Chemistry & Biology 13, 733–743, July 2006 DOI 10.1016/j.chembiol.2006.05.009

Deciphering Indolocarbazole and Enediyne Aminodideoxypentose Biosynthesis through Comparative Genomics: Insights from the AT2433 Biosynthetic Locus

Qunjie Gao,[1,2] Changsheng Zhang,[1,2] Sophie Blanchard,[1,2] and Jon S. Thorson[1,2,3,*]
[1] Laboratory for Biosynthetic Chemistry
Pharmaceutical Sciences Division
School of Pharmacy
University of Wisconsin, Madison
777 Highland Avenue
Madison, Wisconsin 53705
[2] University of Wisconsin National Cooperative Drug Discovery Group Program
Madison, Wisconsin 53705

Summary

AT2433, an indolocarbazole antitumor antibiotic, is structurally distinguished by its aminodideoxypentose-containing disaccharide and asymmetrically halogenated *N*-methylated aglycon. Cloning and sequence analysis of AT2433 gene cluster and comparison of this locus with that encoding for rebeccamycin and the gene cluster encoding calicheamicin present an opportunity to study the aminodideoxypentose biosynthesis via comparative genomics. The locus was confirmed via in vitro biochemical characterization of two methyltransferases—one common to AT2433 and rebeccamycin, the other unique to AT2433—as well as via heterologous expression and in vivo bioconversion experiments using the AT2433 *N*-glycosyltransferase. Preliminary studies of substrate tolerance for these three enzymes reveal the potential to expand upon the enzymatic diversification of indolocarbazoles. Moreover, this work sets the stage for future studies regarding the origins of the indolocarbazole maleimide nitrogen and indolocarbazole asymmetry.

Introduction

Deoxysugars are an essential class of naturally occurring carbohydrates. These deoxygenated and often highly functionalized sugars present distinctive hydrophobic and hydrophilic characteristics critical to their role in dictating specificity on a tissue, cellular, and/or molecular level. Among the deoxysugars, deoxyhexoses are arguably the most diverse and best studied to date [1, 2]. In contrast, while functionalized deoxypentoses append a variety of diverse bioactive bacterial secondary metabolites (Figure 1A), the biosynthesis of these important sugar attachments remains largely undefined. Classical metabolic labeling of esperamicin (Figure 1A, **3**) suggested that the esperamicin aminodideoxypentose (2,4-dideoxy-4-methylamino-α-L-xylopyranoside) derived from glucose [3], presumably by loss of C-6 as CO_2 in a fashion reminiscent of the UDP-D-xylose pathway common to primary metabolism [4, 5]. The elucidation of the calicheamicin (Figure 1A, **1**) gene locus [6], followed by the overexpression and biochemical characterization of CalS8 as a UDP-α-D-glucose dehydrogenase, provided further support for C-6 oxidative decarboxylation as a general point of divergence to commit nucleotide sugars to aminodideoxypentose biosynthesis [7]. While the early studies on the biosynthesis of avilamycin (Figure 1A, **4**) implicated ribose as the progenitor of the deoxypentose [8], recent work by Bechthold and coworkers revealed AviE2 to function as a UDP-α-D-glucuronic acid decarboxylase—also consistent with glucose as the deoxypentose progenitor in orthosomycin biosynthesis [9].

Among the naturally occurring indolocarbazole alkaloids, AT2433 (Figure 1B, **6**) shares the unique aminodideoxypentose common to the enediynes calicheamicin (Figure 1A, **1**) and esperamicin (Figure 1A, **3**) [10, 11]. The indolocarbazoles—bacterial secondary metabolites with potent antitumor and neuroprotective properties—all share a common indolo[2,3-*a*]pyrrolo[3,4-*c*]carbazole core but they are often subdivided into two structural subgroups [12]. The first includes AT2433 (Figure 1B, **6**) and rebeccamycin (Figure 1B, **7**) and is structurally defined by a common β-*N*-glucoside appendage. Rebeccamycin is a potent stabilizer of the topoisomerase I-DNA covalent complex. However, the addition of the cationic aminodideoxypentose in **6** abolishes topoisomerase I inhibition while enhancing DNA affinity [13–15]. The second indolocarbazole structural group, exemplified by staurosporine (Figure 1B, **8**), contains a signature bridging of indole nitrogens by a single glycosyl moiety at C-1′ and C-5′, and members of this structural subgroup are potent inhibitors of protein kinases A, C, and K [16, 17]. The clinical development of several indolocarbazole derivatives from both classes has been pursued, including the potent topoisomerase I poison NB-506, the promising kinase inhibitors UCN-01, CEP-701/751, and K252a, and a water-soluble analog (NSC 655649) which, by virtue of semisynthetic modification, is a potent topoisomerase II inhibitor [18–22].

Much is known regarding the biosynthesis of indolocarbazoles. For example, the gene clusters encoding for the production of **7** (from *Saccharothrix aerocolonigenes* ATCC39243) [23–25] and **8** (from *Streptomyces* sp. TP-A0274) [26] have been elucidated and many of the gene functions subsequently established via gene inactivation or heterologous expression. A recent elegant heterologous combinatorial in vivo reconstitution of components of the **7** pathway also led to the identification of 12 bisindole intermediates [27] and a five-stage biosynthetic pathway for indolocarbazole natural products based upon the **7** pathway has been proposed [23–26]. With respect to enzymes required for **7** biosynthesis, RebG [28], RebH [29], RebO [30, 31], RebD [31], and RebM [28, 32] have been confirmed biochemically as the *N*-glucosyltransferase, tryptophan halogenase, 7-chloro-tryptophan oxidase, chromopyrrolic acid synthase, and sugar-4-*O*-methyltransferase, respectively. Recent efforts toward the enzymatic diversification of indolocarbazoles revealed RebG to *N*-glucosylate unnatural aglycons and also provided indolocarbazole

*Correspondence: jsthorson@pharmacy.wisc.edu
[3] Lab address: http://www.pharmacy.wisc.edu/SOPDir/PersonDetails.cfm?&ID=63

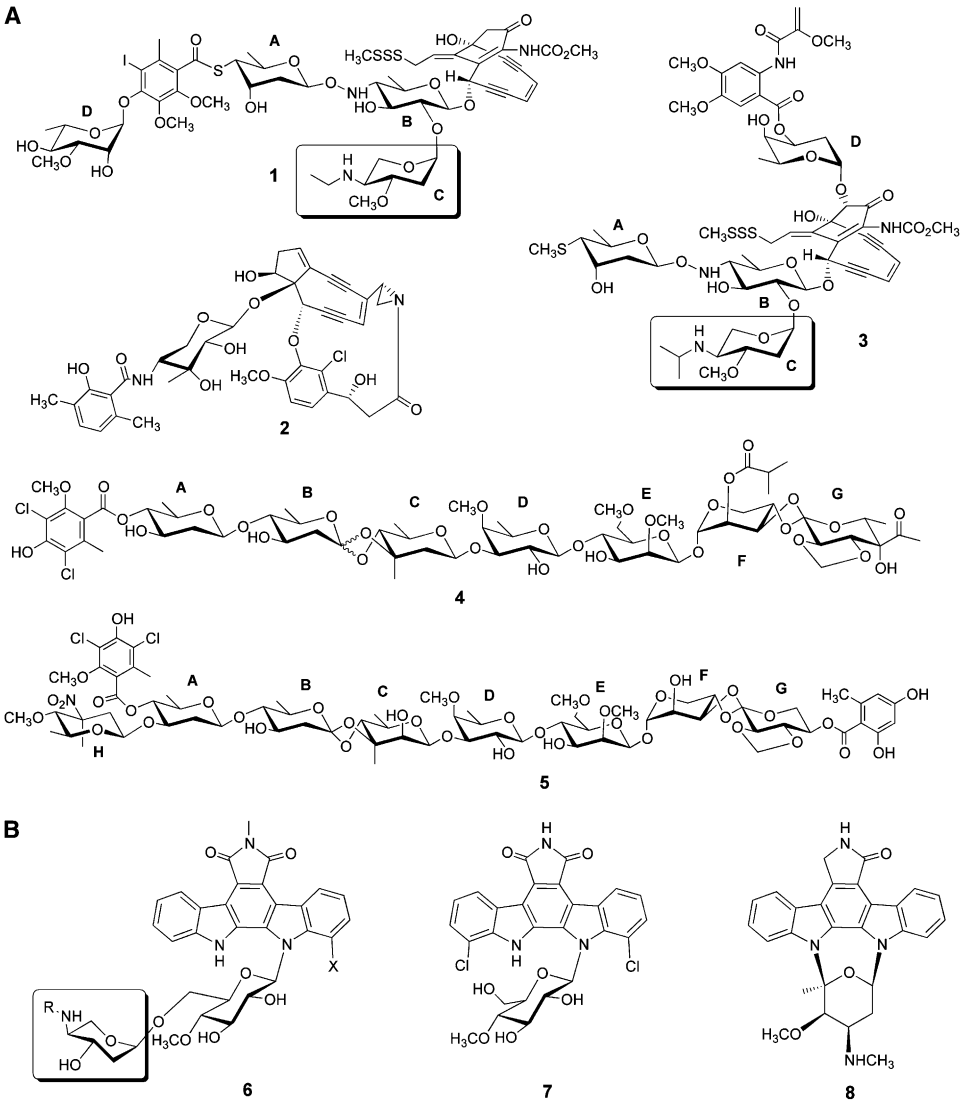

Figure 1. The Structures of Pentose-Containing and Indolocarbazole Natural Products

The structures of (A) pentose-containing natural products calicheamicin (**1**), maduropeptin (**2**), esperamicin (**3**), avilamycin (**4**), evernimicin (**5**), and (B) indolocarbazoles AT2433 (**6**, X = H or Cl), rebeccamycin (**7**), and staurosporine (**8**). The aminodideoxypentose common to **1**, **3**, and **6** has been highlighted by boxes.

N-glucoside regioisomers [28], while RebM could alklyate unnatural substrates and efficiently utilize non-natural cofactor analogs [28, 32]. The late steps of **8** biosynthesis have also been recently elucidated via in vivo strategies to be *N*-12 glycosylation (StaG) followed by a unique P450 (StaN)-catalyzed formation of the second C-5′-carbohydrate-*N*-13-indole connection, and StaG was also utilized for indolocarbazole diversification [33].

Despite the remarkable body of work described above, the study of **6** biosynthesis remains an attractive target, as this pathway promises a variety of new biosynthetic targets and tools. In comparison to other indolocarbazole alkaloids, **6** is the only naturally occurring indolocarbazole to have an *N*-methylated and asymmetrically halogenated indolocarbazole core. In addition, **6** contains a novel cationic pentose—a structural motif, as described above, also shared by enediynes **1** and **3**, which serves as a key DNA recognition element for these structurally diverse natural products. This unique structural, and presumably genotypic, element shared by **1**, **3**, and **6** suggests that a comparison between the **1**, **6**, and **7** gene clusters should quickly identify genes common for aminopentose biosynthesis and thereby present a basis from which to more accurately predict the biosynthesis of uniquely functionalized pentoses. Here we report the elucidation of the **6** biosynthetic gene cluster from *Actinomadura melliaura*. The locus was confirmed via in vitro biochemical characterization of two methyltransferases (one common to **6** and **7**, the other unique to **6**) as well as via heterologous expression and in vivo bioconversion experiments using the **6** *N*-glycosyltransferase. Preliminary studies of substrate tolerance for these three enzymes reveal the potential opportunity to utilize these enzymes to expand upon the general enzymatic diversification of indolocarbazoles.

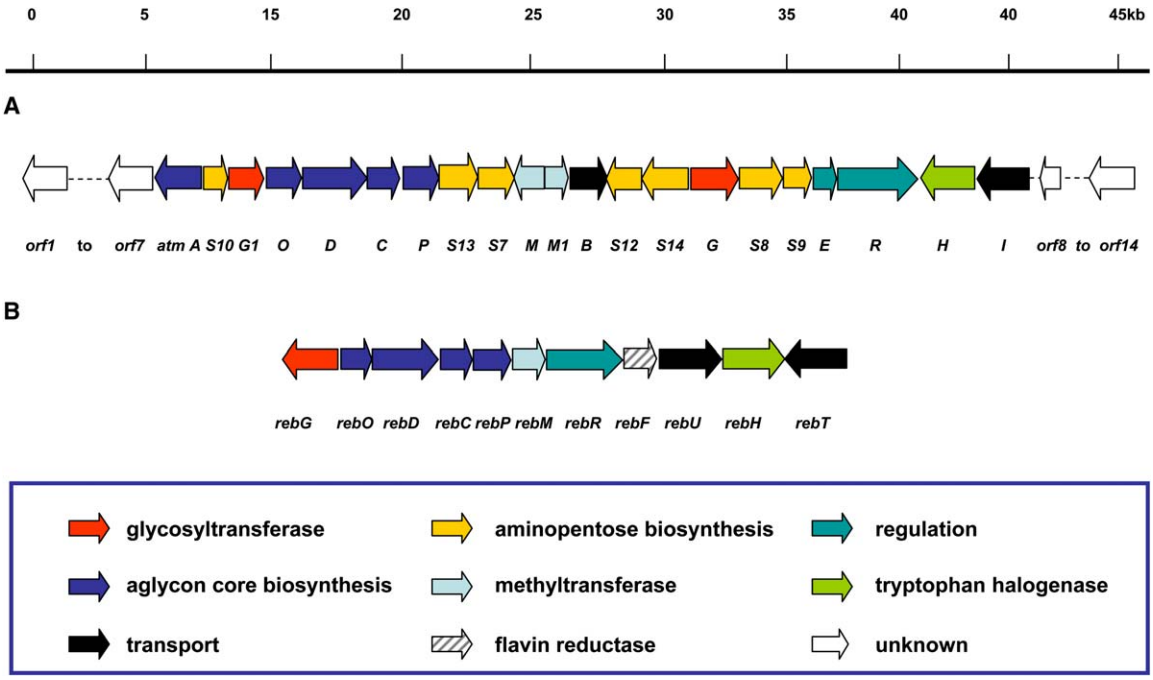

Figure 2. Indolocarbazole Biosynthetic Gene Loci

(A) The organization of AT2433 biosynthetic gene cluster and (B) comparison with the rebeccamycin biosynthetic locus. Arrows represent the relative ORF size and direction of transcription. Colors designate putative or known functionality. The putative or known functions associated with each of the genes and closest homologs are outlined in Table 1.

In addition, a comparison of the **1**, **6**, and **7** gene loci provides, for the first time, a clear biosynthetic model for the aminodideoxypentose common to these structurally diverse natural products.

Results and Discussion

Cloning and Sequencing of the AT2433 Gene Cluster

Given the notable structural and putative biosynthetic similarities between **6** and **7**, degenerate primers designed to amplify two genes essential to biosynthesis of the indolocarbazole core—*rebD*, which encodes the chromopyrrolic acid synthase, and *rebP*, which encodes for a putative P450 oxidase—were used to amplify the corresponding homologs (designated *atmD* and *atmP*, respectively) from *A. melliaura* genomic DNA. The PCR-amplified fragments were confirmed by sequencing and the confirmed *A. melliaura rebD* and *rebP* gene homologs were subsequently employed as DNA probes to screen an *A. melliaura* genomic library.

The *A. melliaura* genomic cosmid library was constructed in SuperCos2 and approximately 4000 clones from this library were screened by colony hybridization using digoxigenin (DIG)-labeled *atmD* and *atmP*. The positive clones were confirmed by PCR amplification with the same degenerate primers originally used to amplify *atmD* and *atmP* from the genomic template. Shotgun sequencing of one positive from this screen, cosmid pJST1004, revealed a 43 kb insert containing five genes highly homologous to genes involved in **7** and **8** biosynthesis and genes encoding for NDP-sugar biosynthesis. Chromosomal walking using DNA probes designed from each end of the cosmid pJST1004 insert, followed by shotgun sequencing of a total of 85 kb of contiguous *A. melliauri* genomic DNA, allowed for the completion of the putative biosynthetic locus for **6**.

The genomic sequence of this entire 85 kb fragment (with an average GC content of 70%) was analyzed by the FRAME program [34] to reveal 48 putative open reading frames (ORFs), the preliminary annotation of which derived from BLAST analysis [35]. The sequence postulated to be relevant to **6** biosynthesis is highlighted in Figure 2 and resides in a 45 kb fragment containing 35 ORFs designated *orf1–orf14*. Consistent with the structure of **6**, the putative ORFs identified consist of six genes (*atmA*, *C*, *D*, *H*, *O*, and *P*) involved in the indolocarbazole core biosynthesis, nine genes (*atmG*, *G1*, *S7–S10*, and *S12–S14*) involved in the construction and attachment of carbohydrates, three methyltransferase genes (*atmM*, *M1*, and *S10*), and four genes (*atmB*, *E*, *I*, and *R*) related to resistance or regulation. The proposed function for each of the 35 ORFs and their closest homologs are listed in Table 1. The sequence upstream (~6 kb) to *orf1* encodes for primarily hypothetical proteins (average GC content of 64.2 ± 1.5%; 86.2 ± 2.6% GC in the wobble position). In contrast, *orf1–7* displayed an average GC content of 73.3 ± 1.6% (90.6 ± 2.1% GC in the wobble position) and the *atm* locus displayed an average GC content of 71.0 ± 2.5% (93.3 ± 3.2% GC in the wobble position). Similar to the *orf1* 5′ region, the sequence downstream of the *atm* locus (*orf8–14*) also primarily encodes for hypothetical proteins (average GC content of 67.6 ± 2.1%; 84.7 ± 3.6% GC in the wobble position). On the basis of

Table 1. Deduced Functions of the ORFs in Figure 2

Gene	Amino Acids	Proposed Function	Protein Homolog (% Identity/% Similarity)	Accession Number
orf1	509	Unknown protein	–	–
orf2	372	Unknown protein	–	–
orf3	275	Unknown protein	–	–
orf4	879	Unknown protein	–	–
orf5	496	Unknown protein	–	–
orf6	618	Hypothetical protein	BL02838 (43/59)	AAU22609.1
orf7	669	Hypothetical protein	BT9727_1156 (25/43)	YP_035491.1
atmA	602	Amidotransferase	OxyD (56/65)	AAZ78328.1
atmS10	238	Putative aminomethylase	CalS10 (56/72)	AAM70334.1
atmG1	400	*O*-glycosyltransferase	MycB (37/52)	AB089954
atmO	478	Putative L-tryptophan oxidase	RebO (64/76)	AAN01208.1
atmD	1020	CCA synthetase	StaD (55/65)	AB088119
atmC	534	Putative FAD-monooxygenase	RebC (63/74)	AB090952
atmP	401	Cytochrome P450	RebP (59/70)	SAE414559
atmS13	369	Aminotransferase	CalS13 (62/76)	AAM94797.1
atmS7	352	Putative dTDP-glucose synthase	StrD (57/72)	AJ862840
atmM	268	Putative D-glucose *O*-methyltransferase	RebM (56/70)	SAE414559
atmM1	195	SAM-dependent methyltransferase	Chlo02004218 (43/53)	NZ_AAAH01001114
atmB	435	Putative antiporter	Orf7 (64/80)	AME16952
atmS12	327	dTDP-4-keto-6-deoxy-L-hexose-2,3-reductase	CalS12 (66/76)	AAM70349.1
atmS14	483	Putative NDP-hexose-2,3-dehydratase	CalS14 (51/65)	AAM70359.1
atmG	445	*N*-glycosyltransferase	RebG (58/71)	SAE414559
atmS8	450	UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase	CalS8 (58/71)	AAM70332.1
atmS9	329	UDP-glucose 4-epimerase; UDP-glucuronate decarboxylase	CalS9 (60/72)	AAM70333.1
atmE	175	Regulatory protein	MarR (33/54)	NZ_AAII01000049
atmR	933	Putative regulatory protein	RebR (43/56)	CAC93719
atmH	535	Tryptophan halogenase	RebH (73/85)	SAE414559
atmI	485	Putative transmembrane efflux protein	SCO3199 (44/63)	SCO939115
orf8	177	Hypothetical protein	MAP2477c (50/66)	NP_961411.1
orf9	207	Transposase	IS4 (52/66)	YP_482468.1
orf10	201	Hypothetical protein	Nfa30620 (38/56)	YP_119273.1
orf11	242	Putative regulator	LysR (68/76)	BAC68895.1
orf12	166	Hypothetical protein	SG7F10.43c (72/81)	CAH94367.1
orf13	529	Hypothetical protein	Nfa39620 (58/72)	YP_120174.1
orf14	520	Unknown protein	–	–

functional and GC content, we postulate the minimal **6** biosynthetic gene cluster to be contained between *atmA* and *atmI*.

Characterization of *N*-Glucosyltransferase AtmG

Of the two putative glycosyltransferase genes (*atmG* and *atmG1*) within **6** gene cluster, *atmG* was proposed to encode the requisite *N*-glycosyltransferase based upon homology to the known indolocarbazole *N*-glycosyltransferase genes *rebG* and *staG*. Overproduction of AtmG in *Escherichia coli* under a variety of conditions led to insoluble protein, consistent with previous work on RebG [28]. Following the strategy previously employed [28, 33, 36], the same AtmG-*E. coli* overexpression strain was subsequently analyzed via bioconversion for its ability to *N*-glycosylate indolocarbazoles. Bioconversion of aglycon surrogates **9** and **12** followed by HPLC analysis (Figure 3A) revealed new products with characteristic indolocarbazole absorption profiles. LC-MS characterization was consistent with the formation of products **10/11** with an estimated bioconversion yield of 99% and trace production of **13/14**, respectively. Similar biochemical studies with RebG revealed the same ability to generate a mixture of regioisomers (**10/11** and **13/14**) [28]. Under the same conditions, only starting material was recovered from control *E. coli* strains containing an expression vector lacking the *N*-glucosyltransferase gene. Notably, the outcome of this cumulative set of experiments is clearly consistent with the gene *atmG* as encoding the requisite *N*-glucosyltransferase involved in **6** biosynthesis.

Characterization of *O*- and *N*-Methyltransferases AtmM and AtmM1

Based upon homology to RebM, AtmM was proposed to function as the required Glc-4′-*O*-methyltransferase in the AT2433 pathway. The recent biochemical characterization of RebM revealed the rebeccamycin Glc-4′-*O*-methylase to utilize both "unnatural" substrates and *S*-adenosyl-methionine surrogates [28, 32]. Paralleling the earlier RebM study, the *atmM* gene was overexpressed in *E. coli* and the corresponding AtmM purified to homogeneity. Figure 3B reveals the purified AtmM could efficiently methylate a variety of indolocarbazoles in vitro. Specifically, compounds **10**, **13**, and **15–19** were converted to compounds **20**, **21**, and **22–26**, with yields ranging from 8% to 99%. As controls, product was not detected in assays lacking AtmM or SAM or in assays with 4′-*O*-methylated substrates (e.g., Figure 3C, **7**). All products were characterized by LC-MS and compounds **20–22** and **24** were shown to coelute with previously characterized standards [28]. Interestingly, AtmM displayed a slightly broader substrate scope in comparison to RebM and allowed for the production of new compounds **25** and **26**. Most important, the outcome of this cumulative set of experiments is also clearly consistent

with the gene *atmM* as encoding the requisite Glc-4′-*O*-methylase involved in **6** biosynthesis.

The sequence analysis of *atmM1* revealed a thiopurine *S*-methyltransferase domain implicating AtmM1 as the putative candidate for the **6** maleimide *N*-methyltransferase. To assess this postulation, *atmM1* was overexpressed in *E. coli* and the catalyst tested both in vivo and in vitro. The *atmM1-E. coli* strain was found to convert **17**β to **27**β via bioconversion (data not shown), while in vitro assays revealed AtmM1 could methylate a variety of indolocarbazoles (Figure 3C). Specifically, compounds **17**β, **24**β, **30**, and **7** were converted to compounds **27**β, **28**β, **31**, and **32** (the complete characterization of which has been previously reported [37–40]), respectively, with yields ranging from 12% to 96% as determined via LC-MS. As controls, products were not detected in assays lacking AtmM1 or SAM. While *N*-methylmaleimido analogs (**16**α, **16**β, and **19**) and derivatives lacking an aglycon carbonyl (**9**, **10/11**) could be processed by AtmG and/or AtmM, neither set could be methylated by AtmM1, further supporting the regiospecificity of AtmM1 and the importance of the C-7 carbonyl for recognition by AtmM1. Notably, the outcome of this cumulative set of experiments is clearly consistent with the gene *atmM1* as encoding the requisite maleimide *N*-methylase involved in **6** biosynthesis. Unlike the confirmed steps catalyzed by AtmG and AtmM described above, which are common to both **6** and **7** biosynthesis, the AtmM1-catalyzed *N*-methylation is unique to the **6** pathway and thus strongly supports our assignment of this locus as encoding for **6** biosynthesis.

A variety of assays were subsequently performed in an attempt to delineate the timing of *N*-glucosylation and *N*- versus *O*-methylation in **6** biosynthesis. Under identical conditions, the *N*-methylation of **7**, **17**β, or **30** led to the production of **32**, **27**β, or **31** in 38%, 96%, and 12.4% yield, respectively (Figure 3C). While not a direct kinetic assessment, this preliminary study implicates *N*-glycosylated indolocarbazoles to be slightly better substrates for AtmM and thus suggests *N*-methylation, like *O*-methylation [41], may occur after *N*-glucosylation. A similar preliminary study designed to probe *N*- versus *O*-methylation (AtmM1 versus AtmM) revealed the AtmM1-catalyzed conversion of **24**β to **28**β to proceed with 86% yield (Figure 3C, iii), while AtmM-catalyzed conversion of **27**β to **28**β proceeded in 67% yield (Figure 3C, iv). This study, again while not a kinetic analysis, suggests the order of *N*- or *O*-methylation to be of fairly indiscriminate nature.

In the context of natural products, simple modifications, such as methylation, can wield remarkable effects upon their pharmacological properties [42–44]. For example, alterations of indolocarbazole sugar-4′-*O*-methylation and/or the aglycon maleimide *N*-methylation modulate cytotoxicity and/or the specific molecular interactions between the indolocarbazole and DNA and/or protein targets [28, 43–45]. Moreover, alterations of these specific molecular interactions between the small molecule and the target are not always predictive of the ultimate cytotoxic potency. The recent biochemical characterization of RebM revealed the rebeccamycin Glc-4′-*O*-methylase to utilize both "unnatural" substrates and *S*-adenosyl-methionine cofactor analogs [28, 32]. Given the slightly broader substrate scope demonstrated by AtmM (in comparison to RebM), and the demonstrated ability of AtmM1 to also *N*-methylate indolocarbazole surrogates, these new indolocarbazole-modifying enzymes may prove to be beneficial new tools for extending the structural diversity of this clinically relevant class of natural product.

Aminopentose Biosynthesis and Attachment

Sequence analysis of the newly confirmed **6** gene cluster revealed seven NDP-sugar biosynthetic genes (*atmS7–10*, *atmS12–14*). Interestingly, homologs for each of these **6** NDP-sugar genes (*calS7–10*, *calS12–14*) were also found in the **1** biosynthetic locus with % identity/% similarity ranging from 51/65 to 66/76. In comparison, the seven genes in **6** gene cluster are more closely colocalized than their counterparts in the **1** locus. Based on the putative functions of this conserved set of gene products, a model for aminodideoxypentose biosynthesis is proposed (Figure 4A). In this set, the deduced product of *atmS7* is similar to a family of glucose-1-phosphate nucleotidylyltransferases which initiate most sugar biosynthetic pathways by presenting the sugar nucleotide. The deduced product of *atmS8* resembles CalS8, a biochemically characterized UDP-α-D-glucose dehydrogenase (UDPGlcDH) previously postulated to be involved in the **1** aminodideoxypentose biosynthetic pathway [7]. Both AtmS9 and CalS9 show considerable similarity to UDP-glucose 4-epimerases of *Bacillus cereus* ATCC 14579 (38% identity, 59% similarity; Table 1) and UDP-glucuronate decarboxylases of *Mus musculus* (33% identity, 50% similarity; Table 1). Several UDP-glucuronate decarboxylases have been characterized in the biosynthetic pathway of UDP-xylose required for primary metabolism in plants [4, 46], vertebrates [47–49], fungi [50], and bacteria [51], and Hofmann et al. recently confirmed AviE2 as the UDP-glucuronate decarboxylase en route to L-lyxose [9].

The final aminodideoxypentose product requires C-2 deoxygenation, a step known to require two enzymes (NDP-4-ketosugar 2,3-dehydratase and subsequent 2,3-reductase) within the biosynthesis of 2,6-dideoxyhexoses [52, 53]. Genes encoding for homologs for a putative 2,3-dehydratase (AtmS14/CalS14) and potential 2,3-reductase (AtmS12/CalS12) are conserved within both the **1** and **6** loci, and thus this comparison presents the first compelling evidence from which to propose a similar strategy for aminodideoxypentose C-2-deoxygenation. The deduced products of *atmS13*/*calS13* resemble NDP-sugar aminotransferases [54–59] and are postulated to catalyze the subsequent amine installation at the C-4 position. Previous studies, in collaboration with Liu and coworkers, identified CalS13 as the TDP-6-deoxy-α-D-glycero-L-threo-4-hexulose-4-aminotransferase in the **1** pathway [60]. In contrast, the current comparative genomics approach clearly reveals CalS13 to be the AtmS13 homolog (62% identity, 76% similarity) and thereby implicates a potential dual activity for this aminotransferase. While this needs to be confirmed biochemically, to our knowledge this would be the first example of an NDP-ketosugar aminotransferase capable of operating upon both NDP-hexose and NDP-pentose scaffolds. The final tailoring event is postulated to be catalyzed by the methyltransferase homologs AtmS10/CalS10. Interestingly, the original labeling studies of **4**

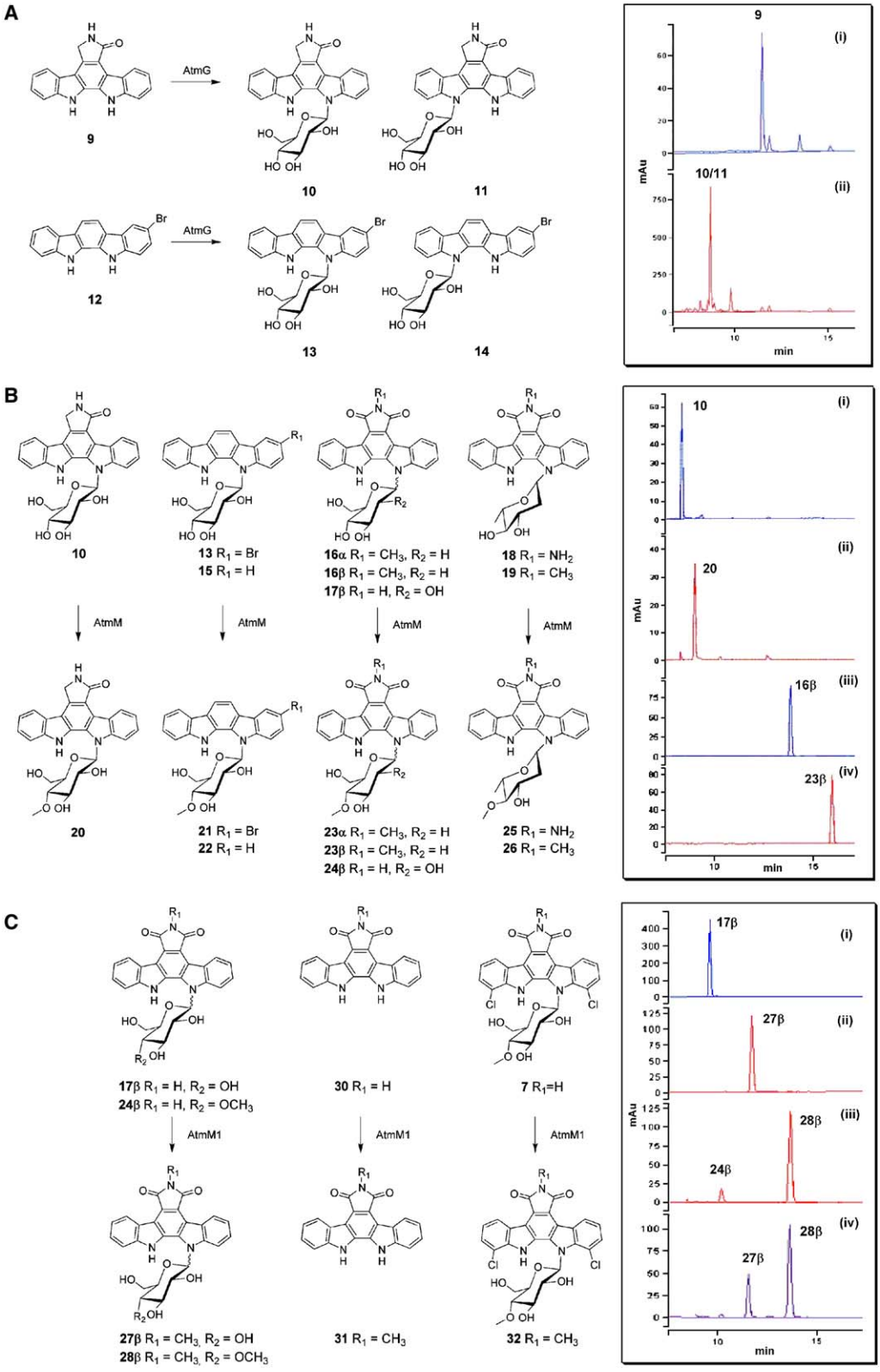

Figure 3. Reaction Schemes and HPLC Chromatographs of Representative Assays for *N*-glycosylation (AtmG), Glucose-4′-*O*-methylation (AtmM), and Maleimide *N*-methylation (AtmM1)

In some cases, anomeric stereochemistry is designated by α and β. Parameters for analytical HPLC and product characterization are described in Experimental Procedures.

(A) AtmG-mediated *N*-glycosylation of **9** and **12** via in vivo bioconversion: (i) control strain pET-28a(+)-*E.coli* BL21(DE3) with **9** (50 μM); (ii) AtmG overexpression strain, pUW-atG221-*E.coli* BL21(DE3), with **9** (50 μM).

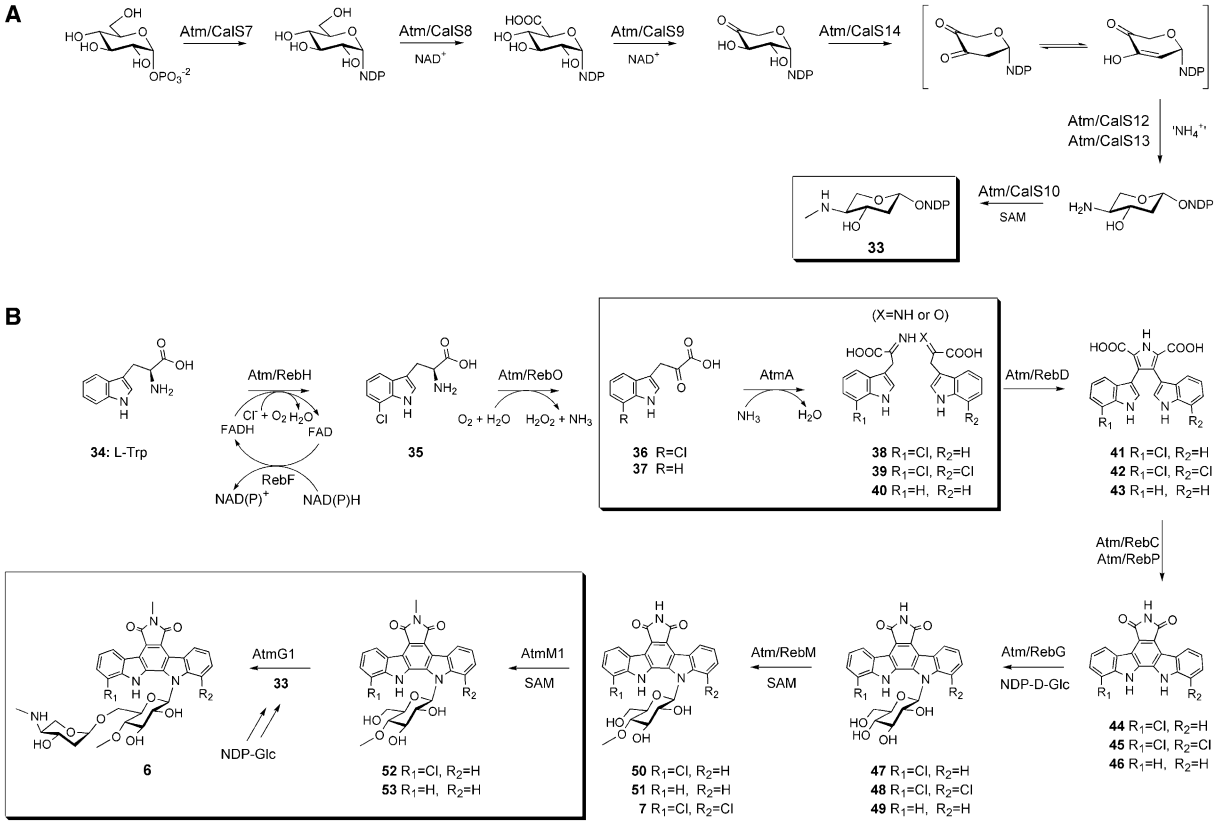

Figure 4. An Overview of Indolocarbazole Biosynthesis

(A) Proposed biosynthetic pathway for the aminodideoxypentose shared by **1**, **3**, and **6**. *N*-methylation (Atm/CalS10) may alternatively occur after the aminodideoxypentose is attached to the aglycon.

(B) Proposed overall biosynthetic pathway for AT2433.

In comparison to other indolocarbazole biosynthetic pathways, the steps unique to AT2433 biosynthesis include the entire aminodideoxypentose pathway highlighted in (A) and the boxed steps in (B).

revealed all alkyl carbons within the **4** *N*-alkyl-aminodideoxypentose (esperamicin A_{1c}, A_{1b}, and A_1 corresponding to *N*-methyl-, *N*-ethyl-, and *N*-isopropyl-, respectively) to derive from methionine [3]. Given what is known about *S*-adenosylmethionine-dependent methyltransferase catalysis, future characterization of AtmS10/CalS10 may expose a novel enzymatic mechanism for this potential sequential alkylation event. Finally, the gene *atmG1* encodes for a 415 amino acid *O*-glycosyltransferase homolog [61, 62] postulated to catalyze the final attachment of the aminopentose, the timing of which remains to be deciphered.

Additional Unique Features—Halogenation and the Source of Indolocarbazole Amine

The first step in indolocarbazole biosynthesis is flavin-dependent halogenation catalyzed by RebH in **7** biosynthesis [23, 24, 27, 29]. The AT2433 locus encodes for a RebH homolog (AtmH), but unlike the **7** loci lacks a gene encoding for the requisite flavin reductase (*rebF*). The next step (the flavin-dependent L-amino acid oxidase RebO/AtmO) catalyzes the oxidative deamination of 7-chloro-L-tryptophan to 7-chloroindole-3-pyruvic acid (Figure 4B, **36**, **37**) [30]. The core fusion proceeds via RebD/AtmD-catalyzed oxidative dimerization of Trp-derived monomers (e.g., **38–40**) to form chromopyrrolic acid (Figure 4B, **41–43**) [31], and recent studies revealed RebO/RebD were able to tolerate nonhalogenated, monohalogenated, or dihalogenated substrates [27, 30, 31]. Although many of the early steps within indolocarbazole core formation are fairly well understood, the source of the maleimide nitrogen within **38–40** remains controversial.

Early staurosporine labeling experiments revealed the core nitrogen to be nontryptophan in origin [63]. Recent in vitro studies revealed the maleimide nitrogen within **41–43** derived from indole 3-pyruvate imine (**38–40**) [31]. However, the in vivo nitrogen source en route to imines **38–40** is unknown. Interestingly, the AT2433 gene cluster contains a gene (*atmA*) directly upstream of *atmS10*, the product of which displays high homology to OxyD (56% identity, 65% similarity)—a glutamine-dependent amidotransferase involved in tetracycline biosynthesis in *Streptomyces rimosus* (Table 1). We

(B) AtmM-catalyzed in vitro *O*-methylation of **10**, **13**, **15–19**: (i) **10** (50 μM) and SAM (100 μM); (ii) **10** (50 μM), SAM (100 μM), and AtmM (15 μM); (iii) **16β** (50 μM) and SAM (100 μM); (iv) **16β** (50 μM), SAM (100 μM), and AtmM (15 μM).

(C) AtmM1-catalyzed in vitro *N*-methylation of **17β**, **24β**, **30**, and **7**: (i) **17β** (50 μM) and SAM (100 μM); (ii) **17β** (50 μM), SAM (100 μM), and AtmM1 (15 μM); (iii) **24β** (50 μM), SAM (100 μM), and AtmM1; (iv) **27β** (50 μM), SAM (100 μM), and AtmM (15 μM). α and β represent stereoisomers.

postulate AtmA may catalyze nitrogen installation (to provide imine **38** or **40**). Furthermore, given the indiscriminate nature of RebD (and presumably AtmD), we propose that the unique substrate specificities of AtmA and AtmO may dictate the final formation of asymmetrically halogenated aglycon in AT2433. This unique feature may also present additional opportunities for the combinatorial biosynthesis of differentially halogenated indolocarbazole aglycons.

Resistance and Regulation

Based on database comparison, two gene products may be related to AT2433 biosynthesis regulation. The first, AtmR, is homologous to RebR and StaR (43% identity, 56% similarity)—annotated as putative transcriptional activators of **7** and **8** biosynthetic pathways, respectively [23, 26]. As putative transporters associated with AT2433 biosynthesis, *atmB* encodes a 435 amino acid protein homologous to a putative antiporter from *Amycolatopsis balhimycina* (64% identity, 80% similarity; Table 1), while *atmI* encodes a product which resembles a putative transmembrane efflux protein from *Streptomyces coelicolor* A3(2) (44% identity, 63% similarity; Table 1). Finally, *atmE* reveals low homology (33% identity, 54% similarity) with a putative regulatory protein from *Frankia sp. EAN1pec* (Table 1).

Significance

The natural product AT2433, an indolocarbazole antitumor antibiotic, is structurally distinguished by its unique aminodideoxypentose-containing disaccharide and asymmetrically halogenated *N*-methylated aglycon. This work reveals the first glimpse of the AT2433 gene locus with locus confirmation via in vivo bioconversion and in vitro biochemical assay using AtmG, AtmM, and AtmM1. Preliminary studies of substrate tolerance suggest the potential opportunity to utilize these three enzymes for future indolocarbazole enzymatic diversification. Comparative genomics based upon the rebeccamycin, calicheamicin, and newly cloned AT2433 gene loci also provides, for the first time, a biosynthetic model for the aminodideoxypentose common to these structurally diverse natural products. Moreover, this work sets the stage for further studies regarding the origin of indolocarbazole maleimide nitrogen and factors which dictate the biosynthesis of asymmetric indolocarbazoles. Cumulatively, this work has both biosynthetic significance and the potential for clinical impact.

Experimental Procedures

Bacterial Strains, Culture Conditions, Vectors, and Reagents

Actinomadura melliaura sp. nov. SCC 1655 was grown on ISP2 agar medium (Bacto Laboratories Pty Ltd, Liverpool, NSW, Australia) and in liquid medium TSB medium (Bacto) at 30ºC. *Escherichia coli* DH5α or NovaBlue (Novagen, EMD Biosciences, San Diego, CA) competent cells were used for standard subcloning, *E. coli* XL-1 Blue MRF′ (Stratagene, La Jolla, CA) for cosmid library construction, and *E. coli* BL21(DE3) (Novagen) for gene expression and in vivo bioconversion strains. All *E. coli* strains were grown and transformed as described previously [64]. SuperCos2, a SuperCos1 (Stratagene) derivative vector lacking the neomycin resistance gene, was used for cosmid library construction. Vectors pEZSeq and pSMART-LCkan (Lucigen, Middleton, WI) were utilized for shotgun library construction for cosmid sequencing, while pGEM-Teasy (Promega, Madison, WI) was used for PCR cloning and sequencing. Vectors pET-30Xa/LIC and pET28a (Novagen) were used for gene expression. Biochemicals, chemicals, media, restriction enzymes, and other molecular biology reagents were from standard commercial sources. All indolocarbazole derivatives employed for enzyme assays and bioconversion experiments have been described elsewhere [11, 28, 45, 65].

DNA Isolation, Manipulation, and Cloning

Plasmid and cosmid DNA were isolated from *E. coli* strains using Qiagen miniprep and large construct kits (Qiagen, Valencia, CA). The isolation of DNA fragments from excised agarose was accomplished with the Qiagen gel clean kit (Qiagen). Restriction endonuclease digestion, ligation, and transformation were performed according to standard procedures [64] or manufacturers' recommendations. Chromosomal DNA isolation from *A. melliaura* SCC 1655 was accomplished via previously described methods [66] and PCR amplifications were carried out on a GeneAmp PCR system 9700 (Perkin-Elmer/ABI, Foster City, CA) using either PfuTurbo DNA polymerase (Stratagene) or TaKaRa LA-Taq DNA polymerase (Takara Mirus Bio, Madison, WI). For Southern analysis, DIG labeling of DNA probes, hybridization, and detection were performed according to the protocols provided by the manufacturer (Roche Applied Science, Indianapolis, IN). For PCR amplification from genomic DNA and cloning of *atmD* gene, the following pair of degenerate primers was used: 5′-TCGCCCRCGAGGAGATGATCC-3′ (forward) and 5′-GTCGATCSCGAACAGCGCRCTG-3′ (reverse); for the *atmP* gene: 5′-TGGCTGGTYTTCMTSGACCCGCC-3′ (forward) and 5′-TTCGTGGTSGTCTCGTGSCCGGC-3′ (reverse). PCR conditions were as follows: 5 min at 95ºC, 30 cycles (30 s at 95ºC, 30 s at 57ºC, 1 min at 72ºC), and 10 min at 72ºC with 1 U LA-Taq DNA polymerase. All primers were synthesized by Integrated DNA Technologies (Coralville, IA).

Genomic Library Construction and Screening

A. melliaura SCC 1655 genomic DNA was partially digested with Sau3AI to give fragments averaging 30–40 kb in size. These fragments were dephosphorylated with calf intestine alkaline phosphatase (CIAP) and ligated into the HpaI- and BamHI-digested cosmid vector SuperCos2. The ligation products were packaged with Gigapack III XL packaging extract (Stratagene) as described by the manufacturer and the resulting recombinant phage was used to infect *E. coli* XL1-Blue MRF′ cells. The resulting genomic library was screened by colony hybridization using the PCR-amplified genes *atmD* and *atmP* as probes, respectively, and the resultant positive clones were further confirmed by PCR amplification directly from positive cosmids.

DNA Sequencing and Analysis

Shotgun libraries of each positive cosmid were generated by Lucigen and shotgun sequencing was accomplished by the UW-Madison Genome Center (University of Wisconsin, Madison). Automated sequencing was done on double-stranded DNA templates from at least 700 shotgun subclones for each cosmid. All sequencing data were subsequently assembled and edited using SeqMan software (DNAStar). ORF searches were done by using the frameplot software available at http://www.nih.go.jp/~jun/cgi-bin/frameplot-3.0b.pl. Database comparison was performed with the BLAST search tools on the server of the National Center for Biotechnology Information (Bethesda, MD). The DNA sequence has been deposited in GenBank under the accession number DQ297453.

Overexpression and Characterization of the *N*-Glucosyltransferase AtmG

To generate an AtmG expression construct, the *atmG* gene was amplified from chromosomal DNA using primers atG-EF21 (5′-GTCAC CATATGGCACGGGTGCTCATG-3′, NdeI site underlined) and atG-ER21 (5′-TCAGGATCCTCAACTGATCGCTGTCCTG-3′, BamHI site underlined). PCR conditions were as follows: 3 min at 94ºC, 30 cycles (30 s at 94ºC, 30 s at 55ºC, 1 min at 72ºC), and 10 min at 72ºC with 1 U Pfu DNA polymerase, 5% DMSO. The gel-purified amplicon was cloned into the NdeI and BamHI sites of pET28a(+) (Novagen, EMD Biosciences) and the construct was confirmed by sequencing to yield plasmid pUWG-atGT221 (designed to express an N-terminal

His tag fusion protein). The bioconversion assays were accomplished in triplicate using pUW-atG221-*E. coli* BL21(DE3) with pET28a(+)-*E. coli* BL21(DE3) as the control. For each assay, a single fresh colony of *E. coli* BL21(DE3) was inoculated into 3 ml Luria Bertani (LB) medium containing 50 μg ml^{-1} kanamycin. After incubation with shaking (250 rpm) at 37°C for 8–10 hr, an aliquot of preculture was transferred to 5 ml LB (1:100 dilution), and the culture was grown at 28°C with shaking until A_{600} = 0.6. Isopropyl-β-D-thiogalactopyranoside was subsequently added (final concentration of 0.4 mM) and the culture was grown for 1.5 hr. Compound **9** (or **12**) (50 μM final concentration) was added and the culture was grown for an additional 4 hr and subsequently extracted with an equal volume of ethyl acetate. After centrifugation, the organic phase was evaporated, resuspended in 100 μl methanol, and analyzed by HPLC and LC-MS as described under Product Isolation and Characterization.

Overexpression and Characterization of the Glc-4′-*O*-Methyltransferase AtmM

To generate an AtmM expression construct, the *atmM* gene was amplified from chromosomal DNA using primers at24M-EF31 (5′-GGTATTGAGGGTCGCATGACGGATATCAGCCAG-3′, ligation site underlined) and at24M-ER31 (5′-AGAGGAGAGTTAGAGCCCT AGCTGCGGCGGGCGGC-3′, ligation site underlined). PCR conditions were as follows: 3 min at 94°C, 30 cycles (30 s at 94°C, 30 s at 55°C, 1 min at 72°C), and 10 min at 72°C with 1 U Pfu DNA polymerase, 5% DMSO. The gel-purified amplicon was treated and annealed with the ligation-independent cloning (LIC) sites of pET30Xa/LIC vector (Novagen) and confirmed by sequencing to yield plasmid pUWG-at24M31 (designed to express an N-terminal His tag fusion protein).

The plasmid pUWG-at24M31 containing the gene of N-His-tagged *atmM* was transformed in *E. coli* BL21(DE3) cells and a fresh colony was used to inoculate a 3 ml culture in LB medium supplemented with 50 μg ml^{-1} kanamycin which was grown at 37°C with 250 rpm shaking for 8–10 hr. A 0.5 ml aliquot of this preculture was transferred to 50 ml LB supplemented with 50 μg ml^{-1} kanamycin in 250 ml flasks and the culture was grown at 37°C with 250 rpm shaking for 8 hr. Approximately 10 ml of this secondary culture was transferred to 1 l LB medium grown under identical conditions until A_{600} = 0.8. AtmM overexpression was induced with addition of isopropyl-β-D-thiogalactopyranoside (IPTG; final 0.4 mM concentration) and the culture was grown for an additional 17 hr at 18°C before being harvested. Cells were harvested by centrifugation, and the cell pellet was washed twice (20 mM phosphate buffer [pH 7.5]) and then stored at −80°C. All subsequent purification steps were accomplished at 4°C.

The thawed cell pellet from 1 l of culture was resuspended in 30 ml of binding buffer (20 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole [pH 7.5]) containing 1 mg ml^{-1} lysozyme and the mixture was incubated on ice for 30 min. The cells were lysed completely by French press (three rounds at 700 psi; Thermo IEC, Waltham, MA) and the cell debris was subsequently removed by centrifugation (16,000 × g, 30 min). The supernatant fraction was filtered (0.45 μm syringe filter; Nalgene, Rochester, NY) and subjected to FPLC purification using a HisTrap HP column (1 ml; GE Healthcare Life Sciences [formerly Amersham Biosciences], Piscataway, NJ). The protein was loaded on the column with binding buffer and eluted with the same buffer using a linear imidazole gradient (10–500 mM, flow rate 1 ml min^{-1}, UV detection 280 nm). The desired protein fractions eluted with 100 mM imidazole were pooled, transferred into assay buffer (25 mM Tris-HCl, 20 mM NaCl, 0.5 mM DTT, 10% glycerol [pH 8.0]) via a PD-10 column (GE Healthcare Life Sciences), and concentrated (VIVASPIN 15R 10,000 MWCO; Sartorius AG [formerly Vivascience], Goettingen, Germany). The protein concentration was measured by the Bradford assay with bovine serum albumin as a standard (Bio-Rad, Hercules, CA). The protein was stored at −80°C.

Typical assays with compounds **10**, **13**, and **15**–**19** as substrates were conducted in a total volume of 100 μl of reaction mixture (50 mM Tris-HCl [pH 8.0]), including substrate (50 μM), S-adenosylmethionine (SAM; 100 μM), and purified protein (15 μM). Reactions were initiated by the addition of AtmM and incubated at 30°C for 3–8 hr. The assay mixture was subsequently extracted with an equal volume of ethyl acetate, the organics evaporated, and the recovered material redissolved in 100 μl of methanol and analyzed by HPLC and LC-MS as described in Product Isolation and Characterization.

Overexpression and Characterization of *N*-Methyltransferase AtmM1

To generate an AtmM1 expression construct, the *atmM1* gene was amplified from chromosomal DNA using primers atM1U2-EF31 (5′-GGTATTGAGGGTCGCATGCGCCCACTTCTTTAT-3′, ligation site underlined) and atM1U2-ER31 (5′-AGAGGAGAGTTAGAGCCTCACG GGTGCCGATCCCG-3′, ligation site underlined). PCR conditions were as follows: 3 min at 94°C, 30 cycles (30 s at 94°C, 30 s at 55°C, 1 min at 72°C), and 10 min at 72°C with 1 U Pfu DNA polymerase, 5% DMSO. The gel-purified amplicon was treated and annealed with the LIC sites of pET30Xa/LIC vector (Novagen) to yield plasmid pUWG-atM2U131 (designed to express an N-terminal His tag fusion protein) which was confirmed by sequencing.

The plasmid pUWG-atM2U131 was transformed in *E. coli* BL21(DE3) cells and a fresh colony was used to inoculate 3 ml of LB medium supplemented with 50 μg ml^{-1} kanamycin at 37°C with 250 rpm shaking for 8–10 hr. A 0.5 ml aliquot of this seed culture was transferred to 50 ml LB supplemented with 50 μg ml^{-1} kanamycin in 250 ml flasks and the secondary culture was grown at 37°C with 250 rpm shaking for 8 hr. Approximately 10 ml of secondary culture was transferred to 1 l LB and the culture was grown under the same conditions to A_{600} = 0.8. AtmM1 overexpression was induced with addition of IPTG (final concentration 0.4 mM) and the culture was grown for an additional 17 hr at 18°C. Cells were harvested by centrifugation, and the cell pellet was washed twice (20 mM phosphate buffer [pH 7.5]) and then stored at −80°C.

The thawed cell pellet from 1 l of culture was resuspended in 30 ml of binding buffer (20 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole [pH 7.5]) containing 1 mg ml^{-1} lysozyme and the mixture was incubated on ice for 30 min. The cells were lysed completely by French press (three rounds at 700 psi; Thermo IEC) and the cell debris was subsequently removed by centrifugation (16,000 × g, 30 min). Upon purification AtmM1 lost activity, and thus assays were routinely conducted at this stage (AtmM1 was estimated to comprise approximately 30% of the soluble protein in the crude extract by SDS-PAGE). For AtmM1 purification, the supernatant fraction was filtered (0.45 μm syringe filter; Nalgene) and subjected to FPLC purification using a HisTrap HP column (1 ml; GE Healthcare Life Sciences). The protein was loaded on the column with binding buffer and eluted with the same buffer using a linear imidazole gradient (10–500 mM, flow rate 1 ml min^{-1}, UV detection 280 nm). The desired protein fractions eluted with 200–300 mM imidazole were pooled, transferred into assay buffer (25 mM Tris-HCl, 20 mM NaCl, 0.5 mM DTT, 10% glycerol [pH 8.0]) via a PD-10 column (GE Healthcare Life Sciences), and concentrated (VIVASPIN 15R 10,000 MWCO; Sartorius AG). The protein concentration was measured by using the Bradford assay with bovine serum albumin as a standard (Bio-Rad). The protein was stored at −80°C.

The in vivo bioconversion experiments with AtmM1 were identical to those described for AtmG except that the overexpression strain was driven by plasmid pUWG-atM2U131 and different aglyons (**10**, **17β**) were employed as substrates. Typical assays with compounds **7**, **17β**, **24β**, and **30** as substrates were conducted in a total volume of 100 μl reaction mixture (50 mM Tris-HCl [pH 8.0]), including substrate (50 μM), SAM (100 μM), and AtmM1 (~15 μM). Reactions were initiated by the addition of AtmM1 and incubated at 30°C for 3–8 hr. The assay mixture was subsequently extracted with an equal volume of ethyl acetate, the organics evaporated, and the recovered material redissolved in 100 μl methanol and analyzed by HPLC and LC-MS as described in Product Isolation and Characterization.

Product Isolation and Characterization

Analytical HPLC utilized a Beckman Coulter Ultrasphere C_{18} analytical column (Beckman Coulter, Fullerton, CA; 5 μm; 4.6 × 250 mm; mobile phase A: H_2O; mobile phase B: CH_3CN; 0–5 min, 90:10 to 50:50 A:B; 5–20 min, 50:50 A:B to 100% B; 20–25 min, 100% B; flow rate 1.0 ml min^{-1}; A_{280} and/or A_{315}) using a Varian Prostar system with a photodiode array detector (Varian Analytical Instruments, Walnut Creek, CA). ESI-MS analysis of products was performed on an Agilent 1000 HPLC-MSD SL instrument (Agilent Technologies, Palo Alto, CA).

Acknowledgments

This work was supported in part by National Institutes of Health grants CA84374, AI52218, and GM70637, and a National Cooperative Drug Discovery Group grant from the National Cancer Institute (U19 CA113297). J.S.T. is a Romnes fellow. The authors want to thank Professors Peng George Wang (The Ohio State University) and David L. Van Vranken (UC Irvine) for graciously providing materials and Drs. Byron R. Griffith and Aqeel Ahmed for technical assistance. The authors also wish to acknowledge the Analytical Instrumentation Center of the School of Pharmacy, UW-Madison, for MS and NMR support.

Received: March 10, 2006
Revised: April 28, 2006
Accepted: May 4, 2006
Published: July 28, 2006

References

1. He, X., and Liu, H.-w. (2002). Mechanisms of enzymatic C-O bond cleavage in deoxyhexose biosynthesis. Curr. Opin. Chem. Biol. *6*, 590–597.
2. Rupprath, C., Schumacher, T., and Elling, L. (2005). Nucleotide deoxysugars: essential tools for the glycosylation engineering of novel bioactive compounds. Curr. Med. Chem. *12*, 1637–1675.
3. Lam, K.S., Gustavson, D.R., Veitch, J.A., and Forenza, S. (1993). The effect of cerulenin on the production of esperamicin A_1 by *Actinomadura verrucosospora*. J. Ind. Microbiol. *12*, 99–102.
4. Pattathil, S., Harper, A.D., and Bar-Peled, M. (2005). Biosynthesis of UDP-xylose: characterization of membrane-bound AtUxs2. Planta *221*, 538–548.
5. Harper, A.D., and Bar-Peled, M. (2002). Biosynthesis of UDP-xylose. Cloning and characterization of a novel *Arabidopsis* gene family, UXS, encoding soluble and putative membrane-bound UDP-glucuronic acid decarboxylase isoforms. Plant Physiol. *130*, 2188–2198.
6. Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B.O., Huang, K., Fonstein, L., Czisny, A., Whitwam, R.E., et al. (2002). The calicheamicin gene cluster and its iterative type I enediyne PKS. Science *297*, 1173–1176.
7. Bililign, T., Shepard, E.M., Ahlert, J., and Thorson, J.S. (2002). On the origin of deoxypentoses: evidence to support a glucose progenitor in the biosynthesis of calicheamicin. ChemBioChem *3*, 1143–1146.
8. Weitnauer, G., Muhlenweg, A., Trefzer, A., Hoffmeister, D., Sussmuth, R.D., Jung, G., Welzel, K., Vente, A., Girreser, U., and Bechthold, A. (2001). Biosynthesis of the orthosomycin antibiotic avilamycin A: deductions from the molecular analysis of the *avi* biosynthetic gene cluster of *Streptomyces viridochromogenes* Tü57 and production of new antibiotics. Chem. Biol. *8*, 569–581.
9. Hofmann, C., Boll, R., Heitmann, B., Hauser, G., Durr, C., Frerich, A., Weitnauer, G., Glaser, S.J., and Bechthold, A. (2005). Genes encoding enzymes responsible for the biosynthesis of L-lyxose and attachment of eurekanate during avilamycin biosynthesis. Chem. Biol. *12*, 1137–1143.
10. Matson, J.A., Claridge, C., Bush, J.A., Titus, J., Bradner, W.T., Doyle, T.W., Horan, A.C., and Patel, M. (1989). AT2433-A_1, AT2433-A_2, AT2433-B_1, and AT2433-B_2 novel antitumor antibiotic compounds produced by *Actinomadura melliaura*. Taxonomy, fermentation, isolation and biological properties. J. Antibiot. (Tokyo) *42*, 1547–1555.
11. Chisholm, J.D., and Van Vranken, D.L. (2000). Regiocontrolled synthesis of the antitumor antibiotic AT2433-A_1. J. Org. Chem. *65*, 7541–7553.
12. Yoshinari, T., Yamada, A., Uemura, D., Nomura, K., Arakawa, H., Kojiri, K., Yoshida, E., Suda, H., and Okura, A. (1993). Induction of topoisomerase I-mediated DNA cleavage by a new indolocarbazole, ED-110. Cancer Res. *53*, 490–494.
13. Long, B.H., Rose, W.C., Vyas, D.M., Matson, J.A., and Forenza, S. (2002). Discovery of antitumor indolocarbazoles: rebeccamycin, NSC 655649, and fluoroindolocarbazoles. Curr. Med. Chem. Anticancer Agents *2*, 255–266.
14. Facompre, M., Carrasco, C., Colson, P., Houssier, C., Chisholm, J.D., Van Vranken, D.L., and Bailly, C. (2002). DNA binding and topoisomerase I poisoning activities of novel disaccharide indolocarbazoles. Mol. Pharmacol. *62*, 1215–1227.
15. Carrasco, C., Facompre, M., Chisholm, J.D., Van Vranken, D.L., Wilson, W.D., and Bailly, C. (2002). DNA sequence recognition by the indolocarbazole antitumor antibiotic AT2433-B_1 and its diastereoisomer. Nucleic Acids Res. *30*, 1774–1781.
16. Osada, H., Satake, M., Koshino, H., Onose, R., and Isono, K. (1992). A new indolocarbazole antibiotic, RK-286D. J. Antibiot. (Tokyo) *45*, 278–279.
17. Takahashi, H., Osada, H., Uramoto, M., and Isono, K. (1990). A new inhibitor of protein kinase C, RK-286C (4′-demethylamino-4′-hydroxystaurosporine). II. Isolation, physico-chemical properties and structure. J. Antibiot. (Tokyo) *43*, 168–173.
18. Pindur, U., Kim, Y.S., and Mehrabani, F. (1999). Advances in indolo[2,3-a]carbazole chemistry: design and synthesis of protein kinase C and topoisomerase I inhibitors. Curr. Med. Chem. *6*, 29–69.
19. Bailly, C. (2000). Topoisomerase I poisons and suppressors as anticancer drugs. Curr. Med. Chem. *7*, 39–58.
20. Pilch, B., Allemand, E., Facompre, M., Bailly, C., Riou, J.F., Soret, J., and Tazi, J. (2001). Specific inhibition of serine- and arginine-rich splicing factors phosphorylation, spliceosome assembly, and splicing by the antitumor drug NB-506. Cancer Res. *61*, 6876–6884.
21. Akinaga, S., Gomi, K., Morimoto, M., Tamaoki, T., and Okabe, M. (1991). Antitumor activity of UCN-01, a selective inhibitor of protein kinase C, in murine and human tumor models. Cancer Res. *51*, 4888–4892.
22. Courage, C., Snowden, R., and Gescher, A. (1996). Differential effects of staurosporine analogues on cell cycle, growth and viability in A549 cells. Br. J. Cancer *74*, 1199–1205.
23. Sanchez, C., Butovich, I.A., Brana, A.F., Rohr, J., Mendez, C., and Salas, J.A. (2002). The biosynthetic gene cluster for the antitumor rebeccamycin: characterization and generation of indolocarbazole derivatives. Chem. Biol. *9*, 519–531.
24. Onaka, H., Taniguchi, S., Igarashi, Y., and Furumai, T. (2003). Characterization of the biosynthetic gene cluster of rebeccamycin from *Lechevalieria aerocolonigenes* ATCC 39243. Biosci. Biotechnol. Biochem. *67*, 127–138.
25. Hyun, C.G., Bililign, T., Liao, J., and Thorson, J.S. (2003). The biosynthesis of indolocarbazoles in a heterologous *E. coli* host. ChemBioChem *4*, 114–117.
26. Onaka, H., Taniguchi, S., Igarashi, Y., and Furumai, T. (2002). Cloning of the staurosporine biosynthetic gene cluster from *Streptomyces sp*. TP-A0274 and its heterologous expression in *Streptomyces lividans*. J. Antibiot. (Tokyo) *55*, 1063–1071.
27. Sanchez, C., Zhu, L., Brana, A.F., Salas, A.P., Rohr, J., Mendez, C., and Salas, J.A. (2005). Combinatorial biosynthesis of antitumor indolocarbazole compounds. Proc. Natl. Acad. Sci. USA *102*, 461–466.
28. Zhang, C., Albermann, C., Fu, X., Peters, N.R., Chisholm, J.D., Zhang, G., Gilbert, E.J., Wang, P.G., Van Vranken, D.L., and Thorson, J.S. (2006). RebG- and RebM-catalyzed indolocarbazole diversification. Chembiochem *7*, 795–804.
29. Yeh, E., Garneau, S., and Walsh, C.T. (2005). Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. Proc. Natl. Acad. Sci. USA *102*, 3960–3965.
30. Nishizawa, T., Aldrich, C.C., and Sherman, D.H. (2005). Molecular analysis of the rebeccamycin L-amino acid oxidase from *Lechevalieria aerocolonigenes* ATCC 39243. J. Bacteriol. *187*, 2084–2092.
31. Howard-Jones, A.R., and Walsh, C.T. (2005). Enzymatic generation of the chromopyrrolic acid scaffold of rebeccamycin by the tandem action of RebO and RebD. Biochemistry *44*, 15652–15663.
32. Zhang, C., Weller, R.L., Thorson, J.S., and Rajski, S.R. (2006). Natural product diversification using a non-natural cofactor analog of S-adenosyl-L-methionine. J. Am. Chem. Soc. *128*, 2760–2761.
33. Salas, A.P., Zhu, L., Sanchez, C., Brana, A.F., Rohr, J., Mendez, C., and Salas, J.A. (2005). Deciphering the late steps in the biosynthesis of the anti-tumour indolocarbazole staurosporine:

sugar donor substrate flexibility of the StaG glycosyltransferase. Mol. Microbiol. *58*, 17–27.
34. Bibb, M.J., Findlay, P.R., and Johnson, M.W. (1984). The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene *30*, 157–166.
35. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
36. Ohuchi, T., Ikeda-Araki, A., Watanabe-Sakamoto, A., Kojiri, K., Nagashima, M., Okanishi, M., and Suda, H. (2000). Cloning and expression of a gene encoding N-glycosyltransferase (*ngt*) from *Saccarothrix aerocolonigenes* ATCC39243. J. Antibiot. (Tokyo) *53*, 393–403.
37. Moreau, P., Anizon, F., Sancelme, M., Prudhomme, M., Bailly, C., Carrasco, C., Ollier, M., Severe, D., Riou, J.F., Fabbro, D., et al. (1998). Syntheses and biological evaluation of indolocarbazoles, analogues of rebeccamycin, modified at the imide heterocycle. J. Med. Chem. *41*, 1631–1640.
38. Anizon, F., Belin, L., Moreau, P., Sancelme, M., Voldoire, A., Prudhomme, M., Ollier, M., Severe, D., Riou, J.F., Bailly, C., et al. (1997). Syntheses and biological activities (topoisomerase inhibition and antitumor and antimicrobial properties) of rebeccamycin analogues bearing modified sugar moieties and substituted on the imide nitrogen with a methyl group. J. Med. Chem. *40*, 3456–3465.
39. Bailly, C., Qu, X., Graves, D.E., Prudhomme, M., and Chaires, J.B. (1999). Calories from carbohydrates: energetic contribution of the carbohydrate moiety of rebeccamycin to DNA binding and the effect of its orientation on topoisomerase I inhibition. Chem. Biol. *6*, 277–286.
40. Messaoudi, S., Anizon, F., Leonce, S., Pierre, A., Pfeiffer, B., and Prudhomme, M. (2005). Synthesis and cytotoxicities of 7-aza rebeccamycin analogues bearing various substituents on the sugar moiety, on the imide nitrogen and on the carbazole framework. Eur. J. Med. Chem. *40*, 961–971.
41. Weidner, S., Kittelmann, M., Goeke, K., Ghisalba, O., and Zahner, H. (1998). 3′-Demethoxy-3′-hydroxystaurosporine-O-methyltransferase from *Streptomyces longisporoflavus* catalyzing the last step in the biosynthesis of staurosporine. J. Antibiot. (Tokyo) *51*, 679–682.
42. Blondeau, J.M., DeCarolis, E., Metzler, K.L., and Hansen, G.T. (2002). The macrolides. Expert Opin. Investig. Drugs *11*, 189–215.
43. Bauer, N.J., Kreuzman, A.J., Dotzlaf, J.E., and Yeh, W.K. (1988). Purification, characterization, and kinetic mechanism of *S*-adenosyl-L-methionine:macrocin *O*-methyltransferase from *Streptomyces fradiae*. J. Biol. Chem. *263*, 15619–15625.
44. Kreuzman, A.J., Turner, J.R., and Yeh, W.K. (1988). Two distinctive *O*-methyltransferases catalyzing penultimate and terminal reactions of macrolide antibiotic (tylosin) biosynthesis. Substrate specificity, enzyme inhibition, and kinetic mechanism. J. Biol. Chem. *263*, 15626–15633.
45. Zhang, G., Shen, J., Cheng, H., Zhu, L., Fang, L., Luo, S., Muller, M.T., Lee, G.E., Wei, L., Du, Y., et al. (2005). Syntheses and biological activities of rebeccamycin analogues with uncommon sugars. J. Med. Chem. *48*, 2600–2611.
46. Bindschedler, L.V., Wheatley, E., Gay, E., Cole, J., Cottage, A., and Bolwell, G.P. (2005). Characterisation and expression of the pathway from UDP-glucose to UDP-xylose in differentiating tobacco tissue. Plant Mol. Biol. *57*, 285–301.
47. Suzuki, K., Watanabe, K., Masumura, T., and Kitamura, S. (2004). Characterization of soluble and putative membrane-bound UDP-glucuronic acid decarboxylase (OsUXS) isoforms in rice. Arch. Biochem. Biophys. *431*, 169–177.
48. Moriarity, J.L., Hurt, K.J., Resnick, A.C., Storm, P.B., Laroy, W., Schnaar, R.L., and Snyder, S.H. (2002). UDP-glucuronate decarboxylase, a key enzyme in proteoglycan synthesis: cloning, characterization, and localization. J. Biol. Chem. *277*, 16968–16975.
49. Kuhn, J., Gotting, C., Schnolzer, M., Kempf, T., Brinkmann, T., and Kleesiek, K. (2001). First isolation of human UDP-D-xylose: proteoglycan core protein β-D-xylosyltransferase secreted from cultured JAR choriocarcinoma cells. J. Biol. Chem. *276*, 4940–4947.
50. Bar-Peled, M., Griffith, C.L., and Doering, T.L. (2001). Functional cloning and characterization of a UDP-glucuronic acid decarboxylase: the pathogenic fungus *Cryptococcus neoformans* elucidates UDP-xylose synthesis. Proc. Natl. Acad. Sci. USA *98*, 12003–12008.
51. Breazeale, S.D., Ribeiro, A.A., and Raetz, C.R. (2002). Oxidative decarboxylation of UDP-glucuronic acid in extracts of polymyxin-resistant *Escherichia coli*. Origin of lipid A species modified with 4-amino-4-deoxy-L-arabinose. J. Biol. Chem. *277*, 2886–2896.
52. Chen, H., Agnihotri, G., Guo, Z., Que, N.L.S., Chen, X.H., and Liu, H.-w. (1999). Biosynthesis of mycarose: isolation and characterization of enzymes involved in the C-2 deoxygenation. J. Am. Chem. Soc. *121*, 8124–8125.
53. Draeger, G., Park, S.H., and Floss, H.G. (1999). Mechanism of the 2-deoxygenation step in the biosynthesis of the deoxyhexose moieties of the antibiotics granaticin and oleandomycin. J. Am. Chem. Soc. *121*, 2611–2612.
54. Zhao, Z., Hong, L., and Liu, H.-w. (2005). Characterization of protein encoded by *spnR* from the spinosyn gene cluster of *Saccharopolyspora spinosa*: mechanistic implications for forosamine biosynthesis. J. Am. Chem. Soc. *127*, 7692–7693.
55. Obhi, R.K., and Creuzenet, C. (2005). Biochemical characterization of the *Campylobacter jejuni* Cj1294, a novel UDP-4-keto-6-deoxy-GlcNAc aminotransferase that generates UDP-4-amino-4,6-dideoxy-GalNAc. J. Biol. Chem. *280*, 20902–20908.
56. Nedal, A., and Zotchev, S.B. (2004). Biosynthesis of deoxyaminosugars in antibiotic-producing bacteria. Appl. Microbiol. Biotechnol. *64*, 7–15.
57. Albermann, C., and Piepersberg, W. (2001). Expression and identification of the RfbE protein from *Vibrio cholerae* O1 and its use for the enzymatic synthesis of GDP-D-perosamine. Glycobiology *11*, 655–661.
58. Chiu, H.T., Hubbard, B.K., Shah, A.N., Eide, J., Fredenburg, R.A., Walsh, C.T., and Khosla, C. (2001). Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. Proc. Natl. Acad. Sci. USA *98*, 8548–8553.
59. Breazeale, S.D., Ribeiro, A.A., and Raetz, C.R. (2003). Origin of lipid A species modified with 4-amino-4-deoxy-L-arabinose in polymyxin-resistant mutants of *Escherichia coli*. An aminotransferase (ArnB) that generates UDP-4-deoxyl-L-arabinose. J. Biol. Chem. *278*, 24731–24739.
60. Zhao, L., Ahlert, J., Xue, Y., Thorson, J.S., Sherman, D.H., and Liu, H.-w. (1999). Engineering a methymycin/pikromycin-calicheamicin hybrid: construction of two new macrolides carrying a designed sugar moiety. J. Am. Chem. Soc. *121*, 9881–9882.
61. Anzai, Y., Saito, N., Tanaka, M., Kinoshita, K., Koyama, Y., and Kato, F. (2003). Organization of the biosynthetic gene cluster for the polyketide macrolide mycinamicin in *Micromonospora griseorubida*. FEMS Microbiol. Lett. *218*, 135–141.
62. Olano, C., Rodriguez, A.M., Michel, J.M., Mendez, C., Raynal, M.C., and Salas, J.A. (1998). Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring. Mol. Gen. Genet. *259*, 299–308.
63. Yang, S.W., and Cordell, G.A. (1997). Origin of nitrogen in the indolocarbazole unit of staurosporine. J. Nat. Prod. *60*, 788–790.
64. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, *Volumes 1–3* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
65. Gilbert, E.J., Chisholm, J.D., and Van Vranken, D.L. (1999). Conformational control in the rebeccamycin class of indolocarbazole glycosides. J. Org. Chem. *64*, 5670–5676.
66. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000). Practical *Streptomyces* Genetics (Norwich, UK: John Innes Foundation).

Accession Numbers

The DNA sequence for the *Actinomadura melliaura* AT2433 biosynthetic gene cluster has been deposited in GenBank under accession number DQ297453.