Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis

Qiu, Yibo; Lian, Yilong; Hemu, Xinya; Wang, Shujing; Nguyen, Giang Kien Truc; Tam, James Pingkwan

2014

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

https://hdl.handle.net/10356/79260

https://doi.org/10.1038/nchembio.1586

Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis

Giang K T Nguyen, Shujing Wang, Yibo Qiu, Xinya Hemu, Yilong Lian & James P Tam

School of Biological Sciences, Nanyang Technological University, Singapore. Correspondence should be addressed to J.P.T. (jptam@ntu.edu.sg)

Proteases are ubiquitous in nature whereas naturally-occurring peptide ligases, enzymes catalyzing the reverse reactions of proteases, are rare occurrences. Here we describe the discovery of butelase 1, the first Asn/Asp (Asx) peptide ligase reported. This highly efficient enzyme was isolated from *Clitoria ternatea*, a cyclic-peptide-producing medicinal plant. Butelase 1 shares 71% sequence identity and the same catalytic triad with legumain proteases, but does not hydrolyze the protease substrate of legumain. Instead, butelase 1 cyclizes various peptides of plant and animal origin with yields greater than 95%. With k_{cat} values of up to 17 s^{-1} and catalytic efficiencies as high as 542,000 M^{-1} s^{-1}, it is the fastest peptide ligase known. Importantly, butelase 1 also displays broad specificity for the N-terminal amino acids of the peptide substrate, thus providing a new tool for C-terminus-specific intermolecular peptide ligations.

Ribosome-derived macrocyclic peptides and proteins are natural products with diverse structures and functions[1]. With their defined conformation and enhanced metabolic stability, they are attractive candidates for drug development[2, 3]. Both prokaryotes and eukaryotes produce macrocyclic peptides. Certain bacteria are known to produce very large macrocycles such as bacteriocins and pilins which consist of 35 to 78 residues[4, 5], whereas cyanobacteria produce much smaller cyanobactins that are usually 6 to 11 residues[6]. Higher organisms typically produce medium-sized ring, with some containing one to three disulfide bonds. An example is the family of θ-defensins and retrocyclins, but this is also the only known family of animal-produced cyclic peptides[7, 8]. In contrast, cyclic peptides occur more commonly in plants, particularly in the families of cyclotides, sunflower trypsin inhibitors (SFTIs), and orbitides[9-11].

A major step in the bioprocessing of macrocyclic peptides from their precursors involves an enzyme that acts as a ligase/cyclase to join the N- and C-termini of the processed mature domain to form a cyclic structure. Although peptide-specific enzymes for nonribosomal cyclization are well-characterized[12, 13], only three peptide cyclases involved in the ribosomal peptide synthesis have been isolated at the protein level: TraF, PATG and PCY1[14-16]. All three are serine proteases and require a C-terminal propeptide. However, they exhibit slow kinetics with a k_{cat} of 1 h^{-1} or 1 d^{-1}, which may limit their applications for chemoenzymatic synthesis of macrocyclic peptides and labeling of macromolecules. For these and many other applications, an efficient enzyme with broad substrate specificity would be desirable.

Hints of natural occurring ligases can be found in the cyclotide-producing plant families. Cyclotides are macrocyclic peptides functioned as host defense in plants[17]. The cyclotide family, estimated to contain over 50,000 members, is believed to be the most diverse family of macrocyclic peptides[18]. A legumain or asparaginyl endopeptidase (family C13, clan CD) has been suggested to be responsible for the backbone cyclization which involves an Asn/Asp (Asx) residue at the ligation site[19], but its identity had remained elusive.

Here we report the isolation and characterization of a novel Asx-specific peptide ligase from *Clitoria ternatea* (Fabaceae), a tropical cyclotide-producing plant with medicinal values[20, 21]. We named this enzyme butelase 1, in accordance to the plant's local name (Bunga Telang + Ligase). Butelase 1 efficiently cyclized non-native peptides from various organisms including plant cyclotide kalata B1 (kB1), sunflower trypsin inhibitor (SFTI), conotoxin, insect antimicrobial peptide thanatin and human salivary antimicrobial peptide histatin. The catalytic kinetics of these reactions were extraordinarily high with catalytic efficiencies of up to 542,000 M^{-1} s^{-1}. Furthermore, butelase 1 appears to diverge evolutionarily from the peptidases of family C13 to function in a reverse direction as an Asx-specific ligase/cyclase.

RESULTS

Screening of asparaginyl endopeptidase and cyclase activity

To assay for asparaginyl endopeptidase (AEP) activity, we incubated the crude extract of *C. ternatea* with Z-Ala-Ala-Asn-AMC (Z-AAN-AMC), a fluorogenic substrate selective for legumains[22, 23]. We observed a large increase in fluorescence intensity at 460 nm which indicated the presence of a putative legumain.

We then assayed the cyclase activity using the peptide substrate kB1-NHV, a 31-residue, linear and oxidatively folded form of the cyclotide kB1 with the His-Val sequence at the C-terminus as the propeptide (**Table 1** and **Supplementary Results, Supplementary Fig. 1**). We selected kalata B1, the prototypic cyclotide found in *Oldenlandia affinis* but not in *C. ternatea*, in our screening assays to distinguish it from native cyclotides produced by *C. ternatea*. The His-Val motif is conserved in the cyclotide precursors of *C. ternatea* and a C-terminal dipeptide has been shown to be sufficient for the biosynthesis of cyclotides[24, 25]. Treatment of kB1-NHV with the crude extract of *C. ternatea* yielded a new peptide which matched the calculated mass of native cyclic kB1 (**Fig. 1a**). We further confirmed this peptide product as cyclic kB1 by (1) co-elution with native cyclic kB1 in RP-HPLC (**Supplementary Fig. 2**), (2) tryptic digestion which resulted in a mass increase of 18 Da suggesting a cyclic backbone, and MS/MS analysis confirming the kB1 sequence and Asn-Gly as the ligation site (**Supplementary Fig. 3**), and (3) 1D NMR which showed identical chemical shifts for the cyclized peptide and native cyclic kB1 (**Supplementary Fig. 4**). These results indicate the presence of a putative ligase capable of peptide macrocyclization in the crude extract of *C. ternatea*.

Purification and identification of butelase 1

Our initial attempts to isolate the peptide ligase guided by Z-AAN-AMC were unsuccessful. Fractions giving strong fluorescence intensity after HPLC separation of the crude extract were unable to cyclize kB1-NHV. Instead, we observed a peptide corresponding to the linear form of kB1 with His-Val being hydrolyzed (data not shown). This result suggests that these fractions contain a putative legumain that is capable of hydrolyzing the asparaginyl bond but is not able to catalyze the backbone cyclization.

As Z-AAN-AMC was not useful in screening of the putative ligase, we tested directly all HPLC-separated fractions using kB1-NHV as the substrate. Interestingly, we found cyclase activity in fractions lacking fluorescence (**Fig. 1b**). This result demonstrates that the cyclase activity is separate from the AEP activity. As a control, commercial jack bean legumain was unable to cyclize kB1-NHV and generated only the linear form of kB1 (**Fig. 1c**). We then pooled all the fractions containing cyclase activity and purified the putative ligase in several chromatographic steps to give a single protein band of 38 kDa on SDS-PAGE (**Fig. 2a** and **Supplementary Fig. 5**). The purified enzyme remained stable with minimal loss of activity (<5%) for 30 days at 4 °C. It was soluble in water at a concentration of 10 mg/ml and displayed an optimal working pH of 5.5 to 6.5.

To determine the identity of the purified ligase, we excised the protein band from the SDS gel and subjected to an in-gel tryptic digestion which gave five dominant tryptic fragments (**Supplementary Fig. 6).** The resulting tryptic sequences obtained by MS/MS were searched against the transcriptome data of *C. ternatea* provided by the Beijing Genomics Institute and found to match a unique sequence of a novel protein designated as butelase 1. Transcriptome

profiling also showed five additional butelase-like partial sequences suggesting the presence of multiple legumains/ligases in *C. ternatea* (**Supplementary Fig. 7**).

Butelase 1 does not hydrolyze Z-AAN-AMC

To determine why Z-AAN-AMC was not useful in assaying for butelase 1, we incubated 0.125 μM purified enzyme with 50 μM Z-AAN-AMC. No apparent increase in the fluorescence intensity was observed after incubating for 30 h, indicating that butelase 1 did not hydrolyze Z-AAN-AMC. RP-HPLC analysis showed that <3% hydrolyzed product was formed (**Supplementary Fig. 8**). As a positive control, jack bean legumain completely hydrolyzed Z-AAN-AMC under the same experimental conditions. This result suggests that butelase 1 has evolved to function as a ligase rather than a protease.

Sequence analysis of butelase 1

BLASTP search against the NCBI non-redundant protein database showed that butelase 1 shares high sequence homology with several members of the legumain family. Butelase 1 has the highest homology with a legumain-like protein from *Glycine max* (NCBI reference sequence: XP_003525979) and VmPE-1 from *Vigna mungo* (GenBank: BAA76744.1) with 71% and 70% sequence identity, respectively. This result strongly suggests that butelase 1 is a novel member of the legumain family. The enzymological classification of butelase 1 was further supported by its labeling with the fluorescent probe LP-1, an aza-Asn epoxide probe specific for legumains (**Supplementary Fig. 9**)[26].

Based on the EST sequences in the transcriptome of *C. ternatea*, we predicted that butelase 1 consists of 482 residues and has a mass of 53 kDa, whereas the purified active enzyme is approximately 38 kDa, suggesting that it is post-translationally modified by proteolytic processing. Incubating butelase 1 with PNGase F or glycopeptidase A produced no change in the molecular weight, indicating that butelase 1 is not N-glycosylated (data not shown). Legumains are known to produce as inactive zymogens that undergo autoproteolytic activation to release the N- and C-terminal prodomains. Edman sequencing revealed VEGTR as the N-terminal sequence of butelase 1. The C-terminal processing site was predicted to occur between Asn383 and Ser384, which is based on the apparent molecular weight of 38 kDa, and the auto-cleavage site of other legumains such as proteinase B from *Vicia sativa* and jack bean legumain (**Fig. 2b**)[27, 28].

Homology modeling of butelase 1

We used MODELLER to construct a homology model of butelase 1 based on the zymogen of human legumain, the only member of the legumain family with a known crystal structure (**Fig. 2c**)[29]. The zymogen of butelase 1 (V42-I468) shares 37.8% sequence identity with human legumain. The constructed model of butelase 1 agreed well with the template structure of human legumain with a RMSD of 0.352 Å for the backbone C_α (**Fig. 2d**).

Previous study defined the zymogen of human legumain into three structural parts: the AEP active domain, the activation peptide region, and the legumain stabilization and activity modulation (LSAM) domain. The latter two domains are auto-cleaved during enzyme activation in human legumain. Similarly, we also divided the modeled structure of butelase 1 into three parts: the putative AEP active domain (marine blue, V42-T318), the activation

peptide region (magenta, D319-N383) and the LSAM domain (gray, S385-I468, which is excluded in the final active form of butelase 1). Overall, the AEP active domain of butelase 1 retains 49.8% sequence identity (V42-T318) and displays a good structural alignment of the catalytic triad (Asn59, His165 and Cys207) with that of human legumain (**Fig. 2d,e**). Of interest and possible relevance to the ligase activity in butelase 1 is the activation peptide region which is cleaved off in human legumain. This domain also displays low sequence identities with human (<10%) and jack bean legumains (35%).

Butelase 1 cyclizes non-native peptide substrates

We determined the kinetics of butelase 1 as a peptide cyclase by HPLC and MS analysis using two non-native linear peptide substrates derived from different plant families, the 31-residue kB1-NHV and the 16-residue SFTI-NHV. Despite being non-native substrates with different lengths and sequences, butelase 1 efficiently cyclized these peptides with excellent yields (**Table 1**).

RP-HPLC traces of the cyclization reaction revealed that butelase 1 converted about 40% of kB1-NHV into cyclic kB1 within 6 min, and reached >95% conversion within 45 min, at an enzyme-to-peptide ratio of 1:400 (**Supplementary Fig. 10**). The apparent kinetic parameters of butelase 1 for kB1-NHV calculated from Michaelis-Menten plot using GraphPad Prism are 2.28 ± 0.05 s^{-1} for k_{cat}, 213 ± 10 μM for K_m and 10,700 M^{-1} s^{-1} for catalytic efficiency (k_{cat}/K_m) (**Fig. 3a**). In addition, a 58-residue, cyclodimer of kB1 (<10%) was observed at high substrate concentrations (> 400 μM), suggesting that butelase 1 is able to perform intermolecular ligation and cyclization of long peptide (**Supplementary Fig. 11**).

With SFTI-NHV, butelase 1 also showed >95% conversion yield with the k_{cat} of 0.6 ± 0.02 s^{-1}, K_m of 51 ± 4 μM, and catalytic efficiency of 11,700 M^{-1} s^{-1} (**Fig. 3b**). These data suggest that butelase 1 could cyclize a wide range of peptide substrates.

Cyclization is independent of disulfide bonds

Since both SFTI and kB1 exist in nature as cyclic peptides stabilized by disulfide bonds, we then determined whether conformational assistance by disulfide bonds is required for the cyclization reaction by butelase 1. *S*-alkylation of the reduced kB1-NHV with iodoacetamide gave SA-kB1-NHV. Treatment of 50 μM *S*-alkylated peptide (SA-kB1-NHV) with 0.125 μM butelase 1 resulted in >95% conversion to its cyclic form within 12 min (**Table 1**). Kinetic analysis showed a 50-fold improvement in the catalytic efficiency of SA-kB1-NHV as compared to kB1-NHV (**Fig. 3c**). This result demonstrates that disulfide bonds are not required for peptide cyclization by butelase 1.

Cyclization is facilitated by a C-terminal NHV or DHV

To investigate the requirement of the P1' and P2' positions of the C-terminal propeptide, we synthesized four analogs of kB1-NHV (**Table 1**). Analogs with longer propeptides than kB1-NHV displayed a small decrease in the cyclization rates, with catalytic efficiencies of 4032 and 2971 M^{-1} s^{-1} for kB1-NHVI and kB1-NHVIA, respectively (**Table 1** and **Supplementary Table 1**). In contrast, butelase 1 was significantly less efficient in cyclizing two truncated analogs lacking either Val or His-Val with <10% cyclic kB1 yield after 4 h, and an

incomplete reaction after 30 h (**Supplementary Fig. 12**). This result indicates that a C-terminal HV dipeptide is necessary for an efficient cyclization reaction by butelase 1.

To determine the substrate specificity of butelase 1 at the P1 position, we prepared analog substrates of kB1-NHV by individually replacing the conserved Asn residue with Ala or closely related residues such as Asp, Glu and Gln (**Table 1**). No cyclization of kB1-AHV, kB1-QHV or kB1-EHV was observed after incubating with butelase 1 for 4 h (**Supplementary Fig. 13a-c**). Butelase 1 was able to cyclize kB1-DHV, but at about a hundred-fold slower than kB1-NHV and with less than 10% cyclized product after 4 h (**Supplementary Fig. 13d**). Similarly, we compared the activity of butelase 1 on SFTI-NHV and SFTI-DHV. Butelase 1 cyclized both peptide substrates, but was significantly more efficient with SFTI-NHV than SFTI-DHV (**Supplementary Fig. 14**). These results demonstrate that a C-terminal NHV or DHV tripeptide motif is sufficient for cyclization by butelase 1.

Butelase 1 cyclizes bioactive peptides of animal origin

To provide the evidence of generality, we examined whether butelase 1 can cyclize non-plant-derived proteins using substrates derived from conotoxin (MrIA), thanatin (insect antimicrobial peptide), and histatin-3 (human saliva antimicrobial proteins) (**Table 1**). Since 95% of known cyclotide sequences contain a Gly residue at the N-terminus, we designed our linear precursors to have an N-terminal Gly. Butelase 1 efficiently cyclized conotoxin and histatin-3 but displayed a slower cyclization rate for thanatin (**Supplementary Figs. 15-17**). The k_{cat}, K_m and catalytic efficiencies were determined as 2.2 ± 0.2 s^{-1}, 7.9 ± 2 μM, and 278,000 M^{-1} s^{-1} for conotoxin, and 0.04 ± 0.003 s^{-1}, 0.8 ± 0.25 μM, and 50,000 M^{-1} s^{-1} for thanatin. The kinetic parameters for histatin-3 were not determined due to co-elution of the linear precursor and the cyclized peptide in RP-HPLC. Instead, we monitored the cyclization of histatin-3 by mass spectrometry (**Supplementary Fig. 16**). At an enzyme-to-peptide ratio of 1:400, butelase 1 cyclized 70% of histatin-3 within 6 min and reached >90% conversion in 12 min, a rate comparable to SA-kB1-NHV and MrIA conotoxin.

Intermolecular peptide ligation and N-terminal specificity

The cyclodimerization of kB1 suggests that butelase 1 is able to mediate intermolecular peptide ligation. It is worthwhile to point out that the high sequence diversity of >24 native cyclotides isolated from *C. ternatea* also provides tantalizing hints that butelase 1 is a promiscuous enzyme with broad substrate specificity. To define the N-terminal specificity of an acceptor nucleophile, we used KALVINHV as a model peptide and evaluated its ligation efficiency with XIGGIR (X= any 20 amino acids) in the presence of 0.1 μM butelase 1, 50 μM KALVINHV and 1 mM XIGGIR. Butelase 1 efficiently mediated the intermolecular peptide ligation with broad specificity, accepting most natural amino acids at the P1" position except for Pro and acidic amino acids such as Asp and Glu (**Fig. 4a**). The expected ligation products gave 60-80% yields within 10 min, together with <5% hydrolyzed products and 15-35% of starting materials.

To define the specificity at the P2" position, we synthesized a second peptide library LXGGIR (X= 20 amino acids). Butelase 1 exhibited a more stringent requirement at the P2" as compared to the P1" position and displayed a high preference for hydrophobic amino

acids, particularly Ile, Leu and Val (**Fig. 4b**). This result also explains the high catalytic efficiency of butelase 1 for conotoxin and histatin-3 but not for thanatin which contains a Ser residue at the P2" position. However, it remains to be determined if the specificity of P2" position would be similar when P1" is not Leu.

To determine whether the acyl-enzyme intermediate of a peptide substrate would react with a thiol or a non-peptidyl nucleophile, we incubated KALVINHV with various alkyl and aryl thiols as well as hydrazine (Supplementary Table 2). HPLC traces showed that no significant amount of the expected thioester or hydrazide of KALVIN was observed. Instead, only the hydrolyzed product KALVIN together with the starting material KALVINHV were observed (data not shown).

Hydrolysis vs ligation

To determine the effect of hydrolysis when the N-terminus of a peptide substrate is blocked and cannot act as an acceptor nucleophile in a cyclization reaction, we synthesized an analog of kB1-NHV with the N-terminal Gly being replaced by pyroglutamic acid (pGlu) as pGlu-kB1-NHV. Treatment of 50 μM pGlu-kB1-NHV with 0.125 μM butelase 1 gave <40% hydrolysis of the asparaginyl bond to generate pGlu-kB1-N after 20 min (**Fig. 4c,d**). The hydrolysis was reduced to 10% and <5% in the presence of a 5- and 20-fold molar excess of GIGGIR (**Fig. 4e,f**), respectively, together with the expected ligated product pGlu-kB1-NGIGGIR (65-85% yields).

Recombinant expression of butelase 1

We clone the encoding region of butelase 1 without the signal peptide into the vector pNIC28-Bsa4 for recombinant expression in *E. coli*. The recombinant protein was induced in LB broth containing 25 μg/ml of kanamycin and 0.5 mM of IPTG at 16 °C for 18 h. Butelase 1 was only expressed in the insoluble form in *E. coli* cells. This result indicates that *E. coli* may not be a suitable host for soluble expression of butelase 1.

DISCUSSION

Peptide ligases are rare occurrences and valuable biochemical tools. Here, we describe the discovery of butelase 1, an Asx-specific ligase that efficiently cyclized various peptides. Our study provides strong evidence that butelase 1 is responsible for the biosynthesis of cyclotides in the Fabaceae, the third-largest family of flowering plants which includes many economically important plants and food crops.

Butelase 1 is a unique ligase displaying three different types of catalytic activities: transamidase (kB1-N-amide to cyclic kB1), cyclase (kB1-NHV to cyclic kB1), and ligase (ligation of KALVINHV). It is C-terminal specific to produce Asx-Xaa bonds, accepting a diverse group of Xaa residues at the P1" position. As a tool, it will complement the widely used sortase A and N-terminus-specific chemical ligation[30-33].

Kinetic analysis showed that butelase 1 is the fastest peptide ligase/cyclase reported. Its turnover numbers (0.04 to 17 s^{-1}) are three to four orders of magnitude faster than the known

peptide cyclases such as PATG (1 d^{-1}) and PCY1 (1 h^{-1})[15, 16]. Furthermore, the cyclization catalyzed by butelase 1 is highly efficient, producing macrocyclic peptides quantitatively, whereas PCY1 generates a mixture of both cyclic and linear peptides at equilibrium[16]. The catalytic efficiency of butelase 1 is comparable to the extensively characterized thioesterase domains of nonribosomal peptide synthetases[12, 34]. However, these thioesterases require a C-terminal thioester and generally function best in the context of a large multidomain protein[34].

Enzymes that do not have native cyclization functions, such as sortase A and inteins, have been exploited to produce macrocyclic peptides[35, 36]. Sortase A is a transpeptidase that anchors surface proteins to bacterial cell walls[35]. It has been used for cyclization of histatin and kalata B1[37, 38]. However, its cyclization reaction requires a high enzyme-to-peptide ratio of 1:3 and an overnight incubation. In comparison, at an enzyme-to-peptide ratio of 1:400, butelase 1 cyclizes over 90% of SA-kB1-NHV, MrIA conotoxin and histatin-3 in <12 min. Inteins are autocatalytic splicing elements that have been used successfully for expression of cyclotides, SFTI-1 and θ-defensin[39-41]. Furthermore, the production of combinatorial cyclic peptide libraries inside living *E. coli* cells using expressed protein ligation and protein trans-splicing has been achieved[42, 43]. These methods also allow efficient production of cyclic peptides containing non-natural amino acids[44]. The intein-mediated cyclization, however, requires genetic fusion of each target protein with the intein domain, and ectopic expression of this fusion protein in bacterial systems. Butelase 1, with a broad substrate scope and hyper-efficient kinetics, could provide an alternative and complementary approach for peptide macrocyclization.

Although the ligase mechanism of butelase 1 remains to be solved, clues can be deduced from PatG and PCY1[15, 16]. Structural study showed that PatG contains a subtilisin-like domain which is modified by insertions as a "latch" to allow its substrate recognition and perhaps, to bind loosely the cleaved substrate at the active site and shield the acyl-enzyme intermediate from premature hydrolysis[45]. Similarly, the ligase activity of PCY1 might be attributed to the insertion or deletion near the active site, based on the structure homology with the porcine muscle prolyl oligopeptidase POP[16]. By analogy, our homology modeling showed that butelase 1 contains an extended C-terminal sequence and may use a similar mechanism as PatG to act as a ligase.

Butelase 1 acted, albeit at a slower rate, as a protease in the absence of a suitable acceptor nucleophile. Blocking the amino terminus of kB1-NHV resulted in hydrolysis, which can be minimized by an acceptor peptide. These results suggest a plausible explanation for cyclization in favor of hydrolysis by butelase 1 since cyclization occurs intramolecularly with a high effective molarity of an acceptor nucleophile. However, hydrolysis is a competing reaction in the absence or a low molar excess of a suitable acceptor in the intermolecular ligation reaction. Interestingly, butelase 1 does not hydrolyze the legumain substrate Z-AAN-AMC which contains an unusual Asn-AMC amide bond.

It is intriguing how butelase 1 has evolved to become a ligase since jack bean legumain, a closely related enzyme, fails to catalyze the cyclization[19]. Although transpeptidation activity has been demonstrated for jack bean legumain[46], it only hydrolyzed but did not cyclize kB1-

NHV. Different plants may independently evolve the ability to cyclize as although legumains are ubiquitous in nature, only a limited number of plants found to produce cyclic peptides in a screening for cyclotide-producing plants[18]. Transgenic expression of kB1 genes in *Arabidopsis thaliana*, *Nicotinana tabacum* and *Nicotiana benthamiana* resulted in mostly linear and <6% cyclic peptides[25, 47]. This observation suggests that legumains in these plants are not optimized for cyclization. This may explain for the sporadic distribution of cyclotides in plants as plant legumains must evolve to obtain butelase-1-like features to produce cyclic peptides.

In conclusion, butelase 1 is a novel and unconventional member of the legumain family that is capable of intermolecular peptide ligation and cyclization of peptides from 14 to 58 residues. Butelase 1 holds promise for biotechnology and protein engineering, especially in its application for chemoenzymatic synthesis of macrocyclic peptides and proteins.

ONLINE METHODS

Materials. Sodium 2-mercaptoethanesulfonate, 2-mercaptoethanol, 3-mercapto-1-propanol, 2-mercaptoacetic acid, methyl 2-mercaptoacetate, 3-mercaptopropionic acid, 4-mercaptophenylacetic acid, benzyl mercaptan and hydrazine were purchased from Sigma-Aldrich (USA) with >98% purity. Z-AAN-AMC (N-carbobenzyloxy-Ala-Ala-Asn-7-amido-4-methylcoumarin) and peptide substrates (**Table 1**) with 80% purity were synthesized by GL Biochem (Shanghai). All the purchased peptides were repurified to obtain >95% purity before using in the subsequent experiments. Except for histatin, all peptide substrates used in the cyclization assays were oxidized to form disulfide bonds. Oxidative folding of each peptide was performed for 18 h at a peptide concentration of 30 μM in the buffer containing 50% acetonitrile, 100 mM ammonium bicarbonate, 3 mM reduced glutathione, pH 8.0. Jack bean legumain was purchased from Takara Bio (Japan). Native kB1 peptide was isolated from aerial parts of *O. affinis* and purified by using RP-HPLC. Legumain-specific probe LP-1 was provided by Matthew Bogyo (Stanford University).

Solid-phase synthesis of peptide libraries. Wang resin (1110 mg, 0.9 mmol/g, 1 mmol) were swelled with anhydrous dimethylformamide (DMF) for 30 min. Subsequently, Fmoc-Arg(Pbf)-OH (4.5 eq., 4.5 mM), with benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBop) (2.34 g, 4.5 eq., 4.5 mmol) was added to the resin in the presence of N,N-diisopropylethylamine (DIEA) (1.19 ml, 6.75 eq., 6.75 mmol) in DMF. The reaction was performed in the shaker at room temperature for 1 h and repeated once to ensure the coupling reaction was completed. This was followed by deprotection reactions in 20% piperidine for 5 min twice. Kaiser test consisting a mixture of ninhydrin, potassium cyanide and phenol (10 μl, 1:1:1, *v/v/v*) was used to detect the presence of free amine. The completion of Fmoc deprotection was indicated by the presence of a blue color on the resin. Peptide elongation was carried out manually using standard protocol with Fmoc amino acid/PyBOP/DIEA (4/4/6 eq.) in DMF for 0.5 h.

The peptide sequence IGGIR or GGIR were synthesized on a 1 mmol scale respectively for the first and second library. Subsequently, the resin was split into 20 vessels for the synthesis of individual peptides with variations of the 20 amino acids at first or second position variants in peptide sequence XXGGIR. Final cleavage from the resins and removal of all the

side chain protecting groups were achieved by a mixture of 5% triisopropylsilane (TIS) and 95% of trifluoroacetic Acid (TFA) for 3 h. The cleaved peptides were precipitated with cold diethylether, and dried *in vacuo* to give the crude peptides. Purification was carried out by reverse-phase high-performance liquid chromatogram (RP-HPLC) using Phenomenex C18 preparative column (250 x 10 mm) with a linear gradient 10-40% acetonitrile in 60 min. The RP-HPLC was monitored at 220 nm absorbance. They were characterized and analyzed by MALDI-TOF MS. The purified peptides were lyophilized and obtained with >95% purity for the ligation experiments.

***S*-alkylation of kB1-NHV and pGlu-kB1-NHV.** Each peptide (~ 50 μM) was dissolved ammonium bicarbonate buffer (100 mM, pH 7.8) containing 50 mM DTT, and incubated for 1 h at 37 °C. Two-fold excess of iodoacetamide over the total thiols was added and incubated for 1 h at 37 °C. *S*-alkylated peptides were purified by RP-HPLC.

Transcriptome Sequencing. Total RNA was extracted from fresh plant materials using the PureLink RNA purification system (Invitrogen). The extracted RNA was shipped in dry ice to Beijing Genomic Institute (BGI) for transcriptome analysis using Illumina HiSeq 2000 (5Gb data).

Isolation and purification of butelase 1. 300 g pods of *C. ternatea* were homogenized with 500 ml of extraction buffer (20 mM sodium phosphate, 1 mM EDTA, 1 mM PMSF, 5 mM β-mercaptoethanol (β-ME), pH 6.0). The extraction was conducted at 4 °C to minimize protein degradation. The homogenate was centrifuged and filtered to remove plant debris. Ammonium sulfate was added to the supernatant to reach 20% saturation. The precipitated proteins were discarded and ammonium sulfate was continually added to the supernatant to reach 85% saturation. After centrifugation, the supernatant was discarded and the precipitated proteins were redissolved in 300 ml of extraction buffer. The dissolved sample was dialyzed overnight against 6 l of extraction buffer using a 10 kDa cut-off dialysis tubing. The dialyzed sample was centrifuged and filtered to give the crude extract of *C. ternatea*. This crude extract was applied to a flash column containing 100 ml slurry of Q-Sepharose Fast Flow anion-exchange resin (GE Healthcare). The column was wash with 800 ml of buffer A (20 mM phosphate buffer, 1 mM EDTA, 5 mM β-ME, pH 6.0) and eluted with 400 ml of buffer B (20 mM sodium phosphate, 1 mM EDTA, 5 mM β-ME, 200 mM KCl, pH 6.0). The eluent was concentrated to a final volume of 3 ml using 10 kDa cut-off centrifugal filter units (Amicon Ultra, Millipore). The concentrated sample was subjected to size exclusion chromatography using a BioSuite HPLC column (300 x 21.5 mm, Waters). Fractions with peptide cyclase activity were pooled and further purified by anion-exchange chromatography using an analytical PolyWAX HPLC column (200 x 4.6 mm, PolyLC) with a linear gradient 0-500 mM sodium chloride in 60 min. The enzyme purity was analyzed by SDS-PAGE and silver staining. Approximate 0.4 mg butelase 1 was obtained from 300 g of plant materials.

1D NMR spectra of kalata B1. Native and butelase-cyclized kalata B1 were prepared in 95% H2O/5%D2O at 0.1 mM concentration, pH 4.3. 1D ^{1}H spectra of both peptides were recorded on a 600 MHz NMR spectrometer (Bruker) equipped with a cryo-probe.

Protein identification of butelase 1 and BLAST analysis. Purified butelase 1 was analyzed by SDS-PAGE under denaturing condition. The gel was silver stained and the protein band was excised and subjected to in-gel tryptic digestion. Tryptic peptides were sequence by

MALDI-TOF MS/MS. The resulting sequences were used for TBLASTN searches again the *C. ternatea* transcriptome database provided by the Beijing Genomics Institute using BioEdit software. To identify homologous proteins, BLASTP searches were performed using the translated protein sequence of butelase 1 against the NCBI non-redundant protein database.

N-terminal Edman Sequencing. Purified butelase 1 (100 picomole) was separated on SDS-PAGE and transferred to a PVDF membrane. The N-terminal Edman degradation of butelase 1 was performed for 5 cycles with a ABI Procise 494 Protein Sequencer (Applied Biosystems).

Homology modeling. The homology model of butelase 1 was constructed by MODELLER (version 9.10) with the template human legumain (PDB ID: 4FGU). The structure of modelled butelase 1 was viewed and analyzed with the software PyMOL.

Determination of AEP activity. The AEP activity was determined by using the fluorogenic substrate Z-AAN-AMC at a concentration of 100 μM in buffer A. Crude extract or HPLC-separated fractions (50 μL) were added at an equal volume of 100 μM Z-AAN-AMC solution and incubated at 37 °C for 30 min. Emitted fluorescence was measured with an excited wavelength of 380 nm and emission wavelength of 460 nm.

***In vitro* cyclization assays and kinetic analysis.** *In vitro* cyclization assays were performed in 50-μl reaction mixtures containing buffer A, 0.125 μM butelase 1 and varying peptide concentrations (0.5 to 400 μM). The enzyme concentration was estimated by UV absorbance at 280 nm. Each reaction was performed in triplicate at 37 °C and quenched by adding 5 μl of 1 M HCl solution. The peptides were separated by using a reversed-phase C18 analytical column (150 x 2.1 mm, Vydac) with a linear gradient from 12% to 48% acetonitrile over 15 min on a Nexera UHPLC system (Shimadzu). The cyclization velocities were calculated by converting the HPLC-peak areas of remained linear precursors or the cyclized products into concentrations. The identity of each HPLC peak was analyzed by MALDI-TOF MS (ABI 4800 MALDI TOF/TOF). All reactions were carried out in triplicates and the velocities were input into GraphPad Prism (GraphPad Software, San Diego) to obtain the Michaelis-Menten curve and the kinetic parameters (k_{cat} and K_m) for each peptide.

Biophysical characterization of butelase 1. The solubility of butelase 1 was determined by concentrating the purified enzyme solution from 100 μg/ml to 10 mg/ml using an Amicon Ultra-15 centrifugal filter device, 10 kDa MWCO (Merck Millipore Ltd, Ireland). No visible precipitate was observed during the concentrating process.
The enzyme stability at 4 °C was determined by comparing the activity of the freshly isolated enzyme and the 30-day-storage sample using SFTI-NHV as substrate. The amount of the product formed was compared and quantified by RP-HPLC.
The optimal pH for the enzyme activity was determined by conducting the cyclization reactions of SFTI-NHV under different pH conditions (from 3.6 to 8.0). Each assay was performed in a 50-μl reaction mixture containing 50 μM STTI-NHV, 0.1 μM butelase 1 and 20 mM sodium acetate buffer for pH <6.0 or sodium phosphate buffer for pH >6.0. The amount of the product formed was compared and quantified by RP-HPLC.

Butelase-mediated ligation of KALVINHV and peptide library XXGGIR. For the substrate specificity determination, the reactions were conducted at the ratio of 1:500:10000

for butelase 1: KALVINHV: XXGGIR, at the final concentration of 0.1 μM: 0.05 mM: 1 mM in presence of 1 mM EDTA, incubated at pH 6.5, 42 °C. At 10 min, 25 μl reaction mixture was quenched with 250 μl 0.1% TFA and analyzed by RP-HPLC. The yields were calculated by converting the HPLC-peak areas of remained peptides or the ligated products into concentrations.

Butelase-mediated ligation of KALVINHV and thiols (or hydrazine). The reactions were conducted in presence of butelase 1: KALVINHV: XXGGIR, at the final concentration of 0.1 μM: 0.05 mM: 10 mM with 1 mM EDTA, incubated at pH 6.5, 42 °C. At 30 min, the reaction was quenched and analyzed by RP-HPLC.

Recombinant expression of butelase 1. The coding region for butelase 1 without the signal peptide was cloned into the expression vector pNIC28-Bsa4. The recombinant construct was confirmed by DNA sequencing. The expression vector was transformed into *E. coli* strain BL21(DE3) Rosetta T1R. The recombinant protein was induced in LB broth containing 25 μg/mL of kanamycin and 0.5 mM of IPTG at 16 °C for 18 h.

ACCESSION CODES

The nucleotide sequence for butelase 1 has been deposited in the GenBank database under the accession number KF918345.

ACKNOWLEDGEMENTS

We thank Professor Matthew Bogyo at Stanford University for providing the legumain-specific probe LP-1 and Rachel Wang at Nanyang Technological University for helpful comments on this manuscript. This work was supported in part by the Singapore National Research Foundation NRF-CRP8-2011-05.

AUTHOR CONTRIBUTIONS

G. K. T. N. designed the experiments, isolated and characterized butelase 1. S. W. performed the 1D NMR and homology modeling of butelase 1. Y. Q. synthesized the peptide libraries. X. H. evaluated the intermolecular ligation efficiency. Y. L. performed the kinetic studies for conotoxin, thanatin and histatin. J. P. T. initiated, planned, supervised and edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Cascales, L. & Craik, D.J. Naturally occurring circular proteins: distribution, biosynthesis and evolution. *Org. Biomol. Chem.* **8**, 5035-5047 (2010).
2. Craik, D.J. Chemistry - Seamless proteins tie up their loose ends. *Science* **311**, 1563-1564 (2006).

3. Wong, C.T. et al. Orally active peptidic bradykinin B1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment. *Angew. Chem. Int. Ed. Engl.* **51**, 5620-5624 (2012).
4. Eisenbrandt, R. et al. Conjugative pili of IncP plasmids, and the Ti plasmid T pilus are composed of cyclic subunits. *J. Biol. Chem.* **274**, 22548-22555 (1999).
5. Jack, R.W., Tagg, J.R. & Ray, B. Bacteriocins of Gram-Positive Bacteria. *Microbiol. Rev.* **59**, 171-200 (1995).
6. Sivonen, K., Leikoski, N., Fewer, D.P. & Jokela, J. Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Appl. Microbiol. Biotechnol.* **86**, 1213-1225 (2010).
7. Tang, Y.Q. et al. A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins. *Science* **286**, 498-502 (1999).
8. Cole, A.M. et al. Retrocyclin: A primate peptide that protects cells from infection by T- and M-tropic strains of HIV-1. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1813-1818 (2002).
9. Luckett, S. et al. High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *J. Mol. Biol.* **290**, 525-533 (1999).
10. Craik, D.J., Daly, N.L., Bond, T. & Waine, C. Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J. Mol. Biol.* **294**, 1327-1336 (1999).
11. Arnison, P.G. et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108-160 (2012).
12. Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A. & Walsh, C.T. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **407**, 215-218 (2000).
13. Sieber, S.A. & Marahiel, M.A. Learning from nature's drug factories: Nonribosomal synthesis of macrocyclic peptides. *J. Bacteriol.* **185**, 7036-7043 (2003).
14. Haase, J. & Lanka, E. A specific protease encoded by the conjugative DNA transfer systems of IncP and Ti plasmids is essential for pilus synthesis. *J. Bacteriol.* **179**, 5728-5735 (1997).
15. Lee, J., McIntosh, J., Hathaway, B.J. & Schmidt, E.W. Using Marine Natural Products to Discover a Protease that Catalyzes Peptide Macrocyclization of Diverse Substrates. *J. Am. Chem. Soc.* **131**, 2122-2124 (2009).
16. Barber, C.J. et al. The two-step biosynthesis of cyclic peptides from linear precursors in a member of the plant family Caryophyllaceae involves cyclization by a serine protease-like enzyme. *J. Biol. Chem.* **288**, 12500-12510 (2013).
17. Craik, D.J. Host-defense activities of cyclotides. *Toxins* **4**, 139-156 (2012).
18. Gruber, C.W. et al. Distribution and evolution of circular miniproteins in flowering plants. *Plant Cell* **20**, 2471-2483 (2008).

19. Saska, I. et al. An asparaginyl endopeptidase mediates in vivo protein backbone cyclization. *J. Biol. Chem.* **282**, 29721-29728 (2007).
20. Nguyen, G.K. et al. Discovery and Characterization of Novel Cyclotides Originated from Chimeric Precursors Consisting of Albumin-1 Chain a and Cyclotide Domains in the Fabaceae Family. *J. Biol. Chem.* **286**, 24275-24287 (2011).
21. Poth, A.G. et al. Discovery of Cyclotides in the Fabaceae Plant Family Provides New Insights into the Cyclization, Evolution, and Distribution of Circular Proteins. *ACS Chem. Biol.* **6**, 345-355 (2011).
22. Kembhavi, A.A., Buttle, D.J., Knight, C.G. & Barrett, A.J. The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* **303**, 208-213 (1993).
23. Sojka, D. et al. IrAE - An asparaginyl endopeptidase (legumain) in the gut of the hard tick Ixodes ricinus. *Int. J. Parasitol.* **37**, 713-724 (2007).
24. Nguyen, G.K., Lim, W.H., Nguyen, P.Q. & Tam, J.P. Novel Cyclotides and Uncyclotides with Highly Shortened Precursors from *Chassalia chartacea* and Effects of Methionine Oxidation on Bioactivities. *J. Biol. Chem.* **287**, 17598-17607 (2012).
25. Conlan, B.F. et al. Insights into Processing and Cyclization Events Associated with Biosynthesis of the Cyclic Peptide Kalata B1. *J. Biol. Chem.* **287**, 28037-28046 (2012).
26. Lee, J. & Bogyo, M. Development of Near-Infrared Fluorophore (NIRF)-Labeled Activity-Based Probes for in Vivo Imaging of Legumain. *ACS Chem. Biol.* **5**, 233-243 (2010).
27. Becker, C. et al. Purification, Cdna Cloning and Characterization of Proteinase-B, an Asparagine-Specific Endopeptidase from Germinating Vetch (Vicia-Sativa L) Seeds. *Eur. J. Biochem.* **228**, 456-462 (1995).
28. Abe, Y. et al. Asparaginyl Endopeptidase of Jack Bean-Seeds - Purification, Characterization, and High Utility in Protein-Sequence Analysis. *J. Biol. Chem.* **268**, 3525-3529 (1993).
29. Dall, E. & Brandstetter, H. Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10940-10945 (2013).
30. Hackeng, T.M., Griffin, J.H. & Dawson, P.E. Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10068-10073 (1999).
31. Mao, H., Hart, S.A., Schink, A. & Pollok, B.A. Sortase-mediated protein ligation: a new method for protein engineering. *J. Am. Chem. Soc.* **126**, 2670-2671 (2004).
32. Tam, J.P. & Wong, C.T. Chemical synthesis of circular proteins. *J. Biol. Chem.* **287**, 27020-27025 (2012).

33. Tam, J.P., Lu, Y.A., Yang, J.L. & Chiu, K.W. An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8913-8918 (1999).
34. Kohli, R.M., Trauger, J.W., Schwarzer, D., Marahiel, M.A. & Walsh, C.T. Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry (Mosc.)* **40**, 7099-7108 (2001).
35. Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12424-12429 (1999).
36. Levesque MP, T.R., Motley T, Katari MS, Dedhia NN, O'Shaughnessy AL, Balija V, Martienssen RA, McCombie RW, Benfey P, Stevenson D.
37. Bolscher, J.G.M. et al. Sortase A as a tool for high-yield histatin cyclization. *FASEB J.* **25**, 2650-2658 (2011).
38. Jia, X. et al. Semienzymatic Cyclization of Disulfide-rich Peptides Using Sortase A. *J. Biol. Chem.* (2014).
39. Kimura, R.H., Tran, A.T. & Camarero, J.A. Biosynthesis of the cyclotide Kalata B1 by using protein splicing. *Angew. Chem. Int. Ed. Engl.* **45**, 973-976 (2006).
40. Austin, J., Kimura, R.H., Woo, Y.H. & Camarero, J.A. In vivo biosynthesis of an Ala-scan library based on the cyclic peptide SFTI-1. *Amino Acids* **38**, 1313-1322 (2010).
41. Gould, A. et al. Recombinant production of rhesus theta-defensin-1 (RTD-1) using a bacterial expression system. *Mol. Biosyst.* **8**, 1359-1365 (2012).
42. Austin, J., Wang, W., Puttamadappa, S., Shekhtman, A. & Camarero, J.A. Biosynthesis and biological screening of a genetically encoded library based on the cyclotide MCoTI-I. *Chembiochem* **10**, 2663-2670 (2009).
43. Young, T.S. et al. Evolution of cyclic peptide protease inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11052-11056 (2011).
44. Jagadish, K. et al. Expression of fluorescent cyclotides using protein trans-splicing for easy monitoring of cyclotide-protein interactions. *Angew. Chem. Int. Ed. Engl.* **52**, 3126-3131 (2013).
45. Koehnke, J. et al. The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nat. Struct. Mol. Biol.* **19**, 767-772 (2012).
46. Min, W. & Jones, D.H. In-Vitro Splicing of Concanavalin-a Is Catalyzed by Asparaginyl Endopeptidase. *Nat. Struct. Biol.* **1**, 502-504 (1994).
47. Gillon, A.D. et al. Biosynthesis of circular proteins in plants. *Plant J.* **53**, 505-515 (2008).

Table 1 Cyclization yields of peptide substrates mediated by butelase 1

Peptide Substrate	Sequence	Time (h)	Yield (%)
kB1-NHVIA	GLPVCGETCVGGTCNTPGCTCSWPVCTRNHVIA	3	>95
kB1-NHVI	G ——— NHVI	2	>95
kB1-NHV	G ——— NHV	0.8	>95
kB1-NH	G ——— NH	4	<5
$kB1_{NH2}$	G ——— N_{NH2}	4	<10
kB1-DHV	G ——— DHV	4	<10
kB1-AHV	G ——— AHV	4	<1
kB1-QHV	G ——— QHV	4	<1
kB1-EHV	G ——— EHV	4	< 1
SA-kB1-NHV[a]	G ——— NHV	0.2	>95
SFTI-NHV	GRCTKSIPPICFPNHV	0.8	>95
SFTI-DHV	G ——— DHV	4	<10
MrIA	GVCCGYKLCHPCAGNHV	0.2	>95
Histatin-3	GLDSHAKRHHGYKRKFHEKHHSHRGYRSNYLYDNHV	0.2	>90
Thanatin	GSKKPVPIIYCNRRTGKCQRMNHV	4	59

Assays were performed at 37 °C and an enzyme-to-peptide ratio of 1 : 400 (0.125 μM butelase 1 : 50 μM peptide). $kB1_{NH2}$ indicates the linear, amidated form of kalata B1 at the C-terminus. [a]*S*-carbamidomethylated kB1-NHV. Yields are means of triplicate experiments.

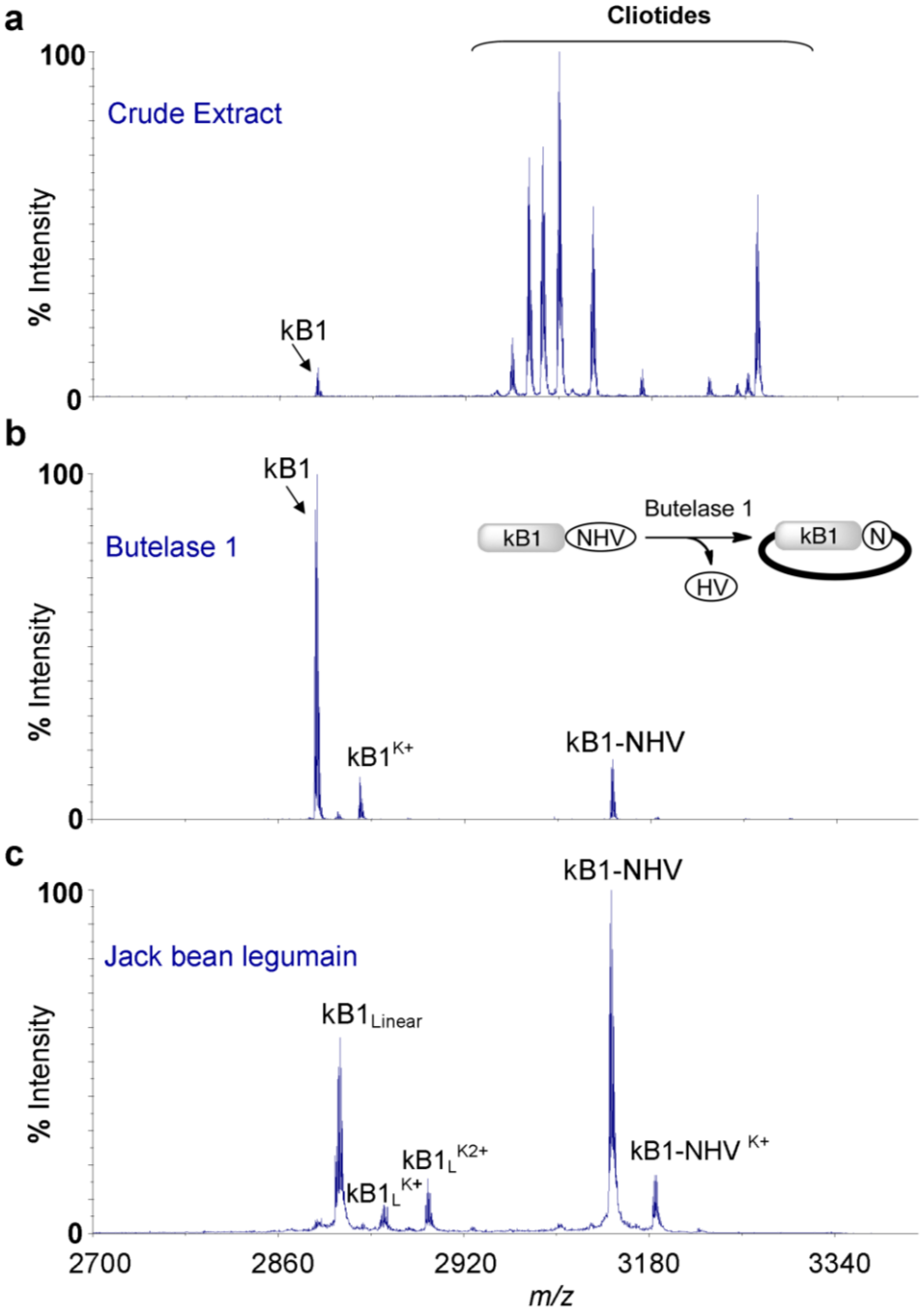


Figure 1 | MS characterization of peptide cyclase activity. (**a,b**) MS profiles of kB1-NHV cyclization mediated by the crude extract of *C. ternatea* and purified butelase 1, respectively. Peptides in the bracket are cliotides, naturally occurring cyclotides in *C. ternatea*. The product, cyclic kB1, is indicated by the arrows. (**c**) Jack bean legumain was used as a control. Jack beak legumain hydrolyzed the asparaginyl bond in kB1-NHV to give linear form of kB1 ($kB1_L$). Peaks labeled with K^+ or K_2^+ are ion adducts corresponding to the binding of one or two potassium ions, respectively.

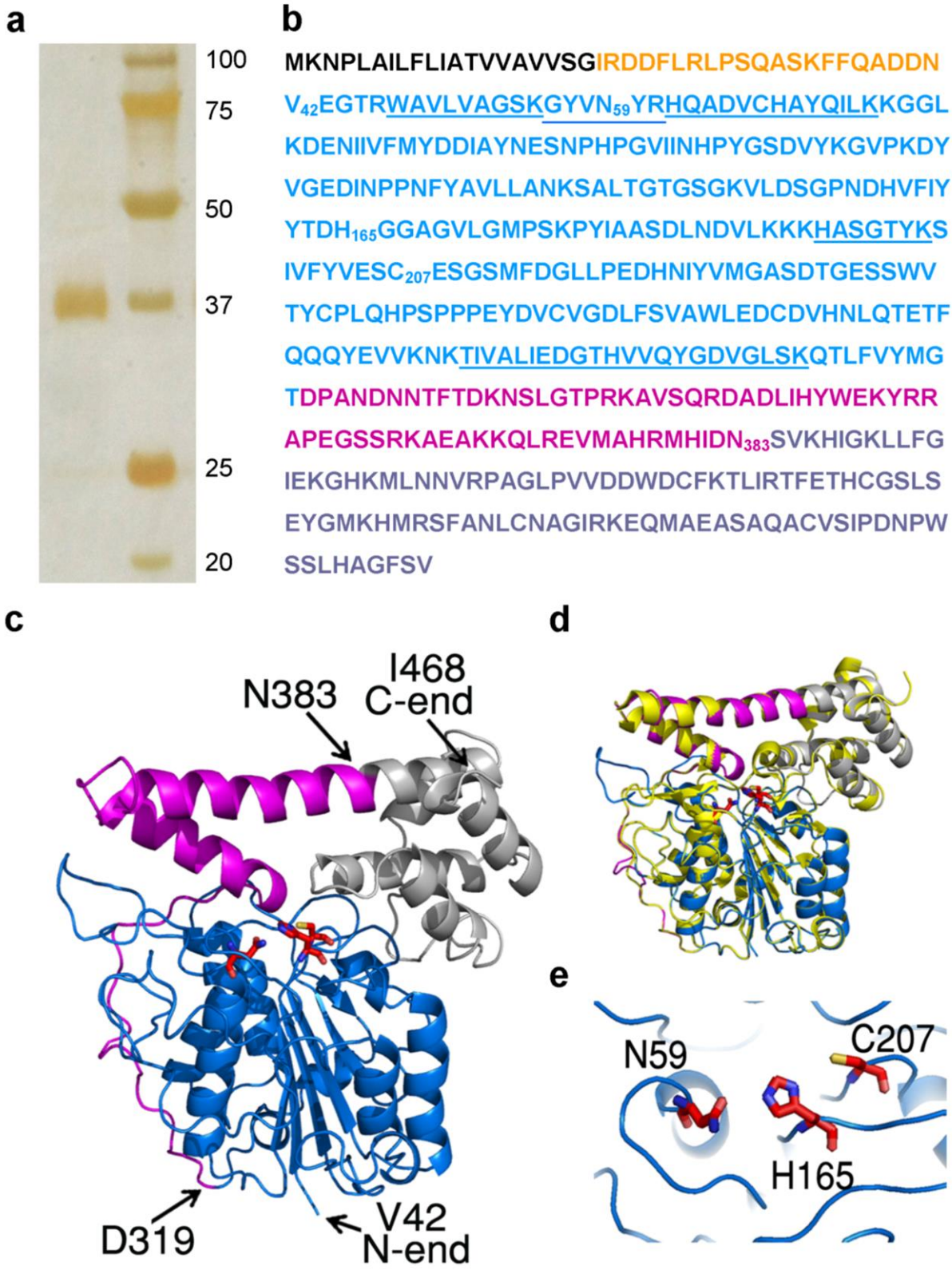


Figure 2 | Isolation, characterization and homology modeling of butelase 1. (**a**) SDS-PAGE analysis of purified butelase 1. Proteins were visualized by silver staining. The left lane is purified butelase 1 and the right lane is a protein ladder with molecular weights (kDa) indicated. (**b**) Translated sequence of butelase 1 precursor deduced from the EST sequences.

The sequence is color-coded with the endoplasmic reticulum signal shown in black, N-terminal prodomain in orange, the AEP domain in blue, the activation peptide region in magenta, and the LSAM domain in gray. The first and the last residues (V42 and N383) of the purified active enzyme are labeled. The conserved residues of the catalytic triad (Asn59, His165 and Cys207) are numbered. Peptide sequences obtained from the in-gel tryptic digestion are underlined. (**c**) Modeling structure of zymogenic butelase 1 based on the structure of human legumain with the AEP domain shown in blue, activation peptide region in magenta and LSAM domain in gray. Residues of the catalytic triad are shown in stick. (**d**) Structure alignment of modeled butelase 1 and the template human legumain (PDB ID: 4FGU; yellow). (**e**) The enlarged view of the catalytic triad residues.

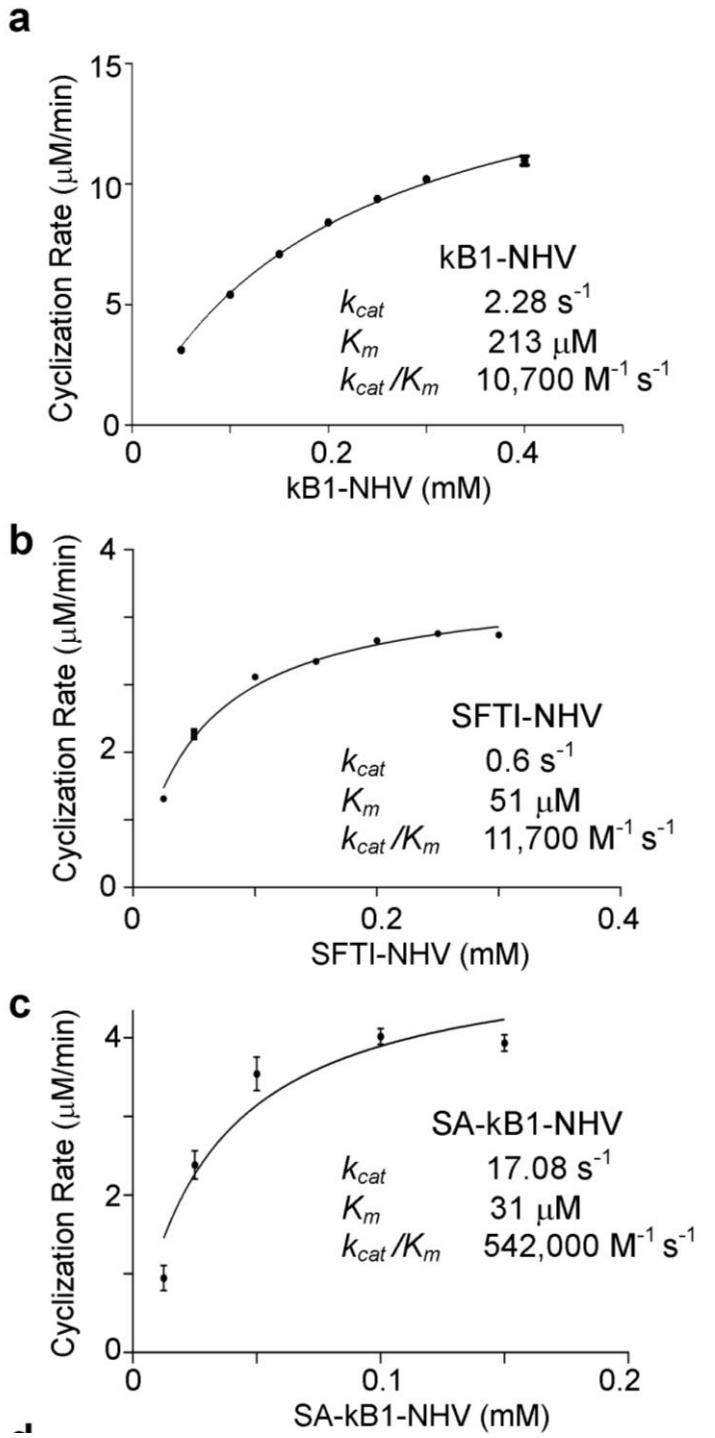


Figure 3 | Kinetic characterization of butelase 1 as a peptide cyclase. (**a-c**) Michaelis-Menten plots of butelase 1 kinetics for kB1-NHV, SFTI-NHV and SA-kB1-NHV. The cyclization rates were calculated by converting the HPLC-peak areas of the products into concentrations. For the kinetic measurements of kB1-NHV and SFTI-NHV, the assays were performed at 37 °C for 12 min in the presence of 0.125 μM butelase 1 and varying substrate concentrations. For SA-kB1-NHV, due to much faster cyclization rate, the enzyme concentration was used at 5 nM instead of 0.125 μM and the incubation time was reduced to 6 min.

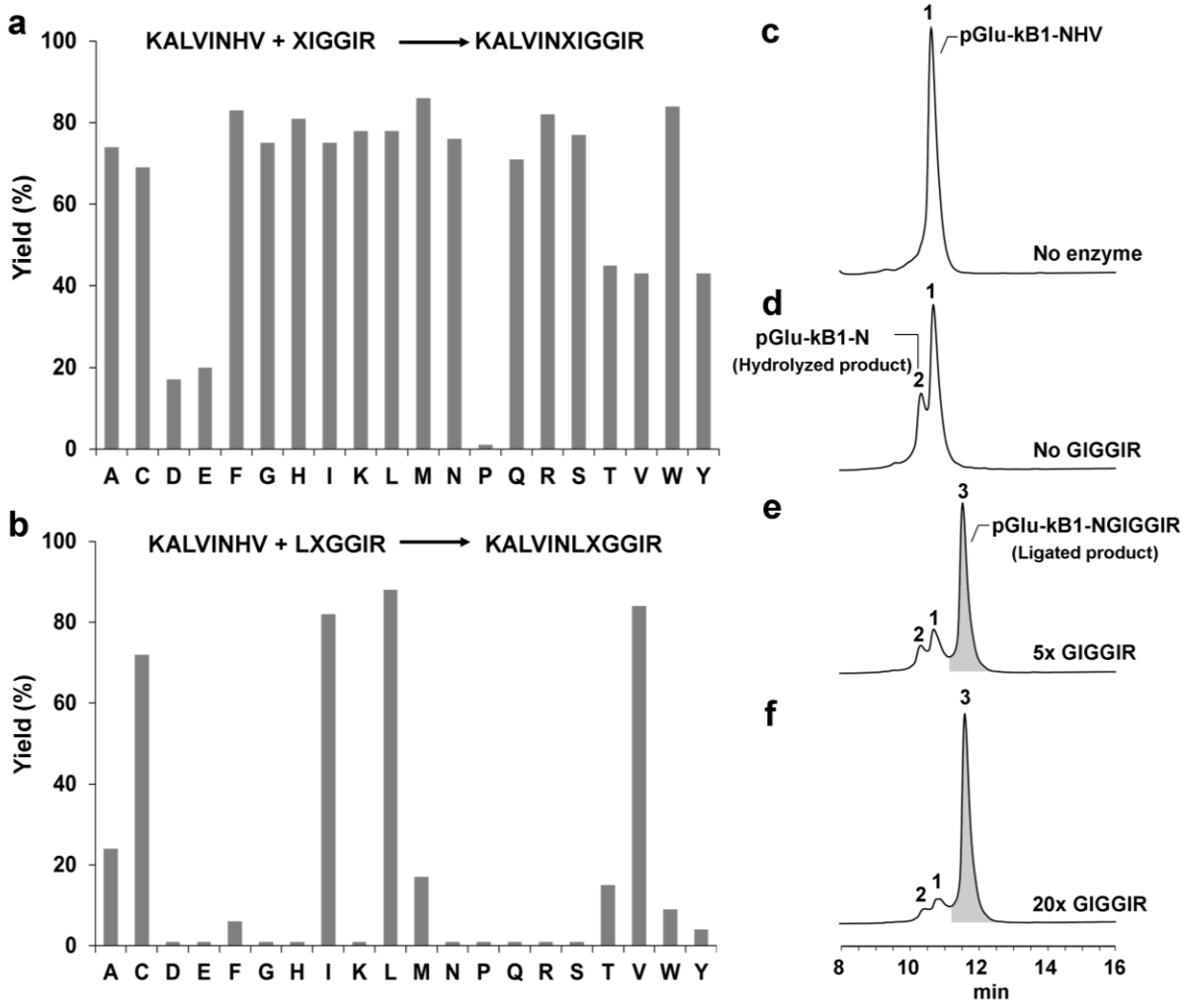


Figure 4 | Acceptor specficity of butelase-mediated peptide ligation. (**a**) Intermolecular ligation of KALVINHV and XIGGIR facilitated by butelase 1. (**b**) Intermolecular ligation of KALVINHV and LXGGIR facilitated by butelase 1. The reactions were performed in the presence of 0.1 μM butelase 1, 50 μM KALVINHV, 1 mM XIGGIR/LXGGIR and incubated for 10 min. Yields are means of triplicate experiments. (**c-f**) HPLC profiles of pGlu-kB1-NHV ligation with GIGGIR. The ligation reactions were performed at 37 °C for 20 min in the presence of 0.125 μM butelase 1 and 50 μM pGlu-kB1-NHV and varying concentrations of GIGGIR (0, 250 μM and 1 mM). pGlu-kB1-NHV used in the ligation reactions was *S*-carbamidomethylated. Peaks labeled 1, 2, and 3 are corresponding to pGlu-kB1-NHV, hydrolyzed product pGlu-kB1-N, and ligated product pGlu-kB1-NGIGGIR (shaded), respectively.

Supplementary Information

Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis

Giang K T Nguyen, Shujing Wang, Yibo Qiu, Xinya Hemu, Yilong Lian & James P Tam

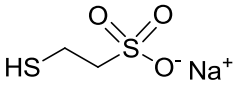
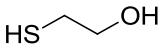
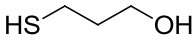
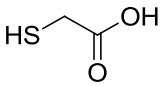
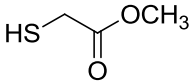
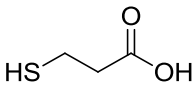
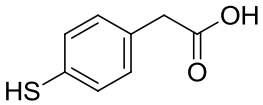
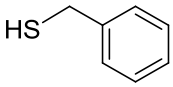
Supplementary Results

Supplementary Table 1 Kinetic parameters of butelase 1 for various peptide substrates

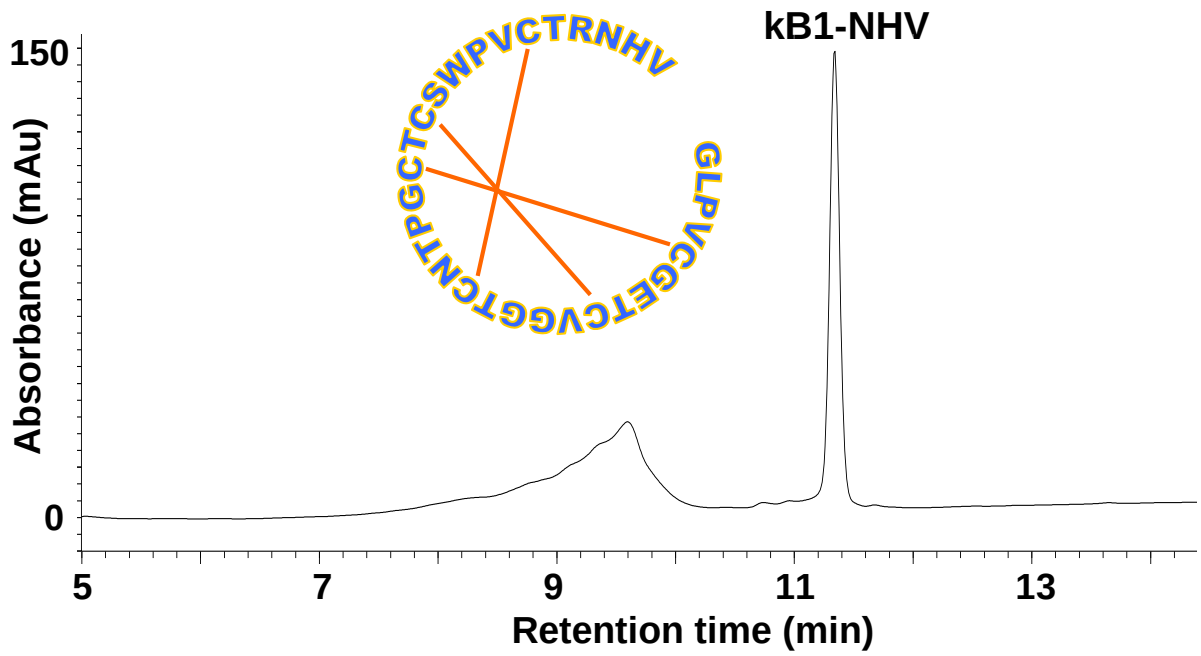
Peptide Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
SA-kB1-NHV	17.08 ± 5	31.5 ± 8	542,000
kB1-NHV	2.28 ± 0.05	213 ± 10	10,700
kB1-NHVIA	0.38 ± 0.01	129 ± 9	2971
kB1-NHVI	0.25 ± 0.01	62 ± 5	4032
SFTI-NHV	0.6 ± 0.02	51 ± 4	10,700
MrIA conotoxin	2.2 ± 0.2	7.9 ± 2	278,000
Thanatin	0.04 ± 0.003	0.8 ± 0.25	50,000

Data are means of triplicate experiments ± SD

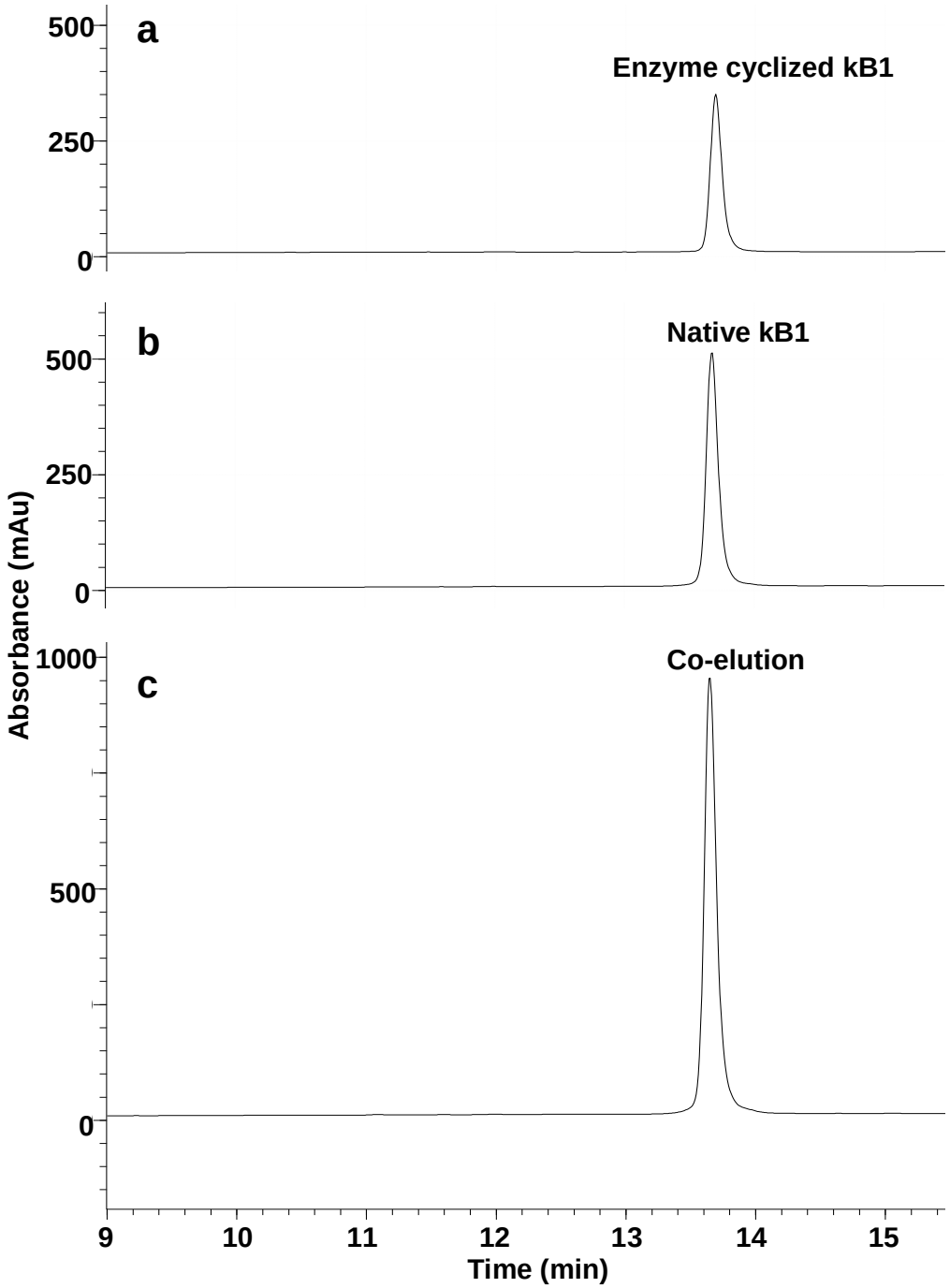
Supplementary Table 2 List of thiols and non-peptidyl nucleophiles tested for the N-terminal specificity of butelase 1

Compound	Chemical Formula
Sodium 2-mercaptoethanesulfonate	
2-Mercaptoethanol	
3-Mercapto-1-propanol	
2-Mercaptoacetic acid	
Methyl 2-mercaptoacetate	
3-Mercaptopropionic acid	
4-Mercaptophenylacetic acid	
Benzyl mercaptan	
Hydrazine	NH_2NH_2

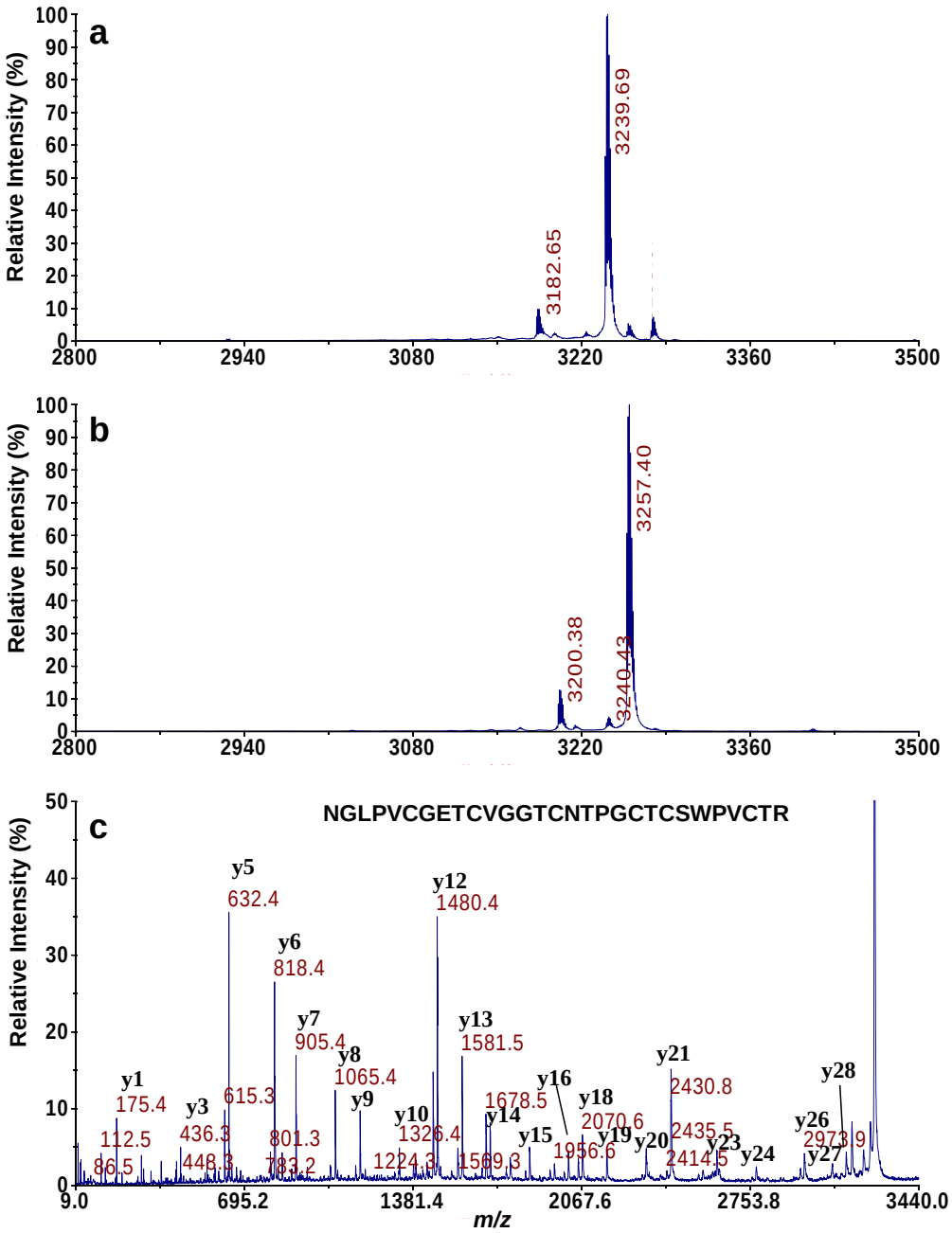
The assays were performed in the presence of 50 μM KALVINHV, 0.1 μM butelase 1 and 10 mM of various thiols and hydrazine for 30 min.



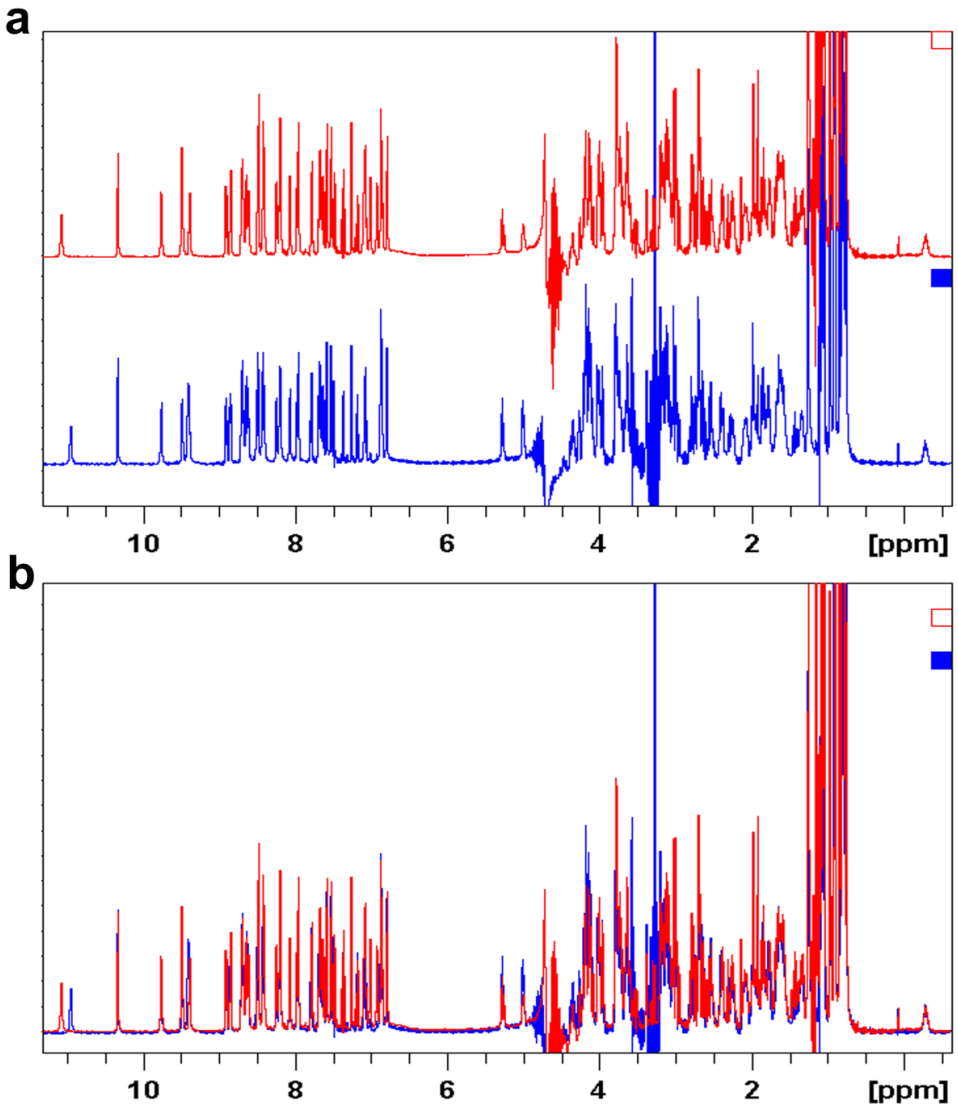
Supplementary Figure 1. Oxidative folding of kB1-NHV. Peptide was folded for 18 h at a 30 μM concentration in buffer containing 50% acetonitrile, 100 mM ammonium bicarbonate, 3 mM reduced glutathione, pH 8.0. The folded peptide eluted last in the RP-HPLC.



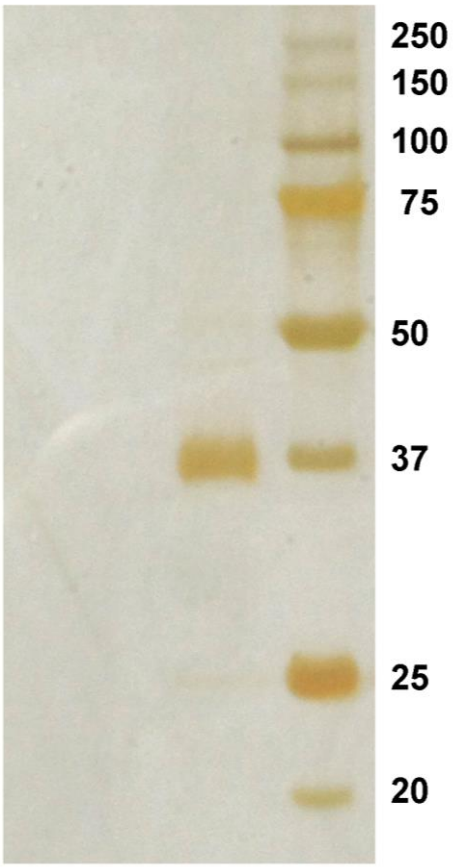
Supplementary Figure 2. Co-elution of enzyme-cyclized kB1 and native peptide. (**a**) HPLC profile of enzyme-cyclized kB1. (**b**) HPLC profile of native kB1 extracted from *O. affinis*. (**c**) Co-elution profile of enzyme-cyclized and native kB1.



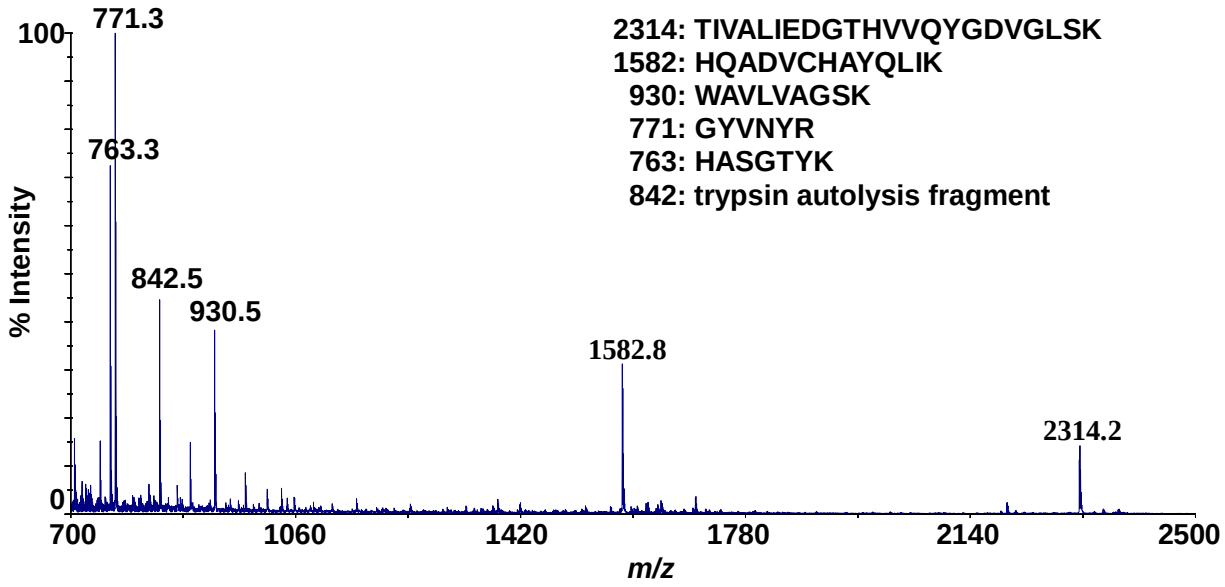
Supplementary Figure 3. MS evidence of a cyclic backbone in the enzyme-cyclized kB1 obtained from conversion of kB1-NHV by butelase 1. (**a**) MS profile of the cyclized kB1 after *S*-carbamidomethylation. The cyclized kB1 has a *m/z* value of 2891, which became 3239 after *S*-alkylation. A minor peak at 3182 is observed due to incomplete alkylation where only 5 cysteines were modified. (**b**) MS profile of *S*-alkylated kB1 after tryptic digestion. A mass increase of 18 Da was observed, which indicated the addition of a water molecule and a cyclic backbone. (**c**) MS/MS profile of the 3257-Da tryptic fragment. The peptide sequence is shown at the top of the MS/MS spectrum. The y-ions are labeled at the top of corresponding peaks.



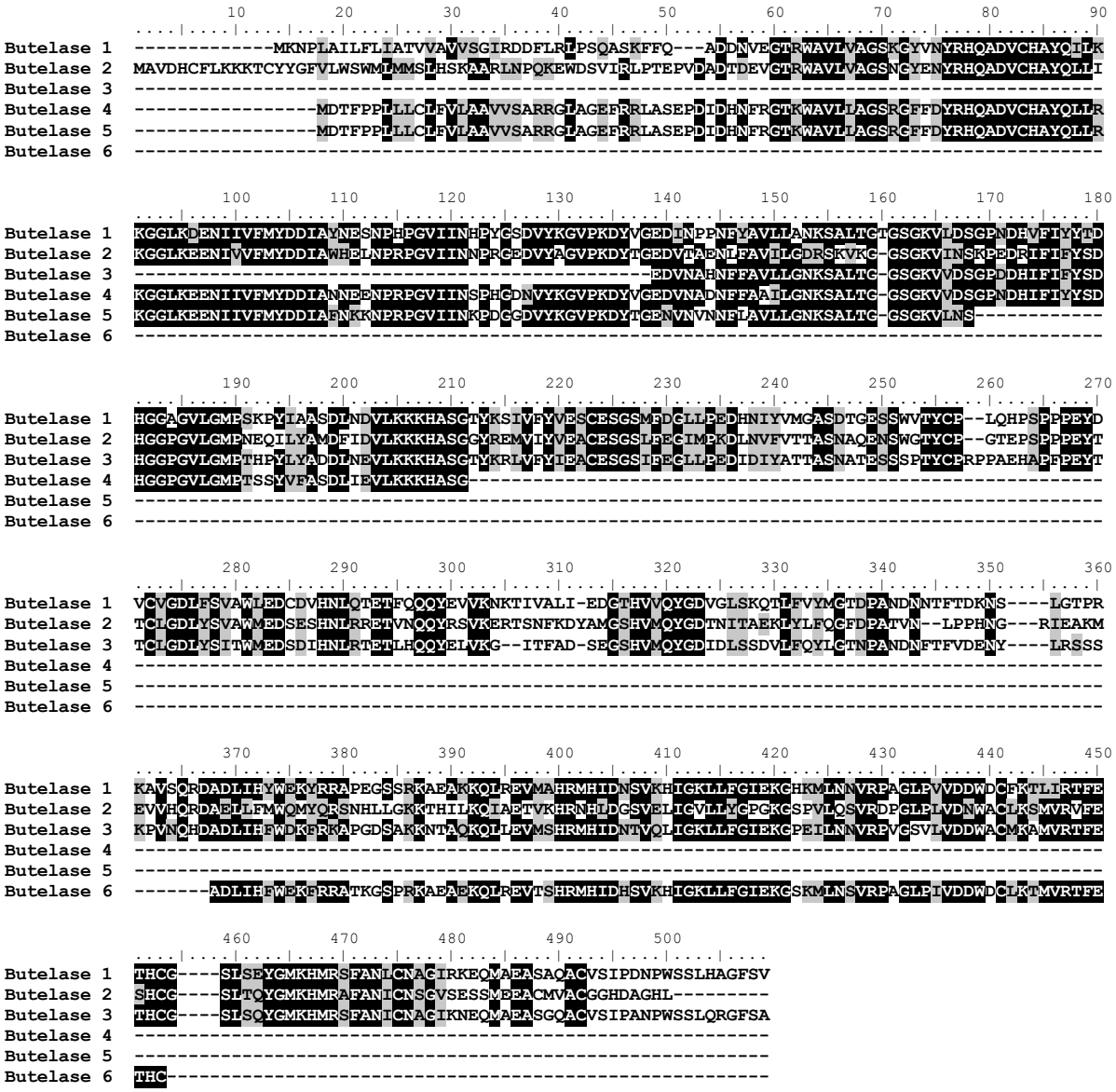
Supplementary Figure 4. 1D NMR characterization of kalata B1. (**a**) 1D NMR spectra of native kB1 (red) and butelase-cyclized kB1 (blue). (**b**) Overlay spectra of native and butelase-cyclized kB1. Peptides were dissolved in 95% H_2O/5%D_2O at pH 4.3. The spectra were recorded at 298 K.



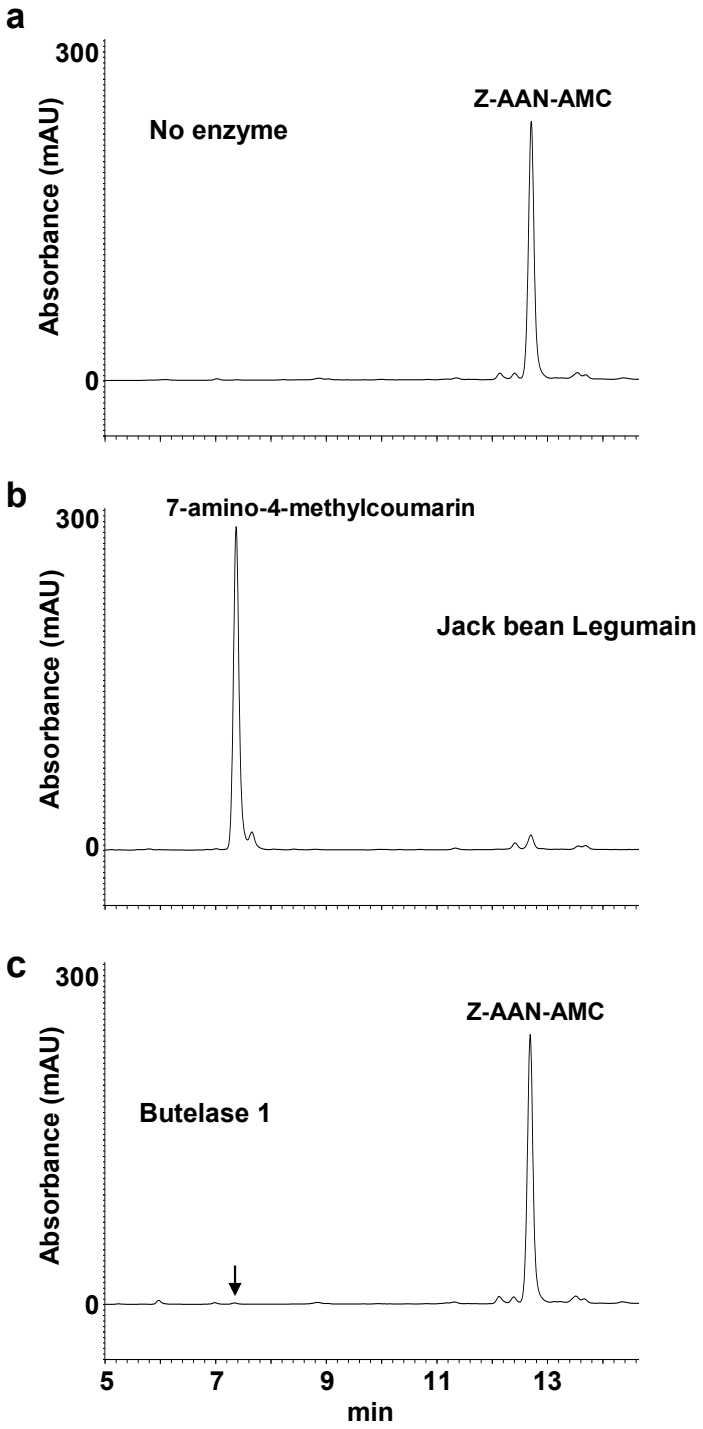
Supplementary Figure 5. SDS analysis of purified butelase 1. Proteins were visualized by silver staining. The left lane is the purified enzyme and the right lane is the protein ladder with molecular weights (kDa) indicated.



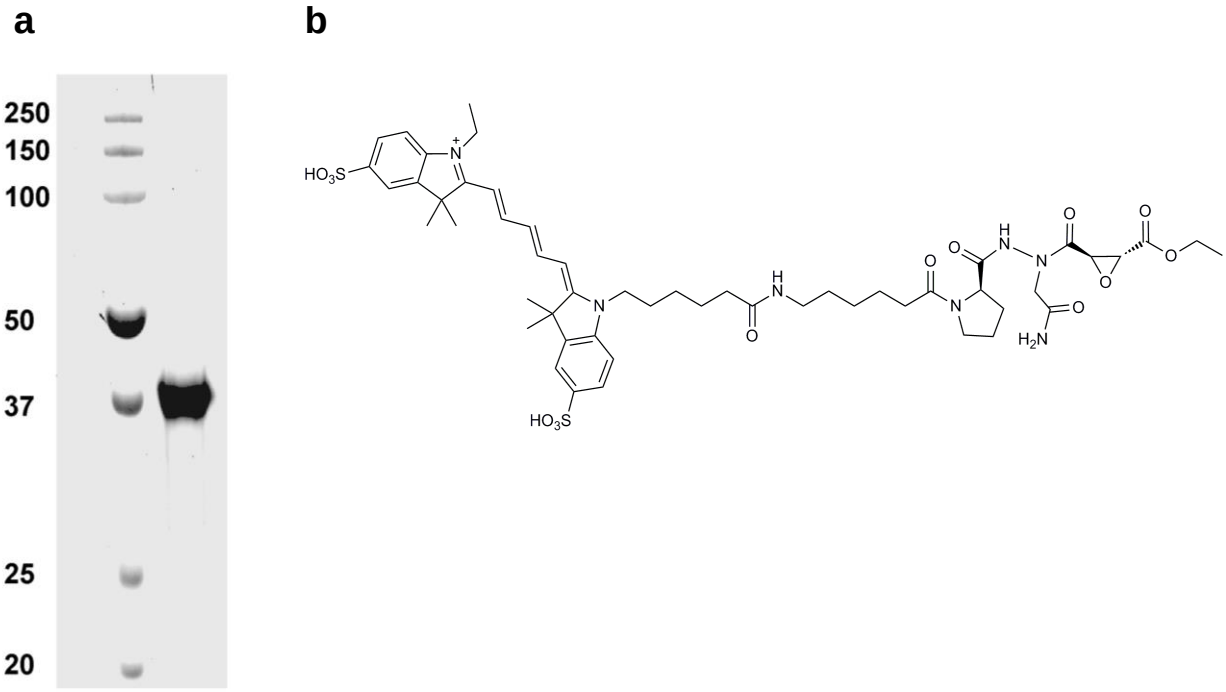
Supplementary Figure 6. Protein identification of butelase 1 by in-gel tryptic digestion. Five dominant tryptic fragments were sequenced by MS/MS and are shown at the top of the MS profile.



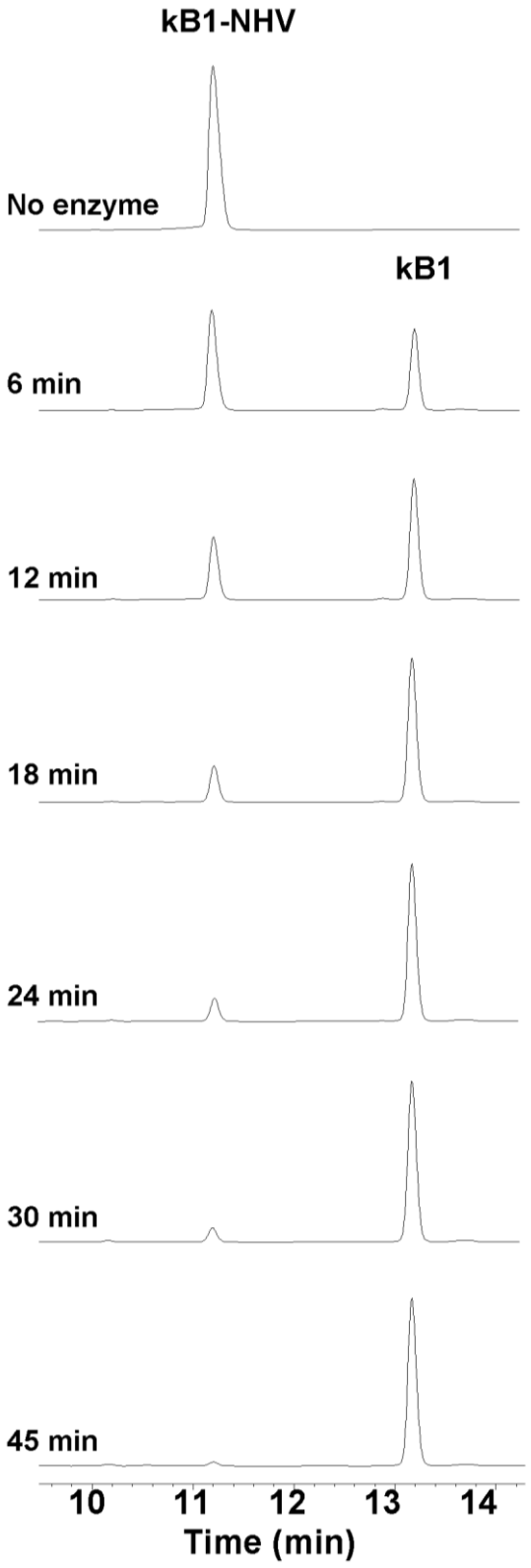
Supplementary Figure 7. Sequence alignment of butelase-1-like proteins from transcriptome of *C. ternatea*. Identical or similar residues were shaded. Only partial sequences were obtained for butelase 3-6.



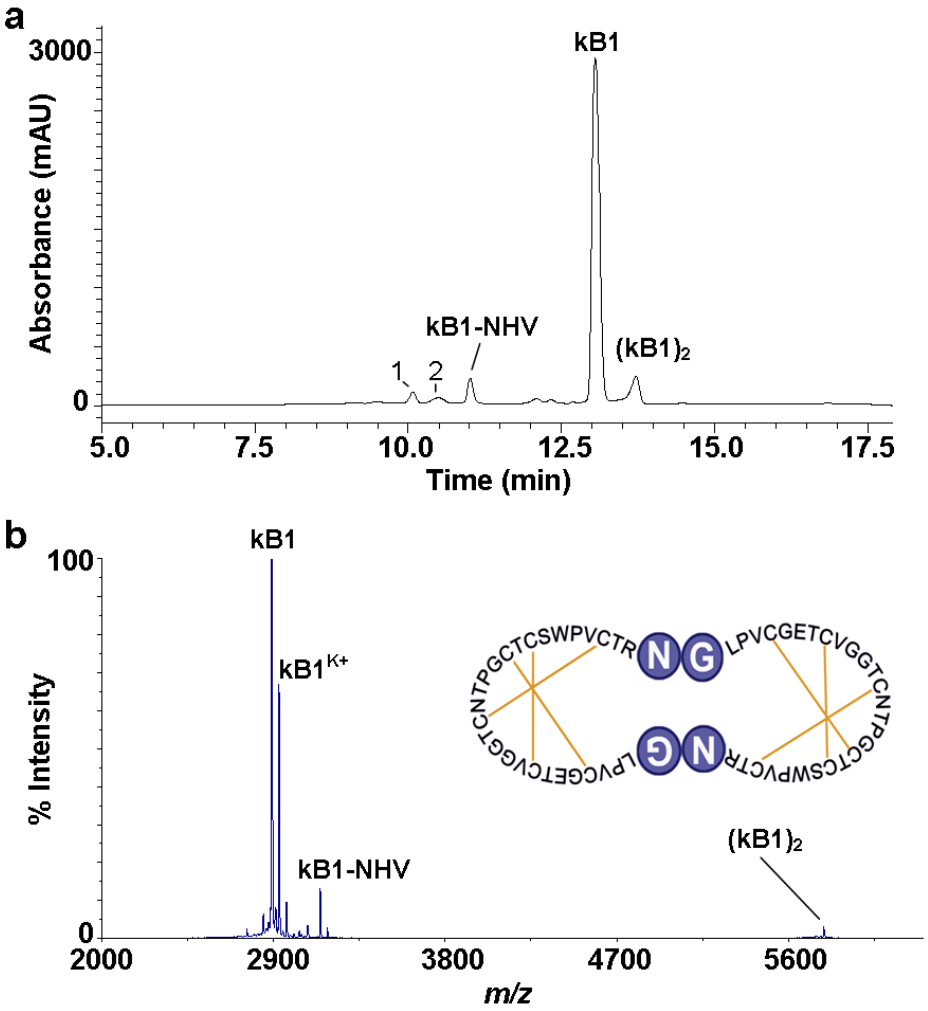
Supplementary Figure 8. HPLC profiles of Z-AAN-AMC treatment with jack bean legumain and butelase 1. (**a**) HPLC trace of Z-AAN-AMC as a control. (**b**) HPLC trace showing the hydrolyzed product 7-amino-4-methylcoumarin catalyzed by jack bean legumain. The assay was performed at 37 °C for 30 h in the presence of 8 μU jack bean legumain and 50 μM Z-AAN-AMC. (**c**) HPLC trace showing the effect of butelase 1 on Z-AAN-AMC. The assay was performed at 37 °C for 30 h in the presence of 0.125 μM butelase 1 and 50 μM Z-AAN-AMC. No significant hydrolyzed product (indicated by the arrow) was observed after 30 h incubation. The absorbance was monitored at a wavelength of 254 nm.



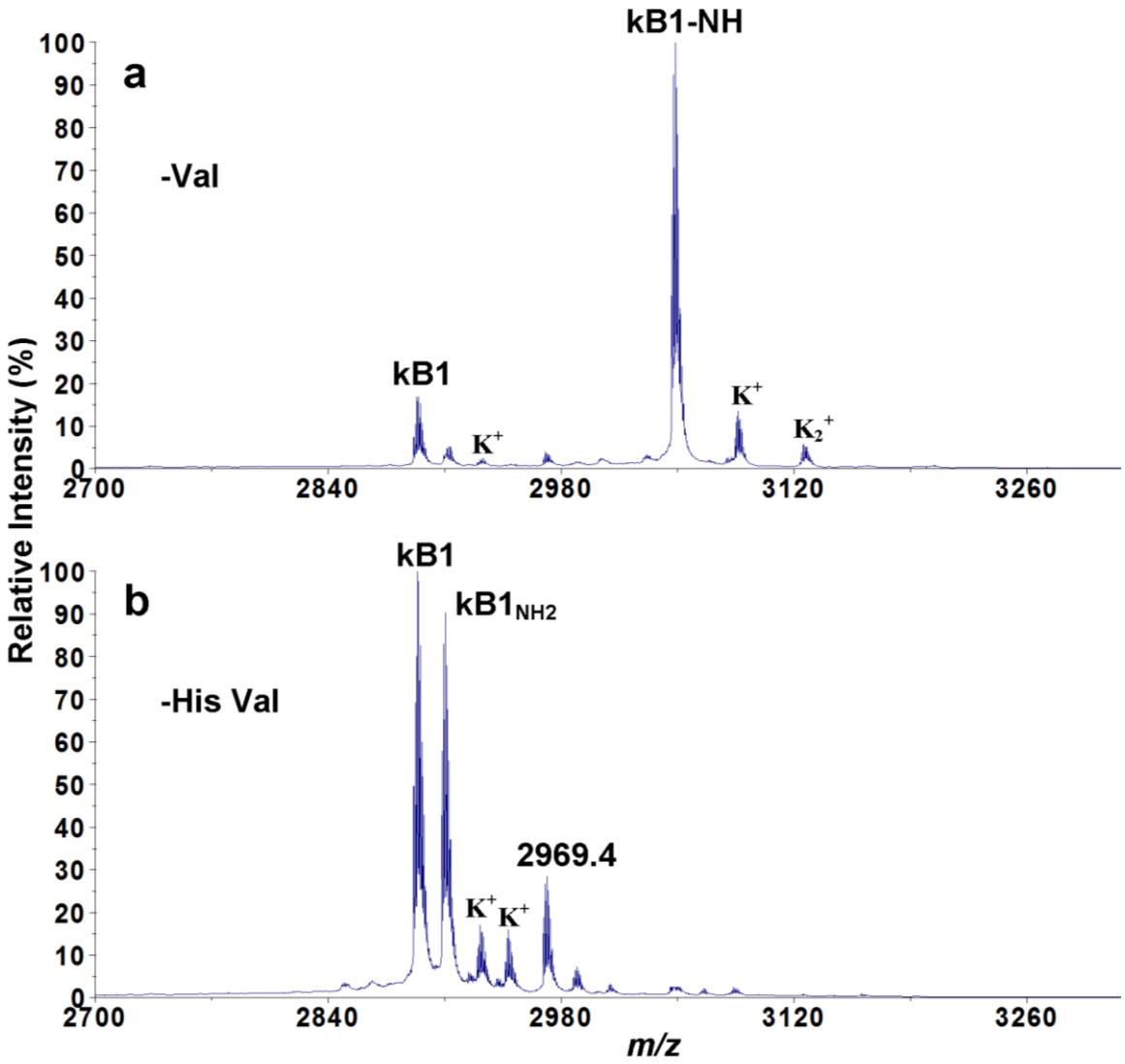
Supplementary Figure 9. Labeling of butelase 1 by the legumain-specific probe LP-1. (a) SDS analysis of purified butelase 1. Proteins were resolved by SDS-PAGE and visualized by a Typhoon scanner (GE Health Care) with a Cy5 filter. **(b)** Chemical structure of the legumain specific probe LP-1.



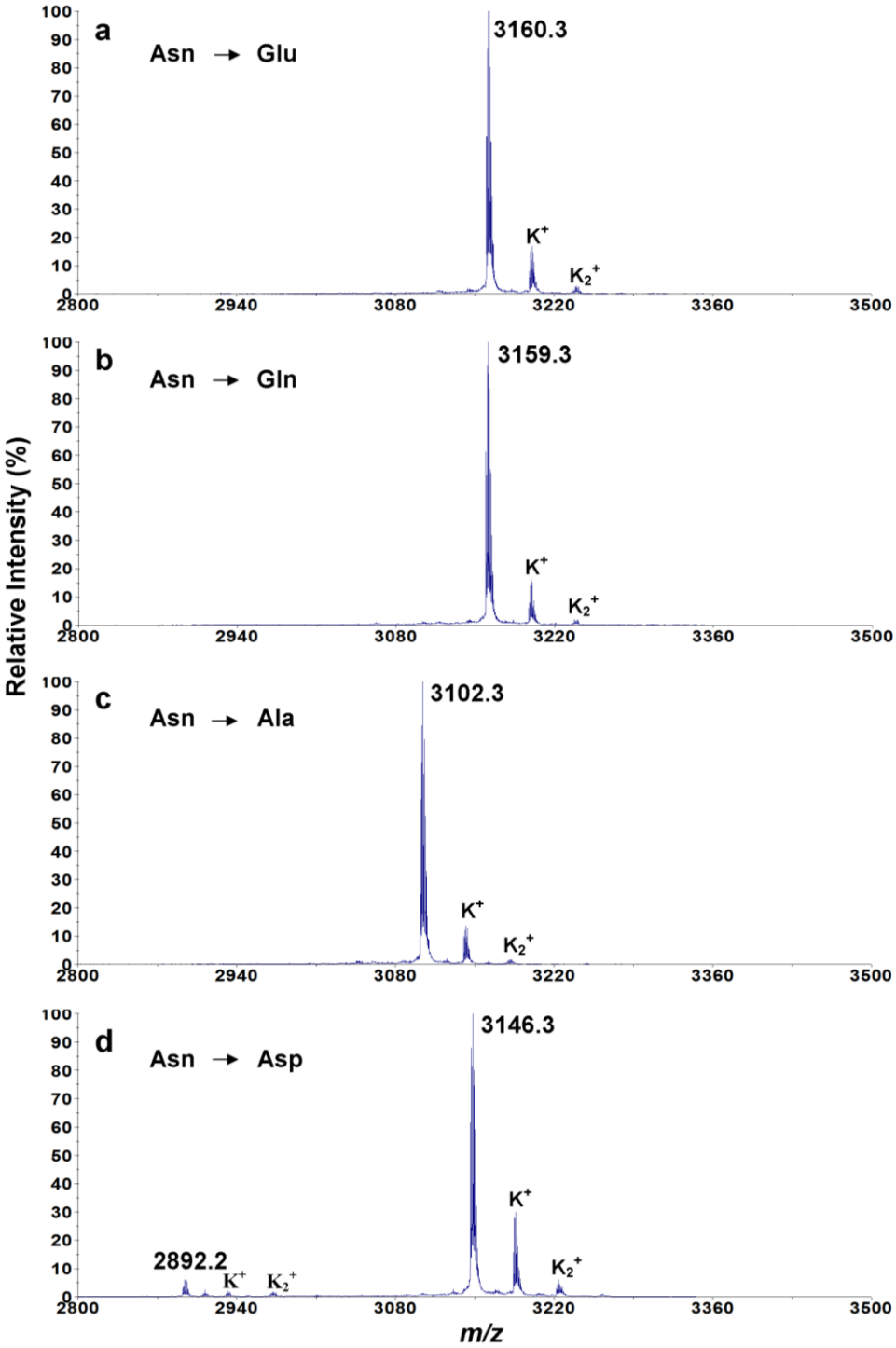
Supplementary Figure 10. RP-HPLC traces of the cyclization reaction over a time course of 45 min. The substrate kB1-NHV and the product kB1 are labeled. The assays were performed at 37 °C in the presence of 0.125 μM butelase 1 and 50 μM kB1-NHV. The absorbance was monitored at a wavelength of 220 nm.



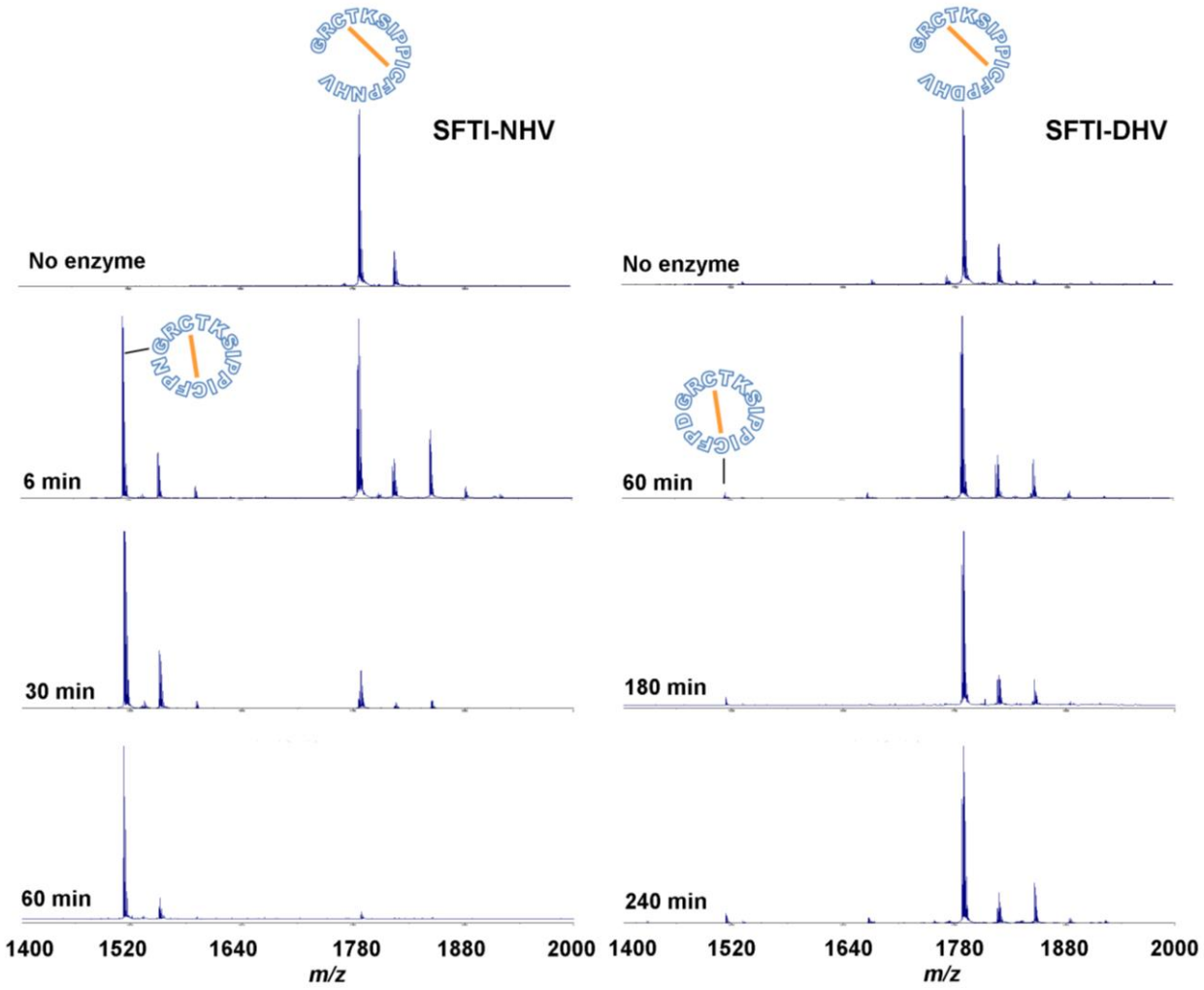
Supplementary Figure 11. Cyclodimer formation of kB1. (a) RP-HPLC profile illustrates the cyclodimer formation of kB1. The reaction was performed at 37 °C, for 4 h in the presence of 0.125 μM butelase 1 and 500 μM kB1-NHV. Peak 1 is the isomer of kB1 with the same *m/z* value of 2891 as native cyclic kB1. Peak 2 has the *m/z* value of 2893, which indicates the reduction of one disulfide bond of kB1. The substrate kB1-NHV and the products, kB1 and $(kB1)_2$ are labeled at peak apexes (**b**) MS profile illustrates the cyclodimer formation of kB1. Peak labeled with K^+ is the potassium adduct.



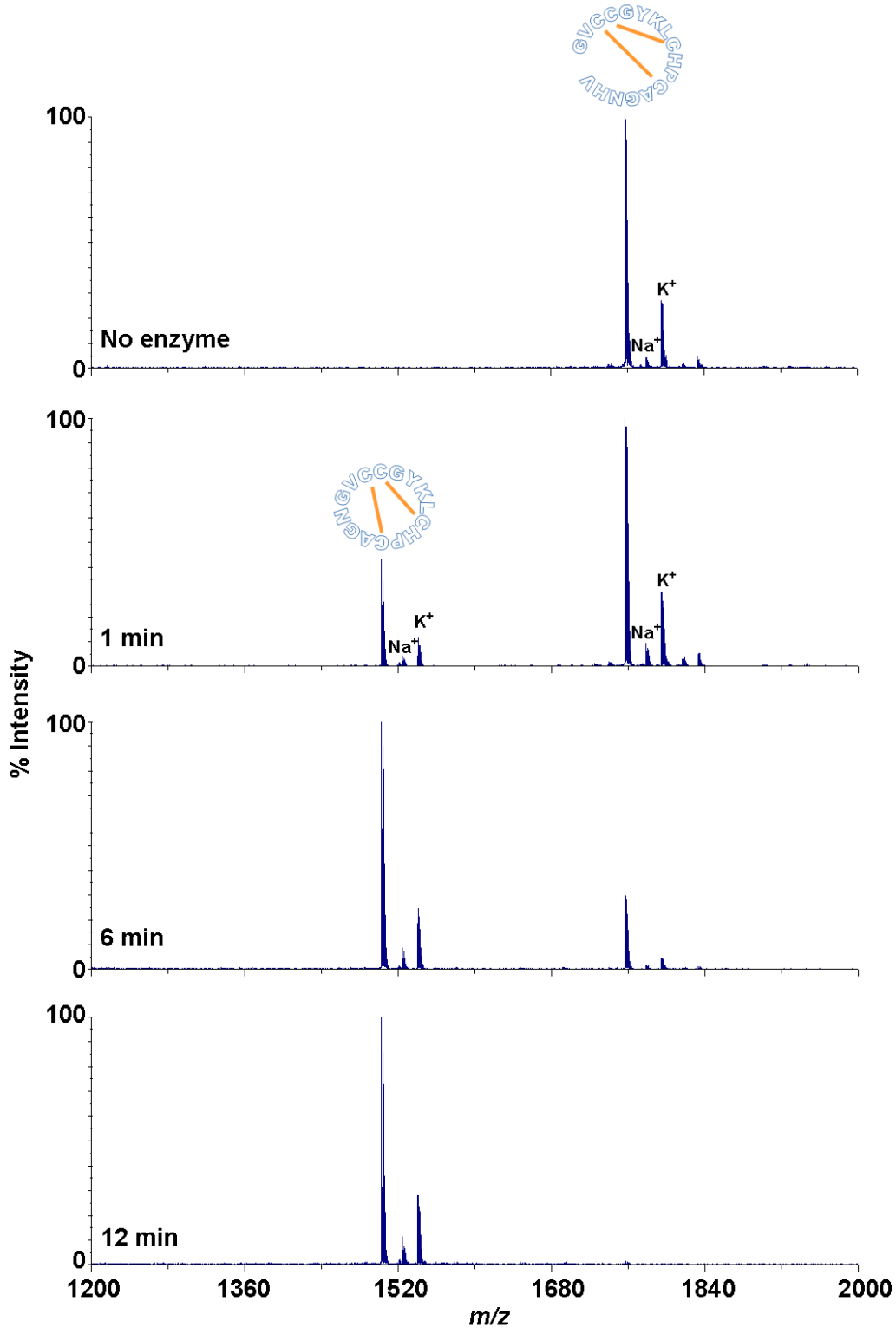
Supplementary Figure 12. MS profiles of kB1-NHV truncated analogs after treatment with butelase 1. Assays were performed at 37 °C for 30 h in the presence of 0.125 μM butelase 1 and 50 μM of each substrate. (**a**) MS profile of kB1-NH treated with butelase 1. (**b**) MS profile of $kB1_{NH2}$ treated with butelase 1. The kB1 product and truncated analogs, kB1-NH and $kB1_{NH2}$, were labeled at the peak apexes. The peak with an *m/z* value of 2969.4 represents a modified cyclic kB1 where one cysteine residue is *S*-alkylated with β-ME. Peaks labeled with K^+ and K_2^+ are ion adducts that correspond to the binding of one or two potassium ions, respectively.



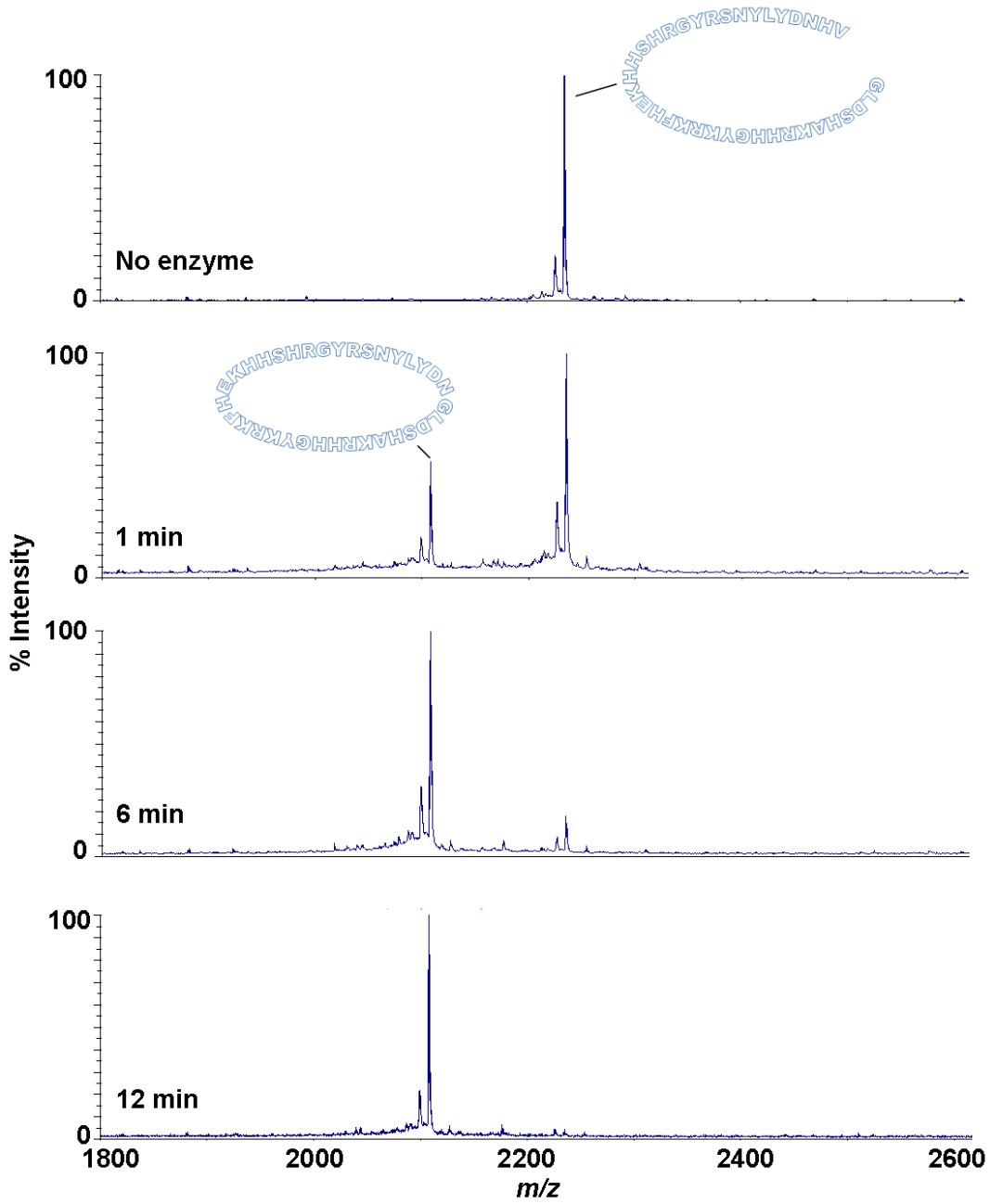
Supplementary Figure 13. MS profiles of the kB1-NHV analogs after treatment with butelase 1. (**a-d**) The conserved Asn at the P1 position was replaced by Glu, Gln, Ala or Asp residues in the kB1-NHV substrate. Assays were performed at 37 °C for 4 h in the presence of 0.125 μM butelase 1 and 50 μM of each substrate. The masses of the kB1-NHV analogs are labeled at the peak apexes. No cyclized products of kB1-EHV, kB1-QHV and kB1-AHV with the expected masses of 2906.2, 2905.2 and 2848.2, respectively, were observed after 4 h incubation. In the case of kB1-DHV, a cyclized product was observed with a *m/z* value of 2892. Peaks labeled with K^+ and K_2^+ are ion adducts that correspond to the binding of one or two potassium ions, respectively.



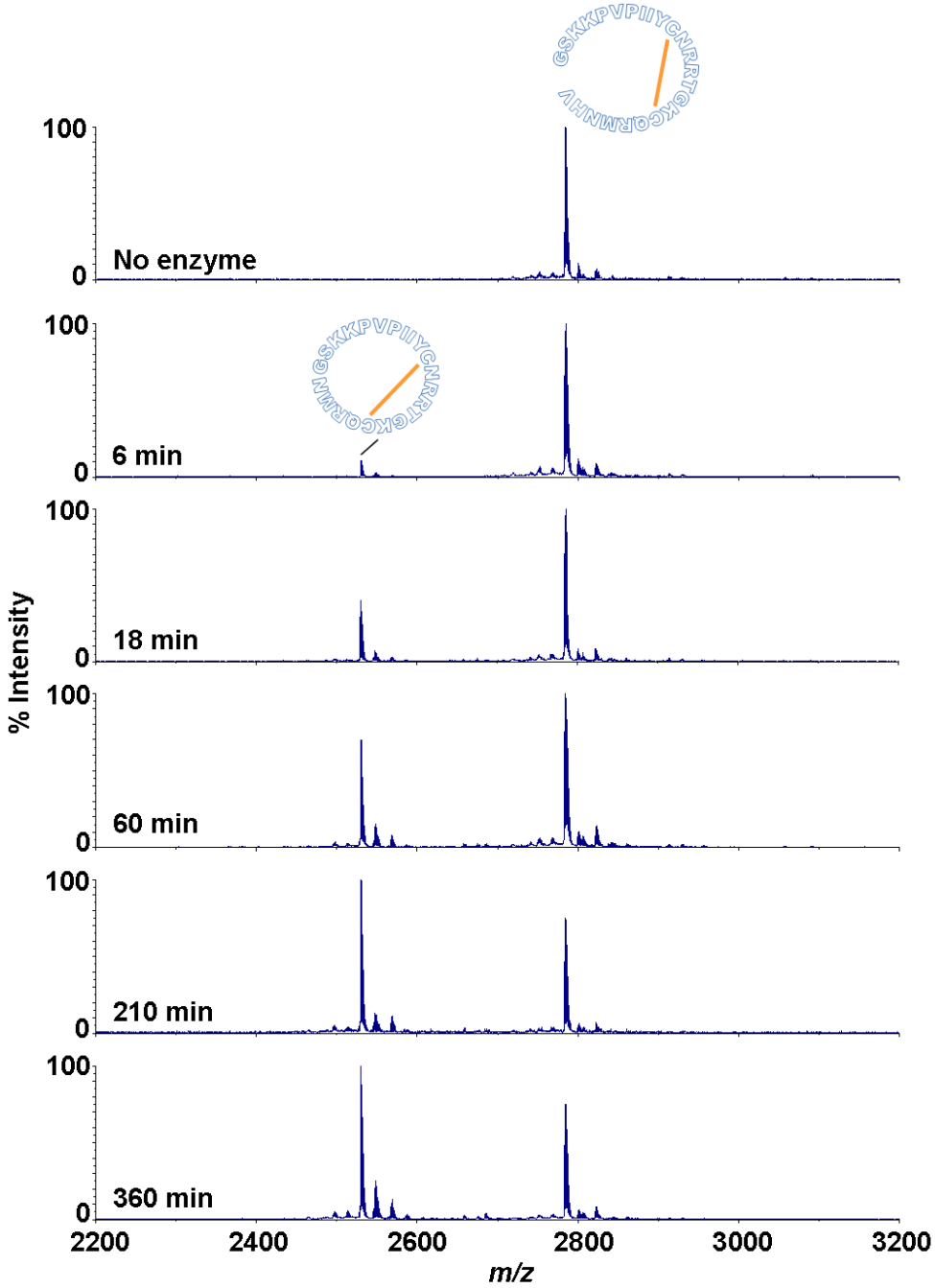
Supplementary Figure 14. Butelase-mediated cyclization of SFTI-NHV and SFTI-DHV. The cyclization reaction was performed at 37 °C in the presence of 0.125 μM butelase 1 and 50 μM peptide subtrate. The reaction was monitored by MS over a time course of 12 min.



Supplementary Figure 15. Butelase mediated cyclization of MrIA conotoxin. The cyclization reaction was performed at 37 °C in the presence of 0.125 μM butelase 1 and 50 μM peptide subtrate. The reaction was monitored by MS over a time course of 12 min. Peaks labeled with Na^+ and K^+ are the sodium and potassium adducts, respectively.



Supplementary Figure 16. Butelase mediated cyclization of histatin-3. The cyclization reaction was performed at 37 °C in the presence of 0.125 μM butelase 1 and 50 μM peptide substrate. The reaction was monitored by MS over a time course of 12 min. Peptides shown in the MS profiles carried a net charge of +2.



Supplementary Figure 17. Butelase mediated cyclization of thanatin. The cyclization reaction was performed at 37 °C in the presence of 0.125 μM butelase 1 and 50 μM peptide substrate. The reaction was monitored by MS over a time course of 360 min.