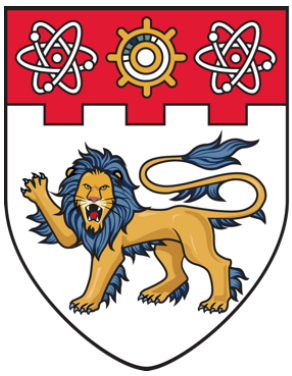


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

Broaden the Applications of Peptidyl Asparaginyl Ligases

Yiyin Xia

SCHOOL OF BIOLOGICAL SCIENCES

2022

Broaden the Applications of Peptidyl Asparaginyl Ligases

Yiyin Xia

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.

17/08/2022

. . . . . .

Date

. . . . . .

[Yiyin Xia]

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.

17/08/2022

. . . . . .

Date

. . . . . .

[Prof. Chuan-Fa Liu]

Authorship Attribution Statement

*(B) This thesis contains material from paper published in the following peer-reviewed journal(s) which I am listed as an author.

Please amend the typical statements below to suit your circumstances if (B) is selected.
Chapter 2 is published as Xia, Y.; To, J.; Chan, N.-Y.; Hu, S.; Liew, H. T.; Balamkundu, S.; Zhang, X.; Lescar, L.; Bhattacharjya, S.; Tam, J. P.; Liu, C.-F. Nγ-Hydroxyasparagine: A Multifunctional Unnatural Amino Acid That is a Good P1 Substrate of Asparaginyl Peptide Ligases. *Angew. Chem. Int. Ed*. **2021**, 60, 22207–22211.

The contributions of the co-authors are as follows:

- Prof. Chuan-Fa Liu provided the initial project direction and edited the manuscript drafts.
- I prepared the manuscript drafts. The manuscript was also revised by Prof. James. P. Tam and Dr Xiaohong Zhang.
- I performed all the laboratory work at the School of Biological Science, and I also analyzed the data.
- Dr. Balamkundu assisted in the collection of the 1D-NMR data, and Prof. Bhattacharjya assisted in the collection of 2D-NMR data.
- Dr..Zhang Xiaohong assisted in the synthesis of peptides.
- Dr. Janet To, Dr. Heng-Tai Liu and Dr. Side Hu assisted in the expression of butelase 1, butelase 2 and VyPAL2.
- Ningyu Chan assisted in the collection of trypsin inhibitory assay data.

17/08/2022
. . . . . .
Date

Xia Yiyin
. . . . . .
[Xia Yiyin]

Acknowledgement

I would like to express my immense gratitude to Prof. Liu Chuan-Fa who gave me the opportunity to join this amazing group and work under his supervision. I would like to thank Prof. Liu for his inspiration and supports through my entire PhD study. Under his encouragement, I have been given a lot of opportunities to explore new frontiers and upgrade my research skills. As a result, I become a more confident and independent researcher.

I would like to express my appreciation to my Thesis Advisory Committee members, Prof. James P. Tam and Prof. Julien Lescar, for their insightful advice and very valuable comments through each Thesis Advisory Committee meeting.

I would like to thank my dear colleagues, especially Dr. Zhang Xiaohong, Dr. Li Fupeng, Dr. Seetharamsing Balamkundu and Dr. Hemu Xinya for their generous supports, constant consultation, and tremendous help. It has been a pleasure to work with them in the same team.

I would like to thank Dr. Antony Kam and Dr. Loo Shining for their useful tips on cell assays. I would also like to thank Dr. Janet To, Dr. Side Hu and Mr. Tang Fan for their help in my experiments, Dr. Abdul Rashid B.M.M (SBS facility manager) who taught me how to use flow cytometry, Dr. Lawrence Sie Eng Kean (SBS facility managers) who taught me how to use Circular Dichroism spectroscopy and Prof. Surajit Bhattacharjya who taught me how to interpret 2D-NMR data.

Lastly, I would like to give special thanks to my parents and my husband who have always stayed by my side. Without their supports, I would have never progressed this far and achieved so much.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
ADC	Antibody-Drug Conjugate
AEP	Asparaginyl Endopeptidase
APP-IP	β-Amyloid Precursor Protein derived Inhibitor Peptide
CD	Circular Dichroism
DARPin	Designed Ankyrin Repeat Protein
Dmab	α-4-{N-[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)-3-methylbutyl]-amino} benzyl ester
Dmb	2,4-dimethoxybenzyl
DMF	Dimethylformamide
DMSO	Dimethyl Sulfoxide
DIEA	N,N-Diisopropylethylamine
EGFR	Epidermal Growth Factor Receptor
ESI-MS	Electrospray Ionization Mass Spectrometry
FAM	5(6)-Carboxyfluorescein
FBS	Fetal Bovine Serum
Fc	Fragment crystallizable
Fmoc	9H-fluoren-9-yl methoxy carbonyl
GFP	Green Fluorescent Protein
kB1	Kalata B1
LAD	Ligase-activity Determinant
IMAC	Immobilized Metal Affinity Chromatography
MALDI	Matrix Assisted Laser Desorption/Ionization
McoTI-II	Momordica cochinchinensis Trypsin Inhibitor-II
MMP	Matrix Metalloproteinase
MTT	4-methyltrityl
MVB	Prevacuolar Multivesicular Body
NMR	Nuclear Magnetic Resonance

OmpA	Outer membrane protein A
PAL	Peptidyl Asparaginyl Ligase
PBS	Phosphate-buffered Saline
PML	PAL-mediated Ligation
pGlu	pyroglutamyl
QC	Glutaminyl Cyclase
RBC	Red Blood Cell
RP-HPLC	Reverse Phase High-performance Liquid Chromatography
RTD-1	Rhesus Theta Defensin-1
SDS-PAGE	Sodium Dodecyl Sulfate PAGE
SEM	Standard Error of Mean
SFTI	Sunflower Trypsin Inhibitor
SPPS	Solid Phase Peptide Synthesis
TFA	Trifluoroacetic acid
TEV	Tobacco Etch Virus
TIS	Triisopropylsilane

Abstract

Peptidyl asparaginyl ligases (PALs) belong to a family of cysteine proteases that recognize and process substrates with a specific Asx-Xaa-Yaa tripeptide motif, where Asx can be Asn or Asp and Xaa can be most natural amino acids, while Yaa is usually a large hydrophobic amino acid. PALs can cleave the amide bond after the Asx residue in an acyl donor substrate and form a new amide bond between the Asx residue and the N-terminal residue of an acyl acceptor substrate. If the acyl acceptor substrate is from the same peptide as the acyl donor substrate, the catalysed reaction is an intramolecular cyclization. On the other hand, if the acyl acceptor is from another peptide, the PAL-mediated reaction is a intermolecular ligation. Owning their short recognition motifs and high catalytic efficiency, PALs have been extensively used in a wide range of applications, such as peptide/protein macrocyclization, protein C- or N-terminal conjugation and cell-surface protein labelling. PALs, however, can only catalyze ligation reactions at asparaginyl or aspartyl peptide bonds. Furthermore, PALs prefer Asn over Asp as the P1 residue, and as a result, the ligation efficiency at Asp is much lower than at Asn, making cyclization or ligation reactions at the Asp site difficult to happen. To broaden the scope of PAL-mediated ligation, it is desirable to design novel amino acids that can be recognized by PALs. Furthermore, PAL-mediated intermolecular ligation is reversible because the cleaved Xaa-Yaa leaving group can return and act as the acyl acceptor to attack the intermolecular ligation product, which is recognized by the same PAL as the acyl donor substrate. The reversibility issue is the major factor that limits the yield of a PAL-mediated ligation reaction.

In my research, I discovered that PALs can accept N^{γ}-hydroxyl-asparagine, also known as Asn(OH), as the P1 residue of acyl donor substrates after screening peptide substrates containing unnatural Asn analogues. In fact, Asn(OH) is a

much better P1-substrate than Asp. Therefore, linear peptide substrates containing a P1-Asn(OH) can be cyclized much more efficiently than those containing a P1-Asp. Using this approach, we synthesized several bioactive Asp-containing cyclic peptides (MCoTI-II, kB2, SFTI, and integrin-targeting RGD peptides) that are otherwise difficult targets for PAL-catalyzed cyclization due to unfavourable reaction kinetics of the P1-Asp substrates. I also prepared cyclic peptides as new matrix metalloproteinase 2 (MMP2) inhibitors using Asn(OH)-mediated cyclization, with the hydroxamic acid moiety of Asn(OH) serving as the key pharmacophore to chelate the Zinc ion in the metalloprotease. It was shown that Asn(OH)-containing cyclic peptides prepared by this method can be potent inhibitors of MMP2. One such cyclic peptide has the Ki in the single-digit nano-molar range (Ki = 2.8 ± 0.5 nM).

By coupling the PAL-mediated ligation reaction to glutaminyl cyclase (QC)-catalyzed pyroglutamyl formation, I was able to overcome the reversibility of PAL-mediated intermolecular ligation. In this method, the acyl donor substrate of PALs is designed to have a glutamine at the P1' position of the Asn-P1'-P2' tripeptide PAL recognition motif. The QC is introduced as a catalyst to selectively quench the nucleophilic α-amine of N-terminal glutamine of the released leaving group into a lactam. By masking the nucleophilicity of the cleaved Gln-Leu dipeptide using QC, PAL-mediated ligation can achieve near-quantitative yields even at an equal molar ratio between two large-size ligation partners. As a result, this cascade enzymatic reaction scheme greatly improves the efficiency of PAL-mediated intermolecular ligation, making it especially useful in the precision biomanufacturing of protein-based biologics such as antibody-drug conjugates and radio-immunoconjugates.

In my studies, we made PAL a more powerful and versatile tool for bioconjugation by (1) broadening the substrate scope of P1 Asx to unnatural Asn analogues and (2) overcoming the reversibility of PAL-mediated ligation by coupling it to QC-catalyzed pyroglutamate formation.

Chapter 1. Introduction

1.1. Ligases

Recent years have seen an increasing use of bioconjugation techniques in protein structure-function studies and for the therapeutic development of biologics. Compared to traditional chemical conjugation methods [1-8], the bioconjugation reactions performed by enzymes are highly specific and operate under mild, physiological conditions. There are a number of naturally existing or engineered enzymes (**Figure 1. 1**) that have the ability to catalyze the formation of an amide bond between an acyl donor and an acyl acceptor. These enzymatic ligation reactions can be generally described as a two-step process: (1) formation of an acyl-enzyme intermediate (acyl-E) with the enzyme's active-site residue (Cys or Ser), and (2) aminolysis of the acyl-E by the amine nucleophile of the acyl acceptor. These ligases are powerful tools for protein bioconjugation through N- or C-terminal labelling. Among all the ligases, a few of them stand out due to their robustness in ligation reaction, which are subtiligases, trypsiligases, sortases, and peptidyl asparaginyl ligases (PALs) (**Table 1. 1**).

Subtilisin is a serine protease from *Bacillus subtilis* [9]. The double mutant (S221C, P225A) of subtilisin, namely subtiligase, is a predominant ligase. The Ser-to-Cys (S221C) [10] and Pro-to-Ala (P225A) [11] mutations enable the formation of an acyl-enzyme thioester intermediate from peptide ester substrates and increase the aminolysis to hydrolysis ratio of the acyl-E intermediate. In addition to these two mutations, other mutations are found to further enhance the ligase activity [12], increase the stability [11], and alter the substrate specificity of subtiligases [13]. These engineered subtiligases have found broad use for protein N-terminal labeling as they have a high selectivity for the N-terminal amine over the sidechain amine of lysine, recognizing yet a wide range of amino acid residues at the N-termini of proteins. Nevertheless, subtiligase can only use a peptide ester or thioester as the acyl donor substrate. The requirement for this

unnatural substrate largely limits subtiligase-mediated protein modification to N-terminal labeling [14-16] unless a protein C-terminal thioester can be generated through the use of the intein splicing process [17]. Besides, hydrolysis of the peptide ester substrate is a side reaction which can lower the yield of subtiligase-mediated ligation.

Trypsiligases are Zn^{2+}-dependent ligases that are engineered from the serine protease trypsin [18] after four mutations [19]. Trypsiligases have a high specificity for the YRH tripeptide motif. They cleave the peptidyl bond between Y (Tyr) and R (Arg) and perform transpeptidation between C-terminal YRH-containing peptides and N-terminal RH containing peptides (usually in excess). Trypsiligases have been used for both N and C-terminal protein bioconjugations [20] and for live cell labelling [21]. Nevertheless, due to the high specificity of the trypsiligase, ligated products will contain an extra YRH motif in its sequence. Furthermore, this YRH motif in the ligated product can be re-recognize by trypsiligase and cause secondary hydrolysis.

Sortases are a class of naturally occurring enzymes that catalyze covalent attachment of bacterial proteins to bacterial cell-wall peptidoglycans [22]. Sortase is an essential enzyme for bacteria as it is involved in bacteria virulence and colonization [23]. Sortase A is from *Staphylococcus Aureus* and it recognizes a LPXTG motif whereas X can be any amino acid [23]. Sortase A first cleaves the bond between T (Thr) and G (Gly) to generate a thioacyl-enzyme intermediate. The intermediate is then attacked by an N-terminal pentaglycine peptide to from a new amide bond. Sortase-mediated ligation, or sortagging, has been used to generate a wide range of modified proteins, such as antibody-drug conjugates (ADCs) [24-26]. However, the high substrate specificity and low catalytic efficiency of sortases limit their broad applications.

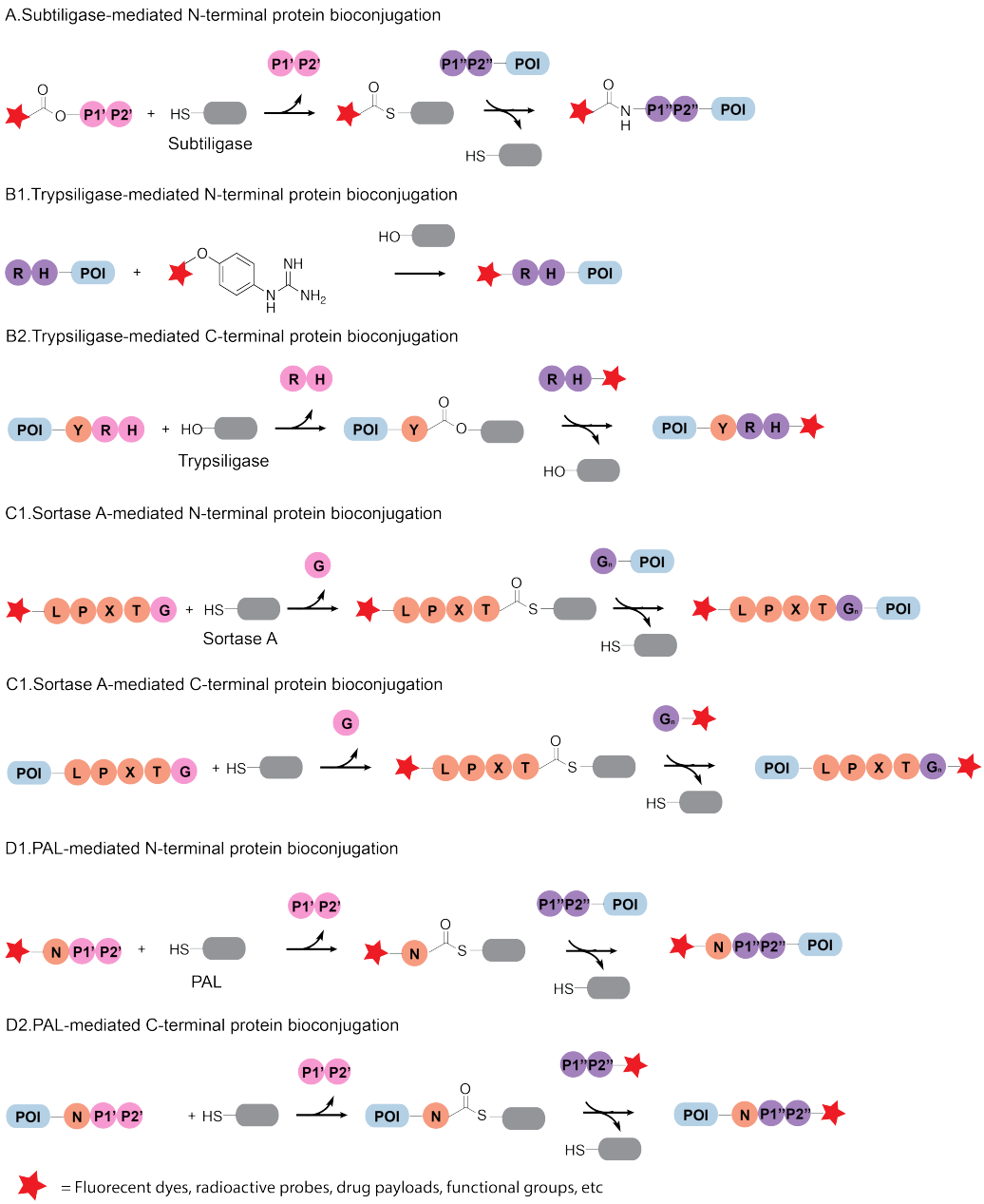


Figure 1. 1. Reactions catalyst by peptide ligases.
(A) Subtiligase-meidated N terminal protein labelling. (B) Trypsiligase-mediated N&C terminal protein labelling. (C) Sortase A-mediated N&C terminal protein labelling. (D) PAL- mediated N&C terminal protein labelling.

Table 1. 1. Comparison of different ligase-mediated reactions.

Enyzme	Advantages	Disadvantages
Subtiligase	• Broad substrate scope • High enzyme efficiency	• Requirement for peptide ester substrates • Largely limited to N-terminal labelling
Trypsiligase	• High enzyme efficiency • Applicable for both N & C-terminal labelling	• Requires RYH motif • Secondary hydrolysis
Sortase A	• Applicable for both N & C-terminal labelling	• Strict requirement for LPXTG motif • Low enzyme efficiency
PAL	• High enzyme efficiency • Applicatable for both N & C-terminal labelling	• Requirement for Asn at the P1 position • Reversibility of the reaction

In recent years, a novel class of plant-derived enzymes, peptidyl asparaginyl ligases (PAL), have been found to have high catalytic efficiency for peptide bond formation [27]. The first PAL, named as butelase 1, was isolated from *Clitoria ternatea* by the Tam lab who reported that butelase 1 was around 24,000 times faster than sortase A in catalyzing peptide cyclization reactions [28]. Butelase 1 and other PALs recognize a tripeptide motif, Asx-Xaa-Yaa (P1-P1'-P2'), to cleave the Asx-Xaa peptide bond and then re-join the Asn/Asp residue with the N-terminal amino acid of an incoming peptide chain in a transpeptidation reaction. PALs accept a wide range of amino acids at P1'-P2'; therefore, the ligation mediated by PALs is almost traceless with only Asx at the ligation junction. Standing out due to their high catalytic efficiency and their minimal recognition motifs, butelase-1 and other PALs are useful tools for protein bioconjugations. However, intrinsic to transpeptidation, PAL-mediated ligation

is reversible, which limits the yield of this reaction.

1.1.1. Asparaginyl endopeptidases (AEPs) and peptidyl asparaginyl ligases (PALs)

Asparaginyl endopeptidases (AEPs), also named legumains or vacuolar processing enzymes, are a family of enzymes that play an important role in programmed cell death [29] and maturation of seed storage proteins in plants [30]. They are cysteine endopeptidases that specifically recognize and cleave the peptide bond after an Asx residue (*i.e.* Asn or Asp) [31]. Their hydrolytic activity was firstly reported in 1980s by Lord's lab and Nishimura's lab [32, 33]. In recent years, it has been reported that many enzymes from this family actually have dual functions [34] – they act as proteases at acidic pH but have ligase activity at neutral or basic pH [35-37]. Both proteolysis and ligation reactions involve the formation of the similar acyl-enzyme intermediates. However, as the basic N-terminal amine nucleophile will be protonated at acidic pH, it is unavailable for the aminolysis of the acyl-enzyme thioester intermediate.

Despite sharing a conserved structure at the catalytic site, a small population of AEPs have predominant ligase activity. These AEPs are able to catalyze transpeptidation at the P1 Asx residue thereby forming new peptide bonds at weakly acidic to weakly basic pH, instead of hydrolyzing the Asx–Xaa bonds as AEPs usually do. The transpeptidation activity of these enzymes makes them qualify as peptidyl asparaginyl ligases or PALs. They are reported to be responsible for generating naturally occurring cyclic peptides in plants, such as the cyclic kalata-type peptides and cyclic trypsin inhibitors [27, 38, 39].

To date, butelease-1 [27], OaAEP1b-C247A [40], VyPAL2 [36] and a few others that catalyze transpeptidation at the asparaginyl peptide bonds with little hydrolase activity are peptidyl asparaginyl ligases (PALs) known to date.

1.1.2. Activation of PALs

In plants, PALs/AEPs are first generated as inactive zymogens consisting of a ~36kDa active domain and a ~13kDa prodomain or cap, they are auto-

proteolytically maturated in the acidic environment of the prevacuolar multivesicular bodies (MVB) and protein storage vacuole to afford the active enzyme. For recombinant preparation of PALs, the full-length zymogens or proenzymes (~49 kDa) are first expressed, followed by an in vitro autoactivation step in acidic conditions to cleave off the prodomain (with cap) from the active enzyme (**Figure 1. 2**). The *in vitro* activation conditions for different PALs vary slightly. In general, the pro-enzyme needs to be treated at pH 4-5 and 37 °C for 2-12 h [41, 42]. Presumably, under the acidic conditions, the electrostatic interactions between the cap and active domain are weakened, loosening the cap from the core domain. This exposes the catalytic surface of the core domain, allowing it to cleave peptide bonds at Asn/Asp residues of the linker peptide between the cap and the core domain. This in vitro process of PAL activation at acidic pH mimics their autoproteolytic maturation in plant.

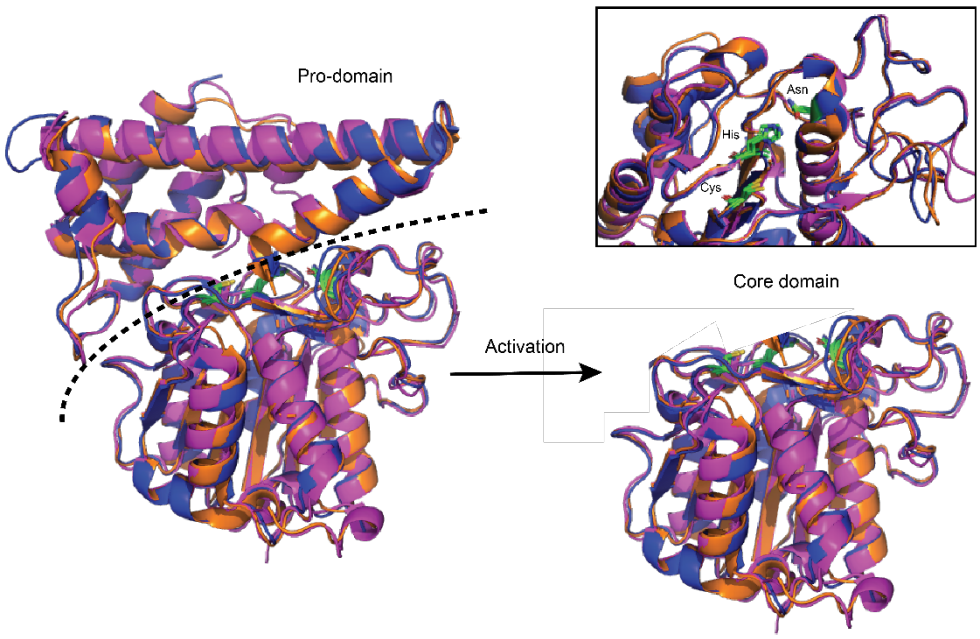


Figure 1. 2. Overlay structures of Butelase 1 (PDB:6DHI, orange), VyPAL2 (PDB: 6IDV, blue) and OaAEP1b (PDB: 5H0I, magenta) [36, 40, 43].
The dotted line separates the prodomain and core domain of PALs. Auto-activation cleaves off the cap and exposes the Cys/His/Asn triad on the catalytic surface of the core domain.

1.1.3. The active site of PALs

Most of the AEPs and PALs have a similar overall architecture with conserved catalytic residues. The active site of AEPs and PALs contains three catalytic residues: Asn, His and Cys. The Cys residue at the active site plays a key role in the enzymatic reaction (**Figure 1. 3**). The cysteine free thiol acts as a nucleophile and attacks the amide carbonyl group after a P1-Asx residue of the substrate to form an acyl-enzyme thioester intermediate. For an acyl-PAL intermediate, it is later attacked by the N-terminal amine of an acyl acceptor peptide to form a new amide bond. The His in the enzyme active site is involved in proton transfer with the catalytic Cys, while the function of the Asn in the active site is less clear. Asn is engaged in hydrogen bonding with the His, which may also contribute to the catalytic proton transfer mechanism. Furthermore, the residues in the S1 pocket for P1 Asx recognition and binding are also well conserved in PALs and AEPs (**Figure 1. 3**). However, the residues that form other binding pockets are less conserved and more variable.

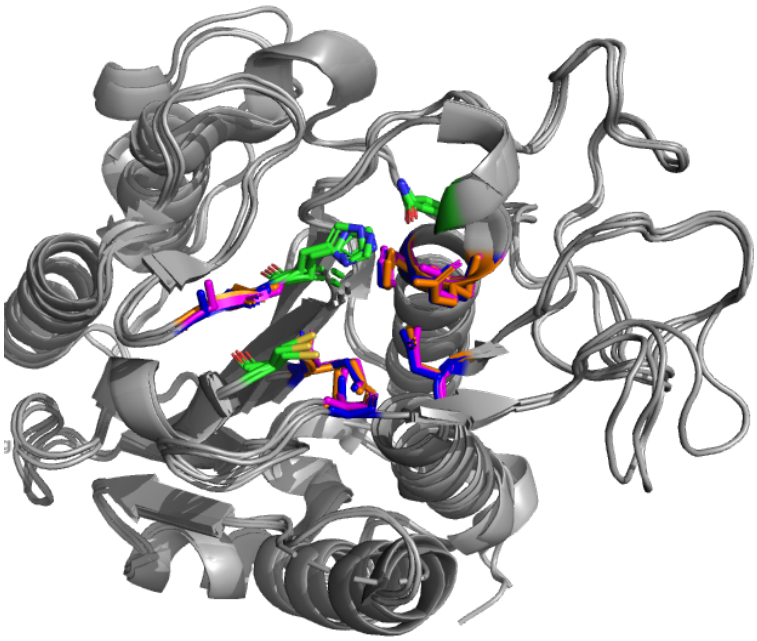


Figure 1. 3. Overlay structure of butelase 1, VyPAL2 and OaAEP1b.
The catalytical triads (Cys, His, and Asn) are indicated as green, and other 5 residues (Ser, Asp, Glu, Arg, His) that form the S1 pockets of butelase 1, VyPAL2 and OaAEP1b are indicated as orange, blue and magenta, respectively.

1.1.4. Mechanism of catalysis of PALs

The reaction of PAL-mediated ligation occurs in two steps (**Figure 1. 4**). During the first step, which is also the rate-determining step, PAL attacks and cleaves the asparaginyl peptide bond in the acyl donor peptide substrate to form an acyl-enzyme thioester intermediate. In the second step, the thioester is attacked by the N-terminal amine of an incoming nucleophile peptide substrate to form a new Asn-Xaa peptidyl bond. This nucleophile can be either from the same peptide, resulting in an intramolecularly cyclised product or from another peptide, resulting in an intermolecularly ligated product. The aminolysis is what makes PAL a peptide ligase rather than an AEP-type protease which, on the other hand, utilizes a catalytic water molecule to hydrolyze the acyl-enzyme thioester in the second step.

Harris *et a.l* studied the *r*OaAEP1b reaction mechanism using ^{18}O-labelled water

($H_2{}^{18}O$). Their result shows that ^{18}O incorporation only occurs in the hydrolyzed product but not in the cyclized product [35]. This indicates that similar to the *AtLEGγ*-mediated reaction [44], water is excluded during *r*OaAEP1b-mediated cyclization. The study confirms that the hydrolysis and aminolysis reactions of *r*OaAEP1b are through two different and independent pathways.

The different behaviours during the second step between AEPs and PALs have been investigated further. In a study by Yang *et al*., a key residue located around the S2 site [40] has been shown to have important effect in controlling the catalytic efficiency of OaAEP1 and the ratio of aminolysis versus hydrolysis although in their study this site was mistakenly recognized as the binding site of the nucleophile substrate. This position has been identified as a gatekeeper [40] or Ligase-Activity Determinant 1 (LAD1) in other studies [36, 37] which binds the acyl-side substrate. Studies show that G (Gly) at LAD1 leads to predominant hydrolysis activity and a bulky residue (V, I, L, M) at LAD1 leads to predominant aminolysis activity. Additionally, a second position close to the S1' site is found to also play a role in altering the aminolysis/hydrolysis ratio and named as Ligase-Activity Determinant 2 (LAD2) [36, 37]. LAD2 is constituted of two residues, and in ligases these two residues are often AP or AA. LAD1 and LAD2 may modulate the ligase activity by (1) stabilizing the thioester intermediate and (2) controlling the accessibility of water or amine nucleophile to the thiolester intermediate. The impact of the PAL's so-called Makers of Ligase's Activity (MLA) region on the ligase activity has also been investigated. However, the MLA region is less conserved among different PALs (**Figure 1. 4**). Modifications on MLA of OaAEP11b tend to reduce the catalytic efficiency rather than shift the hydrolysis-to-aminolysis ratio [45]. To date, there is no crucial evidence showing that MLA can determine the ligase activity.

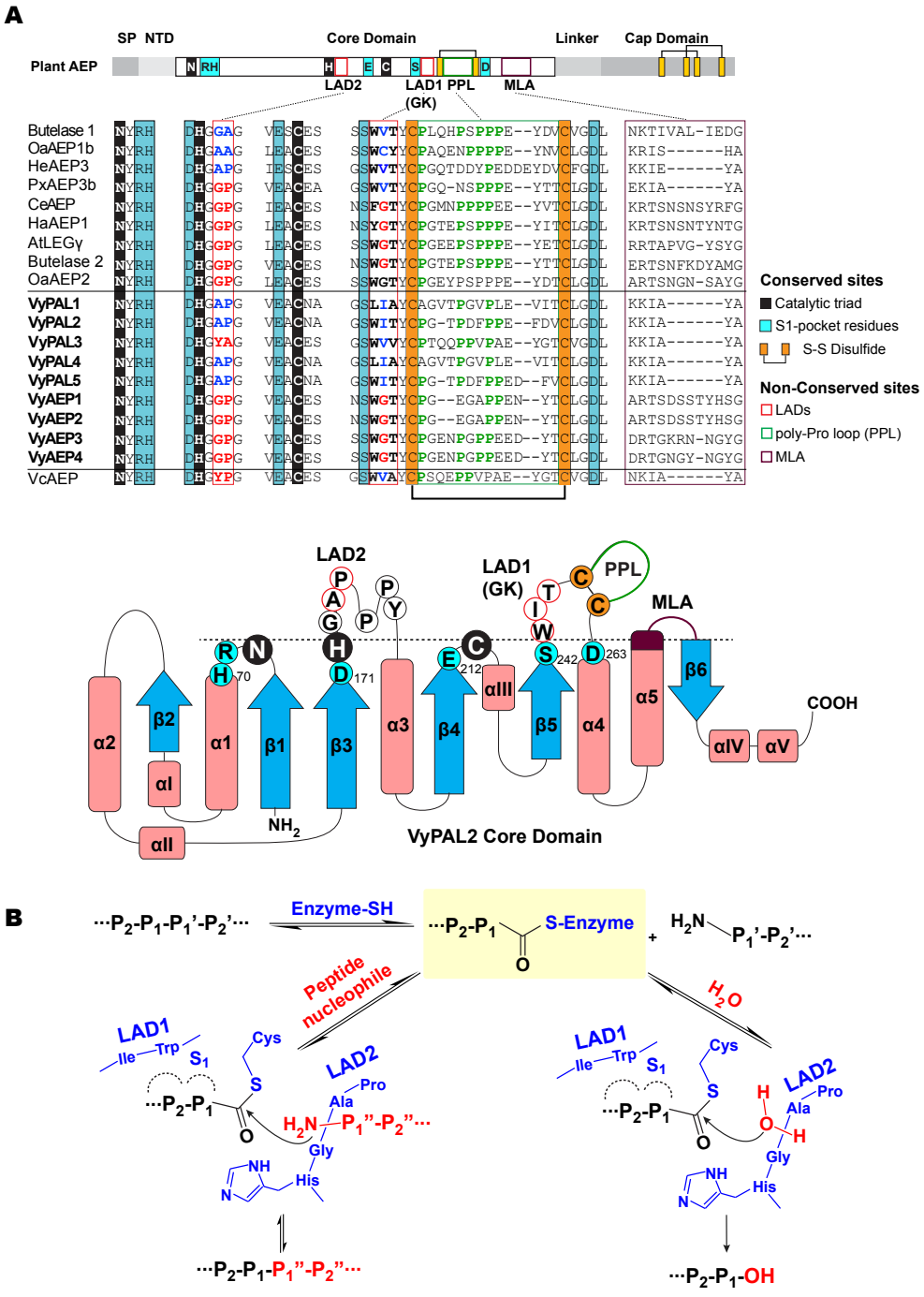


Figure 1. 4. Structural determinants of PALs and proposed catalytic mechanism for PAL/AEP-mediated reaction.
(A) LAD1 and LAD2 for determining ligase activity. (B) The hydrolysis and aminolysis reaction mechanisms [36].

1.1.5. Substrate specificity of PALs

PALs recognize a short tripeptide motif (P1-P1'-P2'), in which the P1 residue must be Asx (**Figure 1. 5**). During catalysis, the peptide bond between P1 and P1' is cleaved to release P1'-P2' and the P1 residue is re-joined with a P1"-P2" motif of an incoming peptide chain. Each PAL has a slightly different preference at the P1'-P2' and P1"-P2" positions. Generally, PALs have a broad substrate scope for P1' and P1'' (except Pro) residues, while it prefers bulky hydrophobic residues (Leu, Ile, Val) or cysteine at the P2' (similar for P2'') position.

Although PAL is an Asx-specific enzyme, its activity towards P1-Asp peptides is drastically lower than the P1-Asn peptides (over 100-fold difference) at the optimal pH 6.5 [27]. For example, a linear precursor of kalata B1 with the NHV tripeptide motif at the C terminus, kB1-NHV, was cyclized by butelase 1 (at 0.0025 eq of the enzyme) in 0.8 h. In contrast, when kB1-DHV was treated with butelase 1 under the same conditions, only 10% cyclic product was obtained after 4 h. Both peptides have exactly the same sequence except for the P1 residue (Asn *vs* Asp).

Therefore, most of the protein substrates designed for C-terminal labelling have an engineered C-terminal NXY motif. Meanwhile, for protein N-terminal labelling, the protein is usually engineered to contain an N-terminal dipeptide motif, such as GI, as the acyl acceptor.

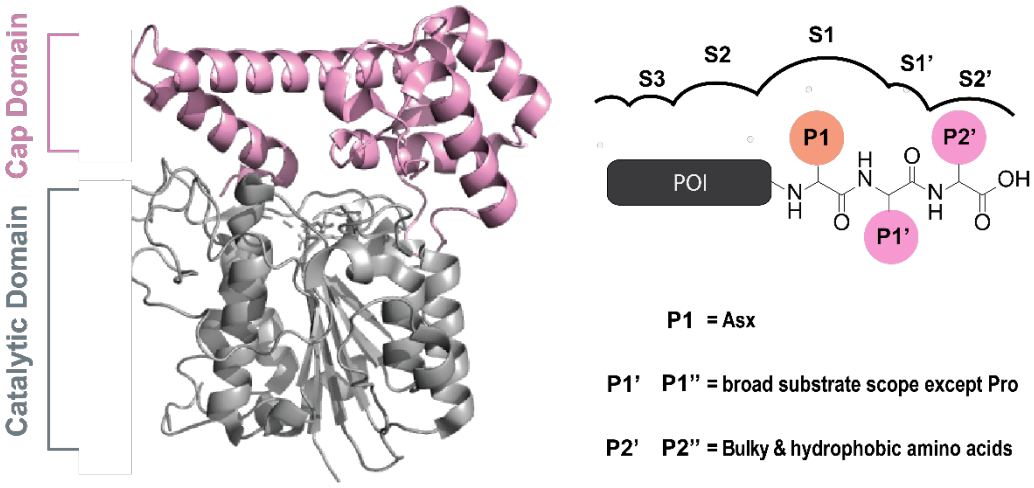


Figure 1. 5. Substrate scope of PALs at P1-P1'-P2'.

1.1.6. Individual PALs

1.1.6.1. Butelase 1

Butelase 1 is the first enzyme characterized as a PAL. It was isolated from *Clitoria ternatea*, a blue butterfly pea locally also known as Bunga Telang [27, 28, 46, 47]. Butelase 1 is structurally very close to plant AEPs as it shares high sequence homology and the same catalytic triad with these proteases. However, butelase 1 functions as a pure PAL with no detectable AEP (*i.e.,* protease) activity. The tripeptide motif in the native substrate for butelase 1 is NHV; however, other amino acids at P1'P2' are also acceptable for butrelase 1. The k_{cat}/K_m of butelase 1 in cyclizing a short peptide (GISTKSIPPISYRNSLAN) is 971,936 $M^{-1} \cdot s^{-1}$ [36]. With its high catalytic efficiency and high tolerance for the N-terminal amino acids of the nucleophile peptides, butelase 1 is the fastest and most versatile peptide ligase known to date.

1.1.6.2. OaAEP1b-C247A

Following the discovery of butelase 1, OaAEP1b was discovered by Anderson et al, [38], which was shown to have a high transpeptidation activity over hydrolysis. The C247A mutant of OaAEP1b has an around 160-fold increase in catalytic rate [40]. OaAEP1b-C247A can be easily expressed using the T7-Shuffle *E. coli* system. The robustness of *E. coli* expression makes OaAEP1b-C247A a powerful

ligation tool.

1.1.6.3. VyPAL2

Using bioinformatic tools for sequence homology analysis of AEPs/PALs, Hemu *et al* identified a new PAL family from Viola yedoensis (Vy) and named them as VyPALs [36]. Having an Ile at LAD1 and the Ala-Pro dipeptide at LAD2, VyPAL2 was found to be a prefect ligase. VyPAL2 has a slightly lower kinetic of peptide cyclization than butelase 1 (k_{cat}/K_m ~ 274,325 $M^{-1} \cdot s^{-1}$), but it can be easily produced by insect cells in a very high yield (10-20 mg/L).

1.2. PALs: versatile tools for bio-modifications

PALs have been utilized from head-to-tail cyclization, protein modification and live cell labelling. The minimal substrate recognition motif and high enzymatic efficiency indeed make PALs versatile biotechnological tools. In this section, several examples of the PAL-mediated reactions are shown and described in detail.

1.2.1. General conditions of PAL-mediated reactions

PAL-mediated reactions are typically conducted at pH 6.5-7 (the optimal pH for ligase activity) and room temperature or 37 ºC using a substrate-to-enzyme ratio of 1000:1. The reactions usually take a few minutes to a few hours to complete, depending on the reactivity and accessibility of the substrates. PALs have been used to cyclize peptides/proteins of various lengths with a near-quantitative yield. PALs are also very useful for protein N and C-terminal labelling. Due to the reversibility nature of this reaction, typically a large excess of one of the two ligation partners is used to shift the reaction equilibrium to the product side. Additionally, PALs are compatible with living cells and can modify living cell surfaces with various moieties including large proteins. PAL-mediated cell surface labelling is often conducted using a large excess of the labelling substrate in PBS at 37 ºC.

1.2.2. Macrocyclization of peptides and proteins by PALs

Macrocyclic peptides/proteins are promising drug candidates in the development of therapeutics. The large surface area of macrocycles not only allows them to inhibit protein-protein interactions which is usually un-druggable when using small molecules, also provides them with good selectivity. Macrocyclization allows peptides/proteins to achieve a certain level of structural pre-organization. This conformational restriction of macrocycles minimizes the enthropic cost in binding to the targets, resulting in a high potency. Additionally, macrocyclization increases the proteolytic resistance and therefore increases the pharmacological activity of the cyclic peptides or proteins.

Many backbone-cyclized cystine-rich peptides (CRPs) occur naturally in plants/animals and have multiple disulfide bonds. The multiple disulfide bonds further stabilize the cyclic peptides and makes them suitable therapeutic candidates. Yap *et al* used OaAEP1b to cyclize three recombinant CRPs (SFTI-KLK5, [G22N]cVc1.1 and MCoTI-II) in *vitro* and obtained a high yield (> 90%) for a 100-mg scale reaction (**Figure 1. 6** a) [48]. This study indicates that PALs are ideal tools for large-scale preparation of cyclic peptides.

Hemu *et al.* reported that butelase-1 can perform one-step oligomerization of antimicrobial peptide motifs (**Figure 1. 6** b) [49]. When used at high concentrations, a short peptide substrate (< 9 aa) can be turned into a cyclo-dimer and cyclo-trimer by the catalysis of butelase 1. The MIC results show that the cyclo-dimer and cyclo-trimer of the RLYR tertrapeptide motif have an increased antimicrobial activity compared to the original linear monomer.

Yang *et al.* [40] made the observation of backbone-to-side chain cyclization between a Lys sidechain amine and the C terminus by OaAEP1b-C247A when there is no N-terminal amine available (**Figure 1. 6** c). This reaction shows the PALs can tolerate unconventional nucleophile substrates.

Hemu *et al.* later reported a more convenient methods to perform PAL-mediated cyclizaion using immobilized PALs [50]. Both VyPAL2 and butelase-1 have been non-covalently immobilized on ConA beads or avidin beads. Studies shows that immobilization stabilizes the PALs. The immobilized PALs can be used multiple times with no significant loss of enzyme activity and be stored in TCEP-containing buffer for a long period of time.

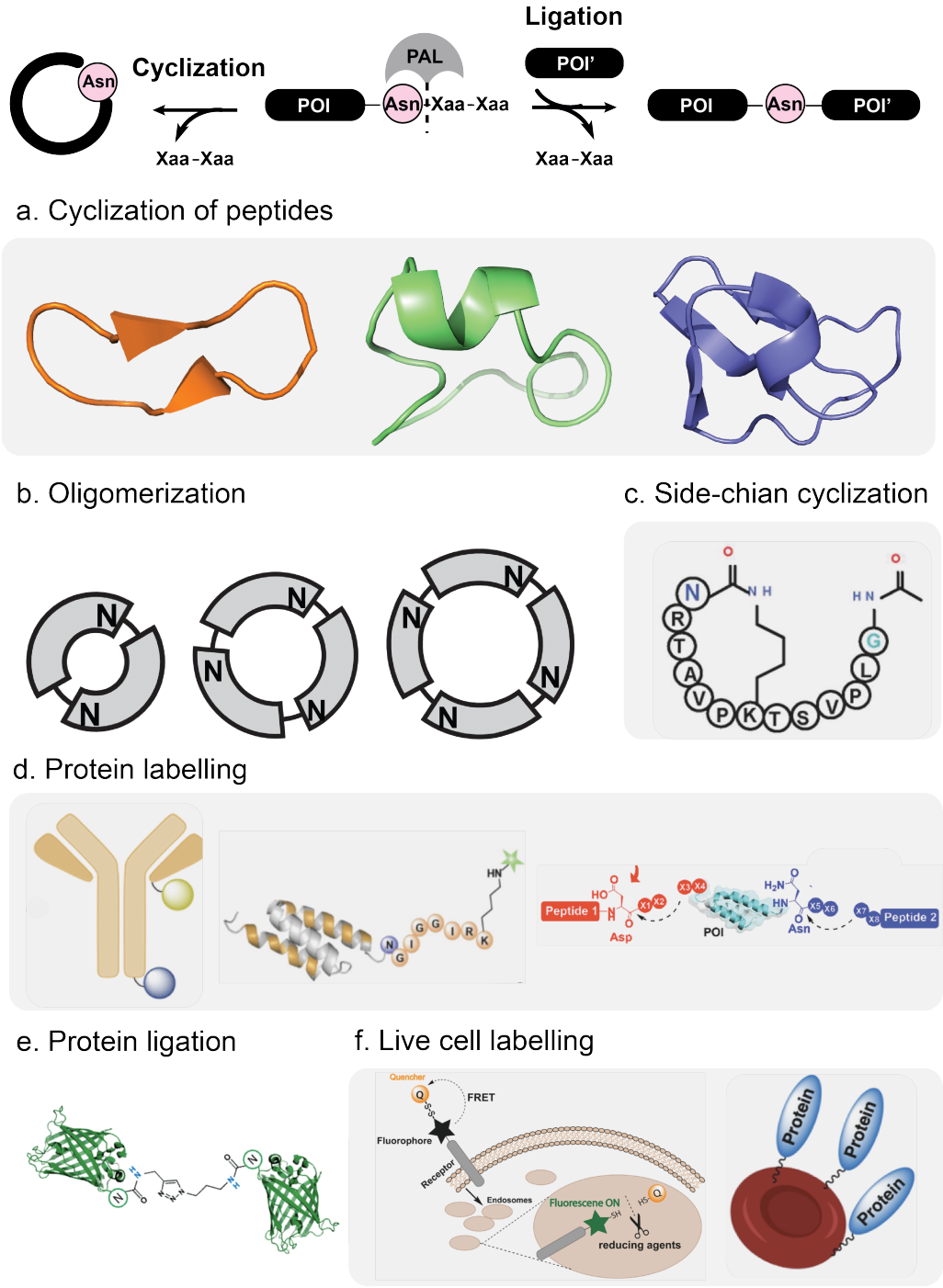


Figure 1. 6. PAL mediated reactions.
(a). Cyclization of cysteine-rich peptides: SFTI-KLK5 (PDB: 6NOX), [G22N]cVc1.1 (PDB: 4TTL) and MCoTI-II (PDB:1HA9) [48]. (b). Oligomerization of antimicrobial peptide motifs [49]. (c) Backbone to Side-chain cyclization [40]. (d) Protein labelling [52, 54]. (d) Protein ligation [53]. (f) live cell labelling [57, 58].

1.2.3. Protein labelling by PALs

Site-specific protein labelling is an effective tool for protein function studies and development of protein therapeutics. The N and C-termini of proteins are suitable ligation sites for PALs. For instance, the N or C-terminus of a protein can be engineered to have a GI dipeptide or a NGL tripeptide motif, respectively, to allow N or C-terminus specific labelling. A wide range of small molecules (fluorophores, biotin, drug payloads, etc) have been labelled onto proteins of interest (ubiquintin, DARPins, nanobodies, antibodies, etc) by PALs [36] [51]. With the absolute selectivity of PAL-mediated ligation, the labelling of a protein provides a good way to study its properties and functions.

Harmand *et al*. ligated two nanobodies (15 kDa) consisting of a C-terminal LEPTGG motif and NHV motif, respectively, to an oligonucleotide linker with dual N termini (H_2N-GGG-linker-VG-NH_2) using sortase A and butelase 1 in a one-pot reaction (**Figure 1. 6** d) [52]. This study demonstrates C-to-C protein fusion based on the different substrate specificities of sortase A and butelase 1. Later, Durek et al. also performed enzymatic C-to-C protein ligation (**Figure 1. 6** e) [53]. They reported that OaAEP1b can recognize and accept simple amines as the nucleophile substrates such as a C-terminal L-Eda. Using a GLHG-Anp-L-Eda peptide, OaAEPlb can ligate GFP-NGLH and nanobody-NGLH to the peptide to form a C-to-C fused protein heterodimer.

Wang *et al*. reported that butelase 1 and VyPAL2 have a distinguishable substrate specificity towards the P1'-P2' (P1''-P2'') dipeptide [54]. The study shows that butelase 1 prefers HV while VyPAL2 prefer GF at P1'-P2'. This makes them orthogonal ligases for protein dual labelling. Therefore, tandem ligation can be achieved to label a protein at both its N and C-terminal ends in either N-to-C or C-to-N direction using butelase1 and VyPAL2 consecutively. Zhang *et al*. found that PALs such as VyPAL2 and OaAEP1 have appreciable ligase activity towards P1-Asp at acidic pH (**Figure 1. 6** d) [55]. The kinetic for cyclization of

P1-Asp substrates at pH 4-5 is much higher than at neural pH. By changing pH, protein orthogonal ligation can be achieved using both P1-Asn and P1-Asp substrates.

Deng *et al*. conducted a step-controlled polyprotein assembly on a glass coverslip using OaAEP1b and tobacco etch virus (TEV) protease [56]. The protein monomer construct has a C-terminal Asn-Gly-Leu motif for OaAEP1b recognition and a N-terminal sequence of Glu-Asn-Leu-Tyr-Phe-Gln-Gly-Leu for TEV cleavage. The initial step is to immobilize the monomeric protein-NGL onto the GL-functionalized glass during OaAEP1b-meidated ligation. After washing off excess protein and OaAEP1b, TEV was used to cleave the masked N-terminus of the protein. By repeating these two steps, a pentameric ubiquitin can be obtained with around 30% yield.

1.2.4. Live cell labelling by PALs

Labelling the cell surface is very crucial for cellular visualization and the study of receptor trafficking. PALs have minimal cytotoxicity towards live cells, therefore can perform live cell labelling on both engineered and non-engineered cells.

Bi *et al.* has modified the surface of E. coli with a fluorophore using butelase 1. In this work, the C terminal of the overexpressed OmpA-NHV was exposed on E.coli cell surface, which acted as the ligase recognition site [57]. In addition, Bi *et al.* also managed to ligate a disulfide FRET probe to the human transferrin receptor 1 on HEK 293T cells to investigate the intracellular redox state (**Figure 1. 6** f) [58].

Red blood cells (RBC) can serve as vascular carriers for bio-active compounds. Harmand et al. used OaAEP1b to modify RBC surfaces with functional proteins of interest [59]. The authors show that, in mice, the surface-engineered RBCs

have a normal half-life and possess desired therapeutic properties (**Figure 1. 6** f) [59].

1.3. Aims of my study

Owing to their high catalytic efficiency, PALs have become a very useful bioconjugation tool. However, their wider applications are still limited by two inherent features of these enzymes. First, PALs have a strict requirement for an Asx at the P1 position, and they prefer a P1-Asn to P1-Asp for acyl donor substrate recognition with the ligation efficiency at P1-Asp being much lower than that at P1-Asn. Second, the reversibility nature of PAL-mediated ligation is the major factor that limits the yield of the reaction. Thus, the general goal of my study was to improve the efficiency of PAL-mediated reaction and to increase the applications of PALs. Specifically, my study aimed at (1) broadening the substrate scope at P1 position of PAL substrates and (2) overcoming the reversibility problem that occurs during PAL-mediated ligation.

Chapter 2. N^{γ}-Hydroxyasparagine: a Multifunctional Unnatural Amino Acid That is a Good P1 Substrate of Asparaginyl Peptide Ligases

2.1. Abstract

Peptidyl asparaginyl ligases (PALs) are powerful tools for peptide macrocyclization. Herein, we report that a derivative of Asn, namely $\mathbf{N^{\gamma}}$-hydroxyasparagine or Asn(OH), is an unnatural P1 substrate of PALs. By Asn(OH)-mediated cyclization, we prepared cyclic peptides as new matrix metalloproteinase 2 (MMP2) inhibitors displaying the hydroxamic acid moiety of Asn(OH) as the key pharmacophore (**Figure 2. 1**). The most potent cyclic peptide (Ki = 2.8 $\pm$ 0.5 nM) was built on the hyperstable tetracyclic scaffold of rhesus theta defensin-1. The Asn(OH) residue in the cyclized peptides can also be readily oxidized to Asp. By this approach, we synthesized several bioactive Asp-containing cyclic peptides (MCoTI-II, kB2, SFTI, and integrin-targeting RGD peptides) that are otherwise difficult targets for PAL-catalyzed cyclization owing to unfavorable kinetics of the P1-Asp substrates. This study demonstrates that substrate engineering is a useful strategy to expand the application of PAL ligation in the synthesis of therapeutic cyclic peptides.

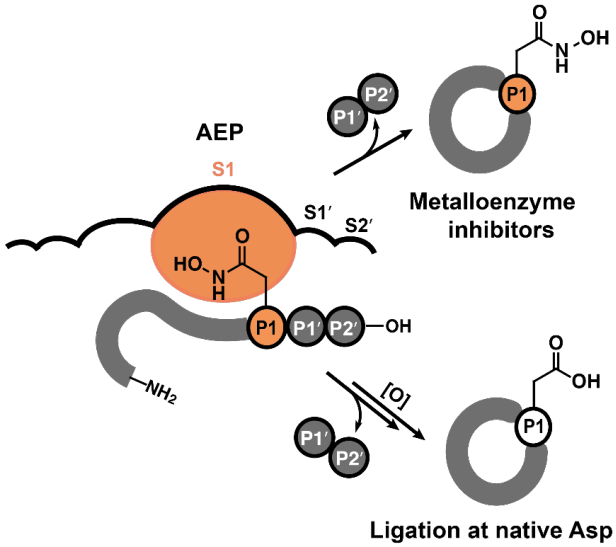


Figure 2. 1. An unnatural amino acid, Asn(OH), mimics P1-Asn in the substrates of asparaginyl ligases.

2.2. Introduction

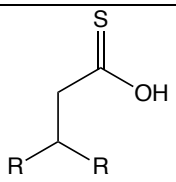
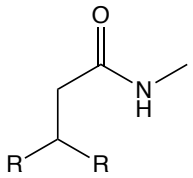
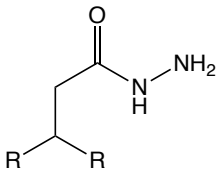
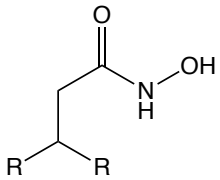
Macrocyclization is an effective strategy to restrict the conformations of linear peptides [60, 61]. The large surface areas of cyclic peptides confer high binding affinity and selectivity to inhibit protein– protein interactions that are often not easily druggable using small-molecule compounds [62-64]. Responsible for cyclotide synthesis in plants, asparaginyl endopeptidases (AEPs) are a useful tool to cyclize synthetic peptides [27, 38, 65, 66]. AEPs are cysteine endopeptidases that typically recognize and cleave peptide bonds after an Asn or Asp residue [31, 43, 67-72]. Some members of the AEP family catalyze transpeptidation at the asparaginyl peptide bonds with little hydrolase activity, which qualifies them as peptidyl asparaginyl ligases (PALs) [27, 35-37, 73]. Most PALs have a strong preference for P1-Asn over P1-Asp substrates. In fact, the rate of cyclization reactions for these two types of substrates can differ by 100 to 800 folds [27, 39]. The optimal pH for the hydrolysis of Asp/Asn–peptide bonds by AEPs is around 4.5–5.5 [35-37]. Structural characterization of AEP–aspartyl peptide inhibitor complexes shows that the hydroxy group of side chain γCOOH of P1-Asp acts as a hydrogen bond donor to bind to a key residue in the S1 pocket of the enzyme [68, 74]. Given the low pKa value of the carboxyl group, an acidic pH is required to maintain the neutral, protonated form needed for enzyme binding [68, 74]. However, at acidic pH, the N-terminal amine of an acyl-acceptor substrate is mostly protonated, making it largely unavailable in a ligation reaction. This dilemma is likely the main reason that limits the rate of PAL-catalyzed ligation at aspartyl bonds. Therefore, most applications of PAL-mediated cyclization reported to date have been limited to P1-Asn substrates [54, 55, 57, 58, 66, 75-77]. Applications of PALs will be expanded if one can broaden their substrate scope.

2.3. Results and discussion

2.3.1. Asn analogues

Based on the previously reported crystal structure of S1 pocket of PAL, we rationally designed a few unnatural Asn catalogues containing peptides and tested PAL's recognition towards these substrates (**Table 2. 1**). Results show only P1-Asn(NH_2) (**Figure S2. 3**) and P1-Asn(OH) (**Table 2. 2**) containing substrate can be recognized by PALs. Nevertheless, Asn(NH_2) is unstable in neutral to basic solution, as we observed the degradation of the Asn(NH_2) containing peptide (Ac-KYSN(NH_2)GL) after incubation in pH 7-8 buffer for 12h at RT (**Figure S2. 4.** Stability of Ac-KYSN(NH_2)GL.). We observed a loss of 32 Da and 16 Da in mass which could be correlated to the N-Succinimidyl ring by-product and hydrolyzed by-product (**Figure S2. 4.** Stability of Ac-KYSN(NH_2)GL.). Therefore, in this study, we will mainly focus on Asn(OH).

Table 2. 1. P1-X analogues.

P1-X Peptide X =	Structure	Synthesis	Recognition by PAL
Asp(S)		X	X
Asn(CH_3)		√	X
Asn(NH_2)		√	√
Asn(OH)		√	√

R = amide bond that connects two amino acids. Asn(CH_3) containing peptide: Ac-KYS**N(NH_2)**GL. Asn(CH_3) containing peptide: AIYRRGRLYRR**N(CH_3)**HV. Asn(OH) containing peptides: AIYRRGRLYRR**N(OH)**HV.

2.3.2. Proposed P1-Asx residue binding towards S1 pocket of PAL

We used the asparagine analog $\mathbf{N^{\gamma}}$-hydroxyasparagine or Asn(OH) as the P1 amino acid of PAL substrates (**Figure 2. 1**). Based on previous articlse and crystal structures of PAL-substrate complex [68, 74], we hypothesized that the $\mathbf{N^{\gamma}}$- hydroxy group has a higher pKa than COOH, therefore it can remain protonated as a hydrogen bond donor at near neutral pH which favors the ligation reaction (**Figure 2. 2**). We show that P1-Asn(OH) peptides are well recognized by butelase-1 and VyPAL2 in the cyclization reaction (**Table 2. 2**, **Figure 2. 1**).

This unnatural Asn-(OH) residue can function as a good chelator for metal cations to inhibit metalloenzymes or be further converted to natural Asp in a mild oxidation reaction (**Figure 2. 1**).

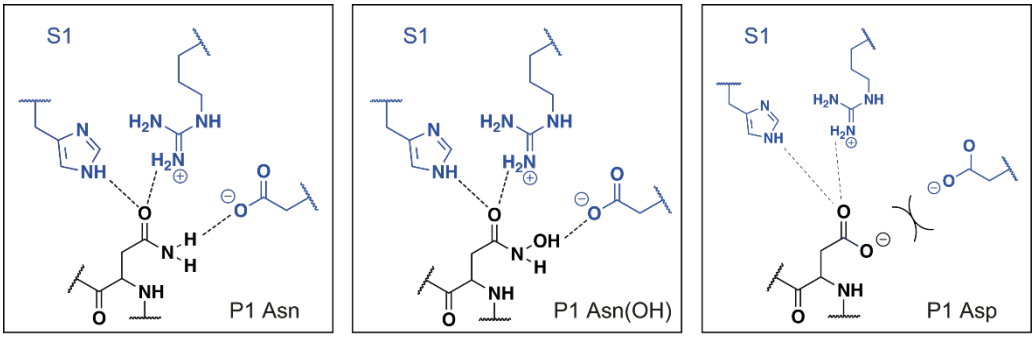


Figure 2. 2. Modeled H-bond interactions within S1 pocket of AEP-substrate complex based on published PAL-substrate complex [68, 74].
Modeled H-bond interactions between the S1 pocket of AEPs and the P1 amino acid of the substrates at neutral pH. Residues in the S1 pocket and P1 amino acids of the substrates are shown in blue and black, respectively.

3.3.3. Kinetic of the cyclization of P1-Asn, P1-Asp and P1-Asn(OH) peptides

Compounds **1–6** (**Table 2. 2**, **Table 2. 3**) were prepared by Fmoc solid-phase peptide synthesis (SPPS) and used as substrates for butelase-1 or VyPAL2-catalyzed cyclization. As expected, the P1-Asn(OH) peptides were good substrates of these PALs and exhibited a much higher binding affinity (K_m) and catalytic turnover (k_{cat}) than the native P1-Asp peptides (**Table 2. 2**). The catalytic efficiency of butelase-1 on the N(OH)HV substrate **2** is only approximately 6-fold lower than that on the NHV peptide **1**, but 48-fold higher than that on the DHV peptide **3**. Similarly, the catalytic efficiency of VyPAL2 on the N(OH)SL peptide **5** is approximately 14-fold lower than that on the NSL peptide **4**, but 28-fold higher than that on the DSL peptide **6** . As a control, the P1-Asn(CH_3) peptide **7** was not recognized by butelase-1 (**Table 2. 2**). Therefore, we conclude that the **N$^{\gamma}$**-hydroxy group of Asn(OH) is involved in H-bonding to the S1 pocket of PALs (**Figure 2. 2**).

Table 2. 2. Kinetics of peptide cyclization by butelase-1 and VyPAL2

Sequence	$k_{cat}(s^{-1})$	K_m (μM)	k_{cat}/K_m (M^{-1}/s^{-1})
AIYRRGRLYRR**N**HV **1** [a]	14.86 ± 0.34	67.5 ± 2.8	220,148
AIYRRGRLYRR**N(OH)**HV **2** [a]	3.46 ± 0.27	98.8 ± 17.8	35,020
AIYRRGRLYRR**D**HV **3** [a]	0.454 ± 0.02	629.7 ± 19.5	721
AIYRRGRLYRR**N**SL **4** [b]	4.56 ± 0.24	42.5 ± 8.83	107,294
AIYRRGRLYRR**N(OH)**SL **5** [b]	0.942 ± 0.082	119.9 ± 19.2	7,857
AIYRRGRLYRR**D**SL **6** [b]	0.581 ± 0.09	2069 ± 407	281
AIYRRGRLYRR**N(CH_3)**HV **7**	NA	NA	NA

[a] Peptides for cyclization by butelase-1; [b] Peptides for cyclization by VyPAL2. The cyclization reactions were performed at 37 ˚C in 20 mM PBS (pH 6.5), with various substrate concentrations. Cyclization was monitored by analytical RP-HPLC. NA = not applicable.

2.3.4. Generation of cyclic Asn(OH) peptides as MMP2 inhibitors

Asn(OH) contains a hydroxamic acid moiety that has interesting properties [78-80],including the ability to bind to metal ions and reactivity toward an S-aryl thio-ester for peptide ligation. A variety of hydroxamic acid-derived inhibitors of metalloenzymes have been designed as pharmacologic agents [78, 81].Our method provides a convenient approach to prepare hydroxamic acid displaying cyclic peptides as potential inhibitors of pathogenic metalloenzymes such as matrix metalloprotease 2 (MMP2) [82]. We used peptide **8**, a β-amyloid precursor protein derived inhibitor peptide (APP-IP) [83-85], to demonstrate the utility of this method (**Figure 2. 4**). APP-IP **8** is an MMP inhibitor and is largely selective for MMP2 (K_i ~ 60 nM) [83]. Asp6 in APP-IP **8** mediates Zn^{2+} bonding. Simple replacement of Asp6 with Asn(OH) resulted in a 2.6-fold improvement in inhibitory activity (**Figure 2. 4** b) that could be attributed to the stronger Zn^{2+}-chelating ability of the Asn(OH) hydroxamate group compared to the carboxylate in Asp [81]. Nevertheless, the linear APP-IP peptide **8** has poor proteolytic

stability and its half-life in human serum is short [85]. Cyclization is a useful strategy to improve the stability and selectivity of MMP2 inhibitory peptides [81]. Therefore, we first generated linear peptides **10-15** which containing ALMP-linker-ISYGQN(OH)AL for PAL-mediated cyclization reaction (see **Figure 2. 3**). We performed head-to-tail cyclization of APP-IP derivative peptides of various lengths using the Asn(OH) ligation method in PBS buffer (20 mM PBS, pH 6.5, 0.5 mM TCEP) at 37 °C. However, the cyclized products, peptides **10–15**, had decreased inhibitory activities toward MMP2 (**Table 2. 3**, **Figure 2. 4** b), likely because simple cyclization restrained the peptides into undesirable conformations. RTD-1 has a highly compact, tetracyclic Cys-ladder structure which is extremely constrained and stable [86]. We grafted Ile1Ser2Tyr3, Leu8 and Asn(OH)6 from APP-IP peptide **9** into rhesus theta defensin-1 (RTD-1) **16** (**Figure 2. 4**) to form Asn(OH)-RTD1 **17**. Asn(OH)-RTD1 **17** was successfully cyclized at the Asn(OH) residue by incubating the linear substrate with butelase-1 (**Figure 2. 4** a). RP-HPLC monitoring indicated that the reaction was nearly completed in 1 h. After enzymatic cyclization, the cyclic product (reduced form) was oxidatively refolded into the mature form by incubation in 0.1 M Tris·HCl buffer (pH 8.5, 2 M urea) at 4 °C for 8 h (**Figure 2. 5**). Enzyme inhibition results showed that the folded backbone-cyclic Asn(OH)-RTD1 peptide **17** had a K_i value of (2.8 ± 0.5) nM toward MMP2 (**Figure 2. 4** b), representing an 8.2-fold improvement over the linear Asn(OH)-APP-IP peptide **9**.

Figure 2. 3. Cyclization of linear peptides **10-15.**
Linear peptides **10-15** were cyclized by treated with 0.001 eq of Butelase 1 for 2-3h at 37 °C in PBS pH 6.5 buffer (0.5mM TCEP).

Table 2. 3. Expected, observed masses and amino acid sequences of peptides derived from APP-IP.

No	Peptide sequence	Calc mass (m/z)	Found Mass $M+H^+$ (m/z)	IC_{50} (μM)	K_i (μM)
8	ISYGNDALMP	1079.2	1080.3	0.092	0.06
9	ISYGNN(OH)ALMP	1094.2	1095.0	0.035	0.023
10	cyclo[ISYGQN(OH)ALMP]	1090.5	1091.5	32.53	21.26
11	cyclo[SISYGQN(OH)ALMPG]	1234.6	1235.6	>100	>100
12	cyclo[SGISYGQN(OH)ALMPSG]	1378.1	1379.6	10.68	6.98
13	cyclo[GSGISYGQN(OH)ALMPSGA]	1506.9	1507.9	18.36	12.00
14	cyclo[ATSGISYGQN(OH)ALMPSGYM]	1845.3	1846.2	3.338	2.18
15	cyclo[ATRSGISYGQN(OH)ALMPSGQYM]	2129.6	2030.6	0.354	0.231

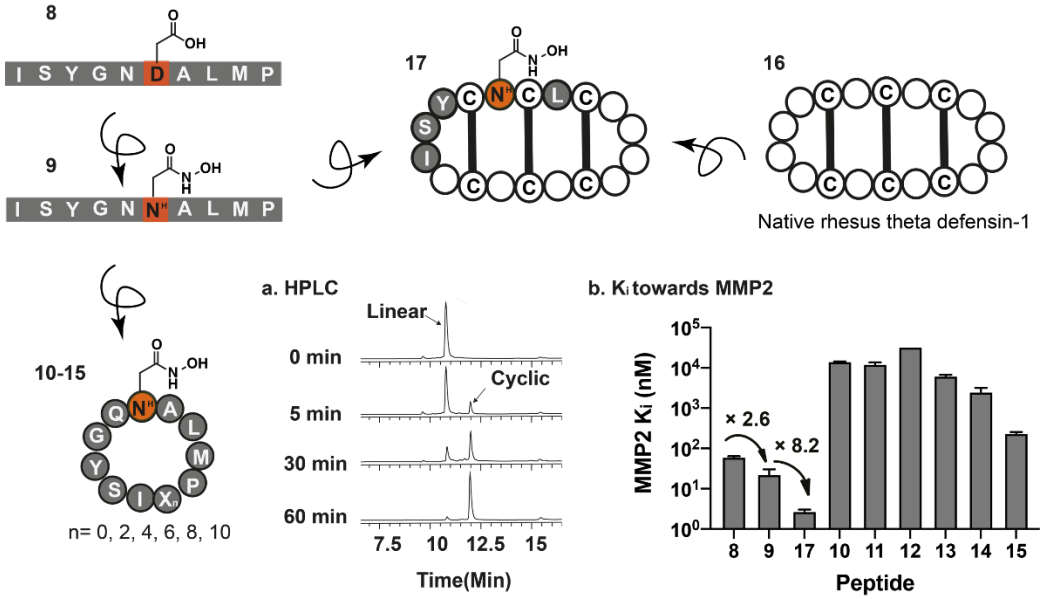


Figure 2. 4. Design of MMP2 peptide inhibitors based on APP-IP and theta defensin. (a) HPLC and ESI-MS monitoring of butelase-1-catalyzed cyclization of the linear precursor CLCRRGVCRCICTISYC**N**(OH)-AL for the preparation of **17** (reduced form); (b) K_i of peptides **8**-**17** towards MMP2.

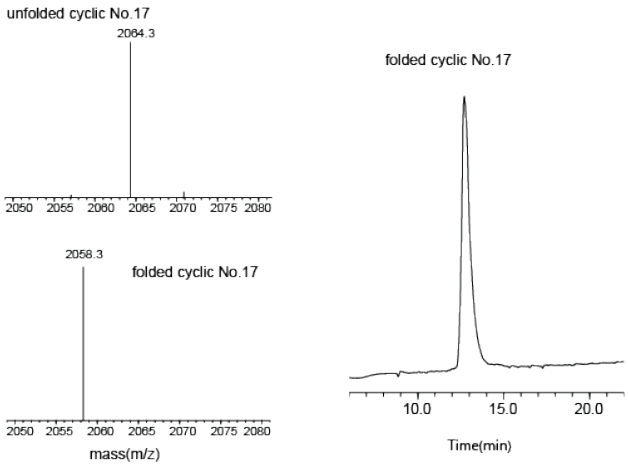


Figure 2. 5. Oxidative folding of peptide No.**17**.
Peptide No. **17** was folded in 0.1 M Tris-HCl (pH 8.5), 2 M urea with GSSG/GSH (10 eq/100 eq), for around 12 h at 4 °C. The folded products were purified by RP-HPLC and characterized by ESI-MS, then was purified and immediately lyophilized and stored at -20 °C.

NMR results showed that peptide 17 has the same 3D structure as RTD-1 **16** (PDB:1HVZ) [86]. The Tyr3 and Asn(OH)5 residues in peptide **17** are constrained in a b-sheet structure with the two side chains pointing parallel in the same direction (**Figure 2. 6**, **Figure 2. 7**, **Table 2. 4**) similar to the orientation adopted by Tyr3 and Asp6 of APP-IP **8** in binding to MMP2 [84]. This preorganized conformation of peptide **17** may have contributed to its improved inhibitory activity. For comparison studies, Asp-RTD1 **18** was obtained by oxidation of Asn(OH)-RTD1 **17** with $NaIO_4$ [87] (1 equiv) in 20 mM PBS (pH 7.2) at 0 °C, which quickly converted Asn(OH) to Asp in 5–10 min. Previously, Pentelute and co-workers showed that $NaNO_2$ oxidation can also be used to convert Asn(OH) to Asp [80]. Asn-RTD1 **19** was prepared in the same manner as Asn(OH)-RTD1 **17** by cyclization at the Asn residue using butelase-1. Although peptides **17**, **18**, and **19** have the same conformation and nearly the same sequences (**Figure S2. 6**, **Table S2. 2**), MMP2 inhibition assay shows that Asn(OH)-RTD1 **17** has a 2.3- and 13-fold improvement in binding affinity over Asp-RTD1 **18** and Asn-RTD-1 **19**, respectively (**Figure S2. 5**, **Figure S2. 6**, **Table S2. 2**). This result further corroborates that Asn(OH) is an excellent pharmacophore for metalloenzyme inhibition. Additionally, all three peptides are more potent MMP2 inhibitors than the native RTD-1 **16** (**Table S2. 2**).

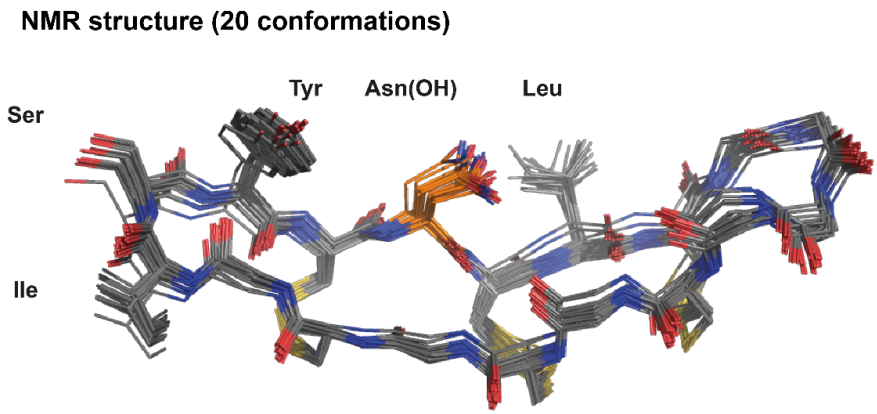


Figure 2. 6. Superposed 20 low-energy structures of peptide **17** (PDB: 7F32) highlighting spatial disposition of the sidechains of residues.

Table 2. 4. Structural statistics for peptide 17

Distance restraints	
Intraresidue (i–j=0)	44
Sequential (|i–j|=1)	52
Medium-range (2≤|i − j|≤4)	32
long-range (|i − j| ≥ 5)	10
Total NOE constraints	138
Angular restraints	
ϕ	15
ψ	15
Distance restraints violations	
No. of violations	22
Average NOE violation (Å)	<0.31
Maximum NOE violation (Å)	<0.48
Average target function values	4.74 (4.65-4.92)
Deviation from mean structure	
Backbone atoms (N, Cα, C') (Å)	0.22 ± 0.09
Heavy atoms (Å)	0.94 ± 0.18
Ramachandran plot analysis	
Residues in the most favorable region (%)	53.3
Residues additionally allowed region (%)	46.7
Residues in the generously allowed region (%)	0
Residues in the disallowed region (%)	0

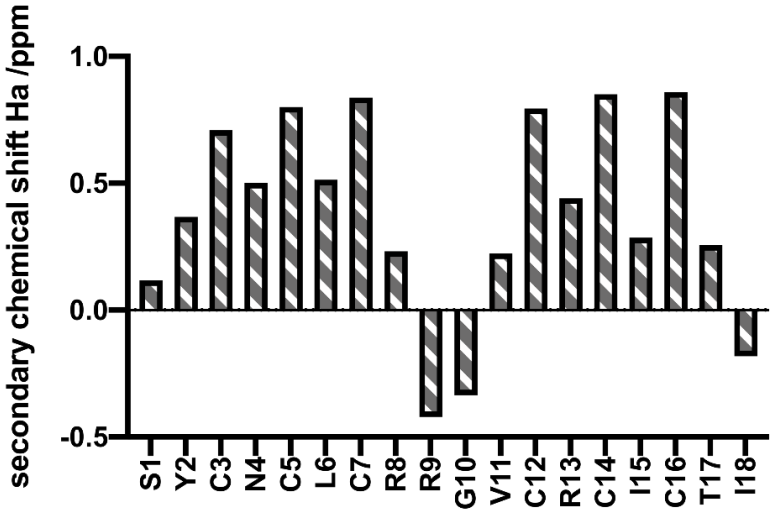


Figure 2. 7. Secondary chemical shifts of H_{α}.
The secondary chemical shift of Asn(OH)-RTD1 **17** shows high similarity with the original RTD1. Using the Hα chemical shifts of random coil as reference, C^3-C^7 and C^{12}-C^{16} regions of Asn(OH)-RTD1 **17** seem to have a well-defined β-sheet structure and are joined by 2 loops [88].

Similar to APP-IP **8**, Asn(OH)-RTD1 **17** is also selective for MMP2 over MMP9 (ca. 1,000-fold difference in K_i) (**Figure 2. 8**). The stability of the linear APP-IP and the cyclic Asn(OH)-RTD1 were examined by incubating the peptides in human serum at 37 ºC. The cyclic peptide **17** remained intact after 16 h, while over 50 % of the linear APP-IP **8** was degraded within 32 min (**Figure 2. 9**). We also did not observe the loss of the $\mathbf{H}^{\boldsymbol{\gamma}}$-hydroxy group on Asn(OH), which is different from what observed by the Heinis group on their bicyclic peptide [81]. This $\mathbf{H}^{\boldsymbol{\gamma}}$ -hydroxy group may be stabilized by the highly constrained conformation of the tetracyclic peptide **17**. The high selectivity and stability of the Asn(OH)- RTD-1 peptide **17** thus makes it a useful pharmacological tool to study the physiological and pathological roles of MMP2.

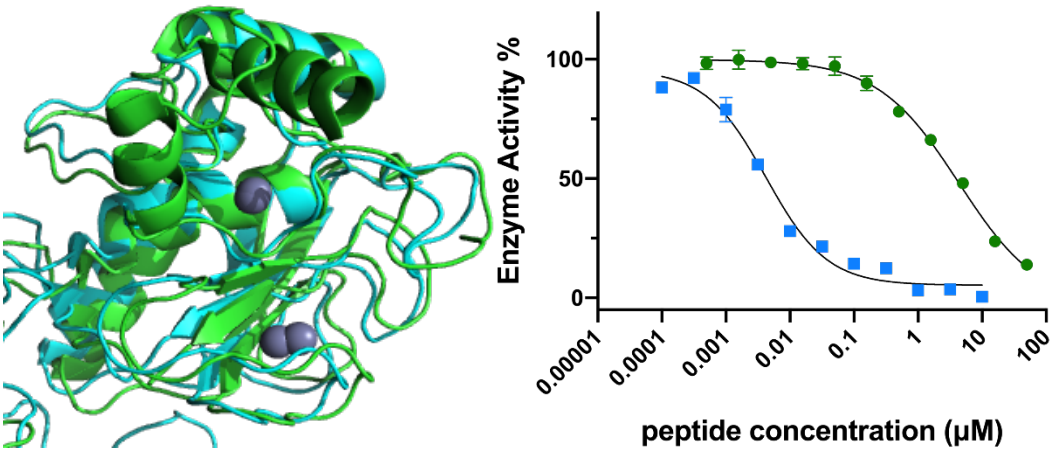


Figure 2. 8. Selectivity of Asn(OH)-RTD-1 17 over MMP2.
Green colour indicates MMP9 and blue colour indicate MMP2.

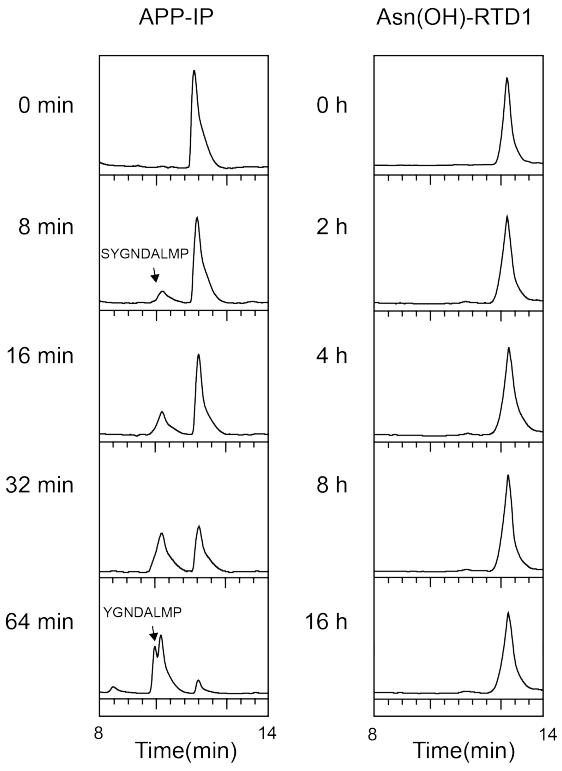


Figure 2. 9. Stability of linear APP-IP **8** and cyclic Asn(OH)-RTD1 **17**.
Both peptides were incubated in human serum at 37 ºC for different time courses and analysed by RP-HPLC. The identities of the remaining peptides were verified by ESI-MS.

2.3.5. Generation of Asp containing peptides

Facile conversion of Asn(OH) to Asp makes Asp cyclization more attainable through PAL ligation, albeit indirectly. To show the general applications of this methodology, we further performed cyclization of P1-Asn(OH) peptides of 6–34 amino acids, followed by $NaIO_4$-mediated oxidation of Asn(OH) to Asp (**Table 2. 5**). The first peptide prepared was MCoTI-II. Isolated from *Momordica cochin-chinensis* seeds, MCoTI-II is a potent trypsin inhibitor [89, 90]. Cyclization of MCoTI-II has been achieved using various methods [90-95]. Linear MCoTI-II-N(OH)IV **20** was prepared by SPPS and oxidatively folded. Backbone cyclization of acyclic, folded MCoTI-II-N(OH)IV **20** was done by incubation with butelase-1 (0.01 equiv) for 3 h at 37 ºC (**Figure 2. 10** a, b). After enzymatic cyclization, the unnatural Asn(OH) residue was then converted to Asp in 10 min by $NaIO_4$ oxidation at 0 ºC (**Figure 2. 11**). To study the folding states, CD spectra for the linear MCOTI-II, acyclic MCoTI-II (folded), and cyclic MCoTI-II (folded) were collected (**Figure 2. 10** c). The spectra for acyclic and cyclic MCoTI-II (both folded) are comparable to the reported CD spectrum [96, 97].

Table 2. 5. Peptide cyclization at the P1 site (residue in bold) and oxidation of Asn(OH) to Asp.

Peptide	Sequence	Cyclization time (h)	Yield (%)	Oxidation time (min)	Yield (%)
MCoTI-II-(OH)IV **20**	GGVCPKILKKCRRDSDCPGACICRGNGYCGSGS**N(OH)**IV[a]	3	95	10	>95
MCoTI-II-DIV **21***	GGVCPKILKKCRRDSDCPGACICRGNGYCGSGS**D**IV[a]	3	ND	NA	NA
kB2-N(OH)IV **22**	GLPVCGETCFGGTCNTPGCSCTWPICTR**N(OH)**IV[b]	1	80	10	>95
kB2-DIV **23**	GLPVCGETCFGGTCNTPGCSCTWPICTR**D**IV[b]	2	ND	NA	NA
SFTI-N(OH)IV **24**	GRCTKSIPPICFP**N(OH)**IV[a]	3	85	10	>95
SFTI-DIV **25**	GRCTKSIPPICFP**D**IV[a]	3	ND	NA	NA
RGD-N(OH)HV **26**	FLaRG**N(OH)**HV	1	85	10	>95
RGD-defensin-N(OH)SL **27**	ACRCLCRRGDCRCICRG**N(OH)**SL[b]	1	95	15	>95

Backbone cyclization was conducted [a]after and [b]before oxidative refolding. The reason to cyclize peptide **22**, **23**, and **27** before oxidative folding is because PAL cannot recognize/cyclize the oxidative folded structure of peptide **22**, **23**, and **27**. Cyclization reactions were performed at 37 ˚C using 400 μM of peptide substrates and 0.01 eq of butelase-1 or VyPAL2 in 20 mM PBS (pH 6.5) for 1-3 h. Peptide **27** was cyclized by VyPAL2 and the others were cyclized by butelase-1. *Cyclization was also conducted at pH 4.5 or 6.5 at room temperature for overnight, and no cyclic product was observed. Asn(OH) oxidation was performed in 20 mM PBS (pH 7.4) at 0 ˚C, with 1eq of $NaIO_4$. ND = not detectable; NA = not applicable.

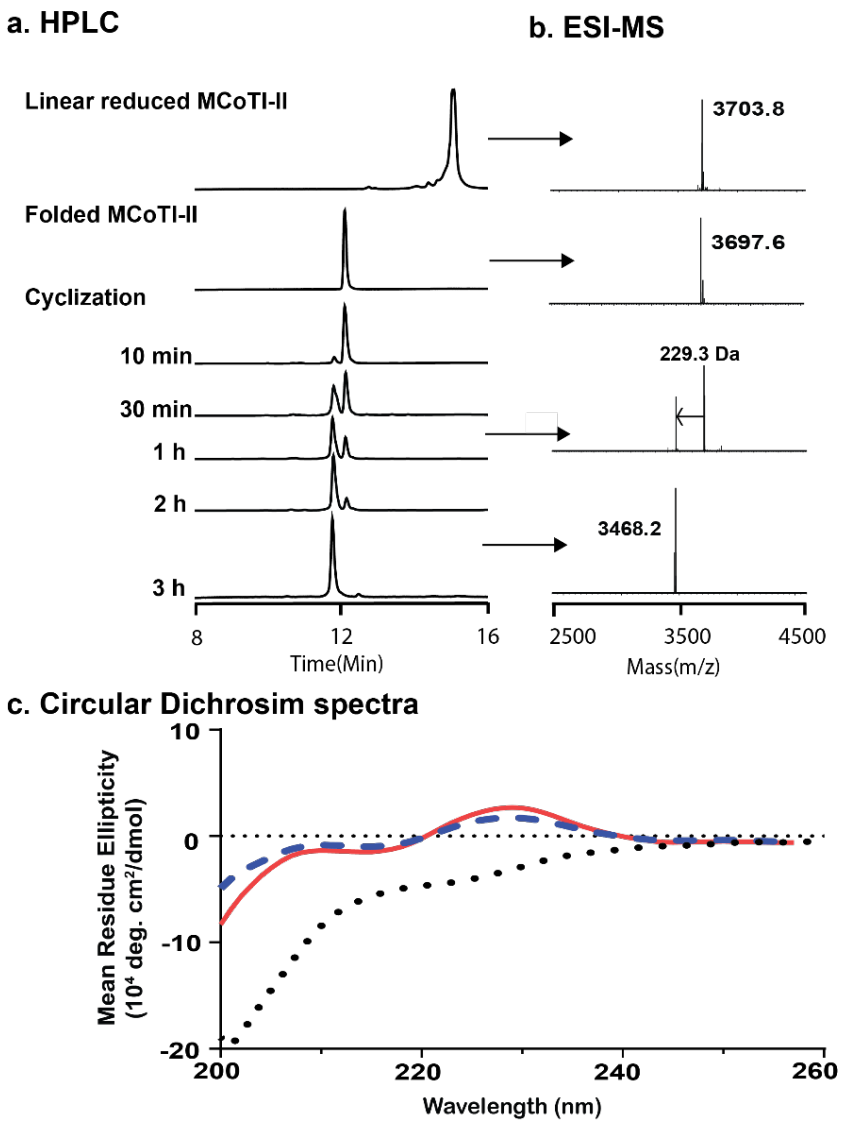


Figure 2. 10. Butelase-1 mediated cyclization of MCoTI-II **20**.
(a) RP-HPLC trace (10% buffer B to 50% buffer B in 20 mins, UV absorption at 220 nm) and (b) ESI-MS for linear, refolded acyclic and refolded cyclic MCoTI-II-N(OH)IV. The cyclization reaction was performed at 37 ºC using 400 µM of peptide substrate and 0.01 eq of butelase-1 in 20 mM PBS (pH 6.5) for a 3-h time frame. (c) Circular dichroism (CD) spectra of linear reduced MCoTI-II (black dot), refolded acyclic MCoTI-II (Blue dash), and cyclic MCoTI-II (red line).

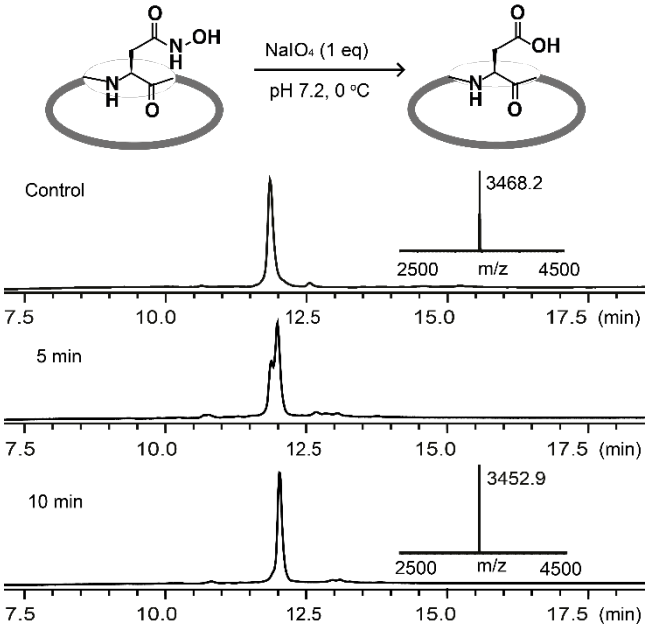


Figure 2. 11. Converting Asn(OH) to Asp by $NaIO_4$-mediated oxidation.
RP-HPLC and ESI-MS monitoring of the oxidation reaction of Asn(OH)-MCoTI-II.

Butelase-1 cyclized P1-Asn(OH) peptides (**20**, kB2 **22** [98], and SFTI **24** [99]) within 1–3 h at 37 °C despite a sterically hindered Asn(OH)-Ile bond (**Table 2. 5**, **Figure S2. 8**, **Figure S2. 9**, **Figure S2. 10**, **Figure S2. 11**, **Figure S2. 13**). In contrast, direct cyclization of the corresponding P1-Asp peptides **21**, **23** and **25** by butelase-1 was extremely inefficient as cyclic products were not detectable after 3–24 h (**Figure S2. 7**, **Figure S2. 8**, **Figure S2. 9**, **Figure S2. 10**). This is most likely because, 1) Asp is a poor P1-amino acid, 2) Ile is a bulky P1' residue and 3) the N-terminal motifs in **21** and **25** (GG- and GR-) are unfavorable nucleophilic substrates of butelase-1. Nevertheless, all these problems were overcome by the higher reactivity of P1-Asn(OH) during cyclization.

We also prepared cyclic peptides containing RGD (for integrin targeting) [100, 101] using Asp as the cyclization site (**Table S2. 2**). The RGD-theta-defensin linear precursor **27** was cyclized in 1 h by VyPAL2 (**Figure S2. 17**). Interestingly, FLaRGN(OH)HV **26** was converted to a cyclic dimer in 1h, likely because

the high angle strain prevented the formation of the small six-residue ring (**Figure S2. 15**), similar to a previous finding by Hemu et al [49].

Next, trypsin inhibition assays were performed (**Table 2. 6**, **Figure S2. 19**). The K_i values of the synthetic MCoTI-II and SFTI toward trypsin were 0.18 and 0.44 nM, respectively, which are consistent with reported values [91, 92, 102] Therefore, both cyclic trypsin inhibitor peptides prepared by our method have the expected biological activities.

Table 2. 6. Trypsin inhibitory activity (IC_{50} and K_i) of MCoTI-II **20** and SFTI **24**.
All triplicated data are presented as means ± SEM.

Inhibitor	**Peptide sequence**	**IC_{50} (nM)**	**K_i (nM)**
[MCoTI-II]	[GGVCPKILKKCRRDSDCPGA CICRGNGYCGSGSD]	2.01 ± 0.28	0.18 ± 0.03
[SFTI]	[GRCTKSIPPICFPD]	4.85 ± 0.23	0.44 ± 0.02

2.4. Conclusion

In conclusion, we have used an unnatural amino acid, Asn(OH), as the P1 substrate for PAL-catalyzed peptide cyclization. The P1-Asn(OH) residue is an excellent mimic of P1-Asn in PAL substrates, and the P1-Asn(OH) peptides exhibit a much higher affinity and turnover rate than the P1-Asp peptides. A wide range of cyclic peptides were prepared to demonstrate the utility of our method for P1-Asn(OH)-enabled cyclization and oxidative Asn(OH)-to-Asp conversion. The structure and bioactivity of these peptides were confirmed by CD spectroscopy, NMR and enzyme inhibition assays. More importantly, we demonstrate that cyclization of Asn(OH)-containing peptides is well-suited for generating inhibitors that have high potency and selectivity toward MMP2 as well as proteolytic stability. Taken together, the engineered Asn(OH) serves three functions: 1) mediator of PAL-catalyzed ligation and a site of peptide cyclization, 2) a chelator for metal ions to inhibit metalloenzymes, and 3) an Asp precursor. Our study shows that substrate engineering can expand the substrate scope of PALs and increase the therapeutic value of the cyclized peptides.

2.5. Materials and methods

2.5.1. General methods

All the solvents and reagents were purchased from commercial suppliers and used without further purification. The synthesis of Fmoc-Asn(OTrt)-COOH and Fmoc-Asn(Me)-COOH were done under a positive pressure of N_2 atmosphere. Evaporations of solvents were performed using a Buchi rotary evaporator. The completion of the reactions was monitored using 0.25 cm silica thin-layer chromatography (TLC). Since the Asn compounds contain the Fmoc group, the spots on TLC can be visualised under UV light. The synthesized Asn compounds were purified using 200 mesh silica column chromatographs, eluting with ethyl acetate/hexanes or MeOH/DCM solvent system. Peptides were synthesized following standard Fmoc solid phase synthesis protocols. Synthesized peptides were purified using semi-preparative RP-HPLC. Semi-preparative RP-HPLC

was preformed using a Shimadzu HFLC system equipped with a Phenomenex jupiter-C18 RP column (10 × 250 mm, 5 μm) with a flow rate of 2.5 mL per minute, eluting using a gradient of buffer B (90% acetonitrile, 10% H_2O, 0.045% TFA) in buffer A (H_2O, 0.045% TFA). All the synthesized compounds were stored at 4 ºC or -20 ºC.

For analysis, ^{1}H and ^{13}C NMR spectra were recorded using a Bruker 400 MHz spectrometer. Samples were dissolved in deuterated solvents and the spectra were obtained at 298 K. All chemical shifts were quoted in ppm and coupling constants were measured in Hz. Mass spectra for peptides were obtained using a Bruker Ultraflex Extreme Matrix Assisted Laser Desorption/Ionization (MALDI) Tandem TOF or electrospray ionization (ESI) mass spectroscopy (Thermo Fisher LTQ XL). Data from MALDI was analysed using Data Explorer software and data from ESI was analysed using Thermo Xcalibur Qual Browser software. For characterization of Asn compounds, HRMS (ESI) were recorded using a 6520 QTOF mass spectrometer, equipped with a dual-spray electrospray ionization source (Agilent Technology, Santa Clara, CA). Analytical reverse-phase HPLC (RP-HPLC) was performed on a Shimadzu HPLC system equipped with a Phenomenex jupiter-C18 RP column (4.6 × 250 mm, 5 μm) with a flow rate of 1.0 mL per minute, eluting with a gradient of buffer B (90% ACN, 10% H_2O, 0.045% TFA) in buffer A (H_2O, 0.045% TFA).

2.5.2. Solid phase peptide synthesis (SPPS)

All the peptides were synthesized as C-terminal amides using Rink amide MBHA resin or 2-chlorotrityl chloride resin by standard Fmoc chemistry. Before use, the resin was pre-swelled in DCM for 20 min. The resin was then washed with DMF, DCM and DMF successively. The appropriate Fmoc-protected amino acids (4 eq) and PyBop (4 eq) were dissolved in DMF/DCM (50%/50%). DIEA (8 eq) was added into the mixture to activate the Fmoc-amino acid. The solution was added to the resin and shaken for one hour at room temperature. The resin was then washed with DMF (3 times), DCM (3 times), successively. The completion of

coupling was tested using the Ninhydrin reagent. To deprotect Fmoc, the resin was treated with 20% piperidine in DMF (2 × 10 min). The resin was then washed with DMF (3 times), DCM (3 times), successively. For peptide cleavage from resin and deprotection of sidechain protecting groups at the end of SPPS, the peptidyl-resin was treated with a cocktail of TFA/H_2O/TIS (95%/2.5%/2.5%) and the mixture was shaken for 1-3 hours. For Cys-containing peptides, TFA/H_2O/TIS/EDC (92.5%/2.5%/2.5%/2.5%) was used for cleavage. The cleavage solution was separated from the resin by filtration and the cleaved peptide was precipitated in the cold Et_2O (20-30 mL). The crude product was isolated by centrifugation and purified by RP-HPLC. The peptide fractions after HPLC purification were lyophilized to afford the peptide in powder form. For Asn(NH_2) containing peptides, Ac-KYSD(Odmab)(Dmb)GL was synthesised using standard Fmoc solid-phase synthesis protocol. Odmab was removed by 5% NH_2NH_2 in DMF (3 x 5min), then PyBop(1.2eq), DIEA(1.2eq) and NH_2NH_2 (10eq) in DMF was added into the resin and the reaction mixture was shaken for 2 X 3h, followed by a ON coupling. The peptide was cleaved from resin by incubating in 100% TFA for 1h with shaking.

2.5.3. Conditions of enzyme-mediated reactions

Enzyme-meditated cyclization and ligation reactions were performed in 20 mM PBS buffer (pH 6.5) at 37 ^oC with 0.5 mM TCEP. The reaction was quenched by adding 20% TFA. The enzyme concentration was determined by the absorbance at 280 nm using Nanodrop. The reactions were monitored by analytical RP-HPLC, and the products were characterized by MALDI-MS or ESI-MS (see **Figure S2. 8-2. 11, S2. 13, S2. 15, and S2.17**)

2.5.4. Kinetic studies

Model peptides (**1**-**6**) were used in the kinetic studies. The cyclization reaction was performed in 20 mM PBS buffer (pH 6.5) containing 0.5 mM TCEP, at 37 ^oC. The reaction was monitored by analytical RP-HPLC (5% buffer B- 45%

buffer B in 20 min). Kinetic constants for these peptides were obtained from the initial reaction rates. To ensure that the measurement of reaction rate was in the range of the initial velocity, the reaction was quenched using 20% TFA, when less than 25% of the substrate has been converted. Each recorded value is an average of measurements in triplicate. All the numbers were input into GraphPad Prism (GraphPad Software, San Diego) to obtain the Michaelis-Menten curve and calculate the kinetic parameters (k_{cat} and K_m).

2.5.5. PAL-mediated cyclization of peptide 1-15, 17-19

Linear form of peptides **1-15** and **17-19** containing C-terminal N(OH)AL were synthesized using SPPS. The cyclic form of peptides **1-15** and **17-19** were generated through PAL-mediated reaction in PBS buffer (20 mM PBS, pH 6.5, 0.5 mM TCEP) at 37 °C using 0.04-0.001 eq of Butelase 1.

2.5.6. Chemical cyclization of RTD-1 16

The RTD-hydrazide peptide was synthezied by SPPS method on 2-chlorotrityl resin. The synthetic RTD-hydrazide peptide was treated with 2 eq of $NaNO_2$ in 6 M Guanidine·HCl buffer (pH 3) at -20 °C for 10 min. 100 eq of MESNa in PBS buffer (pH 7) was added to the reaction mixture, and the pH was adjusted to pH 7 [103]. After leaving the solution for 30 min at RT, 10 mM TCEP was added in the solution and the thiolester was purified by RP-HPLC. The thiolester peptide was then dissolved in a PBS buffer (20 mM PBS, pH 7, 1 mM GSH) to a final concentration of 25 µM. The solution was left for 16 h to afford the cyclic and folded RTD-1 **16**.

2.5.7. Oxidative refolding

Peptides were folded in 0.1 M NH_4CO_3 buffer (pH 8) with 10% DMSO and GSSG/GSH (10 eq/100 eq), for around 8-12 hours at 4 °C MS (see **Figure S2. 12, S2. 14, S2. 16, S2. 18**). The Asn(OH)-RTD1, Asp-RTD1 and Asn-RTD1 peptides were folded in 0.1 M Tris-HCl (pH 8.5), 2 M urea with GSSG/GSH (10 eq/100 eq), for around 12 h at 4 °C. The folded products were purified by RP-HPLC and

characterized by ESI-, then were purified and immediately lyophilized and stored at -20 °C.

2.5.8. Converting Asn(OH) to Asp

$NaIO_4$ (1-2 eq) was added into a 20 mM PBS buffer (pH 7.4) containing 200 μM of Asn(OH) peptide and 10 eq of methionine was added as a scavenger to prevent over-oxidation of sensitive residues. The reaction mixture was left on ice for 5-10 min, then subjected to purification by RP-HPLC. The completion of the reaction was confirmed by peak shift on analytical RP-HPLC and mass spectra analysis. The product was afforded as a white powder after lyophilization, and was stored in -20 °C.

2.5.9. NMR measurement and structure calculation

The NMR of the Asn(OH)-RTD1 cyclic peptide was measured in H_2O/ D_2O (90:10), pH ~4, at a concentration of 0.6 mM. Two-dimensional ^1H-^1H TOCSY (mixing time: 70ms) and ^1H-^1H NOESY (mixing time: 300 ms) were acquired on a Bruker Avance-600 spectrometers at 298K. 10 μM Sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used as an internal reference. Sequence specific resonance assignments were achieved by combined analyses of ^1H-^1H TOCSY and NOESY spectra using NMRFAM sparky [104]. Prediction of the torsion angles was performed on a web-base server PREDITOR [105]. The distance constraints were obtained from intensity of NOE cross peaks. Based on the distance and the torsion angle constraints, structure calculations were carried out using CYANA [106]. Twenty lowest-energy structures were selected for further analyses. Note, residue Asn was used for NMR structure calculation.

2.5.10. Measurement of Circular Dichroism (CD) spectra

The samples for CD measurement were confirmed to be over 95% pure by HPLC. Around 0.1 mg/mL of peptides were dissolved in 20 mM PBS buffer (pH 7.4), placed in a cell with 1 mm path length. The CD spectra were recorded from 200 nm (190 nm) to 260 nm at 25 °C. The final spectra were obtained by averaging 3

independently measured CD spectra of the same sample and then subtracting the CD spectrum of pure buffer.

2.5.11. MMP inhibition assay

The inhibitory activity of the peptides was determined by measuring the MMPs' protease activity towards a FRET substrate. 0.8 nM MMP2 or MMP9 (Sigma-Aldrich), 10μM substrate (FAM-Lys-Pro-Leu-Gly-Leu-Lys(Dabcyl)-Ala-Arg-NH2, λex 490 nm; λem 520 nm) and various concentrations of the peptides (0.0001 μM - 10 μM or 0.001 μM - 100 μM) were used. The reactions were performed in 50 mM Tris-HCl buffer (pH 7.5) containing 150 mM NaCl, 10 mM $CaCl_2$, 0.05% Brij35, at 37 °C. MMP2 was pre-incubated with the peptide inhibitors for 10 min before adding the substrate. All the numbers were input into GraphPad Prism (GraphPad Software, San Diego) for IC_{50} calculation. The inhibition constants (K_i) were calculated according to the equation: $K_i = IC_{50}/(1+[S]_o/K_m)$, where IC_{50} is the peptide concentration when 50% of the enzymatic activity is inhibited, $[S]_0$ is the initial substrate concentration (10 μM), and K_m is the Michaelis-Menten constant. MMP2 has a K_m of 18.83 ± 4.58 μM, and MMP9 have a K_m of 16.23 ± 0.8 μM towards the FRET substrate.

2.5.12. Trypsin inhibition assay

The inhibitory activity of MCoTII and SFTI was determined by measuring trypsin's protease activity towards a FRET substrate. 0.1 nM trypsin, 12 μM substrate [GISTGPE(Edans)RVNSL-QQLVRK(Dabcyl)G-NH2, λex 360 nm; λem 490 nm] and various concentrations of the peptides (0.01 nM - 500 nM) were applied. The reactions were performed in 50 mM Tris-HCl (pH 8, 100 mM NaCl) solution at 37 °C. Trypsin was pre-incubated with the peptide inhibitors for 10 min before adding the substrate. All the numbers were input into GraphPad Prism (GraphPad Software, San Diego) for IC_{50} calculation. The inhibition constants (K_i) were calculated according to the equation: $K_i = IC_{50}/(1+([S]_o/K_m)$, where IC_{50} is the peptide concentration when 50% of the enzymatic activity is inhibited, $[S]_0$ is the initial substrate concentration (10 μM), and K_m is the

Michaelis-Menten constant of the FRET substrate, which is 1.19 ± 0.23 μM.

2.5.13. Plasma stability determination by HPLC analysis

Peptides were mixed with human serum to a final concentration of 100 μM and incubated at 37 ºC. Aliquots of the sample were taken out at different time points and 10% TFA was added to quench all protease activities, followed by addition of 10% of DMSO. The solution mixture was centrifuged (11,000 g, 5 min, 4 °C) and the supernatant was analysed by analytical RP-HPLC (15% buffer B-65% buffer B in 25 min) and ESI-MS. The quantity of the remaining peptides was determined by comparing their HPLC peak area with the initial HPLC peak area.

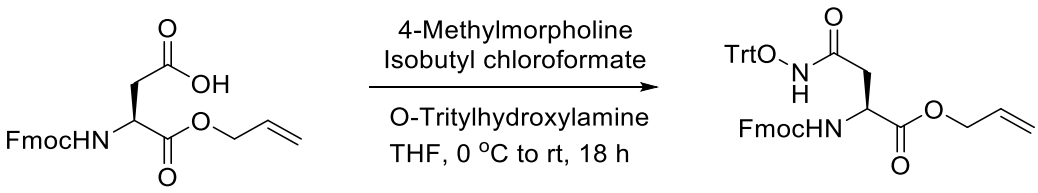
2.5.14. Butelase-1 extraction

Plant flowers of *Clitoria ternatea* grown in the Nanyang Technological University Herb Garden were collected. Butelase-1 was extracted using a chromatography method described previously [27]. Purified butelase-1 was obtained and stored at 4 ºC or -80 ºC in 50 mM sodium phosphate buffer (pH 6.0) containing 1 mM EDTA, 5 mM β-mercaptoethanol, 0.01% Tween 20 and 20% sucrose.

2.5.15. VyPAL2 expression and activation

VyPAL2 was expressed using sf9 insect cells as described previously [36]. 100 mL of the viral vector containing VyPAL2 gene was used to infect sf9 cells at cell density of 2.5 x 10^6 cells/mL. MOI for infection was set between 1-10 for protein expression. The culture was incubated at 27 °C shaker for 3 days (72 hours) at 135 rpm. Protein purification was performed in three steps: Immobilized Metal Affinity Chromatography (IMAC), Ion-Exchange Chromatography (IEX), and Size-Exclusion chromatography (SEC). Pro-VyPAL2 was activated at pH 5 in 50 mM sodium citrate buffer (0.1 M NaCl, 1mM DTT, 0.5 mM LS) for 2-3 h at 37 ºC. After activation, the activated enzyme was purified by SEC at pH 6.5 (20 mM PBS, 0.1 M NaCl, 1mM DTT).

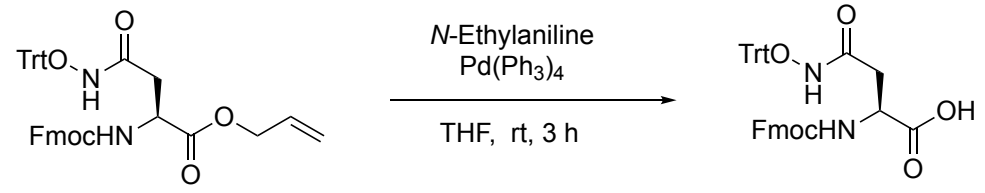
2.6. Appendices



a

(*S*)-3-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-4-(allyloxy)-4-oxobutanoic acid (a): 4-Methylmorpholine (0.72 mL, 5.57 mmol) was added to a solution of Fmoc-Asp-OAll (2.0 g, 5.06 mmol) in anhydrous THF (20 mL), followed by isobutyl chloroformate (0.62 mL, 5.57 mmol) at 0 °C. The solution was stirred at 0 °C for 10 min and O-tritylhydroxylamine (1.39 g, 5.06 mmol) was added. Reaction mixture was allowed to stir at rt for 18 h under N_2 at room temperature. The reaction was quenched by saturated aqueous NH_4Cl (30 mL) and the reaction mixture extracted with ethyl acetate (3 x 30 mL). The combined organic layer was dried over anhydrous Na_2SO_4, evaporated under vacuum. The crude product was purified using column chromatography (10% EA to 40% EA in Hexanes) to afford compound **a** (2.7 g, 85%) as off-white solid.

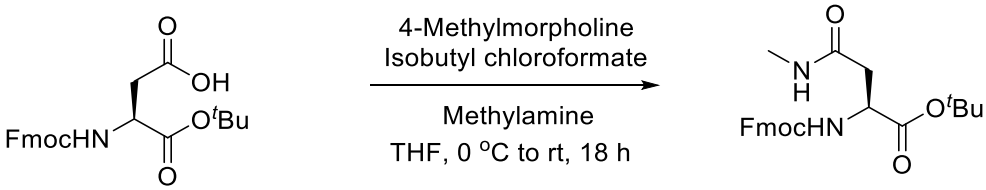
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.80 (d, *J* = 7.6 Hz, 2H), 7.64 (m, 2H), 7.44 (d, *J* = 7.5 Hz, 2H), 7.35 (m, 15H), 5.88 (m, 1H), 5.82 (dd, *J* = 26.1 Hz, 13.6 Hz, 2H), 4.61 (m, 2H), 4.47 (m, 1H), 4.40 (m, 1H), 4.30 (m, 2H), 2.44 (d, *J* = 15.6 Hz, 1H), 1.95 (d, *J* = 17.1 Hz, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 174.1, 170.7, 155.9, 143.8, 141.4, 141.3, 140.8, 131.8, 129.1, 128.3, 127.8, 127.1, 125.2, 120.0, 118.4 93.9, 67.2, 66.1, 49.5, 47.2, 34.6; HRMS(ESI): *m/z* calculated for $C_{41}H_{36}N_2NaO_6$ $[M+Na]^+$ 675.24711, founded 675.24702.



a b

N^2-(((9H-Fluoren-9-yl)methoxy)carbonyl)-N^4-(trityloxy)-L-asparagine (b):

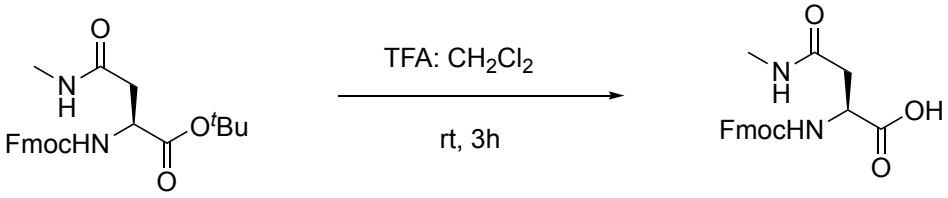
N-Ethylaniline (1.62 mL, 12.9 mmol) was added into a solution of Fmoc-Asn(OTrt)-All (2.7 g, 4.3 mmol) in anhydrous THF (20 mL) and degassed with N_2 for 10 min. $Pd(PPh_3)_4$ (0.35 g, 0.43 mmol) was added and the reaction mixture was stirred for 3 h under N_2 at room temperature. The reaction was quenched using H_2O (10 mL) and the reaction mixture extracted with ethyl acetate (3 x 20 mL). The combined organic layer was dried over anhydrous Na_2SO_4, evaporated under vacuum. The crude product was precipitated in diethyl ether to afford compound **b** (2.0 g, 77%) as pale-yellow solid. HRMS(ESI): *m/z* calculated for $C_{38}H_{32}N_2NaO_6$ $[M+Na]^+$ 635.21581, founded 635.21545. Melting point: 127.8 °C -130 °C.



c

***tert*-Butyl N^2-(((9H-fluoren-9-yl)methoxy)carbonyl)-N^4-methyl-L asparaginate (c):** 4-Methylmorpholine (0.15 mL, 1.34 mmol) was added to a solution of Fmoc-Asp-O^{t}Bu (0.50 g, 1.22 mmol) in anhydrous THF (20 mL), followed by isobutyl chloroformate (0.17 mL, 1.34 mmol) at 0 °C. The solution was stirred at 0 °C for 10 min and methylamine in H_2O (1 mL, 3.66 mmol) was added. Reaction mixture was allowed to stir at rt for 18 h under N_2 at room temperature. The reaction was quenched by saturated aqueous NH_4Cl (30 mL) and the reaction mixture extracted with ethyl acetate (3 x 30 mL). The combined organic layer was dried over anhydrous Na_2SO_4, evaporated under vacuum. The

crude product was purified using column chromatography (10% EA to 40% EA in Hexanes) to afford compound **c** (0.42 g, 81%) as off-white solid. ^{1}H NMR(400 MHz, $CDCl_3$) δ 7.77 (d, *J* = 7.13 Hz, 2H), 7.62 (d, *J* = 6.48 Hz, 2H), 7.41 (d, *J* = 7.13 Hz, 2H), 7.32 (d, *J* = 6.48 Hz, 2H), 4.48 (m, 1H), 4.37 (m, 2H), 4.23 (t, *J* = 6.84 Hz, 1H), 2.89 (m, 1H), 2.78 (d, *J* = 3.87 Hz, 3H), 2.71 (m, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.5, 170.1, 156.3, 143.9, 143.8, 141.28, 127.7, 127.1, 125.2, 120.0, 82.4, 67.1, 51.5, 47.1, 37.9, 27.9, 26.3; HRMS(ESI): *m/z* calculated for $C_{24}H_{29}N_2O_5$ $[M+H]^+$ 425.20765, founded 425.20680.



c d

<u>N^2-(((9H-Fluoren-9-yl)methoxy)carbonyl)-N^4-methyl-L-asparagine (d):</u>

Fmoc-Asn(Me)-O^{t}Bu (0.42 g, 0.99 mmol) was dissolved in 1:1 CH_2Cl_2/TFA and stirred for 3 h. Reaction mixture was evaporated and the crude solid was washed with diethyl ether/Hexane to afford compound **d** (0.35 g, 95%) as off-white solid. HRMS(ESI): *m/z* calculated for $C_{20}H_{21}N_2O_5$ $[M+H]^+$ 369.13722, founded 369.14514. Melting point: 160.1 ºC – 160.6 ºC.

Table S2. 1. Expected, observed masses and amino acid sequences of peptides synthesized for enzyme kinetic studies.

Model peptides	Sequence	calculated mass (m/z)	Founded mass $[M+H^+]$ (m/z)
1	AIYRRGRLYRRNHV	1830.1	1830.9
2	AIYRRGRLYRRN(OH)HV	1846.1	1846.5
3	AIYRRGRLYRRDHV	1831.1	1831.3
4	AIYRRGRLYRRNSL	1794.1	1793.9
5	AIYRRGRLYRRN(OH)SL	1810.1	1810.7
6	AIYRRGRLYRRDSL	1795.1	1794.6
7	AIYRRGRLYRRN(Me)HV	1844.1	1844.1

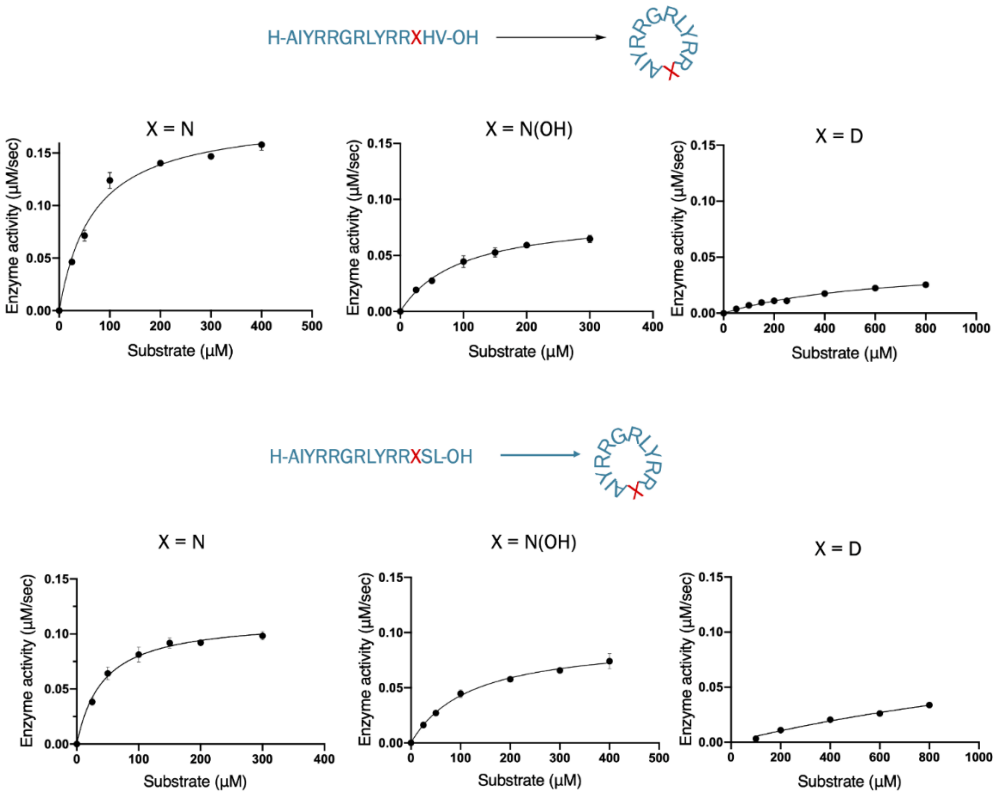


Figure S2. 1. Kinetics of the cyclization of peptides **1-6** by the catalysis of butelase-1 or VyPAL2.

The cyclization reactions were performed at 37 °C in the presence of various concentrations of the peptide substrates in 20 mM PBS (pH 6.5) for 10 min. peptides **1-3** were cyclized by butelase-1 and peptides **4-6** were cyclized by VyPAL2. The amount of the ligase used was 12.5 nM (for peptide **1**), 25 nM (for peptide **2** and **4**), 100 nM (for peptide **3** and **5**) or 200 nM (for peptide **6**). The reactions were monitored by analytical RP-HPLC. All data are presented as means ± SEM.

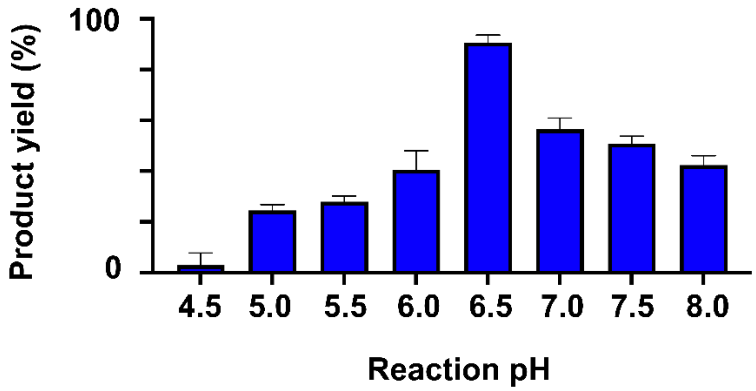


Figure S2. 2. Ligase activity of Butelase-1 towards P1-Asn(OH).
200 μM peptide **2** was treated with butelase-1 (0.001 eq) under pH 4.5-8 at 37 ºC for 12 min. The reactions were monitored by analytical RP-HPLC. All data are presented as means ± SEM.

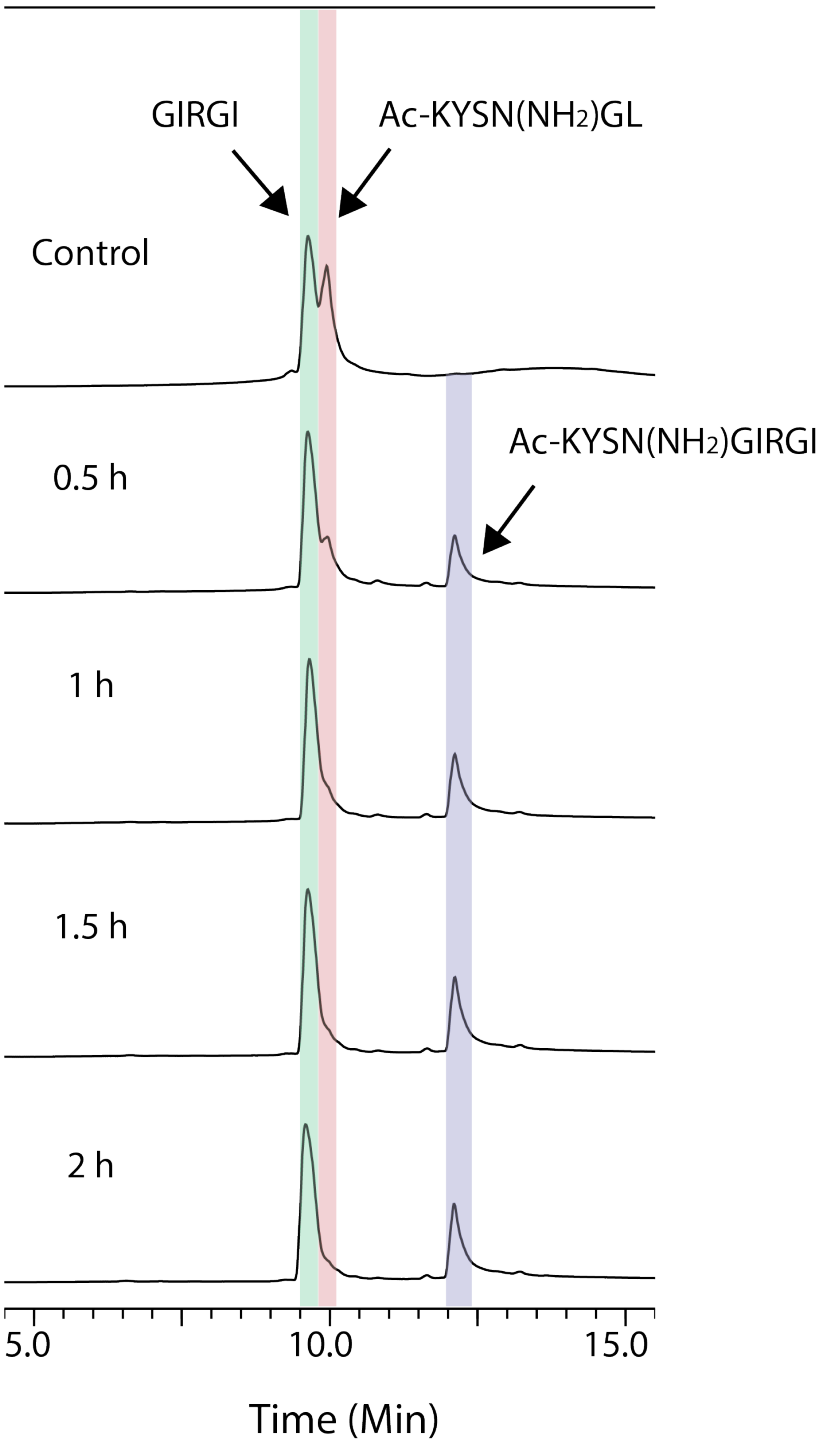


Figure S2. 3. Ligation between Ac-KYSN(NH_2)GL and GIRGI.
The ligation reaction was performed at 37 ºC in the presence of 400 μM of the peptide substrate and 8eq of incoming peptide and 0.01 eq of butelase-1 in 20 mM PBS (pH 6.5) for different time points.

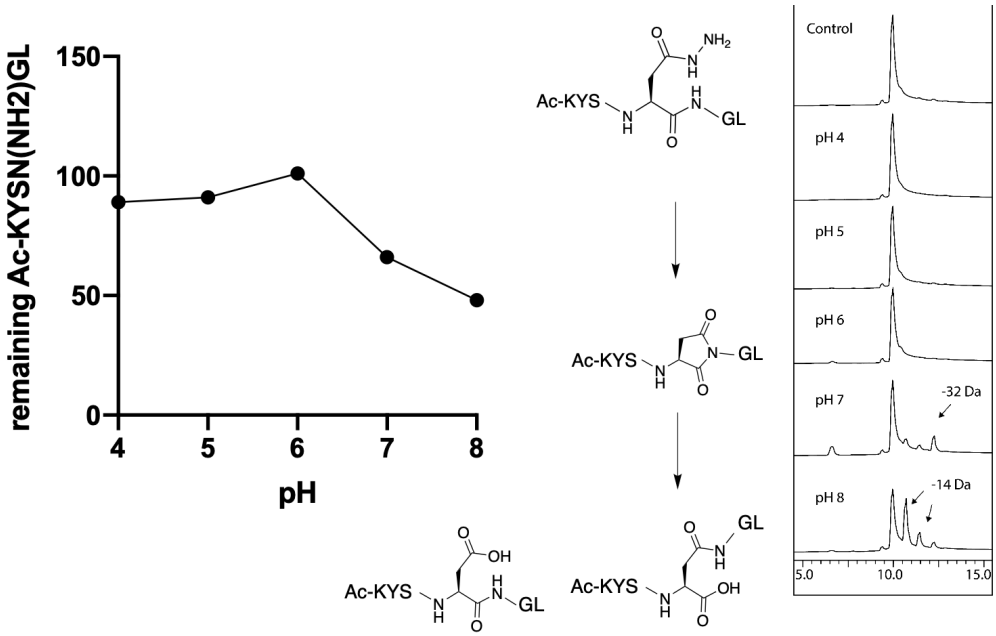


Figure S2. 4. Stability of Ac-KYSN(NH_2)GL.
Ac-KYSN(NH_2)GL have been left in different pH buffer for 12h under RT.

A. Linear peptide precursors used for butelase-1-mediated cyclization to prepare the cyclic Asn(OH)-RTD1 peptide **17** and Asn-RTD1 peptide **19**:

Linear peptide precursor of **17**: CLCRRGVCRCICTISYC**N**(OH)AL

Linear peptide precursor of **19**: CLCRRGVCRCICTISYC**N**AL

B

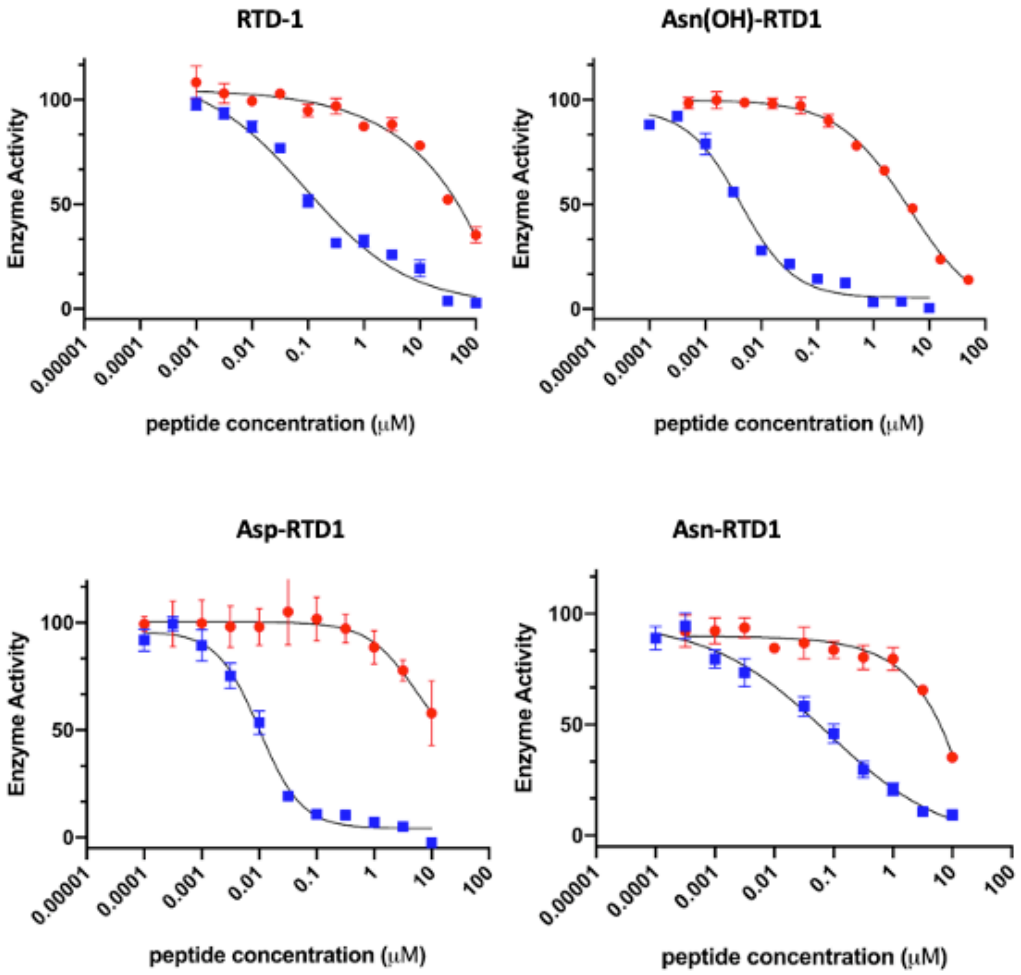


Figure S2. 5. Inhibitory activity towards MMP2 and MMP9.
A) Linear peptide precursors of the cyclic Asn(OH)-RTD peptide **17** and Asn-RTD peptide **19**; B) Inhibitory activity of RTD1 **16**, Asn(OH)-RTD1 **17**, Asp-RTD1 **18** and Asn-RTD1 **19** towards MMP2 (blue square) and MMP9 (red round dot). All values are mean ± SEM from triplicated measurements.

Table S2. 2. Selectivity of RTD1-derived cyclic peptides for MMP2 over MMP9.

No	Peptide sequence	Calc mass	Found Mass M+H$^+$ (m/z)	K_i (nM)	
				MMP2	MMP9
16	cyclo[GFCRCLCRRGVCRCICTR]	2081.6	2082.5	60 ± 10.4	>10000
17	cyclo[SYC**N(OH)**CLCRRGVCRCICTI]	2058.6	2059.7	2.8 ± 0.5	2680
18	cyclo[SYC**D**CLCRRGVCRCICTI]	2043.5	2044.7	6.5 ± 1.1	3290
19	cyclo[SYC**N**CLCRRGVCRCICTI]	2042.5	2043.6	36.5 ± 6.5	5588

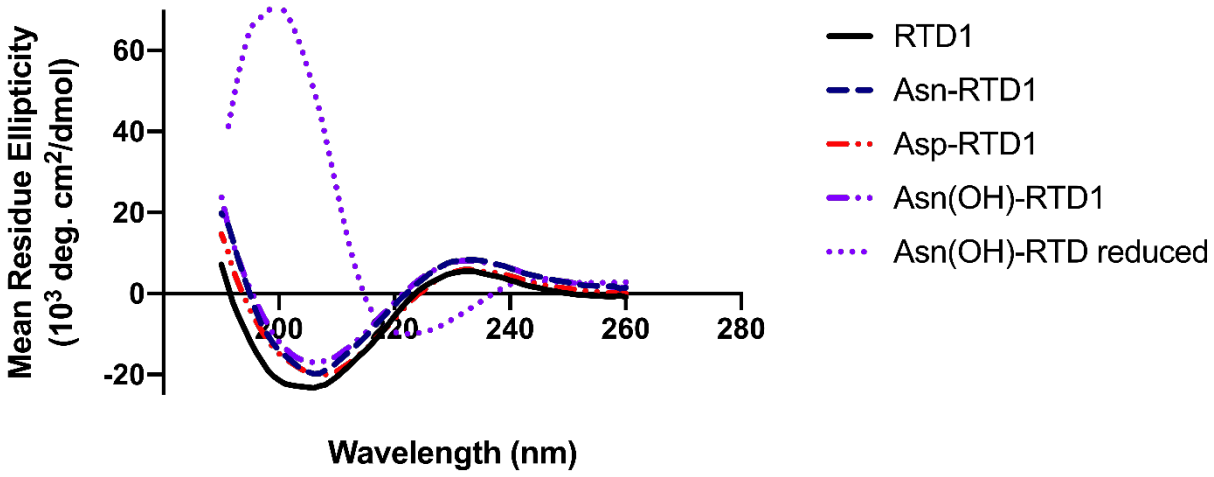


Figure S2. 6. CD spectra of folded RTD1 **16**, folded Asn(OH)-RTD1 **17**, reduced Asn(OH)-RTD1 **17**, folded Asp-RTD1 **18**, and folded Asn-RTD1 **19**.

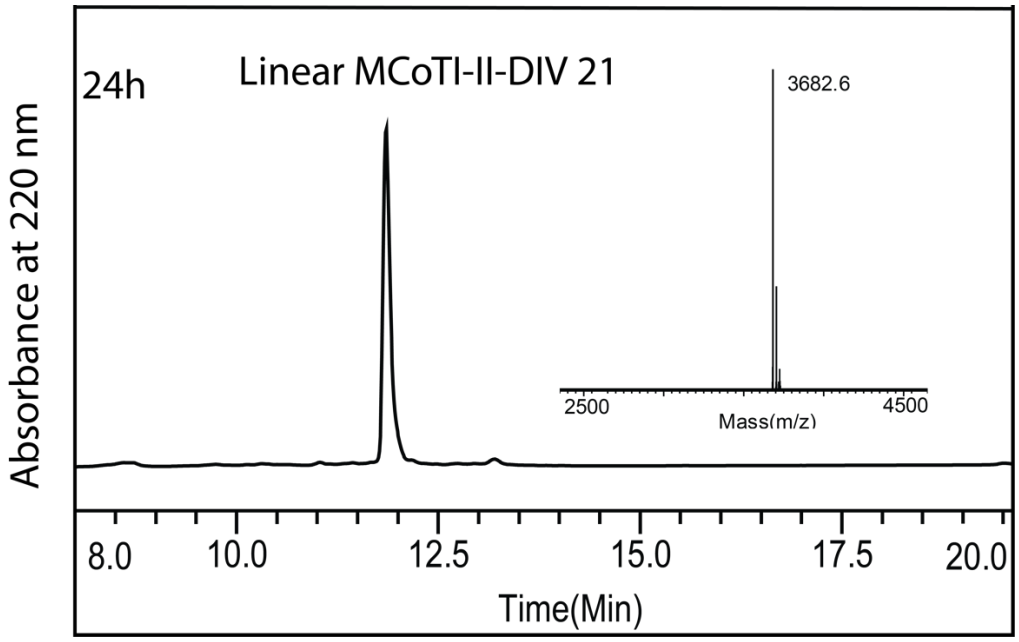


Figure S2. 7. Butelase-1 mediated cyclization of MCoTI-II-DIV **21**.
HPLC trace (10% buffer B to 50% buffer B in 20 mins, UV absorption at 220 nm) of MCoTI-II-DIV **21** cyclization over 24 h. The cyclization reaction was performed in the presence of 400 μM of the peptide substrate and 0.01 eq of butelase-1 in 20 mM PBS (pH 6.5) at RT. There was no cyclic product observed. Reaction at pH 4.5 for overnight also did not yield any cyclized product.

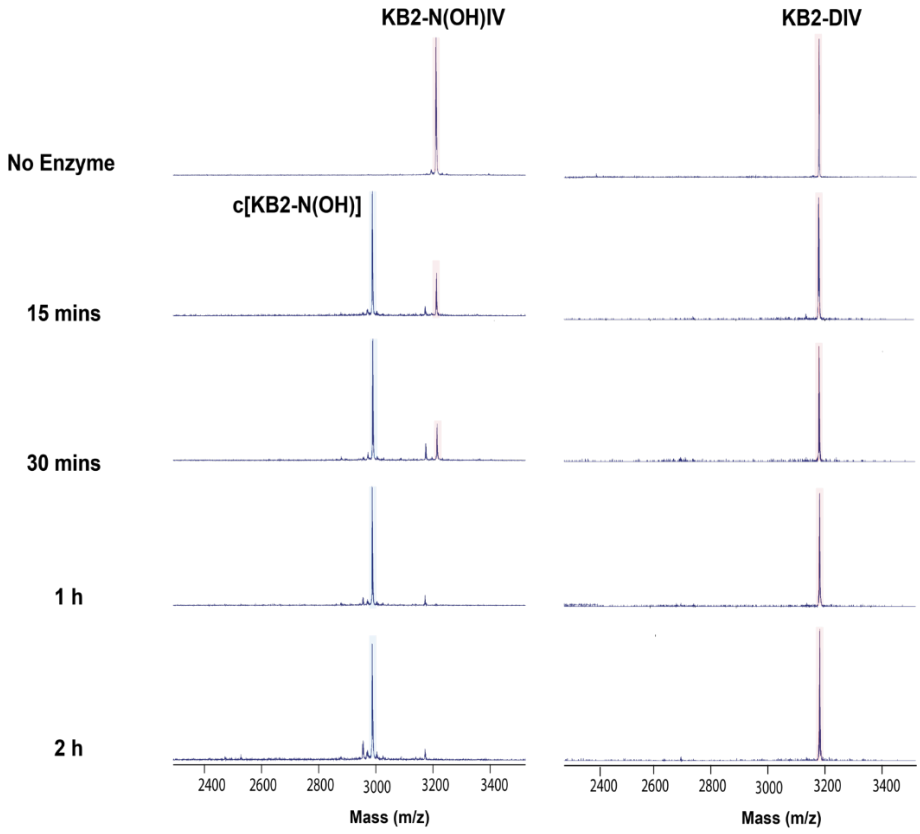


Figure S2. 8. Butelase-1 mediated cyclization of kB2-N(OH)IV **22** and kB2-DIV **23**. The cyclization reactions were performed at 37 °C using 400 μM of the peptide substrates and 0.01 eq of butelase-1 in 20 mM PBS (pH 6.5). The reactions were monitored by MALDI-TOF MS.

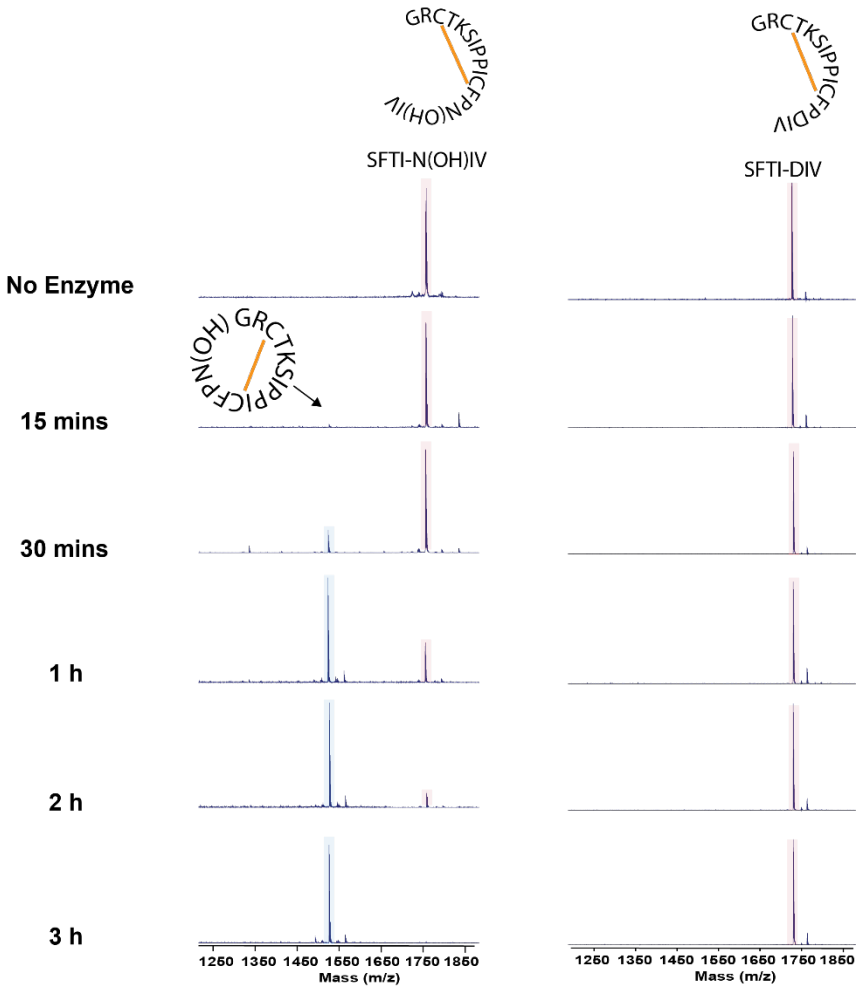


Figure S2. 9. Butelase-1 mediated cyclization of SFTI-N(OH)IV **24** and SFTI-DIV **25**. The cyclization reactions were performed at 37 ºC using 400 µM of the peptide substrates and 0.01 eq of butelase-1 in 20 mM PBS (pH 6.5). The reactions were monitored by MALDI-TOF MS.

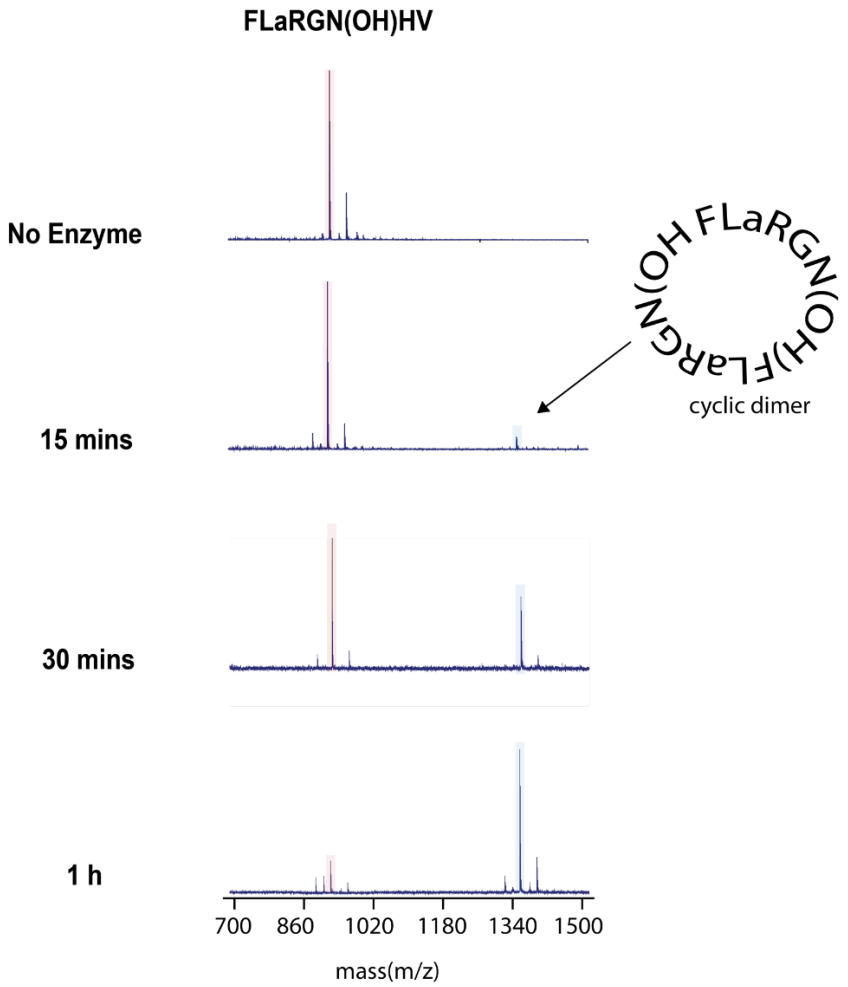


Figure S2. 10. Butelase-1 mediated cyclization of FLaRGN(OH)HV **26**.
The cyclization reaction was performed at 37 ºC using 400 μM of peptide substrates and 0.01 eq of butelase-1 in 20 mM PBS (pH 6.5). The reactions were monitored by MALDI-TOF MS. A mass corresponds to the cyclic dimer of FLARGN(OH) was observed.

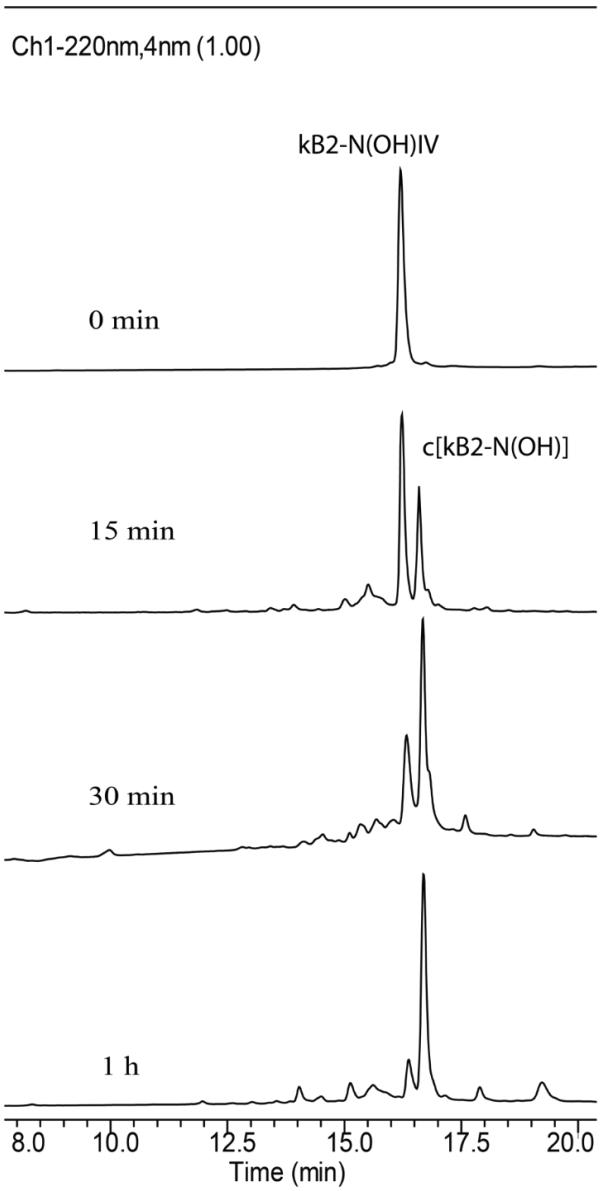


Figure S2. 11. Cyclizaiotn of kB2-N(OH)IV **22**.
HPLC trace (20% buffer B to 60% buffer B in 20 mins, 220 nm) of kB2-N(OH)IV **22** cyclization by butelase-1 over 1-h time course. The linear kB2-N(OH)IV and its cyclic product are labeled as kB2-N(OH)IV and c[kB2-N(OH)], respectively. The cyclization reaction was performed at 37 ºC using 400 μM of the peptide substrate and 0.01 eq of butelase-1 in 20 mM PBS (pH 6.5).

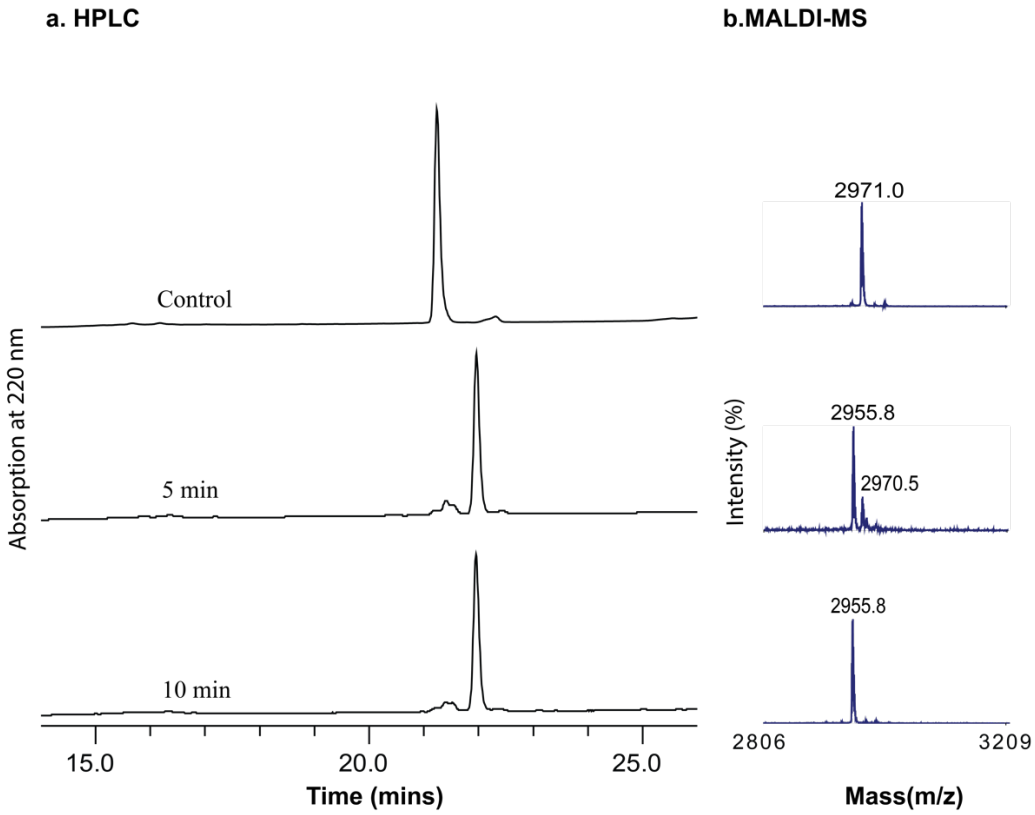


Figure S2. 12. Oxidation of cyclic kB2 **22**.
HPLC (20% buffer B to 80% buffer B in 20 mins, UV absorption at 220 nm) and MALDI-TOF MS monitoring of the hydroxamic acid oxidation reaction of refolded cyclic kB2 **22**. The oxidation was performed at 0 ºC with 200 μM peptide **22**, 1 eq $NaIO_4$, 10 eq methionine in 20 mM PBS (pH 7.2) and monitored at 5 min and 10 min.

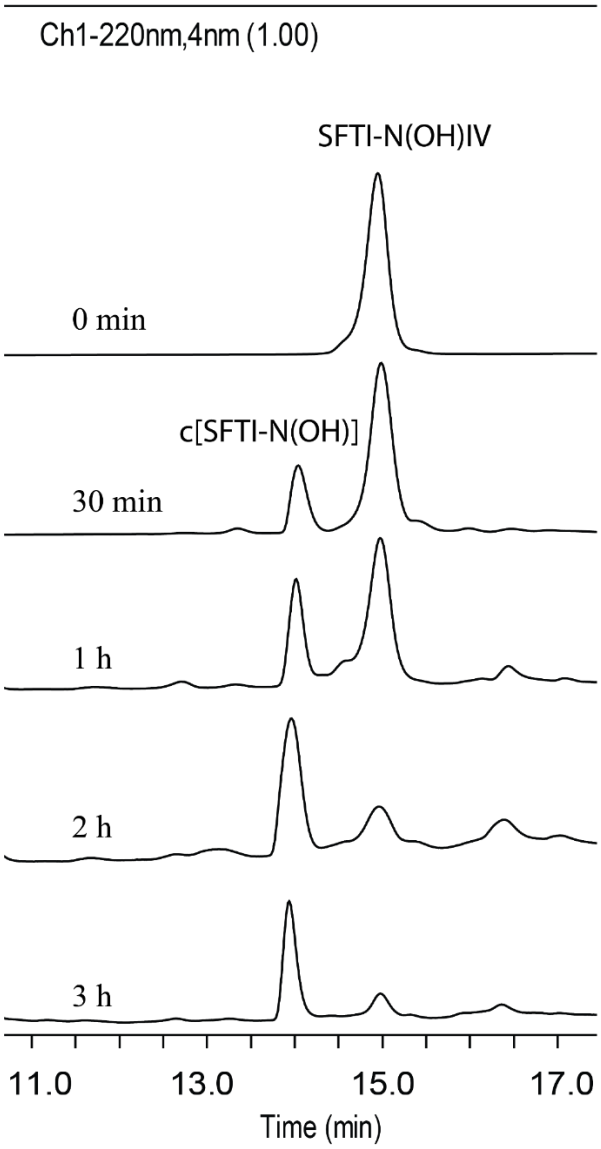


Figure S2. 13. Cyclization of SFTI-N(OH)IV **24**.
HPLC trace (20% buffer B to 60% buffer B in 20 mins, UV absorption at 220 nm) of SFTI cyclization over 3h time course. The linear SFTI-N(OH)IV **24** and its cyclic product are labelled as SFTI-N(OH)IV and c[SFTI-N(OH)], respectively. The cyclization reaction was performed at 37 ^{o}C using 400 μM of the peptide substrate and 0.01 eq of butelase-1 in 20 mM PBS (pH 6.5).

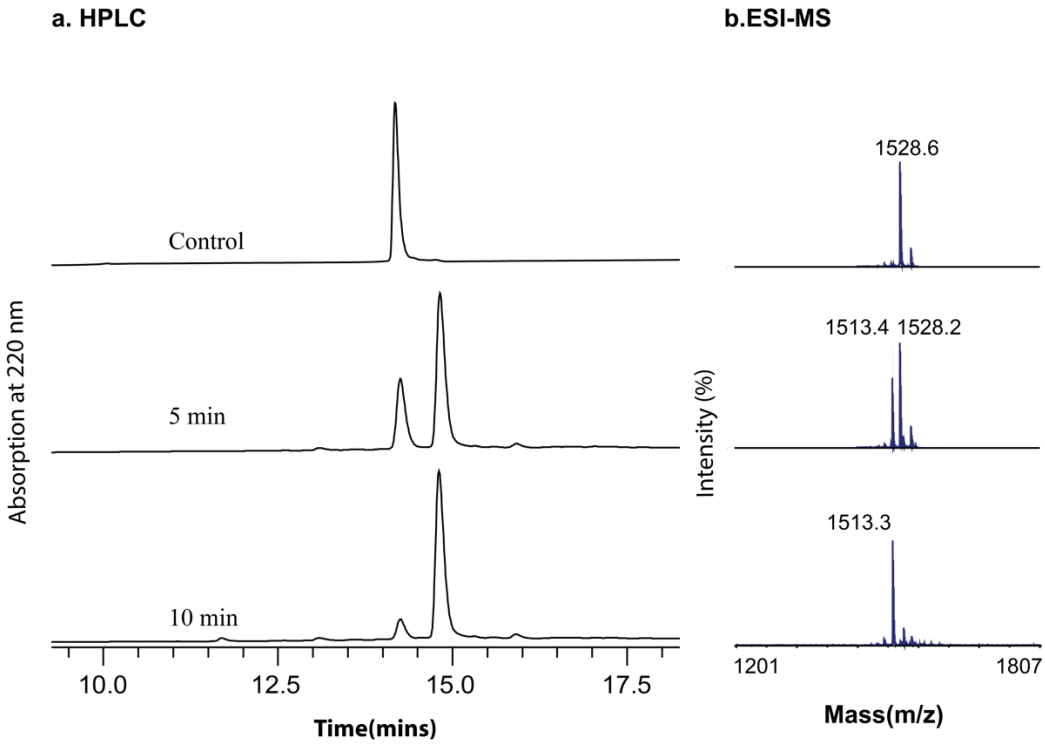


Figure S2. 14. Oxidation of cyclic SFTI **24**.
HPLC (20% buffer B to 60% buffer B in 20 mins, UV absorption at 220 nm) and MALDI-TOF MS monitoring of the hydroxamic acid oxidation reaction of refolded cyclic SFTI **24**. The oxidation reaction was performed at 0 ºC with 200 μM peptide, 1 eq $NaIO_4$, 10 eq methionine in 20 mM PBS (pH 7.2) and monitored at 5 and 10 min.

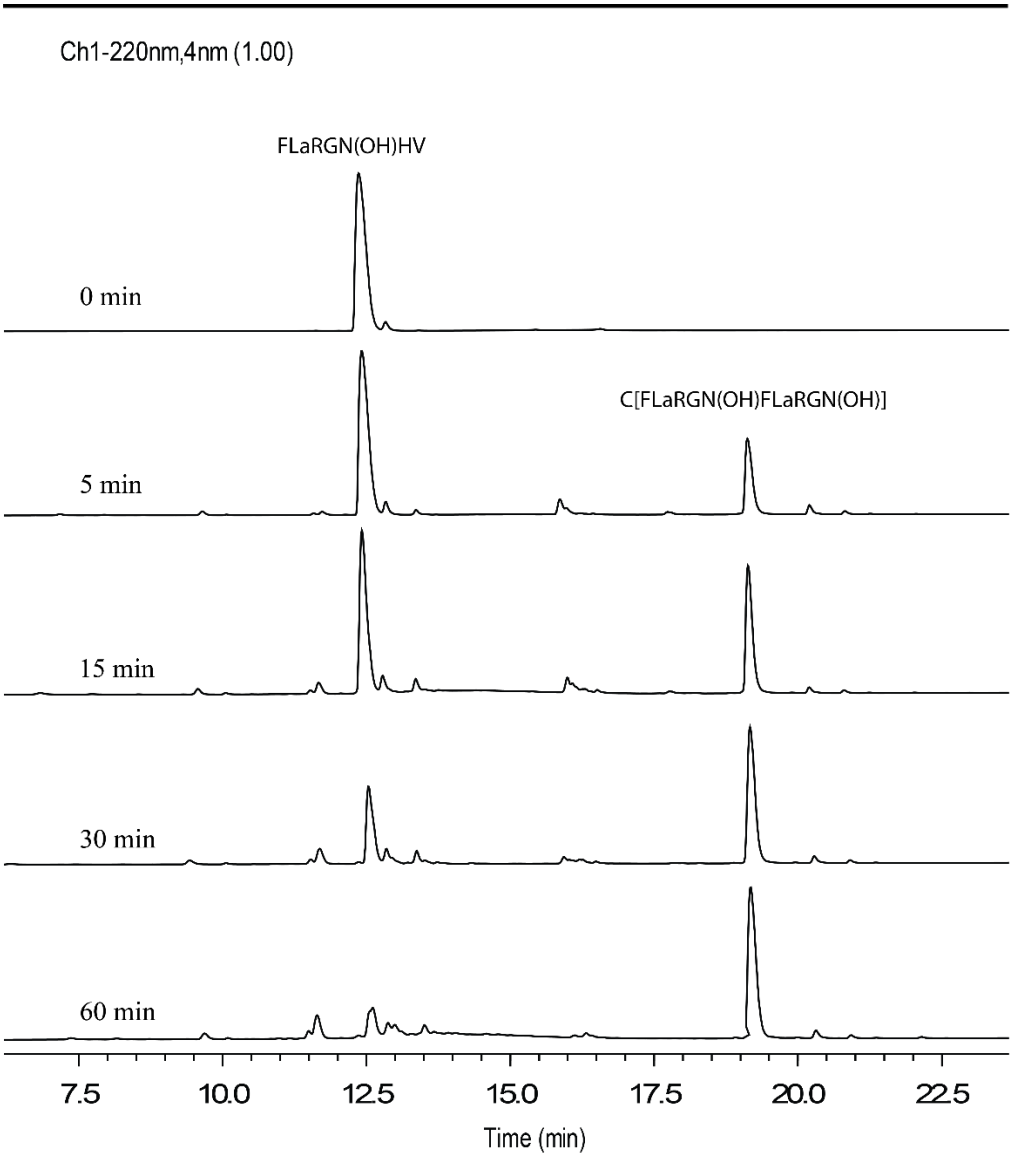


Figure S2. 15. Cyclization of FLaRGN(OH)HV **26**.
HPLC trace (10% buffer B to 60% buffer B in 25 mins, UV absorption at 220 nm) of FLaRGN(OH)HV **26** cyclization over 1-h time course. The linear FLaRGN(OH)HV **26** and its cyclic product are labelled as FLaRGN(OH)HV and c[FLaRGN(OH)FLaRGN(OH)], respectively. The cyclization reaction was performed at 37 ºC using 400 μM of the peptide substrate and 0.01 eq of butelase-1 in 20 mM PBS (pH 6.5).

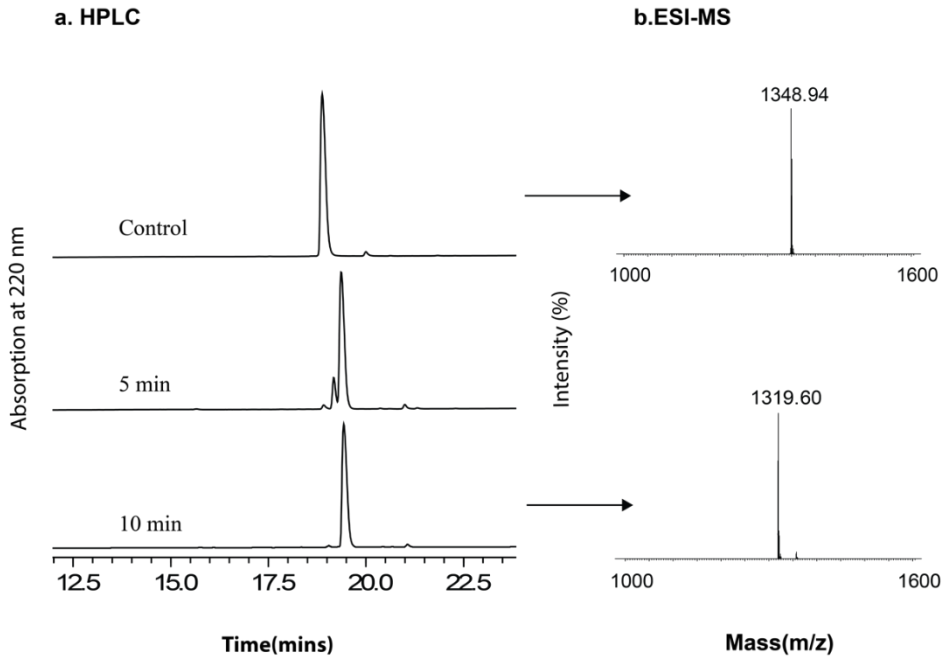


Figure S2. 16. Oxidation of c[FLaRGN(OH)FLaRGN(OH)] **26**.
HPLC (10% buffer B to 60% buffer B in 25 mins, UV absorption at 220 nm) and ESI MS monitoring of the hydroxamic acid oxidation reaction of c[FLaRGN(OH)FLaRGN(OH)] **26**. The oxidation was performed at 0 ^{o}C with 200 μM peptide, 2 eq $NaIO_4$, 10 eq methionine in 20 mM PBS (pH 7.2) and monitored at 5 min and 10 min.

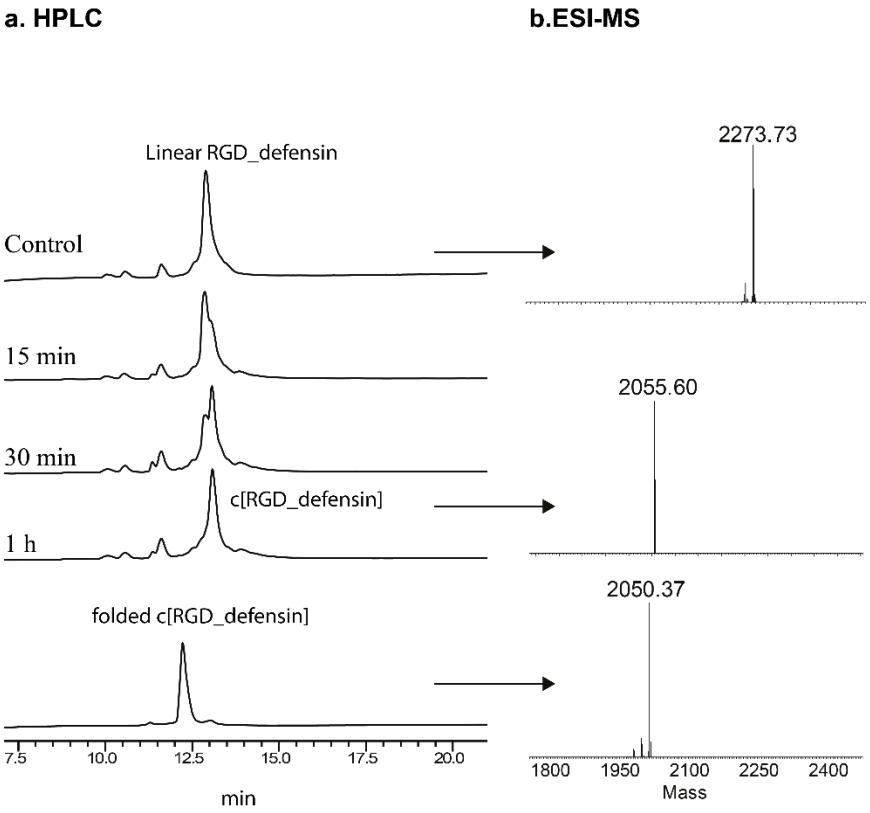


Figure S2. 17. Cyclization of RGD defensin **27**.
HPLC trace (10% buffer B to 50% buffer B in 20 mins, 220 nm) and ESI MS profile of RGD-defensin peptide **27** cyclization over 1h time course. The linear RGD defensin and its cyclic product are labelled as linear RGD_defensin and c[RGD defensin], respectively. The cyclization reaction was performed at 37 ºC using 400 μM of the peptide substrate and 0.01 eq of butelase-1 in 20 mM PBS (pH 6.5).

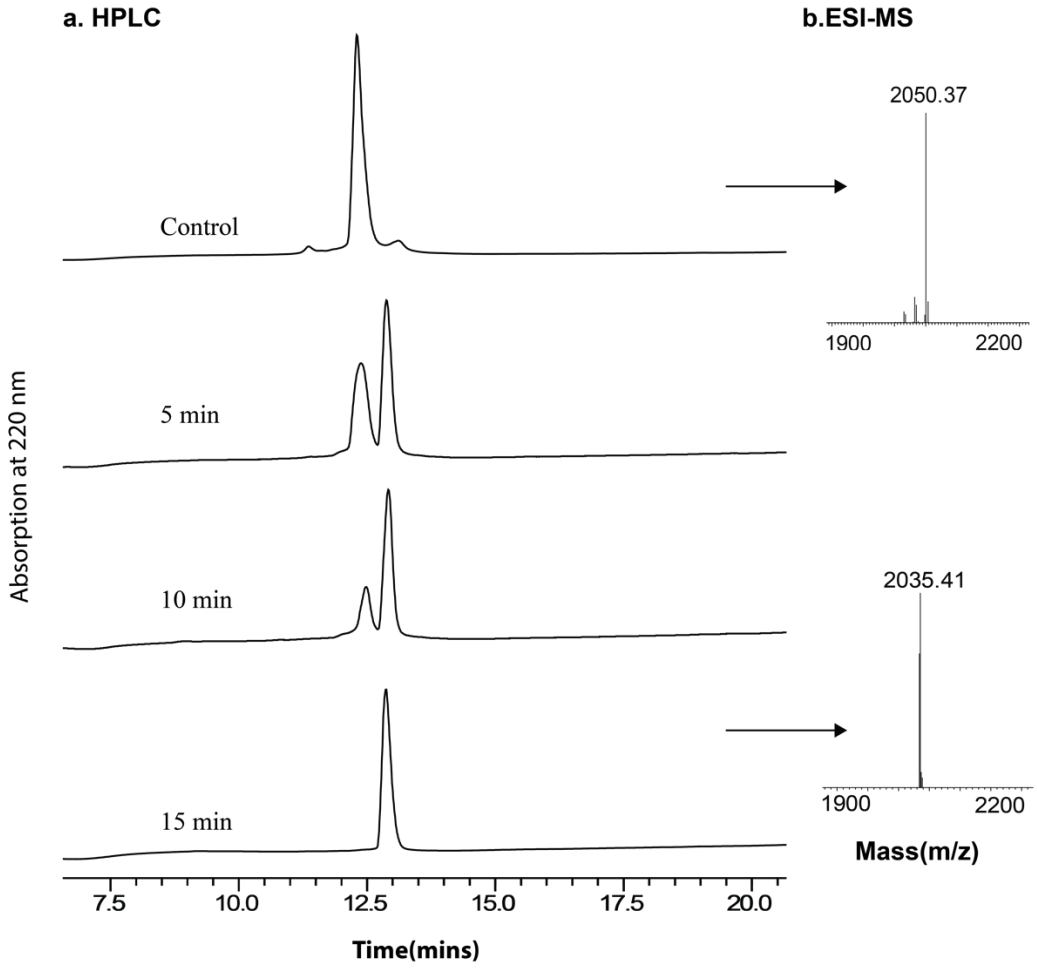


Figure S2. 18. Oxidation of cyclic RGD defensin **27**.
HPLC (10% buffer B to 50% buffer B in 20 mins, UV absorption at 220 nm) and MALDI-TOF MS monitoring of the hydroxamic acid oxidation reaction of refolded cyclic RGD defensin **27**. The oxidation was performed at 0 ºC with 200 μM peptide, 1 eq $NaIO_4$, 10 eq methionine in 20 mM PBS (pH 7.2) and monitored at 5 min, 10 min and 15 min.

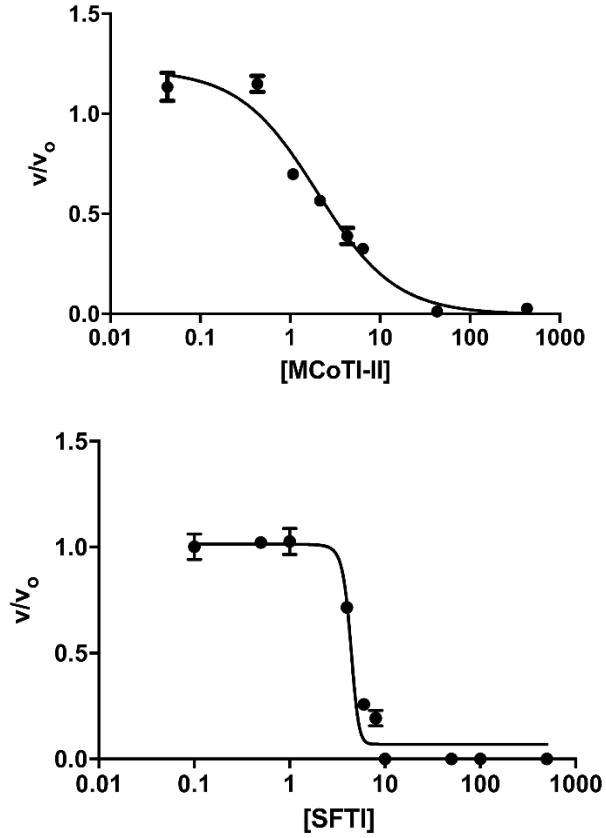


Figure S2. 19. Trypsin inhibition assay of MCoTI-II **20** and SFTI **24**.
All triplicated data are presented as means ± SEM.

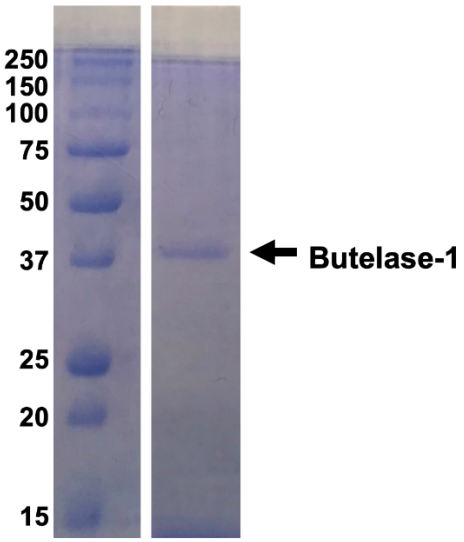


Figure S2. 20. SDS-PAGE of butelase-1.

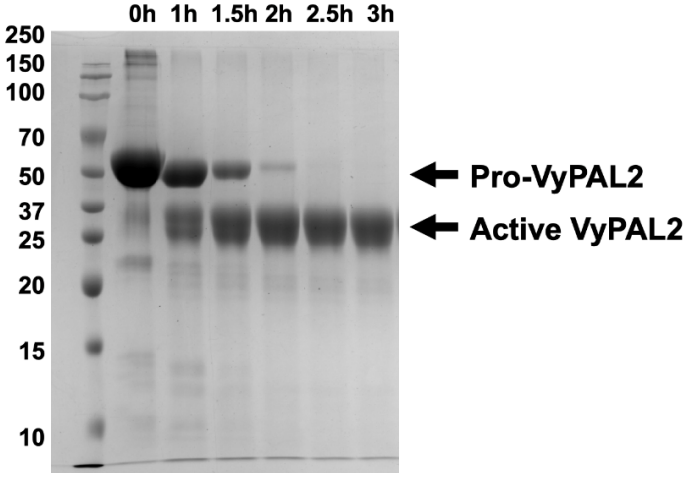


Figure S2. 21. Activation of pro-VyPAL2.
After 2.5-h activation, the activated enzyme was purified by SEC at pH 6.5 and stored at -20 ºC

Figure S2. 22. ^{1}H spectrum of compound **a** (400 MHz, $CDCl_3$)

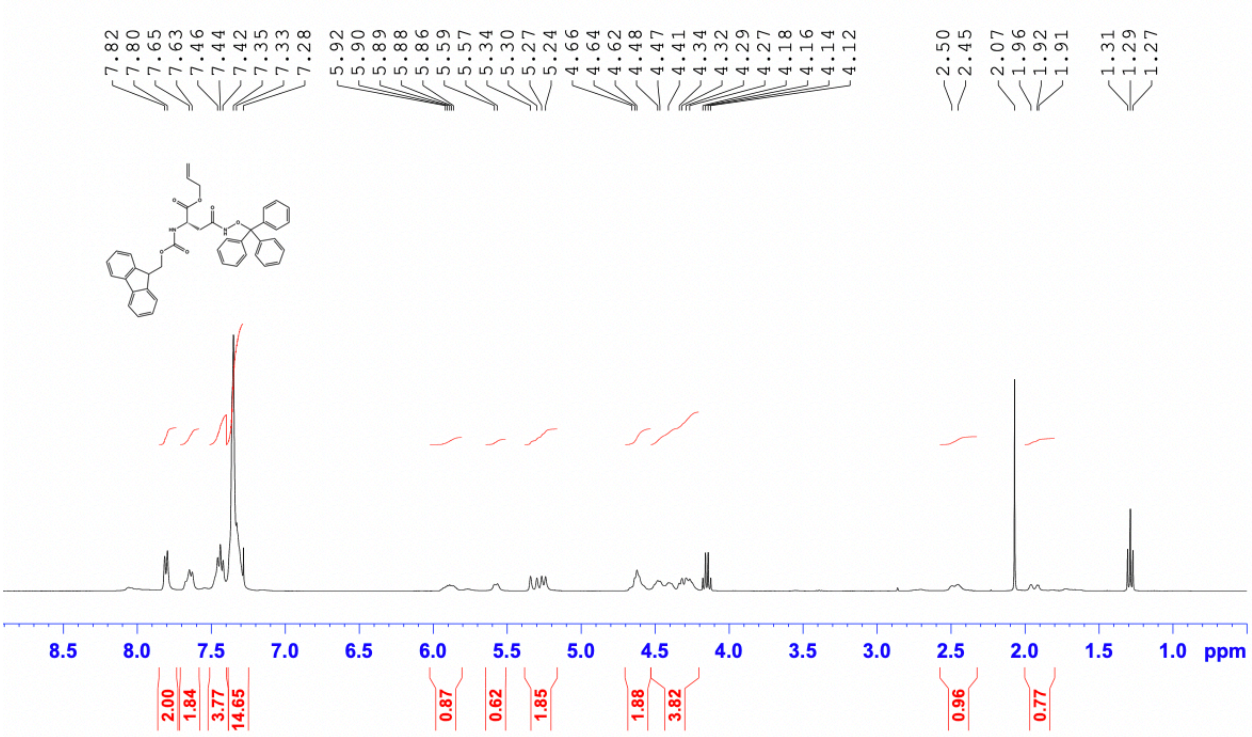


Figure S2. 23. ^{13}C spectrum of compound **a** (100 MHz, $CDCl_3$)

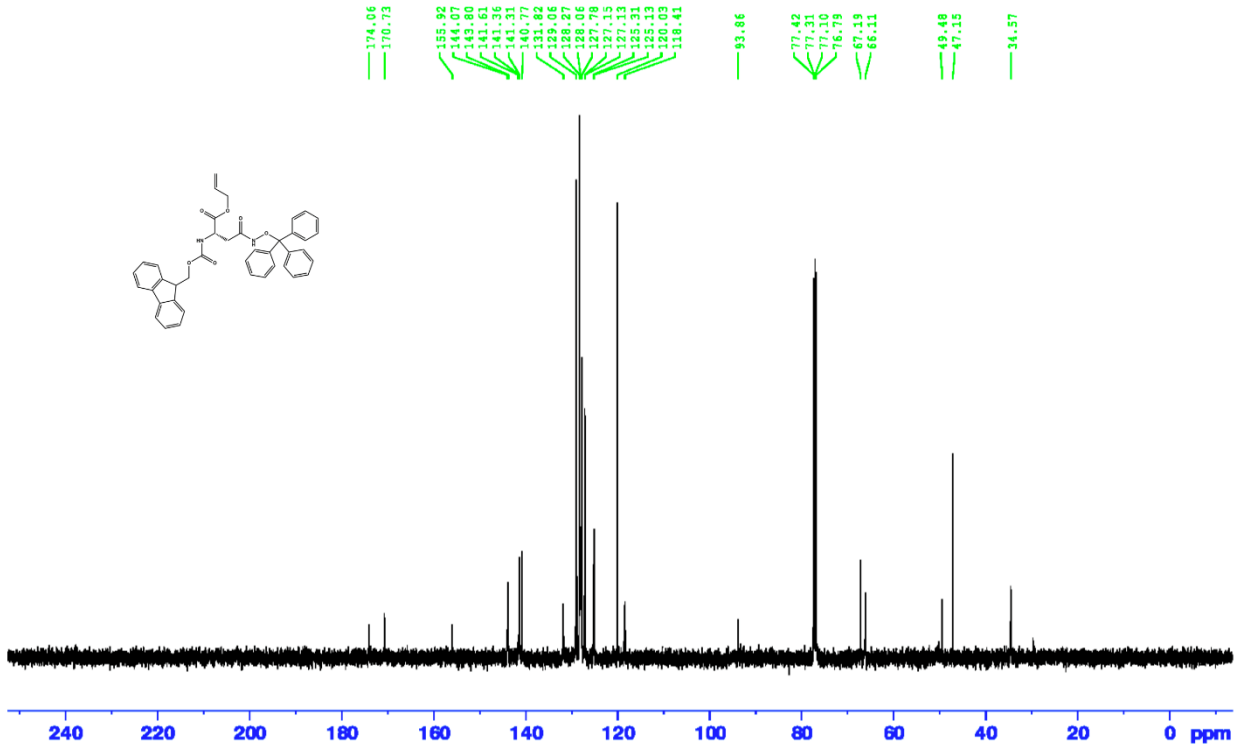


Figure S2. 24. ^{1}H spectrum of compound **c** (400 MHz, $CDCl_3$)

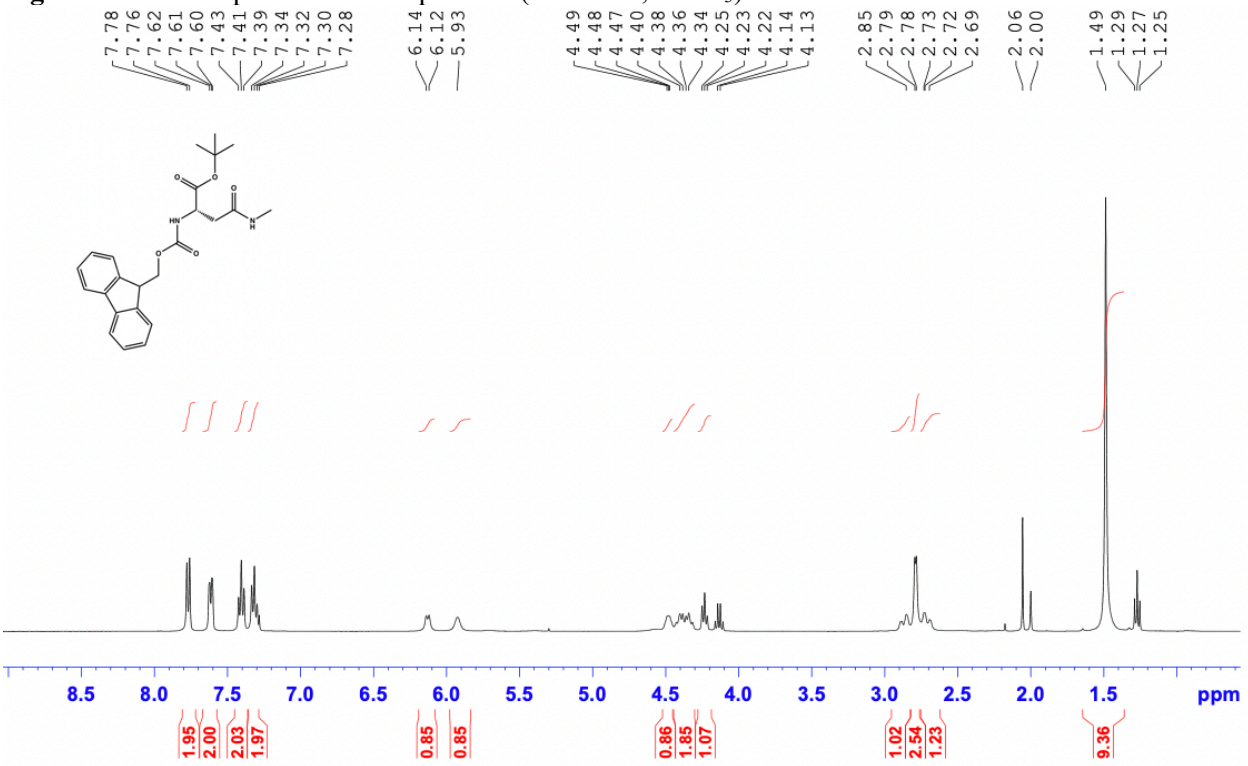
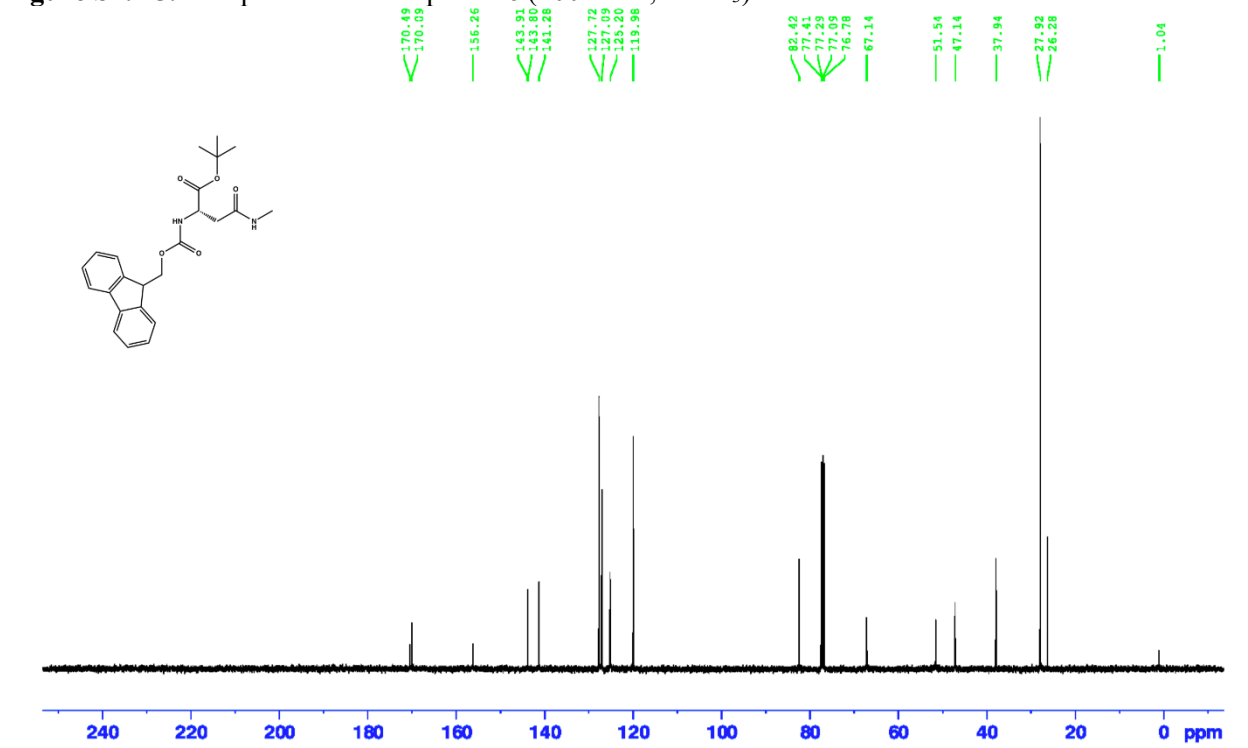


Figure S2. 25. ^{13}C spectrum of compound **c** (100 MHz, $CDCl_3$)



Chapter 3. A Cascade Enzymatic Reaction Scheme for Irreversible Transpeptidative Protein Ligation

3.1. Abstract

Enzymatic peptide ligation holds great promise in the study of protein functions and development of protein therapeutics. Owing to their high catalytic efficiency and a minimal tripeptide recognition motif, peptidyl asparaginyl ligases (PALs) are particularly useful tools. However, as an inherent limitation of transpeptidases, PAL-mediated ligation is reversible, requiring a large excess of one of the ligation partners to shift the reaction equilibrium in the forward direction. Herein, we report a method to make PAL-mediated intermolecular ligation irreversible by coupling it to a glutaminyl cyclase (QC)-catalyzed pyroglutamyl formation (**Figure 3. 1**). In this method, the acyl donor substrate of PALs is designed to have a glutamine at the P1' position of the Asn-P1'-P2' tripeptide PAL recognition motif. Upon ligation with an acyl acceptor substrate, the acyl donor substrate releases a leaving group in which the exposed N-terminal glutamine is cyclized by QC, quenching the Gln N$^{\alpha}$-amine in a lactam. Using this method, PAL-mediated ligation can achieve near-quantitative yields even at an equal molar ratio between the two ligation partners. We have demonstrated this method for a wide range of applications, including protein-to-protein ligations. We anticipate that this cascade enzymatic reaction scheme will make PAL enzymes well suited for numerous new uses in biotechnology.

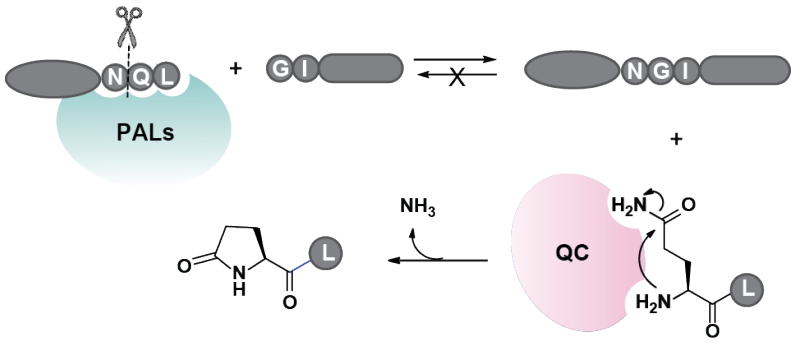


Figure 3. 1. PAL-mediated ligation coupled with QC-mediated pyroglutamine formation.

3.2. Introduction

Site-specific protein ligation facilitates protein function studies and enables the development of protein therapeutics. Besides the well-established chemical ligation methods [107-111], recent years have seen an increasing use of peptide ligases for protein modification [15, 112-114]. For example, a popular peptide ligase has been sortase A, which recognizes a sorting sequence LPXTG and ligates after Thr [113]. Nevertheless, the catalytic efficiency of sortase A is low, requiring a stoichiometric amount of the enzyme for a practicable ligation reaction. In contrast, the recently discovered peptidyl asparaginyl ligases (PALs) – as represented by butelase-1 [27] – have a catalytic efficiency that is 10^4 times that of sortase A. Furthermore, PALs require a minimal asparaginyl tripeptide recognition motif, Asn-P1'-P2', for ligation after Asn, where P1' can tolerate a broad range of amino acid residues and P2' is usually a large hydrophobic residue [27, 28, 55, 57, 58, 66, 73, 77, 115-117]. As members of the asparaginyl endopeptidase (AEP) family, PALs utilize the catalytic cysteinyl thiol to cleave the Asn-P1' peptide bond in an acyl donor substrate, and the resultant asparaginyl thioester intermediate is then resolved by the amine nucleophile of an acyl acceptor substrate. Therefore, the ligation product is formed through transpeptidation. Like butelase-1, VyPAL2 and OaAEP1b-C247A are two other PALs that have excellent transpeptidase activity [36, 38, 40]. However, reversibility is an inherent feature of a transpeptidation reaction because both the ligated product and the cleaved leaving group can be reused by the PAL enzyme in the reverse reaction. Therefore, an excess amount of one ligation partner is needed for a high-yielding intermolecular ligation reaction

Several methods have been developed to shift the equilibrium of PAL-mediated ligation to the product side [47, 76, 118-120]. We first developed the thiodepsipeptide method to use an asparaginyl thioester peptide as the acyl donor substrate (peptide-Asn-thioglc-Xaa), which turned the ligation reaction irreversible [77, 118]. Nevertheless, a limitation of this method lies with the need

to prepare the thioester substrates. In a second method [76], Rehm *et al*. used a glycinyl-valinyl acyl acceptor to ligate with an NGL-containing acyl donor, resulting in an NGV motif in the product which was more resistant to OaAEP-C247A. This reduced both product hydrolysis and reaction reversibility. However, NGV is only relatively more stable than NGL. And as a poorer acceptor substrate, the GV-peptide must also be used in a large excess (>20-fold) to the acyl donor. No protein–protein ligation was demonstrated. Another method, developed by Tang *et al.*, used Asn-Cys-Leu as the P1-P1'-P2' tripeptide motif for OaAEP1-C247A [119]. Quenching the N-terminal cysteine of the Cys-Leu leaving group by the aldehyde compound, 2-formyl phenylboronic acid (FPBA), reduced reversibility of the ligation reaction. However, FPBA was found to slow down the ligation reaction significantly, likely because the aldehyde could form imines with the amine substrates and/or the enzyme. Lastly, a method reported by Rehm *et al.* made use of Ni^{2+} to chelate a GLH tripeptide leaving group to mask the nucleophilicity of its N-terminal amine [120]. This approach gave a great increase in the yield of protein N- and C-terminal labelling reactions with small synthetic peptides. However, for protein–protein ligations, only a modest yield increase was observed – from 24% to 36% in one case and 28% to 39% in another [120]. The first two methods (using peptide-Asn-thioglc-Xaa and glycinyl-valinyl acyl acceptor) certainly have very limited applications. Nevertheless, the two most recent methods are aimed to improve the ligation yield by adding 1 (or 2) to 1 molar ratio of quenching chemicals to the leaving group in order to overcome the reversibility. The high concentration of the added chemicals (Ni^{2+} and FPBA) may reduce the activity of PAL itself. The use of a metal ion like Ni^{2+} also complicates product purification in a manufacturing process because of its potential to form complexes with the protein substrate. Therefore, the quenching reaction should be more orthogonal without introducing any interfering chemicals. The use of another enzyme in catalytical amount to specifically quench the nucleophilicity of the leaving group may meets this requirement.

3.3. Results and discussion

3.3.1. Demonstration of the PAL-QC cascade scheme in model peptide ligation reactions.

By catalyzing N-terminal pyroglutamine (pGlu) formation, QC is involved in the maturation of many bioactive peptides and proteins [121-123]. The efficiency of QCs from different organisms at catalyzing the unimolecular lactamization reaction is ~10^5 $M^{-1}\cdot S^{-1}$ [123], while that of PALs in catalyzing the bimolecular ligation reactions is ~10^4 $M^{-1}\cdot S^{-1}$ [54]. This makes QC a particularly attractive enzyme to trap the released glutaminyl leaving group. We first showed that QC efficiently converted the N-terminal Gln to pGlu in four synthetic peptides of the sequence QXGSA (X = L, I, F or V, which are favored by PALs as the P2' amino acid) (**Figure S3. 1**). Consistent with previous studies, the presence of a large hydrophobic residue like X at the second position of a glutaminyl peptide did not negatively affect QC on its ability to catalyze the lactamization reaction. Indeed, at 0.0001 eq to the substrate, QC was able to complete the reaction in less than 30 min (pH 7) (**Figure 3. 2** b). The reactions of QI-, QL- and QF-peptides had similar rates while that of QV-peptide was about 30% slower (**Figure S3. 1**). Since Leu is the most favored P2' residue of PALs, we chose Asn-Gln-Leu as the tripeptide recognition motif in all acyl donor substrates used in this study.

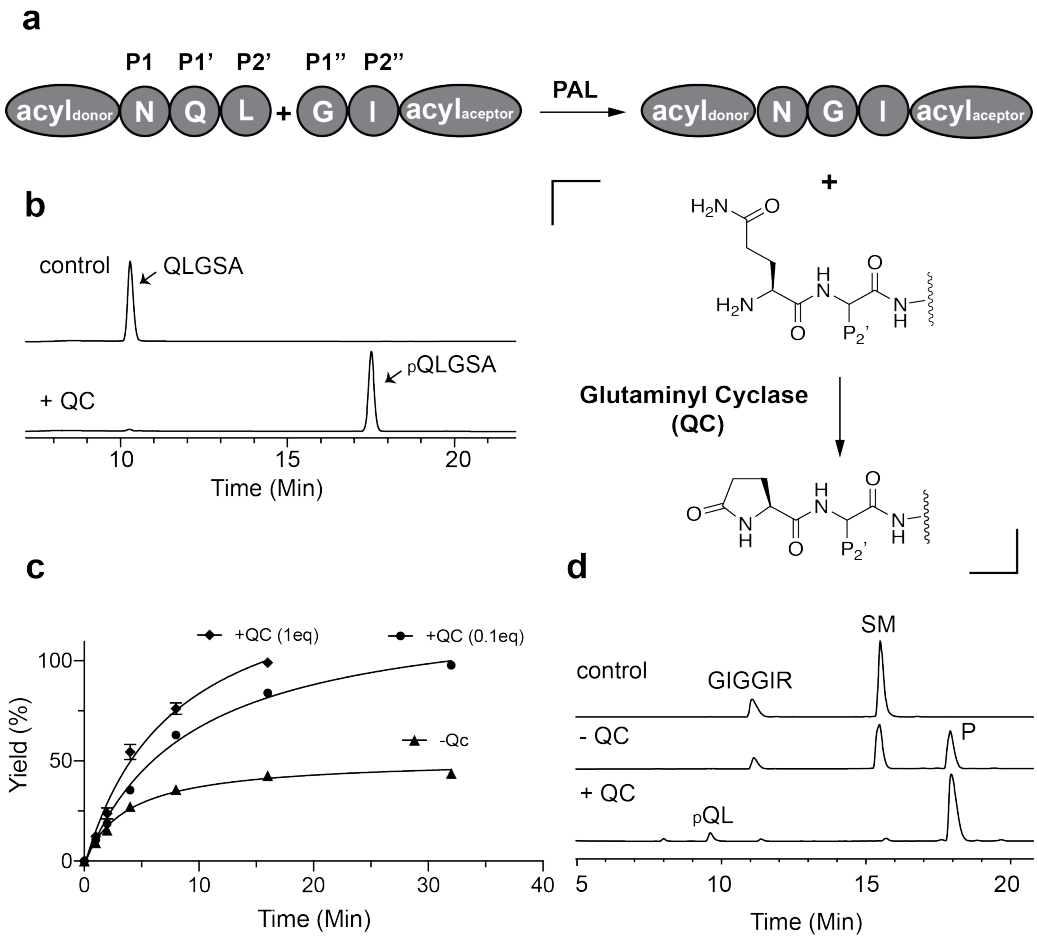


Figure 3. 2. The PAL mediated ligation enhances by addition of QC.
a) A scheme illustrates that QC quenches the nucleophilicity of cleaved QL by catalyzing the glutamine into para-glutamine. b) RP-HPLC monitoring the conversion of QLSGA to $_{p}$QLSGA. The reactions were conducted at 37 ºC using 5mM of substrate and 0.0001 eq of QC in 20 mM PBS (pH 7) for 30 min. c) The reaction yield between Ac-SYRNQL and GIGGIR in the presence or absent of QC. The reactions were conducted at 37 ºC using 5mM of acyl acceptor and 1 eq of acyl donor, and 0.0005 eq of OaAEP1b -C247A, (with or without 0.00005 eq of QC) in 20 mM PBS (pH 7) for various time points. The yield of the ligation is determined from HPLC (280 nm) peak integration. All data are presented as means ± SEM. d). HPLC trace of Ac-SYRNQL and GIGGIR Ligation (220 nm absorbance).

Then we set out to test our cascade enzymatic scheme in a model ligation reaction using Ac-SYRNQL (5 mM) as the acyl donor and GIGGIR (1 eq) as the acyl acceptor. At a PAL-to-substrate molar ratio of 0.0005:1, the reaction by OaAEP1b-C247A at pH 7 gave the product in ~45% yield when the reaction

reached equilibrium at ~30 min (**Figure 3. 2** c and d). In contrast, the addition of 0.00005 eq of QC increased the ligation yield to >95% in 32 min (**Figure 3. 2** c and d). This drastic yield increase indicated that a very low amount of QC (0.005 mol% to the substrate and 1/10 eq to PAL) was sufficient to quench the released QL dipeptide through pGlu formation. This provided a clear validation of our PAL-QC coupled reaction scheme. When performing the ligation reaction using 0.05 mol% QC (i.e, QC:PAL = 1:1), the reaction reached equilibrium faster, but the final yield was almost the same. Because the reaction was fast enough at 1/10 QC to PAL, this ratio was used in all following studies. We also conducted the ligation reaction using VyPAL2 or butelase-1 under the same conditions. A similar drastic increase of the product yield was observed when the reaction was done in the presence of QC, demonstrating that QC is compatible with all the three most useful PALs (**Figure S3. 2**).

3.3.2. Use of the cascade enzymatic scheme for protein N- and C-terminal labelling.

Next, we used ubiquitin as a model protein to demonstrate our method in protein labelling reactions. Two recombinant ubiquitin variants, GI-ubiquitin and ubiquitin-NQL-His_6, were prepared for N- and C-terminal labelling with two biotinylated synthetic peptides, biotin-GRSNQL and GIGGIRK(biotin), respectively. 500 μM of the ubiquitin substrate protein and 1.2 eq of the biotin peptide were used in both ligation reactions which were conducted at pH 7 and 37 ^{o}C with 0.5 μM OaAEP1b-C247A (0.001 eq). For the ligation reaction of ubiquitin-NQL-His_6 with GIGGIRK(biotin), the yield increased from 40 to 94% when 0.0001 eq of QC was added (**Figure 3. 3**). Similarly, the yield of ligation between biotin-GRSNQL and GI-ubiquitin increased from 49% to 88% with the addition of QC. The huge enhancement of the protein C- and N-terminal labelling efficiency by QC further validated our method in protein-to-peptide ligation reactions.

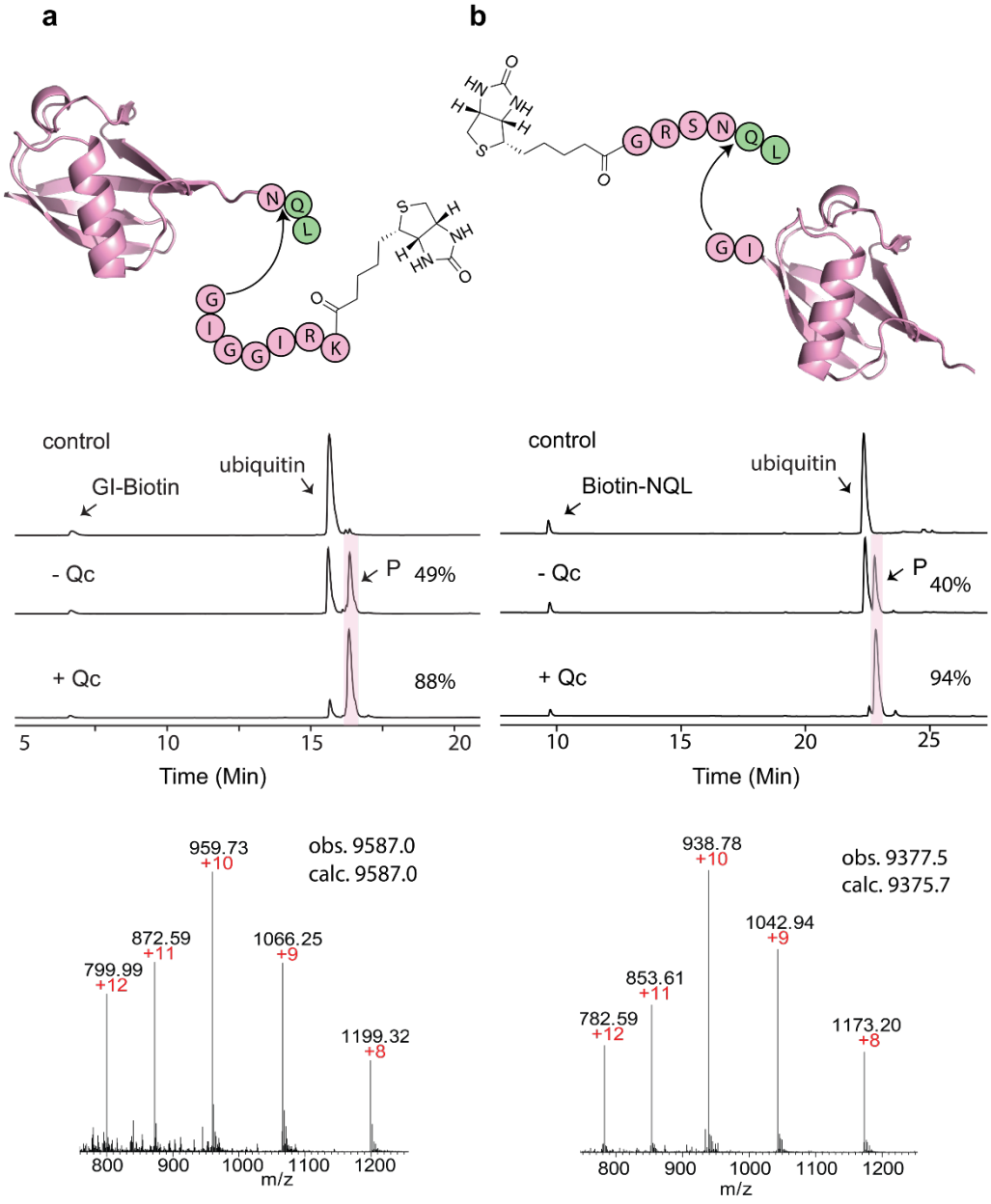


Figure 3. 3. Use of the PAL-QC coupled cascade scheme for protein-peptide ligation. (a) Ligation between ubiquitin-NQL-His_6 and GIGGIRK(biotin). (b) ligation between biotin-GRSNQL and GI-ubiquitin. Both reactions were conducted at 37 ºC using 500 µM of ubiquitin, 1.2 eq of peptide, 0.001 eq of OaAEP1b-C247A with or without 0.0001 eq of QC in 20 mM PBS (pH 7) for 1 h. Both reactions were monitored using HPLC (middle panel) and the labelling products were characterized by ESI-MS (lower panel).

3.3.3. Use of the cascade enzymatic reaction scheme for protein–protein ligation

We then selected several proteins – GFP, ubiquitin, DARPin9-26 [124, 125] and an anti-EGFR affibody Z_{EGFR} [126]– to determine whether our method could be further extended to protein–protein ligation reactions. We first conducted ligation of DARPin-NQL (400 μM) with GI-ubiquitin (1.8 eq) using VyPAL2 (0.001 eq) at pH 7 and 37 ºC. Without QC, the reaction yielded the product in 37% in 3 h, whereas the addition of QC increased the yield to 95% (**Figure 3. 4** a). Interestingly, we found that, for this DARPin-to-ubiquitin ligation reaction, VyPAL2 was a better ligase than OaAEP1b-C247A because the latter produced a small amount of hydrolysis product DARPin-N-OH (**Figure S3. 3**, **Figure S3. 4**). This phenomenon was not observed in neither peptide–peptide ligation nor peptide–protein ligation reactions. It seems that OaAEP1b-C247A has residual hydrolase activity which could be exacerbated in the difficult, entropically demanding protein–protein ligation reaction. Therefore, VyPAL2 was used for all subsequent inter-protein ligation reactions.

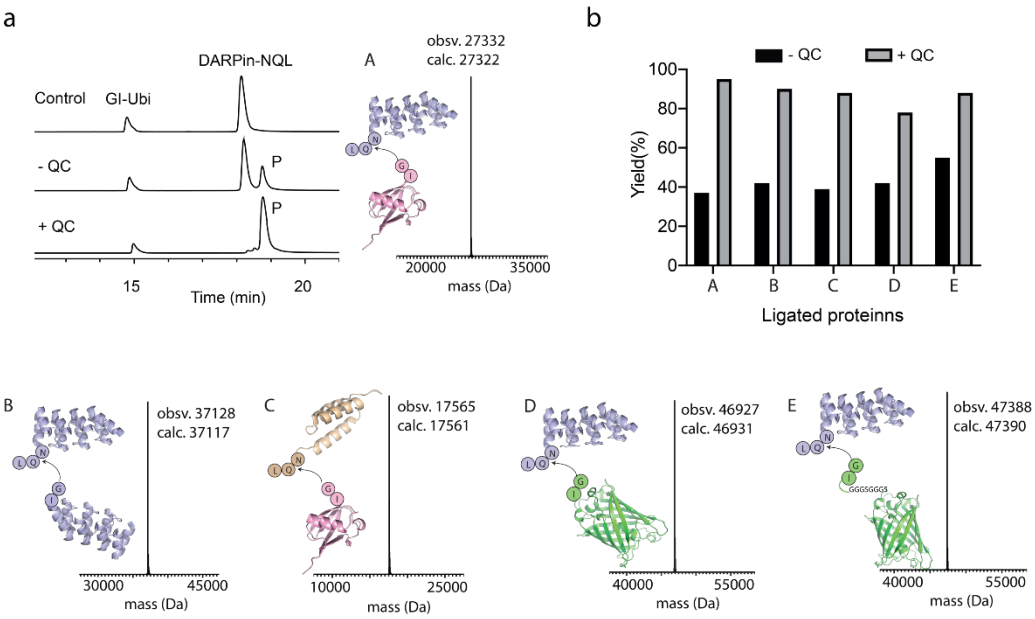


Figure 3. 4. Enhancement of protein–protein ligation efficiency by coupled use of VyPAL2 with QC.

a) HPLC monitoring of VyPAL2-mediated ligation between DARPin-NQL and GI-ubiquitin with or without QC. b) Yields of various protein–protein ligations in the presence or absence of QC (A: DARPin-NQL and GI-ubiquitin, B: DARPin-NQL and GI-DARPin, C: Z_{EGFR}-NQL and GI-ubiquitin, D: DARPin-QL and GI-GFP, E: DARPin-NQL and GI-GGGSGGGS-GFP). All reactions were conducted at 37 °C using 400 μM of acyl donor protein, 1.8 eq of acyl acceptor protein, 0.001 eq of VyPAL2 with/without 0.0001 eq of QC in 20 mM PBS (pH 7) for 3-4 h. The reactions were monitored using HPLC and ESI-MS (**Figure S3. 3**, **Figure S3. 5**, **Figure S3. 6**, **Figure S3. 7**,**Figure S3. 8**).

Similarly, ligation of DARPin-NQL (1 eq) with GI-DARPin (1.8 eq) afforded the tandem-linked DARPin-NGI-DARPin in 42% (without QC) and 91% (with QC) (**Figure 3. 4** b). Ligation of Z_{EGFR}-NQL with GI-uibiquitin afforded the product in 39% (without QC) and 88% (with QC). DARPin-NQL was also ligated with the much larger GFP protein, GI-GFP, to produce DARPin-NGI-GFP in 42% (without QC) and 78% (with QC).

The proximity of the N-terminal GI nucleophile to the rigid β-barrel structure of the GFP protein might have hindered its accessibility, leading to the slightly lower ligation yield and some hydrolysis in this reaction (**Figure S3. 7**). Adding a longer spacer after the GI dipeptide improved the accessibility and consequently the ligation yields between DARPin and GI-GGGSGGGS-GFP – 55% (without QC) and 88% (with QC) (**Figure S3. 8**). In all cases, the product yields were improved by 33-53% when using QC together with VyPAL2. This improvement is remarkably comparable to that observed in the model peptide ligations, which makes our method distinctly advantageous to all previous methods [47, 76, 118-120].

Z_{EGFR}-Fc-NQL, a large dimeric fusion protein (MW ~68 kDa) composed of the affibody Z_{EGFR} and the Fc domain of IgG, was used to ligate with GI-GGGSGGGS-GFP (29 kDa) to get a very large protein product with a mass of ~126 kDa. The ligation reaction between Z_{EGFR}-Fc-NQL (200 μM) and the GFP protein (500 μM) reached *ca*. 90% yield in the presence of QC (**Figure 3. 5** a).

This is remarkable considering the large sizes of the proteins and that the acceptor GFP protein substrate was used only at a 1.25 molar equivalence to the donor substrate Z_{EGFR}-Fc-NQL which, as a dimer, has two reaction sites. Furthermore, as seen from the non-reducing SDS-PAGE gel (**Figure S3. 9**), in the presence of QC, the final dual ligated product was predominant, whereas in the absence of QC, the mono-ligated product was predominant. **Figure 3. 5** b and c show the specific binding of dual-ligated product Z_{EGFR}-Fc-GFP towards A431 cells which overexpresses EGFR, indicating that the receptor binding activity of Z_{EGFR} and fluorogenicity of GFP were preserved after the ligation reaction. This example demonstrates that the VyPAL2-QC coupled method is also suitable for the preparation of large therapeutic protein conjugates.

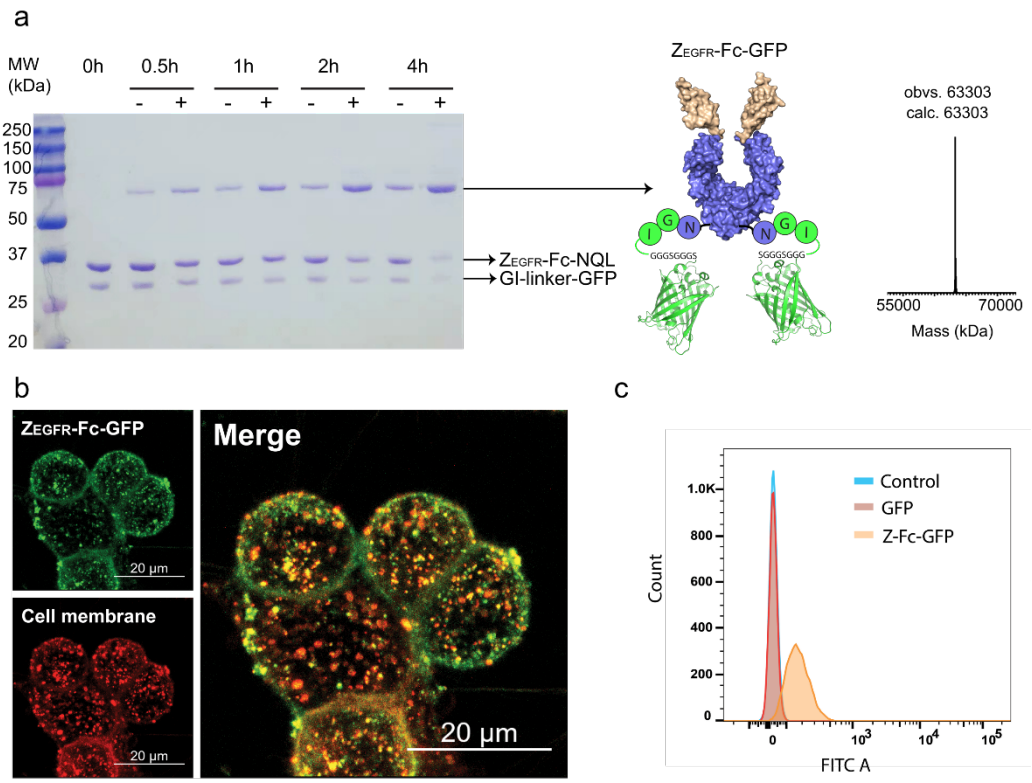


Figure 3. 5. Ligation between ZEGFR-Fc-NQL and GI-GGGSGGGS-GFP.
a) SDS-PAGE of the ligation reaction of ZEGFR-Fc-NQL (200 µM) with GI-GGGSGGGS-GFP (500 µM) using 0.4 µM VyPAL2 with or without 0.04 µM QC in 20 mM PBS (pH 7). The reaction mixture was treated with 50 mM DTT and analyzed by SDS-PAGE gel. The bands at ~63 kDa and 34 kDa correspond to the reduced forms of the ligation product and ZEGFR-Fc-NQL, respectively. The ligation product was also characterized by LC-MS (ESI) after reduction with DTT. b). Confocal microscopy image of ZEGFR-Fc-GFP binding to A431 cells c). Flow cytometry analysis of ZEGFR-Fc-GFP and GFP targeting A431 cells.

3.4. Conclusion

As the most powerful transpeptidases known to date, PALs have previously been shown to catalyze peptide and protein cyclization reactions very efficiently, with a k_{cat}/K_m that is at least one order of magnitude higher than that of intermolecular ligation reactions [54, 55]. This is attributed to the entropically favorable nature of the intramolecular reaction. Moreover, the rigid conformation of the cyclized products often makes them resistant to PALs. Therefore, despite being also a transpeptidation reaction, PAL-catalyzed cyclization is usually irreversible. This

is not the case for the bimolecular ligation reactions. Their reversibility generally limits the product yields to ≤ 50% at a 1:1 ratio between the two reaction partners. We show that this problem can be overcome by using a P1' Gln in the acyl donor substrates since its α-amine can be quenched by lactamization upon cleavage of the Asn-Gln peptide bond. Pyroglutamyl formation can occur spontaneously, but it is a slow process. The reported rate constant of spontaneous pGlu formation of an N-terminal Gln is 1.7 x 10^{-6} s^{-1} at pH 6, which corresponds to a half-life of about 4.7 days [123]. Human QC-catalyzed pGlu formation has a k_{cat} of 30 s^{-1}, representing a rate enhancement by seven orders of magnitude [123]. The high efficiency of QC makes it ideally suited for coupled use with PALs. In our cascade enzymatic scheme, QC was used at one tenth equivalence to the PAL enzyme which was used at 1:1000 or 1:2000 molar ratio to the substrate. Using this scheme, the yield of intermolecular ligations was greatly improved at equal or moderately higher molar equivalence of the acyl acceptor substrate to the acyl donor substrate. Our method is generally applicable with all PALs and to substrates of various sizes ranging from small peptides to large recombinant proteins. Compared to existing methods which utilize metal ions, synthetic chemicals or unnatural elements in the substrates to address the reversibility problem, this method uses an innocuous enzyme. The high reaction yields and need for very low quantities of the enzymes also facilitate product purification and make the process cost-effective. Overall, this robust cascade enzymatic scheme greatly increases the applicability of PAL-mediated ligation in the precision manufacturing of large protein conjugates.

3.5. Materials and methods

3.5.1. general methods

All the solvents and reagents were purchased from commercial suppliers and used without further purification. Human Glutaminyl Cyclase (QC) was purchase from Abcam (ab206806), aliquoted and stored at -80 °C.

Peptides were synthesized following standard Fmoc solid phase synthesis protocols. Synthesized peptides were purified using semi-preparative RP-HPLC. Semi-preparative RP-HPLC was performed using a Shimadzu HPLC system equipped with a Phenomenex-C18 RP column (10 × 250 mm, 5 μm) with a flow rate of 2.5 ml/min, eluting using a gradient of buffer B (90 % acetonitrile, 10 % H_2O, 0.045 % TFA) in buffer A (H_2O, 0.045 % TFA). All the synthesized compounds were stored at 4 ^oC or -20 ^oC.

Proteins were generated using recombinant DNA methods. For protein purification, Immobilized Metal Affinity Chromatography (IMAC), Protein A affinity chromatography and Size-Exclusion chromatography (SEC) were used. SEC was performed on the ÄKTA FPLC UPC-900 using HiLoadTM 16/600 SuperdexTM 200pg column. Protein A and NiNTA affinity chromatography was conducted on ÄKTAstart using HiTrap 5ml MabSelectTM column or HisTrap HP 5ml column, respectively.

For analysis, mass spectra for peptides were obtained using a Bruker Ultraflex Extreme Matrix Assisted Laser Desorption/Ionization (MALDI) Tandem TOF or electrospray ionization (ESI) mass spectroscopy (Thermo Fisher LTQ XL). Data from MALDI was analysed using Data Explorer software, and data from ESI was analysed using Thermo Xcalibur Qual Browser and Magtran software. The deconvolution of protein mass spectra was done using MagTran. Analytical reverse-phase HPLC (RP-HPLC) was performed on a Shimadzu HPLC system equipped with a Phenomenex-C18 RP column (4.6 × 150 mm, 5 μm, 100 Å) or a Phenomenex Jupiter-C4 column (4.6 × 150 mm, 3.6 μm, 200 Å) with a flow rate of 1.0 mL per minute, eluting with a gradient of buffer B (90% ACN, 10% H_2O, 0.045% TFA) in buffer A (H_2O, 0.045% TFA).

3.5.2. Solid phase peptide synthesis (SPPS)

All the peptides were synthesized as C-terminal amides using Rink amide MBHA resin by standard Fmoc chemistry using Liberty Blue Peptide Synthesizer. For

5(6)-carboxyfluorescein and biotin coupling to Lys(MTT) sidechain, the MTT protecting group was first removed using TFA/TIS/DCM (2.5%/2.5%/95%), followed by 5(6)-carboxyfluorescein (or biotin) coupling to the Lys sidechain amine using 2.5eq 5(6)-carboxyfluorescein (or biotin), 2.5eq Oxyma, 2.5 eq DIC in NMP for 3 h. For peptide cleavage from resin and deprotection of sidechain protecting groups at the end of SPPS, the peptidyl-resin was treated with a cocktail of TFA/H_2O/TIS (95%/2.5%/2.5%) for 1-3 hours. The cleavage solution was separated from the resin by filtration and the cleaved peptide was precipitated in the cold Et_2O. The crude product was isolated by centrifugation and purified by RP-HPLC. The peptide fractions after HPLC purification were lyophilized to afford the peptide in powder form.

3.5.3. OaAEP1$_b$-C247A expression, purification, and activation

OaAEP1$_b$-C247A was cloned into vector pET28a and expressed using T7 SHuffle E. *coli*. Pro-OaAEP1$_b$-C247A was activated at pH 4 in acetic buffer (0.1 M NaCl, 0.5 mM TCEP) for 2 h at 37 ^oC. After activation, the activated enzyme was purified by size-exclusion chromatography (SEC) at pH 7 (20 mM PBS, 0.1 M NaCl). Purified enzyme was stored at -80 °C in 5% glycerol, pH 7 (20mM PBS, 0.5 mM TCEP).

3.5.4. VyPAL2 expression, purification, and activation

VyPAL2 was expressed using sf9 insect cells. 100 mL of the viral vector containing VyPAL2 gene was used to infect sf9 cells at cell density of 2.5 x 10^6 cells/mL. MOI for infection was set between 1-10 for protein expression. The culture was incubated in a 27 °C shaker for 3 days (72 hours) at 135 rpm. Protein purification was performed in three steps: Immobilized Metal Affinity Chromatography (IMAC), Ion-Exchange Chromatography (IEX), and Size-Exclusion chromatography (SEC). Pro-VyPAL2 was activated at pH 4.5 in 50 mM sodium citrate buffer (0.1 M NaCl, 1mM DTT, 0.5 mM LS) for 2-3 h at 37 ^oC. After activation, the activated enzyme was purified by SEC at pH 6.5 (20 mM PBS, 0.1 M NaCl, 1mM DTT). Purified enzyme was stored at -80 °C in 5%

glycerol, pH 7 (20mM PBS, 0.5 mM TCEP).

3.5.5. DARPin9_26, GFP, ubiquitin, Z_{EGFR} and Z_{EGFR}-Fc expression and purification

All protein genes were cloned in pET28a, pET3a, pTxB1 or pETDuet, and expressed in E. coli (DE3) or T7 SHuffle. The expressions (except for GI-ubi) were induced with 0.1-0.4 mM IPTG after the OD_{600} of bacteria reached 0.4-0.6 in Luria Bertani broth (Kana or Amp) at 37 °C or 30 °C. After induction, the cells were incubated at 16 °C for 18 h. The cells were harvested by centrifugation (4000 x g, 10 min) and resuspended in lysis buffer (50 mM PBS, 0.1 M NaCl, 10 mM imidazole, 0.01% 100X triton, pH7.5). The solution mixture was lysed using ultrasonicator probe (Vibra cell ™) with alternative cycles of 3 s pulse after every 8 s interval for 15 min on ice. The protein solution was then centrifuged at 15000 x g (20 min) at 4 °C, filtered using 0.2 µm membrane, and bound to NiNTA beads or protein A beads for 1 h at 4 °C. The Ni beads were washed with 20 mM imidazole, 0.1 M NaCl, 20 mM PBS buffer (pH 7.5), then protein was eluted using 500 mM imidazole, 0.1 M NaCl, 20 mM PBS buffer (pH 7.5). The protein A beads were washed with 20mM PBS (pH7.5), then the protein was eluted using 30 mM citrate buffer (pH 3.5). For GI-ubi, it was expressed as a C-terminal intein fusion protein in E.coli (DE3), the protein solution was bound to chitin beads and the GI-ubi was cleaved from bounded intein by incubating in 50 mM DTT, 20 mM PBS (pH 8) ON, at RT. The masses (ESI-MS) of all the proteins have been measured (**Figure S3. 12**). All the proteins were exchanged into 20mM PBS (pH 7) and stored at 4°C for short term and -20 °C for long term. T

3.5.6. Ligation of model peptides

Enzyme-meditated ligation reactions were performed in 20 mM PBS buffer (pH 6.5 or pH 7) at 37 ºC for various time courses with or without QC. The ratio of QC to ligase to substrate (NQL peptide) is 0.1:1:2000. The reactions were quenched by 10% TFA and monitored by analytical RP-HPLC. The ligated products were characterized by MALDI-MS or ESI-MS.

3.5.7. Protein N or C terminal labelling

The ligation reactions were conducted at pH 7 under 37 °C for various time courses with or without QC. The ratio of QC/ligase/protein substrate is 0.1/1/1000. The reactions were quenched by 6 M Guanidine-HCl (pH 3) and the reaction was monitored by analytical RP-HPLC. The ligated products were characterized by ESI-MS.

3.5.8. Protein-protein ligation

The ligation reactions were conducted at pH 7 under 37 °C for various time courses with or without QC. The ratio of QC/ligase/protein substrate is 0.1/1/1000 (500). The reactions were quenched by 6M Guanidine-HCl (pH 3) and the completion reaction was monitored by analytical RP-HPLC. The ligated products were characterized by ESI-MS. The ligation reaction of Z_{EGFR}-Fc-NQL and GI-GGGSGGGS-GFP was analysed by SDS-PAGE under reducing or non-reducing conditions (reducing condition: 50 mM DTT, pH 8.8 for 20 min).

3.5.9. Purification of ligated proteins

The ligated Z_{EGFR}-Fc -GFP protein was purified by Size-Exclusion chromatography (SEC) at pH 7 (20 mM PBS, 0.1 M NaCl). The purified protein was stored at - 20 °C.

3.5.10. Flow cytometry assay

To study binding capacities of Z_{EGFR}-Fc-GFP, A-431 (ATCC, USA) and MCF-7 (ATCC, US) live cells were washed with PBS (HyClone, USA) for three times, trypsinized by 0.05% Trypsin-EDTA (Gibco, USA), and then resuspended in chilled DMEM (Gibco, USA) with 10% FBS (Gibco, USA). One million cells of each cell line were then incubated with Z_{EGFR}-Fc-GFP (100 nM) and GFP (100 nM) on ice for 30 min. After incubation, the cells were washed with chilled PBS for three times and analyzed by the Fortessa X-20 flow cytometer (BD, USA) (**Figure S3. 10**). The cytometer was set to record 10,000 events per sample, to excite the fluorophore with 488 nm laser, and to collect emitting fluorescent

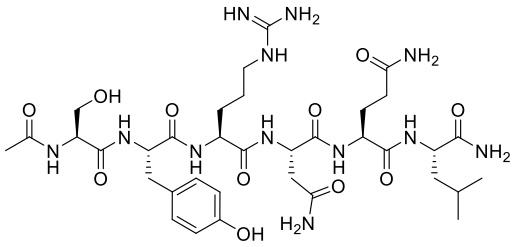
signals in 530/30 nm. The generated raw data were analyzed by Flowjo™10 (BD, USA).

3.5.11. Live cell confocal imaging

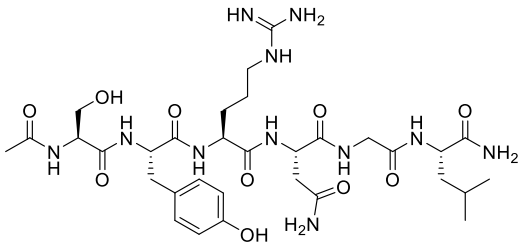
To visualize Z_{EGFR}-Fc -GFP binding activities, A431 and MCF-7 cells were seeded on an 8-well chamber slide (ibidi, USA) and incubated at 37 ºC under 5% CO_2 overnight. The cells were stained with 2 μM PKH26 red-fluorescent dye (Sigma, USA) for 10 min at 37 ºC. The stained cells were then incubated with Z_{EGFR}-Fc-GFP (100nM) and GFP (100nM) on ice for 30 min. After incubation, the cells were washed with chilled PBS for three times and fixed with cold 4% formaldehyde for 15 min. The fixed cells were imaged by the LSM 980 confocal microscope (Zeiss, Germany) (**Figure S3. 11**). Microscopic key settings are as follows: 1) excitation laser wavelengths: 488 nm and 561 nm; 2) emission fillers: 507-552 nm and 575-620 nm; 3) imaging mode: Z-stack. The 3D Z-stack images were processed into 2D images by the technique MIP (maximum intensity projection) in the Zen software (Zeiss, Germany).

3.6. Appendices

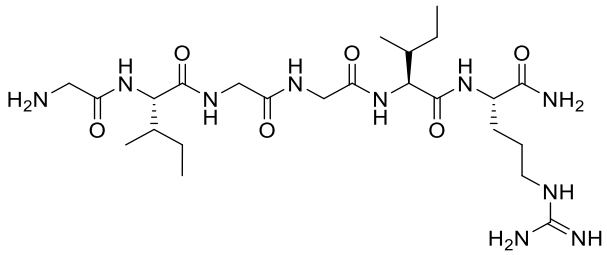
Protein sequences



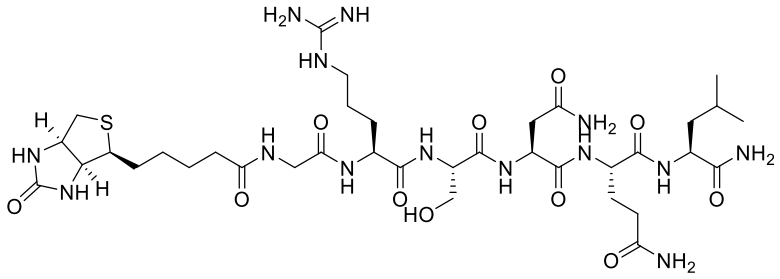
Ac-SYRNQL (calc. 820.4 , obvs. 821.5[H+M$^+$])



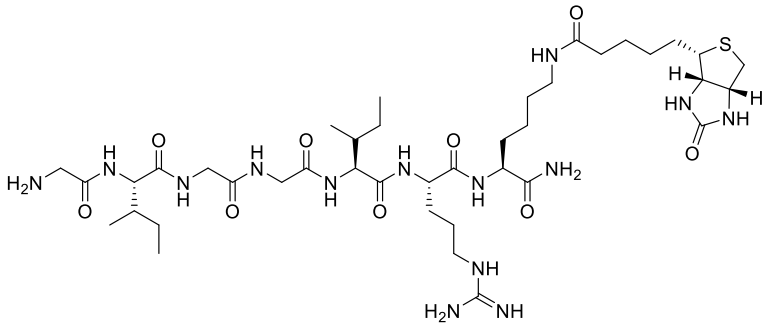
Ac-SYRNGL (calc. 749.4, obvs. 750.5[H+M$^+$])



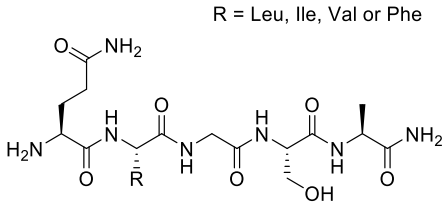
GIGGIR (calc. 570.4, obvs. 571.5[H+M$^+$])



Biotin-GRSNQL (calc. 989.4, obvs.899.8[H+M$^+$])



GIGGIRK(biotin) (calc. 925.2, obvs. 925.7[H+M$^+$])



QLGSA (calc. 473.5, obvs. 474.5[H+M$^+$])

QFGSA (calc. 507.5, obvs.508.3[H+M$^+$])

QIGSA (calc. 473.5, obvs. 474.3[H+M$^+$])

QVGSA (calc. 459.5, obvs. 460.5[H+M$^+$])

Ubi-NQL-His (calc. 9743, obvs. 9746)

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLE
DGRTLSDYNIQKESTLHLVLRLRGGNQLHHHHHH

GI-Ubi (calc. 8735, obvs. 8737)

GIMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQ
LEDGRTLSDYNIQKESTLHLVLRLRGG

GI-GFP (calc. 28344, obvs. 28341)

MGIGSKKVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGK
LTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPE
GYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH
KLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTP
IGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDE
LYKGSGHHHHHH

GI-GGGSGGGS-GFP (calc. 28803, obvs. 28802)
MGIGSGGGSGGGSKKVSKGEELFTGVVPILVELDGDVNGHKFSVRGEG
EGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRH
DFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDF
KEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQL
ADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTA
AGITHGMDELYKGSHHHHHH

DARPin-NQL (calc. 18846, obvs. 18853)
MHHHHHHGSDLGKKLLEAARAGQDDEVRILMANGADVNAKDFYGIT
PLHLAAAYGHLEIVEVLLKHGADVNAHDWNGWTPLHLAAKYGHLEIV
EVLLKHGADVNAIDNAGKTPLHLAAAHGHLEIVEVLLKYGADVNAQD
KFGKTPFDLAIDNGNEDIAEVLQKAAKLGSGSNQL

GI-DARPin (calc. 18530, obvs. 18538)
MGISSHHHHHHGSDLGKKLLEAARAGQDDEVRILMANGADVNAKDFY
GITPLHLAAAYGHLEIVEVLLKHGADVNAHDWNGWTPLHLAAKYGHL
EIVEVLLKHGADVNAIDNAGKTPLHLAAAHGHLEIVEVLLKYGADVNA
QDKFGKTPFDLAIDNGNEDIAEVLQKAAKLN

Z_{EGFR}-NQL (calc. 9085, obvs. 9086)
MKKGSSHHHHHHLQVDNKFNKEMWAAWEEIRNLPNLNGWQMTAFIA
SLVDDPSQSANLLAEAKKLNDAQAPKVDGSGSNQL

Z_{EGFR}-Fc-NQL (calc. 34759, obvs. 34769)

MKKGSSHHHHHHLQVDNKFNKEMWAAWEEIRNLPNLNGWQMTAFIA
SLVDDPSQSANLLAEAKKLNDAQAPKVDGSGSDKTHTCPPCPAPELLG
GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEV
HNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAP
IEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEW
ESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMH
EALHNHYTQKSLSLSPGKGSNQL

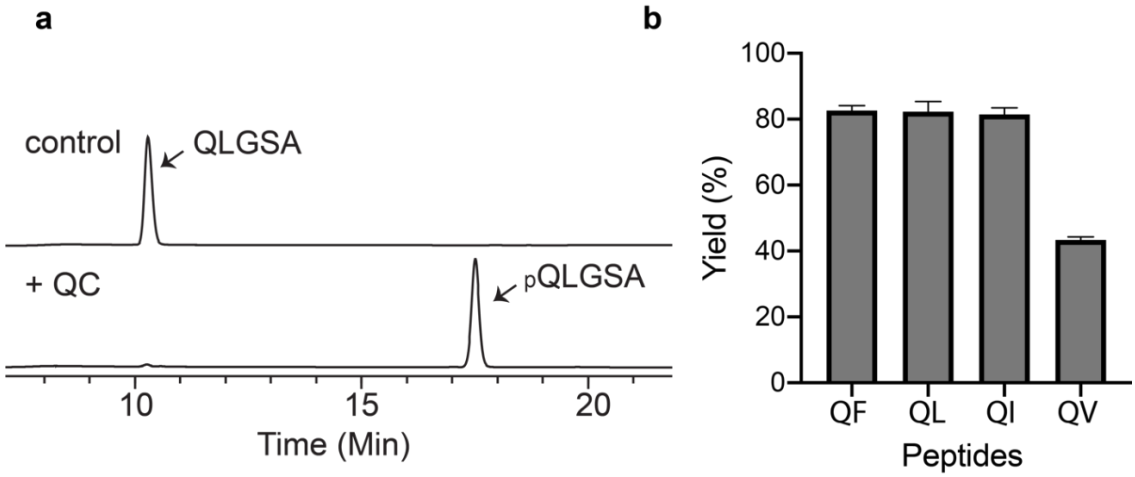


Figure S3. 1. QC-catalyzed pyro-glutamate (pGlu) formation for 4 substrates (QFGSA, QLGSA, QIGSA and QVGSA).
The reactions were performed using 5 mM of Q**X**GSA, 0.0001eq of QC at 37 ºC for 15 min in 20 mM PBS (pH 7) and monitored by RP-HPLC. All Bar charts represent mean ± SEM from triplicated measurements.

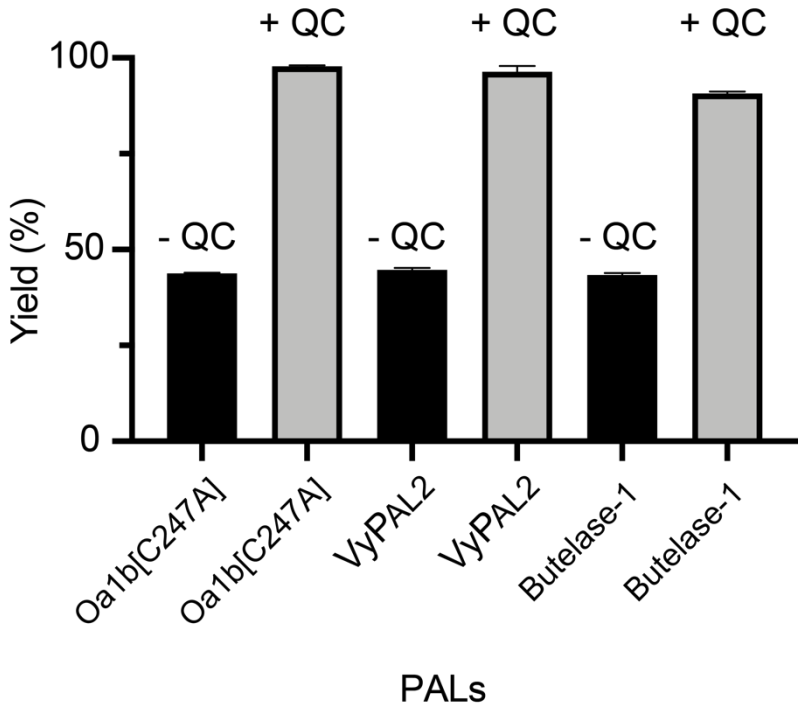


Figure S3. 2. Yields of ligation between Ac-SYRNQL and GIGGIR catalyzed by different PALs (OaAEP1$_b$-C247A, VyPAL2 or butelase-1) in the absence or presence of QC.

The ligation reactions were performed using 5 mM Ac-SYRNQL, 5 mM GIGGIR, 0.0005 eq of PAL, with or without 0.00005 eq of QC, at 37 ºC for 30 min in 20 mM PBS pH 7 (OaAEP-C247A) or pH 6.5 (VyPAL2, butelase-1). All Bar charts represent mean ± SEM from triplicated measurements.

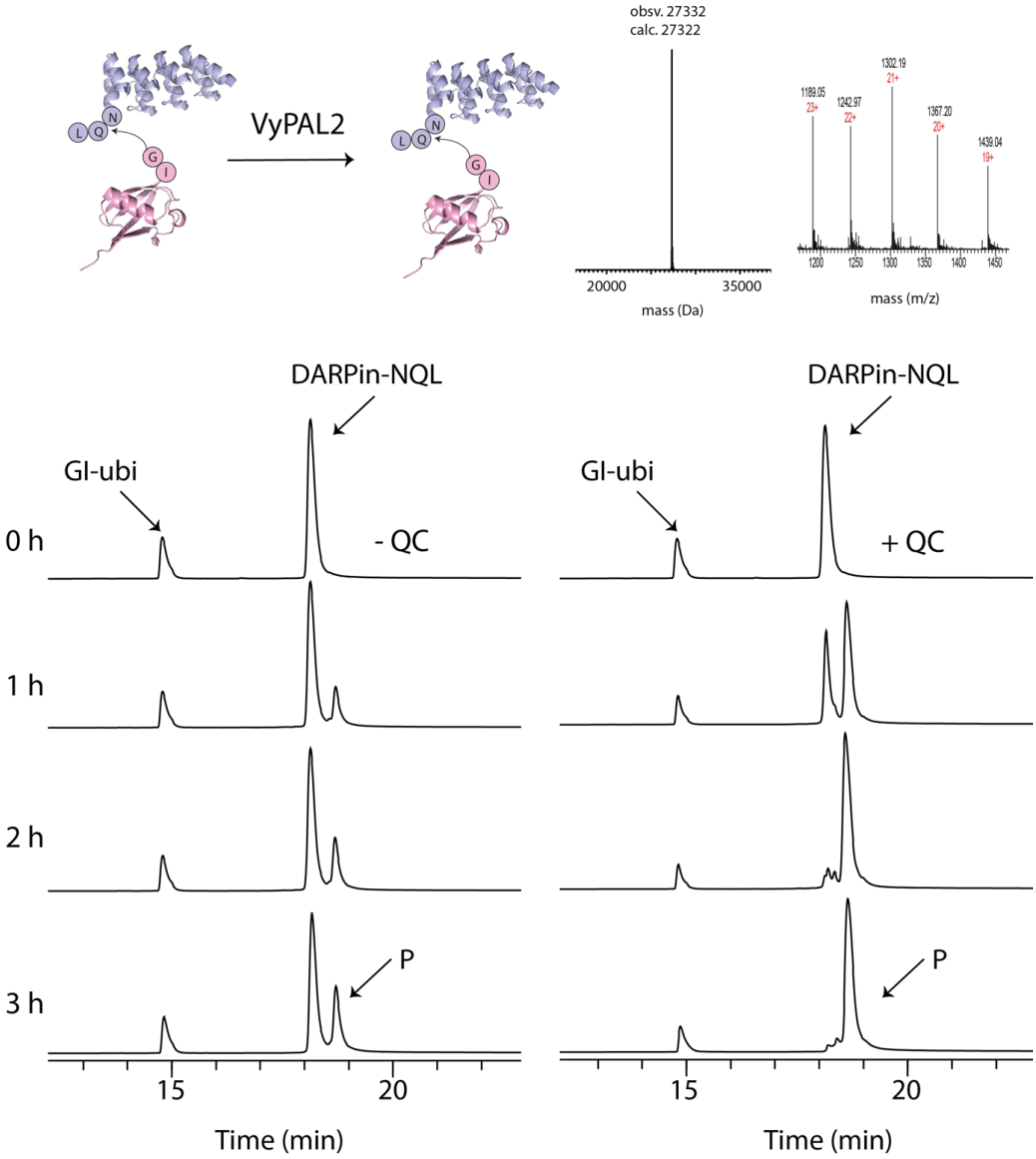


Figure S3. 3. RP-HPLC monitoring of the ligation reaction between DARPin-NQL and GI-ubiquitin at different time points.
The ligated product was analyzed using ESI-MS. The ligation reaction was performed at 37 °C using 400 µM DARPin-NQL, 1.8 eq of GI-ubiquitin, 0.001 eq of VyPAL2 in 20 mM PBS (pH 7), in the absence or presence of 0.0001 eq of QC.

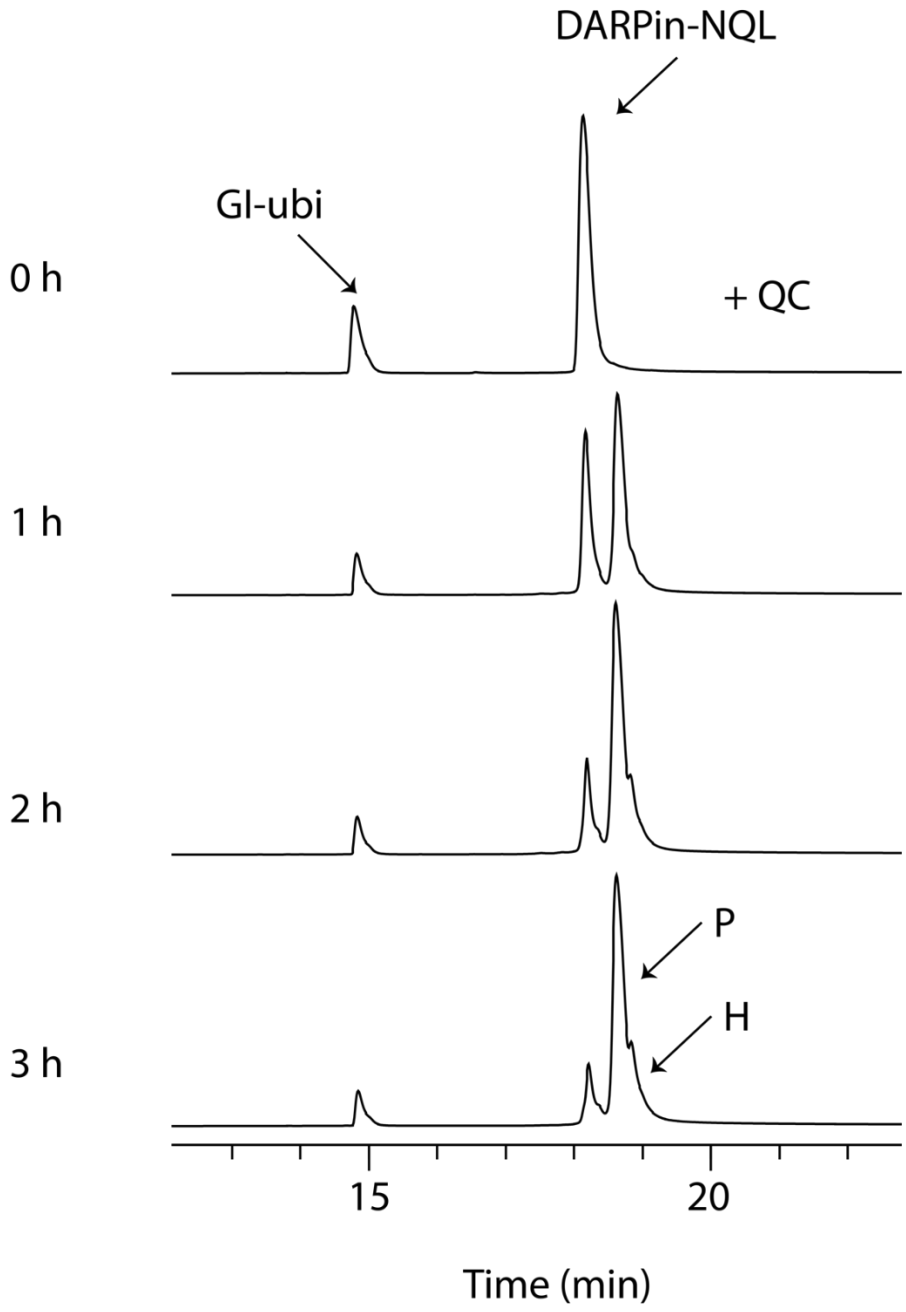


Figure S3. 4. RP-HPLC monitoring of the ligation reaction between DARPin-NQL and GI-GFP at different time points.
The ligated product was analyzed using ESI-MS. The ligation reaction was performed at 37 ºC using 400 µM DARPin-NQL, 1.8 eq of GI-GFP, 0.001 eq of $OaAEP1_b$-C247A and 0.0001 eq of QC in 20 mM PBS (pH 7). The reaction led to about 15% of the hydrolysis product DARPin-N-OH (Peak H).

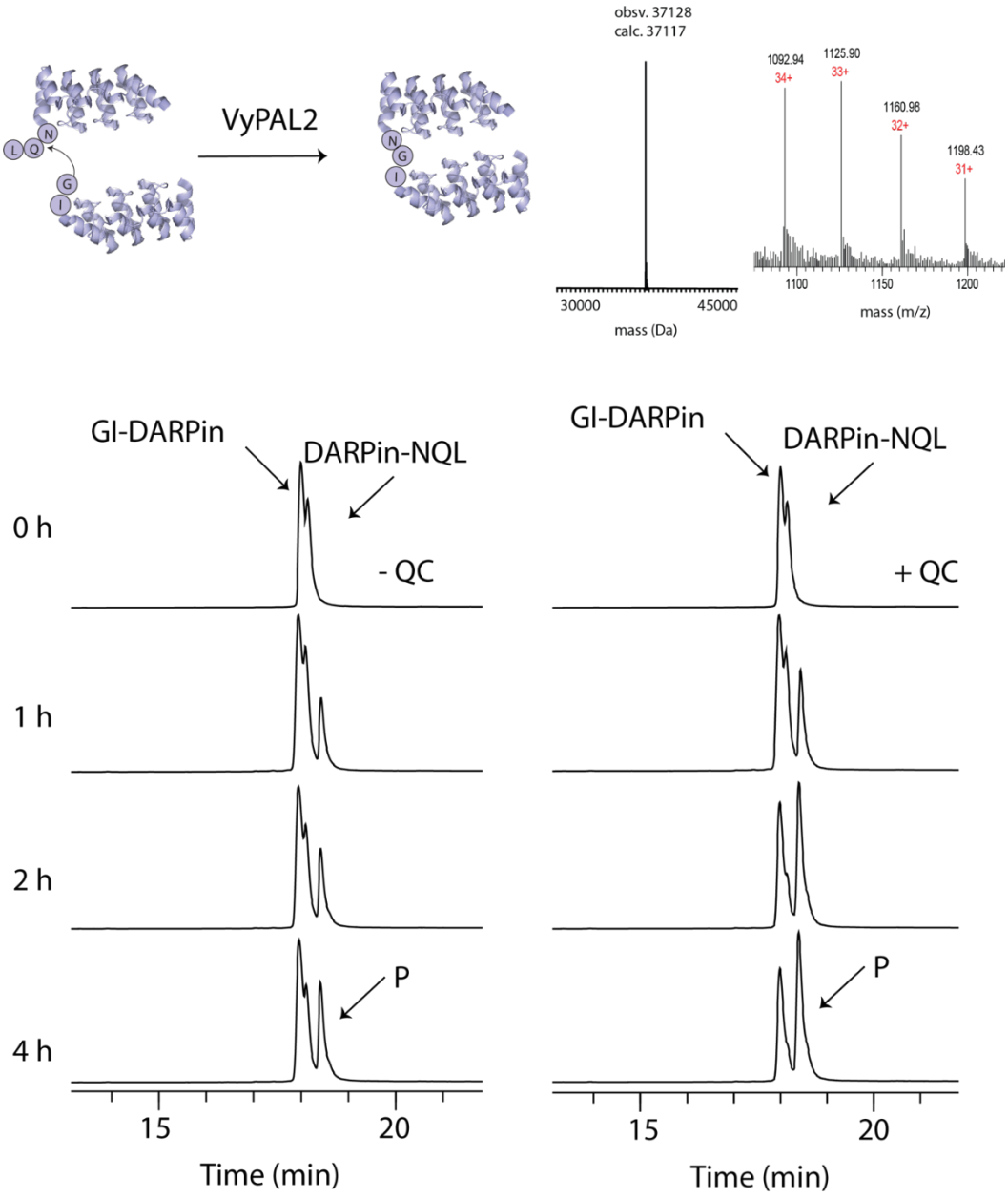


Figure S3. 5. RP-HPLC monitoring of the ligation reaction between DARPin-NQL and GI-DARPin at different time points.
The ligated product was analyzed using ESI-MS. The ligation reaction was performed at 37 ºC using 400 µM DARPin-NQL, 1.8 eq of GI-DARPin, 0.001 eq of VyPAL2 in 20 mM PBS (pH 7), in the absence or presence of 0.0001 eq of QC.

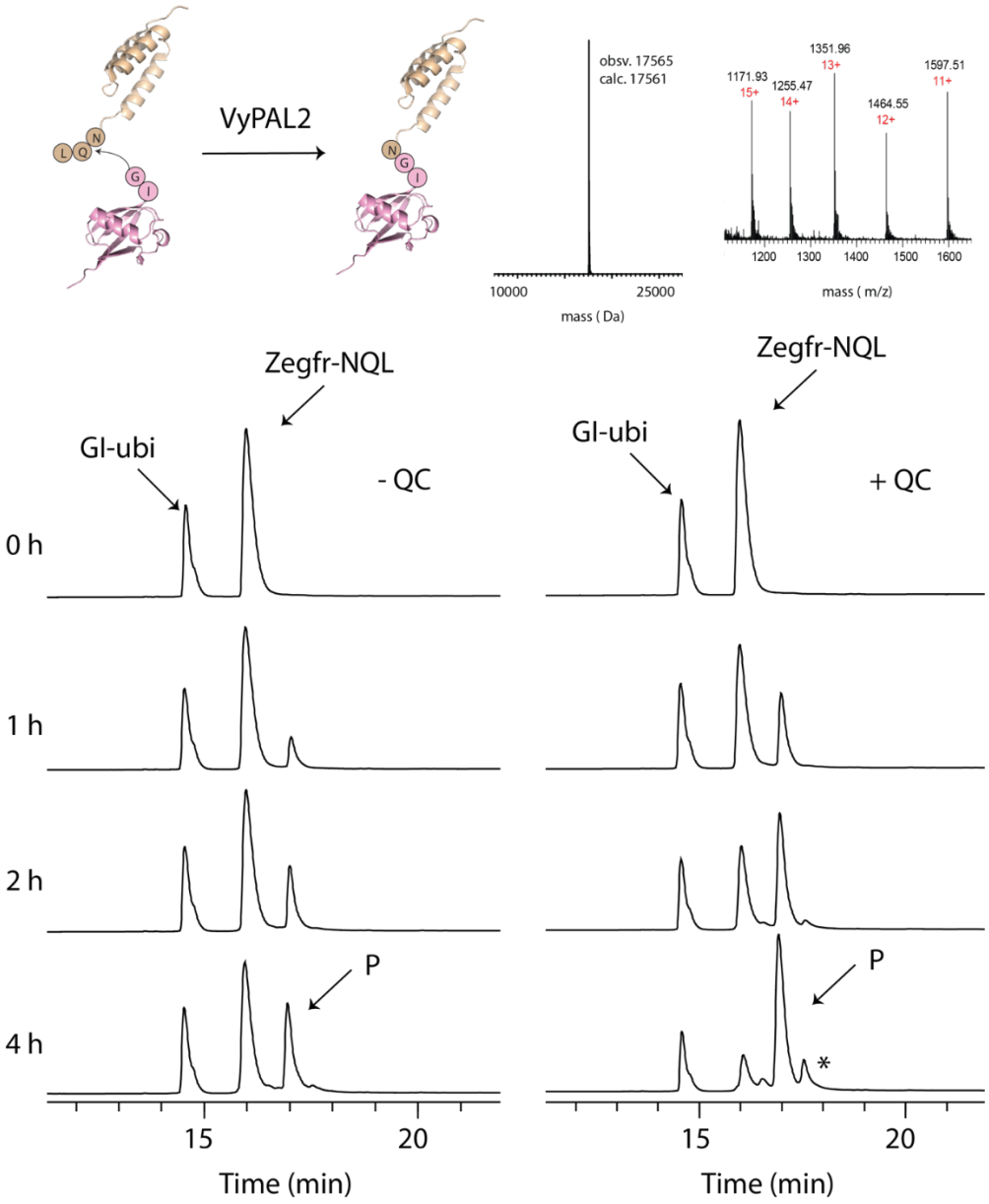


Figure S3. 6. RP-HPLC monitoring of the ligation reaction between Z_{EGFR}-NQL and GI-ubiquitin at different time points.
The ligated product was analyzed using ESI-MS. The ligation reaction was performed at 37 °C using 400 µM DARPin-NQL, 1.8 eq of GI-ubiquitin, 0.001 eq of VyPAL2 in 20 mM PBS (pH 7), in the absence or presence of 0.0001 eq of QC. *peak = the $[Z_{EGFR}]_2$-ubi by-product.

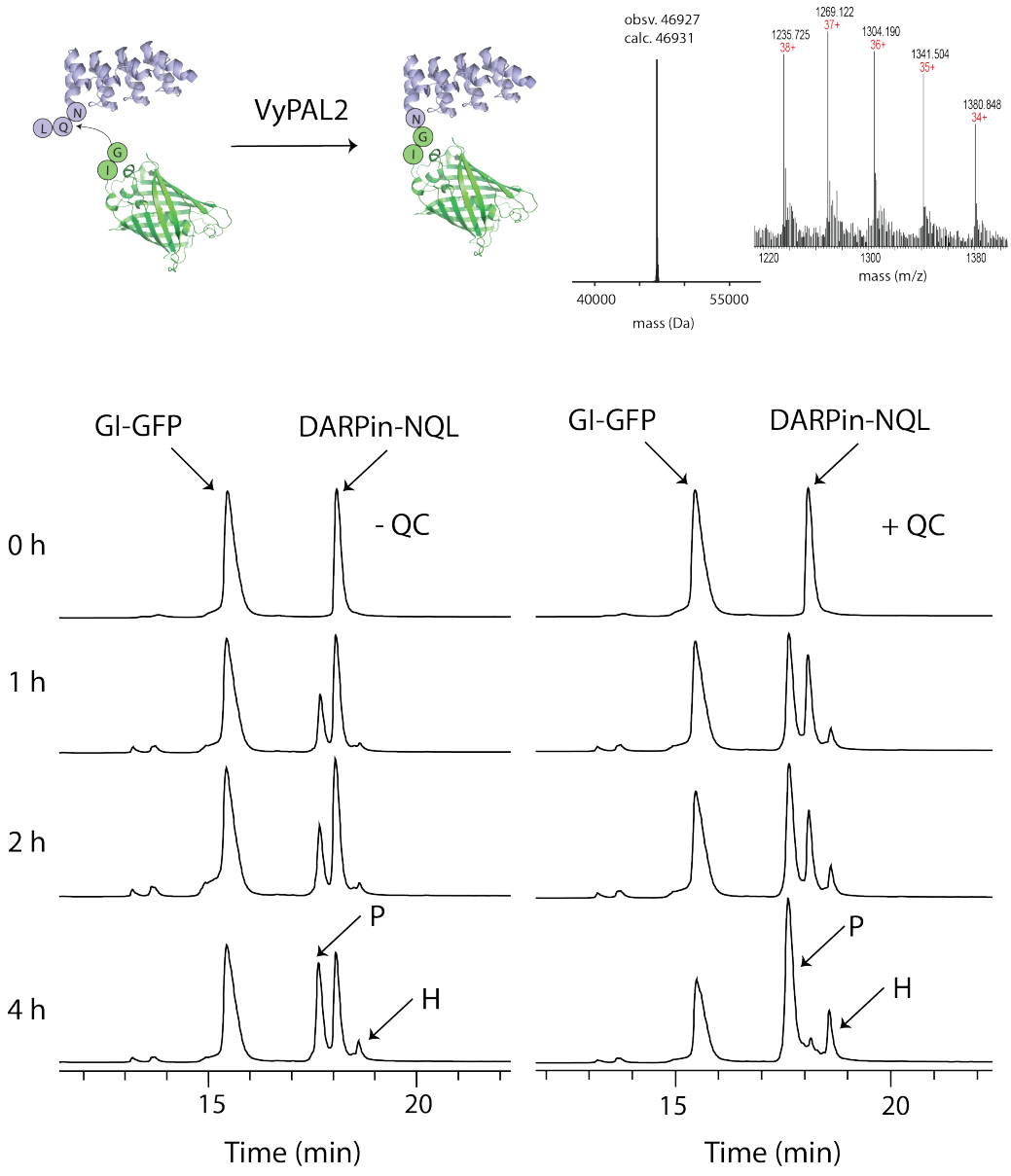


Figure S3. 7. RP-HPLC monitoring of the ligation reaction between DARPin-NQL and GI-GFP at different time points.
The ligated product was analyzed using ESI-MS. The ligation reaction was performed at 37 ºC using 400 µM DARPin-NQL, 1.8 eq of GI-GFP, 0.001 eq of VyPAL2 in 20 mM PBS (pH 7), in the absence or presence of 0.0001 eq of QC. Peak H = hydrolysis product DARPin-N-OH.

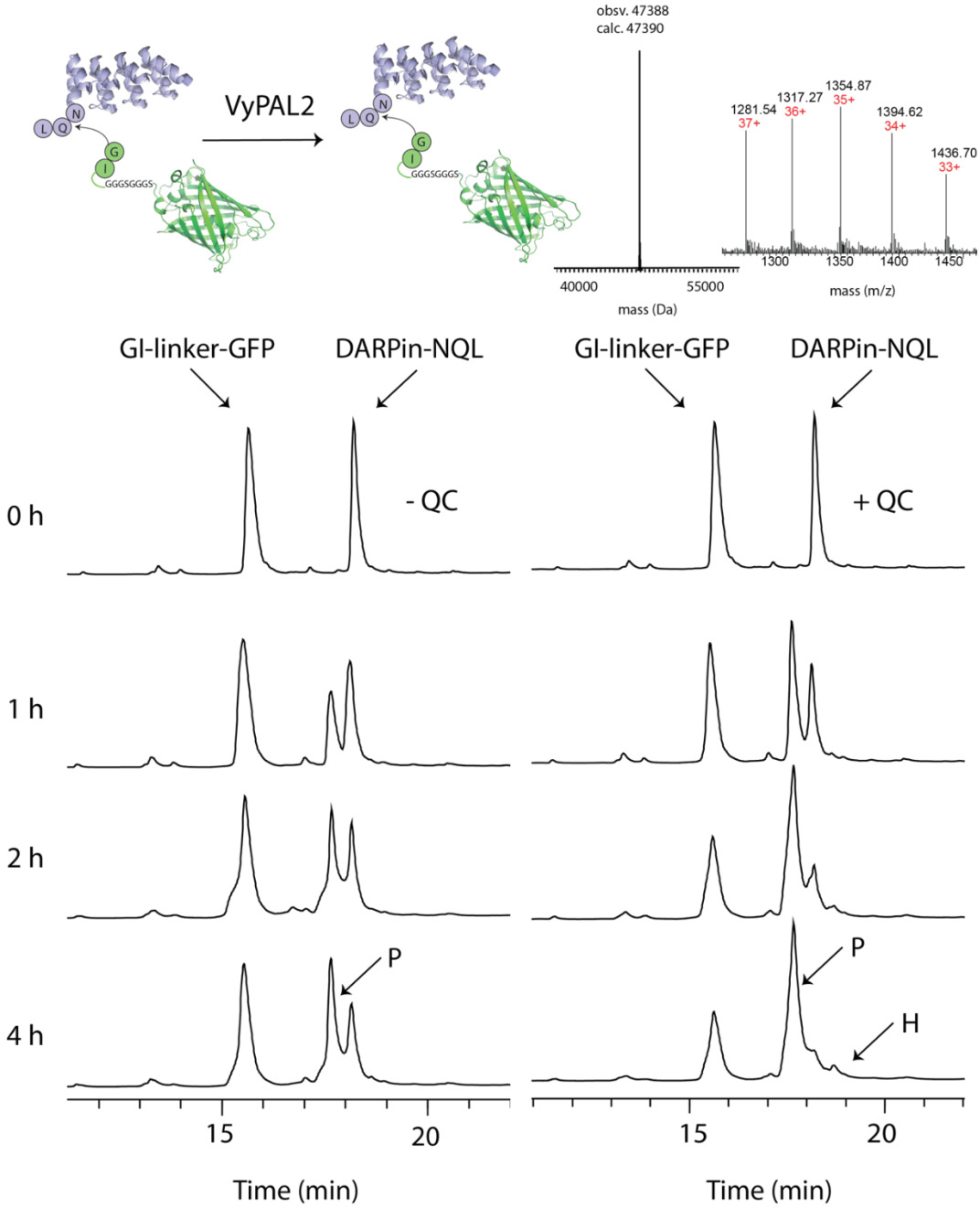


Figure S3. 8. RP-HPLC monitoring of the ligation reaction between DARPin-NQL and GI-linker-GFP ligation at different time points.
The ligated product was analyzed using ESI-MS. The ligation reaction was performed at 37 ºC using 400 µM DARPin-NQL, 1.8 eq of GI-linker-GFP, 0.001 eq of VyPAL2 in 20 mM PBS (pH 7), in the absence or presence of 0.0001 eq of QC.

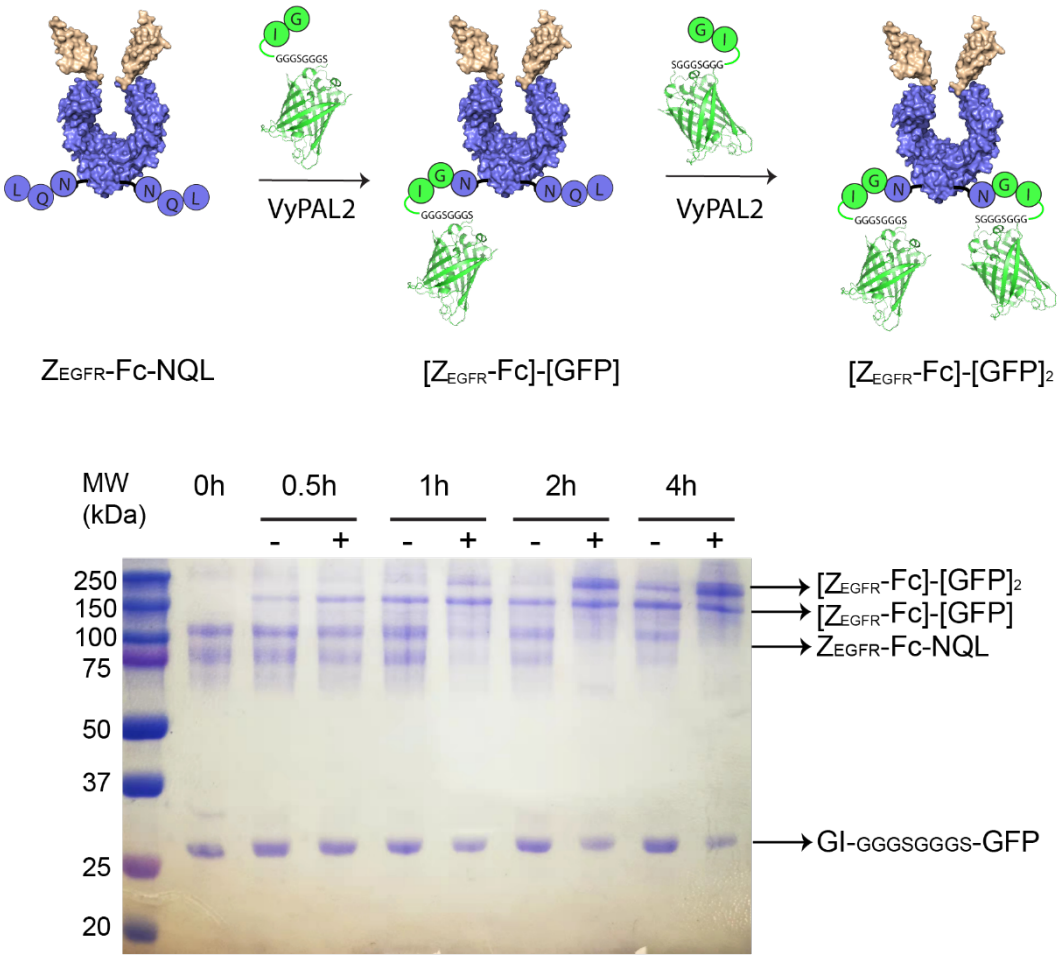


Figure S3. 9. Monitoring the ligation reaction between Z_{EGFR}-Fc-NQL and GI-GGGSGGGS-GFP at different time points by non-reducing SDS-PAGE gel.
The ligation reaction was performed at 37 °C using 200 μM Z_{EGFR}-Fc-NQL, 500 μM GI-GGGSGGGS-GFP, and 0.4 μM VyPAL2 in 20 mM PBS (pH 7), in the absence or presence of 0.04 μM QC.

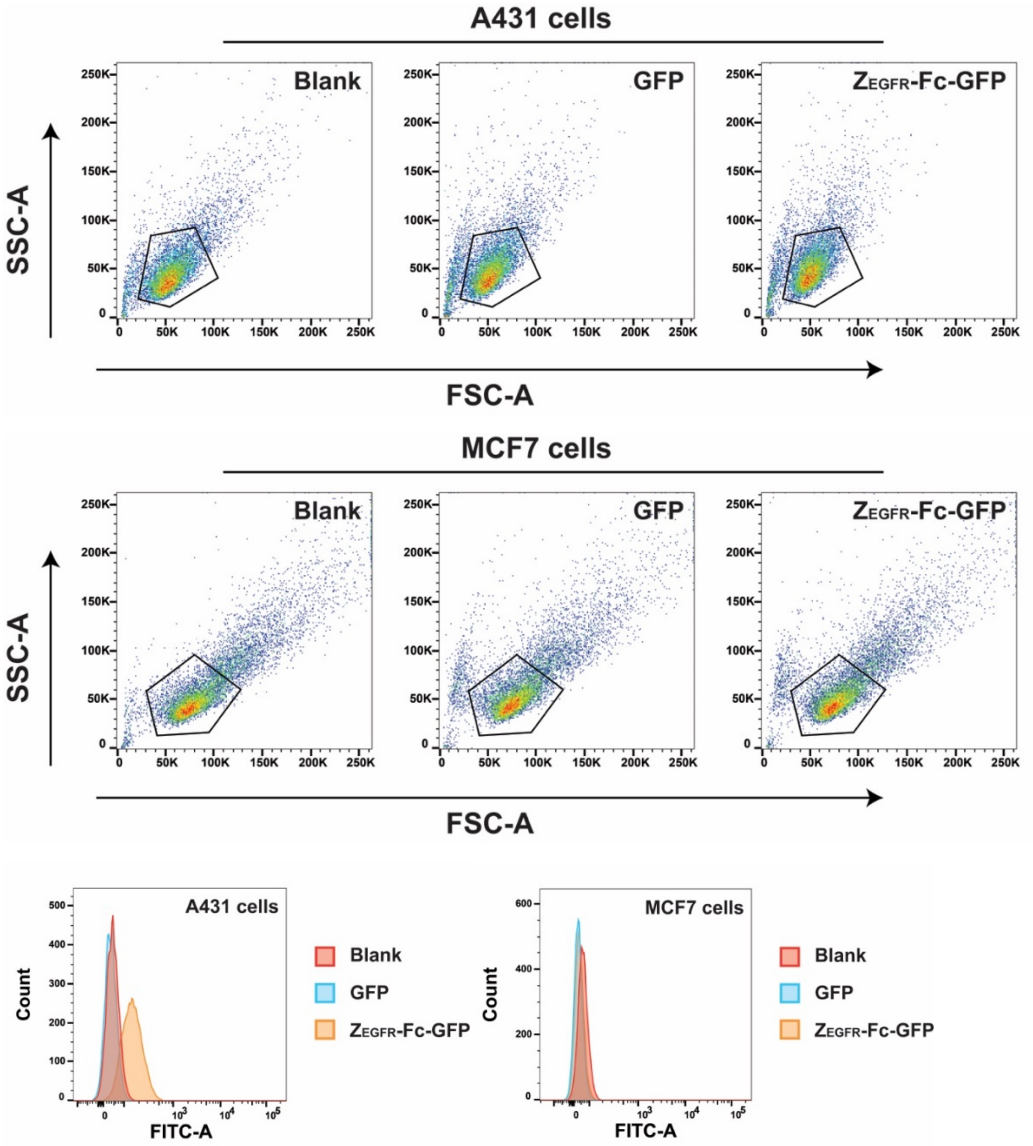


Figure S3. 10. Z_{EGFR}-GFP binding to A431 and MCF7 cells.
Fluorescence intensity (measured by flow cytometry, λex = 488 nm) of untreated A431 and MCF7 cells (ctrl) shown in red, cells treated with either GFP (100 nM, 30min) shown in blue, or with Z_{EGFR}-Fc-GFP (100 nM, 30min) in orange, respectively. Forward scatter (FSC-A) versus side scatter (SSC-A) were used to gate intact cells.

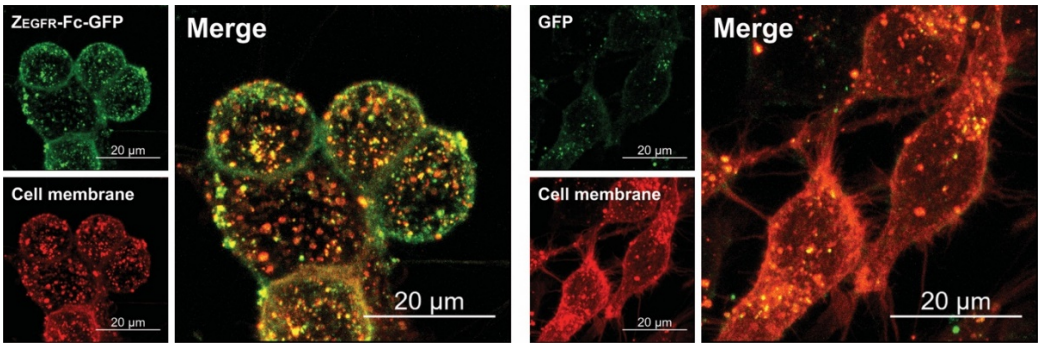
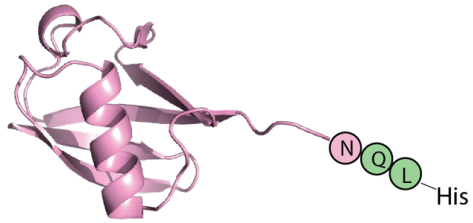
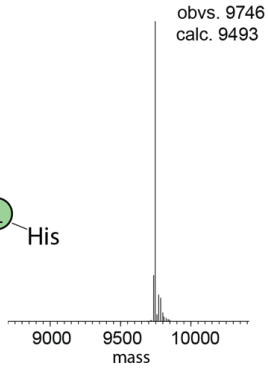
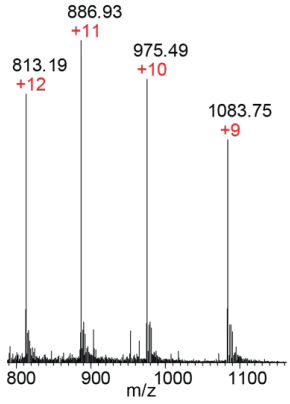


Figure S3. 11. Confocal fluorescence microscopy images.
Confocal fluorescence microscopy images of EGFR-positive A431 cells after incubating with Z_{EGFR}-Fc-GFP or GFP, respectively. Both Z_{EGFR}-Fc-GFP and GFP are shown in green, and the plasma membrane was stained with PKH26 red-fluorescent dye. Scale bar = 20 µm.

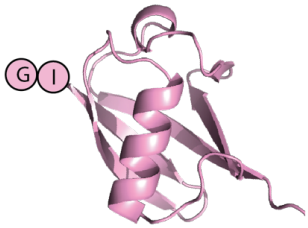
Ubi-NQL-His

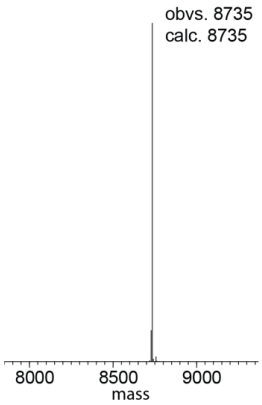


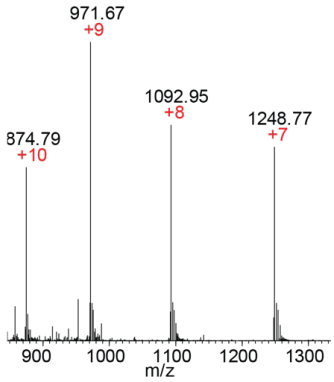




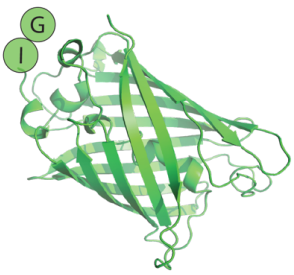
Gl-Ubi

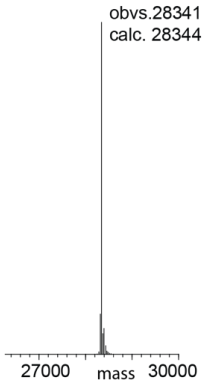


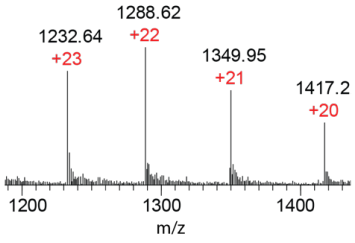




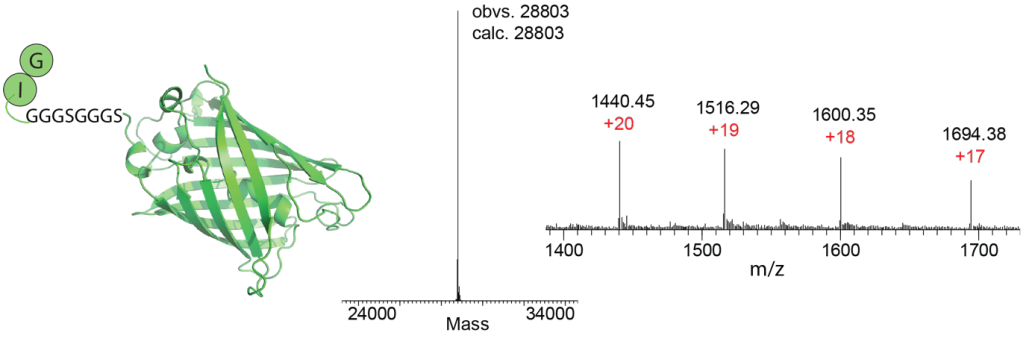
Gl-GFP



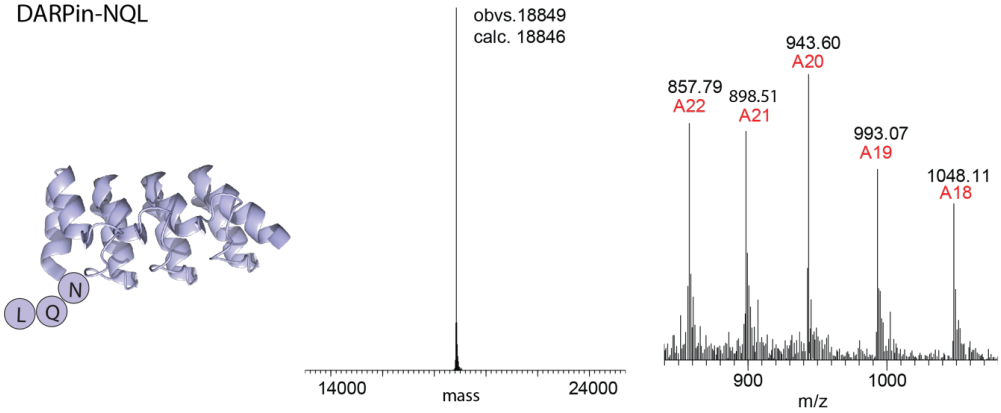




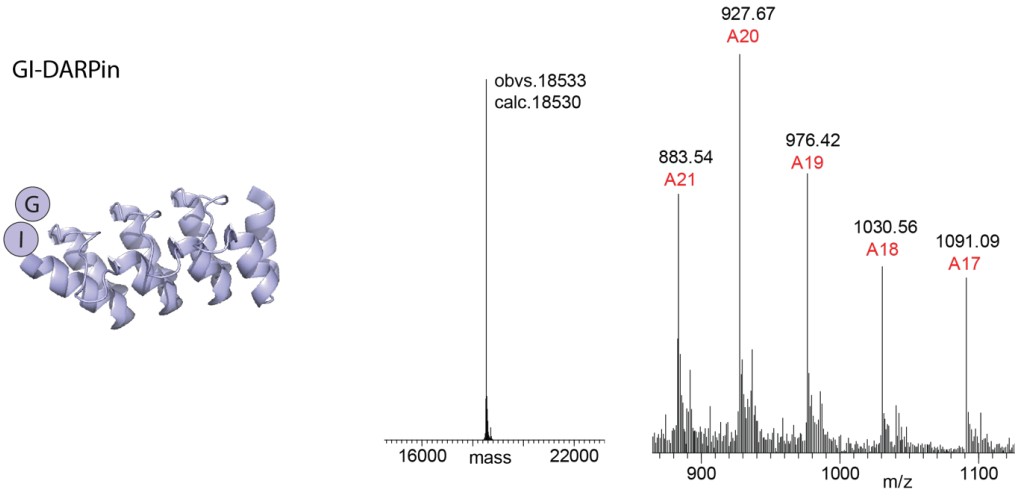
GI-GGGSGGGS-GFP



DARPin-NQL



GI-DARPin



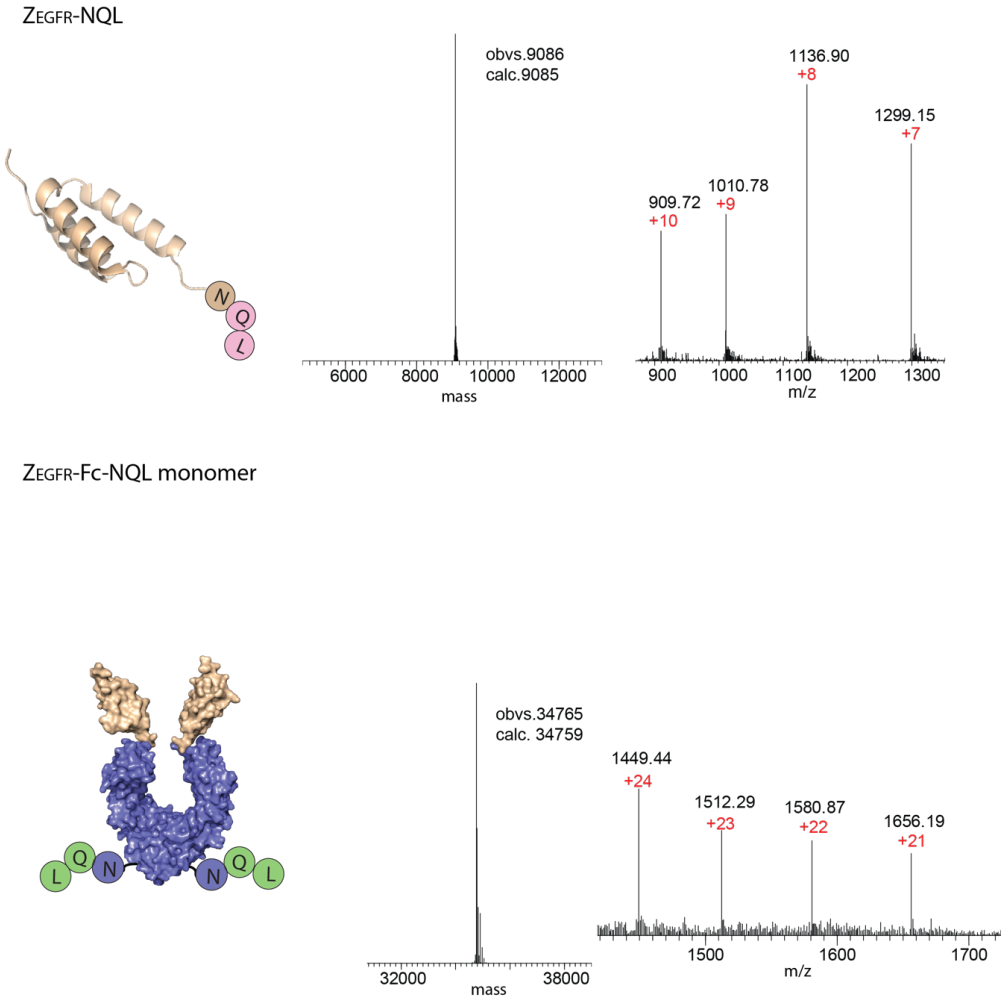


Figure S3. 12. Mass spectra of protein substrates.

Chapter 4. Conclusion and outlooks

In this thesis, I described two projects which were aimed to further improve PAL-mediated ligation (PML) and increase its applications in biotechnology and drug discovery. PALs are already a very powerful tool when used to perform peptide/protein macrocyclization and site-specific bioconjugation. Nevertheless, further improvements are warranted to overcome certain limitations of PML. One limitation is related to PALs' preference for Asn at the P1 residue in their substrates, which makes cyclizing linear substrates at the Asp residue difficult. Another limitation lies with the reversibility of intermolecular ligation catalysed by these transpeptidases, which hinders the reaction to proceed with high efficiency.

To overcome the first limitation, I have designed derivatives of asparagine that may be recognized by PALs as a P1-site substrate. Using a peptide cyclization enzymatic assay, Asn(OH) was found to be a good Asn mimic for PAL recognition. It has been shown that linear peptides containing a P1-Asn(OH) residue are much better substrates of PALs than the corresponding P1-Asp peptides. Using P1-Asn(OH)-enabled cyclization, a wide range of cyclic bioactive peptides were prepared through the catalysis of PALs. We have also developed a method that allows facile conversion of Asn(OH) to Asp using a mild oxidation reaction. Many of these peptides could not be cyclized by the PAL enzyme when Asp was used directly as the P1 residue due to very slow reaction kinetics. More importantly, we have demonstrated that cyclization of Asn(OH)-containing peptides is well-suited for generating highly potent and selective inhibitors for MMP2. The hydroxamic acid group in Asn(OH) is known as a chelator for metal ions, which accounts for the high inhibitory activity of these cyclic peptides against MMP2. Our backbone-cyclic MMP2 inhibitors also have excellent proteolytic stability due to the presence of three disulphide bonds.

Therefore, my work in this project has broadened the substrate scope of PALs from natural Asx to unnatural Asx derivatives, Asn(OH) in particular. We envision that this method will find important applications in the discovery and development of cyclic peptides for therapeutic use. We also believe that similar substrate engineering strategies can be used to expand the substrate scope of other enzymes.

To overcome the second limitation of PML, I have developed a cascade enzymatic scheme to suppress the reverse reaction of transpeptidation catalysed by PALs. We show that by using a P1' Gln in the acyl donor substrates and glutaminyl cyclase (QC) to quench its α-amine in a lactam upon release of the glutaminyl leaving group, PAL-mediated ligation becomes irreversible. QC is a highly active and selective enzyme for pyroglutamyl formation. In our study, QC was used at one tenth equivalent to the PAL enzyme which was used at 1:1000 or 1:2000 molar ratio to the substrate. By coupling QC-catalyzed pyroglutamyl formation with PAL-mediated transpeptidation, near quantitative yields of intermolecular PAL ligations were achieved even at an equal ratio between the acyl acceptor substrate and the acyl donor substrate. The generality of this method is demonstrated with all PALs and with substrates of various sizes ranging from small peptides to large recombinant proteins. Existing methods which utilize metal ions and reactive chemicals or unnatural elements in the substrates to address the reversibility problem. Our method stands out for being much more practical and efficient since it involves the use of QC, an innocuous enzyme. Metal ions and reactive chemicals are not very biocompatible when it comes to sensitive substrates like large proteins and live cells. On the other hand, human QC is a naturally occurring enzyme and poses no biocompatibility problems. Overall, our cascade enzymatic method provides a clean, benign and cost-effective process for bioconjugation using PAL-mediated ligation.

References

1. Futaki, S., et al., *Embodying a stable alpha-helical protein structure through efficient chemical ligation via thioether formation.* Bioorg Med Chem, 1997. **5**(9): p. 1883-91.
2. Thapa, P., et al., *Native chemical ligation: a boon to peptide chemistry.* Molecules, 2014. **19**(9): p. 14461-83.
3. Agouridas, V., et al., *Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations.* Chem Rev, 2019. **119**(12): p. 7328-7443.
4. Zhang, L. and J.P. Tam, *Orthogonal coupling of unprotected peptide segments through histidyl amino terminus.* Tetrahedron Letters, 1997. **38**(1): p. 3-6.
5. Tam, J.P. and Q. Yu, *Methionine ligation strategy in the biomimetic synthesis of parathyroid hormones.* Biopolymers, 1998. **46**(5): p. 319-27.
6. Yan, L.Z. and P.E. Dawson, *Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization.* Journal of the American Chemical Society, 2001. **123**(4): p. 526-533.
7. Crich, D. and A. Banerjee, *Native Chemical Ligation at Phenylalanine.* Journal of the American Chemical Society, 2007. **129**(33): p. 10064-10065.
8. Yang, R., et al., *Dual Native Chemical Ligation at Lysine.* Journal of the American Chemical Society, 2009. **131**(38): p. 13592-13593.
9. Wells, J.A., et al., *Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtilisin in Bacillus subtilis.* Nucleic Acids Res, 1983. **11**(22): p. 7911-25.
10. Nakatsuka, T., T. Sasaki, and E.T. Kaiser, *Peptide segment synthesis catalyzed by the semisynthetic enzyme thiolsubtilisin.* Journal of the American Chemical Society, 1987. **109**(12): p. 3808-3810.
11. Chang, T.K., et al., *Subtiligase: a tool for semisynthesis of proteins.* Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12544-8.
12. Atwell, S. and A. Wells James, *Selection for improved subtiligases by phage display.* Proceedings of the National Academy of Sciences, 1999. **96**(17): p. 9497-9502.
13. Groen, H., M. Meldal, and K. Breddam, *Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching.* Biochemistry, 1992. **31**(26): p. 6011-6018.
14. Yoshihara, H.A., S. Mahrus, and J.A. Wells, *Tags for labeling protein N-termini with subtiligase for proteomics.* Bioorg Med Chem Lett, 2008. **18**(22): p. 6000-3.

15. Weeks, A.M. and J.A. Wells, *Subtiligase-Catalyzed Peptide Ligation.* Chemical Reviews, 2020. **120**(6): p. 3127-3160.
16. Weeks, A.M. and J.A. Wells, *N-Terminal Modification of Proteins with Subtiligase Specificity Variants.* Current Protocols in Chemical Biology, 2020. **12**(1): p. e79.
17. Henager, S.H., et al., *Enzyme-catalyzed expressed protein ligation.* Nat Methods, 2016. **13**(11): p. 925-927.
18. Riechmann, L. and V. Kasche, *Peptide synthesis catalyzed by the serine proteinases chymotrypsin and trypsin.* Biochim Biophys Acta, 1985. **830**(2): p. 164-72.
19. Liebscher, S., et al., *N-terminal protein modification by substrate-activated reverse proteolysis.* Angew Chem Int Ed Engl, 2014. **53**(11): p. 3024-8.
20. Meyer, C., S. Liebscher, and F. Bordusa, *Selective Coupling of Click Anchors to Proteins via Trypsiligase.* Bioconjugtae Chem, 2016. **27**(1): p. 47-53.
21. Liebscher, S., et al., *Trypsiligase-Catalyzed Labeling of Proteins on Living Cells.* ChemBioChem, 2021. **22**(7): p. 1201-1204.
22. Marraffini, L.A., A.C. Dedent, and O. Schneewind, *Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria.* Microbiol Mol Biol Rev, 2006. **70**(1): p. 192-221.
23. Mazmanian, S.K., et al., *Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall.* Science, 1999. **285**(5428): p. 760-3.
24. Paterson, B.M., et al., *Enzyme-mediated site-specific bioconjugation of metal complexes to proteins: sortase-mediated coupling of copper-64 to a single-chain antibody.* Angew Chem Int Ed Engl, 2014. **53**(24): p. 6115-9.
25. Beerli, R.R., et al., *Sortase Enzyme-Mediated Generation of Site-Specifically Conjugated Antibody Drug Conjugates with High In Vitro and In Vivo Potency.* PLOS ONE, 2015. **10**(7): p. e0131177.
26. Massa, S., et al., *Sortase A-mediated site-specific labeling of camelid single-domain antibody-fragments: a versatile strategy for multiple molecular imaging modalities.* Contrast Media Mol Imaging, 2016. **11**(5): p. 328-339.
27. Nguyen, G.K., et al., *Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis.* Nat Chem Biol, 2014. **10**(9): p. 732-8.
28. Nguyen, G.K., et al., *Butelase 1: A Versatile Ligase for Peptide and Protein Macrocyclization.* J Am Chem Soc, 2015. **137**(49): p. 15398-401.
29. Hatsugai, N., et al., *Vacuolar processing enzyme in plant programmed cell death.* Frontiers in Plant Science, 2015. **6**.
30. Hara-Nishimura, I. and M. Nishimura, *Proglobulin processing enzyme in vacuoles isolated from developing pumpkin cotyledons.* Plant Physiol, 1987. **85**(2): p. 440-5.

31. Abe, Y., et al., *Asparaginyl endopeptidase of jack bean seeds. Purification, characterization, and high utility in protein sequence analysis.* J Biol Chem, 1993. **268**(5): p. 3525-9.
32. Harley, S.M. and J. Michael Lord, *In vitro endoproteolytic cleavage of castor bean lectin presursors.* Plant Science, 1985. **41**(2): p. 111-116.
33. Hara-Nishimura, I. and M. Nishimura, *Proglobulin Processing Enzyme in Vacuoles Isolated from Developing Pumpkin Cotyledons 1.* Plant Physiology, 1987. **85**(2): p. 440-445.
34. Min, W. and D.H. Jones, *In vitro splicing of concanavalin A is catalyzed by asparaginyl endopeptidase.* Nat Struct Biol, 1994. **1**(8): p. 502-4.
35. Harris, K.S., et al., *A suite of kinetically superior AEP ligases can cyclise an intrinsically disordered protein.* Scientific Reports, 2019. **9**(1): p. 10820.
36. Hemu, X., et al., *Structural determinants for peptide-bond formation by asparaginyl ligases.* Proc Natl Acad Sci U S A, 2019. **116**(24): p. 11737-11746.
37. Hemu, X., et al., *Turning an Asparaginyl Endopeptidase into a Peptide Ligase.* ACS Catalysis, 2020. **10**(15): p. 8825-8834.
38. Harris, K.S., et al., *Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase.* Nature Communications, 2015. **6**(1): p. 10199.
39. Bernath-Levin, K., et al., *Peptide macrocyclization by a bifunctional endoprotease.* Chem Biol, 2015. **22**(5): p. 571-82.
40. Yang, R., et al., *Engineering a Catalytically Efficient Recombinant Protein Ligase.* Journal of the American Chemical Society, 2017. **139**(15): p. 5351-5358.
41. Hemu, X., et al., *Characterization and application of natural and recombinant butelase-1 to improve industrial enzymes by end-to-end circularization.* RSC Adv, 2021. **11**(37): p. 23105-23112.
42. Zhao, J., et al., *Study on activation mechanism and cleavage sites of recombinant butelase-1 zymogen derived from Clitoria ternatea.* Biochimie, 2022. **199**: p. 12-22.
43. James, A.M., et al., *The macrocyclizing protease butelase 1 remains autocatalytic and reveals the structural basis for ligase activity.* Plant J, 2019. **98**(6): p. 988-999.
44. Zauner, F.B., et al., *Structural analyses of Arabidopsis thaliana legumain γ reveal differential recognition and processing of proteolysis and ligation substrates.* Journal of Biological Chemistry, 2018. **293**(23): p. 8934-8946.
45. Jackson, M.A., et al., *Molecular basis for the production of cyclic peptides by plant asparaginyl endopeptidases.* Nat Commun, 2018. **9**(1): p. 2411.
46. Hemu, X., et al., *Butelase 1-Mediated Ligation of Peptides and Proteins.* Methods Mol Biol, 2019. **2012**: p. 83-109.

47. Cao, Y., et al., *Butelase-Mediated Ligation as an Efficient Bioconjugation Method for the Synthesis of Peptide Dendrimers.* Bioconjugate Chem, 2016. **27**(11): p. 2592-2596.
48. Yap, K., et al., *An environmentally sustainable biomimetic production of cyclic disulfide-rich peptides.* Green Chemistry, 2020. **22**(15): p. 5002-5016.
49. Hemu, X., X. Zhang, and J.P. Tam, *Ligase-Controlled Cyclo-oligomerization of Peptides.* Org Lett, 2019. **21**(7): p. 2029-2032.
50. Hemu, X., et al., *Correction to Immobilized Peptide Asparaginyl Ligases Enhance Stability and Facilitate Macrocyclization and Site-Specific Ligation.* J Org Chem, 2020. **85**(6): p. 4581.
51. Zhang, D., et al., *Vypal2: A Versatile Peptide Ligase for Precision Tailoring of Proteins.* Int J Mol Sci, 2021. **23**(1).
52. Harmand, T.J., et al., *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 Chem, 2018. **29**(10): p. 3245-3249.
53. Rehm, F.B.H., et al., *Enzymatic C-Terminal Protein Engineering with Amines.* J Am Chem Soc, 2021. **143**(46): p. 19498-19504.
54. Wang, Z., et al., *Engineering protein theranostics using bio-orthogonal asparaginyl peptide ligases.* Theranostics, 2021. **11**(12): p. 5863-5875.
55. Zhang, D., et al., *pH-Controlled Protein Orthogonal Ligation Using Asparaginyl Peptide Ligases.* Journal of the American Chemical Society, 2021. **143**(23): p. 8704-8712.
56. Deng, Y., et al., *Enzymatic biosynthesis and immobilization of polyprotein verified at the single-molecule level.* Nature Communications, 2019. **10**.
57. Bi, X., et al., *Enzymatic Engineering of Live Bacterial Cell Surfaces Using Butelase 1.* Angew Chem Int Ed Engl, 2017. **56**(27): p. 7822-7825.
58. Bi, X., et al., *Tagging Transferrin Receptor with a Disulfide FRET Probe To Gauge the Redox State in Endosomal Compartments.* Analytical Chemistry, 2020. **92**(18): p. 12460-12466.
59. Harmand, T.J., et al., *Asparaginyl Ligase-Catalyzed One-Step Cell Surface Modification of Red Blood Cells.* ACS Chem Biol, 2021. **16**(7): p. 1201-1207.
60. Vinogradov, A.A., Y. Yin, and H. Suga, *Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges.* Journal of the American Chemical Society, 2019. **141**(10): p. 4167-4181.
61. Baeriswyl, V. and C. Heinis, *Polycyclic Peptide Therapeutics.* ChemMedChem, 2013. **8**(3): p. 377-384.
62. Arkin, M.R., Y. Tang, and J.A. Wells, *Small-molecule inhibitors of protein-protein interactions: progressing toward the reality.* Chem Biol, 2014. **21**(9): p. 1102-14.
63. Fosgerau, K. and T. Hoffmann, *Peptide therapeutics: current status and future directions.* Drug Discov Today, 2015. **20**(1): p. 122-8.

64. Lau, J.L. and M.K. Dunn, *Therapeutic peptides: Historical perspectives, current development trends, and future directions.* Bioorg Med Chem, 2018. **26**(10): p. 2700-2707.
65. Saska, I., et al., *An asparaginyl endopeptidase mediates in vivo protein backbone cyclization.* J Biol Chem, 2007. **282**(40): p. 29721-8.
66. Nguyen, G.K., et al., *Butelase-mediated cyclization and ligation of peptides and proteins.* Nat Protoc, 2016. **11**(10): p. 1977-1988.
67. Yamada, K., et al., *Vacuolar processing enzymes in the plant life cycle.* New Phytol, 2020. **226**(1): p. 21-31.
68. Dall, E. and H. Brandstetter, *Structure and function of legumain in health and disease.* Biochimie, 2016. **122**: p. 126-50.
69. Müntz, K. and A.D. Shutov, *Legumains and their functions in plants.* Trends Plant Sci, 2002. **7**(8): p. 340-4.
70. Min, W. and D.H. Jones, *In vitro splicing of concanavalin A is catalyzed by asparaginyl endopeptidase.* Nature Structural Biology, 1994. **1**(8): p. 502-504.
71. Zauner, F.B., et al., *Crystal Structure of Plant Legumain Reveals a Unique Two-Chain State with pH-Dependent Activity Regulation.* Plant Cell, 2018. **30**(3): p. 686-699.
72. Haywood, J., et al., *Structural basis of ribosomal peptide macrocyclization in plants.* eLife, 2018. **7**: p. e32955.
73. Tam, J.P., et al., *Peptide asparaginyl ligases—renegade peptide bond makers.* Science China Chemistry, 2020. **63**(3): p. 296-307.
74. Dall, E. and H. Brandstetter, *Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation.* Proc Natl Acad Sci U S A, 2013. **110**(27): p. 10940-5.
75. Hemu, X., et al., *Total Synthesis of Circular Bacteriocins by Butelase 1.* J Am Chem Soc, 2016. **138**(22): p. 6968-71.
76. Rehm, F.B.H., et al., *Site-Specific Sequential Protein Labeling Catalyzed by a Single Recombinant Ligase.* Journal of the American Chemical Society, 2019. **141**(43): p. 17388-17393.
77. Cao, Y., et al., *Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation.* Chem Commun (Camb), 2015. **51**(97): p. 17289-92.
78. Muri, E.M., et al., *Hydroxamic acids as pharmacological agents.* Curr Med Chem, 2002. **9**(17): p. 1631-53.
79. Agrawal, Y.K., *Hydroxamic Acids and Their Metal Complexes.* Russian Chemical Reviews, 1979. **48**(10): p. 948-963.
80. Dunkelmann, D.L., et al., *Amide-forming chemical ligation via O-acyl hydroxamic acids.* Proc Natl Acad Sci U S A, 2018. **115**(15): p. 3752-3757.
81. Maola, K., et al., *Engineered Peptide Macrocycles Can Inhibit Matrix Metalloproteinases with High Selectivity.* Angew Chem Int Ed Engl, 2019. **58**(34): p. 11801-11805.

82. Brooks, P.C., et al., *Localization of matrix metalloproteinase MMP-2 to the surface of invasive cells by interaction with integrin alpha v beta 3.* Cell, 1996. **85**(5): p. 683-93.
83. Higashi, S. and K. Miyazaki, *Identification of amino acid residues of the matrix metalloproteinase-2 essential for its selective inhibition by beta-amyloid precursor protein-derived inhibitor.* J Biol Chem, 2008. **283**(15): p. 10068-78.
84. Hashimoto, H., et al., *Structural basis for matrix metalloproteinase-2 (MMP-2)-selective inhibitory action of β-amyloid precursor protein-derived inhibitor.* J Biol Chem, 2011. **286**(38): p. 33236-43.
85. Higashi, S., et al., *Molecular design of a highly selective and strong protein inhibitor against matrix metalloproteinase-2 (MMP-2).* J Biol Chem, 2013. **288**(13): p. 9066-76.
86. Trabi, M., H.J. Schirra, and D.J. Craik, *Three-dimensional structure of RTD-1, a cyclic antimicrobial defensin from Rhesus macaque leukocytes.* Biochemistry, 2001. **40**(14): p. 4211-21.
87. Emery, T.F. and J.B. Neilands, *Further Observations Concerning the Periodic Acid Oxidation of Hydroxylamine Derivatives1.* The Journal of Organic Chemistry, 1962. **27**(3): p. 1075-1077.
88. Wishart, D.S., et al., *1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects.* Journal of Biomolecular NMR, 1995. **5**(1): p. 67-81.
89. Hernandez, J.-F., et al., *Squash Trypsin Inhibitors from Momordica cochinchinensis Exhibit an Atypical Macrocyclic Structure.* Biochemistry, 2000. **39**(19): p. 5722-5730.
90. Daly, N.L., et al., *Structural Insights into the Role of the Cyclic Backbone in a Squash Trypsin Inhibitor *.* Journal of Biological Chemistry, 2013. **288**(50): p. 36141-36148.
91. Thongyoo, P., E.W. Tate, and R.J. Leatherbarrow, *Total synthesis of the macrocyclic cysteine knot microprotein MCoTI-II.* Chemical Communications, 2006(27): p. 2848-2850.
92. Thongyoo, P., et al., *Chemical and biomimetic total syntheses of natural and engineered MCoTI cyclotides.* Organic & Biomolecular Chemistry, 2008. **6**(8): p. 1462-1470.
93. Camarero, J.A., et al., *Biosynthesis of a Fully Functional Cyclotide inside Living Bacterial Cells.* ChemBioChem, 2007. **8**(12): p. 1363-1366.
94. Stanger, K., et al., *Backbone cyclization of a recombinant cystine-knot peptide by engineered Sortase A.* FEBS Letters, 2014. **588**(23): p. 4487-4496.
95. Du, J., et al., *A bifunctional asparaginyl endopeptidase efficiently catalyzes both cleavage and cyclization of cyclic trypsin inhibitors.* Nature Communications, 2020. **11**(1): p. 1575.
96. Hider, R.C., et al., *Origin of the positive 225–230 nm circular dichroism band in proteins: Its application to conformational analysis.* Biophysical Chemistry, 1988. **31**(1): p. 45-51.

97. D'Souza, C., et al., *Using the MCoTI-II Cyclotide Scaffold To Design a Stable Cyclic Peptide Antagonist of SET, a Protein Overexpressed in Human Cancer.* Biochemistry, 2016. **55**(2): p. 396-405.
98. Gran, L., F. Sandberg, and K. Sletten, *Oldenlandia affinis (R&S) DC: A plant containing uteroactive peptides used in African traditional medicine.* Journal of Ethnopharmacology, 2000. **70**(3): p. 197-203.
99. Luckett, S., et al., *High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds1 1Edited by I. A. Wilson.* Journal of Molecular Biology, 1999. **290**(2): p. 525-533.
100. Conibear, A.C., et al., *The Cyclic Cystine Ladder of Theta-Defensins as a Stable, Bifunctional Scaffold: A Proof-of-Concept Study Using the Integrin-Binding RGD Motif.* ChemBioChem, 2014. **15**(3): p. 451-459.
101. Pfaff, M., et al., *Selective recognition of cyclic RGD peptides of NMR defined conformation by alpha IIb beta 3, alpha V beta 3, and alpha 5 beta 1 integrins.* Journal of Biological Chemistry, 1994. **269**(32): p. 20233-20238.
102. Avrutina, O., et al., *Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTI-II.* 2005. **386**(12): p. 1301-1306.
103. Zheng, J.-S., et al., *Synthesis of Cyclic Peptides and Cyclic Proteins via Ligation of Peptide Hydrazides.* ChemBioChem, 2012. **13**(4): p. 542-546.
104. Lee, W., M. Tonelli, and J.L. Markley, *NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy.* Bioinformatics, 2014. **31**(8): p. 1325-1327.
105. Berjanskii, M.V., S. Neal, and D.S. Wishart, *PREDITOR: a web server for predicting protein torsion angle restraints.* Nucleic Acids Research, 2006. **34**(suppl_2): p. W63-W69.
106. Güntert, P., C. Mumenthaler, and K. Wüthrich, *Torsion angle dynamics for NMR structure calculation with the new program Dyana11Edited by P. E. Wright.* Journal of Molecular Biology, 1997. **273**(1): p. 283-298.
107. Spicer, C.D. and B.G. Davis, *Selective chemical protein modification.* Nature Communications, 2014. **5**(1): p. 4740.
108. Hoyt, E.A., et al., *Contemporary approaches to site-selective protein modification.* Nature Reviews Chemistry, 2019. **3**(3): p. 147-171.
109. Zhang, G., et al., *Illuminating biological processes through site-specific protein labeling.* Chemical Society Reviews, 2015. **44**(11): p. 3405-3417.
110. Sletten, E.M. and C.R. Bertozzi, *Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality.* Angewandte Chemie International Edition, 2009. **48**(38): p. 6974-6998.
111. Rosen, C.B. and M.B. Francis, *Targeting the N terminus for site-selective protein modification.* Nature Chemical Biology, 2017. **13**(7): p. 697-705.
112. Zhang, Y., et al., *Recent progress in enzymatic protein labelling techniques and their applications.* Chemical Society Reviews, 2018. **47**(24): p. 9106-9136.
113. Lotze, J., et al., *Peptide-tags for site-specific protein labelling in vitro and in vivo.* Mol Biosyst, 2016. **12**(6): p. 1731-45.

114. Pishesha, N., J.R. Ingram, and H.L. Ploegh, *Sortase A: A Model for Transpeptidation and Its Biological Applications.* Annu Rev Cell Dev Biol, 2018. **34**: p. 163-188.
115. Rehm, F.B.H., et al., *Asparaginyl Ligases: New Enzymes for the Protein Engineer's Toolbox.* ChemBioChem, 2021. **22**(12): p. 2079-2086.
116. Bi, X., et al., *Immobilization and Intracellular Delivery of Circular Proteins by Modifying a Genetically Incorporated Unnatural Amino Acid.* Bioconjugate Chemistry, 2018. **29**(7): p. 2170-2175.
117. Xia, Y., et al., *Nγ-Hydroxyasparagine: A Multifunctional Unnatural Amino Acid That is a Good P1 Substrate of Asparaginyl Peptide Ligases.* Angewandte Chemie International Edition, 2021. **60**(41): p. 22207-22211.
118. Nguyen, G.K., et al., *Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide.* Angew Chem Int Ed Engl, 2015. **54**(52): p. 15694-8.
119. Tang, T.M.S., et al., *Use of an asparaginyl endopeptidase for chemo-enzymatic peptide and protein labeling.* Chemical Science, 2020. **11**(23): p. 5881-5888.
120. Rehm, F.B.H., et al., *Improved Asparaginyl-Ligase-Catalyzed Transpeptidation via Selective Nucleophile Quenching.* Angew Chem Int Ed Engl, 2021. **60**(8): p. 4004-4008.
121. Busby, W.H., et al., *An enzyme(s) that converts glutaminyl-peptides into pyroglutamyl-peptides. Presence in pituitary, brain, adrenal medulla, and lymphocytes.* Journal of Biological Chemistry, 1987. **262**(18): p. 8532-8536.
122. Schilling, S., et al., *Heterologous Expression and Characterization of Human Glutaminyl Cyclase: Evidence for a Disulfide Bond with Importance for Catalytic Activity.* Biochemistry, 2002. **41**(35): p. 10849-10857.
123. Seifert, F., et al., *Glutaminyl Cyclases Display Significant Catalytic Proficiency for Glutamyl Substrates.* Biochemistry, 2009. **48**(50): p. 11831-11833.
124. Steiner, D., P. Forrer, and A. Plückthun, *Efficient Selection of DARPins with Sub-nanomolar Affinities using SRP Phage Display.* Journal of Molecular Biology, 2008. **382**(5): p. 1211-1227.
125. Jost, C., et al., *Structural Basis for Eliciting a Cytotoxic Effect in HER2-Overexpressing Cancer Cells via Binding to the Extracellular Domain of HER2.* Structure, 2013. **21**(11): p. 1979-1991.
126. Ståhl, S., et al., *Affibody Molecules in Biotechnological and Medical Applications.* Trends in Biotechnology, 2017. **35**(8): p. 691-712.

Appendix: Publication list

(1) Chapter 2 is based on the article below:

Xia, Y.; To, J.; Chan, N.-Y.; Hu, S.; Liew, H. T.; Balamkundu, S.; Zhang, X.; Lescar, L.; Bhattacharjya, S.; Tam, J. P.; Liu, C.-F. Nγ-Hydroxyasparagine: A Multifunctional Unnatural Amino Acid That is a Good P1 Substrate of Asparaginyl Peptide Ligases. *Angew. Chem. Int. Ed*. **2021**, 60, 22207–22211.

(2) Chapter 3 is based on the manuscript below:

Xia, Y.; Li, F.; Zhang. X.; Balamkundu. S.;Tang, Fan.; Hu, S.; Lescar, J.; Tam, J. P.; Liu, C.-F. A Cascade Enzymatic Reaction Scheme for Irreversible Transpeptidative Protein Ligation. *J. Am. Chem. Soc.* **2023**, 145, 6838-6844.