GENERATION OF PEPTIDES

FILING DATA

[0001] This application is associated with and claims priority from Australian Provisional Patent Application No. 20159036918, filed on 25 September, 2015, entitled "Generation of peptides", the entire contents of which, are incorporated herein by reference.

BACKGROUND

FIELD

[0002] The present disclosure relates generally to generation of a recombinant enzyme with cyclization activity and its use for generating cyclic peptides as well as linear peptide conjugates.

DESCRIPTION OF RELATED ART

[0003] Bibliographic details of the publications referred to by author in this specification are collected alphabetically at the end of the description.

[0004] The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgement or admission or any form of suggestion that the prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavor to which this specification relates.

[0005] Proteases are abundant throughout nature and are essential for a wide range of cellular processes. They typically serve to hydrolyze polypeptide chains, resulting in either degradation of the target sequence or maturation to a biologically active form. Less

frequently, proteases can act as ligases to link distinct polypeptides, producing new or alternately spliced variants. This unusual function has been reported for processes such as the maturation of the lectin, Concanavalin A (Sheldon *et al*. (1996) *Biochem. J. 320***:**865–870), peptide presentation by major histocompatibility complex class I molecules (Hanada *et al*. (2004) *Nature 427***:**252-256) and anchoring of bacterial proteins to the cell wall (Mazmanian *et al*. (1999) *Science (80) 285***:**760–763). This enzymatic transpeptidation has also been implicated in the backbone-cyclization of ribosomally synthesized cyclic peptides (Barber *et al.* (2013) *J. Biol. Chem. 288***:**12500–12510; Nguyen *et al.* (2014) *Nat. Chem. Biol. 10***:**732–738; Luo *et al.* (2014) *Chem. Biol.* 1–8 doi:10.1016/j.chembiol.2014.10.015; Lee *et al*. (2009) *J. Am. Chem. Soc. 131***:**2122–2124).

[0006] Gene-encoded cyclic peptides have been identified in a range of organisms including plants, fungi, bacteria and animals (Arnison *et al.* (2013) *Nat Prod Rep 30***:**108–160). In plants, they are divided into four classes: cyclotides (e.g. the prototypical cyclotide kalata B1 [kB1]) (Gillon *et al.* (2008) *Plant J. 53***:**505–515; Saska *et al.* (2007) *J. Biol. Chem. 282***:**29721–29728), PawS-derived trypsin inhibitors (e.g. sunflower trypsin inhibitor (SFTI)) [Mylne *et al.* (2011) *Nat. Chem. Biol. 7***:**257–259], knottin trypsin inhibitors (e.g. *Momordica cochinchinensis* trypsin inhibitor (MCoTI-II)) [Mylne *et al.* (2012) *Plant Cell 24***:**2765–2778] and orbitides (e.g. segetalins) [Barber *et al*. (2013) *supra*].

[0007] Cyclotides were first identified in the African plant *Oldenlandia affinis* and exhibit insecticidal, nematocidal and molluscicidal activity against agricultural pests (Jennings *et al*. (2001) *Proc. Natl. Acad. Sci. U. S. A. 98***:**10614–10619; Plan *et al*. (2008) *J. Agric. Food Chem. 56***:**5237–5241; Colgrave *et al.* (2008) *Biochemistry 47***:**5581–5589; Colgrave *et al.* (2009) *Acta Trop. 109***:**163–166). Other reported activities include neurotensin antagonism (Witherup *et al*. (1994) *J. Nat. Prod* 57:1619–1625), anti-HIV activity (Gustafson *et al.* (2000) *J. Nat. Prod 63:*176–178), anti-microbial activity (Tam *et al*. (1999) *Proc. Natl. Acad. Sci. U. S. A. 96*:8913–8918), cytotoxic activity (Lindholm *et al.* (2002) *Mol. Cancer Ther. 1:*365–369), uterotonic activity (Gran (1973) *Acta pharmacol. toxicol. 33*:400–408), and hemolytic (Tam *et al*. (1999) *supra*) and anti-fouling properties

(Goransson *et al*. (2004) *J. Nat. Prod. 67*:1287–1290). Cyclotides are characterized by a cystine knot motif that, together with backbone cyclization, confers exceptional stability. This has generated much interest in the cyclotide framework as a pharmaceutical scaffold; a potential heightened by the successful grafting of bioactive sequences into both Möbius and trypsin inhibitor cyclotides (Poth *et al*. (2013) *Biopolymers 100*:480–491). Backbone cyclization can also increase the stability and facilitate the oral administration route for bioactive linear peptides, suggesting that this modification will find broad application (Clark *et al*. (2005) *Proc. Natl. Acad. Sci. United States Am. 102*:13767-13772; Clark *et al.* (2010) *Angew. Chem. Int. Ed. Engl.* 49*:*6545–8; Chan *et al.* (2013) *Chembiochem 14*:617–624). Elucidating the mechanism of enzymatic cyclization intrinsic to cyclotide biosynthesis is important not only for the realization of the pharmaceutical and agricultural potential of cyclotides, but also for increasing the cyclization efficiency of unrelated, bioactive peptides.

[0008] Cyclotides are produced from precursor molecules in which the cyclotide sequence is typically flanked by N- and C-terminal propeptides. The first processing event is the removal of the N-terminal propeptide, producing a linear precursor that remains linked to the C-terminal prodomain (Gillon *et al*. (2008) *supra*). The final maturation step involves enzymatic cleavage of this C-terminal region and subsequent ligation of the free C- and N-termini. However, only four native cyclases have been identified to date (Barber *et al*. (2013) *supra*; Nguyen *et al*. (2014) *supra;* Luo *et al*. (2014) *supra;* Lee *et al*. (2009) *supra;* Gillon *et al*. (2008) *supra*). The best characterized of these is the serine protease PatG, which is responsible for maturation of the bacterial cyanobactins (Lee *et al*. (2009) *supra*). In plants, the serine protease PCY1 reportedly facilitates cyclization of the segetalins, cyclic peptides from the Caryophyllaceae (Barber *et al*. (2013) *supra*). In the other three classes of plant-derived cyclic peptides, strong Asx sequence (where x is N (asparagine) or D (aspartic acid)) conservation at the P1 residue of the C-terminal cleavage site suggested involvement of a group of cysteine proteases known as vacuolar processing enzymes (VPEs) or asparaginyl endopeptidases (AEPs) in this process (Gillon *et al*. (2008) *supra*).

[0009] Of the small number of AEPs which have been demonstrated to preferentially act

as peptide ligases, only one of these, butelase 1, has been shown to be an efficient cyclase (Bernath-Levin *et al*. (2015) *Chemistry & Biology 22*:1-12; Nguyen *et al*. (2014) *supra*; Sheldon *et al.* (1996) *supra*). The structural basis for the preferential ligase activity of this subset of AEPs remains unknown.

[0010] Butelase-1 was isolated from the cyclotide producing plant *Clitoria ternatea* and shown to cyclize a modified precursor of kB1 from *O. affinis*, confirming the ability of this group of enzymes to mediate cyclization *in vitro* (Nguyen *et al.* (2014) *supra*) provided that the appropriate recognition sequences are added to the ends of the polypeptide precursor to be cyclized. However, recombinant butelase-1 from *E. coli* was only expressed in insoluble form and thus unable to mediate cyclization. Only one AEP with any cyclizing ability has been produced recombinantly, and this was highly inefficient, producing mainly hydrolyzed substrate (Bernath-Levin *et al.* (2015) *supra*). There is a need to develop methodology to generate a functional recombinant AEP so that it can be used to more efficiently generate cyclic peptides from polypeptide precursors as well as linear peptide conjugates.

SUMMARY

[0011] The present disclosure teaches the production of a functional recombinant asparaginyl endopeptidase (AEP) and its use in an efficient method for producing a cyclic peptide or linear peptide conjugate. The term "cyclic peptide" includes but is not limited to a cyclotide. The cyclic peptide may be naturally cyclical or may be artificially cyclized to confer, for example, added stability, efficacy or utility. A linear peptide conjugate is the ligation of two or more peptides together in linear sequence. The term "peptide" is not to exclude a polypeptide or protein. For brevity, the term "peptide" is used to avoid any doubt, the present invention covers a cyclic peptide, cyclic polypeptide and cyclic protein as well as a linear peptide, linear polypeptide or linear protein. All encompassed by the term "cyclic peptide" or "linear peptide".

[0012] The cyclic peptide or linear peptide can be used in a variety of applications relevant to human and non-human animals and plants. Included are agricultural applications such as the generation of topical agents for treatment of infection or infestation by a pathogen and pharmacological applications such as the treatment of cancer, cardiovascular disease, infectious disease, immune diseases and pain. Therapeutic agents may be delivered topically or systemically. In addition, both naturally cyclic peptides in linear form and naturally linear peptides can be subject to cyclization as well as linear polypeptide precursors comprising non-naturally occurring amino acids and/or modified side chains or modified cross-linkage bonds. The cyclization of a naturally linear peptide can lead *inter alia* to a longer half life and/or increased stability and/or the ability to be orally administered.

[0013] The cyclization process may be conducted in various ways and can employ prokaryotic or eukaryotic organisms and can act on a polypeptide precursor containing a non-naturally occurring amino acid residue or other modification. In essence, an asparaginyl endopeptidase (AEP) with cyclizing ability is employed to cyclize a linear polypeptide precursor or ligate together peptides including polypeptides and proteins. The term "polypeptide" includes a "protein". The polypeptide precursor includes a precursor to a naturally cyclic peptide as well as a polypeptide which is naturally linear and is converted into a cyclic peptide.

[0014] The linear polypeptide precursor comprises a C-terminal AEP processing site. Generally, but not exclusively, the C-terminal processing site is an amino acid sequence defined as comprising P3 to P1 prior to the actual cleavage site and comprising P1' to P3' after the cleavage site towards the C-terminal end. In an embodiment, P3 to P1 and P1' to P3' have the amino acid sequence:

$X_2X_3X_4X_5X_6X_7$

wherein X is an amino acid residue and:

X_2 is optional or is any amino acid;

X_3 is optional or is any amino acid;

X_4 is N or D;

X_5 is G or S;

X_6 is L or A or I; and

X_7 is optional or any amino acid.

[0015] In an embodiment, X_1 through X_6 comprise the amino acid sequence:

$X_2X_3NGLX_7$

wherein X_2, X_3 and X_7 are as defined above.

[0016] The N-terminal end of the linear polypeptide precursor may contain no specific AEP processing site or may contain a processing site defined by any one of P1" through P3" wherein P1" to P3" is defined by:

$X_9X_{10}X_{11}$

wherein X is an amino acid residue:

X_9 is optional and any amino acid or G, Q, K, V or L;

X_{10} is optional or any amino acid or L, F or I or an hydrophobic amino acid residue;

X_{11} is optional and any amino acid.

[0017] In an embodiment, X_9 through X_{11} comprise the amino acid sequence:

GLX_{11}

wherein X_{11} is defined as above.

[0018] In an embodiment, the AEP processing site comprises N- and C-terminal end sequences comprising the sequence:

GLX_{11} $[X_n]$ $X_2X_3NGLX_7$

wherein X_{11}, X_2, X_3, and X_7 are as defined above and $[X_n]$ is absent (n=0) or any amino acid residue in a sequence of from 1 to 2000 amino acids.

[0019] In an embodiment, the C-terminal processing site comprises P4 to P1 and P1' to P4' wherein P1 to P4 and P1' to P4' comprise $X_1X_2X_3X_4X_5X_6X_7X_8$ wherein X_2 to X_7 are as defined above and X_1 is optional or any amino acid and X_8 is optional or any amino acid.

[0020] In the case of a prokaryotic system, the AEP is produced in the cell and isolated before it is used *in vitro* with a linear polypeptide precursor to be cyclized. The linear polypeptide precursor may also be produced in the cell then separated or otherwise isolated from the cell and cyclized *in vitro* using the recombinant AEP. A polypeptide precursor produced by synthesis, including polypeptides with non-naturally occurring amino acids or a recombinant polypeptide with post-translational modification can also be cyclized *in vitro* using a recombinant AEP. The AEP and polypeptide precursor may also be co-expressed in a compartment of a prokaryotic cell such as but not limited to the periplasmic space. In which case, the resulting cyclic peptide is isolated from the cell.

[0021] A similar protocol is adapted when a eukaryotic organism is employed, such as a yeast (e.g. *Pichia* sp., *Saccharomyces* sp. or *Kluyveromyces* sp.). Genetic material encoding AEP is expressed enabling generation of cyclic peptides *in vitro* from a precursor polypeptide or *in vivo* if both the AEP and polypeptide are co-expressed. In either event, the resulting cyclic peptide is subject to isolation and purification from a vacuole or other cellular compartment in the eukaryotic cell or from the reaction vessel. Alternatively, the AEP and polypeptide precursor are produced in separate eukaryotic cells or in different compartments within the same cell, extracted and then co-incubated *in vitro* to generate the cyclic peptide. In yet

another aspect only one or other of the AEP or polypeptide precursor is produced in the eukaryotic cell; the other component is supplied from a different source and the two are then incubated *in vitro* to generate a cyclic peptide.

[0022] Just to re-emphasize, the term "peptide" includes a polypeptide or protein as well as a peptide.

[022a] Disclosed herein is a method for producing; (a) a cyclic peptide; or (b) a peptide conjugate, wherein: (i) said cyclic peptide is generated by a method comprising generating a recombinant asparaginyl endopeptidase (AEP) vacuolar processing enzyme with peptide cyclization activity in a prokaryotic or eukaryotic cell and co-incubating the recombinant AEP with a linear polypeptide precursor of the cyclic peptide wherein the polypeptide precursor comprises N-terminal and/or C-terminal AEP processing site(s) for a time and under conditions sufficient to generate the cyclic peptide; and (ii) said peptide conjugate is generated by a method comprising co-incubating at least two peptides wherein at least one peptide comprises a C-terminal AEP recognition amino acid sequence and at least one other peptide comprises an N-terminal AEP recognition amino acid sequence with a recombinant AEP with peptide ligase activity for a time and under conditions sufficient to generate a linear peptide conjugate; wherein the recombinant AEP comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:1, 2, 4, 39, 41 and amino acid sequences having at least 80% similarity to any of the foregoing after optimal alignment; and wherein the recombinant AEP comprises the presence of 5 or more of residues or absence of residues at 139K, 161D, 186K, 192D, 247C, 248Y, 253Q, 255A, 263V, 293H, Gap, Gap, Gap, Gap, Gap (between residues 299 and 300), 314E and 316G when optimally aligned to SEQ ID NO:1, wherein Gap means the absence of a residue.

[0023] Enabled herein is a method for producing a cyclic peptide, the method comprising introducing into the prokaryotic or eukaryotic cell genetic material which, when expressed, generates a recombinant AEP with cyclization ability, isolating the AEP and incubating the AEP with a linear polypeptide precursor optionally modified to introduce a non-naturally occurring amino acid, the incubation being for a time and under conditions sufficient to generate a cyclic peptide from the polypeptide precursor. Alternatively, genetic material encoding the AEP with cyclization ability is co-expressed with genetic material encoding a

linear polypeptide precursor in a cell for a time and under conditions sufficient to generate a cyclic peptide in a vacuole or other cellular compartment of the cell. This process can also occur in a membranous compartment of a prokaryotic cell such as a periplasmic space. In addition, the AEP can catalyze a ligation reaction to conjugate two or more peptides wherein at least one peptide comprises a C-terminal AEP recognition amino acid sequence and another peptide comprises an N-terminal AEP recognition amino acid sequence. The eukaryotic cell can also be used to generate one or both of the AEP and/or polypeptide precursor for use in the generation of a cyclic peptide *in vitro*. A cyclic peptide can also be produced in the prokaryotic cell. In an embodiment, the cyclic peptide is produced in the periplasmic space of a prokaryotic cell. As indicated above, reference to "peptide" includes a polypeptide or protein. No limitation in size or type of proteinaceous molecule is intended by use of the term "peptide", "polypeptide" or "protein".

[0024] In an embodiment, a linear peptide is generated using ligase activity of an AEP. In this embodiment, a first peptide comprising the C-terminal AEP recognition sequence is co-incubated with a second peptide comprising an N-terminal AEP recognition sequence which may or may not have a tag and an AEP. The AEP catalyses a ligation between the first and second peptides to generate a linear peptide conjugate. This may then subsequently be cyclized into a cyclic peptide or used as a linear peptide.

[0025] In an embodiment, regardless of the manner the cyclic peptide or peptide conjugate is generated, it is subject to isolation which includes purification.

[0026] Enabled herein is a method for producing a cyclic peptide, the method comprising co-incubating an AEP with peptide cyclization activity with a linear polypeptide precursor of the cyclic peptide for a time and under conditions sufficient to generate the cyclic peptide. Reference to "cyclic peptide" includes a "cyclotide". By "co-incubation" means either *in vitro* in a reaction vessel or in a cell or in a compartment of a cell. Multiple peptides or repeat forms of the same peptides may also be cyclized *in vitro* or *in vivo*. Again, it is emphasized that the term "peptide" includes a polypeptide and a protein.

[0027] Hence, taught herein the AEP is generated in a prokaryotic cell or eukaryotic cell and

used *in vitro* or *in vivo* to generate a cyclic peptide from a linear polypeptide precursor. The AEP and linear polypeptide precursor may also be co-expressed in a prokaryotic cell or eukaryotic cell. Alternatively, the linear polypeptide precursor may be produced by synthetic chemistry. In an embodiment, a recombinant AEP is produced in a prokaryotic or eukaryotic cell, isolated from the cell and used *in vitro* on any polypeptide precursor to generate a cyclic peptide.

[0028] Generally, the genetic material comprises nucleic acid which may be expressed in two respective nucleic acid constructs. Alternatively, the recombinant nucleic acid encoding each of the AEP and the polypeptide precursor is expressed in a single nucleic acid construct. Multiple repeats of the same peptide or of different peptides may also be subject to cyclization processing *in vivo* or *in vitro*. Notwithstanding, a key aspect is the production of a recombinant form of AEP which is functional having peptide cyclization activity which can either be used *in vitro* with a precursor polypeptide or a cell expressing an AEP can be used as a recipient for a genetic molecule encoding the precursor polypeptide.

[0029] Enabled herein is a set of rules to enable prediction of whether an AEP is a cyclase. The set of rules is based *inter alia* on the presence or absence of residues or gaps in at least 25% of 17 predictive sites. This equates to 5 or more. The sites encompass an activity preference loop (APL), active sites and sites proximal thereto and non-active surface residues. Predictive sites are summarized in Table 2. Hence, taught herein is a method for determining whether an AEP is likely to have cyclization activity, the method comprising determining the amino acid sequence of the AEP, aligning the sequence with a best fit to the amino acid sequence of OaAEP1$_b$ (SEQ ID NO:1) and screening for the presence of 5 or more of residues or absence of residues at 139K, 161D, 186K, 192D, 247C, 248Y, 253Q, 255A, 263V, 293H, Gap, Gap, Gap, Gap, Gap (between residues 299 and 300), 314E and 316G wherein Gap means the absence of a residue wherein the presence of 5 or more of the listed residues or absence of residues is indicative of an AEP which is a cyclase. Further enabled herein is a method for determining whether an AEP is unlikely to have cyclization activity, the method comprising determining the amino acid sequence of the AEP, aligning the sequence with a best fit to the amino acid sequence of OaAEP1$_b$ (SEQ ID NO:1) and screening for the presence of 13 or more of the residues 139D, 161N, 186G, 192N, 247G, 248T, 253E, 255P, 263T, 293L,

residues aligning between residues 299 and 300 of OaAEP1$_b$ - N, G, N, Y and S, 314K and 316K wherein the presence of 13 or more of the listed residues is indicative of an AEP which is not a cyclase. The AEP may, therefore, be from any source such as but not limited to from the genus *Oldenlandia*. The AEP can be readily tested for cyclase activity. One such species is *Oldenlandia affinis*. Examples include OaAEP1$_b$ (SEQ ID NO:1), OaAEP1 (SEQ ID NO:2), OaAEP3 (SEQ ID NO:4) or a variant, derivative or hybrid form thereof which retains cyclizing activity. In an embodiment, the AEP has an amino acid sequence having at least 80% similarity to any one of SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:4 after optimal alignment and wherein the AEP comprises the presence of 5 or more of residues or absence of residues at 139K, 161D, 186K, 192D, 247C, 248Y, 253Q, 255A, 263V, 293H, Gap, Gap, Gap, Gap, Gap (between residues 299 and 300), 314E and 316G wherein Gap means the absence of a residue when optimally aligned to SEQ ID NO:1. An example of a non-cyclase AEP excluded as a cyclase under this definition is OaAEP2 (SEQ ID NO:3). It is a proviso that statements encompassing cyclase AEPs do not include OaAEP2 (SEQ ID NO:3).

[0030] When the linear precursor is produced in a prokaryotic cell the first N-terminal residue in the construct is necessarily methionine. In the event that an N-terminal methionine precludes cyclization, alternative approaches are utilized. For example:

- The endogenous methionine amino peptidase expressed by some *E. coli* strains is harnessed to remove the initiating methionine *in vivo*, revealing an N-terminus appropriate for cyclization (Camarero *et al*. (2001) *Bioorganic Med Chem 9*:2479-2484).
- A recognition sequence for a protease that cleanly releases the additional residues (e.g. TEV protease, Factor Xa) is added N-terminal to a polypeptide precursor, exposing an appropriate N-terminus for cyclization following cleavage.

[0031] In an embodiment, the cyclic peptide has one of a number of activities such as exhibiting pharmaceutical activity and includes an antipathogenic, therapeutic or uterotonic property. Examples of therapeutic activities include anticancer, protease inhibitory, antiviral or immunomodulatory activity and the treatment of pain. The cyclic peptide may also comprise a functional portion fused or embedded in a backbone framework of a cyclotide or other cyclic scaffold (Poth *et al*. (2013) *supra*). The cyclic peptide may also be generated to be topically applied to a plant or seed of a plant to protect it from pathogen infection or infestation

such as against a fungus, bacterium, nematode, mollusc, helminth, virus or protozoan organism. Alternatively, it is topically applied to human or non-human animal surfaces such as a nail, hair or skin. The polypeptide precursor may be a natural precursor for the generation of a cyclic peptide or it may not naturally become cyclic but is adapted to generate a cyclic peptide. Such a non-naturally occurring cyclic peptide may, for example, have a longer half life in a composition or when used *in vivo* or may have greater stability efficacy or utility.

[0031a] Disclosed herein is an agronomical composition or pharmaceutical composition comprising the cyclic peptide or peptide conjugate generated by the method as defined herein.

[0032] Further enabled herein is a kit comprising an AEP and a receptacle adapted to receive a polypeptide precursor and means to admix the AEP with the polypeptide precursor. Reagents may also be included to facilitate conversion of the polypeptide precursor into a cyclic peptide. Alternatively, the kit contains a eukaryotic or prokaryotic cell comprising genetic material encoding an AEP. Further genetic material encoding a polypeptide precursor to be cyclized is then introduced to that cell. An example is a yeast cell such as a *Pichia* sp.

[0033] The kit enables a useful business model for generating cyclic peptides from any linear polypeptide precursor.

[0034] A summary of sequence identifiers used throughout the subject specification is provided in Table 1.

Table 1
Summary of sequence identifiers

SEQUENCE ID NO:	DESCRIPTION
1	*Oldenlandia affinis* OaAEP1$_b$
2	*Oldenlandia affinis* OaAEP1
3	*Oldenlandia affinis* OaAEP2
4	*Oldenlandia affinis* OaAEP3
5	Amino acid sequence of model peptide R1
6	OaAEPdegen-F, 5' forward primer
7	OaAEP1-R, 5' reverse primer
8	OaAEP2-R, 5' reverse primer
9	OaAEP3-R, 5' reverse primer
10	C-terminal pro-hepta-peptide
11	Kalata B1mature + CTPP protein sequence
12	C-terminal flanking sequence for target peptide
13	Ligation partner
14	Ligation partner
15	Ligation product
16	Ligation product
17	Linker
18	Linker
19	Leaving group
20	6xHis-ubiquitin-*Oa*AEP1$_b$ fusion protein
21	Internally quenched fluorescence peptide wt
22	Nucleotide sequence encoding kalata B1 precursor protein
23	Amino acid sequence of kalata B1 precursor protein
24	Amino acid sequence of model peptide Bac2A
25	Internally quenched fluorescence peptide L31A

SEQUENCE ID NO:	DESCRIPTION
26	R1 peptide derivative
27	R1 peptide derivative
28	R1 peptide derivative
29	R1 peptide derivative
30	R1 peptide derivative
31	R1 peptide derivative
32	C-terminal AEP recognition sequence
33	N-terminal AEP recognition sequence
34	C-terminal AEP recognition sequence
35	OaAEP1b nucleic acid sequence
36	OaAEP1 na seq
37	OaAEP2 na seq
38	OaAEP3 na seq
39	OaAEP4 aa seq from transcriptomics
40	OaAEP4 na seq codon optimized for E. coli expression
41	OaAEP5 aa seq from transcriptomics
42	OaAEP5 na seq codon optimized for E. coli expression
43	OaAEP6 aa seq from transcriptomics
44	OaAEP7 aa seq from transcriptomics
45	OaAEP8 aa seq from transcriptomics
46	OaAEP9 aa seq from transcriptomics
47	OaAEP10 aa seq from transcriptomics
48	OaAEP11 aa seq from transcriptomics
49	OaAEP12 aa seq from transcriptomics
50	OaAEP13 aa seq from transcriptomics
51	OaAEP14 aa seq from transcriptomics
52	OaAEP15 aa seq from transcriptomics
53	OaAEP16 aa seq from transcriptomics

SEQUENCE ID NO:	DESCRIPTION
54	OaAEP17 aa seq from transcriptomics
55	Nicotiana tabacum NtAEP1b
56	Petunia hybrida PxAEP3a
57	Petunia hybrida PxAEP3b
58	Clitoria ternatea CtAEP1
59	Clitoria ternatea CtAEP2
60	EcAMP1 peptide derivative
61	R1 peptide derivative
62	R1 peptide derivative
63	R1 peptide derivative
64	R1 peptide derivative
65	R1 peptide derivative
66	R1 peptide derivative
67	R1 peptide derivative
68	R1 peptide derivative
69	R1 peptide derivative
70	R1 peptide derivative
71	SFTI-I10R peptide product
72	SFTI-I10R peptide + Ubiquitin + His tag
73	Kalata B1 peptide product
74	Kalata B1 + Ubiquitin + His tag
75	Vc1.1 peptide + linker product
76	Vc1.1 + linker + Ubiquitin + His tag
77	Kalata B1 + OaAEP1b aa seq
78	Kalata B1 + OaAEP1b na seq codon optimized
79	Target peptide
80	Ligation partner peptide
81	Ligated peptide product

SEQUENCE ID NO:	DESCRIPTION
82	Ligated peptide product
83	Target peptide
84	Ligated peptide product + C-terminal biotin
85	Ligated peptide product + N-terminal biotin
86	R1 peptide derivative
87	Bac2A derivative
88	Kalata B1 derivative
89	R1 peptide derivative
90	R1 peptide derivative
91	R1 peptide derivative
92	*Cicer arietinum*
93	*Medicago truncatula*
94	*Hordeum vulgare*
95	*Gossypium raimondii*
96	*Chenopodium quinoa*
97	CtAEP6
98	NaD1
99	Ligated peptide
100	Ligated peptide
101	R1 peptide derivative
102	Ligation peptide
103	R1 peptide derivative
104	Ligated peptide

BRIEF DESCRIPTION OF THE FIGURES

[0035] Some figures contain color representations or entities. Color photographs are available from the Patentee upon request or from an appropriate Patent Office. A fee may be imposed if obtained from a Patent Office.

[0036] Figure 1A is a schematic representation of the Oak1 gene product. The precursor protein encoded by the Oak1 gene (SEQ ID NO:23 encoded by SEQ ID NO:22) is proteolytically processed to produce mature kB1. The domains shown in order are: ER signal peptide (ER SP), N-terminal propeptide (NTPP), N-terminal repeat (NTR), cyclotide domain, C-terminal propeptide (CTPP). Dashed lines indicate the N- and C-terminal processing sites and a bold asterisk denotes the r*Oa*AEP1$_b$ cleavage site. The C-terminal P3-P1 and P1'-P3' sites are indicated. P1"-P3" denote the N-terminal residues that replace the P1'-P3' residues upon release of the C-terminal propeptide and subsequent backbone cyclisation. **Figure 1B** is a schematic representation of a synthetic kalata B1 precursor carrying the native C-terminal pro-hepta-peptide (GLPSLAA – SEQ ID NO:10).

[0037] Figure 2 A and B is a Clustal Omega (Sievers *et al*. (2011) *Mol. Syst. Biol 7*: 539) alignment of the full-length protein sequences of *Oa*AEP1$_b$, *Oa*AEP3, *Oa*AEP4 and *Oa*AEP5.

[0038] Figure 3 is a graphical representation showing expression of active r*Oa*AEP1$_b$ in *E. coli*. (A) Pooled r*Oa*AEP1$_b$-containing anion exchange fractions pre-and post-activation at low pH were diluted 1:14 and tested for activity against the wildtype internally quenched fluorescence (wtIQF) peptide (11 μM) [SEQ ID NO:21]. Baseline fluorescence from a no-substrate control has been subtracted and the relative fluorescence intensity (RFU) at t = 90 minutes is reported. A single representative experiment of two technical replicates is shown. (B) Activated r*Oa*AEP1$_b$ was captured by cation exchange and the final product analyzed by SDS-PAGE followed by (i) Instant blue staining and (ii) Western blotting with anti-AEP1$_b$ polyclonal rabbit serum.

[0039] Figure 4 is a representation of the amino acid sequence encoded by the *Oa*AEP1$_b$ gene isolated from *O. affinis* genomic DNA (SEQ ID NO:1). Predicted ER signal sequence shown in grey; N-terminal propeptide shown in italics; the putative signal peptidase cleavage site is indicated by an open triangle and autocatalytic processing sites are indicated by filled triangles. The mature *Oa*AEP1$_b$ cyclase domain is underlined. Cys217 and His175, presumed to be important for catalytic activity, are shown in bold and labeled with an asterisk. The dotted underline indicates possible processing sites for generation of the mature enzyme.

[0040] Figure 5A shows an alignment of the sequence region containing the activity preference loop (APL) for three AEP sequences which act preferentially as proteases (NtAEP1b (SEQ ID NO:55), PxAEP3a (SEQ ID NO:56) and OaAEP2 (SEQ ID NO:3)) and two which act preferentially as cyclases (PxAEP3b (SEQ ID NO:57) and OaAEP1$_b$ (SEQ ID NO:1). **Figure 5B** shows an alignment of OaAEP1$_b$ (preferentially a cyclase) and OaAEP2 (preferentially a protease) indicating the positions of the 17 cyclase predictive residues (or sites).

[0041] Figure 6 is a graphical representation showing the MALDI MS profile of the enzymatic processing products of a linear kB1 precursor (kB1$_{wt}$) containing the C-terminal propeptide in the presence of r*Oa*AEP1$_b$. Pre, linear precursor; Cyc, cyclic product. The +6 Da peak corresponds to the reduced form of the cyclic product.

[0042] Figure 7 is a graphical representation showing the kinetics of r*Oa*AEP1$_b$-mediated cyclisation. Varying concentrations of substrate (kB1$_{wt}$ precursor) were incubated with enzyme (19.7 µg mL^{-1} total protein) for 5 min. The amount of product formed was inferred by monitoring depletion of the precursor by RP-HPLC. A Michaelis-Menten plot shows the mean of three technical replicates and error bars report the standard error of the mean (SEM). The kinetic parameters derived from this plot are listed (±SEM).

[0043] Figure 8A is a graphical representation of the cyclization by r*Oa*AEP1$_b$ (12 µg mL^{-1} total protein) of Bac2A (RLARIVVIRVAR – SEQ ID NO:24), a linear peptide derivative

of bactenecin. The product was analysed by MALDI MS 22 hours post-addition of r*Oa*AEP1$_b$ (+ enzyme) or water (- enzyme). Bold residues, added flanking enzyme recognition sequences. Asterisk, r*Oa*AEP1$_b$ cleavage site. Observed monoisotopic masses (Da; $[M+H]^+$) are listed. +22 Da peaks likely represent Na^+ adducts. Cyc, cyclic product; Pre, linear precursor. **Figure 8B** is a graphical representation showing the MALDI MS profile of the enzymatic processing products of target peptides with additional AEP recognition residues after 5 h. The target peptides shown are (A) the R1 variant **GLP**VFAEFLPLFSKFGSRMHILK**STRNGL** (SEQ ID NO:86), and (B) the Bac2A variant **GLP**RLARIVVIRVAR**TRNGLP** (SEQ ID NO:87) with bold residues indicating additional AEP residues. The enzymes used were (i) r*Oa*AEP1$_b$, (ii) r*Oa*AEP3 , (iii) r*Oa*AEP4 and (iv) r*Oa*AEP5 and all were at a final concentration of 19.7 µg mL^{-1} total protein. A no enzyme control (v) is also shown. The expected monoisotopic mass of the cyclized variants are 3074.7 and 2042.3 Da $[M+H]^+$ for the R1 variant and the Bac2A variant respectively. The observed monoisotopic masses are listed in the figure (Da; $[M+H]^+$]). The +22 Da peak likely represents a sodium adduct.

[0044] Figure 9 is a graphical representation showing the ESI MS profile of the enzymatic processing products of EcAMP1 with additional AEP recognition residues (**GLP**GSGRGSCRSQCMRRHEDEPWRVQECVSQCRRRRGGGD**TRNGLP** (SEQ ID NO:60), bold residues indicate additional AEP recognition residues) after 5 h. The enzymes used were (i) r*Oa*AEP1$_b$, (ii) r*Oa*AEP3, (iii) r*Oa*AEP4 and (iv) r*Oa*AEP5 and all were at a final concentration of 19.7 µg mL^{-1} total protein. A no enzyme control (v) is also shown. The expected monoisotopic mass of cyclic EcAMP1 is 4892.3 Da. The observed monoisotopic masses are listed in the figure (Da).

[0045] Figure 10 is a graphical representation of the cyclisation of the R1 model peptide with various flanking sequences by bacterially expressed, recombinant AEPs. The proportion of cyclic product is displayed after cyclisation by (A) OaAEP1$_b$ (1 h incubation), (B) OaAEP3 (5 h incubation), (C) OaAEP4 (5 hr incubation) or (D) OaAEP5 (1 h incubation). In all cases, the enzyme was added at a final concentration of 19.7 µg mL^{-1} total protein. --- represents the model peptide, R1 (VFAEFLPLFSKFGSRMHILK) and

additional flanking residues are as indicated R1 Peptides: GLP---STRGLP (SEQ ID NO:26), GL---NGL (SEQ ID NO:27), GL---NG (SEQ ID NO:28), ---NGL (SEQ ID NO:29), GL---GHV (SEQ ID NO:61), GL---NHV (SEQ ID NO:62), GL---NHL (SEQ ID NO:63), GL---NGH (SEQ ID NO:64),GL---NGF (SEQ ID NO:65), GL---NFL (SEQ ID NO:66), GL---DGL (SEQ ID NO:67), LL---NGL (SEQ ID NO:89), QL---NGL (SEQ ID NO:30), KL---NGL (SEQ ID NO:31), GK---NGL (SEQ ID NO:90), GF---NGL (SEQ ID NO:91)**.** The average of three technical replicates are shown and the error bars report the standard error of the mean (SEM).

[0046] Figure 11 is a schematic representation of polypeptide ligation catalyzed by AEPs between a target peptide and a ligation partner peptide. The AEP cleavage site is indicated by ▼. For C-terminal labelling, an AEP cleavage site is incorporated into the target peptide and the ligation partner peptide contains an AEP-compatible N-terminus. For N-terminal labelling, an AEP cleavage site is incorporated into the ligation partner peptide and the target peptide contains an AEP-compatible N-terminus. AEP recognition residues added to the target peptides are shown in bold and the leaving groups are underlined.

[0047] Figure 12 is a graphical representation showing the ESI MS profile of the enzymatic processing products of a target peptide (140 μM; GLP-NaD1-TRNGLP (SEQ ID NO:79)) and ligation partner peptides (700 μM) after 6 - 22 h, as indicated. The enzymes used were (A) r*Oa*AEP1$_b$, (B) r*Oa*AEP3, (C) r*Oa*AEP4 and (D) r*Oa*AEP5 and all were at a final concentration of 19.7 μg mL^{-1} total protein. In panel (i) the ligation partner was GLPVSGE (SEQ ID NO:14). In panel (ii) the ligation partner was PLPVSGE (SEQ ID NO:80). In panel (iii) no ligation partner was added. The labelled NaD1 product has the ligation partner peptide added to the C-terminus. The expected monoisotopic mass of labelled NaD1 is 6641.3 Da when the ligation partner is GLPVSGE and 6681.3 Da when the ligation partner is PLPVSGE. The observed monoisotopic masses are listed in the figure (Da).

[0048] Figure 13 is a graphical representation showing the MALDI MS profile of the enzymatic processing products of a target peptide (140 μM; R1 variant

GKVFAEFLPLFSKFGSRMHILK**NGL (SEQ ID NO:90)**) and a ligation partner peptide (700 µM; GLK-biotin) after 6 h. The enzymes used were (A) r*Oa*AEP1**b**, (B) r*Oa*AEP3, (C) r*Oa*AEP4 and (D) r*Oa*AEP5 and all were at a final concentration of 19.7 µg mL^{-1} total protein. In panel (i) the ligation partner peptide was added. In panel (ii) no ligation partner peptide was added. The ligated product has a C-terminal biotin. The expected average mass of the biotin labelled product is 3192.9 Da $[M+H]^+$ and the observed average masses are listed in the figure (Da; $[M+H]^+$]).

[0049] Figure 14 is a graphical representation showing the MALDI MS profile of the enzymatic processing products of a target peptide (140 µM; R1 variant **GL**VFAEFLPLFSKFGSRMHILK**GHV** (SEQ ID NO:61)) and a ligation partner peptide (700 µM; biotin-TRNGL) after 6 h. The enzymes used were (A) r*Oa*AEP1**b**, (B) r*Oa*AEP3, (C) r*Oa*AEP4 and (D) r*Oa*AEP5 and all were at a final concentration of 19.7 µg mL^{-1} total protein. In panel (i) the ligation partner peptide was added. In panel (ii) no ligation partner peptide was added. The +22 Da peak is likely a sodium adduct. The ligated product has an N-terminal biotin. The expected average mass of the biotin labelled product is 3430.1 Da $[M+H]^+$ and the observed average masses are listed in the figure (Da; $[M+H]^+$]).

[0050] Figure 15 is a graphical representation of the activity of recombinant *O. affinis* AEPs (~5 µg mL^{-1} total protein) and rhuLEG (1 µg mL^{-1} total protein) over time against the fluorogenic substrate Z-AAN-MCA (100 µM). Activity is tracked at 1 minute intervals at 37°C for 60 minutes using excitation and emission wavelengths of 360 and 460 nm respectively. A single representative experiment is shown. RFU, relative fluorescence units.

[0051] Figure 16 is a graphical representation of r*Oa*AEP1$_b$ activity against the IQF peptide Abz-STRNGLPS-Y(3NO_2) [SEQ ID NO:21] in the presence of protease inhibitors. r*Oa*AEP1$_b$ (4.4 µg mL^{-1} total protein) was allowed to cleave the IQF peptide (11 µM) for 90 minutes. Enzyme activity against the IQF peptide in the presence of either the Ac-YVAD-CHO or Ac-STRN-CHO inhibitors is reported relative to a no inhibitor control at the 90 minutes time point.

[0052] Figures 17A and 17B are graphical representations of substrate specificity of plant and human AEPs for wt (SEQ ID NO:21) and L31A (SEQ ID NO:25) IQF peptide substrates. Initial velocity of recombinant *O. affinis* AEPs (~10 µg mL^{-1} total protein) (**17A**) and rhuLEG (1.1 µg mL^{-1} total protein) (**17B**) against 50 µM IQF peptide substrates is shown. The assay was conducted at 37°C. The average of two technical replicates are shown and the error bars report the range.

[0053] Figure 18A is a diagrammatic representation of a cyclotide construct for expression in *E. coli* comprising a cyclotide domain joined *via* a short linker to ubiquitin-6xHis. Filled triangle, AEP cleavage site. **Figure 18B** is a diagrammatic representation of an alternative cyclotide construct for expression in E. coli comprising a methionine followed by the kalata B1 N-terminal repeat (NTR), cyclotide domain, short linker and ubiquitin-6xHis.

[0054] Figure 19A is a graphical representation showing the MALDI MS profile of the enzymatic processing products of target peptides fused to ubiquitin. The target peptides are (A) SFTI1-I10R-ubiquitin (SEQ ID NO:72) (1 mg mL^{-1} total protein), (B) kB1-ubiquitin (SEQ ID NO:74) (0.9 mg mL^{-1} total protein) and (C) Vc1.1-ubiquitin (SEQ ID NO:76) (0.24 mg mL^{-1} total protein). The masses produced after incubation for 22 h with (i) r*Oa*AEP1$_b$ (19.7 - 98.5 µg mL^{-1}), (ii) r*Oa*AEP4 (19.7 – 30 µg mL^{-1}) or (iii) no enzyme are shown. Cyc denotes cyclic product. The +22 Da peak is likely a sodium adduct, the -16 Da peak is likely oxidized methionine, the +60 Da peak is likely cyclic product carrying both sodium (+22 Da) and potassium (+38 Da) adducts or may derive from an impurity in the preparation. **Figure 19B** is a graphical representation showing enzymatic processing of the kalata B1-ubiquitin fusion protein (SEQ ID NO:74) (260 µg mL^{-1} total protein) by different AEPs (19.7 µg mL^{-1} total protein) after a 22 h incubation. Approximately 2 µg of starting material was analysed by SDS-PAGE followed by Western blotting with an anti-6xHis mouse monoclonal antibody.

[0055] Figure 20A is a diagrammatic representation of constructs for *Pichia pastoris* transformation. Construct 1 contains the elements in a single construct and comprises, in sequence, an ER signal sequence, a vacuolar targeting signal (Vac), a cyclotide domain, a short linker and a pro-AEP domain. Construct 2 comprises an ER signal sequence, a vacuolar targeting signal, a cyclotide domain and a short linker. Construct 3 comprises an ER signal sequence, a vacuolar targeting domain and a pro-AEP domain. Constructs 2 and 3 are to be co-transformed. Filled triangles denote AEP cleavage sites; open triangles denote cleavage of the vacuolar targeting signal. **Figure 20B** is a diagrammatic representation of alternative constructs for *Pichia pastoris* transformation. Constructs 4 and 5 are identical to Constructs 1 and 2 respectively (**Figure 20A)** except for the addition of a kalata B1 N-terminal repeat (NTR) between the vacuolar targeting signal and the cyclotide domain.

[0056] Figure 21 is a graphical representation showing expression of *Oa*AEP1$_b$ in *Pichia pastoris* when kalata B1 and AEP were expressed from the same transcriptional unit (SEQ ID NOs: 77 and 78). Samples were analysed by SDS-PAGE followed by Western blotting with anti-AEP1$_b$ polyclonal rabbit serum. The negative control shows an unrelated protein expressed and extracted under the same conditions. T, total protein; L, total protein after lysis; S, soluble protein after lysis; C, concentrated soluble protein after lysis; +ve, positive control, r*Oa*AEP1$_b$ prior to activation.

[0057] Figure 22 is a schematic representation of polypeptide ligation catalyzed by r*Oa*AEP1$_b$ between a first peptide (NaD1) having a C-terminal flanking sequence incorporating the r*Oa*AEP1$_b$ cleavage site and a 6xHis tag and a second peptide containing an N-terminus compatible with r*Oa*AEP1$_b$. The leaving group on the first peptide is underlined.

DETAILED DESCRIPTION

[0058] Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or method step or group of elements or integers or method steps but not the exclusion of any other element or integer or method step or group of elements or integers or method steps.

[0059] As used in the subject specification, the singular forms "a", "an" and "the" include plural aspects unless the context clearly dictates otherwise. Thus, for example, reference to "a cyclic peptide" includes a single cyclic peptide, as well as two or more cyclic peptides; reference to "an AEP" includes a single AEP, as well as two or more AEPs; reference to "the disclosure" includes a single and multiple aspects taught by the disclosure; and so forth. Aspects taught and enabled herein are encompassed by the term "invention". All such aspects are enabled within the width of the present invention.

[0060] The present specification teaches a method of producing a cyclic peptide and a peptide conjugate. The term "cyclic peptide" encompasses but is not limited to a "cyclotide". A cyclic peptide is a peptide that is cyclic by virtue of backbone cyclization. It may be naturally cyclic or derived from a non-naturally cyclic linear polypeptide precursor. Hence, the polypeptide precursor from which the peptide is derived may be a natural substrate for cyclization or it may be a naturally linear peptide which is adapted for cyclization. The term "peptide" includes a polypeptide and a protein. For the avoidance of doubt, reference, for example, to a "cyclic peptide", "polypeptide precursor", "conjugate peptide" and the like is not to exclude a "cyclic polypeptide" or "cyclic protein", a "precursor peptide" or "precursor protein" or a "conjugate polypeptide" or "conjugate protein".

[0061] The method comprises the co-incubation either in a receptacle or in a cell of: (i) an AEP with cyclization activity; and (ii) a linear polypeptide precursor of the cyclic peptide. The AEP catalyzes the processing of the polypeptide precursor to facilitate excision and

circularization of the cyclic peptide. If in a receptacle, the cyclic peptide is purified. If cyclization is catalyzed in a cell, the cyclic peptide is isolated from a vacuole or other compartment within the cell. The term "peptide conjugate" means two or more peptides ligated together wherein at least one peptide comprises a C-terminal AEP recognition sequence and another peptide comprises an N-terminal AEP recognition sequence.

[0062] The linear polypeptide precursor comprises a C-terminal AEP processing site. Generally, but not exclusively, the C-terminal processing site is an amino acid sequence defined as comprising P3 to P1 prior to the actual cleavage site and comprising P1' to P3' after the cleavage site towards the C-terminal end. In an embodiment, P3 to P1 and P1' to P3' have the amino acid sequence:

$X_2X_3X_4X_5X_6X_7$

wherein X is an amino acid residue and:

X_2 is optional or is any amino acid;

X_3 is optional or is any amino acid;

X_4 is N or D;

X_5 is G or S;

X_6 is L or A or I; and

X_7 is optional or any amino acid.

[0063] In an embodiment, X_2 through X_7 comprise the amino acid sequence:

$X_2X_3NGLX_7$

wherein X_2 X_3 and X_7 are as defined above.

[0064] The N-terminal end of the linear polypeptide precursor may contain no specific AEP processing site or may contain a processing site defined by any one of P1" through P3" wherein P1" to P3" is defined by:

$X_9X_{10}X_{11}$

wherein X is an amino acid residue:

X_9 is optional and any amino acid or G, Q, K, V or L;

X_{10} is optional or any amino acid or L, F or I or an hydrophobic amino acid residue;

X_{11} is optional and any amino acid.

[0065] In an embodiment, X_9 through X_{11} comprise the amino acid sequence:

GLX_{11}

wherein X_{11} is defined as above.

[0066] In an embodiment, the AEP processing site comprises N- and C-terminal end sequences comprising the sequence:

GLX_{11} $[X_n]$ $X_1X_2NGLX_6$

wherein X_{11}, X_2, X_3, and X_7 are as defined above and $[X_n]$ is absent (n=0) or any amino acid residue in a sequence of from 1 to 2000 amino acids. Reference to "1 to 2000" includes 1 to 1000 and 1 to 500 such as but not limited to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250,

251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808,

809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100, 1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122, 1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1156, 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166, 1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199, 1200, 1201, 1202, 1203, 1204, 1205, 1206, 1207, 1208, 1209, 1210, 1211, 1212, 1213, 1214, 1215, 1216, 1217, 1218, 1219, 1220, 1221, 1222, 1223, 1224, 1225, 1226, 1227, 1228, 1229, 1230, 1231, 1232, 1233, 1234, 1235, 1236, 1237, 1238, 1239, 1240, 1241, 1242, 1243, 1244, 1245, 1246, 1247, 1248, 1249, 1250, 1251, 1252, 1253, 1254, 1255, 1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265, 1266, 1267, 1268, 1269, 1270, 1271, 1272, 1273, 1274, 1275, 1276, 1277, 1278, 1279, 1280, 1281, 1282, 1283, 1284,

1285, 1286, 1287, 1288, 1289, 1290, 1291, 1292, 1293, 1294, 1295, 1296, 1297, 1298, 1299, 1300, 1301, 1302, 1303, 1304, 1305, 1306, 1307, 1308, 1309, 1310, 1311, 1312, 1313, 1314, 1315, 1316, 1317, 1318, 1319, 1320, 1321, 1322, 1323, 1324, 1325, 1326, 1327, 1328, 1329, 1330, 1331, 1332, 1333, 1334, 1335, 1336, 1337, 1338, 1339, 1340, 1341, 1342, 1343, 1344, 1345, 1346, 1347, 1348, 1349, 1350, 1351, 1352, 1353, 1354, 1355, 1356, 1357, 1358, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1367, 1368, 1369, 1370, 1371, 1372, 1373, 1374, 1375, 1376, 1377, 1378, 1379, 1380, 1381, 1382, 1383, 1384, 1385, 1386, 1387, 1388, 1389, 1390, 1391, 1392, 1393, 1394, 1395, 1396, 1397, 1398, 1399, 1400, 1401, 1402, 1403, 1404, 1405, 1406, 1407, 1408, 1409, 1410, 1411, 1412, 1413, 1414, 1415, 1416, 1417, 1418, 1419, 1420, 1421, 1422, 1423, 1424, 1425, 1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435, 1436, 1437, 1438, 1439, 1440, 1441, 1442, 1443, 1444, 1445, 1446, 1447, 1448, 1449, 1450, 1451, 1452, 1453, 1454, 1455, 1456, 1457, 1458, 1459, 1460, 1461, 1462, 1463, 1464, 1465, 1466, 1467, 1468, 1469, 1470, 1471, 1472, 1473, 1474, 1475, 1476, 1477, 1478, 1479, 1480, 1481, 1482, 1483, 1484, 1485, 1486, 1487, 1488, 1489, 1490, 1491, 1492, 1493, 1494, 1495, 1496, 1497, 1498, 1499, 1500, 1501, 1502, 1503, 1504, 1505, 1506, 1507, 1508, 1509, 1510, 1511, 1512, 1513, 1514, 1515, 1516, 1517, 1518, 1519, 1520, 1521, 1522, 1523, 1524, 1525, 1526, 1527, 1528, 1529, 1530, 1531, 1532, 1533, 1534, 1535, 1536, 1537, 1538, 1539, 1540, 1541, 1542, 1543, 1544, 1545, 1546, 1547, 1548, 1549, 1550, 1551, 1552, 1553, 1554, 1555, 1556, 1557, 1558, 1559, 1560, 1561, 1562, 1563, 1564, 1565, 1566, 1567, 1568, 1569, 1570, 1571, 1572, 1573, 1574, 1575, 1576, 1577, 1578, 1579, 1580, 1581, 1582, 1583, 1584, 1585, 1586, 1587, 1588, 1589, 1590, 1591, 1592, 1593, 1594, 1595, 1596, 1597, 1598, 1599, 1600, 1601, 1602, 1603, 1604, 1605, 1606, 1607, 1608, 1609, 1610, 1611, 1612, 1613, 1614, 1615, 1616, 1617, 1618, 1619, 1620, 1621, 1622, 1623, 1624, 1625, 1626, 1627, 1628, 1629, 1630, 1631, 1632, 1633, 1634, 1635, 1636, 1637, 1638, 1639, 1640, 1641, 1642, 1643, 1644, 1645, 1646, 1647, 1648, 1649, 1650, 1651, 1652, 1653, 1654, 1655, 1656, 1657, 1658, 1659, 1660, 1661, 1662, 1663, 1664, 1665, 1666, 1667, 1668, 1669, 1670, 1671, 1672, 1673, 1674, 1675, 1676, 1677, 1678, 1679, 1680, 1681, 1682, 1683, 1684, 1685, 1686, 1687, 1688, 1689, 1690, 1691, 1692, 1693, 1694, 1695, 1696, 1697, 1698, 1699, 1700, 1701, 1702, 1703, 1704, 1705, 1706, 1707, 1708, 1709, 1710, 1711, 1712, 1713, 1714, 1715, 1716, 1717, 1718,

1719, 1720, 1721, 1722, 1723, 1724, 1725, 1726, 1727, 1728, 1729, 1730, 1731, 1732, 1733, 1734, 1735, 1736, 1737, 1738, 1739, 1740, 1741, 1742, 1743, 1744, 1745, 1746, 1747, 1748, 1749, 1750, 1751, 1752, 1753, 1754, 1755, 1756, 1757, 1758, 1759, 1760, 1761, 1762, 1763, 1764, 1765, 1766, 1767, 1768, 1769, 1770, 1771, 1772, 1773, 1774, 1775, 1776, 1777, 1778, 1779, 1780, 1781, 1782, 1783, 1784, 1785, 1786, 1787, 1788, 1789, 1790, 1791, 1792, 1793, 1794, 1795, 1796, 1797, 1798, 1799, 1800, 1801, 1802, 1803, 1804, 1805, 1806, 1807, 1808, 1809, 1810, 1811, 1812, 1813, 1814, 1815, 1816, 1817, 1818, 1819, 1820, 1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828, 1829, 1830, 1831, 1832, 1833, 1834, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1842, 1843, 1844, 1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856, 1857, 1858, 1859, 1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869, 1870, 1871, 1872, 1873, 1874, 1875, 1876, 1877, 1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888, 1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903, 1904, 1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918, 1919, 1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948, 1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999 and 2000.

[0067] In an embodiment, the C-terminal processing site comprises P4 to P1 and P1' to P4' wherein P1 to P4 and P1' to P4' comprise $X_1X_2X_3X_4X_5X_6X_7X_8$ wherein X_2 to X_7 are as defined above and X_7 is optional or any amino acid and X_8 is optional or any amino acid.

[0068] The present invention comprises various aspects in relation to the co-incubation of the AEP with cyclization activity and the linear polypeptide precursor which include:

(i) introducing into a prokaryotic or eukaryotic cell a genetic vector encoding an AEP which is expressed then the AEP isolated and used in an *in vitro* cyclization reaction to generate a cyclic peptide from a linear polypeptide precursor;

(ii) introducing into a prokaryotic or eukaryotic cell a genetic vector encoding a linear polypeptide precursor which is expressed and purified, optionally post-translationally modified to introduce a non-naturally occurring amino acid residue and then subject to cyclization *in vitro* using an AEP to form a cyclic peptide, this includes modifications in the cell such as the production of isotopically-labeled peptides; and

(iii) introducing into a prokaryotic or eukaryotic cell, single or multiple genetic vectors encoding an AEP and a polypeptide precursor which enables production of a cyclic peptide in a vacuole or other cellular compartment of the cell.

[0069] Aspect (ii) can be modified whereby the linear polypeptide precursor is synthetically produced or isolated from a particular source. A linear peptide conjugate can be generated *in vitro* or in *in vivo*. In the case of eukaryotic cells, the AEP and linear polypeptide precursor may be produced in different cells or different cellular compartments of the same cell, isolated then used *in vitro*. In the case of Aspect (ii), in a prokaryotic cell, in a non-limiting embodiment, the cyclic peptide is generated by co-expression with an AEP in the periplasmic space. The polypeptide precursor may be a natural substrate for cyclization or may normally be a linear peptide that is rendered cyclic. Making a cyclic form of a linear peptide can improve stability, efficacy and utility.

[0070] By "co-incubation" is meant co-incubation *in vitro* in a receptacle or reaction vessel as well as within a cell. In addition, the AEP also has ligase activity enabling the generation of peptide conjugates of at least two peptides wherein at least one peptide comprises a C-terminal AEP recognition sequence and at least one other peptide comprising an N-terminal AEP recognition sequence.

[0071] Hence, enabled herein is a method for producing a cyclic peptide, the method comprising introducing into the prokaryotic or eukaryotic cell genetic material which, when expressed, generates an AEP with cyclization ability, isolating the AEP and then incubating the AEP with a polypeptide precursor, optionally incorporating a post-translational modification to introduce a non-naturally occurring amino acid residue or cross-linkage bond or other modification for a time and under conditions sufficient to

generate a cyclic peptide from the polypeptide precursor; or co-expressing genetic material encoding the AEP with cyclization ability and a linear polypeptide precursor in a prokaryotic or eukaryotic cell for a time and under conditions sufficient to generate a cyclic peptide in a vacuole or other cellular compartment of the cell. In addition, the AEP can catalyze a ligation reaction to conjugate two or more peptides wherein at least one peptide comprises a C-terminal AEP recognition sequence and another peptide comprises an N-terminal AEP recognition sequence. The cell can also be used to generate one or both of the AEP and/or polypeptide precursor for use in the generation of a cyclic peptide *in vitro*. In an embodiment, the cyclic peptide is produced by co-expression of an AEP with cyclization ability and a target polypeptide in the periplasmic space of a prokaryotic cell.

[0072] Further enabled herein is a method of generating a linear peptide conjugate the method comprising co-incubating two or more peptides wherein at least one peptide comprises a C-terminal AEP recognition sequence and at least one other peptide comprises an N-terminal AEP recognition sequence with an AEP for a time and under conditions sufficient for at least two peptides to ligate together to form a peptide conjugate.

[0073] As indicated above, reference to a "peptide" includes a polypeptide and a protein. No limitation in the size or type of proteinaceous molecule is intended by use of the terms "peptide", "polypeptide" or "protein".

[0074] A "vector" refers to a recombinant plasmid or virus that comprises a polynucleotide to be delivered into a host cell. The polynucleotide to be delivered comprises a coding sequence of AEP and/or the polypeptide precursor or multiple forms of the same or different peptides. The term includes vectors that function primarily for introduction of DNA or RNA into a cell and expression vectors that function for transcription and/or translation of the DNA or RNA. Also included are vectors that provide more than one of the above functions.

[0075] A vector in relation to a prokaryotic or eukaryotic cell includes a multi-gene

expression vehicle. Such as a vehicle consists of a polynucleotide comprising two or more transcription unit segments, each segment encoding an AEP or linear polypeptide precursor, each segment being joined to the next in a linear sequence by a linker segment encoding a linker peptide, the transcription segments all being in the same reading frame operably linked to a single promoter. Multiple polypeptide repeats or multiple different polypeptides may also be generated. A vector also includes a viral expression vector which comprises a viral genome with a modified nucleotide sequence which encodes a protein and enable stable expression. Alternatively, multiple vectors are used each encoding either an AEP or linear polypeptide precursor.

[0076] A "transcription unit" is a nucleic acid segment capable of directing transcription of a polynucleotide or fragment thereof. Typically, a transcription unit comprises a promoter operably linked to the polynucleotide that is to be transcribed, and optionally regulatory sequences located either upstream or downstream of the initiation site or the termination site of the transcribed polynucleotide. Alternatively, as a multigene expression vehicle, a single promoter and terminator is used to produce more than one protein from a single transcription unit A transcription unit includes a unit encoding either an AEP or a polypeptide precursor, or both.

[0077] A eukaryotic cell includes a yeast, a filamentous fungus and a plant cell. A "yeast cell" includes a species of *Pichia* such as but not limited to *Pichia pastoris* as well as *Saccharomyces or Kluyveromyces*. Other eukaryotic cells include non-human mammalian cells and insect cells. A prokaryotic cell includes an *E. coli* or some other prokaryotic microorganism suitable for production of recombinant proteins.

[0078] A "host" cell encompasses a prokaryotic cell (e.g. *E. coli*) or eukaryotic cell (e.g. a yeast cell such as a species of *Pichia*).

[0079] The terms "nucleic acid", "polynucleotide" and "nucleotide" sequences are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. The following are non-

limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from the lineage of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. The polynucleotide encodes an AEP or linear polypeptide precursor including a linear precursor of a protein to be cyclized or two linear peptides to be ligated or any selectable marker.

[0080] A "gene" refers to a polynucleotide containing at least one open reading frame that is capable of encoding an AEP or polypeptide precursor after being transcribed and translated.

[0081] As used herein, "expression" refers to the process by which a polynucleotide transcription unit is transcribed into mRNA and/or the process by which the transcribed mRNA (also referred to as "transcript") is subsequently translated into an AEP or polypeptide precursor. The transcripts and the encoded polypeptides are collectedly referred to as a "gene product".

[0082] In the context of a linear polypeptide precursor, a "linear" sequence is an order of amino acids in the polypeptide in an N- to C-terminal direction in which amino acid residues that neighbour each other in the sequence are contiguous in the primary structure of the polypeptide. The "precursor" means it is a substrate for the AEP to generate a cyclic peptide. A linear peptide conjugate is generated following ligation of at least two peptides wherein at least one peptide comprises a C-terminal AEP recognition amino acid sequence and at least one peptide comprises an N-terminal AEP recognition amino acid sequence.

[0083] A "pathogen" includes a plant or animal or human pathogen selected from a fungus, insect, bacterium, nematode, helminth, mollusc, virus and a protozoan organism.

[0084] Enabled herein is a method for producing a cyclic peptide the method comprising

co-incubating an AEP with peptide cyclizing activity and a linear polypeptide precursor of the cyclic peptide for a time and under conditions sufficient to generate the cyclic peptide. The co-incubation may occur in a receptacle (*in vitro*) or in a cell such as the vacuole or other cellular compartment of a cell. If the co-incubation is *in vitro*, then the AEP or the linear polypeptide precursor is produced in a prokaryotic or eukaryotic cell. The linear polypeptide precursor may also be produced in a cell and isolated and optionally post-translationally modified or synthetically generated to incorporate a non-naturally occurring amino acid residue or a non-naturally occurring cross-linkage bond or to be isotopically labeled. If co-incubation occurs in a cell, this may occur in a vacuole or other compartment of a eukaryotic cell or in a periplasmic space of a prokaryotic cell.

[0085] AEPs from cyclotide producing plants have been identified that, when expressed with the precursor gene for the cyclotide kalata B1 (oak1), and other peptides, are effective at backbone cyclization. By comparing the amino acid sequences of ligation competent AEPs with those favouring proteolysis, a differential loop region, termed the activity preference loop (APL), has been identified that contributes to the specificity. In ligase competent AEPs, the APL either has several residues missing or is replaced by hydrophobic stretch of amino acids (Figure **5A**).

[0086] Additional residues linked to cyclase function are identified by machine learning (protein sequence space analysis) using a set of experimentally determined cyclase and non-cyclase sequences. The following residues are found to be highly predictive of cyclase function in the currently known cyclases and non-cyclases. All numbering is given relative to *OaAEP1$_b$* (Figure 4; SEQ ID NO:1).

1. **APL** - The absence of residues in the region between **299 - 300** of OaAEP1 is predictive of a higher likelihood of cyclase activity.
2. **Set 1** - The presence of the following active site residues is also predictive of a higher likelihood of cyclase activity:

- **D161**
- **C247**

- **Y248**
- **Q253**
- **A255**
- **V263**

3. **Set 2** - The presence of the following active site-proximal residues is also predictive of a higher likelihood of cyclase activity:

- **K186**
- **D192**

4. **Set 3** - The presence of the following non-active site surface residues is also predictive of a higher likelihood of cyclase activity:

- **K139**
- **H293**
- **E314**
- **G316**

[0087] Overall it is highly predictive of cyclase activity if the sequence contains either:

- The shortened APL
- 3 of the 6 Set 1 active site residues
- Both of the Set 2 active-site-proximal residues
- 3 of the 4 Set 3 non-active-site residues

[0088] The most predictive are the APL and set 1. The more of these criteria that it hits, the more likely that it is to be a cyclase. Predictive residues for cyclase activity are shown in Table 2. Residue numbering is relative to $OaAEP1_b$ (Figure 4; SEQ ID NO:1). Residue properties that strongly predict cyclase activity are disorder propensity (DISORD), net static charge (CHRG), molecular weight of R group (RMW), and hydropathy index (HPATH).

[0089] An AEP having at least 25% or 5 or more of the 17 predictive residues set forth in Table 2 is considered likely to act preferentially as a cyclase. A requirement for at least

25% of the predictive residues to be present enables 100% of the known cyclases to be correctly identified while excluding known non-cyclases at least 80% of the time including at least 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93 or 94% of the time. In an embodiment, the rule established herein enables exclusion of non-cyclases 94% of the time. One AEP excluded for being a non-cyclase is OaAEP2 (SEQ ID NO:3).

[0090] Accordingly, taught herein is a method for determining whether an AEP is likely to have cyclization activity, the method comprising determining the amino acid sequence of the AEP, aligning the sequence with a best fit to the amino acid sequence of $OaAEP1_b$ (SEQ ID NO:1) and screening for the presence of 5 or more residues or absence of residues at 139K, 161D, 186K, 192D, 247C, 248Y, 253Q, 255A, 263V, 293H, Gap, Gap, Gap, Gap, Gap (between residues 299 and 300), 314E and 316G wherein gap means the absence of a residue wherein the presence of 5 or more of the listed residues or absence of residues is indicative of an AEP which is a cyclase.

[0091] In an embodiment, the from 5 to 17 residues or gaps screened at the listed sites include 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16 or 17 residues or gaps. In an alternative representation, if the listed residues have at least 25% of the residues or gaps listed, then the AEP is deemed a cyclase. By "at least 25%" means 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100%.

[0092] In yet a further alternative, the presence of 13 or more or 75% of the residues 139D, 161N, 186G, 192N, 247G, 248T, 253E, 255P, 263T, 293L, residues aligning between residues 299 and 300 of $OaAEP1_b$ - N, G, N, Y and S, 314K and 316K is indicative of an AEP which is a non-cyclase. Reference to "13 or more" means 13, 14, 15, 16 and 17. Refernece to "at least 75%" means 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100%. Accordingly, enabled herein is a a method for determining whether an AEP is unlikely to have cyclization activity, the method comprising determining the amino acid sequence of the AEP, aligning the

sequence with a best fit to the amino acid sequence of OaAEP1$_b$ (SEQ ID NO:1) and screening for the presence of 13 or more of the residues 139D, 161N, 186G, 192N, 247G, 248T, 253E, 255P, 263T, 293L, residues aligning between residues 299 and 300 of OaAEP1$_b$ - N, G, N, Y and S, 314K and 316K wherein the presence of 13 or more of the listed residues is indicative of an AEP which is not a cyclase.

[0093] The present invention extends to any AEP with peptide cyclization activity such as those defined above. Encompassed, herein, is any other AEP such as, but not limited to, OaAEP1$_b$ (SEQ ID NO:1), OaAEP1 (SEQ ID NO:2) and OaAEP3 (SEQ IDNO:4) from *Oldenlandia affinis*. Other AEPs include an AEP having at least 80% amino acid similarity to SEQ ID NO:1 (OaAEP1$_b$), SEQ ID NO:2 (OaAEP1) or SEQ ID NO:4 (OaAEP3) after optimal alignment and which retains AEP and peptide cyclization activity and when the AEP comrpises the presence of 5 or more of residues or absence of residues at 139K, 161D, 186K, 192D, 247C, 248Y, 253Q, 255A, 263V, 293H, Gap, Gap, Gap, Gap, Gap (between residues 299 and 300), 314E and 316G wherein Gap means the absence of a residue when optimally aligned to SEQ ID NO:1. The AEP may also have ligase activity to facilitate generation of peptide conjugates. OaAEP2 (SEQ ID NO:3) is an example of an AEP which is not a cyclase. It is a proviso that statements encompassing cyclase AEPs do not include OaAEP2 (SEQ ID NO:3).

[0094] In a prokaryotic cell, the first N-terminal residue in a construct is necessarily methionine. In the event that an N-terminal methionine precludes cyclization, alternative approaches are utilized. For example:

- The endogenous methionine amino peptidase expressed by prokaryotic cells is harnessed to remove the initiating methionine *in vivo*, revealing an N-terminus appropriate for cyclization (Camarero *et al*. (2001) *supra*).
- A recognition sequence for a protease that cleanly releases the additional residues (e.g. TEV protease, Factor Xa) is added N-terminal to the polypeptide precursor, exposing an appropriate N-terminus for cyclization following cleavage.

[0095] Reference to "at least 80%" includes 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91,

92, 93, 94, 95, 96, 97, 98, 99 and 100%.

[0096] The term "similarity" as used herein includes exact identity between compared sequences at the amino acid level. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. In a particularly preferred embodiment, amino acid sequence comparisons are made at the level of identity rather than similarity.

[0097] Terms used to describe sequence relationships between two or more polypeptides include "reference sequence", "comparison window", "sequence similarity", "sequence identity", "percentage of sequence similarity", "percentage of sequence identity", "substantially similar" and "substantial identity". A "reference sequence" includes from at least 10 amino acid residues (e.g. from 10 to 100 amino acids). A "comparison window" refers to a conceptual segment of typically 10 contiguous amino acid residues that is compared to a reference sequence. The comparison window may comprise additions or deletions (i.e. gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerized implementations of algorithms (BLASTP 2.2.32+, GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA) or by inspection and the best alignment (i.e. resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as for example disclosed by Altschul *et al.* (1997) *Nucl. Acids. Res. 25:* 3389-3402). A detailed discussion of sequence analysis can be found in Unit 19.3 of Ausubel *et al.* (*In: Current Protocols in Molecular Biology*, John Wiley & Sons Inc. 1994-1998).

[0098] The terms "sequence identity" and "sequence similarity" as used herein refers to the extent that sequences are identical or functionally or structurally similar on an amino acid-

by-amino acid basis over a window of comparison. Thus, a "percentage of sequence identity", for example, is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical amino acid residue (e.g. Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e. the window size), and multiplying the result by 100 to yield the percentage of sequence identity. For the purposes of the present invention, "sequence identity" will be understood to mean the "match percentage" calculated by the BLASTP 2.2.32+ computer program using standard defaults. Similar comments apply in relation to sequence similarity.

[0099] In an embodiment, taught herein is a method for producing a cyclic peptide the method comprising co-incubating an AEP with peptide cyclization activity having an amino acid sequence with at least 80% similarity to a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:4, after optimal alignment and wherein the presence of 5 or more of residues or absence of residues at 139K, 161D, 186K, 192D, 247C, 248Y, 253Q, 255A, 263V, 293H, Gap, Gap, Gap, Gap, Gap (between residues 299 and 300), 314E and 316G wherein Gap means the absence of a residue and a linear polypeptide precursor of the cyclic peptide for a time and under conditions sufficient to generate the cyclic peptide.

[0100] In an embodiment, enabled herein is a method for producing a cyclic peptide *in vitro* the method comprising introducing into a prokaryotic cell an expression vector encoding an AEP, enabling expression of the vector to produce a recombinant AEP, isolating the AEP from the cell and co-incubating in a reaction vessel the recombinant AEP with a polypeptide precursor for a time and under conditions sufficient to generate the cyclic peptide.

[0101] Taught herein is a method for producing a cyclic peptide *in vitro* the method comprising introducing into a prokaryotic or eukaryotic cell an expression vector encoding

one or other of an AEP with peptide cyclization activity and a linear polypeptide precursor, enabling expression of the vector to produce a recombinant AEP and recombinant linear polypeptide precursor in the cell or component of the cell or a periplasmic space and isolating a cyclic peptide generated from the polypeptide precursor.

[0102] Enabled herein is a method for producing a cyclic peptide *in vitro* the method comprising introducing into a prokaryotic or eukaryotic cell an expression vector encoding an AEP with peptide cyclization activity, isolating the AEP and co-incubating in a reaction vessel the AEP with a polypeptide precursor for a time and under conditions sufficient to generate the cyclic peptide.

[0103] The polypeptide precursor may be recombinant or synthetically produced. The recombinant polypeptide may be post-translationally modified to introduce, or the synthetic form may incorporate, a non-naturally occurring amino acid.

[0104] Enabled herein is a method for producing a cyclic peptide *in vivo* the method comprising introduction into a prokaryotic or eukaryotic cell an expression vector encoding an AEP with peptide cyclization activity and a linear polypeptide precursor, enabling expression of the vector to produce the AEP and linear polypeptide precursor to produce a cyclic peptide. In an embodiment, this may occur in a periplasmic space or in a cellular compartment such as a vacuole.

[0105] In an embodiment, taught herein is a method for producing a cyclic peptide *in vitro* the method comprising introducing into a prokaryotic or eukaryotic cell an expression vector encoding one or other of an AEP with peptide cyclization activity or a linear polypeptide precursor, enabling expression of the vector to produce a recombinant AEP or recombinant linear polypeptide precursor and isolating the AEP or polypeptide from the cell and co-incubating in a reaction vessel the recombinant AEP with a polypeptide precursor or a post-translationally modified or synthetically modified form thereof for a time and under conditions sufficient to generate the cyclic peptide.

[0106] In an embodiment, the AEP comprises an amino acid sequence having at least 80% similarity to any one or more of SEQ ID NOs:1, 2 and/or 4 after optimal alignment.

[0107] As indicated above, reference to "at least 80%" means 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100%.

[0108] In another embodiment, a linear peptide is generated using the ligase activity of an AEP. In this embodiment, a first peptide comprising the C-terminal AEP recognition amino acid sequence is co-incubated with a second peptide with an N-terminal AEP recognition amino acid sequence and which may or may not have a tag and an AEP. The AEP catalyzes a ligation between the first and second peptides to generate a linear peptide conjugate. This may then subsequently be cyclized into a cyclic peptide or used as a linear peptide. This may occur *in vitro* or *in vivo*.

[0109] The polypeptide precursor may be a recombinant molecule generated by expression of nucleic acid encoding same in a cell or a combination of being produced by recombinant means followed by a post-translational modification (e.g. isotopically labeled) or produced by synthetic means. In relation to a post-translation modification or synthetic form, a non-naturally occurring amino acid may be introduced. A "cell" includes a prokaryotic (e.g. *E. coli*) or eukaryotic (e.g. a yeast) cell. The nucleic acid encoding the AEP and the polypeptide precursor may be present in two separate nucleic acid constructs or be part of a single construct such as a multi-gene expression vehicle. In either event, the nucleic acid is operably linked to a promoter which enables expression of the nucleic acid to produce the AEP and/or a linear form of the polypeptide precursor which is then processed into the cyclic peptide either *in vitro* or *in vivo* in a vacuole or other cellular compartment. In another embodiment, cells are maintained which are genetically modified to produce the AEP and these cells are then hosts for any given nucleic acid encoding a polypeptide precursor.

[0110] Taught herein is a method for producing a cyclic peptide in a cell, the method comprising introducing a genetic vector into the cell, the genetic vector comprising

polynucleotide segments each encoding either an AEP with peptide cyclization activity or a polypeptide precursor, the polynucleotide segments separated by a polynucleotide linker segment wherein all polynucleotide segments are in the same reading frame operably linked to a single promoter and terminator wherein the eukaryotic cell is grown for a time and under conditions sufficient for a cyclic peptide to be generated which is then isolated from the vacuole or other cellular compartment.

[0111] Further taught herein is a method for producing a cyclic peptide in a cell, the method comprising introducing two genetic vectors in the cell, one encoding an AEP with peptide cyclization activity and the other encoding a polypeptide precursor, each genetic molecule comprising a promoter and terminator operably linked to polynucleotides encoding either the AEP or the polypeptide precursor wherein the cell is grown for a time and under conditions sufficient for a cyclic peptide to be generated which is then isolated from the vacuole or other cellular compartment.

[0112] In another embodiment, the vector encodes multiple repeats of the same polypeptide to be cyclized or multiple forms of different polypeptides to be cyclized.

[0113] In an embodiment, the AEP includes an AEP having at least 80% similarity to one or more of SEQ ID NOs:1, 2 and/or 4 after optimal alignment and wherein the presence of 5 or more of residues or absence of residues at 139K, 161D, 186K, 192D, 247C, 248Y, 253Q, 255A, 263V, 293H, Gap, Gap, Gap, Gap, Gap (between residues 299 and 300), 314E and 316G wherein Gap means the absence of a residue. Again, reference to "at least 80%" means 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100%. OaAEP2 (SEQ ID NO:3) is an example of a non-cyclase AEP. A eukaryotic cell includes a yeast cell such as a species of *Pichia* or a *Saccharomyces* sp. or *Kluyveromyces* sp.

[0114] Further taught herein is a method for generating a peptide conjugate comprising two or more peptides, the method comprising co-incubating at least two peptides with an AEP wherein at least one peptide comprises a C-terminal AEP recognition amino acid

sequence and at least one other peptide comprises an N-terminal AEP recognition amino acid sequence. This may occur *in vitro* or *in vivo*.

[0115] Techniques and agents for introducing and selecting for the presence of a vector in cells are well-known. Genetic markers allowing for the selection of the vector cells are well-known, e.g. genes carrying resistance to an antibiotic such as kanamycin, tetracycline or ampicillin. The marker allows for selection of successfully transformed cells growing in the medium containing the appropriate antibiotic because they will carry the corresponding resistance gene. Eukaryotic cell selection of transformed cells is often accomplished through the inclusion of auxotrophic markers in the vector such as HIS4 or URA3 which encode enzymes involved in synthesis of essential amino acids or nucleotides. These vectors are then transformed into a yeast strain that is unable to synthesize specific amino acids or nucleotides that are required for growth, such as histidine for HIS4 and uracil for URA3. Cells that have been successfully transformed with the vector are selected by plating on dropout media lacking the specific amino acid or nucleotide as the untransformed cells are not able to synthesize the essential amino acid or nucleotide that is not present in the growth medium whereas cells carrying the vector with the auxotrophic marker survive as they are able to synthesize the missing amino acid or nucleotide. Other common auxotrophic markers are LEU2, LYS2, TRP1, HIS3, ARG4, ADE2.

[0116] Techniques for introducing an expression vector comprising a promoter operably linked to a polynucleotide into cell are varied and include transformation, electroporation, microinjection, particle bombardment or other techniques known to the art.

[0117] The choice of vector in which the nucleic acid encoding the AEP or polypeptide precursor is operatively linked depends directly, as is well known in the art, on the functional properties desired, e.g. replication, protein expression, and the host cell to be transformed, these being limitations inherent in the art of constructing recombinant nucleic acid molecules. For prokaryotic cells, the vector desirably includes a prokaryotic replicon, i.e. a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extra-chromosomally when introduced into a

prokaryotic cell. Such replicons are well known in the art. For eukaryotic cells, for example, the vector could either be maintained extra-chromosomally, in which case the vector sequence would generally comprise a eukaryotic replicon, or could be incorporated into the genomic DNA, in which case the vector would include sequences that would facilitate recombination of the vector into the host chromosome.

[0118] Those vectors that include a prokaryotic replicon also typically include convenient restriction sites for insertion of a recombinant DNA molecule of the present invention. Typical of such vector plasmids are pUC8, pUC9, pBR322, and pBR329 available from BioRad Laboratories (Richmond, CA) and pPL, pK and K223 available from Pharmacia (Piscataway, NJ), and pBLUESCRIPT tm and pBS available from Stratagene (La Jolla, CA). A vector of the present invention may also be a Lambda phage vector as known in the art or a Lambda ZAP vector (available from Stratagene La Jolla, CA). Another vector includes, for example, pCMU (Nilsson *et al*. (1989) *Cell 58:*707-718). Other appropriate vectors may also be synthesized, according to known methods; for example, vectors pCMU/Kb and pCMUII used in various applications herein are modifications of pCMUIV (Nilsson *et al*. (1989) *supra*). The nucleic acid may be DNA or RNA.

[0119] Once introduced into a suitable host cell, expression of the nucleic acid can be determined using any assay known in the art. For example, the presence of a transcribed polynucleotide can be detected and/or quantified by conventional hybridization assays (e.g. Northern blot analysis), amplification procedures (e.g. RT-PCR), SAGE (U.S. Pat. No. 5,695,937), and array-based technologies (see e.g. U.S. Pat. Nos. 5,405,783, 5,412,087 and 5,445,934). The polynucleotide encodes the AEP or polypeptide precursor or in the case of a eukaryotic system the polynucleotide may encode both.

[0120] Expression of the nucleic acid can also be determined by examining the protein product. A variety of techniques are available in the art for protein analysis. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunosorbent assays), "sandwich" immunoassays, immunoradiometric assays, *in situ* immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), Western blot analysis, immunoprecipitation

assays, immunofluorescent assays, and PAGE-SDS. In an embodiment, mass spectrometry is used for cyclic peptides (Saska *et al*. (2008) *Journal of Chromatography B. 872*:107-114).

[0121] In general, determining the protein level involves (a) providing a biological sample containing polypeptides; and (b) measuring the amount of any immunospecific binding that occurs between an antibody reactive to the AEP or polypeptide precursor, in which the amount of immunospecific binding indicates the level of expressed proteins. Antibodies that specifically recognize and bind to AEP or linear polypeptide precursor are required for immunoassays. These may be purchased from commercial vendors or generated and screened using methods well known in the art. See Harlow and Lane (1988) *Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratories.* and Sambrook *et al*. (1989) *Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory, Plainview, New York*. The sample of test proteins can be prepared by homogenizing the prokaryotic or eukaryotic transformants and optionally solubilizing the test protein using detergents, such as non-reducing detergents which include triton and digitonin. The binding reaction in which the AEP or polypeptide precursor is allowed to interact with the detecting antibodies may be performed in solution, or in cell pellets and/or isolated cells, for example, a solid support that has been immobilized with the test proteins. The formation of the complex can be detected by a number of techniques known in the art. For example, the antibodies may be supplied with a label and unreacted antibodies may be removed from the complex; the amount of remaining label thereby indicating the amount of complex formed. Results obtained using any such assay on a sample from a cell transformant is compared with those from a non-transformed source as a control. Other protein quantitation methods such as BCA and nanodrop methodologies may be employed.

[0122] The prokaryotic or eukaryotic host cells of this invention are grown under favorable conditions to effect expression of the polynucleotide. Examples of prokaryotic cells include *E. coli*, *Salmonella* sp, *Pseudomonas* sp and *Bacillus* sp. Examples of eukaryotic cells include yeast such as *Pichia* spp.(e.g. *Pichia pastoris*), *Saccharomyces* spp. or *Kluyveromyces* spp.

[0123] Accordingly, this invention provides genetically modified cells carrying one or two vectors encoding an AEP and/or a polypeptide precursor.

[0124] The present invention further contemplates a business model for producing cyclic peptides. In one embodiment, the business model comprises a prokaryotic cell encoding a heterologous AEP with cyclizing activity or a prokaryotic cell for use in introducing and expressing a vector encoding a desired linear polypeptide precursor. In either case, the polypeptide precursor produced by recombinant or synthetic means and the AEP are co-incubated in a reaction vessel for a time and under conditions sufficient for a cyclic peptide to be generated from the polypeptide precursor. In another embodiment, a prokaryotic or eukaryotic cell is selected for transformation with a vector encoding an AEP and a polypeptide precursor either in the same or separate constructs or the eukaryotic cell already comprises an AEP-encoding vector and is used as a recipient for a selected vector encoding a polypeptide precursor. The cell is then incubated for a time and under conditions sufficient for a cyclic peptide to form which can be isolated from the vacuole of the eukaryotic cell. The eukaryotic cell may be used to generate an AEP and/or polypeptide precursor which is used *in vitro*. In a further embodiment, the business model extends to the generation of linear peptide conjugates.

[0125] The cyclic peptides may have any of a range of useful properties including antipathogen, therapeutic or other pharmaceutically useful properties and/or insecticidal, molluscicidal or nematocidal activity. Examples of therapeutic activities include anticancer, protease inhibitory, antiviral or immunomodulatory activity and the treatment of pain. The cyclic peptide may also be a framework to incorporate a functionality. A normally linear polypeptide may also be subject to cyclization. This can improve stability, efficacy and utility. Alternatively, the polypeptide precursor is a natural substrate for cyclization.

[0126] As contemplated herein, a non-naturally occurring amino acid may be introduced into the polypeptide precursor. These include amino acids with a modified side chain.

[0127] Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with $NaBH_4$; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with $NaBH_4$.

[0128] The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

[0129] The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitization, for example, to a corresponding amide.

[0130] Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline pH.

[0131] Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form a 3-nitrotyrosine derivative.

[0132] Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with

diethylpyrocarbonate.

[0133] Examples of incorporating unnatural amino acids and derivatives during polypeptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids.

[0134] Crosslinkers can be used, for example, to stabilize 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with n = 1 to n = 6, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of C_α and N $_\alpha$-methylamino acids, introduction of double bonds between C_α and C_β atoms of amino acids.

[0135] The polypeptide precursor may also be isotopically labeled by a cell or during *in vitro* synthesis.

[0136] Further enabled herein is a pharmaceutical formulation comprising the cyclic peptide or linear peptide conjugate or a pharmaceutically acceptable salt thereof. Such a formulation has applications in treating human and non-human animal subjects.

[0137] The term "pharmaceutically acceptable salts" refers to physiologically and pharmaceutically acceptable salts of the peptides of the invention: i.e., salts that retain the desired biological activity of the parent compound and do not impart undesired toxicological effects thereto.

[0138] The pharmaceutical compositions of the present invention may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon

the area to be treated. Administration may be topical (including ophthalmic and to mucous membranes including vaginal and rectal delivery), pulmonary, e.g., by inhalation or insufflation of powders or aerosols, including by nebulizer; intratracheal, intranasal, epidermal and transdermal), oral or parenteral. Parenteral administration includes intravenous, intraarterial, subcutaneous, intraperitoneal or intramuscular injection or infusion; or intracranial, e.g., intrathecal or intraventricular administration. Pharmaceutical compositions and formulations for topical administration may include transdermal patches, ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable. Coated condoms, gloves and the like may also be useful.

[0139] The pharmaceutical formulations of the present invention, which may conveniently be presented in unit dosage form, may be prepared according to conventional techniques well known in the pharmaceutical industry. Such techniques include the step of bringing into association the active ingredients with the pharmaceutical carrier(s) or excipient(s). In general, the formulations are prepared by uniformly and intimately bringing into association the active ingredients with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

[0140] The compositions of the present invention may be formulated into any of many possible dosage forms such as, but not limited to, tablets, capsules, gel capsules, liquid syrups, soft gels, suppositories, and enemas. The compositions of the present invention may also be formulated as suspensions in aqueous, non-aqueous or mixed media. Aqueous suspensions may further contain substances which increase the viscosity of the suspension including, for example, sodium carboxymethylcellulose, sorbitol and/or dextran. The suspension may also contain stabilizers.

[0141] Pharmaceutical compositions of the present invention include, but are not limited to, solutions, emulsions, foams and liposome-containing formulations. The pharmaceutical compositions and formulations of the present invention may comprise one or more

penetration enhancers, carriers, excipients or other active or inactive ingredients.

[0142] Emulsions are typically heterogenous systems of one liquid dispersed in another in the form of droplets usually exceeding 0.1 μm in diameter. Emulsions may contain additional components in addition to the dispersed phases, and the active drug which may be present as a solution in either the aqueous phase, oily phase or itself as a separate phase. Microemulsions are included as an embodiment of the present invention.

[0143] The pharmaceutical formulations and compositions of the present invention may also include surfactants. The use of surfactants in drug products, formulations and in emulsions is well known in the art.

[0144] In one embodiment, the present invention employs various penetration enhancers to effect the efficient delivery of cyclic peptides such as to treat onychomycosis of the nails. In addition to aiding the diffusion of non-lipophilic peptides across cell membranes, penetration enhancers also enhance the permeability of keratin. Penetration enhancers may be classified as belonging to one of five broad categories, i.e., surfactants, fatty acids, bile salts, chelating agents, and non-chelating non-surfactants. Penetration enhancers and their uses are further described in U.S. Patent 6,287,860, which is incorporated herein in its entirety.

[0145] One of skill in the art will recognize that formulations are routinely designed according to their intended use, i.e. route of administration.

[0146] Compositions and formulations for oral administration include powders or granules, microparticulates, nanoparticulates, suspensions or solutions in water or non-aqueous media, capsules, gel capsules, sachets, tablets or minitablets. Thickeners, flavoring agents, diluents, emulsifiers, dispersing aids or binders may be desirable.

[0147] The formulation of therapeutic compositions and their subsequent administration (dosing) is believed to be within the skill of those in the art. Dosing is dependent on

severity and responsiveness of the disease state to be treated, with the course of treatment lasting from several days to several months, or until a cure is effected or a diminution of the disease state is achieved. Optimal dosing schedules can be calculated from measurements of drug accumulation in the body of the patient. Persons of ordinary skill can easily determine optimum dosages, dosing methodologies and repetition rates.

[0148] Optimum dosages may vary depending on the relative potency of individual cyclic or linear peptides, and can generally be estimated based on EC_{50}'s found to be effective in *in vitro* and *in vivo* animal models. In general, dosage is from 0.01 µg to 100 g per kg of body weight, and may be given once or more daily, weekly, monthly or yearly, or even once every 2 to 20 years. Persons of ordinary skill in the art can easily estimate repetition rates for dosing based on measured residence times and concentrations of the peptide in bodily fluids or tissues. Following successful treatment, it may be desirable to have the patient undergo maintenance therapy to prevent the recurrence of the disease state, wherein the peptide is administered in maintenance doses, ranging from 0.01 µg to 100 g per kg of body weight, once or more daily, to once every 20 years.

[0149] Compositions and formulations for parenteral, intrathecal or intraventricular administration may include sterile aqueous solutions which may also contain buffers, diluents and other suitable additives such as, but not limited to, penetration enhancers, carrier compounds and other pharmaceutically acceptable carriers or excipients.

[0150] The cyclic peptide or linear peptide conjugate may also be formulated into an agronomically acceptable composition for topical application to plants or seeds. Agronomically acceptable carriers are used to formulate a peptide herein disclosed for the practice of the instant method. Determination of dosages suitable for systemic and surface administration is enabled herein and is within the ordinary level of skill in the art. With proper choice of carrier and suitable manufacturing practice, the compositions such as those formulated as solutions, may be administered to plant surfaces including above-ground parts and/or roots, or as a coating applied to the surfaces of seeds.

[0151] Agronomically useful compositions suitable for use in the system disclosed herein include compositions wherein the active ingredient(s) are contained in an effective amount to achieve the intended purpose. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the disclosure provided herein.

[0152] In addition to the active ingredients, these compositions for use against plant pathogens may contain suitable agronomically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used in the field, in greenhouses or in the laboratory setting.

[0153] Anti-pathogen formulations include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the cyclic peptides may be prepared as appropriate oily suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Further components can include viscosifiers, gels, wetting agents, ultraviolet protectants, among others.

[0154] Preparations for surface application can be obtained by combining the active peptides with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain powders for direct application or for dissolution prior to spraying on the plants to be protected. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose or starch preparations, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

EXAMPLES

[0155] Aspects disclosed herein are further described by the following non-limiting Examples.

Materials and Methods

Peptide substrates and inhibitors

[0156] Two internally-quenched fluorescent (IQF) peptides (Abz-STRNGLPS-Y($3NO_2$) (SEQ ID NO:21) and Abz-STRNG**A**PS-Y($3NO_2$) (SEQ ID NO:25) where Abz is o-aminobenzoic acid and Y[$3NO_2$] is 3-nitrotyrosine) were synthesized by Genscript or GL Biochem at >90% purity and solubilized in 25% (v/v) acetonitrile:water. The fluorogenic peptide substrate Z-AAN-MCA (where Z is carboxybenzyl and MCA is 7-amido-4-methylcoumarin) was obtained from the Peptide Institute and solubilized in DMSO. The inhibitors Ac-YVAD-CHO and Ac-STRN-CHO (where Ac is acetyl and CHO is aldehyde) were synthesized by the Peptide Institute and Mimotopes respectively. The linear cyclotide precursor of kalata B1 was chemically synthesized and folded as described previously (Simonsen *et al*. (2004) *FEBS Lett* 577(3):399–402). This precursor was solubilized in ultrapure water and synthesized with a terminal free acid or amine. Figure 1 provides a representation of a linear cyclotide polypeptide precursor. Bac2A, EcAMP1 and R1 and its derivatives were synthesized with added AEP recognition residues by Genscript or GL Biochem at >85% purity with a terminal free acid or amine and solubilized in ultrapure water, except for one R1 derivative (**LL**VFAEFLPLFSKFGSRMHILK**NGL**; SEQ ID NO:89) which was solubilized in 25% DMSO. The ligation partner peptides (GLK-biotin; biotin-TRNGL; GLPVSGE, SEQ ID NO: 14; PLPVSGE, SEQ ID NO: 80) were synthesized by GL Biochem at >85% purity with a terminal free acid or amine and solublilized in ultrapure water.

Cyclization assay

[0157] Linear target peptides (280 µM, unless otherwise indicated) were incubated with r*Oa*AEP1_b, r*Oa*AEP3, r*Oa*AEP4 or r*Oa*AEP5 (total protein concentration as indicated in the description of figures) in activity buffer (50 mM sodium acetate, 50 mM NaCl, 1 mM

ethylenediaminetetraacetic acid [EDTA], 0.5 μM Tris(2-carboxyethyl)phosphine hydrochloride [TCEP], pH 5). The reaction was allowed to proceed for up to 22 hours at room temperature and was analysed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS), high performance liquid chromatography (HPLC) or nuclear magnetic resonance (NMR) as appropriate.

Intermolecular ligation assays

[0158] Target peptides (140 μM) were incubated with a ligation partner peptide (700 μM) and a recombinant AEP (r*Oa*AEP1$_b$, r*Oa*AEP3, r*Oa*AEP4 or r*Oa*AEP5 at 19.7 μg mL^{-1} total protein) in activity buffer (50 mM sodium acetate, 50 mM NaCl, 1 mM EDTA, 0.5 μM TCEP, pH 5). The reaction was allowed to proceed for up to 22 hours at room temperature and was analysed by MALDI MS or electrospray ionisation (ESI) mass spectrometry as appropriate.

MS to track AEP-mediated processing of linear peptides

[0159] Cyclization or inter-molecular ligation of linear target peptides was monitored by MALDI or ESI MS. In both cases, the reaction mixture (5 – 50 μL) was de-salted using C18 zip tips (Millipore) and eluted in 4 μL 50-75% v/v acetonitrile, 0.1% v/v trifluoroacetic acid (TFA).

[0160] For MALDI MS, a saturated MALDI matrix solution (α–cyano-4-hyroxycinnamic acid, CHCA, Bruker) prepared in 95% v/v acetonitrile, 0.1% v/v TFA was diluted 1:22 such that the final matrix solution comprised 90% v/v acetonitrile, 0.1% v/v TFA and 1 mM $NH_4H_2PO_4$. Eluted samples were mixed 1:4 with the MALDI matrix, spotted onto a MALDI plate and analyzed by an Ultraflex III TOF/TOF (Bruker) in positive reflector mode. Data were analyzed using the flexAnalysis program (Bruker).

[0161] For ESI MS, 96 μl of 75% v/v acetonitrile, 01% v/v TFA was added to the de-salted sample. The sample was then injected into a MicroTOF Q (Bruker) and data was collected in positive ionisation mode. The mass of ligated or cyclized product was determined by charge deconvolution using the Compass DataAnalysis program (Bruker).

Assaying protease activity against IQF and fluorescent peptides

[0162] To assay activity of recombinant AEPs (r*Oa*AEP1$_b$, r*Oa*AEP3, r*Oa*AEP4 or r*Oa*AEP5) against both internally-quenched and other fluorescent peptides, substrate and enzyme were diluted as appropriate in activity buffer (50 mM sodium acetate, 50 mM NaCl, 1 mM EDTA, 0.5 µM TCEP, pH 5). To assay activity of recombinant human legumain (rhuLEG; R&D systems) against the same substrates, the enzyme was first activated by incubation in 50 mM sodium acetate, 100 mM NaCl, pH 4 (4 µL buffer/1 µL enzyme) for 2 hours at 37ºC. Substrates and activated rhuLEG were diluted in 50 mM acetate, 0.1% v/v triton X 100, 0.5 µM TCEP pH 5.5 or in 50 mM MES, 250 mM NaCl, pH 5 as required. Diluted enzyme and substrate were added to black, flat bottomed microtiter plates in a total assay volume of 100 - 200 µL. The change in fluorescence intensity over time was monitored on a SpectraMax M2 (Molecular Devices) using excitation/emission wavelengths of 320/420 nm (IQF peptides) or 360/460 nm (other fluorescent peptides).

Inhibition assays

[0163] To investigate the impact of inhibitors on enzyme activity against the wild type IQF peptide, Abz-STRNGLPS-Y(3NO_2), r*Oa*AEP1$_b$ (4.4 µg mL^{-1} total protein) was incubated with Ac-YVAD-CHO (500 µM) or Ac-STRN-CHO (409 µM) for 40 minutes prior to addition to the substrate (11 µM). Enzyme activity against the wt IQF peptide was then assessed as described above.

Antibodies

[0164] Polyclonal anti-AEP1b rabbit serum was generated by immunizing a New Zealand White rabbit with a denatured, inactive form of *O. affinis* AEP1b (residues D47 – P474) that was produced recombinantly in *E. coli*. The rabbit received three doses, four weeks apart of 150 µg of antigen in 50% (v/v) phosphate-buffered saline (PBS) and Freund's incomplete adjuvant (Sigma). Serum was obtained two weeks after the final dose.

O. affinis transcriptome

[0165] Total RNA was extracted from *O. affinis* root, leaf and seedling tissues using a phenol extraction method. Plant material was frozen in liquid nitrogen and ground to a fine powder, which was then resuspended in buffer (0.1 M Tris-HCl pH 8.0, 5 mM EDTA, 0.1 M NaCl, 0.5% SDS, 1% 2-mercaptoethanol), extracted twice with 1:1 phenol:chloroform and precipitated by addition of isopropanol. The pellets were dissolved in 0.5 ml water and RNA was precipitated overnight at 4 °C by addition of 4 M lithium chloride. The extracted RNA of each tissue was analysed by GeneWorks using the Illumina GAIIx platform. In total, 69.3 million 75 bp paired-end reads were generated. Reads were filtered with a phred confidence value of Q37 and assembled into contigs using Oases (Schulz *et al.* (2012) *Bioinformatics 28:* 1086-1096) with k-mer ranging from 41–67. The assemblies were merged using cd-hit-est (Li *et al.* (2006) *Bioinformatics 22:* 1658-1659), resulting in 270,000 contigs. Statistics on the depth of sequencing were made by aligning the reads of each tissue on the contigs using BWA (Li *et al.* (2009) *Bioinformatics* 25: 1754-1760). All the sequences, including one AEP, previously identified from an EST library of *O. affinis* were present among the contigs (Qin *et al.* (2010) *BMC Genomics 11*: 111). Homologues of this AEP sequence were searched using BLAST (Altschul *et al.* (1990) *J Mol Biol 215:* 403-410) in the contig library using a maximum E-value of 1e-20, resulting in the identification of 371 putative AEP transcripts. These sequences could then be clustered in 13 groups sharing at least 90% sequence identity using cd-hit (Li *et al.* (2006) *supra*). Coding sequences identified were *Oa*AEP4-17 (SEQ ID NOs: 39 to 54).

Oa*AEP cloning*

[0166] Full length AEP transcripts from the *O. affinis* transcriptome assembly were used to design a set of primers. A single degenerate forward primer (*Oa*AEPdegen-F, 5`-ATG GTT CGA TAT CYC GCC GG-3` - SEQ ID NO:6) was manually designed to amplify all sequences due to the variability at a single nucleotide position within the 5` region of each full length transcript at the start codon. Three reverse primers, designed with the aid of Primer3, successfully amplified AEP sequences (*Oa*AEP1-R, 5`- TCA TGA ACT AAA TCC TCC ATG GAA AGA GC -3` - SEQ ID NO:7; *Oa*AEP2-R, 5`- TTA TGC ACT

GAA TCC TTT ATG GAG GG -3` - SEQ ID NO:8; *Oa*AEP3-R 5`- TTA TGC ACT GAA TCC TCC ATC G -3` - SEQ ID NO:9).

[0167] To clone expressed *Oa*AEPs, total RNA was extracted from *O. affinis* leaves and shoots using TRIzol (Life Technologies) and was reverse transcribed with SuperScript III reverse transcriptase (Life Technologies) according to the manufacturer's instructions. Target sequences were amplified from the resulting cDNA using Phusion High Fidelity Polymerase (New England BioLabs) and the primers described above under the recommended reaction conditions. Gel extracted PCR products were dA-tailed by incubation with Invitrogen Taq Polymerase (Life Technologies) and 0.5 µL 10 mM dA in the supplied buffer. The processed products were cloned into pCR8-TOPO (Life Technologies) and transformed into *E. coli*. Purified DNA from clones that were PCR positive for an AEP insert were sent for Sanger sequencing at the Australian Genome Research Facility. Coding sequences have been deposited in Genbank (accession codes: *Oa*AEP1 (KR259377), *Oa*AEP2 (KR259378), *Oa*AEP3 (KR259379).

[0168] In a different approach, genomic DNA was extracted from *O. affinis* leaf tissue using a DNeasy Plant Mini Kit according to the manufacturer's instructions. PCR amplification from this DNA used primers specifically targeting the *Oa*AEP1 nucleotide sequence. Gel extracted product was dA tailed as above, cloned into TOPO (Life Technologies) and transformed into *E. coli*. DNA from PCR-positive clones was sent for sequencing to the Australian Genome Research Facility. The AEP sequence identified using this method (*Oa*AEP1$_b$) was subsequently expressed as a recombinant protein.

Prediction of cyclase activity

[0169] AEPs are identified from cyclotide producing plants which, when expressed with the precursor gene for the cyclotide kalata B1 (oak1), and other peptides, effect backbone cyclization. By comparing the amino acid sequences of ligation competent AEPs with those favouring proteolysis, a differential loop region, termed the activity preference loop (APL), is identified that contributes to the specificity. In ligase competent AEPs, the APL either has several residues missing or is replaced by hydrophobic stretch of amino acids

(Figure 5A).

[0170] Additional residues linked to cyclase function are identified by machine learning (protein sequence space analysis) using a set of experimentally determined cyclase and non-cyclase sequences. The following residues are found to be highly predictive of cyclase function in the currently known cyclases and non-cyclases (Figure 5B). All numbering is given relative to *OaAEP1b* (Figure 4; SEQ ID NO:1).

1. **APL** - The absence of residues in the region between **299 - 300** of OaAEP1 is predictive of a higher likelihood of cyclase activity.

2. **Set 1** - The presence of the following active site residues is also predictive of a higher likelihood of cyclase activity:

- **D161**
- **C247**
- **Y248**
- **Q253**
- **A255**
- **V263**

3. **Set 2** - The presence of the following active site-proximal residues is also predictive of a higher likelihood of cyclase activity:

- **K186**
- **D192**

4. **Set 3** - The presence of the following non-active site surface residues is also predictive of a higher likelihood of cyclase activity:

- **K139**
- **H293**
- **E314**
- **G316**

[0171] Overall it is highly predictive of cyclase activity if the sequence contains either:

- The shortened APL
- 3 of the 6 Set 1 active site residues
- Both of the Set 2 active-site-proximal residues
- 3 of the 4 Set 3 non-active-site residues

[0172] The most predictive are the APL and set 1. The more of these criteria that it hits, the more likely that it is to be a cyclase. Predictive residues for cyclase activity are shown in Table 2. Residue numbering is relative to OaAEP1b (Figure 4; SEQ ID NO:1). Residue properties that strongly predict cyclase activity are disorder propensity (DISORD), net static charge (CHRG), molecular weight of R group (RMW), and hydropathy index (HPATH).

[0173] An AEP having at least 25% (or 5 or more) of the 17 predictive residues set forth in Table 2 is considered likely to act preferentially as a cyclase. A requirement for at least 25% of the predictive residues to be present enables 100% of the known cyclases to be correctly identified while excluding known non-cyclases at least 80% including 94%.

[0174] Examples of AEPs predicted to be cyclases using this method include OaAEP4 (88%), OaAEP5 (70%), both sequences derived from transcriptome analysis, which have been tested and shown to be cyclases (e.g. Example 4). Other sequences predicted to be cyclases include AEPs from *Cicer arietinum* (SEQ ID NO:92), *Medicago truncatula* (SEQ ID NO:93), *Hordeum vulgare* (SEQ ID NO:94), *Gossipium raimondii* (SEQ ID NO:95 and *Chenopodium quinoa* (SEQ ID NO:96) (Example 10).

Table 2

residue	property	cyclase	non-cyclase
139	CHRG	K	D
161	CHRG	D	N
186	CHRG	K	G
192	CHRG	D	N
247	RMW	C	G
248	RMW	Y	T
253	CHRG	Q	E
255	DISORD	A	P
263	HPATH	V	T
293	HPATH	H	L
299 - 300	GAP	-	N
	GAP	-	G
	GAP	-	N
	GAP	-	Y
	GAP	-	S
314	CHRG	E	K
316	RMW	G	K

EXAMPLE 1

Expression and activation of recombinant O. affinis AEPs (rOaAEPs) in E. coli

[0175] DNA encoding full-length *O. affinis* AEPs without the putative signalling domain (*Oa*AEP1$_b$ residues A_{27}-P_{478}, *Oa*AEP3 residues R_{28}-A_{491}, *Oa*AEP4 residues A_{28}-A_{491} or *Oa*AEP5 residues E_{27}-L_{485}) was inserted into the pHUE vector (Catanzariti *et al*. (2004) *Protein Sci 13:* 1331-1339) to give a 6xHis-ubiquitin-*Oa*AEP fusion protein construct (SEQ ID NO:20 describes the r*Oa*AEP1$_b$ construct and the region containing *Oa*AEP1$_b$ is replaced with *Oa*AEP3, *Oa*AEP4 or *Oa*AEP5 in the other constructs). Residue numbering is as determined by a multiple alignment generated using Clustal Omega (Sievers *et al*. (2011) *supra*) (Figure 2). DNA was then introduced into T7 Shuffle *E. coli* cells (New England BioLabs). Transformed cells were grown at 30°C in superbroth (3.5% tryptone [w/v], 2% yeast extract [w/v], 1% glucose [w/v], 90 mM NaCl, 5 mM NaOH) to mid-log phase; the temperature was then reduced to 16°C and expression was induced with isopropyl ß–D-1-thiogalactopyranoside (IPTG; 0.4 mM; Bio Vectra) for approximately 20 hours. Cells were harvested by centrifugation and resuspended in non-denaturing lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% triton X 100, 1mM EDTA, pH 7). Lysis was promoted by a total of five freeze/thaw cycles and the addition of lysozyme (hen egg white; Roche; 0.4 mg mL^{-1}). DNase (bovine pancreas; Roche; 40 µg mL^{-1}) and $MgCl_2$ (0.4 M) were also added. Cellular debris was removed by centrifugation and the lysate was stored at -80°C until required.

[0176] Lysate containing expressed recombinant AEPs was filtered through a 0.1 µM glass fibre filter (GE Healthcare) before being diluted 1:8 in buffer A (20 mM bis-Tris, 0.2 M NaCl, pH 7) and loaded onto two 5 mL HiTrap Q Sepharose high performance columns connected in series (GE Healthcare; 1.6 – 3.1 mL undiluted lysate mL^{-1} resin). Bound proteins were eluted with a continuous salt gradient (0-30% buffer B [20 mM bis-Tris, 2 M NaCl, pH 7]; 15 column volumes [cv]) and AEP-positive fractions identified by Western blotting (anti-AEP1$_b$ rabbit serum [1:2000]; peroxidase-conjugated anti-rabbit IgG [GE Healthcare; 1:5000]).

[0177] AEPs are usually produced as zymogens that are self-processed at low pH to their mature, active form (Hiraiwa *et al*. (1997) *Plant J 12(4)*:819-829; Hiraiwa *et al*. (1999) *FEBS Lett 447(2-3)*:213-216; Kuroyanagi *et al*. (2002) *Plant Cell Physiol 43(2)*:143-151). To self-activate all AEPs, EDTA (1 mM) and TCEP (Sigma-Aldrich; 0.5 mM) were added, the pH was adjusted to 4.5 with glacial acetic acid and the protein pool was incubated for 5 hours at 37°C. Figure 3A demonstrates that activity of r*Oa*AEP1$_b$ against an IQF peptide (Abz-STRNGLPS-Y(3NO_2)) representing the native C-terminal processing site in kB1 was dramatically increased following this activation step. Protein precipitation at pH 4.5 allowed removal of the bulk of the contaminating proteins by centrifugation. The remaining protein was filtered (0.22 µm; Millipore), diluted 1:8 in buffer A2 (50 mM acetate, pH 4) then captured on a 1 mL HiTrap SP Sepharose high performance column (GE Healthcare). Bound proteins were eluted with a salt gradient (0-100% buffer B2 [50 mM acetate, 1 M NaCl, pH 4]; 10 cv) and fractions with activity against an IQF peptide (Abz-STRNGLPS-Y(3NO_2)) were pooled and used in subsequent activity assays. Figure 3B shows that after activation of r*Oa*AEP1$_b$, a dominant band of ~32 kDa was evident by reducing SDS-PAGE and Instant blue staining (Expedeon) and this was confirmed to be r*Oa*AEP1$_b$ by Western blotting. Experimentally determined self-processing sites of r*Oa*AEP1$_b$ are indicated in Figure 4. The total concentration of protein in each preparation was estimated by BCA assay according to the manufacturer's instructions.

EXAMPLE 2

In vitro cyclization of the cyclotide kalata B1 (kB1)

[0178] The ability of activated, mature r*Oa*AEP1$_b$ to cyclize a synthetic kB1 precursor carrying the native C-terminal pro-hepta-peptide (GLPSLAA) (Figure 1B) was tested using the cyclization assay described in the Materials and Methods followed by MALDI MS. When incubated with the kB1 precursor the active enzyme produced a peptide of 2891.2 Da (monoisotopic, $[M+H]^+$), consistent with the expected mass of mature, cyclic kB1 (Figure 6). This product was confirmed to be identical to native kB1 by HPLC co-elution and 1D and 2D-NMR experiments.

[0179] To determine the kinetics of r*Oa*AEP1$_b$ activity against the wt kB1 precursor (Figure 1; SEQ ID NO: 11), the substrate was assayed at room temperature at a range of concentrations between 75 and 250 μM in a total volume of 20–160 μl of activity buffer. The total protein concentration of the enzyme preparation added to the kinetic assays was 19.7 μg ml^{-1}. The reaction was quenched after 5 min with 0.1% TFA and the volume adjusted to 800 μl. A volume of 700 μl was loaded onto a reversed-phase C18 analytical column (Agilent Eclipse C18, 5 μm, 4.6 × 150 mm) and peptides were separated by HPLC (19 min linear gradient of 12–60% acetonitrile, 0.1% TFA at 1 ml min^{-1}). The identity of eluted peaks was confirmed using MALDI MS. The area under the curve corresponding to the precursor peptide was quantitated by comparison to a standard curve and initial velocities were calculated by converting this to μmoles product formed. Kinetic parameters were estimated using the Michaelis–Menten equation and the curve-fitting program GraphPad Prism (GraphPad Software, San Diego). It was not possible to precisely determine the concentration of active enzyme due to impurities remaining in the preparation and the absence of an inhibitor appropriate for active site titration. However, a conservative turnover rate (k_{cat}) was estimated based on a mass of 32 kDa and the assumption that the total protein concentration reflected active enzyme. Kinetic parameters (±s.e.m.) for the processing of the wt kB1 precursor and r*Oa*AEP1$_b$ were 0.53 (±0.1) s^{-1} for k_{cat}, 212 (±76) μM for K_m and 2,500 M^{-1} s^{-1} for k_{cat}/K_m as determined from a Michaelis–

Menten plot (Figure 7). Differences in purity and proportion of active enzyme in different preparations means these parameters are subject to batch to batch variation.

EXAMPLE 3

In vitro cyclization of non-cyclotide peptides R1, Bac2A and EcAMP1

[0180] The ability of activated AEPs (r*Oa*AEP1$_b$, r*Oa*AEP3, r*Oa*AEP4 and r*Oa*AEP5) to cyclize peptide substrates structurally unrelated to cyclotides was tested in the cyclization assay described in the Materials and Methods. The anti-malarial peptide R1 (Harris *et al*. (2009) *J Biol Chem 284(14)*:9361-9371; Harris *et al*. (2005) *Infect Immun 73(10)*:6981-6989); Bac2A, a linear derivative of the bovine peptide bactenecin (Wu and Hancock (1999) *Antimicrob Agents Ch 43*:1274-1276); and the anti-fungal peptide EcAMP1 (Nolde *et al.* (2011) *J Biol Chem 286(28):*25145-25153) were produced with additional AEP recognition residues and used as the substrates. The appearance of a mass corresponding to cyclic product indicated that in each case the linear precursor peptides were efficiently cyclized following the addition of N- and C-terminal AEP recognition sequences (Figures 8 and 9).

EXAMPLE 4

Sequence requirements for in vitro cyclization

[0181] To investigate the sequence requirements for *in vitro* cyclization, R1 was used as a model peptide. The recognition residues added to this model peptide were sequentially trimmed to determine the minimal requirements for AEP-mediated cyclization. The N- and C-terminal recognition residues were also substituted with alternate amino acids to determine if particular classes of residues were preferred for cyclization by these recombinant AEPs. The ability of activated AEPs (r*Oa*AEP1$_b$, r*Oa*AEP3, r*Oa*AEP4 and r*Oa*AEP5) to cyclize the R1 peptide with varied flanking sequences was then tested in the cyclization assay described in the Materials and Methods (Figure 10; Table 3).

[0182] Sequential trimming of the added recognition residues revealed that all four enzymes tested could cyclize the R1 peptide following the addition of only a C-terminal Asn-Gly-Leu motif (although some linear product was also produced from this precursor). After cleavage C-terminal to the Asn residue, only one residue (Asn) is left behind in the the cyclized peptide. However, more efficient cyclization was generally achieved with an N-terminal Gly-Leu motif as well as the C-terminal Asn-Gly-Leu motif. Subsequent mutations of the flanking residues were made within this format.

[0183] At the N-terminus, Leu, Gln and Lys were all accepted in place of the P1" Gly, although in some cases this decreased the yield of cyclic product. Val was also accepted when presented as the first residue of the model peptide R1. At the P2" position, the positively charged Lys was poorly tolerated in place of Leu but cyclic product could still be produced. At the same position, the aromatic Phe was generally well accepted although again in some cases this decreased the yield of cyclic product at the timepoint tested. Any added amino acids at the N-terminus together with any added C-terminal amino acids up to and including the Asn are retained in the cyclic product. Therefore, for some target peptides where additional N-terminal residues impact function it may be acceptable to reduce the overall yield to minimize the introduction of non-native residues.

[0184] At the C-terminus, most substitutions resulted in a reduced yield of cyclic product under the conditions tested. At the P1' position His and Phe could be accepted but the yield was generally reduced under the conditions tested. At the P2' position Val, His and Phe could be accepted but this reduced yield in some cases. Other residues that could be accepted at this position are Ile, Ala, Met, Trp, Tyr.. Residues C-terminal to the P1 residue are not incorporated into the final product. Therefore, there is little advantage to including sub-optimal residues within this region. All four enzymes tested were able to cyclize substrates presenting either an Asn or Asp in the P1' position and preference was enzyme dependent. Since this residue is incorporated into the final peptide, choice of Asn or Asp at this position will likely be substrate dependent and this may influence which enzyme is selected for use.

[0185] No processing of either the native R1 peptide or a modified R1 carrying the N-terminal Gly-Leu motif but only an Asn at the C-terminus was observed by r*Oa*AEP1$_b$. The cyclic nature of the R1 derivatives presented in Table 3 processed by r*Oa*AEP1$_b$ was confirmed by digestion with endoGlu-C (New England Biolabs; as per manufacturer's instructions). This secondary digestion produced a single linear product (as opposed to two linear peptides) consistent with linearization of a backbone cyclized peptide.

Table 3

The relative percentage of cyclic and linear R1 peptide derivatives following rOaAEP 1$_b$-mediated processing.[a,b,c]

Peptide Sequence	Cyclic product (%)	Linear precursor (%)	Linear product (%)
GLPVFAEFLPLFSKFGSRMHILK**STRNGLP** (SEQ ID NO:26)	78.8 (±6.9)	21.2 (±6.9)	-
GLVFAEFLPLFSKFGSRMHILK**NGL** (SEQ ID NO:27)	92.9 (±2.4)	7.1 (±2.4)	-
GLVFAEFLPLFSKFGSRMHILK**NG** (SEQ ID NO:28)	-	100	-
VFAEFLPLFSKFGSRMHILK**NGL** (SEQ ID NO:29)	49.6 (±14.1)	27.7 (±7.7)	27.7 (±7.3)
QLVFAEFLPLFSKFGSRMHILK**NGL** (SEQ ID NO:30)	93.1 (±2.2)	6.9 (±2.2)	-
KLVFAEFLPLFSKFGSRMHILK**NGL** (SEQ ID NIO:31)	82.2 (±3.8)	17.8 (3.8)	-

[a]The peak area of a given processing variant of R1 is displayed as a percentage of the total peak area attributable to that peptide.

[b]The average of three experiments is reported ± standard error of the mean.

[c]The enzyme concentration used was 12 μg mL^{-1} total protein with an incubation time of 22 hours.

EXAMPLE 5

Polypeptide ligation

[0186] To investigate the ability of recombinantly produced AEPs to perform inter-molecular ligation (as well as the intra-molecular ligation required to produce backbone-cyclized peptides), target peptides were incubated with ligation partner peptides as well as active enzyme (Figure 11). The appearance of new linear peptides were tracked. Ligation of labeled peptides to a target polypeptide provides a generic, targeted protein labeling strategy for a variety of moieties (e.g. fluorescent labels, biotin, affinity tags, epitope tags, solubility tags) that is limited only by the ability of synthetic peptide chemistry or other methods to produce the appropriate ligation partner. This approach can also enable ligation of multiple domains that could be challenging to produce as a single recombinant protein.

[0187] AEP recognition residues were added to the N- and C-termini of the plant defensin NaD1 (Lay *et al*. (2003) *Plant Physiol 131*:1283-1293) to produce a modified defensin (**GLP**-NaD1-**TRNGLP;** SEQ ID NO:79). This was recombinantly expressed in *Pichia pastoris* and purified using a similar method to that described in Lay *et al*. (2012) *J Biol Chem 287*:19961-19972. When AEP-mediated processing of the modified NaD1 (140 µM) was tested using the cyclization assay described in the Materials and Methods section, only a linear product was evident by ESI-MS and there was no evidence of backbone-cyclization (Figure 12). Presumably the disulphide bonded structure of NaD1 is sub-optimal for cyclization. However, when the modified NaD1 was incubated with a ligation partner peptide (GLPVSGE, SEQ ID NO:14 or PLPVSGE, SEQ ID NO:80) and active, recombinant AEPs (r*Oa*AEP1$_b$, r*Oa*AEP3, r*Oa*AEP4 and r*Oa*AEP5) using the ligation assay described in the Materials and Methods, new, linear, peptides were detected using ESI-MS (Figure 12).

[0188] The inter-molecular ligase activity of recombinant AEPs was further explored using using other peptide combinations. A modified R1 peptide (**GK**VFAEFLPLFSKFGSRMHILK**NGL;** SEQ ID NO:83) that was a poor substrate for backbone cyclization was used as a target peptide for C-terminal labelling. The R1

derivative was incubated with a biotinylated ligation partner (GLK-biotin; SEQ ID NO:102) and recombinant AEPs. MALDI-MS showed that AEP-mediated processing created a new linear peptide that incorporated a C-terminal biotin (Figure 13).

[0189] A modified R1 peptide, (**GL**VFAEFLPLFSKFGSRMHILK**GHV**; SEQ ID NO:61) that was not itself an AEP substrate since it does not contain the required Asx residue was used as a target peptide for N-terminal labelling. The R1 derivative was incubated with a biotinylated ligation partner peptide (biotin-TRNGL; SEQ ID NO:104) and recombinant AEPs. MALDI-MS showed that AEP-mediated processing created a new linear peptide that incorporated an N-terminal biotin (Figure 14).

EXAMPLE 6

Identification of cyclizing AEPs by substrate specificity

[0190] AEP activity has traditionally been tracked by monitoring cleavage of the fluorescent substrate Z-AAN-MCA (where Z is carboxybenzyl; MCA is 7-amido-4-methylcoumarin) [Saska *et al*. (2007) *supra*; Rotari *et al*. (2001) *Biol. Chem. 382:*953-959]. Cleavage C-terminal to the Asn liberates the fluorophore which then fluoresces to report substrate cleavage. However, neither butelase-1 (Nguyen *et al*. (2014) *supra*) nor r*Oa*AEP1$_b$, r*Oa*AEP3, r*Oa*AEP4 or r*Oa*AEP5 (Figure 15) had detectable activity against this substrate. Furthermore, two AEP active site inhibitors had limited efficacy against r*Oa*AEP1$_b$ at high concentrations (Figure 16). They are Ac-YVAD-CHO, which is routinely used to identify AEP activity (Hatsugai *et al*. (2004) *Science 305(5685):* 855-858) and Ac-STRN-CHO, which represents the P1-P4 residues of the C-terminal kB1 cleavage site. These traditional routes of identifying AEP activity will therefore likely be ineffective for identification of AEPs with cyclizing ability.

[0191] IQF peptides that incorporate the P1-P4 as well as the P1'-P4' residues are, however, effectively targeted by recombinant *O. affinis* AEPs. These peptides contain a fluorescence donor/quencher pair, with fluorescence observed upon the spatial separation of this pair following enzymatic cleavage. Activity against such IQF reporter peptides without corresponding activity against the generic substrate (Z-AAN-MCA) may allow rapid identification of members of the AEP family likely to have cyclizing ability. In the IQF peptide format, r*Oa*AEP1$_b$ displayed a preference for a bulky hydrophobic residue such as Leu at the P2' position that was not shared by human legumain (rhuLEG), an AEP that preferentially functions as a hydrolase (Figures 17A and 17B). Such P2' specificity could also be used to predict cyclization ability and or to select AEPs with different sequence requirements in the substrate to be cyclized

EXAMPLE 7

In vitro cyclization of bacterially-expressed polypeptides

[0192] DNA encoding a target peptide for cyclisation, a short linker (Glu-Phe-Glu-Leu or Gly-Gly-Gly-Gly-Ser-Glu-Phe-Glu-Leu) and a C-terminal ubiquitin-6xHis was inserted into either the pHUE vector (Catanzariti *et al.* (2004) *supra*) (*XbaI/BamHI*) or the pET23a(+) vector (Invitrogen; *NdeI/XhoI*) to give a target peptide-linker-ubiquitin-6xHis fusion protein construct (Figure 18 A). The linker coding region contains restriction sites for easy substitution of the target peptide domain with other target sequences. In the case of pHUE, the DNA sequence inserted included nucleotides prior to the initiating Met codon to ensure the original vector sequence was reconstituted. If not naturally present in the target peptide, appropriate N-and C-terminal AEP recognition sequences were introduced. Since the first residue of all target peptides was necessarily Met, the N-terminal recognition sequence added was Met-Leu. The C-terminal recognition sequence was Asn-Gly-Leu-Pro. Optionally, the target peptide is preceeded by an initiating Met followed by the kalata B1 N-terminal repeat (NTR) (Figure 18 B) or other cleavable domain.

[0193] The target peptides produced as fusion proteins with ubiquitin were the cyclotide kB1 (SEQ ID NO:74), the modified sunflower trypsin inhibitor SFTI-1 I10R (Quimbar *et al.* (2013) *J Biol Chem 288(19):*13885-13896) (SEQ ID NO:72) and the conotoxin Vc1.1 (Clark *et al.* (2010) *supra*) (SEQ ID NO:76). The constructs were introduced into T7 Shuffle *E. coli* cells (New England BioLabs) and grown at 30°C in 2YT (16% [w/v] tryptone, 10% [w/v] yeast, 5% [w/v] sodium chloride) to mid-log phase. The temperature was then reduced to 16°C and expression was induced with IPTG (0.4 mM; Bio Vectra) for approximately 20 hours. Cells were harvested by centrifugation and resuspended in non-denaturing lysis buffer (50mM sodium phosphate pH 8.0, 300 mM NaCl, 1-10 mM imidazole). Lysis was promoted by up to five freeze/thaw cycles and the addition of lysozyme (hen egg white; Roche; 0.4 – 1 mg mL^{-1}). DNase (bovine pancreas; Roche; 5 µg mL^{-1}) and $MgCl_2$ (5 mM) were also added. Cellular debris was removed by centrifugation. The lysate was then filtered through a 0.1 µM glass fibre filter (GE Healthcare) and passed

over a Ni-NTA resin (QIAgen) to capture 6xHis tag protein. Bound protein was eluted with elution buffer (50mM sodium phosphate pH 8.0, 300 mM NaCl, 250 mM imidazole) and the total protein concentration was estimated by BCA assay according to the manufacturer's instructions. Fusion proteins were then used in enzyme assays. Optionally, the eluted protein is first buffer exchanged into water or appropriate buffer before AEP processing is assayed. Optionally, the eluted protein is first further purified by diluting 1:10 in 20 mM Tris-HCl, pH8 and passing over as second resin (Q sepharose high performance anion exchanger; GE Healthcare). Bound protein is recovered by a continuous salt gradient (0-100% 20 mM Tris-HCl, 1M NaCl pH 8), optionally buffer exchanged into ultrapure water or appropriate buffer and concentrated.

[0194] The ability of AEPs (r*Oa*AEP1$_b$, r*Oa*AEP3, r*Oa*AEP4, r*Oa*AEP5) to release the ubiquitin tag and cyclise target peptides was investigated using the cyclisation assay described in the Materials and Methods followed by MALDI MS (Figure 19 A, Table 4). Estimated substrate and enzyme concentrations are as indicated in the description of the figures. When required, 3.3% v/v glacial acetic acid was also added to the reaction mix to ensure the assay was carried out at acidic pH. When incubated with the fusion proteins, all recombinant AEPs tested released ubiquitin and produced cyclized kB1 (SEQ ID NO: 73), SFTI-1 I10R (SEQ ID NO: 71) and Vc1.1 (SEQ ID NO: 75) in a single step. To estimate the proportion of fusion protein being enzymatically processed, loss of the precursor protein over time was tracked using SDS-PAGE followed by Western blotting (anti-6xHis monoclonal mouse antibody [Genscript; 0.5 μg mL^{-1}]; peroxidase-conjugated anti-mouse IgG [Thermo Scientific; 1:10000] (Figure 19 B). This demonstrated that for r*Oa*AEP1$_b$, r*Oa*AEP3 and r*Oa*AEP5 the bulk of the precursor was being enzymatically processed. In this experiment a smaller proportion of the precursor protein was processed by r*Oa*AEP4.

[0195] Optionally, to separate released ubiquitin and unprocessed fusion protein from cyclized product, the mixture is then diluted 1:5 in non-denaturing lysis buffer and again passed over a Ni-NTA resin (QIAgen). The processed, cyclic, product no longer contains a 6xHis tag and is therefore present in the unbound fraction. This product is then dialysed

into ultrapure water, concentrated and analysed by MALDI MS, HPLC or NMR to confirm its cyclic structure.

Table 4

The expected and observed monoisotopic masses of cyclic products following AEP-mediated processing of target peptides fused to ubiquitin. [a, b, c]

	Monoisotopic cyclic mass (Da; $[M+H]^+$)				
Target peptide	**Expected**	**Observed (r*Oa*AEP1$_b$)**	**Observed (r*Oa*AEP3)**	**Observed (r*Oa*AEP4)**	**Observed (r*Oa*AEP5)**
SFTI1-I10R-ubiquitin	1800.9 (ox) 1802.9 (red)	1803.0 (red)	1802.7 (red)	1803.0 (red)	1802.6 (red)
kB1-ubiquitin	2965.2 (ox) 2971.2 (red)	2965.6 (ox)	2966.6 (ox)	2965.2 (ox)	2966.5 (ox)
Vc1.1-ubiquitin	2460.9 (ox) 2464.9 (red)	2464.7 (red)	2465.3 (red)	2464.7 (red)	2465.3 (red)

[a]Ox, oxidized; red, reduced

[b]Substrate concentrations: SFTI1-I10R-ubiquitin (1 mg mL^{-1} total protein); kB1-ubiquitin (0.9 mg mL^{-1} total protein); Vc1.1-ubiquitin (0.24 mg mL^{-1} total protein)

[c]Enzyme concentrations: r*Oa*AEP1$_b$ and r*Oa*AEP5 (19.7 - 98.5 µg mL^{-1} total protein); r*Oa*AEP3 (19.7 – 21.9 µg mL^{-1} total protein) and r*Oa*AEP4 (19.7 – 30 µg mL^{-1} total protein)

EXAMPLE 8

In vivo cyclization of yeast-expressed polypeptides

[0196] To investigate whether cyclic peptides could be produced *in vivo,* DNA encoding kalata B1 (mature cyclotide domain Gly_1-Asn_{29}; C-terminal tail, Gly_{30}-Pro_{32}) and/or *Oa*AEP1$_b$ (Ala_{24}-Pro_{474}) was introduced into *Pichia pastoris* for co-expression; either from the same or a separate transcriptional unit (Figure 20A).

[0197] For co-expression of kalata B1 and *Oa*AEP1$_b$ from the same transcriptional unit, DNA encoding the ER signal sequence together with the vacuolar targeting sequence (VTR) from *P. pastoris* carboxypeptidase Y (residues Met_1-Val_{107}) (Ohi *et al.* (1996) *Yeast 12*:31–40), kalata B1 and *Oa*AEP1$_b$ were inserted into pPIC9 (Figure 20A, construct 1). The pPIC9 secretion signal was replaced with the vacuolar targeting signal. Optionally an NTR is included between the VTR and the cyclotide domain (Figure 20B, construct 4) or residues Met_1-Lys_{108} of the *P. pastoris* carboxypeptidase Y sequence are included in the construct described above. A linker region (Ala-Ala-Ala-Gly-Gly-Gly-Gly-Gly-Ser – SEQ ID NO:18) was included between kalata B1 and *Oa*AEP1$_b$ to reduce steric hindrance between the cyclotide and AEP domains at the protein level and introduce restriction sites for easy substitution of the cyclotide domain with DNA sequences encoding other target peptides. Alternative linkers could incorporate the MGEV linker (Glu-Glu-Lys-Lys-Asn – SEQ ID NO:17) or an extended sequence (e.g.Ala-Ala-Ala-[Gly-Gly-Gly-Gly-Gly-Ser]$^{2\text{-}5}$). The foreign DNA was then introduced into GS115 *P. pastoris* cells. The vector encoding kalata B1 and *Oa*AEP1$_b$ was then linearized by restriction digestion with SalI and introduced into GS115 cells where it was integrated into the genome at the *his4* locus.

[0198] Kalata B1 and *Oa*AEP1$_b$ were also expressed from separate transcriptional units (Figure 20 A, constructs 2 and 3). DNA encoding an ER signal sequence and a vacuolar targeting sequence (*P. pastoris* carboxypeptidase Y, residues Met_1-Val_{107}) and kalata B1 (including a short C-terminal tail [Gly-Leu-Pro]) (Figure 20 A, construct 2) was inserted into pPICZα (such that the alpha mating factor secretion signal was cloned out and replaced with the ER signal sequence and vacuolar targeting sequence). The vector was

then linearized with SacI and introduced into GS115 cells where it was integrated into the genome at the 5' *AOX1* locus. Optionally the cyclotide domain is preceded by an NTR inserted C-terminally to the vacuolar targeting sequence (Figure 20 B, construct 5). DNA encoding an ER signal sequence and a vacuolar targeting sequence (*P. pastoris* carboxypeptidase Y, residues Met_1-Val_{107}) and *Oa*AEP1$_b$ (Figure 20 A, construct 3) was inserted into pPIC9 (such that the alpha mating factor secretion signal was cloned out and replaced with the ER signal sequence and the vacuolar targeting sequence). The vector was then linearized by restriction digestion with SalI and introduced into GS115 cells already harboring the kalata B1 construct. The *Oa*AEP1$_b$ construct was integrated into the genome at the *his4* locus.

[0199] GS115 cells harboring the appropriate construct/s were grown in 5 mL buffered minimal glycerol medium (BMG; 10 mM potassium phosphate, pH 6, 0.34% w/v yeast nitrogen base, 4 x 10^{-5}% w/v biotin, 1% v/v glycerol) at 30°C, with shaking, for 48 hours. This starter culture was then used to inoculate 40 mL of BMG and grown at 30°C, with shaking, overnight. Cells were harvested by centrifugation and resuspended in 200 mL buffered methanol medium (BMM; 10 mM potassium phosphate, pH 6, 0.34% w/v yeast nitrogen base, 4 x 10^{-5}% w/v biotin, 1% v/v methanol) to induce recombinant protein expression. The culture was incubated at 30°C, with shaking, for 72 hours and methanol was added to 0.5% every 24 hours. After 72 hours, cells were harvested by centrifugation and resuspended in breaking buffer (30 mM HEPES, pH 7.4, 500 mM NaCl) (Visweswaraiah *et al*. (2011) *J. Biol. Chem. 286(42)*:36568-36579) with an equal volume of glass beads. Cells were disrupted by vigorous agitation using a GenoGrinder (AXT) and soluble material was harvested by centrifugation. Samples were analysed by SDS-PAGE followed by Western blotting (anti-AEP1$_b$ rabbit serum [1:2000]; peroxidase-conjugated anti-rabbit IgG [GE Healthcare; 1:5000] (Figure 21). Expression of *Oa*AEP1$_b$ is evident, as judged by antibody reactivity, however the smeared pattern and higher than predicted apparent molecular weight suggests the protein is modified and may be glycosylated or aggregated.

[0200] The vacuolar targeting signal is added to facilitate trafficking of the expressed proteins to the vacuole of *Pichia pastoris* and the *in vivo* cyclization of target peptides. The cyclic target peptides are then directly purified from the cells. This could be aided by isolation of the vacuolar fraction. This is carried out as previously described (Cabrera and Ungermann (2008) *Methods Enzymol 451*:177–196). Volumes relate to a 1 L culture at OD_{600} and are scaled accordingly. Thawed cells are resuspended in 33.3 mL 0.1M Tris-HCl, pH 9.4, 10 mM dithiothreitol (DTT) and incubated at 30ºC for 10 minutes. Cells are then harvested by centrifugation and resuspended in 6.7 mL spheroplasting buffer (0.18 x YPD [0.18% w/v yeast extract, 0.36% w/v bactopeptone, 0.36% w/v dextrose, pH 5.5], 240 mM sorbitol, 50 mM potassium phosphate pH 7.5). A further 3.3 mL of spheroplasting buffer combined with lyticase (Sigma, as per manufacturer's instructions) is then added and cells are incubated at 30ºC, 20 minutes. Cells are harvested by centrifugation and resuspended in 1.67 mL 15% Ficoll (w/v in PS buffer [10 mM PIPES/KOH, pH 6.8, 200 mM sorbitol]). Dextran solution (10 mg mL^{-1} DEAE-dextran, 10 mM PIPES/KOH, pH 6.8, 200 mM sorbitol) is added to 0.4 mg mL^{-1} and cells are incubated on ice (5 minutes), 30ºC (1.5 minutes), and ice again (5 minutes). Cell lysates are transferred to centrifuge tubes and sequentially layered with 3 mL of 8% w/v Ficoll (in PS buffer), 4% w/v Ficoll (in PS buffer) and PS buffer. The lysate is centrifuged at 110,000x*g* at 4ºC for 90 minutes and vacuoles are collected from the 0-4% w/v Ficoll interface.

[0201] Isolated vacuoles are osmotically lysed (Wiederhold *et al*. (2009) *Mol Cell Proteomics 8*:380–392) by addition of a four-fold volume of 20 mM Tris-HCl, pH 8, 10 mM $MgCl_2$, 50 mM KCl (30 minutes, 4ºC with agitation). The lysed vacuoles are filtered through a 0.22 µm filter, further diluted 1:4 with 20 mM Tris-HCl, pH 8 and bound to a Q sepharose high performance anion exchange resin (GE Healthcare). Bound kalata B1 is recovered by a continuous salt gradient (0-100% 20 mM Tris-HCl, 1M NaCl pH 8), buffer exchanged into ultrapure water and concentrated. For further purification, the sample is loaded onto an Agilent Zorbax C18 reversed-phase column (4.6 x 250 mm, 300 Å) and separated using a linear gradient of 5-55% buffer B (90% acetonitrile, 10% v/v H_2O, 0.05% v/v TFA) in buffer A (0.05% v/v TFA/H_2O) over 60 minutes. Fractions containing

kalata B1 are lyophilized, resuspended in ultrapure water and analyzed by MALDI MS, HPLC or NMR to confirm its cyclic structure.

[0202] As cyclized proteins are generally more stable than linear proteins a crude extract could also be heated to 70˚C for 1 hour after cell disruption and centrifuged at 4000g for 20 minutes to denature and remove the majority of non-cyclized cellular protein. Cyclized protein will then be purified from the cleared extract as described below for the vacuolar extract.

EXAMPLE 9

Polypeptide ligation

[0203] The plant defensin NaD1 (Lay *et al*. (2003) *Plant Physiol 131*:1283-1293) with a C-terminal flanking sequence that incorporates an AEP cleavage site and a 6xHis tag (NaD1-**STRNGLP**HHHHHH – SEQ ID NO:12; 280 µM) is incubated with a ligation partner (GLPVSGEK – SEQ ID NO:13-fluorescein isothiocyanate [FITC] or GLPVSGE; - SEQ ID NO:14 - 5.6 mM) and r*Oa*AEP1$_b$ (12 µg mL^{-1} total protein) in activity buffer (50 mM sodium acetate, 50 mM NaCl, 1 mM EDTA, 0.5 µM TCEP, pH 5) for 22 hours at room temperature (Figure 22). The appearance of the ligation product (NaD1-STRNGLPVSGEK-FITC – SEQ ID NO:16 or NaD1-STRNGLPVSGE- SEQ ID NO:15) is tracked by MALDI MS. To separate unprocessed NaD1 from ligated product, the mixture is diluted 1:5 in non-denaturing lysis buffer (without triton X 100) and passed over a Ni-NTA resin (QIAgen). The ligated product does not contain a 6xHis tag and is therefore present in the unbound fraction. This product is then dialyzed into ultrapure water (3 Da molecular weight cut off to ensure the leaving group is also removed), concentrated, and analysed by MALDI MS to confirm the correct ligation product has been generated. Ligation of short, labeled peptides to a larger polypeptide provides a generic, targeted protein labeling strategy for a variety of moieties (e.g. other fluorescent labels, biotin, affinity tags) that is limited only by the ability of synthetic peptide chemistry to produce the appropriate ligation partner.

EXAMPLE 10

Expression and activation of other recombinant AEPs in E. coli

[0204] AEPs from *Cicer arietinum* (SEQ ID NO: 92), *Medicago truncatula* (SEQ ID NO: 93), *Hordeum vulgare* (SEQ ID NO: 94), *Gossypium raimondii* (SEQ ID NO: 95) and *Chenopodium quinoa* (SEQ ID NO: 96) are recombinantly expressed in *E. coli*. DNA encoding these full-length AEPs without the putative signalling domain (*Ca*AEP residues Q_{56}-P_{460}, *Mt*AEP residues E_{54}-N_{497}, *Hv*AEP residues G_{60}-Y_{508}, *Gr*AEP residues Q_{31}-H_{500}, *Cq*AEP residues R_{33}-V_{599}) is inserted into the pHUE vector (Catanzariti *et al*. (2004) *supra*) to give a 6xHis-ubiquitin-AEP fusion protein construct. Residue numbering is as determined by a multiple alignment of the five sequences generated using Clustal Omega (Sievers *et al.* (2011) *supra*). DNA is then introduced into T7 Shuffle *E. coli* cells (New England BioLabs). Transformed cells are grown at 30°C in superbroth (3.5% tryptone [w/v], 2% yeast extract [w/v], 1% glucose [w/v], 90 mM NaCl, 5 mM NaOH) to mid-log phase; the temperature is then reduced to 16°C and expression is induced with isopropyl ß–D-1-thiogalactopyranoside (IPTG; 0.4 mM; Bio Vectra) for approximately 20 hours. Cells are harvested by centrifugation and resuspended in non-denaturing lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% triton X 100, 1mM EDTA, pH 7). Lysis is promoted by a total of five freeze/thaw cycles and the addition of lysozyme (hen egg white; Roche; 0.4 mg mL^{-1}). DNase (bovine pancreas; Roche; 40 µg mL^{-1}) and $MgCl_2$ (0.4 M) are also added. Cellular debris is removed by centrifugation and the lysate is stored at -80°C until required.

[0205] Lysate containing expressed recombinant AEPs is filtered through a 0.1 µM glass fibre filter (GE Healthcare) before being diluted 1:8 in buffer A (20 mM bis-Tris, 0.2 M NaCl, pH 7) and loaded onto two 5 mL HiTrap Q Sepharose high performance columns connected in series (GE Healthcare; 1.6 – 3.1 mL undiluted lysate mL^{-1} resin). Bound proteins are eluted with a continuous salt gradient (0-30% buffer B [20 mM bis-Tris, 2 M NaCl, pH 7]; 15 column volumes [cv]) and AEP-positive fractions are identified by Western blotting (anti-$AEP1_b$ rabbit serum [1:2000]; peroxidase-conjugated anti-rabbit IgG [GE Healthcare; 1:5000]).

[0206] AEPs are usually produced as zymogens that are self-processed at low pH to their mature, active form (Hiraiwa *et al*. (1997) *supra*; Hiraiwa *et al*. (1999) *supra*; Kuroyanagi *et al*. (2002) *supra*). To self-activate all AEPs, EDTA (1 mM) and TCEP (Sigma-Aldrich; 0.5 mM) are added, the pH is adjusted to 4.5 with glacial acetic acid and the protein pool is incubated for 5 hours at 37ºC. Protein precipitation at pH 4.5 allows removal of the bulk of the contaminating proteins by centrifugation. The remaining protein is filtered (0.22 µm; Millipore), diluted 1:8 in buffer A2 (50 mM acetate, pH 4) then captured on a 1 mL HiTrap SP Sepharose high performance column (GE Healthcare). Bound proteins are eluted with a salt gradient (0-100% buffer B2 [50 mM acetate, 1 M NaCl, pH 4]; 10 cv) and fractions with activity against an IQF peptide (Abz-STRNGLPS-Y($3NO_2$) or other target sequence or fluorescent peptide as appropriate) are pooled and used in subsequent activity assays. The total concentration of protein in each preparation is estimated by BCA assay according to the manufacturer's instructions. Enzymes are used in cyclization and ligation assays as described in the Materials and Methods.

[0207] Those skilled in the art will appreciate that aspects of aspects described herein are susceptible to variations and modifications other than those specifically described. It is to be understood that these aspects include all such variations and modifications. These aspects also include all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of the steps or features.

[0208] The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

BIBLIOGRAPHY

Altschul *et al.* (1997) *Nucl Acids Res 25:* 3389-3402

Altschul *et al.* (1990) *J Mol Biol 215:* 403-410

Arnison *et al.* (2013) *Nat Prod Rep 30***:**108–160

Ausubel *et al.* (*In: Current Protocols in Molecular Biology*, John Wiley & Sons Inc. 1994-1998

Barber *et al*. (2013) *J. Biol. Chem 288*:12500-12510

Bernath-Leven *et al*. (2015) *Chemistry & Biology 22*:1-12

Cabrera and Ungermann (2008) *Methods Enzymol 451*:177–196

Camarero *et al*. (2001) *Bioorganic Med Chem 9*:2479-2484

Catanzariti *et al.* (2004) *Protein Sci 13*:1331–1339

Chan *et al.* (2013) *Chembiochem 14*:617–624

Clark *et al*. (2005) *Proc. Natl. Acad. Sci. United States Am. 102*:13767-13772

Clark *et al.* (2010) *Angew. Chem. Int. Ed. Engl.* 49*:*6545–6548

Colgrave *et al.* (2008) *Biochemistry 47***:**5581–5589

Colgrave *et al.* (2009) *Acta Trop. 109***:**163–166

Dall et al. (2015) *Angewandte Chemie (International Ed. in English)* 54: 2917-2921

Gillon *et al.* (2008) *Plant J. 53***:**505–515

Goransson *et al*. (2004) *J. Nat. Prod. 67*:1287–1290

Gran (1973) *Acta Pharmacol. Toxicol. 33*:400–408

Gustafson *et al.* (2000) *J. Nat. Prod 63:*176–178

Hanada *et al*. (2004) *Nature 427***:**252-256

Harlow and Lane (1988) *Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratories.*

Harris *et al.* (2005) *Infect. Immun. 73***:**6981–6989

Harris *et al.* (2009) *J. Biol. Chem. 284***:**9361–9371

Hatsugai *et al*. (2004) *Science 305(5685):* 855-858

Hiraiwa *et al*. (1997) *Plant J 12(4)*:819-829

Hiraiwa *et al*. (1999) *FEBS Lett 447(2-3)*:213-216

Jennings *et al*. (2001) *Proc. Natl. Acad. Sci. U. S. A. 98***:**10614–10619

Kuroyanagi *et al*. (2002) *Plant Cell Physiol 43(2)*:143-151

Lay *et al*. (2003) *Plant Physiol 131*:1283-129

Lay *et al*. (2012) *J Biol Chem 287*:19961-19972

Lee *et al*. (2009) *J. Am. Chem. Soc. 131***:**2122–2124

Li *et al.* (2006) *Bioinformatics 22:* 1658-1659

Li *et al.* (2009) *Bioinformatics* 25: 1754-1760

Lindholm *et al.* (2002) *Mol. Cancer Ther. 1:*365–369

Luo *et al.* (2014) *Chem. Biol.* 1–8 doi:10.1016/j.chembiol.2014.10.015

Mazmanian *et al*. (1999) *Science (80) 285***:**760–763

Mylne *et al.* (2011) *Nat. Chem. Biol. 7***:**257–925

Mylne *et al.* (2012) *Plant Cell 24***:**2765–2778

Nguyen *et al.* (2014) *Nat. Chem. Biol. 10***:**732–738

Nilsson *et al*. (1989) *Cell 58:*707-718

Nolde *et al.* (2011) *J Biol Chem 286(28):*25145-25153

Ohi *et al.* (1996) *Yeast 12*:31–40

Plan *et al*. (2008) *J. Agric. Food Chem. 56***:**5237–5241

Poth *et al*. (2013) *Biopolymers 100*:480–491

Qin *et al.* (2010) *BMC Genomics 11*: 111

Quimbar *et al.* (2013) *J Biol Chem 288(19):*13885-13896

Rotari *et al*. (2001) *Biol. Chem. 382:*953-959

Sambrook *et al*. (1989) *Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory, Plainview, New York.*

Saska *et al*. (2008) *Journal of Chromatography B. 872*:107-114

Saska *et al.* (2007) *J. Biol. Chem. 282***:**29721–29728

Schulz *et al.* (2012) *Bioinformatics 28:* 1086-1096

Sheldon *et al*. (1996) *Biochem. J. 320***:**865–870

Sievers *et al*. (2011) *Mol. Syst. Biol 7*: 539

Simonsen *et al*. (2004) *FEBS Lett.* 577*:*399–402

Tam *et al*. (1999) *Proc. Natl. Acad. Sci. U. S. A. 96*:8913–8918

Visweswaraiah *et al*. (2011) *J. Biol. Chem. 286(42)*:36568-36579

Wiederhold *et al*. (2009) *Mol Cell Proteomics 8*:380–392

Witherup *et al*. (1994) *J. Nat. Prod* 57:1619–1625

Wu and Hancock (1999) *Antimicrob Agents Ch 43*:1274-1276

CLAIMS:

1. A method for producing; (a) a cyclic peptide; or (b) a peptide conjugate, wherein:

(i) said cyclic peptide is generated by a method comprising generating a recombinant asparaginyl endopeptidase (AEP) vacuolar processing enzyme with peptide cyclization activity in a prokaryotic or eukaryotic cell and co-incubating the recombinant AEP with a linear polypeptide precursor of the cyclic peptide wherein the polypeptide precursor comprises N-terminal and/or C-terminal AEP processing site(s) for a time and under conditions sufficient to generate the cyclic peptide; and

(ii) said peptide conjugate is generated by a method comprising co-incubating at least two peptides wherein at least one peptide comprises a C-terminal AEP recognition amino acid sequence and at least one other peptide comprises an N-terminal AEP recognition amino acid sequence with a recombinant AEP with peptide ligase activity for a time and under conditions sufficient to generate a linear peptide conjugate;

wherein the recombinant AEP comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:1, 2, 4, 39, 41 and amino acid sequences having at least 80% similarity to any of the foregoing after optimal alignment; and

wherein the recombinant AEP comprises the presence of 5 or more of residues or absence of residues at 139K, 161D, 186K, 192D, 247C, 248Y, 253Q, 255A, 263V, 293H, Gap, Gap, Gap, Gap, Gap (between residues 299 and 300), 314E and 316G when optimally aligned to SEQ ID NO:1, wherein Gap means the absence of a residue.

2. The method of Claim 1, wherein the recombinant AEP comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:1, 2, 4, 39 and 41.

3. The method of Claim 1 or Claim 2, wherein the recombinant AEP is co-incubated with a linear polypeptide precursor or a post-translationally or synthetically modified form thereof *in vitro* in a reaction vessel for a time and under conditions sufficient to generate the cyclic peptide or peptide conjugate.

4. The method of Claim 1 or Claim 2 comprising introducing an expression vector into a prokaryotic or eukaryotic cell encoding the linear polypeptide precursor, enabling expression of the vector to produce a recombinant linear polypeptide precursor and isolating the polypeptide from the cell and co-incubating in a reaction vessel the

polypeptide precursor with the recombinant AEP for a time and under conditions sufficient to generate the cyclic peptide or peptide conjugate.

5. The method of Claim 1 or Claim 2, wherein the polypeptide precursor is in the form of multiple repeats of the peptide to be cyclized or conjugated or is in the form of multiple different polypeptides to be cyclized or conjugated.

6. The method of Claim 1 or Claim 2, wherein the cyclic peptide or peptide conjugate exhibits antipathogenic or therapeutic properties including for the treatment of infection or infestation by a pathogen or treatment of cancer, cardiovascular disease, immune disease and pain.

7. The method of Claim 1 or Claim 2, wherein the C-terminal AEP recognition amino acid sequence comprises P3 to P1 prior to the actual cleavage site and comprising P1' to P3' after the cleavage site towards the C-terminal ends wherein P3 to P1 and P1 to P3 have the amino acid sequence:

$X_2X_3X_4X_5X_6X_7$

wherein X is an amino acid residue and:

X_2 is optional or is any amino acid;

X_3 is optional or is any amino acid;

X_4 is N or D;

X_5 is G or S;

X_6 is L or A or I; and

X_7 is optional or any amino acid;

and/or wherein the N-terminal recognition amino acid sequence may contain no specific AEP processing site or may contain a processing site defined by any one of the P1" through P3" wherein P1" to P3" is defined by:

$X_9X_{10}X_{11}$

wherein X is an amino acid residue:

X_9 is optional and any amino acid or G, Q, K, V or L;

X_{10} is optional or any amino acid or L, F or I or an hydrophobic amino acid

residue;

X_{11} is optional and any amino acid.

8. The method of Claim 7, wherein X_2 through X_7 comprise the amino acid sequence:

$X_2X_3NGLX_7$

wherein X_2, X_3 and X_7 are as defined in Claim 10 above; and

wherein X_9 through X_{11} comprise the amino acid sequence:

GLX_{11}

wherein X_{11} is optional and any amino acid.

9. The method of Claim 1 or Claim 2, wherein the AEP processing site comprises N- and C-terminal end sequences comprising the sequence:

GLX_{11} $[X_n]$ $X_2X_3NGLX_7$

wherein X_{11}, X_2, X_3, and X_7 are as defined above and $[X_n]$ is absent (n=0) or any amino acid residue in a sequence of from 1 to 2000 amino acids.

10. An agronomical composition or pharmaceutical composition comprising the cyclic peptide or peptide conjugate generated by the method of Claim 1 or Claim 2.