Published in final edited form as:

Chembiochem. 2014 January 24; 15(2): 320–331. doi:10.1002/cbic.201300319.

Identification and characterization of the carbapenem MM 4550 and its gene cluster in *Streptomyces argenteolus* ATCC 11009

Dr. Rongfeng Li^[a], Evan P. Lloyd^[a], Dr. Kristos A. Moshos^[b], and Prof. Craig. A Townsend^[a]

Craig. A Townsend: ctownsend@jhu.edu

^[a]Department of Chemistry, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218 (USA)

Abstract

Nearly 50 naturally-occurring carbapenem β -lactam antibiotics, most produced by *Streptomyces*, have been identified. The structural diversity of these compounds is limited to variance of the C-2 and C-6 side chains as well as the stereochemistry at C-5/C-6. These structural motifs are of interest both for their antibiotic effects and their biosynthesis. While the thienamycin gene cluster is the only active gene cluster publically available in this group, more comparative information is needed to understand the genetic basis of these structural differences. We report here the identification of MM 4550, a member of the olivanic acids, as the major carbapenem produced by *S. argenteolus* ATCC 11009. Its gene cluster was also identified by degenerate PCR and targeted gene inactivation. Sequence analysis revealed that genes encoding the biosynthesis of the bicyclic core and the C-6 and C-2 side chains are well conserved in the MM 4550 and thienamycin gene clusters. Three new genes, *cmmSu*, *cmm17* and *cmmPah* were found in the new cluster and their putative functions in the sulfonation and epimerization of MM 4550 are proposed. Gene inactivation showed that, in addition to *cmmI*, two new genes, *cmm22/23*, encode a two-component response system thought to regulate the production of MM 4550. Overexpression of *cmmI*, *cmm22* and *cmm23* promoted MM 4550 production in an engineered strain. Finally, the involvement and putative roles of all genes in the MM 4550 cluster are proposed based on the results of bioinformatics analysis, gene inactivation, and analysis of disruption mutants. Overall, the differences between the thienamycin and MM 4550 gene clusters are reflected in characteristic structural elements and provide new insights into the biosynthesis of the complex carbapenems.

Keywords

β -lactams; carbapenems; biosynthesis; MM 4550; *Streptomyces argenteolus*

Introduction

It has been reported that 50-60% of the more than two million nosocomial infections in the USA each year are caused by antimicrobial-resistant bacteria ^[1]. β -lactam antibiotics are critically important as pharmaceutical agents to combat these bacterial pathogens. Among these, carbapenems are the most potent subclass of these antibacterials due to their rapid penetration across bacterial cell membranes, high affinity to penicillin-binding proteins and resistance to a broad range of β -lactamases. With the rapid spread of multidrug resistance and abandoned anti-infective research and development programs in pharmaceutical

Correspondence to: Craig. A Townsend, ctownsend@jhu.edu.

^[b]Current address: Cubist Pharmaceuticals, 65 Hayden Avenue, Lexington, Massachusetts 02421 (USA)

companies, there are few antibiotics available for treatment and even fewer drugs in the antibiotic pipeline.

To date, all currently used carbapenems such as Imipenem[®] and Meropenem[®] are manufactured by total synthesis, in part because fermentation efforts have failed and knowledge of their biosynthesis is far from complete. Therefore treatment of severe infections with carbapenems is very costly. An important role could be played by the development of semi-synthetic approaches to provide large quantities of carbapenems at reduced cost.

The complex carbapenems are composed of a (5*R*)-carbapenem-3-carboxylic acid core that is commonly diversified by a sulfur-linked side chain at C-2 and an alkyl substituent at C-6. The nature of these substituents, often highly oxidized, gives rise to approximately 50 known natural carbapenem products; the simplest and best understood of which is (5*R*)-carbapenem-3-carboxylic acid. The biosynthesis of (5*R*)-carbapenem-3-carboxylic acid (Scheme 1) in *Pectobacterium carotovorum* has been discovered to be exquisitely efficient requiring only three enzymes, CarA, CarB and CarC. CarB is a member of the crotonase superfamily and catalyzes the formation of (2*S*,5*S*)-carboxymethylproline (CMP), the first committed intermediate in the pathway, from malonyl-CoA and L-pyrroline-5-carboxylate (L-P5C) [2]. CarA, carbapenam synthetase, cyclizes CMP to yield the (2*S*,5*S*)-carbapenam nucleus [3]. Finally, CarC, a non-heme iron α -ketoglutarate (α -KG)-dependent oxygenase, catalyzes both epimerization at the C-5 bridgehead and C-2/C-3 desaturation of the carbapenam to produce the final simple carbapenem [4].

The discovery of the gene cluster for thienamycin (Scheme 1) in *S. cattleya* led to the assignment of 23 genes (*thnA-V*), 21 of which are not present in the simple carbapenem gene cluster [5]. Biochemical analysis has shown that the activity of ThnE and ThnM mirror that of CarB and CarA, respectively, and are capable of producing the common bicyclic core. Additionally, ThnQ is involved in the hydroxylation of the C-6 side chain of PS-5 to produce *N*-acetyl-thienamycin [6]. Efforts aimed at revealing the biosynthesis of the C-2 assembly have shown that ThnR, ThnH and ThnT incrementally cleave CoA to 4-phosphopantetheine, pantetheine and cysteamine, respectively. ThnF acts as an acetyltransferase to convert thienamycin to *N*-acetyl thienamycin [6b, 7]. In addition, ThnG is responsible for the further modification of the C-2 side chain as a desaturase and an oxygenase [6b]. Gene inactivations have indicated that *thnL*, *thnN*, *thnO*, and *thnP* are absolutely necessary for the biosynthesis of thienamycin. However, reported mutational analyses give seemingly contradictory results as thienamycin production was still observed in *thnG*, *thnR* and *thnT* disruption mutants, suggesting that these genes are not required for thienamycin biosynthesis [8]. Finally, two pathway regulatory proteins, ThnI and CepU, positively regulate the expression of 10 genes in the thienamycin cluster and the biosynthesis of cephamycin in *S. cattleya*, respectively [9].

These recent advances notwithstanding, many questions remain about the biosynthesis of complex carbapenems including how, and in what order, epimerization of the carbapenam bridgehead, introduction of the double bond at carbons C-2/C-3, incorporation of CoASH at C-2 and alkylation at C-6 occur. Enzymatic analysis has demonstrated that the differences in the level of oxidation of the C-6 and C-2 side chains and C-6 stereochemistry give rise to many structural combinations [10]. It remains unknown how these differences are genetically encoded to produce potent antibiotics, distinct from thienamycin. Genetic differences across different species would explain these structural alterations. Apart from the silent gene cluster discovered in the *S. flavogriseus* genomic sequencing project, the thienamycin cluster in *S. cattleya* so far remains the only openly available gene cluster functionally associated with a specific complex carbapenem. Thus, exploration of the comparative genomics of the

structurally advanced carbapenems would not only give greater insight into the biosynthesis of these metabolites, but also provide an opportunity to apply combinatorial biosynthesis to generate new carbapenems with potentially improved properties.

In the early 1980's *S. argenteolus* ATCC 31589 was reported to be a producer of asparenomycins, another group of potent carbapenems [11]. In this article, we establish that MM 4550 rather than asparenomycins is the major carbapenem produced in a related strain of *S. argenteolus* ATCC 11009. In addition, we provide detailed information on the identification, bioinformatic analysis, and mutational analysis of the gene cluster responsible for the biosynthesis of MM 4550.

Results

Identification of the carbapenems produced in *S. argenteolus* ATCC 11009

Asparenomycin A, B and C (Scheme 1) were isolated from two *Streptomyces* strains, *S. tokunonensis* ATCC 31569 and *S. argenteolus* ATCC 31589 [11b]. However, a query of the ATCC and NRRL catalogs did not yield *S. argenteolus* ATCC 31589, but a different strain of *S. argenteolus*, ATCC 11009. *S. argenteolus* ATCC 11009 was also reported to produce carbapenems and carbapenams such as an olivanic acid complex and 17927 D [10c, 12], but identifications have not been published [12b]. As the first step in our investigation, we conducted small-scale fermentation of *S. argenteolus* ATCC 11009 in OYM and ISP4⁺ media, which showed antibiotic activity against the β -lactam super-sensitive *E. coli* ESS and inhibition of the β -lactamase in *Klebsiella pneumoniae* subsp. *pneumoniae* ATCC 29665 (data not shown). Bioassay of the fractions collected from HPLC by direct injection of supernatant broth indicated that the active product(s) resided in only one narrow region of the chromatogram (data not shown).

Initial efforts to identify the β -lactam product failed owing to the low titer from the wild-type *S. argenteolus* ATCC 11009. To improve production, a genetic engineering approach was applied by overexpression of three positive regulatory genes, *cmmI*, *cmm22* and *cmm23* (see below). The expression of each gene was individually controlled by an *ermE***p* constitutive promoter and the final expression construct containing all three genes, pMS82/*cmmI*-22-23 (Figure S1), was integrated into the ϕ BT1 site of the genome. Bioassay against *E. coli* ESS showed an enlarged zone of inhibition from the engineered strain, and HPLC analysis revealed that this altered strain still produced the same product as the wild-type strain (Figure 3), but in a nearly 4-fold improved yield based on comparison of peak areas for the active fraction. In addition, the maximal production in the engineered strain was at 48 h and 56 h in 50-mL and 1.3-L scale fermentations, respectively, compared to the wild-type strain at 72 h and 80 h (Figure S2).

S. argenteolus *cmmI*-22-23 was used for characterization of the carbapenem produced. A large-scale fed-batch fermentation procedure was developed using a modified minimal medium (ISP4⁺⁺) to give maximal production that also minimized impurities during product purification. To make large-scale growth applicable to laboratory purification techniques, a long-chain quaternary ammonium salt was used as a phase transfer catalyst to extract negatively charged carboxylates into a small volume of organic solvent. Back-extraction of the resulting organic mixture with aqueous potassium iodide allowed recovery of the antibiotic into a volume 100 times more concentrated than the original fermentation filtrate. We found that a 5% solution of Aliquat 336 in dichloromethane was highly effective in extracting the antibiotic where only two 300 mL washes were required to extract almost all of the antibiotic from 13L fermentation filtrate as monitored by *E. coli* ESS bioassay. The concentrated KI solution was then lyophilized, and the crude salt obtained was further purified by anion exchange chromatography over Dowex resin. We found that anion

exchange chromatography was critical at this stage to remove the majority of impurities, salts, and residual Aliquat. Most importantly, this procedure resulted in highly concentrated fractions for further purification and analysis by HPLC.

Reverse-phase (C18) HPLC analysis of the active fractions resolved only one bioactive peak, as determined by bioassay on *E. coli* ESS plates, with a retention time of 16.3 min (Figure 3). Co-injection with a synthetic asparenomicin A standard showed a different retention time and UV spectrum, strongly indicating that the asparenomicins originally reported from *S. argenteolus* ATCC 31589 are not produced in *S. argenteolus* ATCC 11009 [10b, 11b]. Reaction with hydroxylamine resulted in the loss of that peak, consistent with the presence of a β -lactam in the active product. Additional bioactive material was collected by HPLC for further characterization.

The UV spectrum showed the product had not only a typical carbapenem absorbance signature at 288 nm but also an additional absorbance band was noted at 245 nm. A search of available UV data across all natural carbapenems resulted in only one potential compound that matched the spectrum of the isolated antibiotic. ^1H NMR spectroscopy identified the compound as MM 4550, as both chemical shifts and coupling constants were consistent with the published characterization of this compound (Figure S3a). Furthermore, two-dimensional ^1H - ^1H COSY, ROESY, and TOCSY experiments confirmed the assignments of all protons made in the original publication (Figure S3b-e) [13]. Electrospray ionization mass spectrometry provided accurate mass of 407.0226 (± 0.5 ppm; calculated 407.0224) for the anion of MM 4550.

Cloning of the MM 4550 biosynthetic gene cluster

Sequence analysis suggests and biochemical studies confirmed that ThnE and ThnM in the thienamycin biosynthetic pathway mirror the chemistry of CarB and CarA in the simple carbapenem pathway [6a, 6c]. Alignment of CarA (*P. carotovorum*), β -LS (*S. clavuli-gerus*), β -LS3 (*S. antibioticus*), CpmA (*P. luminescens*), ThnM (*S. cattleya*), and several Asn-B (asparagine synthetase) revealed two motifs (M1 and M2) highly conserved in the β -lactam forming enzymes but either absent or present at distinctly lower homology in asparagine synthetases (Asn-B). Using the CODEHOP program (<http://blocks.fhcrc.org/codehop.html>) [14], two degenerate primers were designed, thnM-DEG-5 and thnM-DEG-3 (Figure.1), to these regions. A specific PCR product with the expected size of ~ 350 bp was obtained from *S. argenteolus* gDNA. This PCR product showed an incomplete open reading frame (ORF) encoding a truncated protein of 108 amino acids with 74% identity to ThnM of *S. cattleya*, and 32% identity to CarA of *P. carotovorum*. A BLASTP alignment also showed that the two highly conserved motifs, M1 and M2, were present in the partial protein. These results strongly suggested that a partial *thnM* orthologous gene had been cloned from *S. argenteolus*.

A cosmid gDNA library of *S. argenteolus* ATCC 11009 was constructed and screened using the ~ 350 -bp gene fragment as a probe. Fourteen positive clones were obtained from a 2000 clone library. The sequencing analysis revealed that the T₃ and T₇ ends of the insert in cosmid 5D1 and 12B9 encoded a MarR-family transcriptional regulator and an AsnC-family transcriptional regulator, respectively. These proteins are not encoded by genes in the thienamycin gene cluster, indicating that cosmids 5D1 and 12B9 likely contained a whole carbapenem gene cluster. Cosmid 5D1 was chosen for further investigation.

Cosmid 5D1 was sequenced in its entirety. Five gaps between five contigs were bridged by PCR amplification and Sanger sequencing to give the full, circular cosmid. Overall, we ascertained that 5D1 contained the putative MM 4550 biosynthetic genes, which spanned

almost 43.8 kb of the *S. argenteolus* genome. Computer-assisted analysis of the sequence revealed 31 complete open reading frames (ORFs) (Figure 2).

In order to confirm the cloned gene cluster was responsible for the biosynthesis of MM 4550, *cmmE*, whose gene product showed 76% and 40% identity to ThnE and CarB, respectively, was inactivated by a PCR targeting strategy with an *oriT-aac(3)IV* cassette to generate a gene replacing mutant argE8. The intended disruption of *cmmE* was confirmed by Southern hybridization with *cmmE* and *oriT-aac(3)IV* probes. As expected, when *Pml*I-digested gDNA was tested with the *cmmE* probe, a 1.65-kb positive fragment was observed in the wild-type but not the argE8 mutant gDNAs. Correspondingly this fragment was replaced with a 2.2-kb band in the argE8 mutant when hybridized with the *oriT-aac(3)IV* probe, but this band was missing in the wild-type gDNA (Figure 4). Bioassay on *E. coli* ESS plates showed that the argE8 mutant failed to produce antibacterial products when fermented in ISP4⁺ medium (Table 2, Figure 3B). HPLC analysis of the extract from the argE8 fermentation medium further showed that the peak corresponding to MM 4550 in the wild-type *S. argenteolus* was absent (Figure 3A). These results strongly suggested that the gene cluster cloned in cosmid 5D1 is responsible for the biosynthesis of MM 4550 in *S. argenteolus*.

The DNA sequences of MM 4550 gene cluster have been deposited in the NCBI nucleotide sequence database with accession number KF042303.

Sequence analysis of the MM 4550 biosynthetic gene cluster

The putative functions of the deduced gene products are summarized in Table 1. The overall architecture and some gene products of the MM 4550 gene cluster are highly similar to a cryptic cluster in *S. flavogriseus* ATCC 33331 identified by gene scanning [15]. However, because no product is known from this cluster and none of the enzymes encoded by it have been characterized, analysis of the MM 4550 gene cluster focused principally on comparison with the thienamycin gene cluster.

As shown in Figure 2, among the 31 complete ORFs present in the insert of 5D1, 17 clear homologues are also present in the thienamycin gene cluster [5]. Genes *cmmE-Q* show the same organization as *thnE-Q* in the thienamycin gene cluster. Notable genes *cmmR*, *cmmS* and *cmmT* are also shared between the two clusters. However, in the MM 4550 gene cluster, they are organized as *cmmR-cmmT-cmmS* with *cmmR* and *cmmS* transcribed in the opposite orientation to *cmmT*. A new gene, *cmmI7*, is inserted between *cmmQ* and *cmmR* in the MM 4550 gene cluster. In addition to *cmmI7*, four more new genes, *cmmSu*, *cmmPah*, *cmm22* and *cmm23* are present in the MM 4550 cluster with *cmmSu* located upstream of *cmmE* at one end and the other three genes are downstream of *cmmS* at the other end of the cluster. Bioinformatic analysis indicated that genes upstream of *cmmSu* and downstream of *cmm24* encode proteins that are not related to the biosynthesis of carbapenems or are not present in known carbapenem gene clusters (Table 1). Finally, *thnA-D* and *thnV*, whose roles in the thienamycin biosynthetic pathway have not been determined, and *cphU*, which encodes a SARP-like regulator for cephamycin biosynthesis [9], are all absent from the MM 4550 gene cluster.

It has been reported that thienamycin and (5*R*)-carbapenem-3-carboxylic acid share the first two steps in their biosynthetic pathways to make the carbapenam bicyclic core. The orthologous genes in the MM 4550 cluster encode proteins sharing 69%, 37% and 64%, 22% identity to ThnE, CarB and ThnM, CarA, respectively, and active site residues are also conserved, indicating CmmE and CmmM catalyze the same reactions in MM 4550 biosynthesis [3, 6a, 6c, 8, 16].

The three putative radical-SAM enzymes encoded by *cmmK*, *cmmL* and *cmmP* are also highly conserved in the MM 4550 gene cluster and share 79%, 67%, and 83% amino acid identities to ThnK, ThnL and ThnP, respectively. CmmQ shows sequence homology to ThnQ that has been shown to be a Fe(II)/ α -KG-dependent oxygenase responsible for the hydroxylation of the C-6 ethyl side chain of thienamycin [6b]. To explain the difference between the C-8 sulfate of MM 4550 and C-8 hydroxyl of thienamycin, a BLASTP search revealed that CmmSu has 38% identity and 52% positive to the putative sulfotransferase in *S. blastomyceticus* [17]. A conserved 5'-phosphosulfate-binding sequence (5'-PSB), ¹⁷RCGSTM²³, was located in a strand-loop-helix (PSB-loop) motif. A second motif, ⁸⁹GRYAAGEVPAISL¹⁰¹, similar to the conserved 3'-phosphate binding sites in sulfotransferases is also present in CmmSu (Figure S7) [18]. These conserved motifs strongly suggest that CmmSu acts as a sulfotransferase required for sulfonation of the C-6 side chain in MM 4550. Together, we propose that some or all of *cmmK*, *cmmL*, *cmmP*, *cmmQ* and *cmmSu* are involved in the biosynthesis of the C-6 side chain in MM 4550.

We identified a set of genes, designated as *cmmR*, *cmmH*, *cmmT*, *cmmF* and *cmmG*, whose products show high similarity to ThnR, ThnH, ThnT, ThnF and ThnG in the thienamycin biosynthetic pathway. Their roles in the biosynthesis of the C-2 cysteaminy side chain in thienamycin have been demonstrated through stepwise cleavage of CoA by ThnR, ThnH and ThnT [7]. Furthermore, the acetyltransferase ThnF, is responsible for the *N*-acetylation of the cysteaminy side chain from acetyl-CoA [7]. Finally, CmmG shares 70% identity with ThnG in the thienamycin biosynthetic pathway. *In vitro* characterization of ThnG has shown that it catalyzes the desaturation and sulfoxidation of the *N*-acetyl-2-cysteaminy side chain of *N*-acetyl-thienamycin to form the sulfoxide of PS-7 [6b]. Consistent with its structure, the C-2 side chain of MM 4550 is likely to be processed by these five enzymes in the order of CmmR, CmmH, CmmT, CmmF and CmmG.

cmmI encodes a putative LysR-family positive regulator, which shares 50% and 26% identities to ThnI and ClaR in the biosynthetic pathways of thienamycin and clavulanic acid, respectively [9, 19]. Two new genes, *cmm22* and *cmm23*, located at the right end of the gene cluster encode proteins homologous to the two-component regulatory systems in a wide range of microorganisms. The amino acid sequence of Cmm22 (227 aa) shows end-to-end similarity to response regulators, including an aspartate residue (Asp⁵²) typically phosphorylated inside the N-terminal phosphorylation domain, a C-terminal DNA binding domain, and several other key characteristic residues (Asp⁹, Asp¹⁰, Thr⁸⁰ and His¹⁰¹) [20]. As observed in most two-component regulatory systems, *cmm23* encodes a protein (506 aa) with full-length homology to sensory histidine kinases and is located immediately downstream of *cmm22*. Cmm23 contains three conserved domains including the H-box (²⁶⁹HELRTPL²⁷⁵) (His²⁶⁹ is a putative phosphorylation site), the N-box (³⁷⁶NLVGNA³⁸¹), and the G-box (⁴⁰⁵VRDSGPGI⁴¹²) [20c]. Structural prediction using TopPred 0.01 (<http://mobyli.pasteur.fr/cgi-bin/portal.py?#forms::toppred>) [21] identified three possible transmembrane domains in Cmm23: domain I (residues 14-34, score 1.96), II (147-167, score 1.26) and III (181-201, score 1.26). The sensor domain is generally located at the N-terminus in most histidine kinases; therefore, domain I is likely the sensor domain of Cmm23. The organization and likely properties of Cmm22 and Cmm23 suggested that they belong to the OmpR family of two-component response regulators [20b].

Two genes, *cmm17* and *cmmPah*, absent in the thienamycin gene cluster were also identified. *cmm17* is located directly downstream of and in the same orientation as *cmmQ* in the gene cluster and encodes a 292-amino acid protein. BLASTP analysis showed Cmm17 has appreciably high similarity (35-38% identity and 50-55% similarity) to enoyl-CoA hydratases from a number of microorganisms. Cmm17 possesses most residues important for substrate binding and catalysis including Glu¹²⁸. However, a His¹⁴⁸ residue is present

instead of the Glu residue that is typically conserved in most enoyl-CoA hydratases (Figure S8) [16, 22]. *cmmPah* is located between *cmmT* and *cmmS* and is transcribed in the opposite direction to *cmmT* and *cmmS*. Its gene product comprises 298 amino acids and belongs to the arginase-like/histone-like hydrolase superfamily [23]. A BLASTP search showed that CmmPah has significant similarity to agmatinases, a subgroup of the arginase superfamily and a key enzyme in an alternative metabolic pathway of arginine in microorganisms and plants [24]. The highest homology (46-52% identity and 60-67% similarity) is to proclavamine amidino hydrolase (PAH), an enzyme in the clavulanic acid and 5S clavam biosynthetic pathways from *S. clavuligerus* and *S. antibioticus* [25]. However, alignment of CmmPah with three PAH enzymes from *S. clavuligerus* and *S. antibioticus* and two arginases from *A. flavithermus* and *S. viridis* showed that only one of the six Mn²⁺-binding residues (Asp²²²) is conserved in CmmPah. Only one of the three guanidino binding residues (Thr²³²) and one of the four α -amino acid binding residues (Gly¹⁴⁷) conserved in arginases and PAHs is present in CmmPah [26]. The Gly-Gly-Asp-His-Ser motif conserved in all arginases is changed to ¹¹⁶Gly-Gly-Gly-Arg-Ser¹²⁰ in CmmPah (Figure S9). Finally, secondary structure prediction (<http://bioinf.cs.ucl.ac.uk/psipred/>) showed that the N-terminal secondary structure of CmmPah is similar to the PAHs and arginases, but the C-terminal region is different [26].

Mutational analysis of genes in the MM 4550 gene cluster

To better understand the role of each individual gene in the biosynthesis of MM 4550, we generated gene-replacing mutants for twenty genes in the proposed biosynthetic gene cluster using λ -RED-mediated PCR mutagenesis [27]. Each targeted gene was replaced with a FRT-*oriT-aac(3)IV*-FRT cassette in its resultant mutant. Mutants derived from double crossover events with the targeted genes, and their disrupted copies on 5D1 were selected according to their apramycin-resistant/kanamycin-sensitive (Am^R/Km^S) phenotype. To confirm the genotype of the mutants, gDNA was digested with restriction enzymes and hybridized with a *oriT-aac(3)IV* probe. As expected, the gDNA of wild-type *S. argenteolus* did not hybridize to the probe, but positive hybridization was observed from all tested mutants and the size of each resultant band was as predicted by restriction analysis (Figure S6.). The digested gDNA of each mutant was also hybridized against the gene-specific probe. No hybridization was observed from any potential mutant (data not shown), confirming that targeted gene replacement had occurred in each case. One mutant in each disruption of a specific gene was chosen for further characterization.

All disruption mutants were characterized by fermentation in ISP4⁺ medium and analyzed for antibacterial activity on *E. coli* ESS or β -lactamase induction on *B. licheniformis* using Nitrocefin as indicator [28]. We observed that, in addition to antibiotic activity against *E. coli* ESS, MM 4550 also exhibits moderate β -lactamase inducing activity in *B. licheniformis* (Figure S4b). We further analyzed bioactive intermediates in disruption mutants by UPLC-MS (Table 3, Figure S5).

As summarized in Table 2 and explicitly shown in Figure S4a, bioassay on *E. coli* ESS showed that disruption of *cmmE* and *cmmM* completely eliminated production of MM 4550 during 7-day fermentation. Nitrocefin assays showed that β -lactamase induction was only at background levels compared with the wild-type, consistent with no carbapenems being produced in these mutants (Figure S4b). Moreover, HPLC analysis confirmed that the peak corresponding to MM 4550 was completely absent in the *argE8* mutant (Figure 3). Similar results were also obtained from *argK11*, *argL2* and *argP4* mutants, suggesting that these enzymes are also necessary for biosynthesis (Table 2, Figure S4a). Bioassay against *E. coli* ESS showed that the disruption mutant of *cmmQ*, *argQ4*, consistently produced a bioactive product during its fermentation, but in reduced amount or potency in comparison to the

wild-type. The Nitrocefin assay showed that argQ4 still maintains β -lactamase induction activity, but loses detectable antibacterial activity against *B. licheniformis* (Table 2, Figure S4a and b). The active fraction of extracted fermentation broth shared both the same retention time and mass with synthetic PS-5, confirming that PS-5 is accumulated in argQ4 (Table 3, Figure S5) [6b]. BLASTP analysis suggested that CmmSu is the most likely candidate in the gene cluster to catalyze C-6 sulfonation to MM 4550. Interestingly, the disruption mutant of *cmmSu*, argSu13, still produced antibacterial products. However, unlike the wild-type *S. argenteolus*, the supernatant of the argSu3 only showed β -lactamase induction activity rather than antibacterial activity on *B. licheniformis* (Figure S4b), indicating that the product is likely not MM 4550. UPLC-MS analysis of the bioactive material in argSu13 revealed a mass for the non-sulfonated derivative of MM4550 (Table 3, Figure S5).

Mixed results were obtained from genes potentially involved in the biosynthesis of the C-2 side chain. Mutants argR8, argT3 and argF7, disrupted at *cmmR*, *cmmT* and *cmmF*, respectively, were unable to produce MM 4550 at any time point during the fermentation, while the *cmmH* and *cmmG* mutants, argH8 and argG5, still produced bioactive products as indicated by bioassay on *E. coli* ESS (Table 2, Figure S4a). The Nitrocefin assay showed that both the argG5 and argH8 mutants had similar antibacterial and β -lactamase induction activities as the wild type (Figure S4b). UPLC-MS analysis showed the active intermediate in argG5 has the same mass as MM 17880 (Table 3, Figure S5). These results indicate that *cmmH*, *cmmG*, *cmmR*, *cmmT* and *cmmF* are essential for the production of MM 4550).

Gene *cmm22* was disrupted to investigate if the two-component response regulators are involved in the MM 4550 biosynthesis. Both antibacterial and β -lactamase induction activities from the disruption mutant, arg22-6, were negative, suggesting that the two-component regulators encoded by *cmm22/23* are essential for the biosynthesis of MM 4550 (Table 2, Figure S4a and b). Thus, the biosynthesis of MM 4550 is regulated by at least two regulatory mechanisms; LysR-type transcription activation and the two-component response system, in *S. argenteolus* [5, 9].

The involvement of two new genes, *cmm17* and *cmmPah*, in the biosynthesis of MM 4550 was also studied through gene inactivation. Disruption of *cmm17* generated a mutant (arg17-4) that was unable to produce any metabolites active against *E. coli* ESS and *B. licheniformis*. However, the fermentation supernatant showed strong β -lactamase induction activity as shown by the Nitrocefin assay, indicating a β -lactam ring-containing intermediate accumulated in the mutant. On the other hand, bioassays and β -lactamase induction assay showed that the *cmmPah* disruption mutant (argPah9) failed to produce metabolites positive against *E. coli* ESS and *B. licheniformis*, or induce β -lactamase expression. While the role of CmmPah in the biosynthesis is unknown, extensive mass spectroscopic analyses were unable to detect intermediates in the mutants. Notwithstanding, the results demonstrated that these two new genes are essential for the biosynthesis of MM 4550 (Table 2, Figure S4a, and b). Finally, analysis of CFEs of mutants arg17-4 and argPah9 showed that they displayed similar antibacterial and β -lactamase induction activities as their fermentation supernatants (data not shown).

Disruption of *cmm24* resulted in a mutant with the same phenotype of antibacterial and β -lactamase induction activities as the wild-type *S. argenteolus*. Combined with the BLASTP analysis (Table 1), *cmm24* may not be involved in the biosynthesis of MM 4550 and likely to be the 3' boundary of the gene cluster. The ORF analysis also showed that the intergenic region upstream of *cmmSu* is nearly 1 kb and the gene (*orf4*) upstream of *cmmSu* encodes a protein similar to sugar metabolism that is clearly unrelated to carbapenem biosynthesis

(Table 1). Thus, we propose that the MM 4550 gene cluster harbors 21 genes involved in its biosynthesis, resistance and regulation (Figure 2).

Discussion

Olivanic acids, including MM 4550, MM 13902 and MM 17880, were originally identified from *S. olivaceus* ATCC 31126 and *S. olivaceus* ATCC 21379 [10c, 13, 29]. Although it has been mentioned that olivanic acids are also produced by other *Streptomyces* strains, including *S. argenteolus* ATCC 11009, no further information has been reported [10c]. By combining genetic engineering approaches with modified extraction protocols and structural analysis, we have demonstrated, for the first time, that MM 4550 is the major carbapenem produced by *S. argenteolus* ATCC 11009. This finding contrasts with the two *S. olivaceus* strains in which three compounds of the olivanic acid complex are produced in similar titers [13]. NMR and UV spectroscopy, and mass spectrometry support the structure of the final product in *S. argenteolus* as MM 4550.

The MM 4550 gene cluster is only the second gene cluster determined to be involved in the biosynthesis of a complex carbapenem that has been characterized. In comparison with thienamycin, MM 4550 is more highly functionalized, with a sulfate at the C-8 hydroxyl as well as an oxidatively modified C-2 side chain. The syn relationship across the C-5/C-6 bond is another feature that differs, and many carbapenems have been discovered having this stereochemistry. This structural feature suggests a potential fundamental difference in their biosynthetic pathways, despite the similarity of the gene clusters for thienamycin and MM 4550.

Although genes presumed to be involved in the biosynthesis of the bicyclic core and the C-6 and C-2 side chains are conserved between the two gene clusters, there are differences that stand out. First, the genes *thnA-D*, *cepU* and *thnV* in the thienamycin gene cluster are all missing in the MM 4550 cluster. It is known that *cepU* is a regulatory gene for cephamycin biosynthesis, but the functions of the other genes in this group are yet to be determined in *S. cattleya*. Second, we found that the biosynthesis of MM 4550 is regulated by two regulatory mechanisms with proteins encoded by *cmmI* and *cmm22/23*. A helix-turn-helix (HTH) DNA-binding domain and the rare TTA codon are shared by ThnI and CmmI, indicating similar regulatory roles in their biosynthetic pathways [9]. Two-component regulatory systems are generally global regulators in the biosynthesis of secondary metabolites [30]. Disruption of *cmm22* clearly showed that *cmm22/23* is involved in MM 4550 biosynthesis, presumably through phosphorylation of the regulator (Cmm22) by the membrane-bound phosphokinase (Cmm23). In addition, overexpression of *cmmI* and *cmm22/23* improved the production of MM 4550, confirming their positive regulatory roles. Lastly, three putative biosynthetic genes that are unique in the MM 4550 gene cluster were identified.

The first, *cmmSu*, encodes a putative sulfotransferase as predicted by sequence analysis. Sulfotransferases are a superfamily of enzymes catalyzing sulfonation of a variety of endogenous and exogenous substrates. In contrast to humans and other mammals, little is known about bacterial sulfotransferases [18a]. BLASTP analysis suggested that CmmSu is PAPS-dependent owing to the presence of a putative PBS-loop and the 3'-phosphate binding sites. Disruption of *cmmSu* resulted in a mutant with a different antibacterial spectrum from MM 4550. Mass-spectroscopic analysis showed the accumulation of a non-sulfonated derivative of MM 4550 in the argSu13 mutant, as expected. An earlier analysis of the OA-6129 group of carbapenems identified in a series of random mutants of *S. fulvoviridis* processing at C-6 and C-8 that was thought to be ordered as C-8 hydroxylation, C-6 isomerization and C-8 sulfonation. The substrate of the sulfonation enzyme, OA-6129B1, was demonstrated to be bioactive [31]. MM 4550 possesses the same C-6 side chain as

OA-6129C, suggesting that the C-6 side chain of MM 4550 is processed through a similar biosynthetic sequence as the OA-6129 carbapenem group.

Structurally, the major difference that distinguishes MM 4550 from thienamycin is the syn-relationship across C-5 and C-6. The remaining enzymes that could be involved in establishing this alternate stereochemistry are Cmm17 and CmmPah as they are present in *S. argenteolus* but not in *S. cattleya*. Cmm17 shows high sequence homology to enoyl-CoA hydratases. It is known that the acid/base chemistry for this type of enzymes can hydrate α , β -unsaturated double bonds or perform the reverse reaction [22]. Interestingly, disruption of *cmm17* resulted in the loss of antibacterial activities against *E. coli* ESS and *B. licheniformis* but gained stronger β -lactamase induction activity compared to the wild-type and other disruption mutants, indicating that a β -lactam-containing intermediate was accumulated in the mutant. We propose that Cmm17 could be involved in the epimerization of the C-6 side chain and, therefore, produce a different ultimate stereochemical outcome in MM 4550 biosynthesis (Scheme 2). Although CmmPah displays high sequence similarity to arginases and PAHs, almost all the amino acid residues involved in substrate binding and co-factor binding in the comparison enzymes are absent from CmmPah (Figure S8), strongly suggesting that CmmPah catalyzes a different reaction(s) in the biosynthesis of MM 4550 from the arginases and PAHs.

It is important to note that the disruption mutants for *cmmG* and *cmmQ* resulted in the detection of intermediates with masses for MM 17880 and PS-5, respectively, in the mutants (Table 3, Figure S5). As shown previously in *S. cattleya*, ThnQ hydroxylates the bioactive substrate PS-5 to generate *N*-acetyl-thienamycin, and the later is further oxidized by ThnG to produce *N*-acetyl dehydrothienamycin. ThnG also catalyzes the oxidation of PS-5 to produce PS-7 and PS-7 sulfoxide [6b]. Thus, in keeping with this *in vitro* analysis, CmmG is responsible for the oxidation of MM 17880 to produce MM 13902 and further to produce MM 4550 (Scheme 2), CmmQ is likely involved in the C-8 hydroxylation of MM 4550 as observed in the OA-6129 group of carbapenems. Thus, a bioactive intermediate with the mass and UPLC retention of PS-5 is accumulated when *cmmQ* was knocked-out.

Importantly, disruption of *cmmR* and *cmmT* led to MM 4550 non-producers in *S. argenteolus*. This finding contradicts reports in *S. cattleya* in which inactivation of orthologous genes *thnR* and *thnT* retained thienamycin production in the mutants [8]. Biochemical studies have shown that both ThnR and ThnT are involved in the incremental truncation of the C-2 side chain in thienamycin [7]. Our gene inactivation results support the *in vitro* analysis and confirmed that *cmmR* and *cmmT* are essential for the biosynthesis of MM 4550. ThnR and ThnT are similar to Nudix hydrolases and peptidases, respectively. Searching the recently sequenced genome of *S. cattleya* showed five putative Nudix hydrolases and 46 putative peptidases. Thus, the continued production of thienamycin in *thnR* and *thnT* disruption mutants could result from *in trans* complementation.

The questions that remain in the biosynthesis of complex carbapenems include the mechanism and timing of inversion of the C-5 carbapenam bridgehead, the desaturation of the C-2/C-3 bond, and the attachment of the C-2 and C-6 side chains. Like the thienamycin cluster, there are no obvious candidates to catalyze these steps in the MM 4550 gene cluster. As both thienamycin and MM 4550 require these transformations, it is probable that enzymes conducting these steps are among those encoded in common by both gene clusters. Our gene disruption results ruled out the involvement of *cmmQ* and *cmmG* in the epimerization/desaturation as has been hypothesized [6c] because the disruption mutants retain the ability to produce bioactive intermediates.

Based on previous reports and the results of our earlier experiments, we propose a biosynthetic pathway to MM 4550 shown in Scheme 2. Many of the early steps are almost certainly shared with thienamycin biosynthesis, but diverge to achieve stereochemical inversion of the C-6 and C-8 stereocenters and the addition of a sulfate, likely through the action of Cmm17 and CmmSu. The OA-6129 group of carbapenems identified from mutagenesis studies in *S. fulvoviridis* includes the C-2 side chain prior to pantethiene cleavage by a ThnT orthologue [32]. These accumulated products, presumed intermediates, provide possible insight into the sequence of CmmQ hydroxylation, dehydration/rehydration of the C-6 sidechain and *O*-sulfonation. The OA-6129 carbapenems are shown as possible intermediates in Scheme 2. Additionally, the occurrence of the asparenomicin group of carbapenems in *Streptomyces* suggests the possibility of an ene-carbapenem intermediate during the proposed C-6 and C-8 epimerization [10b]. The origin of the structural diversity of many isolated carbapenems across the entire family can in part be rationalized by this scheme.

It has been shown that the stereochemical relationship across C-5 and C-6 influences both antibacterial activity and stability toward β -lactamases where the presence of a sulfate in C-8 also appears to enhance stability toward β -lactamases [33]. Fully understanding the biosynthesis of MM 4550 in *S. argenteolus* will surely provide information for rational development of novel carbapenems in the future.

Conclusions

The isolation and characterization of the carbapenem MM 4550 from *S. argenteolus* ATCC 11009 coupled with the data resulting from DNA sequence and individual gene disruptions provide insights into the biosynthesis of the structurally complex carbapenem antibiotics. The consistencies and inconsistencies apparent in comparisons with the thienamycin gene cluster point to specific gene products to account for the structural differences between thienamycin and MM 4550. The absence of genes homologous to *thnA-D* and *thnV* suggest that their corresponding proteins are not involved in the assembly of the carbapenem, but the additional genes *cmmSu*, *cmm17* and *cmmPah* may well be responsible for the differing stereochemical features of the C-6 side chain. The knowledge provided by the sequencing of multiple carbapenem-producing gene clusters can offer insights into their genetic diversity and how that diversity translates into antibiotic structural variation. Exploration of this diversity may provide cost-effective fermentation methods to produce antibiotics or semi-synthetic strategies for the synthesis of new antibiotics to combat the ever-developing resistance of bacterial pathogens.

Experimental Section

Cloning of *thnM/carA* homologous gene in *S. argenteolus*

Degenerate PCR primers *thnM*-DEG-5 and *thnM*-DEG-3 were designed using the CODEHOP program [14] (Figure 1). The PCR reaction was performed using a “touchdown” program, which consisted of (1) denaturation at 98 °C for 5 min, (2) 7 cycles of 40 s at 98 °C, 40 s at 63–56 °C (decreasing 1 °C every cycle) and 1 min at 72 °C, (3) 30 cycles of 40 s at 98 °C, 40 s at 56 °C and 1 min at 72 °C, and (4) 10min at 72 °C. The 300-bp PCR product was cloned into pBluescript SK(+) and the resulting recombinant was subjected to DNA sequence analysis.

Disruption of *S. argenteolus* genes by λ -RED-mediated PCR targeting mutagenesis

A λ RED mediated PCR mutagenesis system developed by Gust et al. was employed for the inactivation of genes in the targeted gene cluster in *S. argenteolus* [27]. The extended *FRT*-

oriT-aac(3)IV-FRT cassette amplified by PCR was used to target each individual gene on cosmid 5D1 in *E. coli*. For the overlapped ORFs in the gene cluster, primers were designed from outside the overlapped regions but the translation remained in-frame. The PCR products were used to replace the corresponding genes in the cluster in *E. coli* BW25113 (pIJ790) by homologous recombination. The mutated cosmid was introduced into *E. coli* ET12567 (pUZ8002) for conjugation into wild-type *S. argenteolus* [27]. Disruption mutants generated by double crossover events were confirmed by a Km^S/Am^R phenotype and eventually by Southern hybridization using the gene-specific and *oriT-aac(3)IV* probes.

Construction of *S. argenteolus* *cmmI*-22-23

cmmI, *cmm22*, and *cmm23* were cloned by PCR from cosmid 5D1 (see Supporting Information) and cloned into pBluescript SK(+) to give recombinants pBS/*cmmI*, pBS/*cmm22*, and pBS/*cmm23*, respectively. *cmmI* was excised from pBS/*cmmI* with *NdeI*-*HindIII* and inserted into pUWL201PW digested with the same enzymes to generate plasmid pUWL201/*cmmI*. *cmm23* was also inserted into pUWL201PW at *NdeI*-*EcoRI* sites to yield plasmid pUWL201/*cmm23*. pUWL201/*cmm23* was digested with *KpnI* and blunt-ended with Klenow DNA polymerase followed by digestion with *EcoRI*. The blunt-*EcoRI* *ermE***p-cmm23* was cloned downstream of *cmmI* in pUWL201/*cmmI* at *HindIII*/Klenow-*EcoRI* sites to give recombinant pUWL201/*cmmI*-*cmm23*. *cmm22* was excised as a *NdeI*-*HindIII* fragment from pBS/*cmm22* and inserted into pET28(b) to yield plasmid pET28(b)/*cmm22*. The RBS-*cmm22* obtained by digesting pET28(b)/*cmm22* with *XbaI*-*HindIII* was inserted into pIJ4070 to give plasmid pIJ4070/*cmm22*. The *ermE***p*-RBS-*cmm22* was excised as a blunt-end fragment by digestion of pIJ4070/*cmm22* with *EcoRI*-*PstI* and followed by treatment with Klenow DNA polymerase. The *ermE***p*-RBS-*cmm22* was inserted into the blunt-ended *HindIII* site in pMS82 to give plasmid pMS82/*cmm22*. Finally, the *ermE***p-cmmI-ermE***p-cmm23* was excised as a *XhoI*-*EcoRI* fragment and blunt-ended with Klenow DNA polymerase and inserted into pMS82/*cmm22* at the blunt-ended *EcoRI* site to give the final integrative expression vector pMS82/*cmmI*-22-23.

S. argenteolus *cmmI*-22-23 was constructed by conjugating the wild-type *S. argenteolus* with *E. coli* ET12567 (pUZ8002, pMS82/*cmmI*-22-23).

Fermentation and detection of the wild-type, genetic-engineered and disruption mutants of *S. argenteolus*

The wild-type *S. argenteolus* and its derived strains were grown at 30 °C on S/P medium (1% glycerol, 1% yeast extract, 2% agar, pH 7.0) for solid cultivation, and on tomato paste-oatmeal agar for sporulation [34]. For liquid cultivation, mycelia or spores were inoculated in S/P liquid medium and grown for 72 h at 30 °C. Antibiotics were added in cultures if necessary at the following final concentrations: apramycin (50 µg/mL); nalidixic acid (20 µg/mL); hygromycin (50 µg/mL). For the general antibacterial activity test, 2.5 mL seed culture was transferred to 50 mL fermentation medium of either modified ISP4 medium (ISP4⁺: ISP4 plus 6 mg/L CoCl₂ • 6H₂O and pH 7.7) or OMYM medium [12b]. For isolation, purification, and characterization of intermediates or final products, the fermentation was carried out in a 1.3-L × 10 scale as follows: 100 mL S/P medium (containing antibiotics if necessary) was inoculated with 100 µL of spores at a concentration of ~10⁶⁻⁹/mL and grown at 30 °C for 48-56 h and 10 mL of seed culture was transferred into 1.3 L modified ISP4 medium (ISP4⁺⁺: ISP4 plus 6 mg/L CoCl₂ • 6H₂O and 1 mg/L methylcobalamin, pH 7.7). The fermentation was carried out at 30 °C for 52-56 h with 250 rpm rotation.

A 2-mL sample was taken every 24 h from each culture, and supernatant (250 µL) was added to a micro-cylinder on top of nutrient agar seeded with β-lactam supersensitive *E. coli* ESS [35]. The carbapenem MM 4550 and its bioactive intermediates could be visualized as

inhibition zones after incubating at 37 °C for 15 h. The production of carbapenems was also detected with induction of β -lactamases in *Bacillus licheniformis* ATCC 14580 in a colorimetric assay with Nitrocefin [28].

Isolation and identification of MM 4550 in *S. argenteolus*

Isolation and purification steps were performed at 4 °C, with the exception of preparatory HPLC, which was performed at room temperature. K_2HPO_4 buffer (20 mM, pH = 7.0) containing ethylenediaminetetraacetate (EDTA, 0.5 mM) was used to make all aqueous solutions in order to maintain the pH and provide a neutral environment to avoid degradation of carbapenem products.

To extract antibiotics from media, the filtrate from fermentation broth (13 L) was first washed with dichloromethane (600 mL). The resultant washed medium was then extracted with Aliquat 336 (5%, 300 mL) in dichloromethane. The medium was tested for bioactivity by *E. coli* ESS bioassay before and after extraction, which confirmed efficient extraction. The organic phase was then back-extracted with aqueous potassium iodide (KI, 4 × 25 mL), each of 1%, 3%, 5% and 10%. The 3% and 5% KI fractions contained the greatest quantity of antibiotic as shown by bioassay on *E. coli* ESS, and were collected and lyophilized.

The crude powder was dissolved in phosphate buffer and loaded onto a column of anion-exchange resin (Dowex 1 × 2, 200 × 20 mm, 100 mesh). The antibiotic was eluted with NaCl (1 M) and the bioactive fractions were lyophilized and dissolved in a minimum amount of phosphate buffer for purification by reverse-phase HPLC.

HPLC was carried out using an Agilent 1100 series G1311A pump with an Agilent model 7725i injector fitted with a 1-mL injection loop. The chromatogram was monitored by an Agilent 1100 series G1315B DAD detector at 210, 240, 270, 300, and 330 nm wavelengths.

HPLC conditions: Phenomenex Prodigy C18 (2) 250 × 10.00 mm 5 μ micron semi-preparatory column, 2 mL/min; buffer A = 20 mM K_2HPO_4 , 0.5 mM EDTA, pH = 7.4 with HCl, buffer B = acetonitrile. Method 1, t = 0 min 1.5% B, t = 15 min 1.5% B, t = 30 min 30% B, t = 35 min 30% B, t = 40 min 1.5% B, t = 55 min 1.5% B.

For the Hydroxamate assay, a stock solution (100 μ L) of hydroxylamine hydrochloride (50 mM) in phosphate buffer (pH 7.4) was added to the crude antibiotic-containing solution (900 μ L), and was incubated for 1 h at room temperature before it was injected for HPLC analysis using the method described above. The fractions containing a peak in the control that was absent when treated with hydroxylamine were collected and employed for bioassay on *E. coli* ESS.

To prepare samples for NMR analysis, the crude antibiotic-containing solution (50 mL) was HPLC purified using method described above and the peak that reacted with hydroxylamine was collected for all runs. The biologically active peak was collected and lyophilized. In order to remove excess salts from the resulting powder for NMR experimentation, the entire mixture was purified by one run of HPLC. The peak was collected in 1 mL fractions and lyophilized. Samples taken for NMR were dissolved in D_2O (333 μ L, 100%) and placed in a 5 mm D_2O Shigemi tube.

1H NMR (600 MHz, D_2O using Shigemi advanced NMR microtube, Bruker Avance II 600 spectrometer at 298 K, processed using the ACD labs software from Advanced Chemistry Development, Inc.): 7.60 (d, J = 13.7 Hz, 1 H), 6.41 (d, J = 14.7 Hz, 1 H), 4.94 (qd, J = 5.9, 14.7 Hz, 1 H), 4.49 (dt, J = 3.9, 5.9 Hz, 1 H), 3.98 (dd, J = 5.9, 8.8 Hz, 1 H), 3.52 (dd, J = 8.8, 17.6 Hz, 1 H), 3.12 (dd, J = 10.8, 17.6 Hz, 1 H), 2.14 (s, 3 H), 1.54 (d, J = 6.8 Hz, 3 H)

ESI Mass spec of MM 4550: $C_{13}H_{15}N_2O_9S_2^-$ calc: 407.0224 found: 407.0226. The analysis was run on a PerkinElmer AxIon TOF in the negative mode.

For mass-spectroscopic analysis of carbapenem intermediates in mutants, cultures grown in ISP4⁺⁺ were directly injected onto the UPLC-MS. UPLC-MS was carried out using a Waters Acquity H-Class UPLC system with a multi-wavelength UV-Vis diode array detector in tandem with high-resolution MS analysis by a Waters Xevo-G2 Q-ToF ESI mass spectrometer.

UPLC conditions: Waters Acquity BEH UPLC column packed with ethylene bridged hybrid C-18 2.1 mm × 50 mm, 1.7 micron, 0.3 mL/min; solvent A = H₂O, solvent B = acetonitrile. Method: t=0 min 0% B, t= 1 min 0% B, t = 7.5 min 80% B, t = 8.4 min 80% B, t = 10 min 0% B. Samples were run in the negative mode with a 1 min solvent purge after injection.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Ananya Majumdar and Dr. Cathy Moore for assistance with NMR spectroscopic analyses and Joshua Wilhide at the University of Maryland for mass spectral data. We are grateful to Prof. Susan E. Jensen (Univ. of Alberta) and Dr. Margaret C. M. Smith for providing the *E. coli* ESS and plasmid pMS82. We also thank Dr. Michael M. Freeman for informed discussion, Tegan Feehery for assistance with the early PCR experiments, Daniel Marous for the provision of PS-5 and the National Institute of Health (AI 014937) for financial support.

References

1. Suárez C, Lolans K, Villegas M, Quinn J. Expert Rev Anti Infect Ther. 2005; 3:915–922. [PubMed: 16307504]
2. Sleeman M, Schofield C. J Biol Chem. 2004; 279:6730–6736. [PubMed: 14625287]
3. Miller M, Gerrata B, Stapon A, Townsend C, Rosenzweig A. J Biol Chem. 2003; 278:40996–41002. [PubMed: 12890666]
4. a) Li R, Stapon A, Blanchfield J, Townsend C. JACS. 2000; 122:9296–9297. b) Stapon A, Li R, Townsend C. JACS. 2003; 125:8486–8493. c) Clifton I, Doan L, Sleeman M, Topf M, Suzuki H, Wilmouth R, Schofield C. J Biol Chem. 2003; 278:20843–20850. [PubMed: 12611886]
5. Nunez L, Mendez C, Brana A, Blanco G, Salas J. Chem Biol. 2003; 10:301–311. [PubMed: 12725858]
6. a) Bodner M, Li R, Phelan R, Freeman M, Moshos K, Townsend C. ChemBioChem. 2011; 12:2159–2165. [PubMed: 21913298] b) Bodner M, Phelan R, Freeman M, Li R, Townsend C. JACS. 2010; 132:12–13. c) Hamed R, Batchelar E, Mecinovi J, Claridge T, Schofield C. ChemBioChem. 2009; 10:246–250. [PubMed: 19090510]
7. Freeman M, Moshos K, Bodner M, Li R, Townsend C. PNAS. 2008; 105:11128–11133. [PubMed: 18678912]
8. Rodriguez M, Nunez L, Brana A, Mendez C, Salas J, Blanco G. Antimicrob Agents Chemother. 2011; 1638–1649. [PubMed: 21263049]
9. Rodriguez M, Nunez L, Brana A, Mendez C, Salas J, Blanco G. Mol Microbiol. 2008; 69:633–645. [PubMed: 19138192]
10. a) Okamura K, Hirata S, Koki A, Hori K, Shibamoto N, Okumura Y, Okabe M, Okamoto R, Kouno K, Fukagawa Y, Shimauchi Y, Ishikura T, Lein J. J Antibiot. 1979; 32:262–271. [PubMed: 468714] b) Shoji J, Hino H, Sakazaki R, Tsuji N, Nagashima K, Matsumoto K, Takahashi Y, Kosuki S, Hattori T, Kondo E, Tanaka K. J Antibiot. 1982; 35:15–23. [PubMed: 7040325] c) Butterworth D, Cole M, Hanscomb G, Rolinson G. J Antibiot. 1979; 32:287–294. [PubMed: 468715] d) Stapley E, Cassidy P, Tunac J, Monaghan R, Jackson M, Hernandez S, Zimmerman S, Mata J, Currie S. J Antibiot. 1981; 34:628–636. [PubMed: 7275847] e) Tsuji N, Kobayashi M,

- Terui Y, Matsumoto K, Takahashi Y, Kondo E. *J Antibiot.* 1985; 38:270–274. [PubMed: 3997673] f) Nakayama M, Kimura S, Tanabe S, Mizoguchi T, Watanabe I, Mori T. *J Antibiot.* 1981; 34:818–823. [PubMed: 7287584]
11. a) Tsuji N, Kondo E, Mayama M, Kawamura Y, Hattori T, Matsumoto K, Yoshida T. *J Antibiot.* 1981; 34:909–911. [PubMed: 6974725] b) Kawamura Y, Yasuda Y, Mayama M, Tanaka K. *J Antibiot.* 1982; 35:10–14. [PubMed: 7068508]
 12. a) Haneishi T, Nakajima M, Serizawa N, Inukai M, Takiguchi Y, Arai M, Satoh S, Kuwano H, Tamura C. *J Antibiot.* 1983; 36:1581–1584. [PubMed: 6654762] b) Nakamura Y, Ono E, Kohda T, Shibai H. *J Antibiot.* 1989; 42:73–83. [PubMed: 2784131]
 13. Hood J, Box S, Verrall M. *J Antibiot.* 1979; 32:295–304. [PubMed: 468716]
 14. Rose T, Schultz E, Henikoff J, Pietrovski S, McCallum C, Henikoff S. *Nucleic Acid Res.* 1998; 26:1628–1635. [PubMed: 9512532]
 15. Blanco G. *Arch Microbiol.* 2012; 194:549–555. [PubMed: 22200863]
 16. Hamed R, Batchelar E, Clifton I, Schofield C. *Cell Mol Life Sci.* 2008; 65:2507–2527. [PubMed: 18470480]
 17. Yan Y, Zhang L, Ito T, Qu X, Asakawa Y, Awakawa T, Abe I, Liu W. *Org Lett.* 2012; 14:4142–4145. [PubMed: 22861048]
 18. a) Negishi M, Pedersen L, Petrotchenko E, Shevtsov S, Gorokhov A, Kakuta Y, Pedersen L. *Arch Biochem Biophys.* 2001; 390:149–157. [PubMed: 11396917] b) Shi R, Lamb S, Bhat S, Sulea T, Wright G, Matte A, Cygler M. *J Biol Chem.* 2007; 282:13073–13086. [PubMed: 17329243] c) Kakuta Y, Pedersen L, Pedersen L, Niegishi M. *Trends Biochem Sci.* 1998; 23:129–130. [PubMed: 9584614]
 19. a) Paradkar A, Aidoo K, Jensen S. *Mol Microbiol.* 1998; 27:831–843. [PubMed: 9515708] b) Perez-Redondo R, Rodriguez-Garcia A, Martin J, Liras P. *Gene.* 1998; 211:311–321. [PubMed: 9602162]
 20. a) Stock J, Ninfa A, Stock A. *Microbiol Rev.* 1989; 53:450–490. [PubMed: 2556636] b) Hutchings M, Hoskisson P, Chandra G, Buttner M. *Microbiology.* 2004; 150:2795–2806. [PubMed: 15347739] c) Hwang Y, Lee J, Kim E, Choi C. *Biotechnol Lett.* 2001; 23:457–462.
 21. Claros M, von Heijne. *CABIOS, Comput Appl Biosci.* 1994; 10:685–686.
 22. Agnihotri G, Liu H. *Bioorg & Med Chem.* 2003; 11:9–20. [PubMed: 12467702]
 23. Christianson D. *Acc Chem Res.* 2005; 38:191–201. [PubMed: 15766238]
 24. Tabor C, Tabor H. *Annu Rev Biochem.* 1984; 53:749–790. [PubMed: 6206782]
 25. a) Nobary S, Jensen S. *Can J Microbiol.* 2012; 58:413–425. [PubMed: 22435762] b) Wu T, Busby R, Houston T, McIlwaine D, Egan L, Townsend C. *J Bacteriol.* 1995; 177:3714–3720. [PubMed: 7601835]
 26. Elkins J, Clifton I, Hernández H, Doan L, Robinson C, Schofield C, Hewitson K. *Biochem J.* 2002; 366:423–434. [PubMed: 12020346]
 27. Gust B, Challis G, Fowler K, Kieser T, Chater K. *PNAS.* 2003; 100:1541–1546. [PubMed: 12563033]
 28. Sykes R, Wells J. *J Antibiot.* 1985; 38:119–121. [PubMed: 3871760]
 29. a) Brown A, Corbett D, Eglington A, Howarth T. *J Chem Soc Chem Comm.* 1977:523–525. b) Corbett D, Eglington A, Howarth T. *J Chem Soc Chem Comm.* 1977:953–954.
 30. a) Ishizuka H, Horinouchi S, Kieser H, Hopwood D, Beppu T. *J Bacteriol.* 1992; 174:7585–7594. [PubMed: 1339426] b) Lu Y, Wang W, Shu D, Zhang W, Chen L, Qin Z, Yang S, Jiang W. *Appl Microbiol Biotechnol.* 2007; 77:625–635. [PubMed: 17899070]
 31. Kojima I, Fukagawa Y, Okabe M, Ishikura T, Shibamoto N. *J Antibiot.* 1988; 41:899–907. [PubMed: 3417564]
 32. Okabe M, Azuma S, Kojima I, Kouno K, Okamoto R, Fukagawa Y, Ishikura T. *J Antibiot.* 1982; 35:1255–1263. [PubMed: 7174513]
 33. Wise R. *Antimicrob Agents Chemother.* 1986; 30:343–349. [PubMed: 3535659]
 34. Williams, S.; Booth, C. *Actinomycetes.* Vol. 4. Academic Press; 1971.
 35. Aoki H, Sakai H, Kohsaka M, Konomi T, Hosoda J. *J Antibiot.* 1976; 29:492–500. [PubMed: 956036]

A)

AsnB1	318	YVKVILSGEGA	329	-	100aa	-	429	EFAQGCASSTVLE	441
AsnB2	318	YVKVILTGEGA	329	-	100aa	-	429	QFDEGSGTVDLLV	441
AsnB3	318	YVKVVLSGEGA	329	-	100aa	-	429	EFAQGCASSGVLE	441
AsnB4	353	GPLRILTGYGA	363	-	96aa	-	459	GIHEGSGTTSALT	471

		<u>Motif M1</u>					<u>Motif M2</u>		
BLS3	330	SGRRILTGYGA	340	-	96	aa	-	436	GVHEGSGVTAAFT 448
BLS	339	PERRILTGYGA	349	-	96	aa	-	445	GVHEGSGTTSSFS 457
CarA	336	QVSCMLTGYGS	346	-	99	aa	-	445	GIHEGSSVNQAF 457
CpmA	333	KVNSLITGYGS	343	-	99	aa	-	442	GIHEGSSVNKSFA 454
ThnM	295	ARQVMLSGYGA	305	-	96	aa	-	401	GIHEGTAMSRMFA 413
PluM	313	DRQVLLSGYGA	323	-	96	aa	-	419	GIHEGTAMSRMFA 430
		:::***:						*:***:. . *:	

B)

Primers:

thnM-DEG-5: 5'-GCGCCAGTGCATGCTGwsnggntaygg-3'**thnM-DEG-3:** 5'-CATGGCGGTGCCctcrtgdatncc-3'

Figure 1.

(a) The two conserved motifs in the β -lactam ring-closing enzymes. The names and access numbers are as follows: BLS3, *Streptomyces antibioticus* (AFH74298.1); BLS, *Streptomyces clavuligerus* ATCC 27064 (ZP_05002824.1); CarA, *Pectobacterium carotovorum* subsp. *brasiliensis* PBR1692 (ZP_03825848.1); CpmA: *Photorhabdus luminescens* subsp. *laumondii* TTO1 (NP_927548.1); ThnM, *Streptomyces cattleya* NRRL 8057 (YP_004920124.1); PluM, *Streptomyces sulfonofaciens* NRRL B-16438 (unpublished); AsnB1, *Nodularia spumigena* CCY9414 (YP_003135129.1); AsnB2, *Cyanothece* sp. CCY0110 (ZP_01729766.1); AsnB3, *Leptolyngbya* sp. PCC 7375 (ZP_18904938.1). BLS: β -lactam synthetase; AsnB: asparagine synthetase. The beginning and ending numbers indicate the position of amino acids in the proteins, and the numbers between the motifs indicate the separation in amino acid residues. (b) Primers designed for the M1 and M2 motifs according to the CODEHOP program. N: A/T/G/C; Y: C/T; M: A/C; R: A/G.

Figure 2.

Gene organization of the MM 4550 and the thienamycin gene clusters in *S. argenteolus* ATCC 11009 (top) and *S. cattleya* NRRL8057 (below). The direction of transcription and the proposed functions of individual ORFs are indicated.

Figure 3.

(A) HPLC analysis of *S. argenteolus* extracts. **1.** Wild-type *S. argenteolus* ATCC 11009; **2.** *S. argenteolus* cmmI-22-23; **3.** Wild-type *S. argenteolus* ATCC 11009 spiked with isolated peak from *S. argenteolus* cmmI-22-23; **4.** *argE8* disruption mutant. Peaks (absorbance in arbitrary units) corresponding to MM4550 are indicated by a heavy dot (•). (B) Bioassay of the 16.3-min HPLC fractions of *S. argenteolus* cmmI-22-23 (left) and *argE8* (right).

Figure 4.

(**A**) The organization of *cmmSu*, *cmmE*, and *cmmF* in cosmid 5D1 and the construction of the *argE8* disruption mutant. The restriction maps of the wild-type *cmmE* gene and its disrupted copy in the mutant are shown. (**B**) Southern analysis of the wild-type and the *argE8* ($\Delta cmmE$) mutant, showing the positive hybridizing bands to the *cmmE* and the *oriT-aac(3)IV* probes, respectively.

MM 4550

Carbapen-2-em-3-carboxylic acid

Thienamycin

Asparenomyacin A

Scheme 1.
Representative carbapenems.

Scheme 2.

Proposed biosynthetic pathway to MM 4550 in *S. argenteolus* ATCC 11009. Solid arrows indicate steps thought to correspond to those in thienamycin biosynthesis in *S. cattleya*. (Pant = pantetheine).

Table 1
Deduced functions of proteins encoded by the MM 4550 biosynthetic gene cluster

Protein	Amino acids	Proposed function	Sequence similarity (identity/positive)	Identity/ positive to ATCC 33331(%)
Orf1	238	Unknown	No similar sequences	--/-- ^a
Orf2	475	arabinofuranosidase	SMCF_639, <i>Streptomyces coelicoflavus</i> ZG0656 (89/92)	--/--
CmmSu	353	sulfotransferase	Sfla_0165, <i>S. flavogriseus</i> ATCC 33331 (87/92)	87/92
CmmE	263	carboxymethylproline synthase	ThnE, <i>S. cattleya</i> NRRL 8057 (69/80)	93/96
CmmF	338	N-acetyltransferase	ThnF, <i>S. cattleya</i> NRRL 8057 (58/66)	86/89
CmmG	263	oxidoreductase	ThnG, <i>S. cattleya</i> NRRL 8057 (70/80)	87/92
CmmH	229	HAD-superfamily hydrolase	ThnH, <i>S. cattleya</i> NRRL 8057 (57/69)	82/87
CmmI	466	LysR family transcriptional regulator	ThnI, <i>S. cattleya</i> NRRL 8057 (51/64)	83/87
CmmJ	487	Transporter protein	ThnJ, <i>S. cattleya</i> NRRL 8057 (56/68)	89/93
CmmK	680	radical SAM methyltransferase	ThnK, <i>S. cattleya</i> NRRL 8057 (79/86)	95/96
CmmL	394	radical SAM methyltransferase	ThnL, <i>S. cattleya</i> NRRL 8057 (67/72)	82/86
CmmM	459	Carbapenam synthetase	ThnM, <i>S. cattleya</i> NRRL 8057 (64/76)	86/91
CmmN	369	AFD class I superfamily protein, acyl-CoA synthase	ThnN, <i>S. cattleya</i> NRRL 8057 (86/91)	95/97
CmmO	467	aldehyde dehydrogenase	ThnO, <i>S. cattleya</i> NRRL 8057 (75/82)	90/94
CmmP	482	radical SAM methyltransferase	ThnP, <i>S. cattleya</i> NRRL 8057 (83/91)	97/98
CmmQ	259	putative oxygenase	ThnQ, <i>S. cattleya</i> NRRL 8057 (72/81)	93/94
Cmm17	292	enoyl-CoA hydratase	Kyg_13766, <i>Acidovorax</i> sp. NO-1 (34/46)	89/93
CmmR	228	nudix hydrolase	ThnR, <i>S. cattleya</i> NRRL 8057 (55/66)	75/82
CmmT	388	hydrolase	ThnT, <i>S. cattleya</i> NRRL 8057 (69/75)	87/90
CmmS	334	β-lactamase	Orf11, <i>S. clavuligerus</i> ATCC 27064 (63/71)	80/87
CmmPah	298	proclavamate amidino hydrolase	PAH, <i>S. clavuligerus</i> ATCC 27064 (52/65)	86/90
Cmm22	227	two-component system response regulator	SPW_1836, <i>Streptomyces</i> sp. W007 (84/90)	--/--
Cmm23	506	two-component system sensor kinase	Sacte_5700, <i>Streptomyces</i> sp. SirexAA-E (89/93)	--/--
Cmm24	217	unknown	Sacte_5699, <i>Streptomyces</i> sp. SirexAA-E (74/79)	--/--
Orf25	1094	putative peptidase	Sacte_5698, <i>Streptomyces</i> sp. SirexAA-E (81/88)	--/--
Orf26	1029	putative glycosyl hydrolase	Szn_17732, <i>Streptomyces zinciresistens</i> K42 (78/88)	--/--
Orf27	294	putative RNA polymerase sigma factor	Sfla_6134, <i>S. flavogriseus</i> ATCC 33331 (88/93)	--/--
Orf28	441	putative serine/threonine phosphatase	Sfla_6135, <i>S. flavogriseus</i> ATCC 33331 (85/91)	--/--
Orf29	164	putative MarR family transcriptional regulator	S.fl_6136, <i>S. flavogriseus</i> ATCC 33331 (93/95)	--/--

^a --/--: not in the cluster.

Table 2
Analysis of the disruption mutants of *S. argenteolus*

Disruption mutants	<i>E. coli</i> ESS	<i>B. licheni-formis</i>	Nitrocefin
argSu13	+	-	+
argE8	-	-	-
argF7	-	-	-
argG5	+	+	+
argH8	+	+	+
argJ4	-	-	-
argK11	-	-	-
argL2	-	-	-
argM7	-	-	-
argN7	-	-	-
argO2	-	-	-
argP4	-	-	-
argQ4	+	-	+
arg17-4	-	-	+
argR8	-	-	-
argT3	-	-	-
argS6	+	+	+
argPah9	-	-	-
arg22-6	-	-	-
arg24-3	+	+	+
WT	+	+	+

Table 3
HPLC-MS analysis of carbapenems from *S. argenteolus*

Strain	Found (m/z)	Calculated (m/z)	Molecular Formula	Proposed Structure
WT	407.0215	407.0224	$C_{13}H_{15}N_2O_9S_2^-$	MM 4550
argSu13	313.0848	313.0858	$C_{13}H_{17}N_2O_5S^-$	Epithienamycin A
argQ4	297.0909	297.0909	$C_{13}H_{17}N_2O_4S^-$	PS-5
argG5	393.0404	393.0426	$C_{13}H_{17}N_2O_8S_2^-$	MM 17880
argE8	n/a	n/a	n/a	no carbapenem masses found