The macrocyclizing protease butelase 1 remains auto-catalytic and reveals the structural basis for ligase activity

Amy M. James[1,2,†,‡], Joel Haywood[1,2,†], Julie Leroux[1,2], Katarzyna Ignasiak[1], Alysha G. Elliott[3], Jason W. Schmidberger[1], Mark F. Fisher[1,2], Samuel G. Nonis[1,2], Ricarda Fenske[2], Charles S. Bond[1], and Joshua S. Mylne[1,2]

The University of Western Australia, [1] School of Molecular Sciences & [2] The ARC Centre of Excellence in Plant Energy Biology, 35 Stirling Highway, Crawley, Perth 6009, Australia; [3] The University of Queensland, Institute for Molecular Bioscience, St Lucia, Brisbane, QLD 4072, Australia

[†] equal author contribution

[‡] Current address: University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK

Plant asparaginyl endopeptidases (AEPs) have attracted attention for their macrocyclization reactions and are usually classified as cleavage or ligation specialists. Like most proteases, AEPs are expressed as zymogens that self-cleave to become active. Recently, Haywood et al. (2018) showed several AEPs were capable of macrocyclization given favorable pH conditions. One AEP purified directly from *Clitoria ternatea* seeds (butelase 1) is classified as a ligase rather than a protease, presenting an opportunity to test for loss of cleavage activity. Here, making recombinant butelase 1 and rescuing an *Arabidopsis thaliana* mutant lacking AEP, we show butelase 1 retains cleavage functions *in vitro* and *in vivo*. The *in vivo* rescue was incomplete, consistent with some trade-off for butelase 1 specialization toward macrocyclization. Its crystal structure showed an active site with only subtle differences from cleaving AEPs, suggesting the many differences in its peptide binding region are the source of its efficient macrocyclization.

23 # Introduction

24 Proteases are a large family of enzymes that are responsible for peptide bond breakage and are
25 highly conserved in their mode of catalysis, forming acyl-enzyme intermediates with their
26 substrates. The residues that comprise the catalytic triad (or dyad) of proteases are highly
27 conserved and an analysis of naturally existing variation among classes of proteases has defined
28 additional geometric constraints on protease active site architecture (Buller and Townsend,
29 2013).

30 Asparaginyl endopeptidases (AEPs) target Asn (N) and Asp (D) residues and cleave peptide
31 bonds. In plants, AEPs are also known as vacuolar processing enzymes as they process proteins
32 in protein storage vacuoles. In seeds, AEPs have roles in processing storage albumins and
33 globulins for which they perform mainly Asn-specific cleavages of seed storage protein
34 precursors. An *Arabidopsis thaliana* quadruple *aep* knockout line shows misprocessing of the
35 major seed storage globulins and albumins (Gruis et al., 2004; Kuroyanagi et al., 2005).

36 AEPs are expressed as zymogens and require auto-activation by cleavage of N- and C- terminal
37 pro-domains at low pH. Seed storage proteins and AEPs are synthesized on the rough
38 endoplasmic reticulum as precursors before being sorted into separate vesicles. Processing is
39 predicted to occur at the multi-vesicular bodies where both mature and immature seed storage
40 proteins have been shown to co-localize with AEP (Otegui et al., 2006). Ultimately, matured
41 seed storage proteins are trafficked to protein storage vacuoles where they remain until
42 catabolized during germination.

43 In addition to these cleavage functions, AEPs have attracted interest as enzymes that synthesize
44 cyclic peptides (reviewed by (James et al., 2018)). Several structural and molecular studies have
45 helped decipher the structural features necessary for macrocyclization. We have recently
46 shown through structural and biochemical studies of a sunflower AEP that subtle amino acid
47 changes around the active site were able to alter the ratio of cyclized and cleaved reaction
48 products (Haywood et al., 2018) and that all activated AEPs are likely able to perform a
49 macrocyclization reaction at a favorable pH. Mutating a bulky Cys residue located in the
50 substrate channel to a smaller Ala residue greatly increased the rates of macrocyclization

51 activity of an AEP from *Oldenlandia affinis* (Yang et al., 2017). Furthermore, recent sequence
52 analysis and molecular modeling has revealed a range of residues that together contribute to
53 ligation efficiency (Jackson et al., 2018; James et al., 2018; Zauner et al., 2018a).

54 Given the considerable interest in developing efficient macrocyclizing enzymes as tools for
55 producing highly stable bioactive molecules; butelase 1, the most efficient native cyclizing AEP
56 known to date, has attracted a great deal of attention (Nguyen et al., 2015; Nguyen et al.,
57 2016a; Nguyen et al., 2016b; Bi et al., 2017). However, the practicality of butelase 1 as a tool for
58 synthetic production of cyclic peptides has been limited by an inability to produce it as a
59 recombinant protein (Nguyen et al., 2014).

60 Although butelase 1 shares high sequence similarity to other AEPs, it exhibits a strong
61 preference for transpeptidation reactions over hydrolysis. Butelase 1 extracted from seeds of
62 the butterfly pea, *Clitoria ternatea* was proposed to have evolved to function as a ligase rather
63 than a protease and would only perform a cleavage reaction in the absence of a suitable
64 nucleophile for macrocyclization (Nguyen et al., 2014). In the presence of a C-terminal
65 tripeptide motif (Asx/His/Val) butelase 1 will accept most N-terminal amino acids for
66 transpeptidation and has been shown to circularize a number of non-native substrates of
67 various sequence compositions and sizes (Nguyen et al., 2014; Nguyen et al., 2015; Hemu et al.,
68 2016). However, when given a fluorogenic substrate recognized and cleaved by AEPs, butelase
69 1 produced no cleaved product indicating it was incapable of hydrolysis (Nguyen et al., 2014).

70 With butelase 1 suggested to have evolved to function as an Asx-specific ligase, one might
71 expect it to be unable to self-activate. It is conceivable that butelase 1 is activated by another
72 AEP in *trans* such as butelase 2, which has been shown to have only a cleavage function and to
73 lack cyclizing activity (Serra et al., 2016). To test this hypothesis, we produced recombinant
74 butelase 1 for biochemical studies. We also examined its biological activity *in vivo* by placing a
75 transgene encoding butelase 1 in an *A. thaliana* mutant line that lacks AEP activity. We found
76 recombinant butelase 1 readily self-activated at low pH and would also act as an endopeptidase
77 *in vivo*, partly rescuing the seed protein profile of the *aep* null line. A crystal structure we

acquired for butelase 1 showed only subtle differences between its active site and those of AEPs that favor cleavage reactions.

These findings suggest that, although an efficient macrocyclase, butelase 1 either has not yet fully specialized into a macrocyclase or will never fully specialize if the evolutionary constraint of self-processing cannot be overcome in natural systems. Structurally, we propose that differences local to the substrate binding region are the source of its ligase specialty.

Results

Crystal structure of butelase 1

AEPs are expressed as zymogens with a C-terminal pro-domain that 'caps' the active site of the core domain. Dissociation of this cap domain occurs upon a shift to a low pH environment where *in trans* self-cleavage at a flexible linker region occurs. Although previous attempts at producing recombinant butelase 1 failed (Nguyen et al., 2014), by using an *E. coli* strain optimized for recombinant proteins with disulfide bonds (Lobstein et al., 2012) we found we could successfully express and purify recombinant butelase 1 as we have done for four other plant AEPs (Bernath-Levin et al., 2015; Haywood et al., 2018). The recombinant protein possessed a 6-His tag at its N-terminus, followed by a Gly-Ser linker and 462 residues of butelase 1 from Ile21 to Val482 (**Figure 1—figure supplement 1**). This constitutes the full ORF-encoded butelase 1 (482 residues) minus its endoplasmic reticulum signal (20 residues). After purification by immobilized metal affinity chromatography at neutral pH, the purified protein possessed both a small N-terminal pro region and the much larger C-terminal cap that covers the active site.

Crystallization trials of butelase 1 in its inactive form yielded diffraction quality crystals at 3.1 Å resolution. The crystal structure was solved by molecular replacement yielding four molecules in the asymmetric unit each (**Figure 1A**). These molecules were arranged in a manner analogous to that previously reported for *A. thaliana* AEP3, with the asymmetric unit similarly consisting of two dimers formed by interactions between the α6 and α7 helices and stabilized by the C341 specificity loop. Each dimer interaction buries approximately 2500 $Å^2$ (Krissinel and Henrick,

105 2007). The AEP core domain of butelase 1 consists of six β-sheets surrounded by five major α-
106 helices, with several additional short α-helices and β-sheets in the linker regions (**Figure 1—**
107 **figure supplement 2**). The C-terminal pro-domain cap, shielding the active site residues Asn59,
108 His165, and Cys207, consists of five helices. The substrate binding site is flanked by two loops
109 between residues Trp236 to Pro245 and Ile289 to Ile293. The latter loop forms an extended,
110 flexible α5-β6 loop, varies in length among various AEPs (**Figure 1—figure supplement 2**) and
111 only modelled in two molecules of the butelase 1 asymmetric unit due to weak electron density
112 in the other chains.

113 Butelase 1 is structurally similar to AEPs from *Oldenlandia affinis* (OaAEP1, PDB ID: 5H0I), the
114 common sunflower *Helianthus annuus* (HaAEP1, PDB ID: 6AZT), and *A. thaliana* (AtAEP3, PDB
115 ID: 5NIJ) with an r.m.s.d. value of 1.1 Å, 0.9 Å, and 1.0 Å over 397, 271, and 422 alpha carbon
116 residues of the most complete chain (B) of butelase 1, respectively (Hasegawa and Holm, 2009).
117 The butelase 1 core domain structure has approximate dimensions of 53 Å × 48 Å × 40 Å and
118 within the core domain the active sites of butelase 1, OaAEP1, HaAEP1, and AtAEP3 show a high
119 degree of structural similarity (**Figure 1B**). Moreover, comparison between inactive and active
120 forms of AEPs reveal only subtle differences in catalytic residue orientations (**Figure 1—figure**
121 **supplement 3**). The catalytic residues in the active site of butelase 1, Asn59, His165, and
122 Cys207, align closely with the catalytic residues of OaAEP1 (Asn70, His175, and Cys217),
123 HaAEP1 (Asn73, His178, and Cys220), and AtAEP3 (Asn72, His177, and Cys219). Furthermore,
124 we chose to model a succinimide (SNN) residue at position 164 due to the preponderance of
125 this aspartimide at this position (N-terminally adjacent to the catalytic His165) in other AEP
126 crystal structures (Dall et al., 2015; Haywood et al., 2018; Zauner et al., 2018a), along with an
127 associated reduction in steric clashes when compared with the modelled alternative Asp164
128 residue (**Figure 1—figure supplement 4A**). In particular, the short 1.9 Å distance between the
129 OH group of Tyr162 and an Oδ from the side chain of Asp164, which is extended to 2.9 Å with
130 the O5 of SNN, was notable in guiding our decision to model residue 164 as SNN.

131 The substrate binding sites of butelase 1, on the other hand, showed differences in sequence
132 and structure compared to the aforementioned published plant AEP crystal structures (**Figure**
133 **1C**). All four AEPs have a non-polar Leu residue N-terminal to the start of the α4-helix (Leu258

134 in butelase 1, Leu268 in OaAEP1, Leu271 in HaAEP1, and Leu270 in AtAEP3) and a polar residue
135 at the start of the C341-loop (Thr238 in butelase 1, Tyr248 in OaAEP1, Thr251 in HaAEP1, and
136 Thr250 in AtAEP3). However, on one side of the entrance to the substrate binding pocket
137 butelase 1 has a bulky Trp residue and a non-polar Ile residue (Trp236 and Ile289). By contrast
138 HaAEP1 has a smaller Tyr residue and a polar Asn residue (Tyr249 and Asn303). OaAEP1 and
139 AtAEP3 both also have a Trp at the former position (Trp246 and Trp248, respectively) but differ
140 in the latter residue with OaAEP1 exhibiting a polar His residue (His300) and AtAEP3 exhibiting
141 a Pro residue (Pro302) that results in a more rigid loop that is shifted away from the binding
142 site. On the opposite side of the entrance to the substrate binding pocket the cyclization
143 specialist butelase 1 has a positively charged residue (His244), whereas the other AEPs possess
144 a negatively charged Glu residue at the same location. Overall, based on these structural
145 differences, the substrate binding region is likely to be responsible for the differences in
146 macrocyclization efficiencies among AEPs.

147 **Peptide processing by recombinant butelase 1**

148 To be catalytically active, AEPs are thought to undergo auto-catalytic cleavage at an Asx residue
149 *in trans*. However, if the Asx-specific ligase butelase 1 is incapable of cleavage, then it would be
150 unable to self-process. To test this, purified 6-His butelase $1_{[21-482]}$ with its N-terminal pro-region
151 and C-terminal cap domain was dialyzed at pH 4.0 and found to be capable of autocatalytic
152 cleavage, producing active protein with a mass approximately corresponding to removal of
153 both the N- and C-terminal propeptides (~38 kDa, **Figure 2A**). This was consistent with the mass
154 of native butelase 1 purified from seeds (Nguyen et al., 2014). This mass was further verified as
155 the active enzyme by incubating the activated protein with a fluorophore-labelled (BODIPY)
156 activity-based probe JOPD1, which is specific for AEP activity (Lu et al., 2015), and then running
157 the mixture on a gel before imaging it (**Figure 2B**). The ability to self-mature and bind JOPD1
158 demonstrated that recombinant butelase 1 is capable of performing a cleavage reaction. To
159 provide an opportunity for butelase 1 to perform either a cleavage reaction or its efficient
160 transpeptidation reaction, we incubated purified and self-activated recombinant butelase 1
161 with an acyclic sunflower trypsin inhibitor (SFTI) synthetic peptide, the seleno-Cys modified
162 SFTI(D14N)-GLDN, which can be processed into both cyclic and acyclic products by other plant

163 AEPs (Bernath-Levin et al., 2015). As predicted, butelase 1 produced cyclic SFTI(D14N) (**Figure**
164 **2C**). In addition to the cyclic product, we also detected a 1626 Da mass consistent with the
165 acyclic-SFTI(D14N) product that would arise from cleavage only (**Figure 2C**). No mass
166 corresponding to cyclic SFTI(D14N) was detected in a no enzyme control; peaks corresponding
167 to acyclic product were detected in the control that suggested there was some spontaneous
168 cleavage, however these peaks were barely distinguishable above the background (**Figure 2—**
169 **figure supplement 1**) and have been observed previously (Bernath-Levin et al., 2015).

170 As a further control for protease activity we produced recombinant HaAEP2, the second most
171 abundant *AEP* in sunflower seeds based on the number of RNA-seq reads that mapped to
172 sunflower *AEP* transcripts (Bernath-Levin et al., 2015). Only a partial *HaAEP2* sequence was
173 obtained from RNA-seq, however its full-length sequence was obtained by sequencing the
174 cDNA clone that was used to generate an expressed sequence tag found in GenBank
175 (DY926452.1). The construct for recombinant HaAEP2 encoded 440 residues of HaAEP2 with a
176 6-His and Gly-Ser linker in *lieu* of its ER signal (19 residues) and the predicted N-terminal
177 propeptide (21 residues) (**Figure 1—figure supplement 1**). HaAEP2, which was purified and
178 activated alongside butelase 1, processed the SFTI(D14N)-GLDN substrate exclusively to a 1626
179 Da mass with no 1608 Da mass for the cyclic product detectable (**Figure 2D**). A small amount of
180 a 1611 Da product was observed and this is thought to be due to deamidation of the Asn
181 residue to an acyl group. This mass has been observed previously in studies using the
182 SFTI(D14N)-GLDN substrate (Bernath-Levin et al., 2015). This demonstrated that both butelase
183 1 and HaAEP2 are capable of performing cleavage reactions. The mass detected for the acyclic-
184 SFTI(D14N) had a mass spectrum that indicated a sodium adduct formed with the cyclic product
185 causing an additional mass 22 Da larger than the protonated form. Sodium adducts of cyclic
186 masses have been observed in previous studies with OaAEP1 (Harris et al., 2015). The presence
187 of the sodium adduct was verified by desalting the sample using solid phase extraction and
188 seeing a reduction, particularly of the 1629 Da peak within the envelope (**Figure 2—figure**
189 **supplement 2**). The correct mass was also verified by liquid chromatography-mass
190 spectrometry (LC-MS), which desalted the sample while separating it (**Figure 2—figure**
191 **supplement 2**).

192 The sunflower AEP, HaAEP2, was also shown to be capable of self-maturation using the activity
193 based probe JOPD1; however, it showed heterogeneity in the size of its active form when
194 compared with butelase 1 (**Figure 2B**).

195 **Butelase 1 can partially rescue an *aep* null mutant plant**

196 To characterize the activity of butelase 1 *in vivo*, an open reading frame for it under the control
197 of a seed-specific promoter was expressed in an *A. thaliana* mutant lacking endogenous AEP
198 activity. For comparison, the open reading frames of *A. thaliana* AEP2, known be the major
199 seed storage protein processing AEP in *A. thaliana* (Shimada et al., 2003), and *HaAEP2* from
200 sunflower were similarly expressed in the *aep* null line. The seed storage protein profiles of
201 these various transgenic lines were examined by SDS-PAGE and compared to wild type Col-0 or
202 the untransformed *aep* null line. To measure the extent of processing, targeted proteomics was
203 used to quantify the level of tryptic fragments matching either correctly processed or
204 misprocessed seed storage proteins; these values were normalized using tryptic fragments that
205 lack AEP recognition sites, specifically SEED STORAGE ALBUMIN 1 (SESA1, At4g27140), SESA3
206 (At4g27160), and SESA4 (At4g27170), and the globulin CRUCIFERIN3 (CRU3, At4g28520). As
207 expected, the *AtAEP2* transgene fully rescued seed storage protein processing in the *aep* null
208 background (**Figure 3**). It is important to note that this full rescue by the *OLEOSIN:AtAEP2*
209 transgene demonstrates that the expression pattern of the *OLEOSIN* promoter is suitable and
210 that the *AEP* gene being intronless does not complicate full and proper rescue by an *AEP*
211 transgene. The heterologous expression of the sunflower AEP2, *HaAEP2*, similarly fully rescued
212 the phenotype of the *aep* null line with no discernible differences in the protein profiles of the
213 HaAEP2 in *aep* lines versus wild type (**Figure 3**). The analysis of transgenic HaAEP2 in *aep* was
214 performed twice with comparable results. In transgenic lines where the ORF for butelase 1 was
215 similarly placed under control of the *OLEOSIN* promoter, we observed only partial rescue
216 (**Figure 3**). Correctly processed α and β globulins were observed in SDS-PAGE gels; however,
217 several bands corresponding to misprocessed seed storage proteins were seen in extracts from
218 seeds of the *butelase 1* in *aep* line that match bands in untransformed *aep* null (**Figure 3**).

219 Gel images were consistent with the quantification provided by targeted proteomics where
220 *AtAEP2* and *HaAEP2* almost fully rescued all of seed storage protein processing in the *aep* null,

221 whereas *butelase 1* rescued 70-90% of CRU3 globulin processing and only 30-40% of albumin
222 processing. (**Figure 3, Table 2, Table 3**).

223 Discussion

224 Several asparaginyl endopeptidases with dual functions, capable of both endopeptidase and
225 transpeptidase reactions, have now been characterized from plants and animals (Bernath-Levin
226 et al., 2015; Dall et al., 2015; Harris et al., 2015). Butelase 1 is one of few examples of
227 asparaginyl endopeptidases that favor transpeptidation over cleavage. The initial study
228 describing butelase 1 activity against non-native substrates found butelase 1 had seemingly a
229 reduced capacity for cleavage activity and so was purported to have evolved to specialize as a
230 ligase (Nguyen et al., 2014). Our study demonstrates that butelase 1 can in fact perform both
231 reactions and will perform a cleavage reaction both *in vitro* and *in vivo*. Using pure,
232 recombinant protein, we demonstrated that butelase 1 was capable of auto-activation, a
233 process that requires cleavage activity. Furthermore, we demonstrated that in the presence of
234 a non-native substrate for macrocyclization, butelase 1 would still perform a cleavage reaction.
235 When expressed in an *A. thaliana* line lacking AEP, a *butelase 1* transgene could partially
236 recover seed storage processing, again an action requiring a cleavage function. Butelase 1
237 recovered processing of globulins as evidenced by SDS-PAGE analysis and targeted proteomics
238 of seed extracts; however, recovery of albumin processing was minimal with only a partial
239 reduction in the production of misprocessed albumins. Potentially, the endogenous proteases
240 responsible for misprocessing albumins when AEPs are absent are more efficient at cleavage
241 than butelase 1.

242 The crystal structure of the proenzyme form of butelase 1 bears a strikingly similar structure to
243 those of other AEPs, where an unstructured linker region between the core domain and the cap
244 could provide the flexibility required for self-cleavage of the cap domain upon activation.
245 Furthermore, the crystal structure lends further support for the cap domain forming a dimer
246 interaction surface as suggested previously (Yang et al., 2017; Zauner et al., 2018b). Notably, we
247 have chosen to model a succinimide residue adjacent to the catalytic Cys despite the crystal
248 structure being the pro-form of butelase 1. The electron density at residue 164 is not

249 conclusively illustrative of a SNN residue and quite conceivably the crystal structure could
250 contain a mixture of SNN/Asp at this location. However, we chose to model a SNN residue as
251 this residue is more congruous with the electron density and results in reduced steric clashes. In
252 light of this, our own investigation of the electron density from higher resolution AEP crystal
253 structures 4NOK and 5H0I from inactive mouse and inactive *O. affinis* (Zhao et al., 2014; Yang et
254 al., 2017) indicated that a SNN residue might also be present in these pro-AEP forms (**Figure 1—**
255 **figure supplement 4B-C**). Previous observations of SNN residues in the active sites of AEP
256 molecules were made on activated forms of the enzyme which had been subjected to a low pH
257 environment to induce activation. These observations of what appear to be SNN residues in
258 inactive forms of AEP purified at neutral pHs suggests that low pH activation of AEP is not
259 required to form a SNN residue in their active sites.

260 Recent structural studies of AEPs have identified subtle amino acid changes around the active
261 site and substrate binding domain that could influence their ability to perform a
262 transpeptidation reaction (Yang et al., 2017; Haywood et al., 2018). The crystal structure of
263 butelase 1 shows that residues surrounding the catalytic His are conserved in asparaginyl
264 endopeptidases (**Figure 1C**). The substrate binding domain shows less conservation between
265 the cyclizing specialists. In the absence of differences between active sites, it is likely that
266 differences in the substrate binding pockets are what define substrate specificity and catalytic
267 efficiency of AEPs, as has been suggested for the improved ligation efficiency of butelase 1
268 compared to OaAEP1 (Yang et al., 2017). In comparison to previously published crystal
269 structures, butelase 1 exhibits an extended, flexible α5-β6 loop that might become more rigid
270 upon substrate binding and play a role in guiding the N-terminus of a substrate for attack and
271 resolution of a thioacyl intermediate, resulting in peptide cyclization as recently suggested for
272 *A. thaliana* AEP3 (Zauner et al., 2018a). This loop has also recently been highlighted as a
273 'marker of ligase activity' with an extended loop proposed to be absent in ligases, and although
274 butelase 1 does not exhibit such a deletion it was shown to display a more hydrophobic loop
275 than its cleavage-favoring counterpart butelase 2 (Jackson et al., 2018). Moreover, Zauner *et al*.
276 have further substantiated the possible role of a hydrophobic pocket around a conserved Gly
277 residue adjacent to the catalytic His (Gly172 in butelase 1) by molecular dynamics simulations

278 (Zauner et al., 2018a). This region is suspected to play a role in cyclization through the binding
279 of the C-terminal tail of peptide substrates, which is postulated to protect thioacyl
280 intermediates from hydrolysis (Haywood et al., 2018; Zauner et al., 2018a). Further research
281 into these regions with recombinant proteins might discern the subtle amino acid differences
282 that dictate cyclization efficiency and substrate selectivity.

283 Changes to residues around the substrate binding domain are consistent with results from
284 directed evolution studies of proteins aimed at generating promiscuous functions, which show
285 that mutations leading to novel activity commonly do not impact the protein scaffold or the
286 catalytic residues (Aharoni et al., 2005; Khersonsky et al., 2006). As a result, trade-offs between
287 the new activity and the native activity are minimized (Aharoni et al., 2005). This appears to be
288 the case for dual-function AEPs which do not show any differences in their catalytic residues.
289 Numerous laboratory directed evolution studies have shown that mutations leading to
290 significant improvements in a promiscuous enzyme activity can occur with little effect on the
291 native activity (Aharoni et al., 2005; Khersonsky et al., 2006). However, here we have shown
292 that specializing towards a macrocyclizing activity has resulted in a reduction in cleavage
293 activity for butelase 1 *in vivo* as evidenced by failure to fully recover seed storage processing in
294 the *A. thaliana aep* null mutant.

295 Evolutionary theory suggests that pre-existing promiscuous enzyme activity is a prerequisite for
296 specialization of novel enzymes following gene duplication and many examples of naturally
297 existing enzymes with promiscuous function exist (Jensen, 1976; Jacob, 1977; Hughes, 1994;
298 Copley, 2014). As we have recently shown, AEPs have an inherent ability to perform ligation
299 under optimal conditions (Haywood et al., 2018). Furthermore, when the SFTI-1 precursor is
300 expressed in wild type *A. thaliana*, a plant lacking cyclic peptides, cyclic SFTI-1 is detectable
301 (Mylne et al., 2011). Therefore, given the high sequence identity of butelase 1 and butelase 2, it
302 is likely that their progenitor sequence possessed both activities.

303 Following gene duplication, mutations allow specialization. Although enzymes can possess dual
304 functions, single function enzymes are more common to allow for fine tuning of dosage control
305 and, in cases where there is a negative trade-off for dual functionality, to allow for

306 specialization (Khersonsky et al., 2006). In most cases, gene duplication alleviates the issue of
307 functional constraint as one copy is able to retain the original function and each copy can be
308 regulated independently. This is almost certainly the case for butelase 1 as there is a clear
309 negative trade-off in specializing as a ligase as evidenced by only partial rescue of seed storage
310 processing. However, due to the requirement for auto-catalysis, cleavage ability might be an
311 evolutionary constraint. By maintaining its cleavage ability, butelase 1 can be matured
312 independently from butelase 2. Thus, although gene duplication might have allowed for
313 specialization of butelase 1 into a transpeptidase and butelase 2 to remain an endopeptidase,
314 cleavage activity by butelase 1 might remain so that it can be independently expressed and
315 matured.

316 Alternatively, there might not be sufficient selection pressure for butelase 1 to completely lose
317 its cleavage function. As observed in directed evolution experiments, without an appropriate
318 selection pressure, complete removal of the original function is difficult to achieve (Tokuriki et
319 al., 2012; Kaltenbach and Tokuriki, 2014). In these cases, mutations leading to reductions of the
320 original activity, followed by optimization of the new activity, corresponded to only weak
321 improvements in the new activity. In nature, the selection pressure for removal of the original
322 activity might not exist once the new activity has been optimized, which is potentially the case
323 for maintenance of the cleavage activity of butelase 1.

324 In conclusion, this study has demonstrated that butelase 1 has not fully specialized into a
325 cyclizing enzyme. The structural differences between butelase 1 and other AEPs with weak
326 ligation ability suggest that plant AEPs are promiscuous ligases that can readily evolve a
327 macrocyclization capability. This is the first study that has demonstrated successful purification
328 of recombinant butelase 1, whose efficiency makes it useful for the synthetic production of
329 cyclic peptides. Recombinant butelase 1 will expedite future studies as it is now possible to
330 eschew laborious purification of native enzyme from *C. ternatea* seeds (Nguyen et al., 2014;
331 Nguyen et al., 2015; Nguyen et al., 2016b).

332 Mutagenesis of butelase 1 might enhance its ability to perform macrocyclization reactions and
333 would provide important insights into its evolution.

334 # Methods

335 **Recombinant butelase 1 and HaAEP2**

336 A synthetic DNA sequence encoding butelase 1 (Uniprot:A0A060D9Z7), including an N-terminal
337 six-His tag in *lieu* of its ER signal, was designed with codon optimization for *E. coli* (GeneArt) and
338 cloned into pQE30 (Qiagen). Similarly, a synthetic DNA sequence encoding HaAEP2 was
339 designed with codon optimization for *E. coli*, however the sequence included an N-terminal six-
340 His tag in *lieu* of its ER signal, as well as the predicted N-terminal propeptide. The pQE30-
341 HaAEP2 construct was transformed into the T7 SHuffle Express strain of *E. coli* (New England
342 Biolabs) for protein production. The construct encoding butelase 1 was co-transformed into the
343 same *E. coli* strain with the suppressor plasmid pREP4 (Qiagen) for protein production. Proteins
344 were expressed and purified as previously described (Haywood et al., 2018). The proteins were
345 activated by dialysis in activation buffer (20 mM sodium acetate pH 4.0, 5 mM dithiothreitol,
346 100 mM sodium chloride, 1 mM ethylenediaminetetraacetic acid) for 3 hours at room
347 temperature. Following activation, proteins were dialyzed into activity buffer (0.5 mM
348 dithiothreitol, 100 mM sodium chloride, 1 mM ethylenediaminetetraacetic acid, 20 mM 2-(N-
349 morpholino)ethanesulfonic acid pH 5.5) for 3 hours at room temperature. All proteins were
350 used in assays immediately following dialysis into activity buffer.

351 **Butelase 1 crystallization and X-ray data collection**

352 Butelase 1 was purified by size-exclusion chromatography (HiLoad 16/600 Superdex 200) in 50
353 mM Tris (pH 8.0), 50 mM sodium chloride and concentrated to 10-15 mg/mL. Crystal screening
354 was performed using the sitting-drop vapor-diffusion method with 80 µL of reservoir solution in
355 96-well Intelli-Plates at 20°C. Crystals of butelase 1 were obtained in 10% PEG 8000, 0.2 M
356 sodium chloride, and 0.1 M HEPES (pH 7.5). Single crystals were soaked in mother-liquor
357 supplemented with 20% glycerol as a cryoprotectant prior to being flash-frozen and stored in
358 liquid nitrogen. Data collection was performed at 100 K on the Australian MX1 beamline using a
359 wavelength of 0.9537 Å and diffraction data for crystals were collected to a resolution of 3.1 Å
360 (McPhillips et al., 2002).

361 **Structure determination and refinement**

362 Diffraction data were processed using XDS and scaled with AIMLESS from the CCP4 program
363 suite (Battye et al., 2011; Winn et al., 2011) in space group $P2_12_12_1$ with unit cell dimensions a =
364 71.36, b = 147.69, c = 183.33. A sequence alignment of butelase 1 and OaAEP1 was generated
365 using ClustalO and used to create a search model of butelase 1 based on the last common atom
366 of OaAEP1 using CHAINSAW. The structure of butelase 1 was solved by molecular replacement
367 with PHASER using 5H0I as a search model. Manual building and refinement was performed in
368 iterative cycles with Coot and REFMAC5 using the CCP4 program suite. Refinement was
369 performed using jelly-body refinement with sigma 0.02 and non-crystallographic symmetry
370 constraints applied globally. Structural analysis and validation were carried out with Coot and
371 MolProbity (Emsley and Cowtan, 2004; Chen et al., 2010). Crystallographic data and refinement
372 statistics are summarized in (**Table 4**) with Ramachandran plot values calculated from Coot.
373 Coordinates and structure factors were deposited into the Protein Data Bank (PDB) under
374 accession code 6DHI. Figures illustrating structures were generated using PyMol (Pei et al.,
375 2008; Schrodinger, 2010).

376 **BODIPY imaging of butelase 1 and HaAEP2**

377 For activity probe analysis 50 µL of AEP at 10 µg/mL was incubated with 0.5 µL of the BODIPY
378 probe JOPD1 at 100 µM (Lu et al., 2015) at room temperature overnight and protected from
379 light. The labelling reaction was stopped by the addition of 10 µL of 6x SDS-PAGE loading buffer
380 (62.5 mM Tris-HCl pH 6.8, 2.5% SDS, 0.002% Bromophenol Blue, 0.7135 M (5%) β-
381 mercaptoethanol, 10% glycerol). Proteins were separated using 4-12% Bis-Tris SDS-PAGE gels as
382 previously described (Lu et al., 2015; Haywood et al., 2018). Labelled proteins were visualized in
383 gel with excitation and emission wavelengths of 532 and 580 nm respectively, using a Typhoon
384 9500 (GE Healthcare).

385 **Cleavage and macrocyclization by recombinant AEPs**

386 Both cleavage and macrocyclization activities were determined by incubation of purified
387 recombinant protein with synthetic peptide, seleno-Cys SFTI(D14N)-GLDN. Activated
388 recombinant AEPs at a concentration of 1 mg/mL were incubated with 0.25 mM SFTI(D14N)-
389 GLDN possessing a diselenide bond and 25 µM native (with a disulfide bond) SFTI-1 as an

390 internal standard in activity buffer. Reactions were carried out at 37°C for 24 hours. Synthetic
391 peptides were incubated in the absence of enzyme as a negative control. Most AEPs, along with
392 butelase 1, have shown a preference for Asn hence the use of the SFTI(D14N)-GLDN peptide
393 which possesses an Asn residue in place of the native Asp residue (Nguyen et al., 2014;
394 Bernath-Levin et al., 2015). Products were determined by analysis with an UltraFlex III matrix-
395 assisted laser desorption/time of flight/time of flight (MALDI/TOF/TOF) mass spectrometer
396 (Bruker Daltonics) as described by Bernath-Levin *et al*. (2015).

397 The presence of a sodium adduct in an unexpectedly wide peak envelope of the acyclic-
398 SFTI(D14N) in the MALDI/TOF/TOF spectra was verified by LC-MS through comparison with the
399 HaAEP2 acyclic product. Reactions were dried and resuspended in 5% acetonitrile 0.1%
400 trifluoroacetic acid and then analyzed using a method described previously (Fisher et al., 2018).
401 Briefly, 2 µL of each sample was injected onto an EASY spray C18 column (75 µm x 150 mm, 3
402 µm particle size, 10 nm pores; Thermo Scientific) using a Dionex UltiMate 3000 Nano UHPLC
403 system (Thermo Scientific). A 40 min gradient elution was run from 5% acetonitrile to 95%
404 acetonitrile in water with 0.1% formic acid. The UHPLC system was coupled to an Orbitrap
405 Fusion mass spectrometer (Thermo Scientific) for ion detection following electrospray
406 ionization. To further verify the unexpected product, the reaction was desalted using a C18
407 column and the sample was reanalyzed by MALDI-MS.

408 **Seed-specific expression of butelase 1, HaAEP2, and AtAEP2 in an *A. thaliana* line lacking AEP**
409 The complete *HaAEP2* open reading frame was amplified by PCR from a sunflower cDNA
410 template using a forward primer JM603 containing an upstream *ClaI* site (5′-tta tcg atA TG AAT
411 CGC CAC CTG CTT ATT CTG-3′) and downstream, a reverse primer JM604 containing a *Sac*I site
412 (5′-atg agc tcG CTG ATT ATA GCA TTA CAT ACA G-3′). AtAEP2 was similarly cloned from *A.*
413 *thaliana* Columbia (Col-0) cDNA with primers containing the *Cla*I-compatible *Nar*I site upstream
414 and the same *Sac*I restriction enzyme recognition sequence downstream JM384 (5′-aag gcg ccA
415 TGG CTA AGT CTT GCT ATT TCA GAC C-3′) forward; JM385 (5′-aag agc tcA AAA AGT TAG ACG
416 ACC TTA TTG CTA C-3′) reverse. AtAEP2 encodes the AEP primarily responsible for seed storage
417 processing in *A. thaliana* (Shimada et al., 2003; Gruis et al., 2004). A synthetic sequence for

418 *butelase 1* with one silent change (T1071A) to prevent an internal *Sac*I site was designed with a
419 *Cla*I site to the 5' end and a *Sac*I site to the 3' end of the sequence. The synthetic *butelase 1*
420 sequence and those for *HaAEP2* and *AtAEP2* were digested with *Cla*I and *Sac*I. The *OLEOSIN*
421 promoter had been cloned previously from Columbia (Col-0) gDNA and was digested with *Xho*I
422 and *Cla*I (Mylne et al., 2011). The *CaMV35S* promoter was removed from pAOV (Mylne and
423 Botella, 1998) by digestion with *Xho*I and *Sac*I. Each *AEP* sequence, the *OLEOSIN* promoter and
424 pAOV were triple ligated as described in Mylne et al. (2011), to generate *Xho*I-*OLEOSIN*-*Cla*I-
425 *AEP*-*Sac*I-*nos* 3'-EcoRI in the pSLJ755I5 binary vector backbone which contains the *bar* gene
426 conferring resistance to glufosinate ammonium herbicide. The binary construct was tri-parental
427 mated into *Agrobacterium tumefaciens* strain LBA4404 and used to transform an *aep* null line
428 essentially as described by Bechtold *et al*. (1993). The *aep* null line contains mutations in all four
429 *AEP* genes, the alleles being *α-1 β-3 γ-1 δ-1* (Kuroyanagi et al., 2005). Seeds from transformed
430 plants (T_0/T_1) were sown on soil and sprayed twice with Basta® herbicide at a concentration of
431 400 mg/L glufosinate ammonium; once after germination when the cotyledons had expanded
432 and then again 3 days later. The T_2 generation of seeds was collected and sterilized by chlorine
433 gas then sown on MS agar plates containing 80 mg/L glufosinate ammonium. Seeds that
434 showed a 3:1 segregation typical of a single insertion event were collected and sown on soil.
435 Following germination, homozygous plants were selected with several sprays of herbicide.
436 Seeds that showed ~100% herbicide resistance were considered homozygous for a single
437 insertion. Seeds from homozygous plants were used for subsequent analyses. Three
438 independent replicates, representing independent transformation events, for plants expressing
439 *butelase 1* in an *aep* null background and two independent replicates for plants expressing
440 *AtAEP2* in an *aep* null background were analyzed. A single transgenic line was analyzed for
441 HaAEP2 in an *aep* null background.

442 **Butelase1 maturation of seed storage proteins**

443 Seed storage processing has been shown to be disrupted in the *aep* null background (Gruis et
444 al., 2004). We tested the ability of butelase1, AtAEP2 and HaAEP2 to recover seed storage
445 processing by expressing their encoding genes under the control of the seed specific promoter,
446 *OLEOSIN*, in an *aep* null background. Activity levels were compared to wild type Col-0 and *aep*

447 null Col-0. Approximately 50 mg of seeds were ground in liquid nitrogen with a micro-pestle. To
448 extract protein from seeds, 150 µL of extraction buffer (125 mM Tris-HCl pH 7.0, 7% sodium
449 dodecyl sulfate, 15 mM dithiothreitol, Roche cOmplete™ protease inhibitor [1 tablet/50 mL],
450 0.5% [w/v] polyvinylpyrrolidone-40) was added to ground tissue and samples were incubated
451 on ice for 20 min on a rocking platform. Soluble protein was separated by centrifuging samples
452 at 10,000 x g for 5 min at 4°C. To precipitate the protein the following steps were taken: to 25
453 µL of the supernatant (soluble protein fraction), 0.1 mL of chilled methanol was added and the
454 samples were vortexed; then, 25 µL of chilled chloroform was added followed again by
455 vortexing; finally, 75 µL of chilled water was added and the samples were vortexed.
456 Centrifugation at 14,000 x g for 10 minutes at 4°C moved all precipitated protein to the
457 interphase and the aqueous supernatant was discarded. Chilled methanol (0.1 mL) was added
458 and dissolved in the remaining solvent phase, but left the protein interphase as insoluble. To
459 pellet the protein, the sample was mixed by vortexing and centrifuged at 14,000 x g for 10
460 minutes at 4°C. The supernatant was discarded and the protein pellet was washed three times
461 with chilled acetone.. Following removal of the acetone, the pellet was resuspended in
462 resuspension buffer (7 M urea, 2 M thiourea, 50 mM ammonium bicarbonate, and 10 mM
463 dithiothreitol). Protein was quantified with Bradford reagent and samples were run on an SDS-
464 PAGE gel. For multiple reaction monitoring, samples were alkylated and digested with trypsin as
465 follows: Samples were incubated for 30 minutes at 60°C in 50% volume 100 mM ammonium
466 bicarbonate, 10 mM dithiothreitol. Iodoacetamide was added to a final concentration of 25 mM
467 and the reactions were incubated at room temperature in the dark for one hour. Trypsin was
468 added to the protein samples in a mass ratio of 1:25 and samples were incubated overnight at
469 37°C. The reaction was stopped with the addition of formic acid to a final concentration of 1%
470 (v/v). Samples were purified by solid phase extraction on Strata™-X 33 µm polymeric reversed
471 phase columns. The columns were conditioned with acetonitrile and equilibrated with 5% (v/v)
472 acetonitrile 0.1% (v/v) formic acid. After loading the columns with the sample, the column was
473 washed once with 1 mL 5% (v/v) acetonitrile 0.1% (v/v) formic acid and once with 1 mL 10%
474 (v/v) acetonitrile 0.1% (v/v) formic acid. The sample was eluted in 0.5 mL 85% (v/v) acetonitrile
475 0.1% (v/v) formic acid. The eluate was dried under vacuum and resuspended in 5% (v/v)
476 acetonitrile 0.1% (v/v) formic acid to a final concentration of ~1 µg/µL for multiple reaction

477 monitoring. Using an Agilent 1290 Infinity II LC system, 10 µL of each sample were loaded onto
478 an Agilent AdvanceBio Peptide Map column (2.1 × 250 mm, 2.7 µm particle size, P.N. 651750-
479 902), which was heated to 60°C. Peptides were eluted over a 15 minute gradient (0-15 min 3%
480 [v/v] acetonitrile 0.1% [v/v] formic acid to 45% [v/v] acetonitrile 0.1% [v/v] formic acid; 15-15.5
481 min 45% [v/v] acetonitrile 0.1% [v/v] formic acid to 100% [v/v] acetonitrile 0.1% [v/v] formic
482 acid; 15.5-16 min 100% [v/v] acetonitrile 0.1% [v/v] formic acid to 3% [v/v] acetonitrile 0.1%
483 [v/v] formic acid; 16-30 minutes 3% [v/v] acetonitrile 0.1% [v/v] formic acid) directly into the
484 Agilent 6495 Triple Quadrupole MS for detection. Transitions used for multiple reaction
485 monitoring are given in **Table 1**.

486 The following method was used for independent proteomic analysis of seed extracts of *A.*
487 *thaliana* transgenic lines expressing *HaAEP2* in an *aep* null background compared to wild type
488 and *aep* null seed extracts. Three *A. thaliana* albumins, SESA1, SESA3 and SESA4, were targeted.
489 Trypsin digested protein extracts were analyzed via targeted proteomics. LC-MS/MS data were
490 collected on the AB SCIEX Triple TOF 5600 nanospray coupled to a Shimadzu LC system and the
491 resulting mass spectra were analyzed qualitatively by querying SESA sequences using AB SCIEX
492 Proteinpilot Software 4.0.8085 to identify the individual tryptic fragments. Multiple reaction
493 monitoring was used for quantification of the identified SESA tryptic fragments on the Applied
494 Biosystems 4000 QTRAP nanospray mass spectrometer. Transitions used for multiple reaction
495 monitoring are given in **Table 2**.

496 Author contribution statement

497 J.S.M. conceived the study. A.G.E. and J.S.M. cloned *AEP* sequences. A.M.J., J.L. and A.G.E.
498 generated the *A. thaliana* transgenic lines. A.M.J. produced recombinant protein and tested
499 cleavage activity by MALDI analysis. M.F.F. performed LC-MS analysis. A.M.J. and R.F.
500 performed multiple reaction monitoring analysis of seed extracts. K.I. and S.G.N. performed
501 activity based probe assays. A.M.J. and J.H. produced butelase 1 for crystallization. J.H., J.W.S.
502 and C.S.B. solved the crystal structure. A.M.J., J.H., K.I. and J.S.M. wrote the manuscript with
503 contributions from all authors.

Declaration

The authors declare no financial and non-financial competing interests. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgements

The authors thank Dan Tawfik for critical insights during project design and Renier van der Hoorn for the BODIPY probe JOPD1. J.S.M. was supported by an Australian Research Council (ARC) Future Fellowship (FT120100013). A.M.J. and S.N. were supported by the Australian Government's Research Training Program. M.F. was supported by the Australian Government's Research Training Program and a Bruce and Betty Green Postgraduate Research Scholarship. This research was undertaken in part using the MX1 beamline at the Australian Synchrotron, part of the Australian Nuclear Science and Technology Organisation, and made use of the Australian Cancer Research Foundation detector. J.H. and this work were supported by ARC grant DP160100107 to J.S.M. and Dan Tawfik. *Helianthus annuus ASPARAGINYL ENDOPEPTIDASE 2 (HaAEP2)* was deposited in GenBank under accession MH115430.

520 ## Figure Captions

521 **Figure 1. Crystal structure of butelase 1 compared with macrocyclization specialist (OaAEP1)**
522 **and cleavage-favoring (HaAEP1 and AtAEP3) AEPs.** (**A**) Cartoon representation of inactive
523 butelase 1 with core domain and c-terminal cap domain highlighted in purple and gray,
524 respectively. Dashed line represents region of low electron density in linker region between cap
525 and core domain. (**B**) Butelase 1 core domain aligned with OaAEP1, HaAEP1, and AtAEP3 shows
526 high overall structural conservation among AEPs. Catalytic residues and residues in the
527 substrate binding region that may influence catalytic activity shown as sticks. (**C**) Expanded
528 surface and cartoon representations of the predicted catalytic region and substrate binding
529 region highlight subtle changes in residues (shown as sticks) around the substrate binding site
530 that likely contribute to differences in activity. Electron density maps (2 Fobs – Fcalc) for
531 highlighted residues of butelase 1 contoured at 1σ level.

532 **Figure 2. Butelase 1 is capable of performing a cleavage reaction.** (**A**) Butelase 1 is capable of
533 auto-activation at low pH. Protein was purified at pH 8.0 and subsequently activated at pH 4.0.
534 The proteins were then dialyzed into activity buffer at pH 5.5 prior to loading onto an SDS-PAGE
535 gel and performing the activity assay. Proteins were visualized with Coomassie stain. The
536 cleavage-favoring HaAEP2 was used as a control illustrating an auto-activation at low pH. (**B**) To
537 confirm which of the masses corresponded to the active protein, protein was incubated with
538 the BODIPY fluorescent probe, JOPD1, which is specific for AEP activity and visualized in SDS-
539 PAGE gel with excitation and emission wavelengths of 532 and 580 nm, respectively. (**C**)
540 Butelase 1 is capable of cleaving a synthetic peptide. Recombinant protein was incubated with
541 synthetic peptide for 24 hours at pH 5.5. Masses consistent with both cyclic SFTI(D14N) and
542 acyclic-SFTI (D14N) were produced following incubation of butelase 1 with the synthetic
543 substrate, consistent with butelase 1 performing both cyclizing and cleaving reactions. A
544 sodium adduct of cyclic SFTI(D14N) resulted in a peak envelope (*) containing both acyclic-
545 SFTI(D14N) and the +22 sodium adduct of cyclic SFTI(D14N). Evidence that these peaks are a
546 result of a sodium adduct is presented in **Figure 2—figure supplement 2**. (**D**) Recombinant

547 HaAEP2 showed only cleavage ability with the same peptide substrate under the same
548 conditions. The 1611 Da mass (†) is -15 Da from acyclic-SFTI(D14N) so is not cyclic.

549 **Figure 3. Butelase 1 retains its ability to cleave *in vivo*.** Expression of butelase 1 can partially
550 rescue seed storage processing in an *aep* null background. Total protein examined by SDS-PAGE
551 shows expression of *AtAEP2* (lane 2), *HaAEP2* (lane 3), or *butelase 1* (lane 4) under control of
552 the seed-specific *OLEOSIN* promoter can fully or partially rescue the seed protein profile of an
553 *A. thaliana aep* quadruple null mutant (lane 5). Consistent with complete rescue, lanes 1 (Col
554 0), 2 and 3 have comparable protein profiles indicating correct and complete processing of seed
555 storage proteins by AtAEP2 and HaAEP2. Contrastingly, lane 4 shows an intermediate protein
556 profile between wild type Col-0 (lane 1) and the *aep* quadruple null mutant (lane 5) indicating
557 only partial rescue of the phenotype. Misprocessing events indicated by asterisks can be seen in
558 both protein extracts from *aep* quadruple null mutant seeds and *butelase 1* expressed in *aep*.

559 **Figure 4. Targeted proteomics monitors *in vivo* activity of several AEPs in an *aep* null.** (**A**)
560 Sequences of seed storage proteins used for determining misprocessing events in *aep* mutant
561 plants. ER signal peptides are indicated in bold. The test tryptic fragment with AEP processing
562 site is underlined and the control tryptic fragment lacking AEP processing site is indicated with a
563 double underline. The AEP recognized residues are indicated with black boxes. (**B**)
564 Quantification by multiple reaction monitoring of incorrectly processed seed storage albumins
565 (SESA1, SESA3, SESA4) and globulin (CRU3). Two independent lines with *AtAEP2* and one
566 *HaAEP2* line expressed in an *aep* null background and three independent replicates for *butelase*
567 *1* expressed in an *aep* null background are shown. Quantities are given as ratios of test peptide
568 to control peptide and are normalized against *aep* null which is representative of no correct
569 AEP processing events (nd = not detected). Proteomic quantitation of rescue in the *HaAEP2* line
570 was performed in addition to and independently of the data presented in this figure (**Table 2**,
571 **Table 3**).

572

573 # Table

574 **Table 1.** Multiple reaction monitoring transitions used to quantify seed storage protein
575 misprocessing shown in **Figure 4B**.

Protein	Peptide Sequence	Precursor m/z	Charge	Product m/z	Charge	Fragment Ion
Test peptide	fragments					
SESA1	FEEDDATNPIGPK	716.8	2	726.4	1	y_7^+
SESA3	FEEDDASNPVGPR	716.8	2	525.3	1	y_5^+
SESA4	FDEDDASNPIGPIQK	823.4	2	752.5	1	y_7^+
CRU3	SPQGNGLEETICSMR	560.3	3	553.2	1	y_4^+
Control peptide	fragments					
SESA1	QEEPDCVCPTLK	738.3	2	458.3	1	y_4^+
SESA3	QEEPVCVCPTLK	730.4	2	877.4	1	y_7^+
SESA4	QEEPVCVCPTLR	744.4	2	486.3	1	y_4^+
CRU3	VFHLAGNNQQGGFGGSQQQQEQK	829.7	3	660.3	1	y_5^+

576

577

578 **Table 2.** Multiple reaction monitoring transitions used to quantify SEED STORAGE ALBUMIN
579 (SESA) misprocessing shown in **Table 3**. This analysis was performed independently of the data
580 presented in **Figure 4**. Multiple reaction monitoring properties include the precursor ion m/z
581 ($[M+2H]^{2+}$) selected in the first quadrupole (Q1), the fragment ion m/z ($[M+H]^{+}$)selected in the
582 third quadrupole (Q3), the dwell time (DT) for each multiple reaction monitoring transition (ms)
583 and the voltage (V) of collision energy (CE). In the test fragments the underlined N (Asn)
584 indicates the *in vivo* cleavage point for AEP. LSU, large subunit. SSU, small subunit.

Target protein	Peptide sequence of fragment	Sequence ion and charge	Q1	Q3	DT	CE
LSU control	peptide fragments					
SESA1	QEEPDCVCPTLK	y_7+	738.3	878.4	50	41.9
SESA3	QEEPVCVCPTLK	y_7+	730.3	878.4	50	41.5
SESA4	QEEPVCVCPTLR	y_9+	744.3	1101.6	50	42.2
Test peptide	fragments					
SSU N-terminal	cleavage site					
SESA1	FEEDDAT<u>N</u>PIGK	y_{10}+	668.3	1159.5	50	40.8
SESA3	FEEDDAS<u>N</u>PVGPR	y_7+	716.8	726.4	50	40.8
SESA4	FDEDDAS<u>N</u>PIGPIQK	y_9+	823.4	953.5	50	46.2
LSU N-terminal	cleavage site/s					
SESA4	GGGPSLDDEFDMEDDIE<u>**N**</u>PQR	y_7+	1168.5	871.4	50	44.2

585
586

587 **Table 3.** Quantification of SESA tryptic fragments by multiple reaction monitoring. This analysis
588 was performed independently of the data presented in **Figure 4**. The absence of the test SESA
589 fragments in WT and *OLEOSIN:HaAEP2* in *aep* are indicated by a (-). In the test fragments the
590 underlined N (Asn) indicates the *in vivo* cleavage point for AEP. LSU, large subunit. SSU, small
591 subunit.

Target protein	Peptide sequence of fragment	Sequence ion and charge	*aep*	WT	HaAEP2 in *aep*
LSU Control	peptide fragments		Peak area	(counts)	
SESA1	QEEPDCVCPTLK	y_7+	9.70e5	1.97e5	7.52e5
SESA3	QEEPVCVCPTLK	y_7+	1.87e7	1.13e5	2.06e6
SESA4	QEEPVCVCPTLR	y_9+	3.69e7	5.6e5	1.99e6
Test peptide	fragments				
SSU N-terminal	cleavage site		Peak area	(counts)	
SESA1	FEEDDATNPIGK	y_{10}+	3.28e7	-	-
SESA3	FEEDDASNPVGPR	y_7+	3.45e7	-	-
SESA4	FDEDDASNPIGPIQK	y_9+	3.50e6	-	-
LSU N-terminal	cleavage site/s		Peak Area	(counts)	
SESA4	GGGPSLDDEFDMEDDIE<u>N</u>PQR	y_7+	3.92e6	-	-

592
593

594 **Table 4.** Summary of crystallographic data and refinement statistics. Values in parentheses are
595 for the highest resolution shell.

Data collection	596
Space group	$P2_12_12_1$
Unit cell dimensions	
a, b, c (Å)	71.36, 147.69, 183.33
α, β, γ (°)	90.00, 90.00, 90.00
Wavelength	0.9537
Resolution (Å)	3.1
R_{merge} (%)	23.7 (133.6)
$I/\sigma I$	8.2 (1.6)
Completeness (%)	100 (100)
Redundancy	6.8 (7.0)
$CC_{1/2}$	0.991 (0.648)
Refinement	
Resolution (Å)	49.42 – 3.1
No. reflections	34123
R_{work}/R_{free}	25.5/29.7
No. Atoms	13188
Protein	13162
Water	26
Average Wilson *B* (%)	68.8
R.m.s deviations:	
Bond lengths (Å)	0.01
Bond angles (°)	1.418
Ramachandran analysis	
Favored (%)	93.83
Allowed (%)	5.99
Outliers (%)	0.18

597 ## Figure Supplements

598 **Figure 1—figure supplement 1 |** Alignment of the full length native sequences of butelase 1
599 and HaAEP2 and the synthetic sequences designed for generation of recombinant protein. ER-
600 signals are indicated in gray. N-terminal propeptides are indicated in white font with black
601 highlight. The six-His tag are indicated in bold. Asterisks indicate sites of auto-cleavage. The
602 conserved residues of the catalytic triad are indicated with black boxes.

603 **Figure 1—figure supplement 2 |** PROMALS3D protein sequence and structure alignment of
604 AEPs for which protein structure is known. Sequence alignment generated with PDB files and
605 sequences of AEPs from *C. ternatea* butelase 1 (Uniprot:A0A060D9Z7 PDB: 6DHI), *O. affinis*
606 (Uniprot: A0A0N9JZ32 PDB: 5H0I) *A. thaliana* (Uniprot: Q39119 PDB: 5NIJ) *H. annuus* (Uniprot:
607 A0A0G2RI59 PDB: 6AZT). Identical residues shown with white text and red background, similar
608 residues shown with red text. Conserved secondary structure β-sheets and α-helices shown
609 above sequences with green arrows and green helices, respectively, and with major β-sheets
610 and α-helices labelled. Catalytic triad residues highlighted with blue stars. C341-loop region
611 highlighted with black bar above sequence. Regions of amino acid diversity in substrate binding
612 site highlighted with orange bar above sequence (Pei et al., 2008).

613 **Figure 1—figure supplement 3 |** Comparison of available crystal structures of active and
614 inactive forms of AEPs catalytic core domain. Cartoon representation of core domains with
615 residues thought to play a role in catalytic activity shown as sticks. (**A**) *A. thaliana* AEP3, active
616 yellow with bound chloromethylketone inhibitor shown as orange sticks (PDB: 5OBT), inactive
617 gray (PDB: 5NIJ). (**B**) Human AEP, active pink with bound chloromethylketone inhibitor shown
618 as orange sticks (PDB: 4AWB), inactive gray (PDB: 4FGU). (**C**) Mouse AEP, active brown (PDB:
619 4NOJ), inactive gray (PDB: 4NOK).

620 **Figure 1—figure supplement 4 |** Asp residues modelled as SNN. (**A**) Comparison of alternative
621 models for butelase 1 residue 164. Asp164 was chosen to be modelled as a succinimide residue
622 (SNN) in all four chains of the butelase 1 PDB as the modelled SNN residue was more congruous
623 with the electron density and showed reduced steric clashes when compared with the
624 modelled Asp164 residue. Electron density maps (2 F_{obs} – F_{calc}) contoured at 1 σ level. Contacts

625 to the side chain of residue 164 of ≤ 3.5 Å and ≤ 2.0 Å illustrated with yellow and green dashed
626 lines, respectively. Also shown are examinations of electron density for the equivalent Asp
627 residue (left image) in higher resolution structures for inactive AEPs from (**B**) mouse
628 (PDB:4NOK) and (**C**) *O. affinis* (PDB: 5H0I) with additional omit maps (F_{obs} – F_{calc}) contoured at 3
629 σ level (green). Models of SNN residues are presented in the images on the right.

630 **Figure 2—figure supplement 1 |** MALDI/TOF spectra of processing of SFTI(D14N)-GLDN by
631 HaAEP2 and butelase 1. SFTI(D14N)-GLDN was incubated under the same conditions without
632 any enzyme present as a negative control. In each condition a mass of 2027 Da corresponding
633 to the substrate, seleno-Cys modified SFTI(D14N)-GLDN, is detected indicating incomplete
634 processing after 24 hours. A mass of 1626 Da corresponding to acyclic SFTI(D14N) and a mass of
635 1608 Da corresponding to cyclic SFTI(D14N) were detected when the substrate was incubated
636 with butelase 1, whereas only a mass corresponding to acyclic SFTI(D14N) was detected when
637 the substrate was incubated with HaAEP2. No mass corresponding to cyclic SFTI(D14N) was
638 detected in the negative control. A weak signal could be detected for acyclic SFTI(D14N) in the
639 negative control, but was negligible compared to the test samples. Native cyclic SFTI was
640 included as an internal control with a mass of 1515 Da.

641 **Figure 2—figure supplement 2 |** Sodium adduct testing. (**A**) Using a desalting column removes
642 the peaks caused by the sodium adduct of cyclic SFTI(D14N) observed in MALDI-TOF MS traces
643 with butelase 1. A sodium adduct of cyclic SFTI(D14N) resulted in a peak envelope containing
644 both acyclic-SFTI(D14N) and the +22 sodium adduct of cyclic SFTI(D14N); however, following
645 desalting the peak at 1628.6 corresponding to the sodium adduct is reduced. (**B**) Similarly, the
646 expected m/z for the 3+ ion of acyclic SFTI(D14N) is observed in LC/MS due to the salt being
647 removed during separation of the peptides by liquid chromatography resulting in the expected
648 peak envelope.

649

650 ## References

651 **Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S.M., Roodveldt, C., and Tawfik, D.S.** (2005).
652 The 'evolvability' of promiscuous protein functions. Nat. Genet. **37,** 73-76.
653 **Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W.** (2011).
654 iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.
655 Acta Crystallogr D Biol Crystallogr **67,** 271-281.
656 **Bechtold, N., Ellis, J., and Pelletier, G.** (1993). *In planta Agrobacterium*-mediated gene transfer
657 by infiltration of adult *Arabidopsis thaliana* plants. Comptes rendus de l'Académie des
658 sciences. Série III, Sciences de la vie **316,** 1194-1199.
659 **Bernath-Levin, K., Nelson, C., Elliott, A.G., Jayasena, A.S., Millar, A.H., Craik, D.J., and Mylne,**
660 **J.S.** (2015). Peptide macrocyclization by a bifunctional endoprotease. Chem. Biol. **22,**
661 571-582.
662 **Bi, X., Yin, J., Nguyen Giang, K.T., Rao, C., Halim Nurashikin Bte, A., Hemu, X., Tam James, P.,**
663 **and Liu, C.-F.** (2017). Enzymatic engineering of live bacterial cell surfaces using butelase
664 1. Angew. Chem. Int. Ed. **56,** 7822-7825.
665 **Buller, A.R., and Townsend, C.A.** (2013). Intrinsic evolutionary constraints on protease
666 structure, enzyme acylation, and the identity of the catalytic triad. Proc. Natl. Acad. Sci.
667 U. S. A. **110,** E653-661.
668 **Chen, V.B., Arendall, W.B., III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J.,**
669 **Murray, L.W., Richardson, J.S., and Richardson, D.C.** (2010). MolProbity: all-atom
670 structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D Biol.
671 Crystallogr. **66,** 12-21.
672 **Copley, S.D.** (2014). An evolutionary perspective on protein moonlighting. Biochem. Soc. Trans.
673 **42,** 1684-1691.
674 **Dall, E., Fegg, J.C., Briza, P., and Brandstetter, H.** (2015). Structure and mechanism of an
675 aspartimide-dependent peptide ligase in human legumain. Angew. Chem. Int. Ed. **54,**
676 2917-2921.
677 **Emsley, P., and Cowtan, K.** (2004). Coot: model-building tools for molecular graphics. Acta
678 Crystallographica Section D **60,** 2126-2132.
679 **Fisher, M.F., Zhang, J., Taylor, N.L., Howard, M.J., Berkowitz, O., Debowski, A.W., Behsaz, B.,**
680 **Whelan, J., Pevzner, P.A., and Mylne, J.S.** (2018). A family of small, cyclic peptides
681 buried in preproalbumin since the Eocene epoch. Plant Direct **2,** 1-17.
682 **Gruis, D., Schulze, J., and Jung, R.** (2004). Storage protein accumulation in the absence of the
683 vacuolar processing enzyme family of cysteine proteases. Plant Cell **16,** 270-290.
684 **Harris, K.S., Durek, T., Kaas, Q., Poth, A.G., Gilding, E.K., Conlan, B.F., Saska, I., Daly, N.L., van**
685 **der Weerden, N.L., Craik, D.J., and Anderson, M.A.** (2015). Efficient backbone
686 cyclization of linear peptides by a recombinant asparaginyl endopeptidase. Nat.
687 Commun. **6,** 10199.
688 **Hasegawa, H., and Holm, L.** (2009). Advances and pitfalls of protein structural alignment. Curr.
689 Opin. Struct. Biol. **19,** 341-348.
690 **Haywood, J., Schmidberger, J.W., James, A.M., Nonis, S.G., Sukhoverkov, K.V., Elias, M., Bond,**
691 **C.S., and Mylne, J.S.** (2018). Structural basis of ribosomal peptide macrocyclization in
692 plants. eLife **7,** e32955.

693 **Hemu, X., Qiu, Y., Nguyen, G.K.T., and Tam, J.P.** (2016). Total synthesis of circular bacteriocins
694 by Butelase 1. J. Am. Chem. Soc. **138,** 6968-6971.
695 **Hughes, A.L.** (1994). The evolution of functionally novel proteins after gene duplication. Proc
696 Biol Sci **256,** 119-124.
697 **Jackson, M.A., Gilding, E.K., Shafee, T., Harris, K.S., Kaas, Q., Poon, S., Yap, K., Jia, H., Guarino,**
698 **R., Chan, L.Y., Durek, T., Anderson, M.A., and Craik, D.J.** (2018). Molecular basis for the
699 production of cyclic peptides by plant asparaginyl endopeptidases. Nat. Commun. **9,**
700 2411.
701 **Jacob, F.** (1977). Evolution and tinkering. Science **196,** 1161-1166.
702 **James, A.M., Haywood, J., and Mylne, J.S.** (2018). Macrocyclization by asparaginyl
703 endopeptidases. New Phytol. **2118,** 923-928.
704 **Jensen, R.A.** (1976). Enzyme recruitment in evolution of new function. Annu. Rev. Microbiol.
705 **30,** 409-425.
706 **Kaltenbach, M., and Tokuriki, N.** (2014). Dynamics and constraints of enzyme evolution. J. Exp.
707 Zool. B. Mol. Dev. Evol. **322,** 468-487.
708 **Khersonsky, O., Roodveldt, C., and Tawfik, D.S.** (2006). Enzyme promiscuity: evolutionary and
709 mechanistic aspects. Curr. Opin. Chem. Biol. **10,** 498-508.
710 **Krissinel, E., and Henrick, K.** (2007). Inference of macromolecular assemblies from crystalline
711 state. J. Mol. Biol. **372,** 774-797.
712 **Kuroyanagi, M., Yamada, K., Hatsugai, N., Kondo, M., Nishimura, M., and Hara-Nishimura, I.**
713 (2005). Vacuolar processing enzyme is essential for mycotoxin-induced cell death in
714 *Arabidopsis thaliana*. J. Biol. Chem. **280,** 32914-32920.
715 **Lobstein, J., Emrich, C., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M.** (2012). SHuffle, a
716 novel *Escherichia coli* protein expression strain capable of correctly folding disulfide
717 bonded proteins in its cytoplasm. Microb. Cell Fact. **11,** 56.
718 **Lu, H., Chandrasekar, B., Oeljeklaus, J., Misas-Villamil, J.C., Wang, Z., Shindo, T., Bogyo, M.,**
719 **Kaiser, M., and van der Hoorn, R.A.** (2015). Subfamily-specific fluorescent probes for
720 cysteine proteases display dynamic protease activities during seed germination. Plant
721 Physiol. **168,** 1462-1475.
722 **McPhillips, T.M., McPhillips, S.E., Chiu, H.-J., Cohen, A.E., Deacon, A.M., Ellis, P.J., Garman, E.,**
723 **Gonzalez, A., Sauter, N.K., Phizackerley, R.P., Soltis, S.M., and Kuhn, P.** (2002). Blu-Ice
724 and the Distributed Control System: software for data acquisition and instrument
725 control at macromolecular crystallography beamlines. Journal of Synchrotron Radiation
726 **9,** 401-406.
727 **Mylne, J.S., and Botella, J.R.** (1998). Binary vectors for sense and antisense expression of
728 *Arabidopsis* ESTs. Plant Mol. Biol. Report. **16,** 257-262.
729 **Mylne, J.S., Colgrave, M.L., Daly, N.L., Chanson, A.H., Elliott, A.G., McCallum, E.J., Jones, A.,**
730 **and Craik, D.J.** (2011). Albumins and their processing machinery are hijacked for cyclic
731 peptides in sunflower. Nat. Chem. Biol. **7,** 257-259.
732 **Nguyen, G.K.T., Hemu, X., Quek, J.-P., and Tam, J.P.** (2016a). Butelase-mediated
733 macrocyclization of d-amino-acid-containing peptides. Angew. Chem. Int. Ed. **55,** 12802-
734 12806.

735 **Nguyen, G.K.T., Wang, S., Qiu, Y., Hemu, X., Lian, Y., and Tam, J.P.** (2014). Butelase 1 is an Asx-
736 specific ligase enabling peptide macrocyclization and synthesis. Nat. Chem. Biol. **10,** 732-
737 738.
738 **Nguyen, G.K.T., Kam, A., Loo, S., Jansson, A.E., Pan, L.X., and Tam, J.P.** (2015). Butelase 1: A
739 versatile ligase for peptide and protein macrocyclization. J. Am. Chem. Soc. **137,** 15398-
740 15401.
741 **Nguyen, G.K.T., Qiu, Y., Cao, Y., Hemu, X., Liu, C.-F., and Tam, J.P.** (2016b). Butelase-mediated
742 cyclization and ligation of peptides and proteins. Nat. Protocols **11,** 1977-1988.
743 **Otegui, M.S., Herder, R., Schulze, J., Jung, R., and Staehelin, L.A.** (2006). The proteolytic
744 processing of seed storage proteins in *Arabidopsis* embryo cells starts in the
745 multivesicular bodies. Plant Cell **18,** 2567-2581.
746 **Pei, J., Kim, B.H., and Grishin, N.V.** (2008). PROMALS3D: a tool for multiple protein sequence
747 and structure alignments. Nucleic Acids Res. **36,** 2295-2300.
748 **Schrodinger, L.** (2010). The PyMOL Molecular Graphics System, Version 1.4.
749 **Serra, A., Hemu, X., Nguyen, G.K.T., Nguyen, N.T.K., Sze, S.K., and Tam, J.P.** (2016). A high-
750 throughput peptidomic strategy to decipher the molecular diversity of cyclic cysteine-
751 rich peptides. Sci. Rep. **6,** 23005.
752 **Shimada, T., Yamada, K., Kataoka, M., Nakaune, S., Koumoto, Y., Kuroyanagi, M., Tabata, S.,**
753 **Kato, T., Shinozaki, K., Seki, M., Kobayashi, M., Kondo, M., Nishimura, M., and Hara-**
754 **Nishimura, I.** (2003). Vacuolar processing enzymes are essential for proper processing of
755 seed storage proteins in *Arabidopsis thaliana*. J. Biol. Chem. **278,** 32292-32299.
756 **Tokuriki, N., Jackson, C.J., Afriat-Jurnou, L., Wyganowski, K.T., Tang, R., and Tawfik, D.S.**
757 (2012). Diminishing returns and tradeoffs constrain the laboratory optimization of an
758 enzyme. Nat. Commun. **3,** 1257.
759 **Winn, M. D. , Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M.,**
760 **Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., S. Pannu,**
761 **N., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., and Wilson, K.S.** (2011).
762 Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol
763 Crystallogr **67,** 235-242.
764 **Yang, R., Wong, Y.H., Nguyen, G.K.T., Tam, J.P., Lescar, J., and Wu, B.** (2017). Engineering a
765 catalytically efficient recombinant protein ligase. J. Am. Chem. Soc. **139,** 5351-5358.
766 **Zauner, F.B., Elsasser, B., Dall, E., Cabrele, C., and Brandstetter, H.** (2018a). Structural analyses
767 of *Arabidopsis thaliana* legumain gamma reveal the differential recognition and
768 processing of proteolysis and ligation substrates. J. Biol. Chem.
769 **10.1074/jbc.M117.817031**.
770 **Zauner, F.B., Dall, E., Regl, C., Grassi, L., Huber, C.G., Cabrele, C., and Brandstetter, H.** (2018b).
771 Crystal structure of plant legumain reveals a unique two-chain state with pH-dependent
772 activity regulation. Plant Cell **30,** 686-699.
773 **Zhao, L., Hua, T., Crowley, C., Ru, H., Ni, X., Shaw, N., Jiao, L., Ding, W., Qu, L., Hung, L.-W.,**
774 **Huang, W., Liu, L., Ye, K., Ouyang, S., Cheng, G., and Liu, Z.-J.** (2014). Structural analysis
775 of asparaginyl endopeptidase reveals the activation mechanism and a reversible
776 intermediate maturation stage. Cell Res. **24,** 344-358.
777

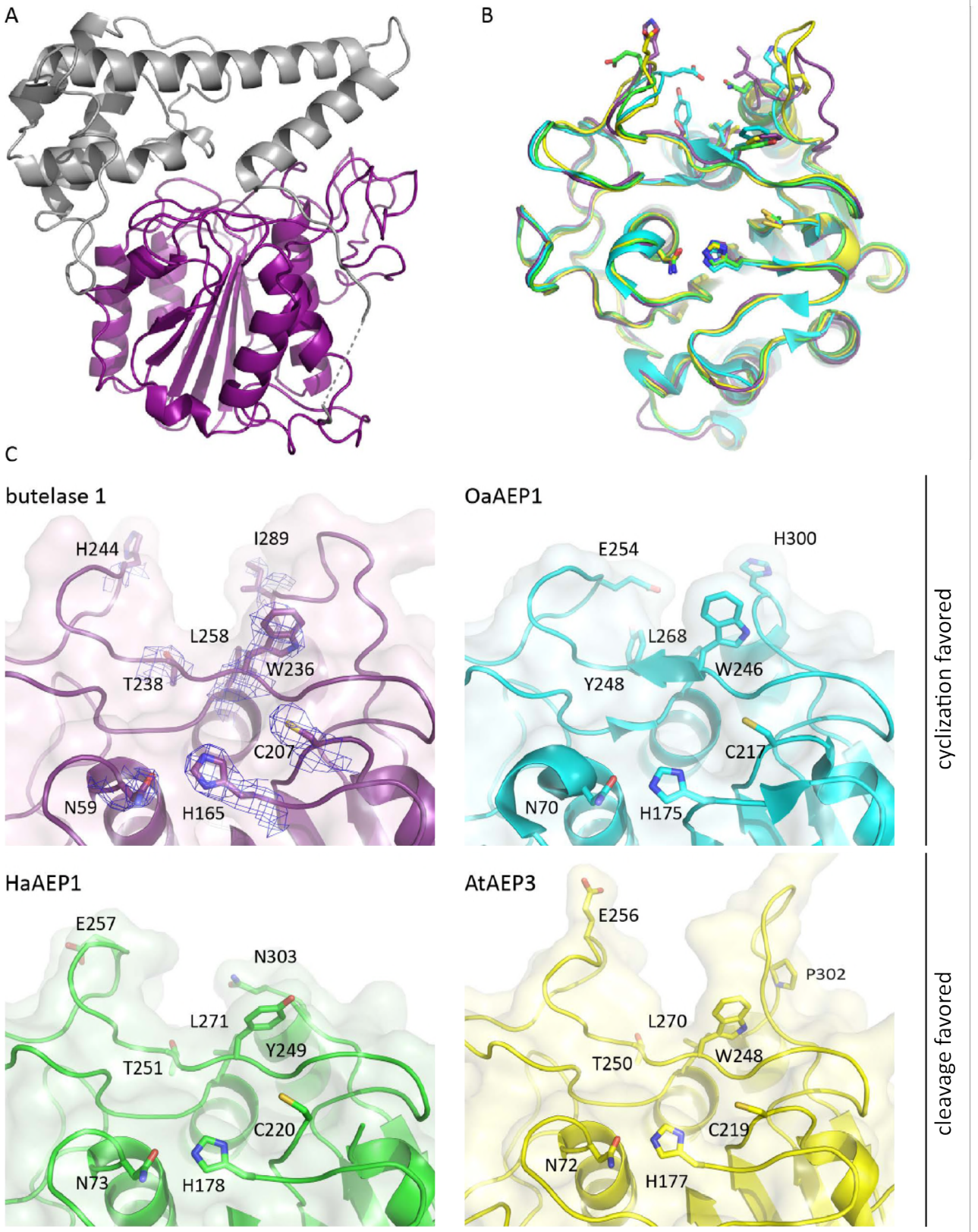

A
B
C
butelase 1
H244
I289
L258
W236
T238
C207
N59
H165
OaAEP1
E254
H300
L268
W246
Y248
C217
N70
H175
HaAEP1
E257
N303
L271
Y249
T251
C220
N73
H178
AtAEP3
E256
P302
L270
W248
T250
C219
N72
H177
cyclization favored
cleavage favored

Figure 1

```
pQE30-HaAEP2                                    MRGSHHHHHHGSSVGTRWAILIAGSNGYSNY
Native HaAEP2     MNRHLLILFALSIIAAVNGRELTEDLIRLPSEKFDFSDDDSVGTRWAILIAGSNGYSNY
pQE30-butelase 1          MRGSHHHHHHGSIRDDFLRLPSQASKFFQADDNVEGTRWAVLVAGSKGYVNY
Native butelase 1 MKNPLAILFLIATVVAVVSGIRDDFLRLPSQASKFFQADDNVEGTRWAVLVAGSKGYVNY

pQE30-HaAEP2    RHQADVCHAYQILKTGGLKDENIVVFMYDDIAYNPENPRRGVIINSPDGDDVYHGVPKDY
Native HaAEP2    RHQADVCHAYQILKTGGLKDENIVVFMYDDIAYNPENPRRGVIINSPDGDDVYHGVPKDY
pQE30-butelase 1 RHQADVCHAYQILKKGGLKDENIIVFMYDDIAYNESNPHPGVIINHPYGSDVYKGVPKDY
Native butelase 1 RHQADVCHAYQILKKGGLKDENIIVFMYDDIAYNESNPHPGVIINHPYGSDVYKGVPKDY

pQE30-HaAEP2    TGDDVNVDNFFAVLLGDKSKVKG-GSGKVVDSGPNDRIFIYYTDHGGPGVLGMPTYPFLY
Native HaAEP2    TGDDVNVDNFFAVLLGDKSKVKG-GSGKVVDSGPNDRIFIYYTDHGGPGVLGMPTYPFLY
pQE30-butelase 1 VGEDINPPNFYAVLLANKSALTGTGSGKVLDSGPNDHVFIYYTDHGGAGVLGMPSKPYIA
Native butelase 1 VGEDINPPNFYAVLLANKSALTGTGSGKVLDSGPNDHVFIYYTDHGGAGVLGMPSKPYIA

pQE30-HaAEP2    ANDLNEVLKQKHASGTYKSLVFYLEACESGSIFDGLLPQGLNIYATTASNPYENSWGTYC
Native HaAEP2    ANDLNEVLKQKHASGTYKSLVFYLEACESGSIFDGLLPQGLNIYATTASNPYENSWGTYC
pQE30-butelase 1 ASDLNDVLKKKHASGTYKSIVFYVESCESGSMFDGLLPEDHNIYVMGASDTGESSWVTYC
Native butelase 1 ASDLNDVLKKKHASGTYKSIVFYVESCESGSMFDGLLPEDHNIYVMGASDTGESSWVTYC

pQE30-HaAEP2    PGDYPSPPPEYDTCLGDLYSVAWMEDCDVHNLRTETIKQQYKLVKERTSSGNYYYGSHVM
Native HaAEP2    PGDYPSPPPEYDTCLGDLYSVAWMEDCDVHNLRTETIKQQYKLVKERTSSGNYYYGSHVM
pQE30-butelase 1 PLQHPSPPPEYDVCVGDLFSVAWLEDCDVHNLQTETFQQQYEVVKNKTIVALIEDGTHVV
Native butelase 1 PLQHPSPPPEYDVCVGDLFSVAWLEDCDVHNLQTETFQQQYEVVKNKTIVALIEDGTHVV

                                        *                   *  *
pQE30-HaAEP2    QYGDLPLSTDNLYVYMGTNPANENFTFSEENSLFRSPKAVNQRDADLLHFWHKYRKAPEG
Native HaAEP2    QYGDLPLSTDNLYVYMGTNPANENFTFSEENSLFRSPKAVNQRDADLLHFWHKYRKAPEG
pQE30-butelase 1 QYGDVGLSKQTLFVYMGTDPANDNNTFTDKNSLGTPRKAVSQRDADLIHYWEKYRRAPEG
Native butelase 1 QYGDVGLSKQTLFVYMGTDPANDNNTFTDKNSLGTPRKAVSQRDADLIHYWEKYRRAPEG

pQE30-HaAEP2    FDKKTEVQEKFTEAMSHRMHIDSSIQLIGKLLFGLEKGPEVLKTVRSAGKPLVDDWTCLK
Native HaAEP2    SDKKTEAQEKFTEAMSHRMHIDSSIQLIGKLLFGLEKGPEVLKTVRSAGKPLVDDWTCLK
pQE30-butelase 1 SSRKAEAKKQLREVMAHRMHIDNSVKHIGKLLFGIEKGHKMLNNVRPAGLPVVDDWDCFK
Native butelase 1 SSRKAEAKKQLREVMAHRMHIDNSVKHIGKLLFGIEKGHKMLNNVRPAGLPVVDDWDCFK

pQE30-HaAEP2    TFVRTFETHCGSLSQYGMKHMRSIANLCNAGVTNEQMAAASSQACTAFPSNPWSSLSRGF
Native HaAEP2    TFVRTFETHCGSLSQYGMKHMRSIANLCNAGVTNEQMAAASSQACTAFPSNPWSSLSRGF
pQE30-butelase 1 TLIRTFETHCGSLSEYGMKHMRSFANLCNAGIRKEQMAEASAQACVSIPDNPWSSLHAGF
Native butelase 1 TLIRTFETHCGSLSEYGMKHMRSFANLCNAGIRKEQMAEASAQACVSIPDNPWSSLHAGF

pQE30-HaAEP2    TA
Native HaAEP2    TA
pQE30-butelase 1 SV
Native butelase 1 SV
```

Figure 1—figure supplement 1

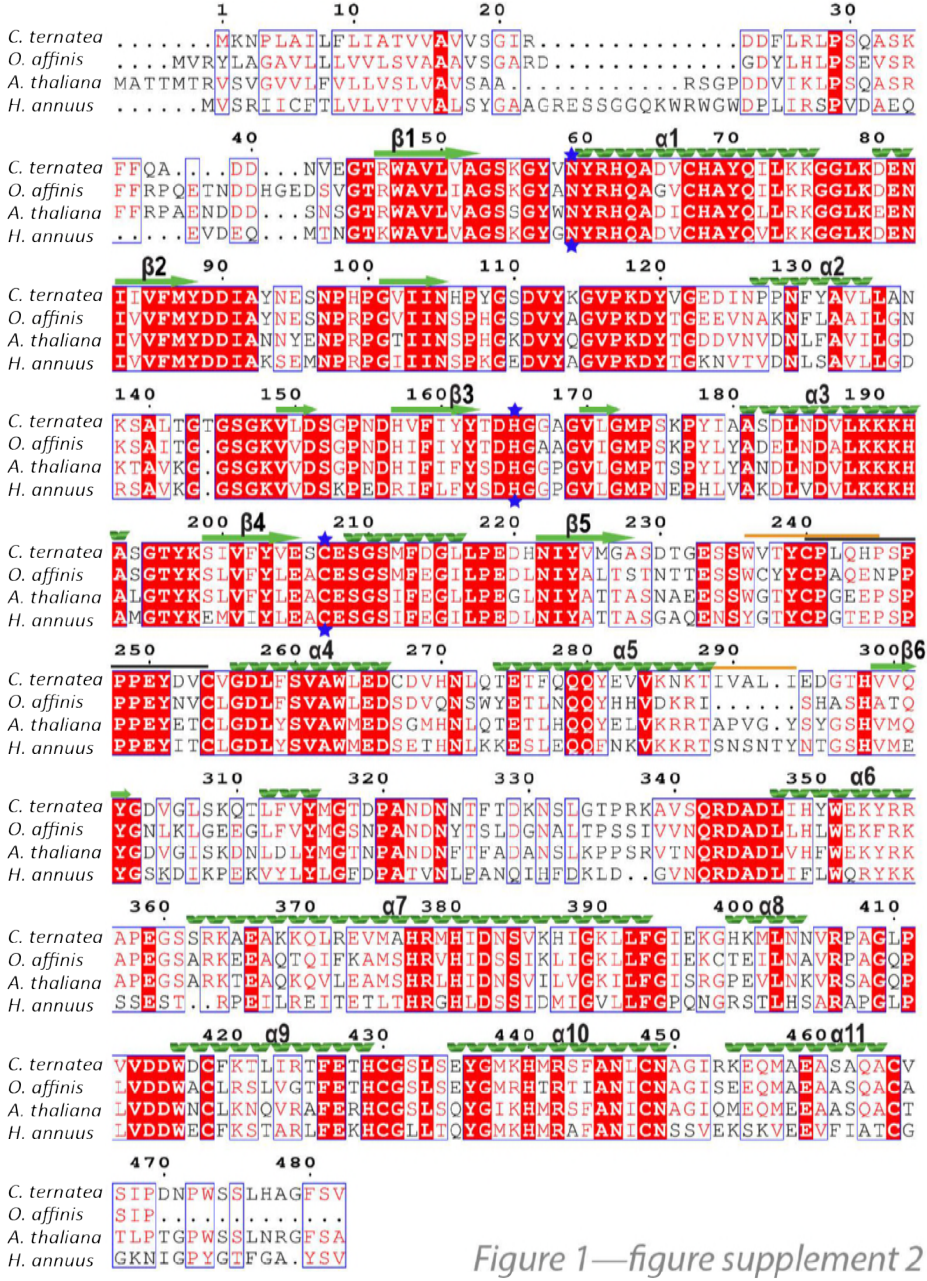


Figure 1—figure supplement 2

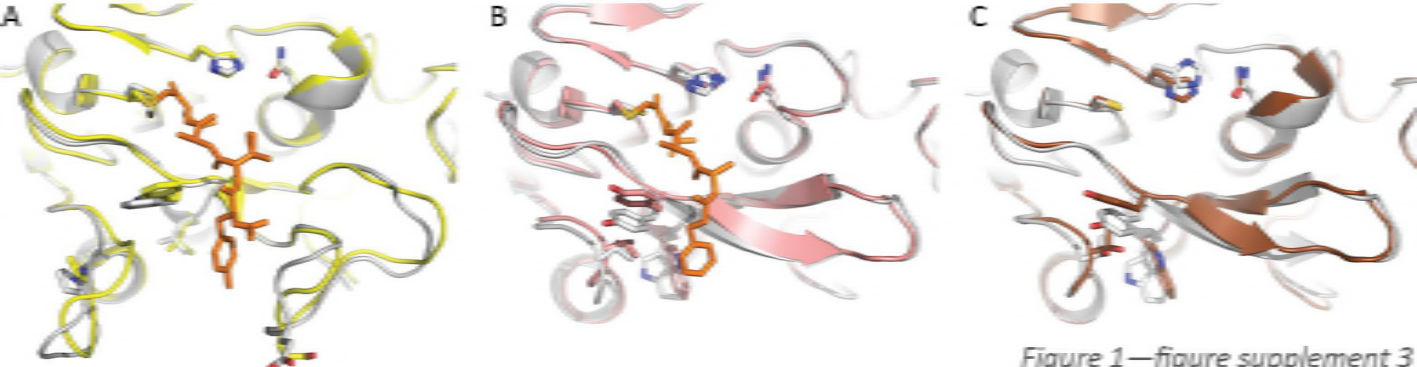


Figure 1—figure supplement 3

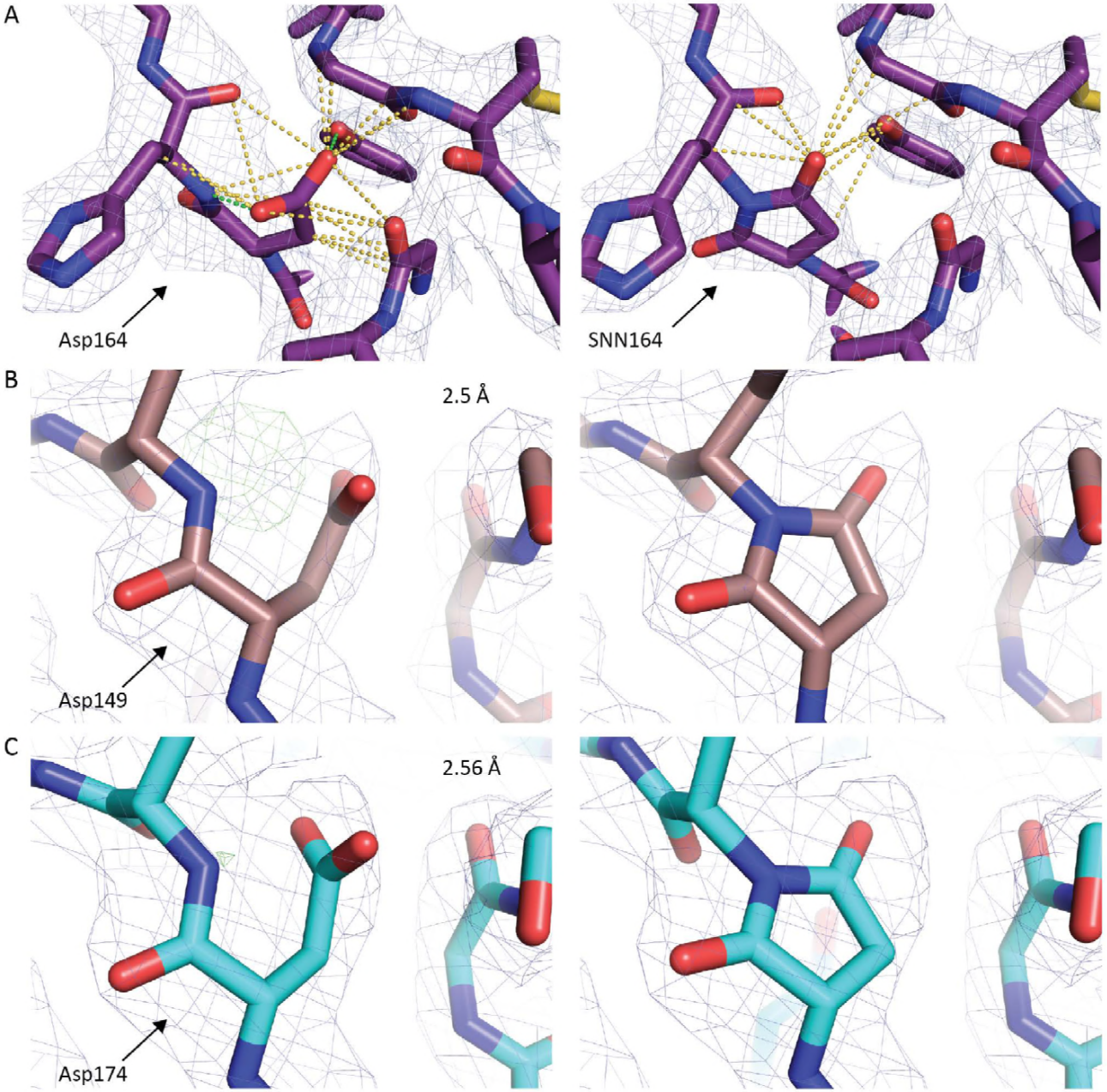
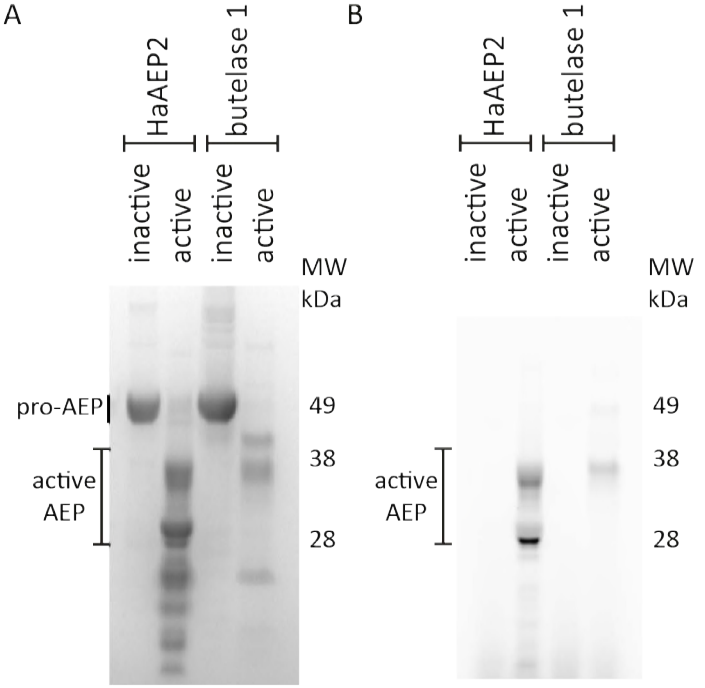


Figure 1—figure supplement 4


A
HaAEP2
butelase 1
inactive
active
inactive
active
MW
kDa
pro-AEP
active
AEP
49
38
28
B
HaAEP2
butelase 1
inactive
active
inactive
active
MW
kDa
active
AEP
49
38
28

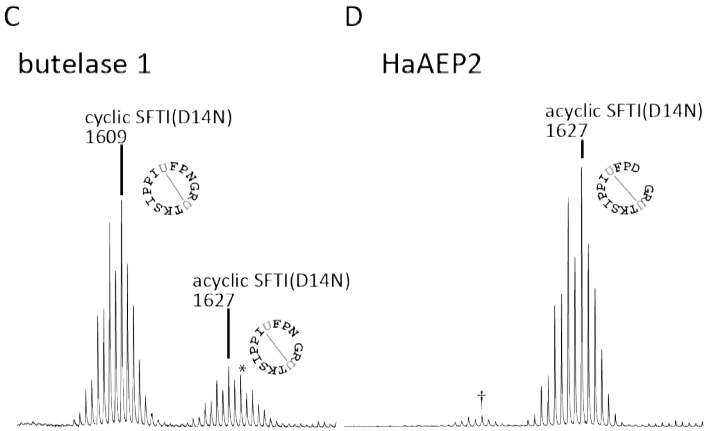

C
butelase 1
cyclic SFTI(D14N)
1609
acyclic SFTI(D14N)
1627
*
D
HaAEP2
acyclic SFTI(D14N)
1627
†

Figure 2

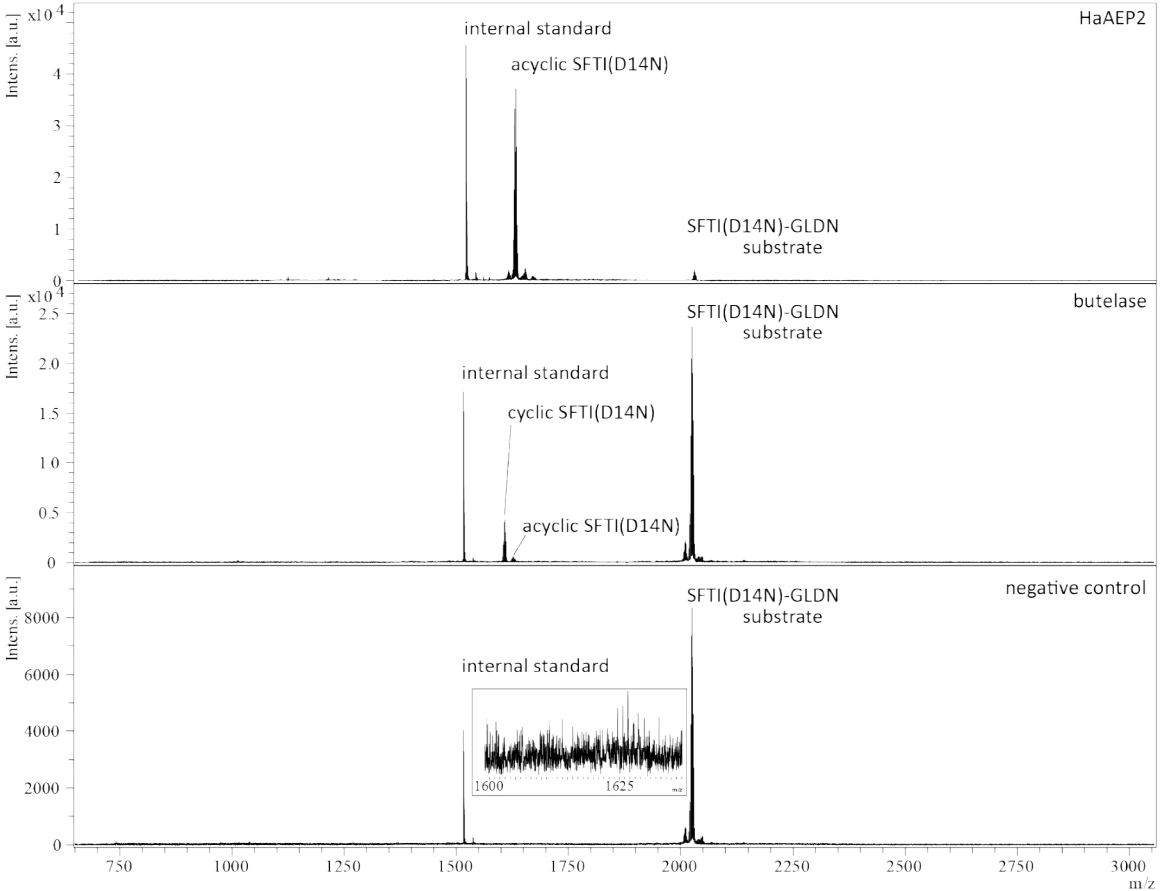


Figure 2—figure supplement 1

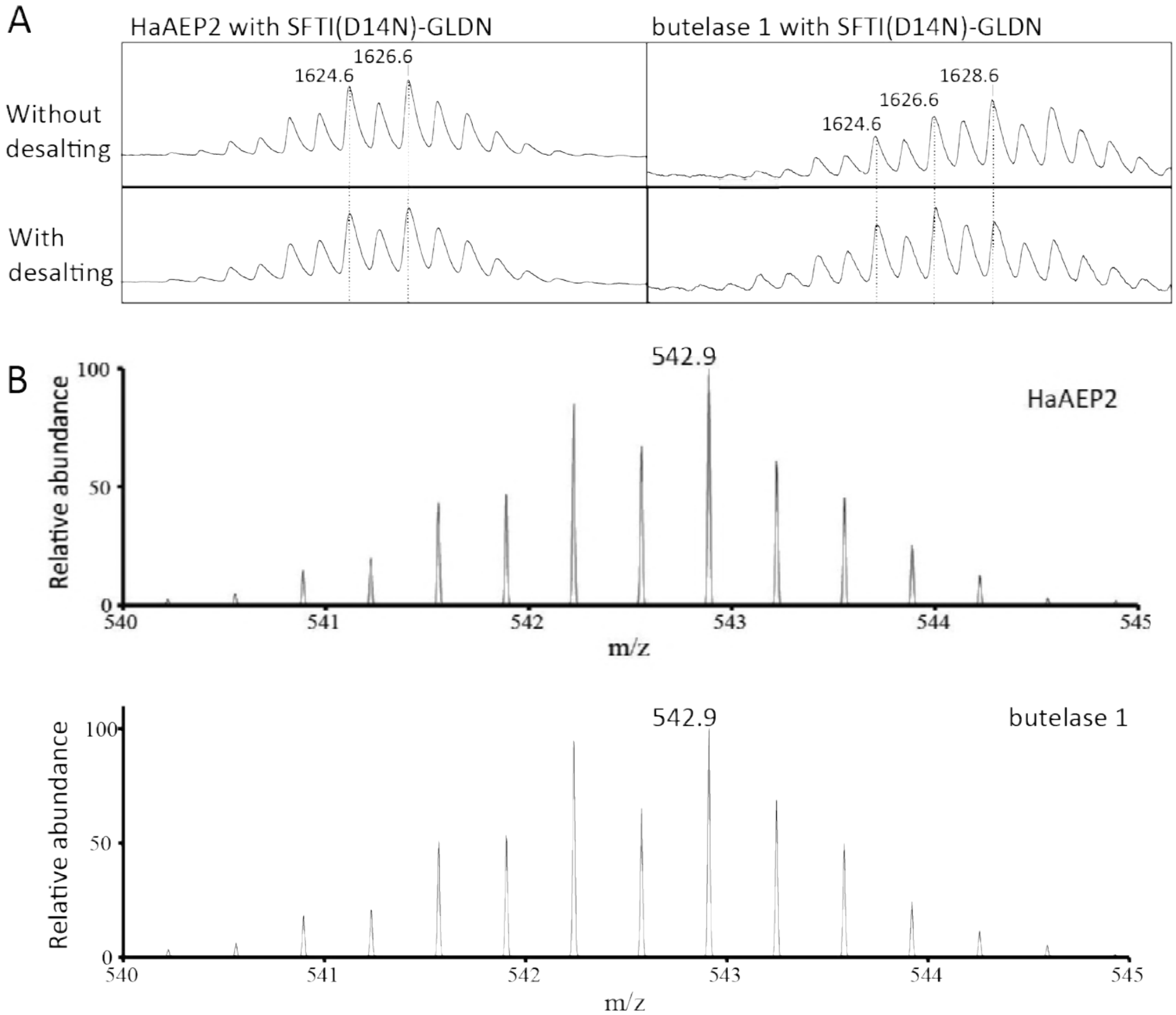


Figure 2—figure supplement 2

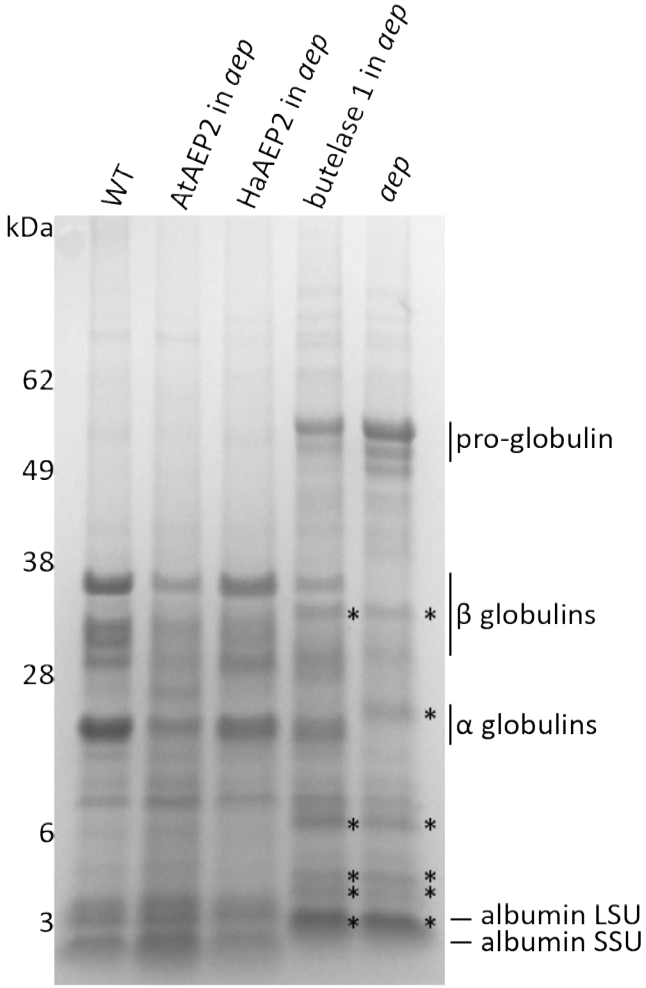

WT
AtAEP2 in aep
HaAEP2 in aep
butelase 1 in aep
aep
kDa
62
49
38
28
6
3
pro-globulin
β globulins
α globulins
albumin LSU
albumin SSU

Figure 3

A

SESA1

MANKLFLVCAALALCFLLTNASIYRTVVEFEEDDATNPIGPKMRKCRKEFQKEQHLRAC
QQLMLQQARQGRSDEFDFEDDMFNPQGQQQEQQLFQQCCNELRQEEPDCVCPTLKQAAK
AVRLQGQHQPMQVRKIYQTAKHLPNVCDIPQVDVCPFNIPSFPSFY

SESA3

MANKLFLVCATLALCFLLTNASIYRTVVEFEEDDASNPVGPRQRCQKEFQQSQHLRACQ
RWMSKQMRQGRGGGPSLDDEFDFEGPQQGYQLLQQCCNELRQEEPVCVCPTLKQAARAV
SLQGQHGPFQSRKIYQSAKYLPNICKIQQVGECPFQTTIPFFPPYY

SESA4

MANKLFLVCAALALCFILTNASVYRTVVEFDEDDASNPIGPIQKCQKEFQQDQHLRACQ
RWMRKQMWQGRGGGPSLDDEFDMEDDIENPQRRQLLQKCCSELRQEEPVCVCPTLRQAA
KAVRFQGQQHQPEQVRKIYQAAKYLPNICKIQQVGVCPFQIPSIPSYY

CRU3

MVKLSNLLVATFGVLLVLNGCLARQSLGVPPQLQNECNLDNLDVLQATETIKSEAGQIE
YWDHNHPQLRCVGVSVARYVIEQGGLYLPTFFTSPKISYVVQGTGISGRVVPGCAETFM
DSQPMQGQQQGQPWQGRQGQQGQPWEGQGQQGQQGRQGQPWEGQGQQGQQGRQGQQGQP
WEGQGQQGQQGFRDMHQKVEHVRRGDVFANTPGSAHWIYNSGEQPLVIIALLDIANYQN
QLDRNPRVFHLAGNNQQGGFGGSQQQQEQKNLWSGFDAQVIAQALKIDVQLAQQLQNQQ
DSRGNIVRVKGPFQVVRPPLRQPYESEEWRHPRSPQGNGLEETICSMRSHENIDDPARA
DVYKPSLGRVTSVNSYTLPILEYVRLSATRGVLQGNAMVLPKYNMNANEILYCTGGQGR
IQVVNDNGQNVLDQQVQKGQLVVIPQGFAYVVQSHGNKFEWISFKTNENAMISTLAGRT
SLLRALPLEVISNGFQISPEEARKIKFNTLETTLTRAAGRQQQQLIEEIVEA

B

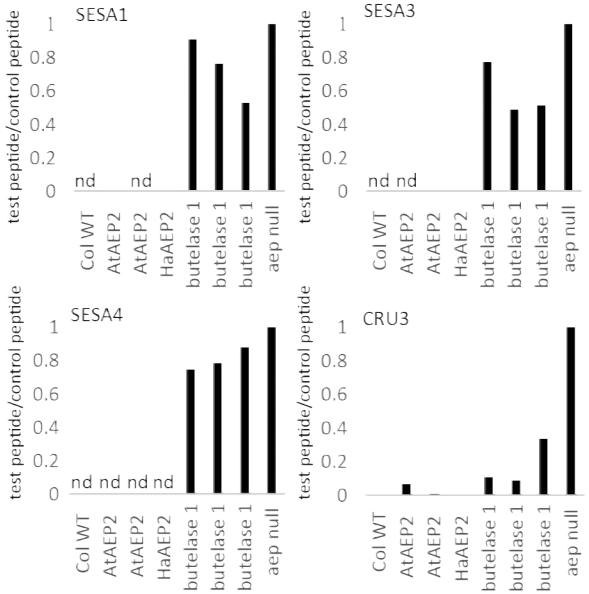


Figure 4