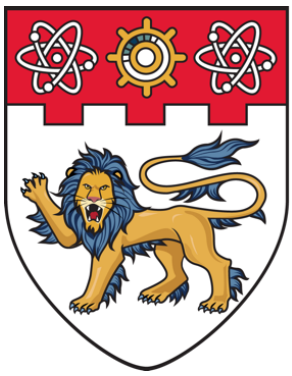


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

PROTEIN ENGINEERING USING ASPARAGINYL PEPTIDE LIGASES

WANG ZHEN

SCHOOL OF BIOLOGICAL SCIENCES

2022

PROTEIN ENGINEERING USING ASPARAGINYL PEPTIDE LIGASES

WANG ZHEN

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University in partial fulfilment of the requirement for the degree of Doctor of Philosophy

2022

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

07-January-2022

. . . . . .

Date

Wang Zhen

. . . . . .

Wang Zhen

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

07-January-2022

. . . . . .
Date

. . . . . .
Liu Chuan Fa

Authorship Attribution Statement

This thesis contains material from 1 paper published in a peer-reviewed journal and 2 manuscripts in preparation of which I am listed as the first author.

(1) **Wang, Z;** Zhang, D; Hemu, X; Hu, S; To, J; Zhang, X; Lescar, J; Tam, J.P. and Liu, C.F. 2021. Engineering protein theranostics using bio-orthogonal asparaginyl peptide ligases. *Theranostics*,*11*(12), p.5863. Content for Chapter 2.

(2) **Wang. Z,** Zhang. D, Hu. S, Simin. Fang, Lescar. J, James. P. T and Liu. CF. 2021. Engineering multi-functional protein biologics through PAL-mediated hydrazide ligation. (Manuscript under preparation) Content for Chapter 3.

(3) **Wang. Z,** Zhang. D, Hu. S, Lescar. J, James. P. T and Liu. CF. 2021. A removable cysteinyl-asparaginyl tag for traceless protein ligation. (Manuscript under preparation) Content for Chapter 4.

The contributions of the co-authors are as follows:

- Prof Liu Chuan-Fa provided the initial project direction and edited the manuscript drafts.
- I prepared the manuscript drafts. The manuscript was revised by Zhang Dingpeng.
- I co-designed the study with Prof Liu Chuan-Fa and performed all the laboratory work at the School of Biological Sciences, NTU. I also analyzed the data.
- All microscopy studies, including sample preparation, were conducted by me and Dr. Zhang Dingpeng in the Facility for Analysis, Characterization, Testing and Simulation.
- Hu Side assisted in providing the VyPAL2 enzyme; Dr. Janet To and Zhang Xiaohong assisted in providing the Butelase-1 enzyme. Dr Seetharam Balamkundu aided on chemical synthesis.
- Prof James P. Tam and Prof Julien Lescar provided advice and suggestions on my projects.

07-January-2022

. . . . . .

Date

. . . . . .

Wang Zhen

Acknowledgements

I would like to thank my academic supervisor, Prof. Liu Chuan-Fa. I thank Prof. Liu for the opportunity to join his research group and for his guidance in conducting research in the field of chemical biology, which is totally new to me. He has been so patient in guiding and supporting me on my projects. Throughout the four years of PhD training, I have enjoyed my work and I have learned so many new things. I love this research field now since I realize that my previous biological knowledge can be combined very well with the new chemistry skills that I have learned from him, which I believe will allow me to solve challenging scientific problems in the future. In his lab, I have not only acquired the ability to work on projects independently, but also the ability to cooperate well with other members, which will be important for my future career development. I am and will always be grateful to him for all his help and kindness.

In addition, I would also like to thank my TAC members, Prof. James P. Tam and Prof. Koh Cheng Gee, who have provided me with valuable suggestions for my projects. I would also like to thank all my lab members who have given me support in my experiments. Additionally, I would like to thank the School of Biological Sciences of NTU for the conducive academic atmosphere and excellent facilities. I would also like to thank the Ministry of Education of Singapore for the Tier 3 program grant that has supported me in my PhD study. The interactions in the Tier 3 teams have helped me gain a much better understanding and knowledge about our research fields. Everyone is very respectful and has good teamwork spirit, as well as an extreme willingness to share data and material. The intellectual debate within the group not only helps me gain a deeper insight into questions related to my PhD study but also helps develop our critical thinking ability and problem-solving skills.

Table of Contents

Abbreviations

ALL, acute lymphoblastic leukemia;

AEP, Asparaginyl endopeptidase;

BML, butelase-mediated ligation;

Boc, tert-butyloxycarbonyl;

CAR-NK, Chimeric antigen receptor-Natural killer cells;

cPDC, cycloprotein-drug conjugate;

CPE: cysteinyl-prolyl ester;

DAPI, 4', 6-Diamidino-2-Phenylindole, Dihydrochloride;

DCM, Dichloromethane;

DIPEA, N,N-Diisopropylethylamine;

DMF, dimethylformamide;

DMEM, Dulbecco's Modified Eagle Medium;

DOX, doxorubicin;

EC_{50}, half-maximal effective concentration;

EDTA, Ethylenediaminetetraacetic acid;

EGFR, epidermal growth factor receptor;

ESI-MS, electrospray ionization mass spectrometry;

FACS: fluorescence-activated cell sorting;

FBS, fetal bovine serum;

FITC, fluorescein isothiocyanate;

Fmoc, Fluorenylmethyloxycarbonyl;

FPLC, fast protein liquid chromatography;

GFP, green fluorescent protein;

GVHD, graft-versus host disease;

HPLC, High-performance liquid chromatography;

IC_{50}, the half-maximal inhibitory concentration;

IL-2: interleukin-2;

IPTG, Isopropyl β-D-1-thiogalactopyranoside;

K_D, equilibrium dissociation constant;

LB: luria broth;

LDH: lactate dehydrogenase;

MBHA, 4-Methylbenzhydrylamine;

MD, molecular dynamic;

MesNa: sodium 2-mercaptoethanesulfonate;

MMAE, monomethyl auristatin E;

MPAA:4-mercaptophenylacetic acid;

mTG: microbial transglutaminase;

MTT, 3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide;

NCL, native chemical ligation;

NHV, Asn-His-Val tripeptide;

NGF, Asn-Gly-Phe tripeptide;

Ni-NTA, nitrilotriacetic acid-nickel;

OD, optical density;

PALs, peptidyl asparaginyl ligases;

PBS, phosphate saline buffer;

PCY1, peptide cycles 1;

PICS, Proteomic Identification of protease Cleavage Sites;

PMHL, PAL-mediated hydrazide ligation;

POI, protein of interest;

PTM, Post-translational modification;

PyBOP, benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate;

SA, surface area;

scFv, single-chain variable fragment;

SPPS, solid phase peptide synthesis;

TCEP: tris (2-carboxyethyl) phosphine;

TFA, Trifluoroacetic acid;

TIS, Triisopropylsilane;

TMP, thrombopoietin mimic peptide;

TPO, thrombopoietin;

TPOR, thrombopoietin receptor;

VML, VyPAL-mediated ligation;

Vy, Viola yedoensis.

Abstract

Peptidyl Asx-specific ligases (PALs) function as molecular gluing machines that join peptides and proteins together at Asn or Asp junctions. Since the discovery of butelase-1 as the first PAL in 2014, these enzymes have found numerous applications in basic and translational research. Due to their highly specific and efficient catalytic activity, PALs continue to attract the attention of researchers from the biotech industry and biomedical research communities. In this thesis, I first report on a bio-orthogonal scheme using two asparaginyl peptide ligases with differential substrate specificities – butelase-1 and VyPAL2. This scheme allows for tandem ligation on the same protein in either the N-to-C or C-to-N direction, making it possible to prepare dually labeled proteins as potential theranostic agents. Second, my study shows that PALs can use non-canonical nucleophilic compounds – such as hydrazides – as acyl acceptor substrates. This expanded substrate scope is explored to engineer multi-functional proteins, including bi-specific protein engagers capable of mediating the killing of cancer cells by Car-NK cells. Finally, I report the design of an auto-processing mechanism of a C-terminal Cys-Asn dipeptide for traceless protein ligation by combining PAL ligation with native chemical ligation or subtiligation. My work further demonstrates PALs as powerful tools of biotechnology for protein engineering.

In chapter 1, I give a general introduction to peptidyl Asx-specific ligases and current research works about protein macrocyclization and modification, as well as cell-surface engineering of chimeric antigen receptor T cells or NK cells. More specific and relevant information is given by a separate introduction to each of the subsequent chapters.

In chapter 2, I report on a bio-orthogonal scheme using two asparaginyl peptide ligases – butelase-1 and VyPAL2 – which allows for tandem ligation on the same protein in either the N-to-C or C-to-N direction, leading to its dual labeling at the C- and N-terminal ends. No protection on the protein substrate is required when performing the first ligation step, although

butelase-1 and VyPAL2 are both asparagine-specific. Thus, a distinct advantage of bio-orthogonal ligation is the use of mild enzymatic reactions under aqueous conditions, which are compatible with biologics, such as proteins, antibodies, and live cells. In addition to N- and C-terminal directed protein dual labeling, our bio-orthogonal tandem ligation strategy can also be used to prepare cycloprotein-drug conjugates or cPDCs. This involves the use of a synthetic intervening peptide designed to join the two termini of a protein. The peptide is trifunctional, containing an N-terminal GF-dipeptide nucleophile substrate for VyPAL2, a C-terminal NHV tripeptide motif as the acyl donor substrate for butelase-1, and an internal aminooxy functionality for oxime conjugation, which would allow consecutive PAL-mediated ligation, cyclization, and doxorubicin attachment. Given the expected thermal and metabolic stability of cycloproteins, cycloprotein conjugates are interesting candidates for theranostics. Using an EGFR-binding affibody as the model protein, I demonstrated the feasibility of our tandem ligation strategy for the preparation of dually labeled proteins and cycloprotein-drug conjugates. Because butelase-1 and VyPAL2 are the two most powerful ligases, such a bio-orthogonal tandem ligation strategy would offer an ideal solution to the challenging problem

In Chapter 3, we report that hydrazide nucleophiles are effective acyl acceptors in ligation reactions catalyzed by peptidyl asparaginyl ligases (PALs). Because hydrazides are easily functionalizable, hydrazide ligation is highly versatile. Interestingly, the linkages formed with hydrazide substrates have varying degrees of liability toward the PAL enzyme, with some being remarkably resistant and thus ensuring nearly irreversible ligation. Using the hydrazide ligation method, we labeled an EGFR-targeting affibody-Fc fusion protein with various functional moieties to generate selective and potent cancer imaging and therapeutic agents. Irreversible hydrazide ligation also allowed a sequential ligation scheme to be conducted on a protein. Using this scheme, quadruple FITC labels were introduced onto the N- and C-termini of the affibody-Fc protein to yield a bispecific engager for CAR-NK cell therapy. Our work

expands the substrate scope of PAL enzymes and further points to their promise as a precision manufacturing tool for multifunctional protein biologics.

In Chapter 4, we report the design of an auto-processing mechanism of a C-terminal Cys-Asn dipeptide for traceless native chemical ligation or subtiligation. Using a PAL enzyme to convert peptide/protein Xaa-Cys-Asn-P1'-P2' into an Xaa-Cys-Asn-hydrazide, oxidation of the hydrazide into the azide group activates the C-terminal Asn for the auto-processing mechanism through diketopiperazine formation, which captures the transient product of N-to-S acyl transfer, driving the formation of the thioester with the thiol group of the Cys residue. In the presence of a thiol compound such as MesNa or MPAA, thiol-thioester exchange leads to a new thioester with the removal of the Cys-Asn dipeptide as diketopiperazine. The new C-terminal Xaa-thioester can be used for native chemical ligation or enzymatically for subtiligation. Essentially, this method would obviate the need for PAL-mediated ligation at the Asn junction. In a model study on small synthetic peptides, we observed formation of the desired Xaa-thioester product. However, this transformation was accompanied by other side reactions. More work needs to be done to demonstrate the practical value of this method. Once this is demonstrated, the traceless nature of this ligation scheme would make it useful for the preparation of homogeneously modified proteins with completely preserved native sequences.

Chapter 1: General introduction

1.1 Peptidyl asparaginyl ligases

Peptidyl Asx-specific ligases (PALs) are members of plant legumains that process their substrates through transpeptidation. Typically functioning as hydrolases, legumains are also known as asparaginyl endopeptidases (AEPs) because these cysteine proteases hydrolyze peptide bonds at the C-terminal side of an Asn or Asp residue. Like AEPs, PAL enzymes also utilize the sulfhydryl on the active-site Cys residue to cleave an Asn/Asp-Xaa peptide bond in a substrate (acyl donor) to first form an acyl-enzyme thioester intermediate, which instead of being hydrolyzed, is then attacked by the N-terminal amino group of another substrate (acyl acceptor) to form a new Asn/Asp-Yaa peptide bond (**Figure 1.1**). This transpeptidation activity of PALs qualifies them as peptide ligases. As the first PAL found in Clitoria ternatea (also known as butterfly pea), butelase-1 catalyzes both the ligation and cyclization reactions of synthetic peptides and recombinant proteins with high efficiency.[1-10] Although butelase-1 can be isolated from the plant in good quantities, the process is time-consuming.1 Thus, recombinant butelase-1 has also been expressed in E. coli, which improves the availability of the enzyme for potential large-scale applications.[11,12] Since the discovery of butelase-1, researchers have spent considerable efforts to discover new plant-derived peptide ligases with similar activities.[13,14,15] As a result of these efforts, VyPAL2 has recently been identified in a collaborative research program between Professors James Tam, Julien Lescar, and Chuan-Fa Liu's laboratories at Nanyang Technological University.[15] VyPAL2 is a fast enzyme and can be easily expressed in insect cells. These characteristics make VyPAL2 a practically very useful ligase.

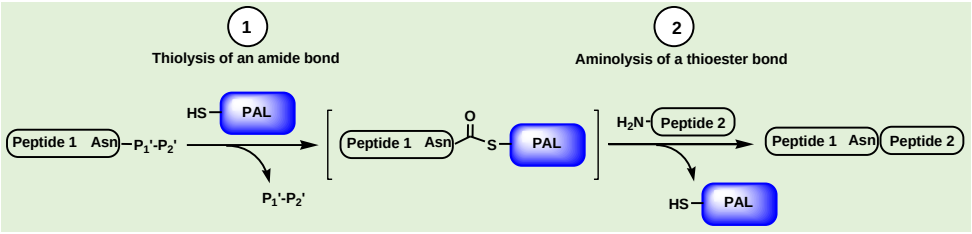


Figure 1.1. The reaction mechanism of PALs. The PALs recognize the Asn-P1′-P2′ tripeptide residue and cleave the amide bond between Asn and P1′ via the sulfhydryl on catalytic Cys. Through the thiolysis reaction on the Asn residue of peptide **1**, the acyl-enzyme thioester intermediate is formed, which is much easier to be attacked by another nucleophile, such as the amine group on peptide **2**, resulting in the ligation of peptides **1** and **2**.

Compared to previously known ligases, such as sortase A,[1,2,16,17] PALs exhibit extremely high catalytic efficiency. This, combined with their mild aqueous conditions as enzymes, brings about tremendous opportunities to develop PAL-based methods for both industrial applications and basic biological research. For example, the flagship PAL, butelase 1, has been reported to be compatible with specific live bacterial and mammalian cell surface labeling.[18,19] Clearly, enzymatic ligation methods using PALs are more desirable than chemical ligation methods because no toxic reagents are used and no undesirable by-products are generated. Moreover, the mild reaction conditions are helpful in maintaining the protein substrates and products in their native folded state.[20]

Numerous reports have described the use of butelase 1-mediated (as well as other PAL-mediated) ligation for protein/peptide labelling and cyclization,[1,3,5-7,9,13,14,21-23] indicating its value as a versatile biotechnological tool. In all these studies, P1-Asn substrates are used in the ligation reactions. In the recent work by Dall et al,[24] Proteomic Identification of Protease Cleavage Sites (PICS) was used to identify the substrate specificity of Arabidopsis thaliana legumain beta (AtLEGβ). The AEP was found to be very efficient at hydrolyzing the

asparaginyl peptide bonds at a broad pH range, with optimum activity usually around pH 6-7.4. On the contrary, pH 4.0-5.5 is the optimum condition for P1-Asp substrate cleavage. Another work by Du et al.[25] also pointed out the effect of pH on the catalytic activity toward the P1-Asp substrate. Together with the previous findings[26-29], these studies suggested AEPs' optimal activity toward the P1-Asp substrates at acidic pH. However, there was no structural or mechanistic study on the ligase-AEPs, i.e., PALs, in catalyzing the ligation reaction at P1-Asp at acidic pH. In spite of this, we have recently established an Asp-specific ligation scheme using PALs for protein modification, and the scheme also allows pH-controlled protein orthogonal ligation at Asn and Asp junctions.[30]

The projects in my thesis are to further exploit the characteristics of PAL enzymes for protein modifications. I have established several new PAL-based methods to modify proteins with various functional moieties with high efficiency. Here, I briefly review protein modification techniques and introduce our work on the use of PALs in the preparation of protein conjugates, including a bi-specific engager protein to mediate the killing of cancer cells by CAR-NK cells.

1.2 Protein modification techniques

Techniques for protein modification are useful for preparing modified proteins as reagents to study protein functions or as diagnostic/therapeutic agents to treat diseases. Tremendous efforts have been made in developing chemo-selective and site-specific methods for protein modification, which are well documented in recent reviews.[31-33] PAL-mediated ligations occur at the amino and/or carboxy ends of a polypeptide backbone, enabling them to install labeling groups specifically onto the termini of protein substrates or effect protein head-to-tail backbone cyclization. Cyclized proteins are resistant to aminopeptidases and carboxypeptidases, and the conformational constraints conferred by cyclization increase their stability against thermal or chemical denaturation as well as stability toward endopeptidases. These often translate into enhanced pharmacokinetic properties in vivo.[34,35] PALs are particularly effective for protein or peptide cyclization, which is not easy to achieve using chemical methods. Protein labeling with a detectable probe, such as a dye, is a valuable technique for biological research.[36] Various materials, such as fluorescent probes, biotin tags, and special chemical functional moieties, can be attached to proteins using innovative chemical and biochemical labeling techniques.[37-39]

Among the many chemical methods reported, native chemical ligation (NCL) is a particularly efficient method for peptidyl cyclization.[40-42] The reaction of NCL (Figure 1.2) is initiated by the nucleophilic attack of the thiol group of an N-terminal cysteine residue on the C-terminal thioester group in a transthioesterification reaction, followed by S-to-N acyl shift to form a new Xaa-Cys peptide bond, joining the peptide's N- and C-termini together.[42] However, the use of NCL for protein cyclization is limited by the requirement for the thioester and thiol functional group. Biochemical approaches based on the use of enzymes typically operate under mild conditions and can be useful for protein cyclization. Peptide ligases such as PatG, PCY1, POPB, and sortase A, as well as the intein-based protein splicing system, have been found capable of mediating peptide/protein backbone cyclization.[43-46] Nonetheless,

certain drawbacks of these approaches hinder their wide applications. Sortase A, for example, requires a long LPXTG peptide motif for recognition and a large enzyme-to-substrate molar ratio for effective catalysis.[42] In contrast, the reactions catalyzed by PALs are nearly traceless, with only an Asn residue left in the cyclic products. The protein substrates of PALs are simple recombinant proteins. Because of the high catalytic efficiency of PALs, a very small enzyme-to-substrate ratio is employed for the cyclization reaction. In this thesis, VyPAL2 and butelase-1 are used together to generate an unusual cyclic protein-drug conjugate consisting of cyclo-affibody and doxorubicin. Such a construct would be difficult to prepare with other conventional methods. This demonstration points to the great promises of PALs for the preparation of protein conjugates of unusual architectures in the future.

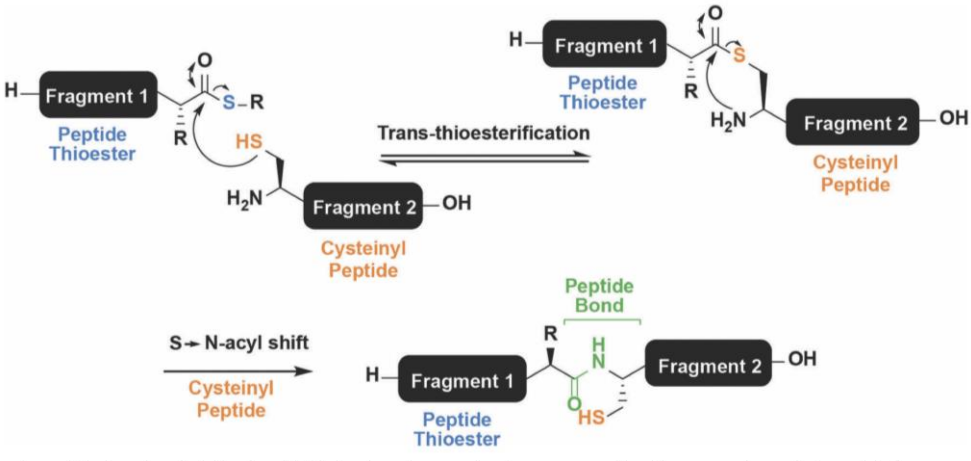


Figure 1.2. The reaction mechanism of native chemical ligation (NCL)[42]. The first step trans-thioesterification is achieved by the free sulfhydryl group on Fragment 2 attacking the thioester group on the C terminus of Fragment 1, forming the novel thioester group and linking Fragment 1 and Fragment 2. The thioester is continuously exchanged with the Cys amine group via S-N acyl shift that will assist to obtain the Cys ligation product with a normal amide backbone.

As for protein labeling, PALs also show advantages over other enzymatic techniques. Many enzymes, including BirA, sortase A, transglutaminase, cutinase, lipoic acid ligase, and

myristoyl transferase, have been adapted for protein labeling, albeit with limitations.[47-55] As described above, PAL-mediated ligations are convenient, efficient, and traceless. All of these desirable features make PALs a versatile tool for protein labeling. In this thesis, I have further advanced the methodology of PAL-mediated ligation for protein labeling. I show that two different modifications can be introduced onto a protein bio-orthogonally using two different PALs. I also show that PALs can use non-peptidic nucleophiles for protein labeling. Finally, I propose a completely traceless protein ligation which combines subtiligation or native chemical ligation, potentially useful for the preparation of native histones containing post-translational modifications.

1.3 Chimeric Antigen Receptor of T cells or NK cells

Chimeric antigen receptor T cells (CAR-T cells) have shown great success in the treatment of certain blood cancers such as acute lymphoblastic leukemia (ALL).[56] However, the production of CAR-T cells for therapeutic use is a cumbersome process that limits their clinical potential.[57] Natural killer (NK) cells are immune effectors with favorable characteristics for practical uses, as they are highly cytotoxic,[56] avoid risks of causing graft-versus-host disease (GVHD),[58] and contain more cytotoxic mechanisms to reduce the risks caused by loss of CAR-specific antigens.[58] These inherent qualities make NK cells a more ideal candidate for immunotherapy. Thus, engineering chimeric antigen receptor-natural killer cell (CAR-NK) with a controllable mechanism of action is a new subject under intense investigation. The detailed CAR-specific antigens of T cells[59] and NK cells[60] are summarized in **Table 1.1** and **1.2**.

Table 1.1. Clinical trials of genetically engineered CAR-T cells.[59]

Target antigen	Disease	ClinicalTrial.gov identifier	Sponsor
CD19	R/R CD19⁺ leukaemia and lymphoma	NCT03166878	Chinese PLA General Hospital
	R/R B-ALL	NCT02746952	Institut de Recherches Internationales Servier
	R/R large B cell or follicular lymphoma	NCT03939026	Allogene Therapeutics
	R/R NHL or B-ALL	NCT03666000	Precision BioSciences
	R/R B cell malignancies	NCT04035434	CRISPR Therapeutics
	R/R B cell haematological malignancies	NCT03229876	Shanghai Bioray Laboratory
	Elderly patients (≥60 years of age) with R/R CD19⁺ B-ALL	NCT02799550	The Affiliated Hospital of the Chinese Academy of Military Medical Sciences
CD19 and CD20 or CD22	R/R leukaemia or lymphoma	NCT03398967	Chinese PLA General Hospital
CD123	Acute myeloid leukaemia	NCT03190278	Cellectis S.A.
BCMA	Multiple myeloma	NCT03752541	Shanghai Bioray Laboratory
	Multiple myeloma	NCT04093596	Allogene Therapeutics
NKG2D ligands	Unresectable metastatic colorectal cancer	NCT03692429	Celyad
Mesothelin	Mesothelin⁺ solid tumours	NCT03545815	Chinese PLA General Hospital

B-ALL, B cell acute lymphoblastic leukaemia; BCMA, B cell maturation antigen; CAR, chimeric antigen receptor; NHL, non-Hodgkin lymphoma; NKG2D, natural killer cell receptor D; R/R, relapsed and/or refractory.

Table 1.2. Clinical trials of genetically engineered CAR-NK cell.[60]

No.	NCT	Start Year	Stage	Tumors	Target	NK source	sponsor locations	CAR structure	Gene transfer
Trials completed									
1	NCT00995137	2009	I	B-ALL	CD19	PB-NK	St. Jude Children's Research Hospital, US	ScFv-CD8αTM-CD137-CD3ζ	mRNA electroporation
2	NCT02944162	2016	I/II	AML	CD33	NK92	PersonGen BioTherapeutics (Suzhou) Co., Ltd., China	ScFv-CD28-CD137-CD3ζ	LV*
Trials actively recruiting									
1	NCT01974479	2013	II	B-ALL	CD19	PB-NK	National University Health System, Singapore	ScFv-CD8αTM-CD137-CD3ζ	mRNA electroporation
2	NCT02742727	2016	I/II	Lymphoma, leukaemia	CD7	NK92	PersonGen BioTherapeutics (Suzhou) Co., Ltd., China	ScFv-CD28-CD137-CD3ζ	electroporation
3	NCT02839954	2016	I/II	Solid tumour	MUC1	NK92	PersonGen BioTherapeutics (Suzhou) Co., Ltd., China	ScFv-CD28-CD137-CD3ζ	LV
4	NCT02892695	2016	I/II	Lymphoma, leukaemia	CD19	NK92	PersonGen BioTherapeutics (Suzhou) Co., Ltd., China	ScFv-CD28-CD137-CD3ζ	LV
5	NCT03056339	2017	I/II	B-lymphoma	CD19	UCB-NK	MD Anderson, US	iCasp9-ScFv-CD28-CD3ζ-IL-15	RV**
6	NCT03383978	2017	I	GBM	HER2	NK92	Johann Wolfgang Goethe University Hospital, Germany	ScFv-CD28-CD3ζ	LV
7	NCT03415100	2018	I	Metastatic solid tumour	NKG2DL	PB-NK	The Third Affiliated Hospital of Guangzhou Medical University China	ScFv-CD8αTM-CD3ζ; ScFv-CD8αTM-DAP12	mRNA electroporation
8	NCT03656705	2018	I	NSCLC	NR	NK92	Xinxiang medical university, China	NR	RV/LV
9	NCT03940833	2019	I/II	R/R multiple myeloma	BCMA	NK92	Asclepius Technology Company Group (Suzhou) Co., Ltd., China	NR	LV
10	NCT03941457	2019	I/II	Pancreatic Cancer	ROBO1	NK92	Asclepius Technology Company Group (Suzhou) Co., Ltd., China	NR	LV
11	NCT03940820	2019	I/II	Solid tumour	ROBO1	NK92	Asclepius Technology Company Group (Suzhou) Co., Ltd., China	NR	LV
12	NCT04245722	2020	I	B-cell lymphoma, CLL	CD19	iPSC (FT596)	Fate Therapeutics, San Diego, USA	scFv-NKG2D-2B4-CD3ζ-IL-15/R-hnCD16	LV
Trials not yet recruiting									
1	NCT03692767		Early I	Refractory B-cell lymphoma	CD22	Unknown	Allife Medical Science and Technology, Beijing, China		
2	NCT03690310		Early I	Refractory B-cell lymphoma	CD19	Unknown	Allife Medical Science and Technology, Beijing, China		
3	NCT03692637		Early I	Epithelial ovarian cancer	Mesothelin	PB-NK	Allife Medical Science and Technology, Beijing, China		
4	NCT03692663		Early I	Castration-resistant prostate Cancer	PSMA	Unknown	Allife Medical Science and Technology, Beijing, China		
5	NCT03824964		Early I	Refractory B-cell lymphoma	CD19/CD22	Unknown	Beijing Cancer Hospital, Beijing, China		

* LV: lentivirus
** RV: retrovirus

A folate-fluorescein isothiocyanate (FITC) conjugate has been reported as a bispecific adaptor molecule for controlled folate-receptor CAR-T cell therapy (Figure 1.3).[61, 62] However, the folate receptor is expressed in many normal tissues, and the endogenous folate might compete with the conjugate, potentially causing problems such as side effects and low efficacy. Undoubtedly, the artificially designed EGFR-targeting affibody Z_{EGFR}-Fc fusion protein has the advantage of being more selective for the tumor antigen.[63, 64] In this thesis, I have synthesized a quadruple FITC-labeled affibody-Fc fusion protein using a two-step PAL ligation scheme developed from this study. When NK cells are engineered with a CAR directed at the FITC component, we show that the FITC-Z_{EGFR}-Fc protein can act as a bi-specific engager to direct the killing of EGFR-expressing cancer cells by the CAR-NK cells. This study demonstrates the power of PAL-mediated ligation in the preparation of large, complex protein conjugates for potential use in cell-based immunotherapy.

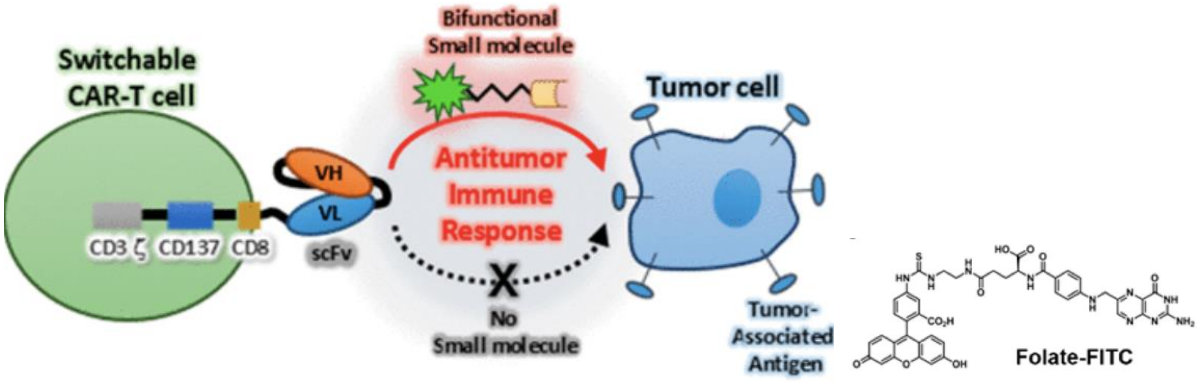


Figure 1.3. Controlled folate-receptor CAR-T cell therapy.[61] This strategy is dependent on the cell-cell interaction between CAR-T expressed with a FITC-specific protein single-chain variable fragment (scFv) and the cancer cells over-expressed with folate receptor on the surface. The compound Folate-FITC is serving as a glue that will capture both CAT-T cells (via scFv binding to FITC) and cancer cells (via folate binding to folate receptor). Therefore, the folate-FITC therapy functioned as a switch to the controlled CAR-T cell therapy.

1.4 Summary

In general, enzymatic methods for protein modification offer distinct advantages over chemical methods. The recently discovered peptidyl Asx-specific ligases (PALs) are particularly useful for applications described herein. In this thesis, I present an orthogonal ligation scheme based on the different substrate specificities of two PALs (butelase 1 and VyPAL2), allowing for dual labeling of proteins for use as theranostic agents. My work also extends the application of PAL ligation from using peptidyl nucleophile substrates to small α-effect nucleophiles, allowing for protein C-terminal labeling with functionalized hydrazide derivatives. This method can be used to prepare complex protein-drug conjugates containing multiple functional moieties. Finally, I also propose a chemoenzymatic scheme based on the use of PALs and NCL or subtiligase for protein ligation at non-asparaginyl peptide bonds. This would provide a traceless ligation method for protein total synthesis. In summary, the work presented in my thesis further demonstrates the versatility of PALs as powerful biotechnological tools for protein engineering.

Chapter 2: Engineering protein theranostics using bio-orthogonal asparaginyl peptide ligases

2.1 Introduction

By combining therapy with the specific diagnostic information of a disease target, theranostics, a new type of medicine combining the procures of diagnosis and treatment, promise to optimize the efficacy and safety of precision medicine.[65-67] This has led to a tremendous interest in the development of theranostic agents for the treatment of cancer.[68, 69] In addition to the use of nanomedicine platforms for the development of theranostics,[70-93] a molecular-based approach involves the attachment of imaging agents and cytotoxic drugs to cancer-targeting proteins and antibodies.[74, 75] In particular, small protein ligands, such as antibody fragments and mimetics, offer the advantages of low production cost, good tissue penetration, and easy maneuverability for designing end products with defined chemical composition. However, a major challenge in developing protein-based theranostic agents lies in the conjugation of the protein ligand with the imaging and treatment moieties.[10] Clearly, the conjugation strategy should be able to introduce at least two modifications onto a protein substrate. Although numerous chemical techniques have been developed for protein labeling, [32, 76] a simple strategy that allows for two consecutive site-specific modifications to be performed on a straight recombinant protein has yet to be developed.

Owing to their high specificity and mild operating conditions, biosynthetic methods that modify proteins through special recognition tags are attractive alternatives to chemical methods.[77] The main advantage of these tag-mediated protein labeling methods is that the tags are themselves a peptide segment or protein domain and thus can be genetically fused to the protein of interest (POI). Of these methods, those that are based on peptide ligases are of utmost

interest. Peptide ligases catalyze the formation of new peptide bonds between ligation partners, which makes them particularly useful bioconjugation tools for protein-based theranostics. Notable examples of peptide ligases include subtiligase[78-82], sortase A[16, 83-86], and butelase-1, [1, 2, 8, 18, 19] which are all tag-recognizing enzymes and can label proteins specifically at the terminal ends. Subtiligase is an artificially engineered ligase that uses an ester or thioester tag for protein labelling.[78-82] Sortase A requires a 5-residue tag, LPETG, and catalyzes transpeptidation at the Thr residue.[16, 83-86] However, the use of the 5-residue tag notwithstanding, the enzymatic activity of sortase A is very low. Butelase-1 is a peptidyl asparaginyl ligase or PAL. So far, the most powerful peptide ligases have been found in the PAL family, and the most efficient PAL is butelase-1. Structurally, butelase-1 is a member of the commonly known asparaginyl endopeptidase (AEP) or legumain family.[87, 88] Depending on the pH or substrate, certain AEPs are also found to display PAL activities.[11, 13-15, 87-95] Butelase-1 is unique in that it functions almost as a pure PAL with no protease activity at weakly acidic to weakly basic pH. It has been shown to catalyze protein and peptide ligation with a high specificity and efficiency.[1 , 2, 8, 18, 19] Like all PALs, butelase-1 recognizes a short tripeptide tag, such as NHV, and cleaves the peptide bond at Asn to rejoin it with the amino terminal residue of another peptide. Thus, only an Asn residue is left in the ligation product, making butelase-mediated ligation (BML) nearly traceless. This is in significant contrast to most of the above-mentioned biosynthetic methods, which leave a large "scar" in the modified protein.[77] Recently, VyPAL2, another plant legumain from the Viola yedoensis family, was identified as a highly active PAL.15 Its catalytic efficiency was 274,325 $M^{-1}\cdot s^{-1}$ in the cyclization of a model peptide, making it one of the fastest PALs reported to date.[15] In addition, the proenzyme of VyPAL2 can be readily expressed in insect cells and can be self-processed at acidic pH to yield the active enzyme.15 These features make VyPAL2 a very attractive ligase for protein labeling.[96] As asparaginyl transpeptidases, both butelase-1 and VyPAL2 use their

active-site cysteinyl thiol group to cleave an Asn-Xaa peptide bond in their acyl-donor substrate, forming an acyl-enzyme thioester intermediate. Instead of being attacked by a water molecule for hydrolysis, as in the case of an asparaginyl endopeptidase, the acyl-enzyme thioester intermediate undergoes aminolysis by a peptidic nucleophile, which results in the formation of a new asparaginyl peptide bond between the acyl-donor substrate and the nucleophile substrate. Intriguingly, there seem to be noticeable differences in substrate specificity between VyPAL2 and butelase-1. VyPAL2 has a relatively low activity towards the tripeptide NHV, which, on the other hand, is one of the preferred recognition motifs of butelase-1.[1, 15] In addition, a nucleophile peptide with a Phe at the P2" position is a weak substrate for butelase-1,[1] but it is favored by VyPAL[2.15] We reasoned that these differential substrate specificities might provide sufficient orthogonality for a tandem ligation strategy for protein dual labeling.

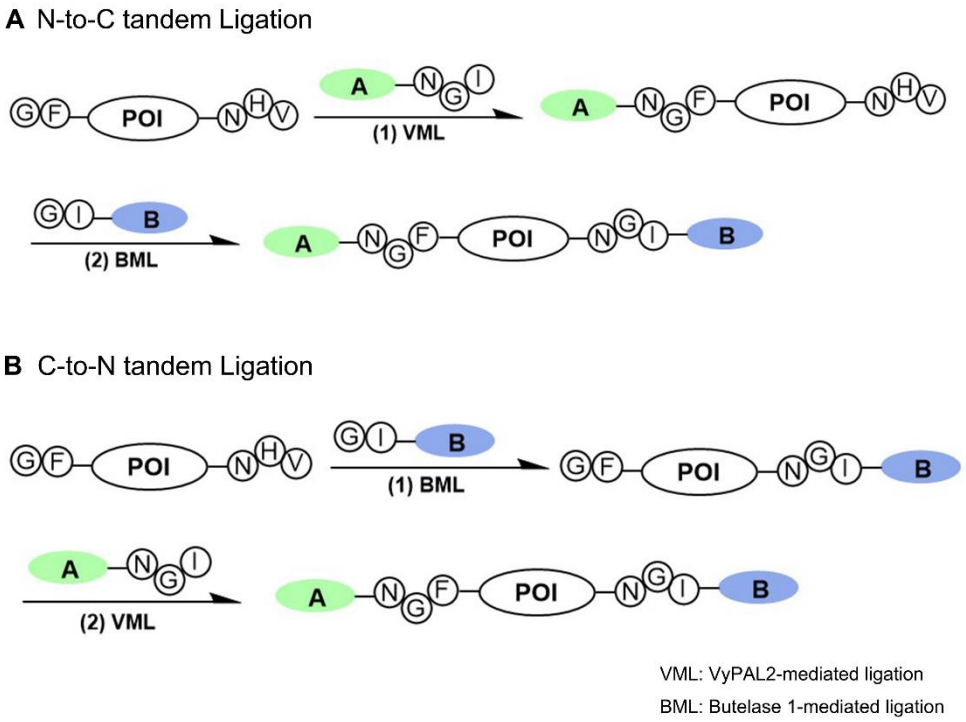
Several protein dual modification methods involving the use of peptide ligases have been reported.[3, 10, 97-100] For example, consecutive protein modifications were achieved chemoenzymatically by combining chemoselective conjugation and sortase A- or butelase-1-mediated ligation.[3, 97, 98] Two sortases of different substrate specificities were used to label a protein at both the N- and C-termini.[100] Butelase-1 was also used together with sortase A for protein dual labeling in a three-step scheme.[3] The two enzymes were also used for one-pot dual labeling of an antibody at the respective C-terminal ends of light and heavy chains.[10] These last two schemes are bio-orthogonal, taking advantage of the distinct substrate specificity of two completely different ligases. However, as discussed above, owing to its extremely slow kinetics and relatively long recognition tag, the use of sortase A has its inherent limitations. Recently, an interesting method was reported that allowed two consecutive ligation reactions on the same protein substrate from the C- to N-terminus direction.[99] However, it should be noted that this

scheme is semi-orthogonal because it requires the protection of the protein's N-terminal amine by a TEV recognition sequence during the first ligation step to avoid the cyclization or self-ligation of the protein substrate.[99] Here, we reported a bio-orthogonal scheme using two asparaginyl peptide ligases – butelase-1 and VyPAL2 – which allows for tandem ligation on the same protein in either the N-to-C or C-to-N direction, leading to its dual labeling at the C- and N-terminal ends (Scheme 1). No protection on the protein substrate is required when performing the first ligation step, although butelase-1 and VyPAL2 are both asparagine-specific. Thus, a distinct advantage of bio-orthogonal ligation is the use of mild enzymatic reactions under aqueous conditions, which are compatible with biologics, such as proteins, antibodies and live cells.

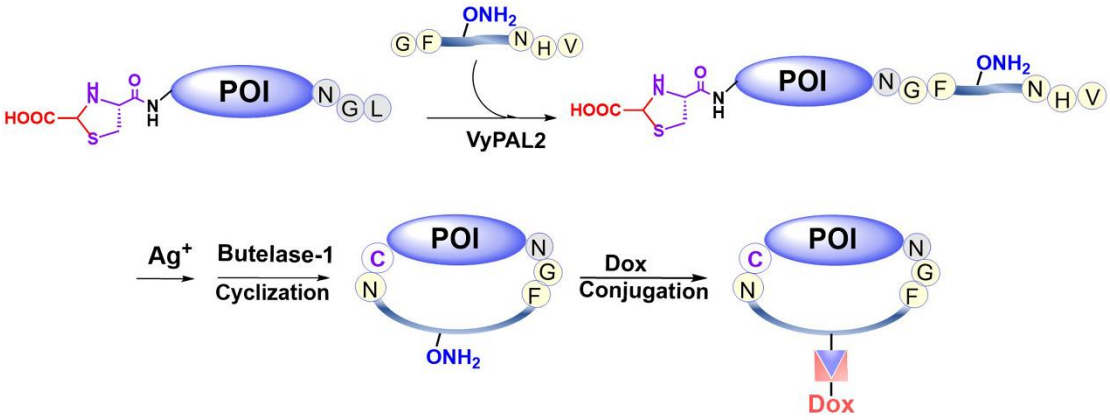
In addition to N- and C-terminal directed protein dual labeling, our bio-orthogonal tandem ligation strategy can also be used to prepare a cycloprotein-drug conjugate or cPDC (Scheme 2). This involves the use of a synthetic intervening peptide designed to join the two termini of a protein. The peptide is trifunctional, containing an N-terminal GF-dipeptide nucleophile substrate for VyPAL, a C-terminal NHV tripeptide motif as the acyl donor substrate for butelase-1 and an internal aminooxy functionality for oxime conjugation, which would allow consecutive PAL-mediated ligation, cyclization, and doxorubicin attachment (Scheme 2). Given the expected thermal and metabolic stability of cycloproteins, cycloprotein conjugates are interesting candidates for theranostics.

Using an EGFR-binding affibody[64, 65] as the model protein, we demonstrated the feasibility of our tandem ligation strategy for the preparation of dually labeled proteins and cycloprotein-drug conjugates. Because butelase-1 and VyPAL2 are the two most powerful ligases, such a bio-orthogonal tandem ligation strategy would offer an ideal solution to the

challenging problem of manufacturing protein-based theranostics and other biologics with unusual architectures and functionalities.



Scheme 1. Bi-directional dual protein labeling by bio-orthogonal tandem ligation in N-to-C (A) or C-to-N (B) direction using butelase-1 and VyPAL2.



Scheme 2. Preparation of a cyclic affibody-drug conjugate by PAL-mediated tandem ligation-cyclization and drug conjugation.

2.2 Results and Discussion

2.2.1 Differential substrate specificity of butelase-1 and VyPAL2 analyzed by kinetic studies

PAL enzymes have been extensively used for protein single site labeling and macrocyclization. However, the use of two PALs with different substrate specificities for bio-orthogonal and dual ligation remains unexplored. Previous studies have revealed noticeable differences in substrate specificity between butelase-1 and VyPAL2.[1, 15] To evaluate these differences quantitatively, we first studied the kinetics of VyPAL2 and butelase-1 towards peptide 1, which has a C-terminal NHV tripeptide motif (**Table 2.1**).

Table 2.1. Kinetics of VyPAL2- and butelase-1-mediated intermolecular ligation.

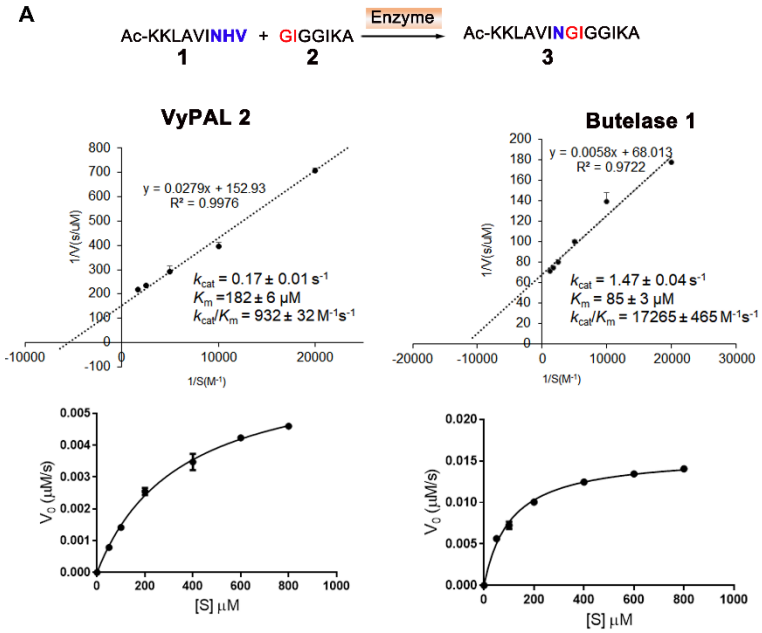
Electrophile substrate	Nucleophile substrate	Enzyme	k_{cat} [s^{-1}]	K_m [μM]	k_{cat}/K_m [$M^{-1}s^{-1}$]
Ac-**KKLAVINHV** **1**	GIGGIKA **2**	VyPAL2	0.17 ± 0.01	182 ± 6	932 ± 32
		Butelase-1	1.47 ± 0.04	85 ± 3	17265 ± 465
YKANGL **4**	**GFGGIKA** **5**	VyPAL2	8.29 ± 0.48	424 ± 26	19559 ± 164
		Butelase-1	1.55 ± 0.01	365 ± 3	4256 ± 52
Ac-**KKLAVINGF** **7**	GIGGIKA **2**	VyPAL2	1.10 ± 0.03	155 ± 15	7219 ± 513
		Butelase-1	0.46 ± 0.04	175 ± 1	2652 ± 218

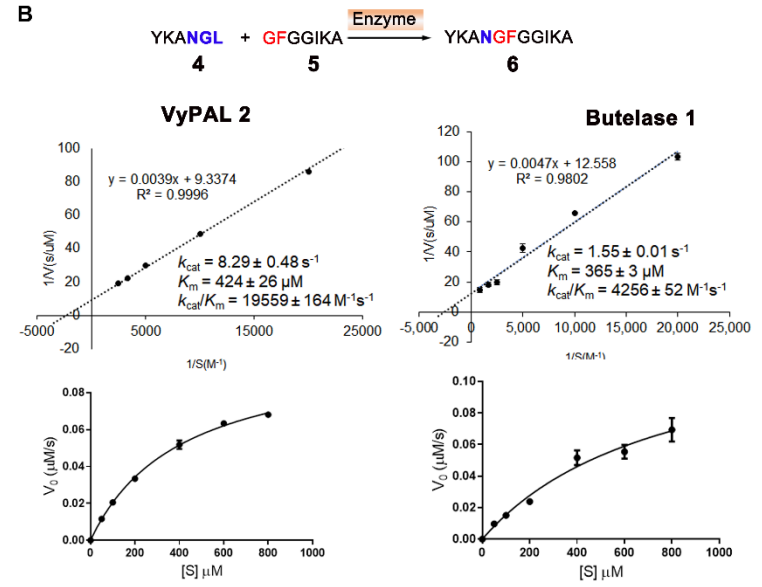
Note: The kinetic parameters for each reaction are for the substrate with the sequence in bold. kcat determines the reaction rate when the enzyme is fully occupied at a saturating concentration of the substrate. K_m is defined as the substrate concentration at which the reaction rate is half of its maximal value. k_{cat}/K_m is a second-order constant rate that determines the reaction rate when the enzyme is mostly free at a very low concentration of the substrate.

2.2.2 Kinetics of VyPAL2 and butelase-1 toward the NHV-ending acyl peptide substrate (peptide 1) and the GF-starting nucleophile peptide substrate (peptide 5)

To demonstrate and quantify the differential specificities of two PAL enzymes, we evaluated the kinetics of peptide ligations between peptides **1** (Ac-KKLAVINHV) and **2** (GIGGIKA) using butelase-1 and VyPAL2, respectively (**Figure 2.1**). The ligation reaction between **1** and **2** yielded peptide **3** Ac-KKLAVINGIGGIKA (ESI-MS: 1423.09 obsv, 1422.28 calc). The reactions were performed by adding different concentrations of Ac-KKLAVINHV 1 (50, 100, 200, 400, 600, and 800 μM) with nucleophile **2** GIGGIKA (kept at 1 mM). Then, 40 nM of VyPAL2 or 10 nM of butelase-1 was added to the above reaction for 30 min at pH 6.5. Aliquots of the reaction mixtures were analyzed by HPLC. The results showed that the catalytic efficiency of butelase-1 was approximately 18-fold higher than that of VyPAL2 toward the "NHV" substrate peptide **1**. Similarly, the catalytic activities of VyPAL2 and butelase-1 towards the GF-nucleophile peptide GFGGIKA **5** were measured. We performed the reactions using different concentrations of **5** (50, 100, 200, 600, and 800 μM) with the acyl-side substrate YKAINGL 4 (kept at 1.5 mM). Butelase-1 and VyPAL2 were added at 50 nM and 13 nM, respectively. The ligation reaction between peptides **4** and **5** yielded peptide **6** YKAINGFGGIKA (ESI-MS: 1124.91 obsv, 1124.15 calc). We found that VyPAL2 was approximately 5 times more efficient than butelase-1 towards the nucleophile substrate GFGGIKA 5 (**Figure 2.1**). Therefore, our results confirm earlier observations that, while the NHV sequence is an excellent substrate of butelase-1, it is a poor substrate of VyPAL2. Furthermore, these results indicate that the N-terminal GF dipeptide is a good substrate of VyPAL2 but not a very good substrate of butelase-1. Similarly, the kinetics of butelase-1 and VyPAL2 towards another acyl donor substrate, peptide **7** (Ac-KKLAVINGF), in ligation with GI-peptide **2** were also determined. The reactions were performed using varying concentrations of **7** and a constant concentration of **2** in the presence of butelase-1 (100 nM) or VyPAL2 (50

nM). We found that the NGF motif was 6.5-fold less active than the NHV motif in butelase-mediated ligation and that VyPAL2 was about 2.6-fold more efficient than butelase-1 towards peptide **7**.


A
Ac-KKLAVINHV + GIGGIKA
1 2
Enzyme
Ac-KKLAVINGIGGIKA
3
VyPAL 2
y = 0.0279x + 152.93
R² = 0.9976
kcat = 0.17 ± 0.01 s-1
Km =182 ± 6 µM
kcat/Km = 932 ± 32 M-1s-1
1/V(s/uM)
1/S(M-1)
Butelase 1
y = 0.0058x + 68.013
R² = 0.9722
kcat = 1.47 ± 0.04 s-1
Km = 85 ± 3 µM
kcat/Km = 17265 ± 465 M-1s-1
V0 (µM/s)
[S] µM


B
YKANGL + GFGGIKA
4 5
Enzyme
YKANGFGGIKA
6
VyPAL 2
y = 0.0039x + 9.3374
R² = 0.9996
kcat = 8.29 ± 0.48 s-1
Km = 424 ± 26 µM
kcat/Km = 19559 ± 164 M-1s-1
1/V(s/uM)
1/S(M-1)
Butelase 1
y = 0.0047x + 12.558
R² = 0.9802
kcat = 1.55 ± 0.01 s-1
Km = 365 ± 3 µM
kcat/Km = 4256 ± 52 M-1s-1
V0 (µM/s)
[S] µM

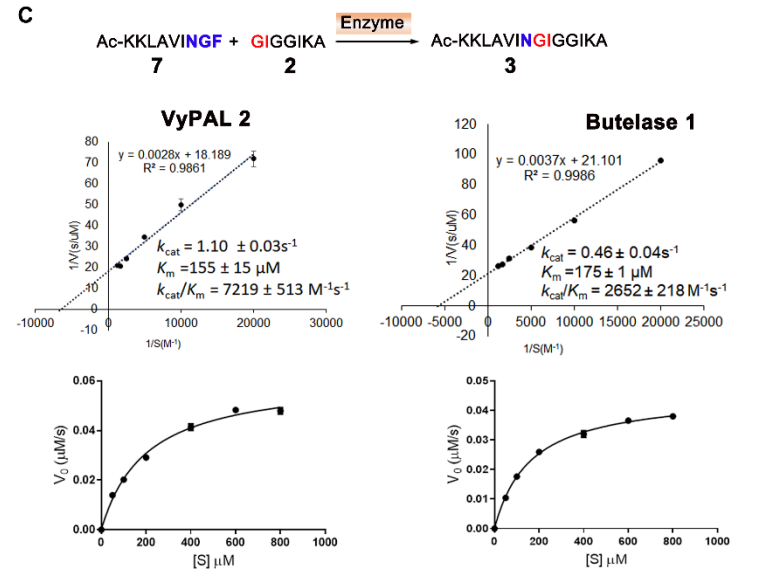

C
Ac-KKLAVINGF + GIGGIKA
7 2
Enzyme
Ac-KKLAVINGIGGIKA
3
VyPAL 2
y = 0.0028x + 18.189
R² = 0.9861
kcat = 1.10 ± 0.03s-1
Km =155 ± 15 µM
kcat/Km = 7219 ± 513 M-1s-1
1/V(s/uM)
1/S(M-1)
Butelase 1
y = 0.0037x + 21.101
R² = 0.9986
kcat = 0.46 ± 0.04s-1
Km =175 ± 1 µM
kcat/Km = 2652 ± 218 M-1s-1
V0 (µM/s)
[S] µM

Figure 2.1. Determination of kinetic parameters of VyPAL2- and butelase-1-catalyzed intermolecular ligations by Michaelis-Menten and Lineweaver-Burk plotting. A) Peptide **1** containing the C-ter NHV sequence was used at varying concentrations (50–800 μM) to react with peptide **2** at a constant concentration (1 mM); B) Peptide **4** at a constant concentration (1.5 mM) was reacted with the GF-peptide **5** at varying concentrations (50–800 μM). C) Peptide **2** at a constant concentration (1.5 mM) was reacted with the -NGF peptide **7** at varying concentrations (50–800 μM). The ligation product **3** or **6** was confirmed by ESI-MS. The reaction rates were calculated from the consumption of the limiting substrate **1**, **5**, or **7**. Initial rates (V_0) at different concentrations of the limiting substrate were used for Michaelis-Menten curve plotting. For kinetic parameter calculation, a Lineweaver-Burk plot was used for the analysis.

The nucleophile substrate used for kinetic studies was peptide **2**, which contains an N-terminal GI dipeptide motif, and was kept at a constant concentration. Reverse-phase analytical HPLC was used to monitor and quantify the ligation reaction. The results indicated that butelase-1 was approximately 18.5 times more catalytically active towards acyl peptide substrate 1 than VyPAL2 (as shown in **Table 2.1** and **Figure 2.1**). Kinetic studies were also conducted on the two ligases towards a GF-starting nucleophile substrate, peptide **5**, where acyl substrate peptide **4** was kept at a constant concentration. Peptide **4** contains NGL at the C-terminus, which is a favorable motif for both VyPAL2 and butelase-1. It was observed that VyPAL2 was 4.6-fold more efficient than butelase-1 towards GF-peptide substrate **5** (as shown in **Table 2.1** and **Figure 2.1**). Additionally, kinetic studies were performed on another acyl donor substrate, peptide **7**, and it was found that butelase-1 exhibited 6.5-fold lower catalytic activity towards this NGF peptide compared to NHV peptide **1**. The difference in catalytic efficiency between VyPAL2 and butelase-1 towards NGF substrate **7** was approximately 2.7-fold.

An analysis of the Butelase-1 and VyPAL2 structures helps to explain their differential substrate specificities, which are likely due to differences in the S1′ and S2′ substrate binding pockets of the two enzymes.[11, 15] For Butelase-1, a glycine residue (Gly167) and a valine residue (Val170) occupy the central positions of its S1′ and S2′ pockets, respectively. However, for VyPAL2, the same positions are occupied by alanine (Ala174) and lysine (Lys177), respectively. The small Gly residue in the S1′ pocket of Butelase-1 makes it possible for the enzyme to tolerate a variety of amino acid residues at the P1′ position of its substrates. The Val residue in its S2′ pocket makes it prefer substrates with a bulky aliphatic side chain at the P2′ position for van der Waals interactions. On the other hand, VyPAL2 has an Ala in its S1′ pocket, which is larger than Gly. This may hinder the binding of a substrate with a larger amino acid (e.g., His) at the P1′ position. The long aliphatic side chain and the positive charge of the Lysine residue in the S2′ pocket of VyPAL2 may explain why the enzyme can accept a P2′ (or P2″) residue like Ile or Leu, which has a large aliphatic side chain for attraction by van der Waals forces, or Phe, which can interact with S1′-Lys through cation-π interactions. In summary, the kinetic studies confirm the differential activities of Butelase-1 and VyPAL2 towards certain substrate sequences, providing strong support for a two-PAL, bio-orthogonal tandem ligation scheme for protein dual labeling.

2.2.3 Bio-orthogonal protein dual-labeling using VyPAL2 and butelase-1

Next, we proceeded to use Butelase-1 and VyPAL2 to dual-label the affibody through tandem enzymatic ligation (**Figure 2.2**). Considering the specificity of the two enzymes, an N-terminal GF dipeptide tag and a C-terminal NHV tripeptide tag were introduced onto Z_{EGFR} to obtain **8**. A new fluorescein-peptide **9** with a C-terminal NGI motif was prepared. We also synthesized peptide **11**, GIGGFKGG-klaklakklaklak, of which the all-D amino-acid sequence

is the mitochondrion-lytic KLA peptide.[101] Phe-Lys is a cathepsin B-sensitive linker[102] and can be cleaved in the lysosomes to release the KLA peptide. Peptides **9** and **11** were used to label the respective N- and C-termini of the Z_{EGFR} **8**. Sequential bio-orthogonal ligations were conducted in both the N-to-C (**Figure 2.2A**) and C-to-N (**Figure 2.2C**) directions. For N-to-C tandem ligation, VyPAL2 was used at the first ligation step, while Butelase-1 was used at the second step (**Figure 2.2A**). C-to-N sequential ligations were performed using the two enzymes in the reverse order (**Figure 2.2C**). Whichever the ligation direction, the same final product **12** was obtained, as characterized by ESI-MS (obs: 10776, calc: 10773). The reactions at each step of the two schemes were remarkably clean with good conversion yields. In N-to-C ligation, 50 μM of Z_{EGFR} **8** and 250 μM of peptide **9** were first reacted in the presence of 150 nM of VyPAL2 at 37 °C for 30 min. The reaction gave approximately 80% of product **9** based on HPLC analysis. After HPLC purification and refolding, the second step was performed by incubating 50 μM of **10** and 250 μM of peptide **11** with 100 nM of Butelase-1 for 20 min at 37 °C. The second step resulted in a conversion yield of around 70% (**Figure 2.2B**).

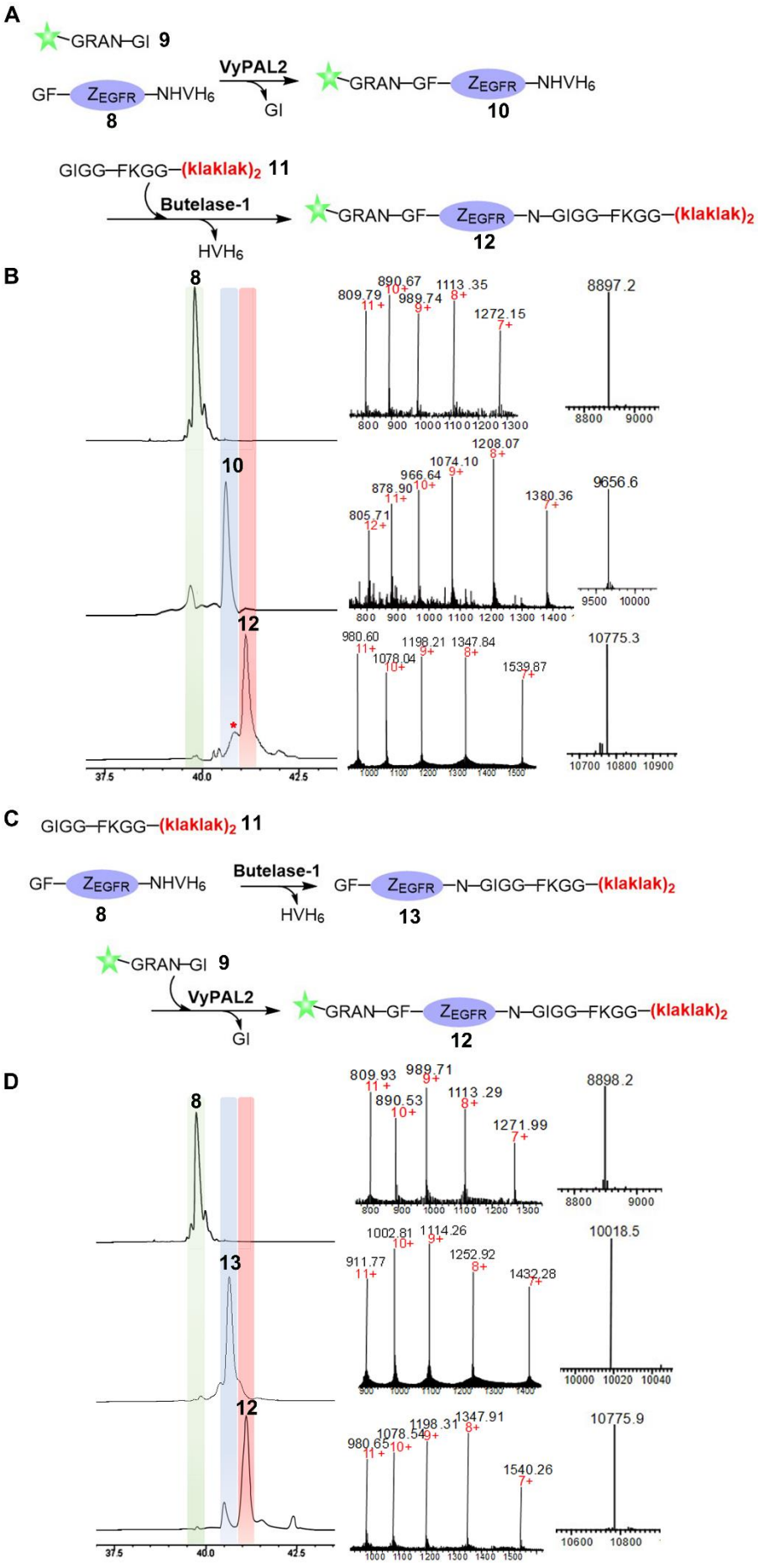

A
GRAN–GI 9
VyPAL2
GF–ZEGFR–NHVH6
8
GI
GRAN–GF–ZEGFR–NHVH6
10
GIGG–FKGG–(klaklak)2 11
Butelase-1
HVH6
GRAN–GF–ZEGFR–N–GIGG–FKGG–(klaklak)2
12
B
8
10
12
809.79 11+
890.67 10+
989.74 9+
1113.35 8+
1272.15 7+
8897.2
805.71 12+
878.90 11+
966.64 10+
1074.10 9+
1208.07 8+
1380.36 7+
9656.6
980.60 11+
1078.04 10+
1198.21 9+
1347.84 8+
1539.87 7+
10775.3
C
GIGG–FKGG–(klaklak)2 11
GF–ZEGFR–NHVH6
8
Butelase-1
HVH6
GF–ZEGFR–N–GIGG–FKGG–(klaklak)2
13
GRAN–GI 9
VyPAL2
GI
GRAN–GF–ZEGFR–N–GIGG–FKGG–(klaklak)2
12
D
8
13
12
809.93 11+
890.53 10+
989.71 9+
1113.29 8+
1271.99 7+
8898.2
911.77 11+
1002.81 10+
1114.26 9+
1252.92 8+
1432.28 7+
10018.5
980.65 11+
1078.54 10+
1198.31 9+
1347.91 8+
1540.26 7+
10775.9

Figure 2.2. Bio-orthogonal protein dual labeling using VyPAL2 and butelase-1. (A) N-to-C tandem ligation scheme. Fluorescein-peptide **9** was first ligated to the N terminus of Z_{EGFR} **8** via VML to give **10** which was then ligated with peptide **11** at C terminus via BML to give **12**; (B) HPLC and LC-MS analysis of N-to-C ligation. The ligation products **8**, **10**, **12** were purified by reverse-phase HPLC and analyzed via ESI-MS; (C) C-to-N tandem ligation scheme. Mitochondrion-lytic peptide **11** is conjugated at C terminus of Z_{EGFR} **8** to give **13** via BML and then the fluorescein-peptide **9** is ligated to the N terminus of **13** to produce **12**; (D) HPLC and LC-MS analysis of N-to-C ligation. The ligation products **8**, **13**, **12** were purified by HPLC and analyzed by ESI-MS (**8:** calcd 8896.8, obsvd 8897.2; **10:** calcd 9652.6, obsvd 9656.6; **12**: calcd 10774.3, obsvd 10775.3 or 10775.9; **13**: calcd 10017.8, obsvd 10018.5). Details of by-product characterization are provided in **Table 2.2**.

Table 2.2. Summary of materials used in the tandem ligation reactions.

	N-to-C				**C-to-N**			
	Step 1		**Step 2**		**Step 1**		**Step 2**	
Protein	Protein **8**	50 μM	Protein **10**	50 μM	Protein **8**	50 μM	Protein **13**	50 μM
Peptide	Peptide **9**	250 μM	Peptide **11**	250 μM	Peptide **11**	250 μM	Peptide **9**	250 μM
Enzyme	VyPAL2	150 nM	Butelase 1	100 nM	Butelase 1	100 nM	VyPAL2	150 nM

2.2.4 Characterization of a by-product in the N-to-C tandem ligation using VyPAL2 and butelase-1

A small amount of a by-product (~10%) was found in the second step (**Figure 2.2 B** and **2.3**). This was because, although the newly formed NGF motif in **10** was not a favored substrate of butelase-1, its 6.5-fold lower reactivity than the NHV motif (**Table 2.1**) meant that it could still be affected in BML, which resulted in the cleavage of the N-G peptide bond for transpeptidation with **11**. In C-to-N ligation, BML was first performed by mixing 50 μM of

Z_{EGFR} **8** and 250 μM of peptide **11** with 100 nM of butelase-1 at 37 °C for 30 min to obtain **13** in approximately 85% based on HPLC analysis. Then, VML was performed by incubating 50 μM of purified **13** and 250 μM of peptide **9** with 100 nM of VyPAL2 at 37 °C for 30 min. The reaction gave **12** in ~70% yield (**Figure 2.2 D**). Notably, the free $^{\alpha}$N-amino group of the N-terminal Gly residue in the affibody was resistant to Butelase-1 at the first ligation step, confirming that the GF dipeptide motif is a relatively poor nucleophile substrate of Butelase-1. Because the two PALs require only a short NXY tripeptide as the recognition tag and ligate at the Asn residue, only minimal traces are left in the modified proteins. These results demonstrate the robustness and neatness of our sequential bio-orthogonal ligation method for protein dual labeling. Detailed information on the reaction protocols and conditions is summarized in **Table 2.2**.

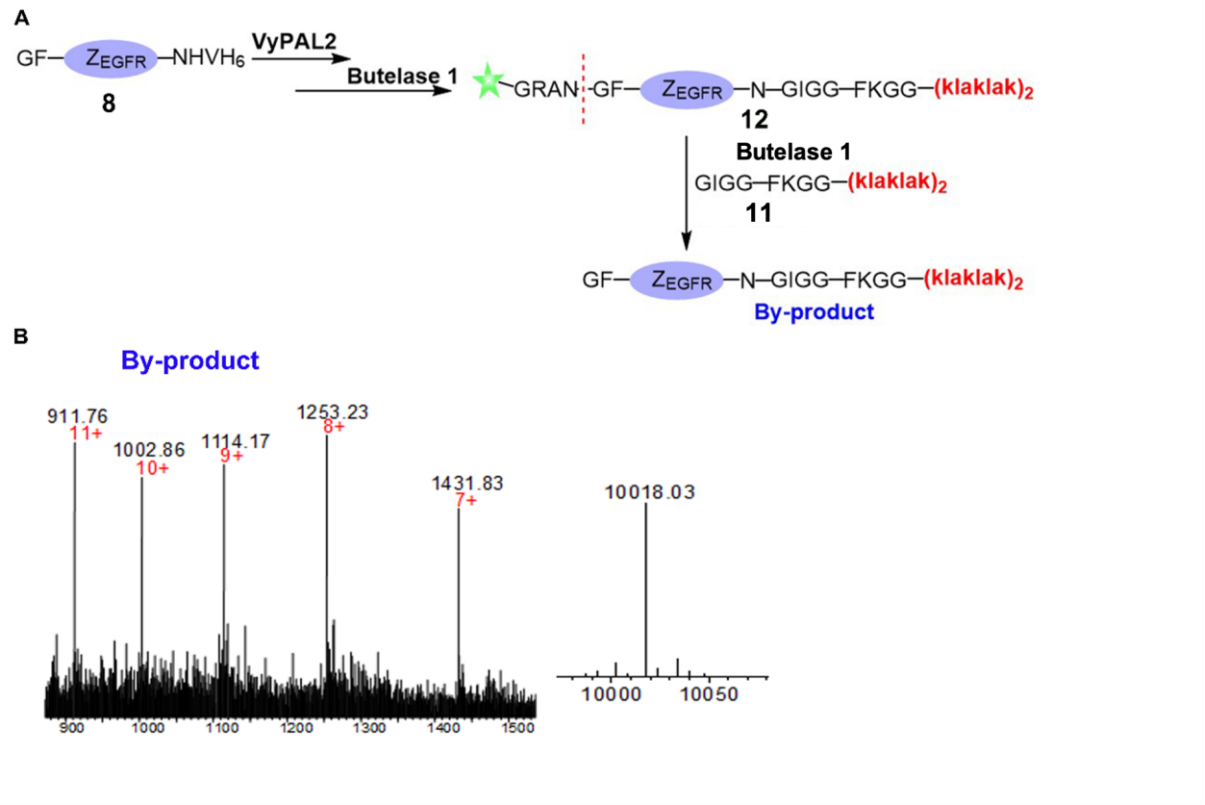


Figure 2.3. Characterization of a minor by-product in the 2nd step of the N-to-C tandem ligation. A) Schematic of the by-product formation resulting from butelase1-mediated nucleophilic

attack by peptide **11** to the N-G peptide bond in Z_{EGFR} **12**; B) ESI-MS characterization of the isolated by-product (calculated mass: 10018; observed mass: 10018).

2.2.5 Butelase-1 and VyPAL2 mediated one-pot reaction

First, we attempted one-pot reactions with simultaneous VML and BML. To ensure good orthogonality, peptide **7** (which has a C-terminal NGF tripeptide motif) and peptide **14** (with an N-terminal HV dipeptide motif) were chosen for N- and C-terminal labeling, respectively. We performed the reaction using 50 μM of affibody **8**, 250 μM of peptide **14**, and 250 μM of peptide **7** in the presence of 250 nM of butelase-1 and 160 nM of VyPAL2. The protein and peptide starting materials were pre-mixed, and then a mixture of butease-1 and VyPAL2 was added. The reaction mixture was incubated at 37 °C for 45 min. As a result, the predominant reaction was found to be inter-peptide ligation. While a significant amount of the N-terminal labeling product resulting from VML was formed, the desired end-product was not detected (data not shown). Next, the amount of peptide **7** (500 μM; peptide **14** was kept at 250 μM; ratio of peptide **14**: peptide **7** = 1:2) in the reaction was increased (**Figure 2.4**). After 45 min of reaction, only the intermediate product **16** was found, and no end-product **15** was formed, or an amount too small for characterization. Again, a large amount of inter-peptide ligation product **17** was formed (**Figure 2.4**). VML for N-terminal protein ligation appears to be much faster than BML. Although adding more butelase-1 would help accelerate C-terminal ligation, this would also increase the rate of inter-peptide ligation. Similarly, decreasing the amount of VyPAL2 may help to balance the two ligations, but this will cause the overall reaction to be too slow to be practically viable. However, inter-peptide ligation is unavoidable. Therefore, a one-pot reaction of simultaneous VML and BML is not recommended.

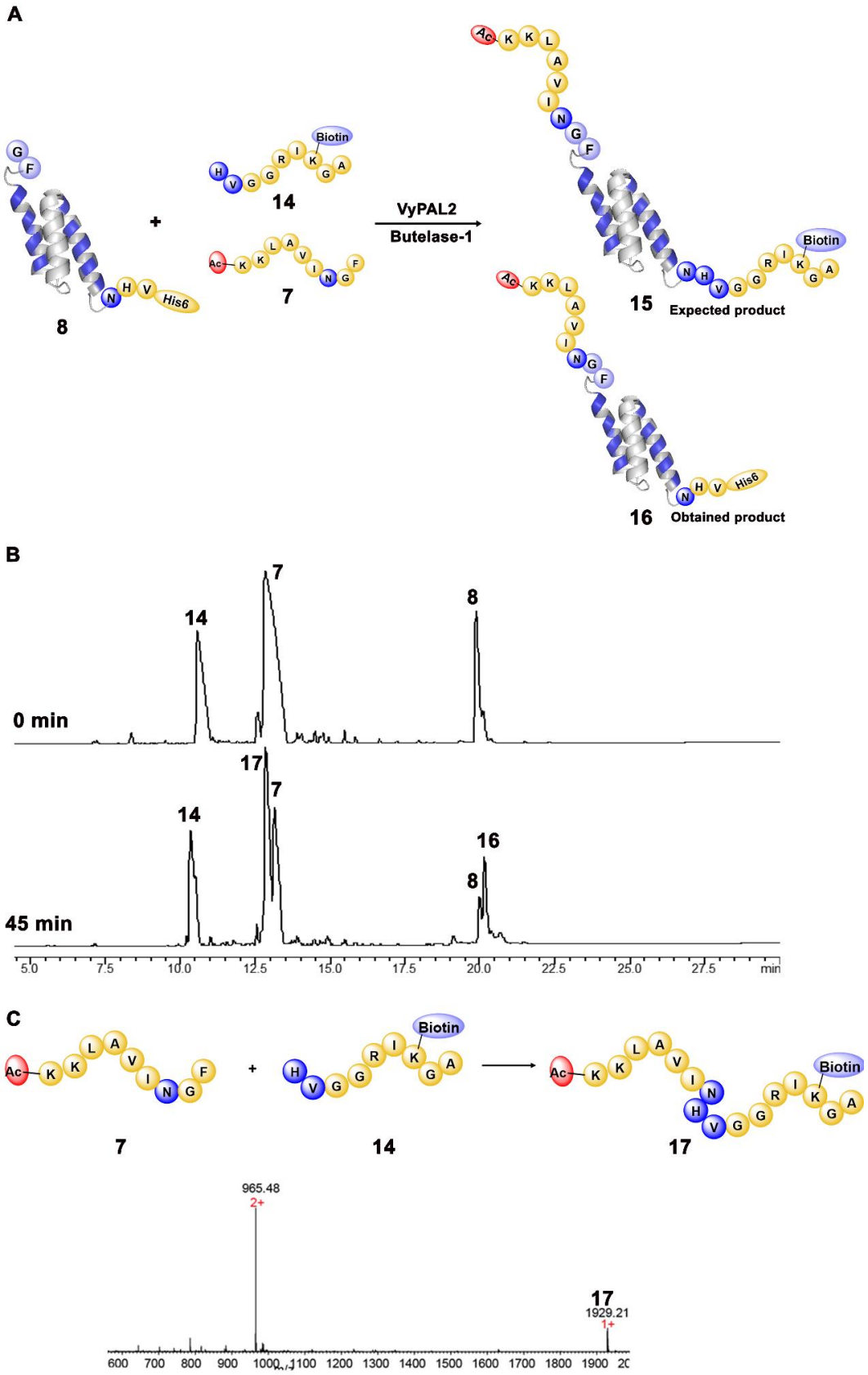
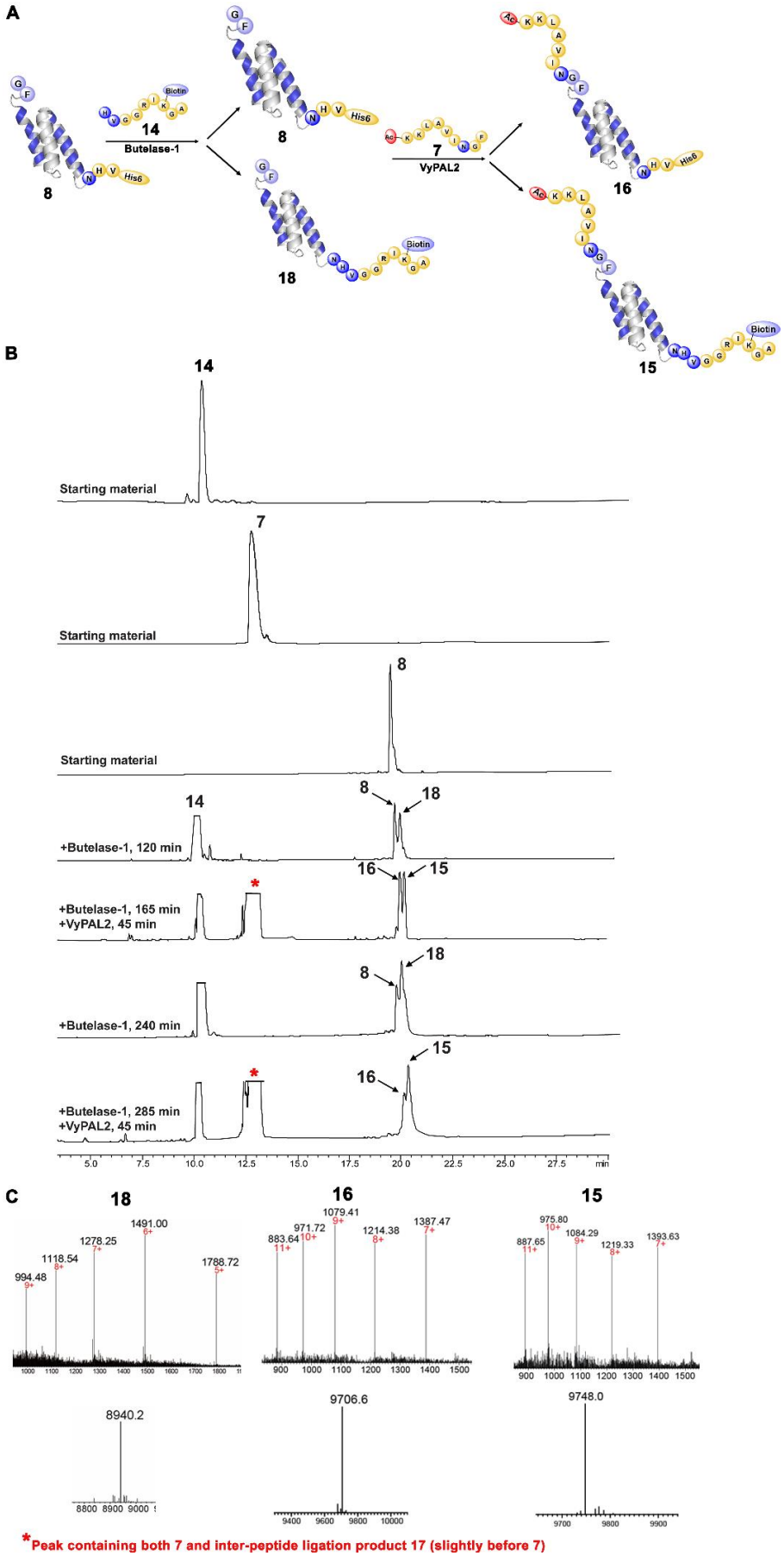

A
8
14
7
VyPAL2
Butelase-1
15
Expected product
16
Obtained product
B
0 min
45 min
14
7
8
17
16
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
27.5
min
C
7
14
17
965.48
2+
1929.21
1+
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900

Figure 2.4. Affibody dual labelling by one-pot reaction of simultaneous BML and VML. A) Reaction scheme of simultaneous one-pot ligations. B) HPLC monitoring of the simultaneous one-pot reaction. Upper panel: mixture of peptide 1**4** (250 μM), **7** (500 μM), and protein **8** (50 μM) before the addition of the enzymes; Lower panel: the reaction mixture at 45 min, after simultaneous addition of butelase-1 (250 nM) and VyPAL2 (160 nM). C) Formation of the inter-peptide ligation product **17** resulting from **7** reacting with **14** in the reaction mixture containing the two PAL enzymes (calcd. 1927.4, obsvd. 1929.2).

Next, we performed sequential ligation reactions in one pot. When the sequence of the reactions was VML first and BML second, a very small amount of the end-product was obtained. This is because firstly, a large amount of peptide **7** was present, which could react with the subsequently added peptide **14** (which was intended for protein C-terminal labeling via BML), and secondly, the N-terminal labeled product **16** could also react with **14** presumably through catalysis by VyPAL2, causing transpeptidation at the newly formed N-G bond in **16** to yield **8**. VML appeared to be much more efficient than BML in this reaction setting. Together, this would cause the unproductive consumption of peptide **14**. Therefore, sequential one-pot VML-BML reactions were not useful.

Next, we performed one-pot sequential ligations in the reverse order. To this end, first, butelase-1 was added to the reaction mixture containing **8** (50 μM) and HV-peptide **14** (250 μM). After incubation (**Figure 2.5**), peptide **7** (50 μM) and VyPAL2 were added. As seen in **Figure 2.5**, the butelase-mediated ligation of the affibody protein **8** with HVGGRIK(Biotin)GA **14** yielded ~40% of the ligation product **18** in 2 h and 60-65% in 4 h (**Figure 2.5**). We found that both the C-terminal labeling product **18** and unreacted **8** from the first step were completely reacted with **7** and cleanly converted to their respective products **15**

and **16** within 45 min. However, significant side reactions involving ligation between the two small peptides **7** and **14** occurred, which gave a large quantity of the inter-peptide ligation product **17** that eluted closely with peptide **7** (**17** slightly before **7**, but not resolved in the profiles shown in **Figure 2.5**).



*Peak containing both 7 and inter-peptide ligation product 17 (slightly before 7)

Figure 2.5. Dual labelling of affibody **8** with an HV-peptide by butelase-1 (C-terminal labelling) and an –NGF peptide by VyPAL2 (N-terminal labelling). A) Schematic illustration of affibody dual labelling by sequential BML and VML in one pot. B) HPLC analysis. The first three HPLC profiles correspond to the starting materials: peptide **14**, peptide **7**, and protein **8**, respectively. For dual labelling of the affibody, butelase-1 was first added to the reaction mixture containing **8** and HV-peptide **14**. At 120 min, an aliquot of the reaction mixture was taken out for HPLC analysis. At the same time, the reaction mixture was divided to two halves. Peptide **7** and VyPAL2 were added to the first half of the reaction mixture. After 45 min, an aliquot of the reaction was analyzed by HPLC. The second half was allowed to continue the reaction under BML for another 120 min, at which time peptide **7** and VyPAL2 were added. After 45 min, an aliquot of the reaction mixture was analyzed by HPLC. **16** was formed by VML of the unreacted starting material **8** (left-over from the BML step) with peptide **7**. **18** was formed from BML of **8** with **14**. **15** is the desired final product formed by VML of **18** with **7**; C) Characterization of **18**, **16**, and **15** by ESI-MS (**18**: calcd. 8937.8, obsvd. 8940.2; **16**: calcd. 9705.8, obsvd. 9706.6; **15**: calcd. 9753.8, obsvd. 9748.0).

Using a GV-peptide such as GVGGRIK(Biotin)GA **19** was also considered for a more orthogonal ligation scheme in one pot. However, the BML reaction with **19** was even slower than that with **14**. Despite a large excess of **19** (400 μM) to **8** (50 μM), the reaction yielded less than 30% of the ligation product in 2 hours. Therefore, because the first BML step would take a very long time to complete, the overall efficiency of using a GV-peptide as the nucleophile substrate for the sequential ligation scheme would be very low, despite the second VML step being relatively fast (**Figure 2.6**). However, it is worth noting that there was significant inter-peptide ligation when the two peptides **7** and **19** were present in the reaction mixture (**Figure 2.6**).

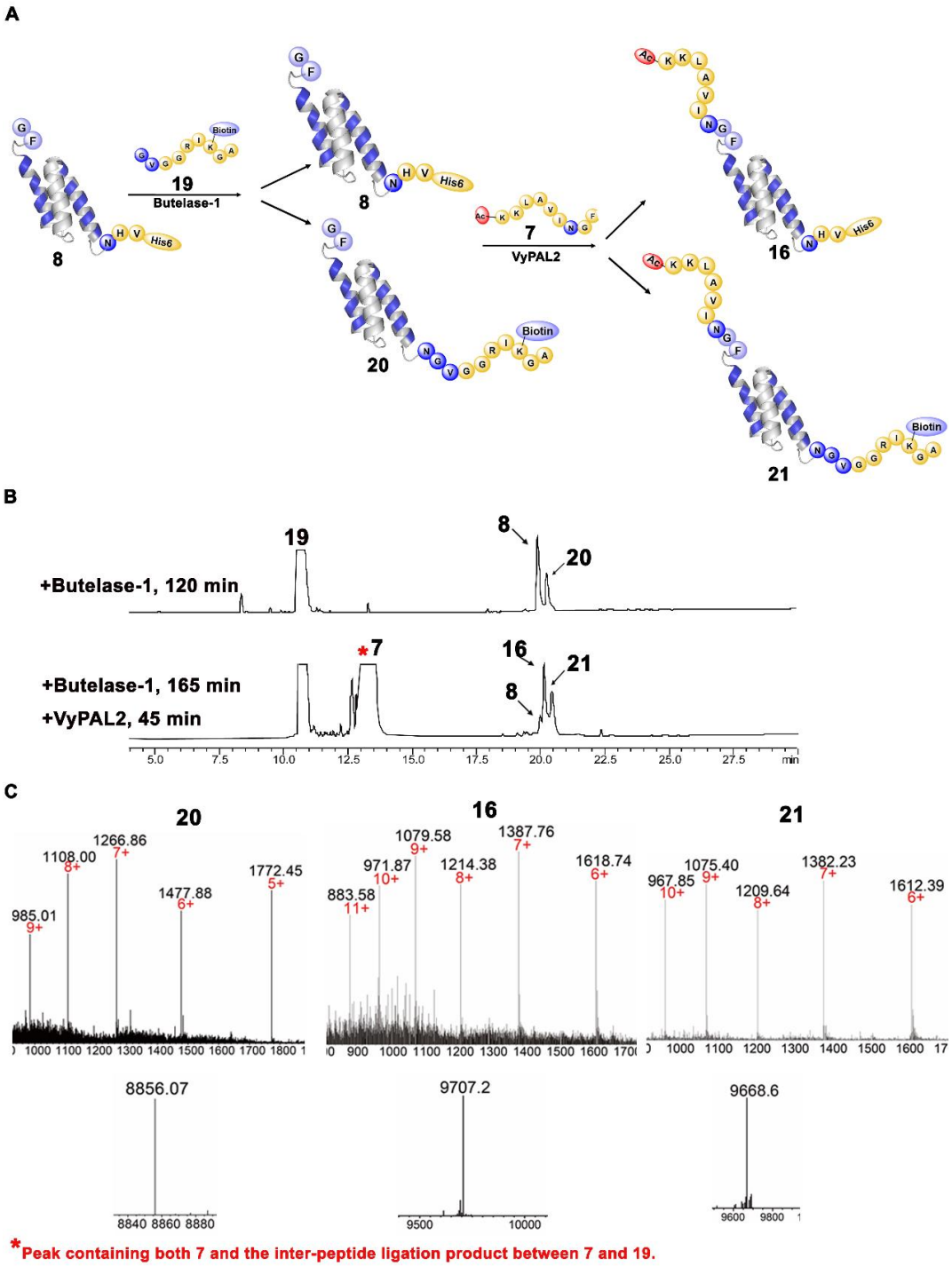


Figure 2.6. Butelase-1- and VyPAL2-mediated affibody dual labelling in one-pot reaction. A) Schematic illustration of the one-pot ligation; B) HPLC profiling of sequential one-pot ligation. Butelase-1 was first added to the reaction mixture containing affibody **8** (50 μM) and the GV-peptide **19**. After 120 min, VyPAL2 and peptide **7** were added. Upper panel: HPLC analysis of affibody **8** ligating with peptide **19** catalyzed by butelase-1 at 120 min. As shown in the

figure, less than 30% of BML product **20** was formed from the first step after 120 min of reaction. Product **20** was isolated and the mass was confirmed using ESI-MS. Lower panel: HPLC profiling of the reaction mixture 45 min after addition of peptide **7** and VyPAL2. **16** was formed by VML of the unreacted starting material **8** with peptide **7**; C) Characterization of **20**, **16**, and **21** via ESI-MS (**20**: cal. 8856.5, obs. 8856.1; **16**: cal. 9706.1, obs. 9707.2; **21**: cal. 9667.9, obs. 9668.6).

Kinetic studies on the GV- and HV-peptides in comparison with the GI-peptide in BML with the NHV peptide **1** were also performed. Indeed, we found that these two nucleophile substrates were inferior to the GI-peptide in the reaction kinetics (**Table 2.3**).

Table 2.3 Kinetics of butelase-1-mediated ligation of peptide **1** with different nucleophile substrates.

Electrophile substrate	Nucleophile substrate	k_{cat} [s^{-1}]	K_m [μM]	k_{cat}/K_m [$M^{-1}s^{-1}$]
Ac-KKLAVINHV **1**	**GI**GGIKA **2**	8.38 ± 0.4	633 ± 41	13253 ± 230
Ac-KKLAVINHV **1**	**GV**GGRIK(Biotin)GA **19**	2.35 ± 0.06	352 ± 13	6671 ± 128
Ac-KKLAVINHV **1**	**HV**GGRIK(Biotin)GA **14**	5.14 ± 0.8	519 ± 17	8714 ± 106

Of note, in the C-to-N scheme, the NGI sequence formed at the first BML step was not affected significantly when conducting VML, likely because **9** – with much faster diffusion kinetics than the much larger molecule **13** – was used in a 5-fold excess to **13**. In addition, the C-terminal NGI in small peptide **9** may be more accessible than the NGI sequence in **13**. Nevertheless, it would be ideal to use an incoming nucleophile peptide sequence at the first BML step that would generate a site that is sub-optimal for recognition by VyPAL2. However,

when testing peptides with an N-terminal HV or GV dipeptide motif for BML in the first step, we found these peptides to be poorer nucleophile substrates than the GI-peptides for butelase-1 recognition. Indeed, the BML reaction of the affibody protein **8** with HVGGRIK(Biotin)GA, peptide **14**, yielded only ~40% of ligation product in 2 h and 60–65% in 4 h (**Figure 2.5**). This reaction was much slower than the reaction with the GI-peptide, **11**, which was nearly complete (at least 85%) in 30 min under the same conditions. The ligation reaction of **8** with a GV-peptide, GVGGRIK(Biotin)GA **19**, was even slower, providing less than 30% in 2 h (**Figure 2.6**). Therefore, although using an NH-peptide could make the C-to-N scheme a potentially more orthogonal method, such a scheme would be significantly less efficient than the one using a GI-peptide.

We also attempted a one-pot reaction using a pair of N-terminal and C-terminal labeling peptides, **7** and **14**, which are supposedly of optimal orthogonality. However, simultaneous one-pot BML and VML reactions did not yield the desired end product (**Figure 2.4**). Conducting BML and VML sequentially in one pot furnished the desired dual labeled product in good yields, albeit with a significant unwanted side reaction of inter-peptide ligation (**Figure 2.5**).

2.2.6 BML for labeling of ubiquitin 22 containing C-ter NHV

Fluorescent ubiquitin **24** was prepared from Ub-NHVH_6 **22** and peptide **23** via butelase-mediated ligation (BML) (**Figure 2.7**). The reaction was performed by mixing 50 μM of **22** and 250 μM of **23** with 100 nM of butelase-1 for 30 min. The product **24** was then purified by HPLC. The purified product **24** was subjected to refolding using the serial dilution method. The lyophilized powder was dissolved in 6 M guanidine HCl (pH 7, phosphate buffer) and dialyzed against decreasing concentrations of guanidine HCl buffers until pure PBS.

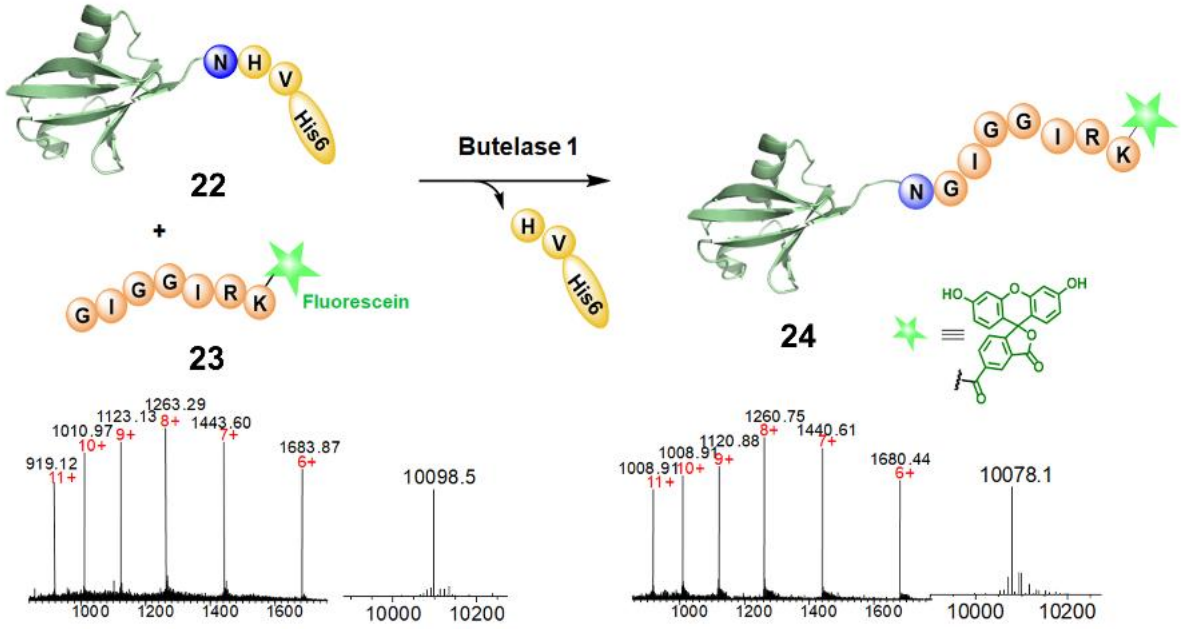


Figure 2.7. Ubiquitin **22** labeling via butelase-mediated ligation. Ubiquitin **22** containing the C-ter NHVHis6 tag was reacted with the fluorescein-peptide **23** using butelase-1. ESI-MS data of ubiquitin **22** calcd mass: 10097, found: 10098; The product **24** calcd mass: 10077 and found: 10078.

2.2.7 Binding affinity (K_D) of the fluorescein-labeled ubiquitin and dual-labeled affibody 12

To study the activities of the dually labeled product **12**, we analyzed its binding to EGFR-overexpressing A431 cells. The cells were treated with 100 nM of **12**, while the fluorescein-tagged ubiquitin **24**, which was prepared via BML with the fluorescent peptide **23** (**Figure 2.7**), was used as a negative control. As shown in **Figure 2.8 B**, strong green fluorescence was observed in A431 cells treated with **12**, whereas no fluorescence was observed in A431 cells treated with **24**. Meanwhile, FACS analysis indicated a remarkable shift in the fluorescence intensity of the **12**-treated cells in reference to ubiquitin **24**-treated cells in the control group (**Figure 2.8 C**), which was consistent with the fluorescence imaging data. To determine the

dissociation constant (K_D) of **12**, these cells were treated with different concentrations of **12** and subjected to FACS analysis after 30 min of incubation. As shown in **Figure 2.8 C**, treatments with different concentrations of **12** resulted in different intensity shifting. The mean fluorescent intensity was analyzed using a non-linear regression function, resulting in a K_D of 18.28 ± 0.48 nM (**Figure 2.8 D**).

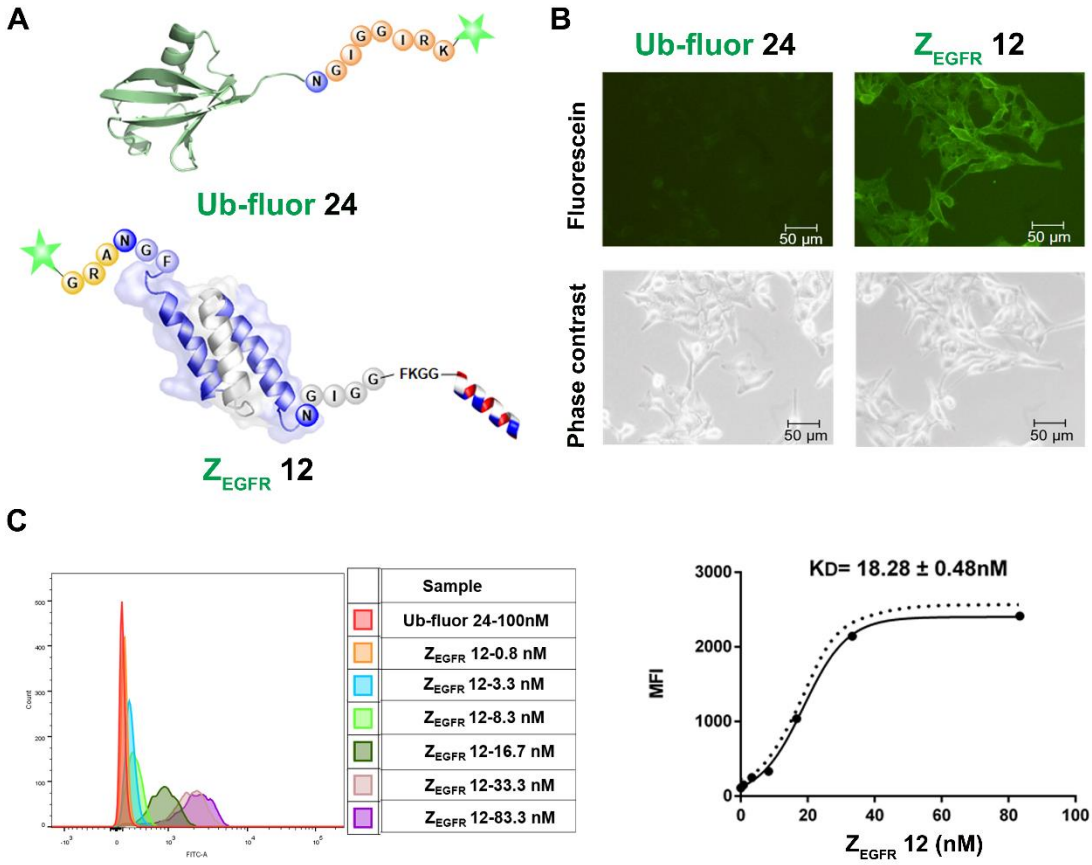


Figure 2.8 Imaging and binding study of **12** on EGFR-overexpressing A431 cells. (A) Schematic structure of ubiquitin tagged with fluorescein and dual labeled affibody **12** with fluorescein on the N-terminus and mitochondrion-lytic peptide at the C-terminal end; (B) Fluorescence microscopy analysis of **12** binding on A431 cells; (C) Determination of K_D of **12** in binding to A431 cells using flow cytometry.

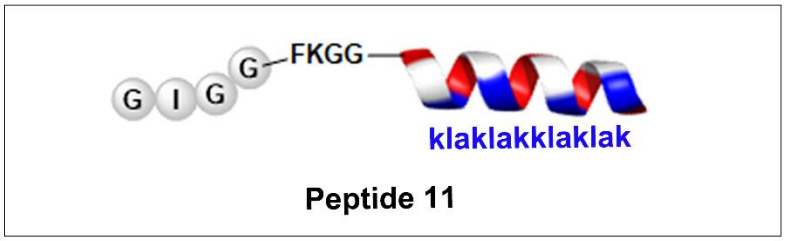
2.2.8 Cytotoxicity evaluation of the dual-labeled affibody

Next, to determine whether compound **12** had any effects on the two cell lines, EGFR-overexpressing A431 cells and MCF-7 cells, which have low EGFR expression levels, we performed the MTT assay.[103] Both cell lines were treated with peptides **11** and **12** at different concentrations: 0, 5, 15, 25, 100, 200, 300, 400, and 1000 μM of **11** and 0, 0.25, 0.5, 1, 5, 20, and 30 μM of **12** for 84 h. Next, an MTT-based viability test was performed, and the optical absorbance was determined. Assays at each concentration point were run in triplicate. The IC_{50} was calculated from the cell survival (%) vs. log (drug concentration) curves via a non-linear regression method using Prism GraphPad.

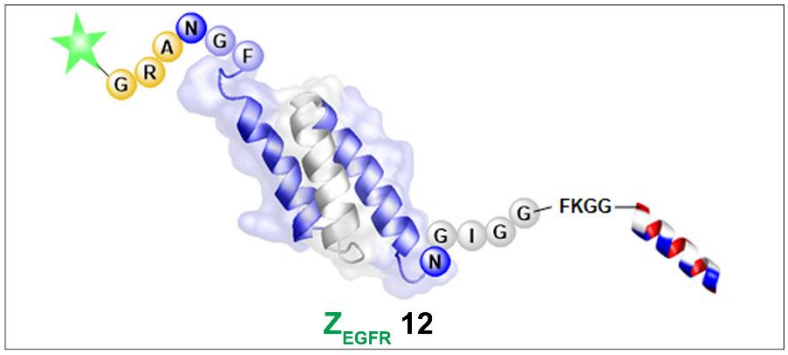
Protein **12** exhibited significant toxicity to A431 cells with an IC_{50} of 11.61 ± 0.96 μM, whereas it showed an IC_{50} of 155.23 ± 3.99 μM for MCF-7 cells. The unconjugated peptide **11** had IC_{50} values of approximately 480 μM and 1300 μM against MCF-7 and A431 cells, respectively (**Figure 2.9** and **Figure 2.10**). Owing to its poor cellular uptake,[104] the low cytotoxicity of **11** itself in the two cell lines was expected.[101] However, conjugating the peptide to the EGFR-targeting affibody drastically enhanced its cytotoxicity against A431 cells, likely because the affibody helped deliver the mitochondrion-lytic peptide intracellularly via EGFR-mediated endocytosis. Thus, as the internalized **12** ended up in the lysosomes, the high proteolytic activity of enzymes, such as cathepsin B, would destroy the peptide linker and even the affibody to release the **KLA** D-peptide, which, after escaping from lysosomes, would disrupt the mitochondrial membrane, leading to apoptosis.

The data clearly indicate that a protein with orthogonal N- and C-terminal recognition tags can be dually labeled by the consecutive action of two PALs with differential substrate specificities. The dually labeled affibody protein has selective imaging and cytotoxic activities.

A

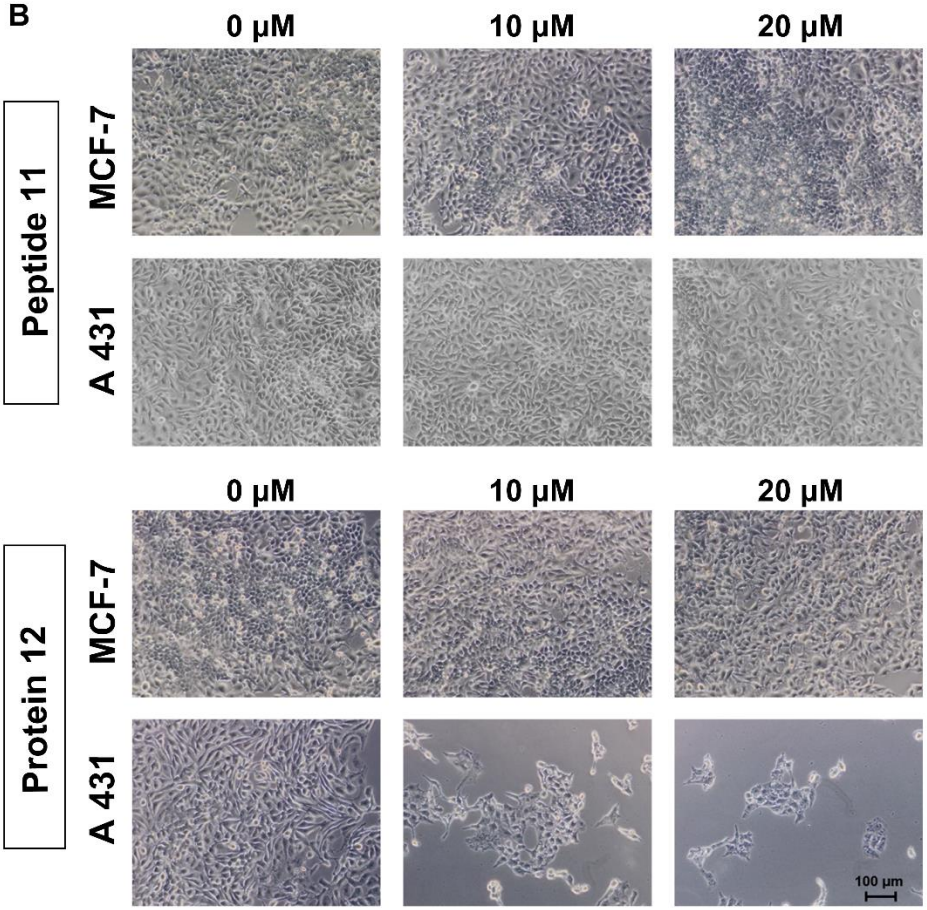


Peptide 11



Z_{EGFR} 12

B



C

	MCF-7	A431
Peptide 11 (μM)	~ 480	~ 1300
Protein 12 (μM)	155.2 ± 4.0	11.6 ± 1.0

Figure 2.9. Cytotoxicity study of the dually labeled protein **12**. (A) Schematic structure of the **KLA** D-peptide **11** and dual labeled affibody **12** with fluorescein on the N-terminus and mitochondrion-lytic peptide at the C-terminal end; (B) Microscopy analysis of **11** and **12** in MCF-7 and A431. Cells were treated with phosphate buffer (as negative control) **11** or **12** for 72 h and then subjected to microscopy analysis after washing 3 times with PBS; (C) IC_{50} of **11** and **12** on MCF-7 and A431 cells. Both cells were treated with **11** or **12** for 84 h, followed by an MTT-based viability test to evaluate optical absorbance and calculate the corresponding IC_{50}.

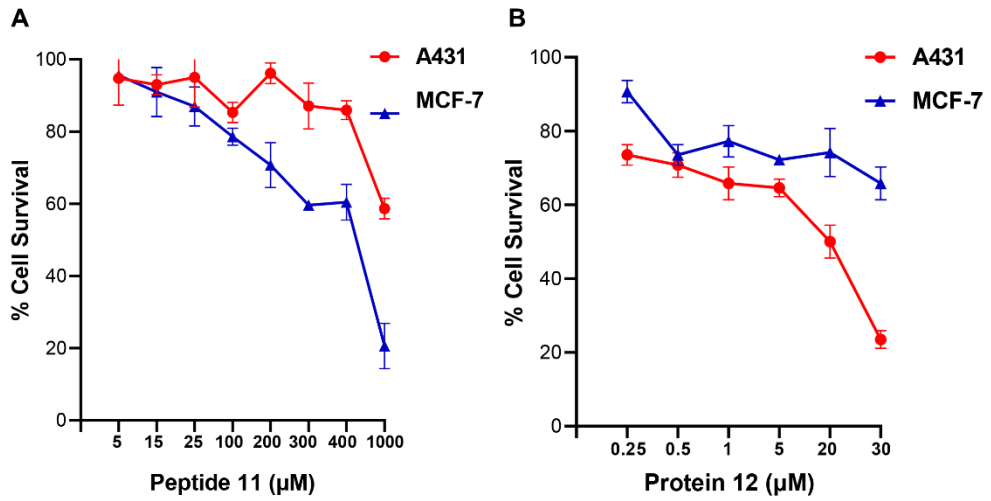


Figure 2.10. IC_{50} calculation of MCF-7 and A431 cells. The cell survival curves of A431 and MCF-7 were both treated with A) peptide **11**; B) protein **12**.

2.2.9 Synthesis of a cyclic affibody–doxorubicin conjugate

To further demonstrate the versatility of the tandem ligation scheme, we proceeded with the synthesis of a cyclic form of the affibody tagged with doxorubicin. For this purpose, peptide **25**, containing an N-terminal GF dipeptide as the nucleophile substrate for VyPAL2 and a C-terminal NHV tripeptide motif at the C-terminus as the electrophile substrate for butelase-1, was prepared using SPPS (**Figure 2.11**). The aminooxy functional group in the peptide would

allow for conjugation with DOX through its ketone group via oxime formation.105 The detailed synthetic route of **25** is shown in **Figure 2.11**. For obvious reasons, **8**, the affibody substrate for dual labeling, could not be used here because the inventing peptide **25** already contains the same respective nucleophile and electrophile substrates for the two PALs (**Figure 2.12A**). Therefore, affibody Z_{EGFR} **26** containing "CG-" at the N terminus and "-NGL" at the C-terminus was prepared recombinantly in E. coli. Interestingly, ESI-MS analysis showed that the N-terminal cysteine residue of **26** was capped during protein expression, presumably as a thiazolidine moiety by the ubiquitous aldehyde metabolite glyoxylic acid in the bacterial cells, effectively blocking it from being used as a nucleophile substrate by the PAL enzymes. Thus, only the C-terminal labeling product Z_{EGFR} **27** would be generated in the first ligation step, without the possibility of cyclization or self-ligation of **26**. As expected, when VML was performed by mixing 50 μM of Z_{EGFR} **26** and 150 μM of peptide **25** with 100 nM of VyPAL2 at 37 °C for 30 min, only the C-terminal ligation product **27** was obtained, as shown clearly by HPLC and ESI-MS analysis (**Figure 2.12B**). The NHV tag in **25** or **27** was not affected, confirming its orthogonality toward VyPAL2.

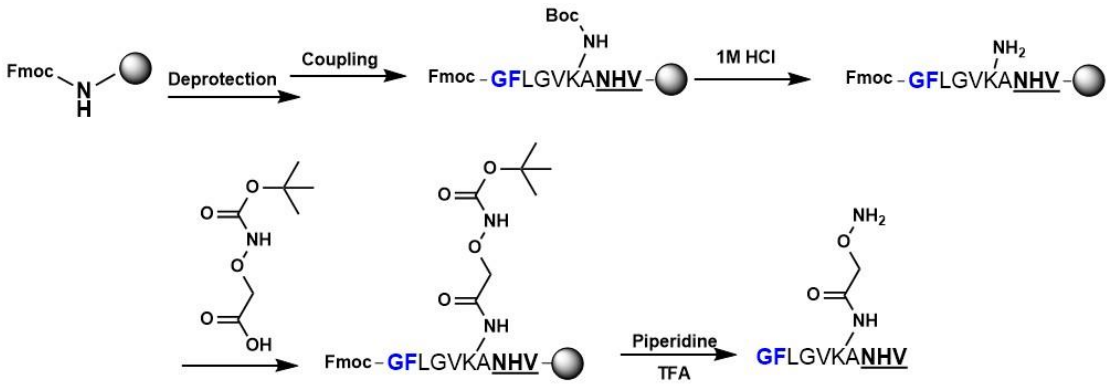


Figure 2.11. Synthesis of peptide **25**.

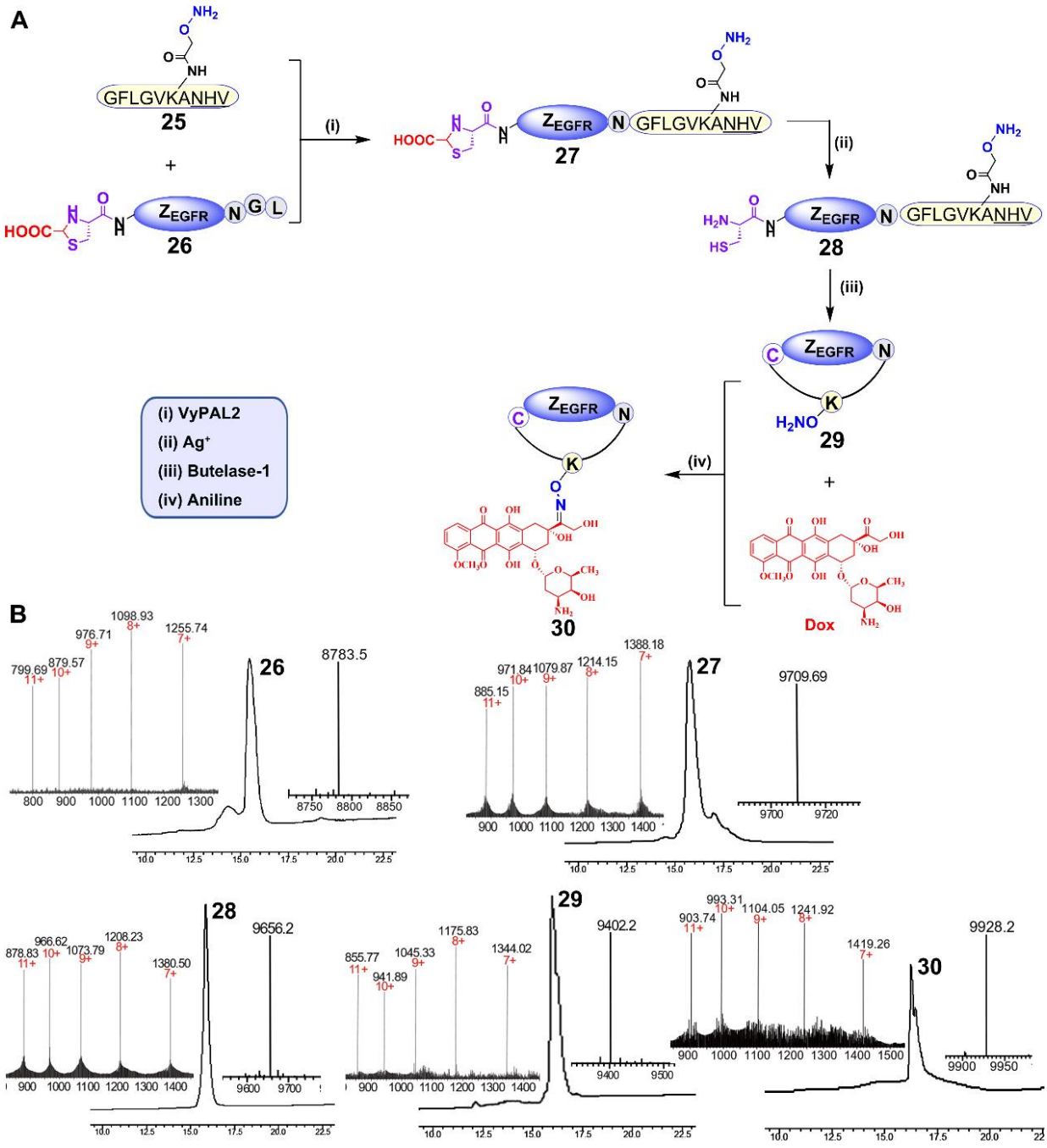


Figure 2.12. Synthesis of a cyclic affibody–drug conjugate using PAL-catalyzed orthogonal ligation and cyclization and oxime conjugation. (A) (i) Peptide **25** was tagged to the C-terminus of Z_{EGFR} **26** via VML to give **27** (90%); (ii) the N-terminal cysteine of Z_{EGFR} **27** was deprotected using silver nitrate to afford **28** (95%); (iii) Z_{EGFR} **28** was cyclized via BML to give **29** (70%); (iv) DOX was attached to Z_{EGFR} **29** via oxime conjugation to obtain final product **30** (80%). (B) HPLC and ESI characterization of purified products (**26:** calcd 8785.6, obsvd 8783.5; **27:**

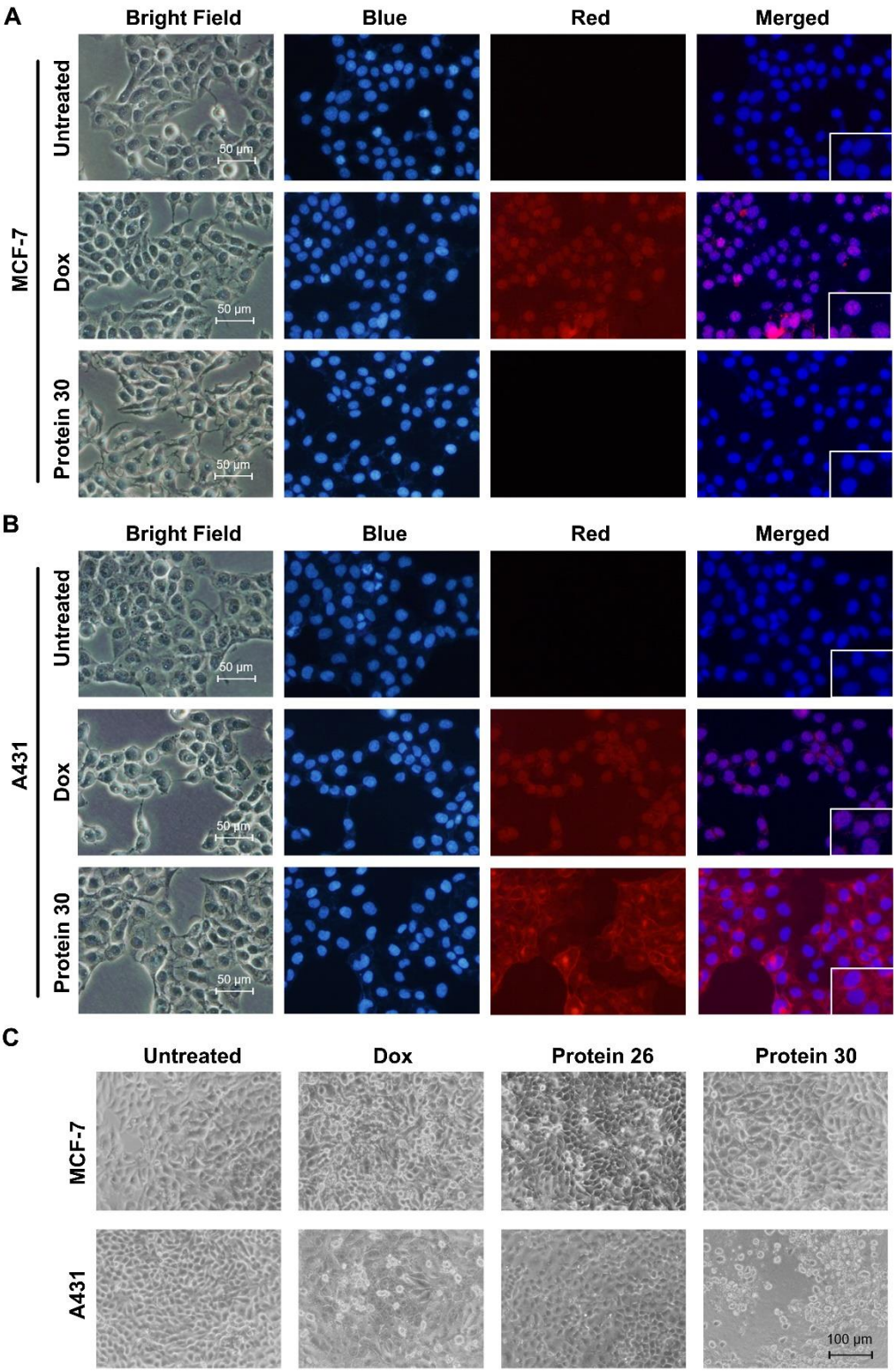
calcd 9708.9, obsvd 9709.7; **28:** calcd 9655.6, obsvd 9656.2; **29:** calcd 9401.7, obsvd 9402.2; **30:** calcd 9927.7, obsvd 9928.2).

To unmask the N-terminal cysteine in **27**, purified **27** (1 mM) was treated with silver nitrate (10 mM) for 30 min, followed by treatment with β-mercaptoethanol (100 mM) for 30 min. The deprotection reaction resulted in product **28**, as confirmed by HPLC and ESI-MS (**Figure 2.12 A** and **Figure 2.12 B**). Butelase-mediated cyclization was performed by mixing **28** (100 μM) with 50 nM butelase-1 for 30 min at 37 °C. Cyclic product **29** was characterized by HPLC and ESI-MS (**Figure 2.12 B**). The aminooxy functional group in Z_{EGFR} **29** can react with the doxorubicin ketone group via Schiff's base formation. Therefore, the oxime ligation reaction was carried out by mixing 100 μM of Z_{EGFR} **29** and 1 mM of DOX in the presence of 10 mM of aniline as the catalyst[106] at pH 6 and 37 °C overnight. The reaction gave rise to final product **30**, as characterized by HPLC and ESI-MS (**Figure 2.12**).

2.2.10 Cell imaging and cytotoxic study of the synthesized cyclic affibody-doxorubicin conjugate 30

Fluorescence imaging, microscopy analysis, and MTT assay were performed to determine the binding and inhibitory effects of cPDC **30** on the MCF-7 and A431 cell lines. The intrinsic fluorescence of doxorubicin serves as an imaging tool to visualize the binding of **30** to the cells (**Figure 2.13 A, B**). As shown in the figures, only the EGFR-overexpressing A431 cells were positively stained after 30 min of treatment with **30**. The same treatment did not yield any staining of EGFR-negative MCF-7 cells. On the other hand, both cell lines were stained by free doxorubicin, which is not surprising as it can enter cells and bind to nuclear DNA. In the cytotoxicity experiments, both cell lines were treated with 0.2 μM of unconjugated affibody **26**, doxorubicin, and **30** for 96 h and subjected to microscopy analysis. At this concentration,

30 exhibited substantial cytotoxic effects on A431 cells, with smaller or no effects observed in the other control settings (**Figure 2.13 C**). Next, an MTT assay was conducted to determine the IC_{50}. Unconjugated affibody **26**, DOX, and cPDC 30 were added at varying concentrations to MCF-7 and A431 cells for 96 h. As shown in Figure **2.13 D**, **30** showed significantly higher toxicity with an IC_{50} of 0.13 ± 0.02 μM for A431 cells than for MCF-7 cells (IC_{50} of 1.51 ± 0.08 μM). The affibody itself had a low level of cytotoxicity, even at very high concentrations, which is consistent with previously published results.[107] The unconjugated DOX showed lower cytotoxicity in terms of IC_{50} to A431 cells compared to **30**, likely due to a lack of receptor-mediated enrichment of the compound in the cells. The measured IC50 of DOX in MCF-7 and A431 was 1.60 ± 0.23 μM and 1.22 ± 0.15 μM, respectively (**Figure 2.13 D**). The enhanced toxicity of the cycloaffibody-DOX conjugate **30** was likely due to the fast enrichment of the conjugate via receptor-mediated endocytosis, which led to the uptake of the conjugate through the endosomal pathway and delivered it to the lysosome. The acidic milieu in this organelle would help cleave the oxime linkage to release DOX.[55] Owing to its hydrophobic properties, doxorubicin could easily escape from the lysosome to bind to nuclear DNA, leading to apoptotic cell death.



D

	MCF-7	A431
Protein 26 (μM)	-	-
Dox (μM)	1.60 ± 0.23	1.22 ± 0.15
Protein 30 (μM)	1.51 ± 0.08	0.13 ± 0.02

Figure 2.13. Cell imaging and cytotoxicity study of the cyclic affibody-DOX conjugate **30**. (A) Fluorescent microscopy analysis of MCF-7 cells after treatment with **30**, DOX, and blank at room temperature for 30 min. (B) Fluorescent microscopy analysis of A431 cells after treatment with **30**, DOX, and blank at room temperature for 30 min. For cell staining experiments in (A) and (B), the nucleus was stained with 700 nM of DAPI; 10 μM DOX and 2 μM **30** were used. Scale bar, 50 μm. (C) Cytotoxicity assay of the cPDC **30**. Microscopy analysis of cells treated with **30**. MCF-7 and A431 cells were treated with 0.2 μM of different molecules: DOX, unconjugated affibody **26**, and **30** for 96 h. Scale bar, 100 μm. (D) Cytotoxic IC_{50} of different compounds against MCF-7 and A431 cells. **30** exhibited a ~10-fold enhanced toxicity on the EGFR-overexpressing A431 cell line compared to doxorubicin.

2.3 Conclusion

Butelase-1 and VyPAL2 are PAL enzymes that recognize short peptide tags for ligation reactions. We exploited the different substrate specificities of these two PALs to develop a new method for the bio-orthogonal dual modification of proteins under mild aqueous conditions at near-neutral pH. We used this novel bio-orthogonal method to prepare a dual-labeled affibody as a selective imaging and cytotoxic agent for cancer cells. Our results have shown that our bio-orthogonal ligation scheme is bi-directional, as it can be executed in both the N-to-C and C-to-N directions, enabling the synthesis of the affibody conjugate **12**. Furthermore, the scheme was extended to the preparation of a cyclic affibody conjugated with the cytotoxic compound doxorubicin. Unlike the hydrophobic free doxorubicin, which is poorly soluble in water, the prepared cycloaffibody-DOX conjugate **30** showed excellent water solubility. Such a conjugate is also expected to have lower cardiotoxicity than free doxorubicin. A backbone-cyclized protein has increased thermal, chemical, and proteolytic stability. As demonstrated by our data, the prepared linear and cyclic affibody conjugates **12** and **30** showed uncompromised

high binding affinity and enhanced cytotoxicity toward EGFR-overexpressing A431 cells. These findings indicate that PALs show promise as a tool for the precision biomanufacturing of complex bioconjugates with multiple functionalities and unusual structures, as well as for the functionalization of protein nanoparticles. Therefore, the methodologies described in this study may pave the way for the development of next-generation protein-based theranostics for the diagnosis, prevention, and treatment of disease in humans.

2.4 Materials and methods

2.4.1 Protein amino acid sequences

The optimized DNA sequences of DARPin and Z_{EGFR} wrere synthesized by Genscript. The following are the amino acid sequences.

Z_{EGFR} **8**

GFGSSLQVDNKFNKEMWAAWEEIRNLPNLNGWQMTAFIASLVDDPSQSANLLAEAKKLNDAQ
APKVDGSGSNHVHHHHHH

Ubiquitin **22**

GGSGSGSQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTL
SDYNIQKESTLHLVLRLRGGNHVHHHHHH

Z_{EGFR} **26**

CGSSHHHHHHLQVDNKFNKEMWAAWEEIRNLPNLNGWQMTAFIASLVDDPSQSANLLAEAKK
LNDAQAPKVDGSGSNGL

2.4.3 Methods

All amino acids, coupling reagents, solvents, and resins were purchased from Sigma and Chemimpex and used as received without further purification. VyPAL2 and butelase-1 were prepared in-house, as previously reported.

HPLC. Analytical RP-HPLC was run on a SHIMADZU (Prominence LC-20AT) instrument using an analytical column (Grace Vydac "Protein C4") (250 × 4.6 mm, 5 μm particle size) at a flow rate of 1.0 mL/min. Analytical HPLC elution was monitored by UV absorption at 214 nm and 254 nm. Semi-preparative RP-HPLC was run on a SHIMADZU (Prominence LC-20AT) instrument using a semi-preparative column (Grace Vydac "Protein C4") (250 × 10 mm, 10 μm particle size) at a flow rate of 2.5 mL/min. Both analytical and semi-preparative HPLC were run at room temperature using a gradient of solvent B in solvent A. Solvent B was 90% acetonitrile in water (0.040% TFA) and solvent A was water (0.045% TFA). Both solvents were filtered through 0.22-μm filter paper and sonicated for 30 min before use.

Protein expression and purification. Genes encoding the desired protein sequences were cloned into the pETDuet vector, and the plasmids were then transformed into E. coli BL21 (DE3) competent cells using the standard 90-second heat shock protocol. The bacterial colonies were transferred to liquid LB medium in a culture flask and shaken in an incubator at 37 °C for 8-12 hours until the OD reached 0.6-0.8, followed by induction with 1 mM IPTG at 37 °C for 4-8 hours for protein expression. Cells were harvested and lysed by sonication in lysis buffer containing 50 mM sodium phosphate and 500 mM NaCl (pH 8.0). After centrifugation, the supernatant was loaded onto a column of Ni-NTA beads and incubated at 4 °C for 1 hour. The beads were washed three times with lysis buffer, and the protein was subsequently eluted with lysis buffer containing 250 mM imidazole. The purified protein was dialyzed in phosphate buffer (pH 6.5) overnight and stored in a freezer at -20°C.

Mass spectrometry. The ESI mass spectrum data of small peptides and proteins were obtained from a Thermo Finnigan LCQ DECA XP MAX (ESI ion source, positive mode). MagTran 1.03 and ESIProt 1.0 software were used for data deconvolution.

Tissue culture and cell imaging. Cells were cultured in DMEM (high glucose) supplemented with 10% FBS at 37 °C in a CO2 incubator. For passaging, cells were detached from the culture plates by washing three times with trypsin-EDTA (0.25%) and neutralizing the trypsin activity with a 3-fold volume of complete DMEM medium. The cells were grown until they reached 40–60% confluency. Peptides or proteins in complete medium were added to the cells and incubated for 30 minutes at 37 °C. After incubation, the cells were washed three times with PBS and subjected to microscopy analysis.

Cell viability assay. MTT assays were carried out following the recommended protocols from Sigma-Aldrich (Cat. No. 11465001001). First, cells were seeded in a 96-well tissue culture plate with 100 μL of medium and allowed to grow until the confluency reached 40–60% of the plate surface. Peptides and proteins were added and incubated for 84 h, followed by the addition of 10 μL of MTT I to each well and further incubation for approximately 4 h. Next, MTT II was added and incubated at 37 °C overnight to solubilize the purple crystals. Spectrophotometric absorbance measurements of the samples were carried out using a microplate reader (Biotek Citation 5) at a wavelength of 575 nm, and the reference wavelength was set at 670 nm.

Cell staining and imaging. MCF-7 and A431 cells cultured in 24-well plates were washed three times with PBS. Formaldehyde (4%, w/v in PBS) was then added to each well for 15 minutes to fix the cells. Then, the cells were washed with PBS three times to remove residual formaldehyde. To permeabilize the cells, Triton X-100 (0.1%, w/v in PBS) was added to the wells for 5 minutes. Then, PBS was used to wash the cells another three times before staining.

To stain the cells, doxorubicin, protein **26**, and DAPI were diluted in PBS to a concentration of 10 μM, 2 μM, and 700 nM, respectively. Then, the solution was added to each well for 30 minutes. Next, the cells were washed with PBS three times and subjected to imaging analysis using inverted fluorescence microscopy (#IX71; Olympus Life Science). To acquire the DAPI fluorescent image, the "Blue" channel (filter cube: 350 nm) was used. Likewise, the "Red" channel (filter cube: 550 nm) was used to obtain the doxorubicin fluorescence, while the "Green" channel (Filter Cube: 450 nm) was used for fluorescein.

Solid-phase peptide synthesis. All peptides used in this study were synthesized as C-terminal amides using Rink amide MBHA resin by standard Fmoc chemistry. Before use, the resin was pre-swelled in DMF for 20 min. Before the first coupling, an Fmoc deprotection procedure was performed using 20% piperidine in dimethylformamide (DMF) for 30 min. The resin was then washed successively with DMF, DCM, and DMF. For the coupling reactions, 3 equiv. of Fmoc-AA-OH and 3 equiv. of PyBOP were dissolved in DMF/DCM (1:1). This mixture was added to the resin, followed by the addition of 6 equiv. of DIEA. Coupling reactions were carried out for 60–90 min. Coupling efficiency was evaluated using the ninhydrin test. For peptides **14** and **19**, Fmoc-Lys(Biotin)-OH was used. For peptide **23**, after peptide assembly on the solid phase, the Boc group on the lysine side chain amine was removed with 1 M HCl in DCM for 30–40 min, then 5(6)-carboxyfluorescein was coupled to the side-chain free amine using PyBOP as the coupling reagent. The peptides were cleaved from the resin with a cocktail containing 95% TFA, 2.5% water, and 2.5% TIS for 2 h. After precipitation with cold diethyl ether, the crude peptides were purified using HPLC. The desired peptides were obtained in powder form after lyophilization. All peptides were characterized by electrospray ionization mass spectrometry.

2.4.4 List of peptides prepared in the study (letters in lower case denote D-amino acids)

Peptide **1**: Ac-**KKLAVINHV**; 1061.01 (observed), 1062.27 (calculated).

Peptide **2**: **GIGGIKA**; 613.68 (observed), 613.74 (calculated).

Peptide **4**: **YKANGL**; 664.26 (observed), 664.67 (calculated).

Peptide **5**: **GFGGIKA**; 648.38 (observed), 648.52 (calculated).

Peptide **7**: Ac-**KKLAVINGF**; 1031.34 (observed), 1031.56 (calculated).

Peptide **9**: Fluorescein-**GRANGI**; 944.52 (observed), 944.97 (calculated).

Peptide **11**: **GIGGFKGG-klaklakklaklak**; 2197.07 (observed), 2197.72 (calculated).

Peptide **14**: **HVGGRIK**(Biotin)**GA**; 1119.89 (observed), 1118.61 (calculated).

Peptide **19**: **GVGGRIK**(Biotin)**GA**; 1039.61 (observed), 1038.58 (calculated).

Peptide **23**: **GIGGIRK**(Fluorescein); 1057.65 (observed), 1057.35 (calculated).

Peptide **25**: **GFLGVK**($COCH_2ONH_2$)**ANHV**; 1113.90 (observed), 1113.29 (calculated).

Ac-KKLAVINHV

1

GIGGIKA

2

Ac-KKLAVINGIGGIKA

3

YKANGL

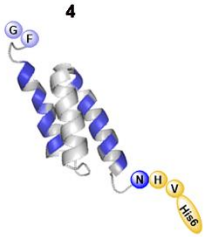
4

GFGGIKA

5

YKANGFGGIKA

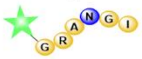
6



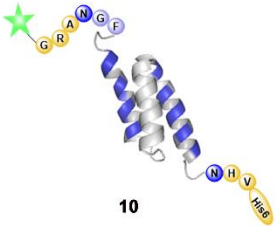
8

Ac-KKLAVINGF

7

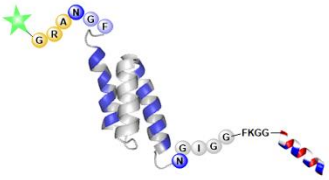


9

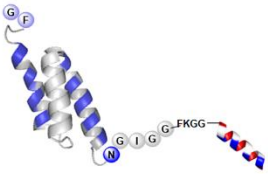


10

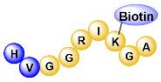
11



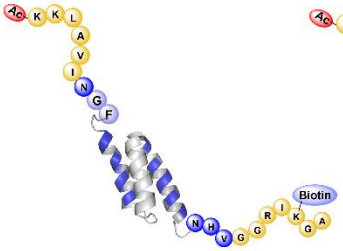
12



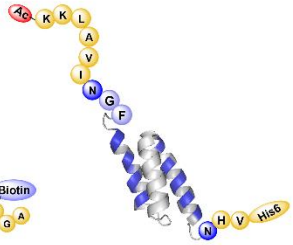
13



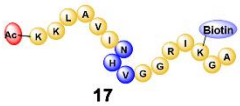
14



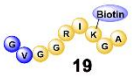
15



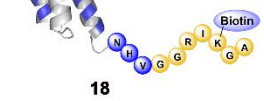
16



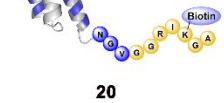
17



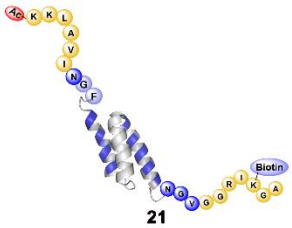
19



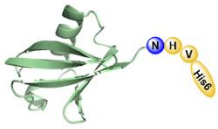
18



20

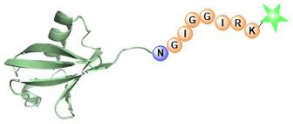


21

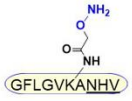


22

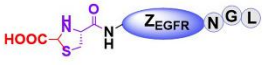
23



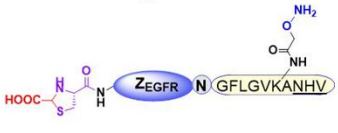
24



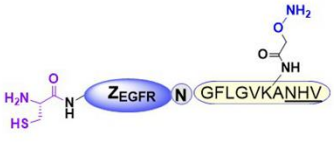
25



26



27



28

29

30

Figure 2.14. Numbering and illustrative structures of compounds used in this study. The green star is 5(6)-carboxyfluorescein coupled through its carboxyl group to the side-chain amine of a lysine residue or to the N-terminal amine.

Chapter 3: Engineering multi-functional protein biologics through PAL-mediated hydrazide ligation

3.1 Introduction

As a technique of increasing importance to study protein functions and develop protein therapeutics, site-specific protein modification allows for a wide range of applications in basic research and translational medicine. [32, 77] In particular, enzymatic methods are gaining traction because of their high specificity and mild reaction conditions.[33, 77] Peptide ligases catalyze the formation of peptide bonds, enabling terminus-directed labeling of a protein of interest (POI) with myriad functional moieties.[8, 86, 82, 108-110] Peptidyl asparaginyl ligases (PALs) are members of the asparaginyl endopeptidase (AEP) family that catalyze peptide ligation through transpeptidation at Asn/Asp-peptide bonds.[8, 110, 111] Compared to sortase A which is also a transpeptidase,[86] PALs exhibit extremely fast catalytic kinetics.[1, 8, 12, 13, 14, 15, 109-111] For this reason, PALs are emerging as very useful tools for protein engineering.[1-3, 4, 7, 9, 10, 13-15, 18, 19, 30, 99, 111-114]

AEPs are cysteine proteases. Their catalytic action involves an acyl-enzyme thioester intermediate formed through the nucleophilic attack of the acyl donor substrate by the cysteinyl sulfhydryl at the enzyme's active site. [8, 11, 25, 27, 93-95, 109, 110] Typically, the acyl-E thioester is hydrolyzed. Atypically, the acyl-E thioester is resolved by a nearby peptidic amine in an aminolysis reaction, forming a new Asx-peptide bond as the result of transpeptidation. PALs are those AEPs that possess predominant transpeptidase activity.[1, 13 - 15, 109-111] PALs recognize an Asx-P1'-P2' tripeptide motif in the acyl donor substrate where the P1' position can tolerate most amino acids and P2' is a bulky hydrophobic amino acid.[1, 13 - 15, 109-111] Interestingly, PALs are relatively promiscuous toward the acyl acceptor substrates. For instance, in catalyzing peptide macrocyclization reactions, butelase-1 tolerates most of the d-amino acid residues at

the P1" and P2" positions of the peptidic nucleophile.[18] During the preparation of this manuscript, an incoming paper reports that certain simple aliphatic amines are effective nucleophile substrates of OaAEP1b – which is also a member of the AEP family (**Figure 3.1 C**).[28] Given the *S*-acyl-E thioester intermediate being a soft electrophile, we envisaged that it could also be intercepted by α-effect nucleophiles, such as hydrazides. Hydrazides are non-basic as evinced with pKa = 3.24 by acethydrazide. However, because of the α-effects the hydrazide amines display significant nucleophilicity, especially in reactions with soft electrophiles.[115, 116]

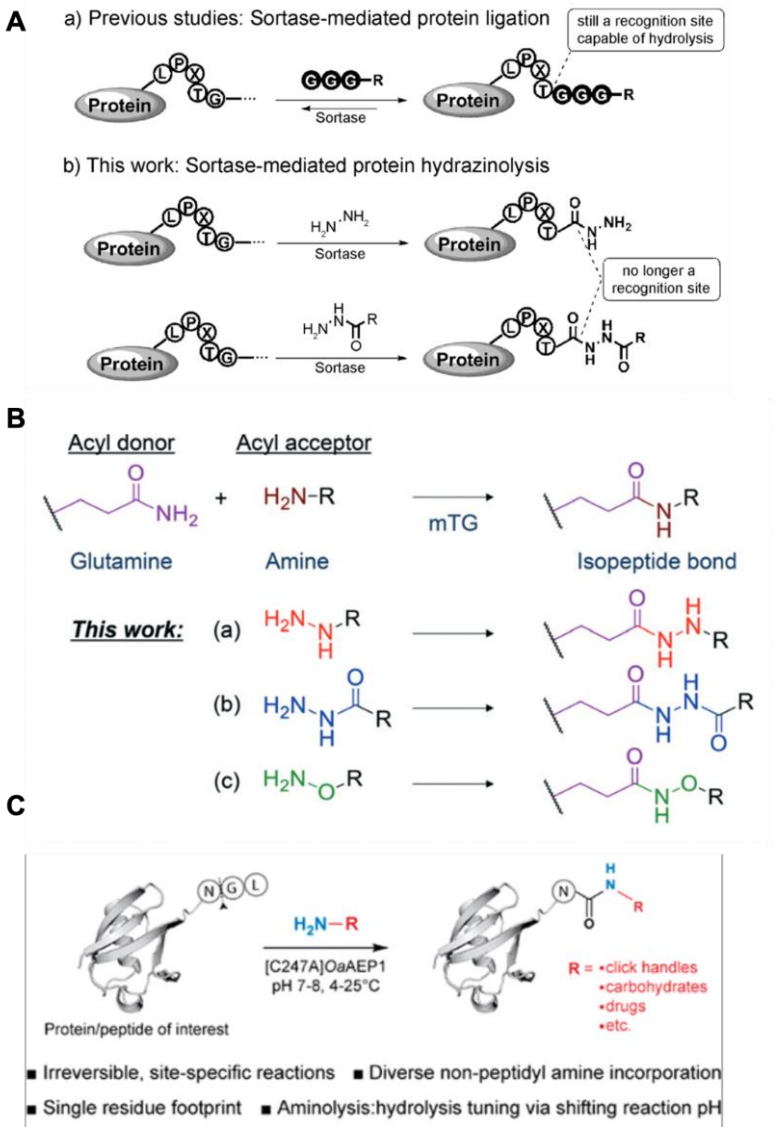


Figure 3.1. Current reported works related with PMHL. A) The Sortase A mediated hydrazinolysis; B) mTG mediated side chain modification of Glu with hydrazide compounds; C) OaAEP1 catalyzed aminolysis reaction.

Previous studies have shown that hydrazine can capture and resolve the (thio)ester intermediates of the proteolytic or transpeptidic reactions catalyzed by certain serine proteases and sortase A to yield hydrazinolysis products (**Figure 3.1 A**)[98, 117-119]. Nevertheless, hydrazine – a rather strong base (pKa-H = 8.1) – is much more nucleophilic than hydrazides. To the best of our knowledge, there has been no report on using hydrazides as the nucleophile partners in aminolysis reactions catalyzed by a peptide ligase or serine/cysteine protease. The only relevant exception is with microbial transglutaminase (mTG), which was shown to accept various hydrazides as the amine substrates and attach them onto the side chain of internal glutamine residues (**Figure 3.1 B**).[120] Unlike mTG, PALs would install a hydrazide group on the C-terminus of the protein backbone. We show that PAL-mediated hydrazide ligation (PMHL) is extremely versatile as the enzymes can use a broad range of hydrazide derivatives as the acyl acceptors (**Figure 3.2**). These hydrazides can be derivatized with various functional moieties (e.g., biotin, fluorescent dyes, and cytotoxic drugs). Even peptide C-terminal hydrazide is tolerated, making it possible to conduct direct C-to-C ligation. Depending on the structure of the hydrazide nucleophile, the amide bond formed in the ligation reaction is of varying stability toward the PAL enzyme. Some of the unnatural peptide bonds are resistant to the enzyme, preventing reversibility of the ligation reaction – a problem seen with conventional peptide nucleophiles. This also makes it possible to perform a second labeling reaction on the same POI in a sequential ligation scheme.

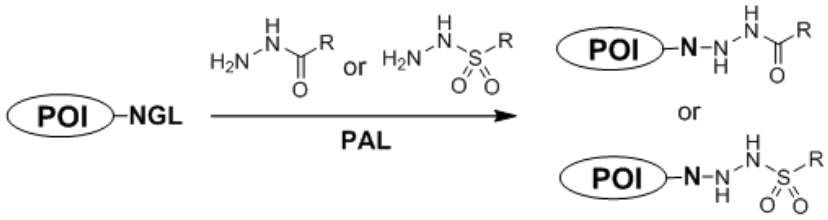


Figure 3.2. General scheme of protein–hydrazide ligation catalyzed by peptidyl asparaginyl ligases. R may carry functional moieties such as amino acids and peptides, biotin, fluorescent or other biophysical tags and drug components.

Attributed to the α-effects, hydrazides are nucleophilic despite being non-basic. Herein, we report that the hydrazide nucleophiles are effective acyl acceptors in the ligation reactions catalyzed by peptidyl asparaginyl ligases (PALs). Because hydrazides are easily functionalizable, hydrazide ligation is highly versatile. Interestingly, the linkages formed with the hydrazide substrates have varying degrees of reversibility toward the PAL enzyme, with some being remarkably resistant and thus ensuring nearly irreversible ligation. Using the hydrazide ligation method, we labelled an EGFR-targeting affibody-Fc fusion protein with various functional moieties to generate selective and potent cancer-imaging and therapeutic agents. Irreversible hydrazide ligation also allowed a sequential ligation scheme to be conducted on a protein. Using this scheme, quadruple FITC labels were introduced onto the N- and C-termini of the affibody-Fc protein (**Figure 3.3**) to yield a bi-specific engager for CAR-NK cell therapy. Our work expands the substrate scope of PAL enzymes and further points to their promise as a precision manufacturing tool for multi-functional protein biologics.

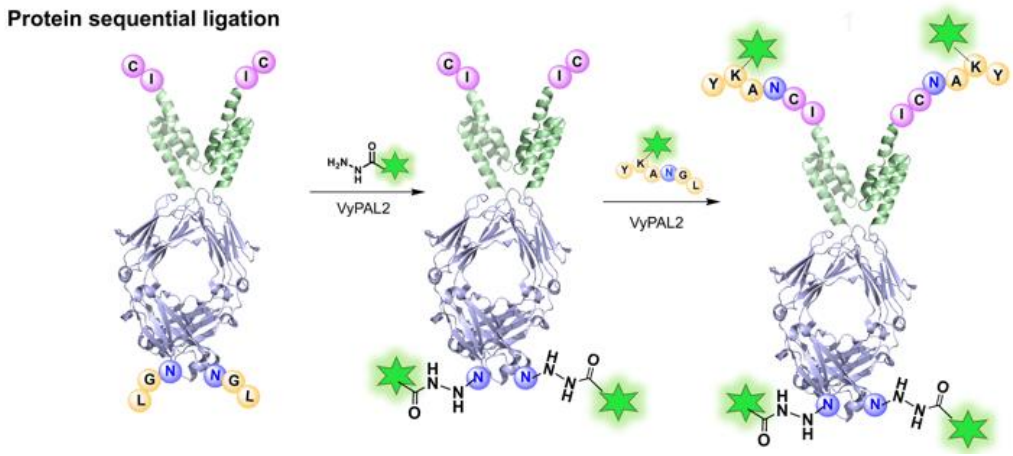


Figure 3.3. The quadruple FITC labels were introduced onto the N- and C-termini of the affibody-Fc protein through sequential ligation to yield a bi-specific engager for CAR-NK cell therapy.

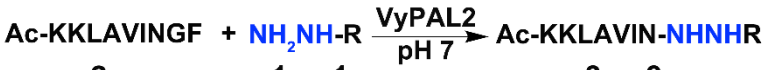
3.2 Results and Discussion

3.2.1 Substrate scope of VyPAL2-catalyzed hydrazide ligation

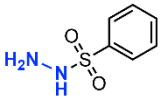
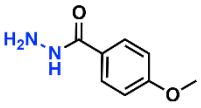
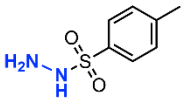
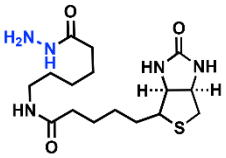
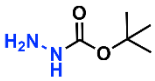
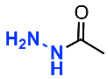
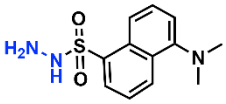
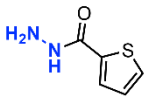
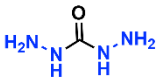
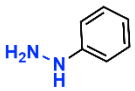
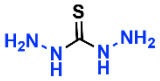
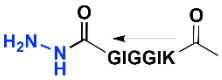
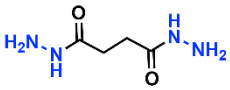
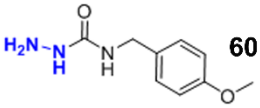
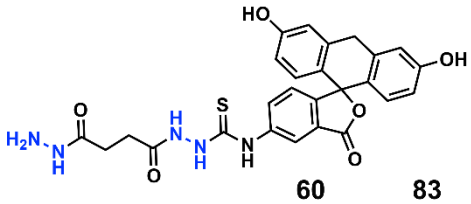
To investigate the reaction scope, we first screened a panel of compounds for PAL-mediated hydrazide ligation. Except hydrazine **1a**, all are hydrazide derivatives, including acyl hydrazide, sulfonyl hydrazide and peptide C-terminal hydrazide (**Table 3.1**). The reactions were performed by adding VyPAL2 (25 nM) to a mixture of 50 μM acyl donor substrate **2** (1 eq) which contains a C-terminal NGF tripeptide tag and 10 mM α -nucleophile **1a**-**1p** (200 eq) at 37 °C and pH 7.0 for 10-60 min. The reactions yielded products **3a**-**3p**, which were analyzed using HPLC and characterized using MS-ESI (**Table 3.1** and detailed data shown in **Appendix B**). As shown in **Table 3.1**, almost all hydrazide derivatives showed highly efficient ligation yields (> 90%), except **1n** which has low solubility in the reaction buffer. According to the study by Chio et al., mTG also displays a broad substrate tolerance toward the hydrazide nucleophiles.[121] However, VyPAL2 is much more efficient since a very small enzyme-to-

substrate ratio (1:2000) is required as compared to mTG (1:10).[121] All products were characterized by ESI-MS. Detail information is found in **Appendix B**. Arrow in **1g** indicates the direction of peptide sequence from the N to C terminus.

Table 3.1 Substrate scope of VyPAL2-catalyzed hydrazide ligation. Reaction yields were conversion yields based on HPLC analysis.



2 1a - 1p 3a - 3p

No	Substrate	Tme (min)	Yield (%)	No	Substrate	Tme (min)	Yield (%)
1a		10	95	1i		60	93
1b		60	73	1j		60	76
1c		60	> 90	1k		30	96
1d		30	> 92	1l		60	81
1e		30	> 90	1m		60	97
1f		60	> 93	1n		60	30
1g		30	> 90	1o		60	89
1h		60	91	1p		60	83

3.2.2 Exploration of conditions for hydrazide ligation reactions catalyzed by VyPAL2

Next, we determined the effects of the concentration of the nucleophile on the reaction. Using peptide **2** as the acyl donor (100 μM) and VyPAL2 as the catalyst (600 nM), the ligation reaction was conducted at room temperature at different concentrations of acethydrazide **1d**. As seen from **Figure 3.4**, the reaction was very fast especially when a high concentration of **1d** was used. At 50 to 200 mM, the reaction was completed in 5 min. In 30 min, a complete

reaction was achieved for a hydrazine concentration at 25 mM. And at 5 mM of **1d**, around 90% yield was obtained in 2 h. Most importantly, no detectable hydrolysis product was seen even after 4 h at a higher concentration (> 25 mM) of **1d** (**Figure 3.4**). However, the hydrolysis product was observed after 4 h when a low concentration (0.5 to 5 mM) of **1d** was used. These results suggest that the new peptide bond surrogate formed with acethydrazide is susceptible to hydrolysis and transamidation under the catalysis of VyPAL2, which is intriguingly similar to the behaviour of mTG.[10] Over extended incubation, hydrolysis could become predominant at low concentrations of acethydrazide whereas the exchange reaction would persist at high concentrations of acethydrazide which would form the same product. As further proofs for the exchange reaction, incubation of **3d** – the product resulted from **1d** – with **1m** in the presence of VyPAL2 indeed formed – albeit slowly – **3m** as the expected exchange product (**Figure 3.5**).

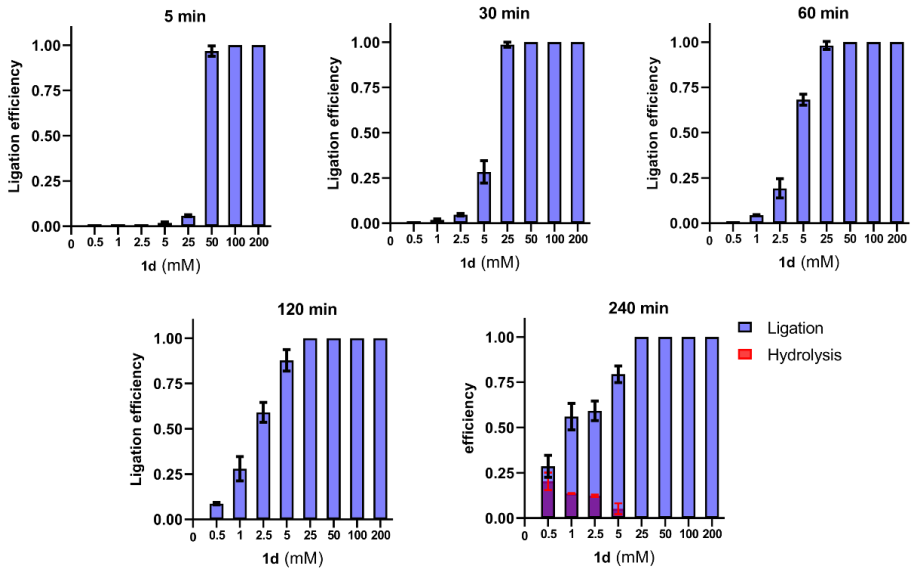


Figure 3.4. ESI-MS characterization of peptide **2** (100 μM) reacting with compound **1d** at different concentrations ranged from 0.5 to 200 mM at pH 7 for 5, 30, 60, 120, 240 min, respectively.

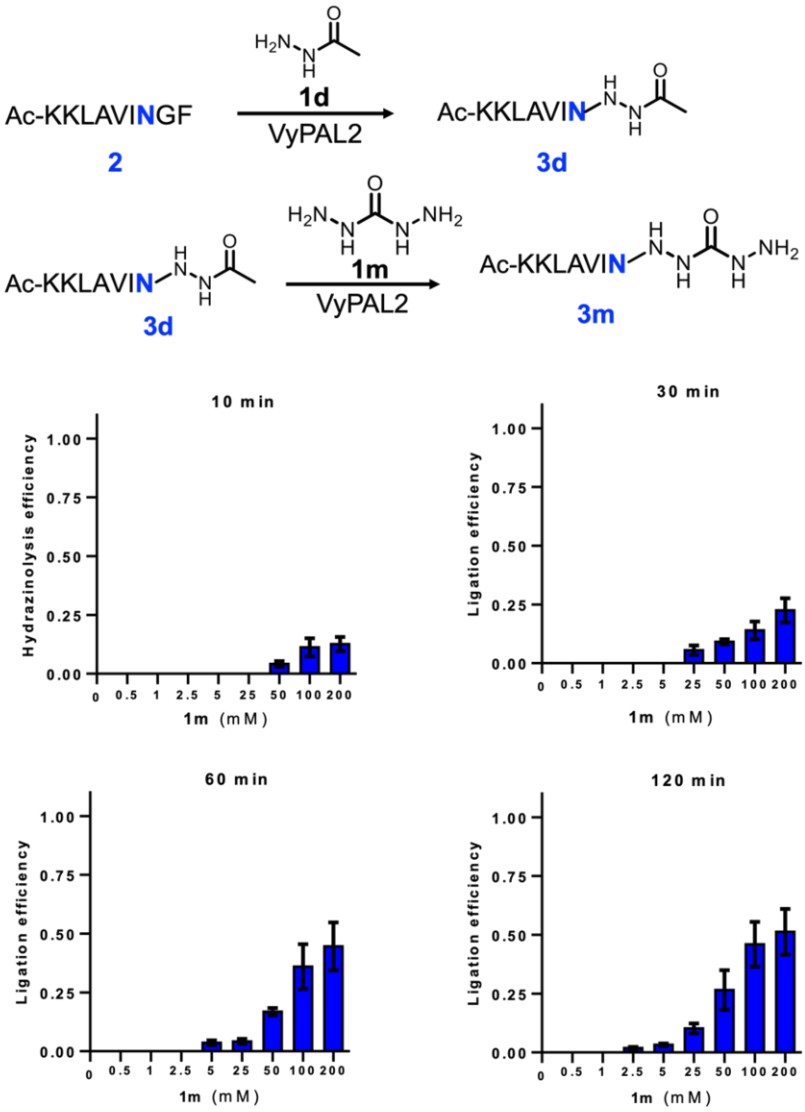


Figure 3.5. Exchange reaction of the hydrazide ligation product with another hydrazide. The reaction was carried out by adding 1 μM VyPAL2 to a mixture of 50 μM peptide **3d** and 0.5 uM – 200 mM compound **1m**. The mixture was maintained at 37 °C for 10-120 min at pH 7, and MALDI-TOF was used to detect the products and determine the exchange and hydrolysis yields.

3.2.3 Synthesis of peptide 5b – 5j

The reactions were performed by adding 25 nM of VyPAL2 to a mixture of 50 μM peptide **4**-NGL substrate (1 eq) and 10 mM nucleophile **1h**, **1g**, **1d**,**1e**, **1j**, **1f**, **1l**, **1o**, or **1p** (200

eq) at 37 °C and pH 7.0 for 30-60 min. The products were purified using HPLC and lyophilized. Peptides were dissolved using DD H_2O to a concentration of 100 µM and 10 mM of $AgNO_3$ in 10% acetic acid were added for 30 min at 25 °C. The mixtures were then subjected to HPLC purification to acquire **5b**-**5j** (**Figure 3.6** and detailed data shown in **Appendix B**).

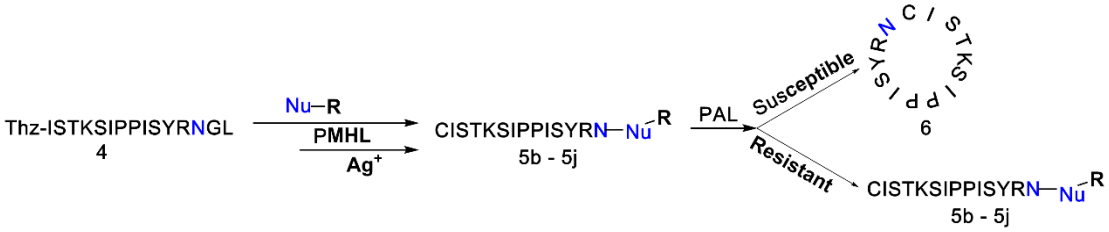
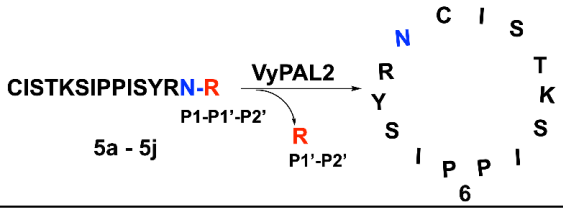


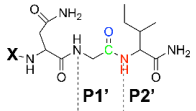
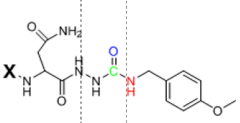
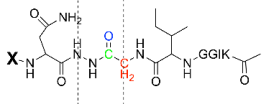
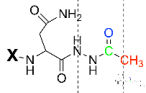
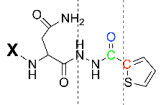
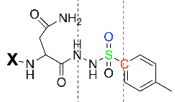
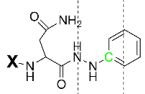
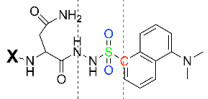
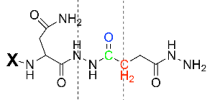
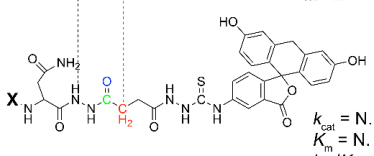
Figure 3.6. The scheme for the peptide synthesis of **5b** – **5j**. Firstly, peptide **4** was synthesized via SPPS with Thz at the N terminus of the peptide. After the ligation of hydrazide derivatives via PMHL and deprotection of Thz, peptide **5b - 5j** with N terminus unprotected Cys were obtained. Next, the cyclization kinetics were detected using PAL enzyme.

The above results prompted us to evaluate the stability of the different linkages formed by the different hydrazide substrates toward the PAL enzyme. Thus, a method based on PAL-mediated cyclization was employed (**Table 3.2**). As an intramolecular ligation reaction, exchange of the C-terminal hydrazide with the N-terminal amine in the same peptide would represent a stringent test. Peptide **4** containing an N-ter Thz-Ile and a C-ter AsnGlyLeu was first synthesized. Using the same conditions as shown in **Table 3.2**, **4** was reacted with selected hydrazide compounds and the resultant hydrazide ligation products were further treated with Ag^+ to deprotect the Thz to give **5b-j** (**Figure 3.6** and detailed data shown in **Appendix B**). Finally, the kinetics of VyPAL2-mediated cyclization of **5b-j** were measured and compared with that of **5a** – a canonical peptide substrate (**Table 3.2**, **Figure 3.7**). As shown in **Table 3.2** and **Figure 3.7**, **5c** displayed a 3-fold decrease in cyclization rate (k_{cat}/K_m) compared to **5a**. **5c** has the P2' hydrophobic Leu residue dislocated by one atom from its normal position in **5a,**

which might make it harder to bind in the S2' pocket of the enzyme. In contrast, peptide **5b** showed a 3-fold increase in cyclization kinetics, which might be explained by the presence of the aromatic ring, which is similar to the side chain of Phe – a preferred P2 residue of VyPAL2. Interestingly, peptides **5e**, **5f**, and **5g**, which contain an aromatic ring at the equivalent P1' position but no P2' component, showed a dramatic decrease in cyclization kinetics. In addition, we found that peptides with particularly small substituted groups, such as acetyl, or large multi-aromatic groups, such as FITC, also showed slow kinetics in VyPAL-mediated cyclization. The kinetic differences observed among these peptides suggest that factors other than the unnatural peptide bond linkages may contribute to their enzymatic stability. Some of these linkages were essentially resistant to the PAL enzyme, which means that the hydrazide ligation reaction leading to the formation of these linkages is nearly irreversible. This represents a new way of overcoming the reversibility of PML ligation. [99, 121-124]

Table 3.2. Stability analysis of the hydrazide ligation products bearing different substitutions on the hydrazide.



R group	P1'-P2'	k_{cat}/K_m ($M^{-1}S^{-1}$)
5a		k_{cat} = 5.0 ± 0.1 s^{-1} K_m = 20 ± 1 μM k_{cat}/K_m = 240,196 ± 4.901 $M^{-1}s^{-1}$
5b		k_{cat} = 9.5 ± 1.9 s^{-1} K_m = 15.1 ± 0.9 μM k_{cat}/K_m = 634,921 ± 15,175 $M^{-1}s^{-1}$
5c		k_{cat} = 3.0 ± 0.2 s^{-1} K_m = 40 ± 1.5 μM k_{cat}/K_m = 75,527 ± 1,452 $M^{-1}s^{-1}$
5d		k_{cat} = 0.19 ± 0.001 s^{-1} K_m = 13.6 ± 1.2 μM k_{cat}/K_m = 13,968 ± 317 $M^{-1}s^{-1}$
5e		k_{cat} = 0.049 ± 0.002 s^{-1} K_m = 7.2 ± 0.2 μM k_{cat}/K_m = 6,819 ± 77 $M^{-1}s^{-1}$
5f		k_{cat} = 0.09 ± 0.007 s^{-1} K_m = 51 ± 5 μM k_{cat}/K_m = 1,783 ± 71 $M^{-1}s^{-1}$
5g		k_{cat} = N. D. K_m = N. D. k_{cat}/K_m = N. D.
5h		k_{cat} = N. D. K_m = N. D. k_{cat}/K_m = N. D.
5i		k_{cat} = 0.72± 0.01 s^{-1} K_m = 17.8 ± 0.3 μM k_{cat}/K_m = 40283 ± 1078 $M^{-1}s^{-1}$
5j		k_{cat} = N. D. K_m = N. D. k_{cat}/K_m = N. D.

For each reaction, VyPAL2 (2.5-50 nM) and the peptide substrates **5a**-**5j** (2.5-400 μM) were mixed and reacted at pH 6.5, 25 °C for 5-60 min. The reaction mixtures were subjected to

MALDI-TOF analyses to allow calculation of the reaction yields. Average yields and SDs were calculated from experiments performed in triplicate. N.D., not determined (too slow).

3.2.3.2 Reversibility analysis for peptides 5a - 5j using the cyclization reaction of these peptides catalysed by VyPAL2

The catalytic kinetics of VyPAL2 towards peptide substrates **5a**-**5j** were done at pH 6.5 (phosphate buffer), 25°C for 5-60 min. The peptides were at a concentration ranging from 2.5-400 µM. The reaction was proceeded, and the initial velocity was calculated based on MALDI-TOF MS signals generated from the consumption of peptide starting materials after formation of the cyclic products. (**Figures 3.7**).

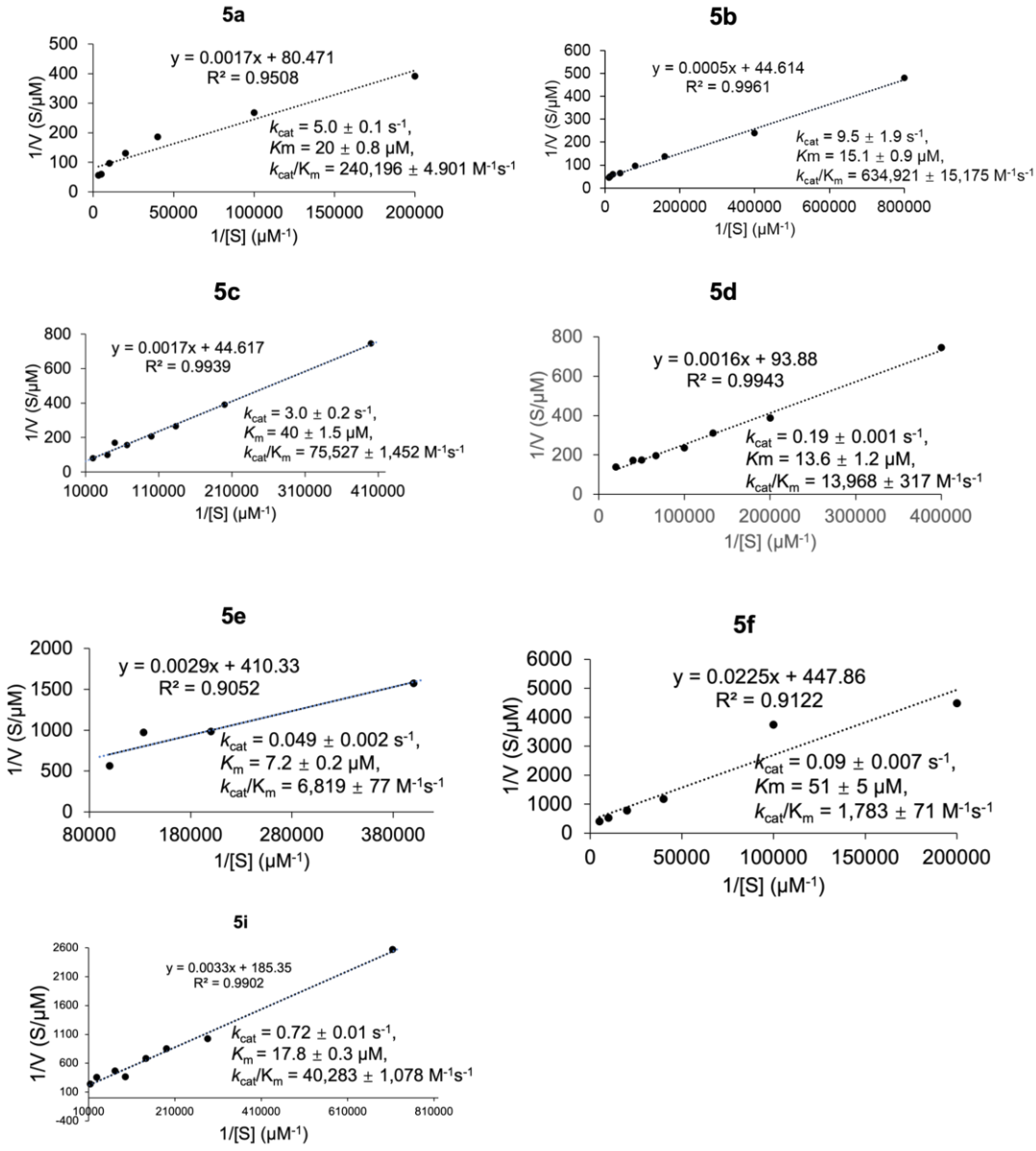


Figure 3.7. Enzymatic activity of the VyPAL2 toward peptides **5a**-**5e**, **5i**.

3.2.5 Molecular Dynamic simulation analysis

Given the difference in the cyclization rate for the different hydrazide substituted groups discussed above, herein we aim to unravel the intrinsic mechanisms of the enzymatic stability or instability. To achieve this, MD simulations were performed using the sequence of peptide **5a** and its variants with the C-terminus changed to -NGF, -NF **11, 12**. To be noted, the C-ter NGL peptide **5a** was used as a natural reference substrate, and -NF **11**, -NF **12** being equivalent

to peptide **5b** and **5g**, respectively (**Figure 3.8 A**). As shown in **Figure 3.8 B**, the -NGF **11** MD simulation result showed a π-aromatic interaction between the C-ter Phe of the substrate and the catalytic His162. This might cause the loss of catalytic activity in the PMHL reaction as His162 is critical to the catalytic activity of the PAL enzyme. While in the MD simulation result of -NGF, the C-ter Phe residue of substrate form intra-molecular π-π interactions with the side chain of Tyr. This prevents the Phe group from interacting with the catalytic His162, thus ensuring the PAL's catalytic activity to the substrate. An alignment between -NGF and -NGL simulation suggested a possible explanation that attributed to the increased catalytic activity on peptide **5b**.

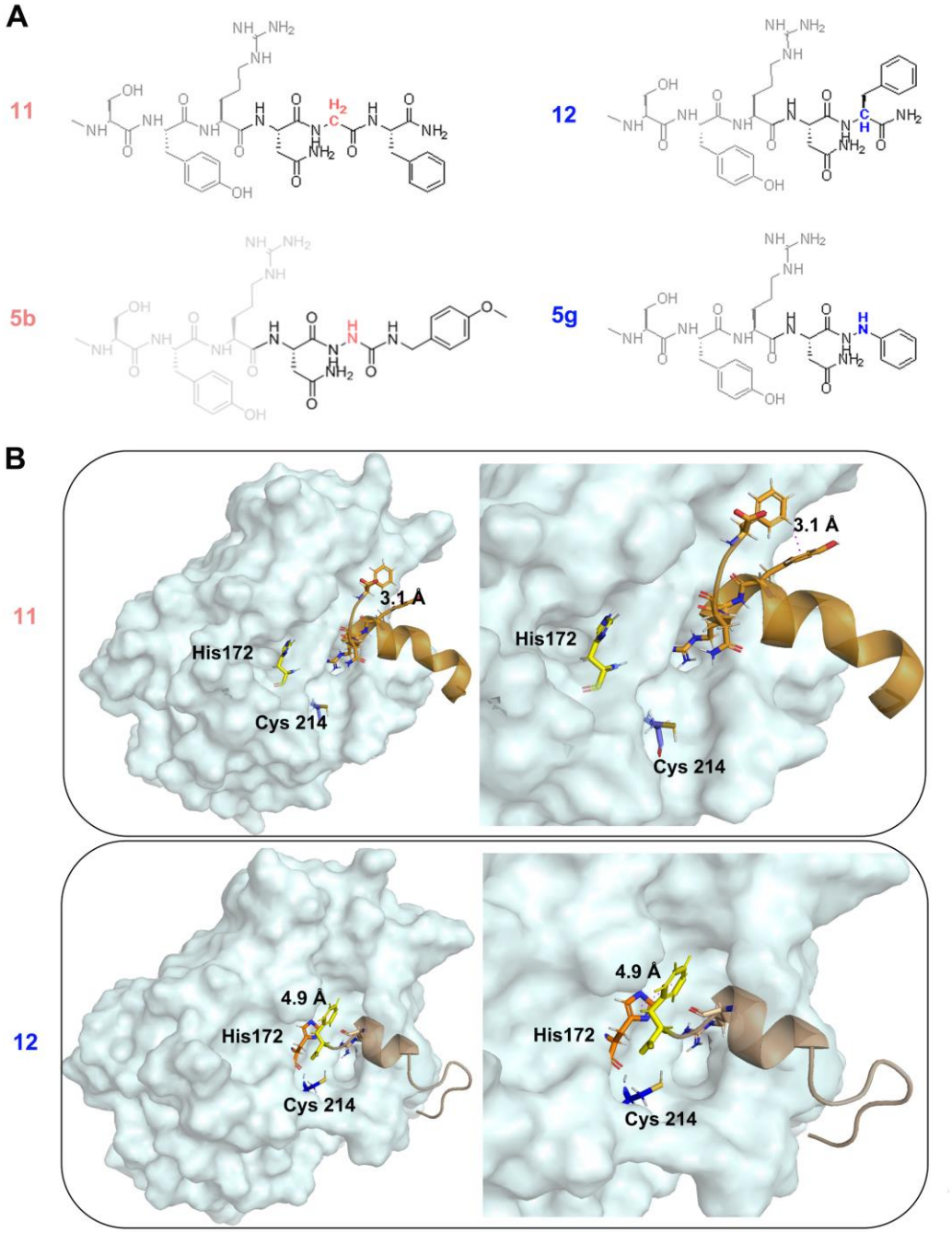


Figure 3.8. MD simulation studying the PAL–substrate interactions. A) Alignments of chemical structure between peptide -NGF **11**, -NF **12** at C-terminus. Note that -NGF and -NF is structurally similar to unnatural peptide **5b** and **5g**, respectively; B) MD simulation of the interactions between the -NF or -NGF substrate and VyPAL2.

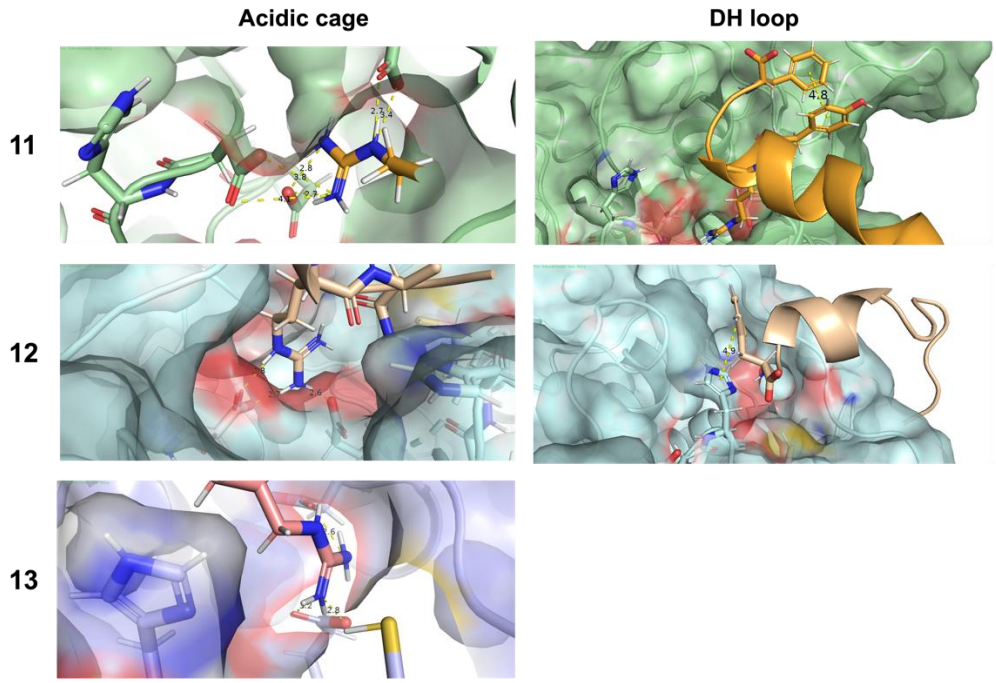


Figure 3.9 Structural analysis of NGF **11**, NF **12** and NGL **13** substrates. MD simulation studying the PAL–substrate interactions between the acid cage from VyPAL2 binding pocket and Arg from the substrate.

3.2.6 pH scan of OaAEP1b-mediated P1-Asx peptide ligation with acethydrazide

Compared with peptide N-terminal amines, the hydrazide-amines have very low pKa-H values (pKa-H << 4). Consequently, they should remain unprotonated, thus nucleophilic for PAL-mediated ligation at a rather low pH. Therefore, we conducted a pH scan experiment on both P1-Asn and P1-Asp peptide substrates. OaAEP1b was chosen as it was previously shown to have good catalytic activity over a broad pH range including acidic pHs. The reactions were performed by adding 10 nM of OaAEP1b to a mixture of 100 µM of the P1-Asp peptide **7** (C-ter DGL) or the P1-Asn peptide **8** (C-ter NGL) and 10 mM acethydrazide **1d** for 120 min at the pH range 4.0-7.5. From the pH profiles of both ligation reactions (**Figure 3.2.5**), we can see that, for P1-Asp substrate, the optimal pH was 4.0-4.5, which is consistent with previous findings that the P1-Asp side-chain COOH must be in its neutral form for effective binding to

its S1 pocket.[27, 30, 95, 113] For the P1-Asn substrate, the optimal pH range was 4.5-5.5 (**Figure 3.10**), which, surprisingly, is significantly lower than that for conventional peptide amines. Overall, we concluded that the pH of OaAEP1b-mediated hydrazide ligation using both Asx substrates has undergone a substantial drift to a more acidic pH. Of particular interest is the excellent catalytic activity of OaAEP1b toward P1-Asp substrates in the hydrazide ligation reaction. The discovery of this new characteristic will shed light on the development of orthogonal ligation schemes using the PAL enzymes. To be noted, both VyPAL2 and butease-1 exhibited very narrow pH range for efficient hydrazide ligation, especially for the P1-Asp substrate (pH 4.0-4.5). Previous OaAEP1b-mediated ligation reactions with normal amine nucleophiles had been conducted at pH 6.5 or slightly above because a lower pH would reduce the availability of the amine nucleophile due to protonation. The unique nucleophilicity of the hydrazides at low pH helped reveal the innate optimal catalytic pH of OaAEP1b as a ligase which is rather acidic. This suggests that, as long as there is an available nucleophile, the enzyme can effectively function as a cyclase at the acidic environment of the vacuoles in the host plant *Oldenlandia affinis*.

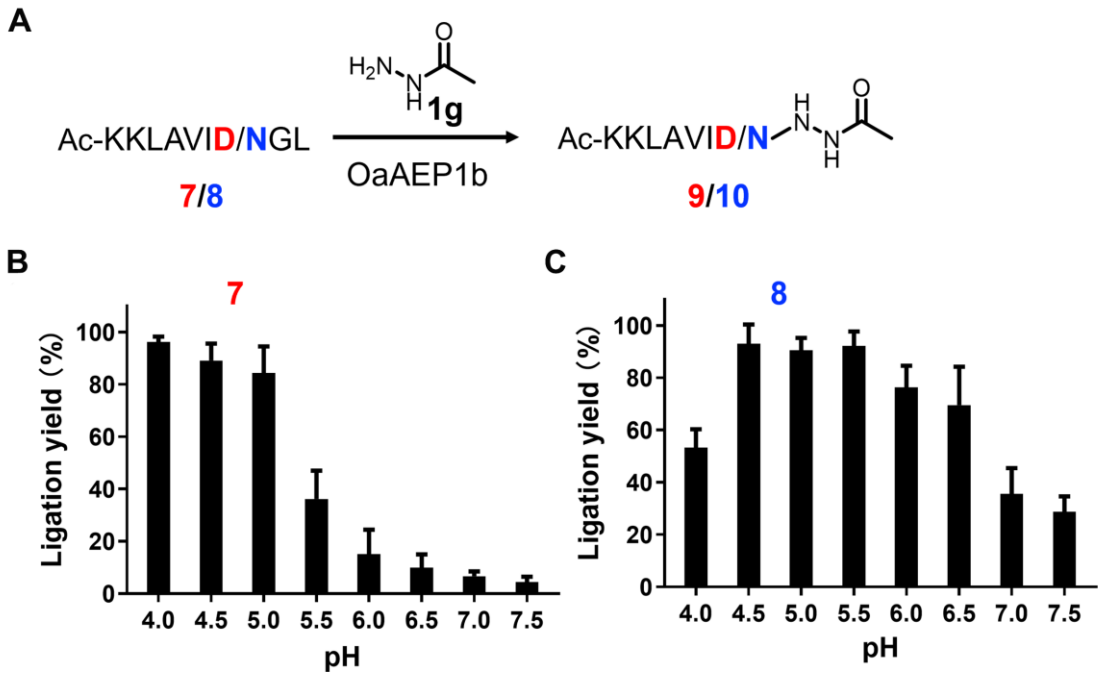


Figure 3.10. pH scan of OaAEP1b-mediated P1-Asx peptide ligation with acethydrazide. A) Scheme of OaAEP1b-mediated ligation with acethydrazide; B) OaAEP1b-mediated ligation of P1-Asp peptide **7** with acethydrazide at different pH; C) OaAEP1b-mediated ligation of P1-Asn peptide **8** with acethydrazide at different pH.

3.2.7 Synthesis of C-C fusion affibody dimer and thrombopoietin mimic peptide (TMP) dimer

The fact that peptide C-terminal hydrazides and bis-hydrazides can be effective nucleophile substrates of PALs allows one to conduct direct C-to-C ligation (**Figure 3.11 A and D**). We first attempted to prepare a C-C linked dimer of the affibody protein **15**. A bivalent peptide (peptide **14**) containing dual C-ter Ile-glycyl-$NHNH_2$ nucleophiles was prepared based on the previous published method.[98] To perform the reaction, 500 nM VyPAL2 was added to a mixture of 100 μM Z_{EGFR} **15** and 1 mM peptide **14** at pH 7.0 and 37 °C. After 1 h, the reaction was subjected to HPLC analysis. We found that an estimated 20% yield of the desired product **16** could be isolated. Besides **16**, two other by-products - the incomplete Z_{EGFR} single peptide ligation **17** and cyclized products **18** could be found in the HPLC profile (**Figure 3.11**). Nonetheless, the pure desired product **16** could still be isolated and characterized using ESI-MS (**Figure 3.11 B** and **Figure 3.11 C**). Next, we used this method for direct C-to-C dimerization of the thrombopoietin mimic peptide (TMP).[126] This could be achieved by the ligation of TMP C-ter hydrazide nucleophile TMP **19** with TMP-NGL **20**, enabling the production of **21** (**Figure 3.11 D**). 50 nM of VyPAL2 was added to a mixture containing 25 μM **19** and 125 μM **20**, and the reaction mixture was kept at 37 °C for 60 min and then subjected to HPLC analysis and LC-MS characterization. As shown in **Figure 3.11 E**, the reaction afforded the product in *ca.* 80% yield. Based on flow cytometry analysis, a 6.5-fold enhancement of binding affinity in terms of disassociation constant (K_D) was observed for the

dimeric **21** over the monomer **20** (**Figure 3.11 G** and **H**). Besides, the other hydrazide peptide ligation with affibody mediated by VyPAL2 was also tested in our work (detailed data shown in **Appendix B**).

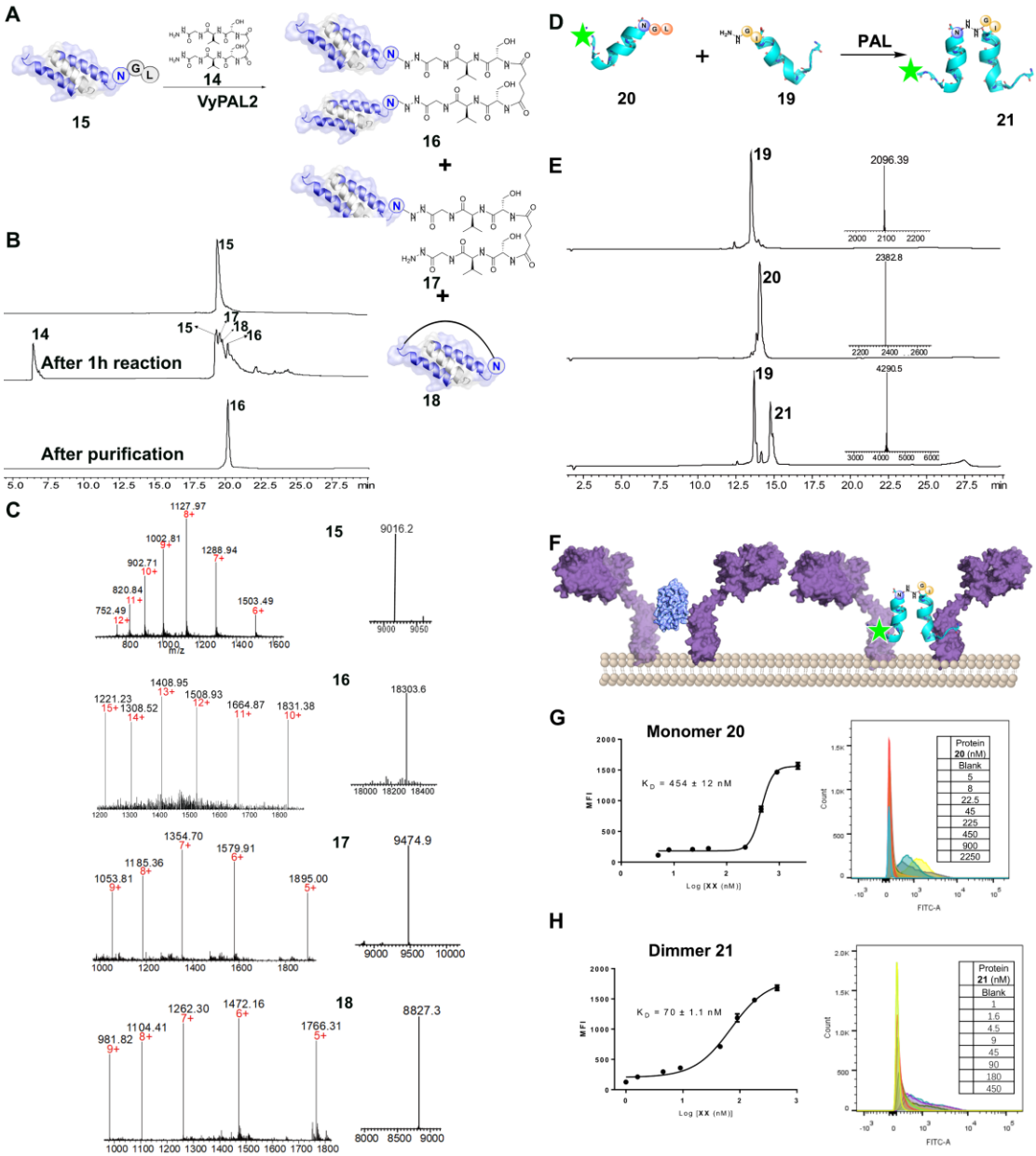


Figure 3.11. Hydrazide-enabled C-to-C ligation. A) Scheme of Z_{EGFR} -NGL**15** ligating with dimeric hydrazide peptide **14** using VyPAL2. The reaction gave the desired C-C fusion product **16** and the incomplete ligation by-product **17** and cyclized by-product **18**. B) HPLC profiles of VyPAL-mediated ligation between Z_{EGFR} **15** and peptide **14** (Upper: starting material **15**;

Middle: VyPAL-mediated ligation between Z_{EGFR} **15** and peptide **14** for 1 h; Lower: purified product **16**). C) ESI-MS characterization of each product in the process of synthesizing C-C fusion affibody. (**16**: calcd 18302.4, obsvd 18303.6; **17**: calcd 9474.3, obsvd 9474.9; **18**: calcd 8828.0, obsvd 8827.3. D) Scheme of PAL mediated C-C ligation between monomer **19** and **20** for the synthesis of C-C fusion TMP dimer. E) HPLC analysis and ESI-MS characterization of the ligation reaction between **19** and **20**. F) Schematic illustration of TMP dimer **21** binding to the TPO receptor; K_D measurement of G) **20** and I) **21** and FACS data of H) **20** and J) **21** using the TPOR over-expressing BF/F3 cell line.

The dimeric TMP peptide was found to enhance cell proliferation by binding to the thrombopoietin receptor (TPOR)[125, 126] (**Figure 3.12**). Upon binding of TPO to the receptor and subsequent dimerization of TPOR, the JACK-STAT/MAPK signaling pathway is activated, leading to downstream signal transduction. Activation of either the STAT or MAPK pathway results in increased production of effectors such as transcription factors in the nucleus, which ultimately promotes megakaryocyte proliferation and differentiation, as well as increased platelet production.[126]

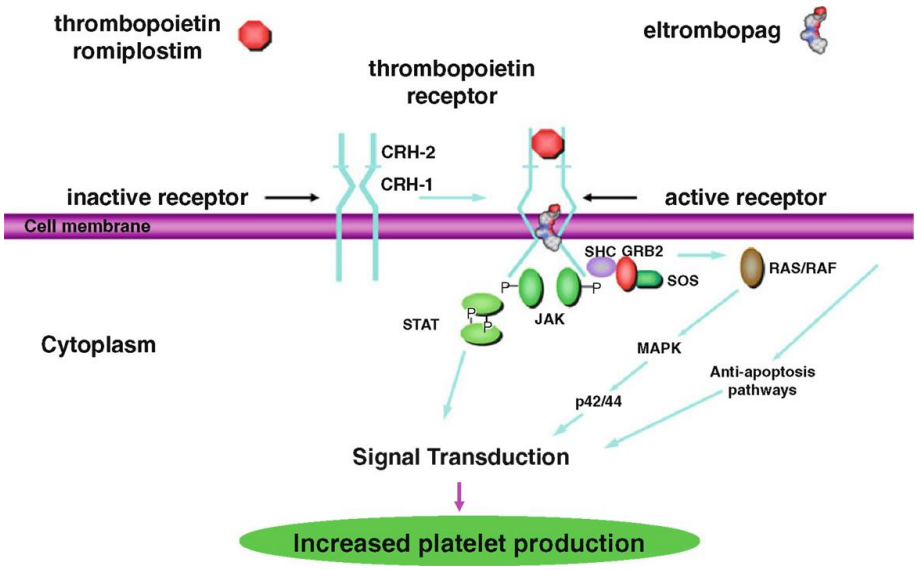


Figure 3.12. The TPOR activation by thrombopoietin via the STAT/MAPK pathways for increased platelet production.[126]

3.2.8 Use of PMHL for conjugating POI with hydrazide nucleophiles containing different payload groups

As an application of our hydrazide ligation scheme, we recombinantly prepared a dimeric affibody-Fc fusion protein **22** and evaluated its ligation with functionalized hydrazides using VyPAL2. These hydrazides **23-25** were prepared from dicarboxylic acids of varying spacer lengths (n = 0, 1, 2, 4) (**Appendix B**).

3.2.9 PAL-mediated ligation of antibody with hydrazide FITC (compound 23)

The reactions were performed by adding 500 nM VyPAL2 to a mixture of 50 μM **22** and 1 mM **23a-d** at pH 6.5 and 37 °C for 2 h. The reactions were then subjected to FPLC purification. To monitor the reactions, aliquots of the reaction mixture were taken and added to 6 M guanidine with 50 mM TCEP and heated in a headblock (95 °C) for 5 min and subjected to HPLC purification and ESI-MS characterization. The bulk of the reaction solution was subjected to Protein A FPLC purification to remove the reactants. It turned out that, when the hydrazides were functionalized with bulky groups (FITC), the hydrazide ligation reaction could not proceed when the spacer length was too short (n = 0 or 1) (**Figure 3.13**, **Figure 3.14**). This is likely due to steric hindrance imposed by the bulky groups. (**Figure 3.15**). Thus, a long spacer led to an excellent yield of hydrazide ligation when the functional group was FITC.

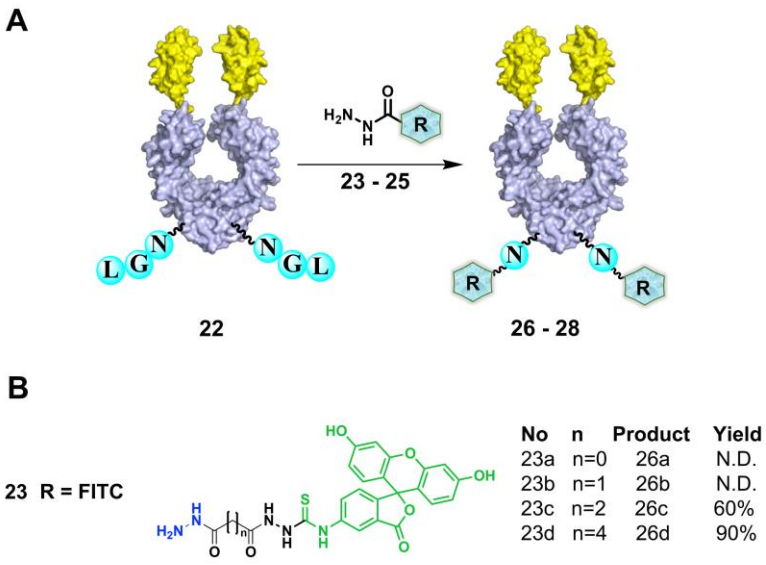


Figure 3.13. Use of PMHL for conjugating the affibody-Fc fusion protein with hydrazide nucleophiles containing different payload groups. A) Scheme of the ligation reaction between the POI (here the dimeric affibody-Fc) and hydrazide-FITC; B) Optimization of the spacer length between the reactive hydrazide and the payload group. Abbreviations: FITC, Fluorescein isothiocyanate. **24**, R = DOX, doxorubicin; **25**, R = MMAE, monomethyl auristatin E. Because of the hydrophobic nature of DOX and MMAE, the corresponding hydrazide compounds **24** and **25** have low solubility in aqueous buffers, making them difficult to handle during the ligation reactions.

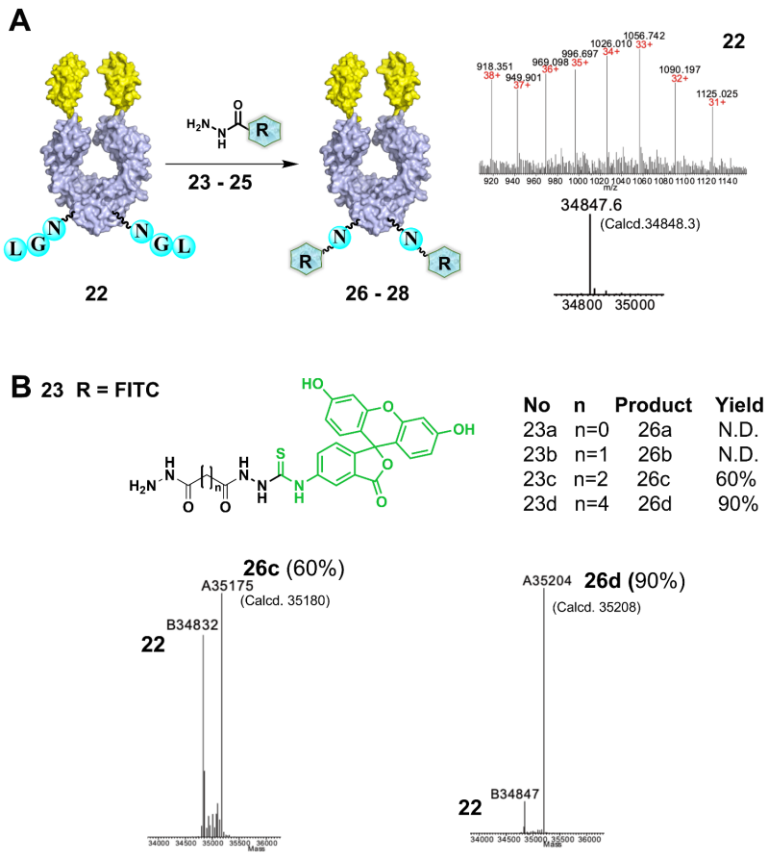


Figure 3.14. Monitoring of PMHL of the POI with hydrazide nucleophile containing different payload groups. A) ESI-MS characterization of the MKK-affibody-Fc-NGL **22**; B) Labeling and ESI-MS characterization of **22** with FITC hydrazide nucleophile **23c** (n = 2) and **23d** (n = 4) to give ligation product **26c** and **26d**, respectively.

Next, we aimed to explore the mechanism that regulates the stability of large molecule substituted groups (peptides **23-25**), as shown in **Figure 3.15**. As FITC is composed of multiple aromatic rings, the MD simulation cannot be mimicked by any amino acid residue. Considering the bulky characteristics of this substitution, we hypothesized that the irreversibility might be related to the structural hindrance of the acyl component in coordinating to the substrate binding pocket on the surface of the enzyme. Thus, we set out to calculate the surface area and volume of the pocket around the catalytic center. The calculation was done by comparing the binding pocket of VyPAL2 before and after substrate binding. As shown in **Figure 3.15**, the

volume of pocket 1 (acyl peptide substrate binding pocket) and pocket 2 (nucleophile substrate entering pocket) at the initial state was calculated as 14.9 and 1.6 SA, respectively. Subsequently, upon the binding of the substrate, the volume of pocket 1 and pocket 2 increased to 79 and 36 SA, respectively, along with a new pocket 3 (electrophile substrate binding pocket) appearing at a volume of 26.4 SA (**Figure 3.15**). Compared to pocket 3, pocket 2 showed a larger volume, which suggested that the nucleophile is more easily able to enter and subsequently attack the acyl-enzyme intermediate. The voluminous difference between the nucleophile substrate entering and electrophile binding pocket might provide a possible explanation for why some nucleophiles with bulky substituted groups are relatively easy to be conjugated to POI but more difficult to be used as a substrate again by the same ligase. In conclusion, this disparity between nucleophile and electrophile binding pockets provides enlightenment for the understanding of the irreversibility of PMHL reaction with hydrazide nucleophiles substituted with bulky groups.

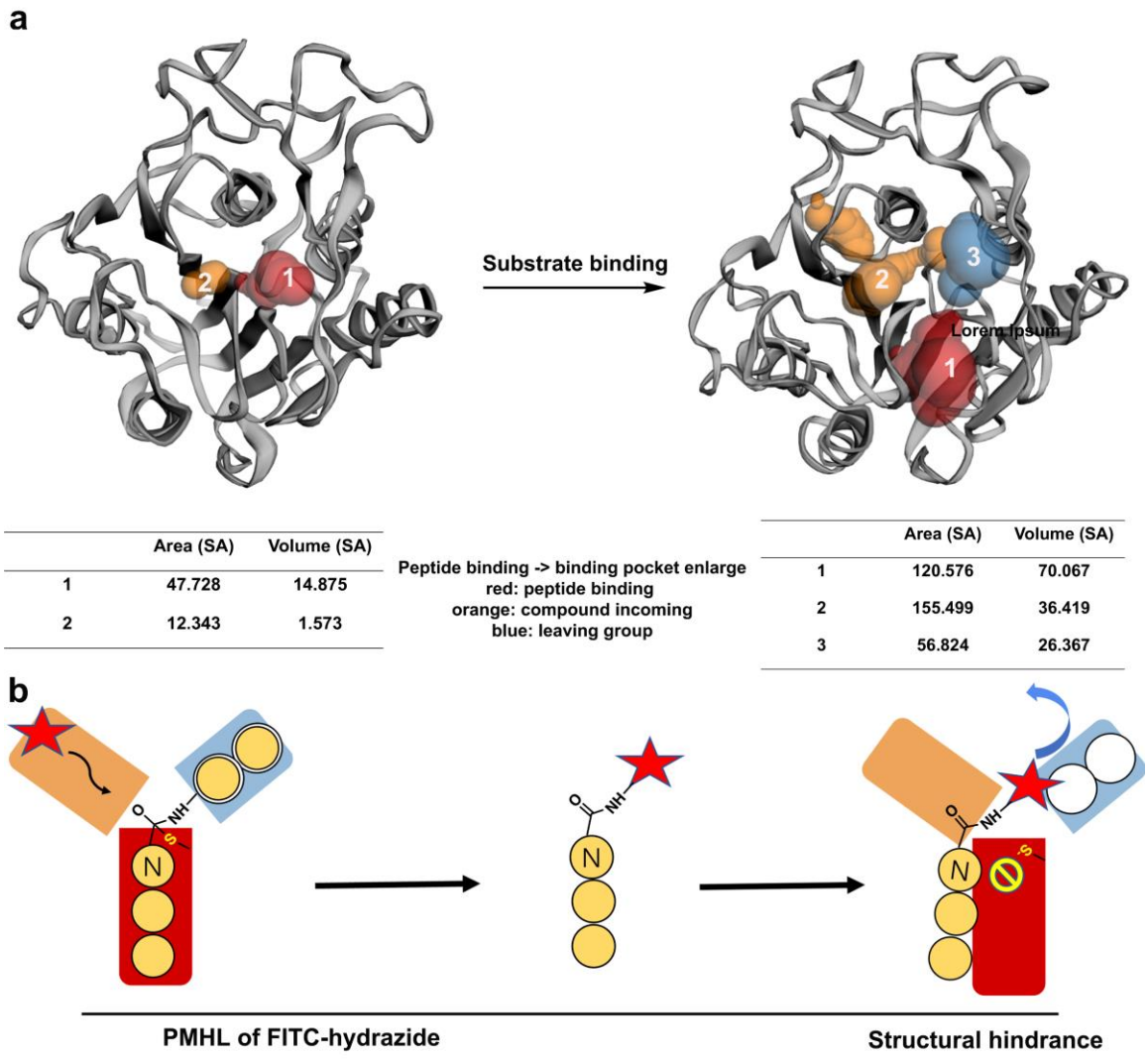


Figure 3.15. Exploration of using PMHL for conjugation of POI with hydrazide nucleophile containing a bulky group. A) Surface calculation of VyPAL2 before and after peptide substrate interaction; B) Scheme of POI ligating with hydrazide-FITC to illustrate a proposed mechanism that governs the irreversibility of the reaction.

3.2.10 A two-step scheme for labelling proteins with the dihydrazide compounds and then conjugation with the desired payload

Given that steric hindrance and low solubility of certain functionalized hydrazides may limit the efficiency of PMHL, we further devise a two-step scheme for labelling proteins with the desired payload compounds (**Figure 3.16**). In this scheme, the POI is first ligated with a highly reactive hydrazide through PMHL; at the second step a conventional conjugation or

click reaction is performed to load the payload compound onto the protein through a preinstalled handle. Therefore, we treated the same Z_{EGFR}-Fc fusion protein **22** (25 µM) with 10 mM succinic dihydrazide in the presence of 200 nM VyPAL2 at 37 °C and pH 7 for 1.5 h. The ligation reaction furnished product **22**-dihydrazide in 95% yield based on mass spec analysis (**Figure 3.17**). The product was purified using FPLC (Protein A) and characterized by ESI-MS. DOX or MMAE was then conjugated with **22**-dihydrazide via hydrazone formation using well-established protocols. 25 µM **22**-dihydrazide was mixed with 1 mM DOX or 250 µM CHO-MMAE in the presence of 5 mM aniline at 37 °C, pH 6 for 4 h to give **27c** and **28c**. Then, 5 mM sodium borohydride was added, and the mixture was kept at 25 °C for 30 min to reduce the hydrazone bond. The product was purified using protein A FPLC and characterized by ESI-MS. The reactions in both steps gave excellent yields (> 95%). For ESI-MS characterization of products in both steps, disulfide reduction and denaturation were done by treating the sample with 5 mM TCEP in 6 M guanidine in a 95 °C heat block for 5 min.

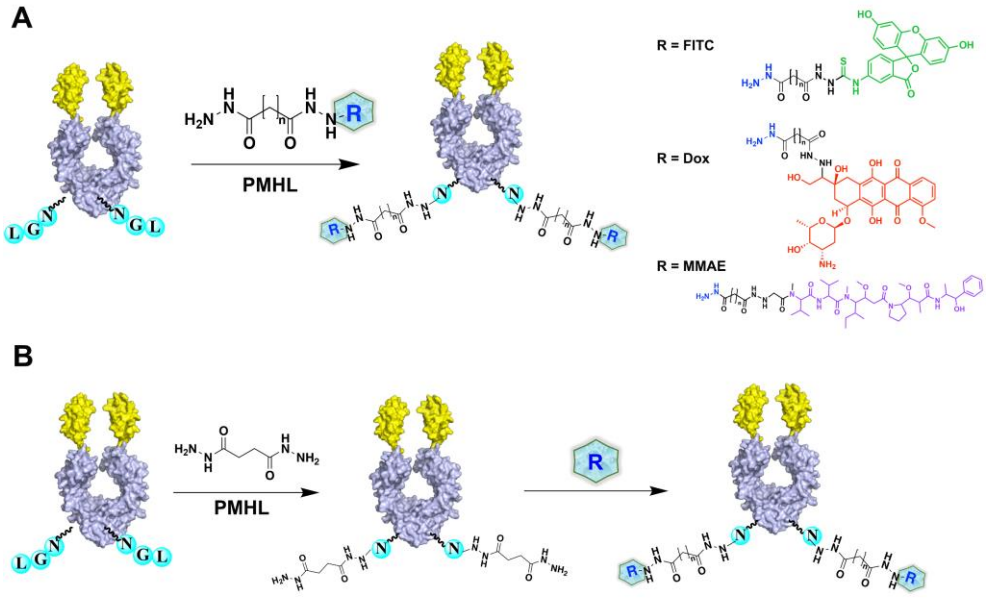


Figure 3.16. Two PMHL strategies: A) One-step scheme for labelling proteins with the desired payload compounds. B) Two-step scheme for labelling proteins with the dihydrazide compounds followed by conjugation with the desired payload.

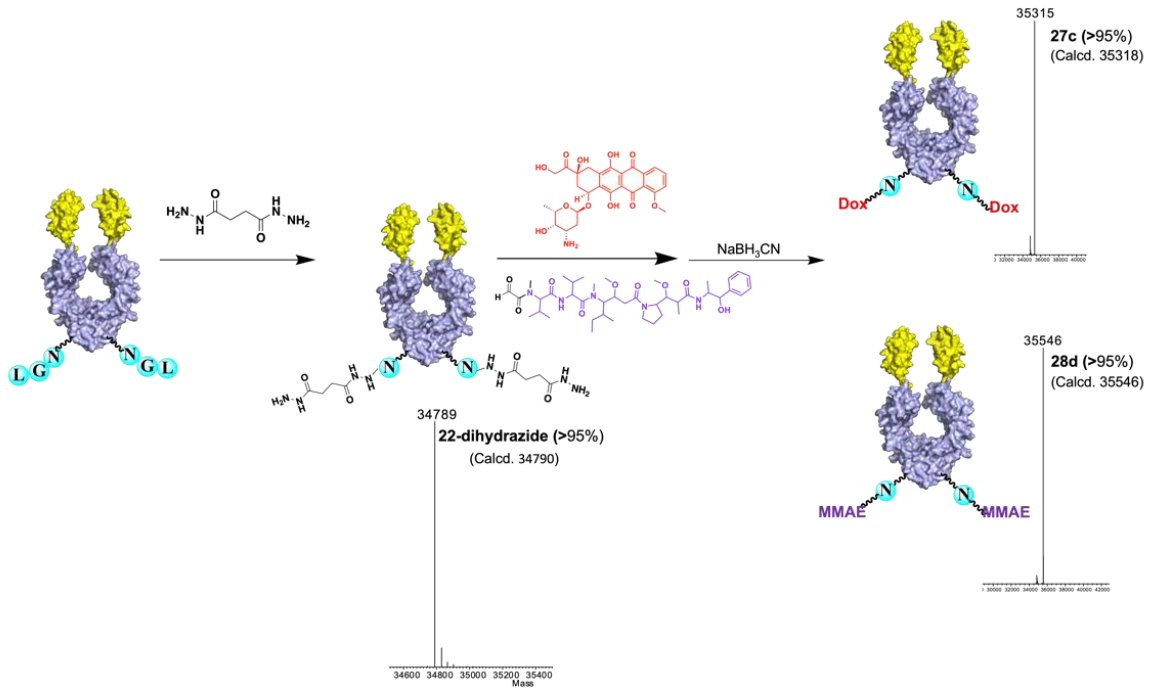


Figure 3.17. Protein-drug conjugation via two-step reactions.

3.2.11 Tandem ligation and bispecific engager for NK-scFv CAR cell therapy

Chimeric antigen receptor T cells (CAR T cells) have shown great success in the treatment of certain blood cancers such as acute lymphoblastic leukemia (ALL).[56] However, the production of CAR-T cells for therapeutic use is a cumbersome process that limits their clinical potential.[57] Natural killer (NK) cells are immune effectors with favorable characteristics for practical use because of their high cytotoxicity [56], lack of risks to cause graft-versus-host disease (GVHD) [58], and the ability to use alternative cytotoxic mechanisms to reduce the risks caused by loss of CAR-specific antigen [58]. These inherent qualities make NK a more ideal candidate for immunotherapy. Thus, engineering CAR-NK cells with a controllable mechanism of action is a new subject under intense investigation.

A folate-FITC conjugate has been reported as a bispecific adaptor molecule for controlled folate-receptor CAR-T cell therapy.[61, 62] However, folate receptor is expressed in many normal tissues, and the endogenous folate might compete with the conjugate. Undoubtedly, the

artificially designed Z_{EGFR}-Fc fusion protein has the advantages of being more selective for the tumor antigen. The larger molecular weight of the fusion protein and the effect of the Fc domain also give it a long half-life in vivo, which reduces the need for frequent drug dosing. Here, we synthesized a quadruple FITC-labeled affibody-Fc using a tandem ligation scheme involving first a C-terminal hydrazide ligation step and second a conventional ligation reaction step at the N-terminal amines.

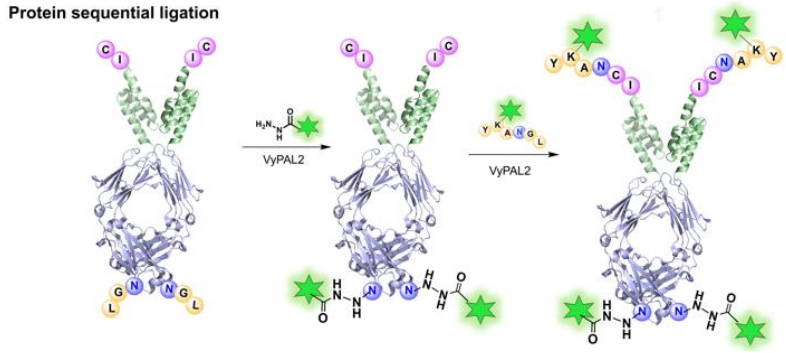


Figure 3.18. Tandem ligation modification of the bispecific engager for NK-scFv CAR cell therapy. Affibody-Fc is first labeled with hydrazide-FITC compound at the C terminus. Then the N terminus of the protein is further modified with FITC peptide containing NGL, leaving the C terminus of affibody-Fc unaffected by VyPAL2 due to the irreversible reaction of the hydrazide-FITC compound. Therefore, both the N and C terminus of affibody-Fc protein will be labeled with FITC, obtaining a quadruple FITC-labelled affibody-Fc.

3.2.12 Synthesis of FITC labeled affibody-Fc 30 and 32 using VyPAL2 for sequential hydrazide and conventional ligations

First, protein C-terminal labelling was achieved by adding 500 nM VyPAL2 to a mixture containing 25 µM CL-affibody-Fc-NGL **29** and 1 mM **23d**. The reaction was kept at pH 6.5 (phosphate buffer) and 37 °C for 2 h and then subjected to protein A-FPLC purification to afford **30** in *ca.* 85% yield. Second, the FITC-functionalized peptide **31** (250 µM) was reacted with **30** (25 µM) in the presence of 200 nM VyPAL2 at pH 6.5 (PBS) and 37 °C for 1 h, which

introduced two additional copies of FITC on the N-termini of the dimeric protein to afford **32** in *ca.* 90% yield. The product was isolated using protein A-FPLC purification. To characterize the products, **30** and **32** were dissolved in 6 M guanidine HCl containing 50 mM TCEP and treated at 95 °C for 5 min and subjected to HPLC purification and ESI-MS characterization (**Figure 3.19)**.

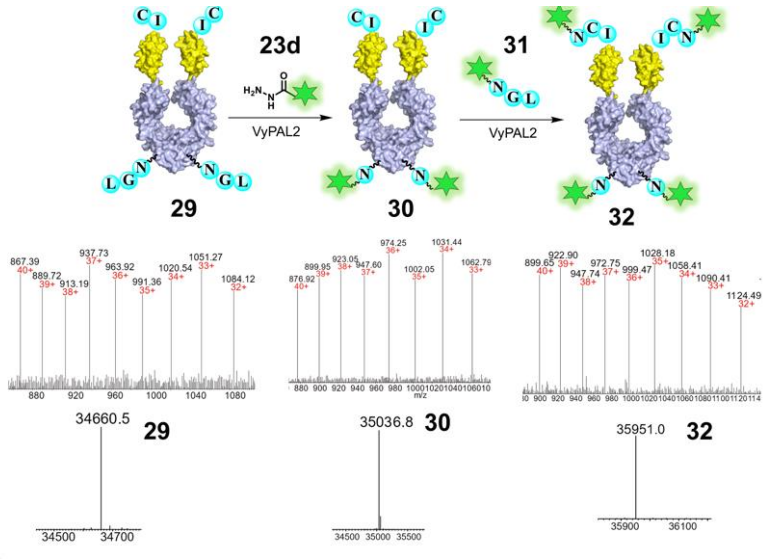


Figure 3.19. ESI-MS analysis of the FITC labeled affibody-Fc **30** and **32**. The tandem ligation was achieved by adding compound **23d** to protein **29** in the first step to label the C terminus of **29**. Then compound **31** was used to label the N terminus of protein 30, generating the sequential ligation product **32**. ESI-MS characterization of the reduced **29, 30** and **32** (**29**: Cacld. 34661.0, Obsvd. 34660.5; **30**: Cacld. 35036.0, Obsvd. 35036.8; **32**: Cacld. 35953.5, Obsvd. 35951.0).

3.2.13 Engineered NK cell line with scFv protein on the cell surface via lentivirus packaging

The NK cells (NK-92MI) were transduced sequentially with scFv-puromycin and pEf1a-mCherry lentivirus vectors using the same protocol as mentioned above. Stable cell lines were obtained using puromycin selection for the scFv-puromycin transduced cell line and then FACS selection for the cell line further transduced with the pEf1a-mCherry lentivirus vector.

The control NK cells were transduced with the pEf1a-mCherry lentivirus vector only. Next, we cultured the cells and carried out the following experiments. To visualize the interaction between A431 cells and NK-92MI cells, A431 and MCF-7 cells were transduced with pEf1a-GFP and lentivirus vector. Three days after the transduction, cells were added with puromycin (2 µg/µL) for 72 h.

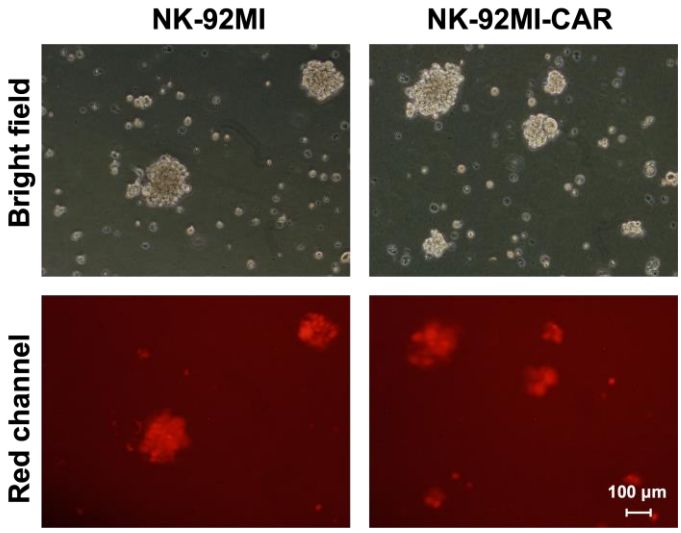


Figure 3.20. The infected control NK cells and CAR-NK cells with pEf1a-mCherry using the same method as mentioned above. CAR-NK cells were obtained after selection by puromycin (0.5 µg/µL). The NK and CAR-NK cells are successfully transfected with mCherry in red shown via 10X view. Around 90% of the cells are selected out and can be used for the next step of experiments. The scale bar is 100 µm.

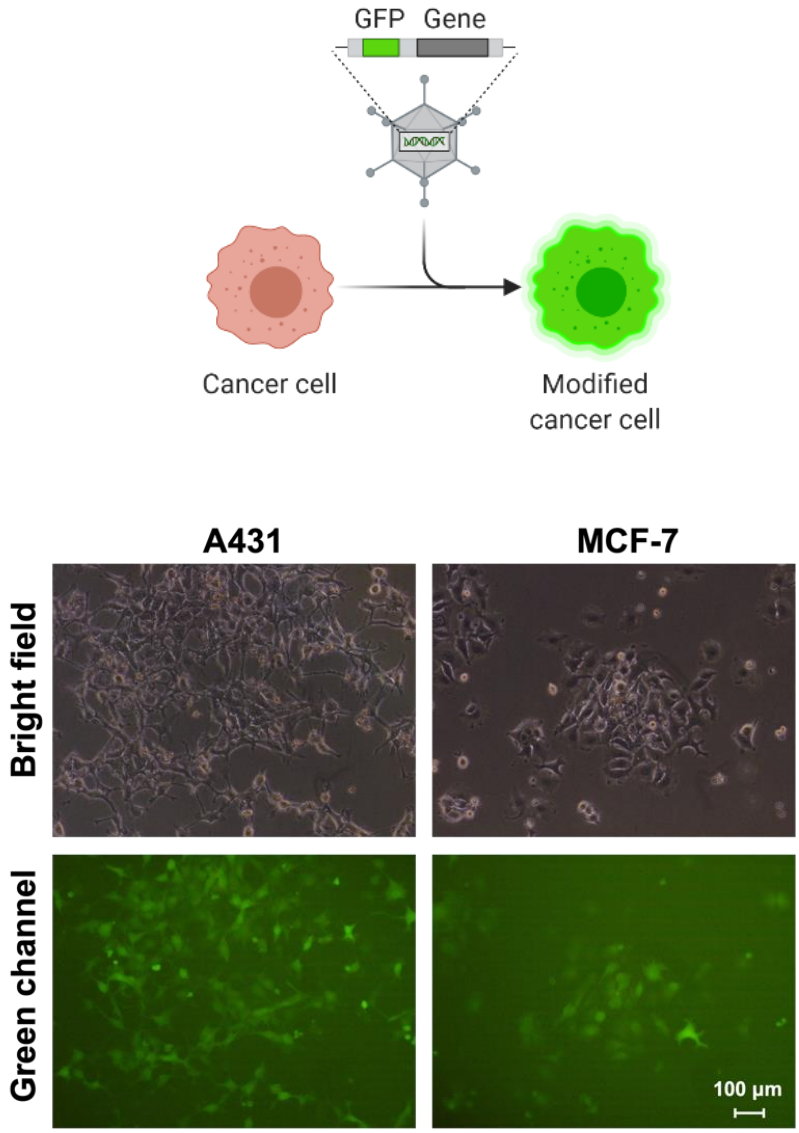


Figure 3.21. The infection of A431 and MCF-7 cells by the pEf1a-GFP after selection of puromycin (2 µg/µL). A431 and MCF-7 cells were transduced with pEf1a-GFP and lentivirus vector. Three days after the transduction, cells were added with puromycin (2 µg/µL) for 72 h.

3.2.14 FITC-labelled affibody-Fc as a bi-specific engager and its bioactivity assessment

A lentiviral cassette containing the gene of the anti-FITC scFV driven by the CMV promoter and the puromycin selection marker was constructed and used to engineer NK cells expressing the anti-FITC scFV as the CAR. The activity of the engineered CAR-NK cells could be controlled via adjusting the concentration of the FITC-labelled affibody-Fc **32**. As a bi-

specific engager, **32** can bind to EGFR-expressing cancer cells and CAR-NK cells simultaneously (**Figure 3.22A**). The flowcytometry-based K_D analysis showed that the dimeric affibody-Fc protein displayed a stronger binding affinity (K_D = 11.7 ± 1.0 nM) than the monomeric counterpart (K_D = 18.9 ± 1.2 nM) towards A431 cell lines (**Figure 3.22B**). To study the interactions between the two types of cells mediated by **32**, CAR-NK and A431 cells were transduced with lentiviral vectors carrying the genes for mCherry and GFP, respectively. After mixing CAR-NK cells with A431 at 5:1 ratio, the drugs **29**, **30** and **32** (25 nM) were added. After incubation for 30 minutes, cells were washed with PBS for 3 times and subjected to fluorescent microscopy analysis. As shown in **Figures 3.22E**, **3.22F,** CAR-NK cells treated with **30** and **32** presented a higher number ratio of NK/A431, indicating a strong interaction of CAR-NK with A431 which was mediated by FITC labelled affibody-Fc proteins. As shown in **Figures 3.22H**, cell specific lysis was also analyzed using FACS after adding 105 nM **32** to a mixture of CAR-NK and cancer cells. To evaluate the efficacy of **30** and **32** in inducing cytokine release, an anti-FITC CAR-jurkat cell line was mixed with A431 cells, followed by treatment with the two proteins at different concentrations. As shown in **Figure 3.22G**, the quadruple-FITC modified **32** presented a better potency (EC_{50} = 6.2 ± 1.2 nM) than their double-FITC modified counterpart **30** (EC_{50} = 15.9 ± 1.8 nM). The cytotoxicity was analysed using LDH-assay, and wild-type NK cells were used as the control. As shown in **Figure 3.22I**, CAR-NK cells showed enhanced cytotoxicity in the presence of 25 nM of **32**. The LDH-releasing analysis was then used to determine the IC_{50} value by treating the cells with different concentration of **32**. As shown in **Figure 3.22J**, the IC_{50} of **32** was determined at 6.7 ± 0.4 nM. These results show that the bi-specific engagers can mediate the killing of EGFR-expressing cancer cells by the engineered CAR-NK cells.

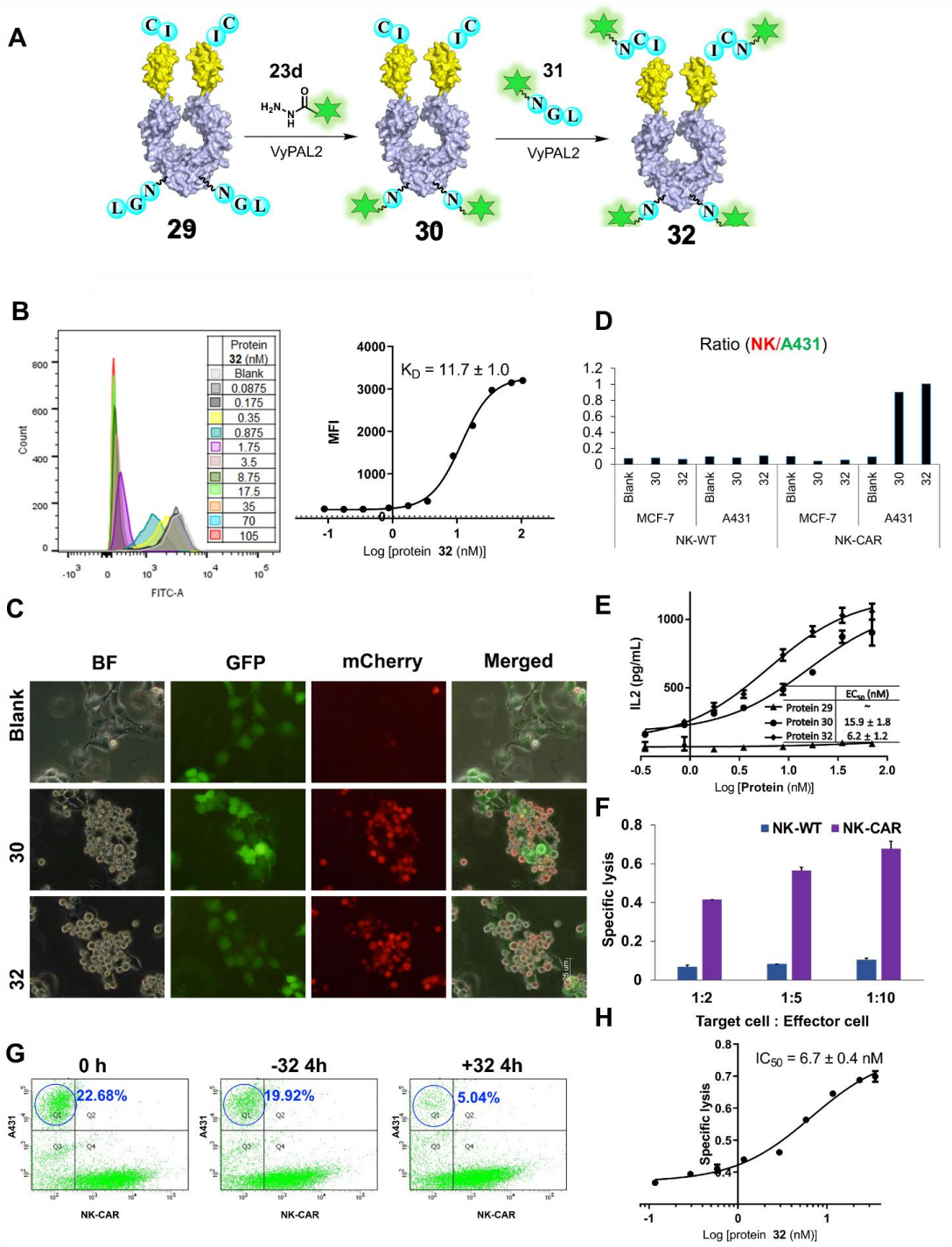


Figure 3.22. FITC-labelled affibody-Fc as a bi-specific engager and its bioactivity assessment. A) Scheme to synthesize **32** using VyPAL2 for sequential hydrazide and conventional ligation; B) K_D analysis of **32** using FACS; C) Analysis of **32**-mediated binding between NK-CAR and A431 cells using fluorescent microscopy (Green: MCF-7 or A431 cells transduced with GFP lentiviral construct; Red: NK-CAR transduced with mCherry lentiviral construct); D)

Calculation of ratio between CAR-NK (GFP) and A431 (GFP) based on image J; E) In-vitro analysis of cytokine secretion by CAR transduced Jurkat T cells stimulated by **30** and **32**; F) Cytotoxicity analysis of **32**-mediated A431 cell killings by NK-CAR cells using the LDH releasing assay; G) Flow cytometry analysis of the A431 cell number after incubated with NK-CAR cells without or with 105 nM protein **32** for 4 h; H) IC_{50} analysis of **32** in mediating CAR-NK cytotoxicity using flow cytometry.

3.2.15 The binding analysis between A431 cells with NK-Car cells

Next, to detect the cell-to-cell interaction, Car-NK cell is transferred into the A431 cell at around 5:1 ratio. Protein **32** was added to the cell mixture at 105 nM for 30 min in 37 °C. However, no protein drug was added to cells in the control group. After incubation and washing the cells using PBS for 3 times, fresh PBS buffer was added, and cell-cell interactions were observed under fluorescent microscopy (**Figure 3.23 and 3.24**) and the number of interacted cells were estimated (**Figure 3.25 and 3.26**).

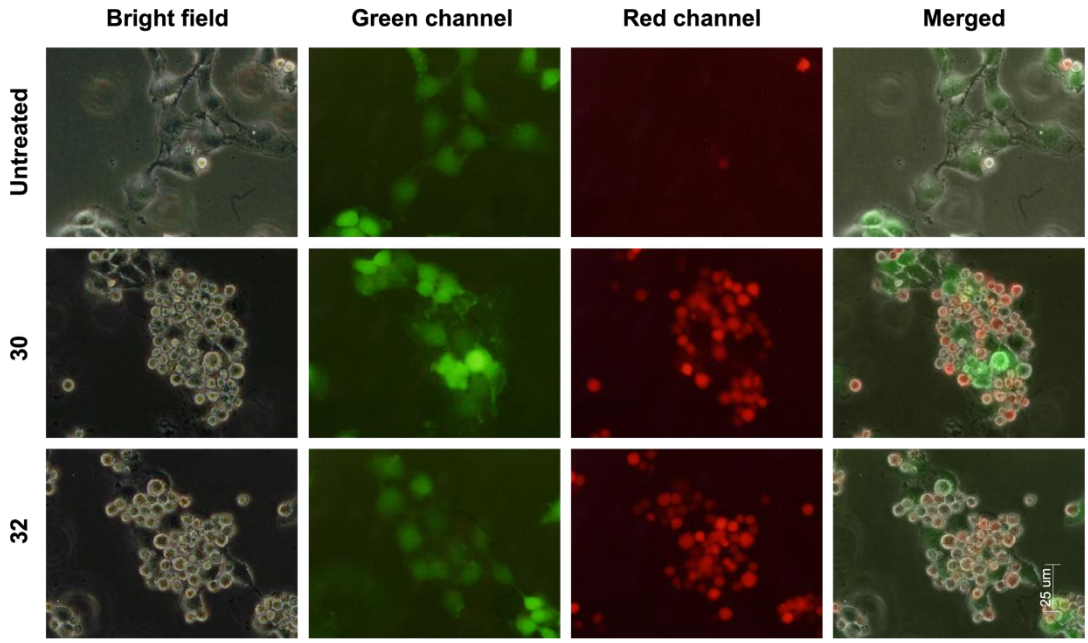


Figure 3.23. Fluorescent microscopy analysis of the interactions of A431 cells with Car-NK cells mediated by the bi-specific engager protein **32** at 105 nM for 30 min in 37 °C.

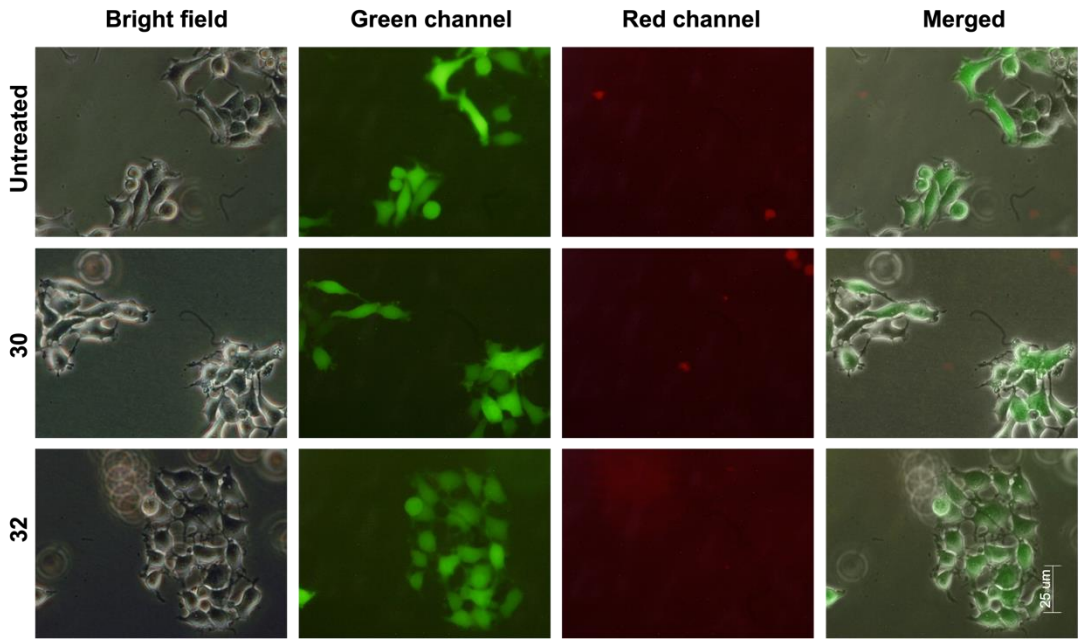


Figure 3.24. Fluorescent microscopy analysis of the interactions of MCF-7 cells with Car-NK cells mediated by the bi-specific engager protein **32** at 105 nM for 30 min in 37 °C.

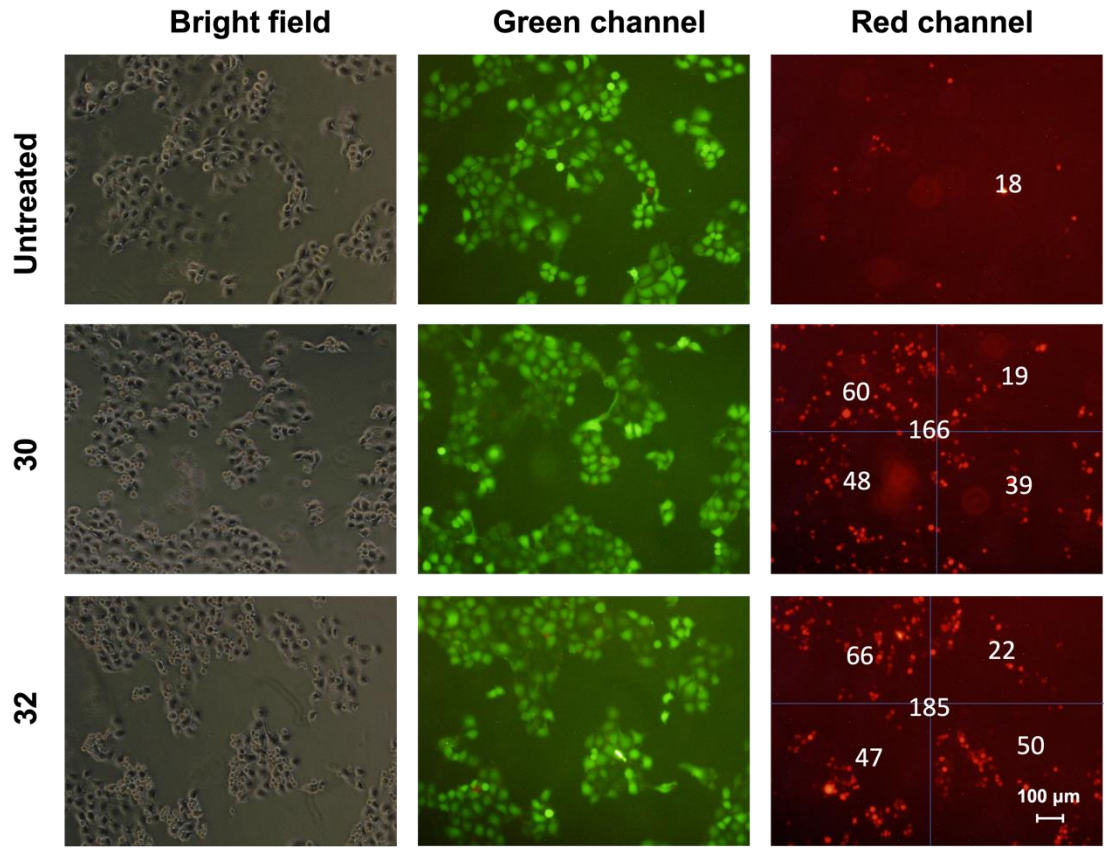


Figure 3.25. Calculation of the number of interacted cells of A431 with Car-NK mediated by the bi-specific engager protein **32** at 105 nM for 30 min in 37 °C at a 10X view.

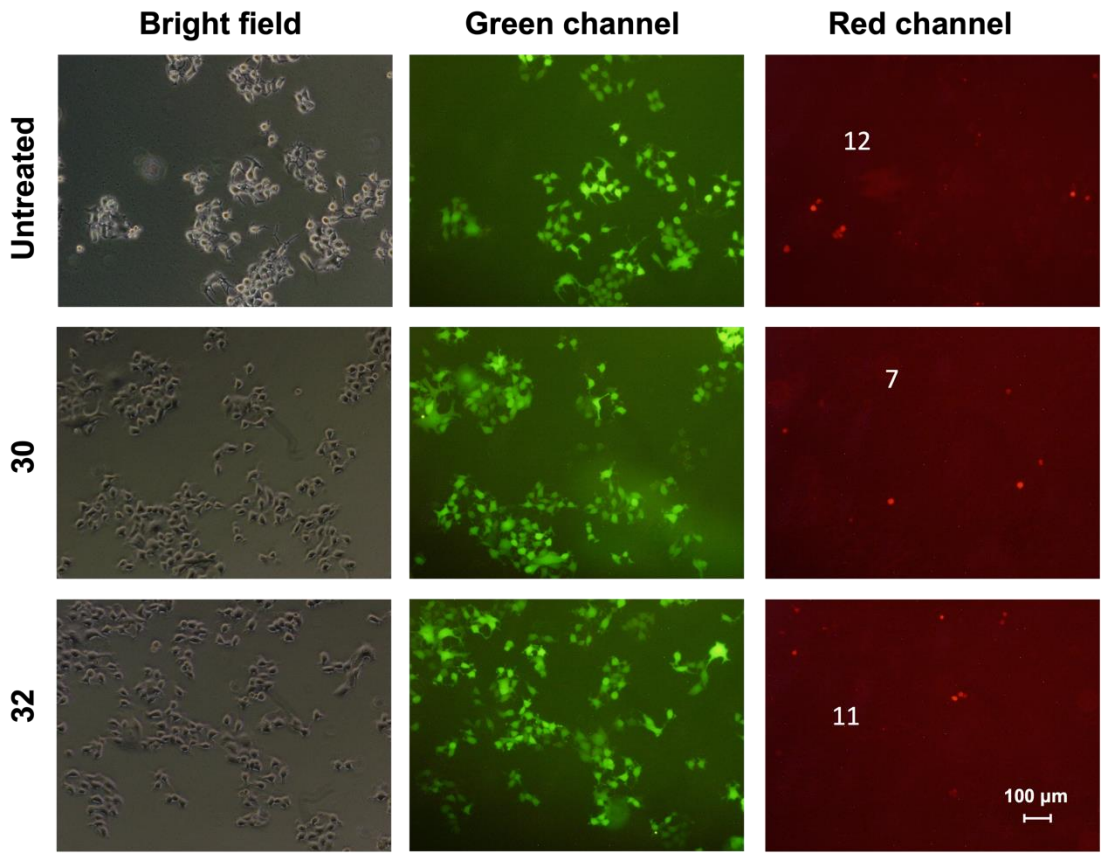
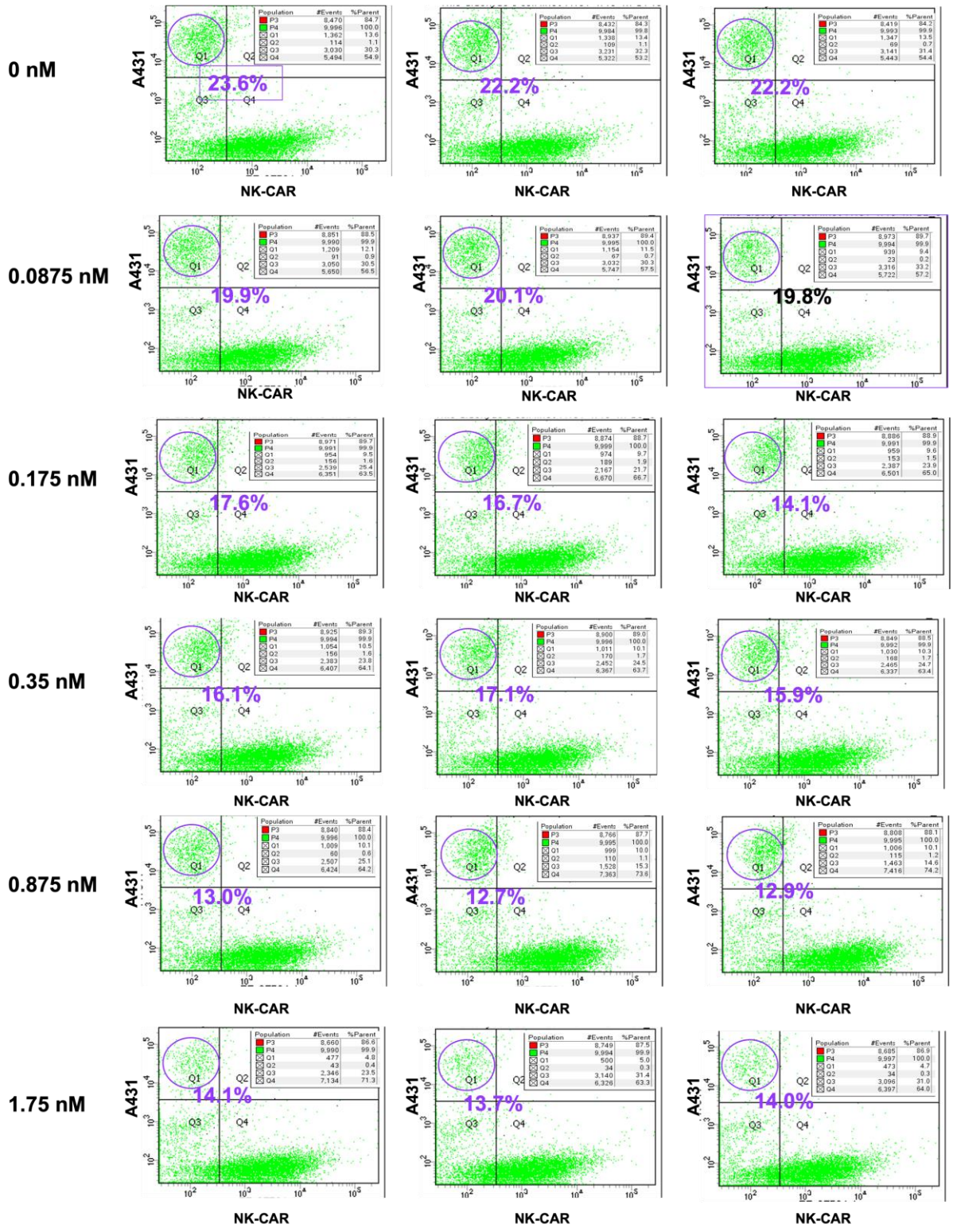


Figure 3.26. Calculation of the number of interacted cells of MCF-7 with Car-NK mediated by the bi-specific engager protein **32** at 105 nM for 30 min in 37 °C at a 10X view.

For calculation of the bi-specific engager mediated NK-92MI lysis. Protein **32** was added at different concentrations for 4 h. Next, we detached the control cells and the cells treated with protein **32** using trypsin. And centrifugation, fresh PBS buffer was added and the suspended cells were transferred into different FACS tubes (**Figure 3.27**).


0 nM
0.0875 nM
0.175 nM
0.35 nM
0.875 nM
1.75 nM
A431
NK-CAR
Q1
Q2
Q3
Q4
23.6%
22.2%
22.2%
19.9%
20.1%
19.8%
17.6%
16.7%
14.1%
16.1%
17.1%
15.9%
13.0%
12.7%
12.9%
14.1%
13.7%
14.0%

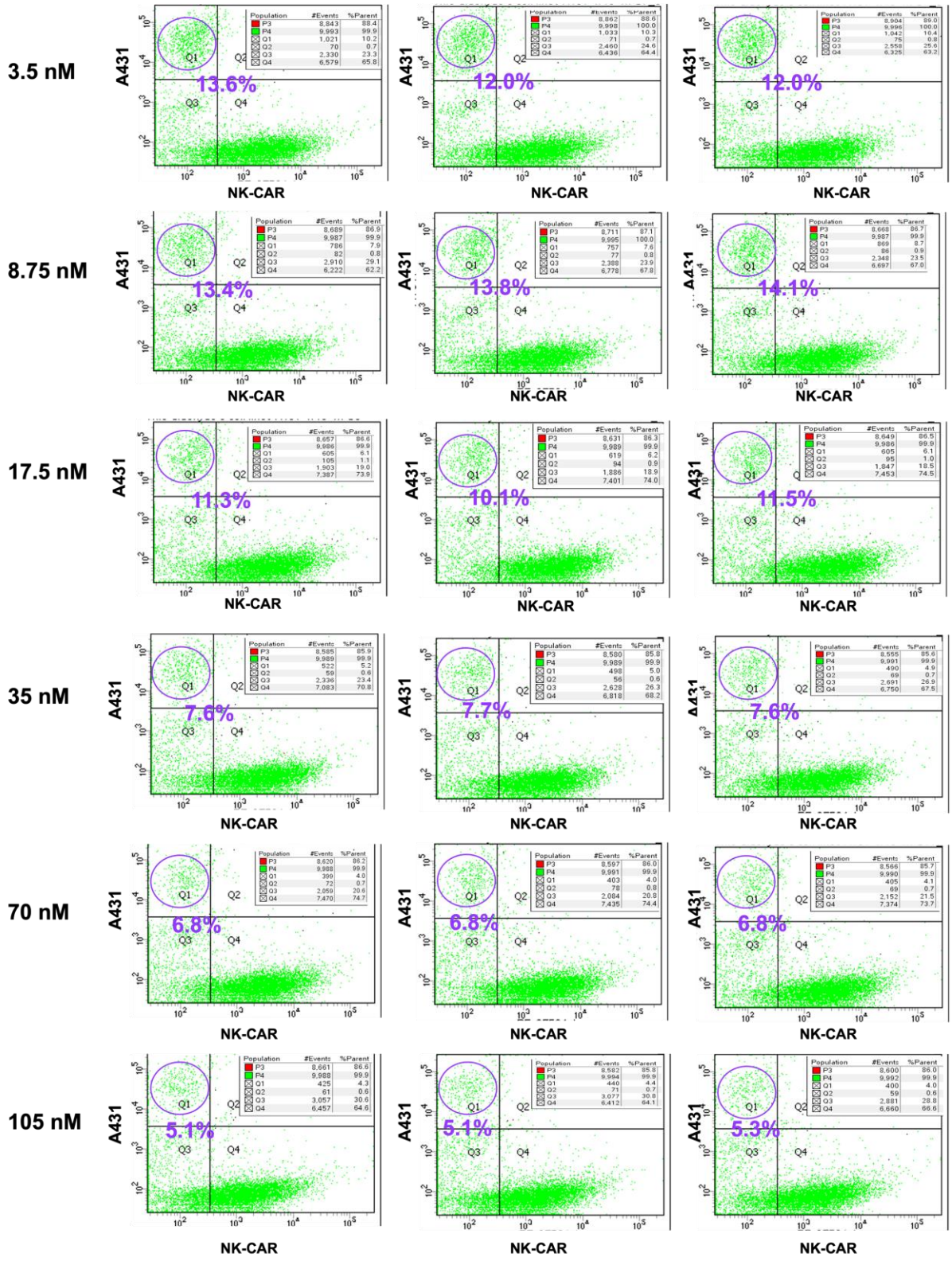


Figure 3.27. The FACS analysis of A431 cells numbers with NK-Car cells pretreated with protein **32** for 4h in different concentrations: 0 nM, 0.0875 nM, 0.175 nM, 0.35 nM, 0.875 nM, 1.75 nM, 3.5 nM, 8.75 nM, 17.5 nM, 35 nM, 70 nM, 105 nM. In this way, the IC_{50} (6.7 ± 0.4

nM) of the CAR-NK specific lysis was also analyzed using FACS after adding different concentrations of **32** to a mixture of CAR-NK and cancer cells.

3.3 Materials and methods

3.3.1 Materials

All amino acids, coupling reagents, solvents and resins were purchased from Sigma and Chemimpex. All commercially available solvents and reagents were used as received without further purification. VyPAL2, butelase-1 and OaAEP1b were prepared in-house as previously reported.[98, 127, 128]

3.3.2 HPLC

Analytical RP-HPLC was run on a SHIMADZU (Prominence LC-20AT) instrument using an analytical column (Grace Vydac "Protein C4", 250 × 4.6 mm, 5 μm particle size) at a flow rate of 1.0 mL/min. Analytical HPLC elution was monitored by UV absorption at 214 nm and 254 nm. Semi-preparative RP-HPLC was run on a SHIMADZU (Prominence LC-20AT) instrument using a semi preparative column (Grace Vydac "Protein C4", 250 × 10 mm, 10 μm particle size) at a flow rate of 2.5 mL/min. Both analytical and semi-preparative HPLC were run at room temperature using a gradient of solvent B in solvent A. Solvent B was 90% acetonitrile in water (0.040% TFA) and solvent A was water (0.045% TFA). Both solvents were filtered through 0.22 μm filter paper and sonicated for 30 min before use.

3.3.3 Solid phase peptide synthesis

All the peptides used in this study were synthesized as C-terminal amides using Rink amide MBHA resin (for normal C-terminal amine peptides) or trityl chloride resin (for hydrazide peptides) by standard Fmoc chemistry. Before use, the resin was pre-swelled in DMF

for 20 minutes. Before the first coupling, an Fmoc deprotection procedure was performed using 20% piperidine in dimethylformamide (DMF) for 30 minutes. The resin was then successively washed with DMF, DCM, and DMF. For trityl resin, 2% hydrazine in DMF should be used. For the coupling reactions, 3 equivalents of Fmoc AA-OH and 3 equivalents of PyBOP were first dissolved in DMF/DCM (1:1). The mixture was added to the resin, followed by the addition of 6 equivalents of DIEA. Coupling reactions were carried out for 60 to 90 minutes. Coupling efficiency was examined by the ninhydrin test. For peptide **23**, after peptide assembly on the solid phase, the Boc on the C-terminal lysine side chain amine was removed with 1 M HCl in DCM for 30-40 minutes, then FITC was coupled to the side-chain free amine. The peptides were cleaved from the resin with a cocktail containing 95% TFA, 2.5% water, and 2.5% TIS for 2 hours. After precipitation with cold diethyl ether, the crude peptides were purified using HPLC. The desired peptides were obtained in powder form after lyophilization. All peptides were characterized by electrospray ionization mass spectrometry.

3.14. Protein expression and purification

Genes encoding the desired protein sequences were cloned into the pETDuet vector, and the plasmids were then transformed into E. coli BL21 (DE3) competent cells using the standard 90-second heat shock protocol. The bacterial colonies were picked and transferred to liquid LB medium in a culture flask. The flask was shaken in the incubator at 37°C for 8-12 hours until the OD reached 0.6-0.8, followed by induction with 1 mM IPTG at 37°C for 4-8 hours to express the protein. Cells were harvested and lysed by sonication in lysis buffer containing 50 mM sodium phosphate and 500 mM NaCl (pH 8.0). After centrifugation, the supernatant was loaded onto a column of Ni-NTA beads and incubated at 4°C for 1 hour. The beads were washed 3 times with the lysis buffer, and the protein was subsequently eluted with lysis buffer

containing 250 mM imidazole. The purified protein was dialyzed in phosphate buffer (pH 6.5) overnight and stored in the freezer at -20°C.

3.1.5. Mass spectrometry

ESI mass spectrum data of small peptides and proteins were obtained from a Thermo Finnigan LCQ DECA XP MAX (ESI ion source, positive mode). The software of MagTran 1.03 and ESIProt 1.0 were used for the data deconvolution.

3.1.6. Tissue culture

The 293T, A431, and MCF-7 cells were maintained in DMEM (high glucose) supplemented with 10% FBS at 37°C in an incubator with 5% CO2. For passaging, cells were washed three times with trypsin-EDTA (0.25%) to detach them from tissue culture plates. Then, a three-times volume of complete DMEM medium was added to neutralize trypsin activity. Jurkat T cells were cultured in RPMI-10 supplemented with 10% FBS. The suspension cells were passaged every three days at a 1:5 split ratio (old: fresh medium). For the culture of NK-92 MI cells, Alpha Minimum Essential Medium (α-MEM) supplemented with 12.5% fetal bovine serum (FBS), 12.5% horse serum (HS), and 200 U/mL recombinant human IL-2 was used. The media should be changed every 2-3 days, and the cells should be subcultured at a ratio of 1:3 to 1:5 every 5-7 days.

3.1.7. Enzymatic kinetic analysis

The catalytic kinetics of VyPAL2 towards peptide substrates **5a-5i** were carried out at pH 6.5 and 25°C for 5-60 minutes. The peptides were added in concentrations ranging from 2.5-400 μM. After mixing peptides with enzymes, the pH was adjusted to 6.5 using 50 mM phosphate buffer. The reaction proceeded for 5-60 minutes, depending on the optimum

MALDI-TOF signal generated from the consumption of peptide starting materials. The reaction yields were quantified using DATA EXPLORER software based on the area of the peptides. The yield was analyzed using the ratio between the starting material **1a-5i** and the product **6**. Average yield and standard deviations (SDs) were calculated from experiments performed in triplicate. The enzymatic kinetics parameters, k_{cat} and K_m, were determined using a Lineweaver-Burk plot.

3.1.8. Cell viability assay

MTT assays were carried out following recommended protocols from Sigma-Aldrich (Cat. No. 11465001001). Cells were first seeded in a 96-well tissue culture plate with 100 µL medium to grow until the confluency reached 40-60% of the plates surface. Compounds were added and incubated for 72 h, followed by adding 10 µL of MTT I to each well and further incubated for approximately 4 h. After this, MTT II was added and incubated at 37 °C for overnight to solubilize the purple crystals. Spectrophotometric absorbance measurement of the samples was carried out by using a microplate reader (Biotek, cytation 5) at the wavelength of 575 nm and the reference wavelength was 670 nm.

3.1.9. LDH assay

Target cells (1×10^4) were mixed with anti-FITC CAR-NK cells (1×10^5) at E:T ratios of 2:1, 5:1, or 10:1 in RPMI media supplemented with 10% FBS and incubated with ZEGFR-FC-FITC **22** or **24** for 24 hours at 37°C. Cytolytic activity was determined by measuring the amount of LDH (lactate dehydrogenase) released into the culture media using the CytoTox 96 Non-radioactive cytotoxicity assay kit (Promega). The absorbance at 490 nm was measured using a Biotek Cytation 5 plate reader. The calculations were done as follows: % Cytotoxicity

= (experimental – effector spontaneous – target spontaneous) / (target maximum – target spontaneous) x 100.

3.1.10. IL-2 releasing assay

1 x 10^5 CAR-Jurkat cells were co-cultured with 1 x 10^4 A431 target cells in 100 μL cultured media per well in 96-well round bottom plates for 24 hr. Cultured supernatants were harvested and assayed for the presence of IL-2 using an ELISA kit, according to manufacturer's instructions (Gibco). Values represent the mean of triplicate wells and error bars represent SEMs derived from triplicate samples.

3.1.11. Virus packaging and transduction

293T cells were transfected with Lenti-CMV-Anti-FITC SCFV, Lenti-CMV-mCherry, or Lenti-CMV-GFP plasmid along with two packaging vectors PMD2.G and Pspax2 using lipo3000 (Thermo Fisher Scientific). After 12 hours of transfection, the old medium was removed from the cell culture, and fresh DMEM complete medium was added. The virus was harvested by collecting the supernatant 48 hours post-transfection. The collected supernatant was added to the cells in a 1:1 ratio with fresh medium, and the cells were incubated for an additional 24 hours (MCF-7, A431, Jurkat) or 12 hours (NK92-MI). After incubation, cells were cultured in fresh medium containing puromycin at a concentration of 2 μg/μL (MCF-7, A431, Jurkat) or 0.5 μg/μL (NK92-MI) to enrich the infected positive cells for 72 hours.

3.1.12. Protein A fast protein liquid chromatography

Protein FPLC purification was performed using a MAbPac column. Two mobile phase solutions were used: elution buffer A containing 50 mM PBS and elution buffer B containing 100 mM citric acid (pH 3.0). After injecting the samples, the column was washed with elution buffer A at a flow rate of 4 mL/min for 20 minutes. Then, the affibody-FC was eluted with elution buffer B at a flow rate of 2 mL/min for 5 minutes. The elution was collected in 0.5 mL fractions and subjected to Nanodrop and PAGE gel analysis to confirm the samples.

3.3.13 Protein amino acid sequence and peptide used in this study

The DNA sequence was optimized and synthesized by Genescript and cloned into the expression plasmid pETduet with ampicillin resistance for bacterial expression or pEF1a-puro with puromycin resistance for mammalian cell expression. The following protein sequences were used in this study.

Z_{EGFR} 1

MKKGSSHHHHHHLQVDNKFNKEMWAAWEEIRNLPNLNGWQMTAFIASLVDDPSQSANLLAEA
KKLNDAQAPKVDGSGSNGL

Z_{EGFR} 5

MKKLQVDNKFNKEMWAAWEEIRNLPNLNGWQMTAFIASLVDDPSQSANLLAEAKKLNDAQAP
KVDGSGSNGLHHHHHH

sfGFP 1

MMSVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPT
LVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVN
RIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQ

QNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSGSNGL

HHHHHH

MPL (TPOR)

MPSWALFMVTSCLLLAPQNLAQVSSAGSEQKLISEEDLQDVSLLASDSEPLKCFSRTFEDLT

CFWDEEEAAPSGTYQLLYAYPREKPRACPLSSQSMPHFGTRYVCQFPDQEEVRLFFPLHLWV

KNVFLNQTRTQRVLFVDSVGLPAPPSIIKAMGGSQPGELQISWEEPAPEISDFLRYELRYGP

RDPKNSTGPTVIQLIATETCCPALQRPHSASALDQSPCAQPTMPWQDGPKQTSPSREASALT

AEGGSCLISGLQPGNSYWLQLRSEPDGISLGGSWGSWSLPVTVDLPGDAVALGLQCFTLDLK

NVTCQWQQQDHASSQGFFYHSRARCCPRDRYPIWENCEEEEKTNPGLQTPQFSRCHFKSRND

SIIHILVEVTTAPGTVHSYLGSPFWIHQAVRLPTPNLHWREISSGHLELEWQHPSSWAAQET

CYQLRYTGEGHQDWKVLEPPLGARGGTLELRPRSRYRLQLRARLNGPTYQGPWSSWSDPTRV

ETATETAWISLVTALHLVLGLSAVLGLLLLRWQFPAHYRRLRHALWPSLPDLHRVLGQYLRD

TAALSPPKATVSDTCEEVEPSLLEILPKSSERTPLPLCSSQAQMDYRRLQPSCLGTMPLSVC

PPMAESGSCCTTHIANHSYLPLSYWQQP

MKK-Z_{EGFR} -FC 22

MKKGSSHHHHHHLQVDNKFNKEMWAAWEEIRNLPNLNGWQMTAFIASLVDDPSQSANLLAEA

KKLNDAQAPKVDGSGSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDV

SHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKAL

PAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENN

YKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGKGSGSN

GL

CI-Z_{EGFR} -FC 29

CIGSSHHHHHHLQVDNKFNKEMWAAWEEIRNLPNLNGWQMTAFIASLVDDPSQSANLLAEAK

KLNDAQAPKVDGSGSDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVS

HEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALP

APIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNY

KTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGKGSGSNG

L

3.3.14 Summary of peptides used in this study

Peptide No	Peptide sequence	Cal. MS $[M+H]^+$	Obs. MS
2	Ac-KKLAVINGF	1029.63	1029.29
3a	Ac-KKLAVIN-1a	841.55	841.62
3b	Ac-KKLAVIN-1b	975.59	975.67
3c	Ac-KKLAVIN-1c	1180.72	1180.74
3d	Ac-KKLAVIN-1d	883.57	883.63
3e	Ac-KKLAVIN-1e	951.54	951.50
3f	Ac-KKLAVIN-1f	917.59	917.64
3g	Ac-KKLAVIN-1g	1408.89	1408.90
3h	Ac-KKLAVIN-1h	1004.28	1004.71
3i	Ac-KKLAVIN-1i	981.55	981.65
3j	Ac-KKLAVIN-1j	995.56	995.61
3k	Ac-KKLAVIN-1k	941.61	941.64
3l	Ac-KKLAVIN-1l	1074.61	1074.63
3m	Ac-KKLAVIN-1m	899.57	899.71
3n	Ac-KKLAVIN-1n	914.55	914.79
3o	Ac-KKLAVIN-1o	954.60	954.28
3p	Ac-KKLAVIN-1p	1344.53	1344.67
4	Thz-ISTKSIPPISRNGL	1759.95	1759.95
5a	CISTKSIPPISYRNGI	1746.95	1746.93
5b	CISTKSIPPISYRN-1h	1768.28	1768.88
5c	CISTKSIPPISYRN-1g	2159.19	1080.76
5d	CISTKSIPPISYRN-1d	1634.86	1634.81
5e	CISTKSIPPISYRN-1e	1702.83	1702.92
5f	CISTKSIPPISYRN-1j	1746.86	1746.98
5g	CISTKSIPPISYRN-1f	1668.88	1669.02
5h	CISTKSIPPISYRN-1l	1825.90	1825.79
5i	CISTKSIPPISYRN-1o	1706.89	1706.90
5j	CISTKSIPPISYRN-1p	2092.95	1047.74
19	Ac-IEGPTLRQWLAARAGSGSIG-$NHNH_2$	2096.13	2096.39
20	(FITC)KIEGPTLRQWLAARANGL	2382.17	2382.80
23a	NH_2NH-FITC	421.07	421.99
23b	$NH_2NHCOCH_2CONHNH$-FITC	521.10	521.12
23c	$NH_2NHCO(CH_2)_2CONHNH$-FITC	536.12	536.18
23d	$NH_2NHCO(CH_2)_4CONHNH$-FITC	564.15	564.25
24a	NH_2NH-Dox	560.22	560.54
24b	$NH_2NHCOCH_2CONHNH$-Dox	660.24	660.56
24c	$NH_2NHCO(CH_2)_2CONHNH$-Dox	674.26	674.06
24d	$NH_2NHCO(CH_2)_4CONHNH$-Dox	702.29	702.38
25a	NH_2NH-MMAE	790.54	790.57
25b	$NH_2NHCOCH_2CONHNH$-MMAE	890.54	890.65
25c	$NH_2NHCO(CH_2)_2CONHNH$-MMAE	904.54	904.41
25d	$NH_2NHCO(CH_2)_4CONHNH$-MMAE	932.61	932.52
31	K(FITC)IARNGL	1105.21	1105.30

3.4 Conclusion

In this study, we have found that PALs can use hydrazides, the non-basic α-effect nucleophiles, as acyl acceptor substrates and attach them to the C-terminal end of an acyl donor peptide or protein. This expanded substrate scope provides further proof that PALs are versatile tools for protein engineering. Although the unnatural linkage formed in the hydrazide ligation reaction is chemically the same, its stability toward the PAL enzyme varies with the hydrazide structures, as the ligation products may have different binding properties toward the enzyme. We have further demonstrated the applications of PMHL in C-to-C ligations and the preparation of complex protein conjugates for potential therapeutic applications, including bi-specific engager molecules as mediators of cell-based immunotherapy. Being highly efficient, the PMHL scheme is a new precision protein engineering tool for the biomanufacturing and development of multi-functional protein biologics, such as antibody-drug conjugates and other immunotherapeutic agents.

Chapter 4: Future work and Perspectives - A removable cysteinyl-asparaginyl tag for traceless protein ligation

4.1 Introduction

The work reported in the previous two chapters demonstrates that PALs are powerful tools for protein bioconjugation. I have shown that, by making use of the differential substrate specificities of butelase-1 and VyPAL2, two different labeling groups can be introduced onto the respective C- and N-termini of a recombinant protein. The dually labeled proteins are useful theranostic agents. I have also shown that the non-basic hydrazide compounds are effective nucleophile substrates of PALs, expanding the substrate scope of this class of peptide ligases for protein engineering. Because hydrazide can be easily derivatized, a variety of functional moieties can be conjugated to a protein substrate through PAL-catalyzed hydrazide ligation. In addition to the preparation of protein conjugates, PALs also hold great promise for the semi-synthesis of proteins, especially those containing post-translational modifications (PTMs). The semi-synthesis of such a protein involves ligating a recombinant truncated bulk of the protein with the remaining peptide fragment, which contains the desired PTM and is prepared by solid-phase peptide synthesis. Ideally, the semi-synthetic protein should be prepared in its native sequence. Although PALs require a minimum tripeptide recognition motif for the ligation reaction, this still poses a significant limitation to the possible junctions that can be used for ligation. In this chapter, I propose the design of a traceless protein ligation scheme as a potential strategy to overcome this limitation. My preliminary results on model peptides show that the scheme can produce some desired product, but the yield is still unsatisfactory because of the occurrence of significant side reactions. If the scheme can be made more feasible in the future, it would become a good method that can help to increase the availability of PTM-containing proteins for structural and functional characterizations.

Histones are perfect examples of proteins undergoing extensive post-translational modifications. Post-translational modifications of histones play important roles in regulating fundamental cellular processes such as gene transcription, DNA replication, and damage repair. Understanding the effects of the modifications at the molecular level is of great importance to the field of epigenetics. One approach to studying this involves the preparation of site-specifically modified histones for incorporation into synthetic chromatin structures. The earlier methods for histone modification include the genetic incorporation of unnatural amino acids into histones. By incorporating the amber codon UAG in genes with orthogonal pairs of amber suppressor tRNAs (transfer RNAs) and their cognate aminoacyl tRNA synthetases (aaRS), the strategy has been demonstrated to be useful in incorporating modified amino acids in histone H3 and, more recently, in histone H4. However, various disadvantages, such as the inability to control the degree and diversity of histone modifications, the process of unnatural amino acid synthesis, and undesired incorporation yield, have restricted its broad applications.

Strategies based on native chemical ligation (NCL) have been developed, enabling the generation of homogeneously site-specifically modified histones.[137,138] This involves the reaction of a C-terminus thioester peptide with an N-terminus cysteinyl moiety to produce a native peptide bond between the two segments.[40, 139-141] In recent years, the development of hydrazide-based NCL has made the method suitable for a wide range of applications (**Figure 4.1 A**).[142-145] Therefore, the introduction of hydrazide at the C-terminus of proteins is frequently needed and highly desired for research and industrial applications. Synthetic C-terminus hydrazide peptides can be readily prepared using solid-phase peptide synthesis; however, it is technically challenging to use this method to prepare peptide C-terminus hydrazides of more than 50 amino acids in length. At present, a broadly useful protein C-terminal hydrazinolysis method is lacking. Several approaches to introduce a hydrazide functionality at the protein C-terminus have been developed in the past, albeit each with certain significant shortcomings.

One of these makes use of the intein-mediated protein splicing mechanism. This method is based on the inherent activity of intein in mediating N,S(O)-acyl transfer when fused to the C-terminal side of a target protein (**Figure 4.1 B**).[146,147] However, the intein approach requires expression of the fusion protein, which is often time-consuming and low-yielding.[148,149] In addition, a large amount of the fusion protein needs to be prepared because the system has no catalytic turnover.

Due to the limitations of the methods described above, enzymatic approaches, which typically can be operated under mild conditions and require only a tiny amount of the enzyme for the catalyzed reaction, are widely accepted as desirable methods to facilitate protein hydrazinolysis. Enzymes proven to be able to mediate hydrazinolysis include Sortase A, CPaseY[118], and DUBs. However, the use of Sortase A would leave a long LPXT peptide "trace" in the product. Since proteins don't usually have such a motif in their sequences, this method cannot be used to prepare PTM-containing proteins such as histones in their native sequences. Also, hydrazinolysis using some of these enzymes requires a large enzyme-to-substrate molar ratio (1:1 for Sortase A), making it difficult to separate the hydrazide products from the enzyme for subsequent applications. CPaseY was reported to be useful for mediating protein hydrazinolysis (**Figure 4.1 C**).[118] The enzyme is a carboxypeptidase, favoring the hydrophobic amino acids while disliking the hydrophilic and prolyl residue.[150] Whereas amino acids with hydrophobic properties could be removed quickly, still, other amino acids not typically favored by CPaseY could still be removed during the hydrazinolysis reaction, forming the over-reaction products with truncated sequences.[118] Also, prolonged time is needed to cleave the C-terminal residue. Those disadvantages might create certain problems in the process of preparing the hydrazide product, especially when a purification tag is needed at the C terminus. The extra processing of residues might lower the yield and require more time to acquire the product. Consequently, a method able to mediate protein hydrazinolysis in a site-specific manner, being

potentially traceless and of high catalytic efficiency, is highly desirable. In recent years, peptidyl asparaginyl ligases (PALs) have been widely studied and demonstrated useful in mediating protein ligation reactions.[11, 13-15, 92-96, 151, 152] PAL-mediated hydrazinolysis would overcome the above disadvantages (**Figure 4D**). Since PALs recognize an Asx-P1'-P2' tripeptide motif and process the substrate at the Asx-peptide bond, the reaction only leaves the Asx in the resulting product.

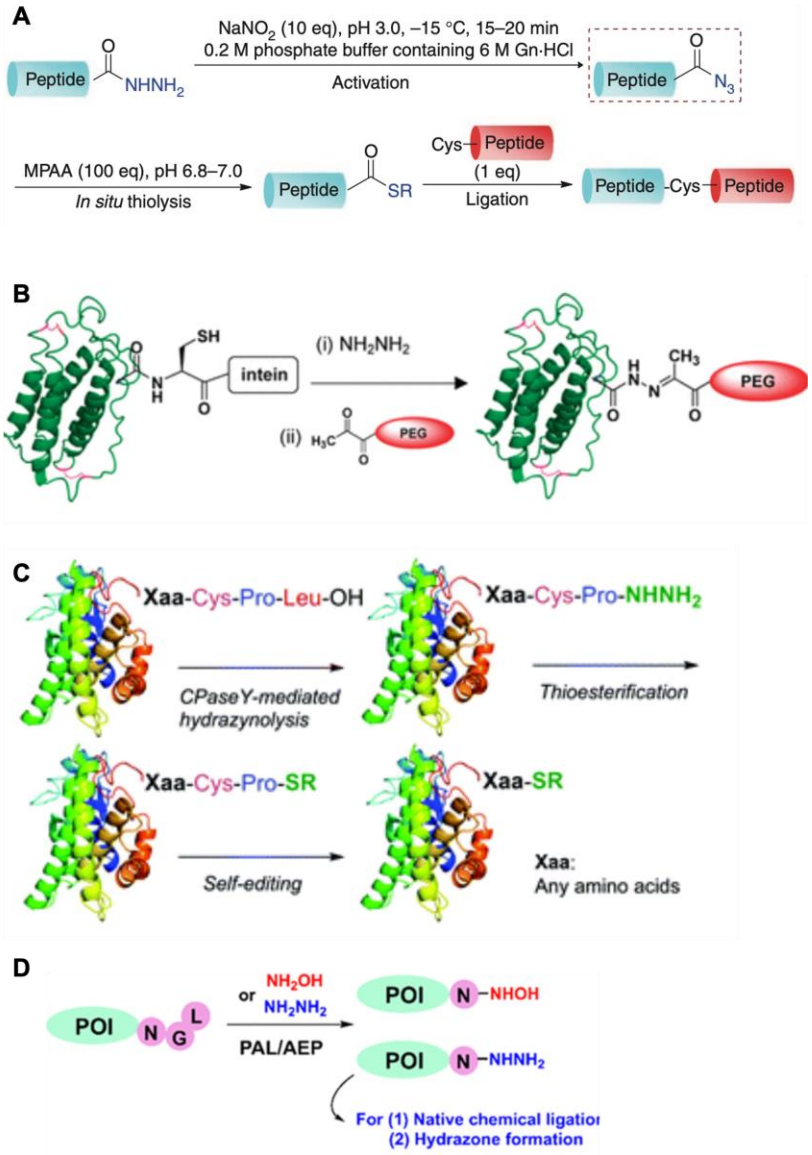


Figure 4.1 A) Mechanisms for hydrazide-based native chemical ligation; B) Intein-based protein hydrazinolysis; C) CPaseY-catalysed hydrazinolysis reaction of proteins; D) PAL/AEP-mediated hydrazinolysis reaction of a protein of interest (POI).

Various enzymes with ligation properties are able to catalyze ligation to form a natural peptide bond between a protein and a functional peptide. With their high specificity and traceless features, PAL enzymes show great potential in protein engineering. However, the advantage of specific recognition also causes a restriction of this method, i.e., the ligation junction residue Asx is left in the product. Although Asn or Asp are relatively common amino acids, they may not be present at the desired site in a target protein for the ligation reaction. Therefore, it is desirable if one can remove the Asn residue from the ligation product.

Kawakami and Aimoto developed the cysteinyl-prolyl ester (CPE) system for the preparation of peptide thioesters. In this system, peptides containing the CPE system (–XaaCysPro-OR) are converted to thioesters (–Xaa-SR) driven by diketopiperazine formation by the transient N-to-S acyl transfer intermediate.[118,153,154] We envisioned that a similar mechanism involving a cysteinyl-asparaginyl thioester might also be able to convert a protein C-terminal asparaginyl thioester (–Xaa-Cys-Asn-SR) to a more general thioester (–Xaa-SR) following removal of the Cys-Asn dipeptide as diketopiperazine. An asparaginyl thioester can be readily generated through PAL-mediated hydrazinolysis, activation of the hydrazide (e.g., by nitrite oxidation) and conversion to thioester.[155] The self-editing system driven by diketopiperazine formation would deviate the need for Asn at the ligation junction.

Previously, Otaka et al. reported a method for the preparation of protein thioesters using CPaseY-mediated hydrazinolysis and subsequent self-editing of a C-terminal Cys-Pro dipeptide through diketopiperazine formation.[118,153,154] However, because CPaseY can progressively process the C-terminal residues, the hydrazinolysis reaction must be conducted in the presence of a ketone compound (cyclohexanone) to prevent over-reaction as cyclohexanone would temporarily protect the so-formed hydrazide in the form of hydrazone. Our method would be more generally useful since, as Asn-specific enzymes, PALs recognize

only an Asn-Xaa1-Xaa2 tripeptide motif where Xaa1 is relatively broad and Xaa2 must be a large hydrophobic amino acid. It would be unlikely to have the problem of over-reaction unless the protein substrate has such motifs in its natural sequences and such motifs are accessible by the PAL, which is not the case because the tightly folded protein structure would protect them from the attack by the enzyme.

The protein thioester so-generated can be further used for native chemical ligation. Coupled with desulfurization of the ligation product, such a strategy can be used to perform NCL at alanine and other amino acids (**Figure 4.2**). Besides, we hypothesized that subtiligase can use the thioester as the acyl-donor substrate for subtiligase-mediated ligation with another nucleophile peptide substrate (**Figure 4.2**). Moreover, this protocol could be applied to the conversion of recombinant proteins to their C-terminal thioesters, allowing traceless synthesis of post-translationally modified histones.

In short, we designed a method that utilizes an auto-removable mechanism of C-terminal cysteinyl-asparaginyl dipeptide for the synthesis of protein thioesters which are useful for NCL and subtiligase-mediated ligation. Our method might greatly expand the scope of PAL-mediated ligation since the ligation site is no longer limited to Asx but expanded to many other amino acid residues.

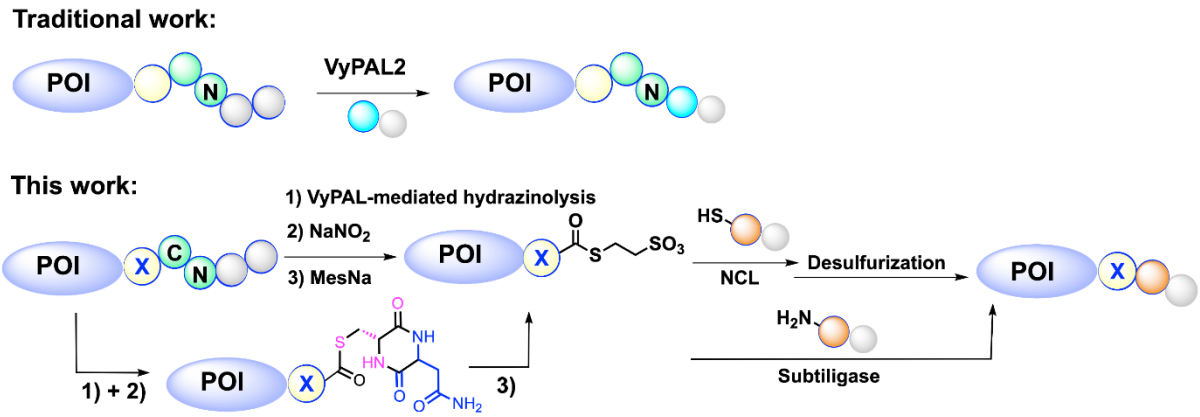
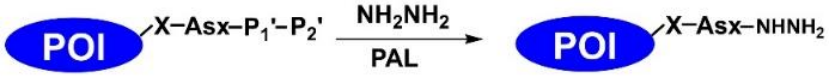


Figure 4.2. Scheme of traceless ligation using an auto-removable mechanism of C-terminal Cys-Asn dipeptide.

4.2 Results and discussion

4.2.1 Evaluation of various PAL-mediated hydrazinolysis conditions

Several factors could influence the efficiency of the PAL-catalysed hydrazinolysis reaction, including the choice of PAL enzyme and residues at P1, P1', and P2' positions like what has been found in the conventional ligation reactions.[15, 96] Earlier studies have shown the importance of selecting a small residue at P1' and large hydrophobic residue at P2' position. Based on these findings and enzymatic preferences of butelase 1 and VyPAL2, we designed 6 peptides in a combination of highly favourable residues at the position. We synthesized peptides **1a-e** containing different substrates, **1a**-NGL, **1b**-NGF, **1c**-NHV, **1d**-DGL, **1e**-NGL. All the peptides had a Cys residue at P2 position. The reaction was done by using 50 µM of peptides **1a-e** and 10 mM hydrazine with 100 nM PALs at pH 4.5-7.4 and 37 °C for 0.25-1 h. The reaction was analysed using HPLC and ESI-MS to calculate the hydrazinolysis yields (**Figure 4.3**). We found several favourable conditions for PAL-mediated hydrazinolysis with yield >95%. As seen from **Figure 4.3**, peptides **1a**-NGL, **1b**-NGF are favoured by VyPAL2 at pH 7.4 and 1**d**-NHV, **1e**-NGL are more suitable for butelase 1 at pH 6.5. The results were consistent with the previous reports and suggesting that the rate limiting step of hydrazinolysis was the formation of the acyl-PAL-acyl intermediate. Therefore, the affinity of each PAL to the substrates are very critical to the enzyme's catalytic efficiency. For the simplicity of our next experiments, we used VyPAL2 at pH 7.4 to perform the following hydrazinolysis reactions.



1a-1e 2a-2d

Entry	No.	Peptide sequence	PAL	pH	RT (Min)	Yield (%)
1	1a	Ac-ASLKAGCNGL	VyPAL2	7.4	15	80
2	1b	Ac-KKLAVINGF	VyPAL2	7.4	15	90
3	1c	Ac-KKLAVINHV	VyPAL2	6.5	60	85
4	1d	Ac-ASLKAGCDGL	VyPAL2	4.5	60	70
5	1e	Ac-ASYKAGCNGL	VyPAL2	6.5	15	89
6	1a	Ac-ASLKAGCNGL	Butelase 1	7.4	15	85
7	1b	Ac-KKLAVINGF	Butelase 1	7.4	15	30
8	1c	Ac-KKLAVINHV	Butelase 1	6.5	60	88
9	1d	Ac-ASLKAGCDGL	Butelase 1	4.5	15	5
10	1e	Ac-ASYKAGCNGL	Butelase 1	6.5	15	95

Figure 4.3. PAL-mediated hydrazinolysis of model peptides **1a-e** in the presence of hydrazine.

4.2.2 Conversion of hydrazide into asparaginyl thioester

In our design, PAL-mediated hydrazinolysis of peptide/protein-Xaa-cysteinyl-asparaginyl-P1-P2 (Xaa-Cys-Asn-P1-P2) was performed to give the hydrazide product -Xaa-Cys-Asn-$NHNH_2$. Following oxidation by $NaNO_2$, a C-terminal azide -Xaa-Cys-Asn-N_3 was generated. In principle, diketopiperazine formation-driven thioesterification with the thiol group of Cys would occur from the azide products because azide is a good leaving group (**Figure 4.4**). However, in the presence of a thiol, the asparaginyl azide would also be quickly converted to a thioester, which would then undergo diketopiperazine formation by the same mechanism to afford the same -Xaa-thioester with the cysteinyl thiol. The formation of this thioester is driven by the formation of diketopiperazine, which quenches the free α-amine transiently formed in the energetically unfavorable N-to-S acyl transfer reaction. Obviously, the presence of a thiol compound such as MesNa or MPAA in the reaction system would allow

further transthioesterification, leading to a new thioester (protein-Xaa-Mes/MPAA) (**Figure 4.4**).

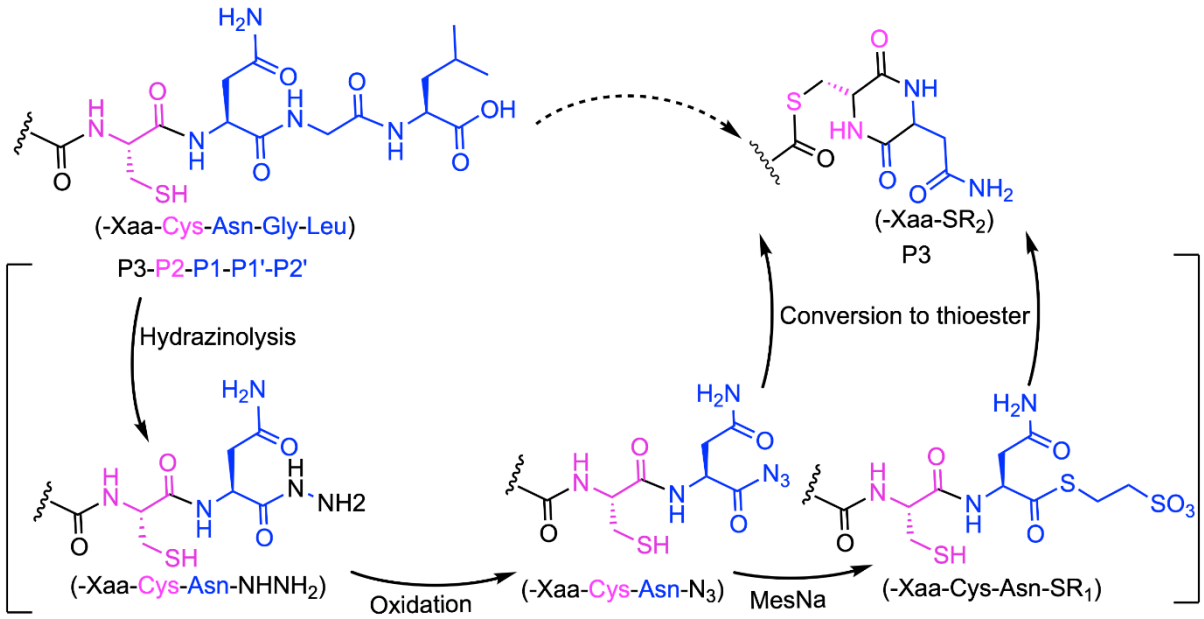


Figure 4.4. Envisioned transformation reactions for synthesis of peptide/protein C-terminal thioester using an auto-removable Cys-Asn dipeptide tag.

4.2.3 Examination of feasibility of the design using Xaa-Cys-Ala-NHNH$_2$ and Xaa-Cys-Asx peptide

To examine the feasibility of our design, we firstly synthesized a model peptide Ac-**ASLKAGCNGL** to explore the transformation during each step. As shown in **Figure 4.5**, we started from the hydrazinolysis experiment using VyPAL2. Starting material **1a** at 1 mM was mixed with 200 mM $NH_2NH_2{\cdot}H_2O$ and the pH was adjusted to 6.5, followed by addition of 200 nM VyPAL2. The reaction was completed after 30 min of incubation at 37 °C, affording **2a**. For the second step, 1.0 mM of **2a** dissolved in 6 M guanidine·HCl (pH 3.0) was reacted with 10 mM $NaNO_2$. The reaction was kept at -10 °C for 20 min. As monitored by HPLC and ESI, the reaction was highly specific and efficient, as almost 90% starting material was converted to product **3** which was further purified by HPLC. Then 50 μM **3** was reacted with 5 mM MesNa at pH 6.5 for 1 h and

12 h, respectively. The HPLC analysis showed that **3** had shifted obviously and ESI data indicated three major products, including the diketopiperazine intermediate product **4** or most likely the aspartimide by-product **6** with m/z 787.47, the MesNa product after removing of Cys-Asn, **7**, m/z 712.56 and the asparaginyl thioester, **5** m/z 929.36 (**Figure 4.5**). However, the formation of the undesired by-product 6 might be a major reason that decreases the transformation efficiency. There were also many other smaller peaks in the HPLC profile, and it was difficult to ascertain the identities of these side products. Nevertheless, the fact that the desired product was observed is encouraging, indicating that the self-editing transformation had taken place, albeit accompanied by the occurrence of other competing reactions. In the future, all side products and intermediates will need to be further analyzed and characterized by NMR and/or Mass Spectrometry, which will help to pinpoint the mechanisms involved in the formation of the side products and to find potential solutions to suppress the side reactions and promote the desired transformation.

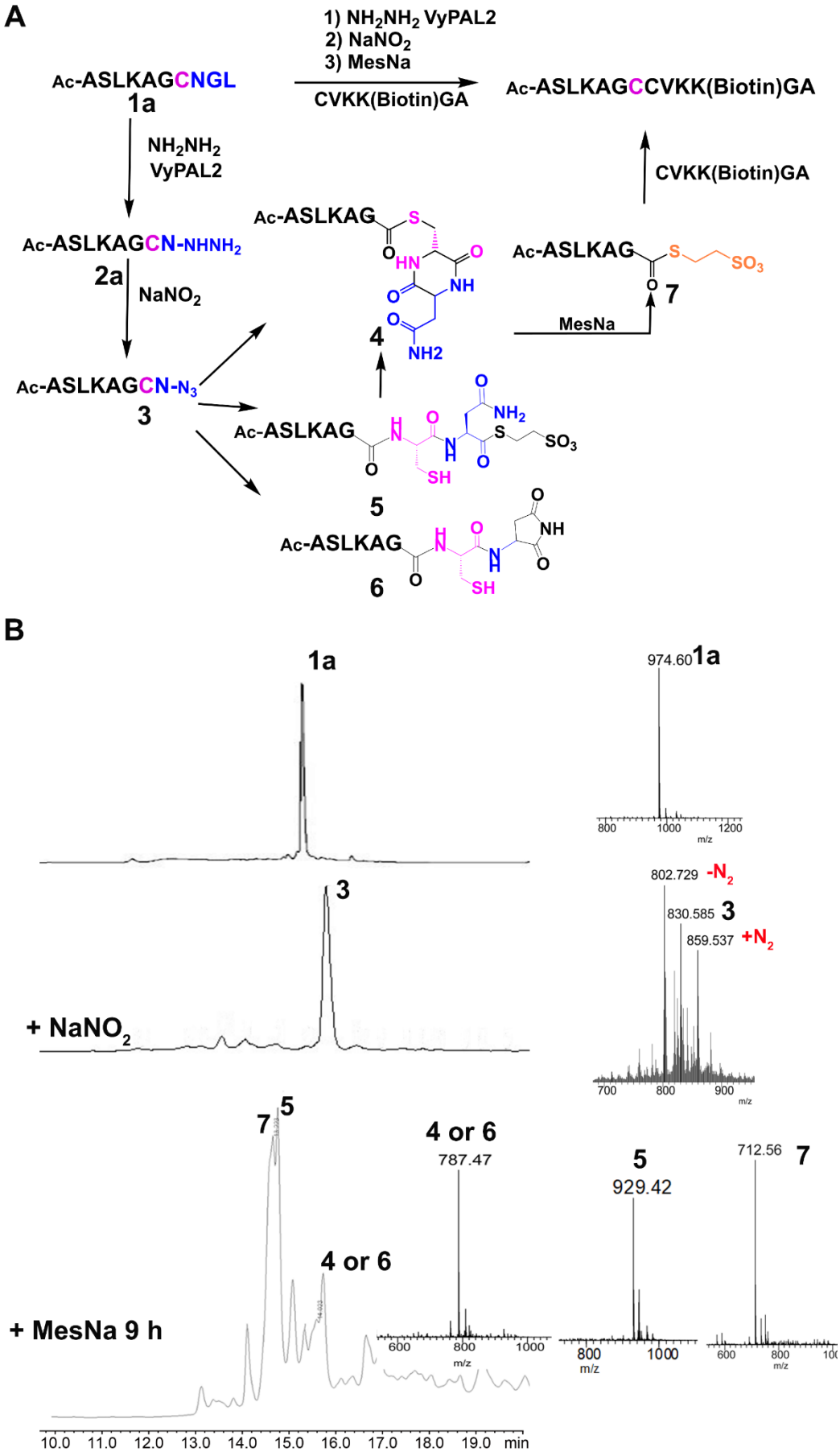

A
1) NH2NH2 VyPAL2
2) NaNO2
3) MesNa
Ac-ASLKAGCNGL
1a
CVKK(Biotin)GA
Ac-ASLKAGCCVKK(Biotin)GA
NH2NH2
VyPAL2
CVKK(Biotin)GA
Ac-ASLKAGCN-NHNH2
2a
NaNO2
Ac-ASLKAG
4
Ac-ASLKAG
7
MesNa
Ac-ASLKAGCN-N3
3
Ac-ASLKAG
5
Ac-ASLKAG
6
B
1a
974.60
1a
3
802.729 -N2
830.585 3
859.537 +N2
+ NaNO2
7
5
4 or 6
787.47
4 or 6
5
929.42
712.56
7
+ MesNa 9 h
m/z
min

Figure 4.5. Conversion of model peptide **Ac-ASLKAGCNGL 1a** to thioester **7** for traceless peptide ligation. A) Proposed mechanism of the conversion reactions. B) HPLC and ESI data the reaction products at each step.

As a parallel of our protocol, the CysAla-$NHNH_2$ peptide **2a'** was synthesized and the same experiments were performed. We monitored the intermediates of transformation from the azide peptide to the thioester peptide using HPLC. As seen in **Figure 4.6**, following azide formation from CA-$NHNH_2$ peptide **2a'**, peptide **3a'** was successfully transformed to the self-edited thioester **7a'** in which the CA dipeptide was removed, demonstrating that this is a feasible reaction. Interestingly, there were two separate HPLC peaks with the MW corresponding to the desired thioester product 7' (**Figure 4.6**), the reason for which is not clear at this stage. Obviously, the cleaner transformation in this example is due to that Ala does not have the –CO-NH_2 functional group on its side chain as does Asn.

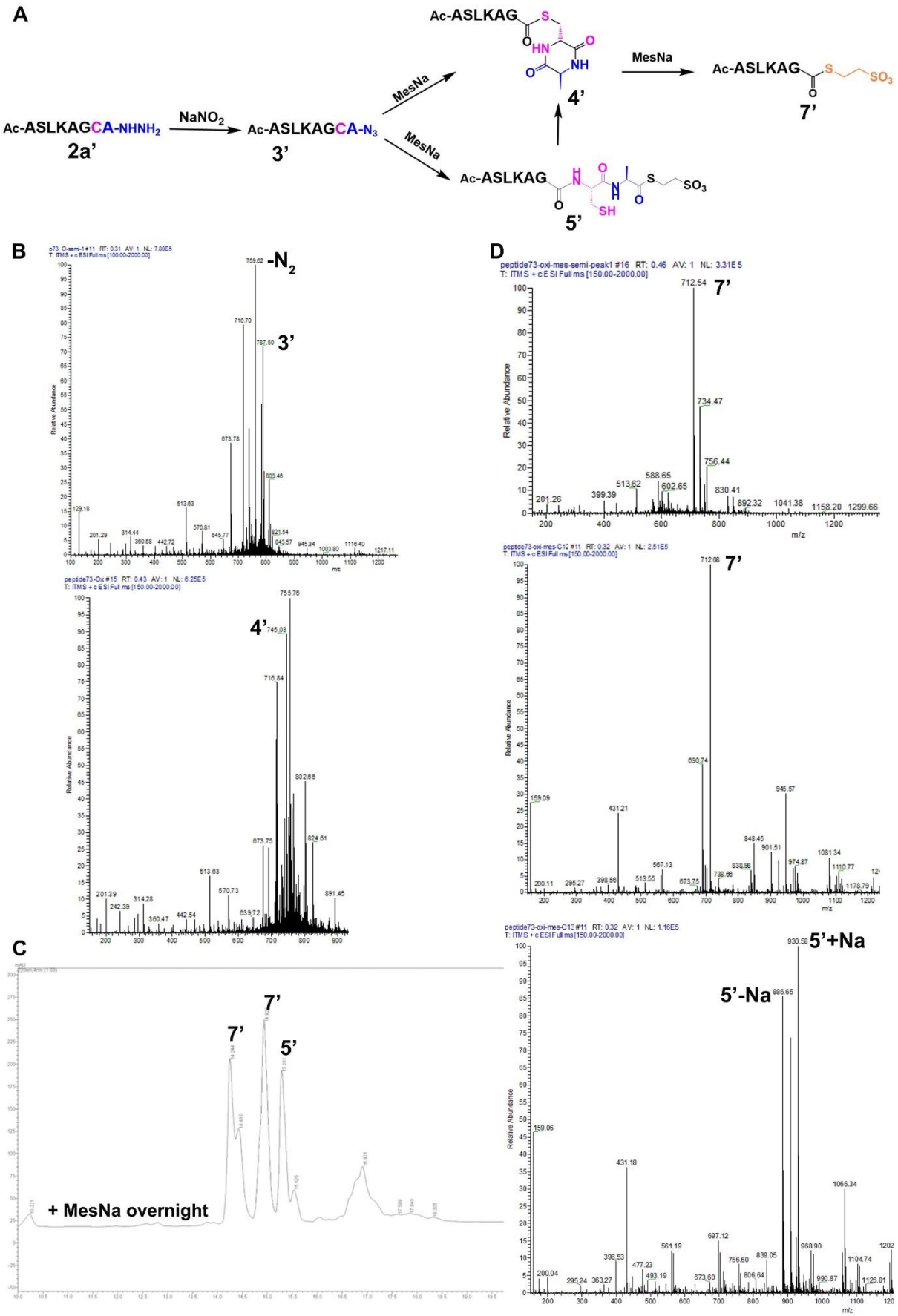

A
Ac-ASLKAGCA-NHNH2
2a'
NaNO2
Ac-ASLKAGCA-N3
3'
MesNa
MesNa
Ac-ASLKAG
4'
MesNa
Ac-ASLKAG
7'
Ac-ASLKAG
5'
B
-N2
3'
4'
C
7'
7'
5'
+ MesNa overnight
D
7'
7'
5'+Na
5'-Na

Figure 4.6. Conversion of model peptide Ac-**ASLKAGCA-NHNH$_2$** to thioester **7'** for traceless peptide ligation. A) Proposed mechanism of the conversion reactions. B) ESI data the reaction products at oxidization of **2a'** to **3'**. C and D) HPLC and ESI data for the reaction of MesNa induced transformation from **3'** to **5'** and **7'**.

As we can see from the two examples above, the self-editing scheme was workable for the two peptides with C-terminal Xaa-Cys-Ala-NHNH$_2$ **2a'** and Xaa-Cys-Asn-NHNH$_2$ **2a**, though significant by-products were formed in the case of **2a**. We also conducted the reaction on an Asp containing peptide as shown in **Figure 4.7**. However, hydrolysis was found to be the predominant reaction during the conversion from azide to thioester. Therefore, the self-editing reaction involving Cys-Asp is not efficient enough under current conditions (**Figure 4.7B**).

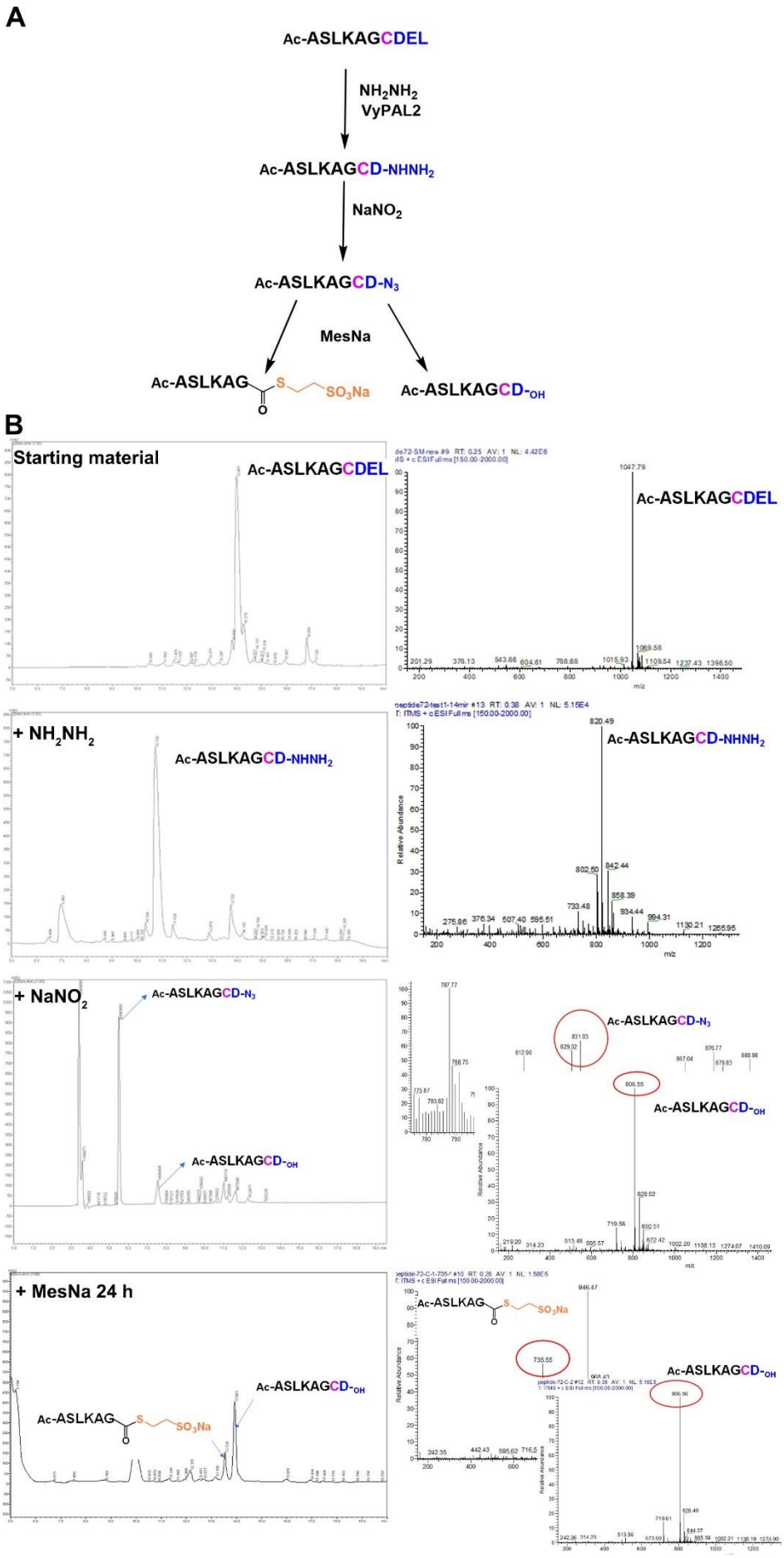

A
Ac-ASLKAGCDEL
NH2NH2
VyPAL2
Ac-ASLKAGCD-NHNH2
NaNO2
Ac-ASLKAGCD-N3
MesNa
Ac-ASLKAG
S
SO3Na
O
Ac-ASLKAGCD-OH
B
Starting material
Ac-ASLKAGCDEL
Ac-ASLKAGCDEL
+ NH2NH2
Ac-ASLKAGCD-NHNH2
Ac-ASLKAGCD-NHNH2
+ NaNO2
Ac-ASLKAGCD-N3
Ac-ASLKAGCD-OH
Ac-ASLKAGCD-N3
Ac-ASLKAGCD-OH
+ MesNa 24 h
Ac-ASLKAG
S
SO3Na
Ac-ASLKAGCD-OH
Ac-ASLKAG
S
SO3Na
Ac-ASLKAGCD-OH

Figure 4.7. Conversion of model peptide Ac-**ASLKAGCDEL** to thioeste **4** for traceless peptide ligation. A) Proposed mechanism of the conversion reactions. B) HPLC and ESI data the reaction products at each step

4.2.4 Effect of P2 position residue to thioesterification

The above result has demonstrated the feasibility of using a Cys-Asn tag to prepare a C-terminal thioester through PAL-mediated hydrazinolysis and subsequent transformations. However, the detailed transformation intermediates need to be confirmed and the current conditions are not optimal as there are significant side reactions. Nevertheless, we next aimed to determine whether the residue Xaa at P3 before P2-Cys would affect the thioesterification to form the thioester Xaa-SR. Therefore, we designed peptides **10a**-**10q** with different amino acids at the P3 position. The reaction was monitored to get the oxidization product **12a** – **12j** as shown in **Appendix C.** The thioesterification needs to be further confirmed in the future due to the by-products in this reaction.

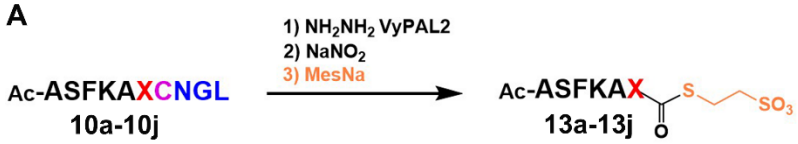


Figure 4.8. Screening of various amino acids at X position (i.e., the P3 position for a PAL substrate) before Cys for the influence on auto-removing efficiency.

4.2.5 Possible application of the auto-removing strategy for affibody modification

We also aimed to assess the applicability of this method to protein modification. An EGFR-targeting protein-Z_{EGFR} with the -CNGL tag at C terminus was firstly prepared. Afterwards, 100 μM of Z_{EGFR} -CNGL **14** was reacted with 1 mM NH_2NH_2 using 200 nM VyPAL2 at pH 6.5 and 25 °C to generate **15** in *ca.* 90% yield. Next, to oxidize the hydrazide into azide, 10 mM $NaNO_2$ was added to 100 μM **15** at pH 3.0 and -10 °C for 20 min. The reaction mixture was subjected to

purification using HPLC and the azide product **16** was obtained in > 90% yield (**Figures 4.9 B**). Up to this step, everything was working well. The next step is the self-editing reaction to acquire the thioester product **17,** which will be performed in future using conditions that are still to be optimized.

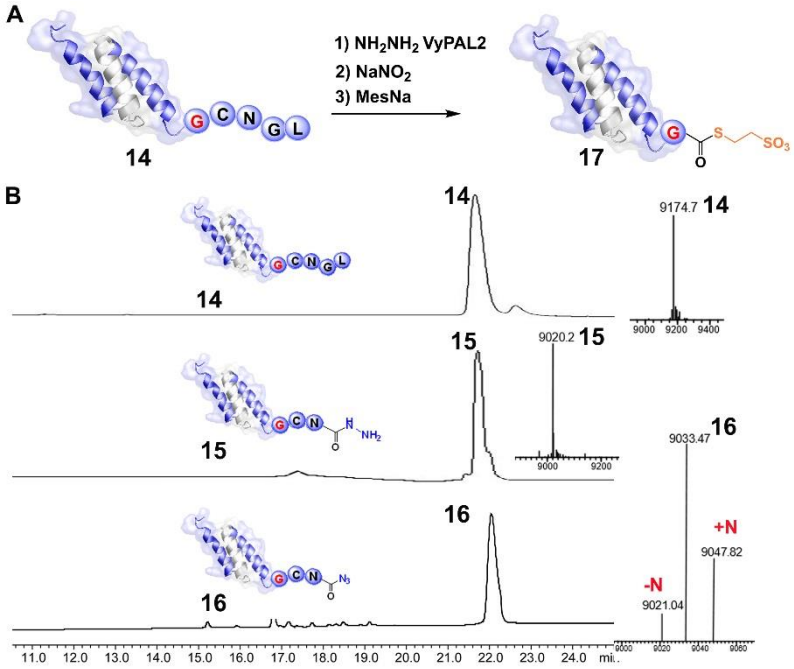


Figure 4.9 Application of the Cys-Asn auto-processing methodology to the synthesis of Z_{EGFR} protein thioester. A) Scheme of Z_{EGFR} protein thioester synthesis. B) HPLC and ESI data of the product at each step. Z_{EGFR} **14** Obs: 9174.7, cal: 79174.3; Z_{EGFR} **15** Obs:9020.2, cal: 9018.1; Z_{EGFR} **16** Obs: 9033.5, cal: 9029.1.

4.2.6 Traceless synthesis of modified histone as a potential application of the self-editing strategy

Synthetic or semi-synthetic histones containing site-specifically installed post-translational modifications are invaluable reagents in investigating the epigenetic mechanisms controlling various DNA-related processes. Although a number of techniques have been developed to prepare such proteins, establishing an efficient and convenient traceless ligation method would be very

beneficiary to this area of research. To demonstrate the usefulness of our new scheme in this setting, a truncated histone H2A (1-113) **21** tagged with a C-ter Cys-Asn-His_6 self-editing tag was recombinantly expressed using E.coli cells. Again, the azide product **23** was obtained in good yield using the established protocol. The clean transformation of **23** to **24** by the self-editing reaction awaits an improved procedure that is able to minimize the side reactions.

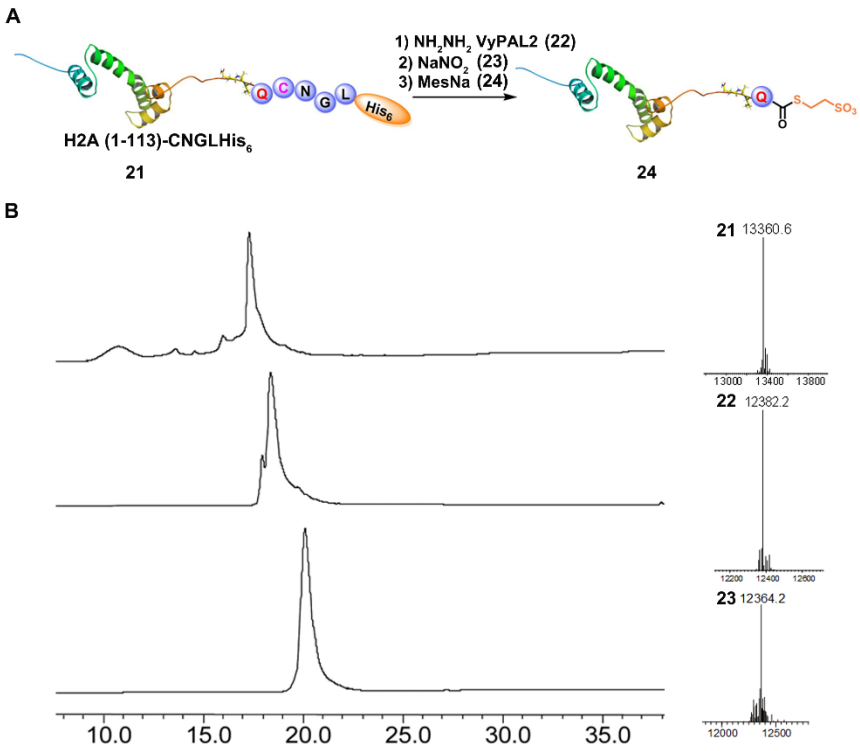


Figure 4.10. Traceless synthesis of modified H2A-Thr121 (PO_3^{2-}). A) Scheme for the semi-synthesis of Thr-phosphorylated H2A; B) HPLC analysis of each reaction during the synthetic process and ESI-MS characterization of the starting material and products (**21**, Obsvd: 13360.6, Calcd: 13360.4; **22**, Obsvd: 12382.2, Calcd: 12381.3; **23**, Obsvd: 12364.2 [M-N_2], Calcd: 12364.3). Please show structure drawings for 22 and 23 in B (just like Figure 4.9B).

4.3 Conclusion

In this project, we proposed a method for preparing peptide or protein C-terminal thioesters using an auto-processing mechanism of the activated C-terminal Cys-Asn dipeptide. Through PAL-mediated hydrazinolysis, followed by oxidation of the C-terminal hydrazide to an azide, the activated Xaa-cysteinyl-asparaginyl moiety undergoes auto-processing reactions to form a C-terminal Xaa-thioester with concomitant removal of the Cys-Asn dipeptide as a diketopiperazine. In future work, the conversion efficiency of the auto-processing reactions will be evaluated and optimized, which will also be used for the conversion of recombinant proteins (such as Z_{EGFR} and histone) to their thioesters with a non-Asn residue at the C-terminus. In conclusion, the newly proposed method has the potential to overcome the limit of PAL-mediated ligation reactions at the Asx site and expand it to many other amino acids. We anticipate many useful applications of this method for protein modification and semi-synthesis of post-translationally modified proteins.

References

1 Nguyen, G.K., Wang, S., Qiu, Y., Hemu, X., Lian, Y. and Tam, J.P., 2014. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. Nature chemical biology, 10(9), 732-738.

2 Nguyen, G.K., Kam, A., Loo, S., Jansson, A.E., Pan, L.X. and Tam, J.P., 2015. Butelase 1: a versatile ligase for peptide and protein macrocyclization. Journal of the American Chemical Society, 137(49), 15398-15401.

3 Cao, Y., Nguyen, G.K., Tam, J.P. and Liu, C.F., 2015. Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation. Chemical Communications, 51(97), 17289-17292.

4 Nguyen, G.K., Cao, Y., Wang, W., Liu, C.F. and Tam, J.P., 2015. Site-specific N-terminal labeling of peptides and proteins using butelase 1 and thiodepsipeptide. Angewandte Chemie, 127(52), 15920-15924.

5 Cao, Y., Nguyen, G.K., Chuah, S., Tam, J.P. and Liu, C.F., 2016. Butelase-mediated ligation as an efficient bioconjugation method for the synthesis of peptide dendrimers. Bioconjugate chemistry, 27(11), 2592-2596.

6 Hemu, X., Qiu, Y., Nguyen, G.K. and Tam, J.P., 2016. Total synthesis of circular bacteriocins by butelase 1. Journal of the American Chemical Society, 138(22), 6968-6971.

7 Nguyen, G.K., Hemu, X., Quek, J.P. and Tam, J.P., 2016. Butelase-mediated macrocyclization of d-amino-acid-containing peptides. Angewandte Chemie International Edition, 128(41), 12994-12998.

8 Nguyen, G.K., Qiu, Y., Cao, Y., Hemu, X., Liu, C.F. and Tam, J.P., 2016. Butelase-mediated cyclization and ligation of peptides and proteins. Nature protocols, 11(10), 1977-1988.

9 Bi, X., Yin, J., Hemu, X., Rao, C., Tam, J.P. and Liu, C.F., 2018. Immobilization and intracellular delivery of circular proteins by modifying a genetically incorporated unnatural

amino acid. Bioconjugate chemistry, 29(7), 2170-2175.

10 Harmand, T.J., Bousbaine, D., Chan, A., Zhang, X., Liu, D.R., Tam, J.P. and Ploegh, H.L., 2018. One-pot dual labeling of IgG 1 and preparation of C-to-C fusion proteins through a combination of sortase A and butelase 1. Bioconjugate chemistry, 29(10), 3245-3249.

11 James, A.M., Haywood, J., Leroux, J., Ignasiak, K., Elliott, A.G., Schmidberger, J.W., Fisher, M.F., Nonis, S.G., Fenske, R., Bond, C.S. and Mylne, J.S., 2019. The macrocyclizing protease butelase 1 remains autocatalytic and reveals the structural basis for ligase activity. The Plant Journal, 98(6), 988-999.

12 Hemu, X., Zhang, X., Nguyen, G.K., To, J., Serra, A., Loo, S., Sze, S.K., Liu, C.F. and Tam, J.P., 2021. Characterization and application of natural and recombinant butelase-1 to improve industrial enzymes by end-to-end circularization. RSC Advances, 11(37), 23105-23112.

13 Harris, K.S., Durek, T., Kaas, Q., Poth, A.G., Gilding, E.K., Conlan, B.F., Saska, I., Daly, N.L., Van Der Weerden, N.L., Craik, D.J. and Anderson, M.A., 2015. Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. Nature communications, 6(1), 1-10.

14 Yang, R., Wong, Y.H., Nguyen, G.K., Tam, J.P., Lescar, J. and Wu, B., 2017. Engineering a catalytically efficient recombinant protein ligase. Journal of the American Chemical Society, 139(15), 5351-5358.

15 Hemu, X., El Sahili, A., Hu, S., Wong, K., Chen, Y., Wong, Y.H., Zhang, X., Serra, A., Goh, B.C., Darwis, D.A. and Chen, M.W., et al., 2019. Structural determinants for peptide-bond formation by asparaginyl ligases. Proceedings of the National Academy of Sciences, 116(24), 11737-11746.

16 Williamson, D.J., Fascione, M.A., Webb, M.E. and Turnbull, W.B., 2012. Efficient N-terminal labeling of proteins by use of sortase. Angewandte Chemie International Edition, 124(37), 9511-9514.

17 Bolscher, J.G., Oudhoff, M.J., Nazmi, K., Antos, J.M., Guimaraes, C.P., Spooner, E., Haney, E.F., Vallejo, J.J.G., Vogel, H.J., Hof, W.V.T. and Ploegh, H.L., 2011. Sortase A as a tool for high-yield histatin cyclization. The FASEB Journal, 25(8), 2650-2658.

18 Bi, X., Yin, J., Nguyen, G.K., Rao, C., Halim, N.B.A., Hemu, X., Tam, J.P. and Liu, C.F., 2017. Enzymatic engineering of live bacterial cell surfaces using butelase 1. Angewandte Chemie International Edition, 56(27), 7822-7825.

19 Bi, X., Yin, J., Zhang, D., Zhang, X., Balamkundu, S., Lescar, J., Dedon, P.C., Tam, J.P. and Liu, C.F., 2020. Tagging transferrin receptor with a disulfide FRET probe to gauge the redox state in endosomal compartments. Analytical Chemistry, 92(18), 12460-12466.

20 Schmidt, M., Toplak, A., Quaedflieg, P.J. and Nuijens, T., 2017. Enzyme-mediated ligation technologies for peptides and proteins. Current opinion in chemical biology, 38, 1-7.

21 Rosen, C.B. and Francis, M.B., 2017. Targeting the N terminus for site-selective protein modification. Nature chemical biology, 13(7), 697-705.

22 Nguyen, K.N.T., Nguyen, G.K.T., Nguyen, P.Q.T., Ang, K.H., Dedon, P.C. and Tam, J.P., 2016. Immunostimulating and Gram-negative-specific antibacterial cyclotides from the butterfly pea (Clitoria ternatea). The FEBS journal, 283(11), 2067-2090.

23 Nguyen, G.K. and Wong, C.T., 2017. Making circles: Recent advance in chemical and enzymatic approaches in peptide macrocyclization. Journal of Biochemistry and Chemical Sciences, 1, 1-13.

24 Dall, E., Zauner, F. B., Soh, W. T., Demir, F., Dahms, S. O., Cebrele, C., Huesgen, P. F., Brandstetter, H., 2020. Structural and functional studies of Arabidopsis thaliana legumain beta reveal isoform specific mechanisms of activation and substrate recognition. Journal of Biological Chemistry, 295, 13047–13064.

25 Du, J., Yap, K., Chan, L. Y., Rehm, F. B., Looi, F. Y., Poth, A. G., Gilding, E. K., Kaas, Q., Durek, T., Craik, D. J., 2020. A bifunctional asparaginyl endopeptidase efficiently catalyzes

both cleavage and cyclization of cyclic trypsin inhibitors. Nature Communication, 11, 1-11.

26 Dall, E., Brandstetter, H., 2012. Activation of legumain involves proteolytic and conformational events, resulting in a context- and substrate-dependent activity profile. Acta Crystallographica Section F, Structural Biology and Crystallization Communications, 68, 24-31.

27 Dall, E., Brandstetter, H., 2013. Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. Proceedings of the National Academy of Sciences, 110, 10940-10945.

28 Haywood, J., Schmidberger, J. W., James, A. M., Nonis, S. G., Sukhoverkov, K. V., Elias, M., Bond, C. S., Mylne, J. S., 2018. Structural basis of ribosomal peptide macrocyclization in plants. Elife, 7:e32955.

29 Dall, E., Fegg, J. C., Briza, P., Brandstetter. H., 2015. Structure and Mechanism of an Aspartimide-Dependent Peptide Ligase in Human Legumain. Angewandte Chemie International Edition, 54, 2917–2921.

30 Zhang, D., Wang, Z., Hu, S., Balamkundu, S., To, J., Zhang, X., Lescar, J., Tam, J. P., and Liu, C. F., 2021. pH-Controlled Protein Orthogonal Ligation Using Asparaginyl Peptide Ligases. Journal of the American Chemical Society,143 (23), 8704-8712.

31 Omar Boutureira O., and Bernardes, G. J. L., 2015. Advances in Chemical Protein Modification. Chemical Reviews, 115 (5), 2174–2195.

32 Hoyt, E. A., Cal, P. M. S., Oliveira, B. L., Bernardes, G. J. L., 2019. Contemporary approaches to site-selective protein modification. Nature Reviews Chemistry, 3, 147-171.

33 Zhang, Y., Park, K. Y., Suazo, K. F., Distefano, M. D., 2018. Recent progress in enzymatic protein labelling techniques and their applications. Chemical Society Reviews, 47, 9106-9136.

34 Antos, J.M., Popp, M.W.L., Ernst, R., Chew, G.L., Spooner, E. and Ploegh, H.L., 2009. A straight path to circular proteins. Journal of Biological Chemistry, 284(23), 16028-16036.

35 Popp, M.W., Dougan, S.K., Chuang, T.Y., Spooner, E. and Ploegh, H.L., 2011. Sortase-catalyzed transformations that improve the properties of cytokines. Proceedings of the National Academy of Sciences, 108(8), 3169-3174.

36 Gonçalves, M.S.T., 2009. Fluorescent labeling of biomolecules with organic probes. Chemical reviews, 109(1), 190-212.

37 Li, C., Plamont, M.A., Sladitschek, H.L., Rodrigues, V., Aujard, I., Neveu, P., Le Saux, T., Jullien, L. and Gautier, A., 2017. Dynamic multicolor protein labeling in living cells. Chemical Science, 8(8), 5598-5605.

38 Alvarez-Dorta, D., Thobie-Gautier, C., Croyal, M., Bouzelha, M., Mével, M., Deniaud, D., Boujtita, M. and Gouin, S.G., 2018. Electrochemically promoted tyrosine-click-chemistry for protein labeling. Journal of the American Chemical Society, 140(49), 17120-17126.

39 Lue, R.Y., Chen, G.Y., Hu, Y., Zhu, Q. and Yao, S.Q., 2004. Versatile protein biotinylation strategies for potential high-throughput proteomics. Journal of the American Chemical Society, 126(4), 1055-1062.

40 Liu, C.F. and Tam, J.P., 1994. Chemical ligation approach to form a peptide bond between unprotected peptide segments. Concept and model study. Journal of the American Chemical Society, 116(10), 4149-4153.

41 Dawson, P.E., Muir, T.W., Clark-Lewis, I. and Kent, S., 1995. BH Science 1994, 266, 776–779;(b) Tam, JP; Lu, C.-F.; Shao. Proceedings of the National Academy of Sciences, 92, 12485-12489.

42 Conibear, A.C., Watson, E.E., Payne, R.J. and Becker, C.F., 2018. Native chemical ligation in protein synthesis and semi-synthesis. Chemical Society Reviews, 47(24), 9046-9068.

43 Barber, C.J., Pujara, P.T., Reed, D.W., Chiwocha, S., Zhang, H. and Covello, P.S., 2013. The two-step biosynthesis of cyclic peptides from linear precursors in a member of the plant family Caryophyllaceae involves cyclization by a serine protease-like enzyme. Journal of Biological

Chemistry, 288(18), 12500-12510.

44 Lee, J., McIntosh, J., Hathaway, B.J. and Schmidt, E.W., 2009. Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. Journal of the American Chemical Society, 131(6), 2122-2124.

45 Luo, H., Hong, S.Y., Sgambelluri, R.M., Angelos, E., Li, X. and Walton, J.D., 2014. Peptide macrocyclization catalyzed by a prolyl oligopeptidase involved in α-amanitin biosynthesis. Chemistry & biology, 21(12), 1610-1617.

46 Xu, M.Q. and Evans Jr, T.C., 2001. Intein-mediated ligation and cyclization of expressed proteins. Methods, 24(3), 257-277.

47 Rashidian, M., Dozier, J.K. and Distefano, M.D., 2013. Enzymatic labeling of proteins: techniques and approaches. Bioconjugate chemistry, 24(8), 1277-1294.

48 Oteng-Pabi, S.K., Pardin, C., Stoica, M. and Keillor, J.W., 2014. Site-specific protein labelling and immobilization mediated by microbial transglutaminase. Chemical communications, 50(50), 6604-6606.

49 Lin, C.W. and Ting, A.Y., 2006. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. Journal of the American Chemical Society, 128(14), 4542-4543.

50 Rachel, N.M. and Pelletier, J.N., 2013. Biotechnological applications of transglutaminases. Biomolecules, 3(4), 870-888.

51 Bonasio, R., Carman, C.V., Kim, E., Sage, P.T., Love, K.R., Mempel, T.R., Springer, T.A. and von Andrian, U.H., 2007. Specific and covalent labeling of a membrane protein with organic fluorochromes and quantum dots. Proceedings of the National Academy of Sciences, 104(37), 14753-14758.

52 Baalmann, M., Best, M. and Wombacher, R., 2018. Site-specific protein labeling utilizing lipoic acid ligase (LplA) and bioorthogonal inverse electron demand diels-alder reaction. In

Noncanonical Amino Acids, Humana Press, New York, NY, 365-387.

53 Hauke, S., Best, M., Schmidt, T.T., Baalmann, M., Krause, A. and Wombacher, R., 2014. Two-step protein labeling utilizing lipoic acid ligase and Sonogashira cross-coupling. Bioconjugate chemistry, 25(9), 1632-1637.

54 Heal, W.P., Wickramasinghe, S.R., Leatherbarrow, R.J. and Tate, E.W., 2008. N-Myristoyl transferase-mediated protein labelling in vivo. Organic & biomolecular chemistry, 6(13), 2308-2315.

55 Heal, W.P., Wickramasinghe, S.R., Bowyer, P.W., Holder, A.A., Smith, D.F., Leatherbarrow, R.J. and Tate, E.W., 2008. Site-specific N-terminal labelling of proteins in vitro and in vivo using N-myristoyl transferase and bioorthogonal ligation chemistry. Chemical communications, (4), 480-482.

56 June, C.H., O'Connor, R.S., Kawalekar, O.U., Ghassemi, S. and Milone, M.C., 2018. CAR T cell immunotherapy for human cancer. Science, 359(6382), 1361-1365.

57 Benmebarek, M.R., Karches, C.H., Cadilha, B.L., Lesch, S., Endres, S. and Kobold, S., 2019. Killing mechanisms of chimeric antigen receptor (CAR) T cells. International journal of molecular sciences, 20(6), 1283.

58 Wang, W., Jiang, J. and Wu, C., 2020. CAR-NK for tumor immunotherapy: Clinical transformation and future prospects. Cancer letters, 472, 175-180.

59 Rafiq, S., Hackett, C. S., and Brentjens, R. J., 2020. Engineering strategies to overcome the current roadblocks in CAR T cell therapy. Nature reviews Clinical oncology, 17(3), 147-167.

60 Xie G. Z., Dong H., Liang Y., et al, 2020. CAR-NK cells: A promising cellular immunotherapy for cancer. EBioMedicine, 59, 102975.

61 Kim, M.S., Ma, J.S., Yun, H., Cao, Y., Kim, J.Y., Chi, V., Wang, D., Woods, A., Sherwood, L., Caballero, D. and Gonzalez, J., 2015. Redirection of genetically engineered CAR-T cells using bifunctional small molecules. Journal of the American Chemical Society, 137(8), 2832-

2835.

62 Yang, P., Wang, Y., Yao, Z., Gao, X., Liu, C., Wang, X., Wu, H., Ding, X., Hu, J., Lin, B. and Li, Q., et al., 2020. Enhanced Safety and Antitumor Efficacy of Switchable Dual Chimeric Antigen Receptor-Engineered T Cells against Solid Tumors through a Synthetic Bifunctional PD-L1-Blocking Peptide. Journal of the American Chemical Society, 142(44), 18874-18885.

63 Nord, K., Gunneriusson, E., Ringdahl, J., Ståhl, S., Uhlén, M., Nygren, P. Å., 1997. Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain. Nature Biotechnology, 15, 772-777.

64 Friedman, M., Orlova, A., Johansson, E., Eriksson, T. L., Höidén-Guthenberg, I., Tolmachev, V., et al, 2008. Directed evolution to low nanomolar affinity of a tumor-targeting epidermal growth factor receptor-binding affibody molecule. Journal of Molecular Biology, 376, 1388-1402.

65 Kelkar, S. S., Reineke, T. M., 2011. Theranostics: combining imaging and therapy. Bioconjugate Chemistry, 22, 1879-1903.

66 Funkhouser, J., 2002. Reinventing pharma: the theranostic revolution. Current Drug Discovery, 2, 17-19.

67 Xie, J., Lee, S., Chen, X., 2010. Nanoparticle-based theranostic agents. Advanced Drug Delivery Reviews. 62, 1064-1079.

68 Luo, S., Yang, X., Shi, C., 2016. Newly emerging theranostic agents for simultaneous cancertargeted imaging and therapy. Current Medicinal Chemistry, 23: 483-497.

69 Langbein, T., Weber, W.A., Eiber, M., 2019. Future of theranosctics: an outlook on precision oncology in nuclear medicine. Journal of Nuclear Medicine, 60, 13S-19S.

70 Sumer, B., Gao, J., 2008. Theranostic nanomedicine for cancer. Nanomedicine, 137–140.

71 Chen, H., Zhang, W., Zhu, G., Xie, J., Chen, X., 2017. Rethinking cancer nanotheranostics. Nature Reviews Materials, 2, 17024.

72 Lim, E.K., Kim, T., Paik, S., Haam, S., Huh, Y. M., Lee, K., 2015. Nanomaterials for theranostics: recent advances and future challenges. Chemical Reviews, 115, 327-394.

73 Zhang, L., Jing, D., Jiang , N., Rojalin, T., Baehr, C. M., Zhang, D., et al, 2020. Transformable peptide nanoparticles arrest HER2 signalling and cause cancer cell death in vivo. Nature Nanotechnology, 15, 145-153.

74 Dammes, N., Peer, D., 2020. Monoclonal antibody-based molecular imaging strategies and theranostic opportunities. Theranostics, 10, 938-955.

75 Moek, K. L., Giesen, D., Kok, I. C., de Groot, D. J., Jalving, M., Fehrmann, R. S., et al, 2017. Theranostics using antibodies and antibody-related therapeutics. Journal of Nuclear Medicine, 58, 83S-90S.

76 Spicer, C. D., Davis, B. G., 2014. Selective chemical protein modification. Nature Communications, 5, 4740.

77 Lotze, J., Reinhardt, U., Seitz, O., Beck-Sickinger, A. G., 2016. Peptide-tags for site-specific protein labelling in vitro and in vivo. Molecular bioSystems, 12, 1731-1745.

78 Abrahmsen, L., Tom, J., Burnier, J., Butcher, K.A., Kossiakoff, A., Wells, J.A., 1991. Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. Biochemistry, 30, 4151-4159.

79 Chang, T.K., Jackson, D.Y., Burnier, J.P., Wells, J.A., 1994. Subtiligase: a tool for semisynthesis of proteins. Proceeding of the National Academy of Sciences, 91, 12544-12548.

80 Henager, S.H., Chu, N., Chen, Z., Bolduc, D., Dempsey, D.R., Hwang, Y., Wells, J., et al, 2016. Enzyme-catalyzed expressed protein ligation. Nature Methods, 13, 925-927.

81 Tan, X., Yang, R., Liu, C.F., 2018. Facilitating subtiligase-catalyzed peptide ligation reactions by using peptide thioester substrates. Organic Letters, 20, 6691-6694.

82 Weeks, A.M., Wells, J.A.. 2020. Subtiligase-catalyzed peptide ligation. Chemical Reviews, 120, 3127-3160.

83 Schneewind, O., Fowler, A., Faull, K.F., 1995. Structure of the cell wall anchor of surface proteins in Staphylococcus aureus. Science, 268, 103-106.

84 Mao, H., Hart, S.A., Schink, A., Pollok, B.A., 2004. Sortase-mediated protein ligation: a new method for protein engineering. Journal of American Chemical Society, 126, 2670-2671.

85 Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E., Ploegh, H.L., 2007. Sortagging: a versatile method for protein labeling. Nature Chemical Biology, 3, 707-708.

86 Pishesha, N., Ingram, J.R., Ploegh, H.L., 2018. Sortase A: a model for transpeptidation and its biological applications. Annual Review of Cell and Developmental Biology, 34, 163-188.

87 Müntz, K., Shutov, A.D., 2002. Legumains and their functions in plants. Trends in Plant Science, 7, 340-344.

88 Dall, E., Brandstetter, H., 2016. Structure and function of legumain in health and disease. Biochimie, 122,126-150.

89 Min, W., Jones, D.H., 1994. In vitro splicing of concanavalin A is catalyzed by asparaginyl endopeptidase. Nature Structural Biology, 1, 502-504.

90 Gillon, A.D., Saska, I., Jennings, C.V., Guarino, R.F., Craik, D.J., Anderson, M.A., 2008. Biosynthesis of circular proteins in plants. Plant Journal, 53, 505-515.

91 James, A.M., Haywood, J., Mylne, J.S., 2018. Macrocyclization by asparaginyl endopeptidases. New Phytologist, 218, 923-928.

92 Bernath-Levin, K., Nelson, C., Elliott, A.G., Jayasena, A.S., Millar, A.H., Craik, D.J., et al, 2015. Peptide macrocyclization by a bifunctional endoprotease. Chemistry & Biology, 22, 571-582.

93 Zauner, F.B., Dall, E., Regl, C., Grassi, L., Huber, C.G., Cabrele, C., et al, 2018. Crystal structure of plant legumain reveals a unique two-chain state with pH-dependent activity regulation. Plant Cell, 30, 686-699.

94 Jackson, M.A., Gilding, E.K., Shafee, T., Harris, K.S., Kaas, Q., Poon, S., et al, 2018.

Molecular basis for the production of cyclic peptides by plant asparaginyl endopeptidases. Nature Communications, 9, 2411.

95 Zauner, F.B., Elsässer, B., Dall, E., Cabrele, C., Brandstetter, H., 2018. Structural analyses of Arabidopsis thaliana legumain γ reveal differential recognition and processing of proteolysis and ligation substrates. Journal of Biological Chemistry, 293, 8934-8946.

96 Hemu, X., To, J., Zhang, X., Tam, J.P., 2019. Immobilized peptide asparaginyl ligases enhance stability and facilitate macrocyclization and site-specific ligation. Journal of Organic Chemistry, 85, 1504-1512.

97 Ling, J.J., Policarpo, R.L., Rabideau, A.E., Liao, X., Pentelute, B.L., 2012. Protein thioester synthesis enabled by sortase. Journal of the American Chemical Society, 134, 10749-10752.

98 Li, Y.M., Li, Y.T., Pan, M., Kong, X.Q., Huang, Y.C., Hong, Z.Y., et al, 2014. Irreversible site-specific hydrazinolysis of proteins by use of sortase. Angewandte Chemie International Edition English, 126, 2230-2234.

99 Rehm, F.B., Harmand, T.J., Yap, K., Durek, T., Craik, D.J., Ploegh, H.L., 2019. Site-specific sequential protein labeling catalyzed by a single recombinant ligase. Journal of the American Chemical Society, 141, 17388-17393.

100 Antos, J.M., Chew, G.L., Guimaraes, C.P., Yoder, N.C., Grotenbreg, G.M., Popp, M.W., et al, 2009. Site-specific N-and C-terminal labeling of a single polypeptide using sortases of different specificity. Journal of the American Chemical Society, 131, 10800-10801.

101 Ellerby, H.M., Arap, W., Ellerby, L.M., Kain, R., Andrusiak, R., Rio. G.D., et al, 1999. Anti-cancer activity of targeted pro-apoptotic peptides. Nature Medicine, 5, 1032-1038.

102 Dubowchik, G.M., Firestone, R.A., Padilla, L., Willner, D., Hofstead, S.J., Mosure, K., et al, 2002. Cathepsin B-labile dipeptide linkers for lysosomal release of doxorubicin from internalizing immunoconjugates: model studies of enzymatic drug release and antigen-specific in vitro anticancer activity. Bioconjugate Chemistry,13, 855-869.

103 Davidson, N.E., Gelmann, E.P., Lippman, M.E., Dickson, R.B., 1987. Epidermal growth factor receptor gene expression in estrogen receptor-positive and negative human breast cancer cell lines. Molecular Endocrinology, 1, 216-223.

104 Nakase, I., Okumura, S., Katayama, S., Hirose, H., Pujals, S., Yamaguchi, H., et al, 2012. Transformation of an antimicrobial peptide into a plasma membrane-permeable, mitochondria-targeted peptide via the substitution of lysine with arginine. Chemical Communications, 48, 11097-11099.

105 Jin, Y., Song, L., Su, Y., Zhu, L., Pang, Y., Qiu, F., et al, 2011. Oxime linkage: a robust tool for the design of pH-sensitive polymeric drug carriers. Biomacromolecules, 12, 3460-3468.

106 Dirksen, A., Dawson, P.E., 2008. Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. Bioconjugate Chemistry, 19, 2543-2548.

107 Lee, S.B., Hassan, M., Fisher, R., Chertov, O., Chernomordik, V., Kramer-Marek, G., et al, 2008. Affibody molecules for in vivo characterization of HER2-positive tumors by near-infrared imaging. Clinical Cancer Research, 14, 3840-3849.

108 Xu, S., Zhao, Z. and Zhao, J., 2018. Recent advances in enzyme-mediated peptide ligation. Chinese Chemical Letters, 29(7),1009-1016.

109 Tam, J.P., Chan, N.Y., Liew, H.T., Tan, S.J., Chen, Y., 2020. Peptide Asparaginyl Ligases — Renegade Peptide Bond Makers. Science China Chemistry, 63, 296-207.

110 Jackson, M.A., Nguyen, L.T.T., Gilding, E.K., Durek, E., Craik, D.J., 2020. Make it or break it: plant AEPs on stage in biotechnology. Biotechnology Advances, 45, 107651.

111 Hemu, X.; El Sahili, A.; Hu, S.; Zhang, X.; Serra, A.; Goh, B. C.; Darwis, D. A.; Chen, M. W.; Sze, S. K.; Liu, C.-F., 2020. Turning an Asparaginyl Endopeptidase into a Peptide Ligase. ACS Catalysis, 10 (15), 8825-8834.

112 Wang, Z., Zhang, D., Hemu, X., Hu, S., To, J., Zhang, X., Lescar, J., Tam, J.P., Liu, C.F., 2021. Engineering protein theranostics using bio-orthogonal asparaginyl peptide ligases.

Theranostics, 11(12), 5863–5875.

113 Xia, Y., To, J., Chan, N.Y., Hu, S., Liew, H.T., Balamkundu, S., Zhang, X., Lescar, J., Bhattacharjya, S., Tam, J.P., Liu, C.F., 2021. Nγ-hydroxyasparagine: a multifunctional unnatural amino acid that is a good P1 substrate of asparaginyl peptide ligases. Angewandte Chemie International Edition English, 60 (41), 22207-22211.

114 Rehm, F.B., Tyler, T.J., Yap, K., de Veer, S. J., Craik, D. J. , and Durek, T., 2021. Enzymatic C-Terminal Protein Engineering with Amines. Journal of the American Chemical Society, 143(46), 19498–19504.

115 Edwards, J. O.; Pearson, R. G., 1962. The Factors Determining Nucleophilic Reactivities. Journal of the American Chemical Society, 84 (1), 16-24.

116 Majumdar, P., Pati,A., Patra, M., Behera, R.K., and Behera, A. K., 2014. Acid Hydrazides, Potent Reagents for Synthesis of Oxygen-, Nitrogen-, and/or Sulfur-Containing Heterocyclic Rings. Chemical Reviews, 114, 2942−2977.

117 Jones, R. M., and Offord, R. E., 1982. The proteinase-catalysed synthesis of peptide hydrazides. Biochemical Journal, 203, 125–129.

118 Komiya, C.; Shigenaga, A.; Tsukimoto, J.; Ueda, M.; Morisaki, T.; Inokuma, T.; Itoh, K.; Otaka, A., 2019. Traceless synthesis of protein thioesters using enzyme-mediated hydrazinolysis and subsequent self-editing of the cysteinyl prolyl sequence. Chemical Communications, 55, 7029-7032.

119 Chu, G.C., Pan, M., Li, J., Liu, S., Zuo, C., Tong, Z.B., Bai, J.S., Gong, Q., Ai, H., Fan, J. and Meng, X., 2019. Cysteine-aminoethylation-assisted chemical ubiquitination of recombinant histones. Journal of the American Chemical Society, 141(8), 3654-3663.

120 Chio, T.I., Demestichas, B.R., Brems, B.M., Bane, S.L. and Tumey, L.N., 2020. Expanding the versatility of microbial transglutaminase using α-effect nucleophiles as noncanonical substrates. Angewandte Chemie International Edition English, 59, 13814–13820.

121 Row, R.D., Roark, T.J., Philip, M.C., Perkins, L.L. and Antos, J.M., 2015. Enhancing the efficiency of sortase–mediated ligations through nickel–peptide complex formation. Chemical Communications, 51(63), 12548-12551.

122 Tang, T.S., Cardella, D., Lander, A.J., Li, X., Escudero, J.S., Tsai, Y.H. and Luk, L.Y., 2020. Use of an asparaginyl endopeptidase for chemo-enzymatic peptide and protein labeling. Chemical science, 11(23), 5881-5888.

123 Reed, S.A., Brzovic, D.A., Takasaki, S.S., Boyko, K.V. and Antos, J.M., 2020. Efficient Sortase-Mediated Ligation Using a Common C-Terminal Fusion Tag. Bioconjugate chemistry, 31(5), 1463-1473.

124 Rehm, F.B., Tyler, T.J., Yap, K., Durek, T. and Craik, D.J., 2021. Improved Asparaginyl-Ligase-Catalyzed Transpeptidation via Selective Nucleophile Quenching. Angewandte Chemie International Edition, 60(8), 4004-4008.

125 Kimura, T., Kaburaki, H., Miyamoto, S., Katayama, J., & Watanabe, Y. (1997). Discovery of a novel thrombopoietin mimic agonist peptide. The Journal of Biochemistry, 122(5), 1046-1051.

126 Garzon, A. M., and William, B. M., 2015. Use of thrombopoietin receptor agonists in childhood immune thrombocytopenia. Frontiers in pediatrics 13(3), 70.

127 Lin, S., Li, J., Shao, J., Zhang, J., He, X., Huang, D., Dong, L., Lin, J., Weng, W. and Cheng, K., 2021. Anisotropic magneto-mechanical stimulation on collagen coatings to accelerate osteogenesis. Colloids and Surfaces B: Biointerfaces, 112227.

128 Ta, M.H., Liuwantara, D. and Rangan, G.K., 2015. Effects of pyrrolidine dithiocarbamate on proliferation and nuclear factor-κB activity in autosomal dominant polycystic kidney disease cells. BMC nephrology, 16(1), 1-13.

129 Venkatesh, S. and Workman, J.L., 2015. Histone exchange, chromatin structure and the regulation of transcription. Nature reviews Molecular cell biology, 16(3), 178-189.

130 Müller, M.M. and Muir, T.W., 2015. Histones: at the crossroads of peptide and protein chemistry. Chemical reviews, 115(6), 2296-2349.

131 Guo, J., Wang, J., Lee, J.S. and Schultz, P.G., 2008. Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. Angewandte Chemie International Edition, 47(34), 6399-6401.

132 Neumann, H., Peak-Chew, S.Y. and Chin, J.W., 2008. Genetically encoding N ε-acetyllysine in recombinant proteins. Nature chemical biology, 4(4), 232-234.

133 Neumann, H., Hancock, S.M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D. and Chin, J.W., 2009. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Molecular cell, 36(1), 153-163.

134 Nguyen, D.P., Garcia Alai, M.M., Kapadnis, P.B., Neumann, H. and Chin, J.W., 2009. Genetically encoding N ϵ-methyl-l-lysine in recombinant histones. Journal of the American Chemical Society, 131(40), 14194-14195.

135 Kallappagoudar, S., Dammer, E.B., Duong, D.M., Seyfried, N.T. and Lucchesi, J.C., 2013. Expression, purification and proteomic analysis of recombinant histone H4 acetylated at lysine 16. Proteomics, 13(10-11), 1687-1691.

136 Zheng, Y.P.G. ed., 2015. Epigenetic technological applications. Academic Press.

137 David, Y. and Muir, T.W., 2017. Emerging chemistry strategies for engineering native chromatin. Journal of the American Chemical Society, 139(27), 9090-9096.

138 Qi, Y.K., Ai, H.S., Li, Y.M. and Yan, B., 2018. Total chemical synthesis of modified histones. Frontiers in chemistry, 6, 19.

139 Dawson, P.E., Muir, T.W., Clark-Lewis, I. and Kent, S.B., 1994. Synthesis of proteins by native chemical ligation. Science, 266(5186), 776-779.

140 Zheng, J.S., Tang, S., Huang, Y.C. and Liu, L., 2013. Development of new thioester

equivalents for protein chemical synthesis. Accounts of chemical research, 46(11), 2475-2484.

141 Johnson, E.C. and Kent, S.B., 2006. Insights into the mechanism and catalysis of the native chemical ligation reaction. Journal of the American Chemical Society, 128(20), 6640-6646.

142 Zheng, J.S., Tang, S., Qi, Y.K., Wang, Z.P. and Liu, L., 2013. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. Nature protocols, 8(12), 2483-2495.

143 Li, J., Li, Y., He, Q., Li, Y., Li, H. and Liu, L., 2014. One-pot native chemical ligation of peptide hydrazides enables total synthesis of modified histones. Organic & biomolecular chemistry, 12(29), 5435-5441.

144 Huang, Y.C., Fang, G.M. and Liu, L., 2016. Chemical synthesis of proteins using hydrazide intermediates. National Science Review, 3(1), 107-116.

145 Lan, H., Wu, K., Zheng, Y., Pan, M., Huang, Y.C., Gao, S., Zheng, Q.Y., Zheng, J.S., Li, Y.M., Xiao, B. and Liu, L., 2016. Total synthesis of mambalgin-1/2/3 by two-segment hydrazide-based native chemical ligation. Journal of Peptide Science, 22(5), 320-326.

146 Thom, J., Anderson, D., McGregor, J. and Cotton, G., 2011. Recombinant protein hydrazides: application to site-specific protein PEGylation. Bioconjugate chemistry, 22(6), 1017-1020.

147 Thompson, R.E. and Muir, T.W., 2019. Chemoenzymatic semisynthesis of proteins. Chemical reviews, 120(6), 3051-3126.

148 Berrade, L. and Camarero, J.A., 2009. Expressed protein ligation: a resourceful tool to study protein structure and function. Cellular and molecular life sciences, 66(24), 3909-3922.

149 Wood, D.W. and Camarero, J.A., 2014. Intein applications: from protein purification and labeling to metabolic control methods. Journal of Biological Chemistry, 289(21), 14512-14519.

150 Hayashi, R., Bai, Y. and Hata, T., 1975. Kinetic studies of carboxypeptidase Y: I. Kinetic parameters for the hydrolysis of synthetic substrates. The Journal of Biochemistry, 77(1), 69-79.

151 Saska, I., Gillon, A.D., Hatsugai, N., Dietzgen, R.G., Hara-Nishimura, I., Anderson, M.A. and Craik, D.J., 2007. An asparaginyl endopeptidase mediates in vivo protein backbone cyclization. Journal of Biological Chemistry, 282(40), 29721-29728.

152 Craik, D.J. and Malik, U., 2013. Cyclotide biosynthesis. Current opinion in chemical biology, 17(4), 546-554.

153 Kawakami, T. and Aimoto, S., 2007. Peptide ligation using a building block having a cysteinyl prolyl ester (CPE) autoactivating unit at the carboxy terminus. Chemistry letters, 36(1), 76-77.

154 Ueda, M., Komiya, C., Arii, S., Kusumoto, K., Denda, M., Okuhira, K., Shigenaga, A. and Otaka, A., 2020. Sequence-independent traceless method for preparation of peptide/protein thioesters using CPaseY-mediated hydrazinolysis. Chemical and Pharmaceutical Bulletin, c20-00674.

155 Dingpeng Zhang, Zhen Wang, Side Hu, Ning-Yu Chan, Heng Tai Liew, Julien Lescar, James Tam, Chuan-Fa Liu. 2021. Asparaginyl endopeptidase-mediated protein C-terminal hydrazinolysis for the synthesis of bioconjugates. Bioconjugate Chemistry, 33, 238–247.

Publication list

(1) **Wang. Z,**† Zhang. D,† Hu. S, Lescar. J, James. P. T and Liu. CF. 2021. A removable cysteinyl-asparaginyl tag for traceless protein ligation. Manuscript under preparation.

(2) **Wang. Z,**† Zhang. D,† Carmen. Pui, Hu. S, Lescar. J, James. P. T and Liu. CF. 2021. Structural-guided engineering of asparaginyl ligases with terminus specific recognition for protein orthogonal ligation. (Manuscript under preparation)

(3) **Wang. Z,**† Zhang. D,† Hu. S, Simin. Fang, Lescar. J, James. P. T and Liu. CF. 2021. Engineering multi-functional protein biologics through PAL-mediated hydrazide ligation. (Manuscript under preparation)

(4) Hu. S, El. Sahili, Kishore. S, Wong. Yee, Hemu. X, Goh. Boon, **Wang**. **Z**, Tam. J, Liu. C, Lescar. J. Structural basis for proenzyme maturation, substrate recognition and ligation by a hyperactive Asn peptide ligase. *The Plant Cell*, 34(12), 4936-4949.

(5) Chen,Y; Zhang, D; Zhang, X; **Wang, Z**; Liu, C.F. and Tam, J. P. 2021. Site-specific Protein Modifications by An Engineered Asparaginyl Endopeptidase from Viola canadensis. *Frontiers in Chemistry.* 9: 768854.

(6) Zhang, D;† **Wang, Z;**† Hu, S; Lescar, J; Tam, J. P; and Liu, C.F. 2022. Asparaginyl endopeptidase-mediated protein C-terminal hydrazinolysis for the synthesis of bioconjugates. *Bioconj. Chem.* 33, 1, 238 - 247.

(7) Zhang, D;† **Wang, Z;**† Hu, S; Balamkundu, S; To, J; Zhang, X; Lescar, J; Tam, J.P; and Liu, C.F. A versatile peptide ligase for precision tailoring of proteins. 2022. *IJMS*. 23(1), 458.

(8) **Wang, Z**;† Zhang, D;† Hu, S; Lescar, J; Tam, J.P; and Liu, C.F. 2022. PAL-mediated ligation for protein and cell surface modification. *Chemical Protein Synthesis*. Humana. New York. NY. 177-193.

(9) Zhang, D;† **Wang, Z;**† Hu, S; Balamkundu, S; To, J; Zhang, X; Lescar, J; Tam, J.P; and Liu, C.F. 2021. pH-Controlled Protein Orthogonal Ligation Using Asparaginyl Peptide Ligases. *Journal of the American Chemical Society,*143 (23), 8704-8712.

(10) **Wang, Z;**† Zhang, D;† Hemu, X; Hu, S; To, J; Zhang, X; Lescar, J; Tam, J.P. and Liu, C.F. 2021. Engineering protein theranostics using bio-orthogonal asparaginyl peptide ligases. *Theranostics, 11*(12), 5863.

(11) Zhang, D; **Wang, Z**; Bi, X; Hemu, X; Sashili, A; Zhang, X; Hu, S; To, J; Lescar, J; Tam, J. P; and Liu, C-F. Peptide asparaginyl ligases: versatile tools in biotechnology. *The 16th Chinese International Peptide Symposium (CPS2020)*. September 25-27, 2020; Hefei, Anhui, China (Conference presentation. Invited talk given by CF Liu).

Appendix A. Supplementary data for Chapter 2.

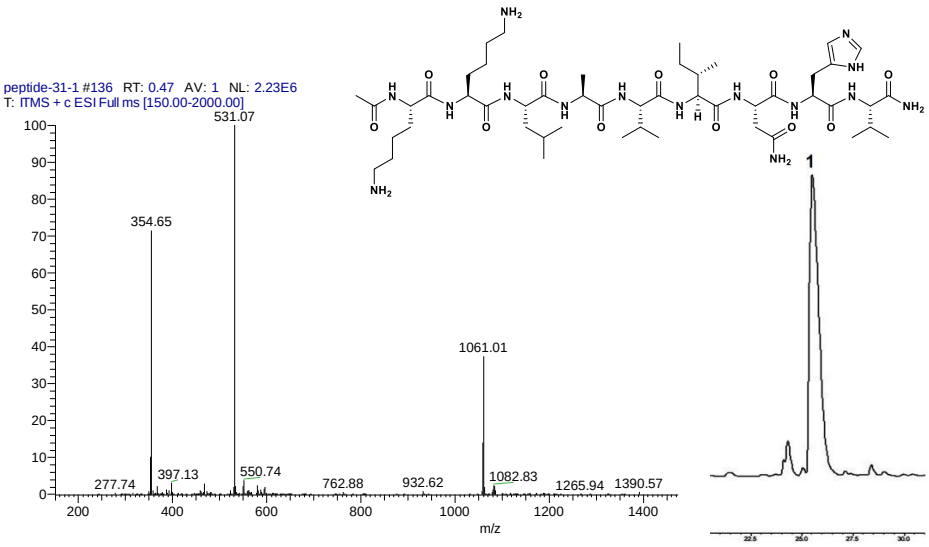


Figure 2.15. HPLC and ESI-MS of peptide **1**. ESI-MS of peptide **1:** 1061.01 (observed), 1062.27 (calculated).

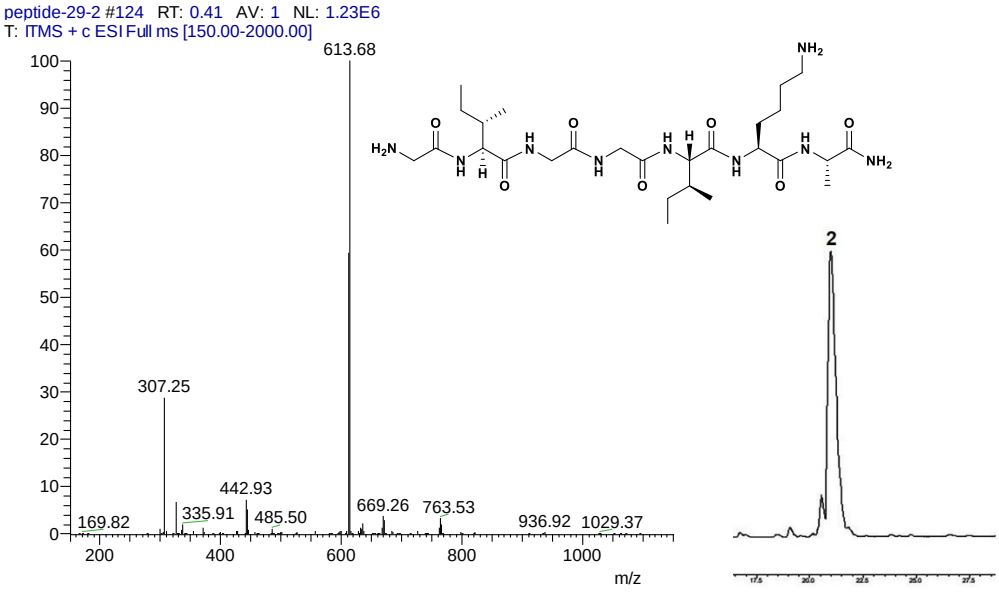


Figure 2.16. ESI-MS and HPLC of peptide **2**. ESI-MS of peptide **2:** ESI-MS: 613.68 (observed), 613.76 (calculated).

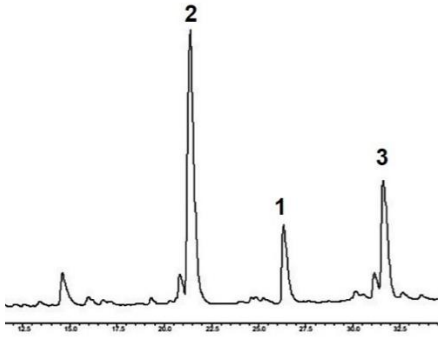


Figure 2.17. HPLC monitoring of the ligation between peptide **1** and **2** using butelase-1 to yield peptide **3** in kinetic studies.

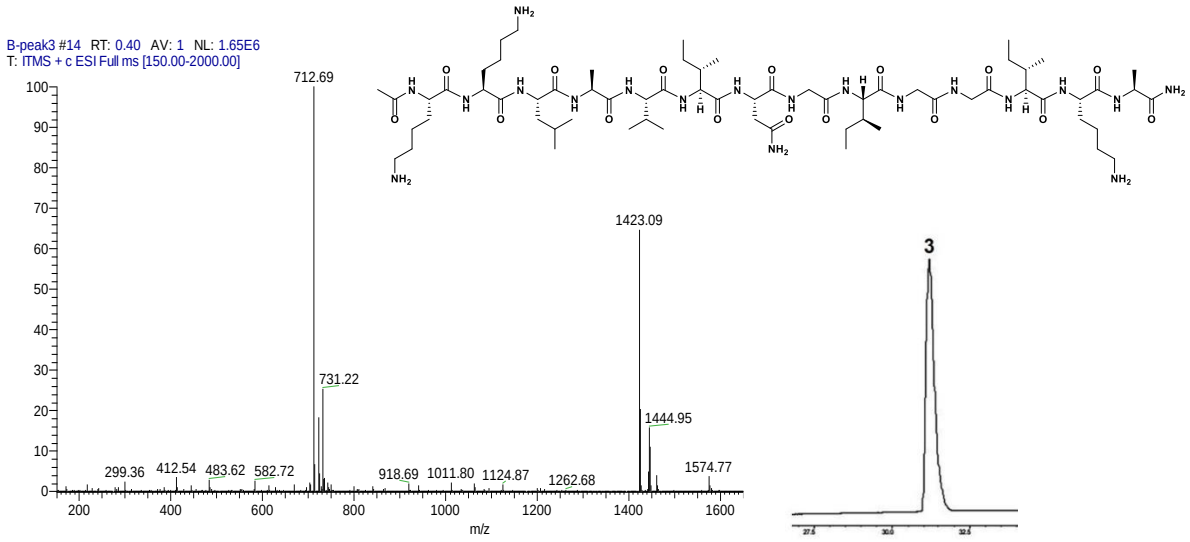


Figure 2.18 ESI-MS and HPLC of Peptide **3**. ESI-MS of peptide **3:** ESI-MS: 1423.09 (observed), 1422.78 (calculated).

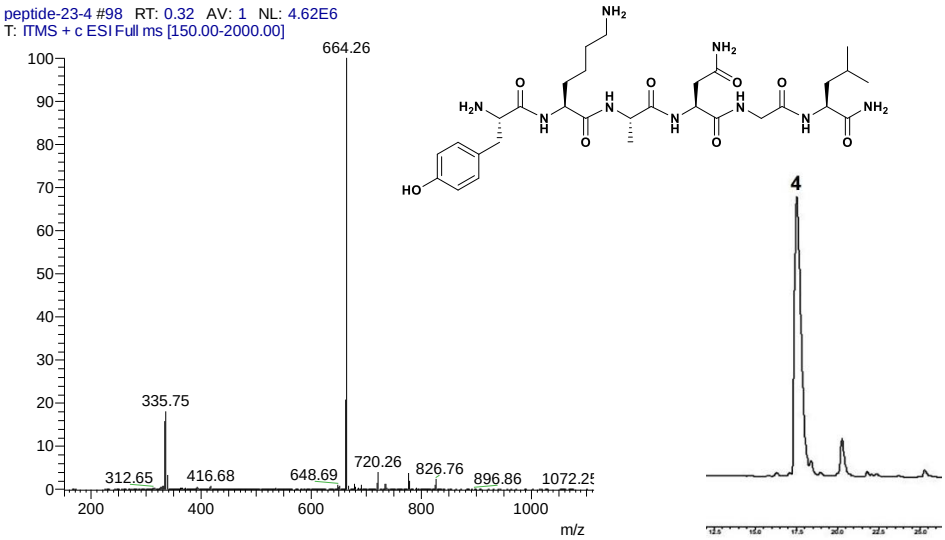


Figure 2.19. ESI-MS and HPLC of peptide **4**. ESI-MS of peptide **4:** ESI-MS: 664.26 (observed), 663.78 (calculated).

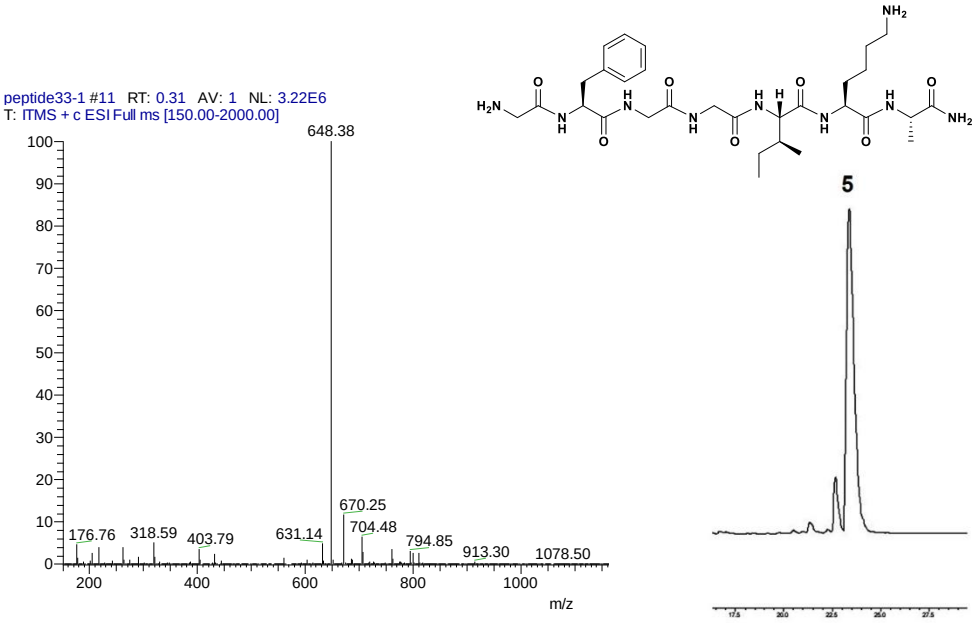


Figure 2.20. ESI-MS and HPLC of peptide **5**. ESI-MS of peptide **5:** ESI-MS: 648.38 (observed), 647.78 (calculated).

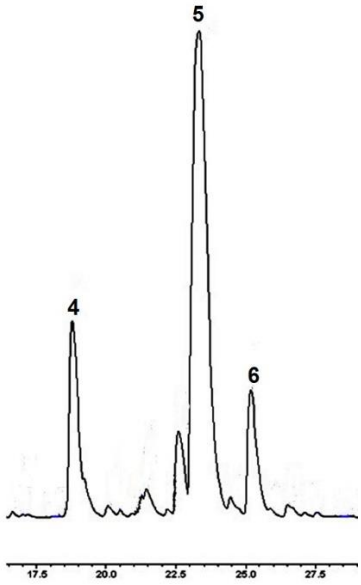


Figure 2.21. HPLC monitoring of the ligation between peptide **4** and **5** to yield peptide **6** using VyPAL2 in 5 min for kinetic studies.

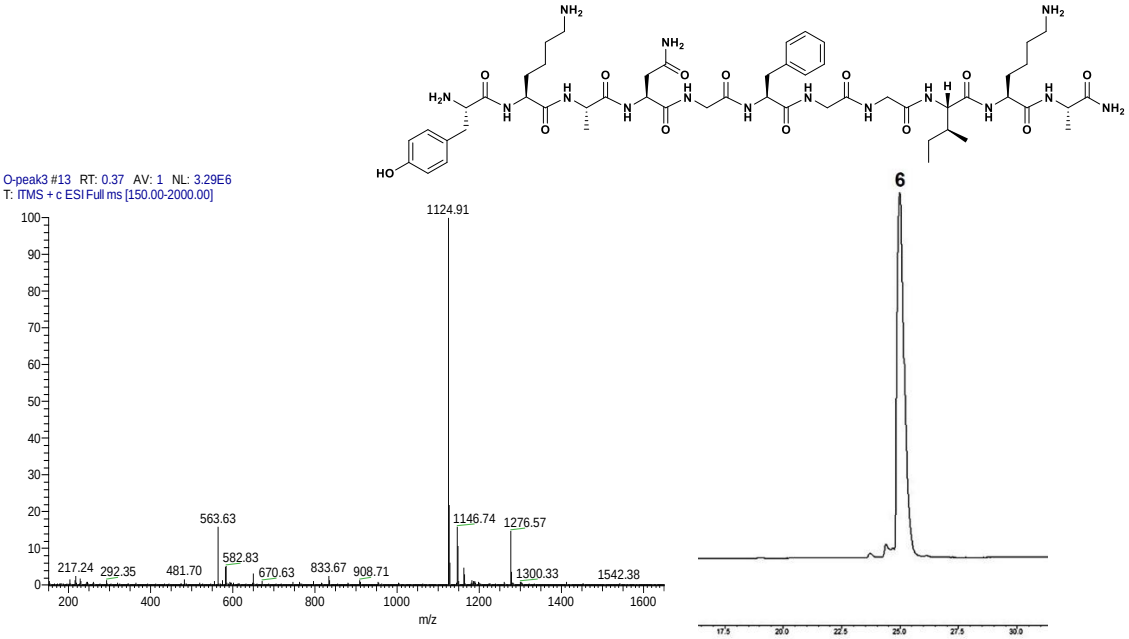


Figure 2.22. ESI-MS and HPLC of peptide **6**. ESI-MS of peptide **6:** ESI-MS: 1124.91 (observed), 1124.31 (calculated).

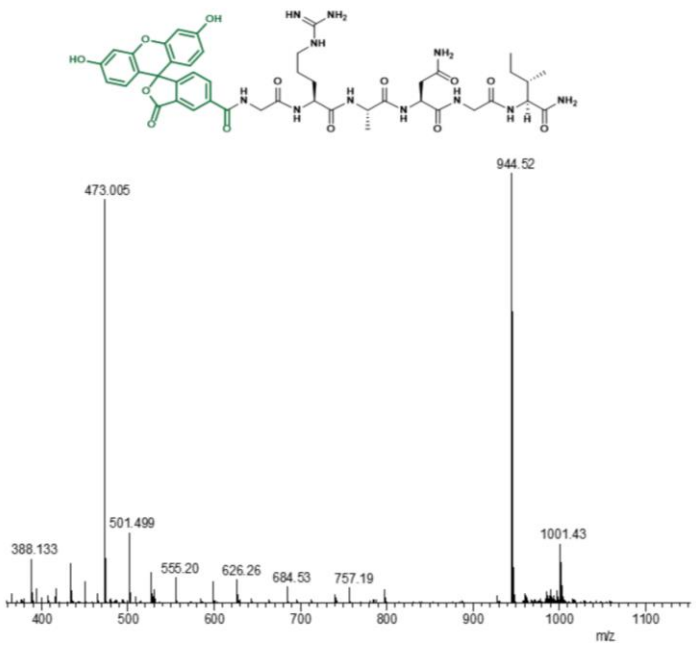


Figure 2.23. ESI-MS and HPLC of peptide **9**. ESI-MS of peptide **9:** ESI-MS: 944.52 (observed), 943.97 (calculated).

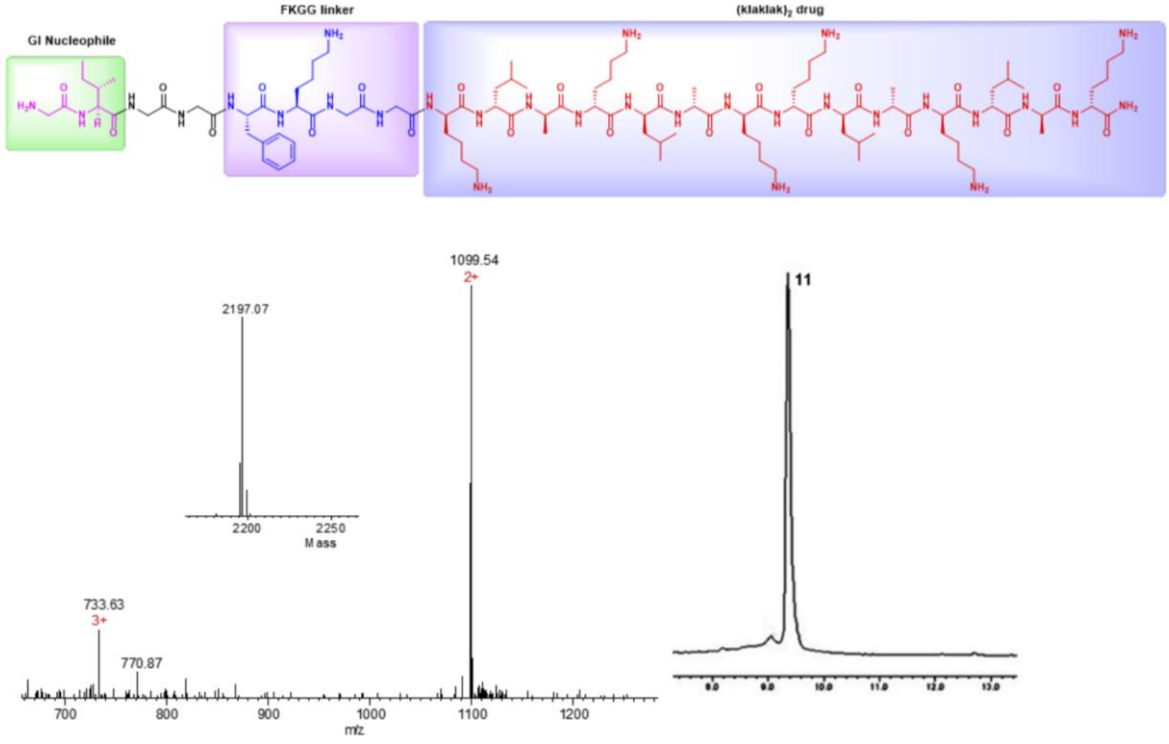


Figure 2.24. ESI-MS and HPLC of peptide **11**. ESI-MS of peptide **11:** ESI-MS: 2197.07 (observed), 2196.81 (calculated).

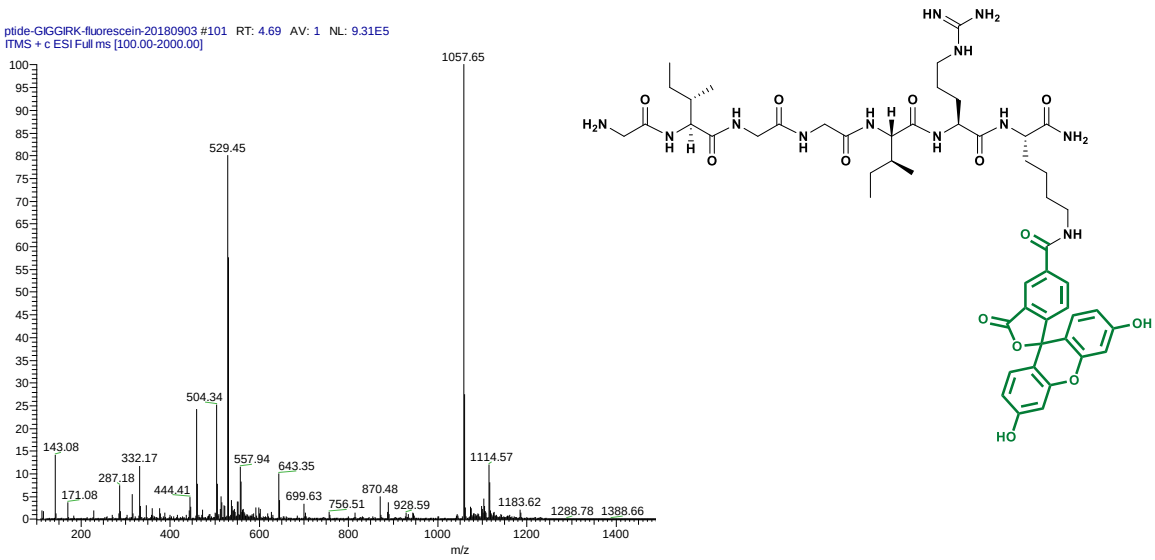


Figure 2.25. ESI-MS and HPLC of peptide **23**. ESI-MS of peptide **23:** ESI-MS: 1057.65 (observed), 1057.18 (calculated).

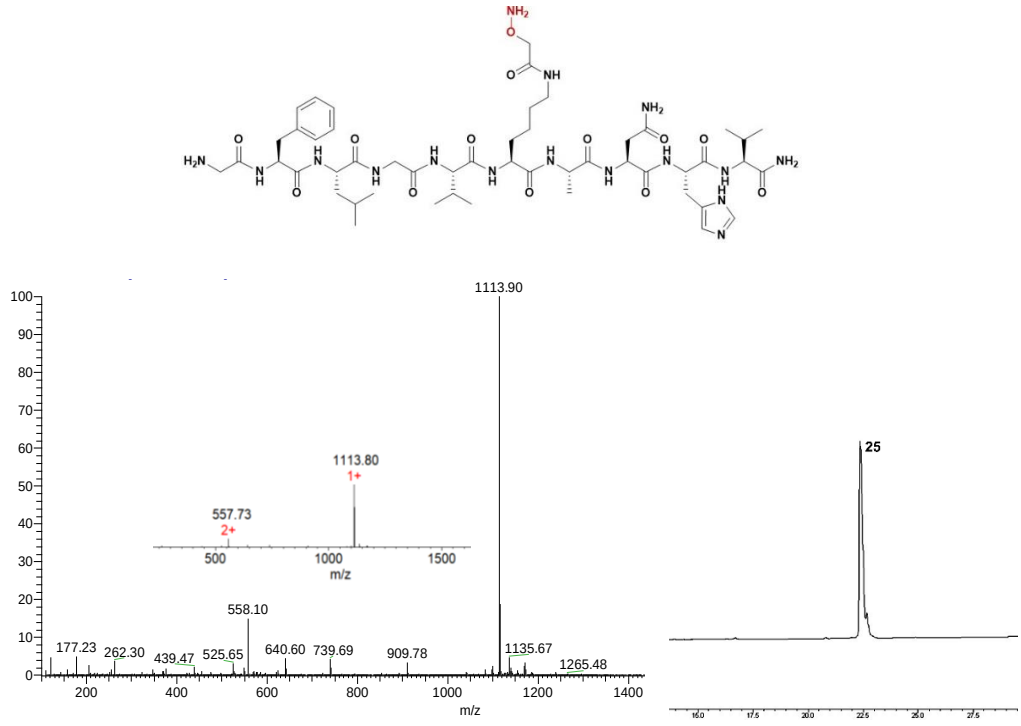


Figure 2.26. ESI-MS and HPLC of peptide **25**; ESI-MS of peptide **25:** ESI-MS: 1113.90 (observed), 1113.29 (calculated).

Appendix B. Supplementary data for Chapter 3.

The reactions were performed by adding 25 nM of VyPAL2 to a mixture of 50 μM peptide **2**-NGF substrate (1 eq) and 10 mM nucleophile **1a**-**1t** (200 eq) at 37 °C and pH 7.0 for 30 - 60 min. The reaction mixtures were analyzed by HPLC after reaction. Besides, the products were collected during HPLC analysis and characterized by ESI-MS as following.

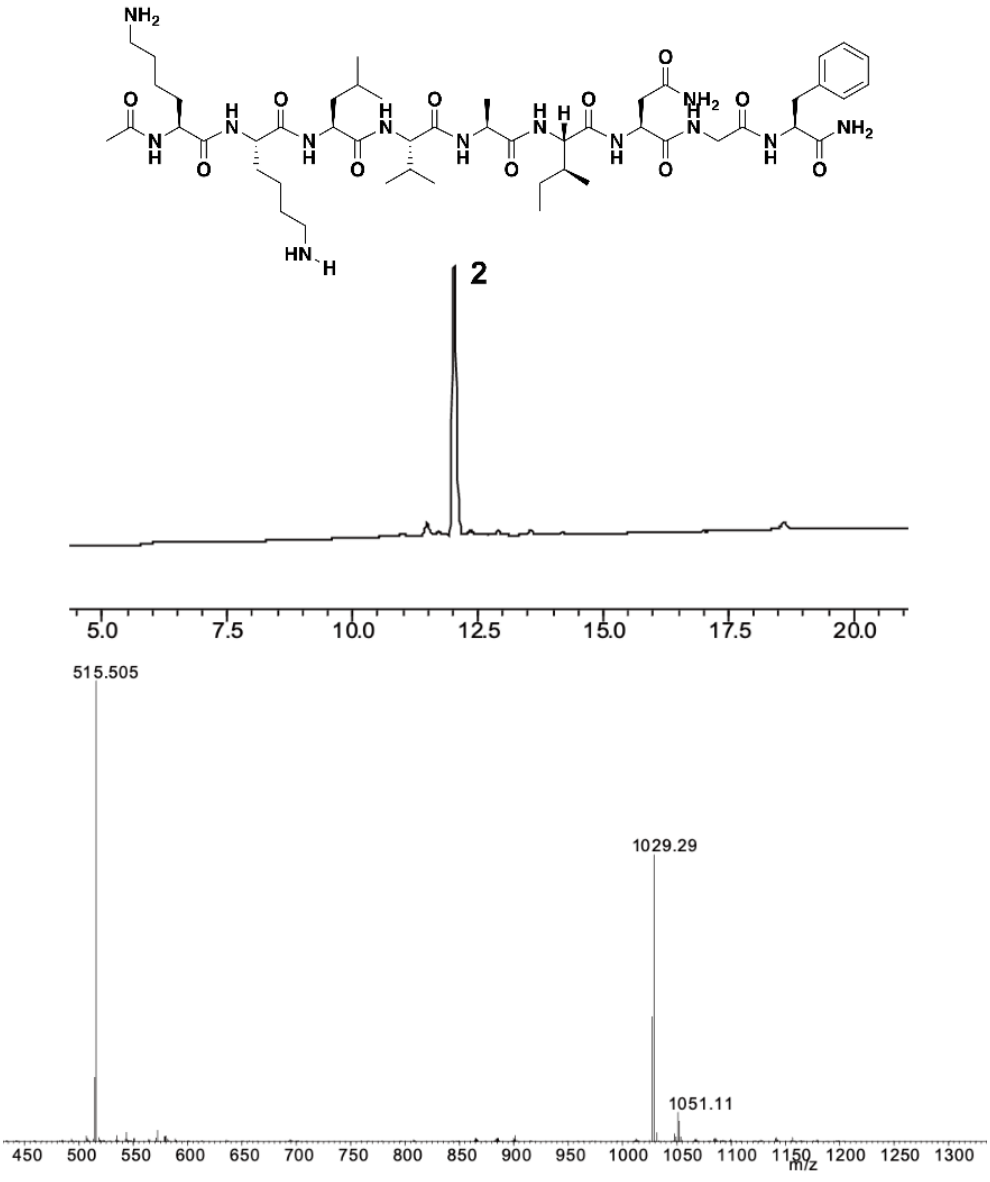


Figure 3.28. HPLC analysis and ESI-MS characterization of the starting material peptide **2** (m/z $[M+H]^+$: calcd 1029.63, obsvd 1029.29).

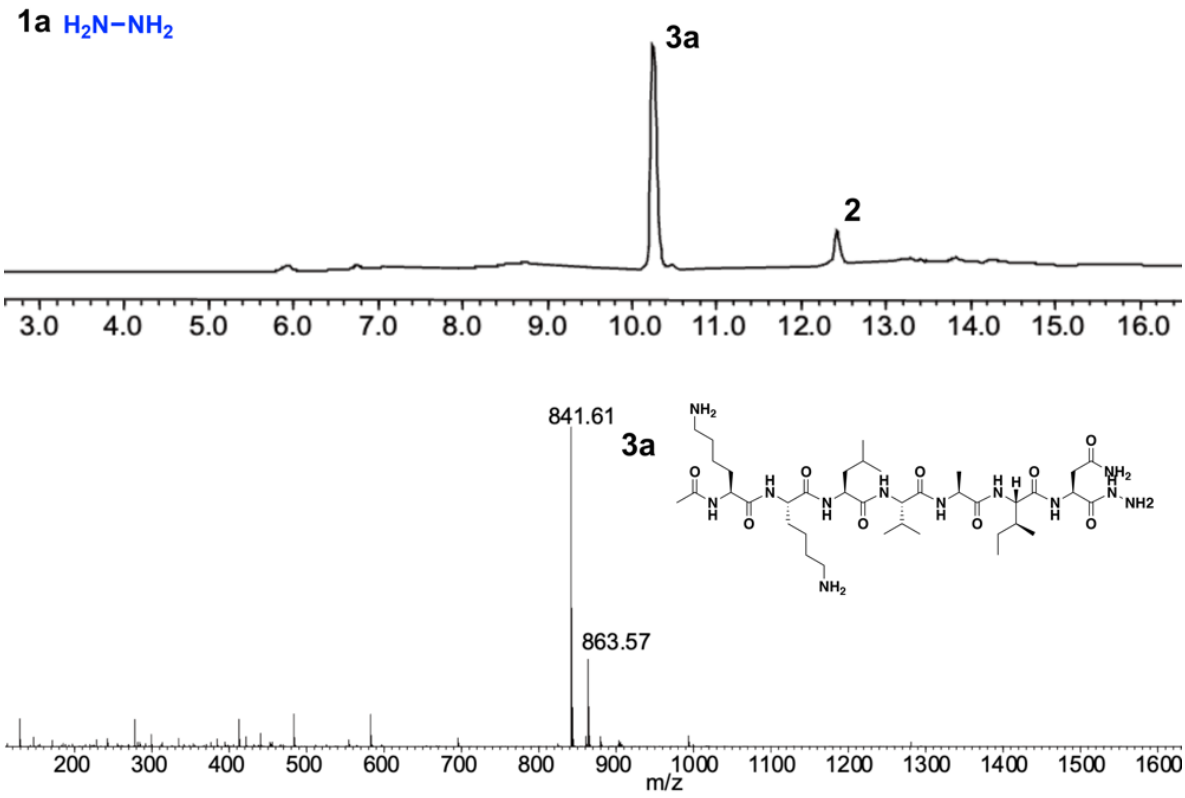


Figure 3.29. HPLC monitoring and ESI-MS characterization of the reaction of peptide **2** with compound **1a** (hydrazine) for 10 min, leading to product **3a** in 95% yield. ESI-MS data for **3a** m/z $[M+H]^+$: calcd 841.55, obsvd 841.62.

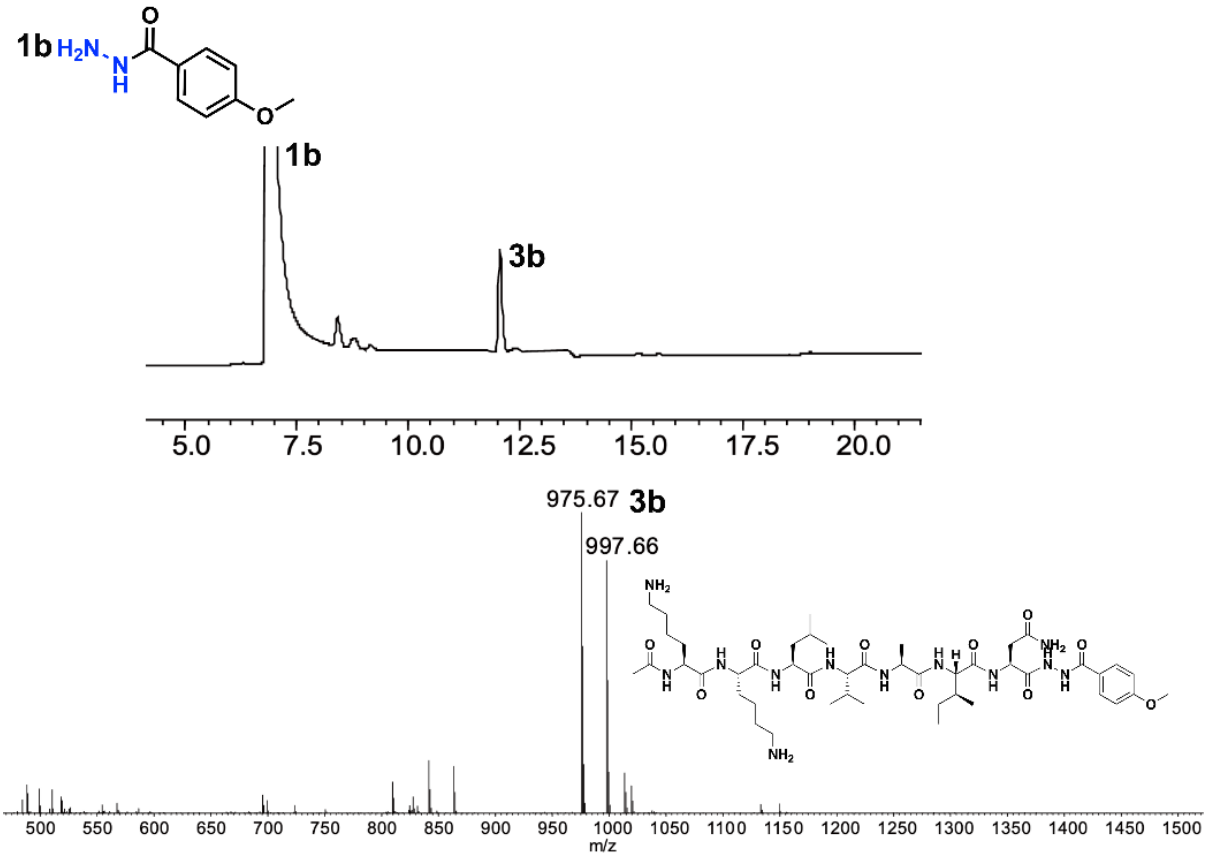


Figure 3.30. The HPLC analysis and ESI-MS characterization of peptide **2** reacting with compound **1b** for 60 min to obtain **3b** in 73% yield and the ESI data for the peptide **3b**, calcd 974.59, obsvd 975.67 $[M+H]^+$.

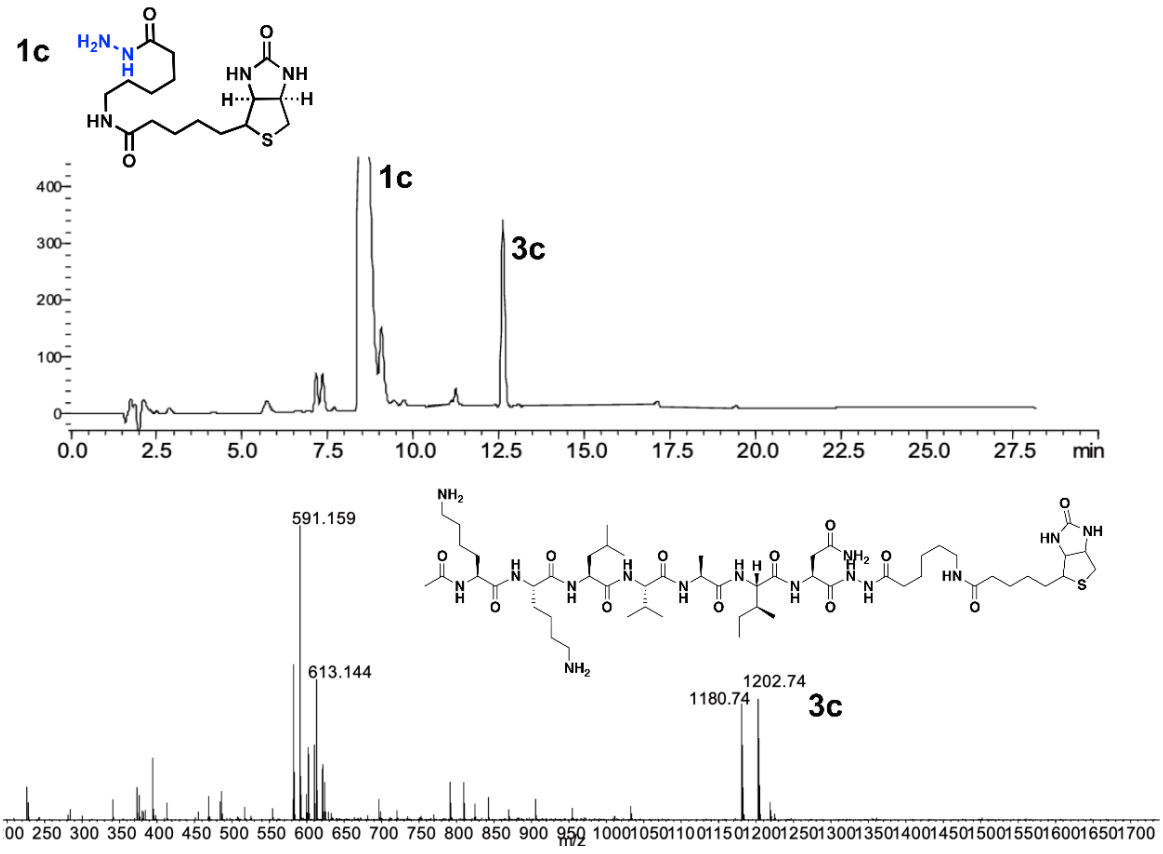


Figure 3.31. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1c** for 60 min, leading to product **3c** in >90% yield. ESI-MS data for **3c** m/z $[M+H]^+$: calcd 1180.72, obsvd 1180.74.

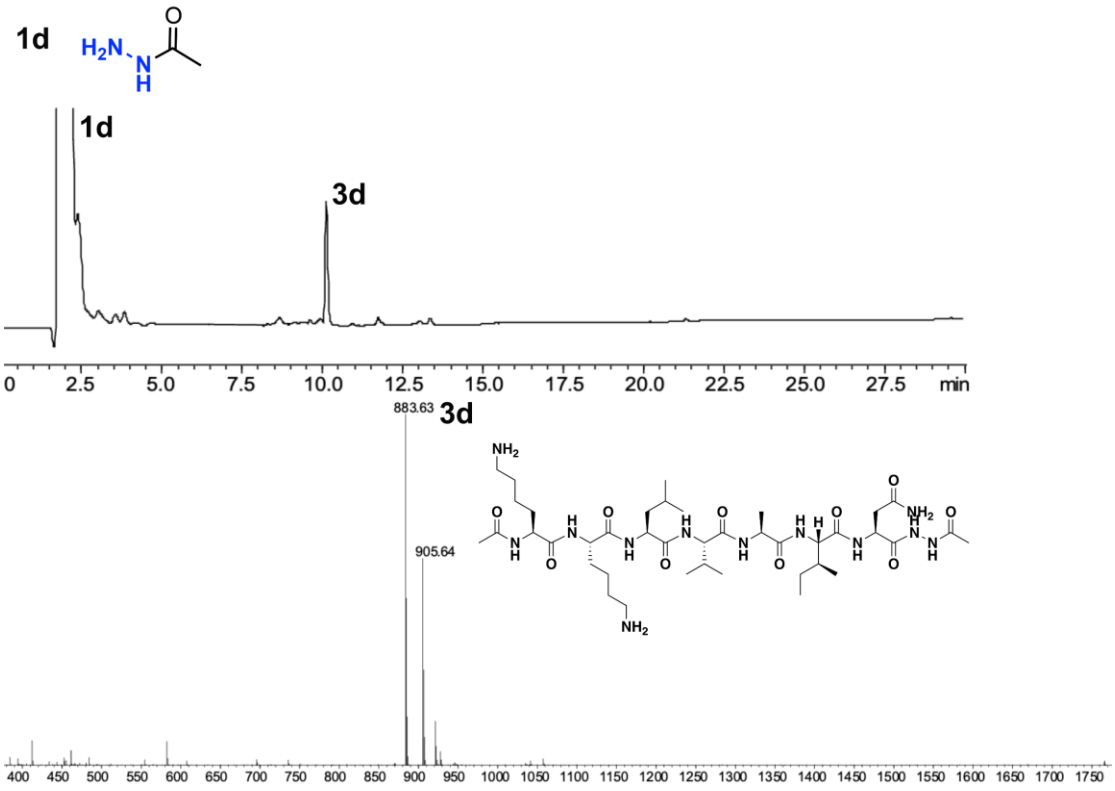


Figure 3.32. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1d** for 30 min, leading to product **3d** in >92% yield. ESI-MS data for **3d** m/z $[M+H]^+$: calcd 883.57, obsvd 883.63.

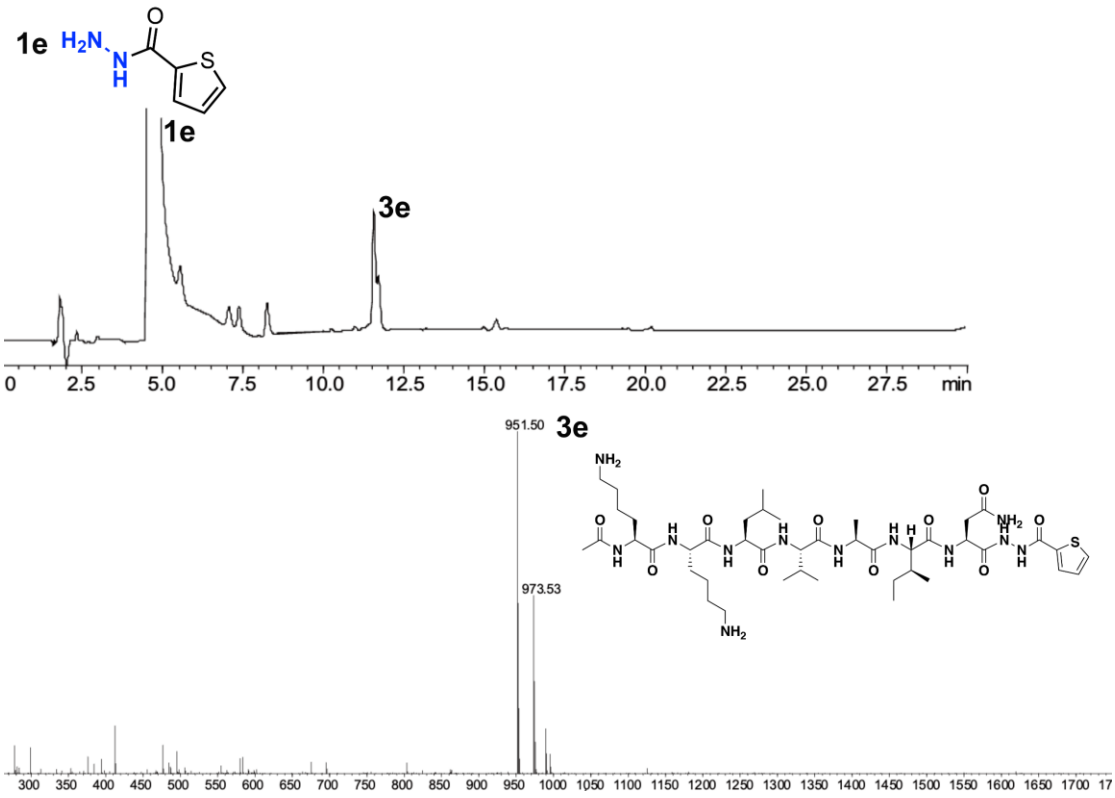


Figure 3.33. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1e** for 30 min, leading to product **3e** in >90% yield. ESI-MS data for **3e** m/z $[M+H]^+$: calcd 951.54, obsvd 951.50.

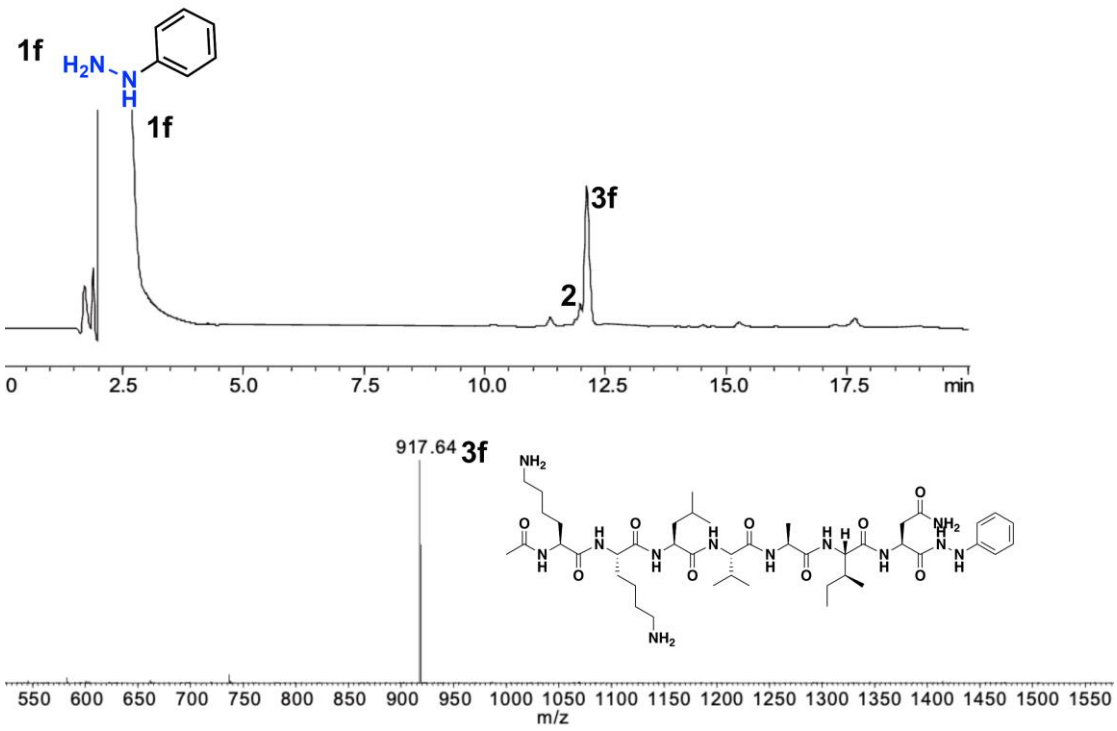


Figure 3.34. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1f** for 60 min, leading to product **3f** in >93% yield. ESI-MS data for **3f** m/z $[M+H]^+$: calcd 917.59, obsvd 917.64.

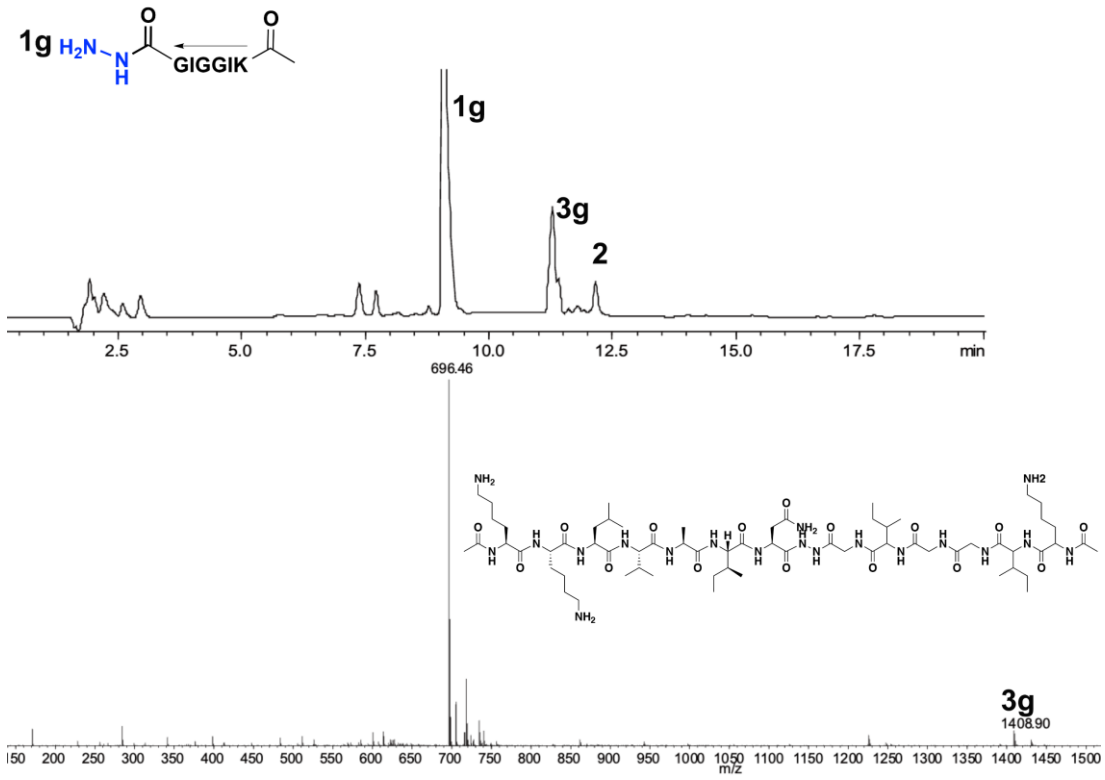


Figure 3.35. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1g** for 30 min, leading to product **3g** in >90% yield. ESI-MS data for **3g** m/z $[M+H]^+$: calcd 1408.89, obsvd 1408.90.

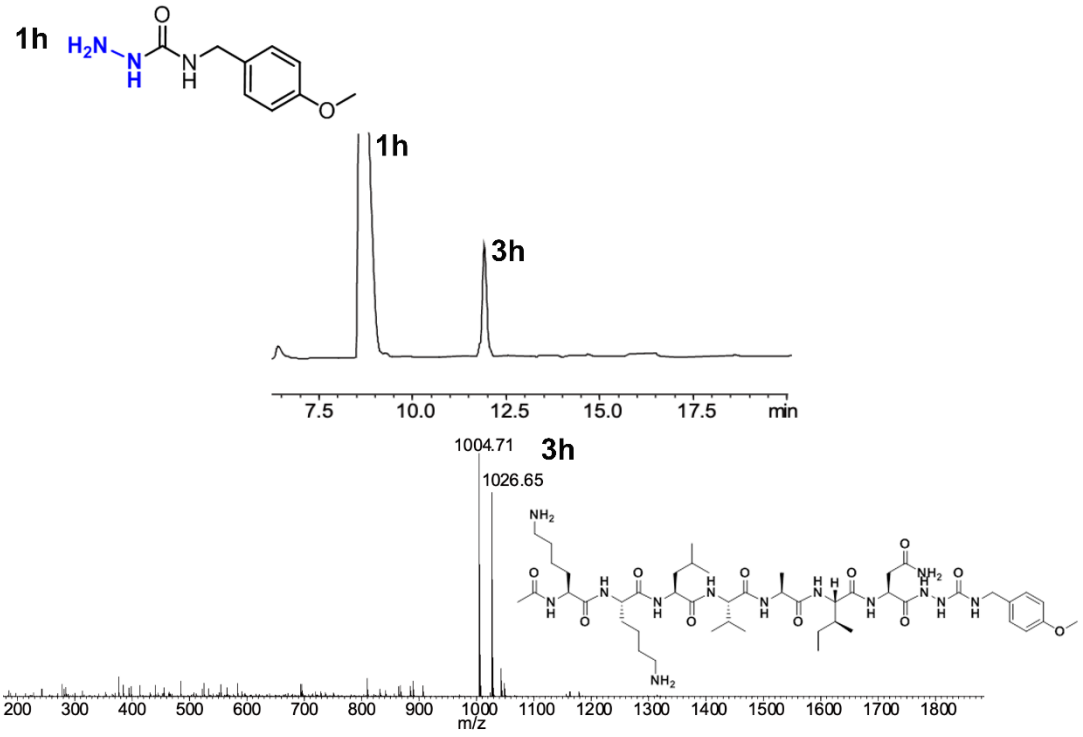


Figure 3.36. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1h** for 60 min, leading to product **3h** in 91% yield. ESI-MS data for **3h** m/z $[M+H]^+$: calcd 1004.28, obsvd 1004.71.

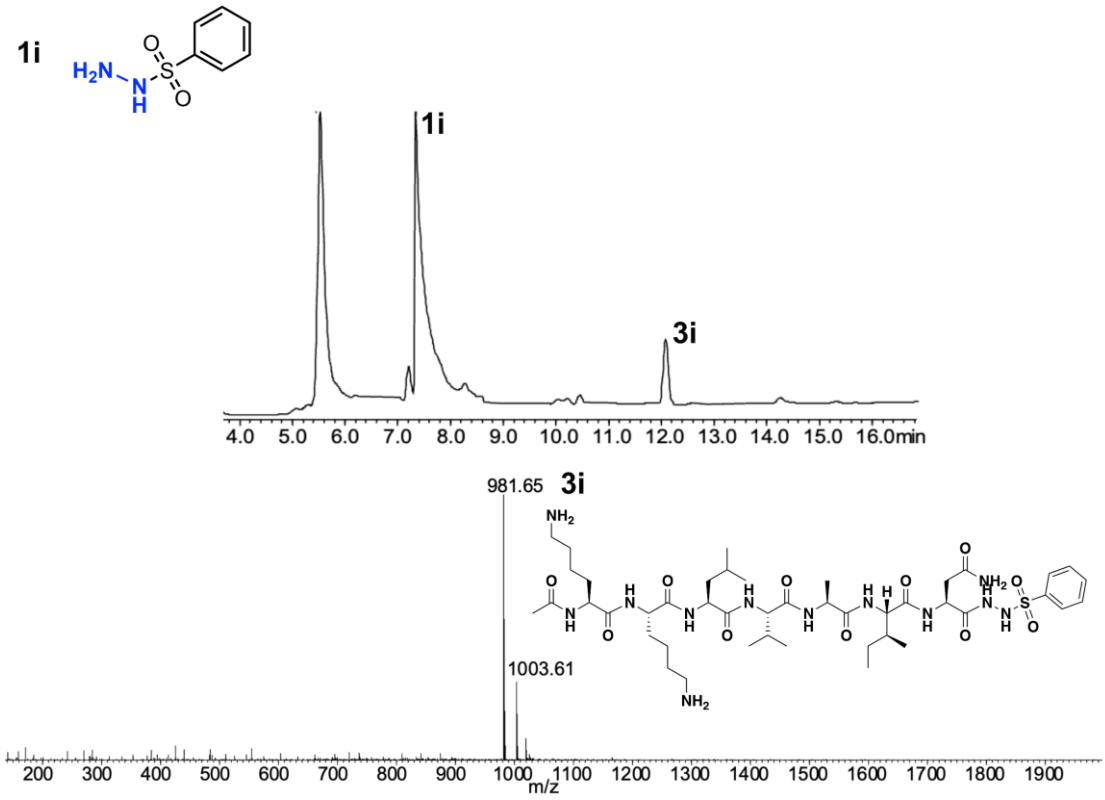


Figure 3.37. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1i** for 60 min, leading to product **3i** in 93% yield. ESI-MS data for **3i** m/z $[M+H]^+$: calcd 981.55, obsvd 981.65.

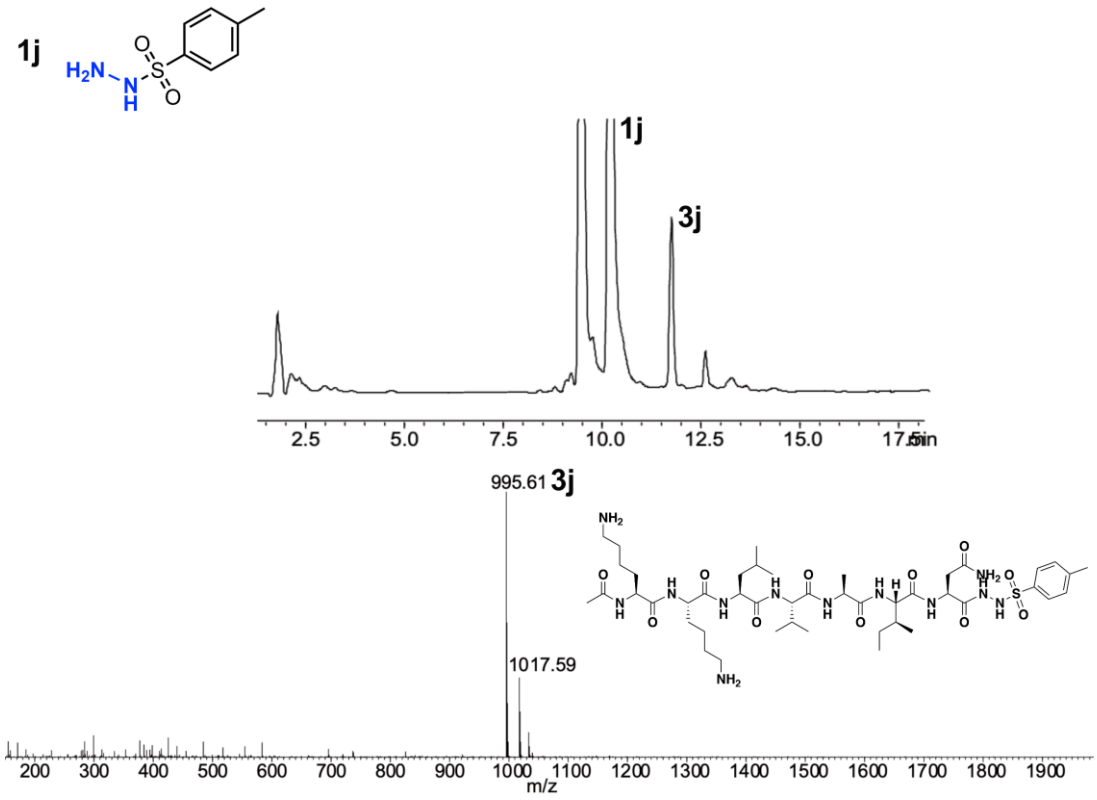


Figure 3.38. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1j** for 60 min, leading to product **3j** in 76% yield. ESI-MS data for **3j** m/z $[M+H]^+$: calcd 995.56, obsvd 995.61.

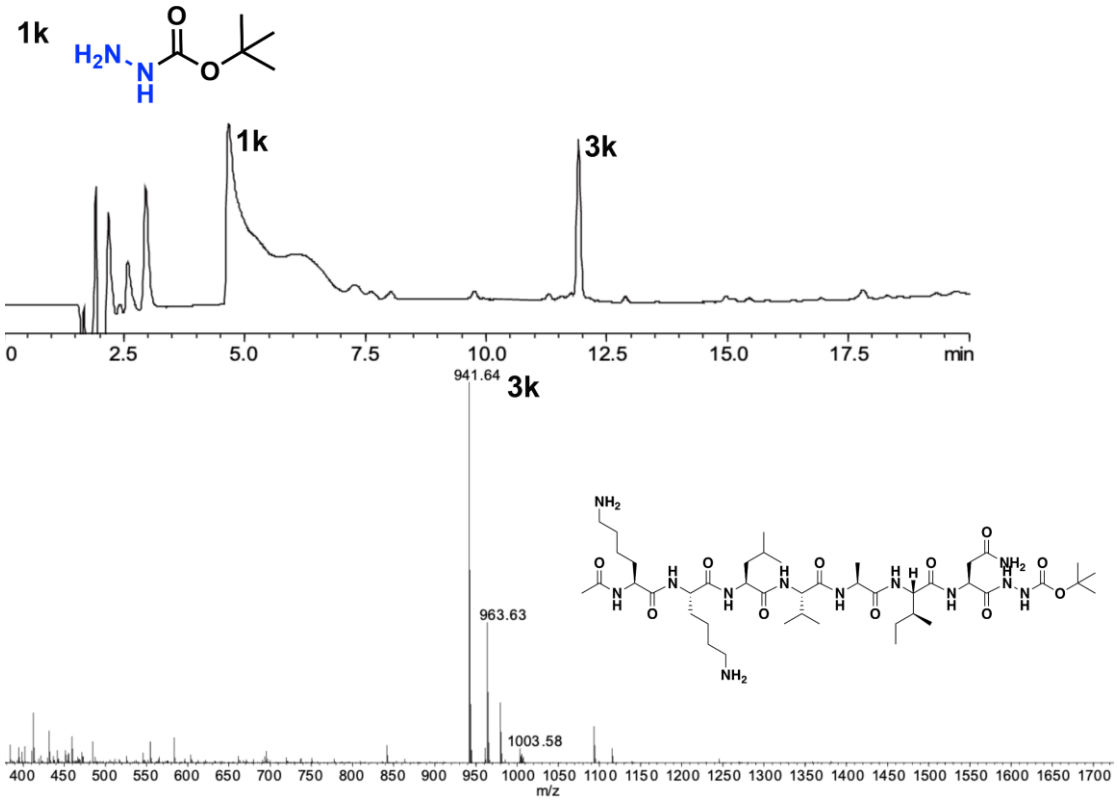


Figure 3.39. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1k** for 30 min, leading to product **3k** in 96% yield. ESI-MS data for **3k** m/z $[M+H]^+$: calcd 941.61, obsvd 941.64.

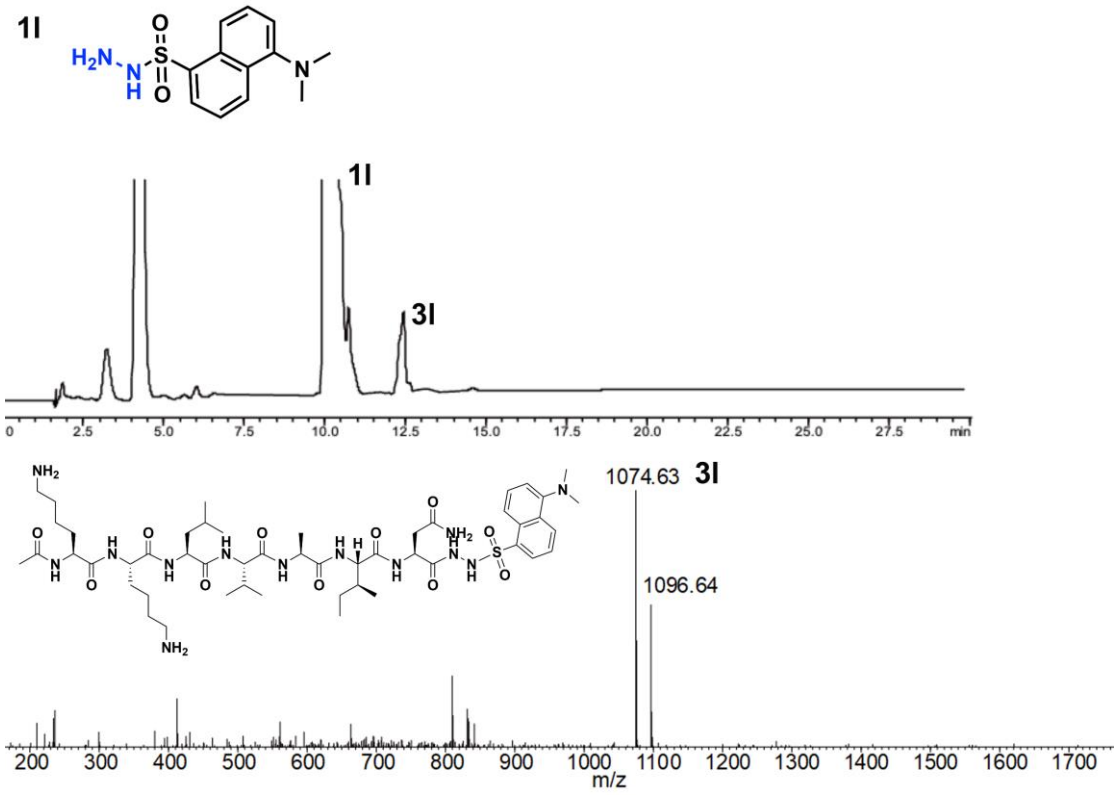


Figure 3.40. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1l** for 60 min, leading to product **3l** in 81% yield. ESI-MS data for **3l** m/z $[M+H]^+$: calcd 1074.61, obsvd 1074.63.

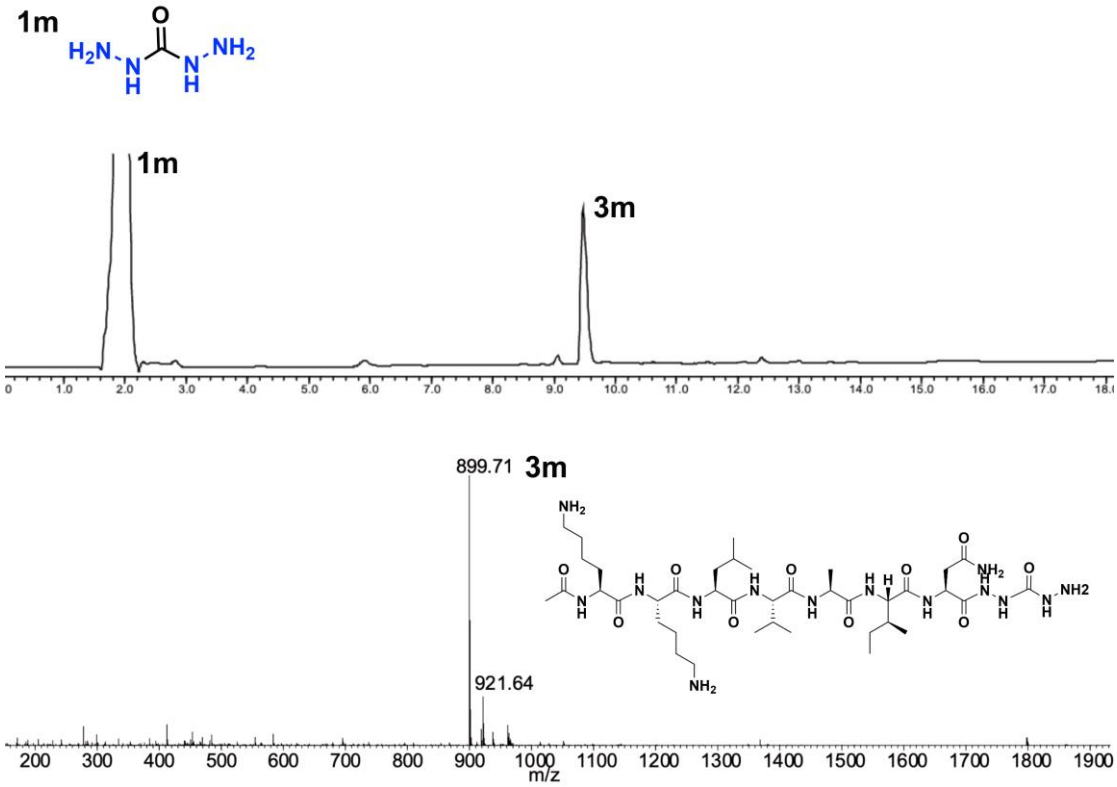


Figure 3.41. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1m** for 60 min, leading to product **3m** in 97% yield. ESI-MS data for **3m** m/z $[M+H]^+$: calcd 899.57, obsvd 899.71.

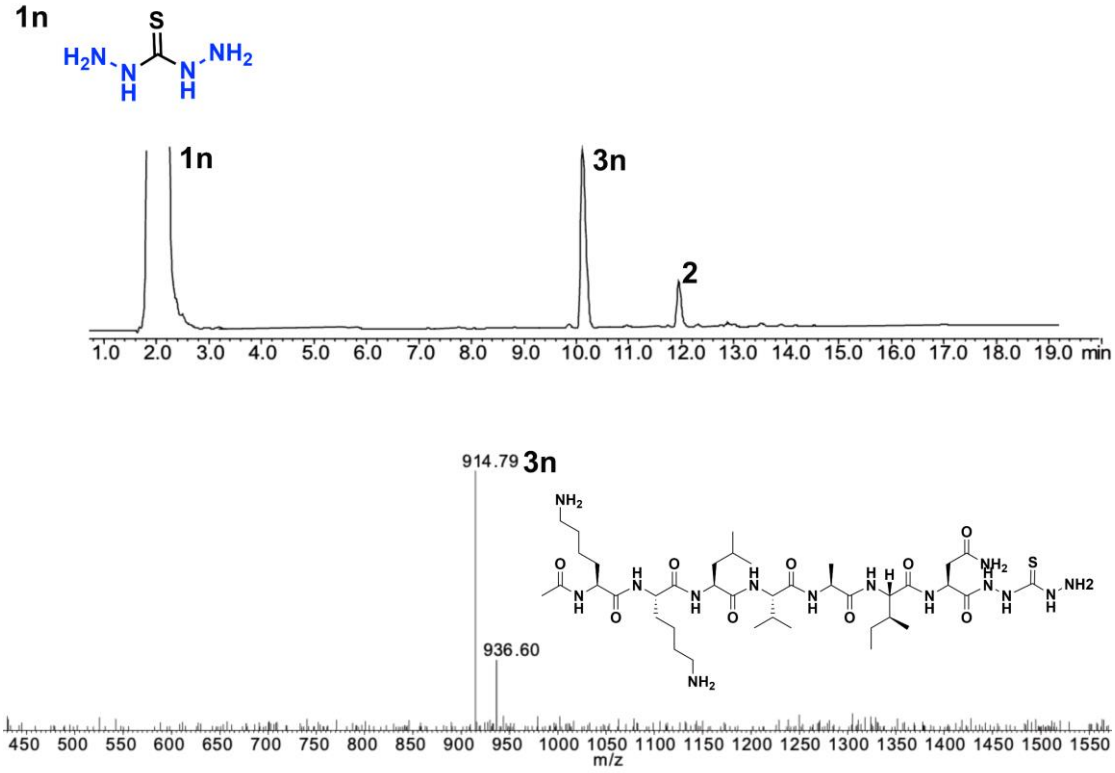


Figure 3.42. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1n** for 60 min, leading to product **3n** in 30% yield. ESI-MS data for **3n** m/z $[M+H]^+$: calcd 914.55, obsvd 914.79.

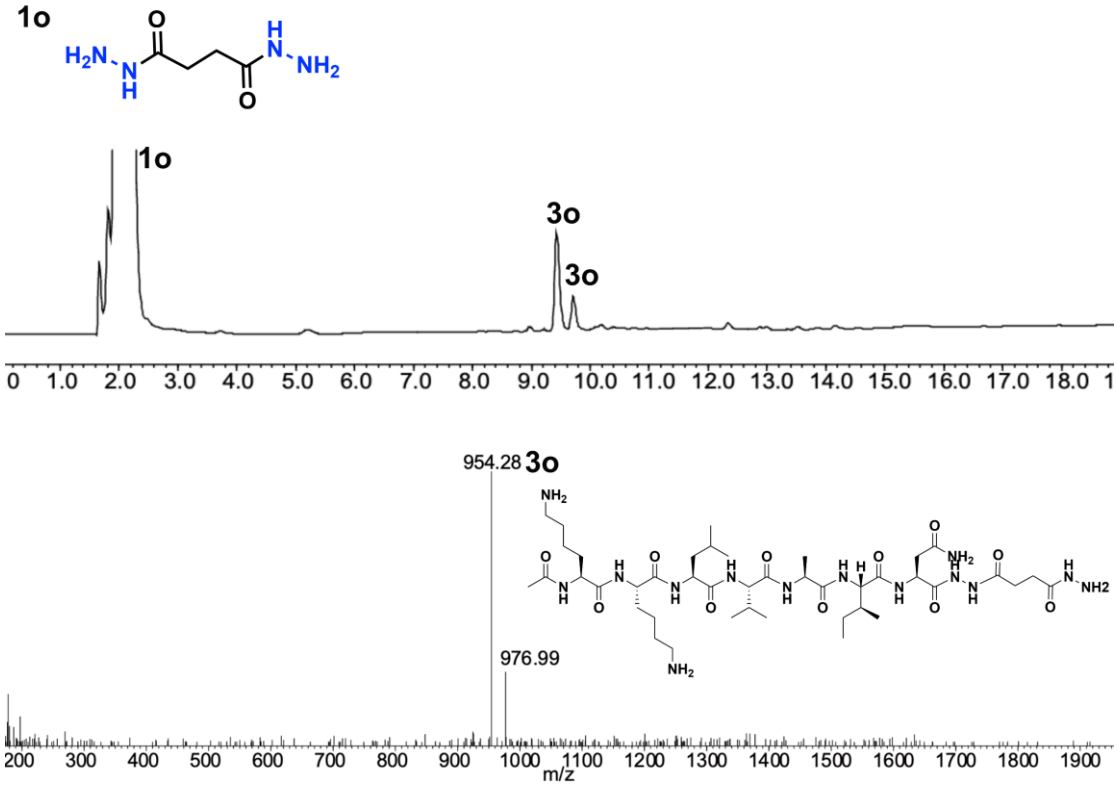


Figure 3.43. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1o** for 60 min, leading to product **3o** in 89% yield. ESI-MS data for **3o** m/z $[M+H]^+$: calcd 954.60, obsvd 954.28.

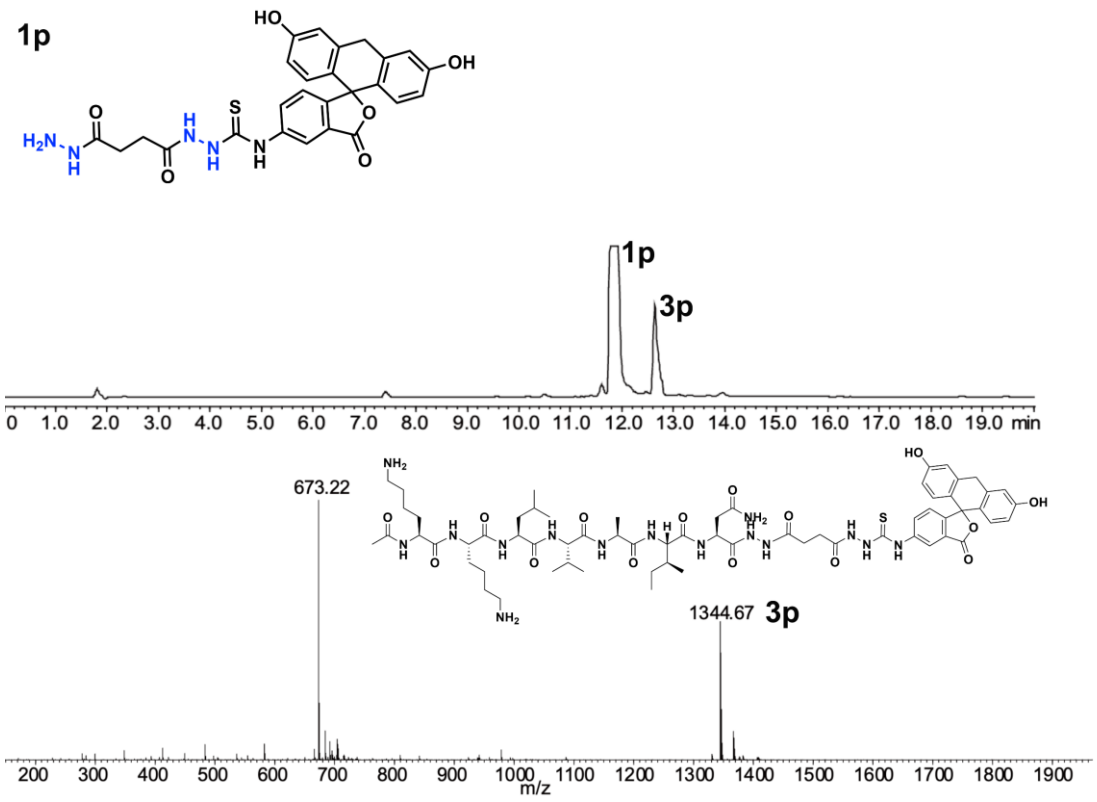


Figure 3.44. HPLC monitoring and ESI-MS characterization of peptide **2** reacting with compound **1p** for 60 min, leading to product **3p** in 83% yield. ESI-MS data for **3p** m/z $[M+H]^+$: calcd 1344.53, obsvd 1344.67.

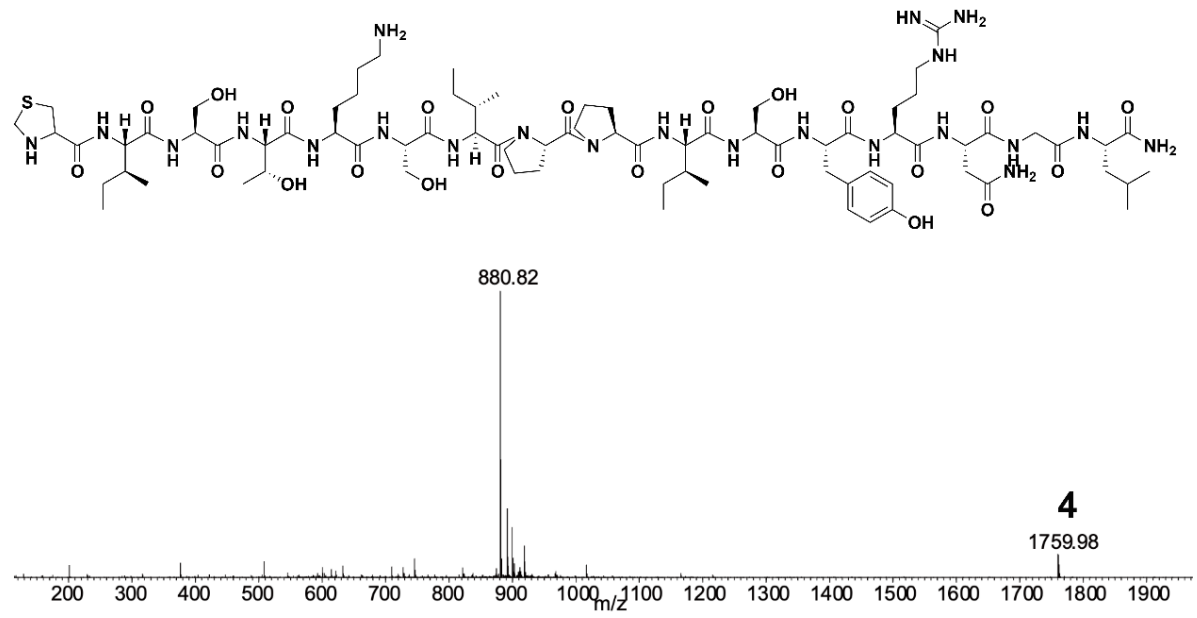


Figure 3.45. ESI-MS characterization of peptide **4,** m/z $[M+H]^+$ calcd 1759.95, obsvd 1759.95.

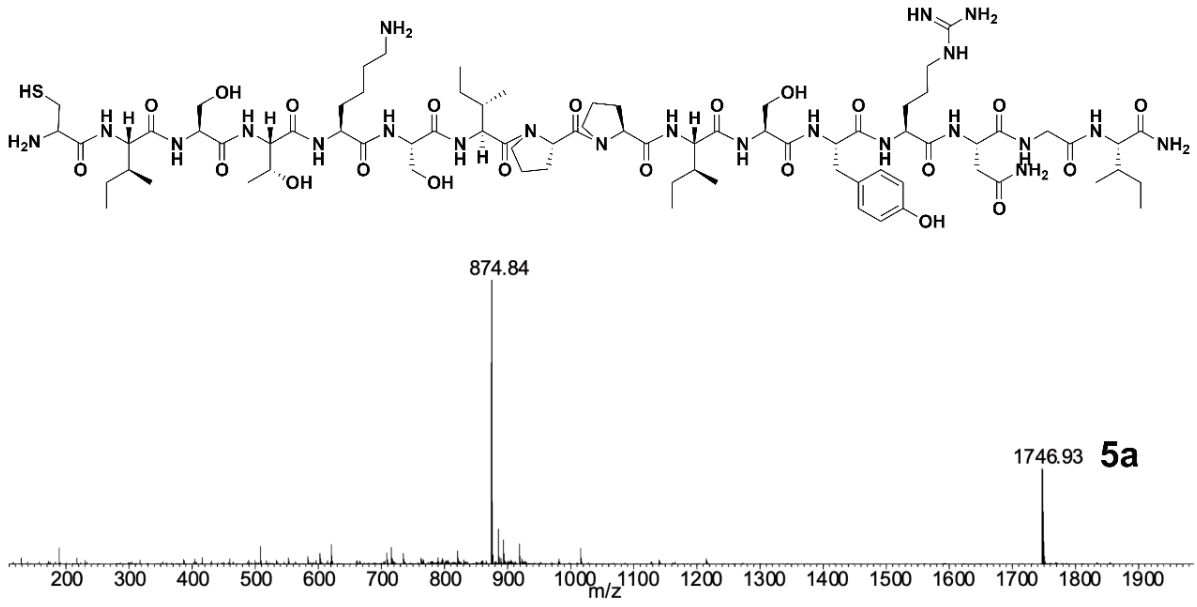


Figure 3.46. ESI-MS characterization of peptide **5a,** m/z $[M+H]^+$ calcd 1746.95, obsvd 1746.93.

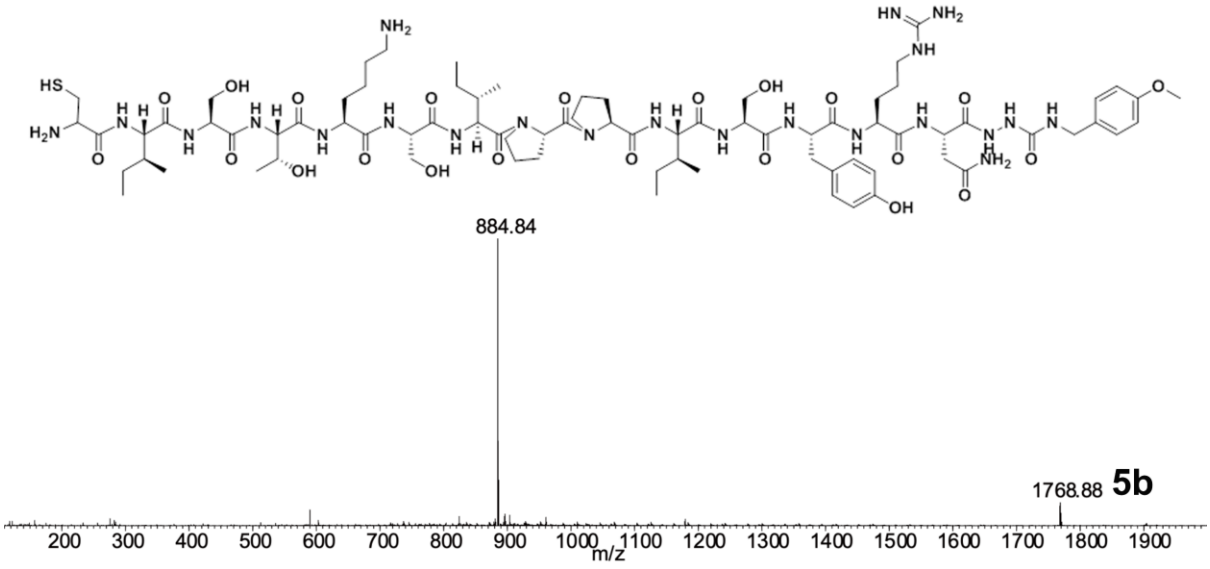


Figure 3.47 ESI-MS characterization of peptide **5b,** m/z $[M+H]^+$ calcd 1768.28, obsvd 1768.88.

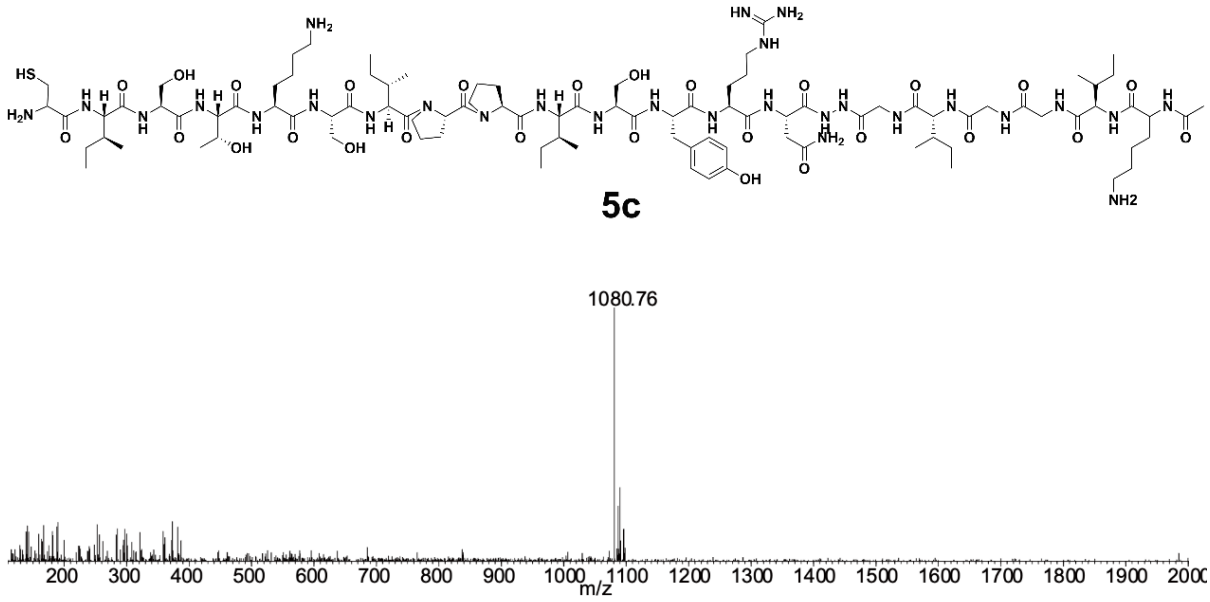


Figure 3.48. ESI-MS characterization of peptide **5c,** m/z $[M+H]^+$ calcd 2159.19, obsvd 1080.76 $[M+2H]^{2+}$.

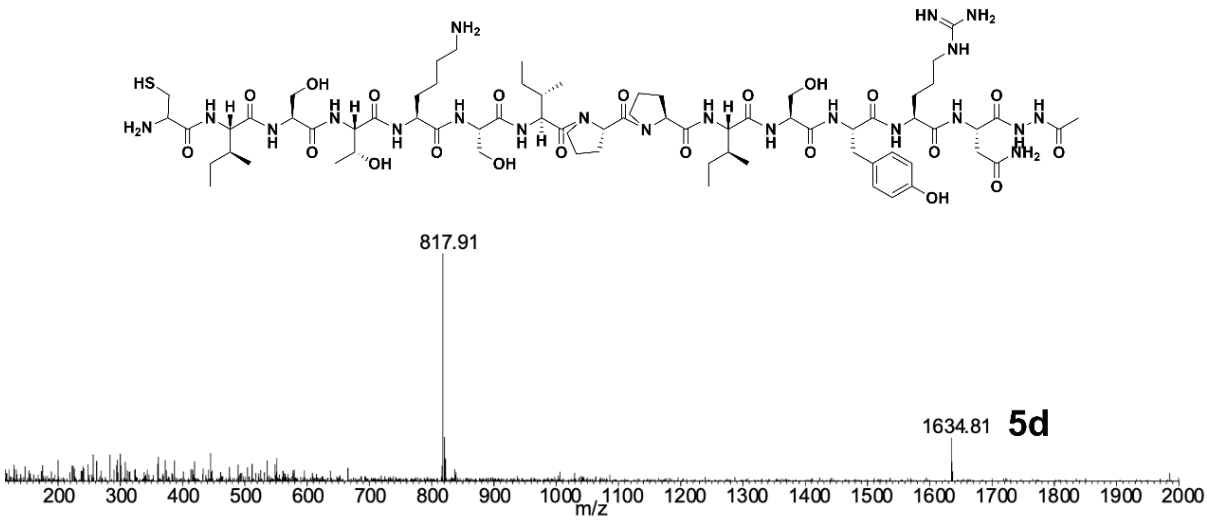


Figure 3.49. ESI-MS characterization of peptide **5d,** m/z $[M+H]^+$ calcd 1634.86, obsvd 1634.81.

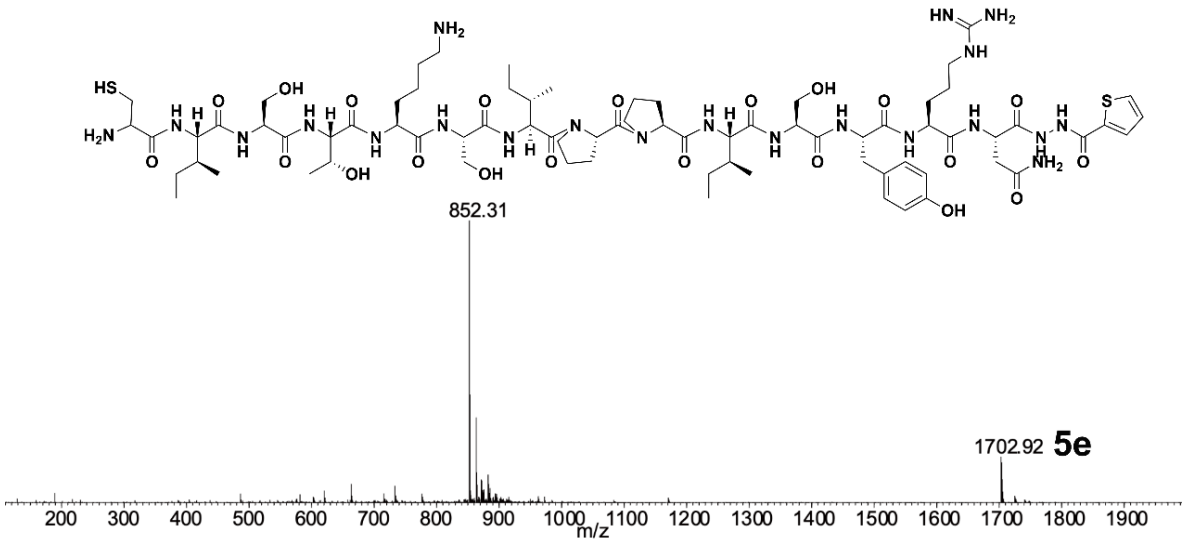


Figure 3.50. ESI-MS characterization of peptide **5e,** m/z $[M+H]^+$ calcd 1702.83, obsvd

1702.92.

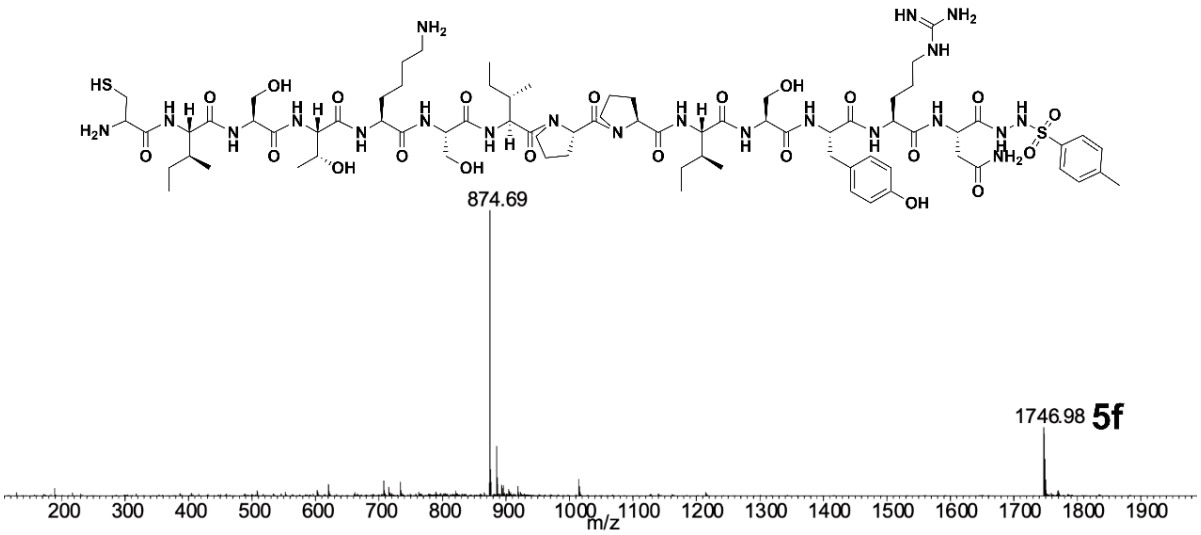


Figure 3.51. ESI-MS characterization of peptide **5f,** m/z $[M+H]^+$ calcd 1746.86, obsvd 1746.98.

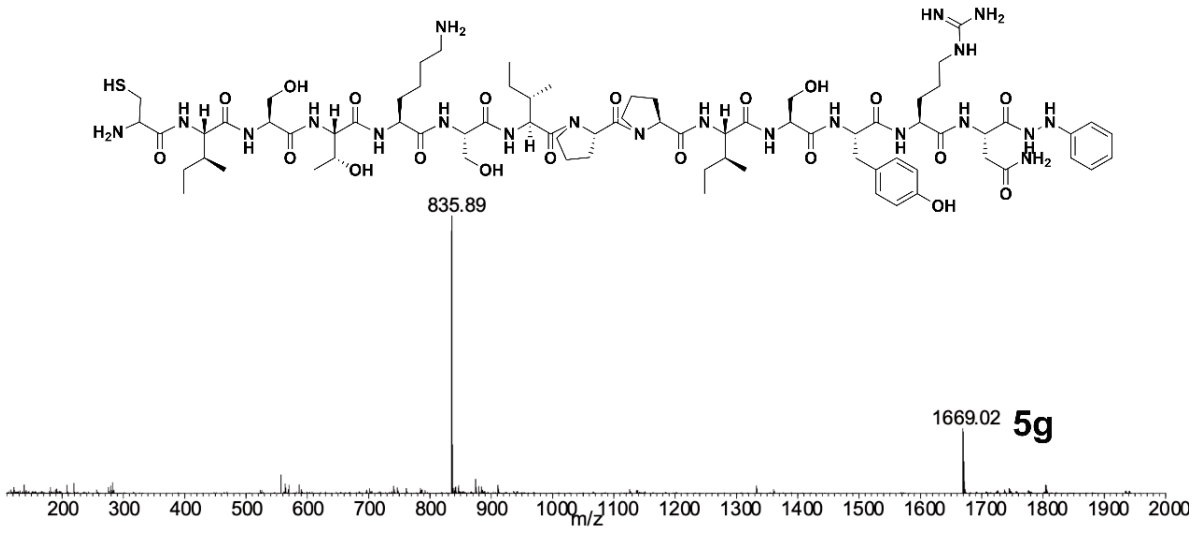


Figure 3.52. ESI-MS characterization of peptide **5g,** m/z $[M+H]^+$ calcd 1668.88, obsvd 1669.02.

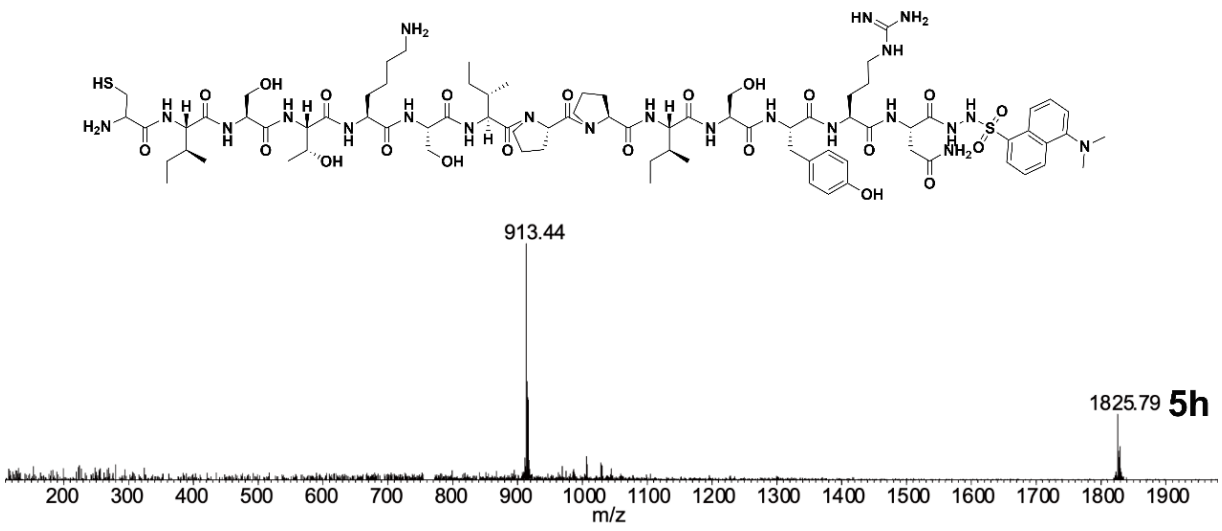


Figure 3.53. ESI-MS characterization of peptide **5h,** m/z $[M+H]^+$ calcd 1825.90, obsvd 1825.79.

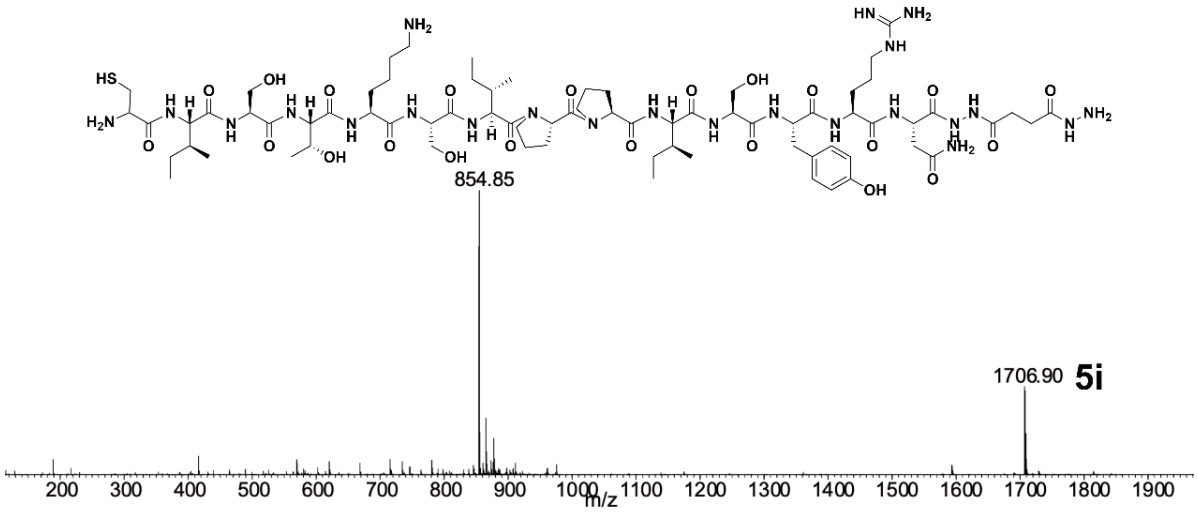


Figure 3.54. ESI-MS characterization of peptide **5i,** m/z $[M+H]^+$ calcd 1706.89, obsvd 1706.90.

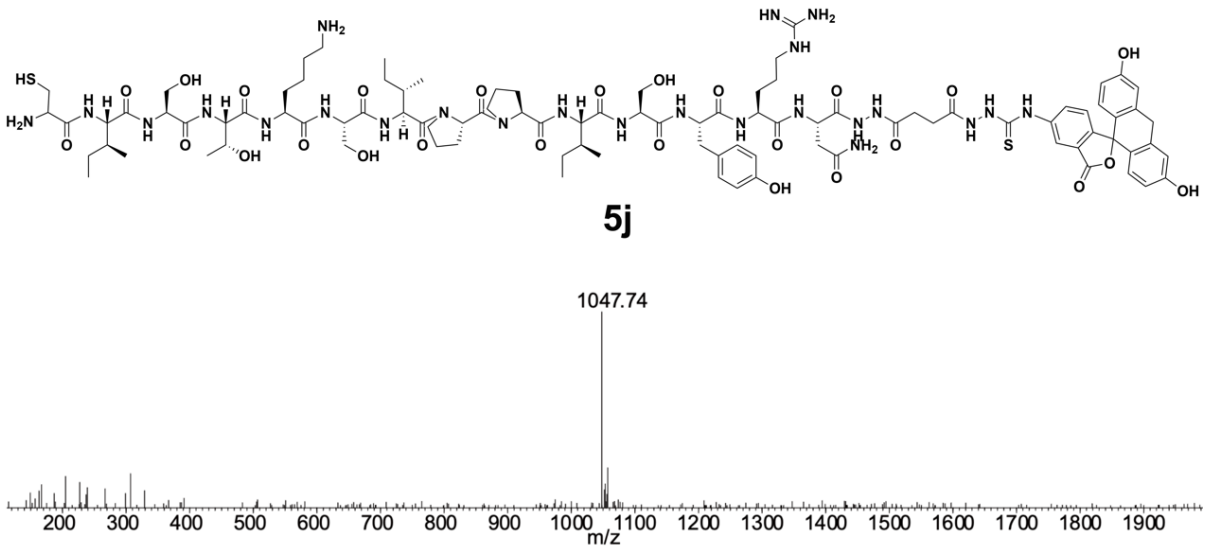


Figure 3.55. ESI-MS characterization of peptide **5j,** m/z $[M+H]^+$ calcd 2092.95, m/z $[M+2H]^{2+}$ obsvd 1047.74.

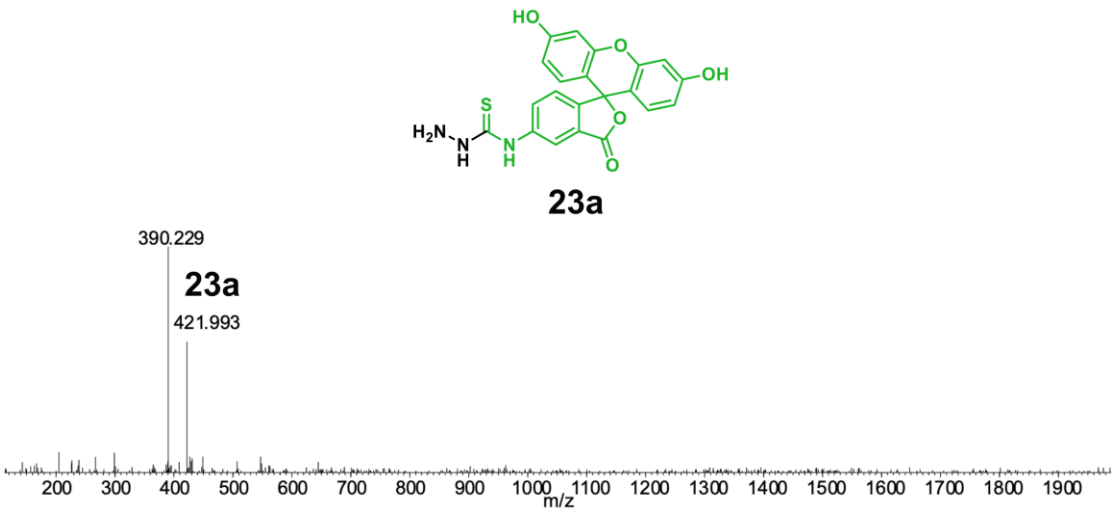


Figure 3.56. ESI-MS characterization of peptide **23a,** m/z $[M+H]^+$ calcd 421.07, obsvd 421.09.

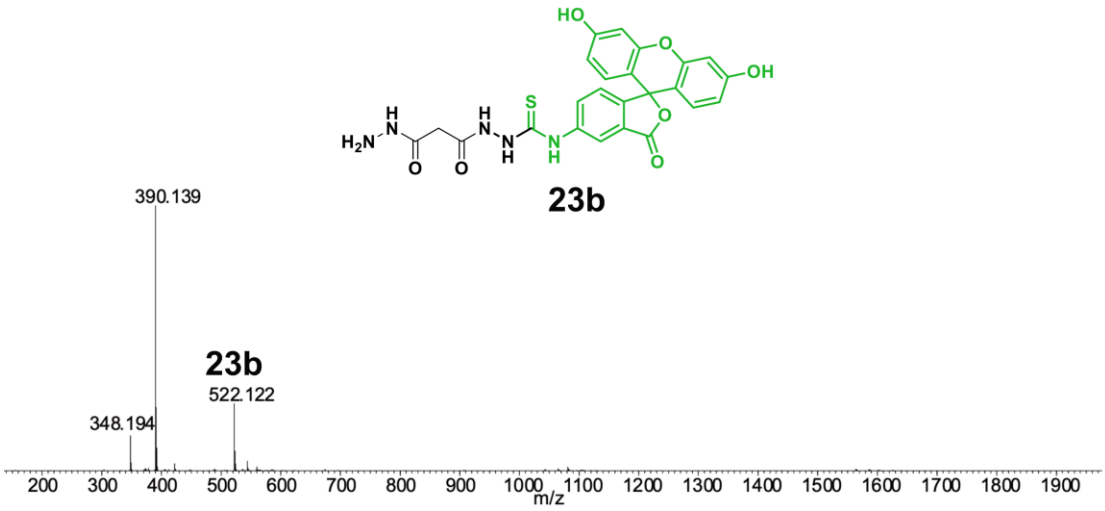


Figure 3.57. ESI -MS characterization of peptide **23b,** m/z $[M+H]^+$ calcd 521.10, obsvd 521.12.

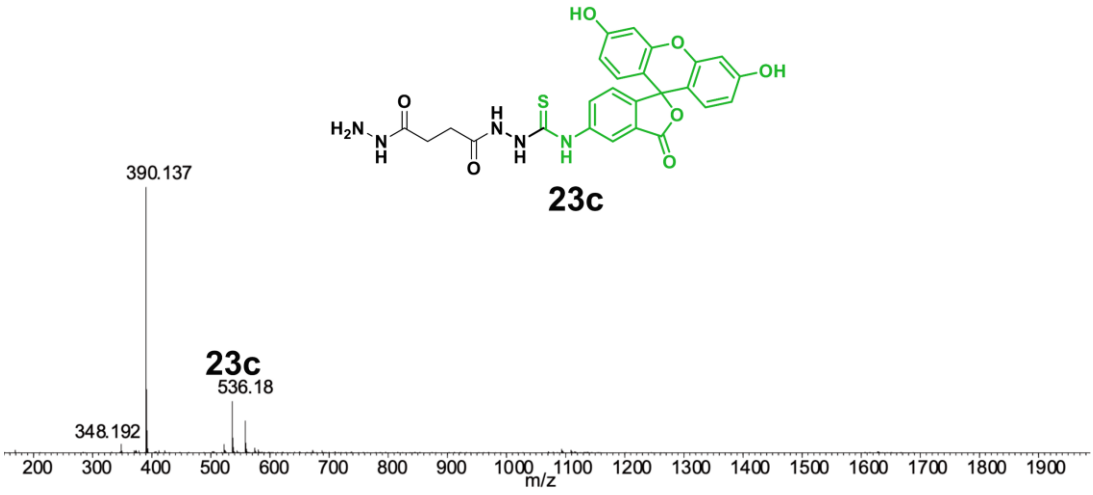


Figure 3.58. ESI-MS characterization of peptide **23c,** m/z $[M+H]^+$ calcd 536.12, obsvd 536.18.

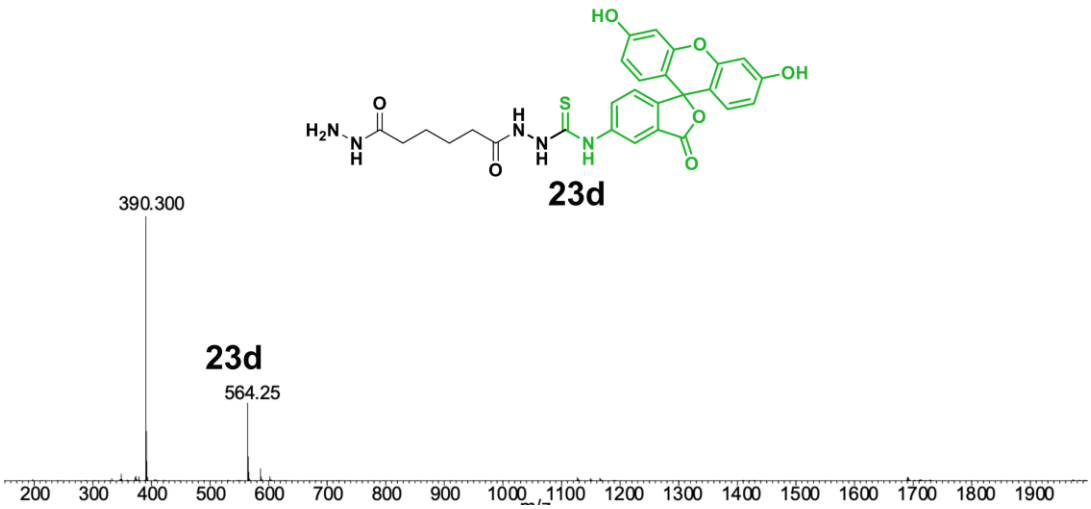


Figure 3.59. ESI-MS characterization of peptide **23d,** m/z $[M+H]^+$ calcd 564.15, obsvd 564.25.

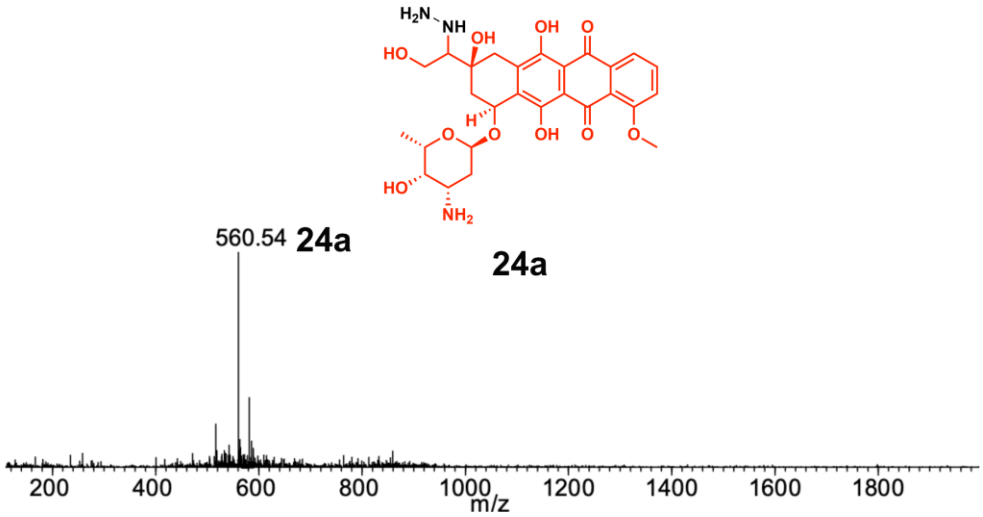


Figure 3.60. ESI-MS characterization of peptide **24a,** m/z $[M+H]^+$ calcd 560.22, obsvd 560.54.

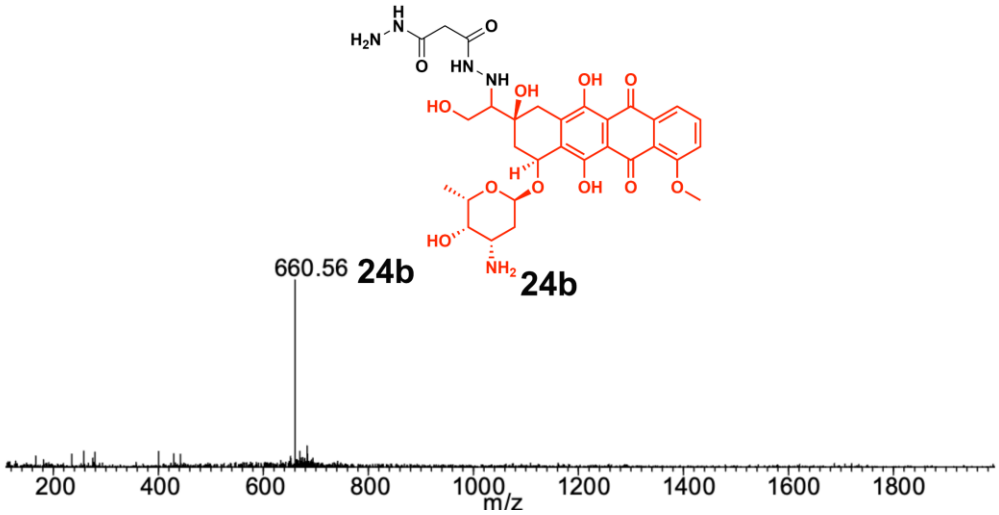


Figure 3.61. ESI-MS characterization of peptide **24b,** m/z $[M+H]^+$ calcd 660.24, obsvd 660.56.

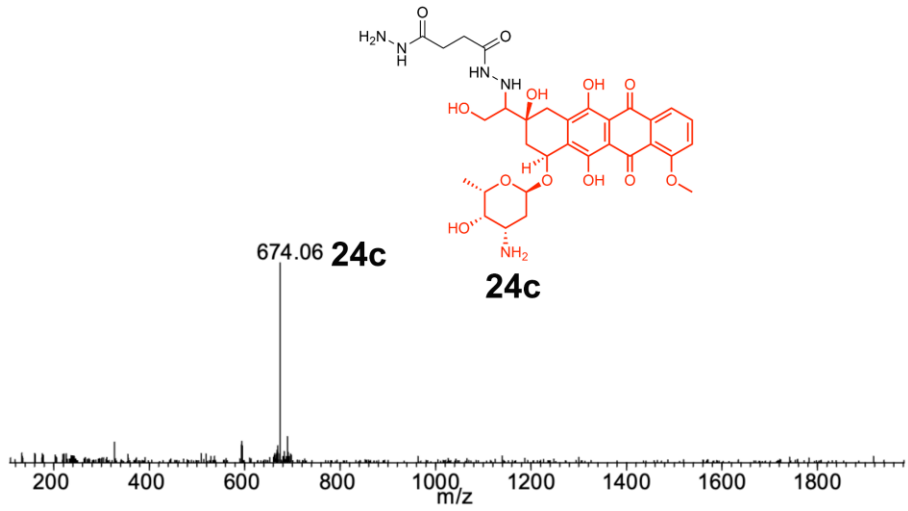


Figure 3.62. ESI-MS characterization of peptide **24c,** m/z $[M+H]^+$ calcd 674.26, obsvd 674.06.

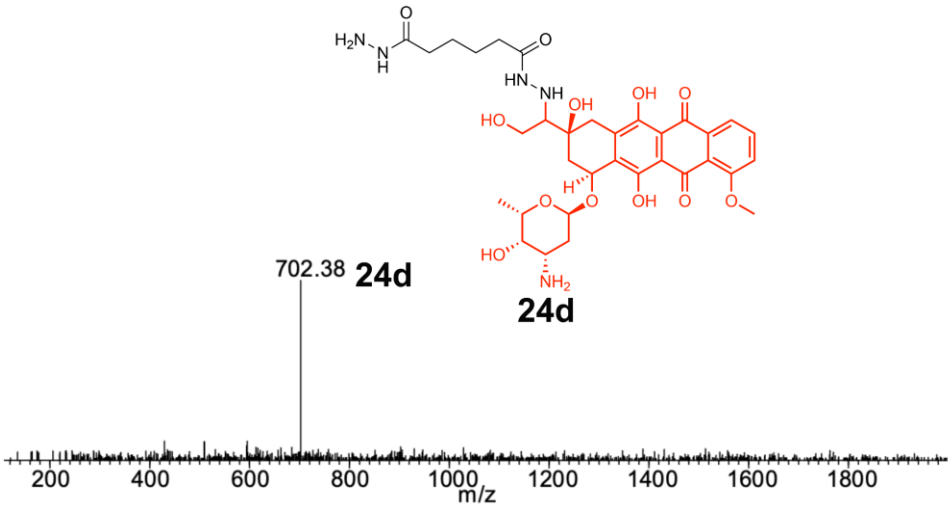


Figure 3.63. ESI-MS characterization of peptide **24d,** m/z $[M+H]^+$ calcd 702.29, obsvd 702.38.

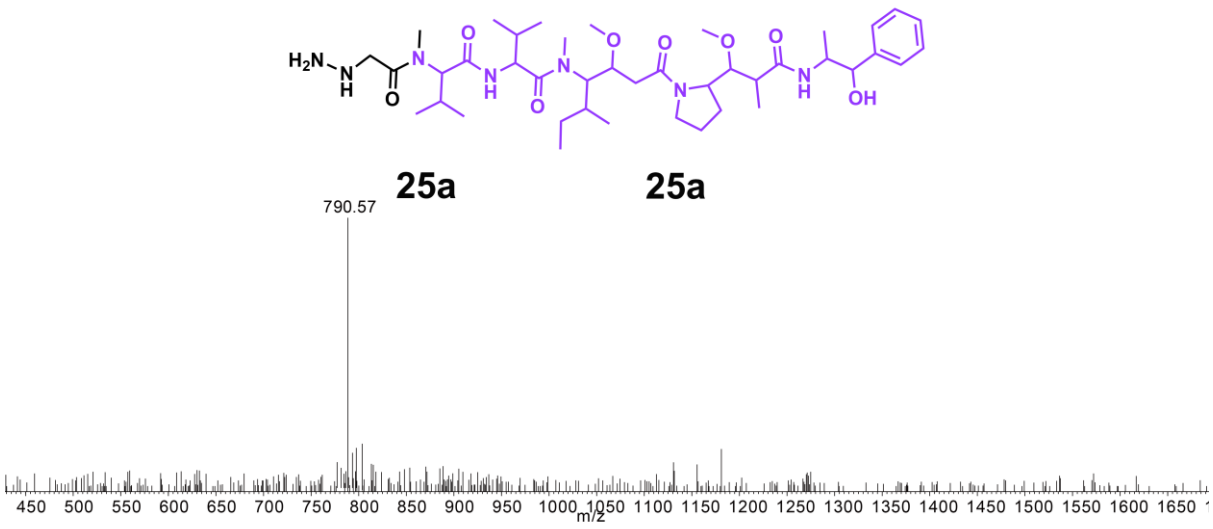


Figure 3.64. ESI-MS characterization of peptide **25a,** m/z $[M+H]^+$ calcd 790.54, obsvd 790.57.

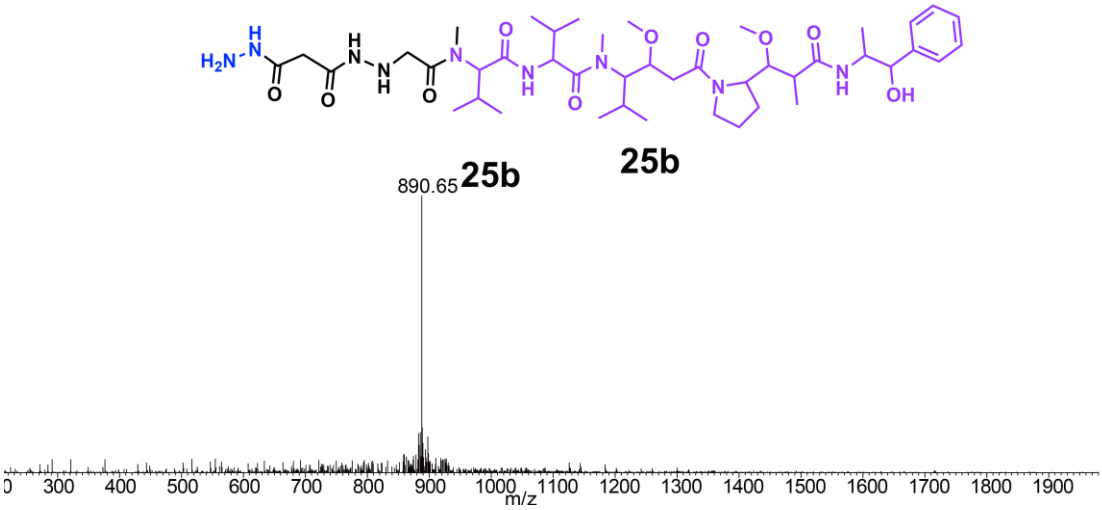


Figure 3.65. ESI-MS characterization of peptide **25b,** m/z $[M+H]^+$ calcd 890.54, obsvd 890.65.

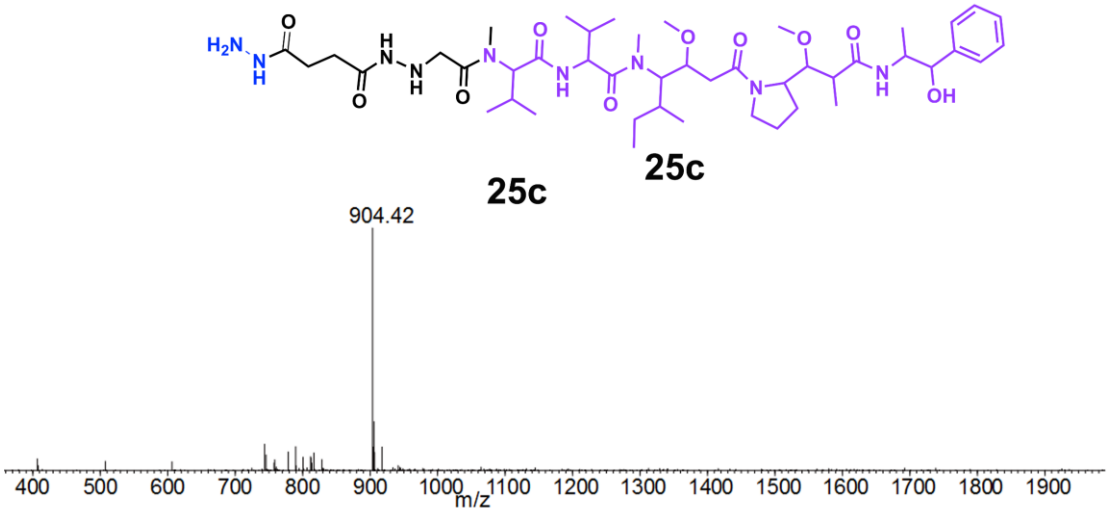


Figure 3.66. ESI-MS characterization of peptide **25c,** m/z $[M+H]^+$ calcd 904.54, obsvd 904.41.

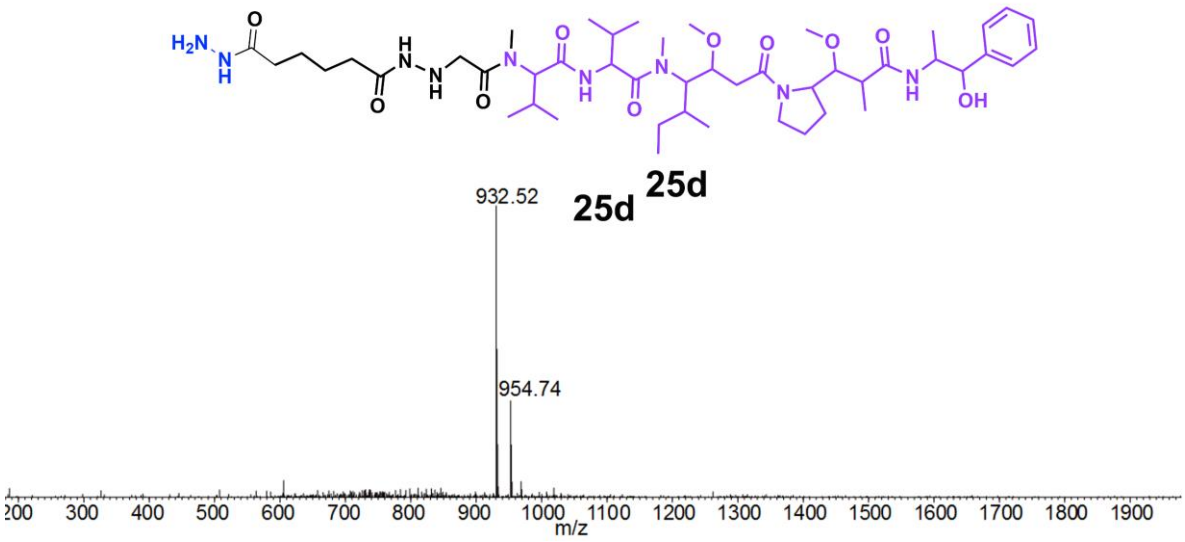


Figure 3.67. ESI-MS characterization of peptide **25d,** m/z $[M+H]^+$ calcd 932.61, obsvd 932.52; m/z $[M+Na]^+$ calcd 954.62, obsvd 954.74.

Appendix C. Supplementary data for Chapter 4.

Table 4.1. Summary of peptides used in this work.

Peptide	Peptide sequence	Starting material		Peptide-$NHNH_2$		Peptide-MesNa	
		Cal. MS $[M+H]^+$	**Obs. MS** $[M+H]^+$	**Cal. MS** $[M+H]^+$	**Obs. MS** $[M+H]^+$	**Cal. MS** $[M+H]^+$	**Obs. MS** $[M+H]^+$
1a	Ac-ASLKAGCNGL	975.15	974.92	819.95	819.54	-	-
1b	Ac-KKLAVINGF	1031.28	1031.13	842.07	841.70	-	-
1c	Ac-KKLAVINHV	1063.33	1063.04	842.07	841.84	-	-
1d	Ac-ASLKAGCDGL	976.13	975.84	819.93	819.53	-	-
1e	Ac-ASYKAGCNGL	1025.16	1024.97	869.96	869.50	-	-
10a	Ac-ASFKAGCNGL	1008.49	1008.63	853.39	853.75	746.27	746.35
10b	Ac-ASFKASCNGL	1038.50	1038.86	883.40	883.84	775.88	776.20
10c	Ac-ASFKAACNGL	1022.50	1022.84	867.41	867.80	760.29	760.47
10d	Ac-ASFKALCNGL	1064.55	1064.93	909.45	909.94	802.33	802.65
10e	Ac-ASFKAFCNGL	1098.53	1099.02	943.44	943.84	836.32	836.39
10f	Ac-ASFKAPCNGL	1048.52	1048.94	893.42	893.66	786.30	786.49
10g	Ac-ASFKAICNGL	1064.55	1064.95	909.45	909.88	802.33	802.47
10h	Ac-ASFKAVCNGL	1050.53	1050.85	895.44	895.78	788.32	788.41
10i	Ac-ASFKAKCNGL	1079.56	1079.98	924.46	924.84	817.34	817.47
10j	Ac-ASFKAQCNGL	1079.52	1079.96	924.43	924.92	817.31	817.45

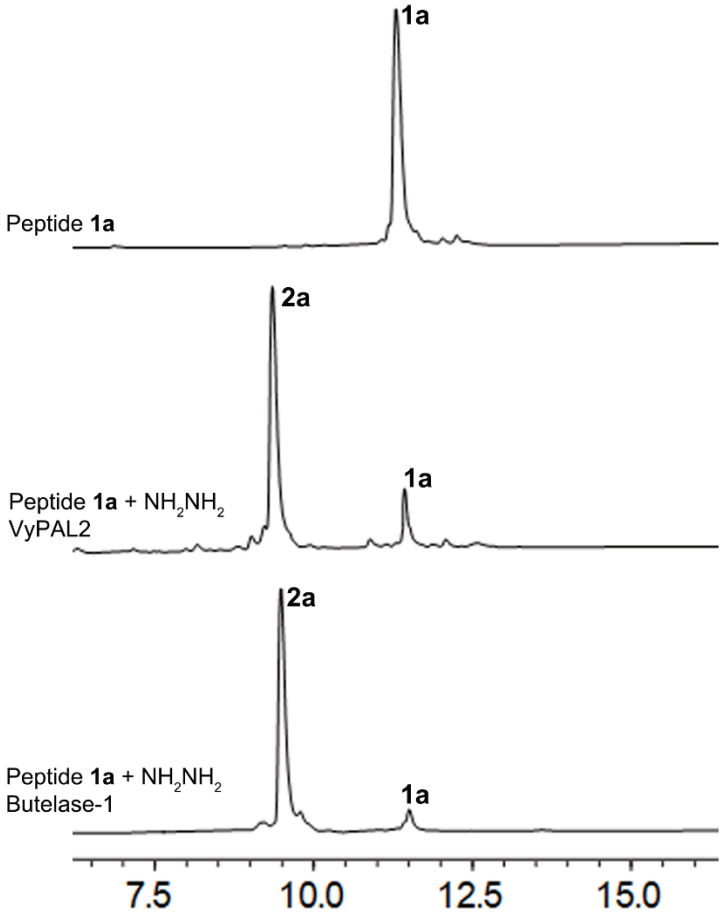


Figure 4.11. HPLC monitoring of VyPAL2 and butelase 1 mediated acyl-peptide 1a hydrazinolysis for 15 min, leading to product 2a in 81% and 86% yield, respectively.

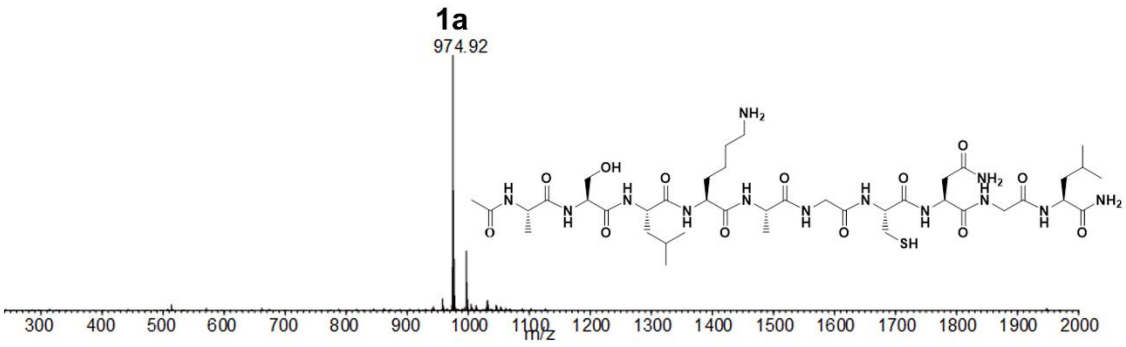


Figure 4.12. ESI-MS characterization of peptide **1a**. ESI-MS data for **1a** m/z $[M+H]^+$: calcd 975.15, obsvd 974.92.

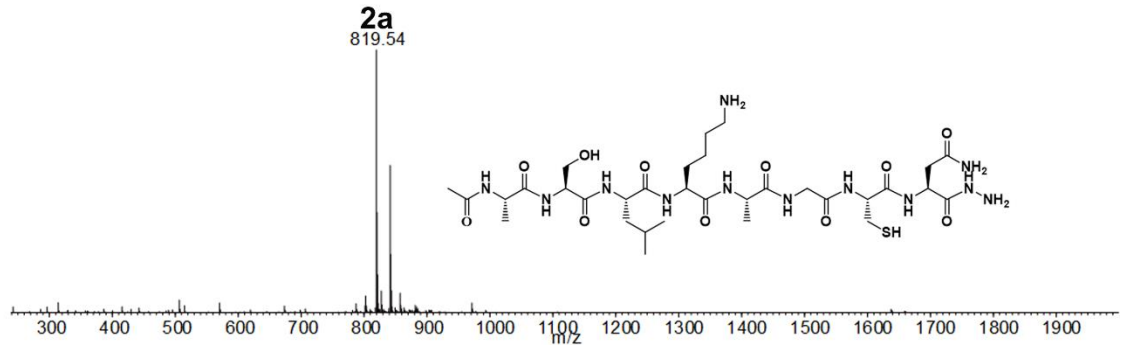


Figure 4.13. ESI-MS characterization of peptide **2a**. ESI-MS data for **2a** m/z $[M+H]^+$: calcd 819.95, obsvd 819.54.

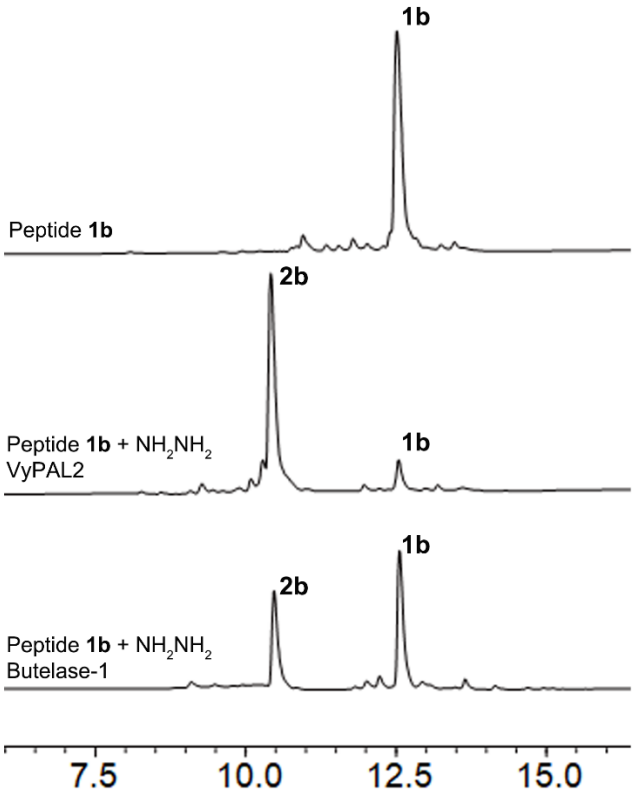


Figure 4.14. HPLC monitoring of VyPAL2 and butelase 1 mediated acyl-peptide **1b** hydrazinolysis for 15 min, leading to product **2b** in 90% and 38% yield, respectively.

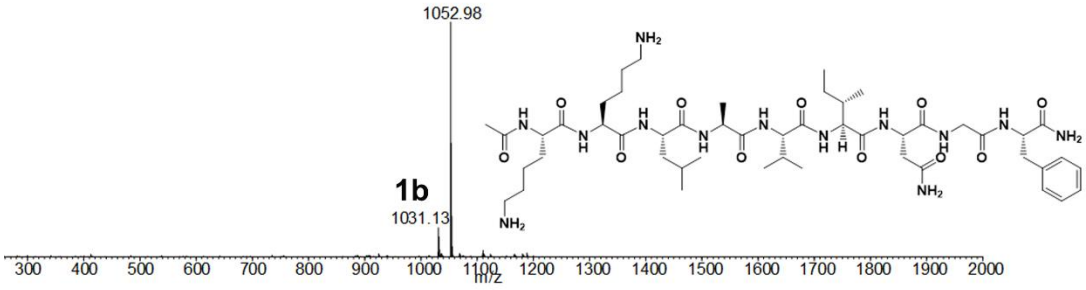


Figure 4.15. ESI-MS characterization of peptide **1b**. ESI-MS data for **1b** m/z [M+H]+: calcd 1031.28, obsvd 1031.13.

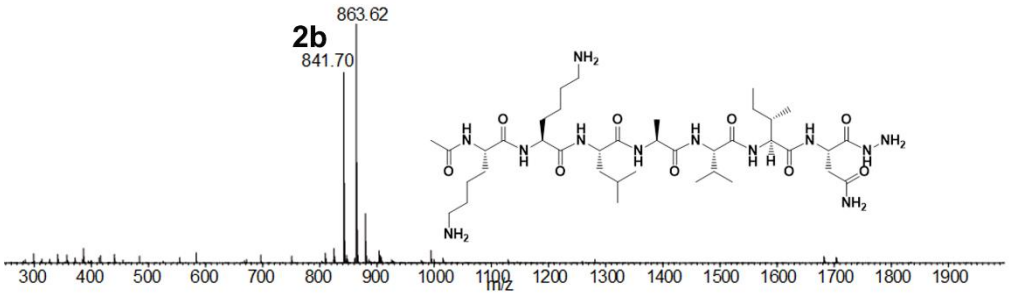


Figure 4.16. ESI-MS characterization of peptide **2b**. ESI-MS data for **2b** m/z $[M+H]^+$: calcd 842.07, obsvd 841.70.

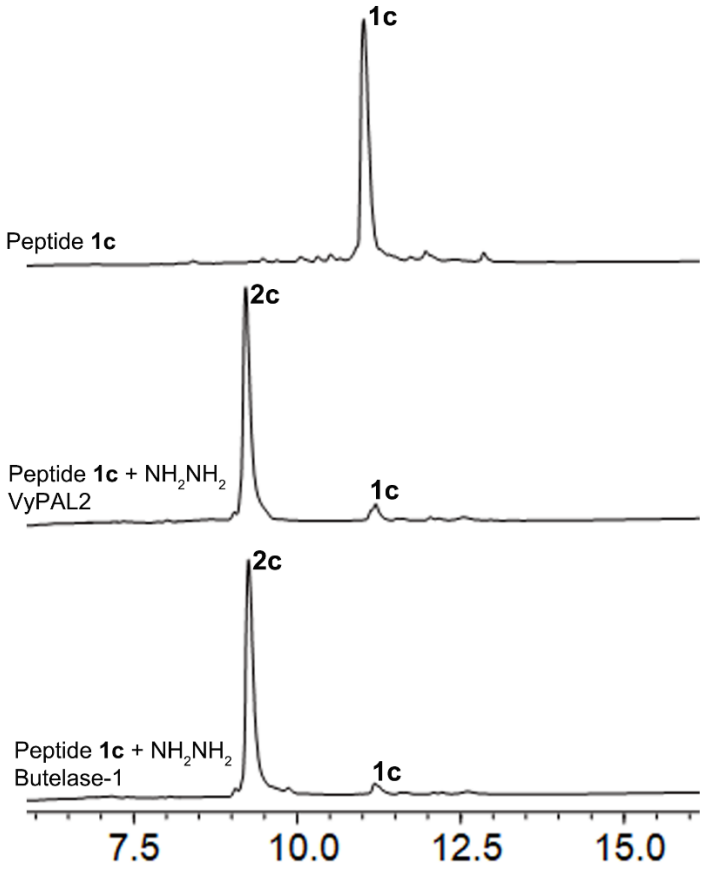


Figure 4.17. HPLC monitoring of VyPAL2 and butelase 1 mediated acyl-peptide **1c** hydrazinolysis for 60 min, leading to product **2c** in 85% and 88% yield, respectively.

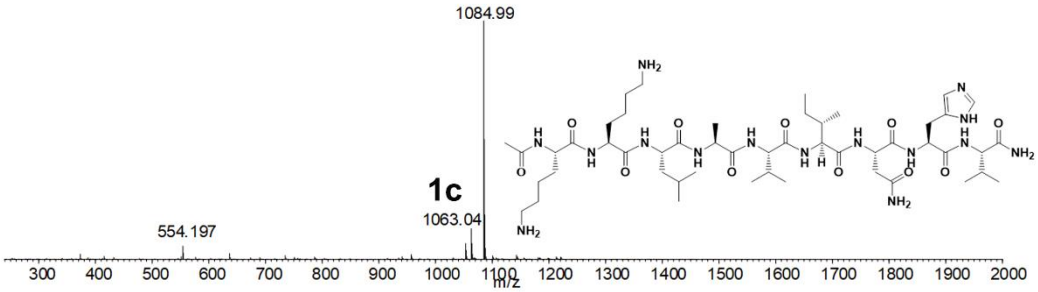


Figure 4.18. ESI-MS characterization of peptide **1c**. ESI-MS data for **1c** m/z $[M+H]^+$: calcd 1063.33, obsvd 1063.04.

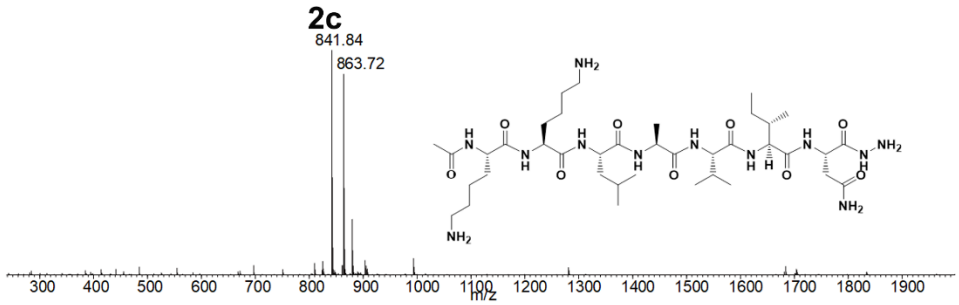


Figure 4.19. ESI-MS characterization of peptide **2c**. ESI-MS data for **2c** m/z $[M+H]^+$: calcd 842.07, obsvd 841.84.

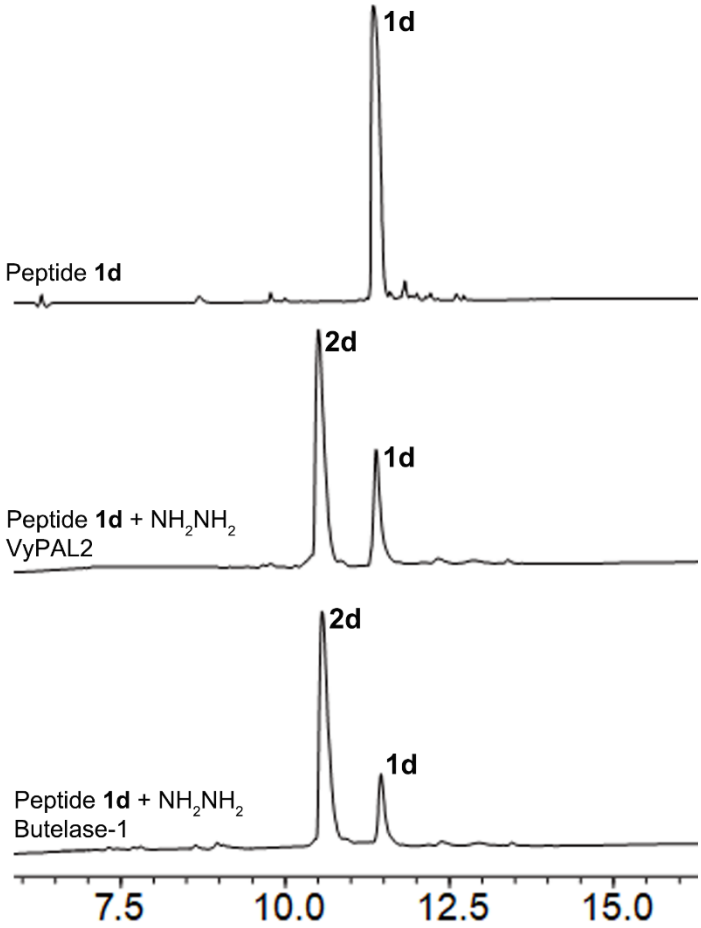


Figure 4.20. HPLC monitoring of VyPAL2 and butelase 1 mediated acyl-peptide **1d** hydrazinolysis for 60 min, leading to product **2d** in 70% and 82% yield, respectively.

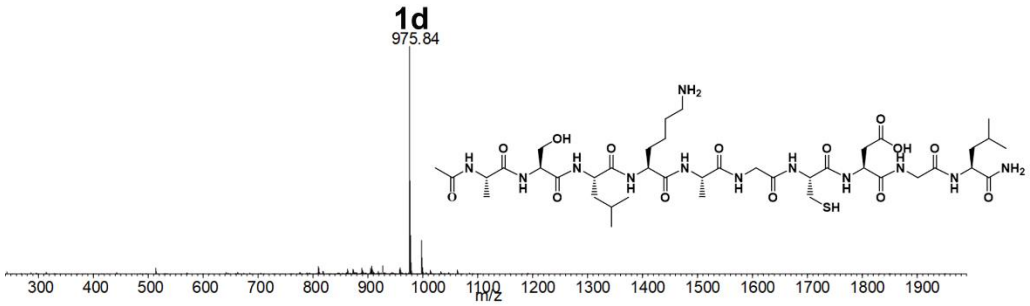


Figure 4.21. ESI-MS characterization of peptide **1d**. ESI-MS data for **1d** m/z $[M+H]^+$: calcd 976.13, obsvd 975.84.

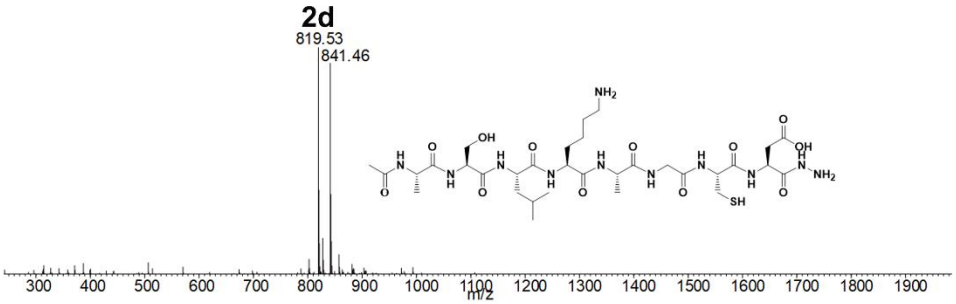


Figure 4.22. ESI-MS characterization of peptide **2d**. ESI-MS data for **2d** m/z $[M+H]^+$: calcd 819.93, obsvd 819.53.

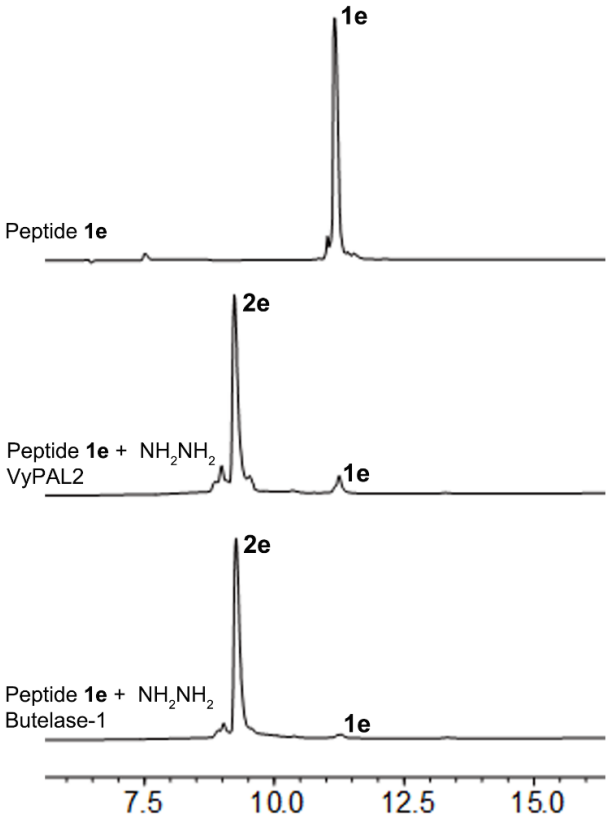


Figure 4.23. HPLC monitoring of VyPAL2 and butelase 1 mediated acyl-peptide **1e** hydrazinolysis for 15 min, leading to product **2e** in 89% and 95% yield, respectively.

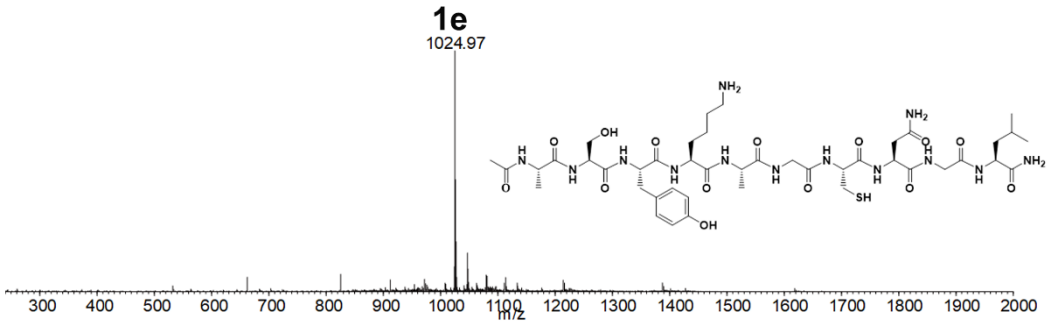


Figure 4.24. ESI-MS characterization of peptide **1e**. ESI-MS data for **1e** m/z $[M+H]^+$: calcd 1025.16, obsvd 1024.97.

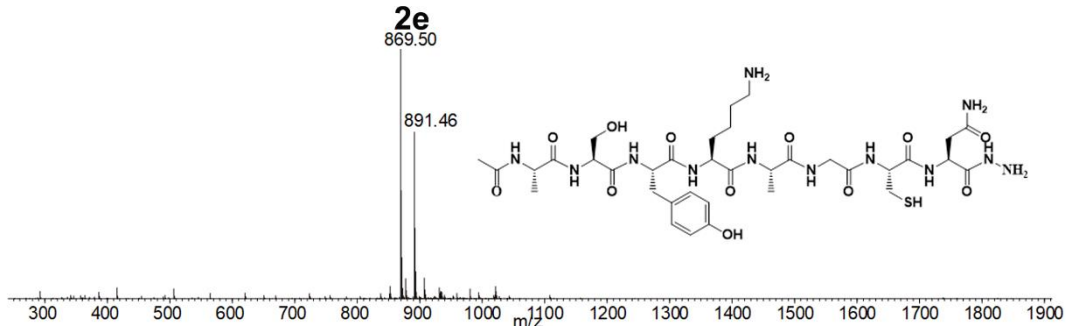


Figure 4.25. ESI-MS characterization of peptide **2e**. ESI-MS data for **2e** m/z $[M+H]^+$: calcd 869.96, obsvd 869.50.

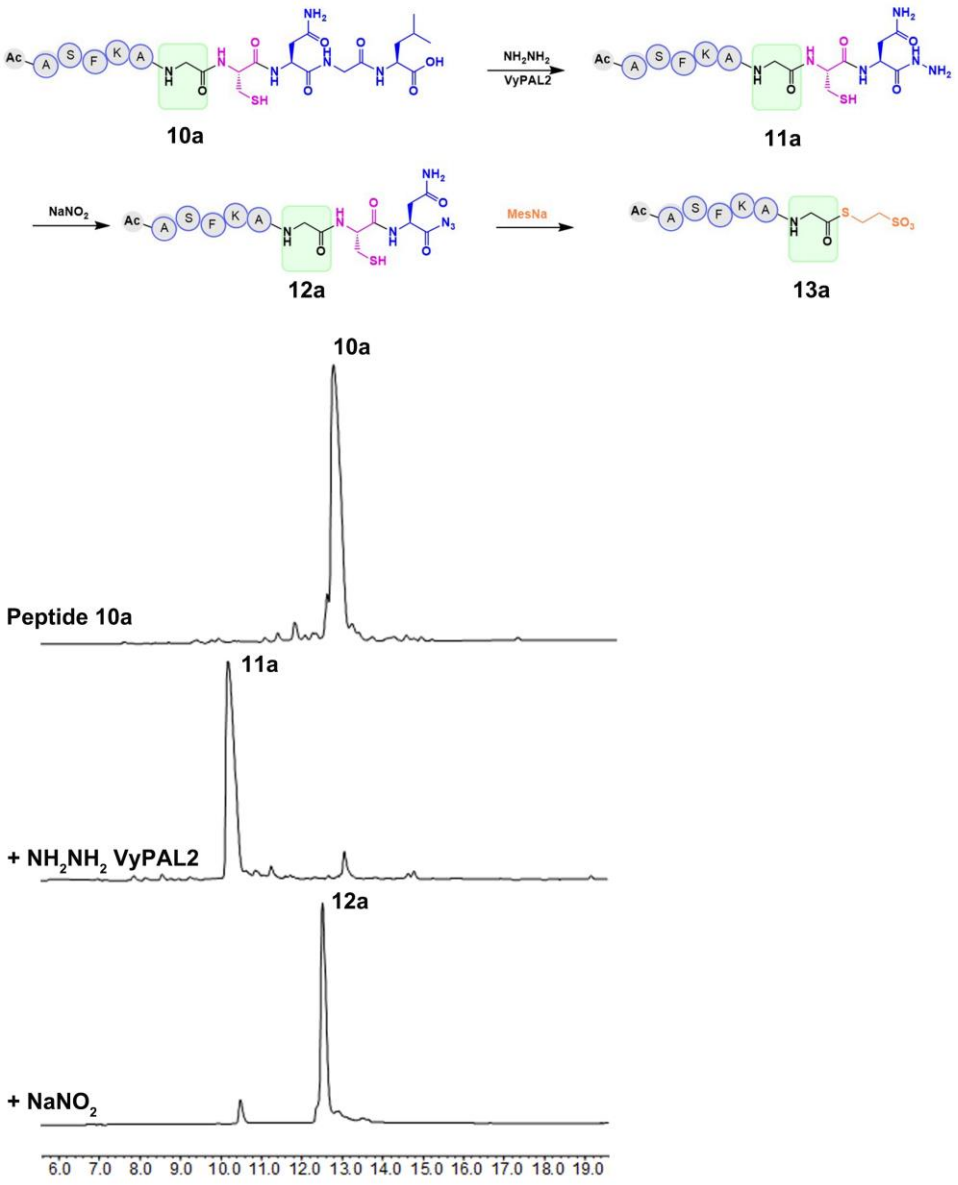


Figure 4.26. HPLC analysis of peptide **10a** (X = G) converting into **12a**.

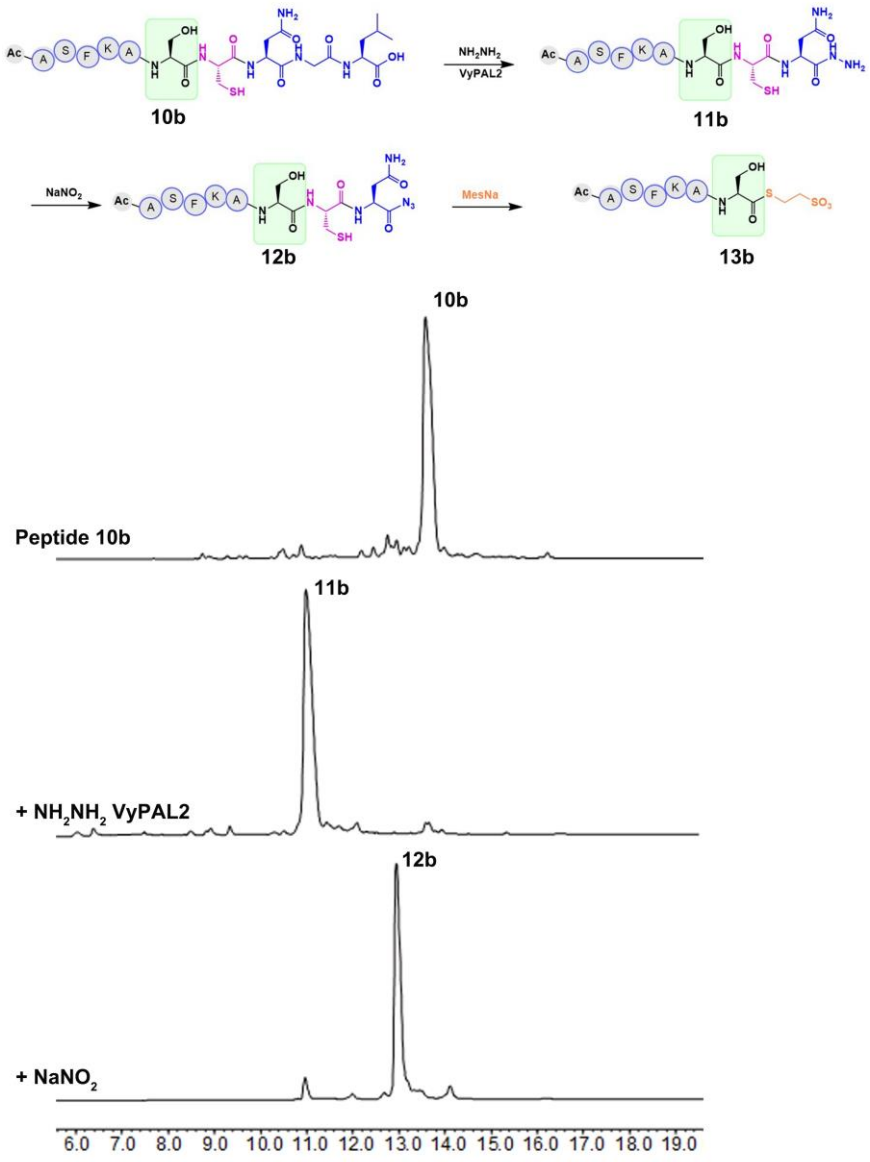


Figure 4.27. HPLC analysis of peptide **10b** (X = S) converting into **12b**.

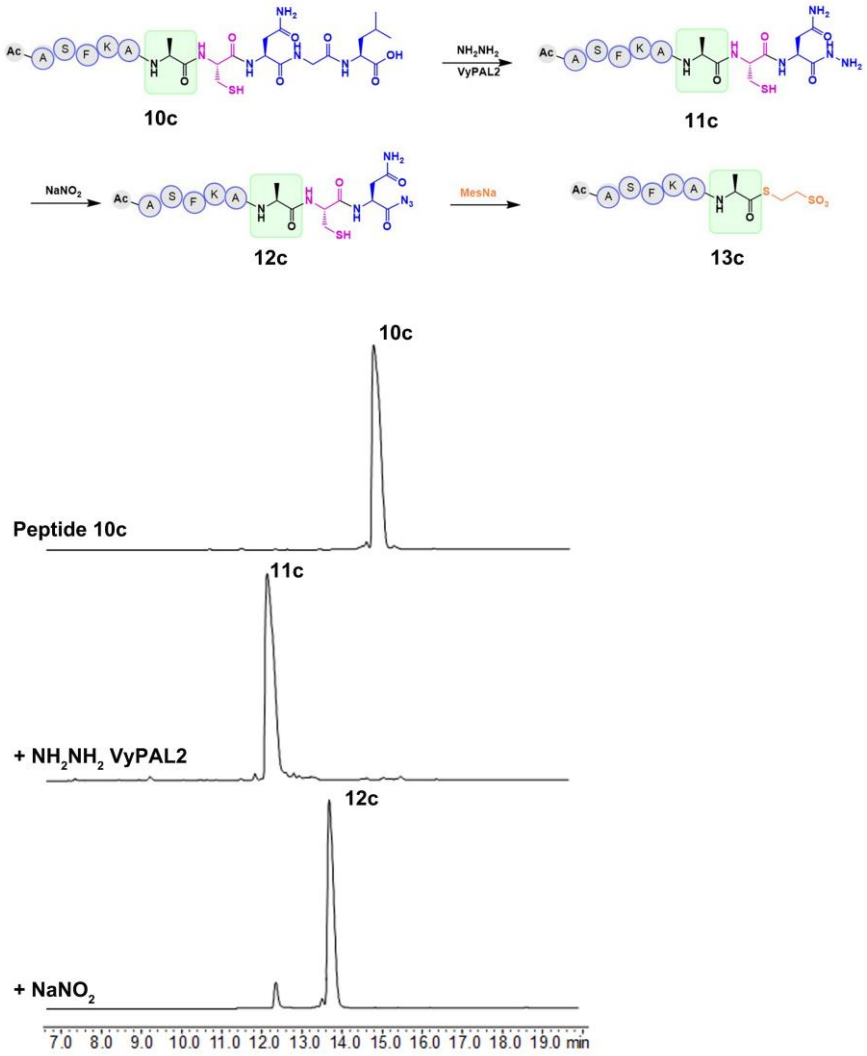


Figure 4.28. HPLC analysis of peptide **10c** (X = A) converting into **12c**.

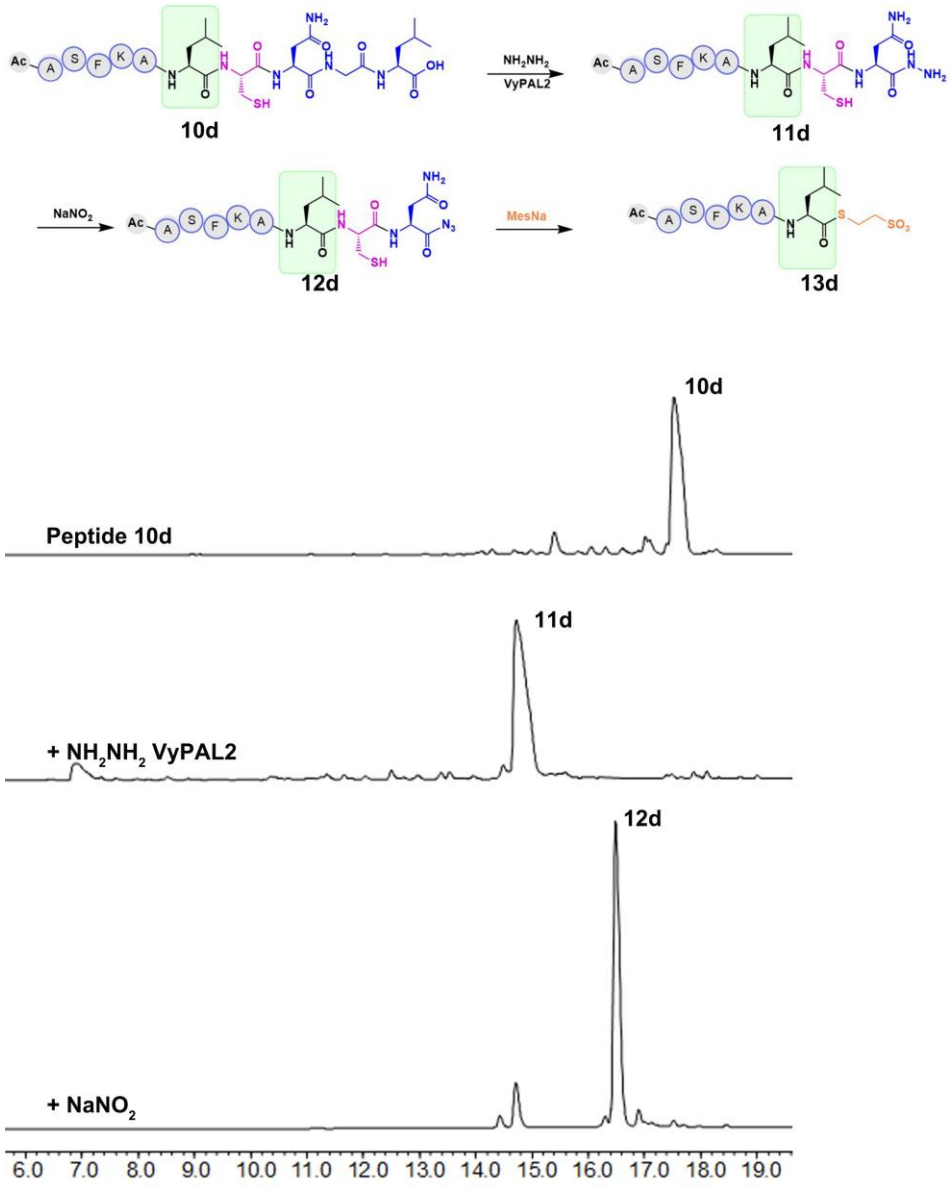


Figure 4.29. HPLC analysis of peptide **10d** (X = L) converting into **12d**.

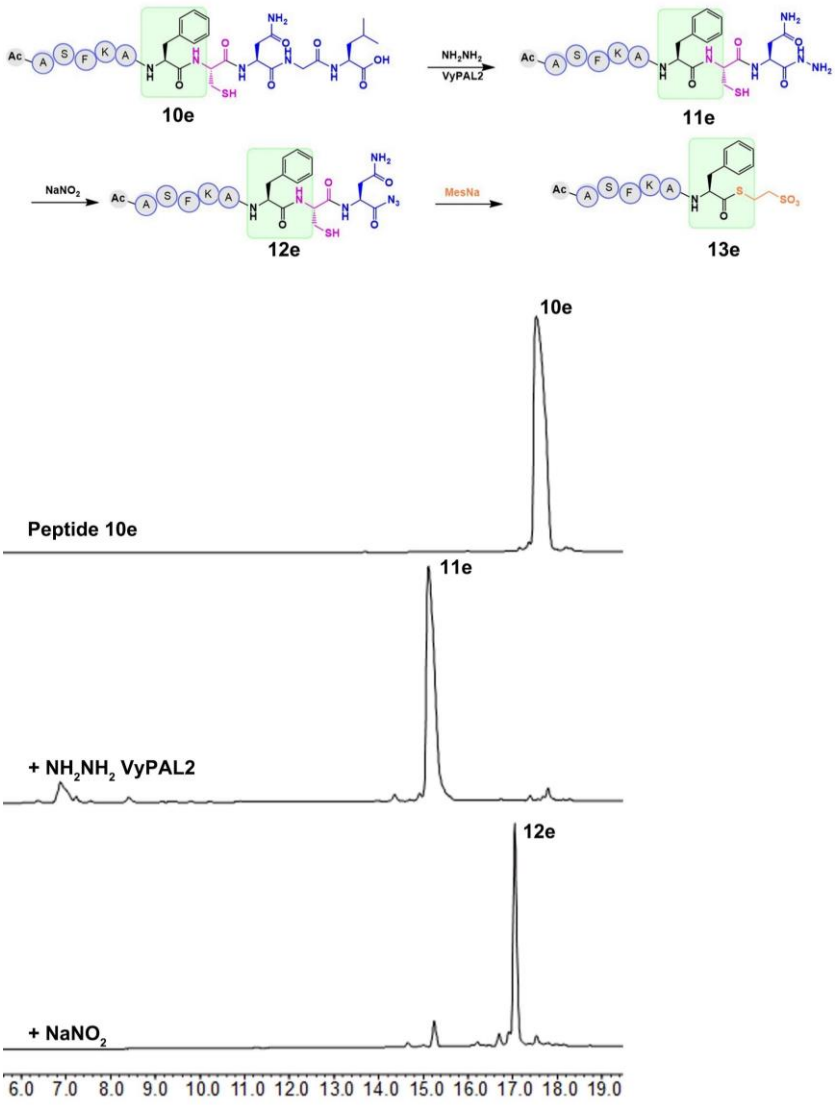


Figure 4.30. HPLC analysis of peptide **10e** (X = F) converting into **12e**.

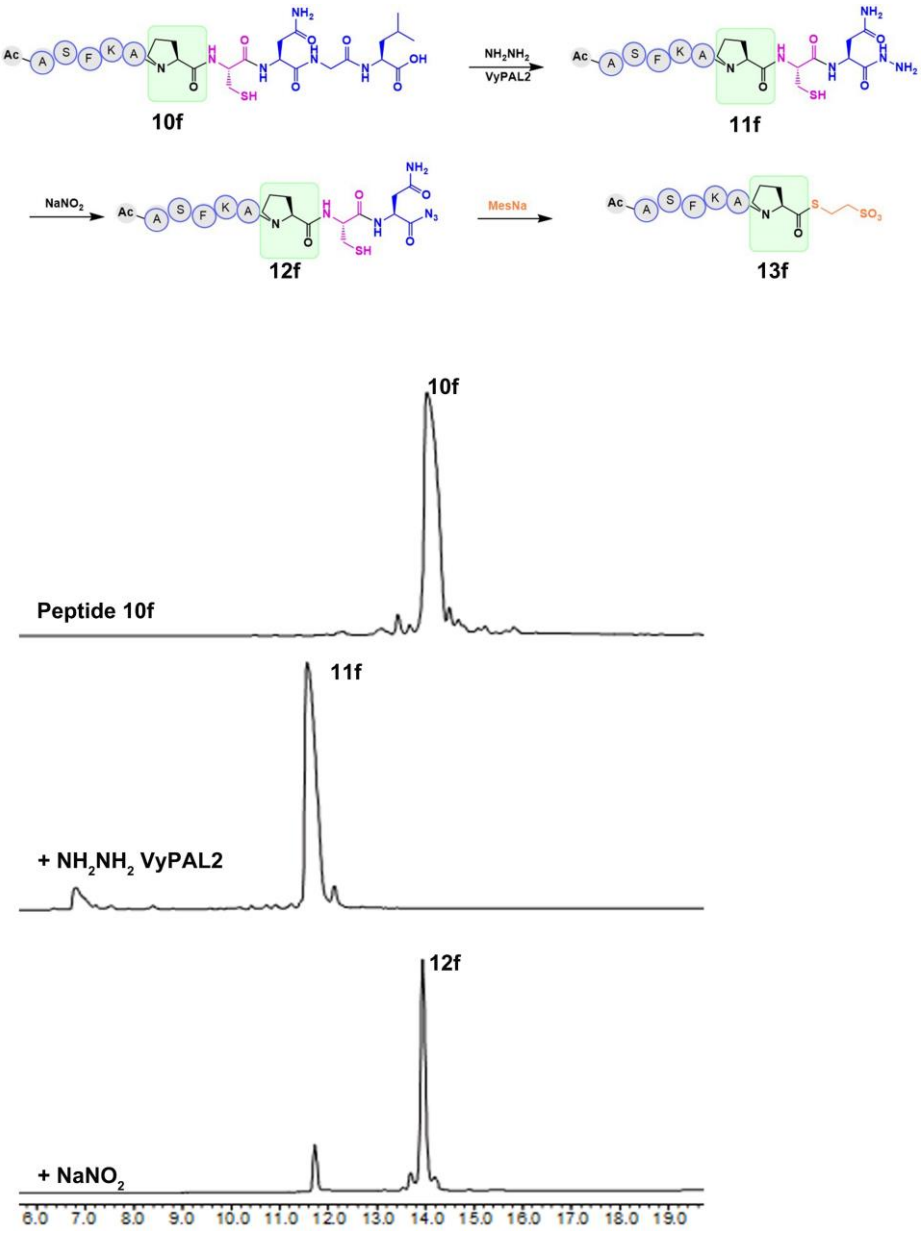


Figure 4.31. HPLC analysis of peptide **10f** (X = P) converting into **12f**.

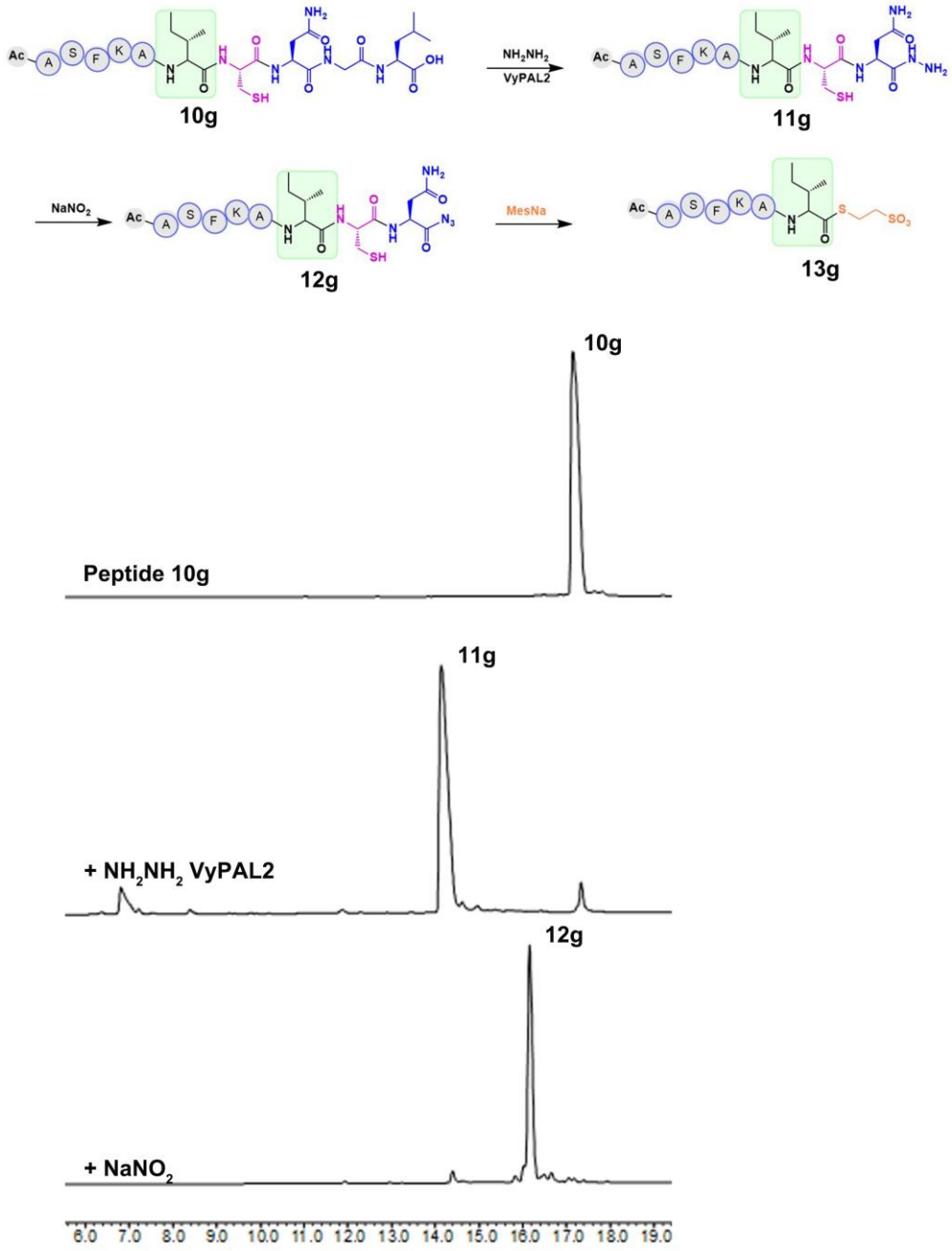


Figure 4.32. HPLC analysis of peptide **10g** (X = I) converting into **12g.**

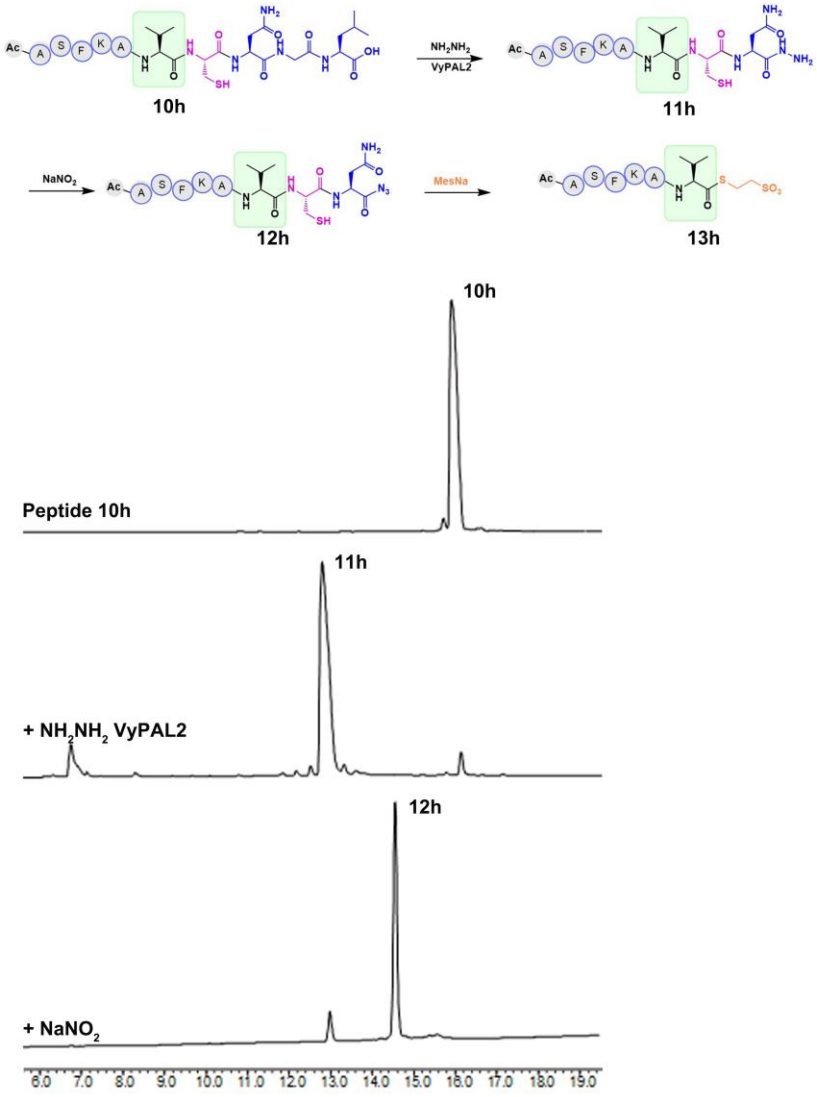


Figure 4.33. HPLC analysis of peptide **10h** (X = V) converting into **12h.**

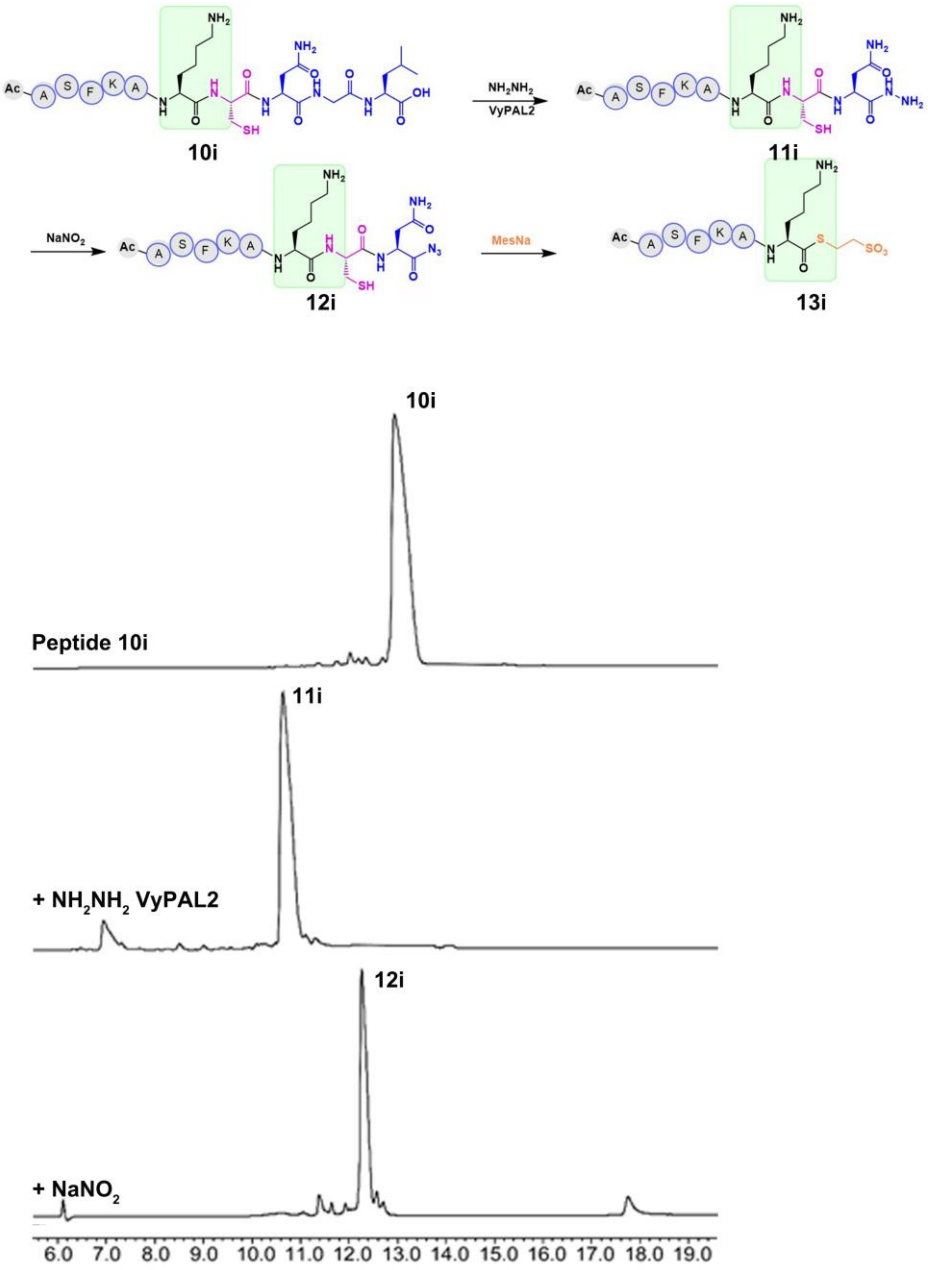


Figure 4.34. HPLC analysis of peptide **10i** (X = K) converting into **12i**.

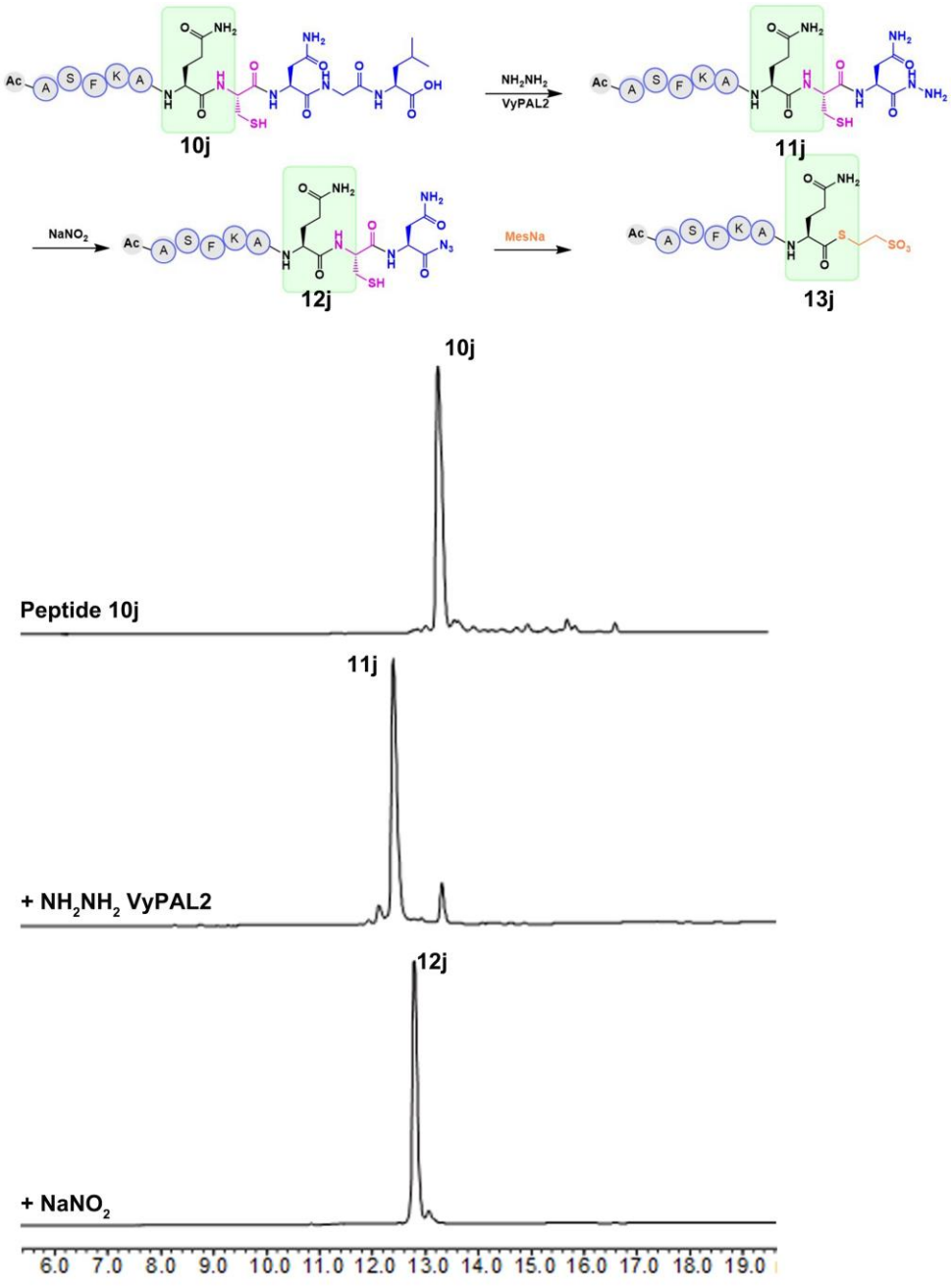


Figure 4.35. HPLC analysis of peptide **10j** (X = Q) converting into **12j**.