

NRC Publications Archive Archives des publications du CNRC

The Two-step biosynthesis of cyclic peptides from linear precursors in a member of the plant family Caryophyllaceae involves cyclization by a Serine Protease-like enzyme

Barber, Carla J. S.; Pujara, Pareshkumar T.; Reed, Darwin W.; Chiwocha, Shiela; Zhang, Haixia; Covello, Patrick S.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.1074/jbc.M112.437947>

The Journal of Biological Chemistry, 288, 18, pp. 12500-12510, 2013-05-03

NRC Publications Archive Record / Notice des Archives des publications du CNRC :

<https://nrc-publications.canada.ca/eng/view/object/?id=faa9a231-9564-4805-8773-fdee91611fdc>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=faa9a231-9564-4805-8773-fdee91611fdc>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

The two step biosynthesis of cyclic peptides from linear precursors in a member of the plant family Caryophyllaceae involves cyclization by a serine protease-like enzyme*

Carla J.S. Barber¹, Pareshkumar T. Pujara¹, Darwin W. Reed¹, Shiela Chiwocha¹, Haixia Zhang¹ and Patrick S. Covello^{1*}

¹From the National Research Council of Canada, 110 Gymnasium Place, Saskatoon, SK, Canada S7N 0W9

*Running Title: Caryophyllaceae cyclic peptide biosynthesis

To whom correspondence should be addressed: Patrick Covello, National Research Council of Canada, 110 Gymnasium Place, Saskatoon, SK Canada S7N 0W9; Tel.: (306) 975-5269; Fax: (306) 975-4839; E-mail: patrick.covello@nrc-cnrc.gc.ca

*The cDNA sequence reported in this paper for Pcy1 has been submitted to the Genbank™/EBI Data Bank with accession number KC588970; nucleotide sequences for two Sanger sequencing-based EST collections from *S. vaccaria* referenced in this paper have been submitted to Genbank™/EBI Data Bank with accession numbers LIBEST_028081 and LIBEST_028082; nucleotide sequences for two additional *S. vaccaria* EST collections have been submitted to the short read archive at NCBI with accession numbers SRX202186 and SRX202889.

Keywords: Peptidases; Peptide biosynthesis; Plant biochemistry; Plant molecular biology; Protease; Caryophyllaceae; *Saponaria vaccaria*; cyclic peptide; peptide cyclase; orbitide

Background: In the Caryophyllaceae, cyclic peptides (CP) are biosynthesized from linear precursors via an unknown pathway.
Results: Two protease-like enzymes are involved in precursor processing.
Conclusion: A serine protease-like enzyme was recruited for the cyclization step in CP biosynthesis.
Significance: This represents a very significant advance in our understanding of the mode and evolution of CP biosynthesis in plants.

SUMMARY

Caryophyllaceae-type cyclic peptides (CP) of 5-12 proteinogenic amino acids occur in ten plant families. In *Saponaria vaccaria* (Caryophyllaceae), these have been shown to be formed from linear peptide precursors derived from ribosomal translation. There is also evidence for such precursors in others members of the Caryophyllaceae, Rutaceae and Linaceae. The biosynthesis of CP in the developing seeds of *S. vaccaria* was investigated with respect to the enzymes involved in

precursor processing. Through biochemical assays with seed extracts and synthetic peptides, an enzyme activity named Oligopeptidase 1 (OLP1) was found which catalyzes the cleavage of intermediates at the N-terminus of the incipient CP. A second enzyme, Peptide Cyclase 1 (PCY1), which was separated chromatographically from OLP1 was found to act on the product of OLP1 giving rise to a cyclic peptide and concomitant removal of a C-terminal flanking sequence. PCY1 was partially purified and, using the methods of proteomics, a full length cDNA clone encoding an enzyme matching the properties of PCY1 was obtained. The substrate specificity of purified recombinant PCY1, believed to be the first cloned plant enzyme whose function is peptide cyclization, was tested with synthetic peptides. The results are discussed in the light of CP biosynthetic systems of other organisms.

In 1959, Kaufmann and Tobschirbel (1) reported the isolation of a cyclic nonapeptide, cyclolinopeptide A, from flax (*Linum*

usitatissimum) oil with the formula *cyclo*[ILVPPFFLI]. This is the first of the class which came to be called Caryophyllaceae-type CP. CP have been discovered in a wide range of taxa, including eubacteria, fungi, plants and animals (2-8). Recently, the name orbitide has been suggested for the Caryophyllaceae-type CP (9). Many CP are known to have interesting biological activities (4,10). For example segetalin F has been reported to have vasorelaxant activity (11). CP can be formed with a variety of chemical linkages beyond the peptide bond. In plants, however, CP tend to be homodetic, in the sense that at least one ring of amino acids involves peptide bond linkages exclusively. There is also a tendency in plants for CP to include only proteinogenic amino acids, or simple derivatives of them. The four major classes of CP in plants include the Caryophyllaceae (orbitide), kalata, PawS and knottin types (9). These differ in typical size, and degree of disulfide bridging, for example.

Given the physiological roles and technological and pharmaceutical interest in CP (12-14), understanding the natural and engineered processes for CP production is currently an active area of research. For many CP in eubacteria and fungi, their biosynthesis involves the linkage and modification of amino acids and other compounds with the help of non-ribosomal peptide synthetases (15,16). More recently, it has become clear that in cyanobacteria, fungi and especially in plants, CP biosynthesis can occur via the processing of ribosomally-produced precursors (17-26). Indeed, for the case of plant CP, in essentially all cases for which there is sufficient evidence, ribosome-derived precursors are involved in CP biosynthesis. As well, there are examples from cyanobacteria and fungi in which CP are formed from ribosome-derived precursors (24,26).

The Caryophyllaceae-type CP occur in members of the Caryophyllaceae, Linaceae, Rutaceae and seven other plant families. They typically consist of 5 to 12 proteinogenic amino acids in a single ring formed from peptide bonds. In some cases, nonproteinogenic amino acids are included in the structure. Within the Caryophyllaceae, *Saponaria vaccaria* (syn. *Vaccaria hispanica*) has been investigated and 8 CP called segetalins (A to H) have been isolated (10,27). More recently, evidence for two

additional segetalins (J and K) has been reported (17).

It appears that these segetalins are biosynthesized via precursors of ~35 amino acids which have conserved N- and C- terminal flanking regions and a variable region representing the sequence of the mature segetalin (see Fig. 1)(17). Expression of cDNAs encoding the presegetalins in transformed roots of *S. vaccaria* gives rise to CP. In this paper, we describe efforts to understand the process by which precursors of Caryophyllaceae-type CP, such as presegetalins, are processed to CP. In an important advance in our understanding of CP biosynthesis, a combination of partial enzyme purification, mass spectrometry and bioinformatics has led to the elucidation of a biosynthetic process involving two peptidase-like enzymes and the molecular cloning and preliminary characterization of the enzyme which catalyzes the cyclization step.

EXPERIMENTAL PROCEDURES

Chemicals-Presegetalin A1 (purity $\geq 75\%$; see Table 1) and presegetalin A1[14,32] (purity $> 75\%$) were chemically synthesized at the Sheldon Biotechnology Centre, McGill University, Montreal, PQ, Canada). The presegetalin A1 was further purified by HPLC fractionation with a C18 column using a water/acetonitrile gradient (with trifluoroacetic acid as a modifier). All other linear peptides (purity, $>90\%$; see Table 1) were chemically synthesized at Bio Basic Canada Inc. (Markham, ON, Canada). Segetalin A isolated from *S. vaccaria* seed by the method of Morita (28) was obtained from John Balsevich (National Research Council of Canada, Saskatoon, Canada)

Plant material and RNA isolation-Seeds of *S. vaccaria* 'White Beauty' were obtained from CN Seeds Ltd (United Kingdom). Plants were grown under greenhouse conditions with a daily regime of 12 h light and 12 h dark at 23°C. *S. vaccaria* developing seeds at stages 1 and 2 were harvested according to the following scheme: Stage 1, seed white, pod green; Stage 2, seed tan, pod green; Stage 3, seed copper, pod partially desiccated; Stage 4, seed dark brown, pod desiccated.

For total RNA isolation from *S. vaccaria* developing seeds, the protocol of Gambino *et al.* (29) was modified. For the rapid CTAB-based procedure, 0.6 mL of extraction buffer, containing

2% cetyltriethylammonium bromide (CTAB), 2.5% polyvinylpyrrolidone ($M_r = 40,000$), 2 M NaCl, 100 mM Tris-HCl, pH 8.0, 25 mM EDTA and 2% of β -mercaptoethanol (added just before use), was heated to 65 °C in a microcentrifuge tube. One hundred and fifty mg of developing seeds were ground in liquid nitrogen and added to the extraction buffer and the tube was incubated at 65 °C for 10 min. The sample was extracted two times with chloroform/isoamyl alcohol (24:1 v/v) and 0.25 volumes of 3 M LiCl was added. The mixture was kept on ice for 30 min and centrifuged at 20,000 g for 20 min at 4 °C. The pellet was dissolved in 0.5 mL of SSTE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1% SDS, 1 M NaCl), extracted with 0.5 mL of chloroform/isoamyl alcohol (24:1, v/v), followed by precipitation with isopropanol.

cDNA sequencing—A collection of *S. vaccaria* developing seed ESTs based on Roche 454 sequencing technology was developed as follows. First-strand cDNA was synthesized from *S. vaccaria* Stage 1 developing seed total RNA using the Omniscript® Reverse Transcription Kit (Qiagen, Mississauga, Canada) according to the manufacturer's instructions. ESTs were generated from cDNA prepared from the isolated RNA using Roche (Indianapolis, IN, USA) GS-FLX Titanium Technology at the McGill and Genome Quebec Innovation Centre (Montreal, Canada) according to the manufacturer's instructions. Within the MAGPIE software system (30), sequences were assembled using MIRA (31) and contigs were annotated based on BLASTX searches of Genbank.

Protein assay—For enzyme assays and purification, protein was measured using a modified micro BCA protein assay (Pierce, <http://www.piercenet.com>). Ten μ L samples were mixed with 100 μ L BCA Working Reagent and incubated at 60 °C for 30 min and after which optical density at 562 nm was recorded.

Oligopeptidase 1 (OLP1) assay—The cleavage of presegetalin A1 (Table 1) by plant extracts was tested using the assay conditions reported by Condie *et al.* (17). The formation of presegetalin A1[14,32] was determined by ion trap LC/MS (see below).

Peptide Cyclase 1 (PCY1) assay of plant extracts—The formation of segetalin A [cyclo(GVPVWA)] and linear segetalin A

(GVPVWA) from presegetalin A1[14,32] (Table 1) by plant extracts was tested as described by Condie *et al.* (17) except that 15 μ g/ml presegetalin A1[14,32] was used as the substrate.

Assay of recombinant PCY1—Unless otherwise noted, the activity of recombinant PCY1 was determined in an optimized assay as follows. The assay contained 20 mM Tris (pH 8.5), 100 mM NaCl, 5 mM DTT, 0.2 mg BSA, and 15 μ g/ml of presegetalin A1[14,32] or other synthetic peptides. The reaction was initiated by the addition of 0.3 μ g of recombinant PCY1, in a total reaction volume of 100 μ L. The assay was incubated at 30°C for up to 1 h and stopped by the addition of 0.9 ml of methanol. Samples were centrifuged and the supernatants were evaporated and resuspended in 50% (v/v) methanol in water. The samples were then analyzed by ion trap LC/MS (see below). Controls were performed with either omission of enzyme or stopping the reaction at 0 h. Where necessary, 0 h peak areas were subtract from 1 h peak areas for the linear product.

Enzyme Purification—Eight grams of Stage 2 developing seeds from *S. vaccaria* (var. White Beauty) were homogenized manually with a plastic pestle in 1.5 mL low protein binding microcentrifuge tubes. The seeds were ground for 2 min in 20 aliquots of 500 μ L 20 mM Tris buffer (pH 8) on ice followed by centrifugation at 13,000 g for 10 min. The supernatants were removed and additional 250 μ L aliquots of buffer were added and the grinding and centrifugation was repeated. The supernatant fractions were pooled to a total volume of 17 ml, and filtered (0.2 μ m 25 mm cellulose acetate membrane, VWR International, Mississauga, Canada). The resulting crude filtrate was used for enzyme assays and chromatographic purification.

Three separate applications of 5 mL each of the crude filtrate (see above) were made to an anion exchange column (Mono Q 10/100, GE Healthcare Life Sciences, Mississauga, Canada) connected to an Agilent 1100 series HPLC equipped with an auto injector, diode array detector and fraction collector cooled to 4°C. The column was held at 4 °C and pre-equilibrated with 20 mM Tris-HCl (pH 8.0). The column was eluted with 180 mL of a linear gradient of NaCl (0-1 M) in 20 mM Tris-HCl (pH 8.0) at a flow rate of 3 mL/min. Six mL fractions were collected, desalted with Sephadex G-25 M PD-10 columns (GE

Healthcare Life Sciences, Mississauga, Canada) and then concentrated to approximately 200 μ L in Amicon Ultra centrifugal filters (Ultracel 30K cellulose, 30 MWCO; Millipore, Bellerica, MA, USA). These fractions were assayed for both OLP1 as well as PCY1 activities. Fractions containing PCY1 activity were combined in a loading buffer containing 3 M ammonium sulfate in 20 mM Tris-HCl (pH 8.0) and applied to a hydrophobic interaction perfusion chromatography column (PerSeptive POROS® 20 HP2, Bio-Rad Laboratories Ltd, Mississauga, Canada) pre-equilibrated with 3 M ammonium sulfate in 20 mM Tris-HCl (pH 8.0). The column was eluted with a 60 mL decreasing linear gradient of ammonium sulfate (3-0 M in 20 mM Tris-HCl, pH 8.0) at a flow rate of 4 mL/min. Four mL fractions were collected, desalted with Sephadex G-25 M PD-10 columns and then concentrated to approximately 200 μ L in Amicon Ultra centrifugal filters and assayed for PCY1 activity. Fractions containing PCY1 activity were combined and concentrated to 100 μ L and applied to a Superose™ 6 10/300 Gel Filtration column (GE Healthcare Life Sciences, Mississauga, Canada) which had been pre-equilibrated with 20 mM Tris-HCl, 100 mM NaCl (pH 8.0). Proteins were eluted with 20 mM Tris-HCl, 100 mM NaCl (pH 8.0) over 30 ml at a flow rate of 0.2 mL/min. One mL fractions were collected and concentrated to ~200 μ L with Amicon® Ultracel-10K membrane centrifugal filter units (Millipore, Bellerica, MA, USA) and assayed for PCY1 activity.

The retention times of standard proteins [thyroglobulin (M_r = 669,000), ferritin (M_r = 440,000), catalase (M_r = 232,000), aldolase (M_r = 158,000), BSA (M_r = 67,000), ovalbumin (M_r = 43,000), chymotrypsinogen (M_r = 25,000) and ribonuclease A (M_r = 14,000); GE Healthcare Life Sciences, Mississauga, Canada] were measured in a separate chromatography experiment under identical conditions. The size exclusion chromatography indicated that the relative molecular mass of the peptide cyclase was approximately 90,000 (data not shown).

Ion trap LC/MS—For ion trap LC/MS analysis of enzyme assays, an Agilent 6320 Ion Trap LC/MS system was used under default Smart Parameter settings. The analyzer and ion optics were adjusted to achieve optimal resolution (Agilent Installation Guide #G2440-90105) using

the ESI Tuning Mix (Agilent #G2431A). The mass spectrometer was scanned in the m/z range of 50 to 2200 at 8100 mass units/s with an expected peak width of ≤ 0.35 mass units. For automated MS/MS, the trap isolation width was 4 atomic mass units. The associated Agilent 1200 LC was fitted with a Zorbax 300 EXTEND-C18 column (150 X 2.1 mm, 3.5 μ m particle size) maintained at 35 C. The binary solvent system consisted of 90:10 v/v water/acetonitrile containing 0.1% formic acid and 0.1% ammonium formate (solvent A) and 10:90 v/v water/acetonitrile containing 0.1% formic acid and 0.1% ammonium formate (solvent B). The separation gradient was 90:10 A/B to 50:50 A/B in 3 mL over 20 min. The detection of segetalin A in assay samples is described previously (17).

MALDI-TOF MS—MALDI-TOF MS of enzyme assays was performed on samples which were purified by adsorption onto and elution from C18 Empore High Performance Disk material (3M, Minneapolis, MN, USA) using the “Stage tip” method (32). Stage tips were prepared by removing the beveled tip from a 20 gauge syringe needle with a tubing cutter. Empore disk material was then cut, cookie cutter style, with this needle and packed into the narrow end of a 10 μ L disposable pipette tip with a piece of fused silica tubing. Methanol (10 μ L) was applied to the tip and expelled slowly with a 1.25 mL syringe. Aqueous trifluoroacetic acid (TFA; 0.1%) was then passed through the tip, followed by assay sample (20 μ L). The disk material was washed with 20 μ L 0.1% TFA and peptides were then eluted with 20 μ L acetonitrile:aqueous 0.1% TFA (1:1). Analysis of the peptides was carried out using an AB Sciex 4800 MALDI TOF-TOF™ Analyzer (Applied Biosystems, LLC., Frederick, MD, USA). The mass spectrometer was operated in positive ion reflectron mode scanning m/z values from 500 to 4000. The default calibration was updated with a standard mixture of peptides containing des-Arg¹ bradykinin (m/z 904.468), Gu¹ fibrinopeptide B (m/z 1570.677), and three ACTH fragments corresponding to amino acids 1-17 (m/z 2093.087), 18-39 (m/z 2465.199), and 7-38 (m/z 3657.929). All samples and calibrants (0.5 μ L) were mixed on the MALDI plate with the matrix α -cyano-4-hydroxycinnamic acid (0.5 μ L). Data were collected and averaged from 800 laser irradiation events. Monoisotopic mass lists were

generated with Data Explorer (Applied Biosystems) and copied into the BioLynx program in MassLynx 4.0 (Waters). Matches to subsequences of presegetalin A1 were investigated using the Find Mass program with an allowed mass deviation of 0.5 Da. Masses within 0.2 Da were considered to be matching.

Q-TOF LC/MS-Q-TOF LC/MS analysis of tryptic peptides was performed using a Quadrupole Time-Of-Flight (Q-TOF) Global Ultima mass spectrometer (Micromass, Manchester, UK) equipped with a nano-electrospray (ESI) source and a nanoACQUITY UPLC solvent delivery system (Waters, Milford, MA, USA). The mobile phase was composed from a binary solvent system of C, 0.2% formic acid and 3% acetonitrile and D, 0.2% formic acid and 95% acetonitrile. Peptides were desalted with an in-line solid-phase trap column (180 μ m x 20 mm) packed with 5 μ m resin (Symmetry C18, Waters) and separated on a capillary column (100 μ m x 100 mm, Waters) packed with BEH130 C18 resin (1.7 μ m, Waters) using a column temperature of 35 °C. Five μ L samples were introduced into the trap column at a flow rate of 15 μ L/min for 3 min, using C:D 99:1 and flow was diverted to waste. After desalting, the flow was routed through the trap column to the analytical column with a linear gradient of 1-10% solvent D (400 nL/min, 16 min), followed by a linear gradient of 10-45% solvent D (400 nL/min, 30 min). Unless otherwise stated, Q-TOF parameter settings consisted of a capillary voltage of 3,850 V, a cone voltage of 120 V and a source temperature of 80 °C.

Samples were analyzed using Data Dependant Acquisition (DDA), which consisted of the detection of multiply charged positive ions ($z = 2-4$) from an MS survey scan. The scan range was from m/z values of 400 to 1900, with a scan time of 1 s. Up to three MS/MS scans were triggered (collision energy ranged from 20 to 80 eV, depending on charge state and precursor m/z) from each MS scan event with a peak detection window of 4 m/z units (signal intensity threshold was 16 counts/s). In MS/MS experiments, data was acquired in continuum mode with a scan time of 1.9 s and dynamic exclusion of previously detected precursors was set at 2 min. Peptide signals corresponding to trypsin and keratin were also excluded from MS/MS data collection. To obtain high mass accuracy, the reference

compound leucine enkephalin (80 nM in 1:1 acetonitrile:0.2% aqueous formic acid, Environmental Resource Associates, Arvada, CO, USA; $m/z = 556.2771$) was continuously introduced to a second ESI source and used for the mass calibration.

Data was processed with ProteinLynx Global Server 2.4 (PLGS 2.4, Waters) using RAW files from Q-TOF LC/MS. The resulting PKL files were submitted to MASCOT (version 2.3.02; Matrix Science Ltd., London, UK) for peptide searches against the NCBI nr database (version 011110) hosted by National Research Council of Canada and local databases containing the sequence information from Sanger (33) and 454 (this work) sequencing of *S. vaccaria* developing seed cDNA. MASCOT search parameters allowed for a maximum of 1 miscleavage for tryptic digestion and a mass tolerance for precursor peptide ions of ± 50 ppm and for fragment ions up to ± 0.2 Da. Carbamidomethylation of cysteine was selected as a fixed modification and oxidation of methionine was used as a variable modification.

Isolation of a full-length Pcy1 cDNA clone from *S. vaccaria*-Fractions which catalyzed the formation of segetalin A from presegetalin A1 [14,32] from the various stages of chromatography were mixed 1:1 with SDS-PAGE Laemmli sample buffer (200 mM Tris-HCl, pH 6.8, 4% SDS, 0.2% bromophenol blue, 200 mM dithiothreitol, 40% glycerol) and heated at 99° C for 5 min. The samples were subjected to SDS-PAGE under denaturing conditions using Electrophoresis Buffer (25 mM Tris-HCl, pH 7.5, 250 mM glycine, 0.1% SDS) and a 10% Ready Gel® pre-cast polyacrylamide mini-gel (Bio-Rad Laboratories Ltd, Mississauga Canada). Precision Plus Protein™ molecular weight standards (Bio-Rad) were loaded on the same gel. The gel was stained with Oriole™ Fluorescent Gel Stain (Bio-Rad Laboratories Ltd, Mississauga Canada) for 15 h. Protein bands were visualized by UV illumination and the most prominent bands, corresponding to the last chromatographic step, were excised from the gel and digested with trypsin using the MassPrep II Proteomics Workstation (Micromass, UK) following a procedure described previously (34). Tryptic peptides were analyzed by Q-TOF LC/MS and the resulting data was used for MASCOT searches of sequence databases (see above).

A DNA plasmid clone containing the full length open reading frame of contig c272 (from Roche 454 sequencing; NCBI accession SRX202186), identified by MASCOT searches as representing a peptide cyclase candidate cDNA, was obtained as follows. First-strand cDNA was synthesized from *S. vaccaria* developing seed Stage 1 total RNA using the Omniscript® Reverse Transcription Kit (Qiagen, Mississauga, Canada) according to the manufacturer's instructions. The contig c272 ORF was amplified by PCR (denaturation at 95° C for 4 min; 35 cycles of 95° C for 20 s, annealing at 54° C for 30 s and extension at 72° for 2.3 min; followed by 10 min at 72° C) using forward (ATG GCG ACT TCA GGA TTC TCG) and reverse (TCA GTC TAT CCA AGG AGC TTC AAG C) oligonucleotide primers and Platinum® *Taq* DNA Polymerase High Fidelity (Invitrogen, Life Technologies, Mississauga, Canada). The resulting PCR product was gel-purified and recombined with pCR8/GW/TOPO® using a TA Cloning® Kit (Invitrogen, Life Technologies, Mississauga, Canada). After transformation of ONE SHOT® TOP 10 competent *E. coli* cells (Invitrogen, Life Technologies, Mississauga, Canada) using spectinomycin for selection, a plasmid clone (named pCB006) containing the contig c272 ORF was identified by colony PCR and DNA sequencing. The corresponding gene was named *Pcy1*. To allow *E. coli* expression of a N-terminal His₆-tagged version of PCY1, the PCY1 ORF of pCB006 was recombined with the Gateway expression vector pDEST17 (Invitrogen-Life Technologies, Carlsbad, CA, USA) and used to transform competent *E. coli* BL21-AI™. The resulting plasmid and *E. coli* strain were called pCB008 and pCB008/ BL21-AI, respectively.

***E. coli* Expression and Purification of PCY1**—An overnight 1 mL LB culture of pCB008/BL21-AI was used to inoculate 100 mL of Overnight Autoinduction Medium (35) which was incubated at 37 °C with shaking until an OD₆₀₀ of 0.4 was reached. Arabinose was then added to the a concentration of 0.2% (w/v) and culture growth was continued at 16 °C with agitation overnight. The cultures were centrifuged at 2,000 g at 4 °C for 10 min and the resulting cell pellets were frozen at -20 °C. The pellets were resuspended in chilled 500 µL of B-Per® Bacterial Protein Extraction Reagent (Pierce Biotechnology,

Rockford, IL, USA), then transferred to two 1.5 mL Eppendorf tubes for cell lysis at room temperature for 20 min. Lysis was promoted with ultrasonication. The lysates were then centrifuged (12,000 g, 4 °C, 8 min) and the supernatants were mixed with an equal volumes of resin binding Equilibration/Wash Buffer and added to 250 µL HisPur™ Cobalt Resin (Pierce Biotechnology, Rockford, IL, USA) for purification of PCY1 by batch adsorption and elution at 150 mM and 300 mM imidazole, respectively, according to the manufacturer's recommendations. Each eluate was concentrated to 150 µL and desalted by spin dialysis (Amicon Ultra-15 devices; Millipore, Bellerica, MA) following the manufacturer's protocol and subsequently assayed for peptide cyclase activity. The purity of recombinant PCY1 was judged by SDS-PAGE to be ~90%.

RESULTS

Detection of separate oligopeptidase and peptide cyclase activities involved in cyclic peptide biosynthesis—In previous work (17), it was shown that extracts of the developing seeds of *S. vaccaria* were capable of catalyzing the conversion of the 32 amino acid linear precursor, presegetalin A1 to the six-membered CP, segetalin A (see Fig. 1, overall reaction). In a further effort to understand the enzyme or enzymes involved in cyclic peptide formation in the Caryophyllaceae, *S. vaccaria* seed extracts were subjected to separation by ion exchange chromatography. The products of assays (see Experimental Procedures, Fig. 2 and (17)) using presegetalin A1 as a substrate were then analyzed by MALDI-TOF MS (see Experimental Procedures and Fig. 3). Assay of fractions eluting at ~0.35M NaCl (see Fig. 2) showed significant loss of substrate and formation of prominent MS peaks corresponding to peptide masses of 1434.8 and 1984.0 which, in turn, correspond to linear peptides with the sequences MSPILAHDVVKPQ and GVPVWAFQAKDVENASAPV, respectively (see Fig. 3). Daughter ion analysis of the peaks was consistent with the above cleavage products (Fig. S0). This suggests that cleavage of the QG peptide bond in presegetalin A1 is an important reaction in the biosynthesis of segetalin A. The formation of MSPILAHDVVKPQ and GVPVWAFQAKDVENASAPV from presegetalin A1 was confirmed by ion trap LC/MS (Fig. 4a and b). Thus, the data are consistent with a peptide

with the sequence GVPVWAFQAKDVENASAPV (called presegetalin A1[14,32]) being an intermediate in segetalin A biosynthesis. Consequently, the pathway from presegetalin A1 to segetalin A shown in Fig. 1 is hypothesized. Presegetalin A1 is suggested to be cleaved initially after position 13, giving rise to presegetalin A1[1,13] and presegetalin A1[14,32]. The latter is then processed, giving rise to segetalin A.

Given the above results, fractions obtained from ion exchange chromatography were tested with synthetic presegetalin A1[14,32] and indeed, specific fractions were found to catalyze the formation of segetalin A, as well as the linear form of segetalin A (Fig. 1, 2 and 4c and d). This activity was distinguishable from that giving rise to the initial cleavage of presegetalin A1 to presegetalin A1[14,32], being highest in fractions eluting at ~0.2 M NaCl (Fig. 2). This indicates that presegetalin A1[14,32] is a true intermediate formed during CP processing and that at least two enzymes are required for CP processing, one for initial cleavage of presegetalin A1 to form the intermediate presegetalin A1[14,32] (named Oligopeptidase 1; OLP1) and a second enzyme for further processing and cyclization of the intermediate to form segetalin A (named Peptide Cyclase 1; PCY1; see Fig. 1).

Partial purification and cloning of Peptide Cyclase 1 (PCY1)-With a view towards cloning and characterization of the peptide cyclase, PCY1, further purification of ion exchange fractions was pursued using hydrophobic interaction and size exclusion chromatography. Fig. 5 shows the results of SDS-PAGE analysis of the active fractions from each stage of the purification. After the final purification step, bands from SDS-PAGE gels were excised and treated with trypsin. The resulting peptides were each analyzed by Q-TOF LC/MS to give peptide mass and fragment ion information. This data was used in MASCOT searches of the predicted tryptic peptides from two EST databases derived from *S. vaccaria* Sanger (36)(Genbank accessions LIBEST_028081 and LIBEST_028082 and Roche 454 sequencing (this work; NCBI Short Read Archive accession SRX202186). When the MS data from an excised gel band corresponding to a M_r of approximately 83,000 was used in searches, matches were found to two Sanger-sequencing derived ESTs

(SVAR04NG UP 023 E08 and SVAR03NG UP 028 E03) and a set of contiguous cDNAs sequences obtained from Roche 454 sequencing called contig c272 (from the SVASD1PC 454 EST collection). The search yielded 13 unique peptide matches with the predicted amino acid sequence of contig c272 (see Table S1). Given the similarity of the predicted amino acid sequence of c272 to serine proteases (see below) and the well-documented propensity for such proteases to catalyze transpeptidation and cyclization reactions (37-39), the corresponding gene product was investigated as a peptide cyclase. The gene corresponding to c272 was named *Pcy1*. A full length cDNA clone was obtained for *Pcy1* (see Experimental Procedures) which encodes a polypeptide of 725 amino acids with a predicted M_r of 82,400 (Genbank Accession KC588970).

Sequence analysis and homology modeling of *S. vaccaria* PCY1-A BLASTP search of Genbank with the predicted amino acid sequence of *Pcy1* revealed greatest sequence identity with members of the enterase lipase superfamily (COG1505) (40). In particular, PCY1 shows highest amino acid sequence identity to predicted gene products from *Vitis vinifera* (Genbank accession number CAN70125) and *Populus trichocarpa* (Genbank accession number XP_002306966; see Fig. S2 for alignment). Further sequence analysis strongly suggests placement of PCY1 within the S9A family of serine peptidases. This family includes porcine (*Sus scrofa*) muscle prolyl oligopeptidase (POP; PDB ID:1QFS, Genbank accession AAA31110), which shows 49% sequence identity to *S. vaccaria* PCY1 and for which a crystal structure has been determined in the presence of a covalently-bound inhibitor (41). The structure of porcine muscle POP was used as a template for the construction of a homology model of PCY1 by MODELLER (version 9.10; Fig. 6)(42). The structural alignment of PCY1 and porcine muscle POP showed a root mean square deviation of C α positions of 0.26 Å. Given the excellent agreement of the model with the structure of porcine muscle POP, PCY1 appears to possess two domains homologous to those of porcine muscle POP - a catalytic α/β hydrolase domain and an unusual β -propeller domain. Also, the model is consistent with the identity of S562, D653 and H695 (PCY1 numbering) as members of a serine-peptidase-like

catalytic triad which shows excellent three dimensional alignment with those of porcine muscle POP (see Fig. 6). The putative catalytic triad faces the β -propeller domain. S562 occurs within the sequence GGSNGG, which is conserved among γ/β hydrolases and likely forms a ‘nucleophilic elbow’ (43). Despite the alignment of catalytic amino acids, there is a notable single amino acid insertion/deletion near H695 (Figure S2).

Recombinant PCY1 shows peptide cyclase activity – Given that serine peptidases have a well documented propensity to catalyze transpeptidation and cyclization reactions (37,38) and the similarity of PCY1 to oligopeptidases, recombinant PCY1 was investigated as a candidate for involvement in CP biosynthesis in *S. vaccaria*. For functional characterization of PCY1, the recombinant enzyme was purified from *E. coli* cells harbouring the plasmid pCB008 which consists of the *Pcy1* ORF fused with an N-terminal His₆ tag. The activity of immobilized metal ion affinity chromatography-purified PCY1 was assayed with presegetalin A1[14,32] using ion trap LC/MS. As did partially purified plant extracts, purified recombinant PCY1 showed the formation of segetalin A and smaller amounts of linear segetalin A in the presence of presegetalin A1[14,32] (Fig. 4c, d and e). Evidence for the formation of the C-terminal fragment presegetalin A1[20,32] was also found by ion trap LC/MS/MS (data not shown). The broad pH optimum of PCY1 was centered between 8.5 and 9.0 (Fig. S3). Optimal production of segetalin A was dependent on the presence of DTT (see Fig. S4). On the other hand the production of *linear* segetalin A was significant in the absence of DTT. The estimated turnover number with presegetalin A1[14,32] is notably low at $\sim 1 \text{ h}^{-1}$.

Substrate specificity of PCY1 – A variety of synthetic linear peptides were tested as substrates for PCY1. First, putative native peptide cyclase substrates from *S. vaccaria* and *D. caryophyllus* (carnation) were tested (Table 1). Evidence for the expected CP formation was found for the putative A-class precursors for segetalins B, D, G, H and L, the F-class precursors of segetalins F and J and for the putative *D. caryophyllus* precursor of *cyclo*[GPIPFYG]. No evidence for CP production from the putative *D. caryophyllus* precursor

GYKDCCVQAKDLENAAVPV (peptide #11) was found. Thus, PCY1 appears to be able to catalyze the production of a number of native *S. vaccaria* CP. On the other hand, the data also suggest important differences among peptide cyclases within the Caryophyllaceae. Interestingly, a P20Q substitution in presegetalin D1[14,31] led to formation of an alternate truncated product, suggesting that position 20 may be important in determining the amino acid positions to be included in the CP product.

Variants of presegetalin A1[14,32] differing in both the C-terminal region and that of the incipient cyclic peptide were tested as substrates with recombinant PCY1. Variants representing C-terminal deletions of 1, 2, 4, 8, 12 and 13 (linear segetalin A) amino acids did not give rise to detectable levels of segetalin A or linear segetalin A in the presence of PCY1 (see Table S1). This suggests that the majority of the C-terminal amino acid sequence of presegetalin A1[14,32] is important for segetalin A production. This was investigated further in what is essentially an alanine scanning study of the substrate specificity of PCY1 with respect to the C-terminus (substitution with V was made for two positions which are A in the wild type substrate; see Table S1 and Fig. 7). All except three of the variants show a very significant drop in production of segetalin A when assayed with PCY1. On the other hand, a conservative substitution of V with A at position 25 resulted in a more than two-fold enhancement of the production of both cyclic and linear segetalin A. The substitutions E26A and S29A both had relatively little effect on both cyclic and linear product formation. With these results in hand, deletion of the amino acids at positions 25-26 in presegetalin A1[14,32] (peptide #18 in Table 1) was tested and found to allow the formation of segetalin A at a rate ~ 170 -fold lower than for the wild type substrate.

Alanine scanning was also performed for the sequence of the incipient cyclic peptide in presegetalin A1[14,32] (the A19 was substituted with V). Given low turnover number of PCY1 and the difficulties in obtaining multiple cyclic peptide standards, the qualitative ability of PCY1 to form linear and cyclic products was simply scored based on the presence of relevant mass spectrometric signals (see Table S2). In all but the case of the A19V substitution, it was possible to

detect such signals suggesting that such substitutions are tolerated, at least qualitatively, in the incipient CP sequence.

Given the pharmaceutical (44) and nanotechnological (45,46) applications for cyclic peptides, it was of interest to further test the range of possible CP which could be produced by PCY1. Two presegetalin A1[14,32] variants representing insertions of A and AAA in the incipient CP sequence were tested (see Table 1). Both of these variants were found to act as substrates. Also, a “D-amino acid scan” of the segetalin A sequence was performed by testing variant presegetalin A1[14,32]’s with single D-amino acid substitutions. These were tolerated by the PCY1 at all chiral positions except for the C-terminal Ala (Table 1). Given this, the single and double substitutions were made in presegetalin A1[14,32] in order to test the production of *cyclo*[Gly-Val-Pro-Val-D-Ala-Ala], *cyclo*[Gly-Val-D-Pro-Val-D-Ala-Ala], and *cyclo*[Gly-Val-D-Ala-Val-D-Ala-Ala]. Indeed, all three products were detectable by MS of enzyme assay samples (Table 1). This is particularly significant in that even numbered CP with alternating D-amino acids are known to form nanotubes (45).

DISCUSSION

Prior to 2011, very little was known about the biosynthesis of Caryophyllaceae-type CP. There had been a preliminary report that the linear form of heterophyllin B could undergo cyclization in the presence of plant extracts (47), however this does not appear to have been followed up. Thus it was still unclear whether ribosomes were involved in Caryophyllaceae-type CP. In a previous report (17), we provided evidence for the existence in the Caryophyllaceae and Rutaceae of genes encoding CP precursors which are formed by translation on ribosomes and processed to shorter CP. In this paper, we have obtained biochemical evidence allowing the elaboration of the process by which CP precursors are converted to CP in *S. vaccaria*. This appears to be a two step process involving two separate enzymes. The first enzyme, which we have called OLP1, is essentially an oligopeptidase which recognizes and cleaves presegetalin A1 at the KPQ-GVP sequence corresponding to the junction between the N-terminal region and the incipient CP sequence. Further work is required to clone and characterize OLP1.

The second enzyme involved in CP precursor processing in the *S. vaccaria* is PCY1, which we have cloned and characterized. Briefly, PCY1 is a serine protease-like enzyme which recognizes the highly conserved C-terminal region of a number of CP intermediates and catalyzes the excision and cyclization of the incipient CP sequence.

It is interesting, particularly from an evolutionary point of view, to compare and contrast CP biosynthesis in the Caryophyllaceae, with other organisms. Within plants, the biosynthesis of the other three classes of CP, the kalata, PawS and cyclic knottin types, show some common features. All appear to be biosynthesized from ribosome-derived precursors which differ greatly in arrangement and sequence. In many cases where evidence is available ribosome-derived biosynthesis of CP in plants, the general pattern appears to be cleavage at the N-terminus of the incipient CP sequence follow by cleavage and cyclization (2,20,48). For the kalata, PawS and cyclic knottins types, the precursor sequences generally share certain features including Gly and Asx at the N- and C-termini of the incipient CP sequence (48). Thus the evidence suggests an asparaginyl endopeptidase is responsible for the final cyclization step for these CP types. In *S. vaccaria*, there is also a tendency for Gly at the N-terminus of the incipient CP sequence. This appears to extend to carnation (17). However, the so-called F class precursors in *S. vaccaria* do not follow this rule, having the rather bulky Phe at that position. In flax (*Linum usitatissimum*), recently reported multidomain precursors of Caryophyllaceae-type CP show a strong tendency for Met at the N-terminus of the incipient CP sequence (49). At the C-terminal position of incipient CP sequences of *S. vaccaria*, *Citrus* spp. and *L. usitatissimum* (17,49), the amino acid is not strictly conserved. Consequently, asparaginyl endopeptidases are not implicated in the processing of Caryophyllaceae-type CP precursors.

Outside of plants, both ribosome-dependent and ribosome-independent modes of CP biosynthesis are prevalent (5,26). In mammals, cyclic defensins appear to be produced by the ligation of two linear peptides (8). In cyanobacteria and fungi there are examples of CP produced from ribosome derived precursors.

Among the fungi, *Amanita* mushrooms produce precursors to amanita toxins. These have a conserved Pro upstream of the C-terminus of the incipient CP and at the N-terminus of it. Although a prolyl oligopeptidase has been implicated in the processing of the precursors, the mechanism of cyclization remains unclear (26). For cyanobacteria, the biosynthesis of patellamides and related CP have been well-studied (19,50). Indeed, the biosynthetic pathway seems to be very similar to the biosynthesis of segetalins in *S. vaccaria*. Precursors containing one or more CP sequences are flanked by conserved regions. A protease cleaves at the “N-terminus” of the incipient CP sequences and a separate protease-like enzyme cyclizes the peptide with removal of C-terminal sequence. For *Prochloron* spp., the latter enzyme is encoded by *patG* (51) and is notably in the S8 family of serine proteases and not closely related to PCY1. It is notable that the products of *patG* and its homologues and that of *Pcy1* are the only cloned enzymes whose main function appears to be peptide cyclization. Thus, among a wide range of taxa there are some commonalities in the mode of biosynthesis of ribosome-derived CP. For systems in which the pathway is reasonably well-characterized, it appears to be common for a precursor with N- and C-terminal flanking regions to be first processed to remove the N-terminus. This is followed by a combined cleavage and cyclization reaction catalyzed by a acyl-intermediate-forming (cysteine or serine type) protease homologue. This is consistent with the known transpeptidation reactions which cysteine and serine proteases are capable of catalyzing (37,38). This certainly suggests a pattern of convergent evolution in which different proteases were recruited in different lineages, giving rise to diverse CP. Further work is required to determine whether the evolution of the cyclases in question was minimal relative to the ancestral proteases, or whether there was significant specialization involved.

The comparison of PCY1 with porcine muscle POP offers some insights into the structure and function of PCY1. Porcine muscle POP is a serine oligopeptidase with an unusual β -propeller domain. This domain is thought to limit access of larger polypeptides to the active site and this is consistent with porcine muscle POP acting on oligopeptide substrates. Indeed, an engineered

disulfide “latch” on the β -propeller lid abolishes even oligopeptidase activity (52). This is consistent with the notion that PCY1 also acts on oligopeptides as shown in this paper, although further work is required to determine the limits of its substrate size. The close agreement of the PCY1 with the porcine muscle POP structure suggests that the structural differences which affect the function of the two enzymes may be quite subtle. On the other hand, the amino acid insertion/deletion near the active site histidine may prove to be important. Clarification of this will likely require a high resolution experimental structure to understand. This is in contrast with the cyanobacterial *patG* homologues whose products have distinctive “capping helices” (reminiscent of the β -propeller lid), which are not found in related proteases (53,54).

As discussed above, the identity of PCY1 as a serine protease homologue is relevant to the mechanism of the cyclization reaction. Serine proteases are known to catalyze proteolysis in a multistep reaction which includes the early formation of an acylserine intermediate formed by nucleophilic attack of the serine at the carbonyl of the scissile peptide bond, with release of the C-terminus of the substrate. This is typically followed by hydrolysis of the acyl intermediate, freeing the N-terminus of the substrate and regenerating the unmodified form of the enzyme. However, in some enzymes or under some conditions aminolysis of the acylserine intermediate can compete with hydrolysis. Where the amine in question is at the N-terminus of a separate peptide, transpeptidation results. If the amine is at the distal end of the acyl group of the intermediate, then cyclization results. Furthermore, Berkers *et al.* (38) point out that “if protein complexation or conformation positions the amine-donating component involved in transpeptidation in close proximity to the acyl ester component, this will facilitate its nucleophilic attack on the acyl-enzyme intermediate and hence favor aminolysis”. Given this last point it is difficult to imagine how this might occur for segetalin formation. The acyl putative intermediates expected for the reaction represented in Table 1 are short, but quite variable in sequence and length. On the other hand, the C-terminus of the PCY1 substrates is quite highly conserved and amino acid substitutions can have a

strong negative effect on cyclization. In principle, the C-terminus of the substrate could be released from the enzyme prior to cyclization, the data presented suggest the possibility that the C-terminus remains bound to the enzyme even after acyl-intermediate formation and somehow directs the aminolysis (cyclization). Further work is required to investigate this. It is notable, however, that Agarwal et al. (54), propose a similar retention of the C-terminal fragment for *patG* homologues.

While the cloning and characterization of PCY1 offers insights into the biosynthesis of CP in plants, the enzyme may also have applications in biotechnology. There is considerable interest in cyclic peptide drugs and vaccines (for example see (55,56)). Biotechnological production of CP allows both the generation of CP libraries for screening as well as the production of individual CP for testing or production. The relative relaxed substrate specificity of PCY1 may be useful in this regard and further investigation of the applications of PCY1 and related enzymes is warranted.

REFERENCES

1. Kaufmann, H. P. and Tobschirbel, A. (1959) Uber ein Oligopeptid aus Leinsamen. *Chem. Ber.* **92**, 2805-2809
2. Cascales, L. and Craik, D. J. (2010) Naturally occurring circular proteins: distribution, biosynthesis and evolution. *Org. Biomol. Chem.* **8**, 5035-5047
3. Conlan, B. F., Gillon, A. D., Craik, D. J., and Anderson, M. A. (2010) Circular proteins and mechanisms of cyclization. *Biopolymers* **94**, 573-583
4. Craik, D. J., Daly, N. L., Mulvenna, J., Plan, M. R., and Trabi, M. (2004) Discovery, structure and biological activities of the cyclotides. *Curr. Protein Pept. Sci* **5**, 297-315
5. Jones, A. C., Gu, L., Sorrels, C. M., Sherman, D. H., and Gerwick, W. H. (2009) New tricks from ancient algae: natural products biosynthesis in marine cyanobacteria. *Curr. Opin. Chem. Biol.* **13**, 216-223
6. Picur, B., Cebrat, M., Zabrocki, J., and Siemion, I. Z. (2006) Cyclopeptides of *Linum usitatissimum*. *J. Pept. Sci* **12**, 569-574
7. Pomilio, A. B., Battista, M. E., and Vitale, A. A. (2006) Naturally-occurring cyclopeptides: structures and bioactivity. *Curr. Org. Chem.* **10**, 2075-2121
8. Selsted, M. E. (2004) Theta-defensins: cyclic antimicrobial peptides produced by binary ligation of truncated alpha-defensins. *Curr. Protein Pept. Sci* **5**, 365-371
9. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Goransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R. D., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108-160
10. Morita, H. and Takeya, K. (2010) Bioactive cyclic peptides from higher plants. *Heterocycles* **80**, 739-764
11. Morita, H., Eda, M., Iizuka, T., Hirasawa, Y., Sekiguchi, M., Yun, Y. S., Itokawa, H., and Takeya, K. (2006) Structure of a new cyclic nonapeptide, segetalin F, and vasorelaxant activity of segetalins from *Vaccaria segetalis*. *Bioorg. Med. Chem. Lett.* **16**, 4458-4461
12. Namjoshi, S. and Benson, H. A. (2010) Cyclic peptides as potential therapeutic agents for skin disorders. *Biopolymers*
13. Sancheti, H. and Camarero, J. A. (2009) "Splicing up" drug discovery. Cell-based expression and screening of genetically-encoded libraries of backbone-cyclized polypeptides. *Adv. Drug Deliv. Rev.* **61**, 908-917
14. White, C. J. and Yudin, A. K. (2011) Contemporary strategies for peptide macrocyclization. *Nat. Chem.* **3**, 509-524
15. Meier, J. L. and Burkart, M. D. (2009) The chemical biology of modular biosynthetic enzymes. *Chem. Soc Rev.* **38**, 2012-2045
16. Walsh, C. T., Chen, H., Keating, T. A., Hubbard, B. K., Losey, H. C., Luo, L., Marshall, C. G., Miller, D. A., and Patel, H. M. (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Chem. Biol.* **5**, 525-534
17. Condie, J. A., Nowak, G., Reed, D. W., Balsevich, J. J., Reaney, M. J., Arnison, P. G., and Covello, P. S. (2011) The biosynthesis of Caryophyllaceae-like cyclic peptides in *Saponaria vaccaria* L. from DNA-encoded precursors. *Plant J.* **67**, 682-690

18. Poth, A. G., Mylne, J. S., Grassl, J., Lyons, R. E., Millar, A. H., Colgrave, M. L., and Craik, D. J. (2012) Cyclotides associate with leaf vasculature and are the products of a novel precursor in petunia (Solanaceae). *J. Biol. Chem.* **287**, 27033-27046
19. Katoh, T., Goto, Y., Reza, M. S., and Suga, H. (2011) Ribosomal synthesis of backbone macrocyclic peptides. *Chem. Commun. (Camb.)* **47**, 9946-9958
20. Mylne, J. S., Colgrave, M. L., Daly, N. L., Chanson, A. H., Elliott, A. G., McCallum, E. J., Jones, A., and Craik, D. J. (2011) Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nat. Chem. Biol.* **7**, 257-259
21. Poth, A. G., Colgrave, M. L., Lyons, R. E., Daly, N. L., and Craik, D. J. (2011) Discovery of an unusual biosynthetic origin for circular proteins in legumes. *Proc. Natl. Acad. Sci. U. S. A* **108**, 10127-10132
22. Poth, A. G., Colgrave, M. L., Philip, R., Kerenga, B., Daly, N. L., Anderson, M. A., and Craik, D. J. (2011) Discovery of cyclotides in the fabaceae plant family provides new insights into the cyclization, evolution, and distribution of circular proteins. *ACS Chem. Biol.* **6**, 345-355
23. Saska, I., Gillon, A. D., Hatsugai, N., Dietzgen, R. G., Hara-Nishimura, I., Anderson, M. A., and Craik, D. J. (2007) An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclization. *J. Biol. Chem.* **282**, 29721-29728
24. Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G., and Ravel, J. (2005) Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc. Natl. Acad. Sci. U. S. A* **102**, 7315-7320
25. Jennings, C., West, J., Waine, C., Craik, D., and Anderson, M. (2001) Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis*. *Proc. Natl. Acad. Sci. U. S. A* **98**, 10614-10619
26. Walton, J. D., Hallen-Adams, H. E., and Luo, H. (2010) Ribosomal biosynthesis of the cyclic peptide toxins of *Amanita* mushrooms. *Biopolymers* **94**, 659-664
27. Tan, N. H. and Zhou, J. (2006) Plant cyclopeptides. *Chem. Rev.* **106**, 840-895
28. Morita, H., Yun, Y. S., Takeya, K., and Itokawa, H. (1994) Segetalin A, a new cyclic peptide from *Vaccaria segetalis*. *Tetrahedron Lett.* **51**, 9593-9596
29. Gambino, G., Perrone, I., and Gribaudo, I. (2008) A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochem Anal.* **19**, 520-525
30. Gaasterland, T. and Sensen, C. W. (1996) Fully automated genome analysis that reflects user needs and preferences. A detailed introduction to the MAGPIE system architecture. *Biochimie* **78**, 302-310
31. Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Muller, W. E., Wetter, T., and Suhai, S. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **14**, 1147-1159
32. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663-670
33. Meesapyodsuk, D., Balsevich, J., Reed, D. W., and Covello, P. S. (2007) Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding β -amyrin synthase and a triterpene carboxylic acid glucosyltransferase. *Plant Physiol* **143**, 959-969
34. Sheoran, I. S., Olson, D. J., Ross, A. R., and Sawhney, V. K. (2005) Proteome analysis of embryo and endosperm from germinating tomato seeds. *Proteomics*. **5**, 3752-3764
35. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207-234
36. Meesapyodsuk, D., Reed, D. W., Covello, P. S., and Qiu, X. (2007) Primary structure, regioselectivity, and evolution of the membrane-bound fatty acid desaturases of *Claviceps purpurea*. *J. Biol. Chem.* **282**, 20191-20199
37. Johnston, R. B., Mycek, M. J., and Fruton, J. S. (1950) Catalysis of transpeptidation reactions by chymotrypsin. *J. Biol. Chem* **187**, 205-211

38. Berkers, C. R., de Jong, A., Ovaa, H., and Rodenko, B. (2009) Transpeptidation and reverse proteolysis and their consequences for immunity. *Int. J. Biochem. Cell Biol.* **41**, 66-71
39. Marx, U. C., Korsinczky, M. L., Schirra, H. J., Jones, A., Condie, B., Otvos, L., Jr., and Craik, D. J. (2003) Enzymatic cyclization of a potent Bowman-Birk protease inhibitor, sunflower trypsin inhibitor-1, and solution structure of an acyclic precursor peptide. *J. Biol. Chem.* **278**, 21782-21789
40. Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H., and Shaw, J. F. (2004) GDSL family of serine esterases/lipases. *Prog. Lipid Res.* **43**, 534-552
41. Fulop, V., Bocskei, Z., and Polgar, L. (1998) Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. *Cell* **94**, 161-170
42. Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815
43. Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolova, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., and . (1992) The alpha/beta hydrolase fold. *Protein Eng* **5**, 197-211
44. Garcia, A. E. and Camarero, J. A. (2010) Biological activities of natural and engineered cyclotides, a novel molecular scaffold for peptide-based therapeutics. *Curr. Mol. Pharmacol.* **3**, 153-163
45. Ghadiri, M. R., Granja, J. R., Milligan, R. A., McRee, D. E., and Khazanovich, N. (1993) Self-assembling organic nanotubes based on a cyclic peptide architecture. *nat* **366**, 324-327
46. Hourani, R., Zhang, C., van der, W. R., Ruiz, L., Li, C., Ketten, S., Helms, B. A., and Xu, T. (2011) Processable cyclic peptide nanotubes with tunable interiors. *J. Am. Chem. Soc.* **133**, 15296-15299
47. Jia, A., Li, X., Tan, N., Liu, X., Shen, Y., and Zhou, J. (2006) Enzymatic cyclization of linear peptide to plant cyclopeptide heterophyllin B. *Science in China: Series B Chemistry* **49**, 63-66
48. Mylne, J. S., Chan, L. Y., Chanson, A. H., Daly, N. L., Schaefer, H., Bailey, T. L., Nguyencong, P., Cascales, L., and Craik, D. J. (2012) Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis. *Plant Cell* **24**, 2765-2778
49. Gui, B., Shim, Y. Y., Datla, R. S., Covello, P. S., Stone, S. L., and Reaney, M. J. (2012) Identification and quantification of cyclolinopeptides in five flaxseed cultivars. *J. Agric. Food Chem.* **60**, 8571-8579
50. Sivonen, K., Leikoski, N., Fewer, D. P., and Jokela, J. (2010) Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Appl. Microbiol. Biotechnol.* **86**, 1213-1225
51. Lee, J., McIntosh, J., Hathaway, B. J., and Schmidt, E. W. (2009) Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *J. Am. Chem. Soc.* **131**, 2122-2124
52. Szeltner, Z., Rea, D., Juhasz, T., Renner, V., Fulop, V., and Polgar, L. (2004) Concerted structural changes in the peptidase and the propeller domains of prolyl oligopeptidase are required for substrate binding. *J. Mol. Biol.* **340**, 627-637
53. Koehnke, J., Bent, A., Houssen, W. E., Zollman, D., Morawitz, F., Shirran, S., Vendome, J., Nneoyiegbe, A. F., Trembleau, L., Botting, C. H., Smith, M. C., Jaspars, M., and Naismith, J. H. (2012) The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nat. Struct. Mol. Biol.* **19**, 767-772
54. Agarwal, V., Pierce, E., McIntosh, J., Schmidt, E. W., and Nair, S. K. (2012) Structures of cyanobactin maturation enzymes define a family of transamidating proteases. *Chem Biol.* **19**, 1411-1422
55. Wong, C. T., Rowlands, D. K., Wong, C. H., Lo, T. W., Nguyen, G. K., Li, H. Y., and Tam, J. P. (2012) Orally active peptidic bradykinin B1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment. *Angew. Chem. Int. Ed Engl.* **51**, 5620-5624
56. Hoogerhout, P., Kamphuis, W., Brugghe, H. F., Sluijs, J. A., Timmermans, H. A., Westdijk, J., Zomer, G., Boog, C. J., Hol, E. M., and van den Dobbelsteen, G. P. (2011) A cyclic undecamer peptide mimics a turn in folded Alzheimer amyloid beta and elicits antibodies against oligomeric and fibrillar amyloid and plaques. *PLoS. One.* **6**, e19110

Acknowledgements—The authors are grateful to Simon Foote and Doug Olson for technical support, to Sean Hemmingsen, Michele Loewen and Jon Page for reviewing the manuscript, useful discussions with Michele Loewen and Randy Purves, and support of members of the PhytoMetaSyn project as funded by Genome Canada and affiliates, McGill Innovation Centre and the Saskatchewan Agriculture Development Fund.

FOOTNOTES

¹To whom correspondence may be addressed: Patrick Covello, National Research Council of Canada, 110 Gymnasium Place, Saskatoon, SK Canada S7N 0W9; Tel.: (306) 975-5269; Fax: (306) 975-4839; E-mail: patrick.covello@nrc-cnrc.gc.ca

²The abbreviations used are: CP, cyclic peptide; PCY1, Peptide Cyclase 1; OLP1, Oligopeptidase 1; EST, expressed sequence tag; POP, prolyl oligopeptidase

FIGURE LEGENDS

FIGURE 1. Proposed pathway for the processing of presegetalin A1 to segetalin A in *S. vaccaria*. Presegetalin A1 is derived from a monocistronic mRNA by translation on ribosomes (17). An oligopeptidase activity (OLP1) and a peptide cyclase activity (PCY1) are involved in removal of the N-terminus and cyclization, respectively.

FIGURE 2. Separation of Oligopeptidase 1 and Peptide Cyclase 1 activities in *S. vaccaria* extracts. Panel a shows a Mono-Q ion exchange chromatogram of stage 2 developing *S. vaccaria* seed proteins using a NaCl gradient. Two minute fractions were collected and assayed for OLP1 activity (b; presegetalin A1[14,32] formation) and PCY1 activity (c; segetalin A formation from presegetalin A1[14,32]; see Experimental Procedures).

FIGURE 3. Products of Oligopeptidase 1 activity in developing seeds of *S. vaccaria*. An ion exchange fraction derived from stage 2 developing seeds of *S. vaccaria* and eluting at ~0.35 M NaCl was assayed using presegetalin A1 (see Experimental Procedures). MALDI-TOF MS analysis provided evidence of the residual substrate as well as its cleavage products, as indicated.

FIGURE 4. Assay of OLP1 and PCY1 activity by ion trap LC/MS. Ion chromatograms are shown for OLP1 assays using presegetalin A1 as a substrate without (a) and with (b) partially purified OLP1, and assays using presegetalin A1[14,32] as a substrate without enzyme (c), and with partially purified plant PCY1 fraction (d) and recombinant PCY1 (e). Products were not detected at significant levels in assays with no incubation time or no peptide substrates (data not shown). Chromatograms consist of the sum of molecular ions for presegetalin A1 ($m/z = 851.5$, $z=4$; $m/z = 681.5$, $z=5$), presegetalin A1[1,13] ($m/z = 718.8$, $z=2$; $m/z = 729.1$, monosodium adduct, $z=2$; $m/z = 740.0$, disodium adduct, $z=2$), presegetalin A1[14,32] ($m/z = 993.5$, $z=2$; $m/z = 662.7$, $z=3$), linear segetalin A ($m/z = 628.5$, $z=1$; $m/z = 650.5$, sodium adduct, $z=1$) and segetalin A ($m/z = 610.5$, $z=1$; $m/z = 632.5$, sodium adduct, $z=1$).

FIGURE 5. Electrophoretic analysis of partially purified peptide cyclase (PCY1) from *S. vaccaria*. Lane 1, crude filtrate from *S. vaccaria* developing seed; lane 2, active fraction from anion exchange chromatography; lane 3, active fraction from hydrophobic interaction chromatography; lane 4, active fraction from gel filtration chromatography. The mobility of relative molecular mass standards of 25,000 and 75,000 are shown on the left. The arrow indicates a band corresponding to a major protein with M_r of approximately 83,000.

FIGURE 6. The predicted structure of PCY1 based on homology modeling using porcine muscle POP. (a) Structural alignment of a homology model of PCY1 (yellow) and porcine muscle prolyl oligopeptidase (blue). The top half of the structure represents the catalytic domain including the putative catalytic triad shown in stick form. The bottom half of the structure is a β -propeller domain which may be involved in controlling access to the active site. (b) Superimposition of the putative catalytic triad (S562, D653, H695) of PCY1 (green) and that of porcine muscle prolyl oligopeptidase (red).

FIGURE 7. Effects of mutations in the C-terminal flanking region of presegetalin A1 on production of linear and cyclic segetalin A production by recombinant PCY1. Wild type (WT) and variants of presegetalin A1[14,32] are indicated. Means and standard deviations (n=3) of 1 h assays are indicated.

TABLE

Table 1. Substrate specificity of PCY1. The synthetic linear peptides indicated were used in 1 h assays with PCY1 (see Experimental Procedures). CP and linear peptide production was determined by ion trap LC/MS and confirmed by ion trap LC/MS/MS (see Table S2). Expected CP sequences are underlined. Variant amino acids are indicated in reverse type. *D. caryophyllus* precursor sequences are derived from Genbank accession nos. AW697819 and CF259478 (17). Quantitation of segetalin A and linear segetalin A production from selected substrates is shown in Fig. 7. Lower case letters denote D-amino acids.

Caryophyllaceae cyclic peptide biosynthesis

Substrate Category	Peptide No.	Peptide Name	Sequence	PCY1 Product	
				Cyclic	Linear
native plant substrates					
	1	Presegetalin A1[14,32]	GVPVWA FQAKDVENASAPV	+	+
	2	Presegetalin B1[14,31]	GVAWA FQAKDVENASAPV	+	-
	3	Presegetalin D1[14,31]	GLSFAFP AKDAENASSPV	+	+
	4	Presegetalin G1[14,31]	GVKYA FQPKDSENASAPV	+	-
	5	Presegetalin H1[14,31]	GYRFS FQAKDAENASAPV	+	-
	6	Presegetalin K1[14,31]	GRVKA FQAKDAENASAPV	+	-
	7	Presegetalin L1[14,32]	GLPGWP FQAKDVENASAPV	+	-
	8	Presegetalin F1[14,38]	FSASYSSKP IQTQVSNMGMDNASAPV	+	-
	9	Presegetalin J1[14,36]	FGTHGLPAP IQVPNGMDDACAPM	+	-
	10	<i>D. caryophyllus</i> Precursor 1	GPIPFYG FQAKDAENASVPV	+	-
	11	<i>D. caryophyllus</i> Precursor 2	GYKDCC VQAKDLENAAVPV	-	-
C-terminal presegetalin A1[14,32] variants					
	12	Presegetalin A1[14,31]	GVPVWA FQAKDVENASAP	-	-
	13	Presegetalin A1[14,30]	GVPVWA FQAKDVENASA	-	-
	14	Presegetalin A1[14,28]	GVPVWA FQAKDVENA	-	-
	15	Presegetalin A1[14,24]	GVPVWA FQAKD	-	-
	16	Presegetalin A1[14,20]	GVPVWA F	-	-
	17	Presegetalin A1[14,19]	GVPVWA	-	n/a
	18	Presegetalin A1[14,32]Δ25-26	GVPVWA FQAKD NASAPV	+	+
	19	Presegetalin A1[14,32] F20A	GVPVWA A QAKDVENASAPV	+	+
	20	Presegetalin A1[14,32] Q21A	GVPVWA F AAKDVENASAPV	+	+
	21	Presegetalin A1[14,32] A22V	GVPVWA F QV K DVENASAPV	+	+
	22	Presegetalin A1[14,32] K23A	GVPVWA FQA A DVENASAPV	+	+

Caryophyllaceae cyclic peptide biosynthesis

	23	Presegetalin A1[14,32] D24A	GVPVWA	FQAKA V ENASAPV	+	+
	24	Presegetalin A1[14,32] V25A	GVPVWA	FQAKD A ENASAPV	+	+
	25	Presegetalin A1[14,32] E26A	GVPVWA	FQAKD V ANASAPV	+	+
	26	Presegetalin A1[14,32] N27A	GVPVWA	FQAKD V E A ASAPV	+	+
	27	Presegetalin A1[14,32] A28V	GVPVWA	FQAKD V EN V SAPV	+	+
	28	Presegetalin A1[14,32] S29A	GVPVWA	FQAKD V EN A AAPV	+	+
	29	Presegetalin A1[14,32] A30V	GVPVWA	FQAKD V EN A S V PV	+	+
	30	Presegetalin A1[14,32] P31A	GVPVWA	FQAKD V EN A S A AV	+	+
	31	Presegetalin A1[14,32] V32A	GVPVWA	FQAKD V EN A S A PA	+	+
	32	Presegetalin D1[14,31] P20Q	GLSFA	F Q AKDAENASSPV	+	+
Variants of incipient segetalin sequences						
	33	Presegetalin A1[14,32] G14A	A VPVWA	FQAKDVENASAPV	+	+
	34	Presegetalin A1[14,32] V15A	G APVWA	FQAKDVENASAPV	+	+
	35	Presegetalin A1[14,32] P16A	G V A VWA	FQAKDVENASAPV	+	+
	36	Presegetalin A1[14,32] V17A	G V P A WA	FQAKDVENASAPV	+	+
	37	Presegetalin A1[14,32] W18A	G V P V A A	FQAKDVENASAPV	+	-
	38	Presegetalin A1[14,32] A19V	G V P V W V	FQAKDVENASAPV	-	-
	39	Presegetalin A1[14,32] V15v	G v VPVWA	FQAKDVENASAPV	+	+
	40	Presegetalin A1[14,32] P16p	G V p VPVWA	FQAKDVENASAPV	+	-
	41	Presegetalin A1[14,32] V17v	G V P v WA	FQAKDVENASAPV	+	-
	42	Presegetalin A1[14,32] W18w	G V P V w A	FQAKDVENASAPV	+	+
	43	Presegetalin A1[14,32] A19a	G V P V w a	FQAKDVENASAPV	-	-
	44	Presegetalin A1 [14,32] ins 16A17	G V P A VWA	FQAKDVENASAPV	+	+
	45	Presegetalin A1 [14,33] ins 16AAA17	G V P AAA VWA	FQAKDVENASAPV	+	+

Figure 1

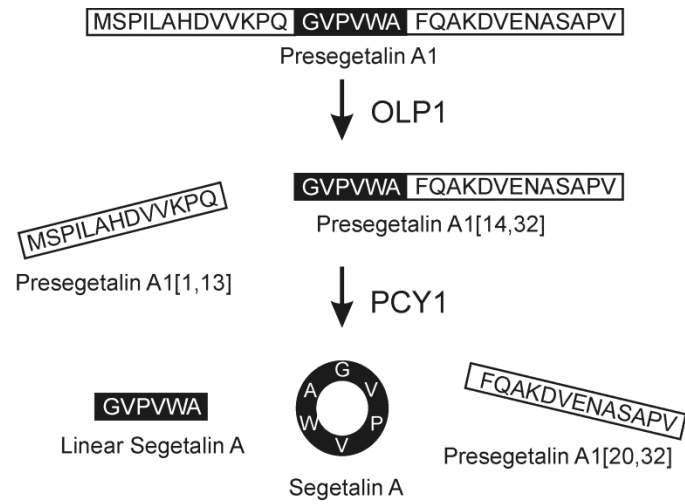


Figure 2

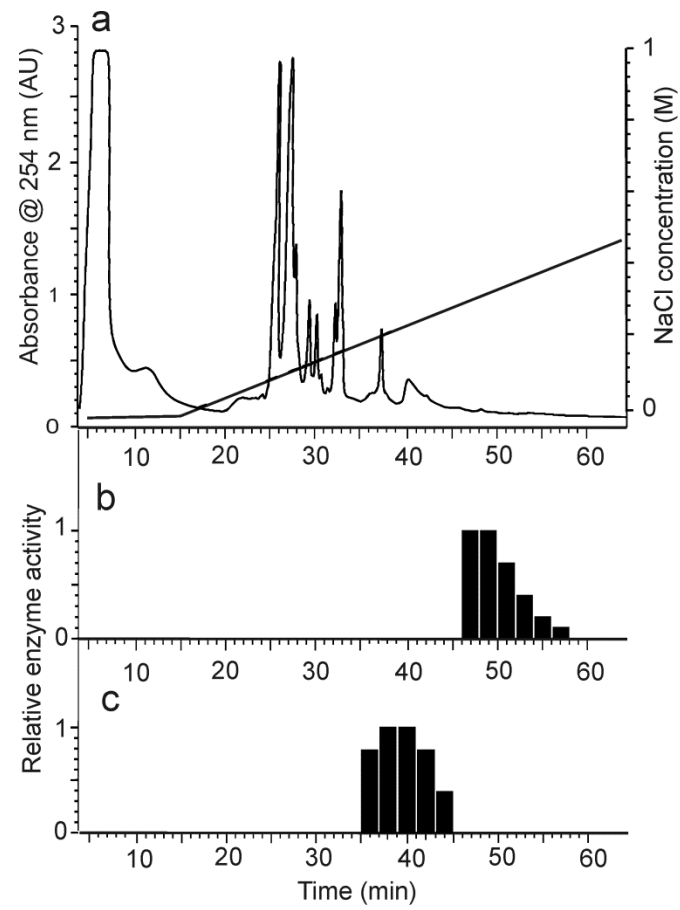


Figure 3

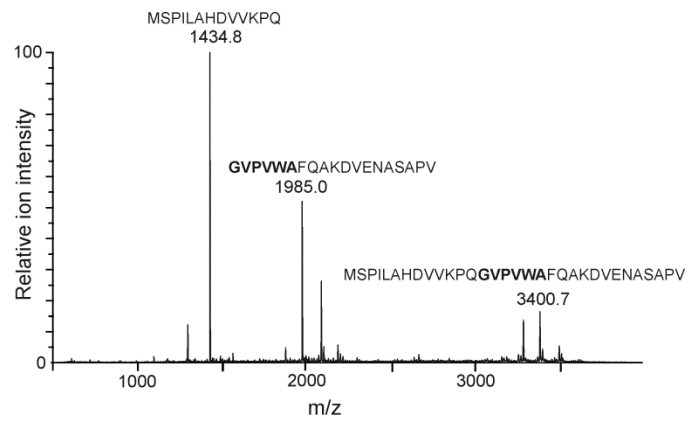


Figure 4

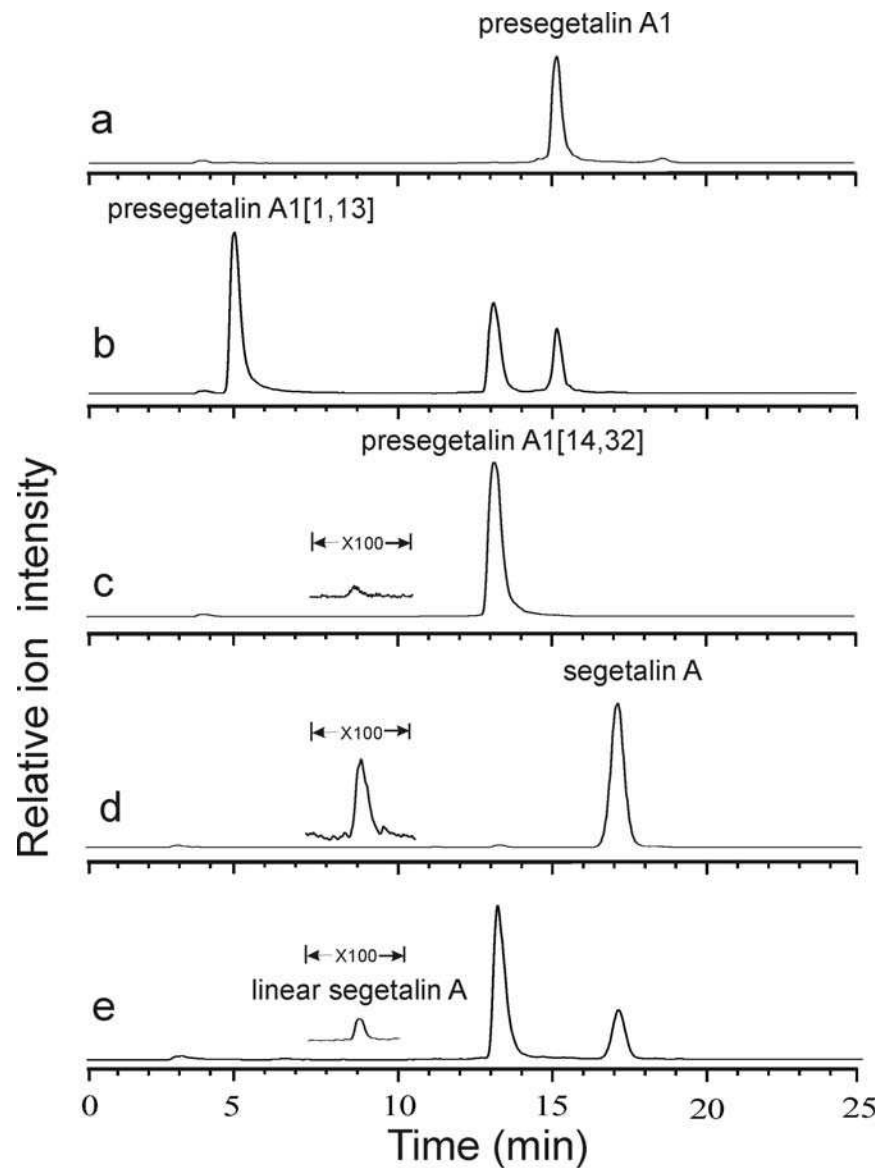


Figure 5

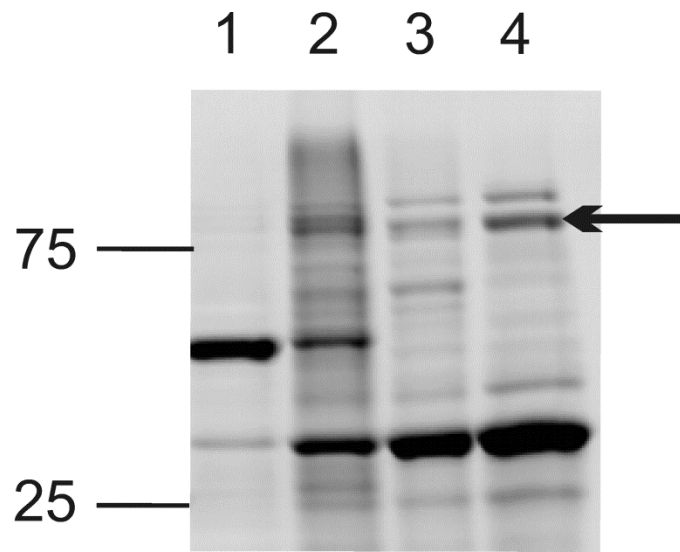


Figure 6

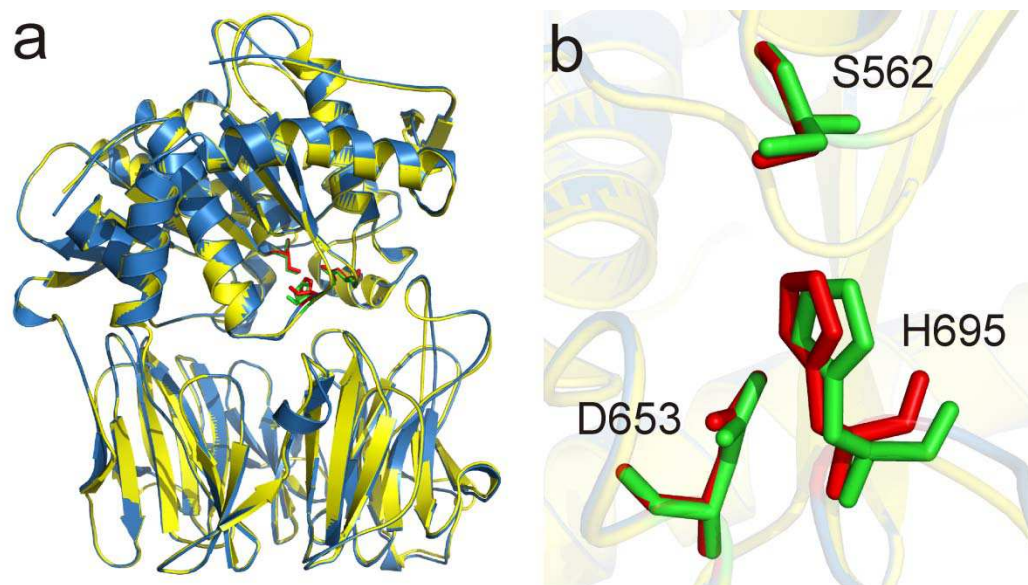


Figure 7

