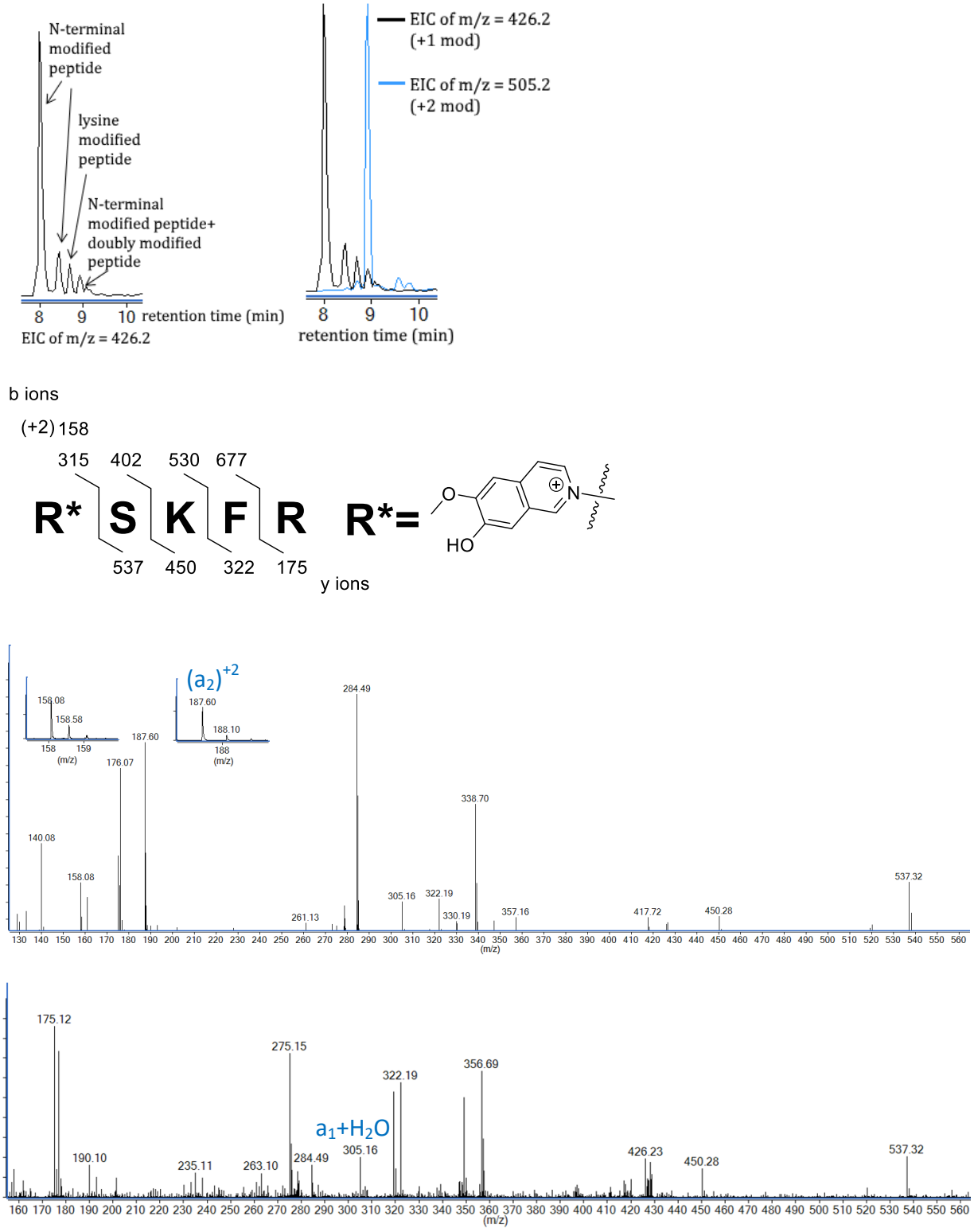
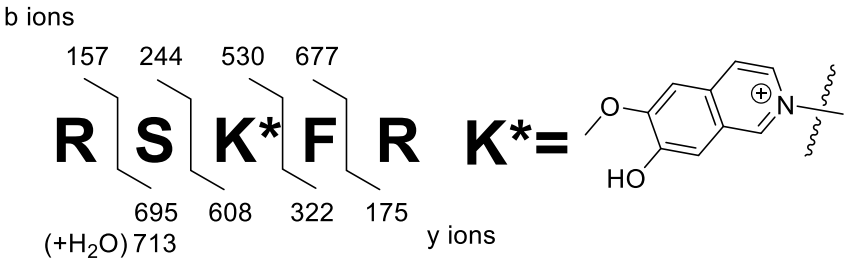
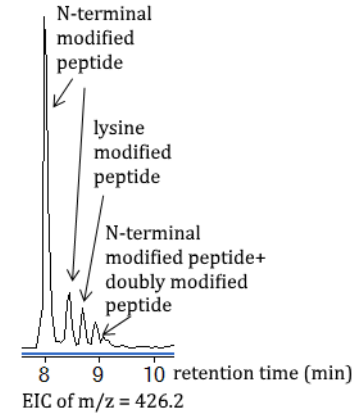


Figure S56. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified RSKFR (Top); *N*-terminal 2b-modified RSKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 426.2 (+1mod), 505.2 (+2mod)). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 426.2 (+1mod), 505.2 (+2mod).





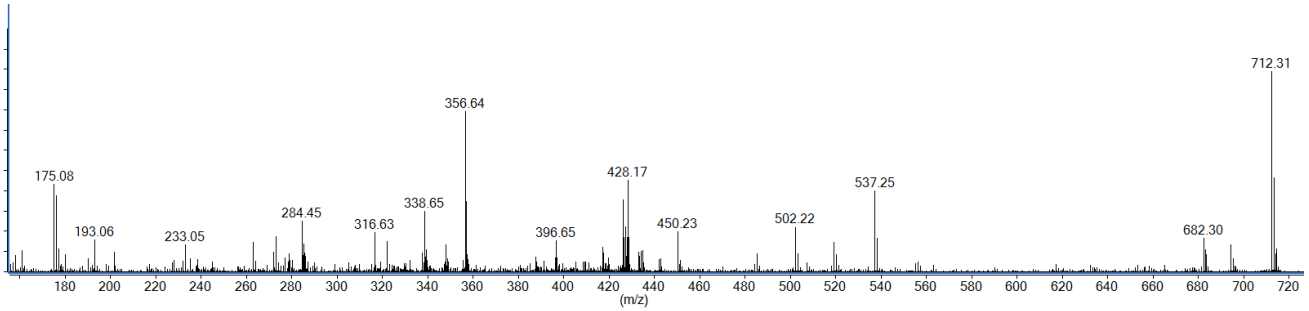


Figure S57. Q-TOF MS/MS spectrum of **Lysine 2b-modified RSKFR** (ESI source, doubly charged ion of m/z = 426.2). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 426.2.

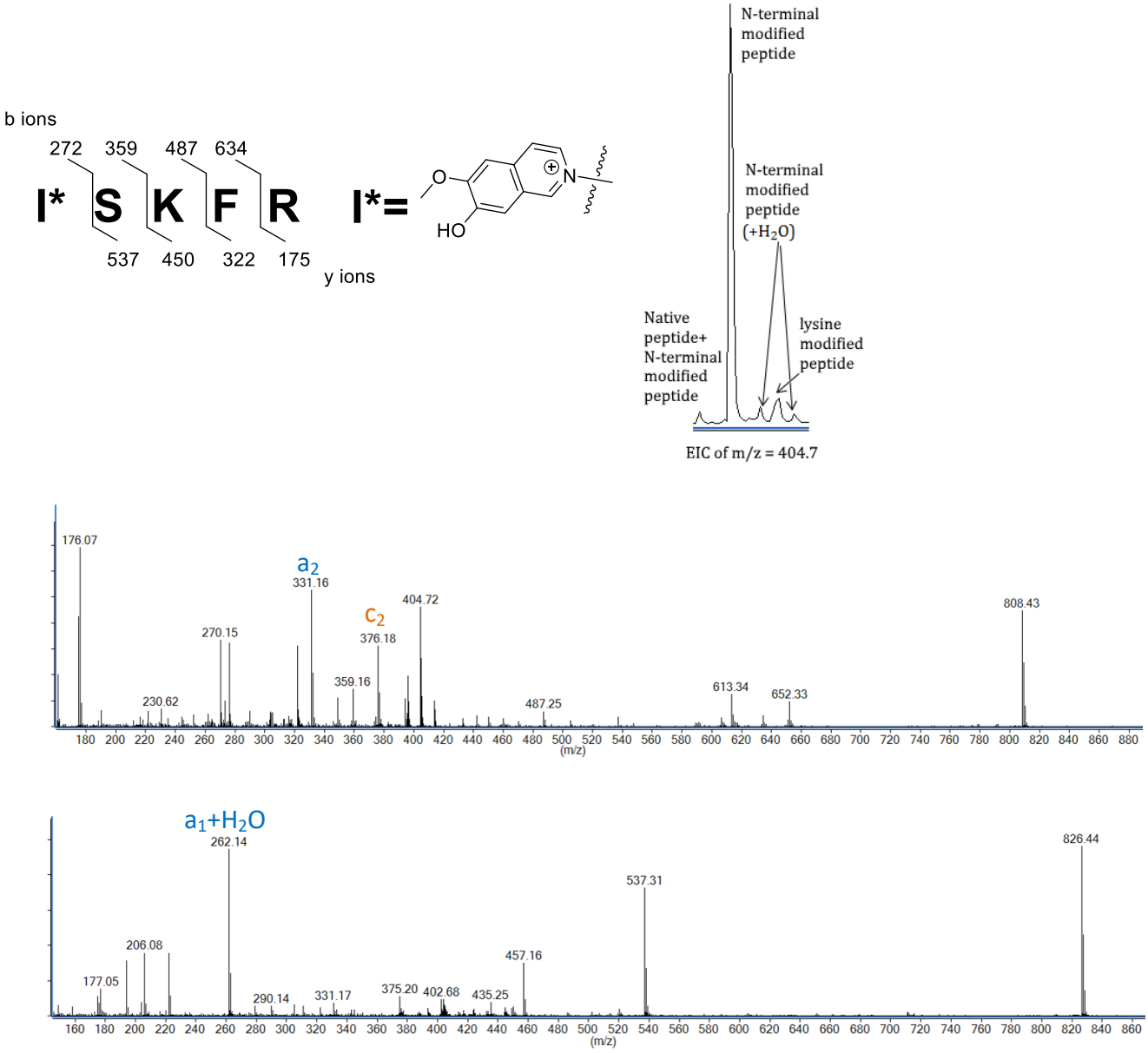


Figure S58. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified ISKFR (Top); *N*-terminal 2b-modified ISKFR + H_2O (Bottom)** (ESI source, doubly charged ion of m/z = 404.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 404.7.

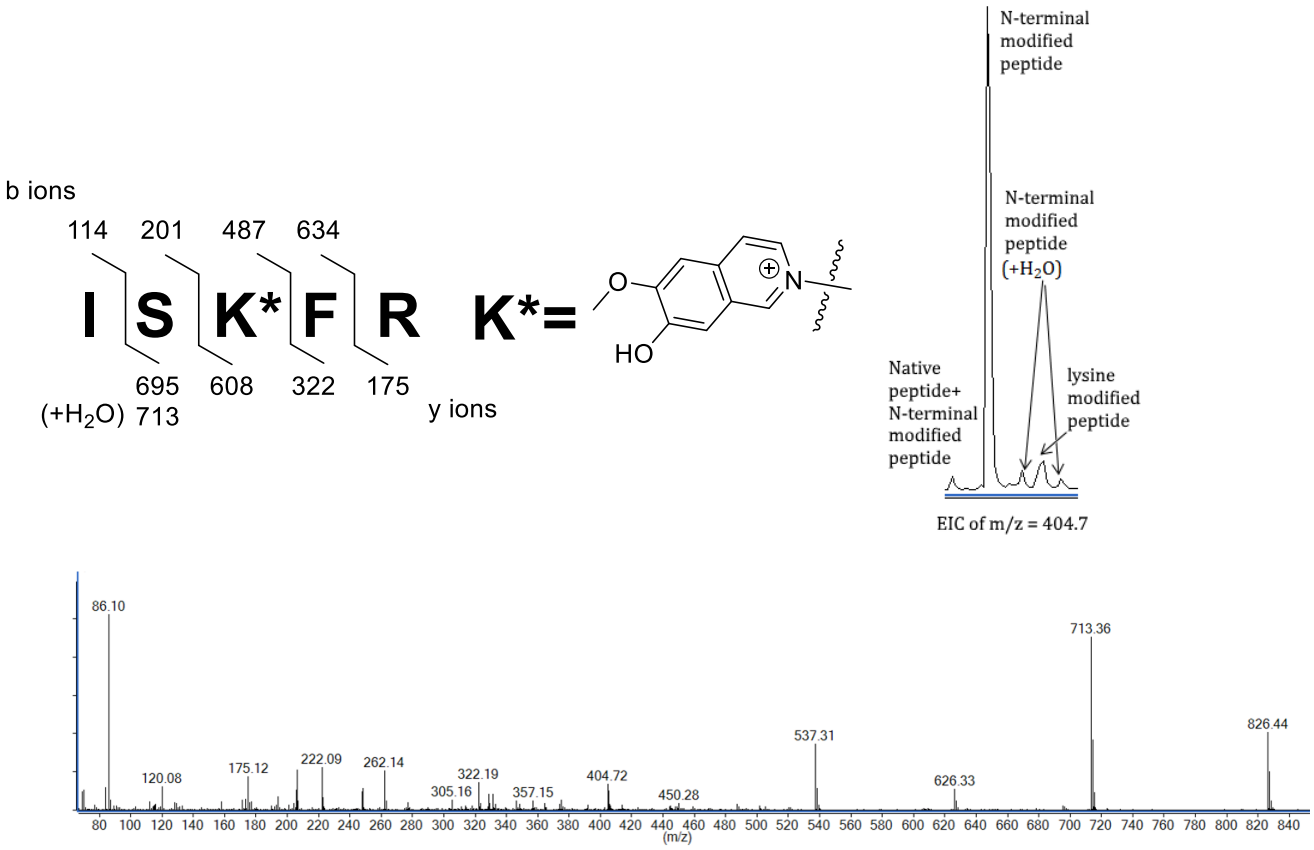


Figure S59. Q-TOF MS/MS spectrum of **Lysine 2b-modified ISKFR** (ESI source, doubly charged ion of m/z = 404.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 404.7.

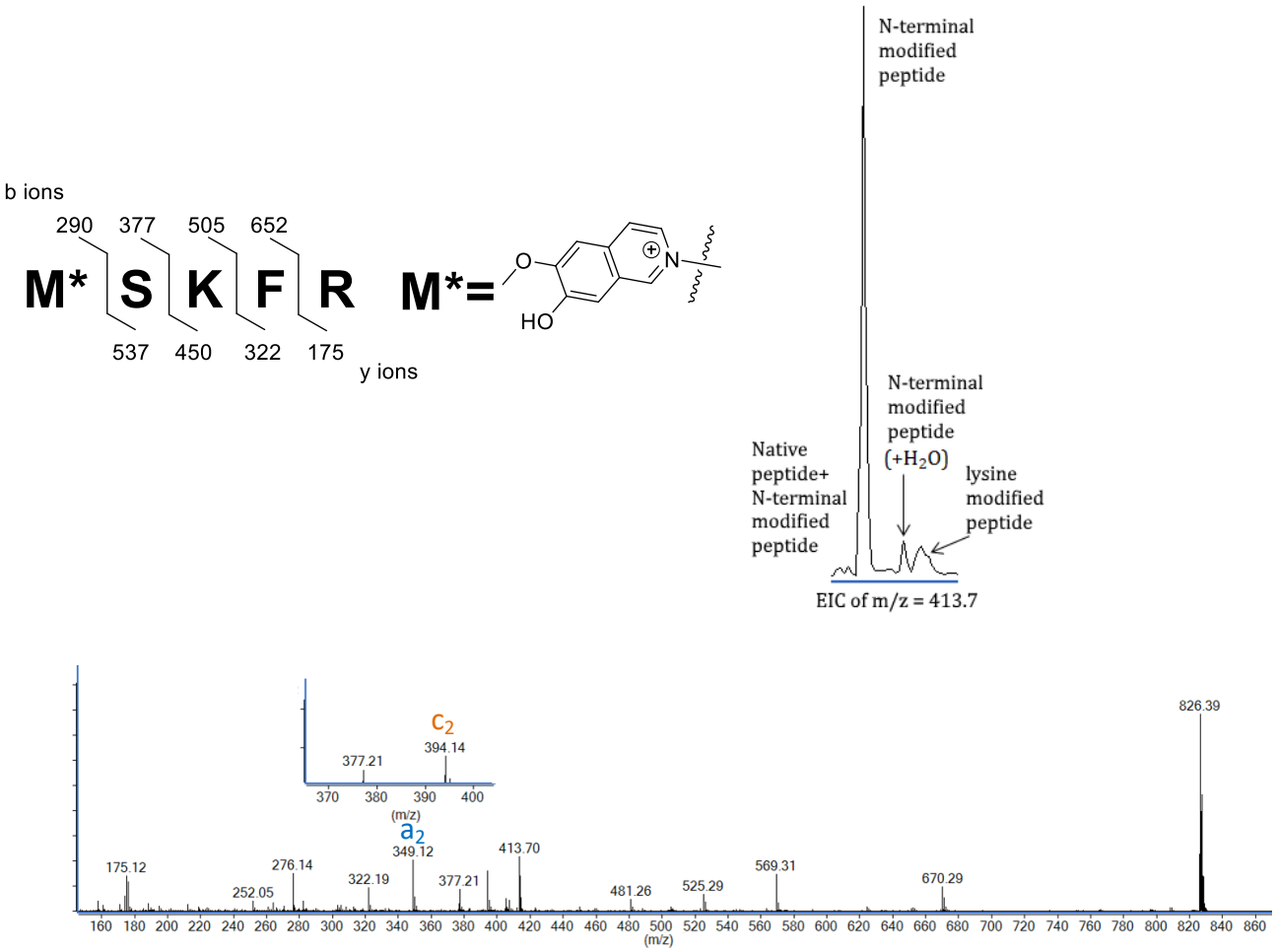


Figure S60. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified MSKFR** (ESI source, doubly charged ion of m/z = 413.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 413.7.

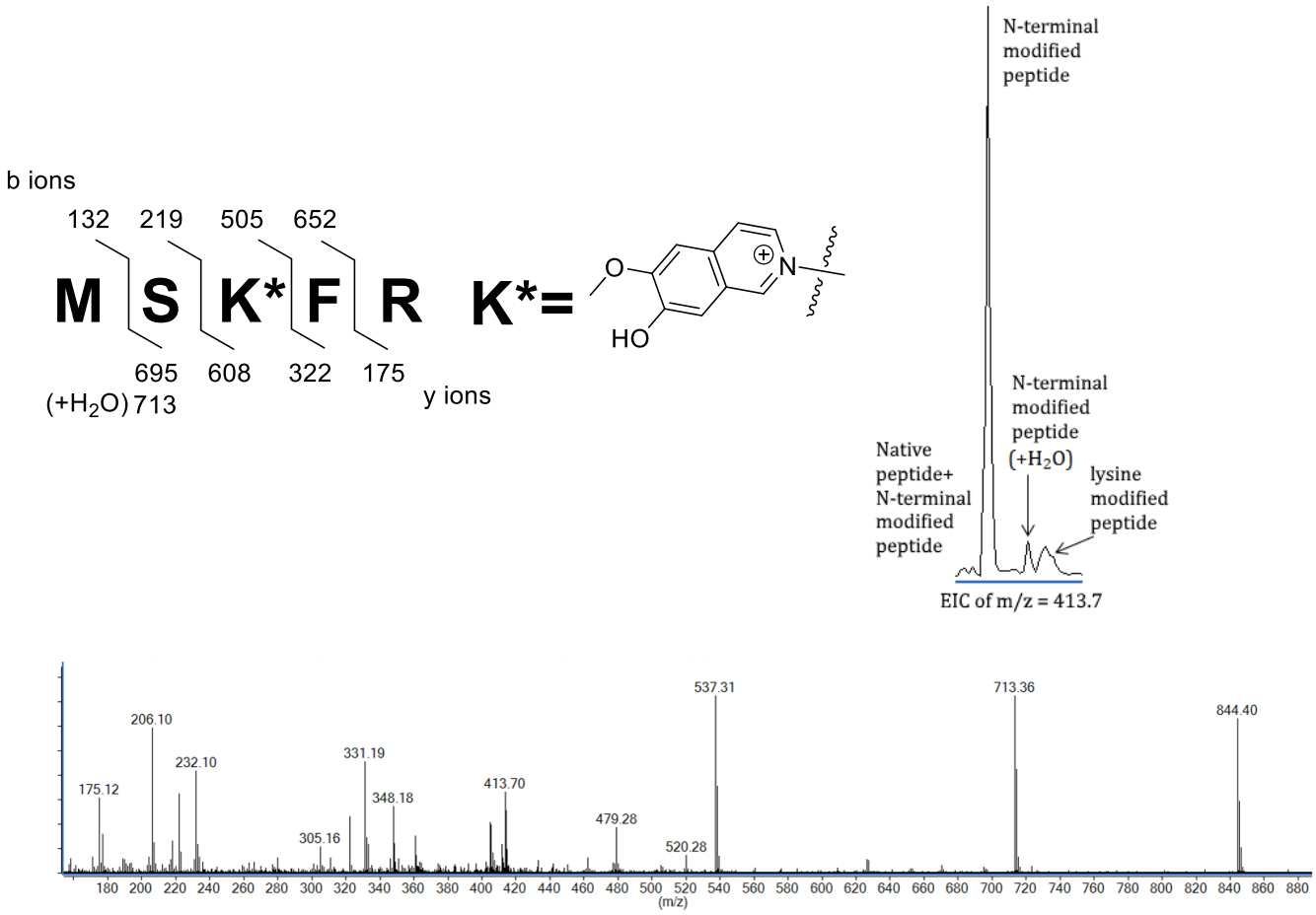


Figure S61. Q-TOF MS/MS spectrum of **Lysine 2b-modified MSKFR** (ESI source, doubly charged ion of m/z = 413.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 413.7.

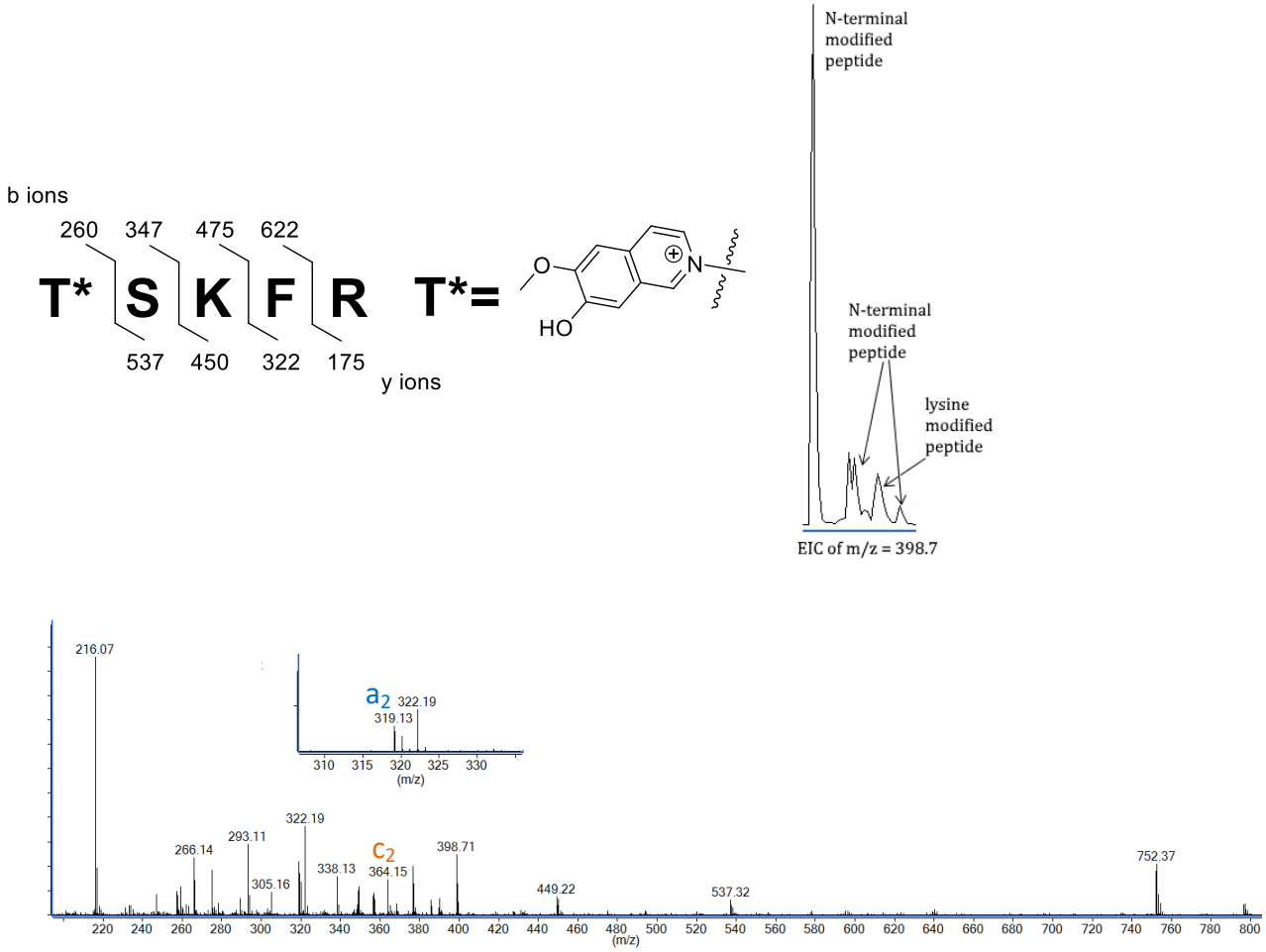


Figure S62. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified TSKFR** (ESI source, doubly charged ion of m/z = 398.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 398.7.

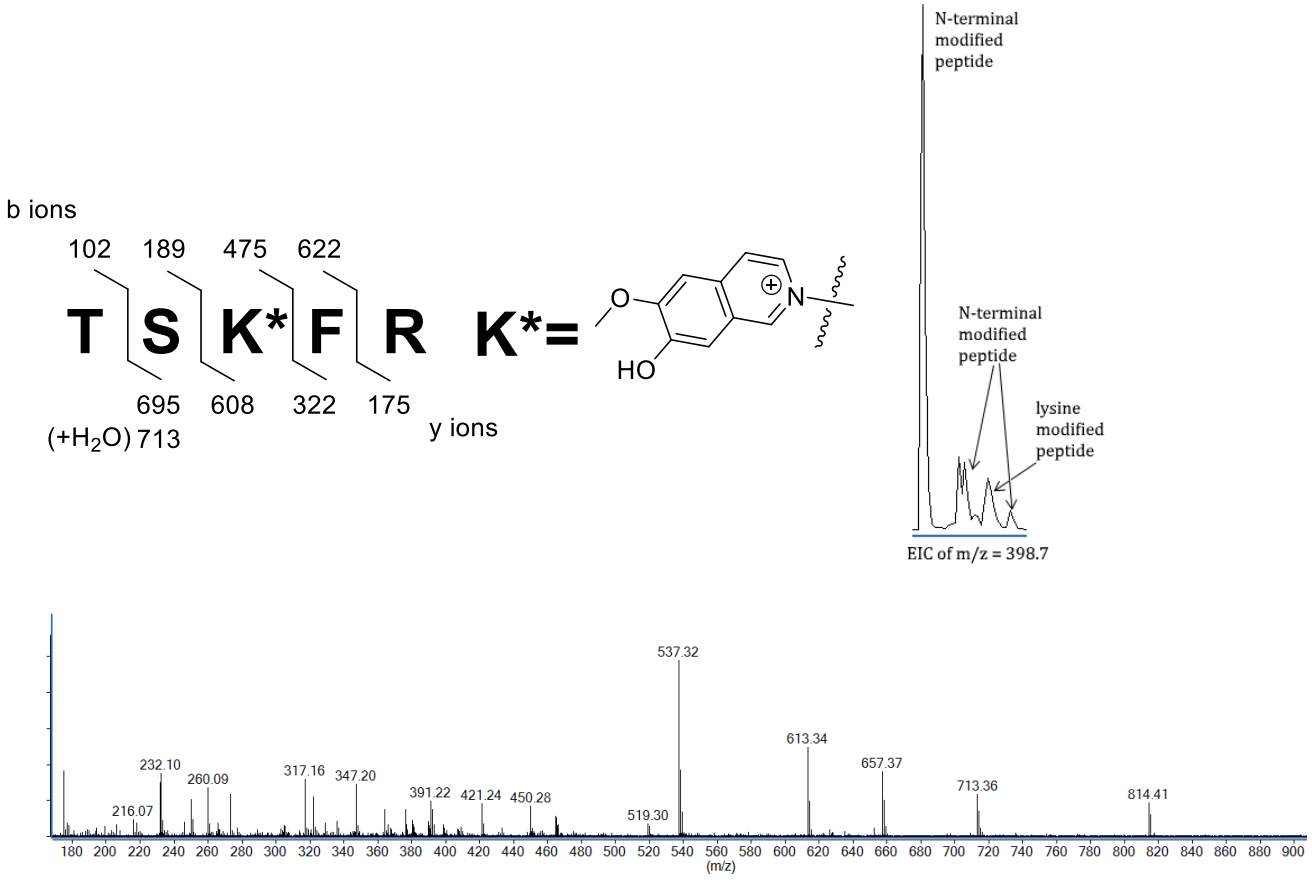


Figure S63. Q-TOF MS/MS spectrum of **Lysine 2b-modified TSKFR** (ESI source, doubly charged ion of m/z = 398.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 398.7.

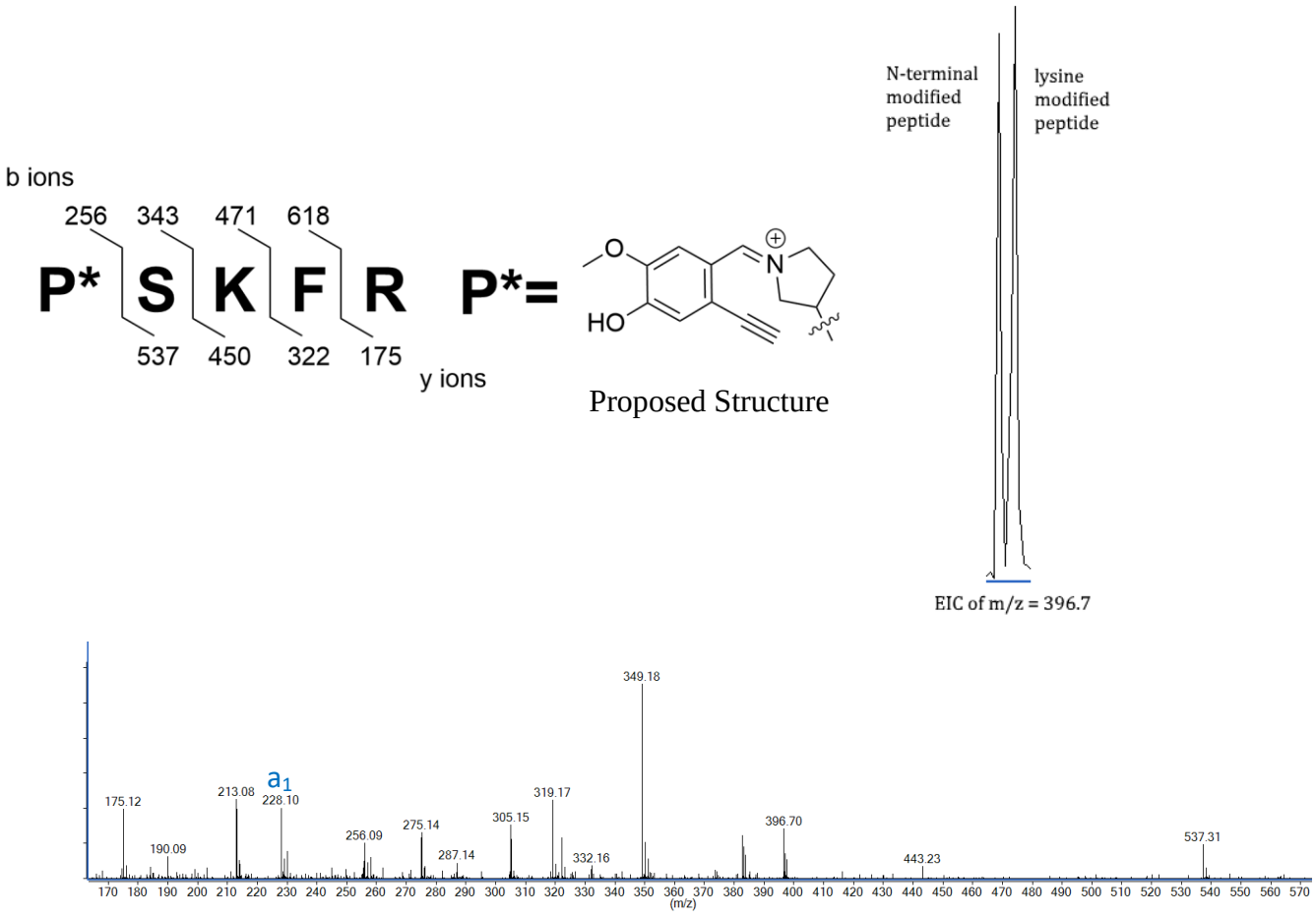


Figure S64. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified PSKFR** (ESI source, doubly charged ion of m/z = 396.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 396.7.

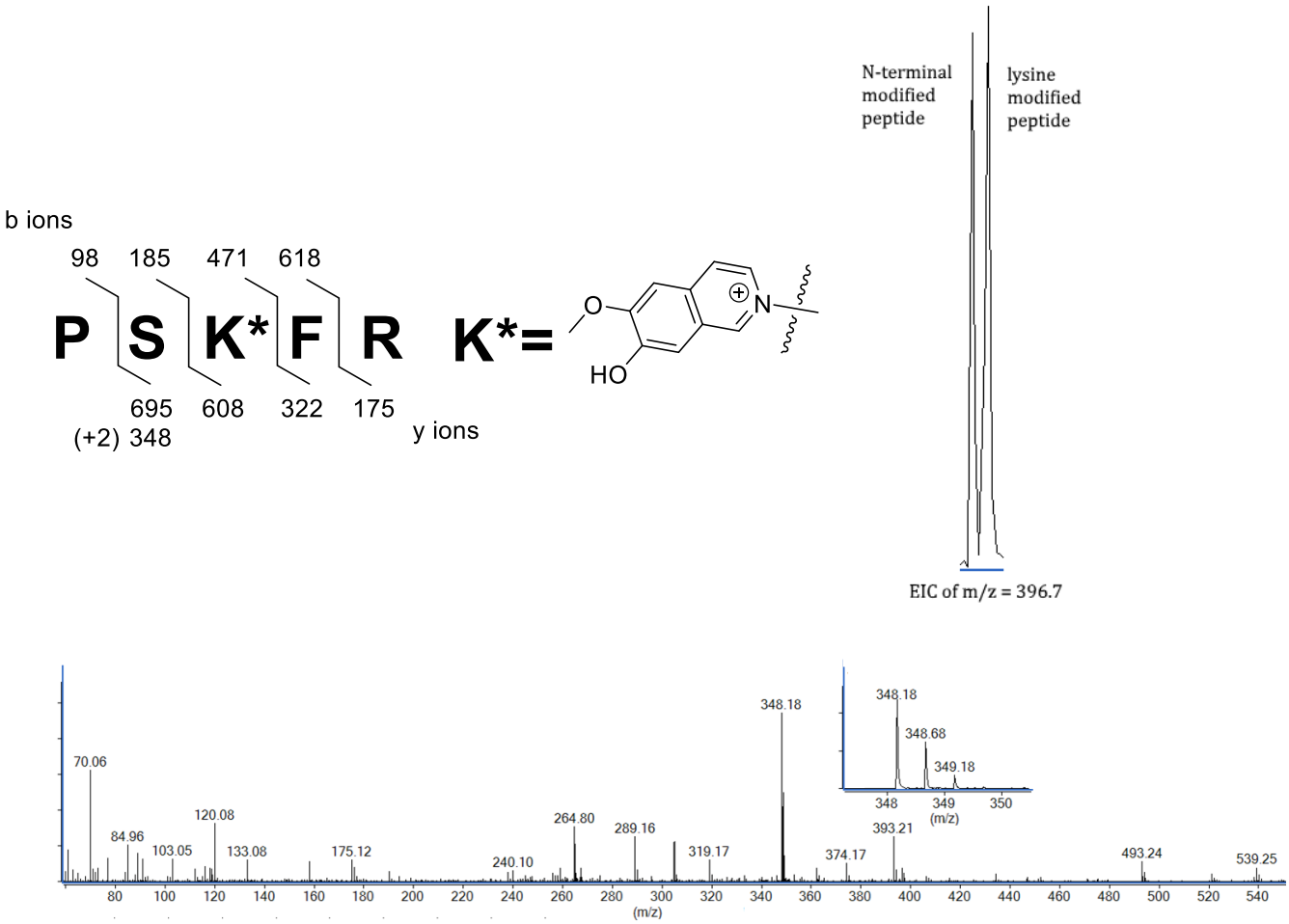


Figure S65. Q-TOF MS/MS spectrum of **Lysine 2b-modified PSKFR** (ESI source, doubly charged ion of m/z = 396.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 396.7.

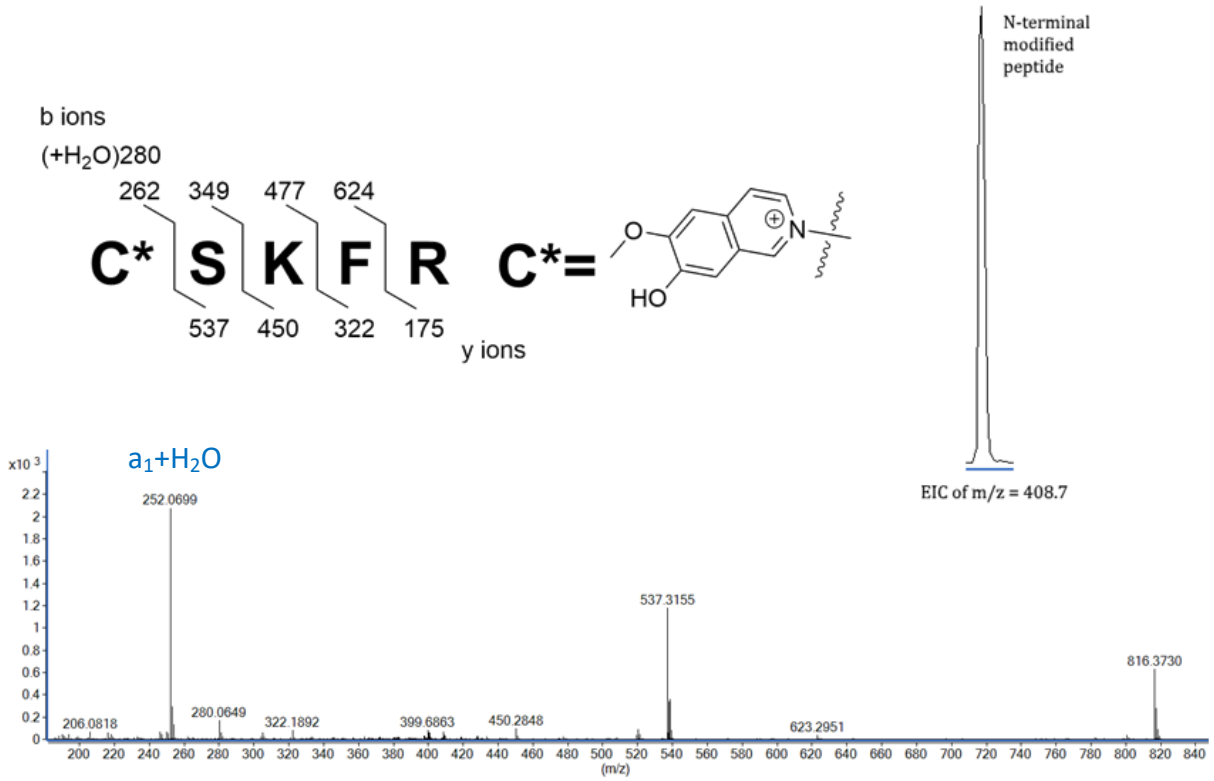


Figure S66. Q-TOF MS/MS spectrum of ***N*-terminal 2b-modified + H_2O CSKFR** (ESI source, doubly charged ion of m/z = 408.7). Inset: the extracted ion chromatogram (EIC) of doubly charged ion of m/z = 408.7. Conditions: CSKFR (0.1 mM), reagent **2b** (2 mM, 20 equiv.), TCEP (5 mM, 50 equiv.) in pH 6.3 50 mM of PBS buffer/DMSO (9:1, 100 μL).

Reference

1. (a) Huang, B. X.; Kim, H. Y. *J. Am. Soc. Mass Spectrom.*, **2004**, *15*, 1237; (b) Temple, A.; Yen, T. Y.; Gronert, S. *J. Am. Soc. Mass Spectrom.*, **2006**, *17*, 1172.

2. (a) Chames, P.; Regenmortel, M. V.; Weiss, E.; Baty, D. *Br. J. Pharmacol*., **2009**, *157*, 220; (b) Pisal, D. S.; Kosloski, M. P.; Balu-lyer, S. V. *J Pharm Sci*., **2010**, *6*, 2557; (c) Bruno, B. J.; Miller, G. D.; Lim, C. S. *Ther Deliv*., **2013**, *11*, 1443.

3. Harris, J. M.; Chess, R. B. *Nat. Rev. Drug. Discov*., **2003**, *2*, 214.

4. Cheng, P. N.; Leung, Y. C.; Lo, W. H.; Tsui, S. M.; Lam, K.C. *Cancer Lett*., **2005**, *224*, 67; (b) Lam, T. L.; Wong, G. K.; Chow, H.Y.; Chong, H. C.; Chow, T. L.; Kwok, S. Y.; Cheng, P. N.; Wheatley, D. N.; Lo, W. H.; Leung, Y. C.; *Pigment Cell Melanoma Res*., **2011**, *2*, 366; (c) Yau, T.; Cheng, P. N.; Chan, P.; Chan, W.; Chen, L.; Yuen, J.; Pang, R.; Fan, S.T.; Poon, R. T. *Invest New Drugs*., **2013**, *1*, 99.

5. Chan, A. O.-Y.; Ho, C.-M.; Chong, H.-C.; Leung, Y.-C.; Huang, J.-S.; Wong, M.-K.; Che, C.-M. *J. Am. Chem. Soc*., **2012**, *134*, 2589.