JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Molecular Insights into the Biosynthesis of Guadinomine: A Type III Secretion System Inhibitor

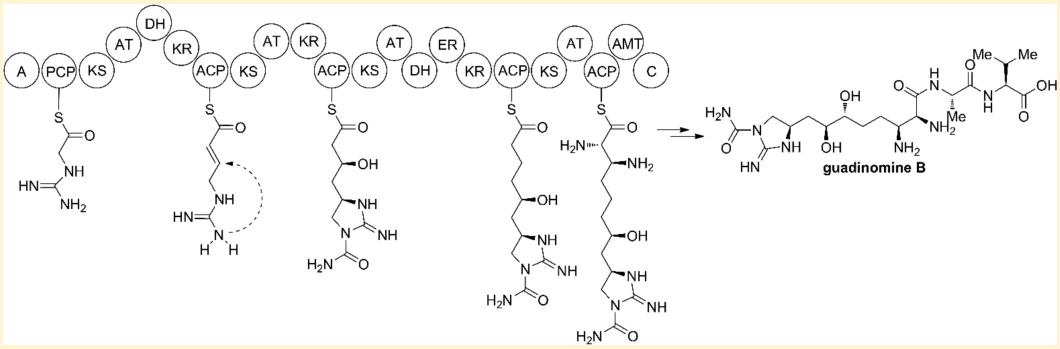
Tracy C. Holmes,[†,▽] Aaron E. May,[‡,▽] Kathia Zaleta-Rivera,[‡] J. Graham Ruby,[⊥] Peter Skewes-Cox,[⊥] Michael A. Fischbach,[∥] Joseph L. DeRisi,[⊥] Masato Iwatsuki,[#] Satoshi Ōmura,[#] and Chaitan Khosla*[,†,‡,§]

[†]Departments of Chemical Engineering, [‡]Chemistry, and [§]Biochemistry, Stanford University, Stanford, California 94305, United States

[⊥]Departments of Biochemistry and Biophysics, and [∥]Computational and Systems Biology, Cellular and Molecular Engineering, University of California, San Francisco, San Francisco, California 94158, United States

[#]Kitasato Institute for Life Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8642, Japan

Supporting Information



ABSTRACT: Guadinomines are a recently discovered family of anti-infective compounds produced by *Streptomyces* sp. K01-0509 with a novel mode of action. With an IC_{50} of 14 nM, guadinomine B is the most potent known inhibitor of the type III secretion system (TTSS) of Gram-negative bacteria. TTSS activity is required for the virulence of many pathogenic Gram-negative bacteria including *Escherichia coli*, *Salmonella* spp., *Yersinia* spp., *Chlamydia* spp., *Vibrio* spp., and *Pseudomonas* spp. The guadinomine (*gdn*) biosynthetic gene cluster has been cloned and sequenced and includes 26 open reading frames spanning 51.2 kb. It encodes a chimeric multimodular polyketide synthase, a nonribosomal peptide synthetase, along with enzymes responsible for the biosynthesis of the unusual aminomalonyl-acyl carrier protein extender unit and the signature carbamoylated cyclic guanidine. Its identity was established by targeted disruption of the gene cluster as well as by heterologous expression and analysis of key enzymes in the biosynthetic pathway. Identifying the guadinomine gene cluster provides critical insight into the biosynthesis of these scarce but potentially important natural products.

INTRODUCTION

The problem of antibiotic resistance has led to an intensive search for new anti-infective agents. Recently an innovative approach to this challenge was highlighted by one of our laboratories in which bacterial components required exclusively for virulence were targeted.[1] Because the type III secretion system (TTSS) is not required for bacterial survival but is indispensable for virulence, it was hypothesized that inhibition of the TTSS would lead to avirulence without imposing selection pressure for mutants that survive antibiotic action. As a preliminary proof of concept, dosing mice with the small molecule TTSS inhibitor, aurodox, was shown to protect against infection with an otherwise lethal inoculation of *Citrobacter rodentium*.[2]

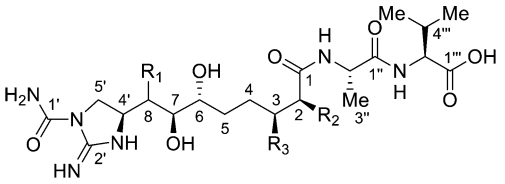
The TTSS is a protein secretion and translocation apparatus utilized by many pathogenic Gram-negative bacteria as their primary virulence mechanism. Common Gram-negative pathogens that use a TTSS include *Escherichia coli*, *Salmonella* spp., *Yersinia* spp., *Chlamydia* spp., *Vibrio* spp., and *Pseudomonas* spp. Virulent strains of these pathogens are responsible for chronic infections and numerous diseases including gastroenteritis, plague, shigellosis, chlamydia, cholera, and pneumonia. The structural components of the TTSS comprise 20−25 proteins, half of which are conserved among the various TTSS-containing species.[3] Hallmarks of the TTSS are a membrane-spanning basal body and a needle protein that forms a pore in the host cell membrane with the help of translocator proteins. Virulence factors, termed effector proteins, are then delivered through the needle protein into the host cell in an ATP-dependent manner. Effector proteins function by a variety of mechanisms to inhibit signaling cascades and block the ability

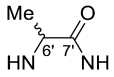
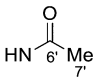
Received: August 30, 2012
Published: October 2, 2012

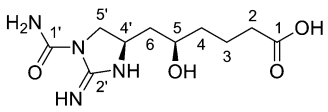
   dx.doi.org/10.1021/ja308622d | J. Am. Chem. Soc. 2012, 134, 17797−17806

of the host to respond to infection, eventually allowing the pathogen to colonize the host.[4] Importantly, engineered pathogens lacking a TTSS or critical effector proteins are often incapable of colonizing a host.[5−8]

In 2008, one of our laboratories reported the discovery of six structurally related compounds isolated from the culture broth of *Streptomyces* sp. K01-0509, named guadinomines A, B, C_1, C_2, D, and guadinomic acid (Figure 1).[9,10] Common to the guadinomines is a carbamoylated cyclic guanidine, a vicinal diamine at C-2/C-3, and a terminal Ala-Val dipeptide. Additionally, the C-6 and C-7 positions of all guadinomines are hydroxylated. The guadinomines were demonstrated to have potent anti-infective activity against enteropathogenic *E. coli* (EPEC) in sheep red blood cell hemolysis assays, with guadinomines A and B showing the strongest anti-TTSS activity reported to date with IC_{50} values of 39 nM (20 ng/mL) and 14 nM (7 ng/mL), respectively.[9]



Guadinomine	R_1	R_2	R_3
A (1)	OH	NH_2	NH_2
B (2)	H	NH_2	NH_2
C_1 (3a), C_2 (3b) (diastereomers)	H		
D (4)	H		NH_2



Guadinomic Acid (5)

Figure 1. Structures of the guadinomines A, B, C_1, C_2, D and guadinomic acid.

To initiate molecular investigations into guadinomine biosynthesis, we sought to clone the biosynthetic gene cluster from *Streptomyces* sp. K01-0509. We speculated that the carbon skeleton of this natural product was synthesized by a multimodular polyketide synthase (PKS). At the same time, we also recognized that the enzymes needed to produce and incorporate the cyclic guanidine and the vicinal diamine, being relatively rare, could be important signatures of the gene cluster. Here we describe the architecture and preliminary biochemical characterization of this gene cluster. Identification of these biosynthetic enzymes provides an opportunity for further studies into their substrate flexibility, while also opening the doors for precursor-directed biosynthesis and combinatorial bioengineering to produce novel analogs of this intriguing anti-infective agent.

RESULTS

Cloning and Sequencing of the Guadinomine Gene Cluster. A combined approach of whole genome shotgun sequencing and cosmid library screening was used to identify the guadinomine gene cluster. Due to the high GC content and complexity of secondary structure in the *Streptomyces* sp. K01-0509 genome, assembly with PRICE *de novo* genome assembly software gave fragmented contigs of the genome. Specifically, 15 400 contigs were obtained, ranging from 70 to 130 000 bp. Basic local alignment search tool (BLAST) was used to identify contigs with homology to proteins expected to be in the guadinomine gene cluster. This search yielded a 3000 bp contig containing a homologue to the zwittermicin A (ZmA) acyl carrier protein, ZmaH, involved in aminomalonyl-acyl carrier protein (ACP) biosynthesis. Aminomalonyl-ACP is an extremely unusual extender unit that, to our knowledge, has only been found in the ZmA biosynthetic pathway from *Bacillus cereus*.[11] It was hypothesized that guadinomine biosynthesis requires a similar aminomalonyl-ACP unit to incorporate an α-amino group into the polyketide backbone. This amine eventually becomes part of the vicinal diamine moiety common to guadinomines. The 3000 bp contig also contained homologues to ZmaG and ZmaI (two ACP-recognizing dehydrogenases) and ZmaJ (a seryl-AMP synthetase), which comprise the balance of the aminomalonyl-ACP cluster from zwittermicin A. Portions of the 3000 bp contig were used to design oligonucleotide probes for screening the *E. coli*-packaged SuperCos I cosmid library by polymerase chain reaction (PCR). The ZmaH probe was used to identify cosmid 7B6. Subsequently, a second cosmid (4A10) was identified having significant overlap with 7B6 (Table S1). Together, these cosmids contained the guadinomine gene cluster (Figure 2). Both were sequenced using primer walking as well as

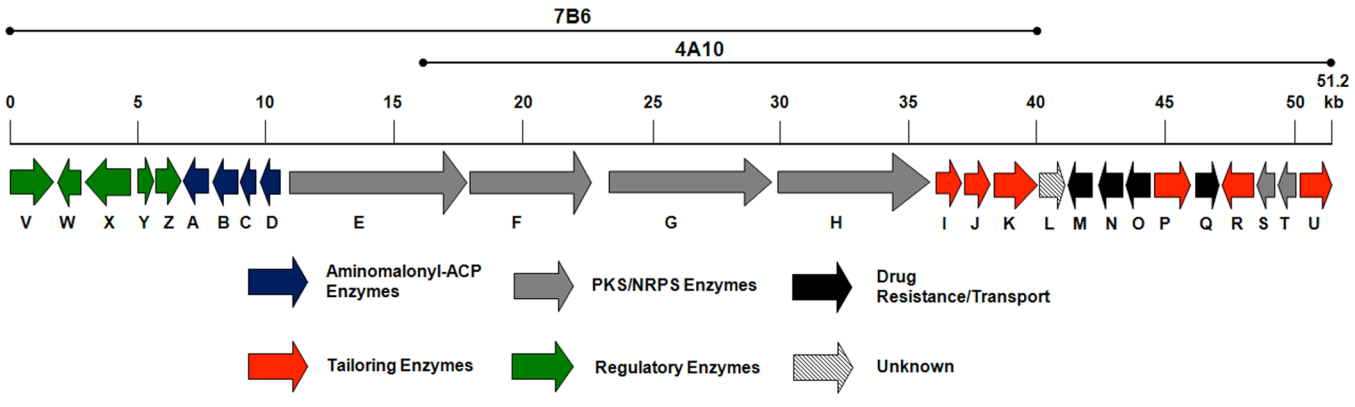


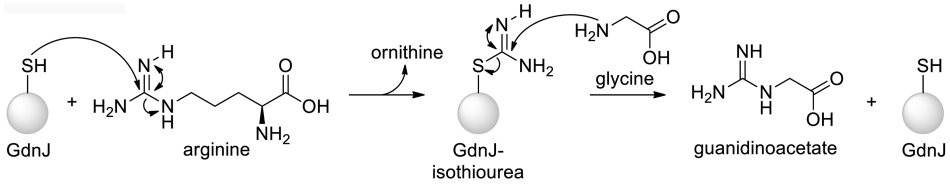
Figure 2. Structural organization of the 51.2 kb guadinomine gene cluster.

dx.doi.org/10.1021/ja308622d | J. Am. Chem. Soc. 2012, 134, 17797−17806

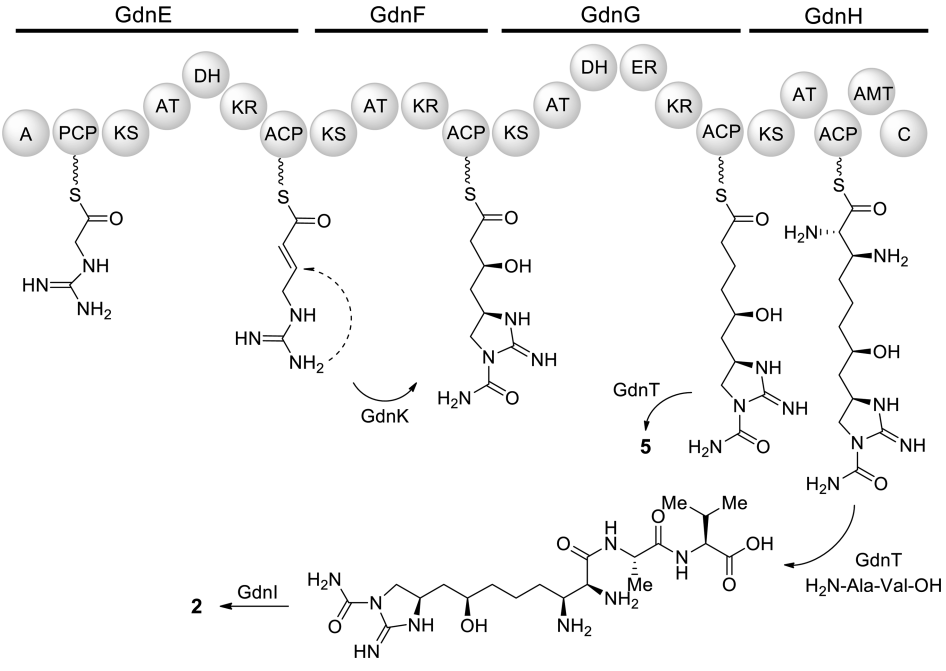
Table 1. Proposed Functions of Open Reading Frames in the Guadinomine Biosynthetic Gene Cluster

gene	size (bp/aa)	homologue (accession number)	% identity/similarity	proposed function
*gdn*A	1125/374	ZmaI (AAR87758)	38/58	ACP dehydrogenase
*gdn*B	1179/392	seryl tRNA synthetase (EIE99121)	46/58	seryl-ACP ligase
*gdn*C	243/80	ZmaD (AAD40107)	39/63	ACP
*gdn*D	870/289	ZmaG (AAR87756)	50/70	ACP dehydrogenase
*gdn*E	7371/2456	MelD(CAD89775)	44/58	NRPS/PKS (A-PCP-KS-AT-DH-KR-ACP)
*gdn*F	4413/1470	NpnA (AEU11005)	44/60	PKS (KS-AT-KR-ACP)
*gdn*G	6300/2099	JamL (AAS98783)	41/58	PKS (KS-AT-DH-ER-KR-ACP)
*gdn*H	6048/2015	McyE (CAE28780)	34/39	NRPS/PKS (KS-AT-ACP-AMT-C)
*gdn*I	792/263	2-oxoglutarate-Fe(II) oxygenase (EAQ64080)	35/55	oxygenase
*gdn*J	1083/360	CyrA/AoaA (ABX60160)	71/83	AGAT
*gdn*K	1728/575	carbamoyltransferase (EDN70596)	55/71	carbamoyltransferase
*gdn*L	969/322	hypothetical protein	35/46	unknown function
*gdn*M	885/294	ABC transporter (EHR60384)	64/79	ABC transporter, permease
*gdn*N	506/252	ABC transporter (EID52982)	60/74	ABC transporter, permease
*gdn*O	954/317	transporter, ATPase (EHR49940)	78/86	ABC transporter, ATPase
*gdn*P	1272/423	carbamoyl phosphate synthetase (EGP34450)	24/42	carbamoyl phosphate synthetase
*gdn*Q	921/306	RNA methylase (ACZ43583)	37/55	rRNA methyltransferase
*gdn*R	1251/416	carbamoyl phosphate synthetase (AEM84317)	40/54	carbamoyl phosphate synthetase
*gdn*S	663/220	4′-phosphopantetheinyl transferase (ADD41185)	47/56	Phosphopantetheinyl Transferase
*gdn*T	783/260	thioesterase (BAF50718)	50/64	thioesterase
*gdn*U	1245/414	biotin carboxylase (EFE81060)	63/72	carbamoyl phosphate synthetase
*gdn*V	1470/489	protein phosphatase (EDY58537)	53/66	protein phosphatase
*gdn*W	666/221	membrane protein (EFL14760)	52/68	integral membrane protein
*gdn*X	1620/539	Lysyl-tRNA synthetase (EFF92412)	69/79	membrane lysyl-tRNA synthetase
*gdn*Y	762/253	two-component system response regulator (AEY87990)	62/74	transcriptional regulator
*gdn*Z	1086/361	two-component system sensor kinase (EFE76207)	51/65	signal transduction kinase

Scheme 1. Proposed Synthesis of Guanidinoacetate Starter Unit

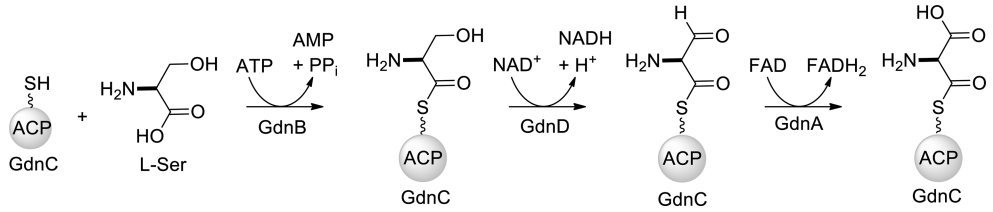


Scheme 2. Proposed Model for Guadinomine Biosynthesis



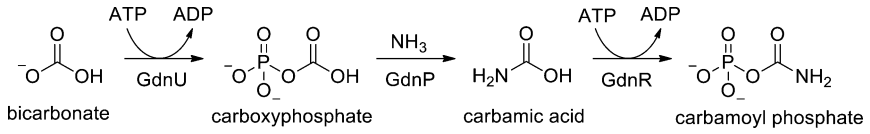
dx.doi.org/10.1021/ja308622d | J. Am. Chem. Soc. 2012, 134, 17797−17806

Scheme 3. Aminomalonyl ACP Biosynthesis[a]



[a]Analogous steps in pathway were biochemically validated by Thomas and co-workers for zwittermicin A.[16]

Scheme 4. Proposed Carbamoyl Phosphate Synthesis from GdnP/R/U



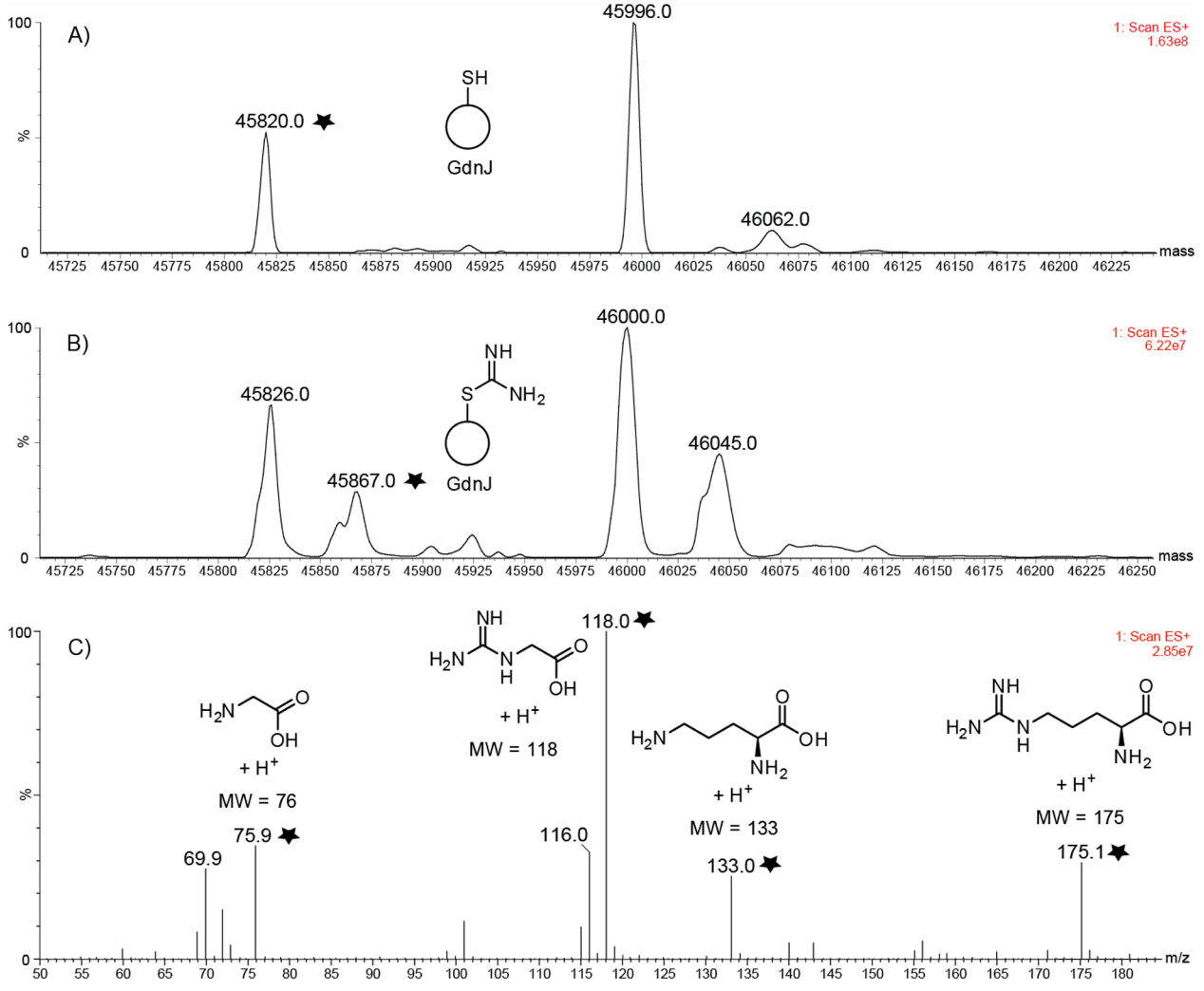


Figure 3. GdnJ functional assays: (A) GdnJ detected at 45820 Da. The ~178 Da shift is a gluconolactone adduct, a commonly seen post-translational modification of His-tagged proteins. (B) GdnJ + Arg + ~42 Da GdnJ–isothiourea intermediate at 45867 Da. (C) GdnJ + Arg + Gly. Guanidinoacetate (m/z = 118) detected in ES+ mode. Ornithine (m/z = 133) byproduct. Glycine (m/z = 76) and arginine (m/z = 175) starting material.

subcloning and PCR amplification to sequence through areas of strong secondary structure.

Sequence Analysis and the Proposed Biosynthetic Pathway. Analysis of DNA sequence from cosmids 7B6 and 4A10 revealed the presence of an nonribosomal peptide synthetase (NRPS)-like loading module and four consecutive PKS modules (Table 1). Additionally, the following tailoring enzymes were identified: a carbamoyltransferase, an arginine:-

dx.doi.org/10.1021/ja308622d | J. Am. Chem. Soc. 2012, 134, 17797–17806

glycine amidinotransferase (AGAT), a pyridoxal phosphate-dependent aminotransferase (AMT), an Fe(II)-dependent oxygenase, and three putative carbamoyl phosphate synthetase subunits. Based on these insights, guadinomine biosynthesis is proposed to begin with the AGAT, GdnJ, which transfers the amidino group from arginine to the acceptor molecule, glycine, forming the starter unit, guanidinoacetate (Scheme 1).[12]

Guanidinoacetate is then activated by the adenylation-peptidyl carrier protein (A-PCP) didomain of GdnE and subsequently transferred to the ketosynthase (KS) domain of module 1, also on GdnE (Scheme 2). There it undergoes condensation with a malonyl extender unit bound to the ACP domain of the same module. The malonyl group is transferred to the ACP by an AT domain with strongest sequence similarity to malonyl-specific AT domains of multimodular PKSs.[13]

GdnE provides insight into the ring closure mechanism that yields the cyclic guanidine. The ketoreductase (KR) and dehydratase (DH) domains reduce the ACP-bound diketide to an α,β-unsaturated diketide that presumably undergoes cyclization with the nucleophilic guanidino group. This cyclization is similar to that seen in module 1 of cylindrospermopsin biosynthesis, which is the only other metabolite known to use guanidinoacetate as a starter unit.[14]

Module 2, encoded by GdnF, incorporates the C-7-hydroxyl group common to guadinomines. Module 3, encoded by GdnG, incorporates a fully reduced ketide unit. This module is also the point at which the shunt product guadinomic acid can be formed, potentially by the thioesterase (GdnT) acting prematurely on the ACP3-bound substrate. As expected, both modules 2 and 3 have AT domains that are homologous to malonyl-specific AT domains.

The fourth module, GdnH, is a minimal PKS that uses the uncommon extender unit aminomalonyl-ACP. This amino-malonyl-specific AT of GdnH shares the closest sequence similarity to the malonyl specific AT of JamM from the jamaicamide gene cluster.[15] The guadinomine cluster has a dedicated set of genes for aminomalonyl-ACP formation (*gdn*ABCD), which are homologous to the four zwittermicin A genes (ZmaIJHG).[16] Serine is presumably activated by GdnB and ligated to GdnC, forming a seryl−ACP intermediate (Scheme 3).

Interestingly, GdnB shares greater homology with seryl tRNA-synthetases than the ZmaJ homologue. Aminoacyl tRNA-synthetases have been shown to acylate carrier proteins and fittingly have been renamed amino acid:[carrier protein] ligases (AMP forming), thus it functions as a standalone A domain.[17] The seryl-ACP then undergoes successive dehydrogenations via GdnA and GdnD to give aminomalonyl-ACP. Incorporation of this extender unit into the polyketide backbone is likely to yield an ACP-bound α-amino-β-ketothioester that subsequently undergoes reductive amination via the AMT of GdnH. Interestingly, there are only a few examples in literature of aminotransferases within a PKS module.[18] The last 1000 residues of GdnH (AMT-C) show strong end-to-end similarity with the corresponding regions of the microcystin and mycosubtilin synthases, McyE and MycA, respectively. The significance of these similarities with regard to chain termination and Ala-Val dipeptide formation will be discussed later.

Finally, the tailoring enzymes GdnK and GdnI facilitate carbamoylation of the cyclic guanidinyl moiety and C-6 hydroxylation, respectively. A third hydroxyl group is found in **1** at C-8 and could also be installed by GdnI or potentially by an enzyme not part of the gene cluster.

Carbamoyl Phosphate Synthesis. Three enzymes with ATP grasp domains (GdnP, GdnR, GdnU) were identified with homology to the three subdomains of the mitochondrial carbamoyl phosphate synthetase.[19] We propose that these enzymes are responsible for forming carbamoyl phosphate that is used by GdnK to carbamoylate the cyclic guanidine. GdnU may form carboxyphosphate from bicarbonate and ATP, with GdnP forming carbamic acid from carboxyphosphate and ammonia. Finally, GdnR may form carbamoyl phosphate from carbamic acid and ATP (Scheme 4).

Functional Analysis of the AGAT (GdnJ). When arginine was incubated with the 45.8 kDa purified GdnJ (Figure 3A, Figure S1A), formation of a GdnJ−isothiourea intermediate was detected at +42 Da (Figure 3B). The +42 Da intermediate peak was lost upon addition of glycine, with concomitant formation of guanidinoacetate, which was observed by formation of a 118 Da peak in ES+ mode (Figure 3C). Kinetic parameters were established for GdnJ, enabling comparison to other characterized amidinotransferases. GdnJ has a k_{cat}/K_m of 18.7 and 17.3 M^{-1} s^{-1} for arginine and glycine, respectively. These kinetic values suggest the guadinomine AGAT is more efficient than the corresponding cylindrospermopsin and human enzymes (Figure S1B, Tables S2, S3).

Functional Analysis of the PKS Module 1 (GdnE). As a control, [2-^{14}C] malonyl-CoA was incubated with purified GdnE (Figure S2A). After 35 min, labeling of GdnE indicated that the AT domain loaded malonyl-CoA and transferred it to the ACP. The 150 mm position on the trace shown in Figure 3A corresponded to the location of the 260 kDa GdnE polypeptide on the SDS-polyacrylamide gel.

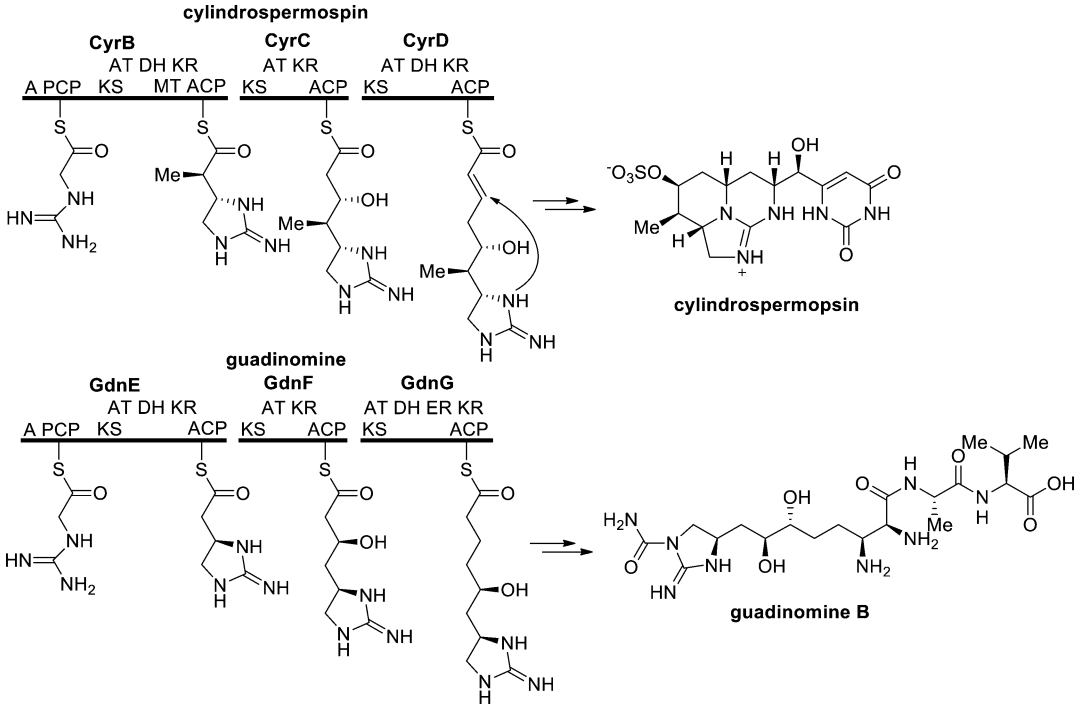
Incubation of [2-^{14}C] glycine, arginine, and GdnJ with GdnE, ATP, and $MgCl_2$ did not lead to labeling (Figure S2B); however, upon addition of unlabeled malonyl-CoA and NADPH to the reaction mixture, GdnE was labeled (Figure 3B), confirming that GdnE was indeed functional. In both cases, guanidinoacetate was likely activated by the A domain and transferred to the KS. However, when malonyl-CoA was absent, the substrate was rapidly hydrolyzed due to the instability of the guanidinoacetyl−PCP and/or guanidinoacetyl−KS intermediates. Once malonyl-CoA was added, decarboxylative acylation proceeded, and the ACP was labeled.

In order to confirm that A domain activity was required for GdnE labeling with an elongated guanidinoacetate starter unit, radiolabeled glycine, arginine, GdnJ, $MgCl_2$, NADPH, and malonyl-CoA were incubated without ATP. As expected, GdnE was not labeled in this scenario (Figure S2C), indicating guanidinoacetate was not directly loaded onto the initiation module.

Like the CyrB A domain, the amino acid activated by the GdnE A domain cannot be predicted by available NRPS databases based on its specificity conferring residues,[20] suggesting that the A domain is not specific to 1 of the 20 common amino acids. Combined with the radiolabeling studies, this evidence indicates that we have identified the second NRPS that activates guanidinoacetate. Furthermore, the results of the GdnJ activity assay combined with the GdnE radiolabeling assay biochemically establish the first two steps of our proposed guadinomine biosynthetic pathway.

Guadinomine Detection by Mass Spectrometry. LC-MS and HRMS analysis of *Streptomyces* sp. K01-0509 fermentation broth indicated ions corresponding to masses

Scheme 5. Comparison of the First Three Modules in *cyr* and *gdn* Biosynthetic Pathways



consistent with those expected for guadinomines: **1** (m/z = 519), **2** (m/z = 503), and **5** (m/z = 259), (Figures S3A wild-type, S4A-C, S5A-E, Table S4). The retention times were as follows: guadinomic acid (~4.8 min), guadinomine A (~12.5 min), and guadinomine B (~13.5 min). MS^2 spectra for guadinomic acid, and guadinomines A and B showed a loss of 43 Da, which corresponds to loss of the carbamoyl group (Figure S4D-F). MS^3 spectra for guadinomines A and B indicate that the decarbamoylation peaks give rise to further losses of 117 and 262 Da (Figure S4G,H). Loss of 117 Da corresponds to cleavage of valine, and loss of 262 Da corresponds to cleavage of the diamine followed by iminium formation.

Verifying the Identity of the Guadinomine Gene Cluster via Targeted Disruption. A derivative of the pJQ200SK disruption vector, containing an internal fragment from the KS portion of *gdn*E, was constructed. Conjugation of an *E. coli* strain harboring this plasmid with wild-type *Streptomyces* sp. K01-0509 yielded a gentamicin-resistant mutant (Figure S3B). Mutants resulting from homologous recombination should produce a 1286 bp amplimer with the T3/F3 primer pair and a 1261 bp amplimer with the T7/R3 primer pair. Amplimers of the expected sizes were identified for the mutant but not for the wild-type *Streptomyces* strain or mutants with random insertion of the pJQ200SK vector. The T7/R3 amplimer of the correct size was then digested with *Bam*HI and *Xba*I, rendering the expected 1079 bp fragment (Figure S3C).

The knockout mutant was cultured in 500 mL R5 media with gentamicin for 5 days, and the culture was extracted for identification of guadinomines. LC-MS analysis of the mutant strain showed loss of all characteristic guadinomine peaks (Figure S3A K01 mutant), confirming the identity of the putative *gdn* cluster. Specifically, loss of the 4.7 min guadinomic acid peak and both 12.5−13.5 min guadinomine A and B peaks was observed.

DISCUSSION

In summary, we have cloned the gene cluster responsible for guadinomine biosynthesis in *Streptomyces* sp. K01-0509. Two features of the cluster are not intuitive and deserve further discussion: (1) A dedicated Ala-Val forming NRPS module has not been identified in the 51.2 kb guadinomine cluster or in the regions immediately surrounding the primary cluster (5 kb preceding 7B6 and 1 kb following 4A10); and (2) the terminal NRPS module consists of a C domain without the corresponding A and PCP domains.

Proposed Ala-Val Dipeptide Origin and Chain Termination. While it is uncommon for natural biosynthetic pathways to be encoded by multiple gene clusters, there is precedent among NRPSs for terminal peptides being encoded outside of the primary cluster. For example, in the virginiamycin M cluster, the NRPS for the terminal Pro residue was not identified and was hypothesized to be located further downstream.[21] In the pristinamycin cluster, the peptide synthetase, *sna*D, is located 20 kb away from the primary cluster.[22]

With regard to the terminal C domain, the AMT-C didomain of GdnH shares the closest end-to-end sequence similarity to the corresponding portions of the microcystin (McyE) and mycosubtilin (MycA) synthetases. These account for two of the very rare instances of an AMT within a PKS module. Interestingly, like the *gdn* cluster, each of these enzymatic assembly lines also has a terminal C domain without accompanying A-PCP domains. In microcystin biosynthesis, the C domain is proposed to form the peptide bond between D-glutamate and *N*-methyl-dehydroalanine.[23] Although the activity of the corresponding C domain was not detailed in Walsh and co-workers' study of mycosubtilin biosynthesis, there too it appears that the domain does not catalyze peptide bond formation but rather helps transfer a β-aminoacyl thioester to the thiol of a downstream PCP.[24] This is in line with the normal function of a C domain, which must bring two

dx.doi.org/10.1021/ja308622d | J. Am. Chem. Soc. 2012, 134, 17797−17806

PCPs together to form a peptide bond. In the cases of GdnH and MycA, we speculate that the downstream PCP is vacant, leading to net transacylation of the β-aminoacyl thioester along the assembly line.

Combining our knowledge of terminal peptide synthesis with the various functions ascribed to terminal C domains, we speculate that there exists a bimodular NRPS responsible for synthesizing the Ala-Val dipeptide. Like MycA, the bimodular NRPS likely contains an N-terminal PCP domain, which receives the guadinomine polyketide precursor from the C domain of GdnH, followed by extension with the two peptides, with GdnT hydrolyzing the final guadinomine product. Alternatively, the C domain may function like McyE to form a peptide bond between the polyketide and a pre-existing Ala-Val dipeptide.

Comparison of Cylindrospermopsin and Guadinomine Clusters. Of particular interest when analyzing the *gdn* gene cluster is its striking similarities and contrasts to the first three modules of the cylindrospermopsin (*cyr*) biosynthetic gene cluster (Scheme 5).

Cylindrospermopsin is a mammalian toxin produced by the cyanobacterium, *Cylindrospermopsis raciborskii* AWT205.[14] The amidinotransferases (CyrA and GdnJ) in the two clusters share 75% amino acid sequence identity, and the first three PKS modules are approximately 40% identical, with the exception that the *cyr* cluster includes a methyltransferase in module 1 (CyrB) and lacks an enoylreductase (ER) domain in module 3 (CyrD) (Table S5). The similarity of the two clusters suggests an uncannily close evolutionary relationship.

While the first two modules are extremely similar in sequence and function, they yield cyclic guanidines (module 1) and C-7 hydroxyls (module 2) with opposite stereochemistry. Guadinomines contain the (*R*)-cyclic guanidine, whereas cylindrospermopsin the (*S*)-cyclic guanidine, as determined by total synthesis of both compounds and an X-ray crystal structure for cylindrospermopsin.[25−27] That only one stereoisomer is seen for guadinomines (rather than a 1:1 mixture) suggests that residues within GdnE catalyze the cyclization. The *cyr* and *gdn* KR domains of module 2 also give rise to opposite configurations of the alcohol. Interestingly, both KR domains contain the Leu-Asp-Asp (LDD) peptide motif, which typically gives rise to (*S*) stereoisomers,[28] however, only guadinomine has the expected (*S*)-configured alcohol. The presence of the methyl group in cylindrospermopsin may factor into how the acyl chain is held in the KR active site leading to the unexpected (*R*) stereochemistry.

Module 3 represents the critical point at which these pathways diverge, as the lack of an ER domain in CyrD leaves an α,β-unsaturated diketide that undergoes cyclization with the cyclic guanidine. After a fourth PKS module (CyrE), having no homologue in the guadinomine gene cluster, yet another cyclization occurs, leading to the tricyclic carbon skeleton. Whereas in guadinomine biosynthesis, the guanidine is used as a chemical handle for cyclization and carbamoylation; cylindrospermopsin uses the same moiety for multiple ring-closing reactions.

When comparing the AGAT's from the two clusters (GdnJ and CyrA), the formation and detection of a GdnJ−isothiourea intermediate along with kinetic data allowed us to deduce the mechanism for GdnJ as a ping-pong or two-step mechanism (Scheme 1), which is common in mammalian AGAT's.[12] In this scheme, the enzyme first binds arginine, transferring the amidino group to the active site cysteine to form the GdnJ−isothiourea intermediate, followed by binding the nucleophile, glycine, which receives the amidino group. This is in contrast to CyrA from cylindrospermopsin that functions by a mechanism consistent with binding both substrates simultaneously.[29]

Aminomalonyl-ACPs: AT Domain Specificity and Potential for Engineering. AT domains play a major role in the structural diversity of polyketides, as they are the principal gatekeepers of extender unit specificity. With the identification of aminomalonyl-ACP biosynthetic enzymes, the ability to modify AT substrate specificity in order to incorporate aminomalonyl(AM)-ACP (and the closely related hydroxymalonyl(HM)-ACP) should be investigated. AM- and HM-ACP elongate polyketide chains with glycyl or glycolyl units, respectively, with the α-carbon substituents (amine and hydroxyl groups) allowing functionality and hydrogen-bonding potential not available with more conventional malonyl- or methylmalonyl-CoA extender units.[30] These more reactive substituents also allow for simpler semisynthetic approaches to couple with combinatorial biosynthetic strategies.

To understand the evolutionary relationship between the aminomalonyl-specific AT's from the *zma* and *gdn* clusters, they were included in a phylogenetic tree consisting of AT domains with specificity for malonyl- , methylmalonyl- , and ethylmalonyl-CoA and methoxylmalonyl-ACP (Figure 4). The tree illustrates that the two aminomalonyl-AT domains (ZmaF and GdnH) cluster closely, along with the ZmaA hydroxymalonyl AT, in the malonyl-specific AT subfamily.

■ CONCLUSION

Guadinomines are the most potent natural products with anti-TTSS activity, and future studies geared toward understanding their precise mode of action will be important. Quite intriguing is the evolutionary and structural similarity of the TTSS and the bacterial flagellum.[31] Studies aimed at understanding any potential link between these two apparatuses and natural product TTSS inhibitors would be very informative.

At the point the full guadinomine synthetase is cloned into an appropriate shuttle vector for heterologous expression and metabolite production, control of guadinomine production as well as several interesting analogs will be within reach. Given that the Ala-Val NRPS has yet to be identified, the cloned 51.2 kb cluster should minimally produce the guadinomine polyketide precursor. Coupling this precursor to commercially available Ala-Val, or other dipeptides, would yield guadinomines and analogs. Such a semisynthetic approach would represent the beginning of producing analogs and analyzing those compounds in TTSS inhibition assays.

■ MATERIALS AND METHODS

Bacterial Culture Conditions. *Streptomyces* sp. K01-0509 was cultured on semisolid R5 media to prepare spore stocks. Spores were grown under two conditions for genomic DNA isolation: (1) in liquid R5 media for whole genome shotgun sequencing; and (2) in liquid YEME media for cosmid library preparation. Both media were prepared according to standard protocols.[32]

Cosmid Library Construction. A cosmid library of *Streptomyces* sp. K01-0509 was prepared using the SuperCos I vector from Stratagene. A library of 30−40 kb fragments was constructed according to the manufacturer's protocol using genomic DNA partially digested with *Mbo*I and dephosphorylated with calf intestine alkaline phosphatase. Ligation reactions were packaged into λ phage for *E. coli* infection. The packaged library was inoculated on semisolid LB media. Approximately 1500 single colonies were picked and inoculated into individual wells of fifteen 96-well plates. This arrayed cosmid

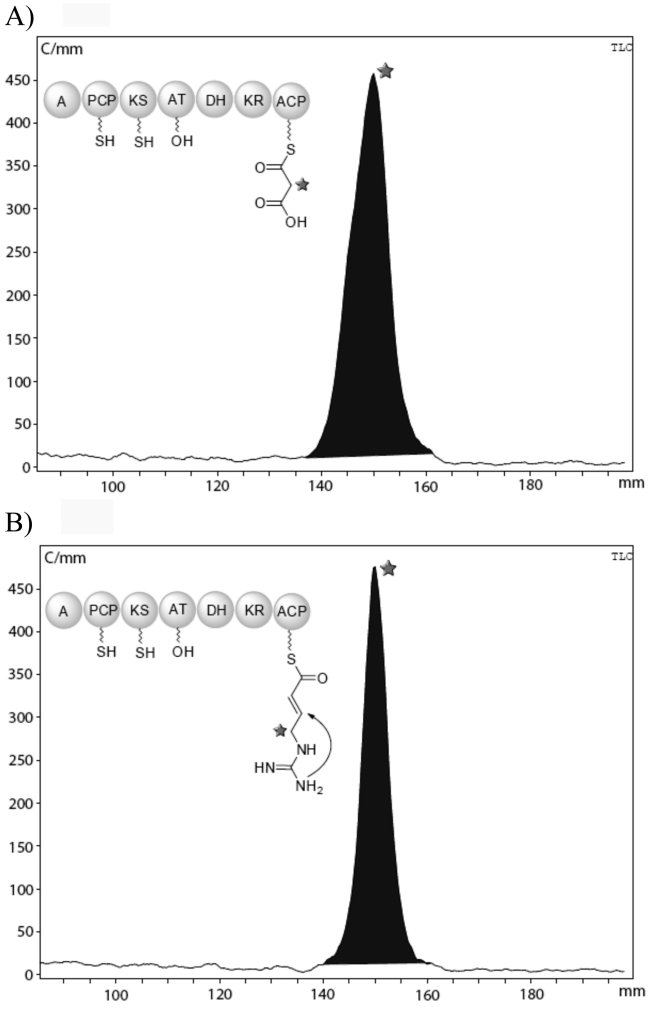


Figure 4. GdnE radiolabeling. Reactions A and B with corresponding structure of expected modification. The position of radiolabeled isotope is indicated with an asterisk. (A) [2-^{14}C] malonyl-CoA + GdnE. (B) [2-^{14}C] Gly + Arg + GdnJ + GdnE + ATP + $MgCl_2$ + NADPH + malonyl-CoA.

library was stored in liquid LB media with 20% glycerol and kept at −80 °C.

Paired-End DNA Library Construction. A paired-end library was constructed for the *Streptomyces* sp. K01-0509 genomic DNA for Illumina whole genome shotgun sequencing following the manufacturer's protocol. The genomic DNA was first fragmented to less than 800 bp using a nebulizer, modified by ligating paired-end adapters, and enriched through PCR amplification to create a DNA library. The quality of the library DNA was verified by agarose gel electrophoresis.

Illumina DNA Sequencing and de Novo Assembly. Illumina sequencing-by-synthesis technology was used for whole genome sequencing of the *Streptomyces* sp. K01-0509 paired-end genomic DNA library. Approximately 35 million reads were obtained, each 65 bp in length, for a total 2.25 Gbp of DNA sequence, corresponding to approximately 280-fold coverage of the genome. Paired-read iterative contig extension (PRICE) *de novo* genome assembly software was used to assemble the genome. Contigs from the assembly were used to develop oligonucleotide probes for screening the cosmid library by PCR amplification to identify the guadinomine gene cluster. The nucleotide sequence of the guadinomine gene cluster has been deposited in GenBank under accession number JX545234.

Cloning and Heterologous Protein Expression. Oligonucleotide primers, engineered with *Nde*I and *Eco*RI restriction sites, were designed to amplify individual guadinomine genes (Table S6). PCR-amplified fragments were first cloned into the pCR-Blunt vector, with the desired inserts being released by double digest with *Nde*I and *Eco*RI. The fragments were then ligated into either the pET28 or pET21 expression vector (Novagen), rendering N- and C-terminal hexahistidine tags, respectively. Following transformation into the BL21-derived *E. coli* BAP1 strain for protein expression,[33] single colonies were inoculated into 5 mL seed cultures with the appropriate antibiotic and incubated at 37 °C overnight. The seed cultures were used to inoculate 500 mL cultures and grown at 37 °C to an OD_{600} between 0.4 and 0.6. Cultures were placed on ice for 15 min, and protein expression was induced with 100 μM isopropyl β-D-1-thiogalactopyranoside (IPTG), followed by incubation in an 18 °C shaker for 8−12 h. Cultures were harvested, lysed by sonication, and purified on a Ni-NTA agarose gravity flow chromatography column. Further purification was performed by fast protein liquid chromatography (FPLC) using a 5 mL HiTrap Q Sepharose ion exchange column.

Amidinotransferase Enzymatic Assay. For 20 h at room temperature in 150 μL, pH 7 phosphate buffer, 28 μM GdnJ was incubated with 11.5 mM arginine and 26.6 mM glycine. Reactions were analyzed for guanidinoacetate formation on a Waters 2795 HPLC system with dual wavelength UV detector and ZQ single quadrupole MS with electrospray ionization source. To establish kinetic parameters, arginine and glycine concentrations were varied from 2 to 16 mM and 1.25 to 10 mM, respectively, while holding the concentration of the other constant. Reactions were run for 15 min at 37 °C, and production of the byproduct, ornithine, was detected via ninhydrin derivatization and monitored at 505 nm absorbance following the method of van Pilsum.[34] Ornithine production correlates to guanidinoacetate production in 1:1 molar stoichiometric relationship. Prism, a nonlinear curve fitting program, was used to determine V_{max}, k_{cat} and K_m, using the relationship: $V = V_{max}*A*B/(K_{m(B)}*A + K_{m(A)}*B + A*B)$, where A corresponds to arginine and B to glycine.

GdnE Radiolabeling Assay. [2-^{14}C] malonyl-CoA was purchased from American Radiolabeled Chemicals, and [2-^{14}C] glycine was purchased from Moravek Biochemicals and Radiochemicals. To study the ability of GdnE to accept the proposed extender unit, malonyl-CoA, GdnE (22 μM) was reacted with [2-^{14}C] malonyl-CoA (45 μM), ATP (10 mM), 5.5 mM $MgCl_2$ (10 mM), and NADPH (1.25 mM). The reaction mixture was incubated at room temperature for 35 min, and the protein was separated by SDS-PAGE. To quantify radioactivity, the gel was scanned on a radio isotope thin layer analyzer (Rita Star; Raytest GmbH). Data was processed with Gina Star software.

To study the ability of GdnE to activate the proposed starter unit, guanidinoacetate, and condense with malonyl-ACP, guanidinoacetate was produced *in situ* by incubating GdnJ (27 μM) with [2-^{14}C] glycine (53 μM), arginine (0.28 mM), ATP (10 mM), 5.5 mM $MgCl_2$ (10 mM), NADPH (1.25 mM), and malonyl-CoA (0.25 mM) for 15 min. To this reaction was added GdnE (22 μM). ATP and $MgCl_2$ are required for the enzymatic activity of the A domain, and NADPH is required for the reducing activity of the KR domain. The reaction mixture was incubated at room temperature for 35 min, and the protein was separated by SDS-PAGE. To quantify radioactivity, the gel was scanned on a radio isotope thin layer analyzer (Rita Star; Raytest GmbH). Data were processed with Gina Star software.

Analysis of Guadinomine Production by LC-MS. *Streptomyces* sp. K01-0509 was cultured in 500 mL R5 liquid media for 5 days at 30 °C. Thereafter, 500 mL acetone was added to the cultures. The mixture was centrifuged to separate mycelium from supernatant, and the acetone evaporated *in vacuo*. The crude extract was passed through a previously activated Dowex 50W x2 [H+] cation exchange column, washed with water, and eluted with 1.5 N NH_4OH. The eluent was neutralized with HCl and lyophilized yielding a salt mixture consisting of the active guanidine-containing compounds. The mixture was resuspended in a minimal amount of water for liquid chromatography mass spectrometry (LC-MS) analysis. Sample extracts were analyzed on a Varian Polaris 5 μM C18-A, 250 × 2.1 mm, HPLC column. The mobile phase consisted of solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile). A gradient elution from 0% to 95% B was used over 31 min at a flow rate of 0.25 mL/min.

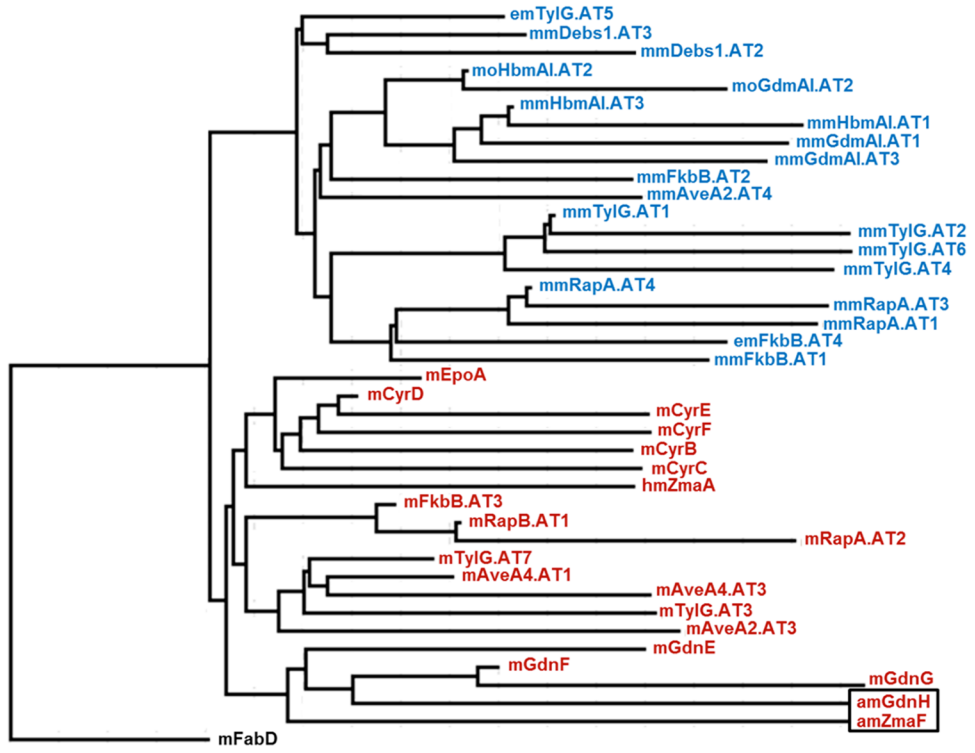


Figure 5. Phylogenetic tree of AT domain sequences. Lowercase letter in front of domain indicates AT specificity (m = malonyl-CoA, mm = methylmalonyl-CoA, mo = methoxymalonyl-ACP, em = ethylmalonyl-CoA, hm = hydroxymalonyl-ACP, am = aminomalonyl-ACP). Blue lettering indicates subfamily 1 (mm, mo, em) and red lettering indicates subfamily 2 (m, hm, am).

Sample analysis were carried out by positive ion ESI-LC/MS using an Agilent 1100 HPLC and LTQ XL ion trap mass spectrometer (Thermo Fisher Scientific) with data-dependent acquisition in dynamic exclusion mode. MS/MS (MS^2) and MS/MS/MS (MS^3) data were generated for the compounds of interest. Full scan high resolution mass spectrometry (HR-MS) traces were collected on an Agilent 1260 HPLC coupled with Bruker MicroTOF-Q II.

Construction of Gene Disruption Vectors and Conjugation with *Streptomyces* sp. K01-0509. To construct the PKS disruption mutant, two primer combinations (F1/R1, F2/R2, Table S7) were used to amplify internal fragments (KS1 and KS2, respectively) from the KS gene of module 1 (*gdn*E). The internal fragments were first cloned into the pCR Blunt vector then transferred to the conjugation vector, pJQ200SK, at the *Xba*I/*Bam*HI cloning site, producing pJQ200SK-KS. pJQ200SK-KS constructs were transformed into the *E. coli* strains, ET12567 [pUZ8002], or S17-1, and positive transformants were selected on LB agar supplemented with gentamicin, chloramphenicol, kanamycin (for ET12567) and gentamicin, and spectinomycin (for S17−1).

The plasmid was then conjugally transferred to *Streptomyces* sp. K01-0509. To accomplish this, a single ET12567, or S17-1, colony was used to inoculate a 5 mL LB culture with the same antibiotics and grown overnight in a 37 °C shaker. A fresh 5 mL LB culture was inoculated with 200 μL of the overnight culture and grown to OD ∼ 0.7−1.0. At the same time, *Streptomyces* sp. K01-0509 cells were collected from semisolid soy-flour and mannitol (SFM) media and transferred to 2 mL TES buffer, heat shocked for 10 min at 50 °C, followed by cooling in a room temperature water bath. Two mL 2xYT media was added, and cells were incubated at 37 °C for 2 h. TES buffer, SFM media, and 2xYT media were prepared according to standard protocols.[32] Both *E. coli* and K01-0509 cells were harvested by centrifugation and resuspended in the residual media. The cells were mixed and plated on SFM plates without antibiotics and incubated at 30 °C. After 18 h, the plates were overlaid with 2 mL of water containing 50 μg/mL nalidixic acid and 50 μg/mL gentamicin , and incubated at 30 °C until the appearance of ex-conjugants. Colonies were restreaked on R5 plates with both antibiotics. Genomic DNA was prepared from drug-resistant K01-0509 colonies that were grown in R5 liquid media with gentamicin, and cointegration of pJQ200SK-KS was confirmed by PCR analysis. Primers flanking the internal fragment (F3 and R3, Table S6) along with T7 and T3 primers from the pJQ200SK vector were used for verifying mutant strains. Figure S3B illustrates the expected disruption mutant resulting from homologous recombination between pJQ200SK-KS and the wild-type K01-0509 strain.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The SI includes details of the BLAST analysis performed on assembled K01 contigs. Figure S1, GdnJ purified protein and kinetic assay plot. Figure S2, GdnE assay. Figure S3, LC-MS comparison of wildtype and mutant; gene disruption illustration. Figure S4, LC-MS analysis of K01 wildtype. Figure S5, HRMS analysis of K01 wildtype. Table S1, S6, S7, lists of primers. Table S2, GdnJ assay data. Tables S3, GdnJ kinetics. Table S4, HRMS data. Table S5, *cyr* and *gdn* comparison. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

khosla@stanford.edu

Author Contributions

▽These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Peter Bernhardt and Bradley Moore at The University of California, San Diego for their assistance in

preparing the paired-end genomic DNA library. We also thank Genetic Chemistry Inc. in Palo Alto, CA for access to an automated colony-picking instrument for arraying the cosmid library. Tiangang Liu provided helpful discussions and insight for gene cluster identification. We especially thank Ludmila Alexandrova at the Stanford University Mass Spectrometry (SUMS) laboratory for help with analyzing wild-type and mutant strains by LC-MS, MS-MS, and HRMS. This research was supported by a grant from the NIH (R01 CA077248 to C.K.). T.C.H. is a recipient of the Diversifying Academia, Recruiting Excellence (DARE) Doctoral Fellowship at Stanford University, the Bell Laboratories Graduate Research Fellowship, and the National Science Foundation Graduate Research Fellowship. A.E.M. is a recipient of a National Institute of General Medical Sciences Postdoctoral Fellowship (GM099248-01) and is a fellow of the Center for Molecular Analysis and Design (CMAD) at Stanford University.

REFERENCES

(1) Ōmura, S. *Jpn. J. Bacteriol.* **1999**, *54*, 795−813.
(2) Kimura, K.; Iwatsuki, M.; Nagai, T.; Matsumoto, A.; Takahashi, Y.; Shiomi, K.; Ōmura, S.; Abe, A. *J. Antibiot.* **2010**, *64*, 197−203.
(3) Ghosh, P. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 771−795.
(4) Cornelis, G. R. *J. Cell. Biol.* **2002**, *158*, 401−408.
(5) Kim, J. G.; Roden, J. A.; Taylor, K. W.; Aakre, C. D.; Su, B.; Lalonde, S.; Kirik, A.; Chen, Y.; Baranage, G.; McLane, H.; Martin, G. B.; Mudgett, M. B. *Plant Cell* **2009**, *21*, 1305−1323.
(6) Wilharm, G.; Lehmann, V.; Neumayer, W.; Trcek, J.; Heesemann, J. *BMC Microbiol.* **2004**, *4*, 27.
(7) Dacanay, A.; Knickle, L.; Solanky, K. S.; Boyd, J. M.; Walter, J. A.; Brown, L. L.; Johnson, S. C.; Reith, M. *Microbiology* **2006**, *152*, 1847−1856.
(8) Zhou, X.; Konkel, M. E.; Call, D. R. *Microbiology* **2009**, *155*, 837−851.
(9) Iwatsuki, M.; Uchida, R.; Yoshijima, H.; Ui, H.; Shiomi, K.; Matsumoto, A.; Takahashi, Y.; Abe, A.; Tomoda, H.; Ōmura, S. *J. Antibiot.* **2008**, *61*, 222−229.
(10) Iwatsuki, M.; Uchida, R.; Yoshijima, H.; Ui, H.; Shiomi, K.; Kim, Y.-P.; Hirose, T.; Sunazuka, T.; Abe, A.; Tomoda, H.; Ōmura, S. *J. Antibiot.* **2008**, *61*, 230−236.
(11) Kevany, B.; Rasko, D.; Thomas, M. G. *Appl. Environ. Microb.* **2009**, *75*, 1144−1155.
(12) Humm, A.; Fritsche, E.; Steinbacher, S.; Huber, R. *EMBO J.* **1997**, *16*, 3373−3385.
(13) Ridley, C. P.; Lee, H. Y.; Khosla, C. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4595−4600.
(14) Mihali, T. K.; Kellmann, R.; Muenchhoff, J.; Barrow, K. D.; Neilan, B. A. *Appl. Environ. Microb.* **2008**, *74*, 716−722.
(15) Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M. A.; Gerwick, W. H. *Chem. Biol.* **2004**, *11*, 817−833.
(16) Chan, Y. A.; Boyne, M. T., II; Podevels, A. M.; Klimowicz, A. K.; Handelsman, J.; Kelleher, N. L.; Thomas, M. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 14349−14354.
(17) Mocibob, M.; Ivic, N.; Bilokapic, S.; Maier, T.; Luic, M.; Ban, N.; Weygand-Durasevic, I. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 14585−14590.
(18) Rounge, T. B.; Rohrlack, T.; Nederbragt, A. J.; Kristensen, T.; Jakobsen, K. S. *BMC Genomics* **2009**, *10*, 396.
(19) *Biochemistry*, 6th ed.; Berg, J. M., Tymoczko, J. L., Stryer, L., Eds.; W. H. Freeman and Co: New York, 2006.
(20) Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. *Nucleic Acids Res.* **2005**, *33*, 5799−5808.
(21) Pulsawat, N.; Kitani, S.; Nihira, T. *Gene* **2007**, *393*, 31−42.
(22) Bamas-Jacques, N.; Lorenzon, S.; Lacroix, P.; De Swetschin, C.; Crouzet, J. *J. Appl. Microbiol.* **1999**, *87*, 939−948.
(23) Tillett, D.; Dittmann, E.; Erhard, M.; von Döhren, H.; Börner, T.; Neilan, B. A. *Chem. Biol.* **2000**, *7*, 753−764.
(24) Aron, Z. D.; Dorrestein, P. C.; Blackhall, J. R.; Kelleher, N. L.; Walsh, C. T. *J. Am. Chem. Soc.* **2005**, *127*, 14986−14987.
(25) Hirose, T.; Sunazuka, T.; Tsuchiya, S.; Tanaka, T.; Kojima, Y.; Mori, R.; Iwatsuki, M.; Ōmura, S. *Chem.—Eur. J.* **2008**, *14*, 8220−8238.
(26) Xie, C.; Runnegar, M. T. C.; Snider, B. B. *J. Am. Chem. Soc.* **2000**, *122*, 5017−5024.
(27) Heintzelman, G. R.; Fang, W. K.; Keen, S. P.; Wallace, G. A.; Weinreb, S. M. *J. Am. Chem. Soc.* **2002**, *124*, 3939−3945.
(28) Caffrey, P. *ChemBioChem* **2003**, *4*, 649−662.
(29) Muenchhoff, J.; Siddiqui, K. S.; Poljak, A.; Raftery, M. J.; Barrow, K. D.; Neilan, B. A. *FEBS J.* **2010**, *277*, 3844−3860.
(30) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. *Nat. Prod. Rep.* **2009**, *26*, 90−114.
(31) Blocker, A.; Komoriya, K.; Aizawa, S.-I. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3027−3030.
(32) *Practical Streptomyces Genetics*; Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., Hopwood, D. A., Eds.;Crowes: Norwich, U.K., 2000.
(33) Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H; Cane, D. E.; Khosla, C. *Science* **2001**, *291*, 1790−1792.
(34) van Pilsum, J. F.; Taylor, D.; Zakis, B.; McCormick, P. *Anal. Biochem.* **1970**, *35*, 277−286.