Identification of the thiazolyl peptide GE37468 gene cluster from *Streptomyces* ATCC 55365 and heterologous expression in *Streptomyces lividans*

Travis S. Young and Christopher T. Walsh

Harvard Medical School, Armenise D1, Room 608, 240 Longwood Avenue, Boston, MA 02115

Contributed by Christopher T. Walsh, June 29, 2011 (sent for review June 6, 2011)

Thiazolyl peptides are bacterial secondary metabolites that potently inhibit protein synthesis in Gram-positive bacteria and malarial parasites. Recently, our laboratory and others reported that this class of trithiazolyl pyridine-containing natural products is derived from ribosomally synthesized preproteins that undergo a cascade of posttranslational modifications to produce architecturally complex macrocyclic scaffolds. Here, we report the genetic cluster responsible for production of the elongation factor Tu (EF-Tu)-targeting 29-member thiazolyl peptide GE37468 from *Streptomyces* ATCC 55365 and its heterologous expression in the model host *Streptomyces lividans*. GE37468 harbors an unusual β-methyl-β-hydroxy-proline residue that may increase conformational rigidity of the macrocycle and impart reduced entropic costs of target binding. Isotope feeding and gene knockout were employed in the engineered *S. lividans* strain to identify the P450 monoxygenase GeTj as the enzyme involved in posttranslational transformation of isoleucine 8 to β-methyl-β-hydroxy-proline through a predicted tandem double hydroxylation/cyclization mechanism. Loss of Ile8 oxygenative cyclization or mutation of Ile8 to alanine via preprotein gene replacement resulted in a 4-fold and 2-fold drop in antibiotic activity, respectively. This report of genetic manipulation of a 29-member thiazolyl peptide sets the stage for further genetic examination of structure activity relationships in the EF-Tu targeting class of thiazolyl peptides.

bacterial resistance | bioengineering | thiopeptide

---

Thiazolyl peptide, or thiopeptide, antibiotics are the most heavily posttranslationally modified ribosomal natural products known to date (1–3). More than a dozen posttranslational modifications transform linear 50–60 residue preproteins, consisting of only canonical amino acids, to molecules rich in thiazole and dehydro amino acids with a pyridine/piperidine core. These rigid scaffolds provide potent inhibition of ribosomal protein synthesis in a broad range of Gram-positive bacteria and malarial parasites. In particular, their activity against Methicillin-resistant *Staphylococcus aureus* (MRSA) has made them attractive candidates for development as clinical agents to fight emerging resistance to the currently used antimicrobial repertoire. The recent discovery that thiazolyl peptides can function as antitumor antibiotics (4) has also heightened interest in exploration of these compounds (5); however, poor pharmacokinetics and low water solubility have kept them from human use. Although laudable synthetic efforts have achieved the total synthesis of several thiazolyl peptides (5–7), their complex scaffolds have prevented synthesis of analogs for screening in oncology and other preclinical studies. An alternate route to diversification, presented in this work, is to exploit the ribosomal synthesis of thiazolyl peptides in a gene replacement approach (8, 9).

Since the discovery of the first thiazolyl peptide, micrococin, more than half a century ago, nearly 100 poly-azole peptides have been discovered in this class of antibiotics; however, it was not until 2009 that it was reported they are the result of extensive posttranslational modifications on ribosomally generated prepro-
modification of isoleucine 8 (Ile8) to the cyclic hemiaminal teriocin biosynthesis that catalyzes the formation of thiazoline heterocycles within the macrocycle. Recent work in our laboratory has shown that these modifications increase antibiotic activity against S. aureus and Moraxella caviae (20). They have also been reported to harbor this unusual residue; however, the modification is stochastic with respect to Ile8, appearing in eight different oxidation states (thiomuracin A-I). Those analogs have similar antibiotic profiles, but notably differ in their inhibition of Streptococcus species (21). They share a triazole-substituted pyridine core and similar primary amino acid sequence with thiocillin (9). Because Ile8 is found exclusively as thiomuracin, but GE37468 harbors several distinct modifications in the macrocycle (SI Appendix, Fig. S1), including oxygenative modification of isoleucine 8 (Ile8) to the cyclic hemiaminal β-methyl-δ-hydroxy-proline (mhP). Thiomuracin has also been reported to harbor this unusual residue; however, the modification is stochastic with respect to Ile8, appearing in eight different oxidation states (thiomuracin A-I). Those analogs have similar antibiotic profiles indicating oxidation in the case of thiomuracin has modest impact on activity. Recent work in our laboratory has shown that these modifications increase conformational rigidity and are required for antibiotic activity in the 26-member thiocillin subclass for comparison with preprotein gene replacement.

Manipulation of the preprotein gene sequence and evaluation of mhp8 in GE37468, it is unknown whether replacement of mhp8 with an acyclic residue would impact antibiotic activity. Manipulation of the precursor gene sequence and evaluation of the tailoring oxygenases that would enable SAR studies in the 29-member subclass for comparison with preprotein gene replacement would be required to determine the role of these modifications in activity.

Results and Discussion

Sequencing of the GE37468 Biosynthetic Cluster from Streptomyces ATCC 55365. To identify the biosynthetic gene cluster for GE37468, a high diversity fosmid library (>10^5 unique clones) of Streptomyces ATCC 55365 genomic DNA was created in Escherichia coli. Thiazolyl peptide gene clusters have previously been identified in fosmid libraries using primers designed to the cyclodehydratase gene from the GE2270 cluster. Primers designed to the GE2270 cyclodehydratase gene were used to screen for the GE37468 cyclodehydratase gene. The fosmids were sequenced using next-generation Illumina sequencing as described in the SI Appendix. Subsequent contig assembly yielded a 28.2-kbp stretch of genomic DNA containing the cluster described below (SI Appendix, Figs. S2 and S3).

Bioinformatic Analysis of the GE37468 Cluster. Thirteen open reading frames—annotated as getA–M for GE37468 thiopeptide—in 17.1 kbp of Streptomyces ATCC 55365 genomic DNA were assigned to the GE37468 cluster. The upstream and downstream ends of the cluster were assigned based on coding regions for a putative HEAT repeat type protein and a putative aldehyde dehydrogenase, respectively, which are not expected to participate in GE37468 biosynthesis (Fig. 2). The structural gene and overall organization of the cluster show good similarity to clusters previously reported for GE2270 and thiomuracin (12). GE37468’s structural gene (getA) is located at the upstream end of the cluster and encodes a 57 amino acid preprotein divided into a 42 amino acid leader peptide (residues +42 to −1) and 15 amino acid coding region (residues +1 to +15). Fourteen residues of the coding region give rise to the mature antibiotic. A single asparagine residue positioned at the C terminus of the coding region is not present in the mature antibiotic. As might be expected, the getA peptide has good homology (56% identity) to the structural peptide for thiomuracin. Following getA are two ABC transporter-like genes (getB,C) that putatively enable host resistance.

Downstream of these genes are a cluster of six genes (getD–I) that encode the hallmark “core” posttranslational modification enzymes responsible for peptide maturation. These genes are similar in protein sequence and arrangement to the tdpB–tdpD genes from the GE2270A and thiomuracin clusters (SI Appendix, Table S2). Genes getD and getE encode lantibiotic-type dehydratases putatively responsible for dehydrating serines +1, +11, +13, and +14 to Dhas. Dhas +1 and +11 are subsequently cyclized to form the pyridine core as shown by isotopically labeled thiopeptide has good homology (56% identity) to the structural peptide for thiomuracin. Following getA are two ABC transporter-like genes (getB,C) that putatively enable host resistance.

Downstream of these genes are a cluster of six genes (getD–I) that encode the hallmark “core” posttranslational modification enzymes responsible for peptide maturation. These genes are similar in protein sequence and arrangement to the tdpB–tdpD genes from the GE2270A and thiomuracin clusters (SI Appendix, Table S2). Genes getD and getE encode lantibiotic-type dehydratases putatively responsible for dehydrating serines +1, +11, +13, and +14 to Dhas. Dhas +1 and +11 are subsequently cyclized to form the pyridine core as shown by isotopically labeled thiopeptide has good homology (56% identity) to the structural peptide for thiomuracin. Following getA are two ABC transporter-like genes (getB,C) that putatively enable host resistance.
serine feeding studies with the thiazolyl peptide nocathiacin (28). The +1 to +11 cyclization generates a 29-membered macrocycle as it forms the central pyridine ring, perhaps through an aza-Diels–Alder reaction catalyzed by the GetF enzyme. This enzyme is homologous to the previously reported TcIM enzyme from the thiocillin biosynthetic cluster (15). Genes getG and getI are homologous to enzymes found in the microcin B17 (McbC dehydrogenase) and cyanobactin (YcaO cyclodehydratase) pathways, respectively, and are putatively responsible for the formation of the six thiazole/oxazole/thiazoline rings in GE37468.

Downstream of the “core” proteins are enzymes involved in auxiliary modification (getLM), transcriptional regulation (getL), and an additional MchC-like dehydrogenase (getK). GetK presumably encodes a P450 monooxygenase enzyme involved in the conversion of isoleucine 8 to β-methyl-δ-hydroxy-proline (see below). This enzyme shows good homology to tpdQ (43% identity) and tpdH (42% identity) from the GE2270 and thiomuracin gene clusters, respectively. The TpdQ enzyme is hypothesized to β-hydroxylate Phe8 of GE2270 while TpdH is expected to function along with an additional P450 enzyme (TpdJ2) to hydroxylate Phe5 and Ile8 of thiomuracin. This yields an array of oxidations including epoxidation and ketonylation on Ile8 of thiomuracin. In contrast, no alternative oxidation states of Ile8 have been identified in the GE37468 framework.

**Transfer of the GE37468 Cluster to Streptomyces lividans TK24.** The GE37468 endogenous producer, *Streptomyces* ATCC 55365, has been reported to be genetically unstable and highly variable in thiopeptide antibiotic expression (18). In order to create a stable, manipulatable platform for GE37468 expression, the gene cluster was transferred to the model host organism *S. lividans* TK24. *S. lividans* has been exploited as a heterologous host for many natural product gene clusters because of its genetic stability, high transformation efficiency, and rapid doubling time (29). To transfer the GE37468 cluster, it was subcloned between the naturally occurring *BglI* and *NheI* restriction sites into the *E. coli*–*Streptomyces* shuttle vector pSET152 to create pSETGE1 (SI Appendix, Fig. S3). This subcloning strategy also transferred part of a tellurium resistance gene, a putative integral membrane protein, and a putative HEAT repeat type protein located upstream of the gene cluster to the pSETGE1 shuttle vector. Because GE37468 production in *E. coli* would also be desirable, expression of the pSETGE1 shuttle vector in DH5α *E. coli* was attempted; however, methanolic extracts from these cultures failed to show any compound by liquid chromatography—mass spectrometry (LCMS) analysis (SI Appendix, Fig. S4).

The pSETGE1 vector was subsequently transferred to *S. lividans* TK24 via bacterial conjugation and expression carried out in shake flasks under conditions previously described for GE37468 expression in *Streptomyces* ATCC 55365 (18). After 72 h of growth, the methanolic extract of *S. lividans* + pSETGE1 was compared with extracts from the endogenous producer (*Streptomyces* ATCC 55365) and an empty cloning vector as a negative control (*S. lividans* + pSET152) (Fig. 3A, B, and F). *S. lividans* + pSETGE1 exhibited a peak in the UV$_{350}$nm trace (2), which coincided with GE37468 from the wild-type producer (1). High-resolution MS confirmed the compound matched the mass of GE37468 from the wild-type producer within 0.5 ppm (M–OH)$^+$ as previously reported in ref. 17 (SI Appendix, Fig. S5). As expected, no corresponding peak or mass was observed in the pSET152 control sample.

$^1$H-NMR of purified GE37468 confirmed the compound produced by *S. lividans* + pSETGE1 was indistinguishable from GE37468 produced by the wild-type host (SI Appendix, Fig. S6). A time-course shake flask expression showed GE37468 production was detectable after 24 h in *S. lividans* and peaked at 72 h (SI Appendix, Fig. S7). Maximum yields of GE37468 from *S. lividans* (approximately 2–3 mg/L) were 40% the yields of

---

**Fig. 3.** LCMS analysis of GE37468 and analogs expression in *S. lividans* and *Streptomyces* ATCC 55365. UV$_{350}$ Chromatogram traces of methanolic extracts from (A) *Streptomyces* ATCC 55365 (GE37468, 1, red), (B) *S. lividans* + pSETGE1 (GE37468, 2, blue), (C) *S. lividans* + pSETGE-getJ::FRT (dihydro GE37468, 3, green), GE37468$_{blue}$, 4, green), (D) pSETGE-getKGE (GE37468$_{blue}$, 5, orange), (E) *S. lividans* + pSETGE-getA::FRT, and (F) *S. lividans* + pSET152 control.

**Streptomyces** ATCC 55365 per gram of cell paste (SI Appendix, Fig. S8). Minor amounts of other GE37468 products were also detectable in the extracts. These alternate products resulted from cleavage of the poly-Dha tail and were previously reported by Stella et al. as GE37468 B and C (30). A product corresponding to a dihydro GE37468 was also detectable in small amounts in the extracts from both *Streptomyces* ATCC 55365 and *S. lividans* (SI Appendix, Fig. S9). Despite the presence of export pumps in the biosynthetic cluster (getB and getC), no compound was detectable in cleared *S. lividans* media (SI Appendix, Fig. S10).

**Biosynthetic Origins of β-Methyl-δ-Hydroxy-Proline.** The mhP residue is found only in two thiazolyl peptides, GE37468 and thiomuracin I (12). The structural genes in both biosynthetic clusters suggest mhP8 arises from hydroxylation and cyclization of Ile8 rather than methylation and hydroxylation of proline as previously hypothesized (17). In thiomuracin, the side chain of Ile8 actually than methylation and hydroxylation of proline as previously hypothesized (17). In thiomuracin, the side chain of Ile8 actually
To investigate the P450 monoxygenase GetJ as the responsible oxygenation/cyclization catalyst for transformation of Ile8 to mhP8, we replaced oxygenation/cyclization catalyst for transformation of Ile8 to mhP8, we replaced the preprotein gene getA with a chloramphenicol antibiotic resistance cassette (cat) in the GE37468 cluster to create pSETGE-getA:cat using λ Red recombination method described above) to create pSETGE-getA::FRT. As expected, pSETGE-getA::FRT failed to yield any thiazolyl peptide compound by LCMS analysis of methanolic extracts (Fig. 3E).

With the GE37468 mhP8 and GE37468 mhP8 analogs in hand, we carried out an initial evaluation of how loss of the mhP residue affects bactericidal potency. Antibiotic activity was quantified against MRSA strain MW2 or Bacillus subtilis 168 using HPLC-purified compounds in serial-dilution liquid culture assays (SI Appendix, Table S3). Activity of wild-type GE37468 from both the endogenous host and our engineered host against MRSA MW2 (0.047 μg/mL, 36 nM) was similar to values reported by Stella and coworkers against the S. aureus strain L165 (0.03 μg/mL, 23 nM) (19, 30). GE37468 was equally active against B. subtilis in our assay. Variants GE37468 mhP8 and GE37468 mhP8 led to a 4-fold (0.19 μg/mL, 144 nM) and 2-fold (0.893 μg/mL, 71 nM) drop, respectively, in antibiotic potency against MRSA.

Conclusions

Whereas genetically tractable host systems such as Bacillus cereus and Streptomyces laurentii have previously been exploited to produce analogs of 26-member thiazolyl peptides, the heterologous expression platform presented in this work enables manipulation of EF-Tu-targeting 29-member variants in S. lividans. Using this system, we have begun to explore how the unique mhP residue is installed in the thiazolyl peptide scaffold and its role in GE37468’s antibiotic activity. GetJ likely carries out a regiospecific tandem double oxygenation of the δ-CH1 of the Ile8 side chain, first to the δ-CH2OH, and then to the δ-CH(OH)2, which is in equilibrium with the aldehyde (Fig. 4). Subsequent intramolecular attack by the amide nitrogen forms the cyclic hemiaminal mhP residue that is the accumulating form of GE37468. The homogenous modification of Ile8 in GE37468 stands in contrast to thiomuracin’s seven oxygenated forms of Ile8. To determine how this oxygenation chemistry is controlled, the tpd11 and tpd2 P450 homologs from the thiomuracin gene cluster could be used to replace getJ and determine if promiscuously oxygenated Ile8 forms of GE37468 are produced.

Here, we show that loss of oxidation of Ile8 to mhP8 in GE37468 leads to a 4-fold decrease in antibiotic potency against MRSA, whereas the corresponding change in thiomuracin (thiomuracin I vs. C, respectively) yields no variation in antibiotic activity against S. aureus (12). It may be that oxygenative transformation of Ile8 to mhP8 is a requirement in GE37468 because of the structural constraint it introduces to the scaffold. Studies
from our group on thiopeptide analogues have shown conformational rigidity imposed by thiazole sp²-α-carbons impart a structural rigidity to the macrocycle that is required for thiopeptide’s activity. This is presumably conveyed through a reduced entropic cost of target binding for a molecule with fewer conformational states (9). It is plausible that in the case of GE37468 the incomplete oxidation of thiazoline 6 introduces conformational flexibility that is compensated for by the mhPSI residue. Conversely, thiomuracin harbors a thiazole at position 6, which may create a rigid molecule that makes stochastic oxygenation of Ile8 permissible with retention of activity. If this is correct, synthetic oxidation (32) of thiazoline 6 in GE37468αNA and GE37468mhPSI should rescue antibiotic activity of these mutants; this is currently under investigation.

Among the most intriguing features of the GE37468 posttranslational maturation is how conversion of Cys6 to a thiazoline is controlled while four other Cys residues are fully desaturated to thiazoles in GE37468. Notably, the 26-member thiazolyl peptide thiostrepton also contains a single thiazoline residue in the presence of four thiazoles. This has been hypothesized to be a result of ring strain driven nonenzymatic epimerization of the thiazoline to the D-isomer, which may protect it from further oxidation by dehydrogenases because the “unnatural” isomer is not expected to be an enzyme substrate (33). It is unknown if this strategy is also employed by the GE37468 cluster to maintain thiazoline 6; however, the putatively reduced ring strain of the mhPSI and mhPSA analogs could be instrumental in yielding insights into how this chemistry is controlled and how macrocyclic architecture controls and affects function.

Methods

Additional details on cluster sequencing, media compositions, reagents, minimal inhibitory concentration assays, and MS parameters used in this work are listed in the SI Appendix.

Construction of Fosmid Library and Sequencing. Streptomyces ATCC 55365 was grown in seed media for 3–5 d, and genomic DNA was isolated via the Pospiech and Neumann “salting out procedure” (34). Fosmid libraries were constructed in the pCC2FOS vector following the “CopyControl Fosmid Library Production Kit” protocol from Epicent Biotechnology. DNA was cleaned with Genomic DNA Clean and Concentrator (Zymo) prior to ligation.

Library pools were screened for the cyclodehydratase gene using primers VNTI F and TSY R (SI Appendix, Table S1). Nine unique fosmids were identified (SI Appendix, Fig. S2), combined, and prepared for Illumina sequencing with the Nextera DNA Sample Prep Kit (Illumina) using the low molecular weight buffer without removal of c<300 bp fragments. Details on cluster assembly can be found in SI Appendix, Fig. S2.

Subcloning in pSET152 and Transfer to S. lividans. To subclone the GE37468 cluster to the integrative pSET152 vector, the Nhel restriction sites were removed using site-directed PCR mutagenesis with primers pSET Nhel QC F and pSET Nhel QC R. The pSET152 backbone was amplified using primers pSET BglII F and pSET Nhel R, and the resulting PCR product was digested with BglII and BglII and ligated to the similarly digested pCC2FOS G8-Δ122 fosmid to create pSETGE1 (SI Appendix, Fig. S3).

geta and getI genetic knockouts were created using J. Red recombination in E. coli following protocols described by Datsenko and Wanner (31). Briefly, the cat resistance cassette was PCR amplified from pKD3 using primers GetA prim 2F and GetA hygroR or primers GetI alt 2F and GetI alt R to knock out the getA and getI genes, respectively. Knockouts were carried out using the plasmid pSETGE1 (Fig. S2) expressing the cyclodehydratase gene in E. coli. BW25113 to create pSETGE-getA::cat and pSETGE-getI::cat. Markers were removed using E. coli BT340 (pcP20) expressing the FLP recombinase, which left the FRT scar and created plasmids pSETGE-getA::FRT and pSETGE-getI::FRT. To create a preprotein gene with the Ile8Aα mutation, a 2.2-kbp fragment of the pSETGE1 plasmid harboring the getA gene was removed using ActI and Scal restriction sites. An identical 2.2-kbp fragment containing getIαNA was inserted using “Gibson/Venter” cloning (described in SI Appendix, Fig. S12) to create pSETGE-getIαNA (35). pSET-derived plasmids were transferred into E. coli ET10267 and transferred to S. lividans TK24 using protocols found in the REDIRECT technology PCR-targeting system in Streptomyces coelicolor (John Innes Center).

Expression and Isolation of GE37468 and Analogs. Briefly, expression was carried out as previously described by Marinelli et al. (18). S. lividans or Streptomyces ATCC 55365 cultures were initiated from spore on soy flour mannikol plates into 30–50 mL starter cultures (seed media and J media) and were grown for 2–3 (3–5 d for Streptomyces ATCC 55365) at 28–30 °C, 250 rpm in baffled flasks containing Colilollers (Novagen) glass beads. Starter cultures were inoculated into AF/MS at 1% culture volume. Expression (100–150 mL culture in 500 mL baffled flasks containing glass beads) was carried out for 3 d. For labeling experiments, cultures were started in J media and inoculum was washed twice with filtered MG Base media prior to inoculation at 2% culture volume. Uniformly labeled 13C-L-l-isoleucine (Cambridge Isotopex) was added to 2 mM at 0, 1, and 2 d of expression.

Isolation of GE37468 was carried out following methods previously described by Bowers et al. (9). Cultures were pelleted, media decanted, and pellets resuspended 10–20% culture volume of methanol containing anhydrous sodium sulfate. After 30 min of frequent vortexing, the slurry was filtered through Whatman filter paper, rinsed with excess methanol, and solvent removed in vacuo. The resulting oil was resuspended with 2.5% acetonitrile/water and filtered through a 0.45-μm filter for LC/MS analysis (SI Appendix, Fig. 55) or further purification by HPLC (described in SI Appendix, Fig. 56).

ACKNOWLEDGMENTS. We thank Michael Acker and Albert Bowers for helpful discussion on thiopeptide biosynthesis and Jared Parker and Steve Malcolmson for careful reading of this manuscript. This work was supported by a New England Regional Center of Excellence Grant and National Institutes of Health Grant GM20011 (to C.T.W.). T.S.Y. is supported by a National Institutes of Health F32 postdoctoral fellowship.


