Turning an asparaginyl endopeptidase into a peptide ligase

Hemu, Xinya; El Sahili, Abbas; Hu, Side; Zhang, Xiaohong; Serra, Aida; Goh, Boon Chong; Darwis, Dina Amallia; Chen, Ming Wei; Sze, Siu Kwan; Liu, Chuan-Fa; Lescar, Julien; Tam, James P.

2020

Hemu, X., El Sahili, A., Hu, S., Zhang, X., Serra, A., Goh, B. C., Darwis, D. A., Chen, M. W., Sze, S. K., Liu, C., Lescar, J. & Tam, J. P. (2020). Turning an asparaginyl endopeptidase into a peptide ligase. *ACS Catalysis*, 10(15), 8825-8834.
<https://dx.doi.org/10.1021/acscatal.0c02078>

<https://hdl.handle.net/10356/148032>

<https://doi.org/10.1021/acscatal.0c02078>

This document is the Accepted Manuscript version of a Published Work that appeared in final form in *ACS Catalysis*, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see <https://doi.org/10.1021/acscatal.0c02078>

Downloaded on 26 Apr 2023 19:06:39 SGT

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Turning an asparaginyl endopeptidase into a peptide ligase

Journal:	ACS Catalysis
Manuscript ID	Draft
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	<p>Hemu, Xinya; Nanyang Technological University, School of Biological Sciences</p> <p>El Sahili, Abbas; Nanyang Technological University, HU, SIDE; Nanyang Technological University College of Science, School of Biological Sciences</p> <p>Zhang, Xiaohong; Nanyang Technological University, School of Biological Sciences</p> <p>Serra, Aida; IMEDA Alimentación</p> <p>Goh, Boon Chong; Singapore-MIT Alliance for Research and Technology Centre, AMR IRG</p> <p>Darwis, Dina; National University of Singapore</p> <p>Chen, Ming Wei; Nanyang Technological University, School of Biological Sciences</p> <p>Sze, Siu Kwan; Nanyang Technological University, Chemical Biology and Biotechnology</p> <p>Liu, Chuan-Fa; Nanyang Technological University, School of Biological Sciences</p> <p>Lescar, Julien; Nanyang Technological University, School of Biological Sciences</p> <p>Tam, James; Nanyang Technological University, School of Biological Sciences</p>

SCHOLARONE™
Manuscripts

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

Turning an asparaginyl endopeptidase into a peptide ligase

Xinya Hemu,1‡ Abbas El Sahili,1,2‡ Side Hu,1,2 Xiaohong Zhang,1 Aida Serra,1,3 Boon Chong Goh,2,4 Dina A. Darwis,1,5 Ming Wei Chen,1 Siu Kwan Sze,1 Chuan-fa Liu,1 Julien Lescar,1,2 James P. Tam1**

1 School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551

2 NTU Institute of Structural Biology, Nanyang Technological University. Experimental Medicine Building, 59 Nanyang Drive, Singapore 636921

3 IMDEA Food Research Institute, Carr. de Canto Blanco, 8, Madrid 28049, Spain

4 Antimicrobial Resistance Interdisciplinary Research Group, SMART, 1 CREATE Way, Singapore 138602

5 Synthetic Biology for Clinical and Technological Innovation, National University of Singapore. 14 Medical Drive, Singapore 117599

‡ These authors contribute equally.

KEYWORDS

1 Peptide asparaginyl ligase (PAL), Asparaginyl endopeptidase (AEP),
2 ligase-activity determinant (LAD), enzymatic ligation, site-specific
3 ligation, protein engineering, enzymatic cyclization
4
5
6
7
8
9
10
11
12

13 **ABSTRACT**

14
15
16
17 Butelase-1 is a peptide asparaginyl ligase (PAL) that efficiently
18 catalyzes peptide bond formation after Asn/Asp (Asx), whereas its
19 homolog butelase-2 is an asparaginyl endopeptidase (AEP) that catalyzes
20 the reverse reaction, hydrolyzing Asx-peptide bonds. Since PALs and
21 AEPs share essentially similar overall structures, we surmised that the
22 S2 and S1' substrate-binding pockets immediately flanking the catalytic
23 S1 site constitute the major ligase activity determinants (LADs) that
24 control the catalytic directionality of these enzymes. Here, we report
25 the successful conversion of butelase-2 into butelase-1-like ligases
26 based on the LAD hypothesis. We prepared 23 LAD-mutants by mutating
27 residues of the S2 pocket (LAD1) and/or the S1' pocket (LAD2) of
28 butelase-2 into homologous residues in PALs. These LAD-mutants markedly
29 diminished protease activity and increased ligase activity. In
30 contrast, substituting 12 non-LAD residues to the corresponding
31 residues in butelase-1 did not change their protease profiles. At
32 physiological pH, mutations targeting both LADs resulted in shifting
33 the catalytic directionality from 95% hydrolysis to >95% peptide
34 ligation. This results in the engineering of efficient recombinant
35 peptide ligases that were demonstrated to be useful for macrocyclization
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

1 and site-specific labeling of bioactive peptides and proteins. Five
2 high-resolution crystal structures of butelase-2 and its engineered
3 mutants reveal subtle changes proximal to the catalytic S1 site that
4 account for the reversal in enzymatic activity. Computational
5 simulations of enzyme-substrate complexes suggest how the S-acyl
6 intermediate is positioned at the S2 site and the substrate orientated,
7 controlling accessibility of either water or incoming nucleophiles from
8 the prime-side (S1' site) of the catalytic center. Together, these
9 features determine the catalytic directionality between hydrolysis and
10 aminolysis in AEPs and PALs. Overall, this work validates the LAD
11 hypothesis as a central guide for making peptide ligases from their
12 corresponding widely available protease counterparts, to produce
13 precision tools for exquisite site-specific conjugation and
14 biomanufacturing of biologics.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

33 INTRODUCTION

34
35 Forming a peptide bond is a highly regulated reaction in biology. In
36 living organisms, it occurs exclusively in ribosomes. Besides
37 ribosomes, ATP-dependent non-ribosomal peptide synthesis is found in
38 microbial organisms^{1, 2} and only a few enzymes are known to catalyze
39 peptide bond formation in an ATP-independent manner.³⁻⁸
40
41
42
43
44
45
46

47 Proteases are known to mediate peptide bond formation via a reverse
48 hydrolysis reaction when an amine nucleophile is proximal and available
49 to compete with water at the active site.⁹⁻¹² However, this reverse
50 reaction by amine nucleophilic attack of the acyl-intermediate, i.e.
51 aminolysis, is thermodynamically disfavored under most physiological
52 conditions. Shifting the equilibrium of protease-catalyzed reactions
53
54
55
56
57
58
59
60

1 from hydrolysis to aminolysis has been studied in diverse angles, such
2 as changing pH that alter the ionization state of reactive
3 functionalities,¹³ using organic solvents to exclude water molecules,¹⁴
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

Recently, a new type of ATP-independent peptide-bond-forming enzymes was isolated from cyclotide-producing plants: peptide ligases mediate macrocyclization of cyclotides by joining their N-terminal residue to a conserved Asn near the C-terminus of the cyclotide linear precursors, via a newly formed peptide bond.^{21, 22} The ligase of this kind that was first discovered and characterized is butelase-1 from *Clitoria ternatea* (Fabaceae), which so far is the most efficient peptide ligase known and has a wide range of applications for peptide synthesis and protein engineering.^{21, 23-27} Other butelase-1-like ligases include OaAEP1b/3/4/5 from *Oldenlandia affinis* (Rubiaceae),^{22, 28} HeAEP3 from *Hybanthus enneaspermus* (Violaceae),²⁹ and VyPALs from *Viola yedoensis* (Violaceae),³⁰ all of which were identified from cyclotide-producing plants as processing enzymes crucial for the cyclotide macrocyclization biosynthetic pathway.^{31, 32}

Butelase-1 or their likes are classified under the C13 subfamily of Cys proteases (EC3.4.22.34),³³ since they possess the same substrate specificity, recognizing Asn at the P1 site and also similar structural features as the representative and founding members, asparaginyl endopeptidases (AEPs). C13 Cys proteases consist of a conserved

1 catalytic core domain and a C-terminal cap domain that entirely covers
2 their active site,³⁴ These precursor enzymes require autolytic
3 activation to dissociate or remove the cap domain to expose their
4 substrate-binding surface to the solvent. This post-translational
5 processing event occurs in the vacuoles or lysosomes, where these
6 enzymes display intrinsic hydrolytic activity at acidic pH.³⁵⁻³⁷ As its
7 name describes, AEPs preferentially catalyze peptide bond hydrolysis
8 after an internal Asn/Asp residue. In contrast, the uncommon ligases
9 found in cyclotide-producing plants preferentially catalyze aminolysis
10 of Asn/Asp even under acidic conditions. Therefore, we named these
11 enzymes "peptide asparaginyl ligases" (PALs) to distinguish them from
12 typical AEPs.³⁰

13
14 AEP-catalyzed hydrolysis and PAL-catalyzed aminolysis share the same
15 initial step: formation of an enzyme-substrate S-acyl intermediate,
16 which covalently links the catalytic Cys and the P1 residue Asn/Asp by
17 a thioester bond that can in principle become accessible to either a
18 water molecule or other nucleophiles. In contrast with AEPs that favor
19 water molecules in the subsequent reaction, PALs have evolved to
20 accommodate N-terminal amine nucleophiles from an incoming peptide
21 either intramolecularly or intermolecularly, which replaces the leaving
22 group within the S' specificity pockets (Figure 1). Protease-mediated
23 reverse hydrolysis can be facilitated by an increase of pH, which shifts
24 the ionization equilibrium of the N-terminal amine of the incoming
25 peptide from the protonated NH₃⁺ to the unprotonated nucleophilic NH₂.
26 Therefore, at near neutral to basic pH, reverse proteolysis has been
27 observed for several AEPs leading to religation of the cap domain to

the catalytic core domain *in vitro*^{37, 38} and weak ligase activity of Asn proteases has been reported in some examples.³⁹⁻⁴¹

FIGURES 1. Putative mechanism of AEP-mediated proteolysis and PAL-mediated ligation. The substrate protein of interest (POI-1) is bound to the active site via the positioning of specific residues in the cognate binding pockets S2-S1-S1'-S2'. The P1 site carbonyl carbon is attacked by a deprotonated sulfur atom of the catalytic Cys, which leads to peptide bond cleavage after P1-Asx and the formation of an S-acyl intermediate that is accessible to nucleophilic attack. The leaving group is replaced either by a water molecule preferred by AEPs, or an incoming peptide or a protein (POI-2) preferred by PALs, resulting in

1 either hydrolysis or aminolysis of the thioester bond, to yield either
2 a hydrolyzed product, or a peptide bond ligating two POIs.
3
4
5
6

7
8 A comparison between the α -carbon backbones of eight unique crystal
9 structures of AEP/PAL returns r.m.s.d. of less than 1.0 Å, following
10 superimposition, ruling out major structural changes between these two
11 types of enzymes (Supplementary Figure S1). We therefore speculated
12 that systematic sequence variations of residues immediately surrounding
13 the catalytic center could exert the most influence on the direction
14 of the enzymatic reaction. While sequence comparisons between PALs and
15 AEPs showed strict conservation of residues that shape the oxyanion
16 hole at the S1 site, we found substantial variations in the residues
17 lining the S2 and S1' substrate binding pockets (Figure 2). Thus, we
18 proposed the term "ligase-activity determinants" (LADs) to account for
19 the structure/function relation between patterns of variations found
20 within the S1-flanking sequences and the associated ligase activity.³⁰
21
22 LAD1 is a tripeptide motif shaping the S2 pocket. The first position
23 of LAD1 is often a conserved aromatic residue Trp or Tyr and the third
24 position is often a hydrophobic residue, such as Thr/Tyr/Val/Ala. The
25 middle residue (also named "Gatekeeper") in particular, displays
26 obvious differences between AEPs and butelase-1 like ligases, as it is
27 commonly a Gly residue in AEPs, while a Val/Ile/Cys is commonly found
28 in PALs. LAD2 is a dipeptide motif located near the S1' pocket and is
29 most commonly Gly-Pro in AEPs but Gly-Ala/Ala-Ala/Ala-Pro in PALs.
30
31 These S1-flanking sequence variations are hypothesized to be the primary
32 determinants of ligase activity presented by PALs.
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

	S2	S1'	S1
Plant AEP	236	166	59
Butelase-1	237	167	61
OaAEP1b	238	168	62
HeAEP3	W V T	G G A	N R H D H E C S D
VyPAL2	W C Y	G A A	N R H D H E C S D
PxAEP3b	W V T	G A P	N R H D H E C S D
VyPAL3	W I T	G A P	N R H D H E C S D
VcAEP	W V T	G G P	N R H D H E C S D
CeAEP	W V V	G Y A	N R H D H E C S D
AtLEGy	W V A	G Y P	N R H D H E C S D
HaAEP1	F G T	G G P	N R H D H E C S D
Butelase-2	W G T	G G P	N R H D H E C S D
	251	181	75
	252	182	77
	253	183	78
			179
			180
			220
			222
			250
			272
	LAD1	LAD2	Catalytic triad
	Variable	Conserved	

FIGURE 2. Amino acid composition of S2-S1-S1' pockets. Residues forming the S1 pocket are strictly conserved in both AEPs and PALs. In contrast, residues lining the S2 and S1' pockets are variable at the LAD1 and LAD2 motifs. Numbers on top and bottom of the sequence alignment refer to residues numbering in butelase-1 and butelase-2, respectively. Residues shaded in grey are proposed to be the sequence variants that influence enzyme activity most between hydrolysis and aminolysis. Residues highlighted in blue are proposed to favor ligase activity.

Here, we describe a success site-directed mutagenesis strategy in converting the potent AEP, butelase-2, into a ligase. Our experiments were guided by the LAD hypothesis. In particular, we focused on mutating the LAD residues in butelase-2 to the corresponding amino acids found in functionally validated PALs reported earlier, such as butelase-1, and butelase-1-like ligases such as VyPAL2, and OaAEP1b. Our results show that this is a successful strategy to produce an efficient recombinant peptide ligase from a protease. In turn, our work validates the critical roles played by LAD residues in controlling the enzyme directionality in AEPs and PALs and presents a potential strategy applicable to the engineering of most C13 Cys proteases.

RESULTS AND DISCUSSION

Design of Butelase-2 Mutants Guided by the LAD hypothesis

Butelase-2 is a potent protease and one of the eight butelase-1 homologs identified from the *C. ternatea* transcriptome.^{21, 42, 43} Based on the LAD hypothesis, we designed three types of PAL-like butelase-2 mutations: LAD1-mutants targeting the middle residue of LAD1 motif Gly252 by substituting it with Val, Ile or Ala to give Bu2a (V/GP), Bu2b (I/GP) and Bu2c (A/GP), respectively; LAD2-mutants targeting both residues in LAD2 motif, Gly182-Pro183 by substituting them with Ala, giving Bu2e (G/GA) (butelase-1-like) and Bu2d (G/AP) (VyPAL2-like). Finally, to study possible synergistic effects between these two sets of mutations, we targeted both LAD1 and LAD2 pockets and produced the following double-site-LAD-mutants: Bu2f (V/AP) (VyPAL2-like), Bu2g (V/GA) (butelase-1-like), and Bu2h (V/AA) (OaAEP1b-like) (Figure 3).

We also designed another sixteen Bu2a analogs with additional mutations targeting twelve non-LAD residues (Figure 3 and Supplementary Figure S2). They include catalytic residues Asn75, Cys222, putative auto-activation sites Asn345, Arg347 and Asn373, and seven residues that show sequence variations between PALs and butelase-2, which may affect enzyme stability. These include core-domain beta-sheet residues Ser178, Met279, and cap-core interacting residues Asn71, Glu74, Asn249, Thr267, and Thr268. In addition, nineteen single mutations on the twelve non-LAD residues alone were constructed to compare with Bu2a-analogs (Supplementary Figure S3a).

LAD Location	S2 (LAD1)	S1' (LAD2)	Bu2a analogs (G252V + non- LAD mutants)
Residue no.	252	182/183	
Wild-type (-/-)			N71K
Butelase-2	G	G P	N71A
LAD1 mutants (+/-)			E74V
Bu2a (V/GP) (Butelase-1-like)	V	G P	N75A
Bu2b (I/GP) (VyPAL2-like)	I	G P	S178A
Bu2c (A/GP) (Ala-substitution)	A	G P	S178T
LAD2 mutants (-/+)			T268S
Bu2d (G/GA) (Butelase-1-like)	G	A P	T267D
Bu2e (G/AP) (VyPAL2-like)	G	G A	T268A
Double-site LAD mutants (+/+)			T268V
Bu2f (V/AP) (VyPAL2-like)	V	A P	T267D/T268V
Bu2g (V/GA) (Butelase-1-like)	V	G A	N249A
Bu2h (V/AA) (OaAEP1b-like)	V	A A	M279L
			N345A
			R347L
			N373A

FIGURE 3. Design of LAD-mutants of butelase-2.

Recombinant expression of butelase-2 mutants

The baculovirus expression system was adopted for soluble expression of glycosylated butelase-2 mutants. The signal peptide Met1-Ala32 of butelase-2 was replaced with a signal sequence from GP64 of baculovirus, which leads to the secretory expression that releases the translated protein into sf9 cell culture media, followed by a His6-tag for immunoblotting and affinity purification (see Methods).

The initial expression level of wild-type butelase-2 zymogen was very low due to protein degradation during cultivation (Supplementary Figure S3b, lane B2). The decrease in pH of culture media from neutral to 5.8–6.2 after 72 h of cultivation was likely to account for the gradual auto-activation and degradation of the expressed enzymes.

We observed stable expression of all eight LAD-mutants Bu2a-h (supplementary Figure S3b) with purified yields of 10–18 mg per liter culture, except for Bu2c (A/GP) and Bu2d (G/GA) which gave 2 mg and 5 mg purified proenzyme per liter culture, respectively. Similarly, all sixteen Bu2a-analogs that carry a G252V point mutation in LAD1 resulted in a comparable yield as Bu2a. In non-LAD mutants, only two mutants,

N373A and N345A were successfully expressed. Asn373 of butelase-2 is homologous to Asn323 of human legumain, which is a confirmed auto-activation site located between the $\alpha 6$ and $\alpha 7$ helices of the cap domain. After auto-activation, the cap domain (including this mutated residue) is removed, yielding an active protein identical to the wild-type active butelase-2. Thus, the N373A butelase-2 mutant is named Bu2*. The Bu2* (N373A) zymogen was obtained with a yield of 10-15 mg per liter of insect cell culture. Likewise, the Ala-substitution of another putative auto-activation site Asn345 located in the linker region also gave similar expression yield of proenzymes as Bu2*. In contrast, other non-LAD single mutants were not stably expressed.

These results demonstrate that a single mutation in LAD1 is sufficient for stabilizing proenzymes by reducing the autolytic activity under mild acidic conditions. The successful expression of both Bu2* (N373A) and N345A suggests that the proenzyme of butelase-2 can be effectively stabilized by suppressing the cleavage and removal of the cap domain, which leads to auto-activation of the proteolytic core domain.

Acid-induced activation of butelase-2 mutants

To characterize the changes in activity, mutants were first activated from the zymogens by acidification at pH 5.0 at 4 °C overnight to remove the C-terminal cap and part of the N-terminus, followed by purification by size-exclusion chromatography using a buffer at pH 4.0 (Supplementary Figure S4). The activated mutants were sequenced using proteomic methods and showed that activation sites were heterogeneous on both the N- and C- termini of the core domain (Supplementary Figure S5 and Table S1). An N-linked glycosylation site at N322 was also identified in the insect

cell-expressed recombinant butelase-2 mutants (Supplementary Figure S6 and Table S2).

PAL-like LAD-mutants showed enhanced ligase activities

The ligase activity of each butelase-2 mutant was examined using a 16-residue peptide substrate GN14-HV 1 (GISTKSIPPISYRNHV, MW 1679 Da), where the C-terminal tripeptide NHV constitutes the recognition motif and the HV dipeptide is the leaving group. Cyclization would yield the 14-residue end-to-end cyclic peptide cGN14 2 (MW 1515 Da), whereas hydrolysis would give the linear hydrolyzed peptide product GN14 3 (MW 1533 Da) (Supplementary Figure S7). All reactions were performed at an enzyme: substrate molar ratio of 1: 500 at 37 °C, pH 6.8 for 10 min and results were examined by MALDI-TOF mass spectrometry.

Overall, all LAD-mutants predominantly display ligase activity and a depleted protease activity with $\geq 80\%$ cyclization, except Bu2c (A/GP) with $< 30\%$ cyclization (Figure 4a). Sixteen Bu2a-analogs displayed similar ligase activity as G252V alone with 61-87% cyclization suggesting the non-LAD residues had a lower impact compared to LAD residues for driving the enzyme directionality towards ligation (Figure 4b). In contrast, the two wild-type-like mutants Bu2* (N373A) and N345A were expressed and displayed predominantly hydrolytic activity. The rest of non-LAD mutants failed to be expressed and hence their activity data could not be acquired.

We also evaluated the change in hydrolysis and ligation activity of PAL-like mutants of butelase-2 in a broader range of pH. Figure 5a summarized the HPLC-quantitation of product yields obtained using wild-type-like mutant Bu2* and six LAD-mutants performed from pH ranging

from 4.5 to 8.0. Under the same experimental conditions, except for changes in pH, Bu2* display a predominant protease activity in all pH values tested; LAD1-mutant Bu2a and LAD2-mutant Bu2d showed similar activity profiles as the dual-functional ligase reported,^{29, 44} which displayed predominantly ligase activity at pH 6-8; LAD2-mutant Bu2e (G/AP) and the three double-site mutants showed similar activity profiles as PALs with largely suppressed hydrolase activity under acidic pH and predominant ligase activity at near-neutral pH. For the double-site PAL-like mutants, hydrolysis yield was significantly reduced from 65% to <5% at pH 7.0, and the isolated yield of cyclization product increased from <5% to >70%, suggesting that the use of LAD-site engineered butelase-2 as PAL-like ligase in physiological conditions is possible.

FIGURE 4. Comparison of cyclization ratios by butelase-2 mutants. (a) Comparison between LAD-mutants (blue dots) and non-LAD mutations (red triangles). (b) Comparison between bu2a and Bu2a-analogs (black diamonds). Peak areas were obtained from MALDI-TOF MS spectra (See details in supplementary figure S7). The cyclization ratios were calculated using product peak areas as $A_{\text{cyclic}} / (A_{\text{cyclic}} + A_{\text{linear}})$.

FIGURE 5. Ligase activity of PAL-like LAD-mutants of butelase-2. (a) Quantitative product analysis of LAD-mutant-mediated reactions under pH ranging from 4.5 to 8.0. The amounts of substrate GN14-HV 1, cyclic GN14 2 and linear GN14 3 were calculated based on peak areas at 220 nm in HPLC profiles. Reactions were performed under pH 4.5–8.0 in triplicate at 37 °C for 10 min with a fixed molar ratio of enzyme: GN14-HV 1: 500. (b) Example HPLC profiles of bu2a and bu2g.

Functional validation of PAL-like LAD-mutants for peptide cyclization and protein labeling

Among the double-site LAD-mutants, Bu2g (V/GA) (butelase1-like mutant) was the most efficient at pH 7.0 with a turnover rate of $k_{cat} = 1.2 \text{ s}^{-1}$ in the cyclization of GN14-HV, which is approximately half the value achieved by butelase-1 (Supplementary Figure S8). Therefore, we chose Bu2g to demonstrate the usefulness of PAL-like Bu2 mutants and compared

its cyclization and ligation activity with Bu2* (negative control) and butelase-1 (a well-established ligase standard). A model peptide, cyclotide kalata-B1 linear precursor kB1-GL, was chemically synthesized with the addition of a PAL-recognition GL sequence at its C-terminus for macrocyclization (Supplementary Figure S9). Bu2g was able to cyclize > 95% of kB1, a ratio comparable to butelase-1 (Supplementary Figure S10a). The negative control Bu2* gave approximately 20% cyclic kB1, which was higher than expected and could be in part attributed to the conformational assistance of oxidatively folded kB1.⁴⁵ A similar conformation-assisted ligation was also reported earlier in the cyclization of sunflower trypsin inhibitor SFTI-1 by Arabidopsis AEP.⁴¹

To show that Bu2g can be exploited for intermolecular ligation, we used a synthetic fluorescent peptide GLE(EDANS)RG to label the recombinantly-expressed anti-Her2 antigen DARPIn9_26⁴⁶ binder, having an additional spacer sequence "GSGS" followed by a PAL-recognition signal "NGL" at its C-terminus (see Methods). The reactions were performed in PBS (pH 7.4) in the presence of 5 mM β -mercaptoethanol. Fluorescence labeling of DARPIn-NGL to give DARPIn-EDANs was completed in 1 h and 3.5 h by butelase-1 and Bu2g, respectively (Supplementary Figure S10b and Table S3). In contrast, Bu2* (negative control) failed to label the recombinant DARPIn-NGL. We conclude that the engineered butelase-2 mutant Bu2g can be practically used for site-specific protein labeling and protein engineering.

Structural study of butelase-2 mutants

To better understand the impact of the introduced mutations on the enzyme structure, we obtained crystal structures of five zymogens,

including Bu2*, a LAD1-mutant Bu2a (V/GP), a Bu2a analog with the additional point mutation at E74V, and two double-site LAD-mutants Bu2g (V/GA) and Bu2h (V/AA) (Table S4). All structures were obtained with resolutions ranging from 1.35–2.64 Å. As expected, they displayed the canonical two-domain fold of AEPs, with the core and cap domains (residues 60 to 481, numbering according to butelase-2) clearly defined in the electron density map. Similar to other plant AEP structures,^{40, 47, 48} the zymogens formed homodimers via interactions between their cap domains (Supplementary Figure S11a). The catalytic domain (residues G61 to N340) is composed of a central β -sheet formed by six parallel strands surrounded by five α -helices, and the C-terminal cap domain (residues Q357 to A481) comprises seven α -helices. Residues L341 to H356 constitute the flexible linker region that connects the catalytic core domain to the cap domain. In all five butelase-2 mutants, D179 is modified to its cyclic form and presents as succinimide (Snn), in agreement with reports of other AEP structures (Supplementary Figure S11b). This Asp-to-Snn modification was previously observed in human legumain and was speculated to be a critical residue in the ligation reaction via a non-canonical mechanism.³⁸ In our studies, the catalytic Cys (C222) in the Bu2a (V/GP)+E74V crystal structure was found to be fully oxidized to C-SO₃ in the absence of reducing agent during purification. This proenzyme could not be auto-activated, suggesting that proteolytic auto-activation is mediated by the catalytic Cys. Thus the precise role of the Snn, which forms part of the oxyanion hole remains to be explored. The N322 N-linked glycosylation site that was revealed in LC-MS/MS sequencing was confirmed in all crystal structures,

where a residual N-acetylglucosamine moiety could be seen. The overall structures of both PAL-like mutants could be superimposed to Bu2*, with r.m.s.d values of 0.43 Å for Bu2a, 0.38 Å for Bu2g, and 0.3 Å for Bu2h, revealing an overall similar conformation.

We docked a substrate peptide ISYRNAGI into the substrate-binding pockets of Bu2*, Bu2a, Bu2g, Bu2h, and butelase-1 and compared the conformations of their LAD1 and LAD2 residues. A close-up view is shown in Figure 6a. Interestingly, the substitution of Gly-to-Val at position 252 of butelase-2 (via mutagenesis targeting LAD1 of the protease Bu2* leading to the ligase Bu2a) results in a shift of 0.8 Å of the backbone of this residue in the S2 pocket. Consequently, the conformation of the mutant enzyme becomes similar to the conformation found in the corresponding S2 pocket of butelase-1 and is in clear contrast with Bu2* (Figure 6b). This shift of the backbone at position 252 is also observed in the experimental 3D structures of mutants Bu2g and Bu2H, which also have predominant ligase activity at neutral pH. (Figure 5a and 6b). The rotamer adopted by Val252 in butelase-1 and also in the Bu2g/Bu2h mutants orients its side-chain such that it is now facing towards the backbone of the P2 residue of the peptide substrate and creates a bulge of hydrophobic character in the S2 (LAD1) pocket. The presence of valine at this position adds molecular volume, pushing its surface by 2.1 Å (Figure 6b). In contrast, in the primed sites, mutagenesis targeting LAD2 does not cause significant backbone changes but appears to increase the molecular surface hydrophobicity in PAL-like mutants (Supplementary Table S5).

Proposed Mechanism of LAD-governed Directionality

1 In 19 LAD1-mutants of butelase-2, changing the middle residue of the
2 S2 pocket (Gatekeeper) from a small to a bulkier amino acid (e.g., Gly
3 to Val and Ile) led to a significant increase in ligation activity. In
4 an attempt to find a structural explanation, we superimposed the Bu2*
5 and Bu2a structures with Arabidopsis AEP AtLEG γ in complex with a YVAD
6 protease inhibitor (Supplementary figure S12). This comparison shows
7 that the presence of the bulkier Val or Ile at the central position of
8 the LAD1 shifts the molecular surface (Figure 6b) towards the incoming
9 peptide substrate, and would lead to steric hindrance with the backbone
10 carbonyl group of the peptide P2 and also with the side chain of its
11 P3 residue. To confirm this, we modeled SYRN substrate acyl-enzyme
12 intermediate with Bu2*, Bu2g, and butelase-1 and subjected them to MD
13 simulation to observe changes in conformations of the substrate (see
14 Movie S1).

15 This small shift at position 252 leads to a rotation of the backbone
16 of 6° for Bu2g and 12° for butelase-1 when compared with Bu2*. This
17 corresponds to a slight displacement of the substrate backbone at the
18 P2 position (0.7 Å for Bu2g and 1.4 Å for Butelase-1) (Figure 7) This
19 rotation and displacement add in a subtle alteration of the orientation
20 of the P1 acyl group. As a result, this movement may prevent an attack
21 by a catalytic water molecule positioned near the catalytic histidine
22 on the other side of the catalytic Cys, and thus disfavors hydrolysis
23 and instead promote the attack of a peptide group bound to the primed
24 site. The simulation data also agreed with the observation that Bu2a
25 series of mutants were dual-functional since the LAD1-252V was not
26 completely superimposed with butelase-1.

On the prime site of the active Cys residue, changing G182 and P183 to Ala near the S1' pocket (LAD2) did not alter the protein conformation but increased local surface hydrophobicity. The first residue of LAD2 seems to play the most important role since Ala-substitution at the first position results in higher proenzyme stability and a more ligation-enhancing effect than in the second position. A bulky residue such as Tyr at the first position of LAD2 strongly alters the reaction towards hydrolysis in both VyPAL3 (Tyr-Ala) and VcAEP (Tyr-Pro), even when their LAD1 contain Ile/Val. Ala-substitution of Tyr in both cases could suppress hydrolysis significantly. These data suggest that LAD2 may play a more direct role in determining enzyme directionality: a PAL-type LAD2 possessing Ala-Pro, Gly-Ala, or Ala-Ala could retain the leaving group for a longer time, particularly the P1' residue until it becomes replaced by the P1'' residue of an incoming peptide. In AEPs possessing Gly-Pro or Tyr-Pro at LAD2, a faster departure of the leaving group and facilitated access of water would favor hydrolysis of the substrate. This speculation is also supported by the earlier finding that changing the LAD1 residue of OaAEP1b, C247 to Ala, increased the catalytic rate 160-fold without much noticeable influence on the enzyme directionality since its LAD2 was Ala-Ala.

We also observed a synergic effect between the two LADs, as the double-site LAD mutants of butelase-2 could be converted more completely into ligases, than single-site mutants. Several LAD1- and LAD2-mutants appear to be similar to some of the reported "dual-functional" plant AEPs, of which one of the LAD motifs in these enzymes does not follow the LAD hypothesis and are reported to display protease activity at

acidic pH, while they act both as protease and ligase at near-neutral pH (Supplemental Table S6).^{29, 40, 41, 44} Finally, we note that the double-site LAD-mutations of butelase-2 did not completely abolish hydrolysis, suggesting the existence of other variations yet to be uncovered between ligases and protease-type AEPs that also contribute to the enzyme directionality.

FIGURE 6. Structural analysis of Bu2 mutants. (a) The active site of butelase-2 mutants. LAD1, LAD2 motifs, and catalytic triad are represented as sticks and respectively colored in gray (Bu2*), cyan (Bu2a), purple (Bu2g), wheat (Bu2h), and red (butelase-1, 6DHI). The molecular surface of the catalytic core domain of Bu2* is represented (brown mesh). A model substrate peptide ISYRNGI was docked at the Bu2* binding site to show the location of the S2-S1-S1'-S2'-S3' binding pockets. (b) A backbone shift of 0.8 Å between G252 in Bu2* and V252 Bu2a is indicated in the left panel. (c) Mapping of plant AEP substrate-binding pockets S2-S1-S1' using butelase-2 amino acid composition. Interaction of the docked substrate to the Bu2* was mapped using Maestro Suite (Shrödinger).

FIGURE 7. Substrate acyl-enzyme conformational change due to the nature of the central LAD1 residue (gatekeeper). The substrate backbone displacement is the measure of the angle between the carbonyl C atom (involved in the acyl-enzyme) and the carbonyl of the P3 residue. This angle rotated by 12° yielding a shift of $1.4\text{\AA}$ in the case of butelase-1 (red sticks) and 6° , yielding a shift of $0.7\text{\AA}$ in the case of Bu2g (purple sticks).

CONCLUSIONS

In summary, we were able to successfully engineer efficient peptide ligases from an asparaginyl endopeptidase butelase-2. The results showed that residues flanking the S1 pocket are remarkable predictors of ligase activity, which largely validates the ligase-activity determinant (LAD) hypothesis. Suppression of hydrolytic activity and increase of ligase activity also enhanced the stability of the recombinant zymogens of butelase-2 against autolytic degradation during expression. The PAL-like mutants having a combination of mutations targeting both their S2 and S1' sites were able to mediate macrocyclization and ligation under neutral conditions. High-resolution crystal structures of the mutant zymogens showed subtle differences including backbone shift and increase of surface hydrophobicity only in the regions targeted by the mutations, confirming that rather subtle structural and hydrophobicity changes near the S1 pockets exert the highest influence on the reaction catalyzed. In conclusion, a possible mechanism is that the mutations at LAD sites with ligase-type amino acids lead to: (i) displacement of the substrate backbone, which stabilizes the S-acyl intermediate by modulating its orientation away

1 from the catalytic water molecule, (ii) increase of local hydrophobicity
2 to exclude water molecules, and (iii) increase of the retention time
3 of the peptide leaving group and to permit selectively an incoming
4 peptide to be ligated, via an increased "hydrophobic stickiness".
5 Overall, the present study could guide future engineering efforts
6 targeting the pool of widely available proteases to obtain versatile
7 peptide ligases for a wealth of exciting applications in biotechnology.
8
9
10
11
12
13
14
15
16
17
18

19 ASSOCIATED CONTENT

22 **Supporting Information**

23
24
25 Supplementary data includes materials and methods, data summary
26 tables, SDS-PAGE gel images, crystal structures, kinetic plots, MS
27 profiles, and MD simulation movie.
28
29
30
31

32
33 The following files are available free of charge.
34
35

36 Supporting information.pdf
37
38

39 Movie S1.mpg
40
41

42 AUTHOR INFORMATION

45 **Corresponding Author**

46
47 *jptam@ntu.edu.sg
48
49

50 *Julien@ntu.edu.sg
51
52

53 **Author Contributions**

The manuscript was written through the contributions of all authors.
All authors approved the final version of the manuscript. ‡These
authors contributed equally.

Funding Sources

This research was supported by the Ministry of Education, Singapore,
under its MOE AcRF Tier 3 Award MOE2016-T3-1-003, NMRC grant
CBRG/0028/2014 and NRF2016NRF-CRP001-063.

ACKNOWLEDGMENT

We thank the scientists and beamline staffs at the Proxima 2A, SOLEIL
(Saint Aubin, France) (Proposal 20180290), and MX2 (Australian
Synchrotron) for their expert assistance. We thank the National
Supercomputing Centre, Singapore (<https://www.nsc.sg>) for providing
ASPIRE-1 for the MD simulations.

ABBREVIATIONS

AEP, asparaginyl endopeptidase; PAL, peptide asparaginyl ligase; LAD,
ligase-activity determinant; HPLC, High-throughput liquid
chromatography; MS, Mass spectrometry; MALDI-TOF, Matrix-Assisted
Laser Desorption/Ionization-Time of Fly.

REFERENCES

1. Lipmann, F., Bacterial Production of Antibiotic Polypeptides by
Thiol-Linked Synthesis on Protein Templates. In *Adv. Microb. Physiol.*,
Rose, A. H.; Morris, J. G., Eds. Academic Press: 1981; Vol. 21, pp 227-
266.

1 2. Kopp, F.; Marahiel, M. A., Macrocyclization strategies in
2 polyketide and nonribosomal peptide biosynthesis. *Nat Prod Rep* 2007,
3 24 (4), 735-49.
4
5

6
7 3. Perler, F. B.; Davis, E. O.; Dean, G. E.; Gimble, F. S.; Jack,
8 W. E.; Neff, N.; Noren, C. J.; Thorner, J.; Belfort, M., Protein
9 splicing elements: inteins and exteins — a definition of terms and
10 recommended nomenclature. *Nucleic Acids Res.* 1994, 22 (7), 1125-1127.
11
12
13
14

15 4. Haase, J.; Lanka, E., A specific protease encoded by the
16 conjugative DNA transfer systems of IncP and Ti plasmids is essential
17 for pilus synthesis. *J. Bacteriol.* 1997, 179, 5728-5735.
18
19
20
21
22
23
24

25 5. Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O.,
26 *Staphylococcus aureus* Sortase, an Enzyme that Anchors Surface Proteins
27 to the Cell Wall. *Science* 1999, 285, 760-763.
28
29
30
31
32

33 6. Lee, J.; Schmidt, E. W.; McIntosh, J.; Hathaway, B. J., Using
34 marine natural products to discover a protease that catalyzes peptide
35 macrocyclization of diverse substrates. *Journal of the American*
36 *Chemical Society* 2009, 131 (6), 2122-2124.
37
38
39
40
41
42

43 7. Barber, C. J. S.; Pujara, P. T.; Reed, D. W.; Chiwocha, S.;
44 Zhang, H.; Covello, P. S., The Two-step Biosynthesis of Cyclic Peptides
45 from Linear Precursors in a Member of the Plant Family Caryophyllaceae
46 Involves Cyclization by a Serine Protease-like Enzyme. *J. Biol. Chem.*
47 2013, 288 (18), 12500-12510.
48
49
50
51
52
53
54

55 8. Luo, H.; Hong, S. Y.; Sgambelluri, R. M.; Angelos, E.; Li, X.;
56 Walton, J. D., Peptide Macrocyclization Catalyzed by a Prolyl
57
58
59
60

1 Oligopeptidase Involved in alpha-Amanitin Biosynthesis. Chem. Biol.
2 2014.
3
4

5 9. Canova-Davis, E.; Jeremy, K. T.; Ling, V. T., Transpeptidation
6 during the analytical proteolysis of proteins. Anal. Biochem. 1991, 196
7
8 (1), 39-45.
9
10
11

12 10. Zheng, W.; Johnston, S. A.; Joshua-Tor, L., The Unusual Active
13 Site of Gal6/Bleomycin Hydrolase Can Act as a Carboxypeptidase,
14
15 Aminopeptidase, and Peptide Ligase. Cell 1998, 93, 103-109.
16
17
18
19

20 11. Vigneron, N.; Stroobant, V.; Chapiro, J.; Ooms, A.; Degiovanni,
21 G.; Morel, S.; Van der Bruggen, P.; Boon, T.; Van den Eynde, B. J.,
22
23 An Antigenic Peptide Produced by Peptide Splicing in the Proteasome.
24
25 Science 2004, 304 (5670), 587-590.
26
27
28
29

30 12. Liepe, J.; Marino, F.; Sidney, J.; Jeko, A.; Bunting, D. E.;
31 Sette, A.; Kloetzel, P. M.; Stumpf, M. P.; Heck, A. J.; Mishto, M.,
32
33 A large fraction of HLA class I ligands are proteasome-generated spliced
34
35 peptides. Science 2016, 354 (6310), 354-358.
36
37
38
39

40 13. Christensen, U., Effects of pH on carboxypeptidase-Y-catalyzed
41 hydrolysis and aminolysis reactions. Eur. J. Biochem. 1994, 220, 149-
42
43 153.
44
45
46
47

48 14. Gaertner, H.; Puigserver, A., Kinetics and specificity of serine
49 proteases in peptide synthesis catalyzed in organic solvents. Eur. J.
50
51 Biochem. 1989, 181, 207-213.
52
53
54
55
56
57
58
59
60

15. Chen, Y.-X.; Zhang, X.-Z.; Zheng, K.; Chen, S.-M.; Wang, Q.-C.;
Wu, X.-X., Protease-catalyzed synthesis of precursor dipeptides or RGD
with reverse micelles. *Enzyme Microb. Techno.* 1998, 23, 243-248.

16. Abrahmsen, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff,
A.; Wells, J. A., Engineering subtilisin and its substrates for
efficient ligation of peptide bonds in aqueous solution. *Biochemistry*
1991, 30 (17), 4151-4159.

17. Wehofsky, N.; Koglin, N.; Thust, S.; Bordusa, F., Reverse
proteolysis promoted by in situ generated peptide ester fragments. *J.*
Am. Chem. Soc. 2003, 125 (20), 6126-6133.

18. Neet, K. E.; Koshland, D. E., The conversion of serine at the
active site of subtilisin to cysteine: a "chemical mutation". *Proc.*
Natl. Acad. Sci. U. S. A. 1966, 56 (5), 1606-1611.

19. Sorensen, S. B.; Raaschou-Nielsen, M.; Mortensen, U. H.;
Remington, S. J.; Breddam, K., Site-Directed Mutagenesis on (Serine)
Carboxypeptidase Y from Yeast. The Significance of Thr60 and Met398 in
Hydrolysis and Aminolysis Reactions. *J. Am. Chem. Soc.* 1995, 117, 5944-
5950.

20. Atwell, S.; Wells, J. A., Selection for improved subtiligases by
phage display. *Proc. Natl. Acad. Sci. U. S. A.* 1999, 96 (17), 9497-502.

21. Nguyen, G. K. T.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J.
P., Butelase 1 is an Asx-specific ligase enabling peptide
macrocyclization and synthesis. *Nat. Chem. Biol.* 2014, 10 (9), 732-738.

22. Harris, K. S.; Durek, T.; Kaas, Q.; Poth, A. G.; Gilding, E. K.; Conlan, B. F.; Saska, I.; Daly, N. L.; van der Weerden, N. L.; Craik, D. J.; Anderson, M. A., Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat. Commun.* 2015, 6, 10199.

23. Nguyen, G. K. T.; Kam, A.; Loo, S.; Jansson, A. E.; Pan, L. X.; Tam, J. P., Butelase 1: A Versatile Ligase for Peptide and Protein Macrocyclization. *J. Am. Chem. Soc. Comm.* 2015, 137 (49), 15398-15401.

24. Hemu, X.; Qiu, Y.; Nguyen, G. K. T.; Tam, J. P., Total Synthesis of Circular Bacteriocins by Butelase 1. *J. Am. Chem. Soc. Comm.* 2016, 138 (22), 6968-71.

25. Cao, Y.; Nguyen, G. K.; Chuah, S.; Tam, J. P.; Liu, C. F., Butelase-Mediated Ligation as an Efficient Bioconjugation Method for the Synthesis of Peptide Dendrimers. *Bioconjugate Chemistry* 2016, 27 (11), 2592-2596.

26. Bi, X.; Yin, J.; Nguyen, G. K. T.; Rao, C.; Halim, N. B. A.; Hemu, X.; Tam, J. P.; Liu, C. F., Enzymatic Engineering of Live Bacterial Cell Surfaces Using Butelase 1. *Angew. Chem. Int. Ed. Engl.* 2017, 56 (27), 7822-7825.

27. Harmand, T. J.; Bousbaine, D.; Chan, A.; Zhang, X.; Liu, D. R.; Tam, J. P.; Ploegh, H. L., One-Pot Dual Labeling of IgG 1 and Preparation of C-to-C Fusion Proteins Through a Combination of Sortase A and Butelase 1. *Bioconjugate Chemistry* 2018.

28. Harris, K. S.; Guarino, R. F.; Dissanayake, R. S.; Quimbar, P.; McCorkelle, O. C.; Poon, S.; Kaas, Q.; Durek, T.; Gilding, E. K.; Jackson, M. A.; Craik, D. J.; van der Weerden, N. L.; Anders, R. F.; Anderson, M. A., A suite of kinetically superior AEP ligases can cyclise an intrinsically disordered protein. *Sci. Rep.* 2019, 9 (1), 10820.

29. Jackson, M. A.; Gilding, E. K.; Shafee, T.; Harris, K. S.; Kaas, Q.; Poon, S.; Yap, K.; Jia, H.; Guarino, R.; Chan, L. Y.; Durek, T.; Anderson, M. A.; Craik, D. J., Molecular basis for the production of cyclic peptides by plant asparaginyl endopeptidases. *Nat. Commun.* 2018, 9 (1), 2411.

30. Hemu, X.; El Sahili, A.; Hu, S.; Wong, K.; Chen, Y.; Wong, Y. H.; Zhang, X.; Serra, A.; Goh, B. C.; Darwis, D. A.; Chen, M. W.; Sze, S. K.; Liu, C.-F.; Lescar, J.; Tam, J. P., Structural determinants for peptide-bond formation by asparaginyl ligases. *Proc. Natl. Acad. Sci. U. S. A.* 2019, 116 (24), 11737-11746.

31. Saether, O.; Craik, D.; Campbell, I.; Sletten, K.; Juul, J.; Norman, D., Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. *Biochemistry* 1995, 34, 4147 - 4158.

32. Simonsen, S. M.; Sando, L.; Ireland, D. C.; Colgrave, M. L.; Bharathi, R.; Goransson, U.; Craik, D. J., A continent of plant defense peptide diversity: cyclotides in Australian *Hybanthus* (Violaceae). *Plant Cell* 2005, 17 (11), 3176-89.

1 33. Rawlings, N. D.; Waller, M.; Barrett, A. J.; Bateman, A., MEROPS:
2 the database of proteolytic enzymes, their substrates and inhibitors.
3
4 Nucleic Acids Res. 2014, 42 (Database issue), D503-509.
5
6

7
8 34. Dall, E.; Brandstetter, H., Mechanistic and structural studies on
9
10 legumain explain its zymogenicity, distinct activation pathways and
11
12 regulation. Proc. Natl. Acad. Sci. U. S. A. 2013, 110 (27), 10940-
13
14 10945.
15
16

17
18 35. Hiraiwa, N.; Nishimura, M.; Hara-Nishimura, I., Vacuolar
19
20 processing enzyme is self-catalytically activated by sequential removal
21
22 of the C-terminal and N-terminal propeptides. FEBS Lett. 1999, 447 (2-
23
24 3), 213-216.
25
26

27
28 36. Okamoto, T.; Yuki, A.; Mitsuhashi, N.; Minamikawa, T., Asparaginyl
29
30 endopeptidase (VmPE-1) and autocatalytic processing synergistically
31
32 activate the vacuolar cysteine proteinase (SH-EP). Eur. J. Biochem.
33
34 1999, 264 (1), 223-32.
35
36

37
38 37. Zhao, L.; Hua, T.; Crowley, C.; Ru, H.; Ni, X.; Shaw, N.;
39
40 Jiao, L.; Ding, W.; Qu, L.; Hung, L. W.; Huang, W.; Liu, L.; Ye,
41
42 K.; Ouyang, S.; Cheng, G.; Liu, Z. J., Structural analysis of
43
44 asparaginyl endopeptidase reveals the activation mechanism and a
45
46 reversible intermediate maturation stage. Cell Res. 2014, 24 (3), 344-
47
48 58.
49
50

51
52 38. Dall, E.; Fegg, J. C.; Briza, P.; Brandstetter, H., Structure and
53
54 Mechanism of an Aspartimide-Dependent Peptide Ligase in Human Legumain.
55
56 Angew. Chem. Int. Ed. Engl. 2015, 54 (10), 2917-2921.
57
58
59
60

1 39. Bernath-Levin, K.; Nelson, C.; Elliott, A. G.; Jayasena, A. S.;
2 Millar, A. H.; Craik, D. J.; Mylne, J. S., Peptide macrocyclization by
3 a bifunctional endoprotease. *Chem. Biol.* 2015, 22 (5), 571-82.
4
5

6
7 40. Haywood, J.; Schmidberger, J. W.; James, A. M.; Nonis, S. G.;
8 Sukhoverkov, K. V.; Elias, M.; Bond, C. S.; Mylne, J. S., Structural
9 basis of ribosomal peptide macrocyclization in plants. *eLife* 2018, 7,
10 e32955.
11
12
13
14
15

16
17 41. Zauner, F. B.; Elsässer, B.; Dall, E.; Cabrele, C.; Brandstetter,
18 H., Structural analyses of Arabidopsis thaliana legumain γ reveal
19 differential recognition and processing of proteolysis and ligation
20 substrates. *J Biol Chem* 2018, 293 (23), 8934-8946.
21
22
23
24
25

26
27 42. Gilding, E. K.; Jackson, M. A.; Poth, A. G.; Henriques, S. T.;
28 Prentis, P. J.; Mahatmanto, T.; Craik, D. J., Gene coevolution and
29 regulation lock cyclic plant defence peptides to their targets. *New*
30 *Phytol.* 2016, 210 (2), 717-30.
31
32
33
34
35

36
37 43. Serra, A.; Hemu, X.; Nguyen, G. K.; Nguyen, N. T.; Sze, S. K.;
38 Tam, J. P., A high-throughput peptidomic strategy to decipher the
39 molecular diversity of cyclic cysteine-rich peptides. *Sci. Rep.* 2016,
40 6, 23005.
41
42
43
44
45

46
47 44. Bernath-Levin, K.; Nelson, C.; Elliott, Alysha G.; Jayasena,
48 Achala S.; Millar, A. H.; Craik, David J.; Mylne, Joshua S., Peptide
49 macrocyclization by a bifunctional endoprotease. *Chem. Biol.* 2015, 22
50 (5), 571-582.
51
52
53
54
55
56
57
58
59
60

45. Daly, N. L.; Love, S.; Paul F Alewood, a.; Craik, D. J., Chemical Synthesis and Folding Pathways of Large Cyclic Polypeptides: Studies of the Cystine Knot Polypeptide Kalata B1. *Biochemistry* 1999, 38 (32), 10606-10614.

46. Jost, C.; Schilling, J.; Tamaskovic, R.; Schwill, M.; Honegger, A.; Pluckthun, A., Structural basis for eliciting a cytotoxic effect in HER2-overexpressing cancer cells via binding to the extracellular domain of HER2. *Structure* 2013, 21 (11), 1979-91.

47. Yang, R.; Wong, Y. H.; Nguyen, G. K. T.; Tam, J. P.; Lescar, J.; Wu, B., Engineering a Catalytically Efficient Recombinant Protein Ligase. *J. Am. Chem. Soc.* 2017, 139 (15), 5351-5358.

48. Zauner, F. B.; Dall, E.; Regl, C.; Grassi, L.; Huber, C. G.; Cabrele, C.; Brandstetter, H., Crystal structure of plant legumain reveals a unique two-chain state with pH-dependent activity regulation. *Plant Cell* 2018, 30 (3), 686-699.

TOC

P1: Asn/Asp

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

**S-Acyl intermediate
Formation**

S2

S1'

S1

Ligase

Protease

Plant AEP
Butelase-1
OaAEP1b
HeAEP3
VyPAL2
PxAEP3b
VyPAL3
VcAEP
CeAEP
AtLEGy
HaAEP1
Butelase-2

236 237 238
W V T
W C Y
W V T
W I T
W V T
W V V
W V A
F G T
W G T
W G T
W G T
W G T
251 252 253

166 167 168
G G A
G A A
G A P
G A P
G G P
G Y A
G Y P
G G P
G G P
G G P
G G P
181 182 183

59 61 62 164 165 205 207 235 257
N RH D **H** E **C** S D
N RH D **H** E **C** S D
N RH D **H** E **C** S D
N RH D **H** E **C** S D
N RH D **H** E **C** S D
N RH D **H** E **C** S D
N RH D **H** E **C** S D
N RH D **H** E **C** S D
N RH D **H** E **C** S D
N RH D **H** E **C** S D
N RH D **H** E **C** S D
75 77 78 179 180 220 222 250 272
↖ ↗ ↗

LAD1

LAD2

Catalytic triad

Variable

Conserved

Page 35 of 40

ACS Catalysis

LAD Location	S2 (LAD1)	S1' (LAD2)	Bu2a analogs (G252V + non-LAD mutants)
Residue no.	252	182/183	
Wild-type (-/-)			N71K
Butelase-2	G	G P	N71A
			E74V
LAD1 mutants (+/-)			N75A
Bu2a (V/GP) (Butelase-1-like)	V	G P	S178A
Bu2b (I/GP) (VyPAL2-like)	I	G P	S178T
Bu2c (A/GP) (Ala-substitution)	A	G P	T268S
			T267D
LAD2 mutants (-/+)			T268A
Bu2d (G/GA) (Butelase-1-like)	G	A P	T268V
Bu2e (G/AP) (VyPAL2-like)	G	G A	T267D/T268V
			N249A
Double-site LAD mutants (+/+)			M279L
Bu2f (V/AP) (VyPAL2-like)	V	A P	N345A
Bu2g (V/GA) (Butelase-1-like)	V	G A	R347L
Bu2h (V/AA) (OaAEP1b-like)	V	A A	N373A

ACS Paragon Plus Environment

a)

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

% Cyclization

b)

ACS Paragon Plus Environment

ACS Catalysis

Page 36 of 40

% Cyclization

a) Ligase activity of PAL-like LAD mutants

b) Examples of HPLC profiles

c) Mapping of butelase-2 substrate binding sites

Bu2* (grey)

Bu2g (purple)

Butelase-1 (red)

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

Substrate backbone displacement

Bu2* → Butelase-1: 12°

Bu2* → Bu2g: 6°

ACS Paragon Plus Environment

**Asparaginyl
Endopeptidase
(AEP)**

ACS Catalysis
Mutation in LADs

**Peptide
Asparaginyl
Ligase (PAL)**

**Asn/Asp-specific
S1 pocket**

“Ligase-Activity Determinants”

**LAD1
in S2 pocket**

**LAD2
in S1' pocket**

Butelase-2

Butelase-2 LAD-mutant