**Characterization of the macrocyclase involved in the biosynthesis of RiPP cyclic peptides in plants**
Chekan, Jonathan R.; Estrada, Paola; Covello, Patrick S.; Nair, Satish K.

National Research Council Canada    Conseil national de recherches Canada

Canada

# Characterization of the macrocyclase involved in the biosynthesis of RiPP cyclic peptides in plants

Jonathan R. Chekan[a], Paola Estrada[a], Patrick S. Covello[b], and Satish K. Nair[a,c,1]

[a]Department of Biochemistry, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [b]National Research Council of Canada, Saskatoon, SK S7N 0W9, Canada; and [c]Center for Biophysics and Computational Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Enzymes that can catalyze the macrocyclization of linear peptide substrates have long been sought for the production of libraries of structurally diverse scaffolds via combinatorial gene assembly as well as to afford rapid in vivo screening methods. Orbitides are plant ribosomally synthesized and posttranslationally modified peptides (RiPPs) of various sizes and topologies, several of which are shown to be biologically active. The diversity in size and sequence of orbitides suggests that the corresponding macrocyclases may be ideal catalysts for production of cyclic peptides. Here we present the biochemical characterization and crystal structures of the plant enzyme PCY1 involved in orbitide macrocyclization. These studies demonstrate how the PCY1 S9A protease fold has been adapted for transamidation, rather than hydrolysis, of acyl-enzyme intermediates to yield cyclic products. Notably, PCY1 uses an unusual strategy in which the cleaved C-terminal follower peptide from the substrate stabilizes the enzyme in a productive conformation to facilitate macrocyclization of the N-terminal fragment. The broad substrate tolerance of PCY1 can be exploited as a biotechnological tool to generate structurally diverse arrays of macrocycles, including those with nonproteinogenic elements.

RiPP | biosynthesis | peptide | plant | orbitide

Macrocyclic peptides have become appealing targets for drug discovery efforts due to the emergence over the past decade of multiple routes for rapid synthesis and screening (1). The drug-like properties of cyclic peptides arise from their constrained rigid structure, improved bioavailability, membrane permeability relative to linear peptides, and resistance to degradation by host proteases (2–4). Organic synthesis of large peptide libraries is constrained by practical concerns, thereby limiting the diversity of sequence variants that would be necessary to truly explore scaffold space. More recent approaches for macrocyclic peptide synthesis have focused on in vivo production, in which the use of genetic templates can provide routes toward the production of libraries of diverse structures (5). Enzymatic routes for macrocycle production include the use of isolated thioesterase domains from non-ribosomal peptide synthetases (6), GST (7), or split inteins (5), as well as technologies based on reprogramming the genetic code (8). However, these approaches either require the use of conjugated biomimetics or are not sufficiently substrate-tolerant to thoroughly sample the diversity of possible structures.

Ribosomally synthesized and posttranslationally modified peptides (RiPPs) represent an abundant source of chemically and structurally diverse natural products (9). RiPPs are translated by the ribosome as linear peptides that are subsequently enzymatically modified to yield compounds with a range of biological activities (10). Many RiPPs undergo enzymatic macrocyclization, as observed in several diverse classes of compounds including bottromycins (11), cyanobactins (12), lanthipeptides (13), streptides (14), lasso peptides (15), and thiopeptides (16). Understanding the biosynthetic pathways for RiPP macrocyclization continues to be important for biotechnological application, for the identification of new gene clusters, and for the production of novel RiPP derivatives (17).

Orbitides encompass a class of plant homodetic macrocyclic natural products, usually 5–12 amino acids in size, which are characterized by the intramolecular condensation between the N and C termini of their linear precursors (10). The chemical structures of orbitides consist solely of α-amide linkages and are distinguished from other macrocyclic plant peptide natural products such as cyclotides, which are further interlocked through multiple disulfide bonds (18), and amatoxins, which contain an unusual tryptathionine moiety formed by a crosslink between a Cys and Trp residue (19). The first macrocyclic orbitide, cyclo-linopeptide A, was isolated in 1959 from flaxseed (*Linum usita-tissimum*) oil and determined to be a cyclic peptide of ILVPPFFLI (Fig. 1*A*) (20). Since then, over 168 Caryophyllaceae-like nonredundant orbitides have been identified, and encoding sequences have been confirmed in Annonaceae, Rutaceae, Euphorbiaceae, and Linaceae (21). Many orbitides demonstrate biological activities including antimalarial (22), vasodilatory (23), immunomodulating (24), and xenoestrogenic activities (25).

Analysis of expressed sequence tag libraries from developing seeds of *Saponaria vaccaria* (*Vaccaria hispanica*) identified genes that likely encode for peptide precursors of orbitides. Expression of candidate precursors in the roots of a nonsegetalin A-producing *S. vaccaria* strain resulted in the production of fully processed cyclic peptide products (26), confirming that orbitides are indeed RiPPs. Subsequent biochemical assays with seed extracts, using synthetic peptide substrates, revealed the order of posttranslational modifications that generate the orbitide segetalin A (Fig. 1 *A* and *B*) (27). Briefly, segetalin A is first produced ribosomally as a 32-aa peptide termed presegetalin A1. A serine protease (oligopeptidase 1 or OLP1) cleaves the first 15 amino acids to yield a product containing an N-terminal glycine (termed presegetalin A1 [14–32]). A second enzyme [peptide cyclase 1 (PCY1)] acts on this product to excise the linear C-terminal

## Significance

The class of bioactive cyclic plant natural products called orbitides was first identified nearly half a century ago. Here we describe how a single enzyme can catalyze the cyclization of a range of ribosomally synthesized linear peptides into the corresponding cyclic products of varying ring sizes. These studies may provide a means for producing large libraries of cyclic peptides without any sequence bias.

BIOCHEMISTRY

**Fig. 1.** (A) Structures of the orbitides cyclolinopeptide A, segetalin A, and ribifolin. (B) Scheme showing the biosynthesis of segetalin A by *S. vaccaria*.

13 residues (termed presegetalin A1 [20–32]), with the concomitant macrocyclization of the remaining N-terminal six residues to yield the cyclo[GVPVWA] segetalin A product (Fig. 1*B*). The biosynthesis of cyclotides (28, 29), amatoxins (30), and cyanobactins (31) uses similar biosynthetic strategies in which a sequence-specific peptidase catalyzes the transamidation of a linear peptide substrate to produce a macrocyclic product.

Primary sequence analysis identifies PCY1 as a member of the S9A protease family that includes several clades of prolyl oligopeptidases (POP). Proteases of the S9A class catalyze the hydrolytic cleavage of peptide substrates at the amide linkage C-terminal to Pro to yield two linear products. However, PCY1 is distinguished from canonical S9A proteases by its unique product profile, as the enzyme generates a macrocyclic product rather than the linear counterpart. In addition, PCY1 does not require a Pro in the sequence of the cyclized peptide. Analysis of the predicted primary sequences of dozens of putative orbitide precursor peptides reveals that residues within the core sequence that is cyclized are highly divergent, whereas residues in the preceding leader sequence (removed by OLP1) and the trailing follower sequence (excised as a linear peptide by PCY1) are highly conserved. Similar patterns of conservation are observed for the precursors of other cyclic peptides including cyanobactins (31) and amatoxins (19, 32). Alanine scanning mutational analysis of presegetalin A1 demonstrates that the necessary elements for substrate recognition by PCY1 reside solely

in the 12-residue follower sequence (hereafter, presegetalin A1 [20–32]) that is excised during the formation of the macrocyclic product (27). Notably, PCY1 can accept a range of substrates leading to rings of different sizes as this single enzyme is capable of processing almost all of the orbitides that are observed in *S. vaccaria*. Consequently, PCY1 may be used as a biotechnology tool for the production of macrocycles of diverse sequences. A similar strategy has been used by employing the peptide cyclase butelase 1 to produce cyclic bacteriocins (33).

The structure of the unrelated cyanobactin transamidating protease (PatG) revealed a peptidase S8 (subtilisin-like) fold that is decorated with a "capping" domain, and biochemical studies suggest that this domain insertion likely facilitates macrocyclization of the peptide substrate (34, 35). The primary sequence of PCY1 fails to identify any such obvious insertions, relative to canonical S9A protease family members, raising the question of how an enzyme from a class of generic proteases can be adapted to produce a macrocyclic product. To address this issue, we have carried out biochemical and crystallographic characterization of PCY1.

## Results

**Kinetic Analysis of PCY1.** Incubation of recombinant PCY1 with a synthetic peptide corresponding to presegetalin A1 [14–32] resulted in a near-complete conversion into cyclic[GVPVWA] segetalin A (and corresponding linear follower peptide) (Fig. 2). The yield of linear core peptide was minimal, demonstrating that the enzyme favors production of macrocyclic products relative to the linear counterpart. To determine steady-state kinetic parameters, we used a LC-MS–based assay to quantify rates of product formation. Complete Michaelis–Menten kinetics were carried out for the presegetalin A1 [14–32] substrate, yielding a $k_{cat}$ value of $6.0 \times 10^{-2} \cdot s^{-1}$ with a $K_M$ of 0.77 µM (*SI Appendix*, Fig. S1 and Table S3). In comparison, the asparaginyl endopeptidases involved in cyclotide macrocyclization (butelase 1 and AEP) have similar $k_{cat}$ values, but with $K_M$ values that are 5- to 10-fold higher (28, 29). Additionally, the transamidative protease involved in the biosynthesis of amatoxin (GmPOPB) was initially shown to have a similar $K_M$ value and a 100-fold faster rate ($k_{cat}$ of $5.7 \cdot s^{-1}$) (30), whereas more extensive kinetic analysis indicated a $K_M$ about 75-fold higher and a $k_{cat}$ approximately 4-fold faster than PCY1 (36).



**Fig. 2.** Extracted-ion chromatograms of products derived upon incubation of recombinant PCY1 with a synthetic peptide corresponding to presegetalin A1 [14–32]. Chromatograms showing the mass to charge ratio (*m/z*) values of the segetalin A product (*m/z* = 610.5, *z* = 1; green trace), of the linear [GVPVWA] product (*m/z* = 628.5, *z* = 1; red trace), of the presegetalin A1 [14–32] substrate (*m/z* = 993.2, *z* = 2; blue trace), and the presegetalin A1 [20–32] by-product (*m/z* = 688.5, *z* = 2; purple trace) are shown as indicated. The wild-type PCY1 used most of the presegetalin A1 [14–32] substrate to produce almost exclusively cyclic segetalin A product after a 1-h reaction.

**Structural Characterization of PCY1.** To elucidate how an S9A protease family can catalyze macrocyclization, we determined the crystal structure of PCY1 that had been incubated with the presegetalin A1 [14–32] substrate (Fig. 3 *A* and *B*) (37). The 1.9-Å resolution structure of this PCY1 complex revealed an overall architecture similar to other prolyl oligopeptidase family members, such as the eukaryotic enzyme from porcine muscle (*Sus scorfa*) (PDB ID code: 1QFS; rmsd of 1.6 Å over 691 aligned Cα atoms) (38, 39) and the bacterial enzyme from *Myxococcus xanthus* (PDB ID code: 2BKL; rmsd of 1.7 Å over 691 aligned Cα atoms) (40) (Fig. 3 *C* and *E*). The structure of PCY1 consists of an α/β hydrolase domain composed of residues Met1-Val77 and Asp437-Asp724 and a β-propeller domain consisting of residues Cys82-Glu432. A hinge region encompassing residues Asn78-Arg81 and Ser433-Pro436 links the two domains and presumably facilitates movement between an open form competent for binding substrate and a closed form where chemistry is carried out (41, 42). The PCY1–presegetalin A1 complex was crystallized in the closed form with a large internal cavity (10,317 Å$^3$ as calculated by CASTp), which can accommodate the binding of large peptides (41, 43). A close-up view of the active sites of canonical prolyl oligopeptidases from *S. scorfa* and *M. xanthus* reveal the protrusion of various loops that significantly decrease the corresponding internal cavity (Fig. 3 *D* and *F*). In PCY1, residues Asp653, His695, and Ser562 make up the active-site catalytic triad, and a hydrophobic pocket enriched in aromatic residues forms a binding pocket for the target peptide (Fig. 4*A*). Finally, Tyr481 acts to stabilize the tetrahedral intermediate oxyanion that is formed following the attack of the Ser562 alkoxide on the peptide substrate.

Although PCY1 crystallization was carried out using presegetalin A1 [14–32], the active site revealed unambiguous electron density only for the terminal six residues (NASAPV) of the follower peptide (Fig. 3*B*). The residues within the follower sequence bind to PCY1 near the hinge region and engage in hydrogen bonds with amino acids in both the α/β hydrolase and the β-propeller domains. Hydrogen-bonding interactions can be observed between PCY1 residues Arg81, Asn97, Asn104, and backbone carbonyl and amide of Ala28 and Ser29, respectively, of the follower peptide. The only sequence-specific interaction with the follower peptide occurs between its side chain of



**Fig. 4.** (*A*) Active-site alignment of PCY1 (red, pink) and porcine muscle POP (cyan, blue) with covalently bound ZPP. All residues are conserved and align well except for the catalytic histidine. Residues are labeled according to PCY1's numbering. (*B*) Sequence alignment of PCY1 and POP enzymes from *S. scorfa* [domestic pig, National Center for Biotechnology Information (NCBI): NP_001004050.1], *Beta vulgaris* (beet, NCBI: XP_010667346.1) and *Morus notabilis* (mulberry, NCBI: XP_010101294.1). (*C*) Changes in structures at regions distal to the active site between PCY1 (red) and *S. scorfa* POP (blue) indicate significant difference in the orientation of the hairpin loop. The hairpin is stabilized in PCY1 by a stacking interaction between His149 and Tyr696.

Asn27 and the side chain of Asp104 and backbone carbonyl of Leu132 of PCY1. The binding of the carboxyl terminus of the peptide is of particular interest, as it forms hydrogen bonds with PCY1 residues Ser495 and Ser493. This observation explains prior alanine-scanning mutational analysis of presegetalin A1 [14–32], which demonstrates that PCY1 is tolerant of mutation along the entirety of the primary sequence of the substrate but cannot process a substrate with the deletion of the C-terminal Val32, indicating the importance of the follower peptide C-terminal carboxylate (27).

To visualize a structure of PCY1 bound to intact presegetalin A1, we carried out cocrystallization with two different active-site variants, His695→Ala and Tyr481→Phe. In both instances, electron density could be observed only for the six-residue follower sequence, despite the presence of the presegetalin A1 [20–32] peptide in the crystals (as determined by mass spectrometry), consistent with disorder/flexibility for the remainder of the substrate (*SI Appendix*, Figs. S2 and S3). Hence, only the follower region of presegetalin A1 is engaged by PCY1, allowing for flexibility in the substrate N-terminal to the cleavage site. This flexibility explains how *S. vaccaria* can produce nine different orbitides of different ring sizes and sequence composition, as only the follower sequences are conserved in the progenitors of the macrocyclic products (*SI Appendix*, Fig. S4).

Most RiPP biosynthetic enzymes use a PqqD domain (RiPP precursor peptide recognition element) (44) to engage the leader precursor peptide through a limited number of hydrophobic interactions. This has been observed in at least some of the enzymes involved in the biosynthesis of pyrroloquinoline quinone (45), lantibiotics (46), lasso peptides (44), and cyanobactins (47). In contrast, the follower peptide binding observed in PCY1 does not rely on hydrophobic interactions, but rather on hydrogen-bonding interactions. Hydrogen-bonding interactions also mediate binding of leader peptides by microvirdin biosynthetic enzymes (48) and of the follower peptide by PatG (34).



**Fig. 3.** (*A*) The 1.9-Å PCY1 structure with the α/β hydrolase domain shown in green and β-propeller domain shown in red. (*B*) The C-terminal six amino acids of the presegetalin A1 [14–32] substrate (orange) is bound near the hinge region of PCY1. A simulated annealing difference Fourier maps (F$_o$-F$_c$), calculated with the coordinates of the peptide omitted, is superimposed and contoured to 2.5σ (blue). (*C*) Structure of the canonical POP enzyme from *S. scorfa* (PDB ID code 1QFS) (38, 39) bound to covalent inhibitor ZPP, with the active site demarcated in the rectangle, and (*D*) a close-up view of the active site. (*E*) Structure of the bacterial enzyme from *M. xanthus* (PDB ID code 2BKL) (40) bound to covalent inhibitor Z-Ala-prolinal, with the active site demarcated, and (*F*) a close-up view of the active site.

**PCY1-Binding Affinity Measurements.** To determine the affinity of PCY1 for various follower peptides, we carried out competition binding measurements (Table 1). A six-residue presegetalin A1 [27–32] peptide conjugated to fluorescein isothiocyanate bound to PCY1 with a dissociation constant ($K_d$) of 4.83 µM as determined by fluorescence polarization (*SI Appendix*, Fig. S5A). Competitive binding between this labeled peptide and various unlabeled peptides was used to determine the $K_i$ values for the unlabeled peptides. These competition experiments demonstrate that unlabeled presegetalin A1 [27–32] bound to PCY1 with a $K_i$ of 31.0 µM (Table 1 and *SI Appendix*, Fig. S5B). The longer, full-length linear by-product of the cyclization reaction, presegetalin A1 [20–32], binds with a much higher affinity with a $K_i$ of 0.40 µM (Table 1 and *SI Appendix*, Fig. S5C). The nearly 80-fold greater binding affinity for the longer presegetalin A1 [20–32] peptide is curious given that only the last six residues of the peptide are involved in contacts with PCY1 in the cocrystal structure. The additional binding affinity, relative to presegetalin A1 [27–32], may be due to solvent exclusion or another nonsequence-dependent mechanism. The amino acids that make up the final cyclic product do not seem to contribute to binding as substrate presegetalin A1 [14–32] was measured to bind to the catalytically inhibited PCY1 H695A with a $K_i$ of 1.10 µM, very similar to that of presegetalin A1 [20–32] (Table 1 and *SI Appendix*, Fig. S5D). This contrasts with binding studies completed with GmPOPB, in which the cyclized region of the peptide enhances binding affinity (36).

To verify the roles of individual interactions observed in the structure, we measured the affinity of PCY1 for variants of the presegetalin A1 [27–32] peptide. Based on the crystal structure, the side chain of Asn27 of presegetalin A1 [27–32] is engaged in specific hydrogen-bonding interactions. Consequently, the Asn27→Ala mutation in presegetalin A1 [27–32] results in a reduction in the PCY1-binding affinity ($K_i$ of 131 µM) (Table 1 and *SI Appendix*, Fig. S5E). The C terminus of the follower peptide forms hydrogen bonds with PCY1 residues Ser495 and Ser493, and the binding affinity of PCY1 for the presegetalin A1 [27–32] variant with a C-terminal amide is diminished and outside the detection range of the assay (Table 1 and *SI Appendix*, Fig. S5F). Mutations were made to PCY1 residues Ser493 and Asn97, both of which bind the follower peptide. Ser493→Ala had minimal changes on catalytic efficiency relative to the wild type, whereas Asn97→Ala greatly diminished activity (*SI Appendix*, Table S3 and Fig. S6).

**Follower Peptide Induces Conformational Change in PCY1.** Based on the kinetic studies, we hypothesized that the high affinity of PCY1 for the presegetalin A1 [27–32] follower peptide may ensure that the enzyme is maintained in the closed conformation throughout catalysis and is displaced only after turnover. To test this, we fluorescently labeled PCY1 with the thiol-specific probe *N*-1-pyrene maleimide. Addition of the presegetalin A1 [27–32] to labeled PCY1 resulted in the quenching of fluorescence compared with a buffer control (*SI Appendix*, Fig. S7). Conversely, addition of a control peptide that does not bind (presegetalin A1 [27–32] amide) showed no change in fluorescence. As there are no cysteines present near the binding site of the

peptide (*SI Appendix*, Fig. S8), these changes in fluorescence have been attributed to changes in the conformation of PCY1. These findings are consistent with the notion that binding of presegetalin A1 [27–32] alone induces a conformational change in PCY1 to the closed state.

As presegetalin A1 [27–32] binding induces closure of the PCY1 structure, addition of exogenous peptide would be expected to inhibit productive turnover of the presegetalin A1 [14–32] substrate. We tested this hypothesis using a 1-h end-point reaction of PCY1 incubated with 30 µM of substrate in the presence of stoichiometric and 10-fold excess of the presegetalin A1 [20–32] linker+follower peptide product, as well as 100-fold excess of the presegetalin A1 [27–32] follower peptide (*SI Appendix*, Fig. S9). Although PCY1 consumed nearly all of the presegetalin A1 [14–32] substrate alone (Fig. 2), addition of the follower peptides resulted in a decrease in turnover in a concentration-dependent manner. These data are consistent with the follower peptide maintaining a closed state for PCY1. The role of the leader/follower peptide in stabilizing active conformations of biosynthetic enzymes is an emerging theme in RiPP biosynthesis. For example, during cyanobactin biosynthesis, binding of the PatE′ leader peptide to the LynD cyclodehydratase stabilizes the enzyme in a catalytically competent form (47).

Using the structural data as a guide, we generated and characterized mutants of several active-site residues to test their roles in catalysis (*SI Appendix*, Table S3 and Fig. S10A). As expected, both the Ser562→Ala and His695→Ala mutation of the catalytic triad residues produced inactive enzyme, confirming the necessity of these conserved residues. Finally, Tyr481 is necessary to stabilize the oxyanion of the tetrahedral intermediate as the Try481→Phe variant is catalytically inactive. We next interrogated the extent to which the enzyme could accommodate primary amines other than a peptide α-amino group for transamidation. First, we used a variant presegetalin A1 [14–32] substrate peptide that contained both the α-amino terminus and a side-chain amine (Val15→Lys). Kinetic characterization of PCY1 using the Val15→Lys presegetalin A1 [14–32] as a substrate demonstrates that the $K_M$ increased by 10-fold relative to the cognate presegetalin A1 [14–32] peptide as a substrate but the $k_{cat}$ also increased by almost 10-fold, resulting in a catalytic efficiency similar to that using the cognate peptide. Although PCY1 was able to completely convert this substrate to the cyclic product (*SI Appendix*, Table S3 and Fig. S10B), tandem mass spectrometric analysis revealed that the α-amino group was used for transamidation (*SI Appendix*, Fig. S11). We did not observe any products consistent with the use of the side chain of Lys15 as part of the macrocycle. We then tested whether PCY1 could generate a macrocyclic product using a nonproteinogenic amine. To this end, we used a presegetalin A1 [14–32] substrate that contained an N-terminal aminohexanoic acid (Ahx) in place of Gly14. PCY1 largely produced a linear product with this substrate, although a minor amount of cyclic [AhxVPVWA] could also be detected (*SI Appendix*, Table S3 and Fig. S12). These data suggest that PCY1-catalyzed macrocyclization is most effective via the use of a peptide α-amine.

**Dual Function of His695 Allows for Cyclization.** Prior studies on other macrocyclases, such as the PatG protease involved in cyanobactin biosynthesis, suggest a plausible route to cyclic peptide production (34, 35). Briefly, following the formation of an acyl-enzyme intermediate during the first half protease reaction, hydrolysis of the intermediate is averted by solvent exclusion from the active site. Instead, the α-amino terminus of the bound peptide substrate is oriented back toward the catalytic Ser, and deprotonation of the amine can then facilitate a transpeptidation reaction to yield the cyclic peptide product. A comparison of the structure of PCY1 with other prolyl oligopeptidases reveals a strong conservation of overall structure and a near-identical alignment of putative active-site residues (Fig. 4A). While nearly all of these features in the PCY1 active site superimpose with the structures of

**Table 1. Binding constants for the follower peptide and mutants**

| Ligand | IC$_{50}$ (µM) | $K_i$ (µM) |
|---|---|---|
| Presegetalin A1 [20–32] | 1.97 ± 0.34 | 0.40 ± 0.007 |
| Presegetalin A1 [14–32]* | 3.03 ± 0.61 | 1.10 ± 0.22 |
| Presegetalin A1 [27–32] | 47.4 ± 1.0 | 31.0 ± 0.65 |
| Presegetalin A1 [27–32] N27A | 197 ± 8 | 131 ± 5 |
| Presegetalin A1 [27–32] amide | >1,000 | >650 |

*Presegetalin A1 [14–32] binding experiments were completed with PCY1 H695A to prevent turnover.

other S9A proteases, a notable exception is the ~2.4-Å displacement of His695 away from the catalytic Ser562.

Studies of hydrolytic serine proteases have established that the catalytic His moves following formation of a tetrahedral intermediate, both to tune pK$_a$ and to serve as an effective proton donor to the leaving group (49–51). To determine if the misaligned His695 is reflective of a catalytically competent active site, we determined the 3.3 Å cocrystal structure of PCY1 covalently bound to an aldehyde inhibitor that mimics a tetrahedral intermediate [Z-Proprolinal (ZPP)] (Fig. 4A). Although the structure of ZPP differs from that of the true presegetalin substrate peptide, PCY1 is sufficiently tolerant for the sequence preceding the scissile bond to accommodate this inhibitor. The cocrystal structure shows that ZPP is covalently bound to Ser562 and adopts a conformation similar to that in canonical S9A enzymes, with the Pro oriented within the aromatic hydrophobic pocket. Surprisingly, the disposition of His695 relative to the unliganded enzyme remained unchanged (0.9-Å movement of Nγ) and was outside the range of a hydrogen-bonding interaction with Ser562.

The obvious change in the orientation of the catalytic His695 suggested that it may play some role in directing the ability of PCY1 to catalyze macrocyclization. A cursory inspection of the PCY1 active site revealed a single residue deletion at a Gly that is normally adjacent to the catalytic His in canonical S9A proteases (Fig. 4B). Surmising that this deletion may force the displacement of His695 away from Ser562, we generated a Gly insertion mutant in PCY1, which would mimic the canonical S9A protease active site. If the misalignment of His695 indeed plays a role in the macrocyclase activity of PCY1, this insertion mutant should generate linear, hydrolytic products, similar to that of canonical S9A proteases. Notably, this PCY1 mutant slowly produces a significant amount of the linear, hydrolytic product (ratio of cyclic:linear product of ~3.7:1), whereas the wild-type enzyme almost exclusively produces the macrocyclic product (Fig. 2 and *SI Appendix*, Table S3 and Fig. S13). These data support the assertion that the altered position of His695 plays some role in macrocycle formation by PCY1.

Although the insertion mutants showed the expected phenotype of production of both linear and cyclic products, the catalytic competency of the enzyme was severely compromised. To determine the basis for the diminished activity, we determined the 2.8-Å resolution structure of the PCY1 Gly696 insertion variant. The structure reveals significant disorder in the loop harboring His695 and could not be confidently modeled. A more detailed comparison of wild-type PCY1 against canonical S9A proteases reveals additional structural deviations near the active site, relative to canonical S9A proteases, in addition to the single Gly deletion (Fig. 4C). In particular, in canonical S9A enzymes, such as porcine prolyl oligopeptidase (PDB ID code 1QFS), the loop of an adjacent β-hairpin (encompassing Leu143 through Val156) is longer and buttresses the orientation of the catalytic His in the active site. In PCY1, the equivalent loop is one residue shorter and is displaced away from the active site. The side chain of Tyr696 (normally an Ala) occupies the volume created by this loop displacement. The Tyr696→Ala/Gly insertion double mutant had a greatly compromised activity, further underscoring that the repositioning of His695 in PCY1 involves a confluence of many local changes (*SI Appendix*, Fig. S14).

Based on the above data, we hypothesized that the His695 in PCY1 may serve two roles. It could serve to deprotonate the catalytic Ser562 (as in other serine proteases), but also may deprotonate the α-amino terminus of the peptide substrate to facilitate transamidation. To further investigate the dual functionality of the catalytic His, we analyzed the PCY1 His695→Ala variant under single turnover conditions (Fig. 5A). If His695 does play a role in the deprotonization of Ser562, this variant should still be unable to yield the hydrolytic, linear presegetalin A1 [20–32] by-product that forms after generation of the acyl-enzyme intermediate. As expected, the His695→Ala variant was completely inactive, indicating that

His695 is necessary for activation of Ser562 despite the long distance between the two residues. Next, we sought to determine if His695 played a role in proton removal from the substrate α-amino group by testing the activity of PCY1 His695→Ala in the presence of imidazole. At the pH of the assay (pH = 8.5), imidazole should be deprotonated and may be able to function as an imperfect, mobile surrogate for His. Surprisingly, the addition of 20 mM imidazole resulted in a gain of function for the His695→Ala variant, which could generate both linear and cyclic [GVPVWA] segetalin A products from the presegetalin A1 [14–32] substrate (Fig. 5B). Competition between the mobile imidazole surrogate and solvent for attack on the acyl-enzyme intermediate may account for the observance of both linear and cyclic products in the His695→Ala variant. This result suggests that imidazole (and presumably His695 in the wild-type enzyme) can both deprotonate Ser562 to facilitate formation of an acyl-enzyme adduct and deprotonate the α-amino terminus of the substrate to facilitate transamidation.

## Discussion

Recent studies of various RiPP biosynthetic enzymes have provided several examples of biocatalysts that can generate a macrocyclic product from a linear peptide substrate (14, 28, 30, 31, 52–54). Such enzymes represent a next-generation technological utility, but many of these catalysts are often slow or require sequence constraints on substrates. The biosynthesis of orbitides, as well as other RiPPs such as cyclotides (28, 29), amatoxins (30), and cyanobactins (31), all use convergent strategies based on Ser proteases that can catalyze the formation of a cyclic product via transamidation, rather than a linear product via hydrolysis. Structural and biochemical studies of the cyanobactin macrocyclase PatG reveal that the enzyme has a helix-turn-helix insertion that sits above the active site, which presumably shields the acyl-enzyme intermediate from solvent. Full enclosure of the PatG active site upon binding of substrate is completed by the AYDG follower peptide sequence necessary for substrate recognition. The structural and biochemical data presented here on the orbitide macrocyclase PCY1 reveals a more nuanced strategy of achieving the same result. Rather than containing obvious insertions, transamidation by PCY1 is facilitated by the binding of the follower peptide, which maintains the enzyme in a closed state that precludes solvent from the active site, potentially limiting the competing hydrolysis reaction. The selectivity for substrates is maintained via recognition of the follower sequence, and the lack of appreciable affinity for the residues to be cyclized (as demonstrated by fluorescence polarization data) allows for product diversity. Specificity dictated by recognition of the follower peptide is not common in RiPP biosynthetic enzymes, and the only other known cases are of enzymes involved in the biosynthesis of cyanobactins and bottromycin (34).
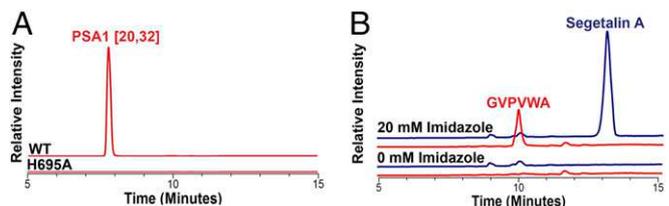


**Fig. 5.** (A) Formation of presegetalin A1 [20,32] was observed under single turnover conditions by LC-MS. After a 5-min incubation, wild-type PCY1 was able to catalyze formation of the acyl-enzyme intermediate, producing the presegetalin A1 [20,32] by-product. The H695A mutant did not catalyze formation of presegetalin A1 [20,32], indicating that it could not form the acyl-enzyme intermediate. (B) An 18-h incubation of H695A with 20 mM imidazole allowed for the production of both linear [GVPVWA] and cyclic segetalin A. However, the H695A variant produces a significantly greater amount of hydrolytic, linear product (ratio of cyclic:linear product of ~3.7:1), whereas wild-type PCY1 almost exclusively produces the macrocyclic product (Fig. 2).

The active site of PCY1 is also altered to more efficiently catalyze the cyclization reaction. The placement of His695 facilitates two roles: the canonical role of activating the Ser nucleophile to help form an acyl-enzyme intermediate and an unexpected role in deprotonating the α-amine of the peptide substrate to facilitate transamidation (*SI Appendix*, Fig. S15). The dual functionality for His695 is aided by its location in a mobile loop, which could accommodate the range of movements necessary for each function. Our biochemical studies provide a foundation for understanding the mechanistic basis for peptide macrocyclization in plants, an activity that is observed in growing classes of RiPP natural products.

## Materials and Methods

Methods describing the cloning, expression, purification, biochemical analysis, kinetic and biochemical characterization, and crystallization of PCY1 and ligand complexes are described in detail in *SI Appendix, SI Materials and Methods*. Chemical data and analytical methods are also provided.

1. Passioura T, Katoh T, Goto Y, Suga H (2014) Selection-based discovery of druglike macrocyclic peptides. *Annu Rev Biochem* 83:727–752.
2. Driggers EM, Hale SP, Lee J, Terrett NK (2008) The exploration of macrocycles for drug discovery: An underexploited structural class. *Nat Rev Drug Discov* 7:608–624.
3. White TR, et al. (2011) On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nat Chem Biol* 7:810–817.
4. Rezai T, Yu B, Millhauser GL, Jacobson MP, Lokey RS (2006) Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *J Am Chem Soc* 128:2510–2511.
5. Scott CP, Abel-Santos E, Wall M, Wahnon DC, Benkovic SJ (1999) Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci USA* 96:13638–13643.
6. Kohli RM, Walsh CT, Burkart MD (2002) Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature* 418:658–661.
7. Zhang C, Dai P, Spokoyny AM, Pentelute BL (2014) Enzyme-catalyzed macrocyclization of long unprotected peptides. *Org Lett* 16:3652–3655.
8. Hipolito CJ, Suga H (2012) Ribosomal production and in vitro selection of natural product-like peptidomimetics: The FIT and RaPID systems. *Curr Opin Chem Biol* 16:196–203.
9. Velásquez JE, van der Donk WA (2011) Genome mining for ribosomally synthesized natural products. *Curr Opin Chem Biol* 15:11–21.
10. Arnison PG, 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.
11. Shimamura H, et al. (2009) Structure determination and total synthesis of bottromycin A2: A potent antibiotic against MRSA and VRE. *Angew Chem Int Ed Engl* 48:914–917.
12. Sivonen K, Leikoski N, Fewer DP, Jokela J (2010) Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Appl Microbiol Biotechnol* 86:1213–1225.
13. Gross E, Morell JL (1971) The structure of nisin. *J Am Chem Soc* 93:4634–4635.
14. Schramma KR, Bushin LB, Seyedsayamdost MR (2015) Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nat Chem* 7:431–437.
15. Hegemann JD, Zimmermann M, Xie X, Marahiel MA (2015) Lasso peptides: An intriguing class of bacterial natural products. *Acc Chem Res* 48:1909–1919.
16. Bagley MC, Dale JW, Merritt EA, Xiong X (2005) Thiopeptide antibiotics. *Chem Rev* 105:685–714.
17. Sardar D, Lin Z, Schmidt EW (2015) Modularity of RiPP enzymes enables designed synthesis of decorated peptides. *Chem Biol* 22:907–916.
18. Craik DJ, Malik U (2013) Cyclotide biosynthesis. *Curr Opin Chem Biol* 17:546–554.
19. Hallen HE, Luo H, Scott-Craig JS, Walton JD (2007) Gene family encoding the major toxins of lethal Amanita mushrooms. *Proc Natl Acad Sci USA* 104:19097–19101.
20. Kaufmann HP, Tobschirbel A (1959) An oligopeptide from flaxseed. *Chem Ber* 92:2805–2809.
21. Gui B, et al. (2012) Identification and quantification of cyclolinopeptides in five flaxseed cultivars. *J Agric Food Chem* 60:8571–8579.
22. Pinto MEF, et al. (2015) Ribifolin, an orbitide from Jatropha ribifolia, and its potential antimalarial activity. *J Nat Prod* 78:374–380.
23. Morita H, et al. (2006) Structure of a new cyclic nonapeptide, segetalin F, and vasorelaxant activity of segetalins from Vaccaria segetalis. *Bioorg Med Chem Lett* 16:4458–4461.
24. Gaymes TJ, Cebrat M, Siemion IZ, Kay JE (1997) Cyclolinopeptide A (CLA) mediates its immunosuppressive activity through cyclophilin-dependent calcineurin inactivation. *FEBS Lett* 418:224–227.
25. Itokawa H, Yun Y, Morita H, Takeya K, Yamada K (1995) Estrogen-like activity of cyclic peptides from Vaccaria segetalis extracts. *Planta Med* 61:561–562.
26. Condie JA, et al. (2011) The biosynthesis of Caryophyllaceae-like cyclic peptides in Saponaria vaccaria L. from DNA-encoded precursors. *Plant J* 67:682–690.
27. Barber CJS, et al. (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.
28. Nguyen GKT, et al. (2014) Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat Chem Biol* 10:732–738.
29. Bernath-Levin K, et al. (2015) Peptide macrocyclization by a bifunctional endoprotease. *Chem Biol* 22:571–582.
30. Luo H, et al. (2014) Peptide macrocyclization catalyzed by a prolyl oligopeptidase involved in α-amanitin biosynthesis. *Chem Biol* 21:1610–1617.
31. Lee J, McIntosh J, Hathaway BJ, Schmidt EW (2009) Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *J Am Chem Soc* 131:2122–2124.
32. Li P, Deng W, Li T (2014) The molecular diversity of toxin gene families in lethal Amanita mushrooms. *Toxicon* 83:59–68.
33. Hemu X, Qiu Y, Nguyen GKT, Tam JP (2016) Total synthesis of circular bacteriocins by butelase 1. *J Am Chem Soc* 138:6968–6971.
34. Koehnke J, et al. (2012) The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nat Struct Mol Biol* 19:767–772.
35. Agarwal V, Pierce E, McIntosh J, Schmidt EW, Nair SK (2012) Structures of cyanobactin maturation enzymes define a family of transamidating proteases. *Chem Biol* 19:1411–1422.
36. Czekster CM, Naismith JH (2017) Kinetic landscape of a peptide bond-forming prolyl oligopeptidase. *Biochemistry* 56:2086–2095.
37. Tsilikounas E, Kettner CA, Bachovchin WW (1992) Identification of serine and histidine adducts in complexes of trypsin and trypsinogen with peptide and nonpeptide boronic acid inhibitors by 1H NMR spectroscopy. *Biochemistry* 31:12839–12846.
38. Holm L, Rosenström P (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res* 38:W545-9.
39. Fülöp V, Böcskei Z, Polgár L (1998) Prolyl oligopeptidase: An unusual beta-propeller domain regulates proteolysis. *Cell* 94:161–170.
40. Shan L, Mathews II, Khosla C (2005) Structural and mechanistic analysis of two prolyl endopeptidases: Role of interdomain dynamics in catalysis and specificity. *Proc Natl Acad Sci USA* 102:3599–3604.
41. Li M, Chen C, Davies DR, Chiu TK (2010) Induced-fit mechanism for prolyl endopeptidase. *J Biol Chem* 285:21487–21495.
42. Polgár L (2002) The prolyl oligopeptidase family. *Cell Mol Life Sci* 59:349–362.
43. Dundas J, et al. (2006) CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34:W116-8.
44. Burkhart BJ, Hudson GA, Dunbar KL, Mitchell DA (2015) A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat Chem Biol* 11:564–570.
45. Tsai TY, Yang CY, Shih HL, Wang AHJ, Chou SH (2009) Xanthomonas campestris PqqD in the pyrroloquinoline quinone biosynthesis operon adopts a novel saddle-like fold that possibly serves as a PQQ carrier. *Proteins* 76:1042–1048.
46. Ortega MA, et al. (2015) Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature* 517:509–512.
47. Koehnke J, et al. (2015) Structural analysis of leader peptide binding enables leader-free cyanobactin processing. *Nat Chem Biol* 11:558–563.
48. Li K, Condurso HL, Li G, Ding Y, Bruner SD (2016) Structural basis for precursor protein-directed ribosomal peptide macrocyclization. *Nat Chem Biol* 12:973–979.
49. Cleland WW (2000) Low-barrier hydrogen bonds and enzymatic catalysis. *Arch Biochem Biophys* 382:1–5.
50. Radisky ES, Lee JM, Lu C-JK, Koshland DE, Jr (2006) Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc Natl Acad Sci USA* 103:6835–6840.
51. Topf M, Richards WG (2004) Theoretical studies on the deacylation step of serine protease catalysis in the gas phase, in solution, and in elastase. *J Am Chem Soc* 126:14631–14641.
52. Duquesne S, et al. (2007) Two enzymes catalyze the maturation of a lasso peptide in Escherichia coli. *Chem Biol* 14:793–803.
53. Yan K-P, et al. (2012) Dissecting the maturation steps of the lasso peptide microcin J25 in vitro. *ChemBioChem* 13:1046–1052.
54. Li B (2006) et al. Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. *Science* 311:1464–1467.

# Characterization of the Macrocyclase Involved the Biosynthesis of RiPP Cyclic Peptides in Plants

Jonathan R. Chekan[1], Paola Estrada[1], Patrick S. Covello[2] and Satish K. Nair[1,3,4*]

## MATERIALS AND METHODS

**Materials.** Peptides were purchased from GenScript with a purity of 90% or greater as verified by HPLC analysis. Chemicals were purchased from Fisher Scientific and Sigma-Aldrich. IPTG and antibiotics were purchased from Gold Biotechnology.

**Enzyme Expression, Purification, and Mutagenesis.** The previously described overexpression plasmid (pCB008) for recombinant production of PCY1 (1) was transformed into a Rosetta *Escherichia coli* expression strain (Novagen). PCY1 was expressed in Terrific Broth (BD Difco) by growing cell cultures to an $OD_{600}$ of 1 at 37 °C. Cell cultures were cooled in an ice bath for 15 minutes prior to the addition of 0.5 mM IPTG. Cultures were further incubated at 18 °C for approximately 18 hours before harvesting. Cells were lysed (Avestin) and purified by $His_6$ affinity purification (GE Healthcare Life Sciences). Fractions containing PCY1, as determined by SDS-PAGE, were collected and further purified using a Superdex 16/60 S200 or S75 size exclusion column (GE Healthcare Life Sciences) equilibrated with a buffer composed of 100 mM KCl, and 20 mM HEPES free acid pH 7.5. PCY1 mutants were generated using the QuikChange methodology with the primers listed in Table S1.

**Protein Crystallization and Data Collection.** Initial crystallization conditions were determined using sparse matrix screens using a Gryphon robot (Art Robbins). Crystals of wild-type PCY1 were further optimized using hanging drop trays (Hampton) by incubating protein (concentration of 14 mg/ml) supplemented with 5 mM boric acid and 1 mM presegetalin A1 [14-32] with an equal volume of precipitant containing 16% PEG 8,000, 0.2 M calcium acetate, and 0.1 M sodium cacodylate (pH 6.5), prior to equilibration against the same solution. Crystallization trays were incubated at 9 °C for at least four days. PCY1 His695➔Ala was crystallized using a protein concentration of 8 mg/mL and precipitant solution of 16 % PEG 8,000, 50 mM calcium acetate, and 0.1 M sodium cacodylate (pH 6.5). Prior to crystallization, the PCY1 His695➔Ala variant was incubated with 1 mM presegetalin A1 [14-32]. For crystals of PCY1 with covalently bound ZPP, the protein was incubated with 1 mM ZPP for 18 hours at 4 °C. Crystals were grown by the hanging drop method at 9 °C with a presegetalin A1 [27-32] concentration of 1 mM and a protein concentration of 10 mg/mL using 20% PEG 8,000, 0.1 M calcium acetate, and 0.1 M

sodium cacodylate pH 6.5 as the precipitant. The PCY1 Gly696 insertion variant was crystallized by pre-incubation with 1 mM ZPP (as before) and 1 mM presegetalin A1 [27-32] at a concentration of 12 mg/mL. A precipitant solution of 18% PEG 8,000, 0.2 M calcium acetate, 0.1 M sodium cacodylate pH 6.5 was used to form crystals at 9 °C. PCY1 crystals were soaked in the crystallization buffer supplemented with 20% PEG 400 immediately prior to vitrification by direct immersion into $LN_2$. The PCY1 Tyr481→Phe mutant crystals were produced by first incubating 1 mM presegetalin A1 [14-32] with protein at a concentration of 7 mg/mL. Crystals formed at 289 K using a 3:1 ratio of the protein sample mixed with a precipitant containing 23% PEG 3,350, 0.1 M magnesium chloride, and 0.1 M HEPES free acid (pH 7.5), prior to equilibration against a solution of the same buffer. The cryo-preservation solution consisted of the crystallization buffer supplemented to a final concentration of 30% PEG 3,350 was utilized. Diffraction data was collected using CCD detectors, at LS-CAT Sector 21 of the Advanced Photon Source (Argonne National Labs, IL).

**Crystal Data Processing and Structure Refinement.** PCY1 data sets were integrated and scaled using the autoPROC software package (2). Using PDB code 1QFS as a search model, phases for wild type PCY1 data set was determined by molecular replacement using PHASER (3), as implemented in the PHENIX software suite (4). Initial model re-building was carried out using a combination of phenix.autobuild and Buccaneer (5). All crystallographic model building was preformed using Coot (6), followed by rounds of refinement using REFMAC5 (7), and phenix.refine. The refined wild type PCY1 structure was subsequently used to determine the phases of all mutants and covalently modified PCY1 structures. Data collection, scaling, and refinement statistics for all data sets can be found in Table S2.

**Fluorescence Polarization.** PCY1 was gel filtered into assay buffer (50 mM Tris base pH 8.5, 100 mM NaCl, and 5 mM DTT) with a Superdex 16/60 S200 column (GE Healthcare). To determine the $K_d$ of the FITC(fluorescein isothiocyanate)-presegetalin A1 [27,32] fluorescent peptide, 100 nM of peptide was incubated with varying concentrations of PCY1 in a 115 μL reaction. After a 1 hour incubation at 4 °C, 100uL of the solution was transferred to a 96-Well Low Fluorescence Assay Plate (Corning) and analyzed a Biotek Synergy H1 Hybrid Reader. Polarization data was fit according to Dunbar *et*. *al*. (8) to determine the $K_d$ of FITC-presegetalin

A1 [27-32]. Competitive binding assays were performed using 100 nM FITC-presegetalin A1 [27-32], 2.3 μM PCY1, and varying concentrations of either presegetalin A1 [27-32], presegetalin A1 [20-32], presegetalin A1 [27-32] amide, or presegetalin A1 [27-32] N27A. Polarization data was fit to a one-site competitive binding curve to derive an $IC_{50}$ value for each peptide. $K_i$ values were derived from the $IC_{50}$ (8).

**Activity assays.** PCY1 activity assays were completed at 30 °C using 175 nM PCY1 and 30 μM of peptide in a 100 μL reaction containing 100 mM NaCl, 5 mM DTT, and 20 mM Tris base (pH 8.5). After the designated incubation time, the reactions were quenched with 1 equivalent (100 μL) of methanol. Single turn over assays were completed using 30 μM PCY1 and quenched after 5 min. Samples were centrifuged to remove protein precipitation. All reactions were completed and analyzed in triplicate.

**LC/MS Analysis.** LC/MS analysis was completed using an Agilent 1200 series HPLC, an Agilent G1956B mass spectrometer, and a Grace Davison Denali 250x4.6 mm 5μ C18 HPLC column. Samples were analyzed using a following method with solvent A (10% ACN, 90% $H_2O$, 0.1% Formic acid, and 0.1% ammonium formate) and solvent B (90% ACN, 10% $H_2O$, 0.1% formic acid, and 0.1% ammonium formate) at 1 mL/min: 1 min at 0% B, linear gradient to 75% B for 14 min, linear gradient to 100% B for 1 min, 3 min at 100% B, linear gradient to 0% B for 1 min, and 5 min at 0% B. 40 μL of each reaction sample was analyzed using this method. Extracted ion chromatograms or single ion mode was utilized to observe the following peptides: segetalin A (*m/z*=610.5, z=1), linear[GVPVWA] (*m/z*=628.5 z=1), presegetalin A1 [14-32] (*m/z*= 993.2, z=2), presegetalin A1 [20-32] (*m/z*=688.5, z=2), cyclic[GKPVWA] (*m/z*= 639.5, z=1), cyclic[Ahx-VPVWA] (*m/z*= 666.6, z=1), and linear[Ahx-VPVWA] (*m/z*= 684.6, z=1).

**Kinetics.** Kinetics assays were completed at 23 °C using 10 nM PCY1 and either 0.5 μM, 1 μM, 5 μM, 15 μM, 30 μM, or 60 μM of the presegetalin A1 [14,32] substrate in a 100 μL reaction containing 100 mM NaCl, 5 mM DTT, and 20 mM Tris base pH 8.5. After 2, 4, or 6 minutes, the reaction was quenched with 100 μL of methanol. After centrifugation, 80 μL of sample was analyzed using the LC/MS method and monitoring the abundance of the linear presegetalin A1 [20-32] byproduct (SIM *m/z*=688.5, z=2). A standard curve was created using synthetic

presegetalin A1 [20-32] concentrations ranging from 10 nM to 400 nM. Reactions were completed triplicate. Kinetics of wild-type and variant PCY1 with native and alternative substrates were conducted in a similar manner.

**Fluorescent Monitoring of Conformational Change.** To monitor conformational changes, PCY1 was first labeled with the cysteine specific fluorophore *N*-(1-pyrene)maleimide. PCY1 was buffer exchanged into 200 mM NaCl, 20 mM HEPES free acid (pH 7.0) and concentrated to 1.25 mL of 50 μM protein. TCEP and *N*-(1-pyrene)maleimide were added to a concentration of 1 mM with the *N*-(1-pyrene)maleimide added drop-wise. The reaction was stirred at 23 $^{o}$C for 2 hours and then quenched with a final concentration of 5 mM DTT. The protein was gel filtered and buffer exchanged into reaction buffer. Modification of PCY1 was confirmed by MALDI analysis to reveal a total of three *N*-(1-pyrene)maleimide added (Figure S16). To monitor the changes in the environment of the pyrenes, 4 μM of labeled PCY1 was incubated with either buffer alone, 1 mM presegetalin A1 [27-32], or 1 mM presegetalin [27-32] amide for 10 minutes. The emission scan was collected using 338 nm excitation wavelength and excitation and emission slit widths of 5 nm.

**Table S1**. **Primers used for generating PCY1 variants.**

| Mutagenesis Primers | Sequence |
| --- | --- |
| Y481F For | AAATGCATGGTTTTGGCGGGTTTGG |
| Y481F Rev | CCAAACCCGCCAAAACCATGCATTT |
| S562A For | TGAAGGTGGTGCTAATGGTGGCCTTCTCGTTGCTGC |
| S562A Rev | GGCCACCATTAGCACCACCTTCAATAGCCACTCTTCTAGC |
| H695A For | CGGATTCAGCGCAAAGCTGCAGCGTACGGACGTGCC |
| H695A Rev | CGCAAAGCTGCAGCGTACGGACGTGCCACAATGACCCAGATTGC |
| G696Ins For | CGCAAAGCTGCACATGGCTACGGACGTGCCAC |
| G696Ins Rev | GTGGCACGTCCGTAGCCATGTGCAGCTTTGCG |
| G696Ins Y697A For | CAAAGCTGCACATGGCGCCGGACGTGCCACAATG |
| G696Ins Y697A Rev | CATTGTGGCACGTCCGGCGCCATGTGCAGCTTTG |
| S493A For | GCGTGCCAACAAGTATTTTCACTTCTATGCTTCCGGCCTTCAAGC |
| S493A Rev | GCTTGAAGGCCGGAAGCATAGAAGTGAAAATACTTGTTGGCACGC |
| N97A For | GCATAAACATGATGCCAACTTTTGCCGCCAGTCGCA |
| N97A Rev | TGCGACTGGCGGCAAAAGTTGGCATCATGTTTATGC |

## Table S2: Data collection and refinement statistics

| | Wild-type | H695A | ZPP complex | 696G –ZPP complex | Y481F |
|---|---|---|---|---|---|
| **Accession code** | 5UW3 | 5UW5 | 5UW6 | 5UZW | 5UW7 |
| **Data collection** | | | | | |
| Space group | P1 | P1 | P1 | P1 | P2$_1$ |
| Cell dimensions | | | | | |
| a, b, c (Å) | 65.5, 85.6, 137.7 | 65.0, 85.6, 137.9 | 65.5, 85.5, 138.1 | 65.0, 85.32, 137.9 | 87.0, 59.6, 134.1 |
| α, β, γ (°) | 87.4, 78.3, 89.3 | 87.5, 78.5, 89.4 | 87.6, 78.2, 89.4 | 87.7, 78.5, 89.6 | 90.0, 93.1, 90.0 |
| Resolution (Å) | 44.9-1.96 (1.967-1.960) | 44.90-2.94 (2.950- 2.940) | 45.2-3.3 (3.311-3.300) | 45.70-2.82 (2.829-2.820) | 86.90-2.37 (2.381-2.373) |
| $R_{sym}$ | 12.9 (52.5) | 12.7 (50.6) | 13.0(27.9) | 12.6 (42.6) | 9.2 (75.1) |
| $I$ / $\sigma I$ | 8.9 (3.3) | 11.0 (2.8) | 10.2 (5.0) | 11.1 (3.3) | 14.0 (2.1) |
| Completeness (%) | 95.8 (96.8) | 98.8 (98.9) | 98.9 (99.5) | 98.3 (98.1) | 99.2 (96.2) |
| Redundancy | 3.8 (3.9) | 3.9 (3.9) | 3.9 (3.9) | 4.3 (4.4) | 6.2 (6.5) |
| **Refinement** | | | | | |
| Resolution (Å) | 44.9-1.96 | 44.90-2.94 | 45.2-3.3 | 45.70-2.82 | 86.90-2.37 |
| No. reflections | 191296 | 58051 | 41469 | 65404 | 52877 |
| $R_{work}$ / $R_{free}$ | 0.224/0.189 | 0.227/0.186 | 0.220/0.172 | 0.270/0.223 | 0.284/0.236 |
| No. atoms | 24731 | 22798 | 22866 | 22661 | 10789 |
| Protein | 22670 | 22734 | 22814 | 22600 | 10754 |
| Water | 2061 | 64 | 52 | 61 | 35 |
| *B*-factors | | | | | |
| Protein | 19.68 | 33.16 | 33.43 | 29.66 | 51.74 |
| Water | 24.87 | 19.19 | 17.08 | 13.30 | 32.58 |
| R.m.s. deviations | | | | | |
| Bond lengths (Å) | 0.0166 | 0.0102 | 0.0107 | 0.0105 | 0.0104 |
| Bond angles (°) | 1.6837 | 1.4070 | 1.4751 | 1.4454 | 1.3242 |

1. Highest resolution shell is shown in parenthesis.
2. R-factor = $\Sigma(|F_{obs}|-k|F_{calc}|)/\Sigma |F_{obs}|$ and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.

**Table S3.** Kinetic parameters for wild-type and variant PCY1 using both cognate and alternative substrates.

| Enzyme/Substrate | $K_M$ (M) x $10^{-6}$ | $k_{cat}$ ($s^{-1}$) | Cyclic/Linear ratio | $k_{cat}/K_M$ ($s^{-1} M^{-1}$) |
|---|---|---|---|---|
| PCY1-WT | 0.77 ± 0.35 | 0.060 ± 0.0048 | 1/0 | 7.8 x $10^4$ |
| PCY1-S493A | 0.923 ± 0.43 | 0.0742 ± 0.006 | - | 8.0 x $10^4$ |
| Ahx-Presegetalin A1 [14-32] | 26.38 ± 5.5 | 0.552 ± 0.043 | 1/48 | 2.1 x $10^4$ |
| V15K-Presegetalin A1 [14-32] | 5.44 ± 0.79 | 0.474 ± 0.017 | 1/0 | 8.7 x $10^4$ |
| PCY1-H695A | Inactive | Inactive | - | Inactive |
| PCY1-H695A + Imidazole | | | 3.7/1 | |
| PCY1-S562A | Inactive | Inactive | - | Inactive |
| PCY1-Y481F | Inactive | Inactive | - | Inactive |
| PCY1-G696 | n.d.* | n.d.* | 2.5/1 | |
| PCY1-N97A | n.d.* | n.d.* | | |

*n.d. indicates that kinetic measurements could not be determined for the variants as they failed to exhibit saturation behavior.
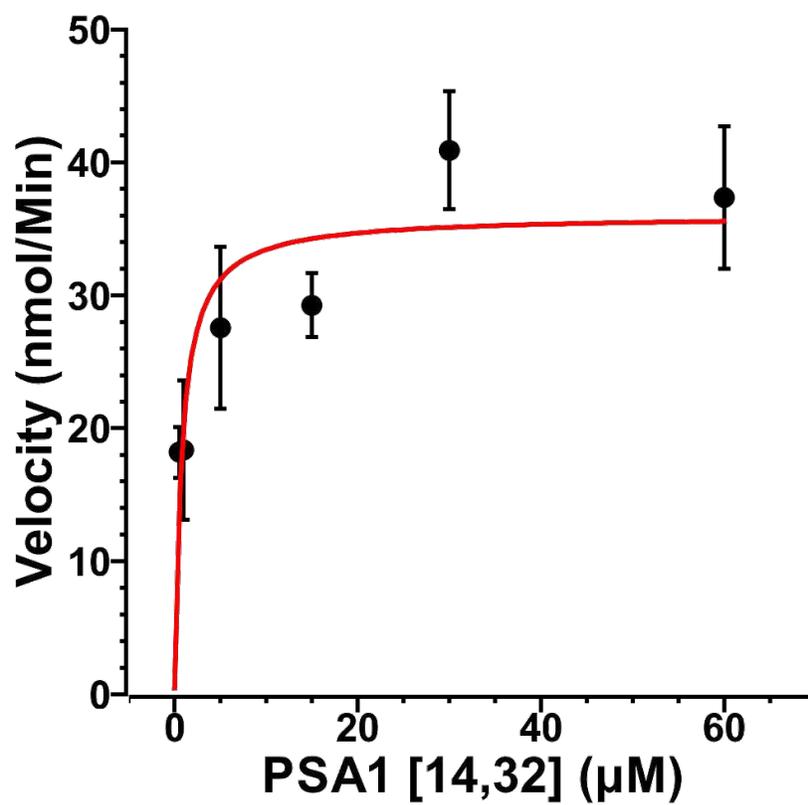
**Figure S1. Michaelis-Menten kinetic curve for PCY1.** Kinetic measurements were carried out using presegetalin A1 (PSA1) [14-32] as the substrate.
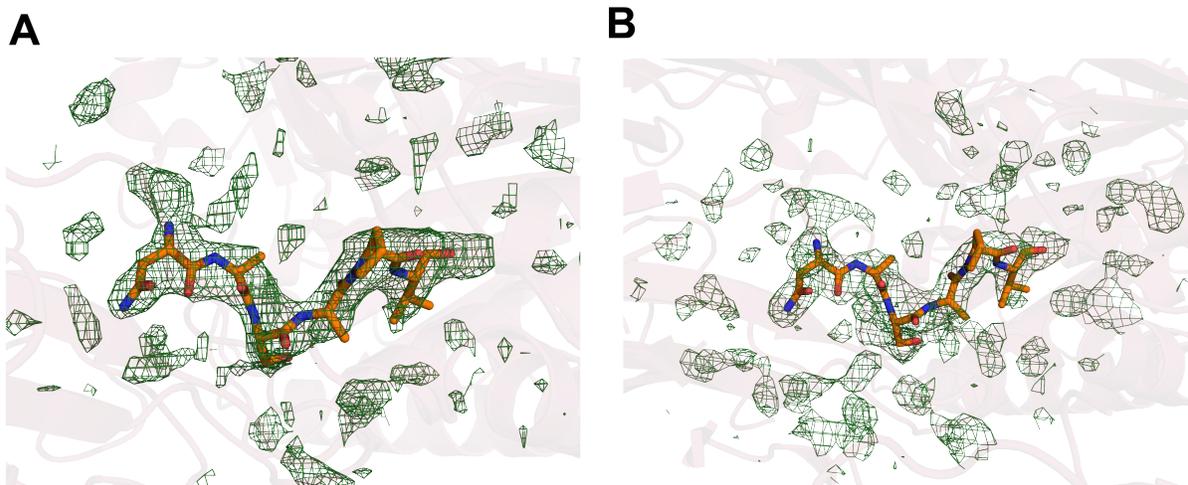
**Figure S2. Difference Fourier maps of PCY1 variant follower binding sites**. Electron density maps calculated with Fourier coefficients $F_o$-$F_c$ and contoured at 2σ over background within a 10 Å radius from the peptide binding sites illustrates that A) His695→Ala and B) Tyr481→Phe PCY1 variants contained only PSA1 [27-32] (orange) visible in the crystal structure.
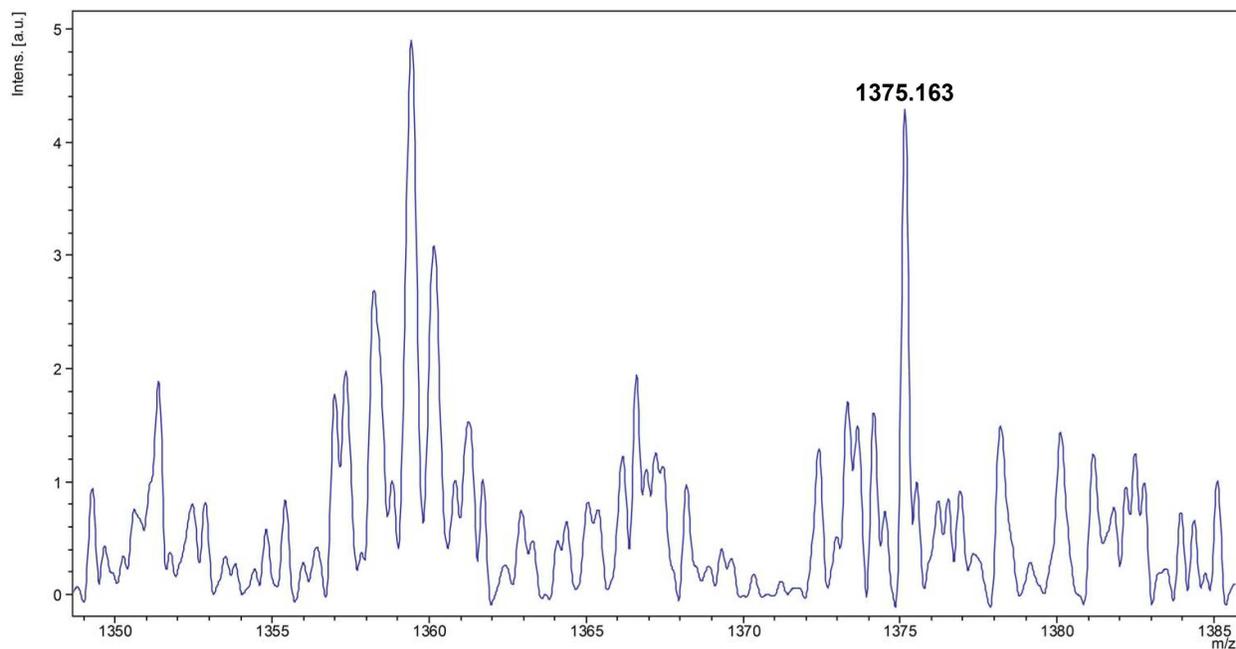
**Figure S3. Integrity of the substrate peptide in PCY1 crystals.** MALDI mass spectrometric analysis of the crystallization drop indicated the presence of the PSA1 [20-32] peptide (*m/z*= 1375.183).    Data was collected on a Bruker UltrafleXtreme MALDI TOFTOF mass spectrometer.

| *D. caryophyllus Precursor 1* | . . . . . G P I P F Y G F Q A K D A E N A S V P V | 20 |
| *Presegetalin D1 [14-31]* | . . . . . G L S F A F P F Q A K D V E N A S A P V | 20 |
| *Presegetalin B1 [14-31]* | . . . . . . G V W A W F Q A K D V E N A S A P V | 18 |
| *Presegetalin G1 [14-31]* | . . . . . . G V K Y A F Q A K D V E N A S A P V | 18 |
| *Presegetalin A1 [14-32]* | . . . . . G V P V W A F Q A K D V E N A S A P V | 19 |
| *Presegetalin L1 [14-31]* | . . . . . G L P G W P F Q A K D A E N A S A P V | 19 |
| *Presegetalin H1 [14-31]* | . . . . . . G Y R F S F Q A K D A E N A S A P V | 18 |
| *Presegetalin K1 [14-31]* | . . . . . . G R V K A F Q A K D A E N A S A P V | 18 |
| *Presegetalin F1 [14-38]* | . . . F G T H G P A P I Q V P N G M D D A C A P M | 22 |
| *Presegetalin J1 [14-36]* | F S A S Y S S K P I Q T Q V S N G M D N A S A P V | 25 |

**Figure S4. Sequences alignments of PCY1 substrates**. Multiple sequence alignment of peptides with confirmed PCY1 activity (1). The follower sequence that is excised is strongly conserved, while the core sequence that is macrocyclized by the enzyme is divergent.
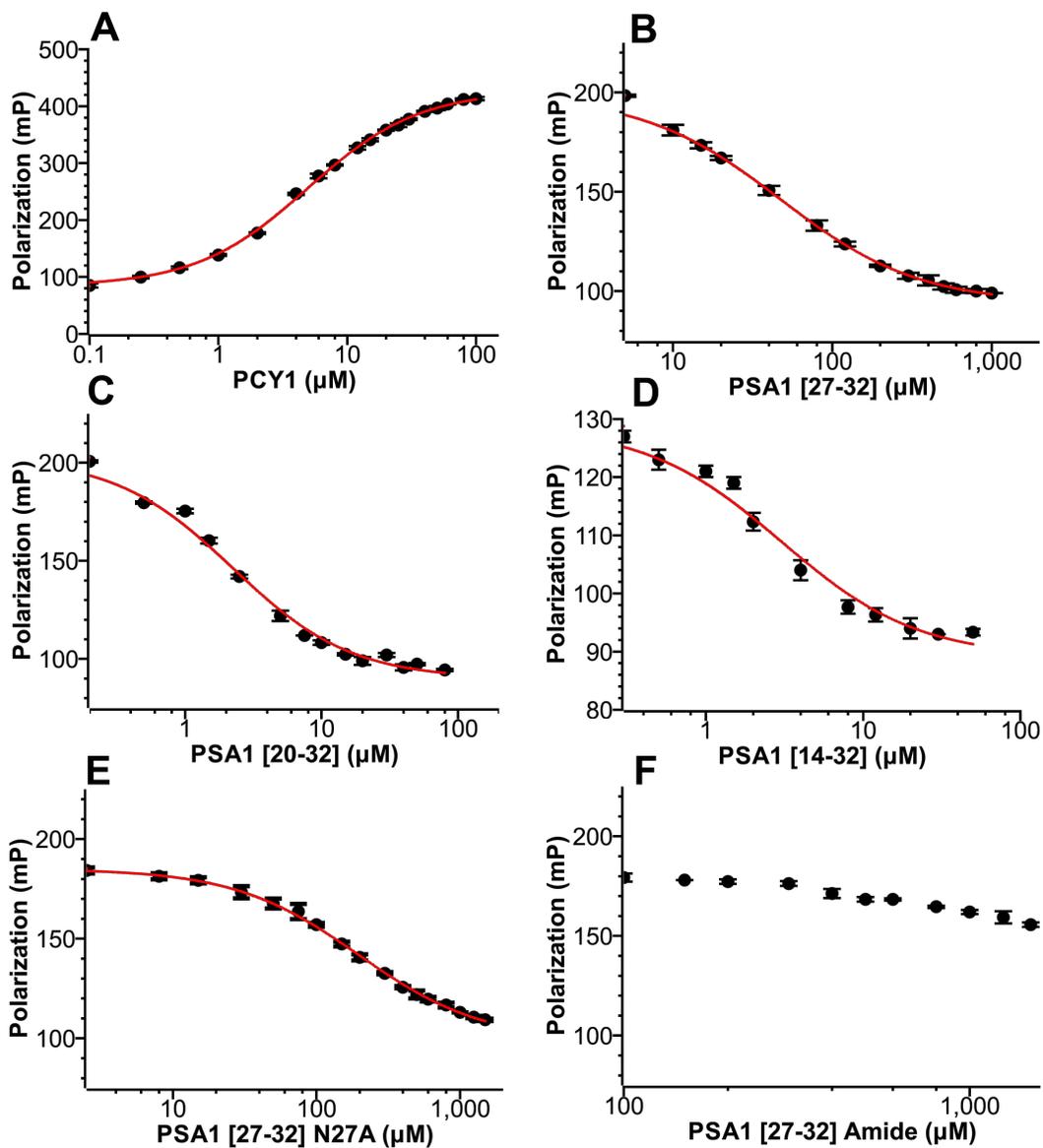
**Figure S5**: **Binding of presegtalin A1 (PSA1) peptides to PCY1**. A) PCY1 was titrated into FITC-presegetalin A1 [27-32]. Competition between B) presegetalin A1 [20-32], C) presegetalin A1 [27-32], D) presegetalin A1 [14-32], E) presegetalin A1 [27-32] N27A, and F) presegetalin A1 [20-32] amide and the FITC- presegetalin A1 [27-32] was used to determine $IC_{50}$ and $K_i$ values for each ligand.
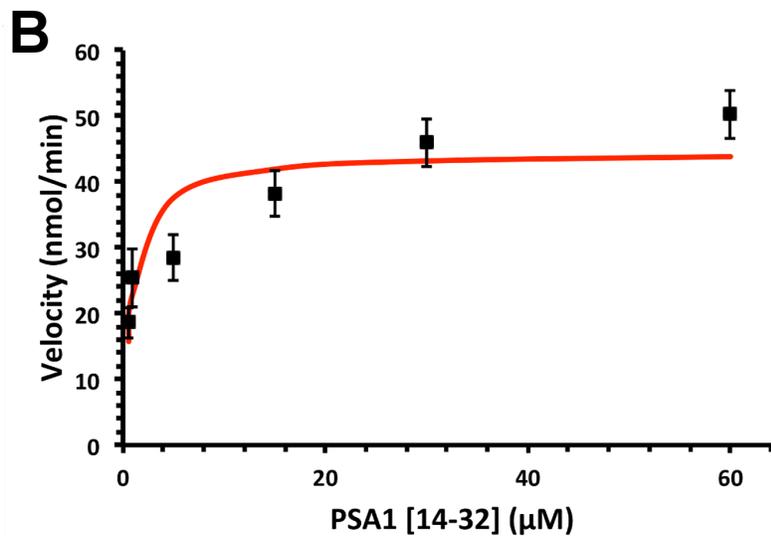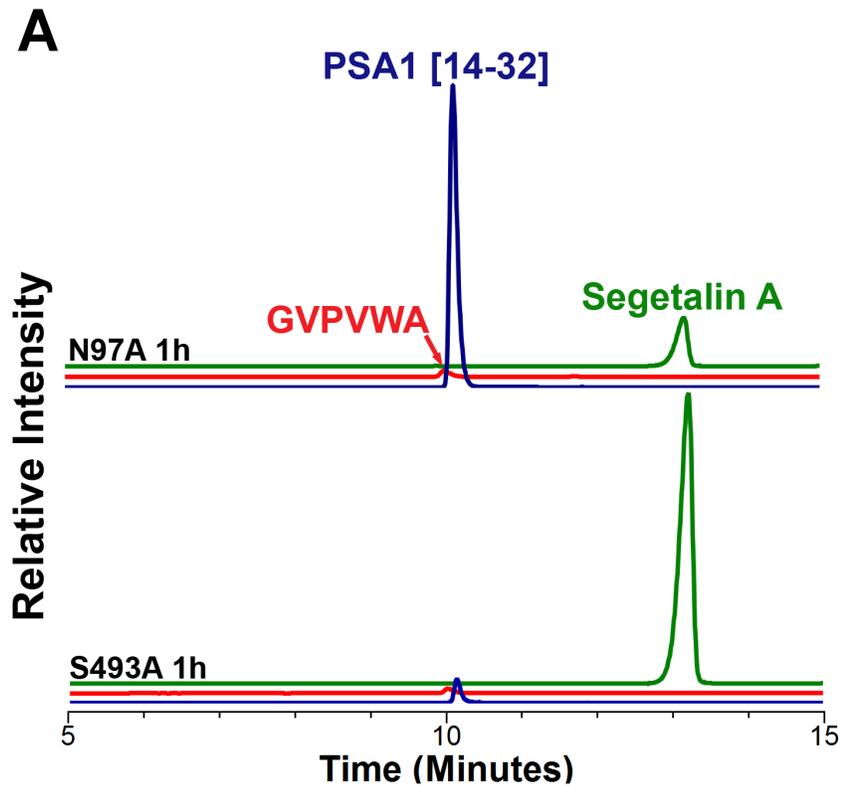
**Figure S6. Activity of PCY1 follower peptide binding variants.** A) Although Ser493 and Asn97 both bind the follower peptide, the Asn97→Ala variant had greatly diminished activity while the Ser493→Ala variant was active in a 1 hour end-point assay. B) Michaelis-Menten kinetic curve for Asn97→Ala PCY1 using PSA1 [14-32] as a substrate.
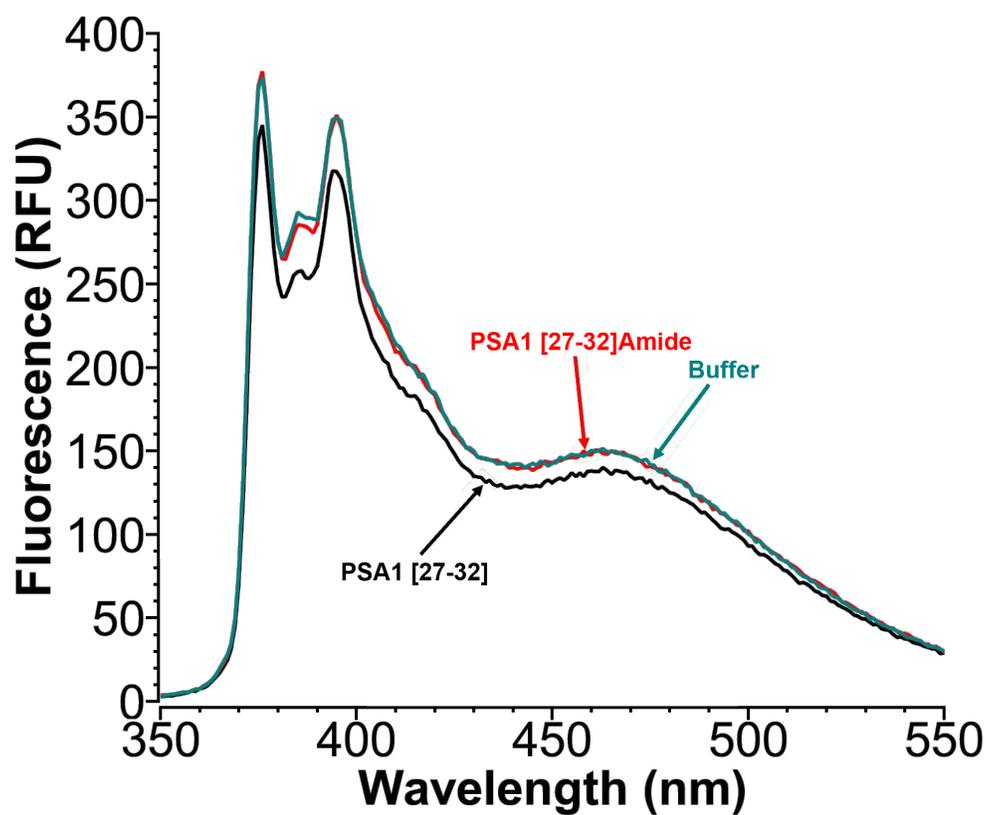
**Figure S7. Binding of the follower peptide alters the conformation of PCY1.** Fluorescence scan of pyrene labeled PCY1 after the addition of presegetalin A1 [27-32] (black), presegetalin A1 [27-32]Amide (red), or buffer (blue). Presegetalin A1 [27-32] induced a quenching of fluorescence, indicative of conformation changes.
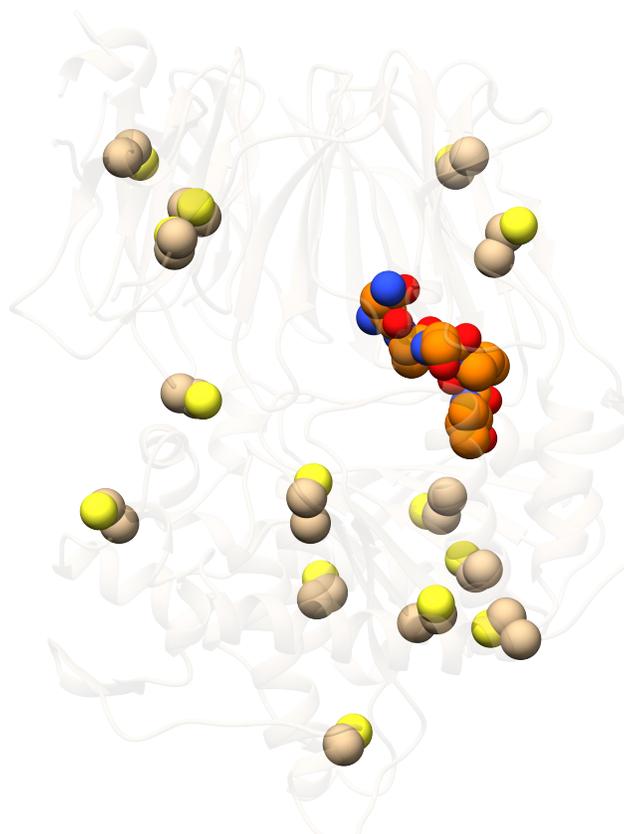
**Figure S8. Possible *N*-(1-pyrene)maleimide labeling locations.** The cysteine residues (cream) of PCY1 are not located near to the binding site of the presegetalin A1 [27-32] peptide (orange).
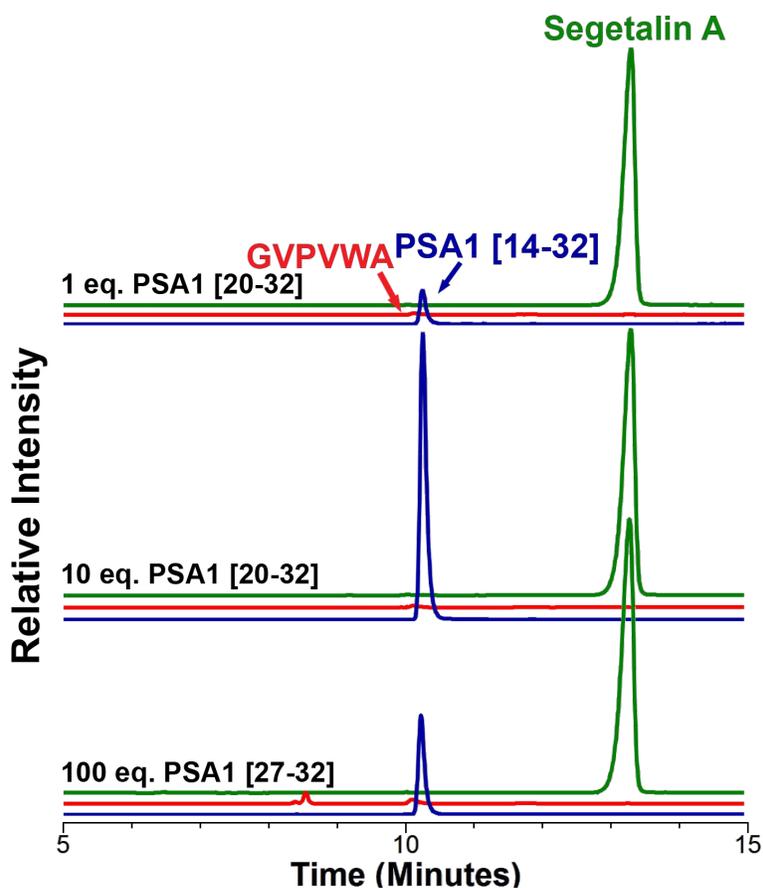
**Figure S9. Product inhibition of PCY1 by the follower peptide.** Competition assays between the presegetalin A1 [14-32] substrate and either 1 equivalent of presegetalin A1 [20-32], 10 eq. of presegetalin A1 [20-32], or 100 eq. of presegetalin A1 [27-32]. In each case substantial amounts of substrate remain, indicating that both presegetalin A1 [20-32] and presegetalin A1 [27-32] can act as inhibitors.
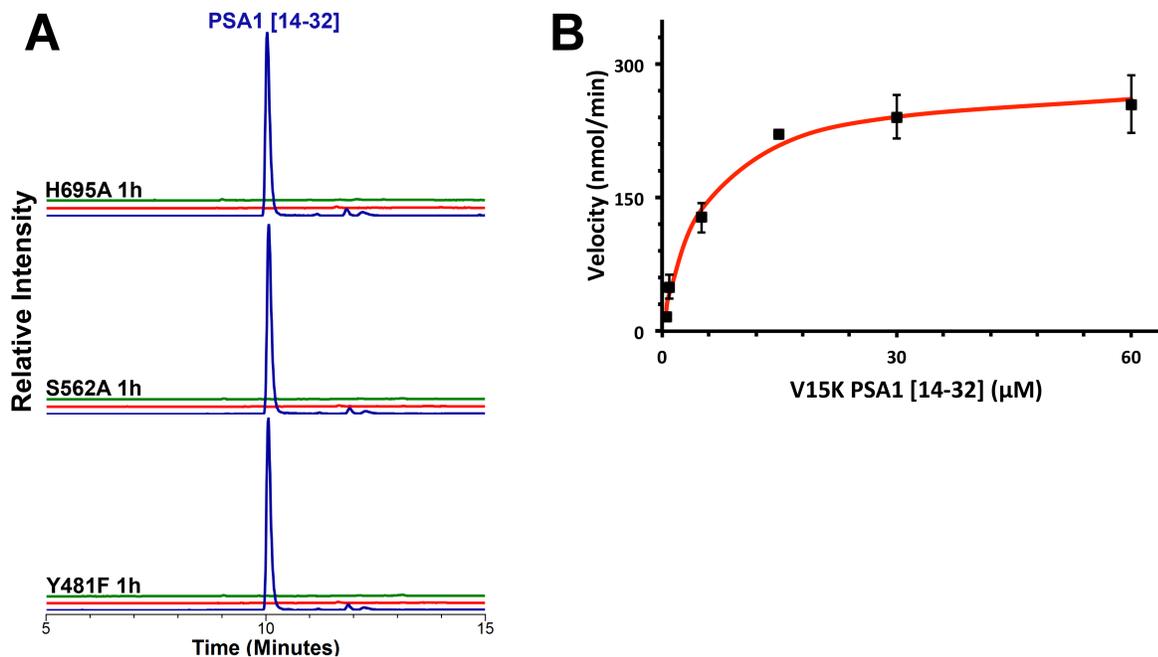
**Figure S10. Biochemical activity of PCY1 variants.** A) Liquid chromatograph/mass-spectrometry (LC-MS) analysis of PCY1 His695→Ala, Ser562→Ala, and Tyr481→Phe variants indicated greatly reduced activity after a 1 hour incubation by monitoring the production of segetalin A1 (green trace), presegetalin A1 [14-32] (blue trace), and linear[GVPVWA] (red trace). B) Michaelis-Menten kinetic curve for PCY1 using PSA1 [14-32] Val15→Lys as a substrate.
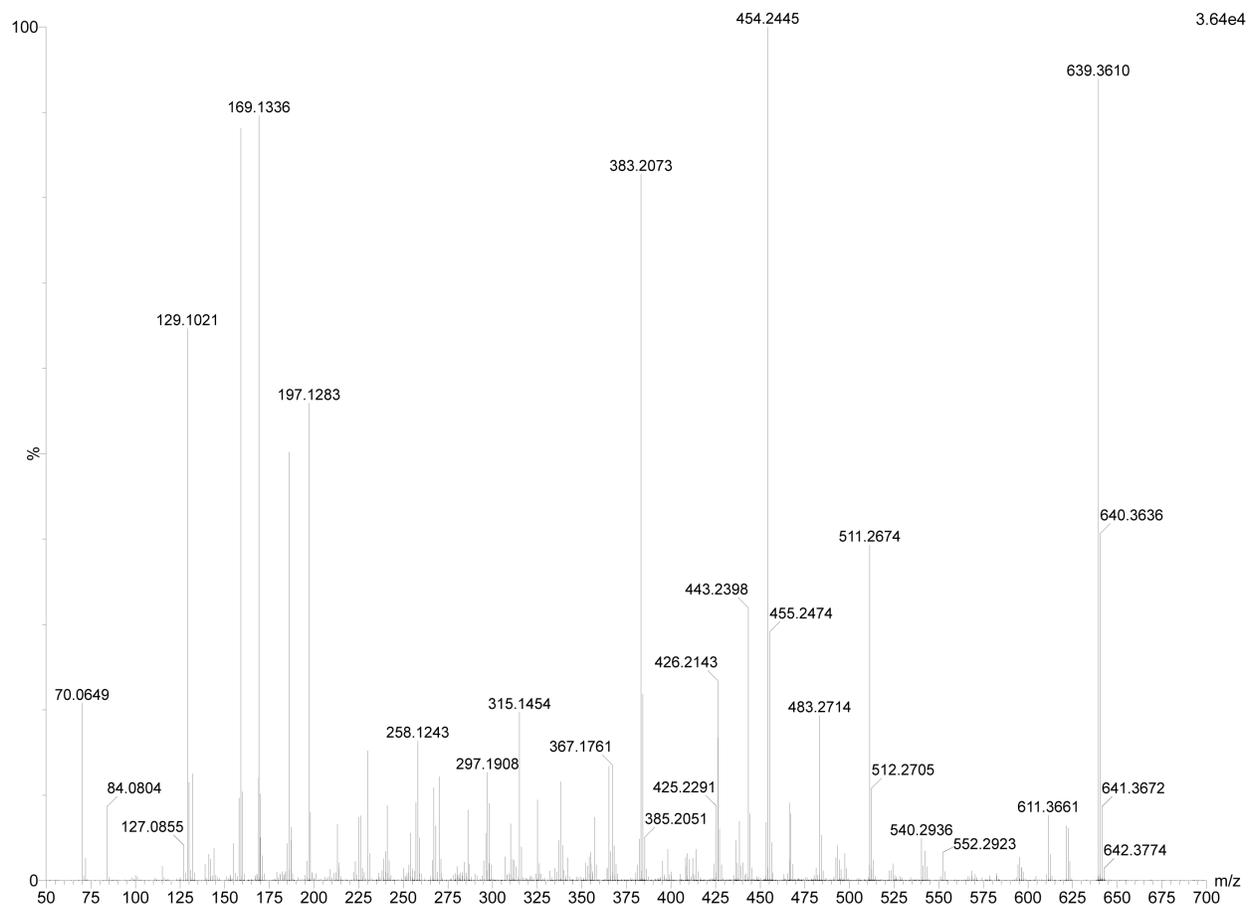
**Figure S11. PCY1 does not cyclize substrates through a Lys side chain amine.** MS/MS analysis of product purified from the presegetalin A1 [14-32] V15K reaction gave no indication of product cyclized using the lysyl amine. Identified fragments are as follows: 639.3610 *m/z* = segetalin A Val15→Lys (cyclic[GKPVWA]), 511.2674 *m/z* = [KPVW] or [PVWAG], 454.2445 *m/z* = [PVWA], 443.2474 *m/z* = [WAGK], 383.2073 *m/z* = [PVW], 197.1283 *m/z* = [PV], 186.1244 *m/z* = [GK], 129.1021 *m/z* = [AG ]or [K].
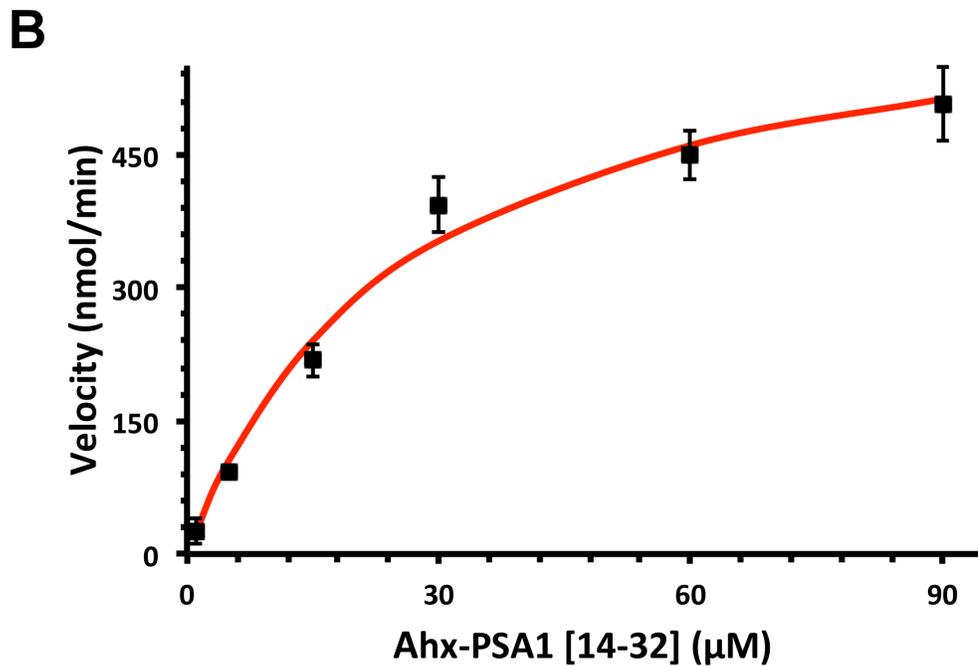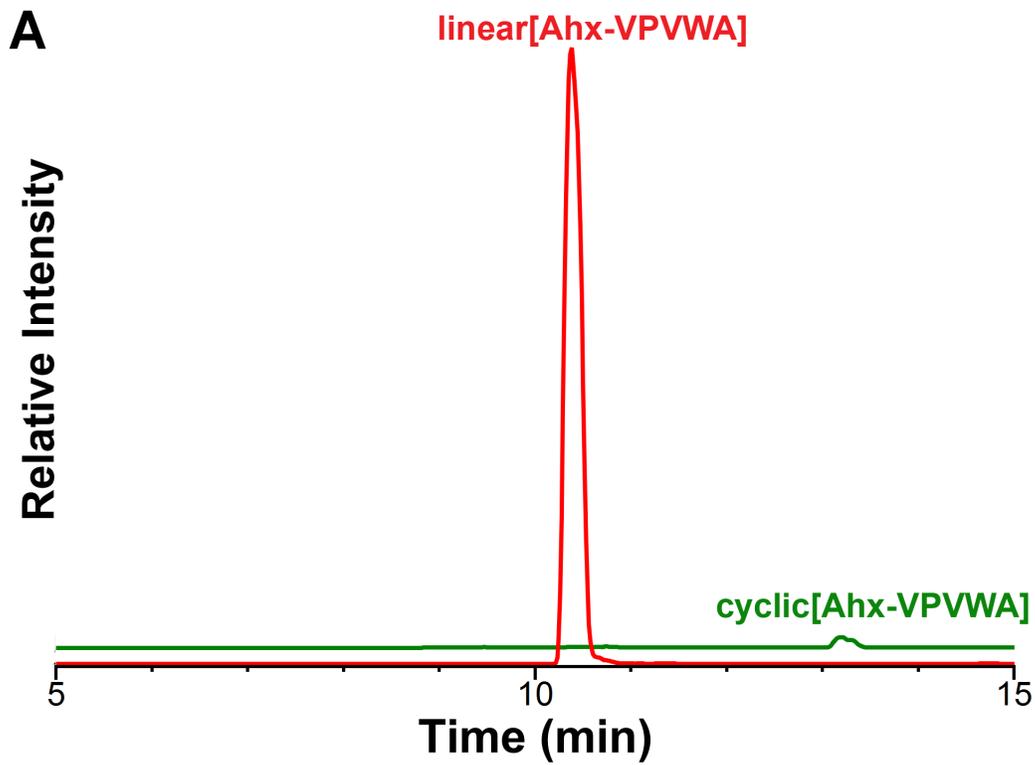
**Figure S12. Macrocyclase activity is dependent on the length of the N-terminal amine.** A) An 18 h incubation of PCY1 with the N-terminal Ahx presegetalin A1 (Ahx-PSA1) [27-32] mostly linear[Ahx-VPVWA] (red trace) and small amounts of the cyclic[Ahx-VPVWA] (green trace). B) Michaelis-Menten kinetic curve for PCY1 using Ahx-PSA1 [27-32] as a substrate.
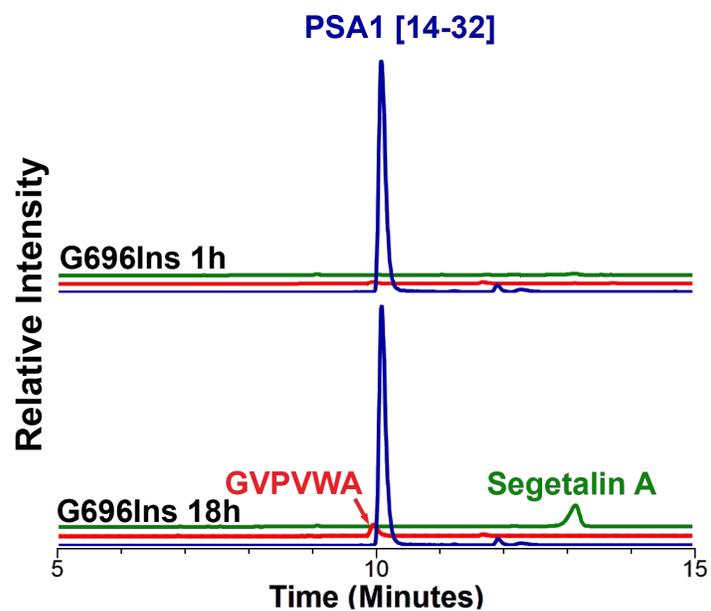
**Figure S13. Activity profile of the PCY1 Gly696 insertion variant.** The Gly696 insertion had minimal turnover after 1 hour. However, an 18 hour reaction produced a cyclic to linear product ratio of closer to one than wild-type PCY1 (Table 2).
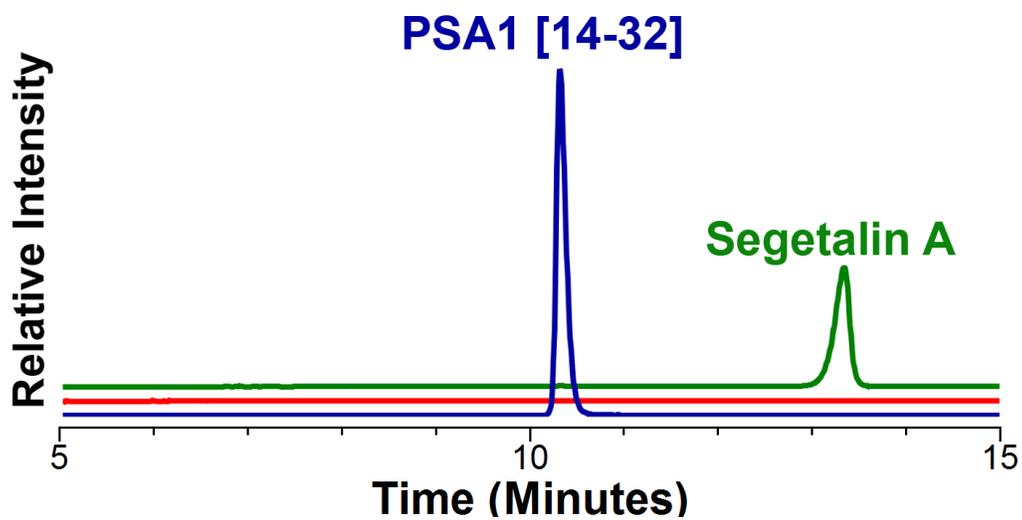
**Figure S14. Activity of the Tyr696→Ala/Gly insertion variants of PCY1.** The Tyr696→Ala/Gly insertion double mutant still has greatly compromised activity. After an 18 h reaction, significant amounts of the PSA1 [14-32] substrate (blue trace) remains and no increase in linear hydrolysis product (red trace) compared to cyclic product (green trace) is observed.
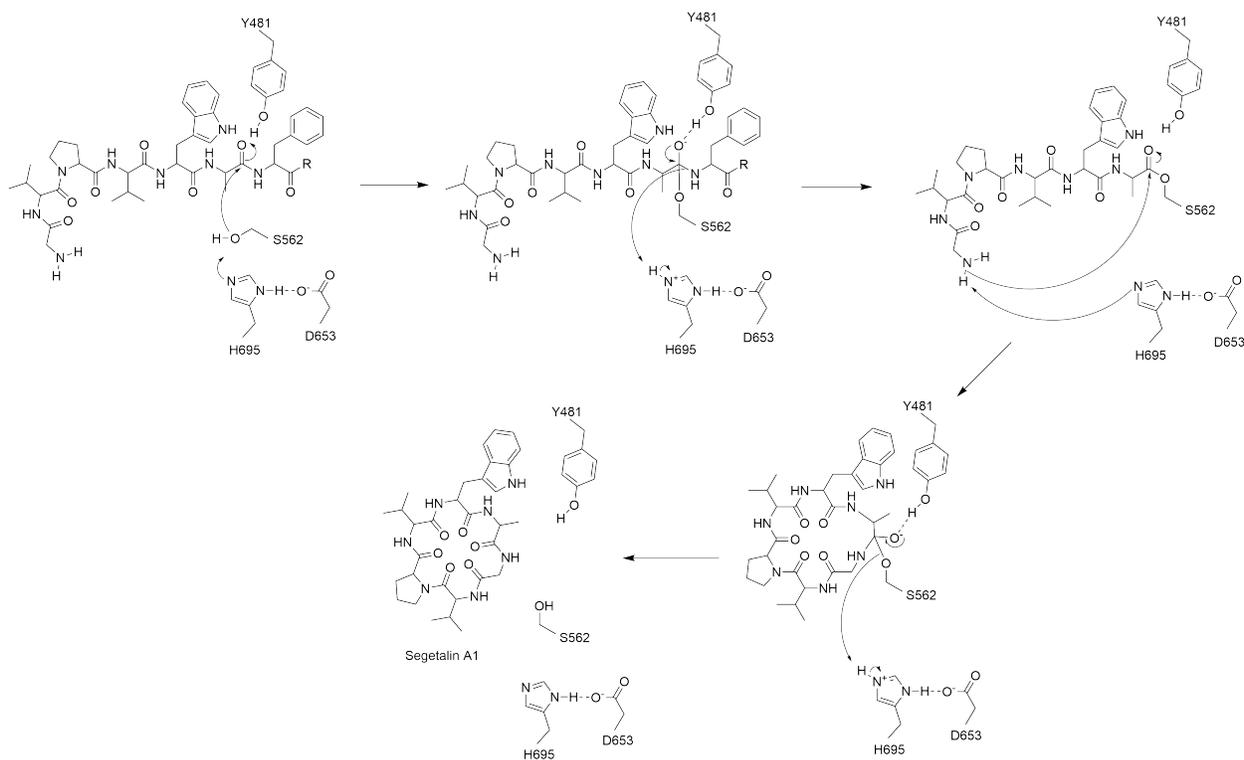
**Figure S15. Proposed mechanism for PCY1.** Unlike canonical S9A proteases, the active site His is mobile in PCY1, and may play a bifunctional role to facilitate the transamidation reaction.
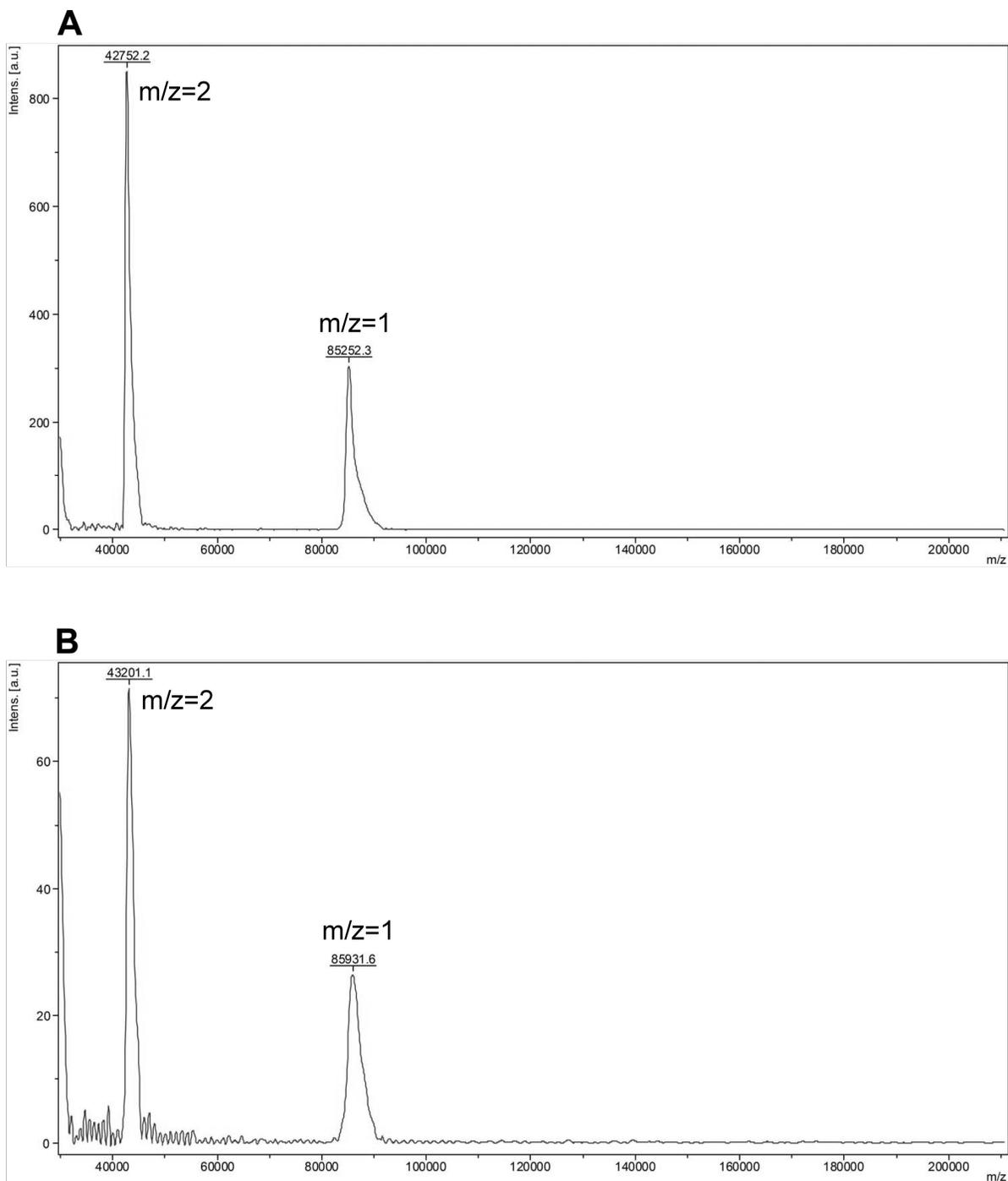
**Figure S16. Extent of modification of PCY1 with fluorescent dye.** MADLI analysis of A) Unlabeled PCY1 and B) PCY1 labeled with *N*-(1-pyrene)maleimide. The 898 *m/z* mass difference, as measured by the z=2 peak, indicates three cysteines had been labeled.

## References

1.      Barber CJS, et al. (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(18):12500–10.

2.      Vonrhein C, et al. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr Sect D Biol Crystallogr* 67(4):293–302.

3.      McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(4):658–674.

4.      Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):213–221.

5.      Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr Sect D Biol Crystallogr* 62(9):1002–1011.

6.      Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.

7.      Vagin AA, et al. (2004) REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr Sect D Biol Crystallogr* 60(12 I):2184–2195.

8.      Dunbar KL, Tietz JI, Cox CL, Burkhart BJ, Mitchell DA (2015) Identification of an Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products. *J Am Chem Soc* 137(24):7672–7.