Chemistry & Biology
Article

Structures of Cyanobactin Maturation Enzymes Define a Family of Transamidating Proteases

Vinayak Agarwal,[1,2] Elizabeth Pierce,[4] John McIntosh,[4] Eric W. Schmidt,[4,*] and Satish K. Nair[1,2,3,*]
[1]Institute for Genomic Biology
[2]Center for Biophysics and Computational Biology
[3]Department of Biochemistry
University of Illinois at Urbana Champaign, Champaign, IL 61820, USA
[4]Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT 84112, USA
*Correspondence: ews1@utah.edu (E.W.S.), snair@illinois.edu (S.K.N.)
http://dx.doi.org/10.1016/j.chembiol.2012.09.012

SUMMARY

Cyanobactins, a class of ribosomally encoded macrocylic natural products, are biosynthesized through the proteolytic processing and subsequent N-C macrocylization of ribosomal peptide precursors. Macrocylization occurs through a two-step process in which the first protease (PatA) removes the amino terminal flanking sequence from the precursor to yield a free N terminus of the precursor peptide, and the second protease (PatG) removes the C-terminal flanking sequence and then catalyzes the transamidation reaction to yield an N-C cyclized product. Here, we present the crystal structures of the protease domains of PatA and PatG from the patellamide cluster and of PagA from the prenylagaramide cluster. A comparative structural and biochemical analysis of the transamidating PatG protease reveals the presence of a unique structural element distinct from canonical subtilisin proteases, which may facilitate the N-C macrocylization of the peptide substrate.

INTRODUCTION

Small molecule peptide natural products continue to represent major avenues towards the development of new antibiotics and antiproliferatives. Many of these natural products exhibit potent biological activities against a number of therapeutic targets, and the range of schemes for the biosynthesis of such molecules gives rise to a diversity of structures (Fischbach and Walsh, 2009; Newman and Cragg, 2007). The structural diversity of ribosomally synthesized peptides has only begun to be appreciated with the realization that such peptides can undergo a range of posttranslational modifications.

One modification that is prevalent in peptide natural products is macrocyclization, which brings together distant parts of an otherwise elongated molecule in close proximity. Examples of such ribosomally synthesized and posttranslationally modified peptides (RiPPs) that contain macrocycles include lantibiotics that contain (methyl)lanthionine-containing thioethers (Figure 1A) (Knerr and van der Donk, 2012), thiopeptides that are constrained through pyridine rings generated by Diels-Alder-type condensations (Figure 1B) (Walsh et al., 2010), and thuricin and subtilosin bearing sulfur-α carbon thioethers (Flühe et al., 2012; Rea et al., 2010) (Figure 1). Cyclic peptides are widely sought in pharmaceutical discovery and development because their constrained structures often provide better target specificity and improved pharmacological and stability properties in comparison to their linear relatives. Macrocyclization also provides rigidity to the structure of the molecules and can also confer thermostability and chemical resistance to the cyclized natural product (Garg et al., 2012).

Cyanobactins (Figures 1C–1E) are a class of RiPPs that are found in greater than 30% of all cyanobacteria. In contrast to many other macrocyclic natural products, cyanobactins are macrocyclized via the peptide backbone and not via amino acid side chains. All cyanobactins are synthesized as precursor peptides ("PatE" and relatives) that are matured by two proteases, exemplified by PatA and PatG in the patellamide biosynthetic pathway (Figure 2) (Lee et al., 2009; Schmidt et al., 2005). In addition to macrocyclization, cyanobactins may contain Ser, Thr, and Cys residues, which are heterocyclized with the main-chain atoms to yield (methyl)oxazoline and thiazoline five-membered rings (McIntosh et al., 2010a; McIntosh and Schmidt, 2010). Further enzymatic oxidation of these rings yields (methyl)oxazole and thiazole rings. Other posttranslational modifications include prenylation at Ser/Thr or Tyr/Trp residues catalyzed by the PatF class of prenyltransferases (McIntosh et al., 2011). Metagenomic studies have revealed the widespread presence of homologs of patellamide biosynthetic genes in the marine microbiota, such as trunkamides, prenylagaramides, and arthrospiramides (Donia and Schmidt, 2011; Schmidt and Donia, 2009; Sivonen et al., 2010). The biosynthetic route for cyanobactins represents one of the most broad substrate RiPP pathways so far characterized and represents an ideal starting point for engineering and genome mining approaches to peptide macrocyclization.

Intriguingly, despite being paralogs that are very sequence similar (greater than 40% similarity), PatA and PatG catalyze quite biochemically distinct reactions. PatA recognizes a five-amino acid sequence on the PatE coding cassette as a site for cleavage, synthesizing short, linear peptides. In turn, these peptides are substrates for PatG, which recognizes a three-amino acid C-terminal sequence (Lee et al., 2009). PatG cleaves this sequence from the core peptide and macrocyclizes the product.

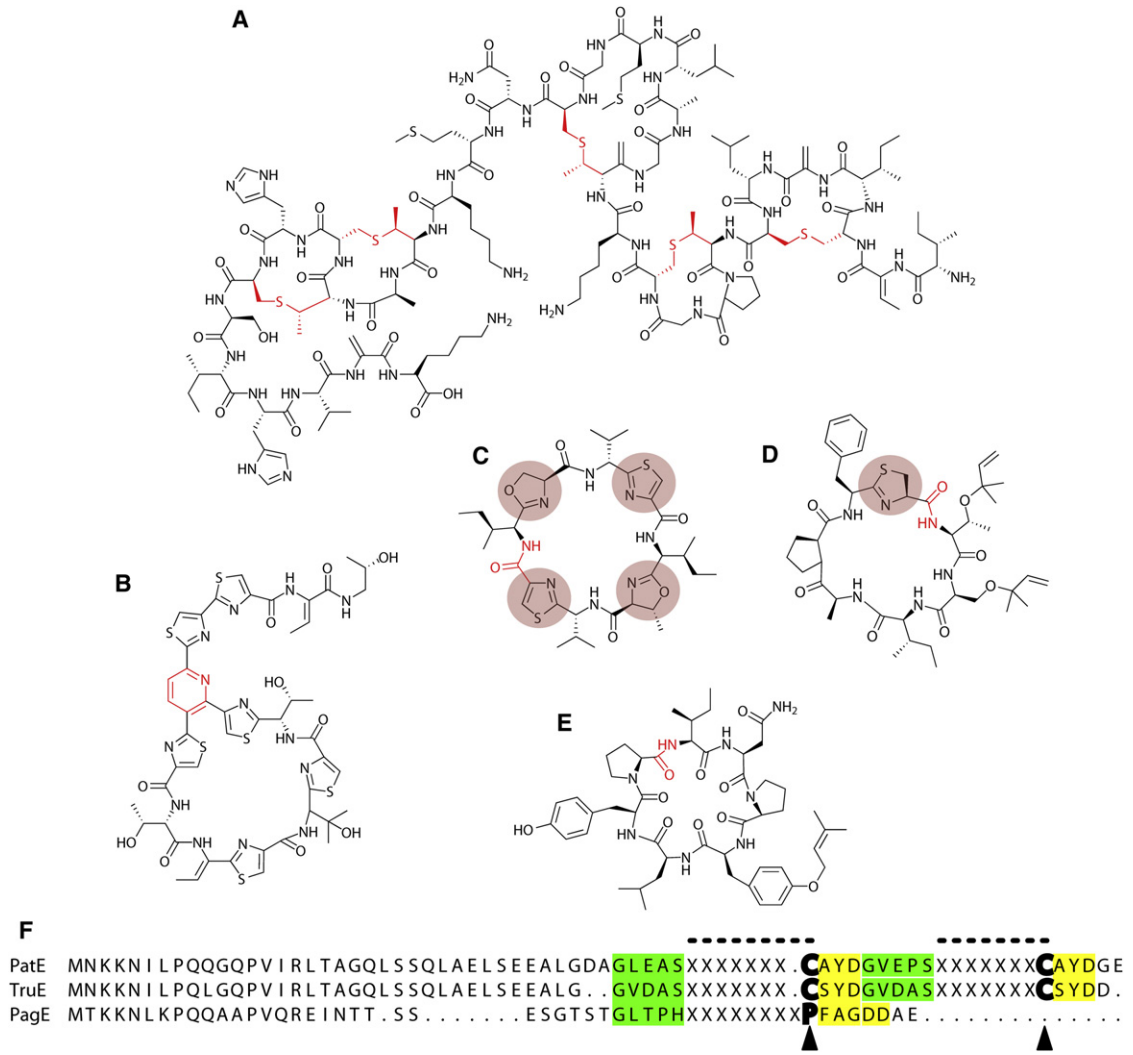

Figure 1. Chemical Structures of Representative Macrocyclized Natural Products

(A) Lantibiotic nisin.

(B) Thiopeptide thiocillin I.

(C) Cyanobactin patellamide A.

(D) Cyanobactin trunkamide.

(E) Cyanobactin prenylagaramide B. The macrocyclizing covalent bonds are highlighted in red. Five membered heterocycles in patellamide A and trunkamide are highlighted. Prenylagaramide is devoid of heterocycles.

(F) Primary sequence alignment of the PatE, TruE, and PagE substrate peptide primary sequences. The hypervariable cyanobactin cassette residues (shown by dashed line above the sequence) are denoted by X. The cyanobactin cassettes are flanked by conserved sequences at the N termini and C termini (shaded green and yellow, respectively). Note that the last residue of the cyanobactin coding cassette (in bold and marked by ▲ under the sequence) is a cysteine residue (for PatE and TruE) which is subsequently heterocyclized, or a heterocycle mimicking proline residue (for PagE).

In contrast to other RiPPs, the core regions of the Pat cassettes are hypervariable, whereas the flanking regions are highly conserved, leading to sequence-diverse macrocycles. The presence of diverse posttranslational modifications, paired with the hypervariable core peptide, has resulted in >200 known cyanobactin natural products and >70 widely sequence-diverse, engineered derivatives that we have previously reported (Donia et al., 2008; Donia and Schmidt, 2011; Tianero et al., 2012). In addition, there are many other homologous cyanobactin pathways that have different recognition elements and lead to different products (Donia et al., 2006, 2008; Donia and Schmidt, 2011).

We previously showed that just the protease domains of PatA and PatG are sufficient to recapitulate activity in vitro (Lee et al., 2009). As discerned from the many known cyanobactin gene clusters, PatA, PatG, and their homologs fall into a single family (clade) within the large subtilisin protease group (Donia and Schmidt, 2011). By carefully manipulating substrates, we showed that PatG follows the normal subtilisin-like proteolytic route, where substrates are cleaved and become covalently bound to the active-site Ser (McIntosh et al., 2010b). Subsequently, displacement by the internal amine nucleophile leads to transamidation and macrocyclization. We also observed the

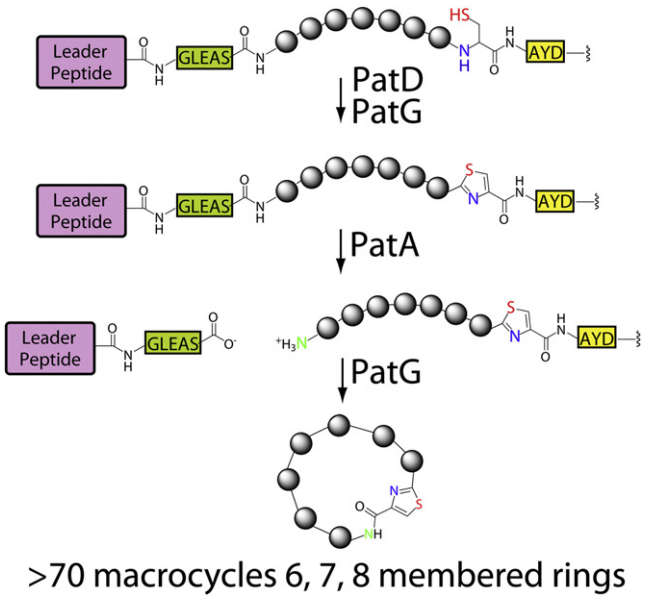

Figure 2. The Proposed Enzymatic Scheme for the Biosynthesis of Patellamide and Related Cyanobactins

For clarity, only one substrate cassette for PatE is shown. The conserved amino acids flanking the amino and carboxy termini of the Pat cassettes are colored as in Figure 1, with the coding cassette hypervariable amino acids shown as spheres. Note that the terminal coding cassette residue is a cysteine, which is heterocyclized and oxidized to a thiozole. A proline residue is present in PagE at this position, and a proline can substitute for cysteine in PatE.

alternative route, hydrolysis to linear peptides, with some substrates. The recent structure determination of the PatG protease domain provided strong evidence in support of this mechanistic hypothesis (Koehnke et al., 2012).

Here, we have biochemically determined the domain boundaries of the protease domains of both of the patellamide processing enzymes and the first protease from the prenylagaramide cluster. We also report high-resolution crystal structures of PatA, PagA, and PatG that elucidate the key features responsible for the different substrate recognition and reactivities of the enzymes. A comparison of the structures reveals the presence of an unexpected insert in the PatG protease domain, which may provide a molecular rationale for the macrocyclization reaction. The structures serve to explain how highly related paralogs promote such different chemical reactions, leading to a broad array of RiPP natural products with therapeutic potential. It should be noted that the in vitro activity for both the PatA protease domain and the PatG protease domain structures described in this study has been experimentally established (McIntosh et al., 2010b, 2011).

RESULTS AND DISCUSSION

Determination of Cyanobactin Protease Domain Boundaries

Cyanobactin gene clusters identified via metagenomic approaches are characterized by the presence of two distinct open reading frames (ORFs) that encode for subtilisin-like proteases. The proteases do not typically exist as independent units but are rather domains of 300–360 residues that are embedded within much larger multidomain ORFs (Lee et al., 2009). For homologs of *patA*, the protease domain is localized at the amino terminus of the corresponding ORF, followed by a linker region and a domain of unknown function (DUF3) at the carboxy terminus. This domain organization is conserved among all *patA* homologs that have been sequenced from homologous gene clusters (Figure 3A). For homologs of *patG*, the protease domain can either be localized to the amino terminus followed by a DUF domain or be sandwiched between an amino terminal flavin-dependent oxidase domain and the carboxy terminal DUF domain (Figure 3B). In protease-containing ORFs that do not have the amino-terminal oxidase domain, the oxidase may be present as a stand-alone ORF (Thc cluster) or be totally absent (Pag cluster) (Donia and Schmidt, 2011).

Attempts at heterologous production and purification of full-length ORFs encompassing the PatA and PatG proteases (as well as other homologs) in *Escherichia coli* yielded only impure and proteolytically degraded protein samples, which could not be resolved by mutation of the active site Ser to Ala to prevent autoprocessing. Consequently, further attempts at protein production focused solely on the respective protease domains, the domain boundaries of which were determined by a combination of sequence analysis and biochemical experiments. Sequence analysis identifies a putative PatA protease recognition motif (Ser299-Val300-Glu301-Ala302-Ser303), which precedes the proline-rich sequence presumed to be the linker that connects the protease and DUF domains. We speculated that this sequence likely constitutes the carboxy terminus of the protease domain and generated an expression construct spanning residues Met1 through Ser303 of PatA. Heterologous expression of this domain in *E. coli* yielded soluble and stable protein samples that were catalytically active. Further refinement of the domain boundaries using mass spectrometric analysis of limited trypsin digests identified a stable domain from which the carboxy terminal 20 residues were removed (i.e., Met1-Ala284). Although diffraction quality crystals could only be obtained for the larger construct, continuous electron density could only be observed up to Ala284, validating residues Met1-Ala284 as the PatA protease domain boundary. The domain boundaries for the PagA homolog from the prenylagaramide cluster were similarly established and a construct encompassing residues Met1 through Ser303 was used for further studies. Alignment of the PatG primary sequence with those of PatA and PagA demonstrates that PatG residues encompassing Lys514 through Gly850 are homologous to the PatA and PagA protease domains. A slightly larger construct that extends through to Thr866 at the carboxy terminus was also generated. The protease domains of both PatA and PatG were demonstrated to be catalytically proficient at levels comparable to those of the corresponding wild-type enzymes (McIntosh et al., 2010b, 2011).

Crystal Structure of the PatA and PagA Protease Domains

Crystals of the PatA protease domain (Met1 through Ser303; hereafter PatA) diffract to 1.7 Å resolution (one molecule in the crystallographic asymmetric unit) at an insertion device synchrotron source (Station 21 ID-D; Argonne National Labs, Lemont, IL,

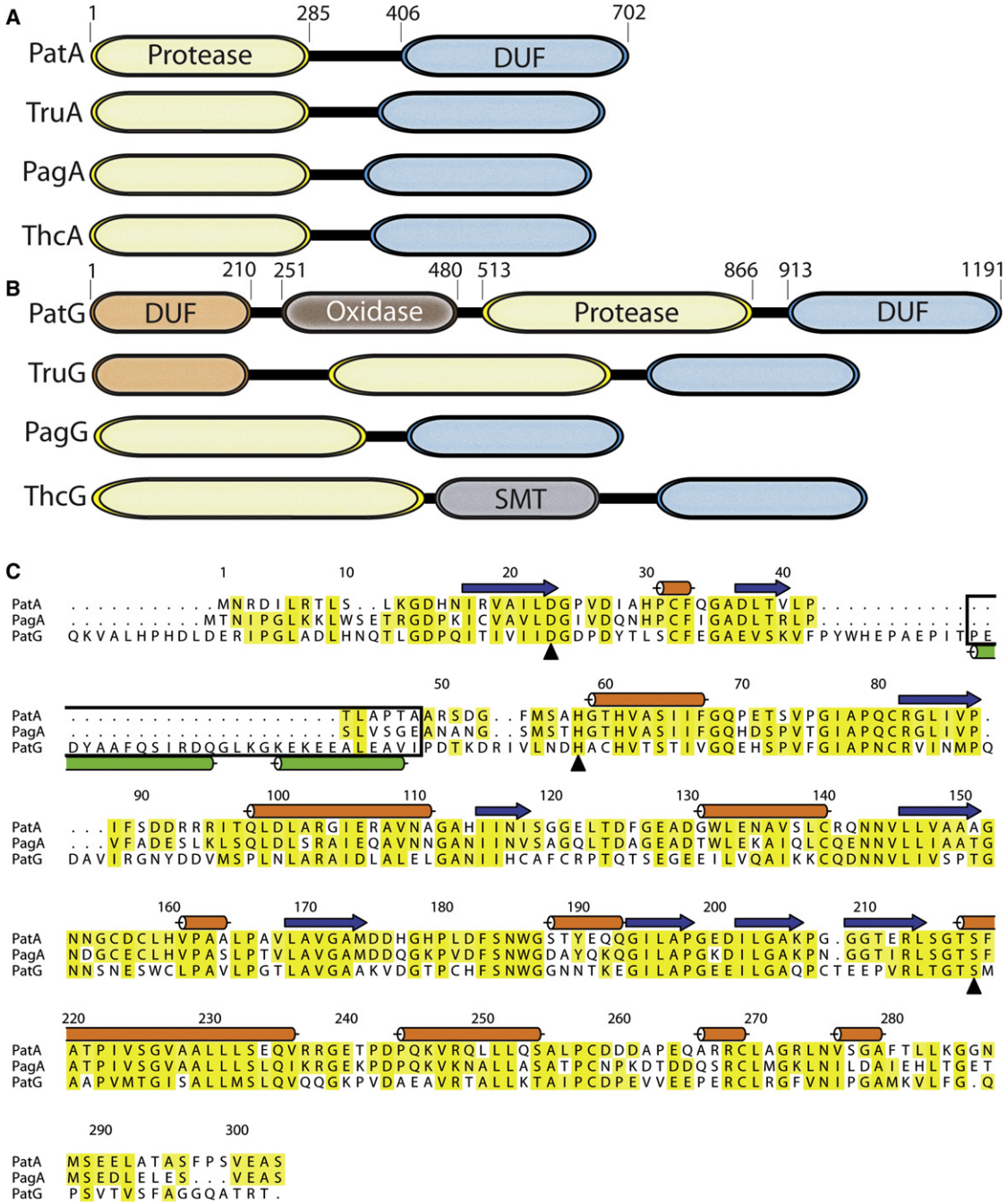

Figure 3. Domain Organization of the Cyanobactin Protease ORFs and Homology between the PatA and PatG Protease Domains

(A) The PatA homologs are organized as an N-terminal protease domain (colored yellow), followed by a linker region, and a C-terminal DUF domain (colored blue). Primary sequence numbering for PatA is shown.

(B) The PatG homologs are organized as a protease domain, followed by the C-terminal DUF domain. Variability is found in the N-terminal organization of PatG homolog sequences, with the PatG sequence harboring an additional DUF domain, followed by a flavin-dependent oxidase domain. In the ThcG sequence, an additional S-adenosyl-methionine-dependent methyltransferase (SMT) domain is present between the protease and DUF domains. Primary sequence numbering for PatG is shown.

(C) Sequence alignment of PatA and PatG identifies homology between the protease and DUF domains. The amino acids corresponding to the capping helices in PatG are boxed. The catalytic triad residues are shown with (▲) under the residues. See Figure S1 for solution oligomerization states of the cyanobactin proteases.

USA) (relevant crystallographic statistics are reported in Table 1). Crystallographic phases were determined by single-wavelength anomalous dispersion methods from data collected on crystals grown from selenomethionine-labeled protein, which grew in a different, but related, crystal form with two molecules in the crystallographic asymmetric unit. The protein behaves as a monomer

Table 1. Data Collection, Phasing, and Refinement Statistics

	PatA	SeMet PatA	PagA	PatG
Data collection				
Space group	$P4_1$	P4	$I2_12_12_1$	C2
a, b, c (Å), β (°)	83.7, 83.7, 42.3	117.4, 117.4, 42.6	49.7, 145.9, 193.5	133.9, 66.9, 105.6
	90	90		112.2
Resolution (Å)[a]	50–1.7 (1.76–1.7)	50–1.9 (1.97–1.9)	50–2.45 (2.54–2.45)	50–2.0 (2.07–2.0)
R_{sym} (%)[b]	8.3 (66.1)	7.7 (38.6)	8.2 (49.1)	8.0 (44.3)
I/σ(I)	43.5 (1.5)	43.2 (5.0)	17.2 (1.7)	22.6 (2.3)
Completeness (%)	97.4 (75.6)	100 (100)	93.2 (67.2)	96.5 (76.9)
Redundancy	12.2 (3.5)	9.3 (8.3)	6.2 (5.3)	7.1 (4.6)
Phasing				
FOM/DM FOM[c]		0.350/0.646		
Refinement				
Resolution (Å)	25.0–1.7	25.0–1.9	25.0–2.45	25.0–2.00
No. reflections	30,147	43,969	23,255	53,640
R_{work} / R_{free}[d]	20.9/24.4	21.3/24.9	18.9/23.1	20.0/24.3
Number of atoms				
Protein	1,992	3,949	4,052	4,541
Water	237	236	156	442
B-factors				
Protein	27.0	28.0	47.8	32.4
Water	40.9	35.5	44.3	42.9
Rmsd				
Bond lengths (Å)	0.009	0.010	0.007	0.009
Bond angles (°)	1.24	1.33	1.27	1.16

[a]Highest resolution shell is shown in parentheses.
[b]$R_{sym} = \Sigma\ |(I_i - < I >|\ \Sigma\ I_i$, where I_i = intensity of the i^{th} reflection and $< I >$ = mean intensity.
[c]Mean figure of merit before and after density modification.
[d]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.

in solution and in the crystal (Figure S1 available online). Crystals of the PagA protease domain (Met1 through Ser303; hereafter PagA) diffract to 2.45 Å resolution, and crystallographic phases were determined by molecular replacement using the refined coordinates of PatA (66% identity) as a search probe.

The overall architectures of PatA and PagA are that of the canonical α/β protease fold first observed in the structure of subtilisin (Siezen and Leunissen, 1997) (Figures 4A and 4C). As the structures of PatA and PagA are essentially identical, further discussion will be based on the higher resolution PatA structure. The overall topology of the PatA protease domain consists of a seven membered central β sheet decorated on either side by α helices. The last strand of this central sheet terminates into two additional β strands, followed by a 22-residue α helix (helix α6; numbered consecutively from the N terminus to the C terminus). Helix α6 is rigidly held in place by numerous hydrophobic interactions with the central β sheet and with the helices around the β sheet.

The catalytic requisite Ser218 resides at the amino terminus of helix α6. The helix dipole moment, in addition to the serine protease catalytic triad, likely aids in the deprotonation of Ser218 to generate the nucleophilic serine alkoxide for catalysis. Two disulfide bonds, between Cys156 and Cys158, and between Cys258 and Cys269, respectively, interconnect the loop emanating from the central sheet β6 and the loop joining the carboxy terminus of helix α7and the amino terminus of helix α8. Both β6 and α7 contact the helix α6 that harbors Ser218. Because of the absence of suitable electron density, a five amino acid flexible loop, comprising residues Ala45 through Ser51 that connects β2 and α1 is not modeled in the final structure.

An alignment of the PatA protease domain structure with various forms of a highly homologous serine protease subtilisin Carlsberg (Neidhart and Petsko, 1988) (rmsd 1.7 Å over 232 aligned residues) provides insights into the positioning of the amino acids proximal to the PatA catalytic triad. The PatA loop bearing residues Val40 through His58 (and containing the disordered residues Ala45 through Ser51) points inward toward the active site, and the positioning of this loop in PatA is identical in all three crystallographically independent copies of the model described above (Figure 4A). By analogy, the equivalent loop in subtilisin Carlsberg apo structure (PDB ID code 1SCA) (Fitzpatrick et al., 1993) and subtilisin Carlsberg structure in complex with the Eglin C inhibitor (PDB ID code 1CSE) (Bode et al., 1987) is directed away from the active site. A number of interactions stabilize the position of the Val40-His58 loop near the PatA catalytic triad, including

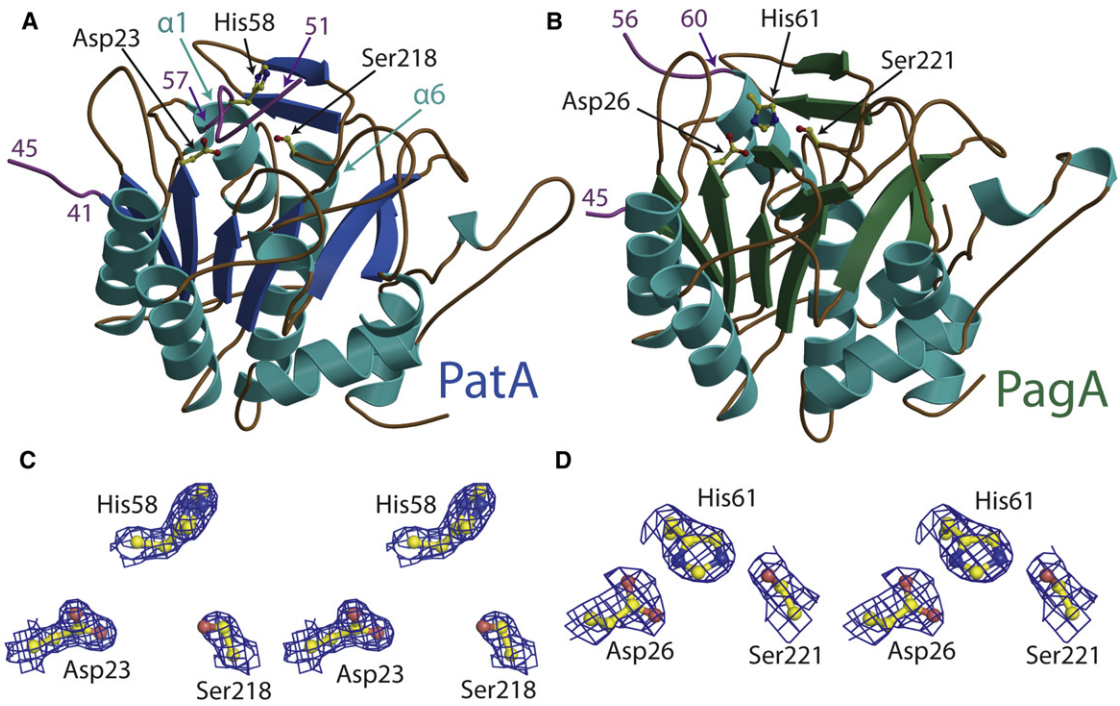

Figure 4. Crystal Structure of the PatA and PagA Protease Domains

(A) The α helices of PatA are colored cyan, with the central beta sheet colored blue. The loop region comprising residues 45–57, which connects the PatA β sheet 2 and α helix 1 is shown in pink with amino acids numbered. Note that the amino acids 46–50 are missing in the final model. The catalytic triad residues are shown in stick-ball representation with the carbon atoms colored yellow.

(B) The α helices of the PagA enzyme are colored cyan, with the central beta sheet colored green. The loop region comprising residues 45–60 is shown in pink with the amino acids numbered. Note that amino acids 46–55 are missing in the final model. The catalytic triad residues are shown in stick-ball representation with the carbon atoms colored yellow. In contrast to PatA residues 51–57, PagA residues 56–60 point away from the PagA catalytic triad.

(C) Zoomed in view of the PatA catalytic triad residue side chains. The residues are labeled. Superimposed is a difference Fourier electron density map (contoured at 2.0σ over background in blue) calculated with coefficients $|F_{obs}| - |F_{calc}|$ and phases from the final refined model with the coordinates of the catalytic triad residue side chain atoms deleted prior to one round of refinement. Note that the positioning of the His58 side chain precludes hydrogen bonding with Asp23 and Ser218 side chains.

(D) Zoomed in view of the PagA catalytic triad residue side chains. The residues are labeled. Superimposed is a difference Fourier electron density map (contoured at 2.0σ over background in blue) calculated as before. Note that the positioning of the His61 side chain allows for hydrogen bonding with Asp26 and Ser221 side chains.

hydrogen bonds between Ser56 and Asp23, between the carbonyl oxygen of Gly53 and main-chain amide of Gly121, and van der Waals contact between Phe54 and Met55.

Active-Site Architecture of the PatA/PagA Protease

By sequence analogy to serine proteases, the catalytic triad in PatA protease domain is comprised of residues Asp23-His58-Ser218 (Figure 4C), and the PagA triad is composed of Asp26-His61-Ser221 (Figure 4D) (analogous to Asp32-His64-Ser221 in subtilisin Carlsberg). Although the PagA triad superimposes nearly perfectly with that of subtilisin, the catalytic triad in the PatA structures is imperfectly positioned. The imidazole side chain of PatA His58, which should be positioned in between and within hydrogen bonding distance of both Asp23 and Ser218, points away from the enzyme active site (Figure 4C). In PatA, Ser56 occupies the site of the catalytic triad histidine, suggesting that the PatA protease domain is not preorganized for efficient catalysis (Figure 4A). In marked contrast, the equivalent loop in PagA is situated away from the active site (Figure 4B), and as a result, His61 in PagA is positioned within hydrogen bonding distance to both Asp26 and Ser221 to form a competently organized catalytic triad (Figure 4D).

The noncanonical position of His58 in PatA is a consequence of the orientation of Val40-His58 loop in the proximity of the active site. In the structures of PatA, this loop spans 19 residues, whereas the equivalent loop is 12 residues in PagA and 13 residues in subtilisin Carlsberg. Consequently, conformational changes must be necessary for the PatA active site to adopt a catalytically proficient state. These conformational changes may be conferred by binding of the substrate, or by the absent PatA C-terminal DUF domain. Nonetheless, the PatA protease domain was fully catalytically active similar to the full-length protein that was previously expressed (Lee et al., 2009). The protease domain construct was completely stable even after extensive enzymatic reactions, whereas the previously reported PatA protein was active but unstable.

Crystal Structure of the PatG Protease Domain

The crystal structure of the PatG protease domain (residues Lys514-Thr866; hereafter PatG) was determined by the

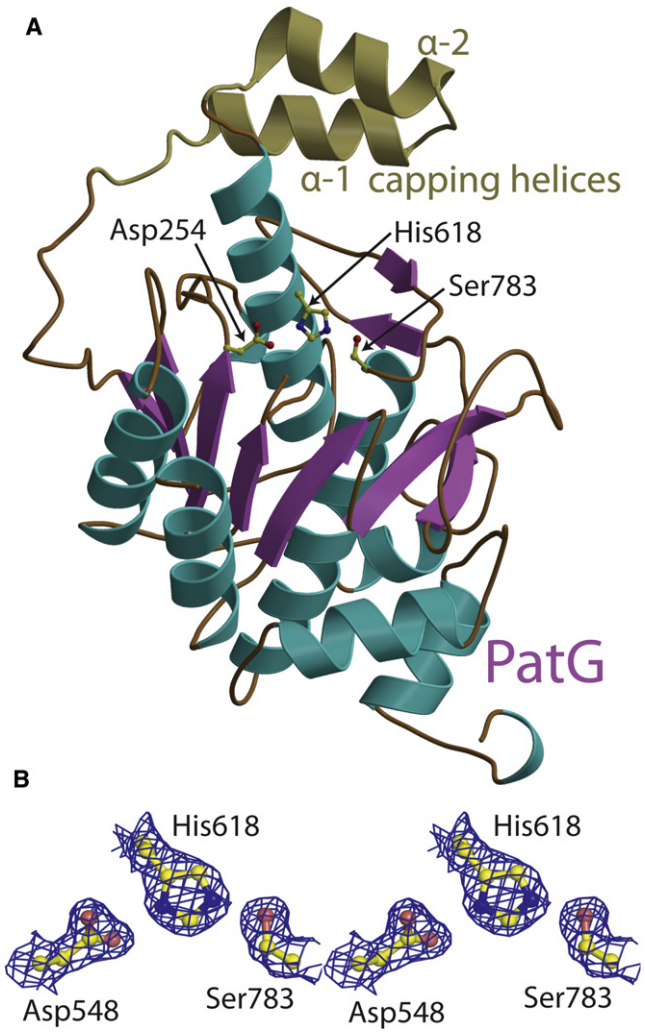

Figure 5. Crystal Structure of the PatG Protease Domain

(A) The α/β core of the enzyme is colored cyan, with the capping helices colored brown. The capping helices are numbered. The catalytic triad residues are shown in stick-ball representation with the carbon atoms colored yellow. (B) Zoomed in view of the PatG catalytic triad residue side chains. The residues are labeled. Superimposed is a difference Fourier electron density map (contoured at 2.0σ over background in blue) calculated as before. See Figure S2 for contribution of capping domain residues Lys594 and Lys598 catalysis.

molecular replacement method. The structures of PatA and PatG are highly homologous (Figure 5A). Almost all of the PatA secondary structural elements can be superimposed directly upon the PatG structure, with even loop residues aligning in near identical fashion. The catalytic nucleophile Ser783 is analogously positioned at the amino terminus of the long α helix, with Asp548 and His618 completing the PatG catalytic triad (Figure 5B).

A superposition of the PatG structure with PatA (and other homologous subtilisin-like proteases) identifies the presence of the two additional α helices that link the fifth and sixth strands of the central β sheet of PatG. These two helices are located directly above the catalytic Ser783 and have hence been termed the "capping helices" (Figure 5A). These helices comprise residues Pro579 to Val605 and are absent in the primary sequences and structures of PatA and other subtilisin proteases. Hence, the presence of the capping helices is unique to PatG and cyanobactin orthologs that presumably catalyze N-C transamidation reactions during the biosynthesis of other cyanobactins. It should be noted that the PatG protease domain construct described here has been experimentally demonstrated to be sufficient to catalyze the macrocyclizing transamidation reaction in cyanobactin biosynthesis (McIntosh et al., 2010b).

The first helix of the capping domain (PatG residues Pro579-Gln591) is rigidly held into place, relative to other secondary structure elements, by a number of hydrogen bonds and hydrophobic interactions. Primary among these is the salt bridge interaction between the side chain of Arg589 and Asp617. Hydrogen bonds also exist between the side chains of Tyr582 and Gln586 and the side chains of Asn616 and Gln770, respectively. The Gln586 side chain amide is also hydrogen bonded to the main-chain carbonyl oxygen atom of Pro771. The side chain of Phe585 makes hydrophobic van der Waals contacts with the side chains of Leu602, Ile613, and Val614. The second helix of the PatG protease-capping domain (residues Lys596-Val605) does not display many interactions with the enzyme core domain as seen for the first helix of the capping domain. A solitary salt bridge exists between the side chains of Glu603 and Lys610. This likely provides for mobility in the positioning of the second helix of the capping domain, which is also reflected in the relatively higher thermal (B) factor values for the residues of this helix.

Probable Mechanism for PatG Catalyzed Transamidation

Formation of amide bonds during the biosynthesis of small molecule natural products is driven by the activation of an amino acid at the carboxy terminus by either phosphorylation (Blasiak and Clardy, 2010; Hollenhorst et al., 2009) or adenylation (Kadi et al., 2007). This activation results in conjugation of the carboxyl with a suitable leaving group, which facilitates nucleophilic condensation with deprotonated primary amine of the incoming amino acid. Another example is the amide bond formation catalyzed by nonribosomal peptide synthetases, in which an amino acid is first adenylated before being covalently tethered via a labile thioester to a phosphopantheinenylated peptidyl carrier protein (PCP). Lastly, covalent tethering of the carboxy terminus to tRNA molecules via an ester linkage has also been demonstrated for amide bond generation in small molecule natural product biosynthesis (Vetting et al., 2010; Zhang et al., 2011). For each of these cases, activation of the carboxy terminus is achieved by hydrolysis of an ATP molecule; hence, such amide-bond-forming reactions are co-factor-dependent.

In contrast, amide bond formation by PatG is not driven by ATP hydrolysis or by tethering to PCP or tRNA molecule. Instead, the carboxy terminus of the cyanobactin cassette is linked by an amide bond to conserved flanking residues and is activated by virtue of formation of an acyl-enzyme intermediate with the catalytic Ser of the macrocyclizing protease.

For all of the cases listed above, there are two mechanistic requirements that need to be satisfied. The first is the spatial positioning of the modified carboxy terminus in close physical proximity to the incoming primary amine within the enzyme

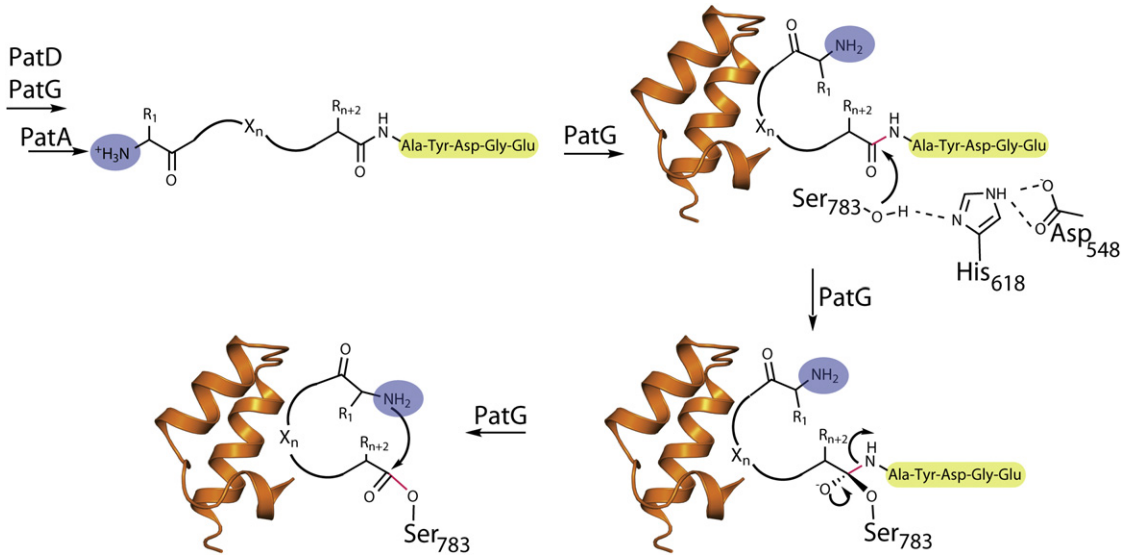

Figure 6. Proposed Mode of Substrate Binding and Reaction Mechanism for the Macrocyclizing Amide Bond Formation by the PatG Protease Domain

The patellamide cassette wraps in the interior of the capping helices, which also serve to deprotonate the amino terminus of the cyanobactin coding cassette generated by the proteolytic action of PatA (highlighted in blue). The PatG catalytic triad catalyzes the first half reaction, causing the scission of the conserved carboxy terminus flanking region (highlighted in yellow) and generating the acyl enzyme intermediate. The capping helices maintain the active site in a desolvated state and position the amino terminus in close proximity to the acyl enzyme intermediate, in order to bias the second half reaction toward aminolysis, rather than hydrolysis. X_n denotes the first (R_1) and last (R_{n+2}) amino acids of the cyanobactin coding cassette.

active site. The second is that the incoming primary amine should be deprotonated by a catalytic base (enzyme Tyr, Lys, or Arg side chain), so as to generate the nucleophile for attack at the activated ester or thioester. The active site can also modulate the pK_a of the primary amine by positioning it in a hydrophobic or a positively charged cavity.

Within the PatG active site, spatial positioning of the carboxy terminus relative to the incoming primary amine may be mediated by the "capping helices" that are unique to this enzyme. The substrate peptide wraps around in the interior of the active site in the pocket formed by the capping helices. Binding would be primarily mediated by the backbone carbonyl oxygen atoms of the substrate peptide with the positively charged interior electrostatic surface of the PatG capping helices. The catalytic Ser783 of PatG would attack the carbonyl carbon of the scissile amide bond to generate a tetrahedral hemiketal intermediate. The intermediate would be resolved by the departure of the C-terminal flanking peptide (Ala-Tyr-Asp-Gly-Glu) of the Pat cassette to yield the covalent acyl-enzyme complex (Figure 6).

Conformational flexibility of the second helix of the capping domain would then allow for the substrate peptide amino terminus to approach the esterified acyl enzyme intermediate. The positively charged interior surface of the PatG capping helices could lower the pK_a of the nucleophile amine, and either of two appropriately positioned Lys side chains—Lys594 or Lys598—could act as a general base. A second nucleophilic attack would lead to the formation of a second hemiketal intermediate, which would be resolved by the departure of the serine side chain to yield an amide bond (Figure 6). In order to facilitate the second nucleophilic addition by the amino terminus of the Pat cassette, the PatG active site would need to be desolvated to protect the ester intermediate from nucleophilic attack by a water molecule to yield a linear Pat cassette as the product. The four heterocycles in the PatE substrate peptide may help to constrict conformational flexibility and offset the entropic cost of binding an otherwise flexible peptide within the constraints of the PatG active site. Notably, the TruG protease (82% sequence identity, 91% sequence similarity with PatG) functions on a peptide substrate that contains only a single heterocycle, and the PatG protease can macrocyclize a variety of peptidic and nonpeptidic substrates (McIntosh et al., 2010b). These data argue that prearrangement of the substrate in a near-cyclical conformation is not a strict requisite for macrocyclization.

The poor catalytic activity of the PatG enzyme (Lee et al., 2009; McIntosh et al., 2010b) precludes determination of the kinetic parameters for the wild-type enzyme and the contribution of Lys594 and Lys598 side chains toward catalysis. The Lys594/Lys598→Ala double mutant was constructed, and its reactivity with two substrates (Lys-Lys-Pro-Tyr-Ile-Leu-Pro-Ala-Tyr-Asp-Gly-Glu and Lys-Gly-Gly-Arg-Gly-Asp-Trp-Pro-Ala-Tyr-Asp-Gly-Glu) is similar to the wild-type enzyme. A closer inspection of the time course of reaction of the first substrate peptide with the PatG 513-866 wild-type enzyme revealed the formation of cyclic product after only 3 hr (Figure S2), whereas no product could be observed with the PatG 513-866 Lys594/Lys598→Ala double mutant enzyme during this time (Figures S2B and S2C). A longer incubation of up to 48 hr converted nearly all of the substrate to a mixture of linear and the cyclic product by both enzymes, which leads us to postulate that mutation of Lys594/Lys598→Ala slows the rate of the reaction but does not alter the overall

catalytic mechanism. This analysis is limited by the nonquantitative nature of MALDI-MS, and a full kinetic analysis needs to be performed to definitively quantitate the contribution of Lys594 and Lys598 residues toward catalysis. However, because of the poor activity of the enzyme, only a semiquantitative approach could be used to characterize site-specific variants.

The recent work of Koehnke et al. (2012) shows two clear effects of mutating Lys594 and Lys598 to Asp. Lys598 was found important for interacting with the Asp residue in the flanking peptide (Ala-Tyr-Asp-Gly). When Lys598Asp PatG was assayed with a substrate with altered recognition sequence Ala-Tyr-Arg-Gly, which would restore the polar interaction between Lys598 and the flanking sequence, there was an order of magnitude improvement in activity over wild-type enzyme with the native Ala-Tyr-Asp-Gly recognition sequence. Lys594Asp reacted very slowly with the altered Ala-Tyr-Arg-Gly substrate. Both lysine residues are also important for cyclization of the product. In wild-type assays, Koehnke and colleagues found that the cyclic product was formed from the substrate Val-Gly-Ala-Gly-Ile-Gly-Phe-Pro-Ala-Tyr-Asp-Gly, whereas mutation of one lysine caused the enzyme to produce a mixture of linear and cyclic products (Lys594Asp) or only linear product (Lys598Asp). They hypothesized that after the first step in the reaction (peptide cleavage), the Ala-Tyr-Asp-Gly product remains in the active site, preventing hydrolysis of the covalently bound intermediate to form the linear product. Our double mutation of both lysines to alanine did not give the same effect on production of cyclic and linear products as the Lys→Asp mutations. In assays with peptide Lys-Lys-Pro-Tyr-Ile-Leu-Pro-Ala-Tyr-Asp-Gly-Glu, we observed similar mixtures of cyclic and linear products from wild-type and mutant enzymes, whereas with substrate Lys-Gly-Gly-Arg-Gly-Asp-Trp-Pro-Ala-Tyr-Asp-Gly-Glu, we observed only cyclic products in assays of both enzymes. In addition to contacts between the lysine residues and aspartate in the recognition sequence Ala-Tyr-Asp-Gly, there are also hydrophobic interactions and hydrogen bonds between tyrosine of the recognition sequence and the enzyme (Koehnke et al., 2012). If the hypothesis that the Ala-Tyr-Asp-Gly product remains in the active site to exclude water and favor production of cyclic products is correct, mutation of Lys594 and Lys598 to Ala may still allow for good binding of the recognition sequence and a small effect on activity, relative to the mutation of these residues to aspartate. A relatively small effect of mutating these lysines on the overall catalytic rate may also be expected in the case that either lysine acts as a general base to deprotonate the N terminus of the peptide, if this step is not rate-limiting.

The slow rate of the PatG reaction makes it difficult to study the kinetic effects of various mutations on the enzyme. We have tried many different buffers without finding conditions that significantly increased the rate. Koehnke et al. (2012) reported a large improvement in product conversion by raising the salt concentration and raising the pH or including DMSO in reactions. When we assayed PatG under these conditions, the relative rate was not greatly improved for one of the best substrates, Lys-Lys-Pro-Tyr-Ile-Leu-Pro-Ala-Tyr-Asp-Gly-Glu, leading us to believe that acceleration may be due at least in part to buffer effects on substrates rather than directly on the enzyme.The poor activity of the wild-type enzyme may be attributed to the absence of additional *cis*-acting domains or *trans*-interaction of the protease domain with other proteins in the patellamide biosynthetic pathway. The possibility that the macrocyclase catalyst has been adapted to produce only modest levels of cyclized cyanobactin natural products under physiological conditions also cannot be discounted because our optimal in vivo production levels require 5 days of fermentation in *E. coli*. The poor activity of the catalyst cannot, however, be attributed to the absence of the four heterocycles, based upon our previous biochemical studies on the substrate promiscuity of PatG protease domain and the fact that highly similar TruG and PatG enzymes employ physiological substrates, which have one or no heterocycles present, respectively. Lastly, prior experiments have characterized PatG and TruG constructs containing the DUF and protease domains, and they were not markedly faster than the protease-only constructs (Lee et al., 2009).

Catalytic Profiles of an Engineered PatG Subtiligase-type Variant

The co-factor-independent peptide bond formation activity of the PatG protease domains may also be considered in the context of prior studies using engineered variants of subtilisin to catalyze peptide ligation (Braisted et al., 1997). Formation of a peptide bond by PatG relies on minimizing hydrolysis (by solvent) and promoting aminolysis (by the substrate peptide α-amine) of the covalent acyl-enzyme intermediate. Hydrolysis of the intermediate yields linear Pat cassettes as products, whereas aminolysis yields macrocyclized cyanobactins. Both linear and macrocyclized products can be observed in in vitro experiments with PatG (Figure 7) (McIntosh et al., 2010b). Previous studies have shown that the equilibrium between hydrolysis and aminolysis can be shifted by changing the chemical nature of the covalent enzyme intermediate in subtilisin proteases. Thioester intermediates, generated by the replacement of active site Ser221 in subtilisin by Cys, are more than 600-fold more efficient for aminolysis over hydrolysis (Nakatsuka et al., 1987). Selenol esters, containing an active-site Se-Cys, were 14,000-fold more efficient for aminolysis (Wu and Hilvert, 1989). The molecular crowding in the active site created by the replacement of the wild-type alcohol Ser side chain by bulkier thiol side chain (in thiolsubtilisin) or selenol side chain (in selenolsubtilisin) could be relieved by an additional Pro225→Ala mutation (Abrahmsén et al., 1991). Thus, Ser221→Cys/Pro225→Ala subtilisin (termed as subtiligase) could be utilized to generate entire functional proteins and enzymes by sequential ligation of peptide fragments (Chang et al., 1994; Jackson et al., 1994) and for the generation of cyclic peptides (Jackson et al., 1995).

In order to explore whether similar substitutions within PatG could offer an increase in the catalytic efficiency for macrocyclization, we generated the PatG protease Ser783→Cys/Pro787→Ala variant (hereafter PatG^{subtiligase}) and carried out in vitro mass spectrometric assays of this variant with different substrates. The crystal structure of PatG establishes that Pro787 is positioned analogously to Pro225 in subtilisin. Both wild-type PatG and PatG^{subtiligase} were first assayed with the substrate Lys-Lys-Pro-Tyr-Ile-Leu-Pro-Ala-Tyr-Asp-Gly-Glu. In order to probe the competing reactions of hydrolysis and aminolysis, the enzymatic reactions of both proteins were challenged using the primary amines of small di-peptides. In the presence

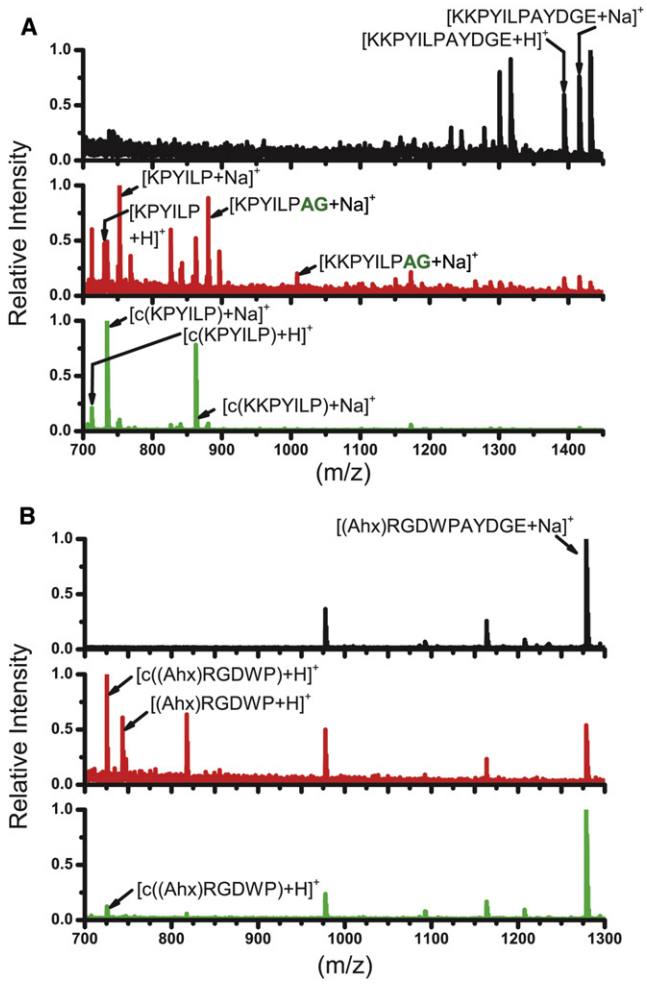

Figure 7. Analysis of Products Generated by the Wild-type PatG, and PatG^{subtiligase} for the KKPYILPAYDGE Peptide in the Presence of Ala-Gly

(A) Analysis of products generated by wild-type PatG and PatG^{subtiligase} for the KKPYILPAYDGE peptide in the presence of Ala-Gly. Top: Black shows the MALDI-MS spectrum for the substrate peptide. Middle: Red shows the reaction of the substrate peptide with wild-type PatG, whereas the bottom (in green) shows the reaction of the substrate peptide with PatG^{subtiligase}. The relevant peaks are labeled. Cyclic peptides are labeled as c. Note that the peaks corresponding to the cyclic products in the bottom are present in the middle. But, the peaks corresponding to linear products or linear Ala-Gly adducts in the middle are absent in the bottom. The reaction has proceeded to completion, as evidenced by the complete disappearance of the substrate peaks.

(B) Analysis of products generated by the wild-type PatG protease domain and PatG^{subtiligase} for the (Ahx)RGDWPAYDGE peptide. The scheme is as before. Note that the peak corresponding to the cyclic product in the bottom is present in the middle, but the peak corresponding to linear product in the middle is absent in the bottom. Also note that the reaction has not proceeded to completion. See Figures S3 and S4 for comparison of PatG and PatG^{subtiligase} reactions in the presence of Gly-Gly di-peptide and formation of cyclo[KKPYILP].

of Ala-Gly, the wild-type enzyme displays a mixture of products, which include linear (Lys)-Lys-Pro-Tyr-Ile-Leu-Pro peptides (note that the first Lys residue of the substrate peptide is hydrolyzed nonenzymatically) (McIntosh et al., 2010b), cyclized peptides, and most interestingly, linear adducts of (Lys)-Lys-Pro-Tyr-Ile-Leu-Pro with the Ala-Gly (Figure 7A). This result is consistent with our hypothesis that the wild-type enzyme can deacylate the ester intermediate by hydrolysis or aminolysis. Further, the aminolysis step can also accept "trans"-primary amines, as provided by the Ala-Gly. This result has been further corroborated by the presence of Gly-Gly adducts with the same substrate peptide under similar assay conditions (Figure S3).

When the PatG^{subtiligase} reaction was challenged with the Ala-Gly dipeptide (Figure 7A) or the Gly-Gly (Figure S3), the predominant products are the cyclized peptides and no Ala-Gly/Gly-Gly adducts can be observed. Our preliminary qualitative data also suggest that PatG^{subtiligase} is a more competent catalyst for cyclization via aminolysis, as compared to the wild-type enzyme, in the absence of competing primary amines (Figure S4). These data are consistent with the idea that the rate-limiting step of the PatG catalytic scheme, as shown in Figure 6, is the deacylation. When the aminolysis of the acyl-enzyme intermediate is hastened in PatG^{subtiligase}, only cyclic peptides are observed as products. The slower aminolytic deacylation of the wild-type enzyme allows for competition from water, as well as small molecule primary amines, for deacylation, which results in the formation of linear products and di-peptide adducts, respectively.

It should be noted that Lys-Lys-Pro-Tyr-Ile-Leu-Pro-Ala-Tyr-Asp-Gly-Glu is one of the best substrates for the PatG protease domain enzyme that we identified in our previous extensive substrate scope studies. We sought to explore whether PatG^{subtiligase} would bias the reaction toward cyclization of the substrate peptide for poorer substrates that are processed very slowly by the enzyme. We assayed the activity of wild-type PatG and the subtligase variant in presence of (Ahx)Arg-Gly-Asp-Trp-Pro-Ala-Tyr-Asp-Gly-Glu (Ahx = aminohexanoic acid). For this substrate, cyclization was demonstrated to occur via the aminohexanoic acid moiety at the N terminus of the peptide. Although the wild-type enzyme displayed a mixture of cyclic and linear products, PatG^{subtiligase} exclusively produced the cyclized product (Figure 7B).

These engineering experiments, analysis of which are compromised by the poor catalytic efficiency of the enzyme, lay the framework for further development of PatG as a general promiscuous, co-factor-independent catalyst for the synthesis of circularized peptides. The range of products that can be generated using wild-type PatG and PatG^{subtiligase} may be of broad interest for pharmaceutical and biotechnological applications.

SIGNIFICANCE

Biosynthesis of microbial genetically encoded small molecule natural products continues to provide numerous examples in which novel enzymatic chemistry can be achieved through subtle changes in the active sites of enzymes with well-established mechanisms. One such example is the co-factor-independent amide bond formation activity of the cyanobactin maturation proteases, which generates a macrocyclic product from a linear substrate. Here, we describe the crystal structures of the subtilisin-like

proteases requisite for cyanobactin maturation. A comparison of the structures reveals the presence of additional structural motifs within the PatG protease that favor transamidation (rather than hydrolysis) of the acyl-substrate complex. We also present biochemical details of an engineered PatG variant with improved macrocyclization activity, while preserving the broad substrate specificity of the wild-type enzyme.

EXPERIMENTAL PROCEDURES

Methods describing the cloning, expression, purification, and crystallization of the various cyanobactin protease domains have been described in detail in the Supplemental Experimental Procedures.

Phasing and Structure Determination

X-ray diffraction data were collected at Life Sciences Collaborative Access Team (LS-CAT), Sector-21, Argonne National Laboratory. Initial crystallographic phases for the PatA protease domain were determined and solved by single-wavelength anomalous diffraction utilizing anomalous scattering from crystals of selenomethionine-labeled PatA protease domain. A 9-fold redundant data set was collected at the selenium absorption edge, to a limiting resolution of 1.9 Å (overall R_{merge} = 0.077, I/σ (I) = 5 in the highest resolution shell) utilizing a Mar 300 CCD detector (LS-CAT, Sector 21 ID-D, Advanced Photon Source, Argonne, IL, USA). Data were indexed and scaled using the HKL2000 package (Otwinowski et al., 2003). A total of 15 selenium sites were identified using HySS (Grosse-Kunstleve and Adams, 2003), and heavy atom parameters were refined using Phaser as implemented in the Phenix suite of programs to yield a figure of merit of 0.418 (acentric/centric = 0.452/0.110). The resultant electron density map was of exceptional quality and permitted most of the main-chain and more than half of the side-chain residues to be automatically built using ARP/wARP. The remainder of the model was fitted using Coot (Emsley and Cowtan, 2004) and further improved through iterative rounds of refinement with REFMAC5 (Murshudov et al., 1997) interspersed with manual rebuilding. Subsequent rounds of model building and crystallographic refinement utilized data from a native crystal of PatA protease domain grown under similar conditions that diffracted to 1.7 Å resolution. Crossvalidation, using 5% of the data for the calculation of the free R factor, was utilized throughout the model-building process in order to monitor building bias (Kleywegt and Brünger, 1996).

X-ray diffraction data from crystals of the PagA and PatG protease domains were collected and processed in a similar manner. The structure of each of these protease domains was determined by the molecular replacement method using the refined coordinates of the PatA protease domain as a search model (McCoy et al., 2007). The resultant solutions were subject to initial rounds of automated model building using ARP/wARP, followed by subsequent rounds of manual rebuilding using XtalView, interspersed with crystallographic refinement using REFMAC5.

ACCESSION NUMBERS

Atomic coordinates have been deposited with the Protein Data Bank (http://www.rcsb.org) with the following accession codes: PatA (4H6V), PagA (4H6W), and PatG (4H6X).

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2012.09.012.

ACKNOWLEDGMENTS

The authors are grateful to Drs. Keith Brister and Joseph Brunzelle and the staff at Life Sciences Collaborative Access Team (LS-CAT, Argonne National Laboratory) for facilitating data collection. We thank Drs. Wilfred A. van der Donk (University of Illinois) and James Wells (University of California at San Francisco) for useful discussions.

Received: August 3, 2012
Revised: September 7, 2012
Accepted: September 18, 2012
Published: November 21, 2012

REFERENCES

Abrahmsén, L., Tom, J., Burnier, J., Butcher, K.A., Kossiakoff, A., and Wells, J.A. (1991). Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. Biochemistry *30*, 4151–4159.

Blasiak, L.C., and Clardy, J. (2010). Discovery of 3-formyl-tyrosine metabolites from Pseudoalteromonas tunicata through heterologous expression. J. Am. Chem. Soc. *132*, 926–927.

Bode, W., Papamokos, E., and Musil, D. (1987). The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech Hirudo medicinalis. Structural analysis, subtilisin structure and interface geometry. Eur. J. Biochem. *166*, 673–692.

Braisted, A.C., Judice, J.K., and Wells, J.A. (1997). Synthesis of proteins by subtiligase. Methods Enzymol. *289*, 298–313.

Chang, T.K., Jackson, D.Y., Burnier, J.P., and Wells, J.A. (1994). Subtiligase: a tool for semisynthesis of proteins. Proc. Natl. Acad. Sci. USA *91*, 12544–12548.

Donia, M.S., and Schmidt, E.W. (2011). Linking chemistry and genetics in the growing cyanobactin natural products family. Chem. Biol. *18*, 508–519.

Donia, M.S., Hathaway, B.J., Sudek, S., Haygood, M.G., Rosovitz, M.J., Ravel, J., and Schmidt, E.W. (2006). Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. Nat. Chem. Biol. *2*, 729–735.

Donia, M.S., Ravel, J., and Schmidt, E.W. (2008). A global assembly line for cyanobactins. Nat. Chem. Biol. *4*, 341–343.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Fischbach, M.A., and Walsh, C.T. (2009). Antibiotics for emerging pathogens. Science *325*, 1089–1093.

Fitzpatrick, P.A., Steinmetz, A.C., Ringe, D., and Klibanov, A.M. (1993). Enzyme crystal structure in a neat organic solvent. Proc. Natl. Acad. Sci. USA *90*, 8653–8657.

Flühe, L., Knappe, T.A., Gattner, M.J., Schäfer, A., Burghaus, O., Linne, U., and Marahiel, M.A. (2012). The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilosin A. Nat. Chem. Biol. *8*, 350–357.

Garg, N., Tang, W., Goto, Y., Nair, S.K., and van der Donk, W.A. (2012). Lantibiotics from Geobacillus thermodenitrificans. Proc. Natl. Acad. Sci. USA *109*, 5241–5246.

Grosse-Kunstleve, R.W., and Adams, P.D. (2003). Substructure search procedures for macromolecular structures. Acta Crystallogr. D Biol. Crystallogr. *59*, 1966–1973.

Hollenhorst, M.A., Clardy, J., and Walsh, C.T. (2009). The ATP-dependent amide ligases DdaG and DdaF assemble the fumaramoyl-dipeptide scaffold of the dapdiamide antibiotics. Biochemistry *48*, 10467–10472.

Jackson, D.Y., Burnier, J., Quan, C., Stanley, M., Tom, J., and Wells, J.A. (1994). A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. Science *266*, 243–247.

Jackson, D.Y., Burnier, J.P., and Wells, J.A. (1995). Enzymic cyclization of linear peptide esters using subtiligase. J. Am. Chem. Soc. *117*, 819–820.

Kadi, N., Oves-Costales, D., Barona-Gomez, F., and Challis, G.L. (2007). A new family of ATP-dependent oligomerization-macrocyclization biocatalysts. Nat. Chem. Biol. *3*, 652–656.

Kleywegt, G.J., and Brünger, A.T. (1996). Checking your imagination: applications of the free R value. Structure *4*, 897–904.

Knerr, P.J., and van der Donk, W.A. (2012). Discovery, biosynthesis, and engineering of lantipeptides. Annu. Rev. Biochem. *81*, 479–505.

Koehnke, J., Bent, A., Houssen, W.E., Zollman, D., Morawitz, F., Shirran, S., Vendome, J., Nneoyiegbe, A.F., Trembleau, L., Botting, C.H., et al. (2012). The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. Nat. Struct. Mol. Biol. *19*, 767–772.

Lee, J., McIntosh, J., Hathaway, B.J., and Schmidt, E.W. (2009). Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. J. Am. Chem. Soc. *131*, 2122–2124.

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. Cryst. *40*, 658–674.

McIntosh, J.A., and Schmidt, E.W. (2010). Marine molecular machines: heterocyclization in cyanobactin biosynthesis. ChemBioChem *11*, 1413–1421.

McIntosh, J.A., Donia, M.S., and Schmidt, E.W. (2010a). Insights into heterocyclization from two highly similar enzymes. J. Am. Chem. Soc. *132*, 4089–4091.

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

McIntosh, J.A., Donia, M.S., Nair, S.K., and Schmidt, E.W. (2011). Enzymatic basis of ribosomal peptide prenylation in cyanobacteria. J. Am. Chem. Soc. *133*, 13698–13705.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Nakatsuka, T., Sasaki, T., and Kaiser, E.T. (1987). Peptide segment synthesis catalyzed by the semisynthetic enzyme thiolsubtilisin. J. Am. Chem. Soc. *109*, 3808–3810.

Neidhart, D.J., and Petsko, G.A. (1988). The refined crystal structure of subtilisin Carlsberg at 2.5 A resolution. Protein Eng. *2*, 271–276.

Newman, D.J., and Cragg, G.M. (2007). Natural products as sources of new drugs over the last 25 years. J. Nat. Prod. *70*, 461–477.

Otwinowski, Z., Borek, D., Majewski, W., and Minor, W. (2003). Multiparametric scaling of diffraction intensities. Acta Crystallogr. A *59*, 228–234.

Rea, M.C., Sit, C.S., Clayton, E., O'Connor, P.M., Whittal, R.M., Zheng, J., Vederas, J.C., Ross, R.P., and Hill, C. (2010). Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. Proc. Natl. Acad. Sci. USA *107*, 9352–9357.

Schmidt, E.W., and Donia, M.S. (2009). Chapter 23. Cyanobactin ribosomally synthesized peptides—a case of deep metagenome mining. Methods Enzymol. *458*, 575–596.

Schmidt, E.W., Nelson, J.T., Rasko, D.A., Sudek, S., Eisen, J.A., Haygood, M.G., and Ravel, J. (2005). Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. Proc. Natl. Acad. Sci. USA *102*, 7315–7320.

Siezen, R.J., and Leunissen, J.A. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci. *6*, 501–523.

Sivonen, K., Leikoski, N., Fewer, D.P., and Jokela, J. (2010). Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. Appl. Microbiol. Biotechnol. *86*, 1213–1225.

Tianero, M.D., Donia, M.S., Young, T.S., Schultz, P.G., and Schmidt, E.W. (2012). Ribosomal route to small-molecule diversity. J. Am. Chem. Soc. *134*, 418–425.

Vetting, M.W., Hegde, S.S., and Blanchard, J.S. (2010). The structure and mechanism of the Mycobacterium tuberculosis cyclodityrosine synthetase. Nat. Chem. Biol. *6*, 797–799.

Walsh, C.T., Acker, M.G., and Bowers, A.A. (2010). Thiazolyl peptide antibiotic biosynthesis: a cascade of post-translational modifications on ribosomal nascent proteins. J. Biol. Chem. *285*, 27525–27531.

Wu, Z.P., and Hilvert, D. (1989). Conversion of a protease into an acyl transferase: selenolsubtilisin. J. Am. Chem. Soc. *111*, 4513–4514.

Zhang, W., Ntai, I., Kelleher, N.L., and Walsh, C.T. (2011). tRNA-dependent peptide bond formation by the transferase PacB in biosynthesis of the pacidamycin group of pentapeptidyl nucleoside antibiotics. Proc. Natl. Acad. Sci. USA *108*, 12249–12253.

