Characterization of the fast and promiscuous macrocyclase from plant PCY1 enables the use of simple substrates

Hannes Ludewig[1], Clarissa M. Czekster[1], Emilia Oueis[1], Elizabeth S. Munday[2], Mohammed Arshad[3], Silvia A. Synowsky[1], Andrew F. Bent[1%] and James H. Naismith[*1,4,5,6]

Funding information
H.L. is funded by the George & Stella Lee Scholarship and EPSRC. This project was funded by the European Research Council project 339367 NCB-TNT and by the BBSRC (J.H.N.). E.S.M. and M.A. are funded by EPSRC. S.A.S. is funded by BSRC mass spec facility.

[1]Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews KY16 9ST UK
[2]EaStChem, School of Chemistry, University of St Andrews, North Haugh, St Andrews, KY16 9ST UK
[3]Institute of Mechanical, Process and Energy Engineering, Heriot-Watt University, Edinburgh, EH14 4AS, UK
[4] Biotherapy Centre, Sichuan University, Chengdu, China.
[5] Research Complex at Harwell, Didcot, OX11 0FA, UK
[6] Division of Structural Biology, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7BN, UK

* To whom correspondence should be addressed
naismith@strubi.ox.ac.uk

% ANDREW BENT IS DECEASED

Abbreviations: POP, prolyl oligopeptidase; PCY1, prolyl oligopeptidase from plant *Saponaria vaccaria*; PatGmac, macrocyclase from patellamide biosynthesis; RiPPs, ribosomally synthesized and post-translationally modified peptides; LC-MS, liquid

chromatography-mass spectrometry; ITC, isothermal titration calorimetry; Preseg, presegetalin. Amino acids from PCY1 are referred to with one letter amino acid code, while amino acids from the peptide substrates are referred to with three letter amino acid code.

Abstract

Cyclic ribosomally derived peptides possess diverse bioactivities and are currently of major interest in drug development. However, it can be chemically challenging to synthesize these molecules, hindering the diversification and testing of cyclic peptide leads. Enzymes used *in vitro* offer a solution to this, however peptide macrocyclization remains the bottleneck. PCY1, involved in the biosynthesis of plant orbitides, belongs to the class of prolyl oligopeptidases and natively displays substrate promiscuity. PCY1 is a promising candidate for *in vitro* utilization but its substrates require an 11 to 16 residue C-terminal recognition tail. We have characterized PCY1 both kinetically and structurally with multiple substrate complexes revealing for the molecular basis of recognition and catalysis. Using these insights, we have identified a three residue C-terminal extension that replaces the natural recognition tail permitting PCY1 to operate on synthetic substrates. We demonstrate that PCY1 can macrocyclize a variety of substrates with this short tail, including unnatural amino acids and non-amino acids, highlighting PCY1's potential in biocatalysis.

Introduction

The class of natural products represented by ribosomally synthesized and post-translationally modified peptides (RiPPs) are of particular interest as novel therapeutics.[1] Several macrocyclic compounds including cyclic RiPPs are orally active despite the fact that they are in disagreement with Lipinskis' Rule of Five.[2] This behavior has lead to the term 'beyond rule of five' to describe peptide macrocycles.[3] The activity and utility of macrocyclic peptides arises from their increased stability (both chemically and to protease degradation), rigidity and hydrophobicity when compared to their linear peptide counterparts.[4] Yet, analogously to linear peptides, macrocyclic peptides can encode complex structural and chemical information and thus may be particularly suitable to tackle difficult targets such as protein-protein interactions, which are traditionally challenging to target using small molecules.[1] The *de novo* chemical synthesis of macrocyclic peptides is well known but has some disadvantages such as the requirement of multiple steps and the optimization of conditions necessary for the ring closing reaction for each peptide variant.[5,6] Moreover, when the active macrocycle is highly modified, as many cyanobactins are, their synthesis is not considered practical.[5] Biotechnology provides an alternative approach utilizing enzymes entirely *in vivo* or *in vitro*. The enzymes involved in RiPPs biosynthesis are particularly attractive as these natural products are genetically encoded, and the enzymes involved are able to carry out highly specific modifications while operating with a very broad substrate range. This allows enzyme activities from different pathways and organisms to be combined.[7]

Cyanobactins possess a wide range of desirable and valuable bioactivities including anti-cancer (cytostatic and cytotoxic), antifouling, immunomodulating and antineoplastic.[8] An understanding of the cyanobactin biosynthetic machinery has underpinned the production of a wide range of novel cyanobactins *in vivo*[9,10] and *in vitro.*[11–14] Whilst *in vivo* approaches - sometimes termed 'cell factory' - hold the advantage of simplicity, *in vitro* technologies have the unique capacity of harnessing the power and diversity of chemical synthesis. This means that not only multiple different non-natural amino acids can easily be accommodated, but hybrid molecules (peptide and non-peptide) that are impossible to be produced *in vivo* can be generated.[13] To date, the majority of reports of *in vitro* macrocyclization have used the macrocyclase domain from the patellamide pathway (PatGmac). Although extremely promiscuous,

the enzyme has extremely slow catalytic rates *in vitro*, a major drawback that limits its application at large scale.[15] A fast and promiscuous macrocyclase enzyme is thus highly desirable. Of the macrocyclases known to date, PatGmac (and other members of its family), the prolyl oligopeptidases (POPs) from Basidiomycete mushrooms (AbPOPB and GmPOPB),[16] and PCY1 from the plant *Saponaria vaccaria* (syn. *Saponaria hispanica*)[17] are known to catalyze the formation of small macrocycles (containing between 5 and 9 amino acids). Although butelase 1 from *Clitoria ternatea* is extremely fast[18] and highly versatile,[19] it does not produce small ring macrocycles efficiently,[20] and a similar situation occurs with the MdnB from *Microcystis aeruginosa* (ATP-grasp superfamily).[21] Further limiting the widespread use of butelase 1 is the difficulty of its overexpression in *E. coli*[18,19] and limited solubility (partially overcome by engineering but with decreased activity).[22]

The POP class of macrocyclases has been underexplored in terms of biocatalysis. GmPOPB has been kinetically and structurally characterized, and although it possesses fast reaction rates, it displays narrow substrate scope.[16,23,24] PCY1, the macrocyclase in the biosynthesis of a range of plant cyclic peptides known as segetalins, has been shown to be a naturally promiscuous enzyme capable to generate 5 to 9 residue segetalins, although the reported turnover rates are slow.[17] A two-step biosynthetic pathway for segetalins has been proposed, starting from a precursor peptide in which the core peptide is flanked by both a leader and tail peptide.[17] Firstly a 10 amino acid *N*-terminal sequence (termed the leader in RiPPs biosynthesis)[25] is removed by an uncharacterized peptidase to give the presegetalin, then PCY1 catalyzes peptide bond formation to yield the corresponding macrocyclic peptide and cleaved *C*-terminal peptide (Figure 1a). Two PCY1 complex structures have been published, the substrate's core sequence organization within PCY1's active site were not resolved, leaving unanswered questions about peptide recognition and macrocyclization.[26]

Here we report the first apo structure of PCY1 as well as complex structures with natural substrates resolving substrate core within the active site for the first time. Combining structure, kinetics and thermodynamics, we have discerned the rules governing substrate recognition. Based on these rules we demonstrate the enzyme's will macrocyclize efficiently with a three residue C-terminal recognition tail. Using this tail we show PCY1 operate on peptides including a peptide alkyl chain hybrid, peptides containing *N*-methylated peptide bond as well as thiazoline containing peptides. PCY1

is an attractive tool for *in vitro* biosynthesis of diverse cyclic peptides.

Results and Discussion

Mechanism of PCY1

PCY1 operates via an acyl enzyme intermediate in a classical serine protease mechanism,[27] with the *N*-terminus of the peptide rather than water acting as the nucleophile. [26] The PatGmac macrocyclase also operates by this mechanism,[28] although possessing an entirely different fold. The putative catalytic triad (D653, H695 and S562) is located in the peptide-binding site of PCY1. The catalytic serine and histidine are further apart than is observed for the triad in other POPs, although a similar 'non-optimal' arrangement is present in GmPOPB.[24] Activity tests using PresegB1 as a substrate for macrocyclization confirmed that PCY1 mutants S562A and H695A were inactive (Figure S1).

Kinetic characterization of PCY1

Barber *et al*. described PCY1 to have a turnover number of ~ 1 h^{-1} for PresegA1.[17] We evaluated the enzyme produced heterologously in *E. coli* by LC/MS-based time course experiments macrocyclizing 200 µM of PresegA1, PresegB1 or PresegF1 with 3.6 µM PCY1. These substrates were chosen as they produce 6, 5 and 9 residue macrocycles, respectively, show variability in the length of the *C*-terminal tail sequence, and have a different *C*-terminal core residue (proline in PresegF1, alanine in the other two) (Figure 1b, Figure S2). PresegB1 was completely macrocyclized in ~ 30 min, PresegA1 in ~ 50 min and PresegF1 in > 3 h (Figure 2a). We chose to investigate the slowest and fastest substrate by steady-state kinetics using an LC/MS assay and $UV_{280\ nm}$ quantification. This resulted in a k_{cat}/K_m value of 830,000 M^{-1} s^{-1} for PresegB1 (Figure 2c and Table 1), which is not only considered above 'average' for a typical enzyme[29] but is also larger than the efficiency using the second-order rate constant of butelase 1 from *Clitoria ternatea* (542,000 M^{-1} s^{-1}), the fastest macrocyclase described to date.[18] The $UV_{280\ nm}$ detection at low substrate concentrations and thus accurate determination of K_m was problematic, especially for PresegF1, which lacks a tryptophan in its sequence. We explored steady-state kinetics with single injection kinetic ITC for PresegB1 (Figure 2d and Table 1) and this allowed more accurate determination of K_m (0.25 ± 0.01 µM). Under the ITC conditions, the measured k_{cat} (0.14 ± 0.01 s^{-1}) was a factor of three lower than the LC/MS derived value. However, to obtain kinetic parameters for PresegF1, a different approach was required since single injection kinetic ITC only showed a small heat change. Therefore, we employed an LC/MS assay followed by

quantification using mass ion counts, relying on a calibration curve employing the cyclic peptide product as a standard (Figure 2b). This approach gave a K_m of 2.40 ± 0.51 μM and k_{cat} of 0.12 ± 0.01 s^{-1} for PresegF1. These kinetic parameters line up with steady-state kinetics performed using PresegA1[26], and show that the difference in efficiency that we observed in the time course experiments is due to large differences in K_m rather than in the turnover rate.

Crystallization of PCY1

PCY1 has been predicted (through homology modelling using a porcine muscle prolyl oligopeptidase (POP) structure)[17] and shown via X-ray crystallography studies to belong to the S9 protease family[26], displaying the overall architecture of the POP family comprising an α/β- hydrolase domain and a seven bladed β-propeller (Figure S3). The structures of PCY1 bound to PresegA1 and to the covalent small molecule inhibitor z-proline-prolinal (ZPP) have been published but disorder meant only the recognition of the *C*-terminal NASAPV motif (conserved in PCY1 substrates[17] Figure S2) was visualised.[26] Our 2.55 Å PCY1 apo structure with a cacodylate molecule at the active site is similar (rmsd: 0.88 Å over 4,853 atoms; Table S1) to the previously reported PCY1 complex adjacent to the active site comprising S562, H696 and D563 (Figure S4). Cacodylate is coordinated through hydrogen bonds by the side chain hydroxyl group of Y481 and the amide of N563; both residues have been proposed to form the oxyanion hole in the POP-family of enzymes (Figure S5).[30,26] The PCY1 structures all have a closed conformation in contrast to the open apo structure of GmPOPB (Figure S3), a structural homologue in the POP family with low sequence identity.[24] Although co-complexes were described for GmPOPB the core peptide was not visualized.[24] To obtain more structural insight into substrate recognition for the POP macrocyclase family we explored other substrates for co-crystallisation with PCY1.[26]

Substrate recognition

We determined complex crystal structures of PCY1:S562A bound to PresegA1 (2.0 Å), PresegB1 (2.17 Å) and PresegF1 (1.86 Å) (Table S2). We did not observe continuous electron density spanning the entire length of the peptide substrate in any of the structures solved (Figure 3a). In all three complexes however, the *C*-terminal peptide (NASA(S)PV) is clearly visible and adopts the same structure (Figure 3b,c and Figure

S6), seen in in previously published PCY1 complex structures with PresegA1 (Figure S6).[26] In the PresegF1 complex, the highest resolution structure, the carboxylic acid group of the substrate's *C*-terminus V25 makes hydrogen bonds to the backbone amide and side chain of S495 as well as the side chain of S493. The carboxyl group is positioned at the *N*-terminal end of the helix thus interacting with the helical dipole (Figure 3b,c). In the substrate, the side chains of Val25, Pro24, Ala23, and Ala21 make only a few van der Waal contacts with the protein, while the main chain of Ala21 makes hydrogen bonds with the protein. Both the main and side chain of Ser22 and Asn20 make either direct or water bridged hydrogen bonds.

In the PresegA1 complex, no other residues could be modeled, as previously observed by Chekan *et al*.[26] However, in PresegF1 and PresegB1 complex structures, it was possible to model the core peptide in at least one of the subunits within the asymmetric unit, although the intervening residues that connect core to the tail were not ordered. In PresegF1, we visualized Phe1 to Gln11 (Figure 3a) but not Thr12 to Met19. In the other three subunits, electron density for the N-terminal region of core peptide has been less well defined, suggesting there could be multiple binding conformations. In the best-ordered subunit, one face of the proline (Pro9) ring in the substrate stacks against W603 whilst the edge of the ring interacts with F484. The Leu8-Pro9 peptide bond (P2-P1) adopts a *trans* configuration. The remainder of the core peptide makes a small number of hydrogen bonds and van der Waal interactions (Figure S7). The scissile bond P1-P1′ (Pro9-Ile10, Figure S8) is positioned such that the modeling of the hydroxyl at S562A results in the distance and orientation expected for the formation of the acyl enzyme intermediate (Figure 3d), with the Y481 and N563 acting to stabilize the negative charge (Figure S9). The chemically sensible positioning of the scissile bond at the active site validates the relevance of the structure. The P1′ side chain (Ile10) sits in a hydrophobic cleft making contacts with the side chains of Y481 and I466 (Figure S8). The amino terminus of the substrate is over 20 Å from the scissile bond, requiring extensive conformational flexibility from the substrate for macrocyclization to occur.

A comparison between complex structures of PCY1-S562A bound to PresegF1 and GmPOPB-S577A bound to a 35 amino acid substrate (GmAMA1) reveal that although the overall protein architecture is remarkably similar, significant differences exist

between the location of the peptide ligand, and positioning of specific regions of each peptide (Figure S10). In GmPOPB, the entire peptide tail or recognition sequence (which is 17 amino acids long) can be seen in the structure, and spans through the □-propeller domain. In contrast, in PCY1 the recognition sequence is not inserted into the propeller domain, and only the last C-terminal residues can be observed, demonstrating greater flexibility in the peptide region that follows the core peptide, different to GmPOPB. The core sequence of the macrocyclization substrate was not experimentally observed in GmPOPB complexes (pdb codes 5N4B and 5N4D).[24]

In the PresegB1 PCY1 complex, the P2-P1-P1′ residues (Trp4, Ala5 and Phe6, Figure S8 and S11) adopt an identical main chain conformation, make similar interactions as PresegF1, and are thus positioned for attack by S562. However, the remaining three residues of the core peptide have a different and more ring-like conformation than PresegF1 and as a result the *N*-terminus is only 11 Å away from the scissile bond (Figure S12). Thus, PresegB1 needs less re-organization than PresegF1 and this may underlie the difference in their macrocyclization turnover rates. Comparing the PCY1 structures in the PresegB1 and PresegF1 complexes reveals the structures are very similar (root mean square deviation 0.8 Å over 707 Cα atoms). The most notable change occurs in the main chain at Y696 which, in the PresegF1 complex, has flipped so that the side chain points into the structure core, and as consequence the loop at E154 has moved to accommodate it (Figure S13). In PresegB1, the residue adopts a different rotamer but the position of the main chain is unchanged. The change at Y696 appears necessary to avoid a clash between this residue and the core of the substrate. In both complexes, R655 makes hydrogen bonds to the substrate in the P2-P1-P1′ region although the precise interactions are different. Informed by the crystal structures, we constructed mutants S493A, S495A, Y696G, R655A and W603A using the reaction with PresgB1 as a test substrate. The relative activities of S495A, S493A, R696G, and R655A were decreased by half, while W603A and H695Q showed around 10% relative activity, and H695A was inactive (Figure S1). No variant resulted in an increase in proteolytic (as opposed to macrocyclase) activity.

To further investigate substrate binding, we performed ITC experiments using PresegA1, PresegB1, PresegD1 and PresegF1 with PCY1:S562A. All four peptides

showed tight binding with K_d values around 200 nM (Figure 4a, Table 2 and Figure S14). The major contribution to binding of all tested peptides is enthalpic rather than entropic (Figure 4b), pointing towards interactions between protein and substrate rather than exclusion of water. The *C*-terminal tail of PresegA1 was previously reported to be essential for activity.[17] The only conserved amino acids in all natural substrates are the *C*-terminal six residues (NASA(S)PV) (Figure S2). The peptide NASAPV binds with a K_d of only 25 μM suggesting that this region although conserved does not dominate binding (Table 2, Figure S14). We measured the binding of PresegA1-NH_2 (an amide rather than carboxylic acid at *C*-terminus) to PCY1:S562A and observed a ~70-fold decrease of affinity (K_d: 12.1 μM) (Table 2, Figure 4).

Structural data point to a key role for W603 in positioning the substrate with the correct geometry for attack by S562. Unlike the PatGmac class of macrocyclase,[27] PCY1 does not require a proline residue or *cis*-like geometry for activity. Instead, we propose the site of macroyclization is controlled by the nature of the residues at P1 and P1′ of the substrate. The structure shows that PCY1 residues W603, F484, Y607, and V558 would clash with the amino acids that have two non-hydrogen substituents at Cβ (Val, Ile, Thr) or larger. Although serine cannot be ruled out on purely steric grounds, the hydrophobic nature of the environment favors proline or alanine. The binding site at P1′ is hydrophobic but requires larger side chains such Ile, Phe, or Leu to make interactions with the protein (Figure S8). This sequence pattern is found at the point of scission for all known substrates for PCY1 (Figure S2c). The K_d for DNASAPV (9.6 μM) and NASAPV (25.2 μM) with PCY1:S562A were larger than observed for full length substrate (Table 2 and Figure S14). A peptide substrate without the *C*-terminus (PresegF1-truncated, FSASYSSKPIQT) was evaluated although its K_d could not be accurately measured but was estimated to ~ 0.6 mM (poor peptide solubility and weak binding prevent more accurate measurement) (Table 2 and Figure S14). The peptide that links P2-P1-P1′ of the core and the *C*-terminus is highly variable in sequence and in length (Figures 1b and S2) consistent with this region acting as a flexible linker rather than a recognition element. Whilst the intact peptide binds very tightly, these two regions on their own bind more weakly, suggesting a model where affinity is generated by linking binding sites (conceptually analogous to the chelate effect).[31]

Exploring substrate promiscuity

It is known that PCY1 is tolerant of significant changes in sequence within the core peptide and competent to make macrocycles as short as five and as large as nine residues.[17] We tested the ability of PCY1 to make larger macrocycles by extending the core sequence of PresegF1 by one amino acid and PCY1 was able to macrocyclize this substrate (Figure S15 and S16). This extends the full range of relevant peptide macrocycles that PCY1 makes to include the upper limit (1000 Da) for 'beyond rule of five' macrocycles.[32] However, the requirement for a long disposable *C*-terminus is a major drawback of the use of PCY1. The molecular recognition of substrate peptide by PCY1 prompted us to devise substrates with a short *C*-terminal extension replacing the native recognition sequence. We analyzed reactions of a peptide containing FSASYSSKP (F1 core) with different truncated recognition sequences derived from PresegB1 or PresegF1 added *C*-terminal to the core. In both PresegB1 and PresegF1 a three residue tail gave optimal results (Figure S17) but a diverse set of extensions (sequence and length) supported catalysis. PresegB1 derived recognition sequences in general yielded more SegF overnight in comparison to the PresegF1 derived recognition sequences. Steady-state kinetic analysis of the variants revealed FQA is the superior truncated recognition sequence over IQT with a ~ 17x larger second-order rate constant, but both show synthetically useful k_{cat} and K_m values (Table 1 and Figure S17). Although GmPOPB can also process short tailed substrates,[24] it is slower (k_{cat} at 20 °C using a shorter substrate is 0.010 s^{-1}), intolerant of changes in core length and restricted in sequence acceptance. We tested the core sequence VTACITWP (related to the natural substrate of PatGmac and unrelated to PCY1 natural substrates) with both short recognition sequences: FQA, and IQT. The peptides (250 µM) were processed with 0.5 µM PCY1 and compared to a 25 µM PatGmac reaction of VTACITWPAYDGE (AYDGE contains the PatGmac recognition sequence). After 90 minutes, PCY1 had turned over ~29 x and ~11 x more substrate for the FQA and IQT variants as analyzed by LC/MS (Figure S18), respectively than the PatGmac reaction (PatGmac was at 50-fold higher concentration). PCY1 therefore outperforms PatGmac on an eight-residue core sequence similar to PatGmac's own substrate. Although known to process a wide range of amino acids within the core peptide, the tolerance of PCY1 for non-natural amino acids is unknown. Capitalizing on our discovery of truncated substrates, we explored the scope of PCY1 with such shortened substrates containing unnatural and modified amino acids. PresegF1 core peptide modified with an amino

octanoic acid (FSA-8Aoc-FP) and VGAG-8Aoc-FPIQT (derived from a known PatGmac substrate) (Figure S19-S21), MALDI spectrum shows an approximate 50% conversion rate overnight for the first peptide FSA-8AOc-SKPIQT with only traces of the linear-cleavage peptide alongside the cyclic peptide (Figure S19), while the other precursor peptide VGAG-8Aoc-FPIQT is almost entirely converted overnight to a mixture of the corresponding cyclic peptide and linear-cleavage peptide (Figure S20). The same reaction with the highly promiscuous PatGmac would have required 60 times the amount of enzyme and an average of 20 days for completion. Nonetheless, PatGmac usually gives rise to either exclusively linear-cleavage peptide (especially for shorter peptides) or none at all, as was the case for the VGAG-8Aoc-FP peptide.[13] Additionally, we showed that PCY1 can macrocyclize FS^{m}ASYSSKPIQT and FS^{m}ASYSSKPFQA (where ^{m}A is *N*-methylated alanine) and produce cyclic FS^{m}ASYSSKP (Figure S22 and S23). *N*-methylated peptidic drugs are known to cross cell membranes more efficiently than compounds lacking this chemical feature. Taken together, our data show that the processing activity of PCY1 is not simply limited to the canonical amino acids within the core. PCY1 failed to process a proline-rich PresegF1 variant yielding only a hydrolysis product (Figure S24). Such proline-rich peptides are also extremely poor substrates for PatGmac, an observation attributed to the conformational rigidity imposed by multiple proline residues.[28] Using LynD$_{fus.}$, an engineered heterocyclase enzyme,[33] we converted VTACITWPIQT into VTAThnITWPIQT (Thn: thiazoline) *in situ* to explore sensitivity to heterocycles which will impose conformational restrictions. Incubation with 1 µM PCY1 produced both macrocyclic and hydrolyzed linear peptides (Figure S25-S26), yielding 10.9 µM Thn containing macrocycle after ~ 20 h incubation, representing a 4.4 % conversion. Although a conversion this low is not synthetically useful, based on well established methods of optimization of enzyme activity including directed evolution this turnover represents a starting point for future improvement.[34] Further investigations are being conducted to study variants of the mutated enzyme and the reaction conditions in order to improve the synthesis of such hybrid cyclic peptides.

Summary

The importance of peptidic and amino acid-containing macrocycles in medicine is now clear. Robust synthetic approaches for these molecules, which can combine synthetic diversity with the ability of enzymes to catalyze challenging chemical steps, are highly desirable. Here we have reported the characterization of PCY1, a macrocyclase from plant. We show substrate recognition occurs in two distinct patches of the substrate, joined by a flexible linker, which underpins the promiscuity of PCY1. Based on understanding substrate recognition, we have demonstrated that the enzyme can utilize simpler synthetically accessible substrates with a short *C*-terminal extension at a synthetically useful rate. PCY1 can process hybrid molecules than are not solely composed of canonical amino acids, suggesting PCY1 could represent a valuable synthetic tool.

Materials and Methods

Materials

All reagents, but peptides, were purchased from Sigma-Aldrich, unless stated otherwise. All peptides were commercially obtained from Biosynthesis, unless otherwise stated. HPLC grade acetonitrile (MeCN) was purchased from Fisher. Aqueous buffers and aqueous mobile-phases for HPLC were prepared using water purified with an Elga® Purelab® Milli-Q water purification system (purified to 18.2 MΩ.cm) and filtered over 0.45 μm filters.

Cloning, expression and protein purification

Codon optimized full-length PCY1 (*Saponaria vaccaria*) including an *N*-terminal His_6-tag and a cleavable Tobacco etch virus (TEV) protease site (sequence ENLYFQG, cleaving between Q and G) encoded in pJ414 plasmid was purchased from DNA2.0. Over-expression of PCY1 was performed in *E. coli* BL21 (DE3) grown in LB media. Cultures were inoculated and grown at 37 °C until $OD_{600\ nm}$: ~ 0.6, after which the temperature was lowered to 16 °C and protein expression was induced with 0.5 mM IPTG (final concentration, Generon). Bacteria were grown for 20 h at 16 °C, and cells were harvested by centrifugation at 8,983*g* for 10 min at 4 °C (Avanti J-26S, rotor JLA-8.1000, New Brunswick Scientific). Wet cell pellets were stored at -80 °C until protein purification. Cell pellets were resuspended in buffer A (50 mM HEPES, 300 mM NaCl, 10 % (v/v) glycerol, 3 mM βME (Fischer Chemical), pH 8.0) supplemented with EDTA-free protease inhibitor tablets (Roche) and 0.4 mg DNase I from bovine pancreas per g wet cells, at 4 °C. Cells were disrupted by double passage through a cell disruptor (Constant Systems Ltd) at 30 kPSI. The lysate was subsequently cleared via centrifugation at 43,667*g* at 4 °C for 20 min (rotor JA 25.50, New Brunswick Scientific). The cleared lysate was loaded onto a 5 mL Co HiTrap TALON (GE Healthcare), pre-equilibrated in buffer A. Protein was eluted in buffer B (50 mM HEPES, 300 mM NaCl, 10 % (v/v) glycerol, 250 mM imidazole, 3 mM βME, pH 8.0). TEV protease was added to the eluted sample (1 mg per 30 mg of eluted protein) and this mixture was dialyzed into 1 L of buffer A at 4 °C overnight. The dialyzed sample was applied onto a 5 mL Co HiTrap TALON, equilibrated in buffer C (50 mM HEPES, 300 mM NaCl, 10 % (v/v) glycerol, 3 mM βME, pH 8.0). The flow-through of this

reverse Ni-affinity chromatography step was pooled and dialyzed into 1 L of buffer D (50 mM HEPES, 5 mM NaCl, 10 (v/v) glycerol, 3 mM βME, pH 8.0) at 4 °C for 1 h and subsequently subjected to anion exchange chromatography using a 5 mL HiTrap Q-FF column (GE Healthcare) and a gradient of buffer D and E (50 mM HEPES, 500 mM NaCl, 10 % (v/v) glycerol, 3 mM βME, pH 8.0). Eluted samples containing PCY1 (verified by SDS-PAGE) were pooled, concentrated and further purified using size-exclusion chromatography (HiLoad 16/600 Superdex 200 pg, GE Healthcare) pre-equilibrated in buffer F (50 mM HEPES, 150 mM NaCl, 10 % (v/v) glycerol, 1 mM TCEP (tris(2-carboxyethyl)phosphate), pH 8.0). Purity and identity of final protein sample were confirmed by SDS-PAGE and mass spectrometry.

Site directed mutagenesis

Site directed mutagenesis was performed using an overlaying mutation carrying primer pair (IDT) of ~30 nucleotides employing *Pfu* polymerase (Thermo Scientific) for amplification (95 °C 3 min initial denaturation; 15 x 95 °C 30s denaturation followed by T_m-5 °C 45 s annealing and 72 °C 2min/kb extension, 68 °C 20 min final extension), according to manufacturer instructions. The amplification product was DpnI digested (Thermo Scientific, according to manufacturer's instructions), and used for transformation of *E. coli* DH5α. Sequencing was performed by Eurofins.

Crystallization, data collection and crystallographic analysis

PCY1 and PCY1:S562A crystals (apo and as complexes) were obtained from hanging drop vapor diffusion crystallization experiments with 500 µL reservoir solution and 2 µL drops (1:1 protein/precipitant ratio). When additive screens were performed, 0.2 µL of additive screen (Hampton Research) was added to the 2 µL crystallization drop. Crystals were harvested, cryo protected and subsequently flash frozen in liquid nitrogen. All data sets were collected at Diamond Light Source (UK).

Apo PCY1 was crystallized at 12.2 mg/mL in 33% (w/v) PEG 2000, 113.75 mM Mg formate and 0.1 M Na cacodylate pH 7.0. For cryo protection 10% (v/v) glycerol were added to reservoir solution. Data was collected at 100 K at beamline i04. Data was processed using the processing pipeline xia2 3d.[35,36] PCY1:S562A complex with PresegA1 was crystallized at 13.3 mg/mL with 163 µM PresegA1 in 28% (w/v) PEG 5000 MME, 140 mM Mg acetate tetrahydrate and 0.1 M Tris pH 7.5. For cryo

protection 7% (v/v) glycerol were added to reservoir solution. Data was collected at 100 K at beamline i04-1. Data was processed using the processing pipeline xia2 3dii.[35,37] PCY1:S562A complex with PresegB1 was crystallized at 13.3 mg/mL with 163 µM PresegB1 in 34.5% (w/v) PEG 3350 and 70 mM Mg sulfate. For cryo protection 5 % (v/v) glycerol were added to reservoir solution. Data was collected at 100 K at beamline i04-1. Data was processed using the processing pipeline xia2 3d.[35,36] PCY1:S562A complex with PresegF1 was crystallized at 13.0 mg/mL with 161.3 µM PresegB1 in 27% (w/v) PEG 3350, 100 mM Calcium chloride and 0.1 M Bis-Tris at pH 6.5. For cryo protection 10% (v/v) glycerol were added to reservoir solution. Data was collected at 100 K at beamline i24. Data processing was performed with xia2 3dii.[35,37] Data processing, structure solution and structure refinement were performed using the CCP4 software suite.[38] Apo PCY1 structure was solved using molecular replacement using PHASER and 1QFS as search model (α/β-hydrolase and β-propeller domain as separated search models).[39] The initial solution was then completed with buccaneer and manual model building and refinement of the model were performed using Coot and refmac5, respectively, including TLS refinement and model validation using MolProbity.[38,40–46] For all co-complex crystal structures, apo PCY1 structure was used as search model.

ITC studies with PCY1 variants and peptide ligands

Equilibrium binding experiments with PCY1:S562A and different substrates were performed using a MicroCal PEAQ-ITC (Malvern) in buffer F (see 'Protein purification'). All samples were buffer-exchanged or solubilized into buffer F and further diluted in the same batch of buffer F, in order to minimize buffer mismatches. ITC binding experiments were performed using a cell solution containing PCY1:S562A and a titrant solution with each peptide. All peptides were used at 200 µM except for presegetalin A1-NH_2 (700 µM), DNASAPV and NASAPV (1.5 mM) as well as FSASYSSKPIQT (2 mM) . The cell contained 20 µM, 18.3 µM, 17.1 µM 15.5 µM, 100 µM, 52.6 µM, 52.5 µM and 59.8 µM PCY1:S562A for Presegetalin A1 , B1 , D1 , F1 and A1-NH_2, DNASAPV, NASAPV and FSASYSSKPIQT, respectively. The measurements were performed at a stir speed of 750 rpm at 25°C using a reference power of 10 µcal/s, except for A1 -NH_2, DNASAPVY, NASAPV and FSASYSSKPIQT (5 µcal/s). A buffer titration control in which peptide was titrated into buffer was performed for each experiment. Data processing, fitting to a non-linear

one-site binding model, and data plotting was performed using MicroCal PEAQ ITC analysis software (Malvern).

Single injection kinetic ITC measurements were performed using MicroCal PEAQ-ITC with 60 nM PCY1 in the syringe and 193 µM presegetalin B1 in the cell. To start the reaction, 2 µL of 9.06 µM PCY1 was injected into the cell (4 s injection time) containing 300 µL of 193.5 µM Presegetalin B1. The measurement was performed in buffer at 30 °C, at 750 rpm using a reference power of 5 µcal/s. Data processing for kinetic ITC measurements was performed using MicroCal PEAQ ITC analysis software. Data fitting and plotting were performed using Prism 6 (GraphPad Software). Steady-state kinetics data were fitted according to Michaelis Menten equation (corrected by enzyme concentration; equation (1)).

$$v_{\mathrm{i}} \cdot [\mathrm{E}]_{\mathrm{t}}^{-1} = \frac{k_{\mathrm{cat}} \cdot [S]}{K_{\mathrm{m}} + [S]} \qquad (1)$$

LC/MS assays: time courses and steady-state kinetics

All kinetic measurements (activity tests, progress curves and initial-rate experiments) were performed as discontinuous assays analyzing samples via LC/MS or MALDI (10-mer activity test). Reactions were performed in 20 mM Tris pH 8.5, 100 mM NaCl, 5 mM DTT (for comparative time courses of PresegA1, PresegB1 and PresegF1 0.2 mg mL^{-1} BSA was included). All reactions (comparative time courses, steady state kinetics, PCY1 and PatGmac comparison and recognition sequence analysis) were performed at 30 °C and at various time points 50 µL samples were quenched with 20 µL of 6 % (v/v) TFA. Steady-state kinetics were performed with 30 nM PCY1 for PresegB1, 0.1 µM PCY1 for PresegF1, 0.25 µM for FSASYSSKPIQT and 50 nM for FSASYSSKPIQT. Recognition sequence experiments were performed at 250 µM [substrate] and 0.5 µM PCY1. Assays comparing PCY1 and PatGmac were performed with 250 µM substrates and 0.5 µM PCY1 or 25 µM PatGmac. Reactions were quenched at different time points with 1.7 % TFA (final concentration), samples were spun down at 20,817g and 50 µL were loaded onto a ZORBAX SB-C18, 5 µm, 9.4 mm × 50 mm (Agilent) column connected to an Agilent LC−MS instrument (G6130B Single Quad, Agilent Technologies). Reactants were separated from products using a gradient from A (H_2O with 0.1% formic acid) to 80% B (acetonitrile), at a flow rate of 1.5 mL/min for 8.75 min. For data analysis, ion count peaks and UV absorbance peaks were integrated using

Agilent ChemStation software. For high resolution mass spectrometry and mass fragmentation of analysis of requirement of recognition sequence in presegetalins, and comparison between PCY1 and PatGmac, 10 µL of quenched reaction mixture was injected onto a ZORBAX SB-C18, 5 µm, 9.4 mm × 50 mm (Agilent) column connected to an UltiMate 3000 LC system (Thermo Scientific) connected to an Orbitrap Velos Pro MS system (Thermo Scientific). Reactants were separated from products using a gradient from A (H_2O with 0.1% formic acid) to 95% B (acetonitrile), at a rate of 1.5 mL/min for 12 min. For data analysis, ion count peaks were integrated using Thermo Xcalibur Qual browser software. For data analysis involving quantitative mass spec integration, calibration curves with cyclic peptide (SegF) standards (treated as reaction products) were performed in triplicate in order to linearly correlate ion counts to cyclic product concentration. SegF1 was biosynthesized using 0.5 µM PCY1 for 50 µM PresegF1, overnight incubation at 30 °C in reaction buffer. Reaction as quenched with 1.7 % TFA (final concentration). Semi-preparative RP-HPLC was performed on an Agilent Infinity 1260 series equipped with an MWD detector using a Macherey-Nagel Nucleodur C18 column (10 µm x 21 x 250 mm at 21 mL/min), fractions were collected automatically by peak detection at the specified wavelength using an Agilent 1260 Infinity preparative-scale fraction collector using the following chromatographic system: MeCN and 0.1 % aqueous TFA [95% TFA (5 min), linear gradient from 5 to 95% of MeCN (35 min), 95% MeCN (40 min)] and UV detection at 280 nm for peak collection, while monitoring at the additional two wavelengths 220 and 254 nm. Fractions of pure SegF were combined and freeze dried before being used for the assays.

Further data fitting and plotting was performed with Prism 6. Comparative time course measurements of PresegA1, PresegB1 and PreseF1 were fitted non-exponentially (equation (2)). Data for steady-state kinetics were fitted to a Michaelis Menten equation (corrected by enzyme concentration; equation (1)). Reactions investigating heterocycle-containing macrocyclic VTACITWP formation have been performed under different conditions: 100 mM NaCl, 100 mM TRIS pH 8.5, 5 mM DTT, 10 mM $MgCl_2$ and 5 mM ATP with 250 µM VTACITWPFQA or VTACITWPIQT and 10 µM $LynD_{fus.}$ and 1 µM PCY1 (added after 19 h). Reactions were stopped via filtration through 30 kDa cut off VIVASPIN 500 (Sartorius) concentrators.

$$y = y_0 + a(1 - e^{-k \cdot x}) \quad (2)$$

y_0: offset; k: apparent rate; a: amplitude

Peptide analysis on the TripleTOF 5600

Peptides were injected on an Acclaim PepMap 100 C18 100 µm × 2 cm trap and washed for 5 min to waste after which the trap was turned in-line with the analytical column (Acclaim Pepmap RSLC 75 µm × 15 cm, both ThermoFisher Scientific) using a nanoLC Ultra 2D plus loading pump and nanoLC as-2 autosampler (Eksigent).

The analytical solvent system consisted of buffer A (2% acetonitrile, 0.1% formic acid in water) and buffer B (2% water, 0.1% formic acid in acetonitrile) at a flow rate of 0.3 µl/min with the following gradient: linear 5-40% of buffer B for 6 min, linear 40-95% of buffer B for 3 min, isocratic 95% of buffer B for 3 min, linear 95-1% of buffer B for 2 min and isocratic 1% solvent buffer B for 6 min for re-equilibration.
The eluate was sprayed into a TripleTOF 5600 electrospray tandem mass spectrometer (ABSciex, Foster City, CA) and analyzed in Information Dependent Acquisition (IDA) positive ion mode. For each MS scan (m/z 150-1250) of 150ms, 20 MSMS (m/z 100-1800) of 80 ms with 10 s dynamic exclusion were acquired. Rolling collision energy with a collision energy spread of 5 eV was used for fragmentation.

To obtain better MSMS spectra of the peptides of interest, a product ion scan (PIS) of the relevant masses were analyzed. The LC condition were kept constant with an adjusted gradient: linear 5-40% of buffer B for 16 min, linear 40-95% of buffer B for 3 min, isocratic 95% of buffer B for 3 min, linear 95-1% of buffer B for 2 min and isocratic 1% solvent buffer B for 6 min for re-equilibration. The mass spectrometer was operated in product ion mode with four selected masses of interest. Each MS and MSMS scan (m/z 100-2400) was acquired for 100ms respectively.

In house peptide synthesis

Automated solid-phase peptide synthesis (SPPS) was carried out on a Biotage® Syro Wave™ system in polypropylene (PP) syringe with a PTFE frit. Final cleavage and deprotection were completed manually.
The different precursor peptides were synthesized by standard automated solid-phase

(SPPS) on a Chem-Matrix Rink amide resin (~ 0.5mmol/g) or a Trityl chloride resin (~ 0.4 mmol/g) using the Fmoc strategy and Fmoc-protected amino acids. A double coupling strategy using a 5-fold excess with 2-(1*H*-benzotriazol-1-yl)-1,1,3,3,-tetramethyluronium hexafluorophosphate (HBTU)/*N*,*N*-Diisopropylethylamine (DIPEA) (HBTU, 0.5M in DMF and DIPEA, 2M in NMP) and DIC/oxyma pure (DIC, 0.5M in DMF, Oxyma, 1M in DMF) was used for all amino acids for 30 minutes at 75 °C, except for the loading of the first amino acid on the trityl resin. In order to achieve that, to the resin (swelled in CH_2Cl_2) was added the amino acid (2 eq.) and DIPEA (4 eq.) in CH_2Cl_2 and the reaction agitated at room temperature for 2 hours. The reaction mixture was discarded, the resin was washed and capped by the addition of MeOH and DIPEA in CH_2Cl_2 (2:1:18) for one hour at room temperature. After washing the resin, the peptide synthesis is completed as described. The Fmoc deprotection was done in 20% piperidine/DMF for 12 min at rt. For the final cleavage and side chain deprotection, the beads (washed with CH_2Cl_2 and dried) were transferred into a flacon tube and the cleavage cocktail was added and left shaking for 2h: 96% TFA, 2.5% H_2O, 1.5% triisopropyl silane. The resin was filtered, washed with CH_2Cl_2, and the filtrate concentrated under reduced pressure. The peptides were then precipitated in cold Et_2O and the precipitate purified by prep-HPLC.

Semi-preparative RP-HPLC was performed on an Agilent Infinity 1260 series equipped with an MWD detector using a Macherey-Nagel Nucleodur C18 column (10 µm x 21 x 250 mm at 21 mL/min) and fractions were collected automatically by peak detection at the specified wavelength using an Agilent 1260 Infinity preparative-scale fraction collector using the following chromatographic system: MeCN and 0.1 % aqueous TFA [95% TFA (5 min), linear gradient from 5 to 95% of MeCN (35 min), 95% MeCN (40 min)] and UV detection at 280 nm for peak collection, while monitoring at the additional two wavelengths 220 and 254 nm. Fractions of the pure peptides were combined and freeze dried before being used for the assays.

Analytical RP-HPLC-MS was performed on an Agilent infinity 1260 series equipped with a MWD detector using a Macherey-Nagel Nucleodur C18 column (10 µm x 4.6 x 250 mm) and connected to an Agilent 6130 single quad apparatus equipped with an electrospray ionization source using the following chromatographic system. 1 mL/min flow rate with MeCN and 0.1 % aqueous TFA [95% TFA (5 min), linear gradient from 5 to 95% of MeCN (35 min), 95% MeCN (40 min)] and UV detection at 220 nm. Peptides with truncated recognition sequences were purified by semi-preparative RP-

HPLC (Agilent infinity 1260 series, Agilent Technologies) using a Jupiter 300 C4 column (Phenomenex, 10 μM x 16 x 250 mm at 10 mL/min) and fractions were collected automatically by peak detection at 220nm using an Agilent 1260 Infinity preparative-scale fraction collector. Solvents were 0.1 % aqueous TFA (solvent A) and acetonitrile (solvent B). Gradients used for purification of different peptides were as follows: Peptides GVAWAI, GVAWAIQ and GVAWAIQT, 5% B (5 minutes), linear gradient from 5 to 40% B (10 minutes), followed by linear gradient from 40-44% acetonitrile (20 minutes); Peptides GVAWAF, GVAWAFQ and GVAWAFQA, 5% B (5 minutes), linear gradient from 5 to 41% B (10 minutes), followed by linear gradient from 41-50% acetonitrile (20 minutes); Peptides FSASYSSKPI, FSASYSSKPIQ, FSASYSSKPIQT, FSASYSSKPF, FSASYSSKPFQ and FSASYSSKPFQA, 5% B (5 minutes), linear gradient from 5 to 15% B (10 minutes), followed by linear gradient from 15-20% acetonitrile (20 minutes). Fractions containing each peptide of interest were pooled, their identity and purity verified by LC-MS and lyophilized for later use.

PDB accession codes: Coordinates have been deposited in the Protein Data Bank, with accession codes 5O3X (apo PCY1), 5O3W (PCY1-S562A mutant bound to PresegA1 peptide), 5O3V (PCY1-S562A mutant bound to PresegB1 peptide), and 5O3U (PCY1-S562A mutant bound to PresegF1 peptide).

ACKNOWLEDGEMENTS

We thank O. K. Miller for crystal handling, and C. H. Botting and S. L. Shirran, for mass spectrometry services. We acknowledge DLS beamtime, R. J. M. Goss for usage of an UltiMate3000 Orbitrap Velos Pro LC/MS & C. Cartmell and C. S. Bailey for assistance with the instrument. H.L. is funded by the George & Stella Lee Scholarship and EPSRC. E.S.M. and M.A. are also funded by EPSRC. This project was funded by the European Research Council project 339367 NCB-TNT (J.H.N.). We dedicate this paper to the memory of our friend and co-author A. F. Bent, who died aged 32 from cancer.

SUPPLEMENTARY INFORMATION

The Supporting Information containing substrates used, LigPlot, additional ITC and x-ray crystallography figures is available free of charge on the ACS Publications website.

COMPETING FINANCIAL INTERESTS

HL, CMC and JHN are named on a patent that utilizes PCY1.

Tables

Table 1. Kinetic parameters for PCY1.

Substrate	K_m **(μM)**	k_{cat} **(s^{-1})**	k_{cat}/K_m **(s^{-1} M^{-1})**	**Method**
PresegB1	0.45 ± 0.17*	0.37 ± 0.02	820,000	LC/MS ($UV_{280\ nm}$)
PresegB1	0.25 ± 0.01	0.14 ± 0.01	560,000	ITC
PresegF1	2.4 ± 0.5	0.12 ± 0.01	50,000	LC/MS (ion count)
FSASYSSKPFQA	83 ± 18	0.1 ± 0.01	1,200	LC/MS (ion count)
FSASYSSKPIQT	227 ± 166	0.016 ± 0.001	70	LC/MS (ion count)

Table 2: Binding parameters of PCY1:s562A and various peptides obtained by ITC.

Peptide	**N (sites)**	K_d **(μM)**	**Δ** ***H*** **(kcal/mol)**	**Δ** ***G*** **(kcal/mol)**	**-T Δ** ***S*** **(kcal/mol)**
PresegA1	1	0.173 ± 0.027	-10.0 ± 0.1	-9.23	0.81
PresegB1	1	0.238 ± 0.022	-12.3 ± 0.1	-9.04	3.25
PresegD1	1	0.250 ± 0.039	-6.8 ± 0.1	-9.01	-2.19
PresegF1	1	0.136 ± 0.019	-11.4 ± 0.1	-9.37	2.03
PresegA1-NH_2	1	12.1 ± 2.0	-5.3 ± 0.2	-6.71	-1.41
FSASYSSKPIQT	1	56.0 ± 76.5	-2.8 ± 3.1	-5.90	-3.14
DNASAPV	1	9.58 ± 0.58	-14.5 ± 0.2	-6.85	7.64
NASAPV	1	25.2 ± 1.2	-14.0 ± 0.1	-6.27	7.70

Figure legends

Figure 1: Reaction scheme of macrocyclization catalyzed by PCY1 and sequence alignment of presegetalins. a) amino acid one letter code is used to abbreviate amino acids, not depicted with the chemical formula. Transparent red circle is highlighting the peptide bond formed via macrocyclization. b) Red dots highlight end of core sequence. Sequence identity is highlighted with a gradient from white to blue. Deep blue showing highest sequence identity.

Figure 2: Kinetic analysis of macrocyclization catalyzed by PCY1.

a) Comparative time course of PresegA1 (green), PresegB1 (blue) and PresegF1 (red) formation with PCY1. Data were fitted mono exponentially. b) LC/MS based steady state kinetics of PresegF1 with PCY1 using ion count quantification and product calibration curve, fitted non exponentially. c) LC/MS based steady state kinetics of PresegB1 with PCY1 using quantification with UV at 280 nm, fitted non exponentially. d) Single injection ITC steady-state kinetics of PresegB1, fitted non exponentially. Reaction temperature for all experiments: 30 °C

Figure 3: Binding studies of PCY1:S562A and presegetalins.

Structure of PCY1:S562A bound to PresegF1. PCY1:S562A is drawn in cartoon style, α-helices are colored orange, β-sheets are colored metallic blue and loops are colored in light grey. PresegF1 is shown as ball-and-stick model in dark red. Nitrogen atoms are colored in blue and oxygen atoms are colored in light red. $2F_o$-F_c electron density map is shown in mesh style (in blue) and rendered at s: 1.0. a) Overview of the structure with peptide shown in spheres showing details into peptide binding site. b) Interactions between PresegF1 *C*-terminus and PCY1:S562A, depicting interactions between the peptide carboxylic acid and the enzyme α-helix. c) Overlay between complex structures, peptides are shown as ball-and-stick model and colored green (PresegA1), blue (PresegB1) and red (PresegF1). Figure was adapted from Ligplot[47], and additional figures depicting interactions with the peptide ligand are available in the supporting information.

Figure 4: Structural analysis of PCY1:S562A complex structures.
a) ITC curves showing equilibrium binding measurements for PresegA1 and PresegA1-NH_2 with PCY1:S562A. b) Thermodynamic parameters derived from equilibrium binding experiments for a selection of presegetalins. c) Chemical structure of PresegA1 variants.

Figure 5: Expanding the substrate scope of PCY1. Native cyclic peptides (blue), cyclic peptides containing non-amino acids and non-native sequences produced by PCY1 via macrocyclization.

Figure 1

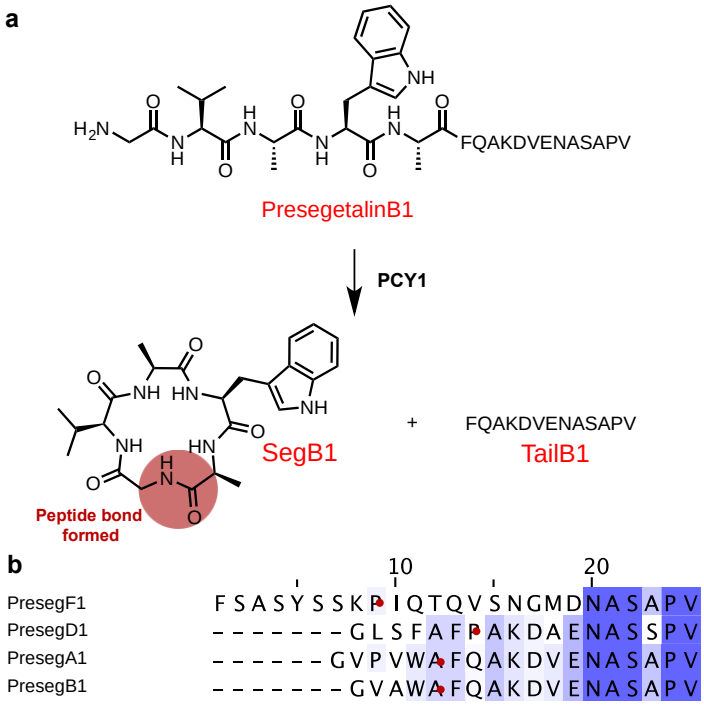

a
FQAKDVENASAPV
PresegetalinB1
PCY1
SegB1
+
FQAKDVENASAPV
TailB1
Peptide bond formed
b
10
20
PresegF1 FSASYSSKPIQTQVSNGMDNASAPV
PresegD1 -------GLSFAFPAKDAENASSPV
PresegA1 ------GVPVWAFQAKDVENASAPV
PresegB1 -------GVAWAFQAKDVENASAPV

Figure 2

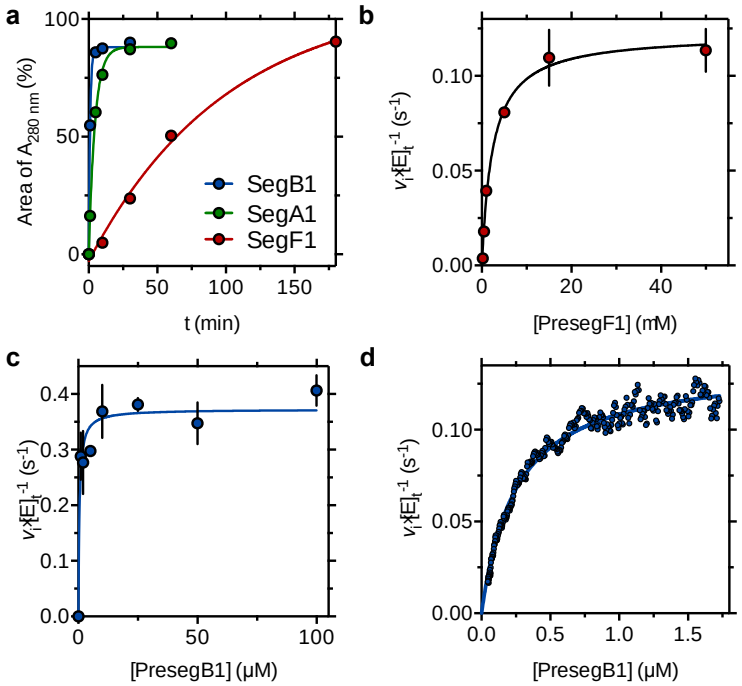
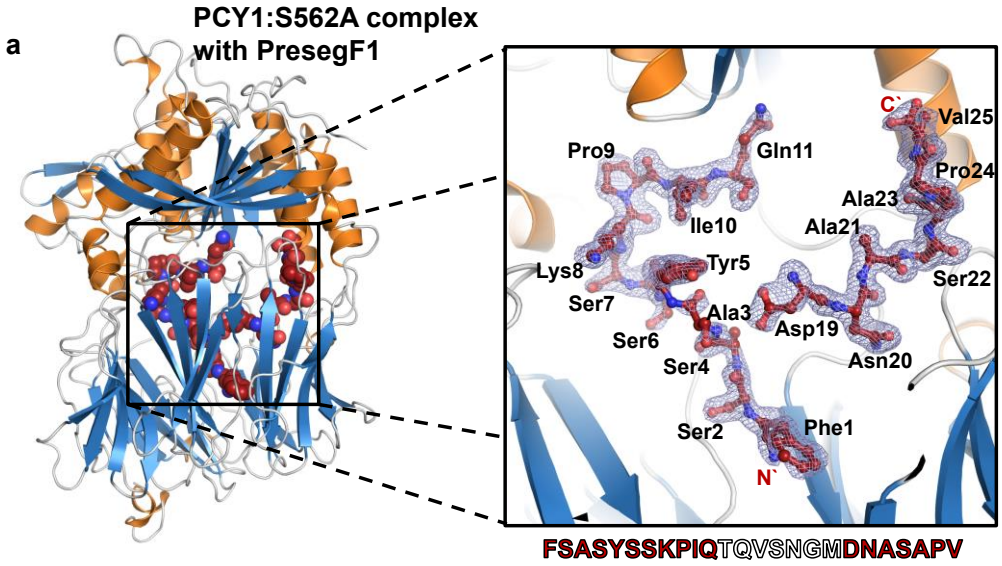
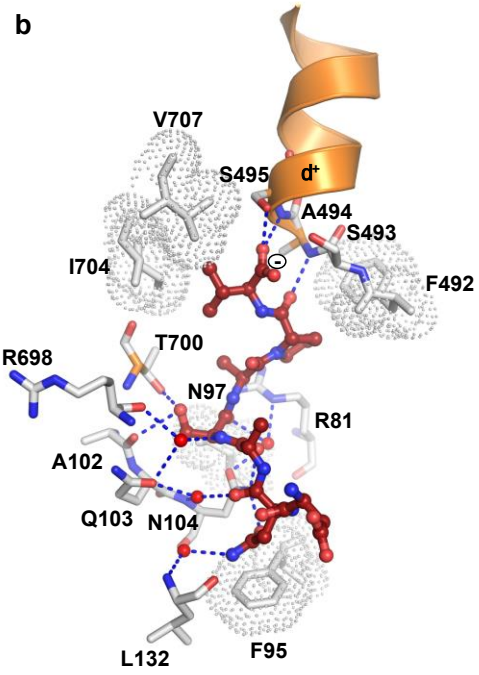
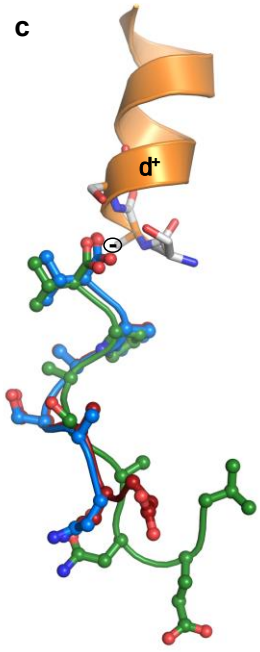

a
b
c
d
Area of A280 nm (%)
t (min)
SegB1
SegA1
SegF1
[PresegF1] (mM)
[PresegB1] (μM)
[PresegB1] (μM)

Figure 3


a
PCY1:S562A complex
with PresegF1
C`
Val25
Pro9
Gln11
Pro24
Ala23
Ile10
Ala21
Lys8
Tyr5
Ser22
Ser7
Ala3
Asp19
Ser6
Asn20
Ser4
Ser2
Phe1
N`
FSASYSSKPIQTQVSNGMDNASAPV


b
V707
S495
d+
A494
S493
I704
F492
T700
R698
N97
R81
A102
Q103
N104
L132
F95


c
d+

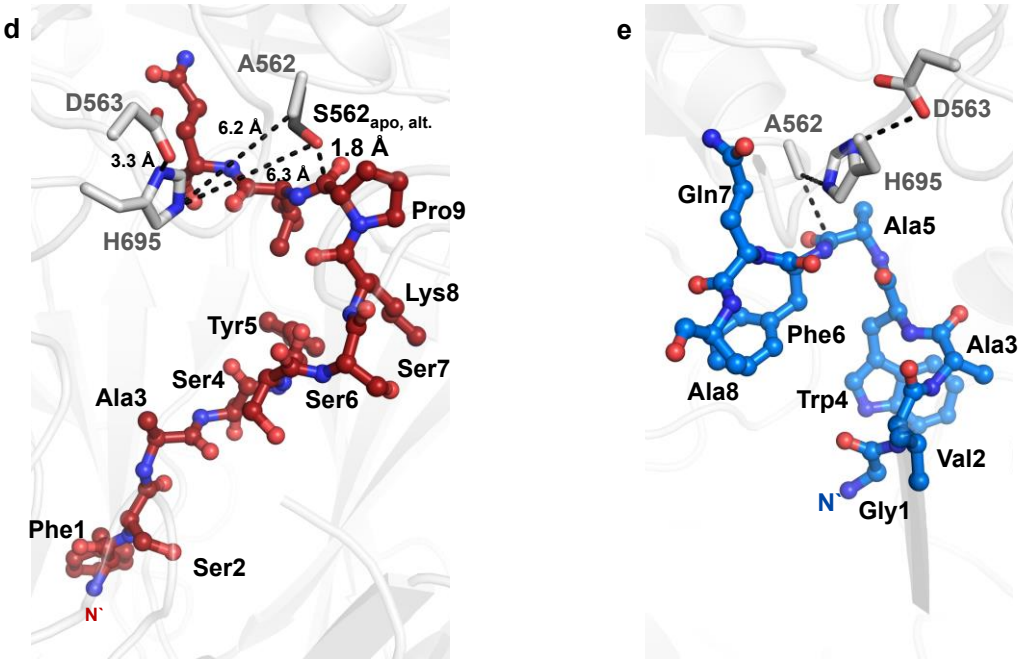

d
A562
D563
S562apo, alt.
6.2 Å
1.8 Å
3.3 Å
6.3 Å
Pro9
H695
Lys8
Tyr5
Ser7
Ser4
Ala3
Ser6
Phe1
Ser2
N`
e
D563
A562
H695
Gln7
Ala5
Phe6
Ala3
Ala8
Trp4
Val2
N`
Gly1

Figure 4

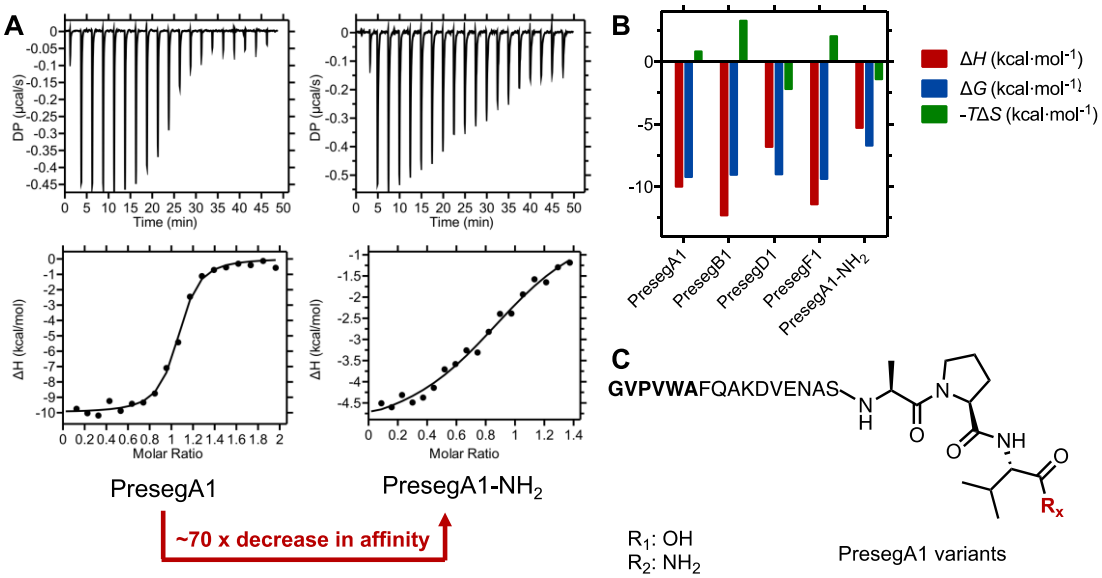
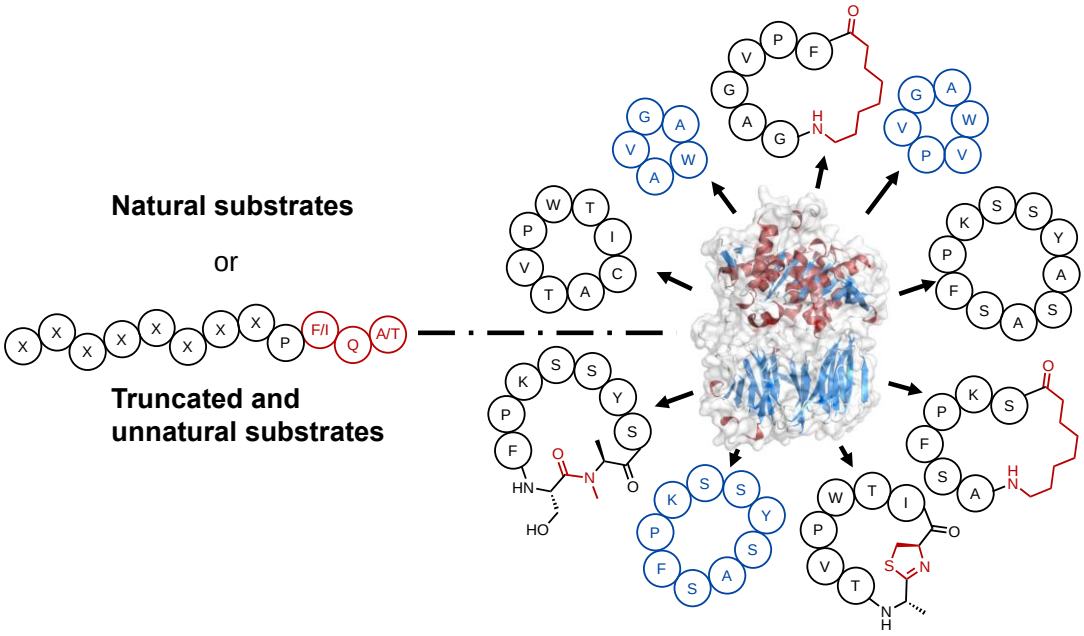
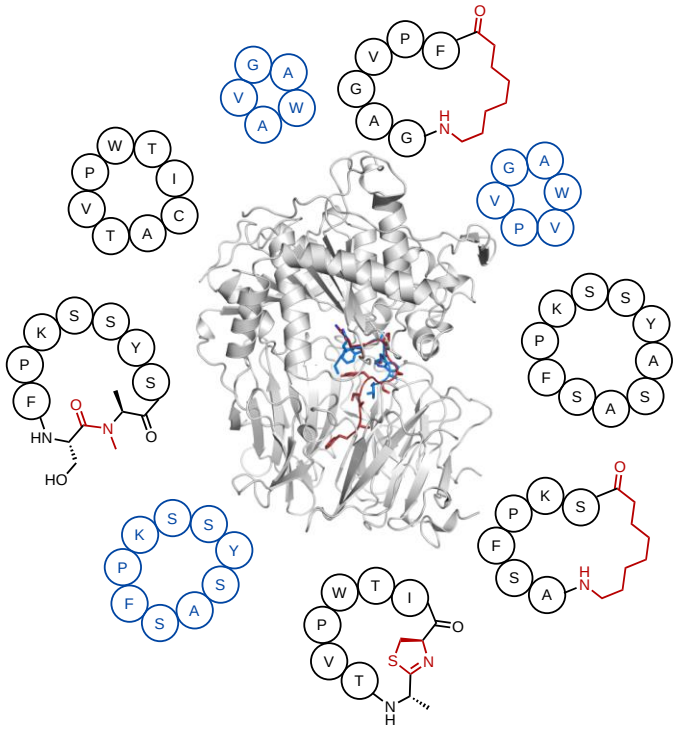

A
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
Molar Ratio
PresegA1
PresegA1-NH2
~70 x decrease in affinity
B
ΔH (kcal·mol-1)
ΔG (kcal·mol-1)
-TΔS (kcal·mol-1)
PresegA1
PresegB1
PresegD1
PresegF1
PresegA1-NH2
C
GVPVWAFQAKDVENAS
R1: OH
R2: NH2
PresegA1 variants

Figure 5


Natural substrates
or
Truncated and
unnatural substrates

TOC figure


G A V A W
P V F G A G
W T I P V T A C
G A V W P V
K S S Y P F S A S
K S S Y P F S A S
K S S Y P F S A S
W T I P V T

REFERENCES

(1) Driggers, E. M., Hale, S. P., Lee, J., and Terrett, N. K. (2008) The exploration of macrocycles for drug discovery — an underexploited structural class. *Nat. Rev. Drug Discovery 7*, 608–624.
(2) Lipinski, C. A., Lombardo, F., Dominy, B. W., and Paul J. Feeney, P. J. F. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev. 23*, 3–25.
(3) Terrett, N. (2013) Drugs in middle space. *MedChemComm 4*, 474–475.
(4) Zorzi, A., Deyle, K., and Heinis, C. (2017) Cyclic peptide therapeutics: past, present and future. *Curr. Opin. Chem. Biol. 38*, 24–29.
(5) Yu, X., and Sun, D. (2013) Macrocyclic Drugs and Synthetic Methodologies toward Macrocycles. *Molecules 18*, 6230–6268.
(6) Martí-Centelles, V., Pandey, M. D., Burguete, M. I., and Luis, S. V. (2015) Macrocyclization Reactions: The Importance of Conformational, Configurational, and Template-Induced Preorganization. *Chem. Rev. 115*, 8736–8834.
(7) Sardar, D., Lin, Z., and Schmidt, E. W. (2015) Modularity of RiPP Enzymes Enables Designed Synthesis of Decorated Peptides. *Chem. & Biol. 22*, 907–916.
(8) Sivonen, K., Leikoski, N., Fewer, D. P., and Jokela, J. (2010) Cyanobactins—ribosomal cyclic peptides produced by cyanobacteria. *Appl. Microbiol. Biotechnol. 86*, 1213–1225.
(9) Sardar, D., Tianero, M. D., and Schmidt, E. W. (2016) Directing Biosynthesis, in *Methods Enzymol.*, pp 1–20. Elsevier.
(10) Tianero, M. D., Pierce, E., Raghuraman, S., Sardar, D., McIntosh, J. A., Heemstra, J. R., Schonrock, Z., Covington, B. C., Maschek, J. A., Cox, J. E., Bachmann, B. O., Olivera, B. M., Ruffner, D. E., and Schmidt, E. W. (2016) Metabolic model for diversity-generating biosynthesis. *Proc. Natl. Acad. Sci. U. S. A. 113*, 1772–1777.
(11) Oueis, E., Adamson, C., Mann, G., Ludewig, H., Redpath, P., Migaud, M., Westwood, N. J., and Naismith, J. H. (2015) Derivatisable Cyanobactin Analogues: A Semisynthetic Approach. *ChemBioChem 16*, 2646–2650.
(12) Oueis, E., Jaspars, M., Westwood, N. J., and Naismith, J. H. (2016) Enzymatic Macrocyclization of 1,2,3-Triazole Peptide Mimetics. *Angew. Chem., Int. Ed. 55*, 5842–5845.
(13) Oueis, E., Nardone, B., Jaspars, M., Westwood, N. J., and Naismith, J. H. (2016) Synthesis of Hybrid Cyclopeptides through Enzymatic Macrocyclization. *ChemistryOpen 6*, 11–14.
(14) Houssen, W. E., Bent, A. F., McEwan, A. R., Pieiller, N., Tabudravu, J., Koehnke, J., Mann, G., Adaba, R. I., Thomas, L., Hawas, U. W., Liu, H., Schwarz-Linek, U., Smith, M. C. M., Naismith, J. H., and Jaspars, M. (2014) An Efficient Method for the In Vitro Production of Azol(in)e-Based Cyclic Peptides. *Angew. Chem., Int. Ed. 53*, 14171–14174.
(15) 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.
(16) Luo, H., Hong, S.-Y., Sgambelluri, R. M., Angelos, E., Li, X., and Walton, J. D. (2014) Peptide Macrocyclization Catalyzed by a Prolyl Oligopeptidase Involved in α-Amanitin Biosynthesis. *Chem. & Biol. 21*, 1610–1617.
(17) Barber, C. J. S., 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.
(18) Nguyen, G. K. T., Wang, S., Qiu, Y., Hemu, X., Lian, Y., and Tam, J. P. (2014) Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol. 10*, 732–738.
(19) Nguyen, G. K. T., Kam, A., Loo, S., Jansson, A. E., Pan, L. X., and Tam, J. P. (2015) Butelase 1: A Versatile Ligase for Peptide and Protein Macrocyclization. *J. Am. Chem. Soc. 137*, 15398–15401.
(20) Nguyen, G. K. T., Qiu, Y., Cao, Y., Hemu, X., Liu, C.-F., and Tam, J. P. (2016) Butelase-mediated cyclization and ligation of peptides and proteins. *Nat. Protoc. 11*, 1977–1988.
(21) Li, K., Condurso, H. L., Li, G., Ding, Y., and Bruner, S. D. (2016) Structural basis for precursor protein–directed ribosomal peptide macrocyclization. *Nat. Chem. Biol. 12*, 973–979.
(22) Yang, R., Wong, Y. H., Nguyen, G. K. T., Tam, J. P., Lescar, J., and Wu, B. (2017) Engineering a Catalytically Efficient Recombinant Protein Ligase. *J. Am. Chem. Soc. 139*, 5351–5358.
(23) Czekster, C. M., and Naismith, J. H. (2017) Kinetic Landscape of a Peptide Bond-Forming Prolyl Oligopeptidase. *Biochemistry 56*, 2086–2095.
(24) Czekster, C. M., Ludewig, H., McMahon, S. A., and Naismith, J. H. (2017) Characterization of a dual function macrocyclase enables design and use of efficient macrocyclization substrates. *Nat. Commun. 8*, 1045.
(25) Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K.-D., Fischbach, M. A., Garavelli, J. S., Göransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Müller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J. T., Rebuffat, S., Ross, R. P., Sahl, H.-G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R. D., Tagg, J. R., Tang, G.-L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep. 30*, 108–160.
(26) Chekan, J. R., Estrada, P., Covello, P. S., and Nair, S. K. (2017) Characterization of the macrocyclase involved in the biosynthesis of RiPP cyclic peptides in plants. *Proc. Natl. Acad. Sci. U. S. A.* 201620499.
(27) Hedstrom, L. (2002) Serine Protease Mechanism and Specificity. *Chem. Rev. 102*, 4501–4524.
(28) 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. M., 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.
(29) Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. (2011) The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry 50*, 4402–4410.

(30) Polgar, L. (2002) The prolyl oligopeptidase family. *Cell. Mol. Life Sci. 59*, 349–362.
(31) Page, M. I., and Jencks, W. P. (1971) Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and the Chelate Effect. *Proc. Natl. Acad. Sci. U. S. A. 68*, 1678–1683.
(32) Doak, B. C., Zheng, J., Dobritzsch, D., and Kihlberg, J. (2016) How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets. *J. Med. Chem. 59*, 2312–2327.
(33) Koehnke, J., Mann, G., Bent, A. F., Ludewig, H., Shirran, S., Botting, C., Lebl, T., Houssen, W. E., Jaspars, M., and Naismith, J. H. (2015) Structural analysis of leader peptide binding enables leader-free cyanobactin processing. *Nat. Chem. Biol. 11*, 558–563.
(34) Khersonsky, O., and Tawfik, D. S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem. 79*, 471–505.
(35) Kabsch, W. (2010) XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 125–132.
(36) Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 62*, 72–82.
(37) Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr. 69*, 1204–1214.
(38) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 235–242.
(39) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr., 40*, 658–674.
(40) Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 62*, 1002–1011.
(41) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126–2132.
(42) Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 355–367.
(43) Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 57*, 122–133.
(44) Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 62*, 439–450.
(45) Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B. 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res. 35*, W375-383.
(46) Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 12–21.

(47) Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model. 51*, 2778–2786.