CHARACTERIZATION OF PEPTIDE CYCLASE 1 (PCY1), A SERINE PROTEASE-LIKE ENZYME INVOLVED IN CYCLIC PEPTIDE BIOSYNTHESIS IN PLANTS

A Thesis Submitted to the College of Graduate Study and Research

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the Department of Biochemistry

University of Saskatchewan

Saskatoon

By

Pareshkumar T. Pujara

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a M.Sc. degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Biochemistry
University of Saskatchewan
Saskatoon, Saskatchewan
S7N 5E5
Canada

ABSTRACT

Plants within the Caryophyllaceae, and certain other families, produce cyclic peptides (CPs) which generally consist of 5–12 proteinogenic amino acids. Until recently, very little was known about the biosynthesis of CPs in the Caryophyllaceae. Recently, in the Covello lab, two enzymes in *Saponaria vaccaria* were found to be involved in the processing of ribosome-derived linear precursors, giving rise to cyclic peptides. Thus, oligopeptidase 1 (OLP1) and peptide cyclase 1 (PCY1) are involved in the biosynthesis of segetalin A (a six-membered CP) from a 32 amino acid linear peptide precursor called presegetalin A1. PCY1 carries out the unusual cyclization reaction to form mature segetalin A from a linear intermediate. The purified recombinant PCY1, the first cloned plant enzyme whose function is peptide cyclization, was identified as a homologue of a prolyl-oligopeptidase from the S9 serine protease family. In principle, PCY1 performs an intra-molecular transpeptidation reaction to produce a CP. A homology-based structural model of PCY1 suggests that it has two domains, a catalytic α/β hydrolase domain and an unusual β- propeller domain. In an effort to define the substrate specificity of PCY1, a wide variety of synthetic peptide precursors were tested in assays and the results are discussed.

PUBLISHED WORK

Barber, C.J., **Pujara, P.T.,** Reed, D.W., Chiwocha, S., Zhang, H., and Covello, P.S. (2013). The two step biosynthesis of cyclic peptides from linear precursors in a member of the plant family Caryophyllaceae involves cyclization by a serine protease-like enzyme. J. Biol. Chem. 288, 12500-12510.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Patrick S. Covello, for providing an exciting research opportunity to work in his lab as graduate student. I am thankful to him for his guidance, constant encouragement, financial support and patience with me throughout my entire research. It has been an honor to complete my master degree under his supervision.

My sincere thanks to Darwin Reed for much needed help in interpreting results and for teaching me required technical skills. He is a good teacher in many ways and he is one the reasons I made it to the end. I would like to thank Carla Barber for her constant support, generosity and advice. The guidance from Darwin and Carla has served me well and I owe them my heartfelt appreciation.

I am grateful to department head Dr. Bill Roesler and my committee members Dr. Jeremy Lee, Dr. Ron Geyer, Dr. Michele Loewen and Dr. Dwayne Hegedus. Special thanks to Dr. Michele Loewen her support and help in analysis of bioinformatics results.

I am indebted to my friends and colleagues Dr. Yeon-Bok Kim, Dr. Xiaoqiang Liu and Fernanda de Costa. They made me laugh on tough days and provided a good ear. Sharing coffee and cracking jokes with them was memorable.

My friends in Canada, India and other parts of the World were sources of fun and motivation. Special load of thanks go to Smitesh, Ketanbhai, Mayank, Darshak, Viral, Priteshbhai, members of ‘Fosla’ and ‘TB boys’ groups. Vrajesh, I am glad that my friendship with you has extended well beyond our shared time.

My greatest appreciation goes to my family for their unconditional support and freedom to take independent decisions in life. I am indebted to my Maa and Papa, their love is true inspiration and driving force for me. My heartfelt appreciation to Kajalben, Chiragkumar, Bipinbhai and Shitalbhabhi; they are strongest pillars of my life. It is difficult to imagine my education without their direct and indirect support. Pratham and Daskh, their funny faces and real stories added comic element in my life.

This thesis is dedicated to

Batuk & Zalak

My Family

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLANT SPECIES

Scientific name	**Common name**
Dianthus superbus	Fringed pink
Dianthus caryophyllus	Carnation and Clove pink
Saponaria vaccaria	Cowherb, Cowcockle and Cow soapwort
Silene vulgaris	Campion bladder, Bird's eggs and Maiden's Tears

LIST OF ABBREVIATIONS

AEP	Asparaginyl endoproteinase
BLAST	Basic Local Alignment Search Tool
CP	Cyclic peptide
DOPE	Discrete optimized protein energy
ESI	Electrospray ionization
EST	Expressed sequence tag
LC/MS	Liquid chromatography / Mass spectrometry
NRPS	Non-ribosomal peptide synthetases
OLP1	Oligopeptidase 1
PCY1	Peptide cyclase 1
POP	Prolyl oligopeptidase
qPCR	Quantitative polymerase chain reaction
RTDs 1-3	Rhesus theta defensins-1-3
SFTI	Sunflower trypsin inhibitor

1. INTRODUCTION

Cyclic peptides (CPs) are one of the important members of the ribosomally synthesized and post-translationally modified peptide natural products (Arnison *et al.,* 2013). They are widely distributed in all kingdoms of life. The circular backbone of CP may be formed by an amide (peptide) bond or non-amide bond with incorporation of proteinogenic and/or non-proteinogenic (modified amino acids) amino acids. CPs exhibit a wide range of potential biological activities. Because of their structural diversity and size, CPs are being explored in drug discovery processes, and they may be useful to develop a novel class of therapeutics in the future (Bockus *et al.,* 2013). There are two principal pathways for the biosynthesis of CPs. First, the biosynthesis of CPs may occur without involvement of the ribosome in which non-ribosomal peptide synthetases (NRPS) participate. Examples of this include biosynthesis of the immunosuppressant cyclosporin-A and the antibiotic gramicidin-S. Second, ribosomes may be involved in the ordering of mRNA-encoded amino acids to form a linear peptide precursor whose processing and cyclization of the peptide is often facilitated by protease-like enzymes Examples of ribosome-derived CPs include cyclotides and Caryophyllaceae-like CPs in plants, and cyanobactins in cyanobacteria.

Most proteins and peptides in a cell are linear . However, CPs have two significant advantages over linear peptides with respect to stability and peptide activity. First, CPs are resistant to protease attack, and second, the ends of linear peptides are often flexible, in contrast to a highly organized core. From an entropic perspective, this flexibility affects the binding affinity and biological activity of proteins. Thus, in principle, both the stability and the activity of proteins and peptides can be improved by "tying up their loose ends" (Craik, 2006; Craik, 2010).

Different CPs, with their native structure and certain chemical modifications, have many applications. There are four possible solutions to obtain a sufficient quantity of CPs: extraction of CPs from their natural source, chemical synthesis, *in vivo* and *in vitro* synthesis with implementation of knowledge about CPs biosynthetic pathways. The first option requires extensive purification skills, and it may not produce an adequate quantity to fulfill

requirements. There has been enormous work carried out to develop chemical methods for synthesis of CPs (White and Yudin, 2011), but chemical synthesis has two main disadvantages: the problem of protecting side chains during synthesis, and a low yield of the CP product (Sieber and Marahiel, 2003). It has been demonstrated by researchers that many CPs are synthesized from their synthetic precursor peptides by a protease-like enzyme (Barber *et al.*, 2013; Lee *et al.,* 2009). These protease-like enzymes also incorporate chemical modification, such as D-amino acid, at certain positions in synthetic precursors and that remain as an integral part of the incipient CP. Application of protein engineering techniques and understanding of enzyme structures and mechanisms offer exciting possibilities for enzymatic production of CPs *in vitro* (Koehnke *et al.,* 2012). The presented research focuses on the mechanism, structure-function relationship and characterization of a serine-protease-like enzyme involved in CP biosynthesis in plants.

2. LITERATURE REVIEW

2.1 Classification of Cyclic Peptides

Over the past few decades, there has been increased research interest in the field of CP synthesis and discovery of CPs that display unique biological activity. From 1970 to 2006, many reviews (Tschesche and Kaubmann, 1975; Schmidt *et al.*, 1985; Tan *et al.*, 1997; Itokawa *et al.*, 1997) were published on the distribution, isolation, biological activities, synthesis, structure, and classification of CPs. There is no specific classification system available to accommodate all known CPs, due to their diverse chemical structures and wide distribution in nature. Tan and Zhou, 2006 proposed the classification of plant CPs based on their structure and distribution in plants. All plant CPs can be grouped into two classes, five subclasses, and eight types. In CP biosynthesis, the ring formation via a peptide bond or non-peptide bond defines the class of CPs as homocyclopeptides and heterocyclopeptides. This Class can be further divided into five subclasses on the basis of the number of rings present in the CP structure. Homocyclopeptides can be divided into three subclasses: homomonocyclopeptides, homodicyclopeptides, and homopolycyclopeptides. Heterocyclopeptides can be divided into two subclasses: heteromonocyclopeptides and heterodicyclopeptides. These subclasses can be further classified into eight types by considering characteristics of the rings and sources. According to this classification, the Caryophyllaceae-type CPs are classified as type VI homomonocyclopeptides.

In 2006, CPs from different taxa were collected in the Cybase database (http://www.cybase.org.au) by The Institute of Molecular Bioscience and University of Queensland, Australia. Cybase is a database of gene-encoded, backbone-cyclized CP sequences and structures, primarily focused on cyclotides. The CPs are divided into several classes based on their origin and sequence similarity. These classes are BBI-like TI (Bowman-Birk-like trypsin inhibitor), squash TI, sunflower CPs, bacterial CPs, cyclic defensins from primates, mushroom toxins, cyclotides, and Caryophyllaceae-like CPs. The database provides search and display capabilities for sequence, structure, and function data. Additionally, the database also provides a range of analytical and predictive tools to handle the challenges of cyclic peptide

characterization and engineering (Mulvenna *et al.*, 2006; Wang *et al.*, 2008; Kaas and Craik, 2010). Mylne *et al.*, 2012 grouped the gene-encoded, backbone cyclized CPs from plant families into four distinct classes: kalata-type CPs, PawS-derived CPs, cyclic knottins, and CPs from Caryophyllaceae and Rutaceae plant families.

2.2 Nomenclature of Cyclic Peptides

Different types of CPs are found with complex chemical structures. The cyclization patterns are the fundamental characteristics which can be use to distinguish different types of CPs, *e.g.* `head-to-tail,` `side chain-to-head (or tail),` and `side chain-to-side chain.` In addition, many CPs have D-amino acids and modified amino acids as part of their structure. As Caryophyllaceae-like CPs have a head-to-tail pattern, their detailed nomenclature is discussed in this section. Spengler *et al.*, 2005 has discussed the nomenclature of other CPs in accordance with IUPAC-IUBMB recommendations.

Recently, Arnison *et al.*, 2013 reviewed the ribosomally synthesized and post-translationally modified peptide natural products and proposed recommendations for universal nomenclature. Additionally, the name ‘orbitides’ was proposed instead of Caryophyllaceae-like CPs, which include all head-to-tail cyclized plant CPs without disulfide crosslinks. In principle, all Caryophyllaceae-like CPs are orbitides. There is a one-line text formula to represent CPs using three-letter codes. Generally, the prefix `cyclo` followed by a hyphen and amino acid sequence in brackets are arranged in the specific order to represent CPs. The use of a hyphen is optional; `cyclo` is written in normal or italic font, and the brackets are round or square. The three-letter codes are standard and represent the proteinogenic amino acid sequences in peptides. At this point, consider segetalin A as an example to understand the nomenclature of head-to-tail cyclized CPs. Segetalin A is a cyclic hexapeptide, GVPVWA. With the nomenclature system described above, segetalin A can be written as *cyclo*[Gly-Val-Pro-Val-Trp-Ala] or *cyclo*[GVPVWA] in text. Despite the simplicity of the nomenclature, there is a disadvantage. Being a CP, there are six possible notations with which to write the one-line formula for segetalin A: *cyclo*[GVPVWA], *cyclo*[VPVWAG], *cyclo*[PVWAGV],

cyclo[VWAGVP], *cyclo*[WAGVPV], and *cyclo*[AGVPVW]. According to Spengler *et al.*, 2005, although all of the segetalin A notations are chemically identical, they create confusion because they produce different results when searching in databases such as SciFinder. There are two options to avoid this confusion. Either the amino acid sequence should follow alphabetical order in the notation, or it should follow the sequence of CPs as it occurs in corresponding precursors. The second option is more logical as CP amino acid sequences are precisely stored in genes for the ribosomally synthesized CPs.

2.3 Distribution of Cyclic Peptides in Nature

CPs have been reported in plants, animals, fungi, bacteria, and cyanobacteria (blue-green algae). They have shown exceptional structure-function diversity depending on their source (Figure 2.1).

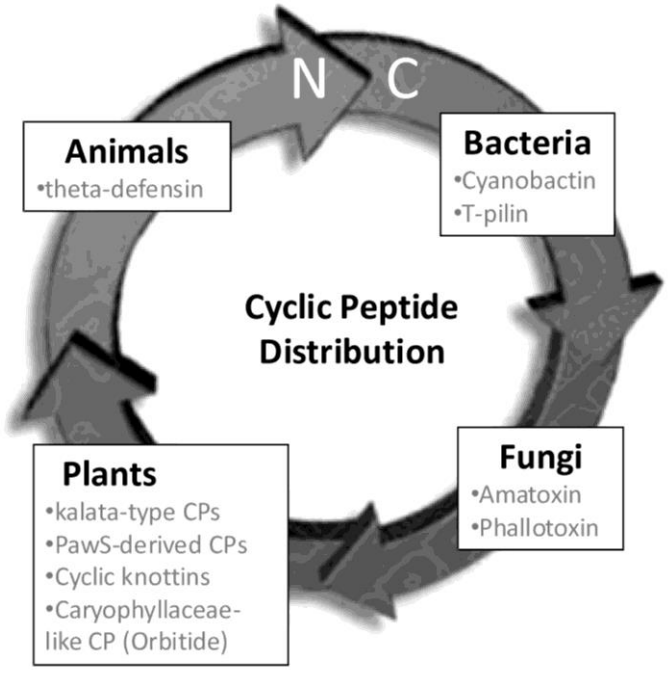

Figure 2.1 Distribution of CPs and their well-known examples in various taxa of life.

2.3.1 Cyclic Peptides from Bacteria

Bacteria are capable of making large circular bacterial peptides that are different in size and biosynthesis. The most studied of these is enterocin AS-48, which is a so-called bacteriocin that is well-known for its antimicrobial activity. AS-48 was isolated from *Enterococcus faecium* (Folli *et al.,* 2003), and in the last decade, nine more bacteriocins were isolated and characterized from Gram-positive bacteria. Bacteriocins are classified into three groups: Type I are the AS-48-like peptides, Type II are the gassericin-like peptides, and Type III are the subtilosin A and subtilosin A variants. Both AS-48 (70 residues) and carnocyclin A (60 residues) are bacteriocins. They have low sequence similarity, but interestingly, X-ray crystallography results have revealed that they have highly similar structures, including an important 'saposin fold' (Martin-Visscher *et al.,* 2009) Despite similar structures, both have different biological functions. The carnocyclin A is a monomeric peptide with anion binding and membrane transport properties, while the AS-48 is dimeric, with the ability to form membrane pores.

Another well-known example of a bacterial CP is T-pilin, which is made up of 74 amino acid residues and is head-to-tail cyclized by amide bond. T-pilin was discovered in *Agrobacterium tumefaciens* and was found to play a key role in transferring genetic material during the process of transformation. T-pilin is processed from full-length VirB2 pro-pilin, a 121-amino-acid-long precursor, into a cyclized peptide by two enzymes, a signal peptidase, and peptidyl cyclase (Lai *et al.,* 2002). The peptidyl cyclase is a key biosynthetic enzyme which performs the final cyclization step. Maksimov *et al.,* 2012 has predicted two putative lasso peptide precursors in archaebacteria by precursor-centric genome-mining approach.

2.3.2 Cyclic Peptides from Cyanobacteria

Cyanobactins comprise a group of CPs discovered in cyanobacteria. They are synthesized as ribosomal precursor peptides that are extensively modified by enzymes to produce mature CPs. Patellamides are well characterized examples of cyanobactins. They are generated from a ribosomally encoded gene cluster (*pat*) in *Lissoclinum patella*. The *pat* gene

cluster is composed of seven ORFs (*patA-patG*) in which *patE* encodes a precursor peptide which includes the amino acid sequences of patellamides A and C (Lee *et al.*, 2009; Jones *et al.*, 2009). The *patA* gene product recognizes the flanking conserved amino acids and cleaves the precursor to release the N-terminal peptide, while the *patG* gene product is involved in the cyclization process and helps to form the amide bond to shape a ring structure by performing an intra-molecular transpeptidation reaction (Figure 2.2). The PatG recognize the (-AYDG) signal, which is immediately downstream to the incipient CP sequence in the precursor protein. The involvement of this gene cluster in patellamide biosynthesis was confirmed through a successful heterologous production of patellamide A in *E. coli* (Schmidt *et al.*, 2005).

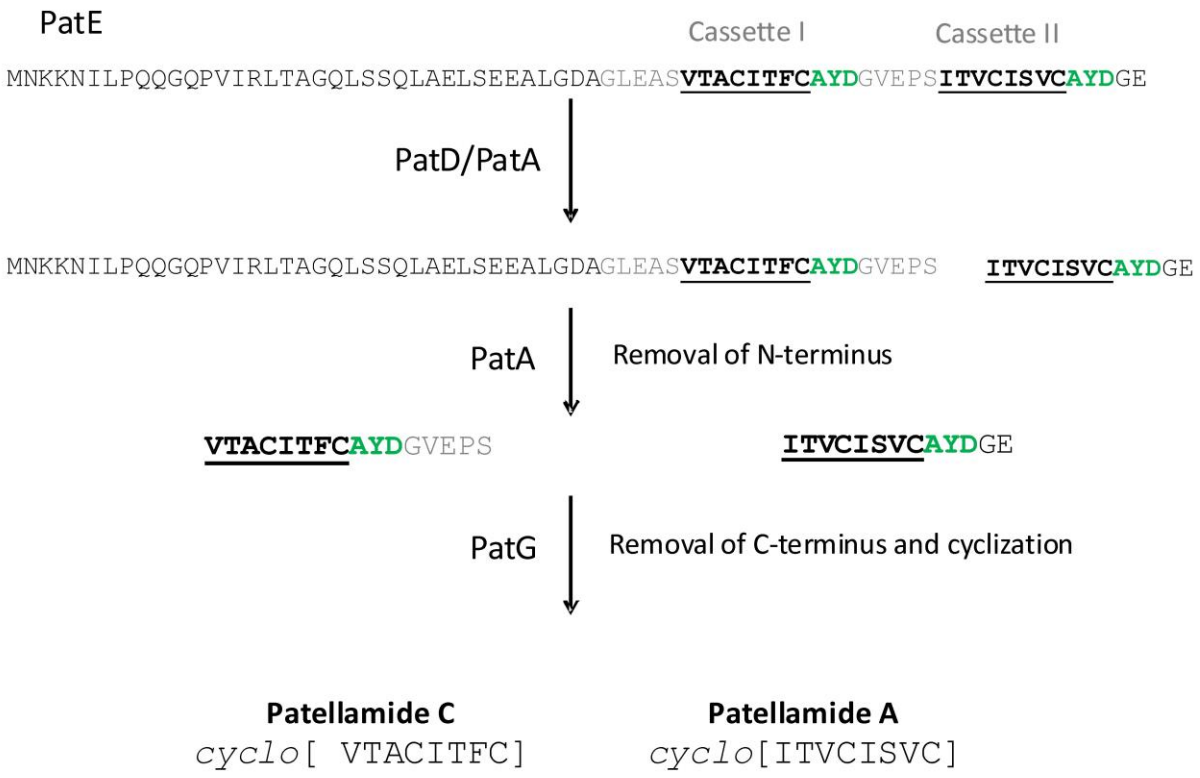

Figure 2.2 Biosynthetic pathway of backbone-cyclized patellamides in cyanobacteria. A multi-domain ribosome-derived precursor (PatE) encodes the patellamide C and patellamide A (underlined). PatA separates two cassettes and cleaves the N-terminus (blue) from a precursor peptide. PatG recognises the C-terminus signal (green), and performs the important cyclization reaction to yield two CPs (modified from Oman and van der Donk, 2010).

PatG was identified as S8 endopeptidase from the serine protease family with the catalytic triad His, Ser, and Asp. The three dimensional structure of patG revealed that it has a subtilisin-like fold. Interestingly, PatG has specific insertions which are not common in subtilisin-like proteases (Koehnke *et al.,* 2012). PatG also has an N-terminal oxidase domain, but it is not required for the cyclization reaction. PatG can act on a wide range of precursors *in vivo* and *in vitro*. PatG was able to cyclize 6 to 11 amino acids effectively. It was also found active on 29 native precursors and other modified precursors which include nonproteinonegenic and D- amino acids. It was also observed that the activity of PatG is low, with a turnover rate of about one per enzyme molecule per day (McIntosh *et al.*, 2010).

2.3.3 Cyclic Peptides from Fungi

Some mushroom species in the genera *Amanita, Galerina, Lapiota, and Conocybe* are extremely poisonous to mammals. These include *Amanita phalloids* (death cap) and *Amanita virosa* (destroying angles) (Goransson *et al.,* 2012). The presence of 'alkaloid-like' toxic compounds in *Amanita* mushrooms has been known for a century. The 'alkaloid-like' toxic elements in the mushrooms are bicyclic amatoxins (octapeptides) and phallotoxins (heptapeptides) -. The potent toxic effect of these CP toxins is mediated through the specific binding interaction of the toxins to the RNA polymerase II, thereby inhibiting the transcription process (Kaplan *et al.,* 2008).

There are nine types of amatoxins found in mushrooms, and they are post-translationally modified product of α-amanitin or β-amanitin. The post-translational modification process incorporates hydroxylation and an unusual crosslink between Trp and Cys residues via sulfoxide. The phallotoxins originate from phallacidin or phalloidin and the internal link between Trp and Cys via sulfide. Amatoxins and phallotoxins share a common genetic origin, and both are products of the MSDIN gene family (Luo *et al.,* 2010; Hallen *et al., 2007*). *AMA1* and *PHA1* encode α-amanitin and phalloidin, respectively. The mature CPs sequence are buried inside the ~35 amino acid precursor peptide as a hypervariable toxin region, which is flanked by Pro residue. The amino acid sequence toward both ends is

conserved. A prolyl oligopeptidase (POP) has been implicated in the maturation of CP, but its exact role and mechanism are still unclear (Luo *et al.*, 2009).

2.3.4 Cyclic Peptides from Animals

A remarkably small number of CPs are reported from animals. In 1999, a new class of defensins was discovered from the leukocytes of rhesus macaques (*Macaca mulatta*) and is called Θ-defensins. Defensins are typically linear proteins with more than 40 amino acids, but the Θ-defensins have a circular backbone. Currently, there are six members in this class, and they are further classified into two groups— rhesus theta defensins-1-3 (RTDs 1-3) and retrocyclins. Both groups of Θ-defensins have different biological activities; the RTD-1 has shown antibacterial and antifungal activity, while retrocyclins have been reported to have anti-HIV activity. The circular RTD-1 has better threefold activity than its linear form. The RTDs-1-3 are synthesized from the ribosome-derived precursor proteins. The precursor proteins have a signal peptide, a pro-domain, and RTD domain. The exact biosynthetic mechanism is not yet known, but it was believed that there are two different genes, *RTD1.1* and *RTD1.2*, that are involved in the biosynthesis of three circular RTDs. RTD-1 is a heterodimer of *RTD1.1* and *RTD1.2*, RTD-2 is a homodimer of *RTD1.1*, and RTD-3 is a homodimer of *RTD1.2* (Selsted, 2004).

2.3.5 Cyclic Peptides from Plants

Plants are the one of the largest known sources of CPs. Many plant CPs have been studied in detail in an effort to understand the CP biosynthesis process.

2.3.5.1 Caryophyllaceae-like Cyclic Peptides (Orbitide)

In 1959, Kaufmann and Tobschirbel discovered cyclolinopeptide A, a cyclic nonapeptide which was isolated from *Linum usitatissimum* (flax) as one of the first natural CPs

isolated from plants. Caryophyllaceae-like CPs are homocyclopeptides with a single ring formed from two or between five and twelve α-amino acids, usually L-isomers (Tan and Zhou, 2006; Morita and Takeya, 2010). As the name suggests, Caryophyllaceae-like CPs are particularly common in the Caryophyllaceae family, but they are also found in nine other families. In the last two decades, considerable efforts have been made to characterize the CPs in various members of the Caryophyllaceae (Ding *et al.,* 1999; Tan and Zhou, 2006). Caryophyllaceae-like CPs provides an excellent model to study ribosome-derived head-to-tail cyclized CPs.

Until recently, very little was known about the biosynthesis of CPs in the Caryophyllaceae plant family. Dr. Covello's laboratory at the National Research Council - Plant Biotechnology Institute (NRC-PBI) provided insight into the biosynthesis of Caryophyllaceae-like CPs. While studying saponin biosynthesis in *S. vaccaria* (Meesapyodsuk *et al.,* 2007), a group of short DNA sequences of ~200 nucleotides was found during the investigation of expressed sequence tags (ESTs) library from the developing seeds. The role of these short DNA sequences was unknown. Further investigation revealed that these sequences encode CP precursors, in which the incipient CP sequences are flanked by highly conserved N- and C-terminal sequences (Condie *et al.*, 2011; Table 2.1).

This information revealed genes that appeared to encode CP precursors (presegetalins) which are subsequently cyclized to mature CPs. To confirm that, presegetalin A1, the precursor for segetalin A, was expressed in transformed *S. vaccaria* roots. The result of this experiment reveals that the transformed *S. vaccaria* roots show the production of segetalin A. In another experiment, an extract of *S. vaccaria* developing seeds was tested with synthetic presegetalin A1, and this lead to segetalin A formation. Thus, under *in vivo* and *in vitro* conditions, formation of mature segetalin A takes place from its linear precursor (presegetalin A1).

Given the observations made on genes encoding CP precursor peptides for CPs found in *S. vaccaria,* it became interesting to investigate related species. Sequence similarity searches with the presegetalin A1 amino acid sequence reveal similar types of precursors from *D. caryophyllus* (Table 2.1b), which may encode the novel CPs *cyclo*[GPIPFYG] and

cyclo[GYKDCC]. Similar types of precursors were also identified from *Citrus* spp. from the plant family *Rutaceae* (Table 2.1c).

Table 2.1 Predicted amino acid sequences of putative CP precursors from (a) *Saponaria vaccaria*, (b) *Dianthus caryophyllus*, and (c) *Citrus* spp. The illustrated CP precursors are devided into three parts N- terminus, incipient CP and C-terminus. The known incipient CP sequences are shown in reverse type, and the predicted CP sequences are in italics and underlined. The presegetalin name and/or GenBank accession numbers are shown in the right-side column (modified from Condie *et al.,* 2011).

N-terminus	Incipeint CP	C-terminus	
a) Saponaria vaccaria			
A class			
MSPILAHDVVKPQ	GVPVWA	FQ AKDVENASAPV	Presegetalin A1
MSPILAHDVVKPQ	GVAWA	FQ AKDVENASAPV	Presegetalin B1
MSPIFAHDVVNPQ	GLSFAFP	AKDAENASSPV	Presegetalin D1
MSPIFAHDVVKPQ	GLSFAFP	AKDAENASSPV	Presegetalin D2
MSPILAHDVVKPQ	GLSFAFP	AKDAENASSPV	Presegetalin D3
MSPIFVHEVVKPQ	GVKYA	FQ PKDSENASAPV	Presegetalin G1
MSPIFAHDIVKPK	GYRFS	FQ AKDAENASAPV	Presegetalin H1
MSPILALDRYKPE	*GRVKA*	FQ AKDAENASAPV	Presegetalin K1
MSPILSHDVVKPQ	*GLPGWP*	FQ AKDVENASAPV	Presegetalin L1
F class			
MATSFQFDGLKPS	FSASYSSKP	IQ TQVSNGMDNASAPV	Presegetalin F1
MATSFQLDGLKPS	*FGTHGLPAP*	IQ VPNGMDDACAPM	Presegetalin J1
b) Dianthus caryophyllus			
MSPNSTRDILKPQ	*GPIPFYG*	FQ AKDAENASVPV	AW697819
MSPNSTRDLLKPL	*GYKDCC*	VQ AKDLENAAVPV	CF259478
c) Citrus spp.			
METTCAGN NWSE	GLLLPPFG	SIADDDVMND NLDFLNVPQYGRNPDYMG	EY850721
MKTLAGAGMSDPSE	GLVLPS	SIADDDVGND NLDLIVIPQYGRNPDYYG	EY848546
MKTLPGAGMSDPSE	GYLLPPS	SIADDDVGND NLDLIVIPQYGRNPDYYG	BB999724
MKIMETTCAGNDDCLE	*GRPWNLA*	SIVDDNVANDVNLDLLAVPQYGRNTDQTG	ABO93452
MKNMETTSAGNDDWLE	*GYVAA*	SIVDDNIANDVNLDLLTVPQYGRNIDQTG	ABO93453
MKNMETTCAGNDDWLE	*GAPWLIAA*	SIVDDNIANDVNLDLLTVPQYGRNIDQTG	ABO93454
MKNMETTSAGNADWLE	*GVPWAIAA*	SIVDDNIANDVNLDLLTVPQYGRNIDQTG	ABO93455

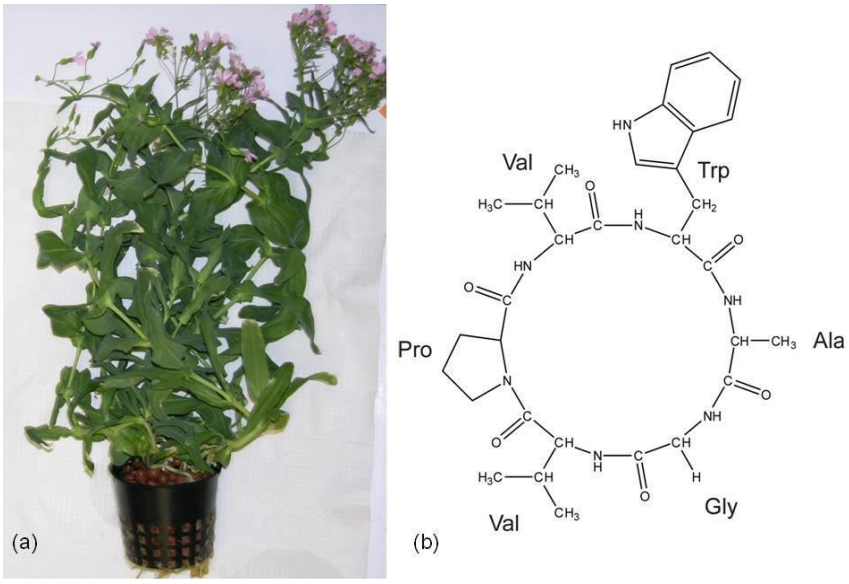

Figure 2.3 (a) *Saponaria vaccaria* plants (var. Pink Beauty, Caryophyllaceae). This plant is also known as cowherb, cowcockle, and cow soapwort. (b) The structure of Caryophyllaceae-like CP segetalin A, which is backbone-cyclized and comprised of six proteinogenic amino acids.

Segetalin A is a homocyclopeptide of the Caryophyllaceae-like class containing six proteinogenic amino acids (Figure 2.3). The segetalins are a group of CPs accumulating in the seeds of *S. vaccaria*. To investigate the biosynthesis of segetalin A, chemically synthesized presegetalin A1 was obtained and tested with various tissue (developing seed, germinating seed, mature seed, root, leaves and flower) extracts of *S. vaccaria* for the enzyme activity in a semi-optimized assay. Amongst all the tissues tested, the developing seeds demonstrated the maximum activity. Based on these results, the developing seeds were chosen for further enzymes purification work.

Recently, Barber *et al.*, 2013 proposed and demonstrated that segetalin A is synthesized from a 32 amino acid linear precursor peptides named presegetalin A1 by two protease-like enzymes called oligopeptidase 1 (OLP1) and peptide cyclase 1 (PCY1) (Figure 2.4). OLP1 is the first enzyme involved in the segetalin A biosynthesis pathway which cleaves presegetalin A1 to form presegetalin A1[1,13] and presegetalin A1[14,32]. The presegetalin A1[1,13] is an

N-terminal peptide and presegetalin A1[14,32] is a substrate for the PCY1. OLP1 has not been identified and characterized yet. PCY1 is the second and most important enzyme in the pathway. It is responsible for catalyzing the cyclization reaction for synthesizing segetalin A from presegetalin A1[14,32]. In principle, PCY1 catalyzes an intra-molecular transpeptidation.

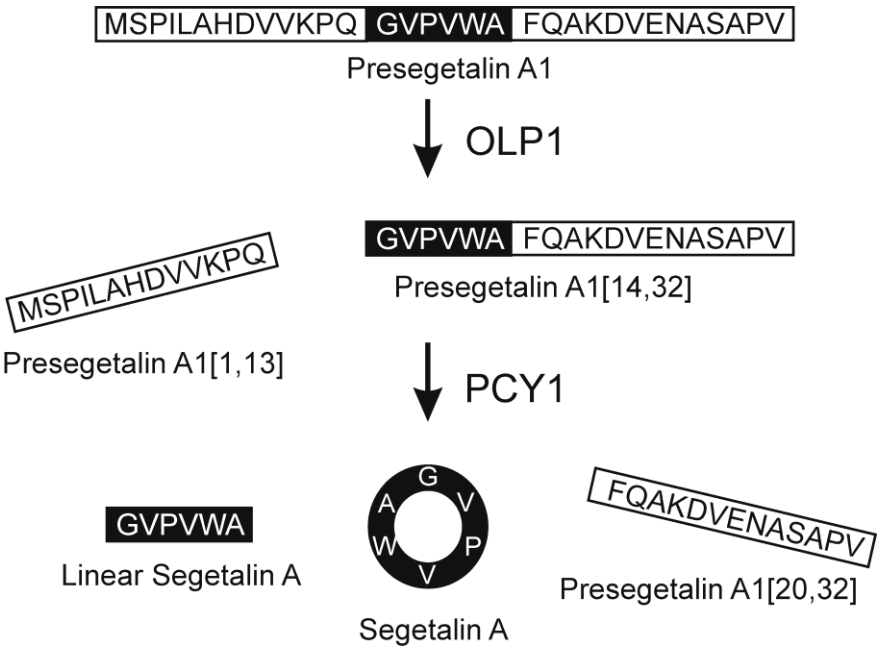

Figure 2.4 Proposed pathway for the processing of presegetalin A1 to segetalin A in *S. vaccaria*. Presegetalin A1 is derived from a monocistronic mRNA by translation on ribosomes. An oligopeptidase activity (OLP1) and a peptide cyclase activity (PCY1) are involved in removal of the N-terminus and cyclization, respectively (Barber *et al.*, 2013). (© National Research Council, Canada; Barber *et al.*, 2013; printed with permission http://www.jbc.org/content/288/18/12500.full?sid=117b51bd-0bc0-4a46-a6ba-38332a2d9ca6).

PCY1 was partially purified from stage-2 developing seeds of *S. vaccaria* by three consecutive chromatography steps: anion exchange, hydrophobic interaction, and gel filtration chromatography. The partially purified enzyme was subjected to SDS-PAGE, and the most prominent bands were excised from the gel and digested with trypsin. The resulting peptides were analyzed by mass spectrometric methods and results were used for MASCOT searches of sequence databases. This information helped to identify the enzyme as prolyl-oligopeptidase-

like enzyme from the serine protease family. A full length cDNA clone encoding an enzyme matching the properties of PCY1 was obtained and expressed in *E. coli*. The purified recombinant PCY1 was tested *in vitro* with its native substrate presegetalin A1[14,32] and found to have cyclase activity. PCY1 is the first cloned plant cyclase of its kind. The present research is an effort toward the characterization of the PCY1.

2.3.5.2 Non-orbitide Cyclic Peptides from Plants

Among all plant CP types, cyclotides are the most studied. They are small disulfide-rich peptides that are characterized by a head-to-tail cyclized peptide backbone and a knotted arrangement of three conserved disulfide bonds. Kalata B1 was the first cyclotide to be discovered. Comprised of 29 amino acids, it was discovered in the 1960s and named after the indigenous African medicine 'kalata-kalata' (Pelegrini *et al.,* 2007; Craik, 2009). Generally, cyclotides are more abundantly expressed in three families: *Violaceae* (violet), *Rubiaceae* (coffee), and *Cucurbitaceae* (cucurbit), where they play a role in plant defense against insect predation (Conlan *et al.,* 2010). Until now, more than 200 sequences have been determined, but it is estimated that these plant families comprise many thousands of members. Cyclotides are classified into two major subfamilies: Möbius and bracelet. A third group of cyclotides is the trypsin inhibitor subfamily. This group is also referred as cyclic knottins, and its members are homologous to a family of non-cyclic trypsin inhibitors from squash plants (Mulvenna *et al.*, 2005).

Cyclotides are gene-encoded proteins, synthesized as precursor proteins and may be processed by a putative protease-like enzyme. The cyclotide precursors from the *Violaceae* and *Rubiaceae* plant families are exclusively dedicated to cyclotide production, but there are other plant families like *Fabaceae* where chimeric precursors encode cyclotides and albumin together, similar to the sunflower trypsin inhibitor (SFTIs) in sunflower seeds (Mylne *et al.,* 2011). A well-characterized member of the cyclotide group is kalata B1, which is produced from its precursor protein *Oak1* (*Oldenlandia affinis* kalata B1 gene). The precursor protein contains an endoplasmic reticulum signal sequence, a conserved N-terminal repeated sequence,

either one or multiple copies of the mature peptide sequences, and a short hydrophobic C-terminus (Conlan *et al.,* 2010; Conlan *et al.,* 2012; Figure 2.5).

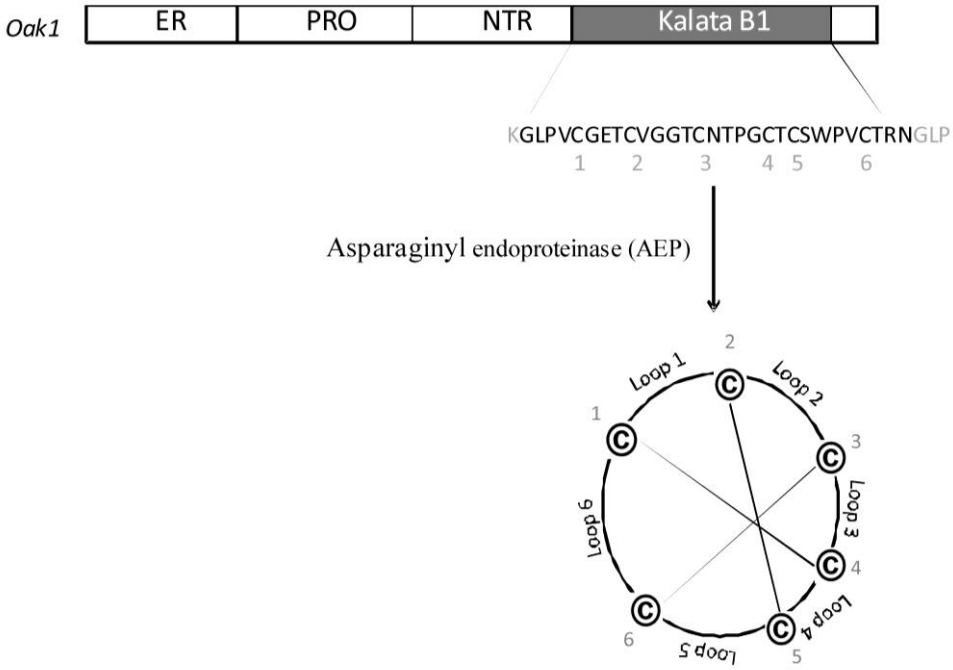

Figure 2.5 Schematic representations of precursor proteins of cyclotide (kalata B1) and its putative processing in *O. affinis*. The amino acid sequence of kalata B1 is shown with the cysteine residues labeled with numbers. The cleavage sites for excision of a mature domain with an N-terminal glycine and a C-terminal asparagine, which are subsequently linked by a putative asparaginyl endopeptidase, are indicated by extended lines.

A putative asparaginyl endoproteinase (AEP), a cysteine protease, is believed to be associated in excision of peptides from a precursor and a ligation of termini (Saska *et al.,* 2007). The AEP performs an exclusive action on highly conserved asparagine or aspartate residue at the C-terminus to excise the mature CP sequence from its precursor. cDNAs encoding cyclotide precursors were isolated from *Oldenlandia affinis*. When a corresponding precursor gene is expressed in tobacco, mature cyclotides are produced (Gillon *et al.,* 2008). The presence of an ER signal on a precursor is believed to be an indication that the folding process takes place in the endoplasmic reticulum before the AEP processing of the precursor. Recently Conlan *et al.,*

2011 reported that the cyclization reaction might take place in vacuoles as cyclotides are targeted to the vacuoles in plant cells.

2.4 Non-ribosomal Peptide Synthetases (NRPSs)

NRPS is a multienzyme complex with the potential of generating structurally diverse CPs. (Lipmann *et al.,* 1971) Lipmann reported a ribosome-independent biosynthesis of the cyclic peptides gramicidin S and tyrocidine by a multienzyme system in *Bacillus brevis*. NRPSs are competent to incorporate proteinogenic, non-proteinogenic and modified amino acids, as well as fatty acids and heterocyclic rings as part of the CP backbone. Apart from their ability to include different types of building blocks, NRPSs perform a variety of cyclization reactions to produce structurally diverse CPs as end products. For example, head-to-tail cyclization (cyclosporin A and tyrocidine), lipo-branched chain cyclization (surfactin and mycosubtilin), amino acid branched cyclization (bacitracin and fengycin), and oligomerization (gramicidin S and enterobactin ; Sieber and Marahiel, 2003). The NRPSs consist of multiple subunits, which can be further categorized by modules and domains.

In general, there are four types of domains within NRPSs that participate in CP biosynthesis. The biosynthetic process begins with the A domain (~550 amino acids) which is primarily responsible for the activation of substrates by adenylation with the help of Mg^{+2}-dependent ATP hydrolysis. The A domain also plays a vital role in the recognition and selection of substrates, as NRPS lack the template to direct the incorporation of building blocks in a sequence like ribosome-dependent peptide synthesis. After analysis of many A domains from different organisms, it was possible to make a so-called non-ribosomal code for the prediction of the substrate binding based on A domain sequences (Stachelhaus *et al.,* 1999; Challis *et al.,* 2000). The activated amino acid from the A domain is transferred to the T domain (~80 amino acids), which is also known as the thiolation domain. The T domain has an invariant serine residue attached to a cofactor phosphopantetheine. The binding of the cofactor with serine activates the T domain and thereby regulates the activity.

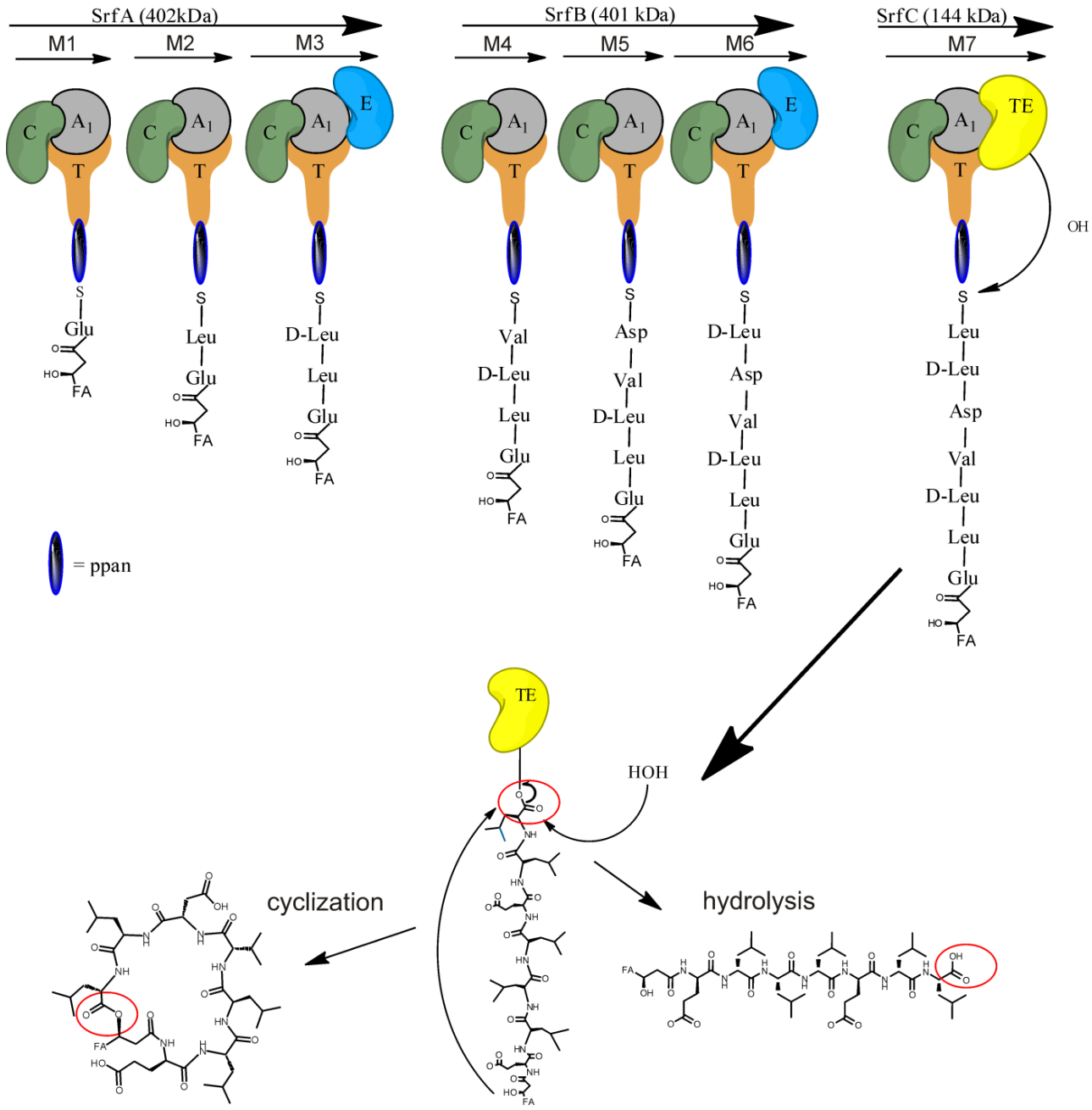

Figure 2.6 Schematic representation of the surfactin biosynthesis by NRPSs. The four domains (A, E, T and TE) of NRPS participate in this biosynthesis process. A phosphopantetheine (ppan) is a cofactor in reaction. SrfA, SrfB and SrfC are three complexes, and M1 to M7 represents the seven modules. The TE domain process the intermediate acyl-*O*-TE enzyme to either linear or cyclic product, depending on attack by the water or the internal nucleophile respectively (modified from Sieber and Marahiel, 2003).

The free thio group of phosphopantetheine binds to the adenylated substrate from the A domain. The structure of the T domain revealed that there is an absence of a binding pocket, so the T domain lacks specificity toward the incoming substrates (Weber *et al.,* 2000). Peptide bond formation is catalyzed by the C domain (~450 amino acids) through condensation of the two adjacent substrates. Detailed analysis of C domains indicates that they are highly specific for certain aminoacyl thioester substrates (Belshaw *et al.,* 1999; Clugston *et al.,* 2003). The TE domain (~280 amino acids) is present in the last module of NRPSs. The C domain transfers the growing peptide chain to the TE domain.The TE domain possesses a catalytic triad of serine, histidine, and aspartic acid, where serine makes a tetrahedral intermediate, acyl-*O*-TE enzyme, by a nucleophilic attack on the growing peptide chain (Kohli *et al.,* 2001; Bruner *et al.,* 2002; Tseng *et al.,* 2002). This acyl-intermediate can be processed in two different directions, either it is attacked by the internal nucleophile to produce a cyclic product or by water (hydrolysis) to yield a linear product (Figure 2.6). The TE domain is also described as a peptide cyclase due to its ability to catalyze the cyclization reaction. In addition, the NRPSs also consist of an optional E domain, also known as an epimerization domain, which is responsible for the incorporation of D-amino acids in the growing peptide chain.

2.5 Artificial Systems for Cyclic Peptide Production

2.5.1 Intein-mediated Cyclization

Inteins are also known as 'protein introns'; they serve a valuable biological function by controlling protein activation. Inteins ensure that an enzyme is first synthesized in an inactive form, then is activated by self-splicing. There are three types of intein-splicing: intermolecular trans-splicing, canonical splicing and cyclization using permuted inteins. The first two types of intein-splicing generate linear products, while a third type is capable of forming a cyclic product. Scott *et al.,* 1999 reported an artificial *in vivo* system based on intein-splicing to produce pseudostellarin F, a Caryophyllaceae-like plant CP. A chemically synthesized DNA fragment encoding pseudostellarin F was inserted into Ssp and a DnaE split intein fragment I_C and I_N to construct a plasmid. This plasmid was expressed in *E. coli,* and the pseudostellarin F

was produced at the level of 30 µg/g wet cell mass. The molecular mechanism in CP formation involves the following steps. In the first step, the expressed precursor folds to form a protein ligase which performs an N-to-S acyl shift to form a TE intermediate. Second, the TE intermediate undergoes a transesterification process to form a lariat intermediate. In the third step, formation of a succinimide releases the CP as a lactone. In the fourth and last step, an X-to-N acyl shift takes place to generate CP (Xu *et al.,* 2011).

Scott *et al.,* 2001 further explored the intein-splicing system to generate CP libraries. According to their reported work, I_{C+1} and I_{N-1} are crucial positions and are occupied by specific amino acids. The I_{C+1} position, which is adjacent to C-intein, can only be occupied by serine, threonine, and cysteine. These amino acids provide the possibility of nucleophilic transesterification. The I_{N-1} position, which is adjacent to N-intein, is a less conservative position compared to the I_{C+1} position. The I_{N-1} position can be occupied by polar (Ser), non-polar (Ala), large (Leu), small (Gly), charged (Lys), or aromatic (Tyr) residues (Scott *et al.,* 2001). With this intein-splicing system, four to nine amino acids CPs were produced (Deschuyteneer *et al.,* 2010).

2.5.2 Sortase-mediated Cyclization

Sortases, classified as cysteine proteases, are an important class of transpeptidases which can be employed for CP production by intra-molecular transpeptidation. They are involved in the cell wall biosynthesis process of Gram-positive bacteria. Sortases are mainly responsible for anchoring surface proteins to the bacterial cell wall by a specific reaction called the 'sorting reaction' (Mazmanian *et al.*, 1999). A specific sorting signal on the C-terminus of the target protein is required to perform the sorting reaction. For example, Sortase A (Srt A) from *Staphylococcus aureus* requires a pentapeptide (LPXTG) as the sorting signal on the C-terminus and Gly on the N-terminus. Wu *et al.,* 2011 reported CP and glyco-CP production from linear peptide precursors by Srt A (Figure 2.7). The linear peptide precursor was converted into five different products by Srt A: linear dimer, linear trimer, CP monomer, CP dimer, and CP trimer. The quantity of each product depended on enzyme concentration,

substrate concentration, and most importantly, the length of the precursor peptide. It was concluded that the linear precursor peptide should be 11 amino acids or longer, excluding the C-terminus signal, for viable monomeric CP production. The 9-amino acid substrate, which was glycosylated with N-acetylgalactosamine at three residues, was also converted into dimeric and trimeric CP by Srt A, but no monomer was observed. Production of CPs by Srt A has two drawbacks. First, this system cannot be utilized for smaller CP production (<10 amino acids), and second, the sorting signal (LPXTG) becomes an integral part of the CP which may affect its structure and biological activity.

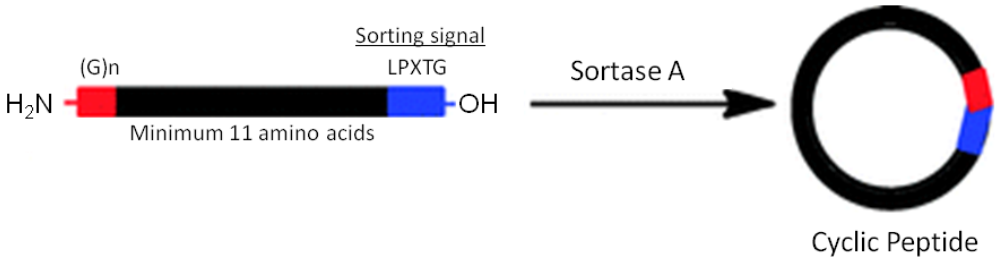

Figure 2.7 Schematic representations of the strategy for CP production with help of sortase A. To produce a CP monomer in relatively high quantity, the sortase A requires a minimum of 11 amino acids excluding the N- terminal Gly and C- terminal sorting signal [LPXTG], (modified from Wu *et al.*, 2011).

2.6 Biological Activities and Applications of Cyclic Peptides

Classes of CPs have shown a wide range of biological activity. The cyclotides are one of the most explored classes of CPs for their biological activity and applications in the agriculture and pharmaceutical industry. Cyclotides have been reported to have anti-HIV, cytotoxic, anti-microbial, immunosuppressant, nematocidal, anti-insecticidal, and haemolytic qualities. The mode of action for antimicrobial activity of cyclotides is not clearly understood yet, but it has been demonstrated that it creates membranes pores which ultimately result in the leakage of vesicular contents (Huang *et al.*, 2009). The superior stability (thermal, chemical,

and biological) and unique structure (cross-linking and other modifications) provide an opportunity for cyclotides to be potential drug scaffolds for protein-based drugs (Daly and Craik, 2009; Craik *et al.,* 2012). Generally, the Caryophyllaceae-like CPs are small in size and have simple chemical structures. *Saponaria vaccaria* from the Caryophyllaceae family possesses a group of cyclic peptide called segetalins. The native biological function of segetalins in plants is not known, but segetalins A, B, G, and H have shown estrogen-like activity (Itokawa *et al.,* 1995). Segetalin F was reported to have vasorelaxant activity (Morita *et al.,* 2006). Recently, two CPs, *cyclo*[GPLTLF] and *cyclo*[GPVTIF], have been reported in *Dianthus superbus* with osteoblastic activity (Tong *et al.,* 2012). Tunicyclin D from *Psammosilene tunicoides* showed a broad spectrum of antifungal activity against *Candida albicans, Candida tropicalis, Candida parapsilosis,* and *Cryptococcus neoformans* (Tian *et al.,* 2010). Cyclolinopeptide A has exhibited immunosuppressive activity by the same mechanism as cyclosporin A and FK-*506* (Siemion *et al.,* 1999). CPs are also potential therapeutic agents for skin disorders (Namjoshi and Benson, 2010), which expands the possibility of CP applications in cosmeceuticals. Labaditin, a CP isolated from the plants of the *Euphorbiaceae* family, has showed an antibacterial effect on Gram-positive bacteria (Barbosa *et al.,* 2011). Three CPs—psychrophilin D, cycloaspeptide A and cycloaspeptide D—have been isolated from psychrotolerant fungus *Penicillium algidum.* Psychrophilin D has shown anticancer activity, while the cycloaspeptide A and cycloaspeptide D have exhibited moderate activity against *Plasmodium falciparum* (Dalsgaard *et al.,* 2005). With information about the advantages of cyclic structure, the ends of a naturally linear conotoxin were connected to design a circular backbone which resulted in improved biological activity, increased resistance to proteases, improved half-life in plasma, and interestingly allowed oral administration (Clark *et al.,* 2010). CPs have been also explored for the treatment of disease like rheumatoid arthritis, multiple sclerosis and Alzheimer's disease, as well as development of immunotherapeutic vaccine for diabetes and autoimmune encephalomyelitis (Katsara *et al.,* 2006).

3. RATIONALE AND OBJECTIVES

3.1 Experimental Rationale

The first Caryophyllaceae-like CP was discovered in 1950s from flax (Kaufmann and Tobschirbel, 1959). Since then, many of them were discovered and their structures have been identified, but their biosynthesis remained unknown until recently. Condie *et al.*, 2011 has provided evidence about the biosynthesis of segetalin A (a six membered backbone-cyclized CP) from a ribosome-derived precursor in *S. vaccaria*, a member of Caryophyllaceae plant family. In continuation to this research, Barber *et al.*, 2013 has suggested that segetalin A1 is synthesized from its ribosome-derived precursor, presegetalin A1, by two protease-like enzymes. The second protease-like enzyme has ability to perform a cyclization reaction. This enzyme was identified, cloned and named PCY1. The rationale behind the present research is that characterization of PCY1 should provide detailed insight into CP biosynthesis in plants, and explore the substrate specificity of PCY1, with a view to understanding its possible biotechnological applications in the future.

3.2 Specific Aims

#1. Development of optimized assay conditions for PCY1.

#2. Prediction of PCY1 structure by homology modeling.

#3. Exploration the substrate specificity of PCY1 with broad range of substrates.

#4. Proposal for catalytic mechanism of PCY1.

4. MATERIALS AND METHODS

4.1 Materials and Distributors

The following table details the materials used in the study. The distributors of the materials used are listed in Table 4.1 and their addresses are listed in Table 4.2. Tables detailing the media and buffer composition are listed in Table 4.3 and 4.4.

Table 4.1 List of reagents and materials used in the study.

Material	**Distributor Name**
Acetonitrile	Fisher Scientific
Amicon Ultra-15 devices	Millipore
Ammonium formate	Sigma-Aldrich Corporation
Ampicillin sodium salt	Sigma-Aldrich Corporation
Bicine	Sigma-Aldrich Corporation
B-Per® bacterial protein extraction reagent	Pierce
Bromphenol blue	Sigma-Aldrich Corporation
BSA	Sigma-Aldrich Corporation
Competent *E. coli* BL21-AI™	Invitrogen
Cupric sulfate	Fisher Scientific
Dithiothreitol (DTT)	Sigma-Aldrich Corporation
Formic acid	Acros Organics
Glycerol	Anachemia Canada Inc.
Glycine	Fisher Scientific
HisPur™ cobalt resin	Pierce
Imidazole	Sigma-Aldrich Corporation

Methanol	Fisher Scientific
Micro BCA protein assay kit	Pierce
3-(*N*-morpholino)propanesulfonic acid (MOPS)	Sigma-Aldrich Corporation
2-(*N*-morpholino)ethanesulfonic acid (MES)	Sigma-Aldrich Corporation
N-Cyclohexyl-2-aminoethanesulfonic acid (CHES)	Sigma-Aldrich Corporation
N-cyclohexyl-3-aminopropanesulfonic acid (CAPS)	Sigma-Aldrich Corporation
Oriole™ fluorescent gel stain	Bio-Rad
pDEST™ 17 expression vector	Invitrogen
Precision Plus Protein™ molecular weight standards	Bio-Rad
10% Ready Gel® precast polyacrylamide mini-gel	Bio-Rad
Segetalin A standard	John Balsevich lab. (NRC)
Sodium chloride	EMD Chemical Inc.
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich Corporation
Synthetic peptides	Bio Basic Inc.
Tris base	Fisher Scientific
Tryptone	BD Biosciences
Yeast extract	BD Biosciences
Zorbax 300 EXTEND-C18 column	Aligent technologies

Table 4.2 List of Distributor name and addresses.

Distributor Name	**Distributor Address**
Acros Organics	Acros Organics, New Jersey, USA
Aligent technologies	Aligent Technologies, Mississauga, Ontario,Canada
Anachemia Canada Inc.	Anachemia Canada Inc., Montreal, Quebec, Canada
BD Biosciences	BD Biosciences, Franklin Lakes, New Jersey, USA
Bio Basic Inc.	Bio Basic Inc., Markham, Ontario, Canada
Bio-Rad	Bio-Rad, Mississauga, Ontario,Canada
EMD Chemical Inc.	EMD Chemical Inc., Gibbstown, New Jersey, USA
Fisher Scientific	Fisher Scientific, Fair Lawn, New Jersey, USA
Invitrogen	Invitrogen-Life Technologies, Carlsbad, CA, USA
John Balsevich lab. (NRC)	NRC, Saskatoon, Saskatchewan, Canada
Millipore	Millipore, Billerica, Massachusetts, USA
Pierce	Pierce, Rockford, Illinois, USA
Sigma-Aldrich Corporation	Sigma-Aldrich, Oakville, Ontario, Canada

Table 4.3 List of media used in the study and their compositions.

Medium	**Composition**
Auto-induction medium	TTB Broth (1L) Tryptone 12 g Yeast Extract 24 g Glycerol 4 ml Adjust with ddH_2O to 1 L and autoclave OnEx Solution 1 (50x) Glucose 2.5% (w/v) α-lactose monohydrate 10% (w/v) Filter sterilize with 0.2 µm filter and store at room temperature OnEx Solution 2 (20x) Na_2HPO_4 1 M KH_2PO_4 1 M $(NH4)_2SO_4$ 0.5 M Autoclave and store at room temperature OnEx Solution 3 (1000x) $MgSO_4$ 1 M Autoclave and store at room temperature Just prior to use, add above OnEx solutions and antibiotics to TTB broth accordingly.
LB Broth	LB Broth (1L) Tryptone 10 g Yeast Extract 5 g NaCl 10 g Adjust with ddH_2O to 1 L and autoclave

Table 4.4 List of buffers used in the study and their composition.

Buffer	**Composition**
Equilibration/Wash buffer	50 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 mM imidazole
Elution buffer-1	50 mM Tris-HCl pH 7.5, 300 mM NaCl, 150 mM imidazole
Elution buffer-2	50 mM Tris-HCl pH 7.5, 300 mM NaCl, 300 mM imidazole
Laemmli sample buffer	200 mM Tris-HCl pH 6.8, 4% SDS, 0.2% bromphenol blue, 200 mM dithiothreitol, 40% glycerol (v/v)
Electrophoresis buffer	25mM Tris-HCl pH 7.5, 250 mM glycine, 0.1% SDS (w/v)

4.2 *E. coli* Expression of PCY1

The *pcy1* gene from *S. vaccaria* was expressed in *E. coli* and purified as follows. As indicated in Barber *et al.*, 2013, the *pcy1* ORF was combined with the Gateway expression vector pDEST^{TM}17 (Figure 4.1) and the resulting plasmid was named pCB008. The pC008 plasmid was transformed into competent *E. coli* BL21-AI™. An overnight 1 ml LB culture of pCB008/BL21-AI was used to inoculate 100 ml of auto-induction medium (Studier, 2005) and incubated at 37 °C with shaking until an A600 of 0.4 was reached. Arabinose was then added to 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 (Barber *et al.*, 2013).

It is noteworthy that, Dr. Covello's lab had identified three homologues of *S. vaccaria* PCY1 from two other species of Caryophyllaceae family. Two homologues were identified from *D. superbus* (contig 250 and contig 1141) and the third homologue was identified from *S. vulgaris* (contig 150). These homologues were also expressed in *E. coli* as explained above. This thesis describes the characterization of *S. vaccaria* PCY1. The other homologues were not

characterized completely, except they were tested for a few substrates as mentioned in section 5.3, 5.4.3.1 and 5.4.3.2. Unless otherwise noted, henceforth the PCY1 refers to *S. vaccaria* PCY1 in the text. The PCY1 homologues from other species are mentioned with their species name in the text, for example, *D. superbus* (c250) PCY1 refers to the PCY1 homologue, corresponding to contiguous cDNA sequence 250 (contig 250) from *D. superbus.*

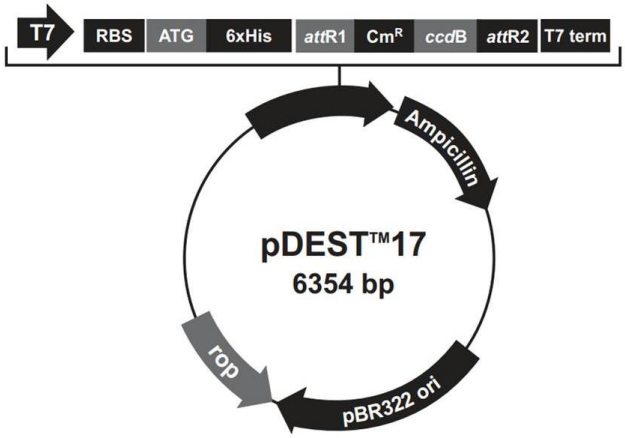

Figure 4.1 The map of pDEST^{TM}17 expression vector. The pDEST^{TM}17 was used for the expression of *pcy1*in *E. coli* BL21-AI™. The *pcy1* ORF was inserted downstream of the T7 promoter for high level expression. The product of the plasmid includes 6xHis sequence on N-terminal to assist purification of expressed protein by affinity chromatography.

4.3 Purification of PCY1 by Affinity Chromatography

The frozen *E. coli* pellets (section 4.2) were resuspended in chilled 500 µl of B-Per® bacterial protein extraction reagent and 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 volume equilibration buffer. Then, 250 µl of HisPur^{TM} cobalt resin (Pierce) was added and allowed for 1 h binding at 4 °C for purification of PCY1 by batch adsorption method. The resin was washed five times by washing buffer and elution at 150 and 300 mM imidazole (Elution buffer 1 and 2), respectively, according to the manufacturer's recommendations. Each eluate

was concentrated to 150 µl and desalted by spin dialysis (Amicon Ultra-15 devices) following the manufacturer's protocol and subsequently assayed for peptide cyclase activity. The metal affinity-purified active PCY1 samples were mixed 1:1 with SDS-PAGE Laemmli sample buffer and heated at 99 °C for 5 min. The samples were subjected to SDS-PAGE under denaturing conditions using electrophoresis buffer and a 10% Ready Gel® precast polyacrylamide mini-gel. Precision Plus Protein™ molecular weight standards were loaded on the same gel. The gel was stained with Oriole™ fluorescent gel stain for 15 h. The protein bands were visualized by UV illumination, and the purity of recombinant PCY1 was determined by SDS-PAGE to be 90% (Figure 5.6; Barber *et al.*, 2013).

4.4 Protein Quantification Assay

For enzyme assays and purification, protein was measured using a modified micro BCA protein assay. Ten µl samples were mixed with 100 µl of BCA working reagent and incubated at 60 °C for 30 min, after which optical density at 562 nm was recorded. BSA solutions (0.5 to 20 µg/ml) were used to plot a standard curve to quantify the unknown amount of PCY1.

4.5 Assay of Recombinant PCY1

The PCY1 assay contained 20 mM Tris (pH 8.5), 100 mM NaCl, 5 mM DTT, 0.2 mg BSA, and 15 µg/ml of presegetalin A1[14,32] or other synthetic peptides. The reaction was initiated by the addition of 0.3 µg of recombinant PCY1, in a total reaction volume of 100 µl. The assay was incubated at 30 °C for up to 1 h and stopped by the addition of 0.9 ml of methanol. The 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. Controls were performed with either omission of enzyme or stopping the reaction at 0 h. Where necessary, 0-h peak areas were subtracted from 1-h peak areas for the linear product (Barber *et al.*, 2013).

4.6 Quantification of Segetalin A and Linear Segetalin A

Ssegetalin A (*cyclo*[GVPVWA]) standard was obtained from J.J. Balsevich, NRC, Saskatoon. The compound was purified by HPLC up to >95%. Known amounts of segetalin A from 0.2 ng to 2.0 ng were subjected to ion trap LC/MS. The obtained peak areas from LC/MS were used to plot the standard curve for segetalin A quantification (Figure 4.2a).

The linear segetalin A standard was obtained from Bio Basic Inc., Ottawa with the purity of >95%. Known amount of linear segetalin A from 0.1 ng to 2.5 ng were subjected to ion trap LC/MS. The obtained peak areas from LC/MS were used to plot the standard curve for linear segetalin A quantification (Figure 4.2b).

4.7 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 electrospray ionization (ESI) Tuning Mix (Agilent installation guide G2431A). The mass spectrometer was scanned in them *m/z* range of 50–2200 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 × 2.1 mm, 3.5-µm 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 cyclic and linear products, from the assays with recombinant PCY1 and synthetic peptides, were detected as $[M+H]^+$, $[M+Na]^+$ and $[M+K]^+$ ions (Barber *et al.,* 2013).

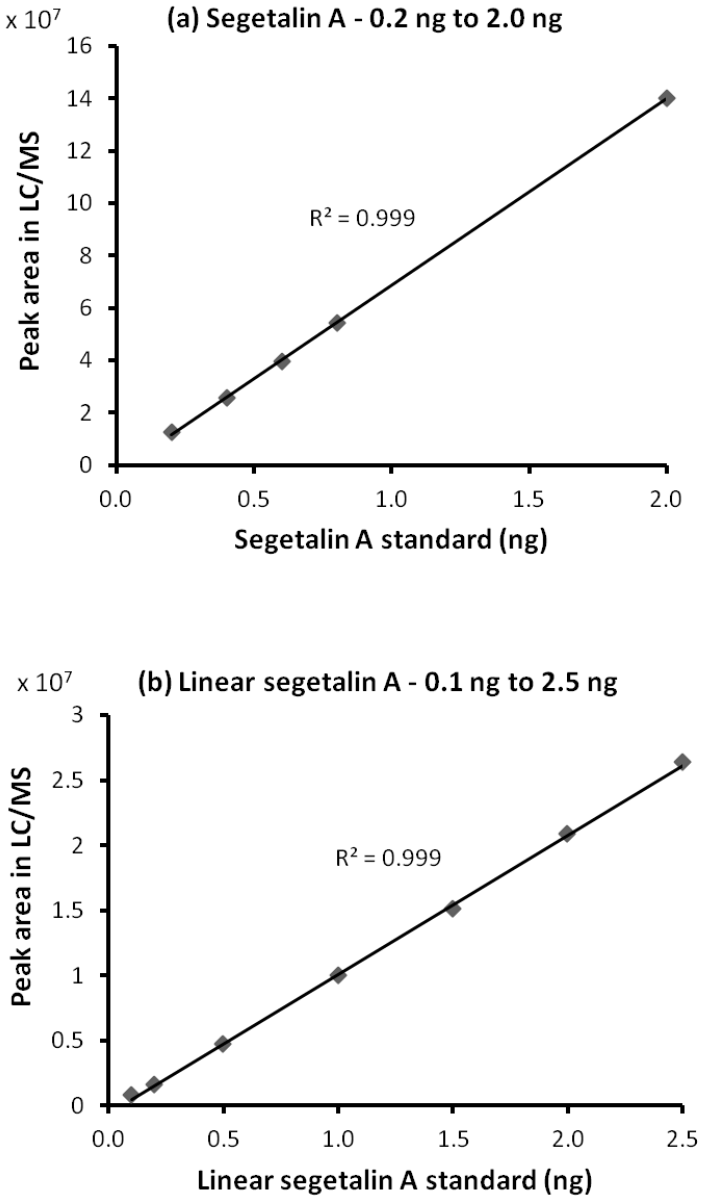

Figure 4.2 Standard curves for quantification of (a) the segetalin A, between 0.2 and 2.0 ng and (b) the linear segetalin A, between 0.1 and 2.5 ng.

4.8 Homology Modelling of PCY1

The three-dimensional structure of PCY1 was predicted using Modeller (version 9.10). The Modeller is a software programme which predicts protein structure by the homology modelling technique, in which an experimentally determined protein structure acts as template to predict the structure of similar protein sequences (Sali and Blundell, 1993). A sequence similarity search was performed using BLASTP with the PCY1 sequence. In BLASTP results, porcine (*Sus scrofa*) muscle prolyl oligopeptidase (POP) was found to be similar to the PCY1. In PDB database, 1QFS (GenBank™ accession number AAA31110) is a porcine muscle POP with 49% sequence identity to *S. vaccaria* PCY1. The POP structure was determined by X-ray crystallography with covalently bound inhibitor Z-Pro-Prolinal (Fulop *et al.*, 1998). To predict the PCY1 structure using Modeller, the input alignment file (.aln) was prepared from BLASTP results. The 1QFS structure was downloaded from the PDB and used for the preparation of other input files (.atm and .py). The automodel program script for 3D model generation was executed with the command >mod9.10 model.py. The final model was selected for further analysis on the basis of discrete optimized protein energy (DOPE) energy evaluation of prepared models (Barber *et al.*, 2013).

4.9 Statistical analysis

Results obtained from the triplicate PCY1 assays (section 5.2.1, 5.2.2, 5.2.3, 5.4.2.2 and 5.5) were analyzed by ANOVA followed by Duncan or Dunnett' C *post-hoc* tests using the statistical package IBM® SPSS® 20.0. These tests perfom multiple comparisons between the variants to determine significant differences. Duncan test and Dunnett' C test were used for homogeneous and non-homogeneous variants, respectively.

5. RESULTS

It is notworthy that the results presented in this thesis were published in Barber *et al.*, 2013. Carla Barber and Darwin Reed, technicians in Dr. Covello's lab, were primarily responsible for identification and cloning of PCY1. The presented results in this section were obtained and analyzed by the author. This section mainly focuses on prediction of *S. vaccaria* PCY1 structure using homology modelling technique and characterization of PCY1. Characterization work includes optimization of assay conditions and also explores the substrate specificity as it relates to enzyme mechanism and biotechnological applications of PCY1.

5.1 Sequence Analysis and Homology Modeling of PCY1

A BLASTP search of GenBank™ with the predicted amino acid sequence of *Pcy1* revealed the greatest sequence identity with members of the enterase lipase superfamily (COG1505) (Akoh *et al.*, 2004). In particular, PCY1 shows the highest amino acid sequence identity to predicted gene products from *Vitis vinifera* (Grape; GenBank™ accession number CAN70125) and *Populus trichocarpa* (Poplar; GenBank™ accession number XP_002306966; see Figure 5.2 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; Protein Data Bank code 1QFS; GenBank™ accession number AAA31110), which shows 49% sequence identity to PCY1 and for which a crystal structure has been determined in the presence of a covalently bound inhibitor (Fulop *et al.*, 1998). In general, POP are involved in important functions such as maturation of peptide hormones and neuropeptides, and also associated with depression and blood pressure maintenance. In a few instances, POP are also known as post-proline cleaving enzymes. POP readily cleave the scissile peptide bond between P1 and P1` positions in small peptides, provided the P1 position is occupied by a proline. 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; Figure 5.1a). The structural alignment of PCY1 and porcine muscle POP showed a root mean square deviation of Cα 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. The putative catalytic triad is located in the α/β hydrolase domain and faces the β-propeller domain. The β-propeller domain is believed to be the entrance of substrate and it is made up of 7 fold repeats of 4-stranded antiparallel β-sheets.

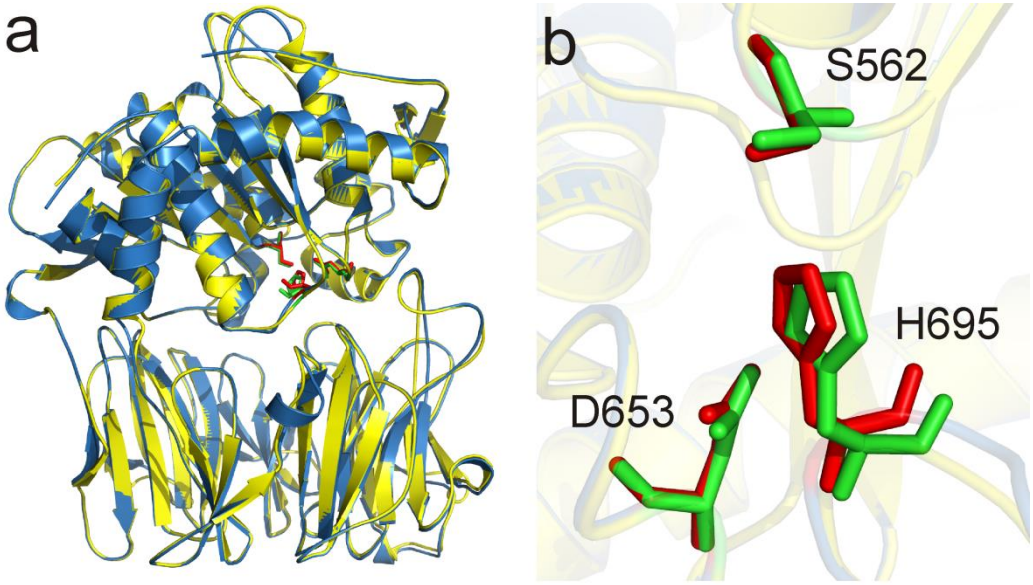

Figure 5.1 The predicted structure of PCY1 based on homology modeling using porcine muscle POP. (a) Structural alignment of a homology model of PCY1 (yellow) and porcine muscle prolyl oligopeptidase (blue). The top half of the structure represents the catalytic domain including the putative catalytic 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 POP (red) (© National Research Council, Canada; Barber *et al.*, 2013; printed with permission http://www.jbc.org/content/288/18/12500.full?sid=117b51bd-0bc0-4a46-a6ba-38332a2d9ca6).

Also, the model is consistent with the identity of Ser^{562}, Asp^{653}, and His^{695} (*S. vaccaria* PCY1 numbering) as members of a serine peptidase-like catalytic triad, which shows excellent three-dimensional alignment with those of porcine muscle POP (Figure 5.1b). Ser^{562} occurs within the sequence GGSNGG, which is conserved among α/β hydrolase and likely forms a "nucleophilic elbow" (Ollis *et al.*, 1992). Despite the alignment of catalytic amino acids, there is a notable single amino acid insertion/deletion near His^{695} (Figure 5.2). The presence of small

residues like glycine in close proximity of Ser^{562} may help to avoid steric hindrance during catalysis involving a bulky nucleophile (the N-terminus of the substrate). The putative structure of catalytic triad also suggests that the –OH group of Ser^{562} is accessible and exposed towards imidazole group of His^{695}, which deprotonates the Ser^{562}. The deprotonated Ser^{562} likely carries out the nucleophilic attack on substrate during catalysis to form an acyl intermediate. The detailed catalytic mechanism of PCY1 is explained in section 6.3.

```
Vvi EYLVSAGYTQPRKLCIEGGSNGGLLVGACINQRPDLFGCALAHVGVMDMLRFHKFTIGHA 608
Ptr EYLVSAGYTQPKKLCIEGGSNGGLLVGACINQRPDLFGCALAHVGVMDMLRFHKFTIGHA 610
Gma EYLVSTGYTQPKKLCIEGGSNGGLLVGACINQRPDLFGCALAHVGVMDMLRFHKFTIGHA 604
Csa EYLISAGYTQPSKLCIEGGSNGGLLVGACINQRPDLFGCALAHVGVMDMLRFHKFTIGHA 608
Aly EYLVSAGYTQPRKLCIEGGSNGGILVGACINQRPDLFGCALAHVGVMDMLRFHKFTIGHA 608
Car KYLISTGYTQPRKLCIEGGSNGGLLVGASINQRPDLFGCALAHVGVMDMLRFHKFTIGHA 608
Hvu EQLISFGYTSSKKICIEGGSNGGLLIAACINQRPDLFGCALAHVGVMDMLRFHKFTIGHA 653
Zma EFLISNGYTSSRRLCIEGGSNGGLLVAASINQRPDLFGCALAHVGVMDMLRFHKFTIGHA 648
Sva EYLISSGYTKARRVAIEGGSNGGLLVAACINQRPDLFGCAEANCGVMDMLRFHKFTLGYL 602
Smo EYLIKEGYTQPRKLGIEGGSNGGLLVAACINQRPDLYGCALAHVGVMDMLRFHKFTIGHA 591
Ppa EYLVKEGYTQSKKLCIEGGSNGGLLVAACINQRPDLYGCALAHVGVMDMLRFHKFTIGHA 614
Bta EYLIKEGYTSPKRLTINGGSNGGLLVATCANQRPDLFGCVIAQVGVMDMLKFHKYTIGHA 594
Ssc EYLIKEGYTSPKRLTINGGSNGGLLVATCANQRPDLFGCVIAQVGVMDMLKFHKYTIGHA 594
Oga EYLIKEGYTSPRRLTINGGSNGGLLVAACANQRPDLFGCVIAQVGVMDMLKFHKYTIGHA 594
Rno EYLIKEGYTTSKRLTINGGSNGGLLVAACANQRPDLFGCVIAQVGVMDMLKFHKFTIGHA 594
Gga KYLIREGYTAPKKLTINGGSNGGLLVAACANQRPDLFGCVIAQVGAMDMLKFHKYTIGHA 594
Xtr EYLVKEGYASAEKITINGGSNGGLLVAACANQRPELFGCAIAQVGVMDMLKFHKFTIGHA 596
Dre EYLIKEGYTSPKKLTINGGSNGGLLVAACVNQRPDLFGCAVAQVGVMDMLKFHKFTIGHA 593
    : *:  **: . :: *:******:*:.:. ****:*:**. *: *.****:***:*:*:

Vvi WTSDYGCSEKEEEFHWLIKYSPLHNVRRPWEQSPDQPSQYPATMILTADHDDRVVPLHSL 668
Ptr WTSDYGCSDKKEEFDWLIKYSPLHNVRRPWEQHPEQPSQYPPTMLLTADHDDRVVPLHSL 670
Gma WTSDYGCSDKEEEFHWLIKYSPLHNVRRPWEQHHDQSFQYPSTMLLTADHDDRVVPLHTL 664
Csa WTSDYGCSDNEEEFKWLIKYSPLHNVKRPWEQHPDRLLQYPSTMLLTADHDDRVVPLHSL 668
Aly WTSEFGCSDKEEEFHWLIKYSPLHNVKRPWEQKTDRFVQYPSTMLLTADHDDRVVPLHTY 668
Car WTSDYGCSEKEKEFHWLFKYSPLHNVRRPWEQSPDQVSQYPATMLLTADHDDRVVPLHSL 668
Hvu WTTDYGCSDKEEEFGWLIKYSPLHNVRRPWEQSFGDNCQYPATMLLTADHDDRVVPLHSL 713
Zma WTTDYGCSDKEEEFQWLIKYSPLHNVRRPWEQSSGNNCQYPATMLLTADHDDRVVPLHSL 708
Sva WTGDYGCSDKEEEFKWLIKYSPIHNVRRPWEQPGNEETQYPATMILTADHDDRVVPLHSF 662
Smo WTTDYGCSEKEEEFKWLIKYSPLHNVLRPWEKPDGLGVQYPPTMLLTADHDDRVVPLHTL 651
Ppa WTSDYGCADNAEEFDWLIKYSPLHNVKRPWEN-GSESIQYPATMLLTADHDDRVVPLHSL 673
Bta WTTDYGCSDNKQHFEWLIKYSPLHNVKLPEAD----DIQYPSMLLLTADHDDRVVPLHSL 650
Ssc WTTDYGCSDSKQHFEWLIKYSPLHNVKLPEAD----DIQYPSMLLLTADHDDRVVPLHSL 650
Oga WTTDYGCSDNKEHFEWLIKYSPLHNVKLPEAD----DVQYPSMLLLTADHDDRVVPLHSL 650
Rno WTTDYGCSDSKQHFEWLLKYSPLHNVKLPEAD----DIQYPSMLLLTADHDDRVVPLHSL 650
Gga WTTDYGCSDHKEQFEWLCKYSPLHNVKLPEED----GIQYPATLLLTADHDDRVVPLHSL 650
Xtr WTTDYGCSDNKEHFEWLIKYSPLHNIKVPEKD----GIQYPSMLLLTADHDDRVVPLHSL 652
Dre WTTDFGCSEIKEQFDWLIKYSPLHNIQVPEGD----GVQYPAVLLLTGDHDDRVVPLHSL 649
    ** ::**::  :.* ** ****:**:  *  .      ***. ::**.**********:
```

```
Vvi KLLATMQYILCTSVEKSPQTNPIIGRIECKAGHGAGRPTQK----MIDEAADRYSFLAKM 724
Ptr KLLATMQYILCTSLENSPQTNPIIGRIECKAGHGAGRPTKK----KIDEAADTYSFMARM 726
Gma KLLATMQYVLCTSLEKSPQTNAIIGRIDCKSGHGAGRPTQK----MIDEAADRYGFMAKV 720
Csa KLLATMQYVLCTSLEKSPQTNPIIGRIECKAGHGAGRPTQK----MIDEASDRYAFMAMM 724
Aly KLLATMQYELGLSLENSPQTNPIIARIEVKAGHGAGRPTQK----MIDEAADRYSFMAKM 724
Car KFLATLQYELCTGLESSPQTNPIISRIERKAGHGAGRPTQK----MIDEAADRYAFMAKV 724
Hvu KLLATMQHVLCTSIENSPQTNPIIGRIDRKSGHGAGRPTKK----LIDEAADRYSFMSKM 769
Zma KLLATLQHVLCTSTEDSPQTNPIIGRIDRKSGHGAGRPTQK----MIDEAADRYSFMAKM 764
Sva KLLATMQHVLCTSLEDSPQKNPIIARIQRKAAH-YGRATMT----QIAEVADRYGFMAKA 717
Smo KLLATLQHVLVKSSKSSNQTNPIVARIDTKAGHGAGRPTKK----MIDEAADRLSFFVKM 707
Ppa KLLATLQYELCTSVENSKQTNPIIARIDTKAGHGSGRPTKKIFIVQIDEMVDAYSFFAKM 733
Bta KFIATLQHLVGRS---RKQNNPLLIHVDTKAGHGAGKPTAK----VIEEVSDMFAFIARC 703
Ssc KFIATLQYIVGRS---RKQNNPLLIHVDTKAGHGAGKPTAK----VIEEVSDMFAFIARC 703
Oga KFIATLQYIVGRS---RKQSNPLLIHVDTKAGHGAGKPTAK----VIEEVSDMFAFIARC 703
Rno KFIATLQYIVGRS---RKQSNPLLIHVDTKAGHGAGKPTAK----VIEEVSDMFAFIARC 703
Gga KFIATLQYVVGRS---RKQTNPLLIHVDTKAGHGAGKPTAK----VIEEVSDMFAFIARC 703
Xtr KFIASLQHIAGQS---PNQTNPLLIHVDTKAGHGAGKPTAK----VIEEVSDMFAFIAQC 705
Dre KYIATLQNVIGQC---PGQKNPLFIYIDTKSGHGAGKPTSK----VIQEVADTYAFIARC 702
    * :*::*           *.*.::  :: *:.*  *:.* .     * *  *  .*:

Vvi LEASWIE 731
Ptr LDASWNE 733
Gma LEVHWIE 727
Csa LAATWID 731
Aly VDASWID 731
Car LGATWVD 731
Hvu LGATWTE 776
Zma LGASWTE 771
Sva LEAPWID 724
Smo TGAEWTE 714
Ppa TDSKWVD 740
Bta LNIDWIQ 710
Ssc LNIDWIP 710
Oga LNIDWIP 710
Rno LNIEWIQ 710
Gga LNLDWIE 710
Xtr LNLQWIE 712
Dre LNLSWLE 709
       *
```

Figure 5.2 Alignment of partial amino acid sequences of selected prolyl oligopeptidases. Sequences were aligned using CLUSTALW using default parameters. Putative active site amino acids are shown in reverse type. ".", ":" and "*" indicate increasing degrees of amino acid conservation. Species and Genbank accession numbers are: Aly, *Arabidopsis lyrata*, XP_002890385; Bta, *Bos taurus*, AAI40508; Car, *Coffea arabica*, AEX58649; Csa, *Cucumis sativus*, XP_004145530; Dre, Danio rerio, AAH95363; Gga, *Gallus gallus*, CAG31056; Gma, *Glycine max*, XP_003523219; Hvu, *Hordeum vulgare*, BAJ96495; Oga, *Otolemur garnettii*, XP_003781615; Ppa, *Physcomitrella patens*, XP_001757912; Ptr, *Populus trichocarpa*, XP_002306966; Rno, *Rattus norvegicus*, EDL99674; Smo, *Selaginella moellendorffii*, XP_002964833; Ssc, *Sus scrofa*, AAA31110; Sva, *Saponaria vaccaria*, KC588970; Vvi, *Vitis*

vinifera, CAN70125; Xtr, *Xenopus tropicalis*, NP_989189; Zma, *Zea mays*, AFW56408 (© National Research Council, Canada; Barber *et al.*, 2013; printed with permission http://www.jbc.org/content/288/18/12500.full?sid=117b51bd-0bc0-4a46-a6ba-38332a2d9ca6).

5.2 Optimization of PCY1 Assay Conditions

For optimum PCY1 activity, the type of buffer, buffer pH, NaCl concentration and DTT concentration were screened and the optimum conditions were determined as explained below.

5.2.1 Effect of pH on PCY1 Activity

Six different buffers (20 mM) and the pH range of 5.5 to 10.5 were screened (Table 5.1) to determine the optimum pH and buffer condition for the PCY1 (100 mM NaCl; 2 mM DTT). The broad range pH screening indicated that the PCY1 is active in the range of pH 7.0-10.0 and the optimum activity was centered between 8.5 and 9.0 (Figure 5.3). In this optimum pH range, there were two buffers, Tris and CHES, showed maximum product formation. At pH 8.5, Tris and CHES showed equal amounts of segetalin A production, but the linear segetalin A produced in CHES buffer was almost double than that seen with the Tris buffer. Thus, 20 mM Tris at pH 8.5 was chosen as optimum buffer and pH condition for the PCY1 assay.

Table 5.1 List of screened buffers with the pKa values, effective pH range and pH values screened for the PCY1 activity to determine optimum pH and buffer condition.

Buffer	pKa	Effective pH range	pH screened for PCY1 activity
MES	6.10	5.5 - 6.7	5.5, 6.0, 6.5
MOPS	7.14	6.5 - 7.9	6.5, 7.0, 7.5
Bicine	8.26	7.6 - 9.0	7.5, 8.0, 8.5
Tris	8.06	7.5 - 9.0	8.0, 8.5, 9.0
CHES	9.50	8.5 - 10.0	8.5, 9.0, 9.5
CAPS	10.40	9.7 - 11.1	9.7, 10.0, 10.5

The PCY1 is a serine protease-like enzyme with the putative catalytic triad of serine, histidine and aspartic acid. The optimum activity at pH 8.5 may be explained based on the chemistry of the catalytic triad. Histidine has the imidazole ring in its side chain and the pKa value for the side chain is 6.0. The ionization state of this side chain plays a key role in the catalysis. At pH 8.5, the side chain is in deprotonate form and histidine acts as general base in the reaction. It allows histidine to deprotonate the serine, which converts serine into a strong nucleophile for further nucleophilic attack on the substrate. The detailed catalytic mechanism of PCY1 is explained in section 6.3.

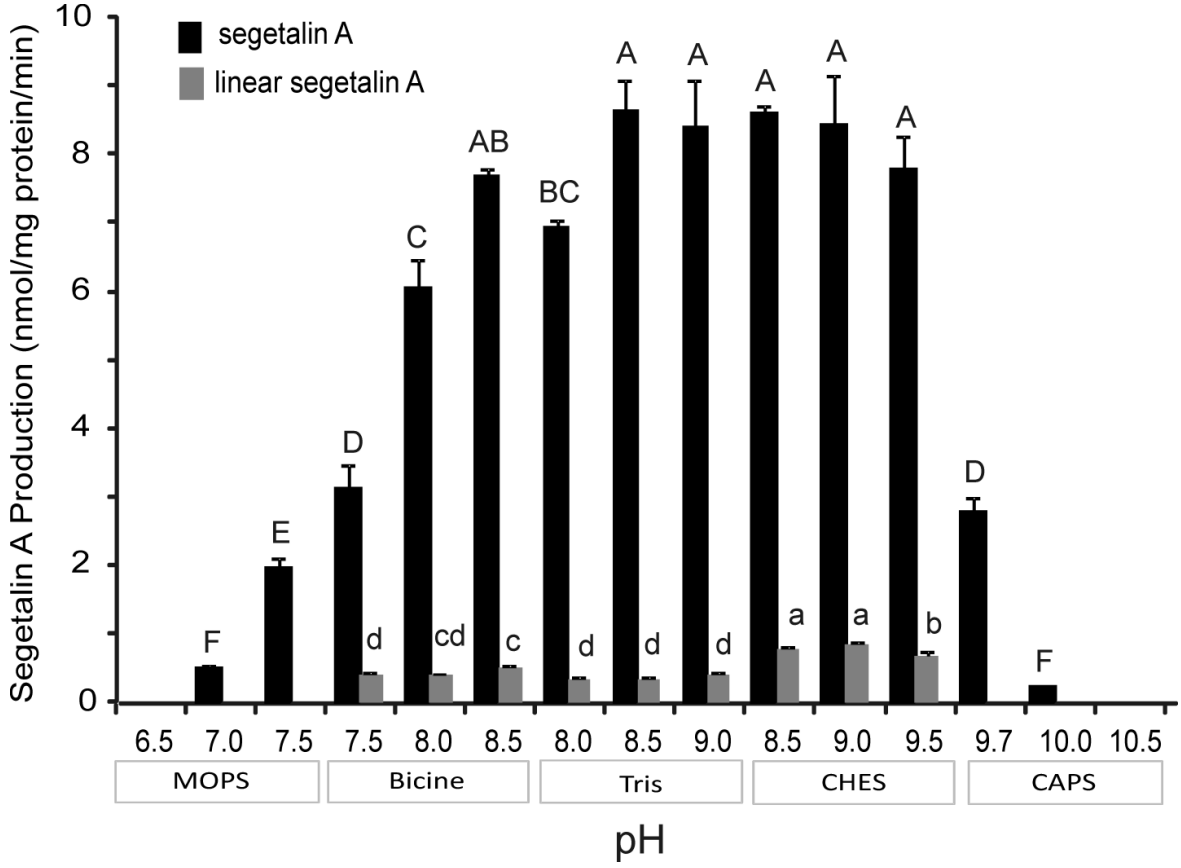

Figure 5.3 Effect of pH on PCY1 activity. Recombinant *S. vaccaria* PCY1 was assayed with presegetalin A1[14,32] at the pH values and buffers indicated. The means and standard deviations (n = 3) of the rate of formation of the linear and cyclic products for 1 h assays is indicated. Different letters indicate significant difference by Duncun test ($P \leq 0.05$). (© National Research Council, Canada; Barber *et al.*, 2013; printed with permission http://www.jbc.org/content/288/18/12500.full?sid=117b51bd-0bc0-4a46-a6ba-38332a2d9ca6).

5.2.2 Effect of Salt Concentration on PCY1 Activity

The PCY1 activity was screened in the range of 0 mM to 500 mM NaCl (20 mM Tris at pH 8.5; 2 mM DTT), which included eight data collecting points. The results revealed that the PCY1 activity was not significantly affected up to a concentration of 200 mM NaCl (Figure 5.4). However, above 200 mM NaCl, the PCY1 activity was gradually inhibited by higher salt concentrations. The effect of other salts on PCY1 activity is yet to be determined.

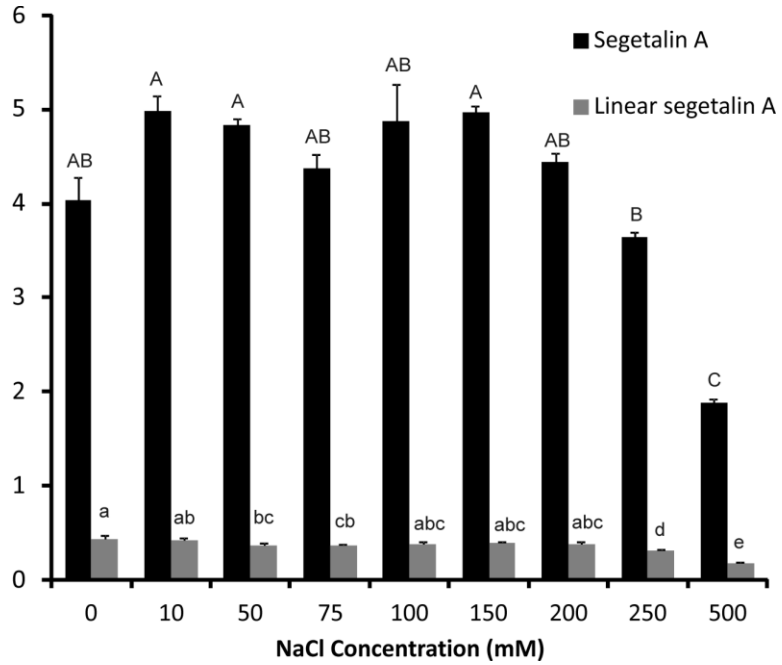

Figure 5.4 Effect of salt concentration on PCY1 activity. Recombinant PCY1 was assayed with presegetalin A1[14,32] at the NaCl concentration indicated. The means and standard deviations (n = 3) of the rate of formation of the linear and cyclic products for 1 h assays is indicated. Different letters for segetalin A and linear segetalin A indicate significant difference by Dunnett C test and Duncun test respectively. ($P \leq 0.05$).

5.2.3 Effect of DTT Concentration on PCY1 Activity

DTT is a reducing agent, and some enzymes require reducing agents for catalysis. Varying DTT concentrations from 0 mM to 10 mM were screened to determine the optimum concentration for the PCY1 activity. There was an approximately equal amount of segetalin A produced in assays within the range of 2 mM to 10 mM DTT. Importantly, all assay samples

with DTT had at least 20 times higher activity than assays performed without DTT (Figure 5.5). This revealed that the presence of DTT is an essential factor for PCY1 to catalyze the enzymatic reaction. It is also important to note that while DTT is critical for the production of segetalin A, but it is not important for the production of linear segetalin A.

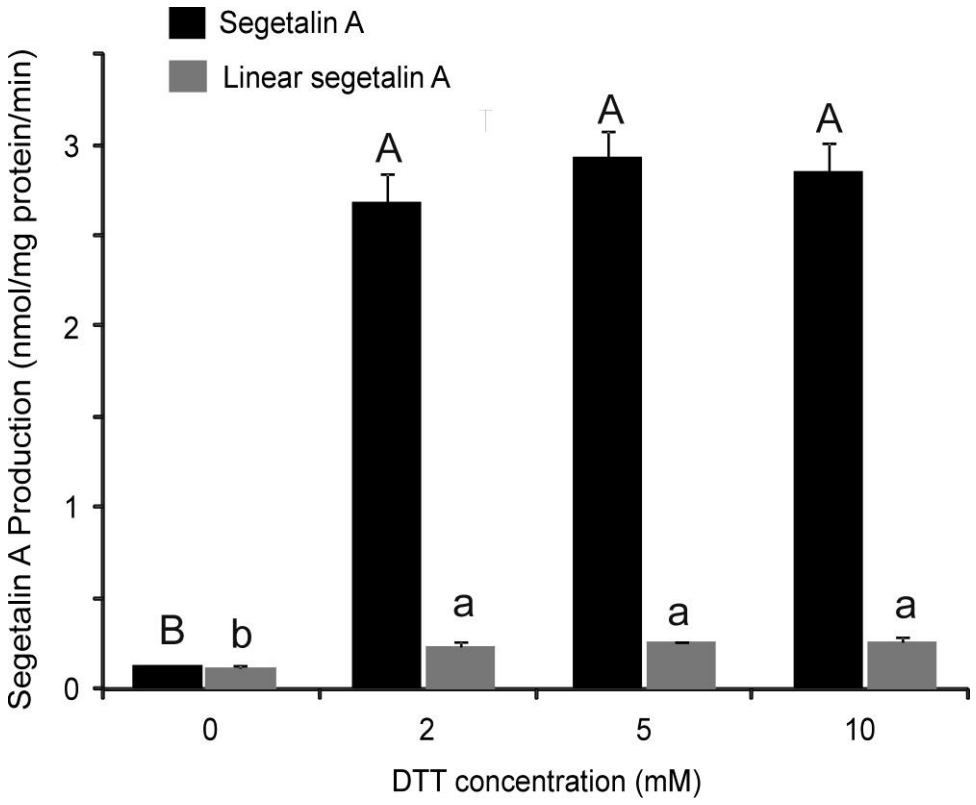

Figure 5.5 Effect of DTT concentration on PCY1 activity. Recombinant PCY1 was assayed with presegetalin A1[14,32] at the DTT concentrations indicated. The means and standard deviations (n = 3) of the rate of formation of the linear and cyclic products for 1 h assays is indicated. Different letters for segetalin A and linear segetalin A indicate significant difference by Dunnett C test and Duncun test respectively. ($P \leq 0.05$).(© National Research Council, Canada; Barber *et al.*, 2013; printed with permission http://www.jbc.org/content/288/18/12500.full?sid=117b51bd-0bc0-4a46-a6ba-38332a2d9ca6).

This result indicates that the PCY1 may have disulfide bonds in its structure which need to be reduced for substrate access to the active site. The experimentally obtained PCY1 structure is required for a better understanding of the relationship between DTT and PCY1 activity.

5.3 Testing Activity of the Homologues of PCY1 from Caryophyllaceae

As explained in section 4.2, Dr. Covello's lab had identified three homologues of *S. vaccaria* PCY1 from two species of Caryophyllaceae family (Table 5.2). Two homologues were identified from *D. superbus* (c250 and c1141) and the third homologue was identified from *S. vulgaris* (c150). These homologues were cloned and expressed in *E. coli*. The purified PCY1 homologues (section 4.3; Figure 5.6) were assayed with presegetalin A1[14,32] (Figure 5.7). The two PCY1 homologues from *D. superbus* were found to be active on presegetalin A1[14,32] and their activities were comparable to the *S. vaccaria* PCY1. The *S. vulgaris* PCY1 was active on presegetalin A1[14,32], but its activity was >500 fold less compared to *S. vaccaria* PCY1.

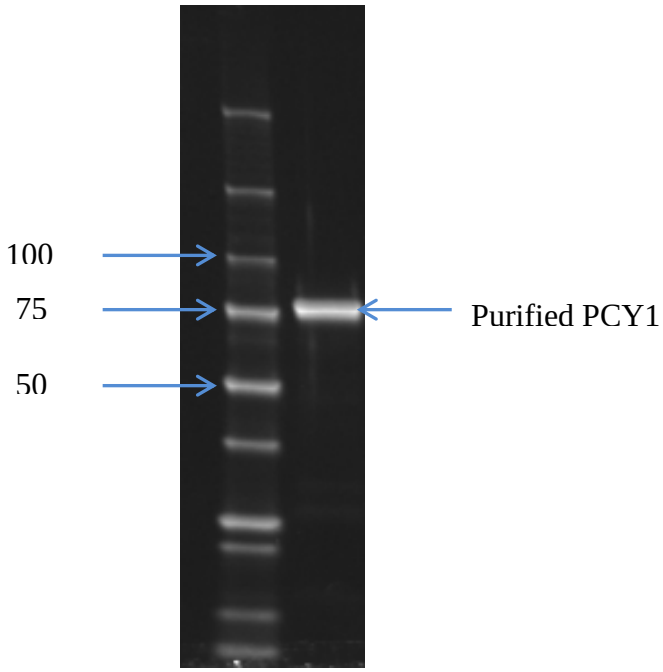

Figure 5.6 Purification of *E. coli* expressed recombinant *S. vaccaria* PCY1 by metal affinity chromatography. The precision plus protein™ unstained molecular weight standards and the purified PCY1 were subjected to SDS-PAGE under denaturing conditions using 10% polyacrylamide gel, and stained with fluorescent Oriole™ stain.

Table 5.2 Amino acid sequence identity (%) between *S. vaccaria* PCY1 and its homologues from *D. superbus* (c250 and c1141) and *S. vulgaris* (c150). The percent sequence identity matrix created by multiple sequence alignment using Clustal 2.1 programme.

	S. vaccaria (c272)PCY1	*S. vulgaris* (c150) PCY1	*D. superbus* (c250) PCY1	*D. superbus* (c1141) PCY1
S. vaccaria (c272)PCY1	100	78.7	79.1	78
S. vulgaris (c150) PCY1	78.7	100	72.3	72.6
D. superbus (c250) PCY1	79.1	72.3	100	88.5
D. superbus (c1141) PCY1	78	72.6	88.5	100

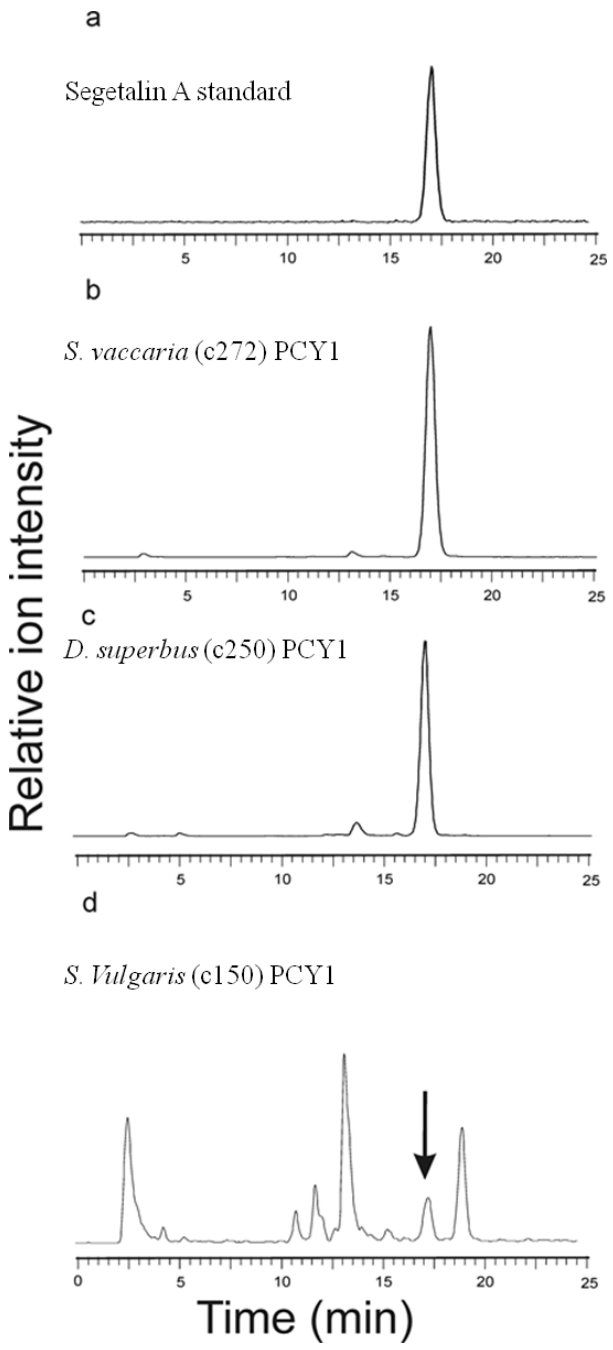

Figure 5.7 Detection of activities of *S. vaccaria* PCY1 and its homologues from *D. superbus* (c250) and *S. vulgaris* (c150), assayed with presegetalin A1[14,32]. Panels a, b and c show single ion monitoring LC/MS chromatograms for (a) segetalin A standard ($[M+H]^+$ at *m/z* 610.5 and retention time 17.1 min), (b) assay of recombinant PCY1 from *S. vaccaria*, and (c) assay of recombinant PCY1 from *D. superbus* (c250). Panel d shows a total ion trap current chromatogram (monitoring *m/z* range 50 to 2200 atomic mass units) of an assay of recombinant PCY1 from *S. vulgaris* (c150) with fragmentation similar to the segetalin A standard.

5.4 Substrate Specificity of PCY1

To determine the substrate specificity of PCY1, presegetalin A1[14,32] and variants of it (Table 5.3), total 48 peptides, were chemically synthesized by Biobasic Inc. with the purity of >90%. Here it is important to understand the nomenclature of peptide substrates listed in Table 5.3. In segetalin A biosynthetic pathway, the first enzyme (OLP1) cleaves 32 amino acid peptide precursor (presegetalin A1) into two smaller peptides named, presegetalin A1[1,13] and presegetalin A1[14,32] (Figure 2.4). In the case of presegetalin A1[14,32], presegetalin is a name of precursor peptide, A represents type of incipient segetalin (here segetalin A), 1 represents type of precursor involve (Table 2.1) and the numbers in brackets symbolize the first and last amino acid number corresponding to its native precursor (presegetalin A1).

5.4.1 Native Plant Substrates

5.4.1.1 Cyclic Peptide Precursors from *Saponaria vaccaria*

Condie *et al.,* 2011 has reported a group of segetalins in *S. vaccaria,* which included known and predicted segetalins. These segetalins are divided into two groups, A- and F-class segetalins. The A-class includes segetalins A, B, D, G, H, K and L while the F-class includes segetalins F and J. The A-class CPs are comprised of 5 to 7 amino acids, of which glycine is the first amino acid in the corresponding presegetalin. In contrast, the F-class CPs are comprised of 9 amino acids, of which phenylalanine is the first amino acid in the corresponding presegetalin.

A-class and F-class presegetalins (Peptides #1 to #9, Table 5.3) were assayed with PCY1. The results of the assays revealed that the cyclic products were detected from all the tested variants in LC/MS (Figure 5.8). The production of cyclic and linear products was confirmed by the presence of expected diagnostic ions $[M+H]^+$ and/or $[M+Na]^+$ in LC/MS. All the segetalins were further confirmed by LC/MS/MS analysis (Table 5.4). Due to unavailability of standards for each of these segetalins, cyclic and linear products produced during *in vitro* assays were not quantified except for segetalin A.

Table 5.3 Substrate specificity of *S. vaccaria* PCY1.
The synthetic linear peptides indicated were used in 1 h assays with *S. vaccaria* PCY1 (see "Materials and Methods"). CP and linear peptide production was determined by ion trap LC/MS and confirmed by ion trap LC/MS/MS. Diagnostic ions from ion trap LC/MS/MS of PCY1 assay samples producing the indicated cyclic and linear peptides are shown. Expected CP sequences are underlined. Variant amino acids are indicated in reverse type. Putative *D. caryophyllus* precursor sequences are derived from GenBank™ accession numbers AW697819 and CF259478 (Condie *et al.*, 2011). Quantitation of segetalin A and linear segetalin A production from selected substrates is shown in Figure 5.9. Lowercase letters denote D-amino acids. (© National Research Council, Canada; Barber *et al.*, 2013; printed with permission http://www.jbc.org/content/288/18/12500.full?sid=117b51bd-0bc0-4a46-a6ba-38332a2d9ca6)

Substrate Category	Peptide No.	Peptide Name	Sequence	PCY1 Product	
				Cyclic	Linear
native plant substrates	Cyclic Peptide Precursors from *Saponaria vaccaria*				
	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 IQTQVSNGMDNASAPV	+	-
	9	Presegetalin J1[14,36]	FGTHGLPAP IQVPNGMDDACAPM	+	-
	Putative Cyclic Peptide Precursors from *Dianthus caryophyllus*				
	10	*D. caryophyllus* Precursor 1	GPIPFYG FQAKDAENASVPV	+	-
	11	*D. caryophyllus* Precursor 2	GYKDCC VQAKDLENAAVPV	-	-

C-terminal presegetalin A1[14,32] variants	Deletion variants of presegetalin A1[14,32]				
	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	+	+
	Alanine/Valine Scanning of the Presegetalin A1[14,32] C-Terminus				
	19	Presegetalin A1[14,32] F20A	GVPVWA AQAKDVENASAPV	+	+
	20	Presegetalin A1[14,32] Q21A	GVPVWA FAAKDVENASAPV	+	+
	21	Presegetalin A1[14,32] A22V	GVPVWA FQVKDVENASAPV	+	+
	22	Presegetalin A1[14,32] K23A	GVPVWA FQAADVENASAPV	+	+
	23	Presegetalin A1[14,32] D24A	GVPVWA FQAKAVENASAPV	+	+
	24	Presegetalin A1[14,32] V25A	GVPVWA FQAKDAENASAPV	+	+
	25	Presegetalin A1[14,32] E26A	GVPVWA FQAKDVANASAPV	+	+
	26	Presegetalin A1[14,32] N27A	GVPVWA FQAKDVEAASAPV	+	+
	27	Presegetalin A1[14,32] A28V	GVPVWA FQAKDVENVSAPV	+	+
	28	Presegetalin A1[14,32] S29A	GVPVWA FQAKDVENAAAPV	+	+
	29	Presegetalin A1[14,32] A30V	GVPVWA FQAKDVENASVPV	+	+
	30	Presegetalin A1[14,32] P31A	GVPVWA FQAKDVENASAAV	+	+
	31	Presegetalin A1[14,32] V32A	GVPVWA FQAKDVENASAPA	+	+
	32	Presegetalin D1[14,31] P20Q	GLSFA FQAKDAENASSPV	+	+

Variants of incipient segetalin sequences	Alanine/Valine Scanning of the Incipient Segetalin A				
	33	Presegetalin A1[14,32] G14A	AVPVWA FQAKDVENASAPV	+	+
	34	Presegetalin A1[14,32] V15A	GAPVWA FQAKDVENASAPV	+	+
	35	Presegetalin A1[14,32] P16A	GVAVWA FQAKDVENASAPV	+	+
	36	Presegetalin A1[14,32] V17A	GVPAWA FQAKDVENASAPV	+	+
	37	Presegetalin A1[14,32] W18A	GVPVAA FQAKDVENASAPV	+	-
	38	Presegetalin A1[14,32] A19V	GVPVWV FQAKDVENASAPV	-	-
	D -Amino Acid Substitution in Incipient Segetalin A				
	39	Presegetalin A1[14,32] V15v	GvPVWA FQAKDVENASAPV	+	+
	40	Presegetalin A1[14,32] P16p	GVpVWA FQAKDVENASAPV	+	-
	41	Presegetalin A1[14,32] V17v	GVPvWA FQAKDVENASAPV	+	-
	42	Presegetalin A1[14,32] W18w	GVPVwA FQAKDVENASAPV	+	+
	43	Presegetalin A1[14,32] A19a	GVPVWa FQAKDVENASAPV	-	-
	44	Presegetalin A1[14,32] P16p, W18A	GVpVAA FQAKDVENASAPV	+	-
	45	Presegetalin A1[14,32] P16p, W18a	GVpVaA FQAKDVENASAPV	+	-
	46	Presegetalin A1[14,32] P16a, W18a	GVaVaA FQAKDVENASAPV	+	-
	Insertions Variants of the Incipient Segetalins				
	47	Presegetalin A1 [14,32] ins 16A17	GVPAVWA FQAKDVENASAPV	+	+
	48	Presegetalin A1 [14,33] ins 16AAA17	GVPAAAVWA FQAKDVENASAPV	+	+

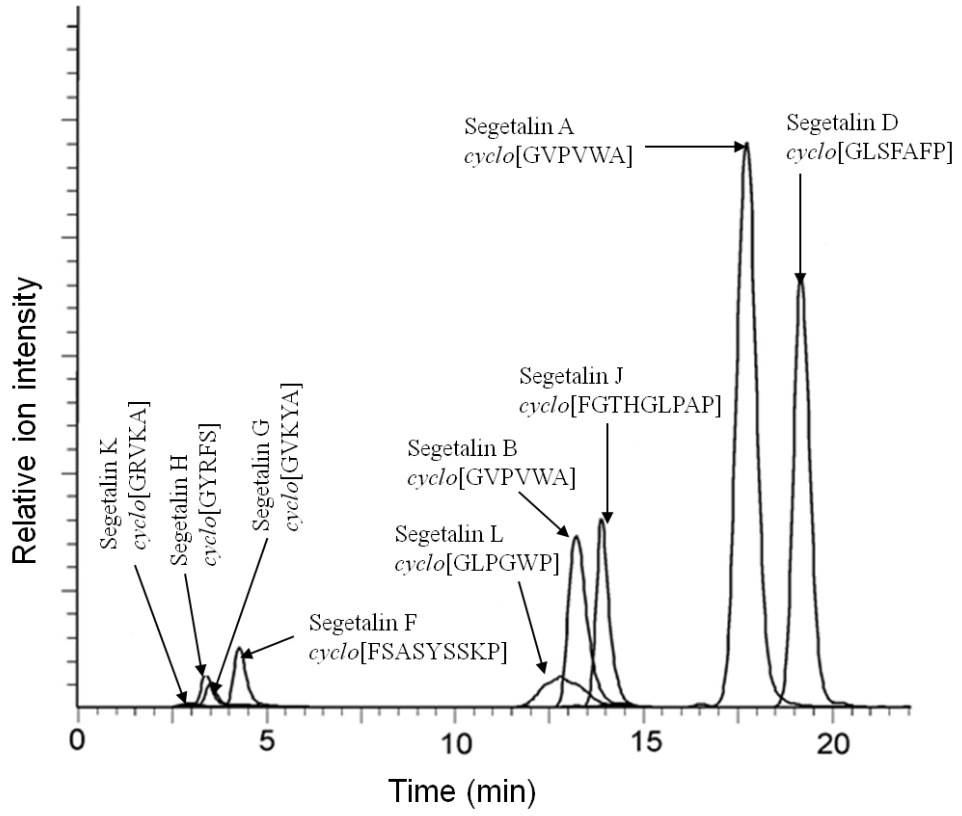

Figure 5.8 LC/MS analysis of native segetalin product of *S. vaccaria* PCY1. Ion chromatograms showing native segetalins derived from *S. vaccaria* PCY1 assays with synthetic linear substrates listed in Table 5.3.

5.4.1.2 Putative Cyclic Peptide Precursors from *Dianthus caryophyllus*

Condie *et al.*, 2011 has reported two putative CP precursors from *Dianthus caryophyllus*. The amino acid sequences of these two precursors (Peptides #10 and #11) appeared to be similar to the A-class segetalin precursors. These precursors were tested in PCY1 assays and the analysis of the assays indicated that PCY1 is weakly active on #10 but inactive on #11. The CP produced from #10 was detected by LC/MS as ions as *m/z* 732.5 $[M+H]^+$ and *m/z* 754.5 $[M+Na]^+$ and its presence was further confirmed by LC/MS/MS analysis (Table 5.4).

5.4.2 C-terminal Presegetalin A1[14,32] Variants

5.4.2.1 Deletion Variants of Presegetalin A1[14,32]

Seven deletion peptide variants of presegetalin A1[14,32] were synthesized by removing various sets of amino acids from the C-terminal (peptides #12 to #17) to explore the importance of the C-terminal region of the substrate in the cyclization reaction. Notably, PCY1 was not able to make segetalin A or linear segetalin A from any of the deletion peptide variants. These *in vitro* assay results with deletion peptide variants suggested that the C-terminal amino acid sequence of presegetalin A1[14,32] is important for segetalin A production.

5.4.2.2 Alanine/Valine Scanning of the Presegetalin A1[14,32] C- Terminus

In vitro assays with truncated mutant peptides suggested the importance of the C-terminal region of presegetalin A1[14,32] in the production of segetalin A. To identify the key residues from the C-terminal region, a total of 13 variants of presegetalin A1[14,32] were synthesized (Peptides #19 to #31) by substitution of each amino acid with alanine. When alanine was present in the sequence, it was substituted with valine. All variants were assayed with PCY1 in optimized assay conditions. In the LC/MS analysis, the segetalin A was detected as diagnostic ions *m/z* 610.5 $[M+H]^+$, 632.5 $[M+Na]^+$ and 648.5 $[M+K]^+$, while linear segetalin A was detected as *m/z* 628.5 $[M+H]^+$ and 650.5 $[M+Na]^+$ diagnostic ions. The segetalin A and linear segetalin A were quantified with a standard curve plotted with known amounts of standards (Figure 4.2).

When alanine/valine variants were assayed with PCY1, all except three of the variants show a very significant reduction in the production of segetalin A. The two substitutions showing the greatest effect in terms of reduced amount of the products were F20A and D24A for which segetalin A production was found to be $\geq$ 42 fold lower than the wild type substrate. Conversely, a substitution of V25A resulted in a more than 2-fold increase in 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 (Figure 5.9).

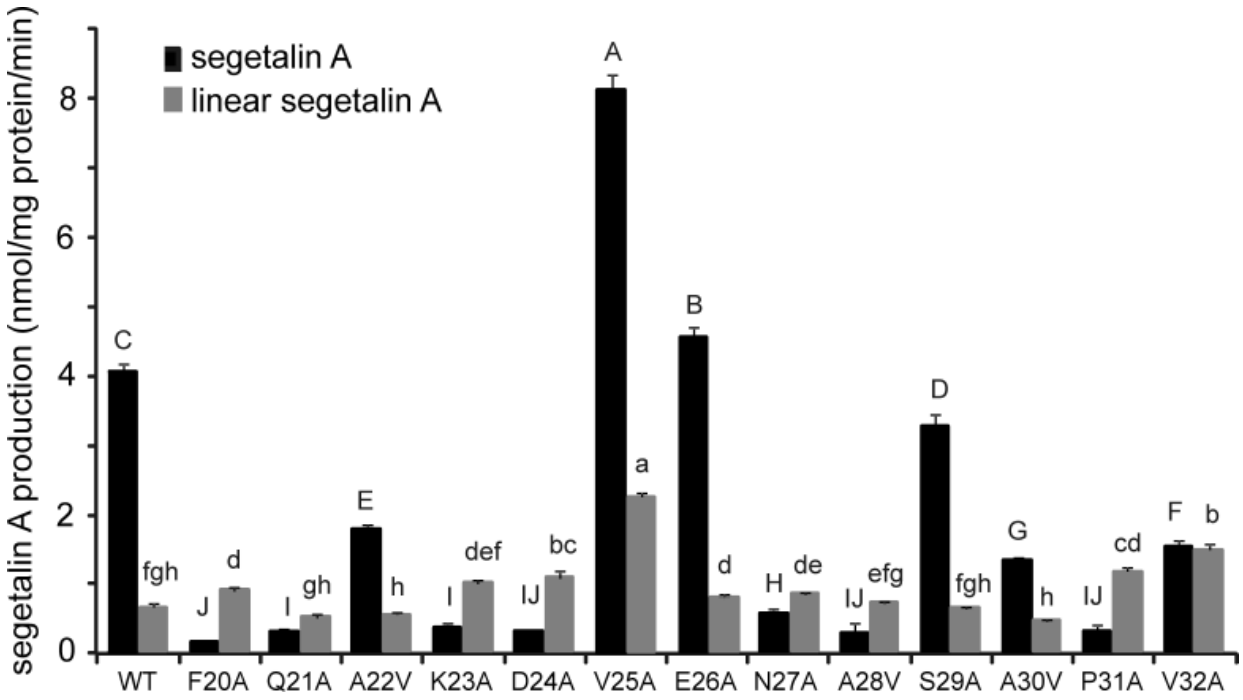

Figure 5.9 Effects of mutations in the C-terminal flanking region of presegetalin A1 on production of linear and cyclic segetalin A production by recombinant PCY1. Wild type (WT) and variants of presegetalin A1[14,32] are indicated. Means and standard deviations (n=3) of 1 h assays are indicated Different letters indicate significant difference by Duncun test ($P \leq 0.05$). (© National Research Council, Canada; Barber *et al.,* 2013; printed with permission http://www.jbc.org/content/288/18/12500.full?sid=117b51bd-0bc0-4a46-a6ba-38332a2d9ca6).

With these results in hand, deletion of the amino acids at positions 25 and 26 in presegetalin A1[14,32] (peptide #18 in Table 3.2) was tested and found to allow the formation of segetalin A at a rate ~170-fold lower than for the wild type substrate.

Presegetalin D1[14,31] is a precursor of seven amino acid segetalin D. Alignment of presegetalin D1[14,31] with other presegetalins suggest that, it would be interesting to substitute the proline at the 20th position in C-terminal region with glutamine (Peptide #32). This modified precursor peptide was tested with *S. vaccaria* PCY1. The proline to glutamine substitution resulted in an unnatural CP, a truncated version of segetalin D, made up of 5 amino acids (*cyclo*[GLSFA]), where the substituted glutamine was not part of the final cyclic product. This

result suggest that the 20th position in presegetalin D1[14,31] is important for CP formation by PCY1.

5.4.3 Variants of Incipient Segetalin Sequences

5.4.3.1 Alanine/Valine Scanning of the Incipient Segetalin A

Variants of the part of presegetalin A1[14,32] corresponding to mature segetalin A sequence were obtained by custom commercial synthesis (Peptides #33 to #38) to determine the importance of each amino acid at particular positions. Each amino acid in segetalin A was replaced with alanine consecutively, and the alanine in the segetalin A sequence was replaced with valine. *In vitro* assays with these variants revealed that the PCY1 was able to tolerate the alanine substitution in all positions of segetalin A, except the A19V substation (Figure 5.10).

D. superbus (c250) PCY1 was also assayed with these variants and the activities were compared with those of PCY1. *D. superbus* (c250) PCY1 activity was comparable to that of PCY1 with two notable differences. Firstly, there was no detectable CP made from G14A and A19V by *D. superbus* (c250) PCY1. Secondly, *D. superbus* (c250) PCY1 appeared to produce more CP from W18A variant, as compared to *S. vaccaria* PCY1 (Figure 5.10).

Given the apparently 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.

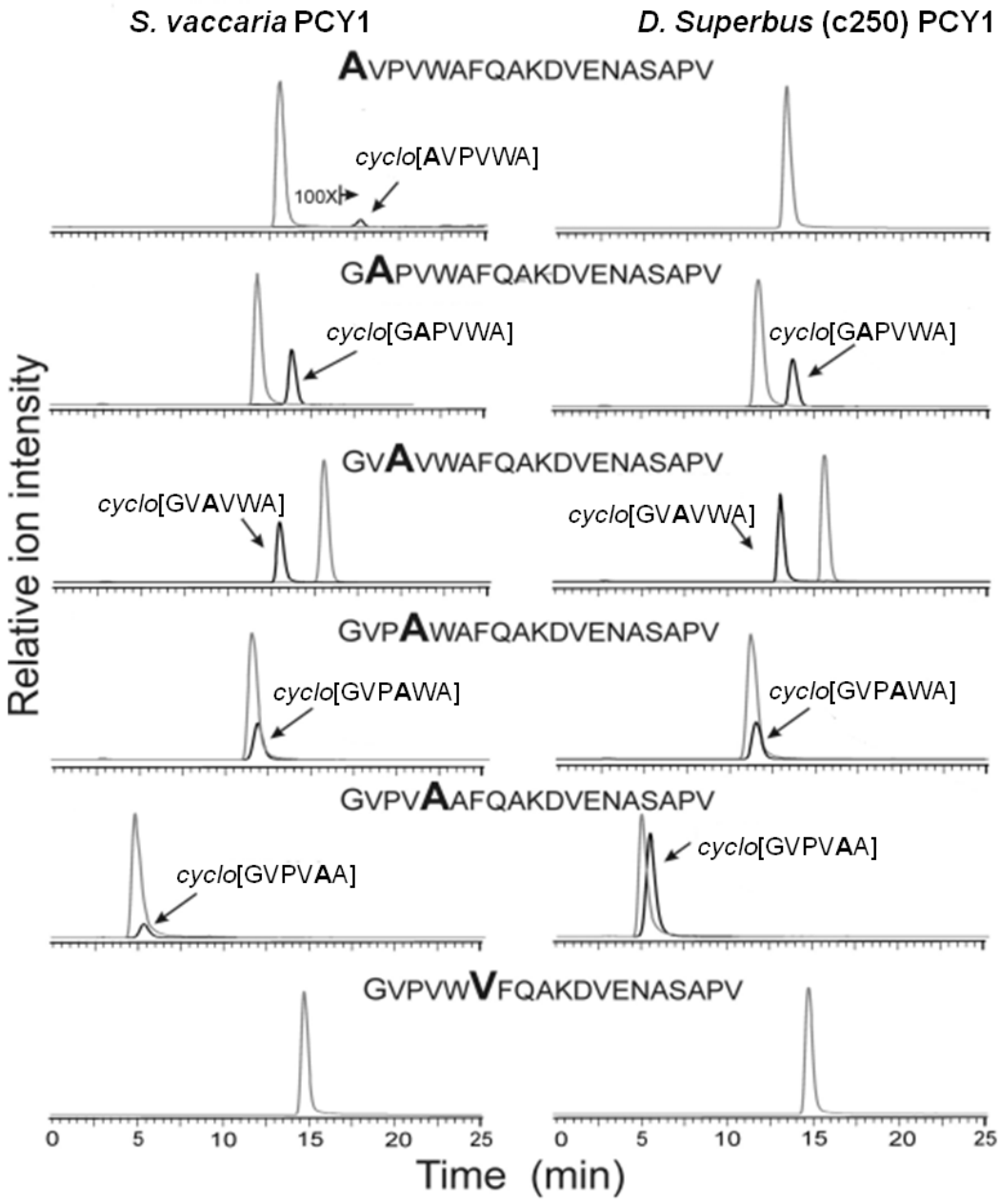

Figure 5.10 Alanine/valine scanning of the incipient segetalin A sequence. LC/MS chromatographs of assays of recombinant *S. vaccaria* PCY1 and *D. superbus* (c250) PCY1 showing single ion traces of alanine/valine scan variants of presegetalin A1[14,32]. The lighter grey traces represent the diagnostic ions for the peptide substrates (multiple charged molecular ions, specifically the sum of $[M+2H]^{2+}$ and $[M+3H]^{3+}$). The darker black traces represent the identification of a peak containing the diagnostic ions for the expected CP product (the sum of $[M+H]^{+}$ and $[M+Na]^{+}$).

5.4.3.2 D-Amino Acid Substitution in Incipient Segetalin A

Gadhiri *et al.*, 1993 and Hourani *et al.*, 2011 have reported that CPs containing an even number of amino acids with the alternating D- and L- chirality are able to form nanotubes, some of which have antimicrobial activity and other interesting commercial properties. Given this, it was of interest to see whether D-amino acids can be tolerated in the presegetalin A1[14,32] so as to give rise to segetalin A with variant stereochemistry.

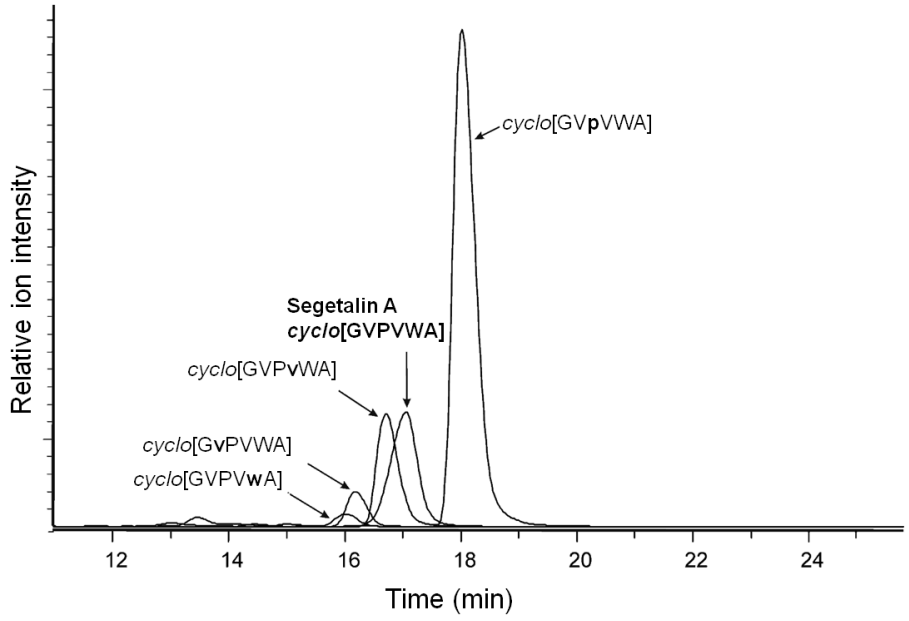

Figure 5.11 LC/MS chromatograms showing detected natural segetalin A and D- amino acid substitution variants of incipient segetalin A based on relative ion intensity. Lower case letters indicate D-amino acids.

Five variants containing D-amino acids (Peptides #39 to #43) were tested with PCY1 under optimized *in vitro* conditions. Glycine at position 14 is achiral and no substitution was required. *In vitro* assay results revealed that the PCY1 can tolerate all L- to D-substitutions, except the positional substitute A19a (lower case letters indicate D-amino acids) for which there was no detectable cyclic or linear product (Figure 5.11).

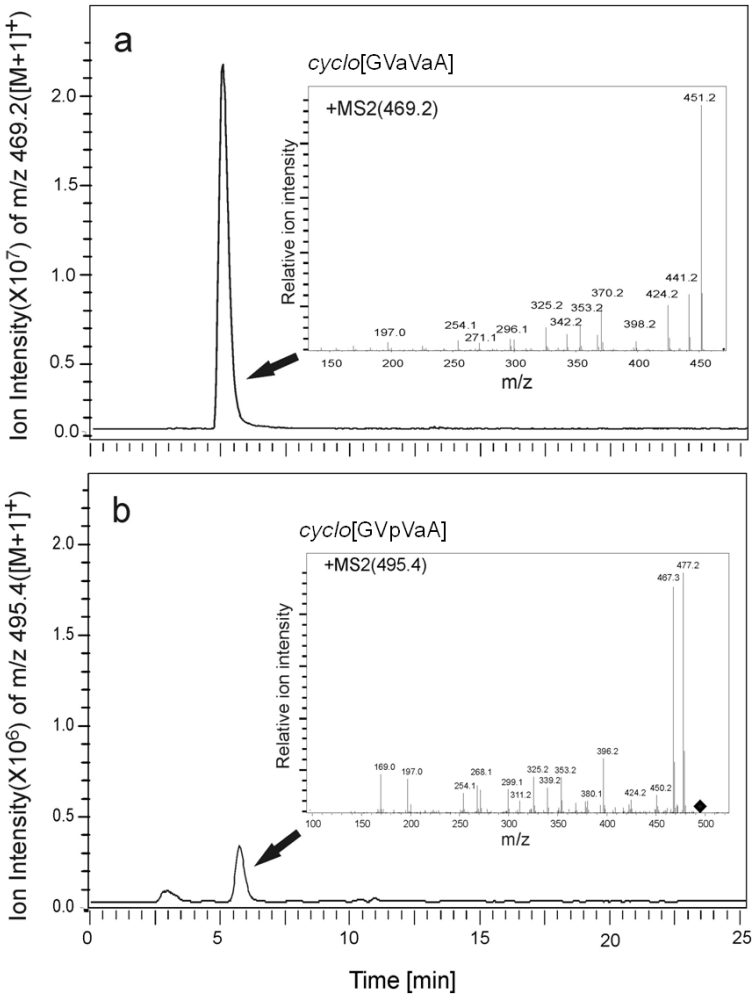

Figure 5.12 LC/MS detection of the incipient CPs with double D-amino acid substitutions, *cyclo*[GVpVaA] and c*yclo*[GVpVaA], produced by *S. vaccaria* PCY1. The CP product was identified by monitoring expected molecular ions $[M+H]^+$ and $[M+Na]^+$ and verified by LC/MS/MS analysis. Insets in (a) and (b) show LC/MS/MS fragmentation of $[M+H]^+$. For the detailed explanation of LC/MS/MS results, see Figure 5.15. Lower case letters indicate D-amino acids.

From these results, it appeared that it may be possible to generate cyclic peptides with alternating D- and L-amino acids using PCY1. To test this possibility, three variants (Peptides #44 to #46) with single and double substitutions with D-amino acids were synthesized and tested for *S. vaccaria* PCY1. In these variants, a tryptophan at the 18th position in presegetalin A1[14,32] was replaced with alanine, because the W18w substitution was apparently weakly converted to CP. PCY1 gives rise to cyclic product from all three variants with single and double

substitutions (Figure 5.12). The CP products *cyclo*[GVpVaA] and *cyclo*[GVaVaA] have alternating D- and L- forms of amino acids (with the exception of the glycine), which gives it the potential to self-assemble into nanotube under appropriate conditions.

5.4.3.3 Insertions Variants of the Incipient Segetalins

According to Tan and Zhou, 2006, the largest known Caryophyllaceae-like CP is made up 11 amino acids, Stelladein A - *cyclo*[PPPLLGPPYYG]. This fact led us to investigate the possibility of making versions of segetalin A with additional amino acids.

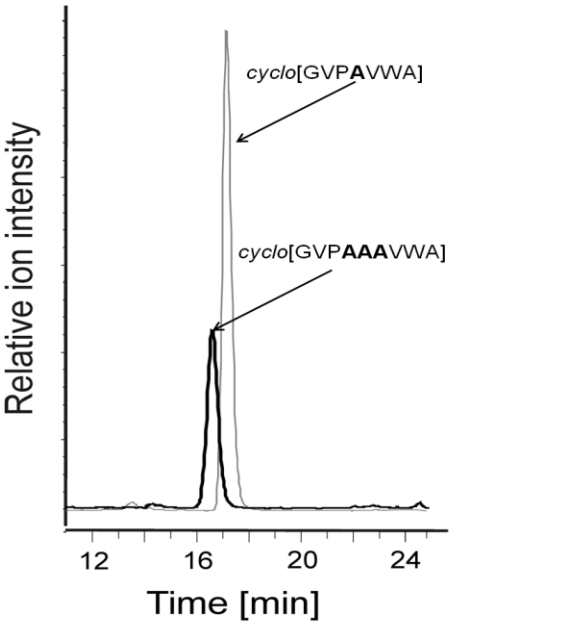

Figure 5.13 LC/MS detection of the incipient CP products with insertion of one (gray) and three (black) alanine in segetalin A sequence, resulted into a CP product of seven and nine amino acids respectively. The inserted alanines are symbolized as bold fonts in the illustrated figure.

Two presegetalin A1[14,32] variants representing insertions of `A` and `AAA` (Peptides #47 and #48) in the incipient CP sequence were tested for PCY1 activity. Both of these variants were found to act as substrates and cyclic and linear products of seven and nine amino acids were detected in LC/MS analysis of the *in vitro* assays (Figure 5.13). The presence of

cyclo[GVPAVWA] and *cyclo*[GVPAAAVWA] were confirmed by LC/MS/MS analysis (Table 5.4). These results and the previously tested substrates suggest that the PCY1 can make CPs from five to nine amino acids.

5.5 Competitive Assays with Inactive Substrates

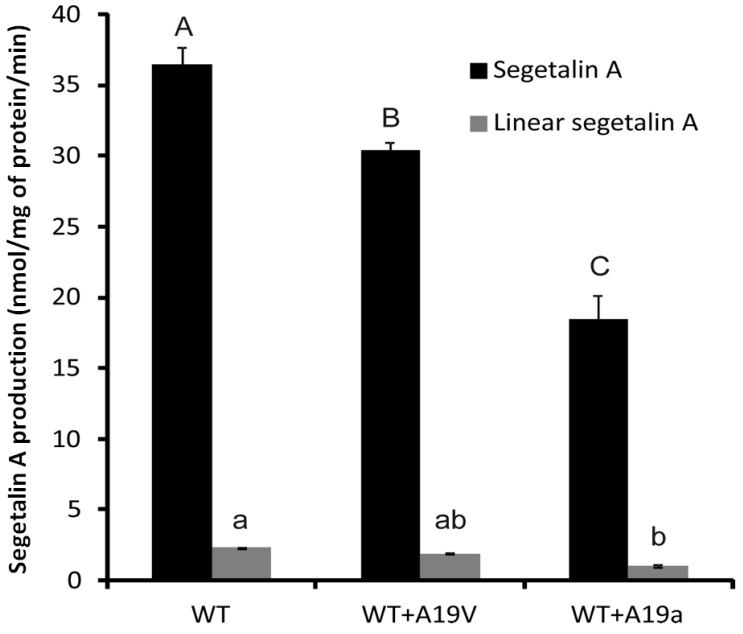

Figure 5.14 Effect of competition (between WT and inactive substrate peptides #38 and #43) on production of linear and cyclic segetalin A production by recombinant PCY1. Wild type (WT) and variants of presegetalin A1[14,32] are indicated. Means and standard deviations (n=3) of 1 h assays are indicated. Different letters indicate significant difference by Duncun test ($P \leq 0.05$).

The substrate specificity experiments revealed that *S. vaccaria* PCY1 was inactive on two substrates, peptides #38 and #43 in Table 5.3. It would be interseting to determine that these inactive substrates compete for the active site or not. In a competitive assay, the activity of PCY1 was tested on presegetalin A1[14,32] (wild type substrate) with equal quantity of inactive substrates in assays. The result revealed that the PCY1 activity was inhibited up to 13% and 51% by peptide #38 and #43, respectively (Figure 5.14). This result suggests that the peptide #43

may be a good candidate to co-crystallize with PCY1 in future structure determination experiments by X-ray crystallography.

5.6 Analysis of Cyclic Peptides by Tandem Mass Spectrometry

As explained in section 5.3, cyclic and linear products from a variety of substrates were detected by LC/MS based on the *m/z* value of $[M+H]^+$ ions. To confirm the presence of particular products, the amino acid sequence was determined by fragmentation of the product in LC/MS/MS analysis. The mass spectrometer was first autotuned on the *m/z* value of the product ion to be fragmented. The autotuned ion trap selects the $[M+H]^+$ of a product ion and performs fragmentation by collision induced dissociation.

There are many possible routes for linear peptide fragmentation. Generally, a linear peptide breaks at an amide bond and follows the b_X-y_Z pathway which generates detectable b and y ions (Qi *et al.,* 2007). In MS/MS, the fragmentation of CPs is more complicated than the linear peptides. In 1982, Gross and co-workers sequenced an unknown CP by using tandem mass spectrometry. Qi *et al.,* 2007 had explained that CP with *n* amino acids will generate *n* series of b ions, but not y ions. It is possible to provide annotations for b ions, a ions (losses of CO), and small neutral losses such as H_2O and NH_3 for the mass spectra produced from the fragmentation of CP (Liu *et al.,* 2009). Ngoka and Gross, 1999 have proposed the nomenclature system for b ions generated from CP fragmentation and mentioned that often there are a few main amide bonds which are the sites of initial cleavage.

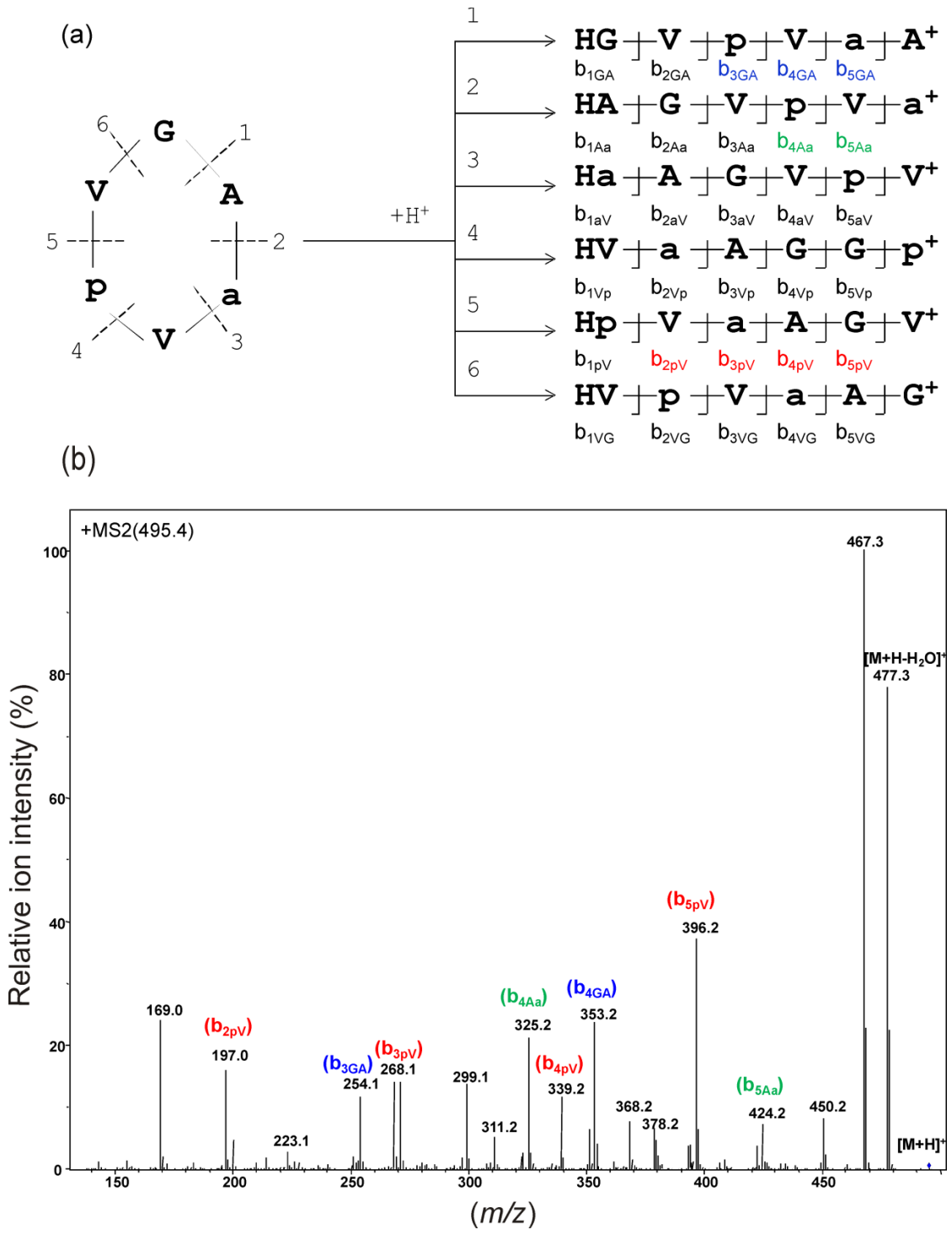

Figure 5.15 Fragmentation of a six membered CP, *cyclo*[GVpVaA]. Panel (a) represents the theoretical possibilities of ring-opening during fragmentation. Panel (b) is the fragmentation pattern obtained from LC/MS/MS. The 495.4 (*m/z*) ion was selected and subjected to collision induced dissociation. The fragmented b ions from the CP have been assigned as indicated. The three different colors represent three ring-opening patterns observed in the mass spectrum.

To understand the fragmentation pattern and annotation of ions, the example of *cyclo*[GVpVaA], a six membered CP produced from peptide #45 is explained in detail, as follows. Theoretically, a CP with *n* numbers of amino acids can form *n* types of linear peptides (known as acylium ions or its isomeric equivalent) of same *m/z* via ring-opening at *n* different positions. In principle, the *cyclo*[GVpVaA] produces six acylium ions which further converts into 30 b series ions (Figure 5.15a). The results suggest that the one stage tandem mass spectrometry (LC/MS/MS) of *cyclo*[GVpVaA] is composed of multiple ring-opening pathways with the set of acylium ions or its isomeric equivalent of the same *m/z* 495.4 and many b ions (Figure 5.15b). Table 5.4 represents detected diagnostic ions in LC/MS/MS to confirm the linear and cyclic products (products mentioned in Table 5.3) formed from a variety of substrate tested for *S. vaccaria* PCY1 activity in section 5.3.

Table 5.4 List of detected diagnostic ions for CPs in tandem mass spectrometry analysis.

Confirmation of products of PCY1 assay by ion trap LC/MS/MS. Diagnostic ions from ion trap LC/MS/MS of PCY1 assay samples producing the indicated cyclic and linear peptides (see Table 5.3) are shown. The *m/z* values and assignments are indicate for prominent daughter ions derived from the $[M+H]^+$ ion. The nomenclature for the assignment of CP-derived ions follows Ngoka and Gross, 1999. Amino acids in reverse type indicate variation from the wild type segetalin A sequence. Lower case letters indicate D-amino acids (Table is modified from Barber *et al.*, 2013; (© National Research Council, Canada; printed with permission http://www.jbc.org/content/288/18/12500.full?sid=117b51bd-0bc0-4a46-a6ba-38332a2d9ca6).

Name of Cyclic Product	Product Sequence	Diagnostic ions	
		Cyclic Product	Linear Product
Segetalin A	GVPVWA	610.5($[M+H]^+$), 511.2(b_{5PV}), 454.2(b_{4PV}), 383.3 (b_{3PV})	628.5($[M+H]^+$), 539.3(b_5), 353.2(b_4), 472.3(y_4)
Segetalin B	GVAWA	485.4 ($[M+H]^+$), 386.1(b_{4AV}), 329.1(b_{3AV}), 258.1(b_{2AV})	
Segetalin D	GLSFAFP	720.5 ($[M+H]^+$), 607.3(b_{6SL}), 453.3(b_{4SL})	738.4($[M+H]^+$), 623.3(b_6), 476.2(b_5)
Segetalin D[1,5]	GLSFA	476.6($[M+H]^+$), 363.1(b_{4SL}), 306.1(b_{3SL}) , 235.0(b_{2SL})	
Segetalin G	GVKYA	519.6($[M+H]^+$), 420.2(b_{4KV}), 363.2(b_{3KV}), 292.1(b_{2KV})	
Segetalin H	GYRFS	611.5($[M+H]^+$), 464.2(b_{4SF}), 308.1(b_{3SF})	
Segetalin K	GRVKA	512.5($[M+H]^+$), 413.2(b_{4KV}), 285.1(b_{3AK})	
Segetalin L	GLPGWP	608.6($[M+H]^+$), 495.2(b_{5PL}), 511.2(b_{5LG}), 341.1(b_{3PL})	
Segetalin F	FSASYSSKP	955.5($[M+H]^+$), 705.3(b_{7KS}), 618.3(b_{6KS}), 547.2(b_{5KS})	
Segetalin J	FGTHGLPAP	878.5($[M+H]^+$), 765.4(b_{8PL}), 674.4(b_{7TG}),	

		710.4(b_{7PA})	
Putative *D. caryophyllus* CP	GPIPFYG	732.5 ($[M+H]^+$), 619.3(b_{6PI}), 522.2(b_{5PI}), 408.1(b_{3PI})	
Segetalin A Variants	AVPVWA	624.5($[M+H]^+$), 525.3(b_{5WV}), 428.1(b_{4WV}), 329.1(b_{3WV})	642.4($[M+H]^+$), 553.3(b_5), 367.2(b_4), 472.2(y_4)
	GAPVWA	582.3($[M+H]^+$), 511.3(b_{5PA}), 454.2(b_{4PA}), 383.2(b_{3PA})	600.3($[M+H]^+$), 511.2(b_5), 325.2(b_4), 472.2(y_4)
	GVAVWA	584.5($[M+H]^+$), 485.2(b_{5AV}), 428.2(b_{4AV}), 357.2(b_{3AV})	602.5($[M+H]^+$), 513.2(b_5), 327.1(b_4), 375.1(y_3)
	GVPAWA	582.3($[M+H]^+$), 483.2(b_{5PV}), 426.2(b_{4PV}), 355.2(b_{3PV})	600.3($[M+H]^+$), 511.2(b_5), 325.1(b_4), 444.2(y_4)
	GVPVAA	495.4($[M+H]^+$), 396.2(b_{5PV}), 339.2(b_{4PV}), 268.1(b_{3PV})	513.5($[M+H]^+$), 424.5(b_5), 353.2(b_4), 357.5(y_4)
	GvPVWA	610.5($[M+H]^+$), 511.2(b_{5PV}), 454.2(b_{4PV}), 383.3 (b_{3PV})	628.5($[M+H]^+$), 539.3(b_5), 353.2(b_4), 472.3(y_4)
	GVpVWA	610.5($[M+H]^+$), 511.2(b_{5PV}), 454.2(b_{4PV}), 383.3 (b_{3PV})	
	GVPvWA	610.5($[M+H]^+$), 511.2(b_{5PV}), 454.2(b_{4PV}), 383.3 (b_{3PV})	
	GVPVwA	610.5($[M+H]^+$), 511.2(b_{5PV}), 454.2(b_{4PV}), 383.3 (b_{3PV})	628.5($[M+H]^+$), 539.3(b_5), 353.2(b_4), 472.3(y_4)
	GVpVAA	495.4($[M+H]^+$), 396.2(b_{5PV}), 339.2(b_{4PV}), 268.1(b_{3PV})	
	GVpVaA	495.4($[M+H]^+$), 396.2(b_{5PV}), 339.2(b_{4PV}), 268.1(b_{3PV})	
	GVaVaA	469.4($[M+H]^+$), 398.2(b_{5VA}), 370.2(b_{5AV}), 299.1(b_{4VA})	
	GVPAVWA	681.6($[M+H]^+$), 582.3(b_{6PV}), 525.3(b_{5PV}), 610.3(b_{6VA})	699.5($[M+H]^+$), 610.3(b_6), 424.2(b_5), 543.3(y_5)
	GVPAAAVWA	823.5($[M+H]^+$), 724.4(b_{8PV}), 667.4(b_{7PV}), 596.3(b_{6PV})	841.5($[M+H]^+$), 752.4(b_8), 566.3(b_7), 685.4(y_7)

6. DISCUSSION

Recent advancement of DNA sequencing techniques have provide significant assistance in the discovery of many ribosome-derived natural products, like CPs. In the past two decades, researchers are actively involved in developing novel peptide-based therapeutics, and many CPs have been considered important scaffolds in the design of novel peptide based drugs. Naturally, plants are major source of variety of CPs. Thus, it is important to study CP biosynthesis in detail for potential applications of CPs in future.

6.1 Comparison of Ribosome-dependent Cyclic Peptide Biosynthesis in Various Taxa

The biosynthesis Caryophyllaceae-like CPs was unclear up to 2011. Jia *et al.,* 2006 had reported a cyclization of the octapeptide heterophyllin B from its linear form by extracts of *Pseudostellaria hetererophylla* (Caryophyllaceae family). This work was not carried forward and the enzyme responsible for cyclization and its mechanism remained unclear. Condie *et al.,* 2011, provided evidence for ribosome-derived precursors from *S. vaccaria,* responsible for the biosynthesis of a group of CPs called segetalins. Similar precursors were evident from *D. caryophyllus* and *Citrus* spp. The ribosome-derived precursors revealed that the CP sequence is buried between highly conserved and N- and C-terminal sequences. The A class CPs from *S. vaccaria* have a propensity to start with Gly; the two F class CPs start with Phe. The reported CPs from carnation and *Citrus* spp. follow the pattern of A class segetalins, starting with Gly. Importantly, all of these incipient CPs have more variation in the C-terminus, as they end with different amino acids, often Ala and Pro. This work was carried forwarded and Barber *et al.,* 2013 explained the biosynthetic pathway for CP biosynthesis in *S. vaccaria.* The 32 amino acid precursor presegetalin A1 was converted into the six amino acid CP segetalin A by two enzymes named OLP1 and PCY1. OLP1 acts as a protease to remove N-terminus from the precursor by performing a cleavage at the junction of N-terminus and incipient CP sequence. The OLP1 has not been identified and characterized yet. PCY1 processes the resulting intermediate to make a CP, by performing a cyclization reaction, specifically, an intra-molecular transpeptidation reaction. Recently Gui *et al.,* 2012

reported multi-domain precursors of Caryophyllaceae-like CPs in *L. usitatissium* (flax). These multi-domain precursors have a propensity to have Met residues at the N-terminus of incipient CP sequence.

It is a matter of interest to compare known biosynthetic pathways from other species with Caryophyllaceae-like CPs. In plants, other known classes of CPs are kalata, PawS and cyclic knottin. These CPs are known to be synthesized from ribosome-derived precursors but their arrangement and sequence differ completely from the precursors of the Caryophyllaceae-like CPs. The precursors of kalata, PawS and cyclic knottin CPs share some common features like Gly and Asx at the N- and C- terminus of incipient CPs. It has been suggested an asparaginyl endopeptidase, a cysteine protease, is responsible for the final cyclization step. These CPs are believed to arise by evolutionary parallelism via asparaginyl-endopeptidase mediated biosynthesis (Mylne *et al.,* 2012). CP biosynthesis in plants appears to have some common steps which include the removal of the N-terminus followed by cyclization step. Moreover, it is evident that the C- terminus is highly conserved and it has specific sequence signal to carry out cyclization reaction. The role of N- terminal sequence is not clear but it is believed that the N- terminus may be involved in the protection of incipient CP sequence from peptidases, post-translational modifications or transport of the precursor into specific parts of the cell (Arnison *et al.,* 2013). Also the N-terminus may be important to allow the `free amino group` on incipient CP instead of N-formylmethionine (fMet). The `free amino group` on the N-terminus of incipient CP sequence is one of the prerequisite conditions in CP biosynthesis.

Other than plants, CPs have been reported in mammals, fungi and bacteria and in some species their biosynthesis is partially or fully understood. The θ-defensins are the only known CP from mammals which are head-to-tail cyclized and by ligation of two nonapeptides (Selsted, 2004). The CP biosynthesis has been studied in fungi, particularly, in *Amanita* mushrooms where there are examples of CP production from ribosome-derived precursors. The mature CP sequence is flanked by conserved Pro residues and a prolyl-oligopeptidase has been suggested to be involved in cyclization (Walton *et al.,* 2010). Patellamides are a group of head-to-tail cyclized CPs found in cyanobacteria (*Prochloron*

didemni) of which the biosynthesis has been well-studied. The CP biosynthetic pathway in cyanobacteria resembles that of Caryophyllaceae-like CP in *S. vaccaria*. The gene encoded precursors contain one or more mature CP sequence flanked by conserved amino acid sequences. The N- terminus is removed by the *patA* product, and *patG* product, a protease-like enzyme, participates in the excision and cyclization to form a CP product (Lee *et al.*, 2009). Comparing biosynthetic pathways, the functions of PatA and PatG in cyanobacteria are similar to OLP1 and PCY1 in *S. vaccaria* respectively. It is noteworthy that the products and homologues of *patG* and *pcy1* are the only cloned enzymes involved in CP biosynthesis and their function is known as peptide cyclization.

Considering all known systems for CP biosynthesis in a wide range of taxa, there are some common patterns for ribosome-derived CP biosynthesis. The gene encoded precursors contain one or more incipient CP sequence flanked by highly conserved N- and C- terminus. The first protease-like enzyme removes the N-terminus of incipient CP from precursor and the second protease-like enzyme participates in cleavage and cyclization of incipient CP. Evidently, the second protease-like enzyme in biosynthetic pathway is a serine or cysteine type, which performs the transpeptidation reaction. The serine and cysteine proteases are known for their possible ability to carry out transpeptidation reaction since 1950s (Johnston *et al.*, 1950, Berkers *et al.*, 2009). This indeed suggests a pattern of convergent evolution of proteases from different lineages to produce diverse ranges of CPs.

6.2 Structure-function Relationships of PCY1

As mentioned in the section 4.1, the homology based PCY1 model was obtained using porcine muscle POP. The comparison of PCY1 model and POP provides structure and function related information. The porcine muscle POP is a S9 serine oligopeptidase with two distinct domains known as α/β hydrolase and β-propeller (Fulop *et al.*, 1998). The comparison of PCY1 and POP structures suggest there may be very subtle differences in the structure responsible for two distinct catalytic activities. In addition, the insertion/deletion in the close proximity of catalytic histidine residue may prove to be important.

The α/β-hydrolase domain possesses the catalytic site and this topology is common in the SC clan of peptidases. It is believed that the α/β-hydrolase fold has yielded enzymes with diverse functions via divergent evolution (Rea and Fulop, 2006). The well-known examples of enzymes with α/β-hydrolase folds are lipases, esterases and thioesterases. The β-propeller domain present in POP and PCY1 has a cylindrical shape with a height of ~ 60 Å and a diameter of ~ 50 Å. It is made up of seven blades (seven fold repeats of four antiparallel β-sheets) where the β-sheets are twisted around the central axis of the β-propeller tunnel. All other known β-propeller structures are overlapped between first and seventh blade (Neer and Smith, 1996; Baker *et al.,* 1997) and makes a closed "velcro-like" structure. The β-propeller found in POP is unusual because the "velcro" is not closed and mostly it is stabilized by hydrophobic interaction (Figure 6.1). Thus, it is less compact, more flexible and more irregular compare to other known closed β-propeller structure. The β-propeller is thought to be the entrance of substrate for catalysis and the size of β-propeller tunnel restrict the size of substrates, by allowing the smaller substrates and excluding the larger ones.

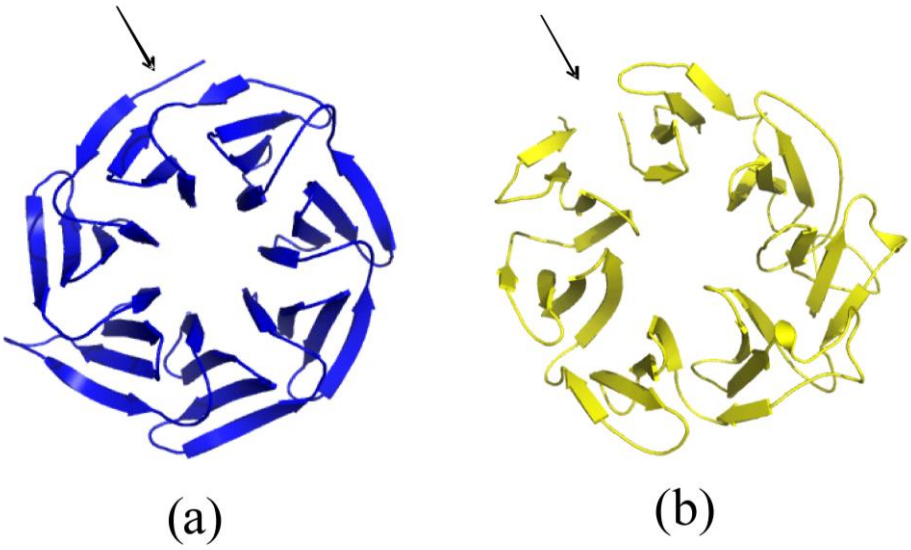

Figure 6.1 Structure comparision of β-propeller domains. (a) A Typical β-Propeller domain of G-protein β subunit (PDB ID: 1tbg) with the closed "velcro-like" (by a hydrogen bond) structure. (b) Non-catalytic, unusual β-propeller domain obtained from the homology based model of PCY1 with the open "velcro-like" structure.

Although, the entry of substrate and how the substrate approach as the catalytic site is not clearly understood, Rea and Fulop, 2006 proposed three possibilities for this event: (1) movements of the side-chains present into β-propeller tunnels widening the tunnel and allow the entry of substrate from the bottom side (2) the first and seventh blades of β-propeller move apart and widen the opening for the substrates (3) peptidase and propeller domains opens at the interface and allow substrate to access the catalytic triad from the side instead of bottom. In future, the high resolution crystal structure of PCY1 may answer the questions related to substrate entry and limit for the size of substrate.

6.3 Intra-molecular Transpeptidation Reaction by PCY1

As discussed earlier, the PCY1 is identified as a S9 serine protease homologue. Serine proteases possess the catalytic triad made up of serine, histidine and aspartic acid, which takes part in a multistep proteolysis reaction on general acid/base catalysis principles. In a typical proteolysis reaction catalyzed by serine proteases, histidine acts as a general base and deprotonates the serine, and thus prepares serine for stronger nucleophilic attack on the carbonyl of the scissile peptide bond. This results in the formation an acylserine intermediate and releases the C-terminus of the substrate. Subsequently, the acylserine intermediate gets hydrolyzed by the entry of water and release the C-terminus of the substrate, regenerating the enzyme in its native form. However, in some enzymes or under certain conditions, the acylserine intermediate can undergo aminolysis in competition with the hydrolysis reaction. When the rate of aminolysis is relatively fast compare to the rate of hydrolysis, a transpeptidation reaction takes place, if the amine in the question is at the N-terminus of the separate peptide. In the CP biosynthetic context, if the amine is located at the other end of acylserine intermediate, this results in a cyclic product via aminolysis, and the reaction is called intra-molecular transpeptidation or cyclization (Figure 6.2). Furthermore, Berkers *et al.,* 2009 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.

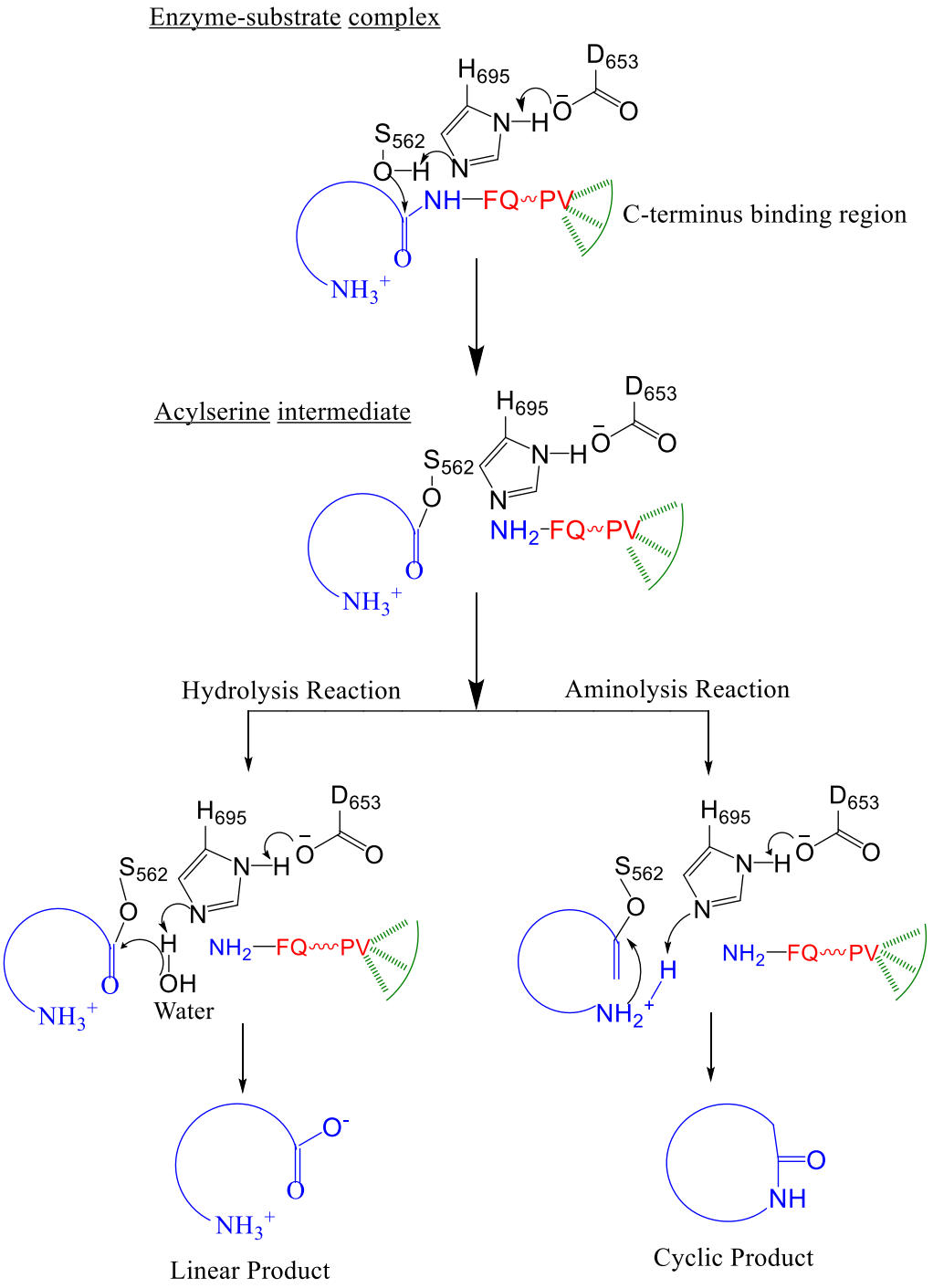

Figure 6.2 Proposed mechanism of PCY1 involved in the final step of Caryophyllaceae-like CP biosynthesis. The catalytic serine residue of PCY1 makes an acylserine intermediate with the substrate. The intermediate is subjected for two possible routes, depending on the competition between hydrolysis and aminolysis which generates linear and cyclic product respectively.

The putative acyl intermediates expected for the reaction represented in Table 5.3 are short but quite variable in sequence and length. However, the C-terminus of the PCY1 substrates is highly conserved. The data from the alanine/valine scanning of the C-terminus suggest that the substitutions had a mostly strong negative effect on cyclization reaction and the cyclic to linear product ratio is also influenced by the subsequent substitutions. In principle, the C-terminus of the substrate could be released from the enzyme prior to aminolysis reaction but the above data indicate that the C-terminus of the substrate remains bound to the enzyme even after acylserine intermediate formation and may influence the cyclization (aminolysis) reaction. Agarwal *et al.,* 2012 has proposed the similar retention of the C-terminal fragment of the substrate on *patG* homologues.

Although, the present research explores the characterization of PCY1 and provides insight into CP biosynthesis in plants, PCY1 has many potential biotechnological applications. The ability of PCY1 to produce CP with alternate D- and L- amino acids unlocks the possibility of enzyme-assisted self-assembled CP nanotube production (Ghadiri *et al.,* 1993). This CP based nanotube can be useful to make many biocompatible materials and nanotechnology based devises like biosensors (Cho *et al.,* 2008), artificial ion channels (Ghadiri *et al.,* 1994), artificial photosystems (Brea *et al.,* 2010), antimicrobial and antiviral agents (Horne *et al.,* 2005) and controlled-release drug delivery carriers. The relatively flexible substrate specificity indicates that the PCY1 can be useful to generate small CP libraries for testing and production purposes (Bourne *et al.,* 2005; Barber *et al.,* 2013).

7. CONCLUSIONS

This study has provided valuable information about the mechanism, substrate specificity and structure of *S. vaccaria* PCY1, a protease-like enzyme involve in CP biosynthesis in Caryophyllaceae family of plants. The important conclusions drawn from this study are as listed below

- *S. vaccaria* PCY1 is a serine protease-like enzyme, which acts on a linear precursor peptide (presegetalin A1[14,32]) and produces a CP (segetalin A) by performing a cyclization reaction in *S. vaccaria*. Technically, the cyclization reaction performed by PCY is an intra-molecular transpeptidation.

- Homology modelling based predicted structure of *S. vaccaria* PCY1 suggests that the catalytic triad is made up of Ser^{562}, Asp^{653}, and His^{695}. The catalytic triad is located at the interface of two distinct domains: a catalytic α/β hydrolase domain and an unusual β-propeller domain.

- *S. vaccaria* PCY1 has broad substrate specificity. It can make CPs of 5-9 amino acids from linear precursors and also able to incorporate non-proteinogenic amino acids (D-amino acids) into CP.

- The C-terminus of the substrate is essential for the cyclization reaction. The size and specificity (amino acid sequence) of C-terminus affects the quantity of cyclic and linear product produced by PCY1.

- *S. vaccaria* PCY1 requires the DTT for catalysis and the pH optimum is centered between 8.5 and 9.0.

- Homologues of *S. vaccaria* PCY1 from other two species *D. superbus* and *S. vulgaris*, are active on presegetalin A1[14,32].

8. FUTURE WORK

This study focused on mechanism, substrate specificity and structure prediction of *S. vaccaria* PCY1. The information from this study provides insight into CP biosynthesis in plants and potential biotechnological applications of PCY1. Although, very important information was obtained from this study, there are still unanswered questions regarding CP biosynthesis in plants. Following are some of the interesting aspects of CP biosynthesis yet to be explored.

The crystal structure of *S. vaccaria* PCY1 is one of the very important aspects to be investigated. It has been proposed that the PCY1 performs an intra-molecular transpeptidation (aminolysis) reaction to make cyclic product. A crystal structure of PCY1 may help to explain the prevention of access of water molecules into the active site, where aminolysis reaction takes place during catalysis. It would also explain the interaction of substrate (presegetalin A1[14,32]) with PCY1, specially the C-terminus. The experimental data suggested that the C-terminus of the substrate is an essential for CP formation by PCY1.

The preliminary results of *S. vaccaria* PCY1 homologue activity on selected substrates (section 5.3, 5.4.3.1 and 5.4.3.2) indicated differences in cyclic and linear product formation. It would be interesting to explore *S. vaccaria* PCY1 homologues substrate specificity with the substrates listed in Table 5.3. Crystal structures of *S. vaccaria* PCY1 homologues would provide an opportunity to access the structure alignment of all PCY1 homologues, which will facilitate identification of the subtle structure differences between the homologues and ultimately help to explain the differences in their activity.

There are 48 substrates that have been tested for *S. vaccaria* PCY1 and CPs from 5-9 amino acid were produced. It would be interesting to determine the upper limit of the amino acids incorporated in CP by PCY1. These results will be important information to make a CPs of various sizes. As explained in McIntosh *et al.*, 2010, substrates with the modified N-termini will helpful to explore the possibility of cyclization via side-chain nucleophiles.

CP precursors were found to be expressed very highly in developing seeds compared to other tissues such as root, leaf and flowers (Condie *et al.,* 2011). The relative gene expression (qPCR) study of *pcy1* in different tissues of *S. vaccaria* will be valuable information to locate the tissues with higher expression for further study. The biological role of CPs (segetalins) in *S. vaccaria* is not understood yet which also an important task to be determined.

As explain in section 2.3.5.1 (Figure 2.4), there are two protease-like enzymes involve in the segetalin A biosynthetic pathway, OLP1 and PCY1. In this thesis, PCY1 was characterized but the identity of OLP1 is still remains unknown. The identification and characterization of OLP1 will provide more information to understand CP biosynthesis in *S. vaccaria*. Knowledge of enzymes involve in pathway will be an excellent resource to construct *in vitro* or *in vivo* systems for CP library to testing and production purposes.

9. REFERENCES

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.

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.

Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, *et al.,* (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep. 30, 108-160.

Baker, S.C., Saunders, N.F., Willis, A.C., Ferguson, S.J., Hajdu, J., and Fulop, V. (1997). Cytochrome cd1 structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to beta-propeller folds. J. Mol. Biol. 269, 440-455.

Barber, C.J., Pujara, P.T., Reed, D.W., Chiwocha, S., Zhang, H., and Covello, P.S. (2013). The two step biosynthesis of cyclic peptides from linear precursors in a member of the plant family Caryophyllaceae involves cyclization by a serine protease-like enzyme. J. Biol. Chem. 288, 12500-12510.

Barbosa, S.C., Cilli, E.M., Dias, L.G., Stabeli, R.G., and Ciancaglini, P. (2011). Labaditin, a cyclic peptide with rich biotechnological potential: preliminary toxicological studies and structural changes in water and lipid membrane environment. Amino Acids 40, 135-144.

Belshaw, P.J., Walsh, C.T., and Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. Science 284, 486-489.

Berkers, C.R., de, J.A., Ovaa, H., and Rodenko, B. (2009). Transpeptidation and reverse proteolysis and their consequences for immunity. Int. J. Biochem. Cell Biol. 41, 66-71.

Bockus, A.T., McEwen, C.M., and Lokey, R.S. (2013). Form and function in cyclic peptide natural products: a pharmacokinetic perspective. Curr. Top. Med. Chem. 13, 821-836.

Bourne, G.T., Nielson, J.L., Coughlan, J.F., Darwen, P., Campitelli, M.R., Horton, D.A., Rhumann, A., Love, S.G., Tran, T.T., and Smythe, M.L. (2005). A convenient method for synthesis of cyclic peptide libraries. Methods Mol. Biol. 298, 151-165.

Brea, R.J., Reiriz, C., and Granja, J.R. (2010). Towards functional bionanomaterials based on self-assembling cyclic peptide nanotubes. Chem. Soc. Rev. 39, 1448-1456.

Bruner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T., and Stubbs, M.T. (2002). Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. Structure. 10, 301-310.

Challis, G.L., Ravel, J., and Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. 7, 211-224.

Cho, E.C., Choi, J.W., Lee, M., and Koo, K.K. (2008). Fabrication of an electrochemical immunosensor with self-assembled peptide nanotubes. Colloid Surface A, 313-314, 95–99.Clark, R.J., Jensen, J., Nevin, S.T., Callaghan, B.P., Adams, D.J., and Craik, D.J. (2010). The engineering of an orally active conotoxin for the treatment of neuropathic pain. Angew. Chem. Int. Ed Engl. 49, 6545-6548.

Clugston, S.L., Sieber, S.A., Marahiel, M.A., and Walsh, C.T. (2003). Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst. Biochemistry 42, 12095-12104.

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.

Conlan, B.F., Colgrave, M.L., Gillon, A.D., Guarino, R., Craik, D.J., and Anderson, M.A. (2012). Insights into processing and cyclization events associated with biosynthesis of the cyclic peptide kalata B1. J. Biol. Chem. 287, 28037-28046.

Conlan, B.F., Gillon, A.D., Barbeta, B.L., and Anderson, M.A. (2011). Subcellular targeting and biosynthesis of cyclotides in plant cells. Am. J. Bot. 98, 2018-2026.

Conlan, B.F., Gillon, A.D., Craik, D.J., and Anderson, M.A. (2010). Circular proteins and mechanisms of cyclization. Biopolymers 94, 573-583.

Craik, D.J. (2006). Chemistry. Seamless proteins tie up their loose ends. Science 311, 1563-1564.

Craik, D.J. (2009). Circling the enemy: cyclic proteins in plant defence. Trends Plant Sci. 14, 328-335.

Craik, D.J. (2010). Discovery and applications of the plant cyclotides. Toxicon 56, 1092-1102.

Craik, D.J., Swedberg, J.E., Mylne, J.S., and Cemazar, M. (2012). Cyclotides as a basis for drug design. Expert. Opin. Drug Discov. 7, 179-194.

Dalsgaard, P.W., Larsen, T.O., and Christophersen, C. (2005). Bioactive cyclic peptides from the psychrotolerant fungus *Penicillium algidum*. J. Antibiot. (Tokyo) 58, 141-144.

Daly, N.L., and Craik, D.J. (2009). Design and therapeutic applications of cyclotides. Future. Med. Chem. 1, 1613-1622.

Deschuyteneer, G., Garcia, S., Michiels, B., Baudoux, B., Degand, H., Morsomme, P., and Soumillion, P. (2010). Intein-mediated cyclization of randomized peptides in the periplasm of *Escherichia coli* and their extracellular secretion. ACS Chem. Biol. 5, 691-700.

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.

Folli, C., Ramazzina, I., Arcidiaco, P., Stoppini, M., and Berni, R. (2003). Purification of bacteriocin AS-48 from an *Enterococcus faecium* strain and analysis of the gene cluster involved in its production. FEMS Microbiol. Lett. 221, 143-149.

Fulop, V., Bocskei, Z., and Polgar, L. (1998). Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. Cell 94, 161-170.

Ghadiri, M.R., Granja, J.R., and Buehler, L.K. (1994). Artificial transmembrane ion channels from self-assembling peptide nanotubes. Nature 369, 301-304.

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. Nature 366, 324-327.

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.

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.

Goransson, U., Burman, R., Gunasekera, S., Stromstedt, A.A., and Rosengren, K.J. (2012). Circular proteins from plants and fungi. J. Biol. Chem. 287, 27001-27006.

Gross M. L., McCrery D., Crow F., Tomer K.B., Pope M.R., Cuifetti L.M., Knoch H.W., Daly J.M. and Dunkle L.D. (1982). The structure of the toxin from *Helminthosporium carbonum*. Tetrahedron. Lett., 23 (1982), pp. 5381–5384

Hallen, H.E., Luo, H., Scott-Craig, J.S., and Walton, J.D. (2007). Gene family encoding the major toxins of lethal *Amanita* mushrooms. Proc. Natl. Acad. Sci. U. S. A 104, 19097-19101.

Horne, W.S., Wiethoff, C.M., Cui, C., Wilcoxen, K.M., Amorin, M., Ghadiri, M.R., and Nemerow, G.R. (2005). Antiviral cyclic D,L-alpha-peptides: targeting a general biochemical pathway in virus infections. Bioorg. Med. Chem. 13, 5145-5153.

Hourani, R., Zhang, C., van der Weegen, R., Ruiz, L., Li, C., Keten, S., Helms, B.A., and Xu, T. (2011). Processable cyclic peptide nanotubes with tunable interiors. J. Am. Chem. Soc. 133, 15296-15299.

Huang, Y.H., Colgrave, M.L., Daly, N.L., Keleshian, A., Martinac, B., and Craik, D.J. (2009). The biological activity of the prototypic cyclotide kalata b1 is modulated by the formation of multimeric pores. J. Biol. Chem. 284, 20699-20707.

Itokawa, H., Takeya, K., Hitotsuyanagi, Y., Morita, H. (1997). Macrocyclic Peptide Alkaloids From Plants The Alkaloids: Chemistry and Pharmacology. In The Alkaloids. Cordell, G. A., Ed.; Academic Press: New York, Vol. 49, pp 301-318.

Itokawa, H., Yun, Y., Morita, H., Takeya, K. and Yamada, K. (1995). Estrogenlike 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. Sci. China Series B Chem. 49, 63–66

Johnston, R.B., Mycek, M.J., and Fruton, J.S. (1950). Catalysis of transpeptidation reactions by chymotrypsin. J. Biol. Chem. 187, 205-211.

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.

Kaas, Q., and Craik, D.J. (2010). Analysis and classification of circular proteins in CyBase. Biopolymers 94, 584-591.

Kaplan, C.D., Larsson, K.M., and Kornberg, R.D. (2008). The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. Mol. Cell 30, 547-556.

Katsara, M., Tselios, T., Deraos, S., Deraos, G., Matsoukas, M.T., Lazoura, E., Matsoukas, J., and Apostolopoulos, V. (2006). Round and round we go: cyclic peptides in disease. Curr. Med. Chem. 13, 2221-2232.

Kaufmann H.P., Tobschirbel A. (1959). An oligopeptide from linseed. Chem. Ber., 92, 2805-2809.

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.

Kohli, R.M., Trauger, J.W., Schwarzer, D., Marahiel, M.A., and Walsh, C.T. (2001). Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. Biochemistry 40, 7099-7108.

Lai, E.M., Eisenbrandt, R., Kalkum, M., Lanka, E., and Kado, C.I. (2002). Biogenesis of T Pili in *Agrobacterium tumefaciens* Requires Precise VirB2 Propilin Cleavage and Cyclization. J. Bacteriol. 184, 327-330.

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.

Lipmann, F., Gevers, W., Kleinkauf, H., and Roskoski, R., Jr. (1971). Polypeptide synthesis on protein templates: the enzymatic synthesis of gramicidin S and tyrocidine. Adv. Enzymol. Relat Areas Mol. Biol. 35, 1-34.

Liu, W.T., Ng, J., Meluzzi, D., Bandeira, N., Gutierrez, M., Simmons, T.L., Schultz, A.W., Linington, R.G., Moore, B.S., Gerwick, W.H., Pevzner, P.A., and Dorrestein, P.C. (2009). Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides. Anal. Chem. 81, 4200-4209.

Luo, H., Hallen-Adams, H.E., and Walton, J.D. (2009). Processing of the phalloidin proprotein by prolyl oligopeptidase from the mushroom *Conocybe albipes*. J. Biol. Chem. 284, 18070-18077.

Luo, H., Hallen-Adams, H.E., Scott-Craig, J.S., and Walton, J.D. (2010). Colocalization of amanitin and a candidate toxin-processing prolyl oligopeptidase in *Amanita basidiocarps*. Eukaryot. Cell 9, 1891-1900.

Maksimov, M.O., Pelczer, I., and Link, A.J. (2012). Precursor-centric genome-mining approach for lasso peptide discovery. Proc. Natl. Acad. Sci. U. S. A 109, 15223-15228.

Martin-Visscher, L.A., Gong, X., Duszyk, M., and Vederas, J.C. (2009). The three-dimensional structure of carnocyclin A reveals that many circular bacteriocins share a common structural motif. J. Biol. Chem. 284, 28674-28681.

Mazmanian, S.K., Liu, G., Ton-That, H., and Schneewind, O. (1999). *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. Science 285, 760-763.

McIntosh, J.A., Robertson, C.R., Agarwal, V., Nair, S.K., Bulaj, G.W., and Schmidt, E.W. (2010). Circular logic: nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst. J. Am. Chem. Soc. 132, 15499-15501.

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.. and Takeya, K. (2010). Bioactive cyclic peptides from higher plants. Heterocycles, 80, 739–764.

Mulvenna, J.P., Foley, F.M., and Craik, D.J. (2005). Discovery, structural determination, and putative processing of the precursor protein that produces the cyclic trypsin inhibitor sunflower trypsin inhibitor 1. J. Biol. Chem. 280, 32245-32253.

Mulvenna, J.P., Wang, C., and Craik, D.J. (2006). CyBase: a database of cyclic protein sequence and structure. Nucleic Acids Res. 34, D192-D194.

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.

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.

Namjoshi, S., and Benson, H.A. (2010). Cyclic peptides as potential therapeutic agents for skin disorders. Biopolymers 94, 673-680.

Neer, E.J., and Smith, T.F. (1996). G protein heterodimers: new structures propel new questions. Cell 84, 175-178.

Ngoka, L.C., and Gross, M.L. (1999). A nomenclature system for labeling cyclic peptide fragments. J. Am. Soc. Mass Spectrom. 10, 360-363.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, 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.

Oman, T.J., and van der Donk, W.A. (2010). Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat. Chem. Biol. 6, 9-18.

Pelegrini, P.B., Quirino, B.F., and Franco, O.L. (2007). Plant cyclotides: an unusual class of defense compounds. Peptides 28, 1475-1481.

Qi, W., Jia, C.X., He, Z. M., and Qiao, B. (2007). Multistep tandem mass spectrometry for sequencing cyclic peptides axinastatin 1 and deducing fragmentation pathways. Acta Chim.Sin. 65, 233-238.

Rea, D., and Fulop, V. (2006). Structure-function properties of prolyl oligopeptidase family enzymes. Cell Biochem. Biophys. 44, 349-365.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815.

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, U., Lieberknecht, A., Haslinger, E. (1985). In The Alkaloids, Chemistry and Pharmacology; Brossi, A., Ed.; Academic Press: New York, Vol. 26, pp 299.

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.

Scott, C.P., Abel-Santos, E., Jones, A.D., and Benkovic, S.J. (2001). Structural requirements for the biosynthesis of backbone cyclic peptide libraries. Chem. Biol. 8, 801-815.

Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C., and Benkovic, S.J. (1999). Production of cyclic peptides and proteins *in vivo*. Proc. Natl. Acad. Sci. U. S. A 96, 13638-13643.

Selsted, M.E. (2004). Theta-defensins: cyclic antimicrobial peptides produced by binary ligation of truncated alpha-defensins. Curr. Protein Pept. Sci. 5, 365-371.

Sieber, S.A., and Marahiel, M.A. (2003). Learning from nature's drug factories: nonribosomal synthesis of macrocyclic peptides. J. Bacteriol. 185, 7036-7043.

Siemion, I.Z., Cebrat, M., and Wieczorek, Z. (1999). Cyclolinopeptides and their analogs--a new family of peptide immunosuppressants affecting the calcineurin system. Arch. Immunol. Ther. Exp. (Warsz.) 47, 143-153.

Spengler, J., Jimenez, J.C., Burger, K., Giralt, E., and Albericio, F. (2005). Abbreviated nomenclature for cyclic and branched homo- and hetero-detic peptides. J. Pept. Res. 65, 550-555.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. 6, 493-505.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. 41, 207-234.

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

Tan, N.H., Zhou, J., and Zhao, S.X. (1997). Progress in the chemistry of plant cyclopeptides. Yao Xue. Xue. Bao. 32, 388-399.

Tian, J., Shen, Y., Yang, X., Liang, S., Shan, L., Li, H., Liu, R., and Zhang, W. (2010). Antifungal cyclic peptides from Psammosilene tunicoides. J. Nat. Prod. 73, 1987-1992.

Tong, Y., Luo, J.G., Wang, R., Wang, X.B., and Kong, L.Y. (2012). New cyclic peptides with osteoblastic proliferative activity from *Dianthus superbus*. Bioorg. Med. Chem. Lett. 22, 1908-1911.

Tschesche R. and Kaubmann E.U. (1975). In The Alkaloidss Chemistry and Physiology. Manske R., editor. XV. HF Ademic Press, New York, pp 165–205.

Tseng, C.C., Bruner, S.D., Kohli, R.M., Marahiel, M.A., Walsh, C.T., and Sieber, S.A. (2002). Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. Biochemistry 41, 13350-13359.

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.

Wang, C.K., Kaas, Q., Chiche, L., and Craik, D.J. (2008). CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. Nucleic Acids Res. 36, D206-D210.

Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A., and Holak, T.A. (2000). Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. Structure. 8, 407-418.

White, C.J., and Yudin, A.K. (2011). Contemporary strategies for peptide macrocyclization. Nat. Chem. 3, 509-524.

Wu, Z., Guo, X., and Guo, Z. (2011). Sortase A-catalyzed peptide cyclization for the synthesis of macrocyclic peptides and glycopeptides. Chem. Commun. (Camb.) 47, 9218-9220.

Xu, W., Li, L., Du, L., and Tan, N. (2011). Various mechanisms in cyclopeptide production from precursors synthesized independently of non-ribosomal peptide synthetases. Acta Biochim. Biophys. Sin. (Shanghai) 43, 757-762.