

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221687450>

# Selectivity, Directionality, and Promiscuity in Peptide Processing from a *Bacillus* sp Al Hakam Cyclodehydratase

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · MARCH 2012

Impact Factor: 12.11 · DOI: 10.1021/ja211675n · Source: PubMed

---

CITATIONS

26

---

READS

17

## 4 AUTHORS, INCLUDING:



Nhat Q Trinh

Loyola University Chicago

4 PUBLICATIONS 31 CITATIONS

SEE PROFILE



Douglas A Mitchell

University of Illinois, Urbana-Champaign

50 PUBLICATIONS 1,584 CITATIONS

SEE PROFILE

Published in final edited form as:

*J Am Chem Soc.* 2012 March 21; 134(11): 5309–5316. doi:10.1021/ja211675n.

## Selectivity, Directionality, and Promiscuity in Peptide Processing from a *Bacillus* sp. Al Hakam Cyclodehydratase

Joel O. Melby<sup>1,2</sup>, Kyle L. Dunbar<sup>1,2</sup>, Nhat Q. Trinh<sup>3</sup>, and Douglas A. Mitchell<sup>1,2,4,\*</sup>

<sup>1</sup>Department of Chemistry, University of Illinois at Urbana-Champaign; Urbana, Illinois, 61801; USA

<sup>2</sup>Institute for Genomic Biology, University of Illinois at Urbana-Champaign; Urbana, Illinois, 61801; USA

<sup>3</sup>School of Molecular and Cellular Biology, University of Illinois at Urbana-Champaign; Urbana, Illinois, 61801; USA

<sup>4</sup>Department of Microbiology; University of Illinois at Urbana-Champaign; Urbana, Illinois, 61801; USA

### Abstract

The thiazole/oxazole-modified microcins (TOMMs) represent a burgeoning class of ribosomal natural products decorated with thiazoles and (methyl)oxazoles originating from cysteines, serines and threonines. The ribosomal nature of TOMMs allows for the generation of derivative products from mutations in the amino acid sequence of the precursor peptide, which ultimately manifest in differing structures and sometimes, biological functions. Employing a TOMM system for the purpose of creating new structures and functions via combinatorial biosynthesis requires processing machinery that can tolerate highly variable substrates. In this study, TOMM enzymatic promiscuity was assessed using a currently uncharacterized cluster in *Bacillus* sp. Al Hakam. As determined by Fourier transform tandem mass spectrometry (FT-MS/MS), azole rings were formed in both a regio- and chemoselective fashion. Cognate and non-cognate precursor peptides were modified in an overall C- to N-terminal directionality, which to date is unique among characterized ribosomal natural products. Studies focused on the inherent promiscuity of the biosynthetic machinery elucidated a modest bias for glycine at the preceding (−1) position and a remarkable flexibility in the following (+1) position, even allowing for the incorporation of charged amino acids and bisheterocyclization. Two unnatural substrates were utilized as the conclusive test of substrate flexibility, of which both were processed in a predictable fashion. A greater understanding of substrate processing and enzymatic tolerance towards unnatural substrates will prove beneficial when designing combinatorial libraries to screen for artificial TOMMs that exhibit desired activities.

### Introduction

The thiazole/oxazole-modified microcins (TOMMs) comprise a recently described class of posttranslationally modified peptide natural products whose thiazole and (methyl)oxazole heterocycles derive from cysteine, serine and threonine residues.<sup>1</sup> Characterized members of this natural product family exhibit a myriad of functions including, but not limited to

\*Corresponding author; (douglasm@illinois.edu), phone: 1-217-333-1345, fax: 1-217-333-0508).

**Associated Content** Molecular biology methods, protein purification methods, supporting tables and supporting figures can be found in the Supporting Information. This material is available free of charge via the Internet at <http://pubs.acs.org>

All authors have given approval to the final version of the manuscript and declare no competing financial interest.

antibacterial compounds, antitumor agents, and cytolytic virulence factors.<sup>2-4</sup> Despite the initial discovery of the requisite heterocycle forming enzymes approximately 15 years ago, biochemical characterization of the TOMM enzymatic machinery has been limited to the microcin B17, thiazole/oxazole-containing cyanobactin, and streptolysin S (SLS) synthetases.<sup>3,5,6</sup> Previous efforts to elucidate the complex underpinnings of substrate processing have been stymied due to poor protein stability, solubility, and difficulties in monitoring heterocycle formation.<sup>7,8</sup> The discovery of a novel TOMM cluster in *Bacillus* sp. Al Hakam (Balh) with more ideal physical-chemical properties made it possible to further explore the factors governing substrate processing with potential implications towards artificial TOMM engineering via a combinatorial biosynthetic approach.

Although TOMMs can display a wide range of posttranslational modifications, their defining features are thiazole and (methyl)oxazole heterocycles. These “azole” rings are installed over two enzymatic steps. First, a complex of the TOMM cyclodehydratase (C-protein) and the docking protein (D-protein) catalyzes the cyclodehydration of cysteine, serine, and threonine residues to azoline heterocycles (Figure 1).<sup>3,5,7,9</sup> The collaborative enzymatic effort of the C- and D-proteins is genetically illustrated in roughly half of all identified TOMMs, where they are produced as a single polypeptide.<sup>1</sup> The azoline rings in many cases undergo a subsequent 2-electron oxidation to the azole heterocycle by the action of a flavin mononucleotide (FMN)-dependent dehydrogenase (B-protein).<sup>7</sup> Many TOMM clusters also contain ancillary tailoring enzymes, which add to the structural complexity of this class of natural products.<sup>1,10</sup>

As with the lantipeptides and other ribosomal peptide natural products, the N-terminal portion of the TOMM precursor peptide, referred to as the leader peptide (Figure 1), has been implicated in substrate recognition and appears to direct the precursor peptide to the biosynthetic complex.<sup>11,12</sup> The C-terminal portion of the substrate, denoted as the core peptide, is extensively modified and converted into a bioactive natural product. Importantly, the primary amino acid sequence of the core peptide directly encodes the structure/function of the mature TOMM. Often the biological function of a previously uncharacterized TOMM can be accurately predicted based on similarity to characterized family members.<sup>1,8</sup>

Early studies of microcin B17 identified and reconstituted the enzymes responsible for the installation of thiazole and oxazoles.<sup>5,13</sup> Later investigations formed the foundation for much of the current knowledge on the selectivity of thiazole/oxazole installation. A combination of cysteine alkylation and MS/MS analysis indicated an overall N- to C-terminal directionality of ring formation on the microcin B17 precursor peptide. However, the exact order of ring formation could never be established using MS/MS, as the thiazole and oxazole heterocycles suppressed fragmentation and prevented the precise localization of some intermediates.<sup>14</sup> Further investigations concluded that the heterocycle-forming enzymes are both regio- and chemoselective.<sup>14-17</sup> Walsh and coworkers also determined that mutating flanking residues around the first bisheterocycle site decreased the rate of heterocycle formation, and combined with its poor biophysical properties render the microcin B17 system a poor candidate for combinatorial engineering.<sup>7,16</sup>

The initial studies of microcin B17 processing have provided insights into substrate processing in additional TOMM clusters; two noteworthy examples are the thiazole/oxazole-containing cyanobactins and SLS. The cyanobactins are macrocyclic ribosomal peptides produced by cyanobacteria, while SLS is a well-known virulence factor from *Streptococcus pyogenes*.<sup>3,18</sup> Studies conducted by Schmidt and coworkers demonstrated that heterocycle formation in the cyanobactins appeared to be regulated mainly by regioselective factors and that a high degree of flexibility existed regarding the cyclized residue and flanking residues (*i.e.* polar, hydrophobic, and charged).<sup>9</sup> In contrast to the relatively well-characterized

cyanobactins, the structure of SLS remains elusive, and as such the finer details of cyclization have not been determined. However, studies using non-cognate SLS substrates established the flexible nature of the biosynthetic machinery.<sup>19</sup> Although these works have been instrumental in the expansion of our knowledge regarding TOMM biosynthesis, no previously characterized enzyme complex had the requisite *in vitro* characteristics for cell-free production of highly variable TOMMs.

In search of a TOMM biosynthetic gene cluster ideally suited for detailed biochemical studies of substrate selectivity and promiscuity, genome mining revealed two biosynthetic clusters in the Balh genome, whose products remain structurally and functionally uncharacterized.<sup>20</sup> The first, cluster 1, which is the focus of the current study, contained all the required biosynthetic genes in neighboring open reading frames and harbored two precursor peptides (BalhA1 and BalhA2, Figure 1). The natural BalhA1/A2 substrates address a variety of shortcomings that plagued in-depth biochemical characterization of previously studied TOMMs. We predicted that the large number of polar amino acids in the precursor peptides would aid solubility and visualization of heterocycles by mass spectrometry. Moreover, the inclusion of tyrosine and tryptophan residues allow for facile UV-detection. Bioinformatic approaches have identified four closely related gene clusters (gene similarity, directionality, and order within the cluster) to Balh cluster 1, all within the *Bacillus cereus* group (Figure S1). The second TOMM cluster in Balh (cluster 2) is found in virtually all *Bacillus cereus* group members (Figure S2). Because of the attractiveness of the cluster 1 precursor peptides towards chromatography, mass spectrometry, and the relative ease of obtaining recombinant protein in good yield and purity, we employed this TOMM system to further probe substrate processing in the natural product family. This work highlights the robustness of the *Bacillus* sp. Al Hakam TOMM enzymes and lays the foundation for larger-scale natural product engineering efforts.

## Results and Discussion

### *In vitro* Reconstitution of the Enzymatic Activity of the Balh TOMM Enzymes

Both substrates and the requisite enzymes in the Balh TOMM cluster 1 were cloned, overexpressed in *E. coli*, and purified as TEV protease cleavable fusions to maltose binding protein (MBP). Although each protein was isolated in good yield and purity (Figure S3), the dehydrogenase (BalhB) purified without the necessary FMN cofactor bound, and varied attempts to increase FMN loading were met with limited success. Similar challenges have been solved in other TOMM systems through the use of a surrogate dehydrogenase from a similar TOMM cluster.<sup>8</sup> Therefore, a highly similar dehydrogenase (78% identity/94% similarity) from *B. cereus* 172560W (BcerB, Figure S1), which purified with the necessary FMN cofactor bound after heterologous expression in *E. coli*, was used for *in vitro* assays with BalhC and BalhD. A synthetase reaction, containing the BalhA1 substrate, BalhC, BalhD, and the non-cognate (surrogate) dehydrogenase, BcerB, was digested with trypsin and analyzed using LC-FTMS. The resulting spectra clearly showed the presence of five azole heterocycles installed on the BalhA1 precursor peptide (Figure 2; -100 Da). To determine if the two heterocyclizable residues at the extreme C-terminus of BalhA1 (not present in the tryptic fragment) were modified, an analogous experiment replacing trypsin with TEV protease was performed. Here, the full-length precursor peptide was analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS and found to contain five azole heterocycles, which indicated the extreme C-terminus of BalhA1 was unmodified under the conditions employed (Figure S4). Omission of the dehydrogenase resulted in a mass loss consistent with the formation of five azoline heterocycles, indicating that cyclodehydration occurred in the absence of the dehydrogenase (Dunbar *et al.*, unpublished).

Further time course studies of azole formation demonstrated the accumulation of multiple intermediates (Figure S5). As such, the Balh thiazole/oxazole synthetase processed substrate in a distributive fashion, reminiscent of trunkamide and microcin B17 processing.<sup>14,21</sup> It is worth noting that the distributive nature of TruD (trunkamide cyclodehydratase) was only seen when an inadequate amount of ATP was supplied.<sup>21</sup> Through LC-FT-MS/MS studies, the five azole heterocycles installed upon BalhA1 were localized to five cysteines (C28, 31, 34, 40 and 45; Figure 2), as determined by the loss of mass and fragmentation pattern surrounding the modified amino acids.<sup>14,15</sup> The discriminating activity towards particular cysteines posed several questions about residue tolerance and prompted a more thorough investigation into the chemo- and regioselectivity of the Balh thiazole/oxazole synthetase, as well as the effects of altering the local environment.

### Chemoselectivity of the Cyclodehydration Reaction

The observation of solely thiazole rings after enzymatic treatment led to the hypothesis that the enzymes were exceptionally chemoselective for cysteine, which is in contrast to many other characterized TOMMs.<sup>22-26</sup> Given the distributive nature of substrate handling, multiple intermediates accumulated at non-equivalent rates, rendering a detailed kinetic analysis exceedingly complicated. To circumvent this complexity, a gene encoding a non-cyclizable (NC) version of BalhA1 was synthesized in which the cysteines/serines and threonines were changed to alanines and valines, respectively (Table S1). Because TOMM heterocyclization has been previously shown to be directly coupled to ATP hydrolysis, and in the case of TruD occurred in a stoichiometric fashion, substrate processing rates were monitored using an assay specific for phosphate detection.<sup>21,27,28</sup>

To determine if the BcerB/BalhCD enzymes were indeed capable of forming (methyl)oxazoles, point mutants of the NC substrate analog containing cysteine, serine, and threonine were generated at three of the five heterocyclized sites (31, 40 and 45). As predicted from inherent nucleophilicity, cysteine was processed most rapidly at all three positions (Table 1, Table S2).<sup>17</sup> Although reactions with the native BalhA1 substrate suggested that serines/threonines were not able to be cyclized, placement of both serine and threonine at position 40 also resulted in ATP hydrolysis (Table 1), and by inference, heterocycle formation. Given that the microcin B17 biosynthetic enzymes were shown to process serine about 100- to 1000-fold slower than cysteine,<sup>17</sup> it is notable that the Balh TOMM biosynthetic enzymes only have a 30-fold slower processing rate for serine compared to cysteine at position 40. As a control, Michaelis-Menten kinetic parameters were obtained for wild type substrate and the amenable mutants (Table 1). The decreased processing efficiency could not be attributed to a decreased apparent  $K_m$ , and indicated the rates measured were true reflections of processing efficiencies. Preliminary efforts to directly confirm that ATP hydrolysis resulted in (methyl)oxazoline formation were not fruitful because BcerB proved unable to oxidize the azoline. Such heterocycles hydrolytically ring open, yielding the original amino acid, under standard LC-based purifications at lower pH values. An alternative approach was developed that involved TEV protease cleavage of the peptide from MBP and organic solvent-based precipitation of all large proteins (MBP and the biosynthetic enzymes). Under the conditions employed, only smaller polypeptides remain soluble, thus allowing for relatively clean MS samples without the need for LC. Our MS methods allowed for visualization and localization of the dehydration to position 40 (NC-A40T and NC-A40S, Figures S6 and S7 respectively). These experiments demonstrated that the enzymatic machinery exhibited moderate chemoselectivity for cysteine; however, our data did not reconcile the lack of (methyl)oxazoline rings in the native substrate, as serine and threonine were accepted as substrates and are found in the core peptide of BalhA1. Thus, we hypothesized that the

synthetase was regioselective for the specific positions where cyclization was confirmed in the native substrate.

### Regioselectivity of the Cyclodehydration Reaction

To probe the regioselectivity of azole formation, a series of alanine to cysteine point mutations were introduced on the non-cyclizable (NC) substrate at the naturally occurring cyclized positions in BalhA1. Based on literature precedent for the N- to C-terminal directionality for ribosomally synthesized peptidic natural products, the most N-terminal mutant (NC-A28C) was expected to have the fastest rate of cyclodehydration.<sup>14,29-31</sup> Surprisingly, NC-A40C was the fastest by a factor of 18 (Table 2). All other single cysteine substrates (NC-A28C, -A31C, -A34C, -A45C) had significantly slower rates of ATP hydrolysis. From these data, we surmised that the rate of ring formation at positions other than 40 might be enhanced upon conformational restriction at position 40. For example, synthetase reactions containing BalhA1 (50  $\mu$ M), BcerB (5  $\mu$ M), BalhC (1  $\mu$ M), and BalhD (1  $\mu$ M) reach completion within 6 hours. Without taking into account C40 or the rate decrease as a function of substrate concentration, the reaction would take over 16 hours if there were no rate enhancement after heterocycle installation. Such synergy between heterocycle sites has been previously reported in microcin B17 processing.<sup>32</sup> We hypothesized that installing a proline, which has previously been shown to functionally substitute for thiazoles and oxazoles, at position 40 would mimic the first thiazole produced and enhance the processing rate of downstream residues.<sup>19</sup> This rate enhancement was most notable with NC-A31C/A40P, which was nearly 6-fold faster than NC-A31C and also suggested that C31 may be the second thiazole (Tables 3 and S2). Even though conformational restriction at position 40 enhanced ring formation elsewhere, it is important from a combinatorial engineering perspective that cyclization at position 40 was not strictly required for processing at secondary sites.

### Order of Ring Formation

The NC substrate analog data suggested C40 and C31 would be the first and second azole heterocycles, respectively, and warranted a more complete investigation into the order of ring formation. The isolation of a one ring species and subsequent MS/MS analysis allowed for the localization of the first thiazole (Figure S8 and Scheme 1). The 20 Da mass loss in the MS/MS spectra of single ring species was only detected in the b-ion series starting at  $b_{26}^{++}$  and the y-ion series starting at  $y_{14}^{+}$ . Moreover, fragmentation was not observed at the amide bonds directly adjacent to C40, which is a well-known behavior for this type of chemical species.<sup>14,15</sup> Together, these observations were consistent with the kinetic results obtained with the aforementioned NC substrates and unequivocally localized the first heterocycle to C40 (Scheme 1). For the two ring species, ions that indicated C31 and C34 as the second heterocycle appeared with equal intensity ( $b_{16}^{++}$  and  $y_{20}^{+}$ ), suggesting a divergent biosynthetic pathway. However, reconvergence was immediate, as fragmentation of the three ring species produced fragment ions indicative of thiazoles only at C31, C34 and C40. The site of the fourth ring and fifth ring were C45 and C28, respectively. These data demonstrated that the Balh thiazole/oxazole synthetase modified BalhA1 in a unique C- to N-terminal overall direction, although not strictly.

To determine if the directionality of thiazole synthesis was also observed with other closely related biosynthetic clusters or an artifact of using the non-cognate dehydrogenase, the cyclodehydratase (BcerC) and docking protein (BcerD) from *B. cereus* 172560W were also cloned and purified. Combined with BcerB, this represented a cognate biosynthetic system for thiazole/oxazole synthesis and could be used to validate the experiments conducted with the Balh system. BcerBCD processed BalhA1 in an identical fashion as that observed with



BcerB, BalhC and BalhD, with the minor exception of cyclizing C31 exclusively as the second thiazole (Figure S9).

### Importance of Flanking Residues on the Cyclodehydration Reaction

An analysis of the local sequence around the cyclized residues in BalhA1, as well as predicted cyclized residues in the other closely related homologs, revealed three generally conserved trends (Figure S10). *i.* Glycine nearly always preceded the putatively cyclized cysteine (−1 position). *ii.* Conversely, glycine rarely occurred directly following the putatively cyclized cysteine (+1), which was largely populated by hydrophobic amino acids. *iii.* A practically invariant glycine was found in the +2 position. As before, we utilized the NC substrate analog as a platform to ascertain the level of tolerance that existed at each of these positions. The necessity of the conserved glycine in the −1 position was first assessed. A preceding glycine could conceivably enhance the rate of cyclization because of the decreased steric bulk at C<sub>α</sub>, allowing for greater Ramachandran space to be sampled.<sup>5</sup> In congruence with this hypothesis, the replacement of glycine with alanine, aspartic acid, or lysine in the −1 position decreased the rate of ATP consumption (Tables 3 and S2). This result was reminiscent of studies performed on microcin B17 and SLS processing. The microcin B17 synthetase did not tolerate substitutions in the −1 position, in relation to the first bisheterocycle site.<sup>16</sup> Moreover, only two azole heterocycles have been detected and localized by mass spectrometry on SagA (precursor peptide of SLS), both of which harbor glycine in the −1 positions.<sup>8</sup> While the presence of a preceding glycine appears to accelerate heterocycle formation, it is by no means a stringent requirement for heterocyclization as thiazole/oxazole-containing-cyanobactins, thiopeptides, and other TOMMs often lack glycines preceding the azole/azoline heterocycles.<sup>22–26</sup> In these systems, we speculate that the presence of a −1 glycine may abolish or diminish biological activity, forcing the evolution of TOMM enzymes to accept residues with restricted Ramachandran space.

Modulating the steric bulk in the +1 position had a lesser effect on the rate of ATP consumption, and this position even tolerated charged residues (Tables 3 and S2). The only amino acid tested that was not tolerated in the +1 position was proline (Table S2). From a combinatorial biosynthesis perspective, flexibility towards charged and hydrophobic residues could be very beneficial when tuning bioavailability/water solubility and membrane permeability. The microcin B17 biosynthetic enzymes did not display such flexibility in the +1 position.<sup>16</sup> The +2 position (glycine) also tolerated substitution (Tables 3 and S2). It is important to note that the +2 glycine may be an artifact of the precursor peptide sequence. In many cases, this glycine precedes another cyclized cysteine or heterocyclizable residue, which may evade *in vitro* processing but be a *bona fide* site of heterocyclization *in vivo*. However, we cannot rule out the possibility that the conserved +2 glycine could also be crucial for biological activity, which would also explain its conserved nature and minimal effect on heterocyclization rate.

### Heterocyclization of BalhA2 and BcerA

As previously mentioned, the Balh TOMM cluster has two precursor peptides (Figure 1); therefore, we sought to characterize the processing of the second peptide, BalhA2. Compared to BalhA1, the BalhA2 substrate had a similar apparent K<sub>m</sub> (17 ± 2 μM) but a reduced *k*<sub>obs</sub> (4.4 ± 0.2 min<sup>−1</sup>). An alignment of these two precursor peptides showed that C28, C31, T34 and C37 of BalhA2 aligned with cysteines that are cyclized in BalhA1 (Figure S11). Many of these positions do not contain the favored −1 glycine, likely resulting in the reduced *k*<sub>obs</sub> for BalhA2. A synthetase reaction of BalhA2 with BcerB, BalhC, and BalhD resulted in three azole heterocycles and 1 azoline heterocycle (Figure S12). Upon trypsin digestion and LC-FTMS analysis, three azole heterocycles remained and were localized to C28, C31, and C37 (Figures S13 and S14). The lone azoline ring eluded exact

localization; however, the MS/MS fragmentation narrowed down the azoline location to T34, T35, or S36 (Figure S12). The presence of a -1 glycine at T34 and its alignment with C34 of BalhA1 left T34 as our predicted site for the azoline heterocycle. The order of azole formation on BalhA2 was determined to be strictly C- to N-directional using BcerB/BalhC/BalhD and BcerBCD (Figures S15 and S16). The substrate from the *B. cereus* 172560W cluster (BcerA) was processed in a highly similar and thus predictable fashion (Figures S17-S19). Taken together with the BalhA1 results, these data represent, to the best of our knowledge, a heretofore-unreported processing direction for a ribosomally synthesized natural product.<sup>14,29-31</sup>

### Effect of Substrate Truncations/Insertions on the Cyclodehydration Reaction

For the purposes of using combinatorial biosynthesis to generate conformationally restrained peptides (TOMMs) with desirable properties, it would be useful for the processing enzymes to also tolerate substrates of varied lengths, increasing the amount of peptide space one could survey. To probe the Balh TOMM biosynthetic ability to accept substrates of different lengths, a series of truncation and insertion substrates were generated. The truncation substrates were prepared by introducing a stop codon into the sequence of *balhA1* via site-directed mutagenesis (*i.e.* G47\* contains a stop codon at position 47 instead of coding for the naturally occurring glycine). The insertion substrates contain an additional “GAG” or “GAGGAG” sequence after the double aspartate motif, which we predict is the end of the leader sequence (Figure 1C, Table S1). All of the substrates were accepted by the cyclodehydratase with varying efficiencies (Table 4). Truncation and insertion substrates that contained C40 (equivalent to C43 or C46 for the insertion substrates) were processed efficiently, and the order of ring formation was not altered greatly or at all for BalhA1-G47\* and BalhA1-C54\* (Figures S20 and S21). The processing rate of the truncated substrates lacking C40 was very similar to the rate of the single cysteine-containing NC substrates (excluding NC-A40C), again highlighting the regioselective, yet promiscuous, nature of the Balh TOMM enzymes.

### Evaluation of Unnatural Chimeric Substrates

To further explore the substrate tolerance of BcerB, BalhC and BalhD, the genes for two chimeric substrates were synthesized. These substrates harbor the BalhA1 leader peptide and an unnatural core peptide. BalhA1-Scramble contained residues 26-46 of BalhA1 in randomized order while BalhA1-McbA harbored a truncated core peptide of the microcin B17 precursor peptide (Figure 1). Based on the substrate processing rules elucidated earlier in this study, BalhA1-Scramble was predicted to have four thiazoles and one methyloxazoline installed after reaction with BcerB/BalhCD. We predicted that all the cysteines, with the exception of the two most C-terminal, would be converted to thiazoles. We also hypothesized that the most C-terminal threonine would be cyclized to a methyloxazoline heterocycle, as BcerB had already been shown to not oxidize (methyl)oxazolines *in vitro* (Figure 1). The cysteine not expected to be cyclized in the scrambled region (C42) had a -1 cysteine, which we predicted upon cyclization would prevent the formation of the bisheterocycle. Synthetase reactions followed by MALDI-MS analysis resulted in five heterocycles on BalhA1-Scramble, as predicted (Figure 3). Moreover, LC-MS/MS allowed the localization of the four azoles to the most C-terminal cysteines of the tryptic fragment and demonstrated that the rings were installed in an overall C- to N-terminal directional fashion (Figure S22). Unexpectedly, the bisheterocycle was indeed formed. Although this modification eluded prediction, the heterocycle at C42 was the last azole heterocycle to be installed. The apparently slow rate of heterocyclization is most likely due to the presence of a thiazole in the -1 position. Such bisheterocycle sites are critical for the activity of bleomycin and microcin B17.<sup>33,34</sup> Thus, combinatorial TOMM engineering efforts may find application for the production of artificial DNA intercalators



and gyrase inhibitors. As with the BalhA2 substrate, it was not possible to localize the azoline ring; however, we presume its location to be T46. It is noteworthy that the one cysteine that remained unmodified had a -1 threonine and +1 glycine, both of which would decrease the rate of heterocyclization. These effects would be expected to be additive, thus providing an explanation for its inability to be cyclized.

The second unnatural substrate tested was the BalhA1-McbA precursor peptide, which was expected to contain four thiazoles after reaction with BcerB/BalhCD (Figure 1). However, only three azole heterocycles were found (Figure 3). Again, LC-MS/MS was utilized to localize the heterocycles to the most C-terminal cysteines, which were installed in an overall C- to N-terminal directional fashion (Figure S23). Similar to the unmodified cysteine in BalhA1-Scramble, the remaining cysteine had a -1 serine and a +1 glycine. As best illustrated with these unnatural chimeric substrates, the Balh TOMM biosynthetic enzymes have shown remarkable tolerance to unnatural core peptides.

## Conclusions

Using a combination of MS and enzyme kinetics, the promiscuity of the thiazole/oxazole forming enzymes from *Bacillus* sp. Al Hakam has been evaluated. *In vitro* reconstitution of the cyclodehydratase and dehydrogenase resulted in five thiazoles on BalhA1. The design and deployment of a simplified substrate allowed a rapid and quantitative assessment of substrate chemo- and regioselectivity. Proper positioning of serine and threonine residues on an unnatural substrate showed that the biosynthetic enzymes were capable of catalyzing the formation of (methyl)oxazolines. The order of ring formation for multiple substrates proceeded with an overall C- to N-terminal directionality, which to the best of our knowledge has not been observed previously in ribosomal peptide natural product biosynthesis. Characterization of a series of substrate mutants with varying flanking residues, including charged amino acids and truncations/insertions again illustrated the substrate flexibility and potential utility in combinatorial engineering. Processing of two artificial core peptides confirmed the promiscuity of these biosynthetic enzymes, which were processed in a highly predictable manner. In this work, we have identified a robust system capable of installing cysteine- and serine/threonine-derived heterocycles onto variable precursor peptides, which was not limited to *in vivo* processing or reliant on other tailoring enzymes. Our findings exemplify the potential of the Balh TOMM synthetase as a platform for employing combinatorial biosynthesis to generate unnatural thiazole and (methyl)oxazoline containing products with desired activities.

## Methods

### Materials and General Methods

Unless otherwise specified, all chemicals were purchased from Fisher Scientific or Sigma-Aldrich. Oligonucleotides were purchased from Integrated DNA Technologies. Restriction enzymes were purchased from New England BioLabs (NEB). dNTPs were purchased from Promega. *Pfu*Turbo DNA polymerase was purchased from Stratagene (now Agilent). DNA sequencing was performed by the Keck Biotechnology Center (University of Illinois at Urbana-Champaign), Eton Biosciences, or ACGT, Inc. Methods regarding bioinformatics, basic molecular biology, protein overexpression and protein purification can be found in the Supporting Information.

### Synthetase Reaction

In general, a precursor peptide (100  $\mu$ M), dehydrogenase (10  $\mu$ M), cyclodehydratase (10  $\mu$ M), and docking protein (10  $\mu$ M) were added together with synthetase buffer [50 mM Tris

pH 7.4, 125 mM NaCl, 20 mM MgCl<sub>2</sub>, 2 mM adenosine triphosphate (ATP), and 10 mM dithiothreitol (DTT)]. When optimizing *in vitro* reaction conditions, supplementing the reaction with 50  $\mu$ M FMN was found to produce a 2-fold rate increase in azole formation. Reaction products were analyzed using two different methods. TEV cleavage was done concurrently (1:50 of  $\mu$ g of TEV protease:  $\mu$ g of total protein) or after the reaction (1:10 of TEV protease: protein at 30 °C for 30 minutes). The samples were desalted by ZipTip (Millipore) according to the manufacturer's specifications. Alternatively, the samples were trypsin digested using Sequencing Grade Modified Trypsin (Promega) and analyzed by LCMS. Trypsin was resuspended according to the manufacturer's specifications. Synthetase reactions (10  $\mu$ L) were used for a 30  $\mu$ L trypsin digestion reaction. TCEP was added to a concentration of 2 mM, and trypsin was used in a 1:40 trypsin:total protein ratio. Trypsin digestions proceeded for at least 30 min at 30 °C. The samples were quenched using formic acid prior to LCMS analysis, with the final concentration being 6.25% (v/v).

### Purine Nucleoside Phosphorylase Coupled Assay

Quantitative kinetic measurements were conducted on a Cary 4000 UV/VIS spectrophotometer (Agilent). The substrate, synthetase buffer (50 mM Tris pH 7.4, 125 mM NaCl, 20 mM MgCl<sub>2</sub>, 2 mM ATP, and 10 mM DTT), 0.2 units purine nucleoside phosphorylase, and 20 nmoles of 7-methyl-6-thioguanosine (Berry and Assoc.) were allowed to equilibrate to room temperature. Immediately before beginning the assay, C- and D-proteins were added to an empty cuvette. For all the NC substrate analogs and other reported substrates, the initial rate was measured using 2  $\mu$ M MBP-BalhC and MBP-BalhD with 100  $\mu$ M substrate. For determination of the kinetic parameters, a range of substrate concentrations was used with 1-5  $\mu$ M of MBP-BalhC and MBP-BalhD. The remainder of the reaction mixture was used to initiate the reaction. Reaction progress was monitored at 360 nm. The concentration of phosphate was calculated based on the molar absorptivity of the product (11,000 M<sup>-1</sup> cm<sup>-1</sup>). Initial rates were determined by plotting the data in Microsoft Excel. Error was calculated as the standard deviation of the mean ( $n \geq 3$ ). Nonlinear Michaelis-Menten regressions were conducted using Igor Pro version 6.1 and the error reported reflects the standard deviation from the curve fitting.

### HPLC MS/MS

All reverse-phase liquid chromatography-Fourier-transform mass spectrometry (FTMS) was carried out on an Agilent 1200 HPLC system with an autosampler connected directly to a ThermoFisher Scientific LTQ-FT hybrid linear ion trap, operating at 11 T. The mass spectrometer was calibrated weekly following the manufacturer's protocol.

Trypsin digested samples were separated using a 1  $\times$  150 mm Jupiter C<sub>18</sub> column (300 Å, 5  $\mu$ M, Phenomenex). Samples were subjected to one of two methods. The first was a full scan followed by data-dependent MS/MS of the four most abundant peaks. Using this method, the MS/MS ions were detected in the FTMS with the following parameters: minimum target signal counts: 5,000; resolution: 50,000;  $m/z$  range detected: dependent on target  $m/z$ , default charge state: 2, isolation width: 5  $m/z$ , normalized collision energy (NCE): 35; activation  $q$  value: 0.40; activation time: 30 ms. The second method was used to determine the ring order, the MS/MS ion detection was carried out in the ion-trap mass spectrometer (ITMS). This method also utilized targeted MS/MS determined for various ring states and substrates. The following parameters were used: isolation width: 3  $m/z$ , normalized collision energy (NCE): 35; activation  $q$  value: 0.25; activation time: 30 ms. Data analysis was conducted using the Qualbrowser application of Xcalibur (Thermo-Fisher Scientific).

## Direct Infusion Analysis by ESI-MS or ESI-MS/MS

Acetonitrile (ACN, 6.75  $\mu$ L) was added to TEV cleaved synthetase reactions (25  $\mu$ L). Precipitate was allowed to form for 5 minutes, which was then removed by centrifugation at 16,060  $\times g$  (Sorvall Fresco Microcentrifuge). The supernatant (20  $\mu$ L) was diluted with ddH<sub>2</sub>O (20  $\mu$ L). The diluted sample was then desalted and concentrated by ZipTip (Millipore) according to the manufacturer's specifications with the omission of acid from all solutions. The samples were eluted into 20  $\mu$ L of 50% ACN (v/v).

Low-resolution analysis was conducted by injecting 15  $\mu$ L onto an Agilent 1200 series HPLC operated at 0.5 ml/min at 50% ACN w/ 0.1% formic acid with no column to a single quadrupole mass analyzer (G1956B). Analysis was conducted in positive ion scan mode, 13 L/min drying gas rate, 30 psig nebulizer pressure, 350  $^{\circ}$ C drying gas temperature and a 4000 V capillary voltage. Basic analysis was done using ChemStation, and the data were plotted using OriginPro 8.5.

High-resolution analysis was conducted using a ThermoFisher Scientific LTQ-FT hybrid linear ion trap, operating at 11 T. Samples were directly infused using an Advion Nanomate 100. MS/MS was conducted using identical settings as described above.

## MALDI-TOF Analysis

TEV cleaved peptides were analyzed using MALDI-TOF analysis on a Bruker Daltonics UltrafleXtreme MALDI TOF/TOF. All samples were desalted using a ZipTip according to the manufacturer's specification and eluted directly onto the target into 4  $\mu$ L of  $\alpha$ -cyano-4-hydroxycinnamic acid. The samples were then allowed to dry under ambient conditions. All analyses were conducted using positive reflector mode. The instrument was calibrated before each use using a peptide calibration kit (AB SCIEX). Data were analyzed using FlexAnalysis.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

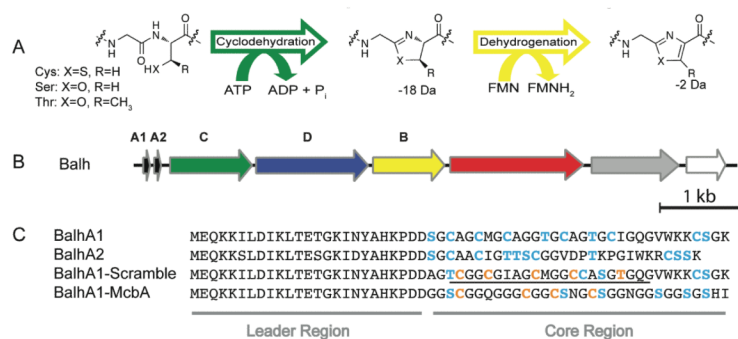
We are grateful for the gift of *B. Al Hakam* and *B. cereus* 172560W from Paul Jackson (Lawrence Livermore National Lab) and Shanmuga Sozhamannan (Naval Medical Research Center), respectively. We thank Brad Evans and Stefanie Bumpus for assistance collecting FTMS data and Joyce Limm for technical assistance. Members of the Mitchell Lab carried out critical review of this manuscript. This work was supported in part by the institutional funds provided by the University of Illinois and the NIH Director's New Innovator Award Program (DP2 OD008463). JOM was supported by a departmental fellowship (Chinoree T. Kimiyo Enta Fellowship). KLD was supported by the NIH Training Program in Chemistry – Interface with Biology (2T32 GM070421-06).

## Abbreviations

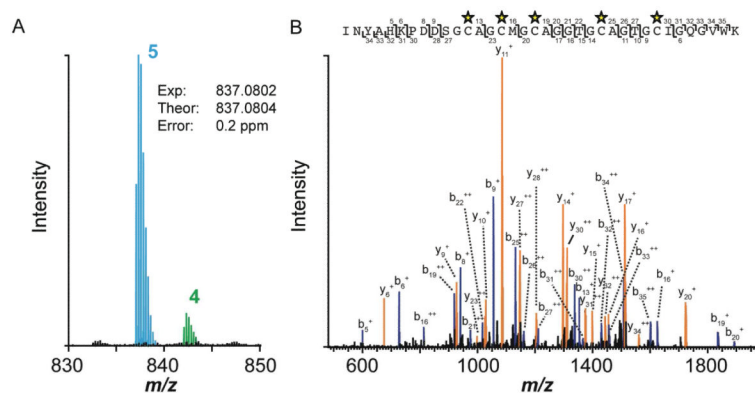
<b>ATP</b>	adenosine triphosphate
<b>Balh</b>	<i>Bacillus</i> sp. Al Hakam
<b>FMN</b>	flavin mononucleotide
<b>TEV</b>	tobacco etch virus
<b>TOMM</b>	thiazole/oxazole-modified microcin
<b>Tz</b>	thiazole

## References

- (1). Melby JO, Nard NJ, Mitchell DA. *Curr Opin Chem Biol.* 2011; 15:369. [PubMed: 21429787]
- (2). Salvatella X, Caba JM, Albericio F, Giralte E. *J Org Chem.* 2003; 68:211. [PubMed: 12530842]
- (3). Lee SW, Mitchell DA, Markley AL, Hensler ME, Gonzalez D, Wohlrab A, Dorrestein PC, Nizet V, Dixon JE. *Proc Natl Acad Sci U S A.* 2008; 105:5879. [PubMed: 18375757]
- (4). Vizan JL, Hernandez-Chico C, del Castillo I, Moreno F. *EMBO J.* 1991; 10:467. [PubMed: 1846808]
- (5). Li YM, Milne JC, Madison LL, Kolter R, Walsh CT. *Science.* 1996; 274:1188. [PubMed: 8895467]
- (6). Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J. *Proc Natl Acad Sci U S A.* 2005; 102:7315. [PubMed: 15883371]
- (7). Milne JC, Roy RS, Eliot AC, Kelleher NL, Wokhlu A, Nickels B, Walsh CT. *Biochemistry.* 1999; 38:4768. [PubMed: 10200165]
- (8). Gonzalez DJ, Lee SW, Hensler ME, Markley AL, Daresh S, Mitchell DA, Bandeira N, Nizet V, Dixon JE, Dorrestein PC. *J Biol Chem.* 2010; 285:28220. [PubMed: 20581111]
- (9). McIntosh JA, Donia MS, Schmidt EW. *J Am Chem Soc.* 2010; 132:4089. [PubMed: 20210311]
- (10). McIntosh JA, Donia MS, Nair SK, Schmidt EW. *J Am Chem Soc.* 2011; 133:13698. [PubMed: 21766822]
- (11). Oman TJ, van der Donk WA. *Nat Chem Biol.* 2010; 6:9. [PubMed: 20016494]
- (12). Madison LL, Vivas EI, Li YM, Walsh CT, Kolter R. *Mol Microbiol.* 1997; 23:161. [PubMed: 9004229]
- (13). Genilloud O, Moreno F, Kolter R. *J Bacteriol.* 1989; 171:1126. [PubMed: 2644225]
- (14). Kelleher NL, Hendrickson CL, Walsh CT. *Biochemistry.* 1999; 38:15623. [PubMed: 10569947]
- (15). Walsh CT, Kelleher NL, Belshaw PJ. *J Am Chem Soc.* 1998; 120:9716.
- (16). Sinha Roy R, Belshaw PJ, Walsh CT. *Biochemistry.* 1998; 37:4125. [PubMed: 9521734]
- (17). Belshaw PJ, Roy RS, Kelleher NL, Walsh CT. *Chem Biol.* 1998; 5:373. [PubMed: 9662507]
- (18). Donia MS, Ravel J, Schmidt EW. *Nat Chem Biol.* 2008; 4:341. [PubMed: 18425112]
- (19). Mitchell DA, Lee SW, Pence MA, Markley AL, Limm JD, Nizet V, Dixon JE. *J Biol Chem.* 2009; 284:13004. [PubMed: 19286651]
- (20). Challacombe JF, et al. *J Bacteriol.* 2007; 189:3680. [PubMed: 17337577]
- (21). McIntosh JA, Schmidt EW. *Chembiochem.* 2010; 11:1413. [PubMed: 20540059]
- (22). Sivonen K, Leikoski N, Fewer DP, Jokela J. *Appl Microbiol Biotechnol.* 2010; 86:1213. [PubMed: 20195859]
- (23). Onaka H, Nakaho M, Hayashi K, Igarashi Y, Furumai T. *Microbiology.* 2005; 151:3923. [PubMed: 16339937]
- (24). Molohon KJ, Melby JO, Lee J, Evans BS, Dunbar KL, Bumpus SB, Kelleher NL, Mitchell DA. *ACS Chem Biol.* 2011
- (25). Kalyon B, Helaly SE, Scholz R, Nachtigall J, Vater J, Borriss R, Sussmuth RD. *Org Lett.* 2011; 13:2996. [PubMed: 21568297]
- (26). Bagley MC, Dale JW, Merritt EA, Xiong X. *Chem Rev.* 2005; 105:685. [PubMed: 15700961]
- (27). Milne JC, Eliot AC, Kelleher NL, Walsh CT. *Biochemistry.* 1998; 37:13250. [PubMed: 9748332]
- (28). Webb MR. *Proc Natl Acad Sci U S A.* 1992; 89:4884. [PubMed: 1534409]
- (29). Lee MV, Ihnken LA, You YO, McClerren AL, van der Donk WA, Kelleher NL. *J Am Chem Soc.* 2009; 131:12258. [PubMed: 19663480]
- (30). Lubelski J, Khusainov R, Kuipers OP. *J Biol Chem.* 2009; 284:25962. [PubMed: 19620240]
- (31). Philmus B, Guerrette JP, Hemscheidt TK. *ACS Chem Biol.* 2009; 4:429. [PubMed: 19445532]
- (32). Roy RS, Allen O, Walsh CT. *Chem Biol.* 1999; 6:789. [PubMed: 10574779]
- (33). Stubbe J, Kozarich JW, Wu W, Vanderwall DE. *Accounts Chem Res.* 1996; 29:322.
- (34). Zamble DB, Miller DA, Heddle JG, Maxwell A, Walsh CT, Hollfelder F. *Proc Natl Acad Sci U S A.* 2001; 98:7712. [PubMed: 11427730]

**Figure 1.**

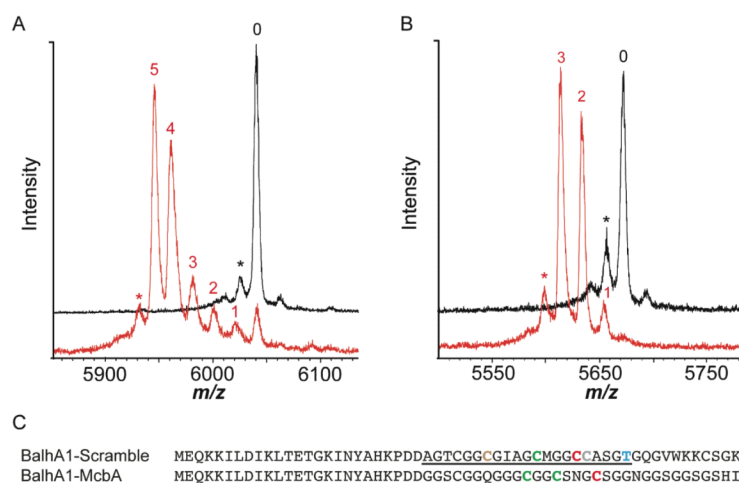
Thiazole/oxazole formation and genetic organization of the TOMM from *Bacillus* sp. Al Hakam. **A.** ATP-dependent cyclodehydration of a peptidic cysteine or serine/ threonine results in the formation of a thiazoline or (methyl)oxazoline ring. These azoline heterocycles can be oxidized to the aromatic azole heterocycles through an FMN-dependent dehydrogenation. **B.** The first TOMM cluster in *B. sp. Al Hakam* consists of two precursor peptides (A1 and A2, black), a cyclodehydratase (C, green), docking protein (D, blue), dehydrogenase (B, yellow), an ABC transporter (red), and two hypothetical proteins (grey and white). **C.** Sequences of precursor peptides used in this study. The blue/orange letters signify heterocyclizable residues. The scrambled portion of BalhA1-Scramble is underlined. The orange residues of the chimera substrates were predicted to be cyclized by the Balh thiazole/oxazole synthetase.



**Figure 2.**

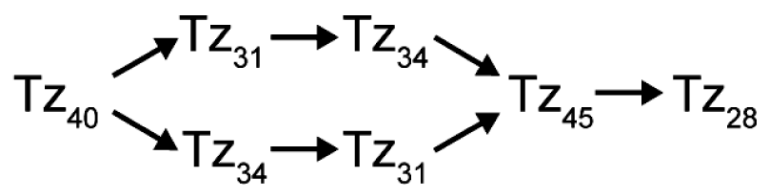
*In vitro* azole synthesis on BalhA1. **A.** A tryptic digestion of a synthetase reaction composed of BalhA1 treated with BcerB/BalhCD produced tryptic fragments with four and five azole heterocycles. The 4<sup>+</sup> charge state is shown and the number next to the peak corresponds to the number of heterocycles. Experimental and theoretical monoisotopic masses are given for the five azole ring species. **B.** MS/MS of the five azole tryptic fragment of BalhA1 shown in **A.** The  $b^+$  and  $y^+$  ions are labeled blue and orange, respectively. The observed fragmentation was annotated based on the residue number within the tryptic fragment. The five azole heterocycles were localized to C28, C31, C34, C40 and C45 (stars).





**Figure 3.**

MALDI-TOF MS of BalhA1-Scramble and BalhA1-McbA reactions. **A.** A synthetase reaction containing MBP-BalhA1-Scramble treated with MBP-tagged BcerB/BalhCD was allowed to react for 16 h and concurrently cleaved with TEV protease (red). The black trace is the substrate only control. The +1 charge state is shown. The numbers over the peaks indicate successive losses of -20 (azole heterocycles), with the exception of ring 5 (-18, azoline heterocycle). The asterisk denotes a laser-induced artifact. **B.** Same as panel A, except BalhA1-McbA was used as the substrate. **C.** The amino acid sequence of the chimera substrates. All colored cysteines contained heterocycles derived from BcerB/BalhCD, and the color indicates the order of formation (Red- 1<sup>st</sup>, Green-2<sup>nd</sup>, Brown-3<sup>rd</sup>, and Grey- 4<sup>th</sup>). The predicted location of the azoline heterocycle is shown in blue.



Scheme 1.

**Table 1**

Chemoselectivity of the Balh cyclodehydratase.

Substrate	$V_o$ ( $\mu\text{M P}_i \text{ min}^{-1}$ )	% WT <sup>a</sup>	$K_m$ , $\mu\text{M}^b$	$k_{\text{obs}}$ $\text{min}^{-1}$	$k_{\text{obs}}/K_m$ , $\text{M}^{-1} \text{ s}^{-1}$
WT	$20.2 \pm 0.7$	100	$16 \pm 2$	$12.9 \pm 0.4$	$(13 \pm 2) \times 10^3$
NC-A40C	$18 \pm 1$	88	$15.3 \pm 0.9$	$10.8 \pm 0.2$	$(11.8 \pm 0.7) \times 10^3$
NC-A40S	$0.63 \pm 0.09$	3.1	n.d. <sup>c</sup>	n.d. <sup>c</sup>	n.d. <sup>c</sup>
NC-A40T	$3.2 \pm 0.1$	16	$16 \pm 2$	$1.38 \pm 0.06$	$(1.4 \pm 0.2) \times 10^3$

<sup>a</sup> % WT = (Substrate  $V_o$  / WT  $V_o$ )  $\times$  100.<sup>b</sup> Apparent  $K_m$ <sup>c</sup> Not determined.

**Table 2**

Regioselectivity of the Balh cyclodehydratase.

Substrate	$V_o$ ( $\mu\text{M P}_i \text{ min}^{-1}$ )	% WT <sup>a</sup>
NC-A28C	$1.01 \pm 0.1$	4.7
NC-A31C	$0.81 \pm 0.09$	4.0
NC-A34C	$0.5 \pm 0.2$	2
NC-A40C	$18 \pm 1$	88
NC-A45C	$0.80 \pm 0.09$	4.0

<sup>a</sup> % WT = (Substrate  $V_o$  / WT  $V_o$ )  $\times$  100.

**Table 3**

Effect of flanking residues on the rate of cyclodehydration.

Substrate	$V_o$ ( $\mu\text{M P}_i \text{ min}^{-1}$ )	% WT <sup>a</sup>
NC-G39A/A40C	$0.67 \pm 0.05$	3.3
NC-G39D/A40C	$0.1 \pm 0.1$	0.3
NC-A40C/A41G	$6.6 \pm 0.1$	33
NC-A40C/A41D	$17 \pm 3$	82
NC-A40C/G42I	$13.1 \pm 0.3$	64.8
NC-A31C/A40P	$4.7 \pm 0.2$	23.3

<sup>a</sup> % WT = (Substrate  $V_o$  / WT  $V_o$ )  $\times$  100.

**Table 4**

Processing Efficiency of BalhA1 Truncations/Insertions.

Substrate <sup>a</sup>	V <sub>o</sub> (μM P <sub>i</sub> min <sup>-1</sup> )	% WT <sup>b</sup>
BalhA1 G30*	0.91 ± 0.02	4.5
BalhA1 G36*	1.0 ± 0.1	4.9
BalhA1 G42*	14.3 ± 0.4	71.0
BalhA1 G47*	20.4 ± 0.8	101
BalhA1 C54*	21.2 ± 0.3	105
BalhA1 GAG ins	17.4 ± 0.9	86.1
BalhA1 GAGGAG ins	19 ± 2	94

<sup>a</sup>The truncations substrates have a stop codon (\*) at the denoted location. Insertion substrates have the denoted amino acid sequence after the predicted end of the leader peptide (PDD).

<sup>b</sup>% WT = (Substrate V<sub>O</sub> / WT V<sub>O</sub>) × 100.