The biosynthesis of Caryophyllaceae-like cyclic peptides in Saponaria vaccaria L. from DNA-encoded precursors

Condie, Janet A.; Nowak, Goska; Reed, Darwin W.; Balsevich, J. John; Reaney, Martin J. T.; Arnison, Paul G.; Covello, Patrick S.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1111/j.1365-313X.2011.04626.x
The Plant Journal, 67, 4, pp. 682-690, 2011-06-10

SUMMARY

Cyclic peptides (CPs) are produced in a very wide range of taxa. Their biosynthesis generally involves either non-ribosomal peptide synthases or ribosome-dependent production of precursor peptides. Plants within the Caryophyllaceae and certain other families produce CPs which generally consist of 5-9 proteinogenic amino acids. The biological roles for these CPs in the plant are not very clear, but many of them have activity in mammalian systems. There is currently very little known about the biosynthesis of CPs in the Caryophyllaceae. A collection of expressed sequence tags from developing seeds of *Saponaria vaccaria* was investigated for information about CP biosynthesis. This revealed genes that appeared to encode CP precursors which are subsequently cyclized to mature CPs. This was tested and confirmed by the expression of a cDNA encoding a putative precursor of the CP segetalin A in transformed *S. vaccaria* roots. Similarly, extracts of *S. vaccaria* developing seed were shown to catalyze the production of segetalin A from the same putative (synthetic) precursor. Sequence analysis indicates the presence of similar CP precursor genes in *Dianthus caryophyllus* and *Citrus* spp. The data support the ribosome-dependent biosynthesis of Caryophyllaceae-like CPs in the Caryophyllaceae and Rutaceae.

INTRODUCTION

Cyclic peptide occurrence and bioactivity

Cyclic peptides (CPs) represent a diverse class of natural products showing broad distribution in bacteria, fungi, plants and animals (Pomilio et al., 2006; Tan and Zhou, 2006; Craik et al., 2007; Morita and Takeya, 2010; Cascales and Craik, 2010). Many CPs

are of pharmaceutical interest. In their simplest form, CPs are cyclic polyamides formed from a small number of proteinogenic amino acids (*e.g.,* see Figure. 1). More complex CPs may be polycyclic and may include non-proteinogenic amino acids and other variations in structure.

CPs occur in 26 families of flowering plants and are particularly common in the Caryophyllaceae, Rhamnaceae and Violaceae (Picur et al., 2006; Tan and Zhou, 2006; Morita and Takeya, 2010; Cascales and Craik, 2010). Plant CPs include the so-called homocyclopeptides, having rings formed from peptide bonds. Two well-studied groups in this category are the cyclotides and the Caryophyllaceae-like CPs. Cyclotides, which appear to serve an anti-insect role, occur mainly in the Violaceae (Craik et al., 2007). They are typically 28-37 proteinogenic amino acids and are polycyclic by virtue of disulfide bonds which give rise to a knotted structure.

Caryophyllaceae-like CPs are homocyclopeptides with a single ring formed from 2 or between 5 and 12 α-amino acids (usually L isomers) (Tan and Zhou, 2006; Morita and Takeya, 2010). As expected from the name, Caryophyllaceae-like CPs are particularly common in the Caryophyllaceae family, but they are also found in 9 other families. In the 1990s, considerable efforts were made to characterize the CPs which occur in various members of the Caryophyllaceae (Ding et al., 1999; Tan and Zhou, 2006). One of the species investigated was *Saponaria vaccaria* (syn. *Vaccaria segetalis*, *Vaccaria hispanica;* cowcockle). Eight CPs, called segetalin A to segetalin H, were isolated and shown to be comprised of 5 to 9 proteinogenic amino acids. The sequences of these are shown in Figure 2A. A number of these were shown to have estrogen-like and/or vasorelaxant activity (Morita et al., 2006; Tan and Zhou, 2006). These

observations are in keeping with the use of *S. vaccaria* in traditional Chinese medicine as a treatment for ammenorhea, regulating blood flow, and increasing lactation (Itokawa et al., 1995). The biological role of segetalins in the plant *per se* is less clear, but may include antibiotic or anti-feedant activity.

Cyclic peptide biosynthesis

There are basically two biosynthetic routes involved in ordering the amino acids in CPs. In one case, the ribosomes are involved in the ordering of mRNA-encoded amino acids, initially, to form a linear peptide precursor (Schmidt et al., 2005). Alternatively, the amino acids are ordered without direct ribosome involvement, by non-ribosomal peptide synthetases (NRPSs; Finking and Marahiel, 2004). While, evidence for ribosome-dependent biosynthesis of cyclotides has been presented, relatively little is known about the biosynthesis of other classes of CPs in plants. Craik and coworkers have found cDNAs from *Oldenlandia affinis* encoding cyclotide precursors (Gillon et al., 2008). When a corresponding precursor gene is expressed in tobacco, mature cyclotides are produced. An asparaginyl protease in tobacco has been implicated in the biosynthetic process (Saska et al., 2007). Thus, cyclotides are biosynthesized from ribosome-derived linear precursors which are apparently trimmed and cyclized with the aid of at least one protease-like enzyme. In terms of Caryophyllaceae-like CPs, Jia and coworkers (Jia et al., 2006) have reported the cyclization of a linear version of heterophyllin B in an incompletely described extract from *Pseudostellaria heterophylla* Caryophyllaceae).

Within the Caryophyllaceae, *S. vaccaria* provides a suitable system with which to study CP biosynthesis. As indicated above, considerable phytochemical work has

elucidated a number of CPs present in *S. vaccaria* seeds. Also, the development of a relevant expressed sequence tag (EST) collection from *S. vaccaria* (Meesapyodsuk et al., 2007) provides a starting point for a molecular genetic investigation of the biosynthesis of Caryophyllaceae-like CPs. In this paper, we report the discovery of cDNAs encoding CP precursors (CPPs) and studies which confirm the role of the corresponding genes in CP biosynthesis.

Is there additional novelty?

RESULTS

Cyclic peptide precursor genes from *S. vaccaria*

In an effort to understand the biosynthesis of natural products in *S. vaccaria*, ~14,000 expressed sequence tags (ESTs) have been generated from cDNA libraries derived from roots and developing seed RNA (Meesapyodsuk et al., 2007). This led to the isolation of one of the first cDNAs encoding a triterpenoid glycosyltransferase, namely UGT74M1 which is involved in saponin biosynthesis. We investigated the *S. vaccaria* EST collection for information about CP biosynthesis. A search for ESTs which matched circular permutations of mature CP amino acid sequences revealed nucleotide sequences encoding short 30-40 amino acid peptides (see Fig. 2A). The ESTs in this group are highly abundant in the developing seed collection, comprising 14% of the total sequences. The corresponding peptide sequences showed highly conserved N- and C-terminal domains which flank the mature cyclic peptide sequences. These data suggest that cyclic peptides in *S. vaccaria* are biosynthesized ribosomally as linear precursors (presegetalins or CPPs) which are then processed to mature cyclic peptides. Thus, it

would appear that segetalin A is formed from the presegetalin A peptide encoded by a presegetalin A gene.

In order to understand the relationships among the CPP cDNA sequences, they were collected and reassembled into contigs. Putative presegetalin cDNAs sequences were first collected based on the presence of nucleotide sequences encoding mature cyclic peptide sequences. Added to this collection was an additional group of sequences which showed a high degree of similarity to members of the above collection. The collection was clustered with parameters which favoured the clustering of sequences encoding the same mature cyclic peptide sequences, but not sequences encoding other mature CP sequences (see Experimental procedures). Due to the large numbers of sequences involved and the possibility of sequencing errors, contigs smaller than 5 ESTs were ignored in the sequence analysis. In general, more than one cluster was obtained for each segetalin. For example, for segetalin D, four clusters were found to have distinct cDNA sequences, which encode three distinct amino acid sequences, all of which include the same circular permutation of the mature segetalin D amino acid sequence. This gave rise to the nomenclature detailed in Table 1. Thus, *sgd3b* is a gene corresponding to the second of 2 cDNAs with distinct nucleotide sequences which encodes the third (preSGD3) of 3 putative segetalin D precursors. PreSGD3 is presumed to give rise to segetalin D (SGD).

Interestingly, the sequence analysis revealed cDNAs which a) showed predicted amino acid sequence similarity to the putative precursors of known segetalins and b) appeared to

encode the precursors of novel segetalins which were named J, K and GLPGWP (see Table 1 and Figure 2A).

Based on the sequence analysis, there appear to be at least 15 *S. vaccaria* genes (or alleles) encoding 12 (precursor) amino acid sequences, which include the sequences of 6 known (mature, cyclic) segetalins and three putative segetalins. The known segetalins represented are A, B, D, F, G and H. This matches well with the segetalins which have been detected chemically in the Pink Beauty variety (A,B,D,F,G,H; J.J. Balsevich, unpublished). Segetalin I has also been found in Pink Beauty plants at low levels. In addition, there appear to be cDNAs in our *S. vaccaria* library which encode precursors of segetalins which have not previously been detected by chemical analysis. By comparison with the precursor sequences of the known segetalins, the unknown segetalins were predicted to have the sequences GRVKA, GLPGWP and FGTHGLPAP (Figure 2A and Table 1).

Based on sequence similarities, the putative precursor sequences can be divided into two classes which we call "A" and "F" (Figure 2A). The class A precursors are distinguished by an initial Gly in the mature CP sequence and a Phe immediately following the mature sequence. The only two class F precursors, presegetalin F and presegetalin J, have Phe and Ile in the respective positions. The flanking sequences differ somewhat in length and sequence between the two classes.

Quantitative RT-PCR analysis indicates that genes encoding certain presegetalins, namely A, H and J, are highly expressed in developing and mature seeds relative to other tissues tested (see Figure 3). This is consistent with the occurrence of segetalins A and H in *S. vaccaria* seed. The *glpgwp1* gene, whose putative CP product has not been reported, did not show high expression in developing seed.

Segetalin A is derived from presegetalin A1 *in vivo* and *in vitro*

To test the hypothesis that *S. vaccaria* cyclic peptides are synthesized from ribosomally-produced precursors, transformed root cultures of *S. vaccaria* (Schmidt et al., 2007) were generated which express presegetalin A. The variety White Beauty was used, since it was found not to produce segetalin A naturally (J.J. Balsevich, unpublished). Segetalin A production was assayed by LC/MS using single ion monitoring. The mass spectrum of segetalin A is shown in Figure 4. Based on LC/MS, hairy root lines which were not engineered to express presegetalin A did not contain detectable amounts of segetalin A (see Figure 5). On the other hand, multiple independent hairy root lines expressing presegetalin A were found to contain segetalin A in the range of 0.1-5 μg/g fresh weight (Figures 5 and 6). Thus, *S. vaccaria* roots appear to have all of the biochemical requirements for CPP processing to mature CPs.

In order to investigate the biochemical machinery involved in CPP processing, extracts of *S. vaccaria* developing seed were examined. Synthetic presegetalin A was incubated with extracts of developing seed and the assay samples were analyzed by LC/MS using single ion monitoring. As indicated in Fig. 7, presegetalin A was converted

to segetalin A. Thus, for presegetalin A, the biochemical machinery for processing to the mature cyclic segetalin is present in developing seeds of *S. vaccaria*.

A general role for ribosome-dependent CPP production in the biosynthesis of Caryophyllaceae-like CPs

Given the evidence for CP biosynthesis from ribosome-generated precursors, it was of interest to investigate related taxa. With this in mind, sequence similarity searches of the Genbank EST database were performed with the presegetalin A1 amino acid sequence. This revealed cDNA sequences of *Dianthus caryophyllus* (carnation) which showed a high degree of similarity (see Figure 2B). Based on sequence similarity and the information presented herein for *S. vaccaria*, it seems very reasonable to interpret the *D. caryophyllus* sequences as representing genes encoding CPPs which give rise to mature cyclic peptides with sequences GPIPFYG and GYKDCC. Thus, we would predict the presence of these CPs in *D. caryophyllus*. Incidentally, the authors are aware of only one report of cyclic peptides in *D. caryophyllus*, which provides evidence for the presence of caryophllusin A [*cyclo*-(GPYFT)] and delavayin B [*cyclo*-(GSIFFA)] (Morita et al., 1997; Li et al., 2008).

Expressed sequence tags from the genus *Citrus*, in the family Rutaceae, also provide evidence for genes encoding CPPs (see Figure 2C). Search of translated ESTs from Genbank for exact matches to the known cyclic peptides of *Citrus aurantium* [*cyclo-(*GLLPPFPG) and *cyclo*-(GLVLPS)] (Matsumoto et al., 2002) and *C. natsudaidai* [*cyclo*-(GYLLPPS)] (Morita et al., 2007), revealed sequences encoding putative CPPs. As with

the putative CPPs from the Caryophyllaceae, the flanking regions of the *Citrus* sequences showed a very high degree of similarity. Further searches revealed gene products from *C. sinensis* which were previously reported as DNA binding proteins for which expression was stress-induced (Mozoruk et al., 2008) (Figure 2C, last four sequences). It is notable that all of the *Citrus* CPP sequences have Gly as the first amino acid of the mature CP and Ser immediately following the mature sequence. This data is consistent with the ribosome-dependent biosynthesis of CPs via CPPs in the genus *Citrus*.

DISCUSSION

Previous phytochemical work has indicated that the so-called Caryophyllaceae-like cyclic peptides occur in, of course, the Caryophyllaceae, but also Rutaceae and Linaceae, for example. These CPs are homomonocyclopeptides generally derived from proteinogenic amino acids. To date, very little is known about their biosynthesis. Certainly, the amino acid composition is consistent with the involvement of translation on ribosomes. Other systems provide some insight into how the process of cyclic peptide formation might occur.

Regarding the biosynthesis of CPs in plants, the cyclotides of the Violaceae, Rubiaceae and Cucurbitaceae are the best understood. For certain cyclotides, cDNA sequences encoding precursors have been isolated. Evidence, particularly from transgenic tobacco studies, indicates that the corresponding mRNAs are translated on ribosomes (Saska et al., 2007; Gillon et al., 2008). The mature cyclotide domain is excised from the precursor and cyclized. This appears to involve an unidentified protease which cleaves at

the N-terminus of the mature domain and an asparaginyl protease which both removes the C-terminus of the precursor and results in cyclization to the mature cyclotide.

A well characterized example from cyanobacteria shows many parallels to cyclotide biosynthesis. Patellamides from the cyanobacterium genus *Prochloron*, are cyclic peptides which are derived from ribosomally-derived precursors (Jones et al., 2009; Lee et al., 2009). The protease PatA is involved in the cleavage of the precursor at the N-terminus of the amino acid sequence destined for the mature peptide. PatG is a protease-like enzyme which catalyzes a transamidation to give a cyclic product.

In the fungal genus *Amanita*, the bicyclic peptide *Amanita* toxins have recently been shown to be biosynthesized ribosomally (Walton et al., 2010). The resulting precursors are cleaved at two proline residues by a prolyl oligopeptidase to give linear intermediates. The mechanism of cyclization by peptide bond formation is unclear, however, it may be related to the formation of a Trp-Cys (tryptathionine) linkage.

The above examples provide precedents for the biosynthesis of cyclic peptides from ribosome-derived precursors through the action of protease-like enzymes. The data from *S. vaccaria* are certainly consistent with this scenario and many of the features of the plant cyclotide, bacterial and fungal systems. There is a high degree of conservation in the flanking regions of the presegetalin sequences. This may be important for the recognition of the processing enzymes. Further work is required to characterize the enzyme activity present in developing seed extracts to determine the number and nature of the enzymes involved in presegetalin processing. Given the similarities to the *D. caryophyllus* and *Citrus* systems, it is likely that this will give further general insight into the biosynthesis of Caryophyllaceae-like CPs.

EXPERIMENTAL PROCEDURES

Chemicals

Presegetalin A1 was chemically synthesized at the Sheldon Biotechnology Centre, McGill University (MW 3400.30 and purity $\geq$ 75%). Segetalin A was isolated from *S. vaccaria* seed by the method of Morita et al, 1994 (Morita et al., 1994).

Plant material

Saponaria vaccaria 'Pink Beauty' and 'White Beauty' seeds were obtained from CN Seeds Ltd (United Kingdom). Plants were grown under a daily regime of 16 h light (150 μEinstein m^{-2} s^{-1}) at 24°C and 8 h dark at 20°C. Stage 2 developing seeds were harvested according to the following scheme: Stage 1, seed white, pod green; Stage 2, seed tan; Stage 3, seed copper, pod partially dessicated; Stage 4, seed dark brown, pod dessicated.

S. vaccaria expressed sequence tag analysis

A *S. vaccaria* Pink Beauty developing seed expressed sequence tag collection developed previously (Meesapyodsuk et al., 2007) was investigated for sequences relating to segetalin biosynthesis. Initially, 6-reading frame translations of the *S. vaccaria* EST database were searched for exact matches to all circular permutations of segetalin amino acid sequences. The presence of numerous cDNA sequences appearing to encode different segetalin precursors, but showing a high degree of similarity, required reclustering using special parameters. Each set of ESTs containing sequences that corresponded to a single circular permutation of a given segetalin amino acid sequence

was first collected and then separately clustered with CAP3 (Huang and Madan, 1999) using a minimum percent identity (p) of 97 and an overlap cutoff (o) of 50. To check the EST database for precursors of previously unknown segetalins, a TBLASTN search was conducted using the consensus amino acid sequence for the precursor of presegetalin A.

Presegetalin gene expression in *S. vaccaria*

Total *S. vaccaria* Pink Beauty RNA was isolated from germinating seed, roots, leaves and flowers, using RNeasy Plant Mini kit (Qiagen). For mature and developing seeds, RNA was isolated by the method of Wang and Vodkin (1994). The first strand cDNA synthesis was performed using Superscript II reverse transcriptase (Invitrogen, http://www.invitrogen.com/site/us/en/home.html) utilizing 2 μg of total RNA as the template cDNA. Quantitative RT-PCR was performed using an Applied Biosystems Step One real-time PCR system with a Power SYBR^{R} Green PCR Master Mix (Applied Biosystems, http://www.appliedbiosystems.com/absite/us/en/home.html). The thermal cycling conditions were as follows: 50 °C for 2 min, 95 °C for 4 min, followed by 40 cycles of 95 °C for 30 s, 57 °C for the indicated annealing time (see Table 2), and 72 °C for 30 s.

Expression of presegetalin genes was compared to an *S. vaccaria Actin1* gene. The clone SVAR04NG 064_H08 was found to contain an insert with nucleotide sequence representing a partial ORF which was highly similar to other plant actins including an actin 7-like protein from *Pelargonium x hortorum* (96% identity over 171 amino acids) and an actin from *Suaeda maritima* (98% over 140 amino acids). The corresponding gene

was named *Actin1*. Table 2 indicates the oligonucleotide primers, PCR product sizes and annealing times used for the reference gene A*ctin1* and the presegetalin genes.

Preparation of pJC003, the Sga1a plant expression plasmid

The preSGA1 ORF was amplified from the clone SVAR04NG_04E02 from a previously prepared *S. vaccaria* 'Pink Beauty' developing seed cDNA library (Meesapyodsuk et al., 2007) using Vent DNA polymerase (New England Biolabs) and the primers JC1 (5'-CACCATGTCTCCAATCCTC-3') and JC2 (5'-TTACACAGGGGCTGAAGC-3'). The 103-bp PCR product was gel-purified using QIAEXII (QIAGEN, http://www.qiagen.com/default.aspx) and cloned into the Gateway entry vector pENTR/D-TOPO (Invitrogen). The DNA sequence was verified using the BigDye terminator cycle sequencing kit (Applied Biosystems Inc.) with an ABI3730 DNA sequencer. LR Clonase II (Invitrogen) was used to transfer the insert into the binary plant transformation vector pK7WG2D (Karimi et al., 2002). After DNA sequence verification, the resultant plasmid, pJC003, was used to transform electrocompetent cells of *Agrobacterium rhizogenes* LBA9402. *A. rhizogenes* LBA9402 was also transformed with pK7WG2D alone. PCR was used to confirm transformation (see below).

Transformed *S. vaccaria* roots

Sterile leaf explants of *S. vaccaria* 'White Beauty' (which does not contain segetalin A; J.J. Balsevich, unpublished) were transformed separately with either pJC003 or pK7WG2D and hairy roots were regenerated as described previously (Schmidt et al., 2007). Rapidly growing lines that showed kanamycin resistance and green fluorescence

with no bacterial contamination were used to establish single hairy root lines. All transformed root lines originated from independent green-fluorescent adventitious roots.

Hairy root DNA extraction and PCR analysis

DNA was extracted from a 100-200 mg sample of each root culture using the DNeasy Plant Mini Kit (Qiagen) and subjected to multiplex PCR analysis to simultaneously score for the presence or absence of the *rolC*, *virD*, *egfp* and *nptII* genes as described previously (Schmidt et al., 2007). To confirm that kanamycin-resistant and *egfp*-positive hairy roots were transformed, the presence of the *Sga1a* XXX???XXXgene was verified by PCR. The PCR reaction mixture (25µl) contained 1 µL of DNA, as prepared above, in 1 x PCR reaction buffer, 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.4 µM of each primer (JC3 5'-CCGACAGTGGTCCCAAAGATG-3'(vector-specific) and JC4 5'GCCTGAAAAGCCCAAACTGG-3' (gene-specific)) and 5 U Taq DNA polymerase (Invitrogen). Amplification was performed in a Stratagene Robocycler Gradient 96 using the following program: 94°C for 10 min, 30 cycles of 94°C for 30 s, 62°C for 40 s, and 72°C for 50 s, followed by 72°C for 10min. The expected size of the PCR fragment was 398 bp.

LC/MS analysis of hairy roots

For each transformed root line, 1.2-2.2 g fresh weight of hairy roots were added to 5 mL methanol in a 10 mL glass screw-top tube and homogenized using a Polytron (Kinematica, http://www.kinematica-inc.com). The sample was sonicated for 20 min using a Branson 2510 ultrasonic cleaner (Branson Ultrasonic Corporation,

http://www.bransonultrasonics.com), centrifuged at 1,400 x g for 3 min and the supernatant transferred to a new tube. An additional 5 mL methanol was added to the pellet and sonicated, centrifuged and decanted, as above. This step was repeated once. A tube containing the combined supernatants was placed in a heating block at 30-35°C and the methanol was evaporated under a nitrogen stream. The sample was resuspended in 1 mL dH_2O, transferred to a 1.5 mL tube, and centrifuged at 12,000 x g for 5 min. The supernatant was then placed in a Costar SPIN-X® (0.22μm cellulose acetate; Corning, http://www.corning.com/index.aspx) centrifuge filter unit and centrifuged at 12,000 x g for 1 min. The filtrate was then used for analysis by LC/MS-using a 2695 Alliance chromatography system, with inline degasser, coupled to a ZQ mass detector and a 2996 photodiode array detector (Waters, Milford MA). MassLynx software was used for data acquisition and analysis. The column used was a Waters Sunfire 3.5-μm RP C-18 150 x 2.1 mm. The flow rate was 0.15 mL/min. The column was maintained at 35°C during analysis. The binary solvent system consisted of 90:10 v/v water/acetonitrile containing 0.12% acetic acid (solvent A) and acetonitrile containing 0.12% acetic acid (solvent B). The gradient program used was 0 – 8 min, 95: 5 A/B; 8 – 31 min, 95:5 to 50:50 A/B; 31 – 33 min, 50:50 to 0:100 A/B; 33 – 48 min, 0:100 A/B. Voltage parameters for negative electrospray ionization (ESI^-) were: capillary, 2.80 kV; cone, ramped from -15 to - 45 V; extractor, -3.00 V; RF lens, -0.5 V; for positive electrospray ionization (ESI^+), they were: capillary, 3.50 kV; cone, ramped from +15 to +45 V; extractor, 6.00V; RF lens, 0.9 V.

***In vitro* processing of presegetalin A**

"Stage 2" developing seed from *S. vaccaria* (var. White Beauty) was homogenized manually with a plastic pestle in a 1.5 mL low protein binding microtubes. One gram of seeds was ground for 2 min in 4 X 250 μL 20 mM Tris buffer (pH 8) on ice followed by centrifugation at 13,000 x g for 5 min. The supernatant was removed and another 250 μL buffer was added and the grinding and centrifugation was repeated. The supernatant fractions were pooled and this crude extract supernatant was used for enzyme assays. The crude extract protein was measured using the micro BCA protein assay (Pierce, http://www.piercenet.com). The *in vitro* assay contained; 15 mM Tris (pH 8), 100 mM NaCl, 2 mM DTT, 0.2 mg BSA, 50 μg / mL presegetalin A. The assay was initiated by the addition of crude extract supernatant, equivalent to 24.4 μg protein, to a total reaction volume of 100 μL. The assay was incubated at 30 ^{O}C XXXhow long???XXXand stopped by placing reactions in dry ice. The assays were lyophilized, re-suspended in methanol, evaporated and re-suspended in 50:50 v/v methanol/water for LC/MS analysis. LC/MS/MS analysis was used to detect production of segetalin A using an Agilent 6320 Ion Trap LC/MS system under standard auto MS/MS conditions fitted with a Zorbax 300 EXTEND-C18 column (150 X 2.1 mm, 3.5 μm particle size) maintained at 35^{O}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 minutes.

ACKNOWLEDGEMENTS

We are grateful to Greg Bishop, Carla Barber, Dustin Cram and the PBI Bioinformatics and DNA Technology Units for technical support, to Sean Hemmingsen and Jon Page for reviewing the manuscript, to the National Research Council of Canada's GHI2, CEHH, PHW and PPHS Programs for funding.

REFERENCES

Cascales, L. and Craik, D.J. (2010) Naturally occurring circular proteins: distribution, biosynthesis and evolution. *Org. Biomol. Chem.* **8**, 5035-5047.

Craik, D.J., Cemazar, M., and Daly, N.L. (2007) The chemistry and biology of cyclotides. *Curr. Opin. Drug Discov. Devel.* **10**, 176-184.

Ding, Z.T., Zhou, J., Tan, N.-H., and Cheng, Y.-X. (1999) Progress in the chemistry of cyclic peptides in the plants of the Caryophyllaceae. *Huaxue Yanjiu Yu Yingyong* **11**, 492-494.

Finking, R. and Marahiel, M.A. (2004) Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **58**, 453-488.

Gillon, A.D., Saska, I., Jennings, C.V., Guarino, R.F., Craik, D.J., and Anderson, M.A. (2008) Biosynthesis of circular proteins in plants. *Plant J.* **53**, 505-515.

Huang, X. and Madan, A. (1999) CAP3: A DNA sequence assembly program. *Genome Res.* 9, 868-877.

Itokawa, H., Yun, Y., Morita, H., Takeya, K., and Yamada, K. (1995) Estrogen-like activity of cyclic peptides from *Vaccaria segetalis* extracts. *Planta Med.* **61**, 561-562.

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.

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.

Karimi, M., Inze, D., and Depicker, A. (2002) GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci.* **7**, 193-195.

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.

Li, F., Zhang, F.M., Yang, Y.B., Yang, X.Q., Cao, Q.E., and Ding, Z.T. (2008) A new cyclopeptide from *Dianthus caryophyllus*. *Chin. Chem. Lett.* **19**, 193-195.

Matsumoto, T., Tashiro, N., Nishimura, K., and Takeya, K. (2002) New cyclic peptides from *Citrus aurantium*. *Heterocycles* **57**, 477-482.

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.

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.

Morita, H., Enomoto, M., Hirasawa, Y., Iizuka, T., Ogawa, K., Kawahara, N., Goda, Y., Matsumoto, T., and Takeya, K. (2007) Cyclonatsudamine A, a new vasodilator cyclic peptide from *Citrus natsudaidai*. *Bioorg. Med. Chem. Lett.* **17**, 5410-5413.

Morita, H., kayahita, T., Uchida, A., Takeya, K., and Itokawa, H. (1997) Cyclic peptides from higher plants. 33. Delavayins A-C, three new cyclic peptides from *Stellaria delavayi*. *J. Nat. Prod.* **60**, 212-215.

Morita, H. and Takeya, K. (2010) Bioactive cyclic peptides from higher plants. *Heterocycles* **80**, 739-764.

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.

Mozoruk, J., Hunnicutt, L.E., Hunter, W.B., and Bausher, M.G. (2008) CsV03-3 is a member of a novel gene family from citrus that encodes a protein with DNA binding activity and whose expression is responsive to defense signals and abiotic stress. *J. Plant Physiol* **165**, 531-543.

Picur, B., Cebrat, M., Zabrocki, J., and Siemion, I.Z. (2006) Cyclopeptides of *Linum usitatissimum*. *J. Pept. Sci* **12**, 569-574.

Pomilio, A.B., Battista, M.E., and Vitale, A.A. (2006) Naturally-occurring cyclopeptides: structures and bioactivity. *Curr. Org. Chem.* **10**, 2075-2121.

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.

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.

Schmidt, J.F., Moore, M.D., Pelcher, L.E., and Covello, P.S. (2007) High efficiency *Agrobacterium rhizogenes*-mediated transformation of *Saponaria vaccaria* L. (Caryophyllaceae) using fluorescence selection. *Plant Cell Rep.* **26**, 1547-1554.

Tan, N.H. and Zhou, J. (2006) Plant cyclopeptides. *Chem. Rev.* **106**, 840-895.

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.

SUPPORTING INFORMATION

Figure S1. Mass spectrum of segetalin A

Tables

Table 1 *S. vaccaria* genes encoding segetalin precursors inferred from EST data. Contig size for the developing seed EST collection (Meesapyodsuk et al., 2007) is indicated.

Segetalin	Segetalin precursor	Gene	Contig Size	Genbank Accession Number
A	A1	*Sga1*	236	XXX
B	B1	*Sgb1*	133	
D	D1	*Sgd1*	205	
	D2	*Sgd2*	191	
	D3	*Sgd3a*	10	
		Sgd3b	9	
F	F1	*Sgf1a*	30	
		Sgf1b	17	
		Sgf1c	5	

G	G1	*Sgg1*	33	
H	H1	*Sgh1*	128	
J	J1	*Sgj*1	28	
K	K1	*Sgk*1	30	
GLPGWP	GLPGWP1	*Glpgwp*1	7	

Table 2 Parameters of qPCR experiments for measurement of presegetalin gene expression.

Gene	Primer 1 Sequence (5' to 3')	Primer 2 Sequence (5' to 3')	PCR product size (bp)	Annealing time (s)
Actin1	GGAGAAGCTGGCCTATGTT	GAGTTGTAGGTAGTCTCATGGATAC	204	60 (30 for *Sga1a*)
Sga1a	CAAGGTGTCCCAGTTTGG	GAAACATGACACTCGACATAAC	151	30
Sgg1	CCCCAAGGCGTGAAATAT	CAGAAACGGGTAATACAACACT	160	60
Sgh1	GGCTACAGATTTAGTTTTCAGG	CTACAGAAATAAGAAACACGACATT	159	60
Sgj1	TATTGCGAGAATGGCCAC	CATTGGGGCACAAGCATC	118	60
GLPGWP1	AAGGTCTCCCTGGTTGGC	AATTACAGAAATAAAAAACACG	164	60

line numbers are indicated. Segetalin A was determined by LC/MS using an external standard and triplicate samples. Means and standard deviations are indicated.

Figure 6. LC/MS analysis of the conversion of presegetalin A (linear) to segetalin A (cyclic) *in vitro* by *S. vaccaria* seed extracts. Segetalin A production was monitored by scanning at m/*z* = 610.3 (M+1). **a**, segetalin A standard; **b**, 5 h incubation of 12.2 μg developing seed protein extract with 2.5 μg presegetalin A; **c**, same as in b except that no substrate was added; **d**, same as in b except the incubation time was 0 h. Insets represent MS/MS fragmentation patterns of the m/z 610.3 ions detected and show typical and identical peptide fragmentation patterns. Chromatograms **b**, **c** and **d** represent the same volume of assay.

Figure S1. Mass spectrum of segetalin A standard under ES^+ conditions showing M+1 (m/z 610) and fragment ions m/z 582 and m/z 511 that were used to verify presence of segetalin A in hairy root samples. Qualitative analysis of Segetalin A was done by confirming the presence of ions with m/*z* = 610, 582 and 511 in an approximate ratio of 6:2:3 and quantitative analysis was done using only the m/*z* = 610 ion

Figures

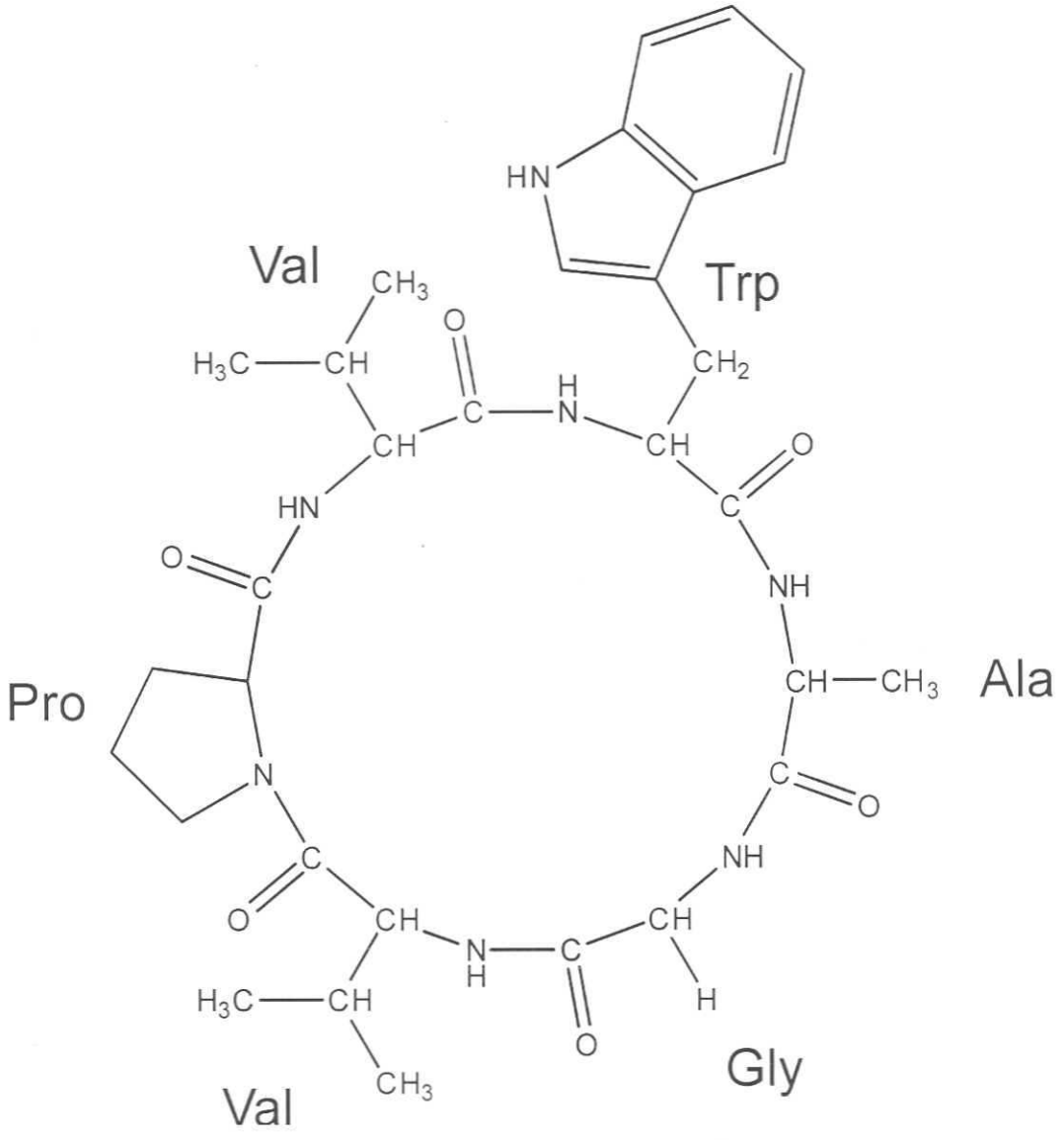

Figure 1

A

S. vaccaria

A class

MSPILAHDVVKPQ	GVPVWA	FQ	AKDVENASAPV	preSGA1
MSPILAHDVVKPQ	GVAWA	FQ	AKDVENASAPV	preSGB1
MSPIFAHDVVNPQ	GLSFA	FP	AKDAENASSPV	preSGD1
MSPIFAHDVVKPQ	GLSFA	FP	AKDAENASSPV	preSGD2
MSPILAHDVVKPQ	GLSFA	FP	AKDAENASSPV	preSGD3
MSPIFVHEVVKPQ	GVKYA	FQ	PKDSENASAPV	preSGG1
MSPIFAHDIVKPK	GYRFS	FQ	AKDAENASAPV	preSGH1
MSPILALDRYKPE	*GRVKA*	FQ	AKDAENASAPV	preSGK1
MSPILSHDVVKPQ	*GLPGWP*	FQ	AKDVENASAPV	preGLPGWP1

F class

MATSFQFDGLKPS	FSASYSSKP	IQTQVSNGMDNASAPV		preSGF1
MATSFQLDGLKPS	*FGTHGLPAP*	IQ	VPNGMDDACAPM	preSGJ1

B

Dianthus caryophyllus

MSPNSTRDILKPQ	*GPIPFYG*	FQ	AKDAENASVPV	AW697819
MSPNSTRDLLKPL	*GYKDCC*	VQ	AKDLENAAVPV	CF259478

C

***Citrus* spp.**

METTCAGN NWSE	GLLLPPFG	SIADDDVMND	NLDFLNVPQYGRNPDYMG	EY850721	*aurantium*
MKTLAGAGMSDPSE	GLVLPS	SIADDDVGND	NLDLIVIPQYGRNPDYYG	EY848546	*aurantium*
MKTLPGAGMSDPSE	GYLLPPS	SIADDDVGND	NLDLIVIPQYGRNPDYYG.	BB999724	*natsudaidai*
MKIMETTCAGNDDCLE	*GRPWNLA*	SIVDDNVANDVNLDLLAVPQYGRNTDQTG		ABO93452	*sinensis*
MKNMETTSAGNDDWLE	*GYVAA*	SIVDDNIANDVNLDLLTVPQYGRNIDQTG		ABO93453	*sinensis*
MKNMETTCAGNDDWLE	*GAPWLIAA*	SIVDDNIANDVNLDLLTVPQYGRNIDQTG		ABO93454	*sinensis*
MKNMETTSAGNADWLE	*GVPWAIAA*	SIVDDNIANDVNLDLLTVPQYGRNIDQTG		ABO93455	*sinensis*

Figure 2

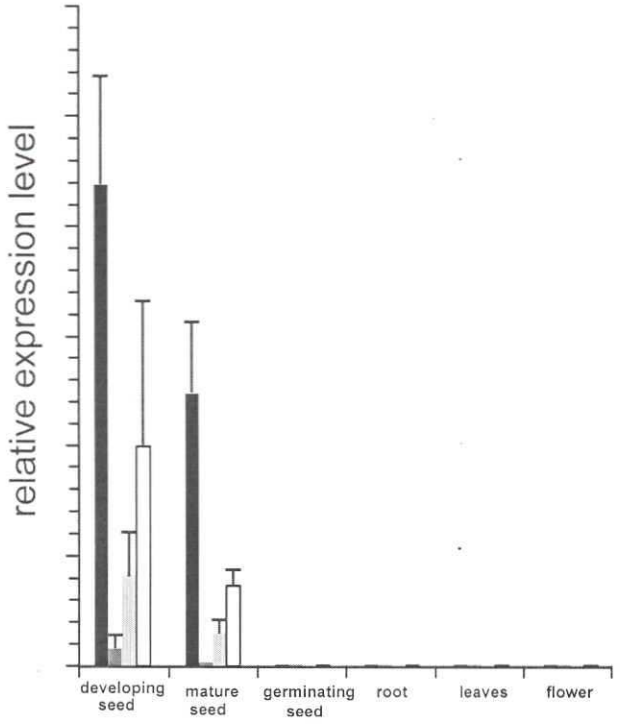

Figure 3

XXX scaleXXX

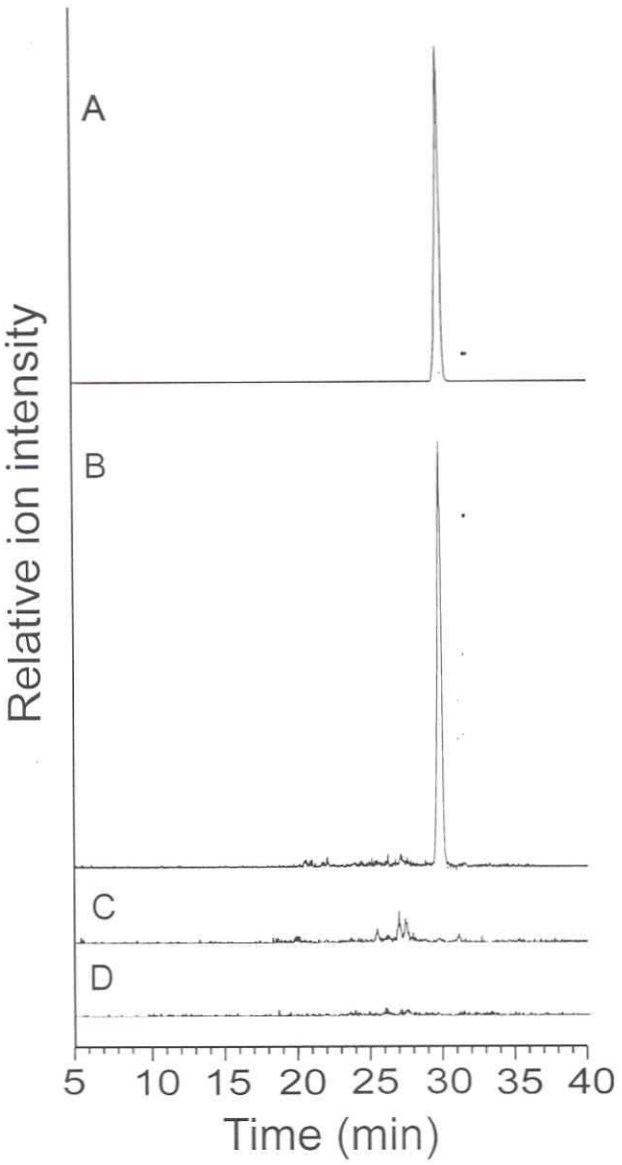

Figure 5

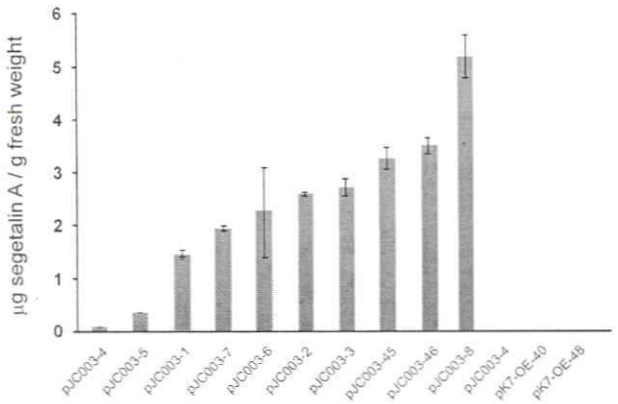
µg segetalin A / g fresh weight
0
1
2
3
4
5
6
pJC003-4
pJC003-5
pJC003-1
pJC003-7
pJC003-6
pJC003-2
pJC003-3
pJC003-45
pJC003-46
pJC003-8
pJC003-4
pK7-OE-40
pK7-OE-48

Figure 6.

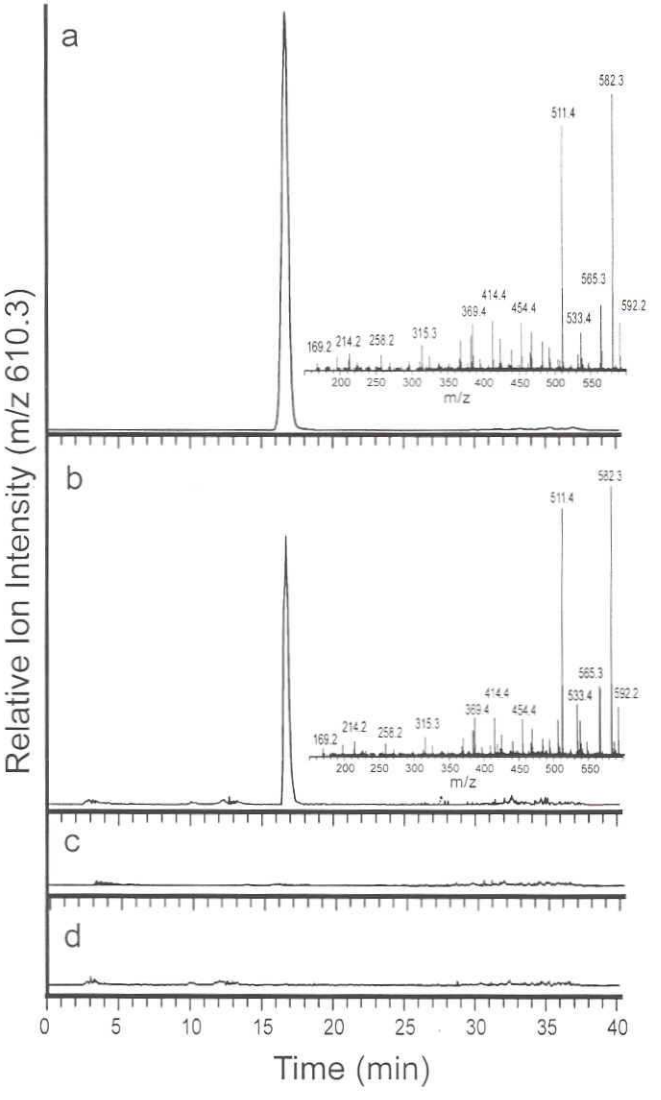
a
b
c
d
582.3
511.4
565.3
592.2
533.4
414.4
369.4
454.4
315.3
214.2
258.2
169.2
200 250 300 350 400 450 500 550
m/z
Relative Ion Intensity (m/z 610.3)
0 5 10 15 20 25 30 35 40
Time (min)

Figure 7

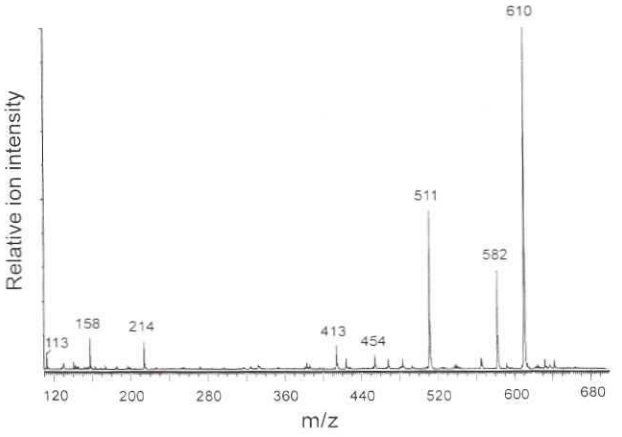
610
511
582
158
214
413
454
113
Relative ion intensity
120
200
280
360
440
520
600
680
m/z

Figure S1