



Consensus design and engineering of an efficient and high-yield peptide asparaginyl ligase for protein cyclization and ligation

Received for publication, November 24, 2022, and in revised form, February 6, 2023. Published, Papers in Press, February 9, 2023.

<https://doi.org/10.1016/j.jbc.2023.102997>

Xinya Hemu^{1,2} , Xiaohong Zhang¹, Hong Yi Chang^{1,3}, Jin En Poh¹, and James P. Tam^{1,*}

¹School of Biological Sciences, Nanyang Technological University, Singapore, Singapore; ²School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing, China; ³Department of Pharmacy, Singapore General Hospital, Singapore, Singapore

Reviewed by members of the JBC Editorial Board. Edited by Phillip A. Cole

Plant legumains are Asn/Asp-specific endopeptidases that have diverse functions in plants. Peptide asparaginyl ligases (PALs) are a special legumain subtype that primarily catalyze peptide bond formation rather than hydrolysis. PALs are versatile protein engineering tools but are rarely found in nature. To overcome this limitation, here we describe a two-step method to design and engineer a high-yield and efficient recombinant PAL based on commonly found asparaginyl endopeptidases. We first constructed a consensus sequence derived from 1500 plant legumains to design the evolutionarily stable legumain conLEG that could be produced in *E. coli* with 20-fold higher yield relative to that for natural legumains. We then applied the ligase-activity determinant hypothesis to exploit conserved residues in PAL substrate-binding pockets and convert conLEG into conPAL1–3. Functional studies showed that conLEG is primarily a hydrolase, whereas conPALs are ligases. Importantly, conPAL3 is a superefficient and broadly active PAL for protein cyclization and ligation.

Legumains are cysteine proteases that are widely distributed among plants and animals. These proteases were named after their discovery in legume seeds in the early 1990s (1). They have also been isolated from parasites (2) and mammals (3). Legumains belong to the same family of enzymes as vacuolar processing enzymes that were discovered in the late 1980s (4). Regardless of their origin, these enzymes share a similar protein fold and are classified in the C13 subfamily of Cys proteases (MEROPS, EC 3.4.22.34) (5). Functionally, legumains are asparaginyl endopeptidases (AEPs) that hydrolyze the peptide bond after an Asn/Asp(Asx) (Fig. 1). Mammalian legumains are lysosomal enzymes that play important roles in antigen presentation and cancer-related events (6). In contrast to animal genomes that have only one gene encoding a legumain, plant genomes contain multiple copies of legumains. The major functions of plant legumains are to regulate senescence processes (7), maturation of seed storage proteins (8), and processing of defensive proteins and peptides.

In addition to cutting peptides at specific Asx sites, legumains can also join peptides (9). An early example of this activity was observed in the maturation of the legume lectin, concanavalin A (conA), which involves a splicing (cut-and-join) mechanism. This protease-mediated peptide ligation is assisted by the correctly folded conformation of the conA precursor that brings two ligating termini in close proximity to facilitate a transpeptidation reaction (10). Recently, we identified a subtype of plant legumains that act as ligases rather than proteases. We termed these enzymes peptide asparaginyl ligases (PALs). Examples of PALs include the prototype butelase-1 from *Clitoria ternatea* (11), which has a catalytic efficiency of $1.3 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$, making it the fastest natural peptide ligase for peptide macrocyclization (12). Other PALs include OaAEP1b from *Oldenlandia affinis* (13, 14) and a dozen more from *Viola* plants (15–17). Compared with widely distributed AEPs in both primitive and higher plants, PALs in nature are much rarer, only accounting for 1% of sequenced plant legumains (18).

Butelase-1 and other PALs serve as bioprocessing enzymes of macrocyclic peptides such as cyclotides and trypsin inhibitors (11, 13, 15, 19–21). They recognize an Asx-Xaa-Yaa tripeptide motif in which Xaa represents any residue except Pro and Yaa represents hydrophobic residues (11, 15). PALs function as efficient, site-specific ligases that do not require ATP for activity. As such, they are invaluable for biochemical and pharmaceutical applications for protein engineering and site-specific modifications of proteins and live cells (22–28).

A challenge to the wider use of PALs in industrial applications is associated with their recombinant productions. Recombinant PAL proenzymes can be expressed in bacteria, but the expressed proteins are often present as misfolded proteins in inclusion bodies (11) and the yield of soluble proteins is fairly low (29). Although incorporation of large fusion tags such as maltose-binding protein can increase the yield of crude fusion proteins, they can also lead to heterogeneous activity of the enzymes after activation (30, 31). Secretory expression in insect cells could produce about 20 mg/L of VyPAL2, but this approach is limited by relatively high costs. Similarly, secretory expression of C-terminal truncated butelase-1 in yeast yielded 16 mg/L active enzyme (32), but the success of this method has

* For correspondence: James P. Tam, jptam@ntu.edu.sg.

Consensus design of high-yield peptide asparaginyl ligases

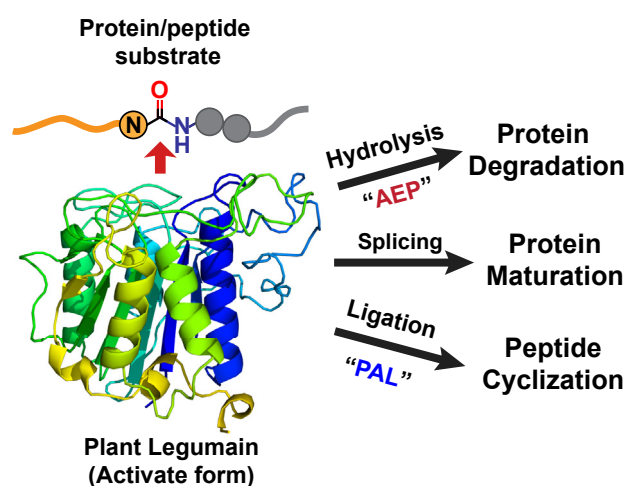


Figure 1. Schematic illustration of enzymatic activity and biological functions of plant legumain.

yet to be repeated for production of other PALs. Attempts to improve expression levels using disulfide-promoting strains, codon modification, truncation of the flexible N-terminal proregion, and incorporation of fusion tags suggest that expression of soluble PALs could be facilitated by increasing the stability of cDNA, mRNA, or peptide chains (28). By using residues that appear most frequently among homologous genes, an engineered functional protein could acquire enhanced stability, or activity, or both (33–37).

Here, we report a strategy to design high-yield PALs. We developed this strategy using bioinformatic analysis of a dataset containing 1500 plant legumains to build a consensus legumain followed by mutation of substrate-binding pocket residues based on the ligase-activity determinants (LAD) hypothesis that was previously developed to convert an AEP to a PAL.

Results

Consensus plant legumain sequence for recombinant protein expression

As recently reported, a dataset containing 1500 plant legumains from 249 plant families was constructed based on BLAST searches of existing protein sequences and transcriptomes in NCBI and 10KP databases, respectively, using representative AEPs and PALs as queries (18). Multiple sequence alignment of these legumains produced a consensus sequence having a typical protein fold of C13 Cys proteases, including an $\alpha 5\beta 6$ core domain carrying the “Asn-His-Cys” catalytic triad and a five-helix cap domain (38) (Fig. 2).

The expression construct consensus plant legumain (conLEG) was designed by removing residues with <10% occupancy in the 1500 plant legumains, substituting the signal peptide and the nonconserved region of the prodomain with a His6 affinity tag, and substituting the C-terminal unstructured loop after helix $\alpha 10$ -Pro457 (corresponding to butelase-1 Pro469) with a C-terminal His6-tag (Figs. 3 and S1A).

Recombinant protein was expressed from the conLEG construct as a proenzyme in *E. coli* Shuffle-T7B. After 48 h induction, 36 to 40 mg/L purified protein was obtained (Fig. S1B). Acid-induced autoactivation of conLEG was carried out under different pHs and temperatures. The optimized conditions for autoactivation were acidification of a 1 mg/ml proenzyme solution to pH 4.0 and incubation for 2 h at 37 °C (Fig. S1C). After size exclusion chromatography purification, a final yield of 12 mg active enzyme per 1 L culture was obtained.

conLEG functions mainly as a protease but has cyclase and splicing activity toward some substrates

To test the innate enzymatic ability of a plant legumain to catalyze hydrolysis, splicing, and ligation, we carried out a functional study of activated conLEG toward four synthetic peptide substrates, including a pair of short peptides GN10-GL (GISYKPAYLN \underline{N} GL, MW 1295 Da) and GD10-GL (GISYKPAYLD \underline{N} GL, MW 1296 Da), the cyclotide-mimicking GN14-SLAN (GISTKSIPPISYRN \underline{N} SLAN, MW 1918 Da), and sunflower trypsin inhibitor (SFTI) (D/N)-HV (GRCTKSIP- \underline{P} ICFPN \underline{H} V, MW 1768 Da) that mimics the precursor of the cyclic trypsin inhibitor SFTI-1 (Fig. 4). Reactions were performed with the same substrate concentration of 20 μ M at eight reaction pHs ranging from 4.5 to 8.0 at 25 °C. Since the catalytic rate was substrate dependent, different enzyme:substrate (E:S) ratios and reaction times were set for each substrate to prevent overreaction. Overall, GN10-GL is the most favored substrate of conLEG with an E:S ratio of 1:2000 and 10 min reaction time, and GD10-GL is the least favored substrate with an E:S ratio of 1:200 and 30 min reaction time. SFTI(D/N)-HV and GN14-SLAN were reacted for 10 min with an E:S ratio of 1:500.

At acidic pH, conLEG acts as an AEP and hydrolyzed all four peptide substrates as monitored by HPLC. At neutral and slightly basic pH, conLEG displayed low ligase activity (Fig. 4A). For example, at pH 5, >60% of the 12-residue peptide GN10-GL was hydrolyzed to the 10-residue GN10 (GISYKPAYLN, MW, 1125 Da) with <3% cGN10 (cyclo-GISYKPAYLN, MW, 1107 Da), resulting in a very low cyclization/hydrolysis (C/H) ratio of 0.04 (Fig. 4B). In contrast, at pH 7.5, 19% of GN10-GL was hydrolyzed and 26% was cyclized, yielding a C/H ratio of 1.4. Similar product distribution profiles in which the C/H ratio increased with pH were obtained for the other three substrates (Fig. 4, C–E).

We also compared the catalytic effect of conLEG on two different peptide substrates: GN10-GL and GD10-GL, which have an Asn and Asp at the P1 recognition site and P1 position, respectively. P1-Asp resulted in higher C/H ratios at all eight pH values tested (Fig. 4C). The generality of cyclic product formation in conLEG-mediated reactions suggest that plant legumains function intrinsically as AEPs under acidic pH conditions but the activity can also be bidirectional and influenced by extrinsic factors such as pH as well as by substrate and sequence.

Consensus design of high-yield peptide asparaginyl ligases



Figure 2. Design of the consensus plant legumain sequence based on multiple sequence alignment and analysis of 1500 plant legumains. The secondary structures, protein domains, and functional features were annotated based on crystal structure of AtLEG β (Protein Data Bank code: 5NUJ).

To determine whether conLEG can act as a splicing enzyme when Asp instead of Asn is present at the P1 position, we synthesized a model substrate AINGLRRGYSGS-DALEG (1735 Da) containing both Asn and Asp to mimic the precursor of *Momordica cochinchinensis* trypsin inhibitor (McoTI)-II (21, 39). At pH 5.0, we observed formation of the Asn-cleaved intermediate GLRRGYSGSD-ALEG (1437 Da) and a cyclic product cyclo-(GLRRGYSGSD) (1049 Da), suggesting that conLEG could catalyze both Asn-

specific hydrolysis and Asp-specific transpeptidation under acidic pH (Fig. 5A). A similar trimodal enzymatic activity was reported for McPAL1, the bioprocessing enzyme that mediates cyclization of McoTI-II at Asp *via* a splicing mechanism (21).

To show conLEG splicing activity toward proteins, we recombinantly expressed a protein substrate, concanavalin A precursor (pro-conA), which was recombinantly expressed, folded by dialysis, and then matured by conLEG-mediated

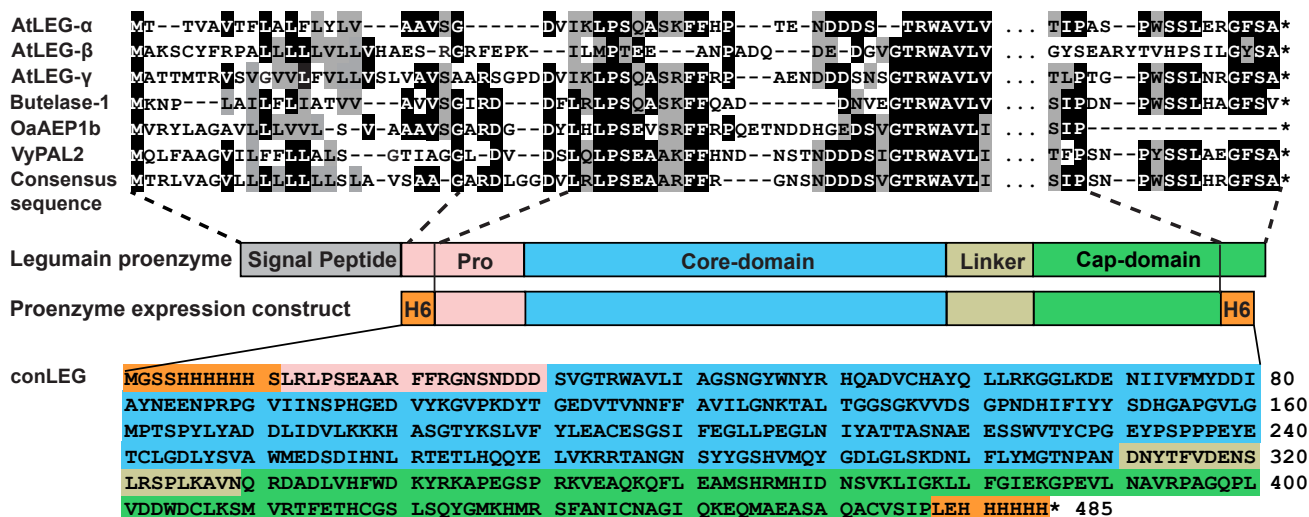


Figure 3. Construction of conLEG sequence for recombinant expression. The construct encoded full-length proenzyme derived from the consensus sequence of 1500 plant legumains, with a His6-tag attached to the N terminus of the conserved region of the prodomain.

Consensus design of high-yield peptide asparaginyl ligases

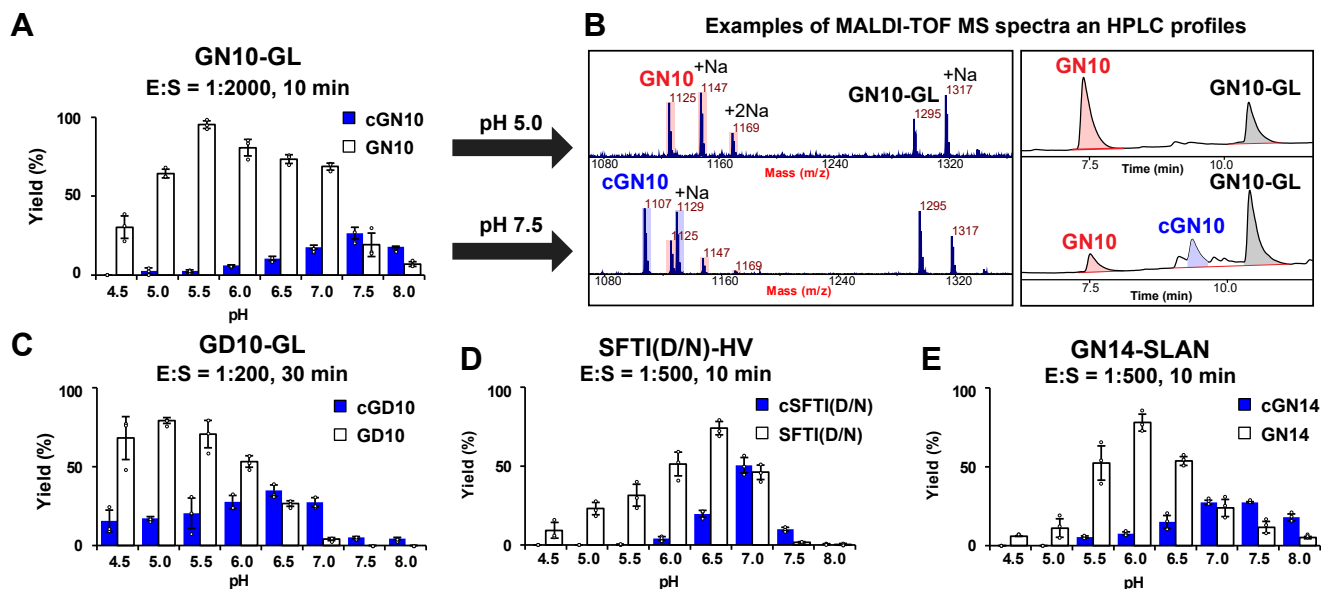
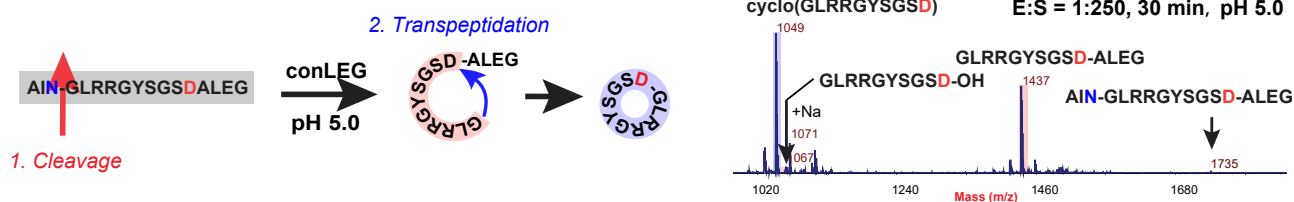


Figure 4. conLEG-mediated hydrolysis and cyclization of peptide substrates. Quantitative product distribution of (A) GN10-GL and its representative examples of mass spectrometry and HPLC data at pH 5 and pH 7.5 in (B), (C) GD10-GL, (D) SFTI(D/N)-HV, and (E) GN14-SLAN. Reactions were carried out at 25 °C in triplicate and monitored by MALDI-TOF mass spectrometry. Quantified data by reverse-phase HPLC are presented in column plots. E:S ratio and reaction time are indicated in each plot.

processing (Fig. S2). Monitoring of conLEG splicing at different time points over 2 h using SDS-PAGE showed that folded pro-conA (28.2 kDa) was rapidly cleaved into smaller fragments within 5 min followed by gradual accumulation of mature conA (25.6 kDa) over time (Fig. 5B). These results were

similar to those reported for conA maturation with its native bioprocessing enzyme CeAEP1 (10). Of note, the *in vitro* splicing reaction or transpeptidation reaction was slow and took hours to reach completion that accompanied with high hydrolysis.

A Substrate(sequence)-guided peptide splicing



B Substrate(conformation)-guided protein splicing

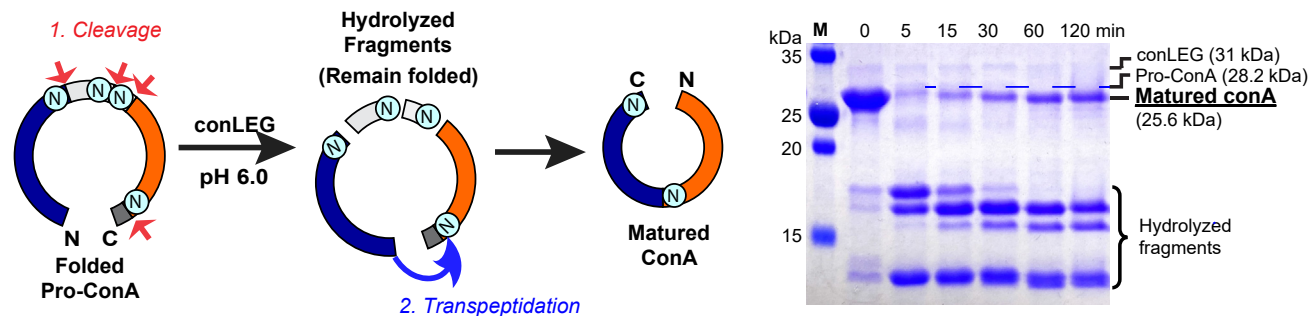


Figure 5. conLEG-mediated Asn/Asp-specific cleavage and transpeptidation. A, conLEG-mediated N-terminal Asn-specific hydrolysis and C-terminal Asp-specific transpeptidation of AINGLRRGYSGSDSLEG to produce cyclo(GLRRGYSGSD). The reaction was carried out at pH 5.0 in an E:S ratio of 1:250 at 37 °C for 30 min to facilitate release of the N-terminal tripeptide cap AIN. B, conLEG-mediated maturation of conA by Asn-specific hydrolysis followed by structural-guided transpeptidation.

Consensus design of high-yield peptide asparaginyl ligases

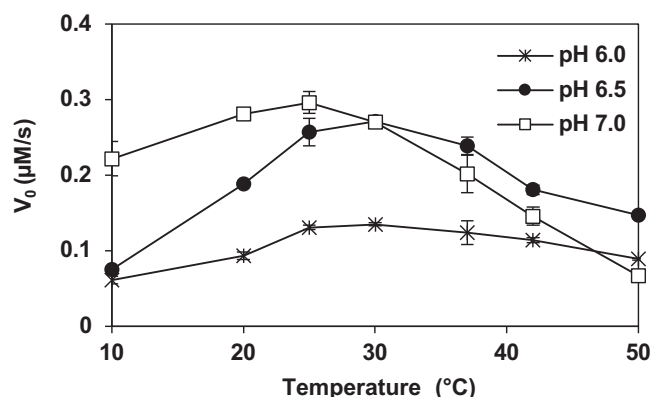


Figure 7. Optimal pH and temperature of conPAL3-mediated GN10-GL cyclization. Reactions were performed in triplicates at seven different temperatures and three different pH values with a 1:10,000 E:S ratio. To calculate the initial catalytic rates, aliquots of the reaction mixtures were collected and quenched every 30 s for quantitative analysis by reverse-phase HPLC.

the substrate specificity screening, a model substrate GFSYKPAYSN-GI (MW 1303.44 Da) was designed and synthesized. At the optimum pH 7.0, catalytic efficiency of conPAL3 toward this model substrate was highly efficient at $2,289,260 \text{ M}^{-1}\text{s}^{-1}$, which is nearly 2-fold faster than the reported efficiency for butelase-1 (Table 1 and Fig. S5C).

The tolerance of conPAL3 for nonaqueous solvents was examined by activity tests performed with various concentrations of polar protic solvents, polar aprotic solvents, and surfactants. conPAL3 retained 90% activity in the presence of 50% acetonitrile, 40% MeOH, 30% EtOH, 10% iPrOH, 30% dimethyl sulfoxide, 10% dimethyl formamide, and 2% Tween-20 but was rapidly inactivated by 0.1% SDS (Table 1C and Fig. S6).

Table 1
Enzymatic and chemical properties of conPAL3

A. Substrate specificity				
Pocket	Site	Substrate	Preference	
S1	P1	GISYKPAYLNGL GISYKPAYLDGL	N>>D	
S2	P2	GISTKSIPPISYXNGI	Any AA except P	
S1'	P1'	GLYRRGRLYRRNXL	Any AA	
	P1''	XLRYRRLYRLNGI XRLYRGRLYRRNHV	Favor G	
S2'	P2'	GLYRRGRLYRRNGX	Any AA	
	P2''	GXLRYRGRLYRRNHV RXARRLYRLNGI	Favor F, I, M, L	
B. Enzyme kinetics				
Substrate	pH	k _{cat} (s ⁻¹)	K _M (μM)	k _{cat} /K _M (M ⁻¹ s ⁻¹)
GISYKPAYLN-GL	7.0	16.52	31.89	518,031
GISYKPAYLD-GL	5.0	0.32	49.54	6431
GFSYKPAYSN-GI	7.0	13.28	5.80	2,289,260
C. Tolerance against nonaqueous solvents and adducts				
Solvent				Tolerance (%)
Polar protic (alcohol)	Methanol			40.0
	Ethanol			30.0
	<i>iso</i> -Propanol			10.0
Polar aprotic	Acetonitrile			50.0
	Dimethyl sulfoxide			20.0
	Dimethyl formamide			10.0
Surfactant	Tween-20			2.0
	SDS			0.1

X, 20 natural amino acids. Letter in bold and underlined, site with a differential residue.

Discussion

In this study, we combined a consensus AEP sequence found in 1500 legumains with conserved LADs found in the substrate-binding sites of a small group of known PALs to generate new PALs having significantly improved biochemical properties. The direct use of a consensus sequence derived from a large set of plant legumain sequences reduced the phylogenetic bias. Compared with natural legumains expressed in the most economically efficient bacterial system, we observed a 20-fold increase in conLEG expression, from an average of 2 mg/L proenzyme to about 40 mg/L. This increase may be due in part to improved mRNA stability and peptide solubility that facilitates translation and folding processes (41). Since the discovery of butelase-1 in 2014, we and others have unambiguously demonstrated PALs for cyclization of peptides and proteins, site-specific modifications of proteins and cell surfaces (Fig. 8) (22–28, 42, 43). Ligation at Asx is a challenge in peptide synthesis due to the unstable Asx-ester/thioester intermediates that are prone to the formation of aspartimide or succinic anhydride (44). Asx-specific PALs fill this role with high efficiency. An interesting outcome of this work is that the new PALs are tolerant to some organic reaction conditions that render them suitable to prepare dendrimers, polymers, and peptide–organic molecule conjugates as probes and for drug development. This high-yield and efficient conPAL3 could serve as a new model for further engineering and applications of PALs.

The designed conLEG shares 64 to 88% core-domain identity with 20 previously described AEPs and PALs (Fig. S7). The consensus design described here also improved the activity of plant legumains. The double-mutant conPAL3 carrying LAD mutations on both sides (S2, and S1'-S2') of the S1 Asx-binding pocket exhibited similar catalytic efficiency toward a model peptide as the fastest natural PAL, butelase-1. In addition, conPAL3 has broad substrate specificity toward P1-flanking sequences and optimal ligase activity at neutral pH and ambient temperature, as well as a wide tolerance to various nonaqueous solvents. We also observed that the reaction pH affects the thermal stability of conLEG, suggesting that the stability of its active conformation is closely associated with intramolecular charge–charge interactions. Together, our results demonstrate that a consensus design based on a large sequence dataset is a feasible approach to provide improved functional proteins.

Protease activity was the first identified and most frequently observed property of legumains. As such, the term AEP is often used interchangeably with legumain as a representative of the C13 subfamily. Our work used an “averaged” sequence to show that plant legumains primarily act as hydrolases, particularly under acidic conditions, but they can be conditionally bidirectional in rare cases, particularly serving as bioprocessors of certain peptides and proteins as well as during seed germination of seed storage proteins. Our results show that the activity of plant legumain is influenced by environmental factors such as pH, substrate composition (such as P1-Asp), and conformation (such as folded conA) or by intrinsic factors such as LAD mutations that led to the

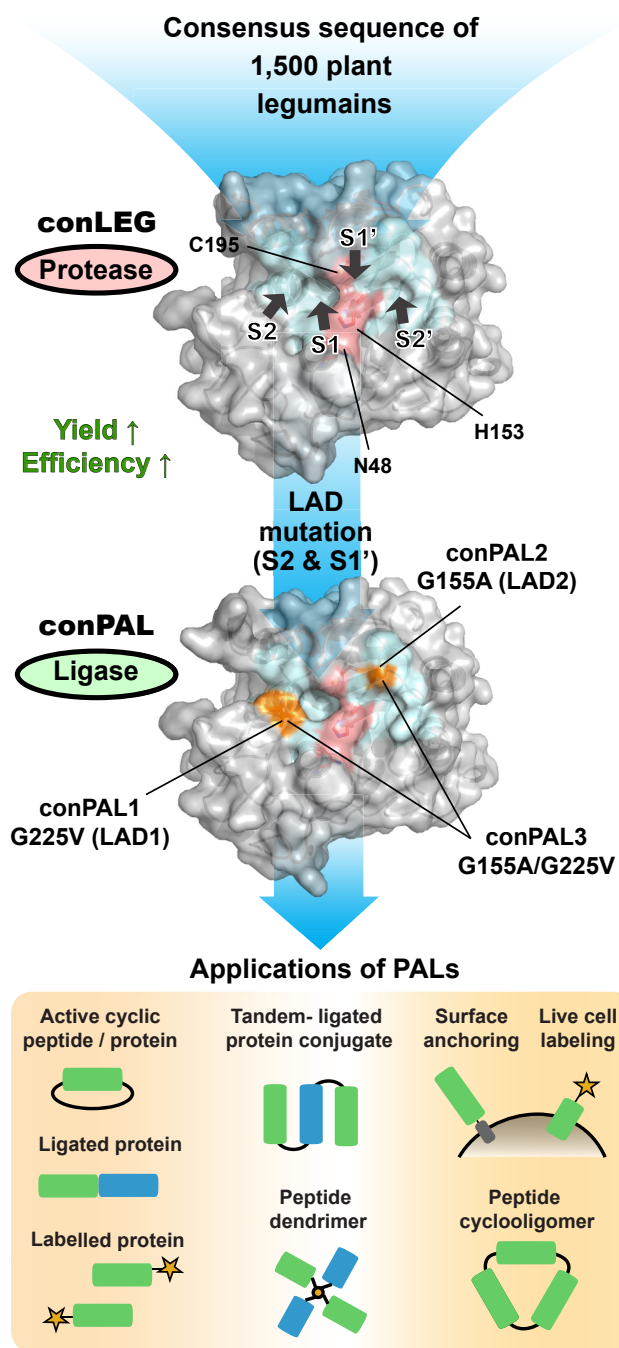


Figure 8. Summary of consensus design and ligase-activity determinant (LAD)-based engineering of high-yield and efficient peptide asparaginyl ligases (PALs) as versatile tools for peptide and protein engineering.

emergence of PALs. Most proteolytic enzymes appear to be unidirectional in nature, likely due to the lack of a ligation-promoting environment or the absence of a coevolved substrate for ligation. We speculate that both intrinsic and extrinsic factors contribute to the ligase activity of AEPs and PALs *via* a general mechanism: a favored coordination of substrates in the catalytic center that increases the thioester intermediate stability and inhibits the access of catalytic water by retaining the prime-side leaving groups in pockets (15, 21, 40). This mechanism could be applied to explain other

protease-based engineering of ligases, such as the subtilisin-BPN'-derived subtiligase and its kinds (45–47), and trypsin-derived tryptiligase that specifically recognizes a YRH motif in its substrates (48). Based on comparison of experimental results, we conclude that the selected intrinsic and extrinsic factors could exert profound influence on the enzymatic directionality toward AEPs or PALs, which may enlighten the design of other protease-derived ligases.

Experimental Procedures

Generation of plant legumain consensus sequence

Full amino acid sequences of butelase-1, OaAEP1b, AtLEG α - γ , and butelase-2 were used as representative plant PALs and AEPs to search homologous sequences in NCBI (BLASTp—nonredundant protein [nr] database, tBLASTx—nr and Transcriptome Shotgun Assembly database) and 10KP database. The search yielded 1500 nonduplicated hits as reported recently (18). Multiple sequence alignment was performed using the ClustalOmega program and analyzed by Jalview to generate visualization data including sequence Logo, consensus sequence, conservation, and occupancy. Regions with <10% occupancy were removed from the consensus sequence.

Plasmid construction and site-directed mutagenesis

The conLEG cDNA sequence was codon-optimized, synthesized, and subcloned into the pET28a(+) vector with restriction cleavage by NdeI/XhoI to carry both N- and C-terminal His6 tags (GenScript). Primers carrying a G155A or G225V mutation were designed for site-directed mutagenesis using a Q5 Site-Directed Mutagenesis Kit (New England Biolabs) to generate conPAL1–3 plasmids.

Recombinant expression of conLEG and mutants

Plasmids were transformed into SHuffle T7 *E. coli* competent cells (NEB, C3029). LB broth (kanamycin+) cultures (1 L) were grown at 30 °C with shaking at 180 rpm until the A_{600} reached 0.4. The cultures were cooled to 16 °C before isopropyl β -D-1-thiogalactopyranoside (IPTG, 0.1 mM) was added, and the cultures were further incubated for 24 to 48 h. Cell pellets were harvested by centrifugation at 6000g for 15 min at 4 °C and resuspended in cold lysis buffer (50 mM Na HEPES, 0.1 M NaCl, 1 mM EDTA, 5 mM β -mercaptoethanol [β -ME], 0.1% TritonX-100, pH 7.5) in a ratio of 20 ml per 1 g pellet. Cells were lysed by sonication (10 min for 50 ml, 2s on/8s off) on ice. Clarified cell lysates were obtained by centrifugation at 12,000g for 15 min and subjected to affinity purification on a 5-ml Excel HisTrap affinity purification column (GE Life Sciences) equilibrated with IMAC-A buffer (20 mM sodium phosphate buffer pH 7.5, 0.1 M NaCl, 5% glycerol, 1 mM EDTA, 5 mM β -ME). The column was then washed with a mixture of IMAC-A and IMAC-B (0.5 M imidazole in buffer IMAC-A, pH 7.5) buffer containing 0.25 M imidazole (95/5 v/v, IMAC-A/IMAC-B). Targeted proteins were eluted with IMAC-B buffer containing 0.5 M imidazole. Fractions containing the targeted proenzymes were further purified on a

Consensus design of high-yield peptide asparaginyl ligases

size exclusion chromatography column (HiLoad 16/600 Superdex 200, Cytiva) equilibrated with SEC-7.5 buffer (20 mM sodium phosphate buffer pH 7.5, 0.1 M NaCl, 5% glycerol, 1 mM EDTA, 5 mM β -ME) to remove imidazole.

Activation of conLEG and mutants

Acid-induced enzyme activation was performed at pH 3.5 to 6.0. Optimal activation was observed with pH 4.0 to 4.2 and incubation for 2 to 2.5 h at 37 °C with approximately 1 mg/ml proenzyme in 20 mM sodium citrate buffer containing 5 mM β -ME and 1 mM EDTA. Activated enzymes were purified by size exclusion chromatography equilibrated with SEC-4 buffer (20 mM sodium citrate buffer, 5 mM β -ME, 5% glycerol, 0.1 M NaCl, pH 4.0). Fractions containing target proteins were neutralized with storage buffer (20 mM sodium citrate, 5 mM β -ME, 1 mM EDTA, 20% sucrose, 0.1 M NaCl, pH 5.5) and kept at 4 °C or stored at -80 °C after rapid freezing in liquid nitrogen.

Preparation of peptide substrates

Peptide substrates (GN10-GL, GD10-GL, GN14-SLAN, SFTI(D/N)-HV, GFSYKPAYSN-GI) were synthesized by Fmoc chemistry using an automated microwave synthesizer (Liberty Blue, CEM) and purified by preparative reverse-phase HPLC (LC-20A, Shimadzu). Peptide substrate libraries (GISTKSIP-PISYXNGI, XLYRRLYRLNGI, RXARRLYRLNGI, GLYRRGR-LYRRNXL, XRLYGRGLYRRN-HV, GLYRRGRGLYRRNGX, GLYRRGLYRRNHV, X = 20 natural amino acids) were purchased from GL Biochem.

Preparation of recombinant pro-conA and conLEG-mediated maturation of conA

cDNA encoding the His6-TEV-pro-conA sequence was synthesized and subcloned into pET28a(+) in frame with the N-terminal His-tag (GenScript), followed by transformation into Rosetta PLYS competent cells (Novagen). The overnight culture was amplified from 1 ml to 1 L with LB broth at 37 °C with shaking at 180 rpm until the A_{600} reached 0.5. Recombinant expression of pro-conA was induced by addition of 0.5 mM IPTG and incubation for 24 h at 16 °C with shaking at 180 rpm. The cell pellet was harvested by centrifugation at 6000g for 10 min at 4 °C. The pellets were resuspended in 10 \times w/v conA lysis buffer (20 mM Mops, 0.5 M NaCl, 1 mM CaCl₂, 1 mM MnCl₂, 0.1% Triton X-100, pH 7). After 5 min sonication on ice, inclusion bodies were collected by centrifugation at 12,000g for 15 min at 4 °C. The pellets were washed twice with conA buffer (conA lysis buffer without Triton X-100). Purified inclusion bodies were resuspended in 40 ml refolding solution (conA buffer with 6 M guanidine-HCl, pH 7) followed by 1 min sonication. The soluble portion was isolated by centrifugation at 12,000g for 30 min at 4 °C. Recombinant protein was refolded by gradual dilution with 30 \times conA buffer followed by purification using size exclusion chromatography. To remove the N-terminal tag, TEV protease was mixed with recombinant His6-TEV-pro-conA at a 1:10 w/w ratio and incubated at 25 °C for 4 h or 4 °C overnight. Pro-conA was purified from the TEV protease by reverse IMAC. For conLEG-mediated circular permutation,

pro-conA was buffer-exchanged with legumain reaction buffer (20 mM sodium phosphate, 5 mM β -ME, 1 mM EDTA, pH 6.0) using a centrifugal concentrator (Vivaspin Turbo 15, Sartorius) and concentrated to a final concentration of ~1 mg/ml (calculated molar concentration 30 μ M based on 280 nm absorbance). Active conLEG was added to the conA solution to reach a molar ratio of conLEG:pro-conA = 1:100. Reactions were performed at pH 6.0 at 25 °C. Reaction mixtures were analyzed by SDS-PAGE at 0, 5, 15, 30, 60, and 120 min.

Functional studies of active conLEG and conPAL1-3 mutants

Functional studies of conLEG and mutants were carried out with synthetic peptide substrates and peptide libraries. The concentration of peptide substrates was fixed at 20 μ M unless specifically mentioned. All reactions were monitored by MALDI-TOF mass spectrometry (5800 TOF/TOF, Applied Biosystem) and if needed, quantitatively analyzed by reverse-phase HPLC on a C18 analytical column (Aeris WIDEPOR, Phenomenex) after quenching with 1:1 v/v acetonitrile containing 0.1% trifluoroacetic acid.

Reactions in Figure 4 were performed with conLEG at 25 °C at eight pHs ranging from 4.5 to 8.0 in 20 mM citrate or phosphate buffers with 1 mM EDTA and 5 mM β -ME. Since the reaction rate was substrate dependent, different amounts of conLEG were added to give a specific E:S ratio ranging from 1:200 to 1:2000. Reaction times were also set differently as indicated in the figure legends and plots. Reactions in Figure 5A were performed at 37 °C at pH 5.0 for 30 min in an E:S ratio of 1:250. Reactions in Figure 6B were performed with conPAL1-3 in an E:S ratio of 1:1000 at 25 °C in eight pHs ranging from 4.5 to 8.0 for 10, 15, and 5 min, respectively. Substrate specificity study in Fig. S4 was performed in different E:S ratios ranging from 1:100 to 1:2000 at 25 °C for 5 to 30 min at pH 7.0. Detailed conditions for each set of substrates were given in the data plot and figure legend. Kinetic study of conPAL3 in Fig. S5 was performed at 25 °C with 10 nM enzyme in a reaction system of 500 μ l at pH 5.0 for GD10-GL and pH 7.0 for the other two Asn-containing substrates, GN10-GL and GFSYKPAYSN-GI. The concentration of peptide substrates tested ranged from 1 to 100 μ M. Every 30 s, 100 μ l reaction mixtures were quenched and quantitatively analyzed. Reactions for Fig. S6 were performed in triplicates with GN10-GL in nonaqueous solvents topped up with a pH 7.0 reaction buffer (20 mM sodium phosphate, 1 mM EDTA, 5 mM β -ME) at 25 °C with an E:S ratio of 1:1000. The reaction progresses were monitored by MALDI-TOF mass spectrometry analysis at 4 min, 10 min, 30 min, and 1 h.

Data availability

All relevant data are within the article and Supporting Information files and available upon request.

Supporting Information—This article contains supporting information. Supporting information includes 7 figures.

Acknowledgments—This research was supported by the Academic Research Grant Tier 3 (MOE2016-T3-1-003) from the Singapore Ministry of Education and Nanyang Technological University.

Author contributions—X. H., J. P. T. conceptualization; X. H. methodology; X. H., X. Z., H. Y. C., P. J. E. validation; X. H., X. Z., H. Y. C., P. J. E. formal analysis; J. P. T. resources; X. H., X. Z., H. Y. C., P. J. E. data curation; X. H., J. P. T. writing – original draft; X. H., X. Z., H. Y. C., J. P. T. writing – review & editing; X. H. visualization; J. P. T. supervision; J. P. T. funding acquisition.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: β -ME, β -mercaptoethanol; AEP, asparaginyl endopeptidase; conLEG, consensus plant legumain; E:S, enzyme:substrate ratio; PAL, peptide asparaginyl ligase; SFTI, sunflower trypsin inhibitor; LAD, ligase activity determinant.

References

- Kembhavi, A. A., Buttle, D. J., Knight, C. G., and Barrett, A. J. (1993) The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* **303**, 208–213
- Dalton, J. P., Hola-Jamriska, L., and Brindley, P. J. (1995) Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. *Parasitology* **111**, 575–580
- Chen, J. M., Dando, P. M., Rawlings, N. D., Brown, M. A., Young, N. E., Stevens, R. A., et al. (1997) Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase. *J. Biol. Chem.* **272**, 8090–8098
- Hara-Nishimura, I., and Nishimura, M. (1987) Proglobulin processing enzyme in vacuoles isolated from developing pumpkin cotyledons. *Plant Physiol.* **85**, 440–445
- Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucl. Acids Res.* **46**, D624–D632
- Dall, E., and Brandstetter, H. (2016) Structure and function of legumain in health and disease. *Biochimie* **122**, 126–150
- Hatsugai, N., Yamada, K., Goto-Yamada, S., and Hara-Nishimura, I. (2015) Vacuolar processing enzyme in plant programmed cell death. *Front. Plant Sci.* **6**, 234
- Gruis, D., Schulze, J., and Jung, R. (2004) Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases. *Plant Cell* **16**, 270–290
- Min, W., and Jones, D. H. (1994) *In vitro* splicing of concanavalin A is catalyzed by asparaginyl endopeptidase. *Nat. Struct. Mol. Biol.* **1**, 502–504
- Nonis, S. G., Haywood, J., Schmidberger, J. W., Mackie, E. R. R., Costa, T. P. S.d., Bond, C. S., et al. (2021) Structural and biochemical analyses of concanavalin A circular permutation by jack bean asparaginyl endopeptidase. *Plant Cell* **33**, 2794–2811
- Nguyen, G. K. T., Wang, S., Qiu, Y., Hemu, X., Lian, Y., and Tam, J. P. (2014) Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol.* **10**, 732–738
- Nguyen, G. K. T., Kam, A., Loo, S., Jansson, A. E., Pan, L. X., and Tam, J. P. (2015) Butelase 1: a versatile ligase for peptide and protein macrocyclization. *J. Am. Chem. Soc. Comm.* **137**, 15398–15401
- Harris, K. S., Durek, T., Kaas, Q., Poth, A. G., Gilding, E. K., Conlan, B. F., et al. (2015) Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat. Commun.* **6**, 10199
- Harris, K. S., Guarino, R. F., Dissanayake, R. S., Quimbar, P., McCorkelle, O. C., Poon, S., et al. (2019) A suite of kinetically superior AEP ligases can cyclise an intrinsically disordered protein. *Sci. Rep.* **9**, 10820
- Hemu, X., El Sahili, A., Hu, S., Wong, K., Chen, Y., Wong, Y. H., et al. (2019) Structural determinants for peptide-bond formation by asparaginyl ligases. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11737–11746
- Rajendran, S., Slazak, B., Mohotti, S., Stromstedt, A. A., Goransson, U., Hettiarachchi, C. M., et al. (2021) Tropical vibes from Sri Lanka - cyclotides from *Viola betonicifolia* by transcriptome and mass spectrometry analysis. *Phytochemistry* **187**, 112749
- Jackson, M. A., Gilding, E. K., Shafee, T., Harris, K. S., Kaas, Q., Poon, S., et al. (2018) Molecular basis for the production of cyclic peptides by plant asparaginyl endopeptidases. *Nat. Commun.* **9**, 2411
- Hemu, X., Chan, N.-Y., Liew, H. T., Hu, S., Zhang, X., Serra, A., et al. (2023) Substrate-binding Glycine residues are major determinants for hydrolase and ligase activity of plant legumains. *New Phytologist*. <https://doi.org/10.1111/nph.18841>
- Bernath-Levin, K., Nelson, C., Elliott, A. G., Jayasena, A. S., Millar, A. H., Craik, D. J., et al. (2015) Peptide macrocyclization by a bifunctional endoprotease. *Chem. Biol.* **22**, 571–582
- Du, J., Yap, K., Chan, L. Y., Rehm, F. B. H., Looi, F. Y., Poth, A. G., et al. (2020) A bifunctional asparaginyl endopeptidase efficiently catalyzes both cleavage and cyclization of cyclic trypsin inhibitors. *Nat. Commun.* **11**, 1575
- Liew, H. T., To, J., Zhang, X., Hemu, X., Chan, N. Y., Serra, A., et al. (2021) The legumain McPAL1 from *Momordica cochinchinensis* is a highly stable Asx-specific splicing enzyme. *J. Biol. Chem.* **297**, 101325
- Cao, Y., Nguyen, G. K., Tam, J. P., and Liu, C. F. (2015) Butelase-mediated synthesis of protein thioesters and its application for tandem chemo-enzymatic ligation. *Chem. Commun.* **51**, 17289–17292
- Nguyen, G. K., Hemu, X., Quek, J. P., and Tam, J. P. (2016) Butelase-mediated macrocyclization of d-amino-acid-containing peptides. *Angew. Chem. Int. Edition* **55**, 12802–12806
- Hemu, X., Qiu, Y., Nguyen, G. K., and Tam, J. P. (2016) Total synthesis of circular bacteriocins by butelase 1. *J. Am. Chem. Soc. Comm.* **138**, 6968–6971
- Bi, X., Yin, J., Nguyen, G. K. T., Rao, C., Halim, N. B. A., Hemu, X., et al. (2017) Enzymatic engineering of live bacterial cell surfaces using butelase 1. *Angew. Chem. Int. Ed. Engl.* **56**, 7822–7825
- Harmand, T. J., Bousbaine, D., Chan, A., Zhang, X., Liu, D. R., Tam, J. P., et al. (2018) One-pot dual labeling of IgG 1 and preparation of C-to-C fusion proteins through a combination of sortase A and butelase 1. *Bioconjug. Chem.* **29**, 3245–3249
- Rehm, F. B. H., Harmand, T. J., Yap, K., Durek, T., Craik, D. J., and Ploegh, H. L. (2019) Site-specific sequential protein labeling catalyzed by a single recombinant ligase. *J. Am. Chem. Soc.* **141**, 17388–17393
- Hemu, X., Zhang, X., Nguyen, G. K. T., To, J., Serra, A., Loo, S., et al. (2021) Characterization and application of natural and recombinant butelase-1 to improve industrial enzymes by end-to-end circularization. *RSC Adv.* **11**, 23105–23112
- Tang, T. M. S., and Luk, L. Y. P. (2021) Asparaginyl endopeptidases: enzymology, applications and limitations. *Org. Biomol. Chem.* **19**, 5048–5062
- Zhao, J., Fan, R., Jia, F., Huang, Y., Huang, Z., Hou, Y., et al. (2021) Enzymatic properties of recombinant ligase butelase-1 and its application in cyclizing food-derived angiotensin I-converting enzyme inhibitory peptides. *J. Agric. Food Chem.* **69**, 5976–5985
- Zhao, J., Ge, G., Huang, Y., Hou, Y., and Hu, S. Q. (2022) Study on activation mechanism and cleavage sites of recombinant butelase-1 zymogen derived from *Clitoria ternatea*. *Biochimie* **199**, 12–22
- Pi, N., Gao, M., Cheng, X., Liu, H., Kuang, Z., Yang, Z., et al. (2019) Recombinant butelase-mediated cyclization of the p53-binding domain of the oncoprotein MdmX-stabilized protein conformation as a promising model for structural investigation. *Biochemistry* **58**, 3005–3015
- Lehmann, M., Pasamontes, L., Lassen, S. F., and Wyss, M. (2000) The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta* **1543**, 408–415
- Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S. F., Pasamontes, L., et al. (2002) The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng. Des. Select.* **15**, 403–411
- Qian, H., Zhang, C., Lu, Z., Xia, B., Bie, X., Zhao, H., et al. (2018) Consensus design for improved thermostability of lipoxygenase from *Anabaena* sp. PCC 7120. *BMC Biotechnol.* **18**, 57

Consensus design of high-yield peptide asparaginyl ligases

36. Gomez-Fernandez, B. J., Risso, V. A., Sanchez-Ruiz, J. M., and Alcalde, M. (2020) Consensus design of an evolved high-redox potential laccase. *Front. Bioeng. Biotechnol.* **8**, 354
37. Yao, H., Cai, H., and Li, D. (2020) Thermostabilization of membrane proteins by consensus mutation: a case study for a fungal delta8-7 sterol isomerase. *J. Mol. Biol.* **432**, 5162–5183
38. Dall, E., and Brandstetter, H. (2013) Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways and regulation. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10940–10945
39. Heitz, A., Hernandez, J.-F., Gagnon, J., Hong, T. T., Pham, T. T. C., Nguyen, T. M., *et al.* (2001) Solution structure of the squash trypsin inhibitor MCoTI-II. A new family for cyclic knottins. *Biochemistry* **40**, 7973–7983
40. Hemu, X., El Sahili, A., Hu, S., Zhang, X., Serra, A., Goh, B. C., *et al.* (2020) Turning an asparaginyl endopeptidase into a peptide ligase. *ACS Catal.* **10**, 8825–8834
41. Sternke, M., Tripp, K. W., and Barrick, D. (2019) Consensus sequence design as a general strategy to create hyperstable, biologically active proteins. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11275–11284
42. Bi, X., Yin, J., Zhang, D., Zhang, X., Balamkundu, S., Lescar, J., *et al.* (2020) Tagging transferrin receptor with a disulfide FRET probe to gauge the redox state in endosomal compartments. *Anal. Chem.* **92**, 12460–12466
43. Rehm, F. B. H., Tyler, T. J., Yap, K., de Veer, S. J., Craik, D. J., and Durek, T. (2021) Enzymatic C-terminal protein engineering with amines. *J. Am. Chem. Soc.* <https://doi.org/10.1021/jacs.1c08976>
44. Friedrich, M. G., Wang, Z., Schey, K. L., and Truscott, R. J. W. (2019) Mechanism of protein cleavage at asparagine leading to protein-protein cross-links. *Biochem. J.* **476**, 3817–3834
45. Abrahmsen, L., Tom, J., Burnier, J., Butcher, K. A., Anthony, K., and Wells, J. A. (1991) Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry* **30**, 4151–4159
46. Toplak, A., Nuijens, T., Quaedflieg, P. J. L. M., Wu, B., and Janssen, D. B. (2016) Peptiligase, an enzyme for efficient chemoenzymatic peptide synthesis and cyclization in water. *Adv. Synth. Catal.* **358**, 2140–2147
47. Schmidt, M., Toplak, A., Quaedflieg, P. J. L. M., Ippel, H., Richelle, G. J. J., Hackeng, T. M., *et al.* (2017) Omniligase-1: a powerful tool for peptide head-to-tail cyclization. *Adv. Synth. Catal.* **359**, 2050–2055
48. Liebscher, S., and Bordusa, F. (2019) Trypsiligase-catalyzed peptide and protein ligation. *Met. Mol. Biol.* **2012**, 111–133