Discovery and Characterization of an Isopeptidase That Linearizes Lasso Peptides

Mikhail O. Maksimov[†] and A. James Link*[†,‡]

[†]Departments of Chemical and Biological Engineering and [‡]Molecular Biology, Princeton University, Princeton, New Jersey 08544, United States

S *Supporting Information*

ABSTRACT: Lasso peptides are a class of ribosomally derived natural products with diverse bioactivities. The characteristic threaded lasso structure in these peptides derives from an isopeptide bond attaching the N-terminus of the peptide to an acidic side chain. Here we describe the heterologous expression of a lasso peptide gene cluster encoding two lasso peptides, astexin-2 and astexin-3, and solve the solution structure of astexin-3. This cluster also encodes an enzyme annotated as a protease. We show that this enzyme, AtxE2, is a lasso peptide isopeptidase that specifically hydrolyzes astexins-2 and -3, converting them to linear peptides. Astexin-3 is highly thermostable and resists unthreading after extensive heat treatment. In contrast, astexin-2 unthreads upon heat treatment. AtxE2 has no activity toward unthreaded astexin-2, demonstrating that this isopeptidase must recognize a knotted structure in order to function. We also use this isopeptidase as a tool to study evolutionary relationships between lasso peptide gene clusters.
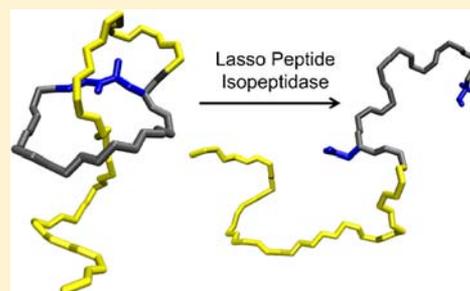
INTRODUCTION

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a rapidly growing superfamily of natural products that originate from gene encoded polypeptides.[1] The rapid growth in the discovery of RiPPs has been facilitated by genome mining approaches that reveal RiPP gene clusters within sequenced organisms.[2] Lasso peptides are a class of RiPPs that consist of ca. 20 amino acid-long chains that are folded into a structure resembling a threaded lasso.[3] An isopeptide bond installed between the N-terminus of the peptide and a Glu or Asp side chain holds the topologically constrained structure together. Our group[4] and others[5–7] have used a combination of genome mining and heterologous expression to discover several new lasso peptides. To date these genome mining studies have focused solely on four genes in lasso peptide gene clusters: the A gene encoding the lasso peptide precursor, the B and C genes encoding maturation enzymes, and the D gene which encodes an ABC transporter.[8–10] Our global lasso peptide genome mining data[4] hinted that there may be other genes associated with lasso peptide regulation or modification beyond these four canonical genes.

This line of inquiry led us to examine the immediate genomic neighborhood of two lasso peptide gene clusters found in the freshwater bacterium *Asticcacaulis excentricus*. We have previously described the heterologous expression and structure determination of astexin-1, the product of the gene cluster found on chromosome 1 of *A. excentricus*. Here we report the heterologous expression of astexin-2 and astexin-3, the two lasso peptides encoded on chromosome 2 of *A. excentricus* (Figure 1). The structure of astexin-3 was determined by NMR.

Both of these clusters lack the ABC transporter found in the gene clusters of the lasso peptides microcin J25, capistruin, and lariatin.[5,9,11] Instead, these clusters include divergently transcribed genes (AtxE1 and AtxE2) annotated as proteases immediately downstream of the genes encoding the lasso peptide maturation enzymes (Figure 1). Here we have demonstrated that the putative protease AtxE2 specifically cleaves the isopeptide bond of lasso peptides encoded in the neighboring cluster. The cleavage reaction proceeds only on threaded lasso peptides. An evolutionary bioinformatics analysis revealed that many known and putative lasso peptide gene clusters segregate into two distinct clades delineated by the presence of either an ABC transporter or a lasso peptide isopeptidase.

RESULTS AND DISCUSSION

Cloning and Heterologous Expression of Astexins-2 and -3. Lasso peptides from Gram-negative species can typically be expressed in greater yield using *Escherichia coli* as a heterologous host than the native organism in which the combined effects of regulatory elements and weak natural promoters can make isolation of these natural products impractical.[4–6,12,13] We made three constructs for the expression of astexins-2 and -3 (Figure 2). In the plasmid pMM37, the natural astexins-2 and -3 cluster is placed under the control of the strong *tet* promoter.[14] Both precursors include a ribosome binding sequence upstream of their respective genes. Inverted repeat (hairpin) sequences are
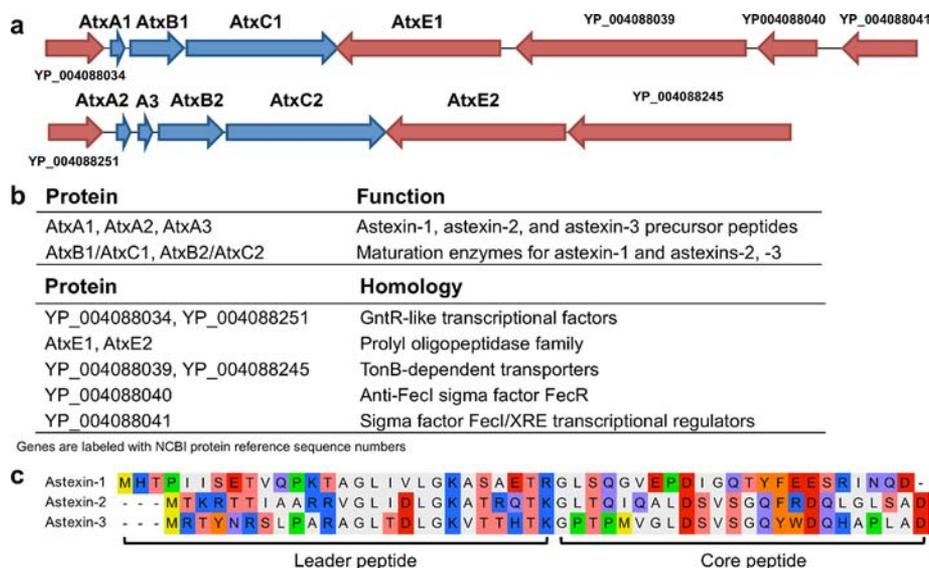
Figure 1. Lasso peptides in *A. excentricus*. (a) The architecture of the astexins-1, -2, and -3 lasso peptide gene clusters. Genes required for the biosynthesis are shown in blue, while other conserved genes are highlighted in red. (b) Homology of lasso peptide cluster associated genes to known protein families. (c) Precursor alignment of astexins-1, -2, and -3. Identity and similarity between pairs of proteins in the astexin gene clusters are as follows: B1/B2 (23.1%, 33.7%), C1/C2 (29.6%, 43.2%), E1/E2 (20.7%, 33.4%).
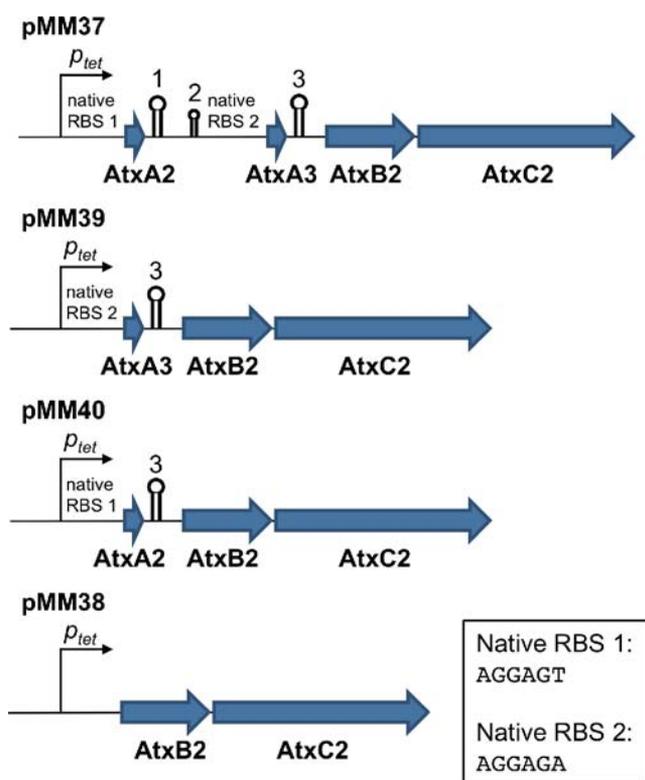

Figure 2. Engineered gene clusters for the production of astexins-2 and -3. DNA inverted repeats are shown as stem-loop structures.

located downstream of each precursor (hairpins 1 and 3) with an additional smaller hairpin (hairpin 2) appearing before the second precursor (Figure 2). We removed the astexin-2 precursor from pMM37 to express astexin-3 on its own (pMM39). To express only astexin-2, we excised the astexin-3 precursor and the intergenic region between the two precursors from pMM37 by overlap PCR to give pMM40. As a consequence, both pMM40 and pMM39 have the same

intergenic region between the precursor and the maturation enzymes. A fourth control construct, containing only the maturation enzymes, was constructed and named pMM38. For convenience, Table S1 includes descriptions of all plasmid constructs used in this study.

We expressed all four constructs in 20 amino acid M9 media at various scales with induction at $OD_{600} = 0.2-0.3$. We then extracted the supernatants, boiled cell lysates, and tested the crude extracts for the presence of astexins-2 and -3 by HPLC and MALDI mass spectrometry (MS) (see Methods in Supporting Information). There are several peaks in the HPLC chromatograms of pMM37, pMM39, and pMM40 cell lysates that are not present in the control (Figure 3). Mass spectra of the extracts confirm the presence of C-terminal truncations of astexin-2 in cells harboring pMM37 and pMM40 as well as full-length and C-terminal truncations of astexin-3 in pMM39 (Figure S1). Only minimal amounts of peptide were found in cell-free supernatants. As expected, each astexin species is singly dehydrated suggesting the presence of an isopeptide bond. To assign astexins-2 and -3 and their truncation variants to individual peaks in the chromatogram, we analyzed the material collected from the 12.24, 14.14, 14.64, 15.09, and 15.45 min peaks by MALDI MS. Only small amounts of astexins-2 and -3 species were detected in culture supernatants, indicating that these peptides are not exported into the extracellular medium (Figure S2).

The combined results indicate that full-length astexin-3 and its ΔC2 and ΔC1 C-terminal truncation products are expressed from pMM39, while pMM40 expression yields mostly ΔC4 and ΔC3 truncations of astexin-2 with essentially no full-length product. The HPLC trace of the pMM37 extract shows that both astexins-2 and -3 are present in the lysate, but the MALDI signal is dominated by astexin-2 (Figure S1a). This can be explained by preferential ionization of the arginine-containing astexin-2 in positive ion mode. Similar amounts of astexin-2 are produced by cells harboring the dual precursor construct pMM37 and the single precursor construct pMM40. In contrast, cells with the single precursor construct pMM39
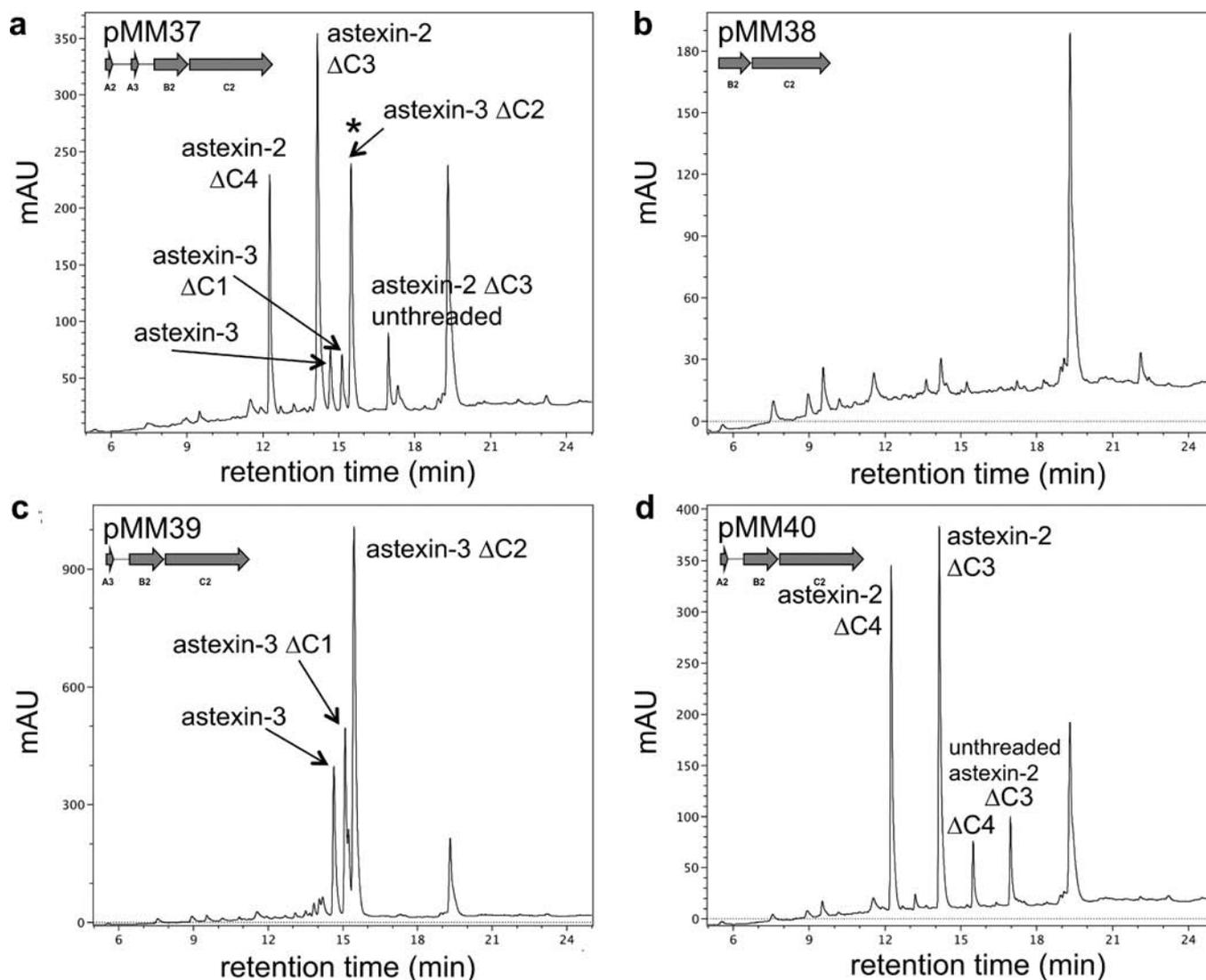
Figure 3. HPLC chromatograms of lysate extracts of pMM37, pMM38, pMM39, and pMM40 bearing cells. C-terminal truncation variants are indicated such that ΔC2 indicates that the 2 C-terminal amino acids have been removed. (a) The pMM37 dual precursor construct produces variants of astexins-2 and -3. The retention times for astexin-3 ΔC2 and unthreaded astexin-2 ΔC4 are identical. The peak containing both species is labeled with a star. (b) No peptides of interest are expressed from the pMM38 control construct. (c,d) The single precursor constructs, pMM39 and pMM40, produce variants of astexins-3 and -2, respectively. Unthreaded ΔC3 and ΔC4 astexin-2 are observed in (a) and (d) and labeled. No species corresponding to unthreaded astexin-3 are observed.

produce ca. 4-fold more astexin-3 than do cells harboring pMM37 (Figure 3). These results suggest that hairpin 3 (Figure 2) does not prevent lasso peptide production, a result that is in contrast to what was observed for astexin-1 where removal of the hairpin led to improved production.[4] Nonspecific cellular proteases are likely responsible for generating C-terminal truncations of astexins-2 and -3, and this process appears to be time dependent for astexin-3. After 48 h of expression, a lysate extract of pMM39 contained mostly ΔC2 astexin-3 (55% of total product), but full-length astexin-3 was still the major product after 12 h of expression (81%) (Figure S3). Finally, we have observed a mass increase of 16 Da in astexin-3 after prolonged incubation in methanol while exposed to the atmosphere. The most probable explanation for this mass increase is oxidation of Met5 within the ring of astexin-3 (Figure S4).

Two additional peaks are present in the pMM40 chromatogram at 15.48 and 16.95 min. MALDI MS analysis of HPLC collected fractions showed that these have the same molecular mass as astexin-2 ΔC4 and ΔC3, respectively. It has previously been observed that in certain lasso peptides, the tail can escape the confines of the ring at high temperatures[6,13] thus "unthreading" the lasso. Since our purification method includes a boiling step, these species are likely unthreaded variants of astexin-2. We tested this hypothesis by incubating an extract of pMM40 at 98 °C for 2 h followed by centrifugation to remove precipitates. The main peaks for astexin-2 ΔC4 and ΔC3 at 12.24 and 14.14 min were nearly gone after the thermal treatment, but the 15.48 and 16.95 min peaks remained and were even enriched (Figure S5a). In contrast, unthreading of astexin-3 was not observed after heating for 3 h at 95 °C (Figure S5b).

Mass Spectrometric Analysis of Astexins-2 and -3. We carried out extensive mass spectrometric studies on both astexins-2 and -3 to confirm their identity and sequence composition. The internal cyclization in lasso peptides

produces a distinct MS^2 fingerprint.[15−17] Specifically, while the macrolactam ring stays intact, the full y- and b-series of fragment ions are usually generated from residues in the loop and the tail. Astexins-2 and -3 have only one possible point of cyclization at Asp9. To confirm this, we subjected astexin-2 ΔC3, astexin-3, and synthetic linear versions of these peptides to MS^2 fragmentation (Figure S6). The ratios of fragment ions in the MS^2 spectra of the lassoed and synthetic linear peptides are strikingly different. Furthermore, the smallest observed b-series fragment (b9) and the largest observed y-series fragment (y15) of astexin-3 correspond to only the ring and only the tail, respectively. This indicates that the point of cyclization is indeed between Gly1 and Asp9. Conversely, strong signals for y21 and y18 fragment ions were observed in the MS^2 spectrum of the linear astexin-3, which are indicative of fragmentation beyond Asp9. Similarly, the cyclization of astexin-2 ΔC3 at Asp9 was confirmed by the fact that the y-series of fragment ions terminates at y12 in the MS^2 spectrum. On the other hand, the MS^2 spectrum of the synthetic linear astexin-2 ΔC3 peptide has y-series fragment ions that include Gln4 through Leu8.

NMR Solution Structure of Astexin-3. Samples of lassoed (3.25 mg) and synthetic linear (1.5 mg) astexin-3 were prepared in 200 μL DMSO-d_6 each for the acquisition of TOCSY, NOESY, and phase-sensitive COSY spectra on a 500 MHz spectrometer at 295 K. Experimental details about spectra acquisition can be found in the Supporting Information. In contrast to the extensive NOESY connectivity of the lassoed astexin-3, we did not detect significant NOESY cross peaks in the spectrum of the synthetic linear astexin-3. The TOCSY spectra of both the linear and lassoed astexin-3 have well-defined signals, yet their distribution is markedly different (Figure S7). As previously noted for MccJ25[15] and other lasso peptides,[18] the band of NH-Hα resonances is narrower in the f2 dimension in the synthetic linear astexin-3 than it is in lassoed astexin-3.

We assigned all proton chemical shifts in the TOCSY and NOESY spectra of lassoed astexin-3 except the amide protons of Leu8 and Gln14 (Table S2) and some exchangeable side chain protons. Chemical shifts of side chain protons of Leu8 were assigned based on cross peaks with its Hα proton and intraresidue resonances further up the side chain. Similarly, we identified cross peaks corresponding to magnetization transfer between side chain protons of Gln14 and its Hα proton. Gln14 has a very strong resonance between its Hε protons in both the TOCSY and NOESY spectrum. The presence of NOE cross peaks between the NH proton of Gly1 and the Hβ protons of Asp9 confirmed the presence of the macrocyclic ring. While the NH-Hα cross peak for Tyr15 was weak in the TOCSY, the cross peak between the amide proton and the protons on the aromatic ring was strong. Trp16 and Tyr15 had extensive connectivity to residues in the ring of astexin-3 (Figure S8). This evidence places the Tyr15/Trp16 dyad as the steric lock that traps the tail of astexin-3 in the ring. Specifically, we observed 19 long-range NOE contacts between Trp16 and Met5, Thr3, Gly7, Asp9, Gly1, and Leu8. Tyr15 had four connections to residues in the ring. Additionally, we observed four contacts between side-chain protons of Gln14, the residue immediately preceding the steric lock, and protons in Leu8 and Val6, which are in the ring.

The volumes of 120 inter-residue and 98 intraresidue cross peaks from the 100 ms NOESY spectrum were measured by integration and calibrated to the Gln14 Hε21-Hε22 crosspeak, yielding a set of upper distance restraints (Table S3). Restraints

on 11 ϕ torsion angles were derived from the vicinal coupling constants $^3J_{HNHα}$. Proper geometry of the Gly1-Asp9 covalent linkage was realized by introducing 8 additional constraints. Based on structures generated in the first round of simulated annealing with CYANA 2.1,[19] constraints that were violated in more than 10 of the 25 lowest energy structures were refined until the weighted sum of the squared violations of conformational restraints (CYANA target function) fell below 20 Å^2. Final simulated annealing was done from an initial set of 200 random structures and yielded an ensemble of top 20 structures with good covalent geometry and an average root-mean-square (rms) deviation of 0.94 ± 0.70 Å. These structures were subsequently energy minimized using TINKER[20] with the AMBER 94[21] force field to an rms gradient of 1.0 kcal/mol/Å. Statistics on structural calculations are summarized in Table S4.

The structure of astexin-3 is presented in Figure 4. The topology of astexin-3 features a relatively short six-residue loop
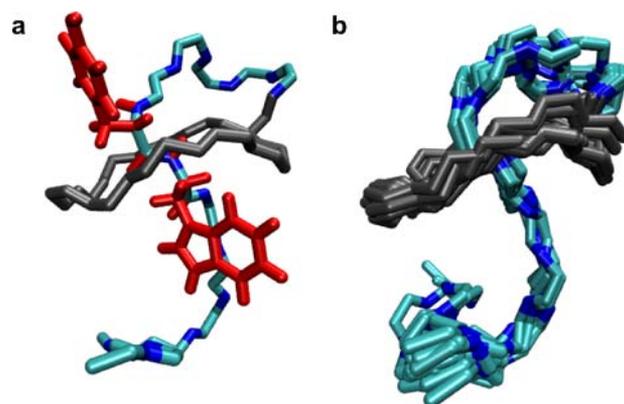

Figure 4. Representative solution structure of astexin-3. (a) Lowest energy structure from the NMR ensemble. The steric lock residues Tyr15 and Trp16 prevent the tail of the peptide from slipping from the ring and are highlighted in red. (b) Superposition of top 20 structures showing the low rms deviation of the ensemble of structures.

and a nine-residue tail. The ring of the peptide is rigid, but there is some flexibility in the loop and in the tail. Most structural flexibility in the tail happens after His19 where it bends toward the ring. Observed long-range NOE contacts between His19, Ala20, Leu22, and Asp24 and residues in the ring support this structural feature and suggest a compacting of the structure by minimization of solvent exposed area. If the sequence of astexin-2 is threaded onto a similar structure, the Phe15/Arg16 dyad would serve as the steric lock. The larger lock residues found in astexin-3 (Tyr15 and Trp16) may explain why astexin-3 is stable upon boiling and astexin-2 is not.

AtxE2 Is an Isopeptide Hydrolase of Astexins-2 and -3. After producing astexins-2 and -3 and solving the astexin-3 structure, we turned our attention to the putative protease in this cluster, AtxE2. Given the low soluble yields and challenges in obtaining pure lasso peptide maturation enzymes,[10,22] we were pleasantly surprised to note that a histidine-tagged AtxE2 expressed well and was readily purified to homogeneity (Figure S9). We investigated whether this enzyme had any activity toward astexins-2 and -3. To this end, we prepared pMM39 and pMM40 lysate extracts containing astexins-2 and -3 and their C-terminal truncation variants (refer to Figure 3). Since most of the cellular proteins and lipids are removed from the lysates prior to C8 extraction by boiling and centrifugation, thermostable lasso peptides are the major products in the extract
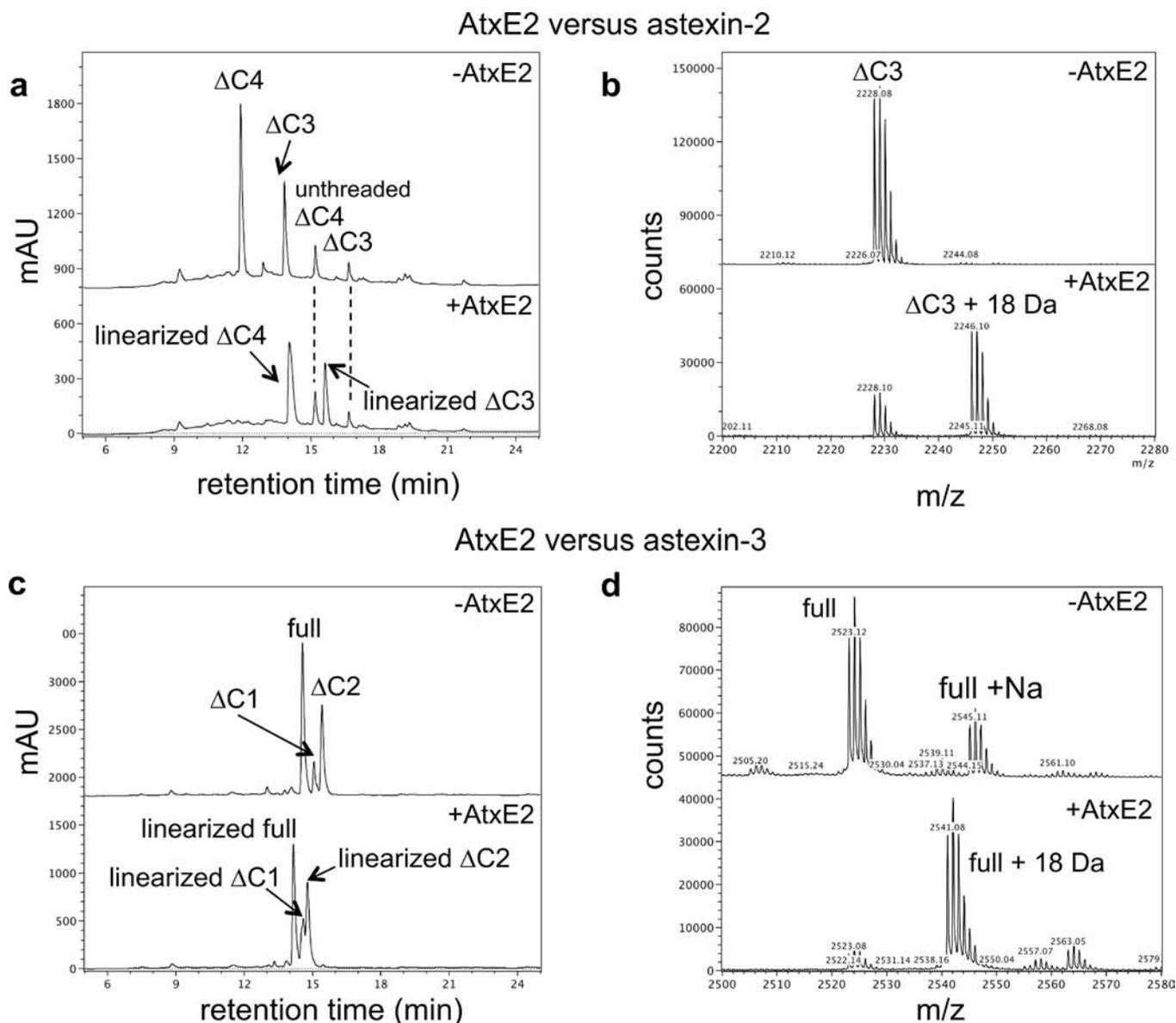
Figure 5. HPLC and MALDI-MS analysis of astexin-2 and -3 extracts before and after treatment with isopeptidase AtxE2. (a) The retention times of lassoed, threaded astexin-2 species increase after incubation with AtxE2. Unthreaded astexin-2 species retention times do not change. (b) Mass spectra confirm hydrolysis of astexin-2 ΔC3. (c) The retention times of astexin-3 species decrease after AtxE2 treatment. (d) The hydrolysis of astexin-3 was confirmed by MS.

without additional purification. We incubated 10 μL of each extract in a phosphate buffered saline solution at pH 7.5 with 860 nM AtxE2 for 6 h at room temperature. The reaction mixture was then quenched by heating to 65 °C for 30 min and cleaned up for HPLC analysis. All threaded astexin-2 species experienced an increase in retention time after AtxE2 treatment, while the retention time of the astexin-3 species decreased (Figure 5a,c). Table S5 summarizes the absolute retention times for all astexin species observed using a water/acetonitrile gradient described previously.[4]

MALDI MS analysis of the reaction mixtures showed an increase of 18 mass units for all astexin-2 and -3 species, corresponding to the addition of a water molecule via the cleavage of a single amide bond (Figure 5b,d). Since the tail of astexin-3 is locked in place on either side of the ring by a tyrosine and a tryptophan residue, it is conceivable that one of the peptide bonds in the ring or the loop of the peptide could

have been cut. However the MS[2] spectra of the linear synthetic astexins-2 and -3 peptides and the lasso peptides treated with AtxE2 are identical (Figure 6) as are their HPLC retention times (Figure S10). This shows that AtxE2 hydrolyzes the isopeptide bond between Gly1 and Asp9 selectively. In contrast to the results obtained for astexins-2 and -3, the retention time of astexin-1 did not shift after incubation with AtxE2, and no molecular weight difference was observed in the MS (Figure S11). This indicates that AtxE2 has specificity toward only the two lasso peptides in its own cluster.

AtxE2 Does Not Hydrolyze Unthreaded Astexin-2. As noted above, we observed unthreaded astexin-2 species in HPLC analyses of extracts of cells harboring pMM37 and pMM40 (Figure 3). Remarkably, the retention times for these unthreaded species did not change after incubation with AtxE2 (Figure 5A). To probe this observation further, we purified astexin-2 ΔC3 in its threaded form and generated an
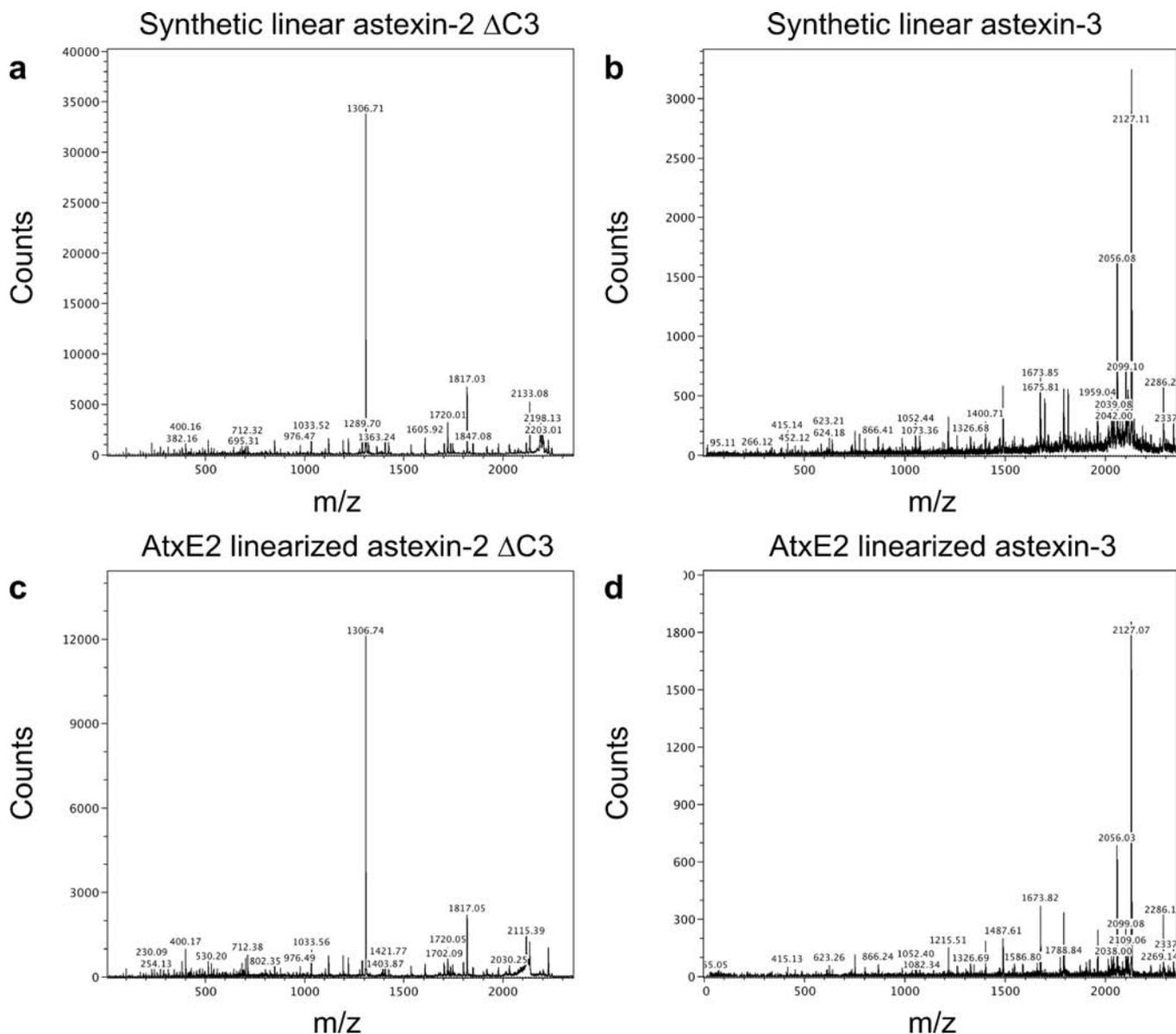
12042

dx.doi.org/10.1021/ja4054256 | J. Am. Chem. Soc. 2013, 135, 12038−12047

Figure 6. MS^2 analysis showing that the synthetic linear astexin-2 ΔC3 (a) is identical to the AtxE2 linearized astexin-2 ΔC3 (c) and that synthetic linear astexin-3 (b) is identical to AtxE2 linearized astexin-3 (d).

unthreaded form by extensive heat treatment. Both the threaded and unthreaded astexin-2 ΔC3 peptides were treated with 424 nM AtxE2, but only the threaded species exhibited a change in retention time consistent with hydrolysis of the isopeptide bond (Figure S12). Collectively these results suggest that AtxE2 can only function on a knotted structure, likely making it a highly specific enzyme. Such preference for a fully processed version of a natural product has been observed for the lanthipeptide flavipeptin, whose leader peptide is cleaved by a prolyl oligopeptidase only after the intramolecular rings have been installed.[23]

AtxE1 Hydrolyzes Astexin-1 *in Vivo*. In contrast to AtxE2, we were unable to solubly express the AtxE1 enzyme. We investigated AtxE1 activity toward astexin-1 *in vivo* by expressing astexin-1 along with AtxE1 in *E. coli*. Plasmid pMM62 was constructed by introducing the gene for AtxE1 downstream of the astexin-1 biosynthesis cluster in pMM32,[4] with its own ribosome binding site. In contrast to the supernatant extracts of pMM32, MALDI results showed both

lassoed and linearized versions of astexin-1 in the supernatant of pMM62 (Figure S13).

Kinetics of Astexin-3 Proteolysis by AtxE2. Tryptophan fluorescence of peptides and proteins has been used extensively to probe changes in their conformation.[24] Astexin-3 contains a single tryptophan positioned right below its ring which prompted us to look for differences in the emission spectra of the lassoed and linear astexin-3 upon excitation with 275 nm light. Figure S14 shows that the emission intensity at the 340 nm of lassoed astexin-3 is ca. 7-fold higher than that of the linearized astexin-3 at the same molar concentration. To calculate k_{cat} and K_m associated with lasso peptide hydrolysis, we carried out protease assays while varying the concentration of astexin-3 at a constant concentration of AtxE2 at 24 °C. Measurements were taken using a fluorescence plate reader at discrete time intervals or as time points from aliquots of reactions quenched after different times (Figure S15). The concentration of AtxE2 (87 nM) was chosen such that the initial astexin-3 concentration would always be in at least 90-

fold molar excess relative to the peptidase. We determined the k_{cat}/K_m value for AtxE2 to be 2.9×10^3 M^{-1}s^{-1}, with $k_{cat} = 0.38$ s^{-1} and $K_m = 131$ μM. The data and fit are shown in Figure 7
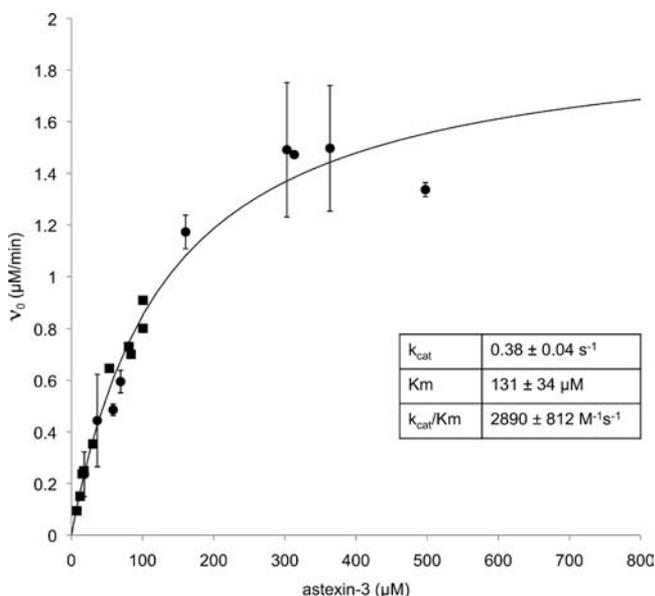

Figure 7. Michaelis−Menten plot detailing the kinetics of the hydrolysis of astexin-3 by AtxE2. Relevant parameters with 95% confidence intervals are summarized in the inset table. Circles indicate data points obtained in triplicate with the corresponding error bars (standard deviation), while squares represent single measurements. The R^2 value for the fit is 0.97.

along with a summary of the relevant parameters. The k_{cat}/K_m for AtxE2 is several orders of magnitude smaller than that of serine proteases acting on simple peptide substrates ($k_{cat}/K_m \sim 10^7$ M^{-1}s^{-1}).[25] From homology modeling (see below), AtxE2 most closely resembles prolyl oligopeptidases, and the kinetics of AtxE2 are on par with characterized enzymes in this family.[26]

Serine 527 Is the Catalytic Residue in AtxE2. Several serine proteases in the prolyl oligopeptidase family (Pfam 00326) were identified as homologues of AtxE1 and AtxE2 through modeling using the I-TASSER server.[27] The putative catalytic triads are located in the C-terminal regions of both proteins with Ser526-Asp550-His639 being the relevant residues in AtxE1 and Ser527-Glu610-His638 in AtxE2. We wanted to confirm that AtxE2 was catalyzing amide bond hydrolysis using Ser527 as the nucleophile. To this end we expressed and purified the S527A mutant of AtxE2 and assayed its activity toward astexin-3 *in vitro*. AtxE2 S527A had no activity toward astexin-3 (Figure S16) supporting the idea that this serine residue is the catalytic nucleophile.

Phylogenetic Analysis of Lasso Peptide Synthetases. With the confirmation that AtxE2 is a *bona fide* lasso peptide isopeptidase, we revisited our global lasso peptide genome mining data to determine how frequently such isopeptidases are observed in the neighborhood of lasso peptide clusters and whether there are other genes that belong to the clusters. We noted the presence of a GntR homologue and a TonB-dependent transporter (TBDT) in both astexin clusters and homologues of FecR and FecI in the astexin-1 cluster (Figure 1). The putative clusters in *Sphingobium japonicum* and the caulosegnin cluster[4,6] also feature these genes. We used MEME[28] to generate conserved domain motifs for the GntR,

isopeptidase, TBDT, FecI, and FecR homologues using protein sequences from *A. excentricus, Caulobacter segnis,* and *S. japonicum* as the training set. As noted above, many previously described lasso peptide gene clusters include an ABC transporter rather than an isopeptidase. Motifs for the lasso peptide ABC transporter were generated from the proteins McjD, CapD, and LarE.[5,9,11] Open reading frame translations within 20 kbp of the lasso peptide biosynthesis genes were queried for the presence of these motifs.

Of the 81 clusters that were analyzed, 17 were found to have an isopeptidase, 25 an ABC transporter, and the remaining clusters had neither (Figure S17). Each of the clusters with an isopeptidase also had a TBDT, and all except the astexins-2, -3 and *Xanthomonas gardneri* clusters had homologues of FecR and FecI. A GntR homologue was identified in 13 of the 17 clusters (Table 1). Of the clusters with an ABC transporter, 17

Table 1. Summary of Lasso Peptide Cluster Analysis

characteristic	no. of clusters
Isopeptidase	**17**
TonB homologue	17
GntR homologue	14
FecR homologue	15
FecI homologue	15
multiple precursors	7
multiple clusters in a single organism	7
ABC Transporter	**25**
split-B	17
split-C	1
fused B-D	4
multiple precursors	3
multiple clusters in a single organism	1
Biosynthesis Only Clusters	**9**
split-B	4
split-C	1
multiple precursors	1
multiple clusters in a single organism	1

had a B gene homologue that was shorter than a typical B gene by ca. 100 residues but still carried the essential trans-glutaminase catalytic triad located near the C-terminus of these enzymes.[10,22,29,30] In all cases, a second smaller protein annotated as being of unknown function was encoded nearby. It has been noted that the lariatin cluster includes such a protein (termed LarC by the authors).[11] However, no comparison of the lariatin maturation enzymes to those of other lasso peptides has been done to determine the role of this unknown gene. By looking globally at all the clusters with and without this maturation enzyme, we determined that the predicted secondary structure of the shorter protein aligns well with the N-terminal portions of McjB, CapB, AtxB1, AtxB2, and other "full-length" B gene homologues, and contains a conserved LDXXXXRYFXL motif (Figure S18). This suggests that the function of the B homologue is split between two proteins in these clusters, an observation supported by the fact that LarC is essential for lariatin production.[11] Another novel aspect of our survey of lasso peptide cluster architectures is the identification of putative "B-D" fusion proteins in *Streptococcus suis, Enterococcus faecalis,* and *Ruminococcus albus.* In these organisms, the N-terminal portion of the B homologue is also encoded as a distinct protein, but the C-terminal catalytic domain appears to be fused to the ABC transporter. In one of
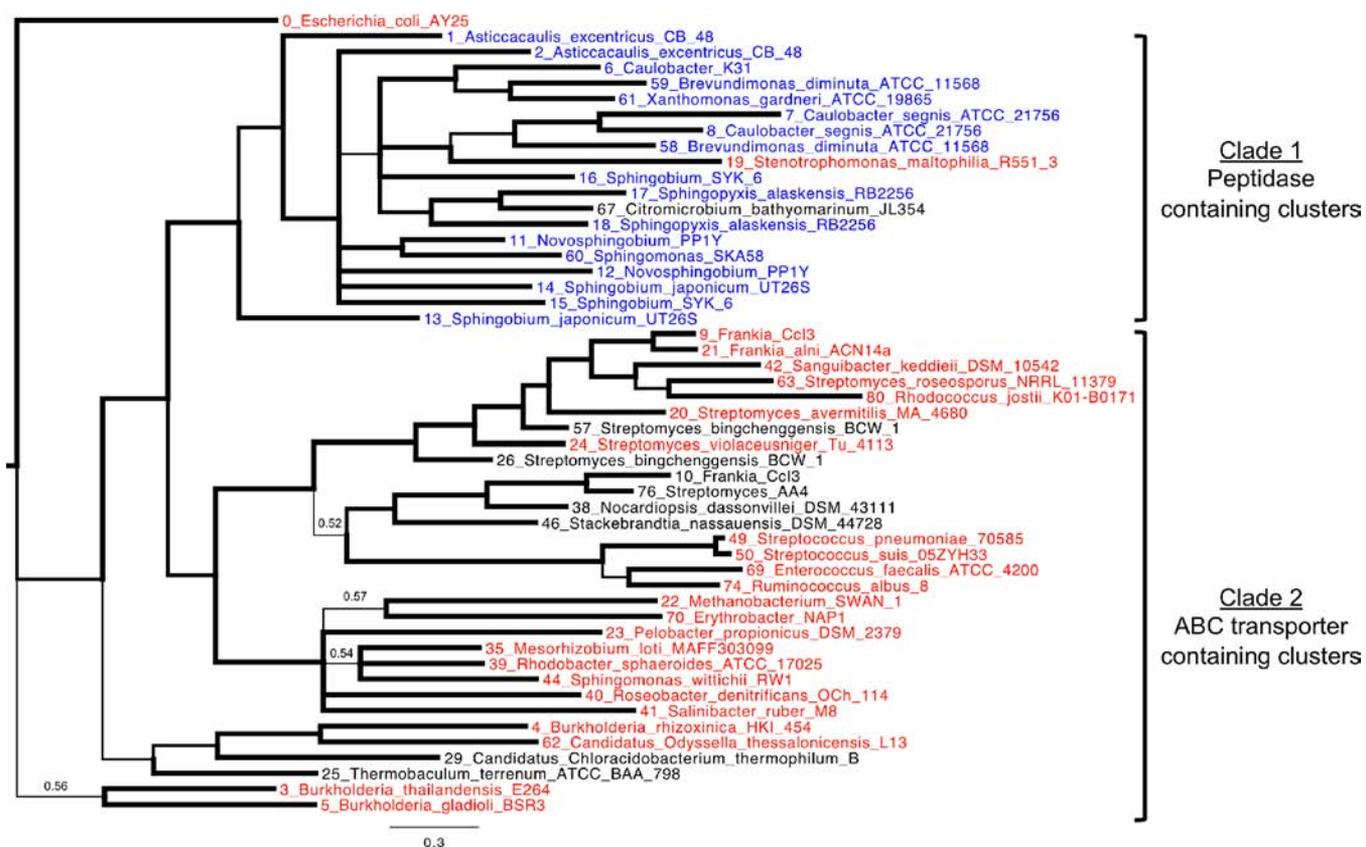
Figure 8. Phylogenetic tree of known and putative lasso B homologues. Clusters containing a peptidase are highlighted in blue, while clusters with an ABC transporter are shown in red. Clusters containing only an A, B, and C homologue are presented in black. Thicker line width indicates higher values of the Bayesian posterior probability. Branches with posterior probabilities <0.6 are indicated in the figure. Refer to Figure S15 for the architectures of all clusters in this tree.

the *Frankia* CcI3 and the *S. suis* clusters, the C homologue appears to be split between two distinct proteins.

The successful classification of lasso peptide clusters based on architecture alone led us to consider whether sequences of B and C homologues also segregate into distinct clades. To investigate their evolutionary relatedness, we performed a Bayesian phylogenetic analysis on the protein sequences using MrBayes 3.2.1,[31−33] a strategy recently used to examine relationships between lanthipeptide clusters.[34] Since there is no structural information on any B and C homologues, we used a secondary structure prediction generated with SPINE-X[35] to assemble an accurate multiple sequence alignment (Figure S19). Thirty putative homologues were discarded from the analysis due to poor alignment. Since the two portions of the "split-B" homologues could experience differing amino acid substitution rates, we chose to use only the C-terminal portion of the proteins for the analysis. Similarly, short regions of the C homologue responsible for ATP binding, Mg^{2+} coordination and two additional conserved domains were chosen as input to MrBayes. All Bayesian MCMC inference analyses were run for at least 5×10^6 generations with 2 sets of 7 chains (1 cold and 6 heated) to convergence, as indicated by a value of the average standard deviation of split frequencies <0.01. A 25% burn-in was accepted before calculating the final statistics.

Phylogenetic trees derived from either the B or the C gene homologues have similar branching (Figures 8 and S20). These trees segregate into two clear clades. Remarkably, Clade 1 consists almost exclusively of clusters containing an isopepti-dase. Clade 2 is made up of canonical lasso peptide clusters

with an ABC transporter as well other "biosynthesis-only" clusters with just A, B, and C gene homologues. Even though the genome of *Citromicrobium bathyomarinum* is incomplete, its membership in Clade 1 suggests that it may be an isopeptidase-containing cluster. Assessing the sequence composition and amino acid conservation of the precursor peptides provided further support to our two-clade model and allowed the identification of general features of lasso peptide precursors in both clades. Clade 1 core peptides are generally polar, negatively charged, and terminate in an aspartic or glutamic acid residue, while Clade 2 peptides are mostly hydrophobic with patches of positive charge (Figure S21). Besides the functionally important Thr residue in the penultimate position of the leader peptide,[12,36] a conserved proline residue is found in Clade 2 leader peptides. The conservation of amino acid composition of lasso peptides in the two clades suggested an evolutionary pressure on the function of these natural products.

■ CONCLUSIONS

In this study we describe the first two examples of isopeptidases that act selectively on lasso peptides produced by *A. excentricus*. We demonstrate activity of AtxE2 toward astexin-2 and astexin-3 *in vitro* and show that AtxE1 can hydrolyze astexin-1 *in vivo*. In addition to elucidating the kinetic parameters of AtxE2, we used NMR to solve the structure of astexin-3. We establish the identity of astexin-2 as a lasso peptide by MS, thus adding to the list of lasso peptides in the astexin family. We present evidence that AtxE1 and AtxE2 catalyze the reverse reaction of lasso peptide synthesis by cleaving the isopeptide bond that was

installed during the maturation of the lasso peptides. At least in the case of astexin-2, AtxE2 works by recognizing the lasso topology rather than a specific amino acid sequence, since unthreaded astexin-2 is left intact after AtxE2 treatment. This makes AtxE2 stand out in relation to intracellular proteases that target unstructured protein regions for degradation.[37] The discovery that AtxE1 and AtxE2 can deconstruct the astexins also opens up new avenues for engineering these lasso peptides.

When heterologously expressed in *E. coli*, astexins-2 and -3 are intracellular lasso peptides in contrast to astexin-1, which can be found both in the culture supernatant[4] and in producer cells.[13] It is conceivable that the astexin family of peptides functions intracellularly, though it is also possible that these peptides are exported from the native producer *A. excentricus* by some protein not present in *E. coli*. Phylogenetic analysis revealed that the astexins belong to the same clade of lasso peptides as the caulosegnins, which is distinct from the clade of capistruin, lariatin, and other clusters with an ABC transporter. Furthermore, differences in the gene architectures between the two clades and the sequence composition of their lasso peptides suggest differences in function. To date, the only activity ascribed to a Clade 1 lasso peptide is the modest narrow-spectrum antimicrobial activity of astexin-1. In contrast, MccJ25, capistruin, and lariatin are all antimicrobial peptides, some with fairly potent activity. The associated ABC transporter in these Clade 2 peptides serves as an exporter and immunity factor, and the sequences of the peptides themselves are evolutionarily attuned to infiltrate susceptible cells by co-opting iron import machinery[38,39] and disabling the RNA polymerase of the target organism.[40−42] Conversely, Clade 1 lasso peptides have a membrane-bound importer protein and a dedicated enzyme to specifically deconstruct the lasso peptide in the cell. Additionally, the expression and degradation of these peptides appears to be tightly controlled by a helix−turn−helix transcriptional repressor and a sigma/antisigma pair. This is remarkably similar to biosynthetic and regulatory system of siderophores.[43] In light of this, the natural function of the astexins (and likely other Clade 1 peptides) is more consistent with a type of scavenging molecule that acts through a catch-and-release mechanism.

Another fascinating aspect of the lasso peptide superfamily that remains to be learned is their evolutionary history. For example, did "whole-B" homologues evolve from two genes by fusion, or did the "split-B" architecture emerge through a duplication event followed by selective degradation of redundant domains? Both gene fusion and fission are commonly observed events in prokaryotes with their own set of evolutionary benefits and drawbacks that are highly context dependent.[44] Parsimony suggests that Clade 1 is more ancient simply because the number of evolutionary changes that would have to occur to transform an ABC-transporter cluster into an isopeptidase-type cluster is much greater than the opposite possibility. Regardless of the origins of these clades, the insights that we have gained about the existence of lasso peptide isopeptidases, astexins-2 and -3, and the various architectures of lasso peptide clusters will facilitate engineering and characterization of known and as-yet unknown lasso peptides.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Detailed methods and supplementary experimental and bioinformatics data. The coordinates for the astexin-3 structure are deposited at BMRB (accession number 19250) and PDB (accession number 2M8F). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

ajlink@princeton.edu

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. A.; Garavelli, J. S.; Goransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Muller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Sussmuth, R. D.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. *Nat. Prod. Rep.* **2013**, *30*, 108.

(2) Velasquez, J. E.; van der Donk, W. A. *Curr. Opin. Chem. Biol.* **2011**, *15*, 11.

(3) Maksimov, M. O.; Pan, S. J.; Link, A. J. *Nat. Prod. Rep.* **2012**, *29*, 996.

(4) Maksimov, M. O.; Pelczer, I.; Link, A. J. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 15223.

(5) Knappe, T. A.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X. L.; Marahiel, M. A. *J. Am. Chem. Soc.* **2008**, *130*, 11446.

(6) Hegemann, J. D.; Zimmermann, M.; Xie, X. L.; Marahiel, M. A. *J. Am. Chem. Soc.* **2013**, *135*, 210.

(7) Kersten, R. D.; Yang, Y.-L.; Xu, Y.; Cimermancic, P.; Nam, S.-J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C. *Nat. Chem. Biol.* **2011**, *7*, 794.

(8) Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Salomon, R. A. *J. Bacteriol.* **1996**, *178*, 3661.

(9) Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Gonzalez-Pastor, J. E.; Moreno, F.; Salomon, R. A. *J. Bacteriol.* **1999**, *181*, 2659.

(10) Duquesne, S.; Destoumieux-Garzón, D.; Zirah, S.; Goulard, C.; Peduzzi, J.; Rebuffat, S. *Chem. Biol.* **2007**, *14*, 793.

(11) Inokoshi, J.; Matsuhama, M.; Miyake, M.; Ikeda, H.; Tomoda, H. *Appl. Microbiol. Biotechnol.* **2012**, *95*, 451.

(12) Pan, S. J.; Rajniak, J.; Maksimov, M. O.; Link, A. J. *Chem. Commun.* **2012**, *48*, 1880.

(13) Zimmermann, M.; Hegemann; Julian, D.; Xie, X.; Marahiel; Mohamed, A. *Chem. Biol.* **2013**, *20*, 558.

(14) Skerra, A. *Gene* **1994**, *151*, 131.

(15) Wilson, K. A.; Kalkum, M.; Ottesen, J.; Yuzenkova, J.; Chait, B. T.; Landick, R.; Muir, T.; Severinov, K.; Darst, S. A. *J. Am. Chem. Soc.* **2003**, *125*, 12475.

(16) Loo, J. A.; He, J. X.; Cody, W. L. *J. Am. Chem. Soc.* **1998**, *120*, 4542.

(17) Zirah, S.; Afonso, C.; Linne, U.; Knappe, T. A.; Marahiel, M. A.; Rebuffat, S.; Tabet, J. C. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 467.

(18) Xie, X. L.; Marahiel, M. A. *ChemBioChem* **2012**, *13*, 621.

(19) Guntert, P.; Mumenthaler, C.; Wuthrich, K. *J. Mol. Biol.* **1997**, *273*, 283.

(20) Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016.

12046

dx.doi.org/10.1021/ja4054256 | *J. Am. Chem. Soc.* 2013, 135, 12038−12047

(21) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(22) Yan, K. P.; Li, Y. Y.; Zirah, S.; Goulard, C.; Knappe, T. A.; Marahiel, M. A.; Rebuffat, S. *ChemBioChem* **2012**, *13*, 1046.

(23) Völler, G. H.; Krawczyk, B.; Ensle, P.; Süssmuth, R. D. *J. Am. Chem. Soc.* **2013**, *135*, 7426.

(24) Eftink, M. R. *Biophys. J.* **1994**, *66*, 482.

(25) Hedstrom, L. *Chem. Rev.* **2002**, *102*, 4501.

(26) Nomura, K. *FEBS Lett.* **1986**, *209*, 235.

(27) Zhang, Y. *BMC Bioinformatics* **2008**, *9*.

(28) Bailey, T. L.; Elkan, C. *Machine Learning* **1995**, *21*, 51.

(29) Pan, S. J.; Rajniak, J.; Cheung, W. L.; Link, A. J. *ChemBioChem* **2012**, *13*, 367.

(30) Severinov, K.; Semenova, E.; Kazakov, A.; Kazakov, T.; Gelfand, M. S. *Mol. Microbiol.* **2007**, *65*, 1380.

(31) Ronquist, F.; Huelsenbeck, J. P. *Bioinformatics* **2003**, *19*, 1572.

(32) Huelsenbeck, J. P.; Ronquist, F. *Bioinformatics* **2001**, *17*, 754.

(33) Altekar, G.; Dwarkadas, S.; Huelsenbeck, J. P.; Ronquist, F. *Bioinformatics* **2004**, *20*, 407.

(34) Zhang, Q.; Yu, Y.; Velasquez, J. E.; van der Donk, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 18361.

(35) Faraggi, E.; Zhang, T.; Yang, Y. D.; Kurgan, L.; Zhou, Y. Q. *J. Comput. Chem.* **2012**, *33*, 259.

(36) Cheung, W. L.; Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.* **2010**, *132*, 2514.

(37) Sauer, R. T.; Baker, T. A. In *Annual Review of Biochemistry*; Kornberg, R. D., Raetz, C. R. H., Rothman, J. E., Thorner, J. W., Eds.; Annual Reviews: Palo Alto, CA, **2011**; Vol. *80*, p 587.

(38) Destoumieux-Garzon, D.; Duquesne, S.; Peduzzi, J.; Goulard, C.; Desmadril, M.; Letellier, L.; Rebuffat, S.; Boulanger, P. *Biochem. J.* **2005**, *389*, 869.

(39) Salomon, R. A.; Farias, R. N. *J. Bacteriol.* **1993**, *175*, 7741.

(40) Adelman, K.; Yuzenkova, J.; La Porta, A.; Zenkin, N.; Lee, J.; Lis, J. T.; Borukhov, S.; Wang, M. D.; Severinov, K. *Mol. Cell* **2004**, *14*, 753.

(41) Kuznedelov, K.; Semenova, E.; Knappe, T. A.; Mukhamedyarov, D.; Srivastava, A.; Chatterjee, S.; Ebright, R. H.; Marahiel, M. A.; Severinov, K. *J. Mol. Biol.* **2011**, *412*, 842.

(42) Mukhopadhyay, J.; Sineva, E.; Knight, J.; Levy, R. M.; Ebright, R. H. *Mol. Cell* **2004**, *14*, 739.

(43) Noinaj, N.; Guillier, M.; Barnard, T. J.; Buchanan, S. K. In *Annual Review of Microbiology*; Gottesman, S., Harwood, C. S., Eds.; Annual Reviews: Palo Alto, CA, **2010**; Vol. *64*, p 43.

(44) Snel, B.; Bork, P.; Huynen, M. *Trends Genet.* **2000**, *16*, 9.