Peptide asparaginyl ligases—renegade peptide bond makers

Tam, James P.; Chan, Ning-Yu; Liew, Heng Tai; Tan, Shaun J.; Chen, Yu

2020

Tam, J. P., Chan, N., Liew, H. T., Tan, S. J. & Chen, Y. (2020). Peptide asparaginyl ligases—renegade peptide bond makers. *Science China Chemistry*, 63(3), 296-307.
<https://dx.doi.org/10.1007/s11426-019-9648-3>

<https://hdl.handle.net/10356/148282>

<https://doi.org/10.1007/s11426-019-9648-3>

© 2020 Science in China Press and Springer-Verlag GmbH Germany, part of Springer Nature. All rights reserved. This paper was published in *Science China Chemistry* and is made available with permission of Science in China Press. The original publication is available at www.scichina.com and www.springerlink.com.

Downloaded on 26 Apr 2023 19:59:53 SGT

Peptide Asparaginyl Ligases — Renegade Peptide Bond Makers

[Tam James](#), [Chan Ning Yu](#), [Liew Heng Tai](#), [Tan Shaun](#) and [Chen Yu](#)

Citation: [SCIENCE CHINA Chemistry](#); doi: 10.1007/s11426-019-9648-3

View online: <http://engine.scichina.com/doi/10.1007/s11426-019-9648-3>

Published by the [Science China Press](#)

Articles you may be interested in

[Influence of peptide bond on photosensitized oxidation of tryptophan, tyrosine and histidine dipeptides](#)

Chinese Science Bulletin **42**, 1624 (1997);

[NEW EVIDENCE FOR THE INHIBITION BY HARRINGTONINE IN THE FORMATION OF THE FIRST PEPTIDE BOND IN PROTEIN SYNTHESIS](#)

Science in China Series B-Chemistry, Biological, Agricultural, Medical & Earth Sciences **28**, 512 (1985);

[Peptide segment ligation: A new method for synthesis of peptide and protein](#)

Chinese Science Bulletin **48**, 1 (2003);

[USE OF HOEC ACTIVE ESTERS IN SOLIDPHASE PEPTIDE SYNTHESIS——THE SYNTHESIS OF GROWTH HORMONE RELEASING PEPTIDE](#)

Chinese Science Bulletin **34**, 43 (1989);

[Peptide primary messengers in plants](#)

Chinese Science Bulletin **46**, 622 (2001);

Peptide Asparaginyl Ligases — Renegade Peptide Bond Makers

Journal:	SCIENCE CHINA Chemistry
Manuscript ID	SCC-2019-0706.R1
Manuscript Type:	Review
Date Submitted by the Author:	02-Nov-2019
Complete List of Authors:	Tam, James; Nanyang Technological University, School of Biological Sciences Chan, Ning Yu; Nanyang Technological University, School of Biological Sciences Liew, Heng Tai; Nanyang Technological University, School of Biological Sciences Tan, Shaun; Nanyang Technological University, School of Biological Sciences Chen, Yu; Nanyang Technological University, School of Biological Sciences
Keywords:	Asparaginyl endopeptidase, Asn-specific ligation, bioorthogonal ligation, Butelase, chemoenzymatic ligation, live-cell labeling, protein engineering, protein modification, site-specific labeling, tandem ligation
Speciality:	Chemical Biology

SCHOLARONE™
Manuscripts

Peptide Asparaginyl Ligases — Renegade Peptide Bond Makers

James P. Tam, Ning-Yu Chan, Heng Tai Liew, Shaun J. Tan, Yu Chen

Abstract

Making peptide bonds is tightly controlled by genetic code and machinery which includes cofactors, ATP, and RNAs. In this regard, the stand-alone and genetic-code-independent peptide ligases constitute a new family of renegade peptide-bond makers. A prime example is butelase-1, an Asn/Asp(Asx)-specific ligase that structurally belongs to the asparaginyl endopeptidase family. Butelase-1 specifically recognizes a C-terminal Asx-containing tripeptide motif, Asn/Asp-Xaa-Yaa (Xaa and Yaa are any amino acids), to form a site-specific Asn-Xaa peptide bond either intramolecularly as cyclic proteins or intermolecularly as modified proteins. Our work in the past five years has validated that butelase-1 is a potent and versatile ligase. Here we review the advances in ligases, with a focus on butelase-1, and their applications in engineering bioactive peptides and precision protein modifications, antibody-drug conjugates, and live-cell labeling.

Keywords Asparaginyl endopeptidase, Asn-specific ligation, bioorthogonal ligation, Butelase, chemoenzymatic ligation, live-cell labeling, protein engineering, protein modification, site-specific labeling, tandem ligation

1 Introduction

Ligases are enzymes that form peptide bonds of peptides and proteins. Compared to the ubiquitous and well-characterized proteases which break peptide bonds, ligases are rare and poorly characterized. Advances in recombinant DNA methodology [1-4] and peptide chemistry [5-8] have led to access to proteins and antibodies for research and development. Chemists and biologists have also started sharing common goals in combining recombinant protein expression and chemoselective ligation to precisely produce proteins for mechanistic studies [9, 10]. These efforts have also led to a rapid expansion of biological drugs [11]. In turn, they have prompted the need for site-specific and aqueous-compatible ligation chemistry for precision biomanufacturing of bioactive peptides, protein drugs, and antibody-drug conjugates [12-14]. Peptide bond-forming ligases stand out as highly suitable for such purposes.

Peptide bond-forming ligases function as aqueous-compatible superglues for a wide range of applications (Scheme 1). As biochemical and biotechnological tools, they enable peptide and protein engineering, macrocyclization [15-18], site-specific protein modification [19, 20] and live-cells labeling [21-24]. Furthermore, they can be used

under physiological conditions, which are necessary for modifying antibodies and live cells. Importantly, they are also pivotal catalysts that make life possible.

Ligases are generally found in large complexes because making proteins is a tightly controlled and gene-encoded translational process, a process localized in ribosomes. The complexity and fidelity of a ribosomal protein synthetic machinery, involving mRNA, tRNA, ATP, proteins, and enzymes, have been revealed in great detail using cryo-electron microscopy [25]. Microbes also make use of a non-ribosomal protein synthesis process known as thiol-template [26, 27]. This process does not involve mRNA and generally produces peptide antibiotics, for example, gramicidin and tyrocidine, but it is still an ATP-dependent process requiring multiple enzymes.

In the past decade, stand-alone ligases which do not require a multi-enzyme complex have been discovered (Table 1) [28-29]. Some of these ligases have been identified as the enzymatic bioprocessors from the biosynthesis of the ribosomally synthesized and post-translationally modified peptides (RiPPs) [30]. RiPPs represent a superfamily of highly-modified and structurally-diverse peptidyl natural products, once thought to be derived from the non-ribosomal pathways, such as biosynthesis of patellamides [31, 32]. Examples of ligases that act on RiPPs can be found in plants [33-40], bacteria [41, 42], and fungi [43]. These ligases are capable of ATP-independent cyclization of peptides and proteins in the presence of a proper recognition signal. Because they are stand-alone and ATP-independent peptide ligases, unhampered by genetic code or a cofactor, and importantly, do not follow the rules of genetic codes, they are renegade peptide-bond makers.

In this review, we describe the occurrences, characterization, and application of a specific family of renegade peptide-bond makers, the peptide asparaginyl ligase (PALs). In particular, we focus on their prototype, butelase-1, which is the most efficient known ligase so far.

2 Naturally-occurring ligases

2.1 Peptide Asparaginyl Ligases (PALs), the AEP-type ligases

Asparaginyl endopeptidases (AEPs), also known as legumains or vacuolar processing enzymes, are endopeptidases that cleave after an Asx residue [33]. Plant AEPs perform crucial roles in a wide range of biological activities, including programmed cell death and the post-translational modification of RiPPs in the biosynthesis of cyclic peptides for plant defense against bacteria or agricultural pests [44, 45]. In humans, overexpression of legumains has been reported to facilitate cancer development, such as promoting the metastasis and tissue invasion of various cancer types [46].

The thiol protease activity of AEPs was first observed in the 1980s [47, 48], and about the same period, ligase activity of AEPs was reported in jack bean [49, 50]. However, it was not until 2014 when our laboratory isolated the first PAL, butelase-1, which acts as a post-translational processing enzyme in the biosynthesis of cyclic peptides [35].

Butelase-1 is capable of macrocyclization without observable protease activity at near-neutral or neutral pH. Consequently, its high ligation efficiency significantly shortens the reaction time to complete reactions in minutes with a minimal amount of enzyme.

Butelase-1 has been and can be extracted efficiently from pods of *Clitoria ternatea* (5 mg/kg), the butterfly pea which is also known by its Malay name 'Bunga telang' from which the name butelase is derived [35]. Recently, recombinant expression of zymogenic butelase-1 in *Escherichia coli* and yeast *Pichia pastoris* has been achieved [51, 52], and crystal structure of zymogen of butelase-1 was solved at 3.1 Å (Figure 1), providing rooms for engineering and understanding of the mechanism of an AEP-like ligase. It is worthwhile to note that the naturally-occurring butelase-1 is heavily glycosylated and more stable than the recombinantly-expressed versions.

A distinctive advantage of butelase-1 over other families of ligases is that it can be used for the synthesis of natural products. Butelase-1 catalyzes macrocyclization of peptides and proteins of various sizes tracelessly, with extraordinarily high efficiency (Scheme 2). The tripeptide recognition signal of butelase-1 requires an Asx with a dipeptide as a leaving group. In later sections, we will use the synthesis of bacteriocins as examples.

A tripeptide recognition Asx-Xaa-Yaa appears to be a general rule for the PALs. However, different PALs have preferences for different Xaa and Yaa amino acids. For butelase-1, the preferred tripeptide recognition motif is Asn-His-Val to enable the ligation to proceed with high site-specificity [35, 53, 54] (Table 2). Also, butelase-1 accepts most N-terminal amino acids for intermolecular peptide and protein ligations. Its capacity to work under aqueous conditions eliminates the usage of organic solvent, making butelase-1 an excellent tool for green chemistry [55]. The biosynthetic precursor or recombinant form of butelase-1, similar to all other precursors of AEPs and PALs, are expressed as zymogens with a large-cap domain covering their active site [56]. The cap domain is removed by autoactivation under acidic conditions [57].

Since the discovery of butelase-1, additional PALs such as VyPAL1-3, HeAEP 3, OaAEP 1b, and 3-5, have been identified [36, 38-40]. Among them, VyPAL2 has a fast kinetic ($K_{cat}/K_m = 274,325$) and can be easily produced by insect cells (10-20 mg/L). While PALs are rare, AEPs are common. Thus, there is a strong incentive to engineer AEPs to ligases. By modifying the S2 and S1' pockets, Hemu *et al.* were able to convert the protease AEP, VcAEP, to PAL [39]. However, an unequivocal and general mechanism-based approach to engineer ligases from proteases remains to be fully developed.

2.2 Serine protease-like ligases

Serine protease-like ligases differ from PALs by having serine instead of cysteine at the catalytic site. Similarly to butelase-1, they were discovered because they are also bioprocessing enzymes of RiPPs responsible for producing many naturally-occurring cyclic peptides with various structures and biological functions. Selected examples of RiPPs include α -amanitin, patellamides, and orbitides [32, 34, 43]. Table 1 lists recent

examples of serine protease-like ligases. They include PatG, a natural cyclase found in cyanobacteria *Lissoclinum patella* [32], it catalyzes macrocyclization of patellamides with a C-terminal AYDG motif [42]. PCY1, a plant ligase first discovered in the Caryophyllaceae family [34], is involved in the biosynthesis of orbitides which are cyclic peptides with 5-12 hydrophobic amino acids. A characteristic of this class of ligases is that they have demanding substrate requirements. For example, the recognition motif of POPB is an eight-residue peptide. A recent report described the use of PCY1 to cyclize peptides with a C-terminal three-residue tail, FQA or IQT [58]. In general, the serine protease-like ligases display slow kinetics and stringent substrate requirement, making it a challenge to be exploited for routine ligation reactions at this time. However, they are also ripe for improvement to complement the AEP-like ligases for bioorthogonal ligation reactions (see section 3.3.2).

2.3 Transpeptidation by sortase A

Transpeptidases catalyze the joining of two peptides, a good example is sortase A, which is adapted for ligation reaction [41, 59]. In Gram-positive bacteria, sortase-mediated transpeptidation is essential for bacteria virulence and colonization. Sortase A recognizes a universal and conserved C-terminal sorting signal LPXTG (X = 20 natural amino acids) for surface protein anchoring. To expand the stringent substrate requirement, Liu DR and coworkers applied direct evolution approach on yeast display libraries to improve substrate specificity of sortase A [60]. Despite its poor kinetics and long recognition motif, sortase A and its mutants have been exploited in various areas ranging from cell biology to structural biology [61].

Recently, Howarth and coworkers developed SpyLigase by breaking the CnaB2 domain of *Streptococcus pyogenes* fibronectin adhesion protein into three parts. Spontaneous isopeptide bond formation occurs when the split parts are recombined [62, 63]. However, this method leaves more than 20 amino acids on the conjugated proteins, and the enzyme cannot be reused, limiting their versatility for ligation reactions.

2.42.5 Modified subtilisin

In the late 1960s, Koshland and then Bender showed that chemical transformation of the catalytic Ser221 to Cys221 of the bacterial protease subtilisin to give thiosubtilisin severely damaged its hydrolase activity [64, 65]. Exploiting the 'damaged' protease, Kaiser and his coworkers applied the thiosubtilisin for the synthesis of peptides with 5-12 amino acids [66, 67]. Later, Wells and his coworkers used recombinant methods and site-directed mutagenesis to further improve the amyolytic activity of the S221C mutant by releasing the steric crowding through mutating the Pro255 to Ala255 to synthesis a protein [68]. The stability of the subtiligase against heat, alkali, and organic solvents was further enhanced by incorporating five other mutations [69], which allowed the enzyme to catalyze reactions in denaturants. An interesting application by Liu CF and his coworker was to exploit the significant increase in esterase activity and reduced hydrolase activity of modified subtilisin to prepare peptide thioesters and

thioacids, which served as the substrates for sequential ligation of unprotected peptides and proteins [70-73]. Recently, Wells and coworkers used the vast proteome-derived peptide libraries to modify N-terminal substrate scope to allow rapid pairing of selected subtiligase mutants for ligation [74]. However, the stringent requirement of an ester or a thioester substrate which is not accessible by recombinant methods for a subtiligase-mediated aminolysis is a limiting factor in certain ligation reactions.

3 Applications

3.1 Head-to-tail cyclization

Cyclization of the N- and C-termini of a peptide or protein confers resistance against degradation by exopeptidases and, at times, structural rigidity to minimize denaturation. Butelase-mediated cyclization of peptides and proteins generally resulted in high yield and completed within minutes. Table 3 shows selected examples of macrocyclization by butelase-1, which ranges from 14 to a few hundred amino acids [35, 54, 55, 75-77].

An advantage of butelase-mediated cyclization is its product is traceless with only the Asx remaining at the ligation site. This advantage was exploited in the total synthesis of the bacteriocins. Bacteriocins represent the largest naturally-occurring cyclic peptides antimicrobials, which include the 70-residue AS-48 and uberlolsin (see Table 3 example 9-10). Both bacteriocins are highly hydrophobic, containing >60% of hydrophobic amino acids in their sequences. Furthermore, these hydrophobic amino acids are found as long stretches, rendering them prone to aggregation which resulted in difficulties for chemical synthesis. Using the butelase-mediated macrocyclization, the cyclization of AS-48 can be accomplished in 1 hour, with more than 85% yield (Figure 2) [76]. Similarly, other AEP-like ligases could also cyclize various proteins, including merozoite surface protein 2 (MSP2), which is recombinantly expressed in the disordered form [40].

All D-antimicrobial peptides are the mirror image of, and often equally active as, their all-L counterparts because their targets are bacterial membranes. However, they have the advantage of being resistant to proteolytic degradation. Different from other ligases, butelase-1 can also cyclize peptide substrates containing entirely D-amino acids except that the S1 site must remain as L-Asx as the key recognition signal [75].

3.2 Site-specific protein modifications

Installing features of interest to a protein devoid of them is appealing. These features could confer desirable functions to the protein for various studies. In particular, exquisite site-specificity conjugation of proteins at their N- or C-termini, or side chains with a cargo of choice (Table 4). Without the requirement of a C-terminal ester and thioester, the butelase-mediated ligation reaction can be performed under mild conditions, utilizing components from recombinantly-expressed proteins or chemically-synthesized peptides or polymers.

3.2.1 N-terminal protein modifications

N-terminal protein modification is perhaps the most straightforward of all site-specific modifications of biopolymers, particularly for proteins expressed by recombinant DNA methods. N-terminal modification by PALs or other ligases is an invaluable tool to understand the sophisticated biological process and post-translational modifications of the target proteins.

Many N-terminal amino acids are bifunctional and can be used to react with other functional groups to form heterocycles, such as Cys, Thr, Ser, His, and Lys. Indeed this area of research has been reviewed frequently. Also these N-terminal bifunctional amino acids have been exploited for chemical ligation of peptide-to-peptide and peptide-to-protein, mainly through entropic chemical ligation, which is proximity-driven for the amide bond formation [6-8, 78, 79].

Figure 3 illustrates a general scheme of N-terminal ligation of protein of interest by butelase-mediated ligation. The N-terminal segment of a peptide or a protein should contain the tripeptide motif NHV, which is then ligated to the second peptide or protein of interest to form a new compound. Table 4 shows selected examples of N-terminal ligation.

Figure 4 shows another general approach for N-terminal ligation using thioesters. The N-terminal segment of a peptide or a protein, in this case, should contain a C-terminal asparaginyl thioester, instead of the tripeptide NHV motif. We used this approach to provide site- and linkage-specific ligation of ubiquitin and green fluorescent protein under physiological conditions [80].

3.2.2 C-terminal protein modification

C-terminal modifications are generally more challenging than N-terminal modifications because it remains difficult to prepare reactive chemical groups at the C-terminus of a peptide or protein by recombinant methods. This is also true for peptides produced by standard solid-phase peptide synthesis.

Currently, butelase-mediated ligation is the method of choice for C-terminal modifications. By expressing the NHV motif at the C-terminus of a target protein, butelase-1 can link the target protein with cargo of choice efficiently (Figure 5). Selected examples include site-specifically ligating designed ankyrin repeat proteins (DARPin) to fluorescein isothiocyanate (FITC) and cytolytic peptide magainin (Figure 6), the DARPin-magainin could be applied to target HER2-positive breast cancer cells specifically. The precisely modified DARPin-FITC was further shown to bind to HER2-positive BT474 cells by live-cell confocal microscopy [81].

Another application of butelase-mediated C-terminal modification is to prepare a protein with a thioester group (Figure 7). This can be achieved by ligating a small peptide thioester to a recombinant protein obtained by recombinant or chemical methods [80, 82, 83].

Peptide or protein thioesters are highly useful building blocks for numerous chemical transformations. These include peptide-segment condensation [84], substrates of subtiligase, and starting materials for C-terminal modifications to other functional

groups [6-8]. For butelase-mediated ligation, a protein or peptide thioester as a substrate increases the ligation efficiency in comparison with the conventional protein or peptide tripeptide recognition motif of NHV or Asn-ester. We have shown that the use of thiopeptides as substrate results in an irreversible ligation reaction with high efficiency. Liu and coworkers proposed that a peptide thioester can promote the formation of acyl-enzyme thioester intermediate and thus accelerate the ligation process. They compared the efficiency of subtiligase-mediated ligation using its model peptide sequence and observed a largely improved $k_{\text{cat}}/k_{\text{m}}$ value by 24-fold using peptide thioester other than ester [73].

3.2.4.5 Side chain modification of proteins

Site-specific side chain modifications, using ligases or chemical methods, are the most challenging among the site-specific modifications for recombinantly expressed proteins. However, advances in genetic code expansion allow the installation of unusual amino acids with desirable features to a protein for site-specific ligation for a diverse range of applications, including protein probing, imaging, and function controlling [85]. Liu CF and coworker incorporated unnatural amino acids in the protein sequence by amber codon suppression technology to enable site-specific modification at the side chain of murine dihydrofolate reductase (mDHFR). Butelase-mediated ligation was then applied to cyclize the biotinylated linear mDHFR within 30 min. With the side chain modification, the biotinylated cyclic mDHFR can be immobilized to reliable support, streptavidin-functionalized agarose beads, with controlled orientation [77].

Intracellular delivery of critical enzymes that are deficient in particular disease is the pivotal step of enzyme-replacement therapy. However, the ability of the drug to penetrate the cell membrane and to remain functional in the cell has been hampering therapeutics targeting specific sites *in vivo*. By conjugating the cyclic biotinylated mDHFR with cell-penetrating peptide Liu CF and coworker successfully delivered mDHFR into the cell, paving the way toward refinement and optimization of approaches for the enzyme-replacement therapy [77].

3.3 Orthogonal Ligation

3.3.1 Chemoenzymatic tandem ligation

Chemoenzymatic ligation combines one or more enzymatic and chemical ligation methods to allow ligation of multiple peptide segments in tandem. The most common approach is a bidirectional modification of both ends of a peptide or protein in tandem. Cao *et al.* showed that protein thioester efficiently prepared by butelase-1 (Figure 8) could be used as substrate for chemical ligation method, the product of which can later be bioorthogonally modified by sortase A [82]. Butelase-mediated protein thioester preparation allows minimum protein structure disturbance as it only leaves a dipeptide trace -NG on the target protein.

The utility of the tandem chemoenzymatic approach was further demonstrated by Thompson *et al.* recently. Applying sortase A-mediated transamidation and intein-mediated protein splicing, they were able to introduce post-translational modifications

(PTMs) and biochemical probes into multiple proteins [86]. We envision that this novel and convenient approaches in tandem will open up a wide array of applications in the field of protein manipulation.

3.3.2 One-pot bioorthogonal ligation of biopolymers

One-pot bioorthogonal ligation exploits multiple ligases under the one-pot condition to reduce time required for the reaction and product yields. Butelase-1 has been used bioorthogonally with sortase A because of their different recognition signals. Consequently, they could be performed under one-pot condition. Figure 9 shows the one-pot protein conjugation using sortase A and butelase-1 to produce C-to-C protein-protein conjugate connected by a two-headed, PEG-based linker [87]. Similarly, in the same study, by appending incoming nucleophiles of butelase, Fmoc-Gly-Val-OH, and sortase A, Fmoc-Gly-Gly-Gly-OH, to strands of the double-stranded oligonucleotide, two different proteins with proper C-terminal recognition motifs can be ligated to the oligonucleotide-based linker (Figure 10). The protein-DNA-protein conjugate can be cleaved by restriction enzyme, relieving potential constraints caused by the linker.

3.3.3 One-pot antibody modification

Because of the bioorthogonality of ligases, we can use them to modify antibodies and biopolymers under one-pot conditions. This method is a useful new biological tool for preparing precise and highly tailored recombinant proteins and antibodies in high homogeneity. Figure 11 illustrates the site-specific modification of full-size antibody IgG1 by the orthogonal combination of butelase-1 and sortase A. The light chain and heavy chain of the antibody were both successfully labeled with probes by sortase A and butelase-1, respectively, with >95% conversion within 4 h [87].

3.4 Live-cell labeling

Labeling live-cell allows the visualization of the trafficking and interaction between cells. However, live-cell labeling is demanding because of its stringent requirement for the cells to survive during and after the ligation reaction. Commonplace in biological research, target proteins are recombinantly expressed with reporter proteins, causing potential disturbance of the protein folding and intracellular distribution. Using ligase-mediated ligation, various functional probes can be appended to the cell surface without compromised functionality.

Butelase-1 is a promising tool as it works under aqueous solution closed to neutral pH, a mild operating condition that allows cells to stay alive during the modification process. By genetically introduce butelase-1 recognition motif NHV to Lpp-OmpA protein at the cell surface, butelase-1 was found efficiently linking probes of various sizes to cell surface within 30 min (Figure 12) [24].

The ability of ligases to conjugate a variety of substrates enables live-cell engineering with non-genetically encoded molecules or those that require post-translational modifications. Several techniques have been exploited for live-cell surface manipulation, such as metabolic labeling [88, 89]. It is conceivable that a combined chemoenzymatic ligation approach for live-cell labeling opens up more possible applications.

3.5 Synthesis of peptides with unusual architectures

3.5.1 End-to-side chain cyclic peptides

Another conformational restricting tool for peptides is end-to-side chain cyclic peptides which improve proteolytic stability [90]. The best known naturally-occurring examples belong to the lasso peptides of the RiPPs superfamily.

Lasso peptides contain a cyclic structure consisting of an N-terminal macrolactam ring threaded with a C-terminal tail [91], creating various unusual architectures. The earliest lasso peptide Anantín was discovered in 1991 from the bacteria *Streptomyces coeruleus*, which bound to the cardiac hormone, which is synthesized in the human atrium [92].

Different than the other RiPPs, the unique scaffold cannot be manipulated using the chemical methods, but its production was shown to involve an ATP-dependent cysteine protease and adenylate-forming enzyme in bioprocessing of the genetically-encoded precursor peptide.

In general, a similar topology can be achieved using butelase-1-like ligase via side chain-to-tail cyclization. By blocking the N-terminus of a linear peptide, the Lys side chain can attack the thioester intermediate and the reaction is proximity-driven (Figure 13) [93].

3.5.2 Peptide dendrimer

Peptide dendrimers are true polypeptides in a branched format. They consist of a central lysine or other trifunctional as a core to tether identical peptides as branches. The original intent is to produce chemically-defined vaccines by increasing the immunogenicity antigenic peptides in a clustered dendrimeric format [94, 95]. Peptide dendrimers have since extended to other applications, including diagnostics, protein mimetics, vaccines, and therapeutics [96].

Peptide dendrimer appeared to be an ideal molecule to be tested by butelase-mediated ligation as many copies would be linked to the core to produce homogeneous products. Previously, we designed antimicrobials using a tetrapeptide RLYR motif [95]. Figure 14 shows that a tetra- or octa-branched RLYR dendrimer can be successfully achieved by ligating lysine cores with N-acetylated thiodipeptide, Ac-RLYRN-thioglc-V. The tetravalent dendrimeric peptide antimicrobial exhibited broad-spectrum antimicrobial activity against six drug-resistant strains [83]. The ability to produce controlled size dendrimers by butelase-1 provides an advance in producing synthetic vaccines and branched biopolymers

3.5.3 Cyclo-oligomerization of peptides

Cyclo-oligomerization combines both multimerization and head-to-tail cyclization of the resulted oligomers in a one-pot reaction. As such, it is an efficient approach to generate a focused cyclic peptide library. An example is the use of small peptide building blocks of 4-8 amino acids to create a library of the cyclic peptide of 12-24 amino acids by butelase-mediated cyclo-oligomerization method (Figure 15) [98].

The product distribution of the one-pot cyclo-oligomerization is controlled by reaction time, substrate concentration, peptide length, and peptide sequence. By manipulating these conditions, cyclo-oligomers of various shapes and sizes can be obtained. This simple and straightforward approach exemplifies the strengths of butelase-mediated, one-pot synthesis of simple peptide building blocks to complex products.

Cyclooligomeric antimicrobials were compared with their linear tetravalent dendrimers containing RLYR motif synthesized by butelase-mediated ligation. Both methods generated broad-spectrum antimicrobial activity with MICs ranging from <1 to $5\ \mu\text{M}$ [98]. However, cyclic antimicrobial peptides improved stability with the undiminished antimicrobial activity [99, 100]. As such, butelase-controlled one-pot cyclo-oligomerization could be useful for preparing potent antimicrobials with enhanced stability.

4 Conclusion and perspectives

Peptide bond-forming ligases are naturally-occurring superglues. These ligases are stand-alone and ATP-independent ligases and can be considered renegades because they catalyze ligation reactions unrestricted by the genetic code or machinery.

AEP-like ligases expand the biochemical tool kits by combining with other ligases and chemical ligation methods as bioorthogonal or chemoenzymatic approaches to work harmoniously in one-pot conditions and chemoenzymatic tandem reactions to modify proteins. These methods enable precision biomanufacturing of complex biologics such as homogeneous antibody-drug conjugates. The ability of ligases to work under mild and aqueous conditions also reduces the use of organic solvent and promotes green chemistry.

Future challenges include the understanding of the ligase mechanisms from their protease counterparts and to systematically identify novel ligases from nature. This novel group of renegade ligases expands the biological and chemical ligation toolbox to enable precise and site-specific modifications of peptides and proteins.

Acknowledgment

We thank every member of Tam's lab for discussions and comments. This work is supported by Academic Research Grant Tier 3 (MOE2016-T3-1-003) from the Singapore Ministry of Education.

Conflict of interest

The authors declare no conflict of interest.

References

- 1 Smith GP. *Science*, 1985, 228, 1315–1317
- 2 Verhoeven M, Milstein C, Winter G. *Science*, 1988, 239, 1534–1536
- 3 Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. *Nat. Biotechnol.*, 1998, 16, 258–261
- 4 Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, Schultz PG. *Proc Natl Acad Sci USA*, 2004, 101 (20): 7566–7571
- 5 Kent SBH. *Ann Rev Biochem*, 1988, 57, 957–89

- 6 Liu CF, Tam JP. *Proc Natl Acad Sci USA*, 1994, 91(14): 6584-8
- 7 Liu CF, Tam JP. *J Am Chem Soc*, 1994, 116(10): 4149-4153
- 8 Muir TW, *Annu Rev Biochem*, 2003, 72, 249-289
- 9 Tam JP, Lu YA, Liu CF, Shao J. *Proc Natl Acad Sci USA*, 1995, 92(26): 12485-12489
- 10 Pan M, Zheng Q, Ding S, Zhang L, Qu Q, Wang T, Hong D, Ren Y, Liang L, Chen C, Mei Z, Liu L. *Angew Chem*, 2019, 131(9): 2653-2657
- 11 Scott JK, Smith GP. *Science*, 1990, 249(4967): 386-390
- 12 Moolten FL, Cooperband SR. *Science*, 1970, 68-70
- 13 McCafferty J, Griffiths AD, Winter G, Chiswell DJ. *Nature*, 1990, 348(6301): 552-554
- 14 Beck A, Goetsch L, Dumontet C, Corvaia N. *Nat Rev Drug Discov*, 2017, 16(5): 315-337
- 15 Tam JP, Lu YA. *Tetrahedron Lett*, 1997, 38(32): 5599-5602
- 16 Tam JP, Lu YA, Yu Q. *J Am Chem Soc*, 1999, 121(18): 4316-4324
- 17 White CJ, Yudin AK. *Nat Chem*, 2011, 3(7): 509-524
- 18 Tam JP, Wong CT. *J Biol Chem*, 2012, 287(32): 27020-27025
- 19 Yoshihara HA, Mahrus S, Wells JA. *Bioorg Med Chem Lett*, 2008, 18(22): 6000-6003
- 20 Rosen CB, Francis MB. *Nat Chem Biol*, 2017, 13(7): 697-705
- 21 Mahal LK, Yarema KJ, Bertozzi CR. *Science*, 1997, 276(5315): 1125-1128
- 22 Chen I, Howarth M, Lin W, Ting AY. *Nat Methods*, 2005, 2(2): 99-104
- 23 Popp MWL, Karssemeijer RA, Ploegh HL. *PLoS pathog*, 2012, 8(3): e1002604
- 24 Bi X, Yin J, Nguyen GKT, Rao C, Halim NBA, Hemu X, Tam JP, Liu CF. *Angew Chem*, 2017, 129(27): 7930-7933
- 25 Frank J. *Nat Protoc*, 2017, 12(2): 209-212
- 26 Lipmann F. *Adv Microb Physiol*, 1981, 21, 227-266
- 27 Kopp F, Marahiel MA. *Nat Prod Rep*, 2007, 24(4): 735-749
- 28 Schmidt, M., et al., Enzyme-mediated ligation technologies for peptides and proteins. *Curr Opin Chem Biol*, 2017. 38: 1-7.
- 29 Xu, S., Z. Zhao, and J. Zhao, Recent advances in enzyme-mediated peptide ligation. *Chin Chem Lett*, 2018. 29(7): 1009-1016.
- 30 Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J. *Nat Prod Rep*, 2013, 30(1): 108-160
- 31 Schmidt EW, Sudek S, Haygood MG. *J Nat Prod*, 2004, 67(8): 1341-1345
- 32 Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J. *Proc Natl Acad Sci USA*, 2005, 102(20): 7315-7320
- 33 Abe Y, Shirane K, Yokosawa H, Matsushita H, Mitta M, Kato I, Ishii S. *J Biol Chem*, 1993, 268(5): 3525-3529
- 34 CJ, Pujara PT, Reed DW, Chiwocha S, Zhang H, Covello PS. *J Biol Chem*, 2013, 288(18): 12500-12510
- 35 Nguyen GKT, Wang S, Qiu Y, Hemu X, Lian Y, Tam JP. *Nat Chem Biol*, 2014, 10, 732-738
- 36 Harris KS, Durek T, Kaas Q, Poth AG, Gilding EK, Conlan BF, Saska I, Daly NL, Van Der Weerden NL, Craik DJ. *Nat Commun*, 2015, 6, 10199-10208
- 37 Haywood J, Schmidberger JW, James AM, Nonis SG, Sukhoverkov KV, Elias M, Bond CS, Mylne JS. *eLife*, 2018, 7, e32955
- 38 Jackson M, Gilding E, Shafee T, Harris K, Kaas Q, Poon S, Yap K, Jia H, Guarino R, Chan L. *Nat Commun* 2018, 9(1): 2411-2422

- 39 Hemu X, El Sahili A, Hu S, Wong K, Chen Y, Wong YH, Zhang X, Serra A, Goh BC, Darwis DA, Chen MW, Sze SK, Liu CF, Lescar J, Tam JP. *Proc Natl Acad Sci USA*, 2019, 116(24): 11737-11746
- 40 Harris KS, Guarino RF, Dissanayake RS, Quimbar P, McCorkelle OC, Poon S, Kaas Q, Durek T, Gilding EK, Jackson MA, Craik DJ, van der Weerden NL, Anders RF, Anderson MA. *Sci Rep*, 2019, 9(1): 10820-10820
- 41 Mazmanian SK, Liu G, Ton-That H, Schneewind O. *Science*, 1999, 285(5428): 760-763
- 42 Lee J, McIntosh J, Hathaway BJ, Schmidt EW. *J Am Chem Soc*, 2009, 131(6): 2122-2124
- 43 Luo H, Hong SY, Sgambelluri RM, Angelos E, Li X, Walton JD. *Chem Biol*, 2014, 21(12): 1610-1617
- 44 Craik DJ. *Toxins (Basel)*, 2012, 4(2): 139-156
- 45 Hatsugai N, Yamada K, Goto-Yamada S, Hara-Nishimura I. *Front Plant Sci*, 2015, 6, 234-244
- 46 Dall E, Brandstetter H. *Biochimie*, 2016, 122:126-150
- 47 Harley SM, Michael Lord J. *Plant Sci*, 1985, 41(2): 111-116
- 48 Hara-Nishimura I, Nishimura M. *Plant Physiol*, 1987, 85(2): 440-445
- 49 Bowles DJ, Marcus SE, Pappin D, Findlay J, Eliopoulos E, Maycox PR, Burgess J. *J Cell Biol*, 1986, 102(4): 1284-1297
- 50 Min W, Jones DH. *Nat Struct Mol Biol*, 1994, 1(8): 502-504
- 51 James AM, Haywood J, Leroux J, Ignasiak K, Elliott AG, Schmidberger JW, Fisher MF, Nonis SG, Fenske R, Bond CS, Mylne JS. *Plant J*, 2019, 98(6): 988-999
- 52 Pi N, Gao M, Cheng X, Liu H, Kuang Z, Yang Z, Yang J, Zhang B, Chen Y, Liu S, Huang Y, Su Z. *Biochemistry*, 2019, 58(27): 3005-3015
- 53 Nguyen GKT, Kam A, Loo S, Jansson AE, Pan LX, Tam JP. *J Am Chem Soc*, 2015, 137(49): 15398-15401
- 54 Schechter I, Berger A. *Biochem Biophys Res Commun*, 1967, 27(2): 157-162
- 55 Nguyen GKT, Qiu Y, Cao Y, Hemu X, Liu CF, Tam JP. *Nat Protoc*, 2016, 11(10): 1977-1988
- 56 Zauner FB, Dall E, Regl C, Grassi L, Huber CG, Cabrele C, Brandstetter H. *Plant Cell*, 2018, 30(3): 686-699
- 57 Zhao L, Hua T, Crowley C, Ru H, Ni X, Shaw N, Jiao L, Ding W, Qu L, Hung L-W. *Cell Res*, 2014, 24(3): 344-358.
- 58 Ludewig H, Czekster CM, Oueis E, Munday ES, Arshad M, Synowsky SA, Bent AF, Naismith JH. *ACS Chem Biol*, 2018, 13(3): 801-811
- 59 Popp MWL, Ploegh HL. *Angew Chem*, 2011, 50(22): 5024-5032
- 60 Dorr BM, Ham HO, An C, Chaikof EL, Liu DR. *Proc Natl Acad Sci USA*, 2014, 111(37): 13343-13348
- 61 Pishesha N, Ingram JR, Ploegh HL. *Annu Rev Cell Dev Biol*, 2018, 34(1): 163-188
- 62 Zakeri B, Howarth M. *J Am Chem Soc*, 2010, 132(13): 4526-4527
- 63 Fierer JO, Veggiani G, Howarth M. *Proc Natl Acad Sci USA*, 2014, 111(13): E1176-E1181
- 64 Neet KE, Koshland DE, Jr. *Proc Natl Acad Sci USA*, 1966, 56(5): 1606-1611
- 65 Polgar L, Bender ML. *Biochemistry*, 1967, 6(2): 610-620
- 66 Nakatsuka T, Sasaki T, Kaiser ET. *J Am Chem Soc*, 1987, 109(12): 3808-3810
- 67 Wu ZP, Hilvert D. *J Am Chem Soc*, 1989, 111(12): 4513-4514

- 68 Abrahmsen L, Tom J, Burnier J, Butcher KA, Kossiakoff A, Wells JA. *Biochemistry*, 1991, 30(17): 4151-4159
- 69 Chang TK, Jackson DY, Burnier JP, Wells JA. *Proc Natl Acad Sci USA*, 1994, 91(26): 12544-12548
- 70 Liu CF, Tam JP. *Org Lett*, 2001, 3(26): 4157-4159
- 71 Tan XH, Wirjo A, Liu CF. *Chembiochem*, 2007, 8(13): 1512-1515
- 72 Tan XH, Zhang X, Yang R, Liu CF. *Chembiochem*, 2008, 9(7): 1052-1056
- 73 Tan X, Yang R, Liu CF. *Org Lett*, 2018, 20(21): 6691-6694
- 74 Weeks AM, Wells JA. *Nat Chem Biol*, 2018, 14(1): 50-57
- 75 Nguyen GKT, Hemu X, Quek JP, Tam JP. *Angew Chem*, 2016, 128(41): 12994-12998
- 76 Hemu X, Qiu Y, Nguyen GK, Tam JP. *J Am Chem Soc*, 2016, 138(22): 6968-6971
- 77 Bi X, Yin J, Hemu X, Rao C, Tam JP, Liu CF. *Bioconjug Chem*, 2018, 29(7): 2170-2175
- 78 Tam JP, Yu Q, Miao Z. *Pept Sci*, 1999, 51(5): 311-332
- 79 Tam JP, Xu J, Eom KD. *Pept Sci*, 2001, 60(3): 194-205
- 80 Nguyen GKT, Cao Y, Wang W, Liu CF, Tam JP. *Angew Chem*, 2015, 127(52): 15920-15924
- 81 Tam JP, Nguyen GKT, Kam A, Loo S. Advances in Site-specific and Linkage-specific Ligation – Precision ligation of DARPin-conjugates targeting the breast cancer HER2 receptor. In: Timmons PB, Hewage CM, Lebl M, Eds. *Proceedings of the 35th European Peptide Symposium*. Chichester, United Kingdom, 2018. 3-7
- 82 Cao Y, Nguyen GK, Tam JP, Liu CF. *Chem Commun*, 2015, 51(97): 17289-17292
- 83 Cao Y, Nguyen GKT, Chuah S, Tam JP, Liu CF. *Bioconjug Chem*, 2016, 27(11): 2592-2596.
- 84 Aimoto S. *Pept Sci*, 1999, 51(4): 247-265
- 85 Chin JW. *Nature*, 2017, 550, 53-60
- 86 Thompson RE, Stevens AJ, Muir TW. *Nat Chem*, 2019, 11(8): 737-743
- 87 Harmand T, Bousbaine D, Chan AI, Zhang X, Liu DR, Tam JP, Ploegh HL. *Bioconjug Chem*, 2018 29(10): 3245-3249
- 88 Saxon E, Bertozzi CR. *Annu Rev Cell Dev Biol*, 2001, 17(1): 1-23
- 89 Siegrist MS, Swarts BM, Fox DM, Lim SA, Bertozzi CR. *FEMS Microbiol Rev*, 2015, 39(2): 184-202
- 90 Knappe TA, Linne U, Robbel L, Marahiel MA. *Chem Biol*, 2009, 16(12): 1290-1298
- 91 Hegemann JD, Zimmermann M, Xie X, Marahiel MA. *Acc Chem Res*, 2015, 48(7): 1909-1919
- 92 Weber W, Fischli W, Hochuli E, Kupfer E, Weibel E. *J Antibiot*, 1991, 44(2): 164-171
- 93 Yang RL, Wong YH, Nguyen GKT, Tam JP, Lescar J, Wu B. *J Am Chem Soc*, 2017, 139(15): 5351-5358
- 94 Posnett DN, McGrath H, Tam JP. *J Biol Chem*, 1988, 263(4): 1719-1725
- 95 Tam JP. *Proc Natl Acad Sci USA*, 1988, 85(15): 5409-5413
- 96 Sadler K, Tam JP. *Mol Biotechnol*, 2002, 90(3): 195-229
- 97 Tam JP, Lu YA, Yang JL. *Eur J Biochem*, 2002, 269(3): 923-932
- 98 Hemu X, Zhang X, Tam JP. *Org Lett*, 2019, 21(7), 2029-2032

- 99 Tam JP, Lu YA, Yang JL. *Biochem Biophys Res Commun*, 2000, 267(3): 783-790
- 100 Tam JP, Wu C, Yang JL. *Eur J Biochem*, 2000, 267(11): 3289-3300

Scheme 1 Concept of ligation and the applications.

Scheme 2 Mechanism of butelase-mediated ligation.

Figure 1 Architecture of butelase-1. The crystal of recombinant butelase-1 (PDB entry: 6HDI) has its N-terminal endoplasmic reticulum signal replaced by 6-His tag and Gly-Ser linker, followed by Ile21-Val482 of native butelase-1. The core domain of butelase-1 exhibited six β -sheet, flanked by five α -helices (colored in gray). The catalytic triad, Asn59, His165 (modeled as succinimide residue), and Cys207 (colored in red), are 'protected' by the cap domain (colored in blue) before acid-induced activation. The orange dashed line represents the linker connecting cap domain and core domain.

Figure 2 Total synthesis of bacteriocin AS-48 adapted from Hemu *et al.* [764]. Linear precursor of AS-48K was assembled by Fmoc chemistry using microwave-assisted synthesizer. The peptide resin was then cleaved and the unprotected precursor was dissolved in 8 M urea and purified by reverse-phase HPLC. The unfolded AS-48K was redissolved then refolded by direct or stepwise dialysis. Butelase-mediated macrocyclization of folded AS-48K was performed with enzyme:peptide ratio of 1:100 at pH 6, 37 °C.

Figure 3 Butelase-mediated N-terminal protein modification. By introducing nucleophile dipeptide GI to the N-terminus of target protein, cargo of choice with the recognition motif NHV will be ligated to the target protein.

Figure 4 N-terminal protein labeling using butelase-1 and thiopeptide. The thiopeptide substrate (colored in green) carrying biotin was ligated to ubiquitin by butelase-1-catalyzed ligation (0.001 molar equivalents).

Figure 5 C-terminal protein modification using butelase-1. By expressing the protein of interest with a tripeptide recognition signal NHV, butelase-1 can ligate the cargo of choice to the protein C-terminus.

Figure 6 Butelase-mediated linkage-specific ligation of designed ankyrin repeat proteins 926 (DARPin 926) for bioimaging and precision bioconjugation. Both substrates were linked to DARPin 926 with >90% yield in 30 min.

Figure 7 Preparation of protein thioester by butelase-mediated ligation. Protein of interest contains NHV tripeptide motif that allows the thioester group to be linked to the protein by butelase-1.

Figure 8 Dual-terminus modification by tandem ligation. A ubiquitin with its N- and C-terminus engineered with recognition signals of sortase A and butelase1, respectively, was prepared. Glycine thioester was ligated to the ubiquitin at the C-terminus by butelase-mediated ligation. By orthogonal ligation, a biotinyl peptide was linked to the ubiquitin with thioester. At the N-terminus, sortase A recognized the LPETG pentapeptide motif and ligated the fluorescein-peptide fluor-YLPET-glc_G to the ubiquitin.

Figure 9 Preparation of C-to-C fusion protein by butelase-1 and sortase A in a one-pot reaction. Two different proteins equipped with recognition motifs of butelase-1 and sortase A, respectively, were linked to two-headed PEG-based linker. The linkers were synthesized by solid-phase peptide synthesis (SPPS) and contained incoming groups of both enzymes, GGG motif for sortase A and VG for butelase-1, at its both ends.

Figure 10 Dual labeling of oligonucleotide-based linker by sortase A and butelase-1. Complementary single-strand oligonucleotides were synthesized on a solid support. Fmoc-Gly-Gly-Gly-OH was attached to one strand, and Fmoc-Gly-Val-OH was attached to the other.

Figure 11 Full-sized IgG1 dual modification by butelase-1 and sortase A in a one-pot reaction. A modified IgG1 molecule containing the C-terminal LPETGG motif at the κ (light) chain and C-terminal NHV motif at the γ (heavy) chain was prepared. Using two fluorescent probes with proper nucleophiles at the N-terminus, the IgG1 was selectively modified by butelase-1 and sortase A.

Figure 12 Labeling of the live bacterial cell by butelase-1. Plasmid encoding the anchoring protein OmpA with C-terminal NHV motif was transformed into BL21 (DE3) *E. coli* strain, allowing the display of recognition signal on the *E. coli* cell surface. The 5(6)-carboxyfluorescein-peptide, GIGGIRK, was prepared by solid-phase peptide synthesis (SPPS) then linked to the cell surface by butelase-1.

Figure 13 Sidechain-to-tail cyclic peptides prepared using butelase-1 adapted from Yang *et al.* [934]. With the N-terminal amine of the peptide substrate protected, butelase1 catalyzed the amide bond formation between C-terminal Asn and the sidechain of Lys, instead of N-terminal Gly.

Figure 14 Synthesis of peptide dendrimer by butelase-1. An N-acetylated thiopeptide was ligated to the bivalent lysyl scaffold. Peptide bonds were formed using butelase-mediated ligation. The enzymatic multimerization can be achieved using octa-branched lysly cores.

Figure 15 One-step cyclo-oligomerization of antimicrobial peptides dendrimer controlled by butelase-1 adapted from Hemu *et al.* [96]. The linear precursors of cyclo-oligomeric AMP contained RLYR motif and NHV recognition signal of butelase-1 were prepared by Fmoc chemistry. AMP precursors with less than 9 residues (excluding His-Val) were linked by butelase-1-catalyzed head-to-tail cyclization.

Table 1 Occurrences of ATP-independent and renegade ligases

Ligase	Selected native product	Species	Refs
1. Thiol protease-like ligases			
butelase-1	Cliotides	<i>C. ternatea</i> ²	[35]
VyPAL1-3	Cycloviolacin Y5	<i>V. yedoensis</i> ²	[39]
OaAEP1b & 3-5	Kalata B1	<i>O. affinis</i> ²	[36, 40]
2. Serine protease-like ligases			
PatG	Patellamides	<i>Prochloron</i> sp. ¹	[32]
PCY1	Orbitides	<i>S. vaccaria</i> ²	[34]
POPB	α -amanitin	<i>G. marginata</i> ³	[43]
3. Transpeptidase			
Sortase A	Cell wall sorting	<i>S. aureus</i> ¹	[41]
SpyLigase	Pillin	<i>S. pyogenes</i> ¹	[62, 63]
Derived from microbe ¹ , plant ² , fungus ³			

Table 2. Recognition motif of ATP-independent ligases

Ligase	C-terminal recognition motif	N-terminal incoming group	Refs
1. Thiol protease-like ligases			
butelase-1	<u>NH</u> ¹ V ²	G ³ I ⁴	[35]
VyPAL2	<u>NS</u> ¹ L ²	G ³ I ⁴	[39]
OaAEP1b & 3-5	<u>NGL</u>	GL	[36, 40]
2. Serine protease-like ligases			
PatG	AYDG	P/thiazoline	[32, 42]
PCY1	FQA / IQT	FSA/VGAG	[34, 59]
POPB	IWGIGCNP	-	[43]
subtiligase	Peptide thioester/ester	-	[68]
3. Transpeptidase			
Sortase A	LPE ¹ <u>TG</u>	GGG	[41]
SpyLigase	SpyTag (AHIVMVDA Y KPTK)	KTag (ATHIKFSK RD)	[62, 63]

*The cleavage site of ligases are underlined. ¹, the P1' position accept almost all 20 amino acid; ², the P2' position prefers hydrophobic amino acids; ³, the P1'' position accepts almost all 20 amino acid; ⁴, the P2'' position prefers hydrophobic amino acids (using nomenclature by Schechter and Berger [54]).

Table 3. Intramolecular ligation mediated by butelase-1

No	Substrate-HV	Sequence	Length (aa)	Time (min)	Yield (%)	Refs
1	SFTI	GRCTKSIPPICFPN	14	45	>95	[35]
2	D-SFTI	GrctksippicfpN	14	60	>95	[75]
3	D-MrIA	GvcegyklchpcagN	15	15	>95	[75]
4	D-θ-defensin analog	GvrcicrrGfcrclcrN	18	60	>95	[75]
5	Thanatin	GSKKPVPIIHCNRRTGKCQRMN	22	240	59	[35]
6	Rat neuromedin	GIKYGVNEYQGPVAPSGGFFLFRP RN	26	5	>95	[53]
7	Kalata-B1	GLPVCGETCVGGTCNTPGCTCSW PVCTRN	29	45	>95	[34]
8	Apelin	GLVQPRGSRNGPGPWQGRRKFR RQRPRLSHKGPMPFN	38	5	>95	[53]
9	AS-48	VVEAGGWVTTIVSILTAVGSGGLS LLAAAGRESIKAYLKKEIKKKGKR AVIAWMAKEFGIPAAVAGTVLN	70	60	>85	[76]
10	UblA	YISRNLKAQAVIWLAGYTGIASGT AKKVVD AIDKGAAAFVIISIISTVIS AGALGAVSASADFIILTVKN	70	1440	>93	[76]
11	p53-binding domain (N-terminal domain) of murine double minute X (N-MdmX)	GLQINQVRPKLPLLKILHAAGAQG EMFTVKEVMHYLGQYIMVKQLY DQQEQHVMVYAGDILLGELLGRQS FSVKDPSPLYDMLRKNLVTLATN	92	40	>95	[52]
12	Human growth hormone (somatropin)	<u>FPTIPLSRLFQNA MLRAHRLHQLA</u> <u>FDTYEEFEEA YIPKEQKYSFLQAPQ</u> <u>ASLCFSESIPTSPNREQAQQKSNLQ</u> <u>LLRISLLLIQSWLEPVGFLRSVFAN</u> <u>SLVYGASDSVDVYDLLKDLEEGIQT</u> <u>LMGRLEDGSPRTGQAFKQTYAKF</u> <u>DANSHNDDALLKNYGLLYCFRKD</u> <u>MDKVETFLRIVQCRSVEGSCGFN</u>	192	15	>85	[53]
13	IL-1Ra	GISYDYMEGGDIRVRRFLFCRTQW YLRIDKRGKVKGTQEMKNYNNIM	143	15	>90	[53]

		EIRTVAVGIVAIAKGVSEFYLAMN				
		KEGKLYAKKECNEDCNFKELILEN				
		HYNTYASAKWTHNGGEMFVALN				
		QKGIPVRGKKTKKEQKTAHFLPM				
		AITN				
14	V44-DHFR	SIAGGVRPLNSIVAVSQNMIGIGN	197	30	N.D.	[77]
		GDLPWPPLRNEFKYFQRM TTTSS-				
		tag-				
		EGKQNLVIMGRKTWFSIPEKNRPL				
		KDRINIVLSRELKEPPRGAHFLAKS				
		LDDALRLIEQPELASKVDMVWIVG				
		GSSVYQEAMNQPGHLRLFVTRMQ				
		EFESDTFFPEIDLKYLKLLPEYPGV				
		LSEVQEEKGIKYKFEVYEKKGSRS				
		GSGN				
15	GFP	GISMSKGEELFTGVVPILVELDGD	242	15	>95	[53]
		VNGHKFSVSGEGEGDATYGLTL				
		KFICTTGKLPVPWPTLVTTLTYG				
		VCFSRYPDHMKQHDFFKSAMPEG				
		YVQERTIFFKDDGNYKTRAEVKFE				
		GDTLVNRIELKGIDFKEDGNILGH				
		KLEYNYNSHNVYIMADKQKNGIK				
		VNFKIRHNIEDGSVQLADHYQQNT				
		PIGDGPVLLPDNHYLSTQSALSKD				
		PNEKRDHMLLEFVTAAGITLGM				
		DELYKN				

*The underlined residue will be ligated to the N-terminal amino acid. Residues in lower-case are in the D-configuration. N.D.: not determined

Table 4. Intermolecular ligation mediated by butelase-1

	N-terminal Substrate	C-terminal Substrate	Refs
N-terminal labeling	YKN-thioglc-V	GIGGIR	[80]
	Biotin-TYKN-thioglv-V	GI-Ubiquitin	[80]
	YKN-thioglv-V	MI-GFP	[80]
C-terminal labeling	DARPin-NHV	GIGKFLHSAKKFGKAFVGEIMNS	[81]
	DARPin-NHV	RIGK- fluorescein amidites	[81]
Live-cell labeling	OmpA-NHV	GIGGIRK-fluorescein	[24]
		GIGGIRK-Biotin	[24]
		Monoglycated Muc1-like peptide with C-terminal Biotin	[24]
Peptide/protein ligation	KALVINHV	XIGGIR (X = 20 natural amino acids)	[35]
		GXGGIR (X = 20 natural amino acids)	[35]
Antibody-fluorescence conjugate	IgG1 γ 1-NHV	AL-Alexa (AlexaFluor 647 fluorescent dye)	[87]
Protein-oligonucleotide-protein conjugate	VHH-Enh-NHV	GV-Fmoc-DNA	[87]
Protein-protein conjugate	VHH-Enh-NHV	Two-headed PEG linker (NH ₂ -GGG-PEG-AL-NH ₂)	[87]
Dendrimer conjugation	Ac-RYRLN-thioglc-V	(RI β A)2KY	[83]
		(RI β A)4K2KY	[83]
		(RI β A)8K4K2KY	[83]
Peptide/protein-thioester preparation	YXN-NHV	YXNG-COSR (X = V, L, S, F, Nle, d-A)	[83]
	Ubiquitin-NHV	GGMQIFVKTLTGKTITLEVEPSDTIENVK AKIQDKEGIPPDQQLIFAGKQLEDGRTL SDYNIQKESTLHLVLRLRGGNXNG- COSR	[82]
	DARPin(ERK)-NHV	SMGSDLGKKLLEAARAGQDDEVRLMA NGADVNAHDDQGSTPLHLAAWIGHPEIV EVLLKHGADVNAARDTDGWTPLHLAADN GHLEIVEVLLKYGADVNAQDAYGLTPL HLAADRGHLEIVEVLLKHGADVNAQDK FGKTAFDISIDNGNEDLAEILQKLNKNXN G-COSR	[82]
	GFP-NHV	MSKGEELFTGVVPILVELDGDVNGHKFS VSGEGEGDATYGKLTCLKFICTTGKLPVP	[82]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

WPTLVTTLTYGVCFSRYPDHMKQHDF
FKSAMPEGYVQERTIFFKDDGNYKTRAE
VKFEGDTLVNRIELKGIDFKEDGNILGHK
LEYNYNSHNVYIMADKQKNGIKVNFKIR
HNIEDGSVQLADHYQQNTPIGDGPVLLP
DNHYLSTQSALSKDPNEKRDHMLLEFV
TAAGITLGMDELYKNXNG-COSR

*GFP is green fluorescence protein, IgG is immunoglobulin, DARPin is designed ankyrin repeat proteins,

TOC figure