

## Tailoring Enzyme-Rich Environmental DNA Clones: A Source of Enzymes for Generating Libraries of Unnatural Natural Products

Jacob J. Banik, Jeffrey W. Craig, Paula Y. Calle, and Sean F. Brady\*

Howard Hughes Medical Institute, Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, 1230 York Avenue, New York, New York 10065, United States

Received July 1, 2010; E-mail: sbrady@rockefeller.edu

**Abstract:** A detailed bioinformatics analysis of six glycopeptide biosynthetic gene clusters isolated from soil environmental DNA (eDNA) megalibraries indicates that a subset of these gene clusters contains collections of tailoring enzymes that are predicted to result in the production of new glycopeptide congeners. In particular, sulfotransferases appear in eDNA-derived gene clusters at a much higher frequency than would be predicted from the characterization of glycopeptides from cultured *Actinomycetes*. Enzymes found on tailoring-enzyme-rich eDNA clones associated with these six gene clusters were used to produce a series of new sulfated glycopeptide derivatives in both in vitro and in vivo derivatization studies. The derivatization of known natural products with eDNA-derived tailoring enzymes is likely to be a broadly applicable strategy for generating libraries of new natural product variants.

### Introduction

In many cases, bacterial natural products can be classified into large families based on the presence of conserved core structures with shared biosynthetic origins.<sup>1</sup> The structural diversity seen within a family of natural products is often attributable almost exclusively to differences in the constellation of functional groups that are appended around the conserved core. Differences in functionalization patterns arise from the action of tailoring enzymes found within the biosynthetic gene clusters that are associated with a given natural product family. Tailoring enzymes often exhibit broad substrate promiscuity, and therefore, enzymes that have evolved to be used in the biosynthesis of one member of a family will likely have the ability to carry out the same reaction on many related structural variants.<sup>2</sup> Tailoring enzymes associated with a particular family of natural products, either alone or in combination, are therefore potentially useful tools for creating libraries of natural products with new functionalization patterns. The utility of this approach for generating collections of novel bioactive metabolites is limited mainly by the size and diversity of the set of tailoring enzymes associated with a specific natural product family. The emerging field of metagenomics provides an unprecedented opportunity to access new collections of tailoring enzymes that could be used to generate additional derivatives of many natural products.

Environmental samples typically contain 2–3 orders of magnitude more bacterial species than is easily cultured in the laboratory.<sup>3–5</sup> Metagenomics relies on the cloning of DNA directly from naturally occurring microbial populations (environmental DNA, eDNA) to circumvent difficulties associated with culturing bacteria and then the subsequent screening of

the resulting clone libraries to identify new genes and gene products of interest.<sup>6–8</sup> All of the genes required for the biosynthesis of a natural product, including genes that encode biosynthetic, regulatory, and self-immunity enzymes, are typically clustered on bacterial chromosomes. In addition to containing completely novel biosynthetic systems, eDNA samples are also likely to contain numerous gene clusters that resemble known systems but that differ in their pathway-specific collections of tailoring enzymes. The construction of 35–40 kb insert cosmid-based eDNA libraries is now routine, and a number of examples exist in the literature where large natural product gene clusters have been recovered on a series of overlapping eDNA clones.<sup>9–12</sup> The construction of larger insert eDNA libraries capable of capturing intact natural product gene clusters that might differ in their pathway-specific collections of tailoring enzymes on individual eDNA clones remains a major

(3) Rappé, M. S.; Giovannoni, S. J. *Annu. Rev. Microbiol.* **2003**, *57*, 369–94.

(4) Torsvik, V.; Goksøyr, J.; Daee, F. L. *Appl. Environ. Microbiol.* **1990**, *56*, 782–7.

(5) Torsvik, V.; Ovreas, L.; Thingstad, T. F. *Science* **2002**, *296*, 1064–6.

(6) Seow, K. T.; Meurer, G.; Gerlitz, M.; Wendt-Pienkowski, E.; Hutchinson, C. R.; Davies, J. J. *Bacteriol.* **1997**, *179*, 7360–8.

(7) Handelsman, J.; Rondon, M. R.; Brady, S. F.; Clardy, J.; Goodman, R. M. *Chem. Biol.* **1998**, *5*, R245–9.

(8) Rondon, M. R.; August, P. R.; Bettermann, A. D.; Brady, S. F.; Grossman, T. H.; Liles, M. R.; Loiacono, K. A.; Lynch, B. A.; MacNeil, I. A.; Minor, C.; Tiong, C. L.; Gilman, M.; Osburne, M. S.; Clardy, J.; Handelsman, J.; Goodman, R. M. *Appl. Environ. Microbiol.* **2000**, *66*, 2541–7.

(9) Piel, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14002–7.

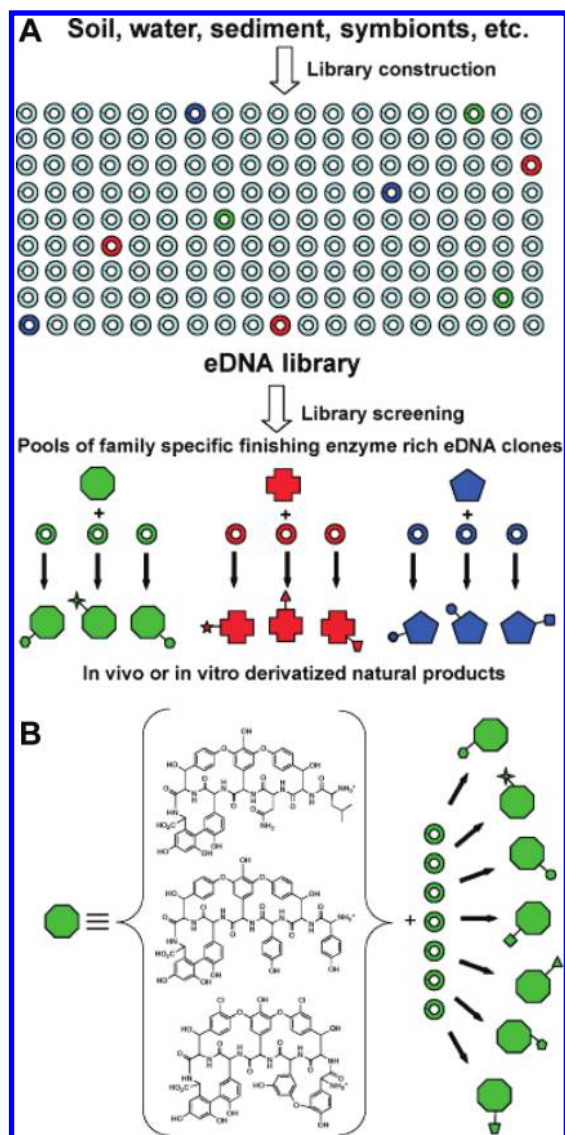
(10) Sudek, S.; Lopanik, N. B.; Waggoner, L. E.; Hildebrand, M.; Anderson, C.; Liu, H.; Patel, A.; Sherman, D. H.; Haygood, M. G. *J. Nat. Prod.* **2006**, *70*, 67–74.

(11) Banik, J. J.; Brady, S. F. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17273–7.

(12) Fisch, K. M.; Gurgui, C.; Heycke, N.; van der Sar, S. A.; Anderson, S. A.; Webb, V. L.; Taudien, S.; Platzer, M.; Rubio, B. K.; Robinson, S. J.; Crews, P.; Piel, J. *Nat. Chem. Biol.* **2009**, *5*, 494–501.

(1) Buckingham, J. *Dictionary of Natural Products*; 1st ed.; Chapman & Hall: London; New York, 1994.

(2) Firm, R. D.; Jones, C. G. *Nat. Prod. Rep.* **2003**, *20*, 382–91.



**Figure 1.** (A) Environmental DNA libraries are predicted to contain thousands of unique biosynthetic gene clusters. Many of these previously inaccessible clusters likely encode the biosynthesis of new members of previously characterized natural product families. Tailoring-enzyme-rich eDNA clones from these clusters will encode collections of enzymes that have evolved to specifically interact with a conserved natural product core structure. These enzymes should therefore be useful for generating new collections of unnatural natural products using either in vitro or in vivo biosynthetic strategies. (B) Glycopeptides are derived from a small number of oxidatively cross-linked heptapeptides. In this paper, we describe the use of eDNA-derived glycopeptide tailoring enzymes to generate a family of new glycopeptide congeners.

challenge for the field. While it is still difficult to regularly clone large intact gene clusters from environmental samples, the systematic examination of eDNA cosmid libraries with probes designed to recognize conserved sequences found in known natural product gene clusters has the potential to uncover individual clones with unique collections of tailoring enzymes that could be used to generate metabolites with new functionalization patterns (Figure 1A).

Glycopeptides are one example of a large family of natural products where the individual members differ by the number and type of functional groups that are arranged around a set of closely related core structures (Figure 1B).<sup>13</sup> Several members of this group of oxidatively cross-linked heptapeptides, including vancomycin and teicoplanin, are the preferred course of

treatment for nosocomial antibiotic-resistant Gram-positive bacterial infections.<sup>14</sup> Although cultured bacteria have been an important source of glycopeptides, there undoubtedly remain numerous previously undetected glycopeptide biosynthetic gene clusters in nature. Environmental samples likely contain both glycopeptide-producing bacteria that have remained recalcitrant to standard culture methods, as well as bacteria that possess cryptic glycopeptide gene clusters that have evaded detection in standard activity-based assays. Here we report the isolation of new glycopeptide biosynthetic gene clusters from eDNA libraries constructed from multiple soil samples and the use of the tailoring enzymes found in these gene clusters to produce a series of new sulfated glycopeptide derivatives. The tailoring of known natural product scaffolds with collections of eDNA-derived tailoring enzymes is likely to be a broadly applicable strategy for generating libraries of new natural product variants.

## Results and Discussion

### Isolation of Glycopeptide Tailoring Enzyme-Rich Clones.

Five of the seven sequenced glycopeptide gene clusters show a very similar overall gene organization (Figure 2A).<sup>11,15–19</sup> In these clusters, the position of the genes that encode each general class of enzyme required for the biosynthesis of a glycopeptide is conserved; resistance genes are grouped together at the beginning of the cluster, followed by nonribosomal peptide synthetase genes, unique collections of genes encoding tailoring enzymes, and ultimately genes that encode the biosynthesis of essential substrates (amino acids and sugars) (Figure 2A).<sup>20</sup> An interesting feature of this conserved architecture is that genes encoding tailoring enzymes tend to be grouped tightly together in the gene cluster. With such an arrangement, we postulated that it would be possible to capture large portions of glycopeptide tailoring-enzyme systems on single eDNA-derived cosmid clones and that the genes found on these clones would be useful for generating glycopeptides with novel functionalization patterns.

The long history of culturing glycopeptide-producing bacteria from soil, coupled with the enormous unexplored bacterial diversity present in most soil environments, makes soil a particularly attractive starting point for the potential discovery of additional glycopeptide biosynthetic diversity using culture-independent strategies. In this study, eDNA libraries were constructed from soil collected in California (CA) and Arizona (AZ), and these libraries were screened for genes encoding glycopeptide biosynthetic machinery. Previous studies have shown that much of the biosynthetic capacity present in a soil sample can be captured in a library containing at least  $1.0 \times 10^7$  cosmid clones, and therefore, each of these libraries was expanded until it contained between  $1.0 \times 10^7$  and  $1.5 \times 10^7$

(13) Nicolaou, K. C.; Boddy, C. N. C.; Bräse, S.; Winssinger, N. *Angew. Chem., Int. Ed.* **1999**, *38*, 2096–2152.

(14) Pace, J. L.; Yang, G. *Biochem. Pharmacol.* **2006**, *71*, 968–80.

(15) Pelzer, S.; Stüssmuth, R.; Heckmann, D.; Recktenwald, J.; Huber, P.; Jung, G.; Wohlleben, W. *Antimicrob. Agents Chemother.* **1999**, *43*, 1565–73.

(16) van Wageningen, A. M.; Kirkpatrick, P. N.; Williams, D. H.; Harris, B. R.; Kershaw, J. K.; Lennard, N. J.; Jones, M.; Jones, S. J.; Solenberg, P. J. *J. Chem. Biol.* **1998**, *5*, 155–62.

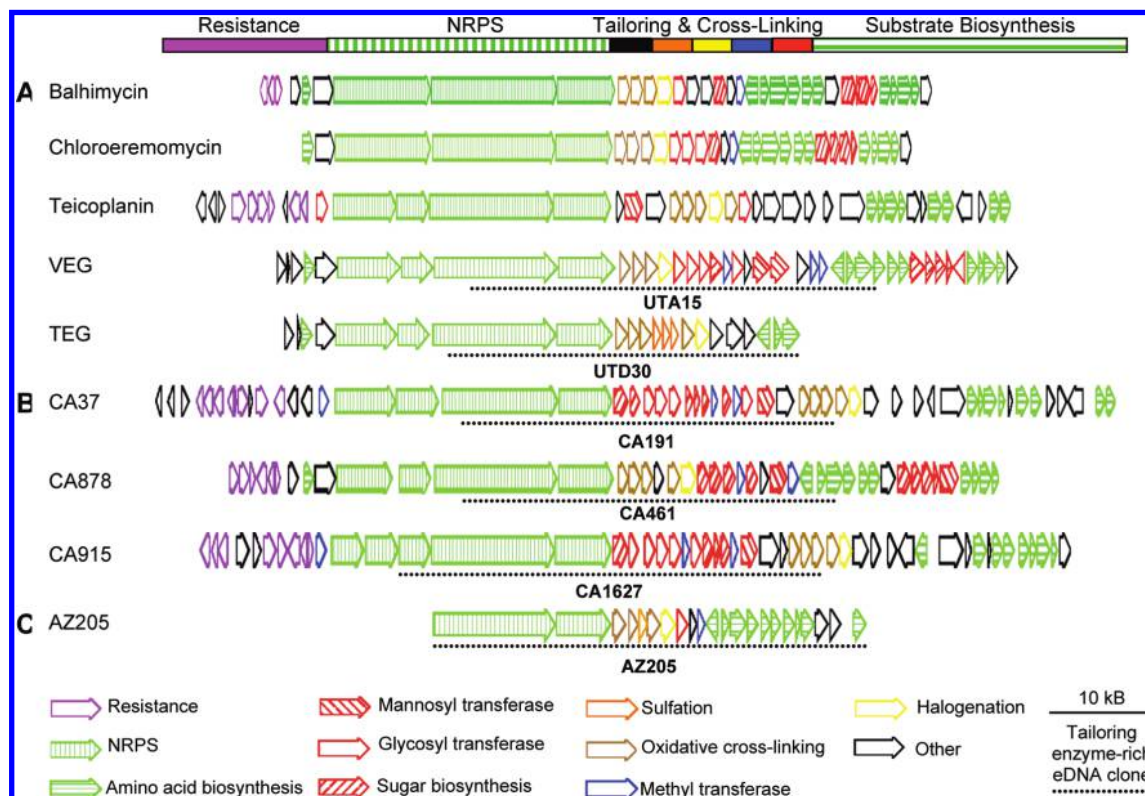
(17) Sosio, M.; Kloosterman, H.; Bianchi, A.; de Vreugd, P.; Dijkhuizen, L.; Donadio, S. *Microbiology* **2004**, *150*, 95–102.

(18) Pootoolal, J.; Thomas, M. G.; Marshall, C. G.; Neu, J. M.; Hubbard, B. K.; Walsh, C. T.; Wright, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8962–7.

(19) Sosio, M.; Stinchi, S.; Beltrametti, F.; Lazzarini, A.; Donadio, S. *Chem. Biol.* **2003**, *10*, 541–9.

(20) Donadio, S.; Sosio, M.; Stegmann, E.; Weber, T.; Wohlleben, W. *Mol. Genet. Genomics* **2005**, *274*, 40–50.





**Figure 2.** (A) Many sequenced glycopeptide gene clusters show a similar overall gene organization. The balhimycin, chloroeremomycin, and teicoplanin gene clusters were cloned from cultured soil bacteria while the VEG and TEG gene clusters were cloned directly from soil. Open reading frames have been color coded according to their predicted functions. (B) Glycopeptide gene clusters recovered from the CA eDNA megalibrary are shown. (C) AZ205, a tailoring-enzyme-rich eDNA cosmid clone, recovered from the AZ library is shown. Dashed lines are used to indicate the region from each gene cluster that corresponds to the tailoring-enzyme-rich clone used in this study.

unique cosmid clones.<sup>21</sup> We have termed these large libraries eDNA megalibraries.

In a previous study, we were able to recover two glycopeptide gene clusters (TEG and VEG) from a Utah (UT) soil eDNA megalibrary by initially screening the library for cosmids containing *oxyC* genes (Figure 2A).<sup>11</sup> The *oxyC* gene product catalyzes the direct C–C coupling of amino acids 5 and 7 within the heptapeptide backbone of a glycopeptide.<sup>22</sup> In known gene clusters, *OxyC* is highly conserved, yet has sufficiently diverged from other monooxygenases to permit the design of *oxyC* gene specific degenerate PCR primers. In most sequenced glycopeptide gene clusters, the *oxyC* gene is positioned directly adjacent to the tailoring-enzyme region, and therefore, we felt it would be a good probe to use in our search for tailoring-enzyme-rich eDNA clones. PCR screening of the AZ and CA megalibraries with two sets of *oxyC* gene specific primers led to the identification of four unique *oxyC* gene sequences, three from the CA library and one from the AZ library. eDNA cosmid clones containing each unique *oxyC* gene sequence were recovered by serial dilution of the individual library aliquots from which the *oxyC* amplicon was originally generated and then fully sequenced using 454 pyrosequencing.

The annotation of the four *oxyC*-containing clones recovered from the CA and AZ libraries, together with two *oxyC*-containing clones associated with the VEG and TEG gene clusters found in the UT library revealed that each eDNA clone

contained genes encoding unique collections of tailoring enzymes (Figure 2A–C). The three clones isolated from the CA library as well as the tailoring-enzyme-rich clone associated with the VEG pathway contain different assortments of genes that encode glycosyltransferases (Figure 2, hollow red) and methyltransferases (Figure 2, blue). Since the bulk of the structural diversity seen in known glycopeptides arises from differences in the arrangement of sugar, methyl, and acyl moieties positioned around the heptapeptide core, the isolation of multiple clones with numerous glycosyl- and methyltransferase genes is not unexpected. The tailoring-enzyme-rich clone associated with the TEG pathway, as well as the clone AZ205 isolated from the AZ library, contain sulfotransferase genes (Figure 2, orange). Only two sulfated glycopeptides have been characterized from cultured bacteria.<sup>23,24</sup> Two of the six eDNA-derived tailoring-enzyme systems contain genes that encode predicted sulfotransferases, enzymes that were previously thought to be rare in glycopeptide biosynthesis. The direct cloning of microbial DNA from environmental samples appears to have provided access to a pool of biosynthetic gene clusters that has not been functionally examined using traditional pure culture-based techniques.

**In Vivo and in Vitro Production of Glycopeptide Derivatives Using eDNA-Derived Tailoring Enzymes.** Telavancin, a semisynthetic derivative of vancomycin recently approved for use by the FDA, differs from its natural counterpart by the addition of an anionic phosphono group and a long chain

(21) Kim, J.; Feng, Z.; Bauer, J. D.; Kallifidas, D.; Calle, P. Y.; Brady, S. F. *Biopolymers* **2010**, *93*, 833–844.

(22) Bischoff, D.; Pelzer, S.; Bister, B.; Nicholson, G. J.; Stockert, S.; Schirle, M.; Wohlleben, W.; Jung, G.; Süßmuth, R. D. *Angew. Chem., Int. Ed.* **2001**, *40*, 4688–4691.

(23) Holdom, K. S.; Maeda, H.; Ruddock, J. C.; Tone, J. (Pfizer, Ltd.) Eur. Patent Appl. 265143, EP-B, 1988.

(24) Boeck, L. D.; Mertz, F. P. *J. Antibiot. (Tokyo)* **1986**, *39*, 1533–40.

hydrophobic moiety.<sup>25</sup> The phosphono group, which we believe likely mimics the naturally occurring sulfate functionality that has been seen in a small number of known glycopeptide congeners, is reported to markedly improve the ADME (absorption, distribution, metabolism, and excretion) characteristics of telavancin. Due to the dearth of naturally occurring anionic glycopeptides characterized to date, we chose to explore the possibility of using the genes found on the six eDNA-derived tailoring-enzyme-rich clones to generate additional sulfated glycopeptides using both in vivo and in vitro derivatization strategies.

For our in vivo derivatization studies, each tailoring-enzyme-rich clone was initially retrofitted with the genetic elements necessary for transfer and integration into *Streptomyces* and then transferred by conjugation into *Streptomyces toyocaensis*.<sup>26</sup> *S. toyocaensis* is the natural producer of the monosulfated glycopeptide congener A47934 (**1**).<sup>24</sup> With the exception of the sulfate found on the N-terminal hydroxyphenylglycine (Hpg), A47934 is devoid of tailoring-enzyme-derived functionality. Thus, A47934 provides numerous sites where new functionality could be added by eDNA-derived tailoring-enzyme genes introduced into *S. toyocaensis*. While it is unlikely that all of the tailoring enzymes from any one clone would be functionally expressed using this strategy, the fact that each clone contains genes for multiple potential modification enzymes significantly increases the likelihood of new functionality being introduced onto A47934.

Cultures of both wild type *S. toyocaensis*, as well as *S. toyocaensis* transformed with each of the six retrofitted eDNA cosmids, were grown in *Streptomyces* antibiotic medium for 7 days, and then crude glycopeptide extracts were generated from the resulting culture broths.<sup>27</sup> Reversed-phase HPLC analysis of the crude extracts indicated that five out of the six *S. toyocaensis* transformants produced new A47934 derivatives. Each new compound was purified by HPLC and then analyzed by HRMS. On the basis of the molecular formulas predicted by HRMS, the introduction of either clone UTA15 (VEG pathway) or clone AZ205 leads to the addition of a new methyl substituent onto A47934, the introduction of clone UTD30 (TEG pathway) leads to the addition of a second sulfate onto A47934, and the introduction of either clone CA191 (CA37 pathway) or clone CA1627 (CA915 pathway) leads to the addition of a sugar onto A47934 (Figure 3). The position of each new functional group was determined by comparing 1D and 2D NMR spectra obtained from A47934 with those obtained from each derivative.

Compound **2**, the C-terminal methyl ester of A47934, is produced by *S. toyocaensis* transformed with clone UTA15. The new oxygen-substituted methyl (53.21 ppm in <sup>1</sup>H–<sup>13</sup>C HMQC spectra) is visible as a deshielded singlet (3.85 ppm) in the <sup>1</sup>H spectrum, and an <sup>1</sup>H–<sup>13</sup>C HMBC correlation from this new methyl to the C-terminal carbonyl carbon at 172.32 ppm allowed us to position the methyl at the C-terminus of the heptapeptide. Clone UTA15 contains three predicted methyltransferase genes (Figure 2A, blue), two of which are predicted to encode N-methyltransferases. The methyltransferase encoded by the

third methyltransferase gene found on this clone is therefore likely to be responsible for the addition of the methyl group to A47934.

The position of the second sulfate found on the disulfated derivative of A47934 (compound **3**) that is produced by *S. toyocaensis* transformed with clone UTD30 was elucidated on the basis of changes in <sup>13</sup>C chemical shifts observed for the Hpg at position 4 in the heptapeptide. In compound **3**, the chemical shift for carbon 4d shifts upfield by 4.57 ppm and the chemical shift for the carbons ortho to this carbon (4c and 4e) shift downfield by 5.12 and 4.76 ppm, compared to the chemical shifts for these same carbons in A47934. This pattern of chemical shift changes is predicted to arise from the sulfation of an aromatic hydroxyl and mimics the changes we observed when the Teg14 sulfotransferase was used to sulfate the teicoplanin aglycon in vitro.<sup>28</sup> The *teg14* gene is found on clone UTD30 and this sulfotransferase is likely responsible for this modification in vivo as well.

Cultures of *S. toyocaensis* transformed with either cosmid CA191 or cosmid CA1627 produce spectroscopically identical glycosylated derivatives of A47934 (**4**). On the basis of the new chemical shifts and observed coupling constants seen in 1D and 2D NMR experiments, this new sugar substituent was determined to be a β-linked D-glucose (see Supporting Information).<sup>28</sup> As is the case with the disulfated derivative of A47934 described above, changes in <sup>13</sup>C chemical shifts for the carbons found in the Hpg at position 4 in the heptapeptide allowed us to place the glucosyl group on this amino acid (Figure 3). Both cosmid CA191 and cosmid CA1627 contain multiple genes encoding glycosyltransferases (Figure 2B, hollow red) that could be responsible for appending the glucose moiety onto A47934.

Compound **5**, the N-terminal methyl amine of A47934, is produced by *S. toyocaensis* transformed with clone AZ205. As seen with compound **2**, the <sup>1</sup>H and <sup>1</sup>H–<sup>13</sup>C HMQC NMR spectra from compound **5** indicate the presence of a new heteroatom-substituted methyl substituent (<sup>1</sup>H 2.48 ppm, <sup>13</sup>C 31.34 ppm) (Figure 3). In this case, however, the chemical shifts suggest that the methyl is bound to a nitrogen atom instead of an oxygen atom. An <sup>1</sup>H–<sup>13</sup>C HMBC correlation from the new methyl to the α carbon (64.25 ppm) of the Hpg at position 1 in the heptapeptide confirmed the N-terminal methylation in compound **5**. ORF10 in clone AZ205 is predicted to encode an N-methyltransferase (Figure 2C, blue) that is likely responsible for the addition of the methyl to the N-terminus of A47934.

Many of the tailoring enzymes encoded on eDNA clones will not be expressed in a particular heterologous system, and the activity of others is undoubtedly thwarted either by incompatibility with the available substrate or by the occupation or absence of the site to be functionalized. In A47934, for example, the hydroxyl on Hpg1 is blocked by a sulfate and Tyr2 is missing the β carbon oxidation that is often seen in glycopeptides. Even with these limitations, five out of six eDNA-derived tailoring-enzyme-rich clones resulted in the production of new A47934 derivatives. The addition of clones in a combinatorial fashion or the cloning of individual eDNA-derived tailoring enzymes under the control of model promoters would likely provide access to even more highly functionalized derivatives, as would the introduction of these same clones into other genetically tractable glycopeptide producers.

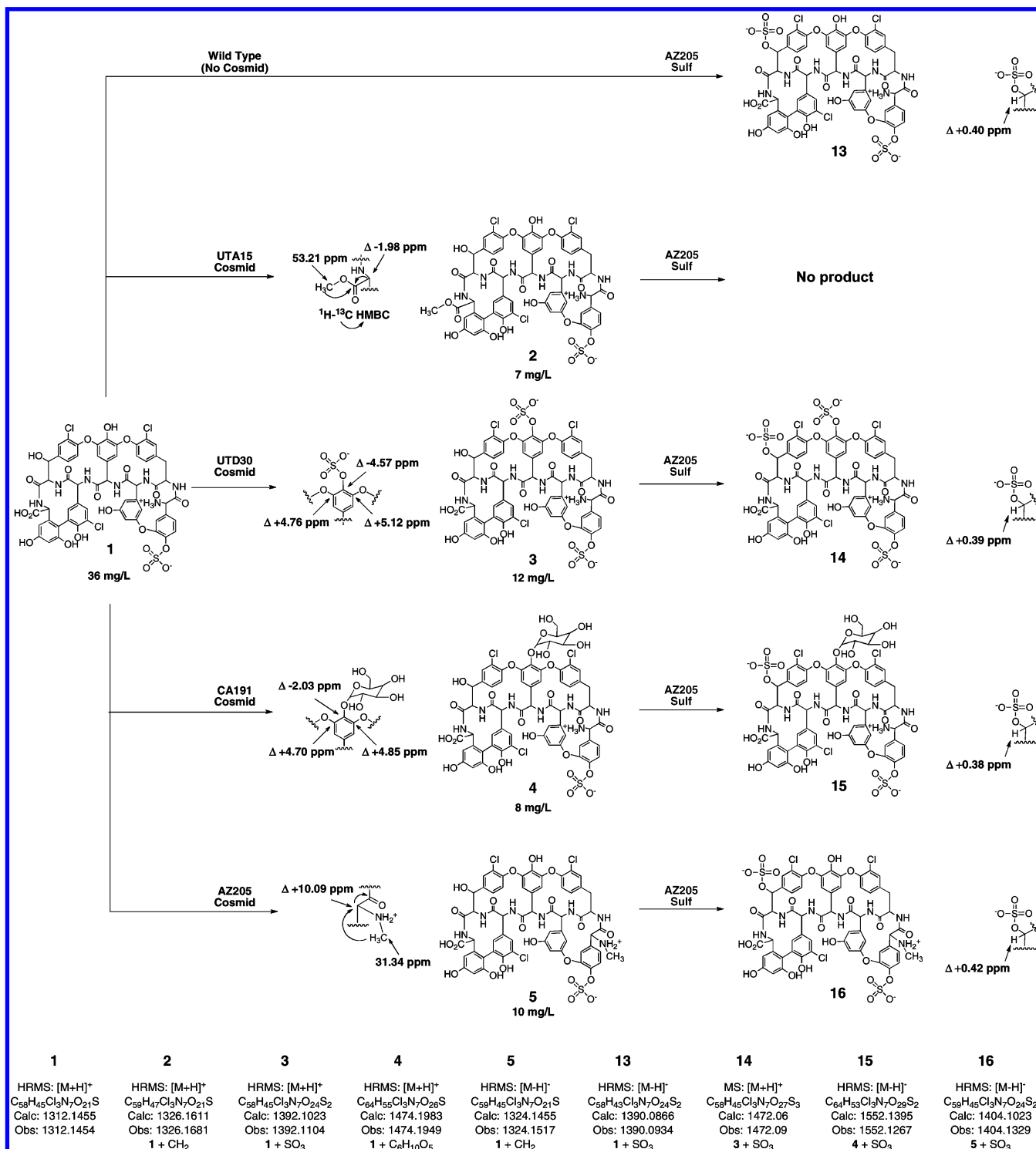
In some instances where tailoring enzymes are not expressed in vivo, it should be possible to gain access to the functionality

(25) Leadbetter, M. R.; Adams, S. M.; Bazzini, B.; Fatheree, P. R.; Karr, D. E.; Krause, K. M.; Lam, B. M.; Linsell, M. S.; Nodwell, M. B.; Pace, J. L.; Quast, K.; Shaw, J. P.; Soriano, E.; Trapp, S. G.; Villena, J. D.; Wu, T. X.; Christensen, B. G.; Judice, J. K. *J. Antibiot. (Tokyo)* **2004**, *57*, 326–36.

(26) Matsushima, P.; Baltz, R. H. *Microbiology* **1996**, *142*, 261–7, Pt 2.

(27) Lamb, S. S.; Patel, T.; Koteva, K. P.; Wright, G. D. *Chem. Biol.* **2006**, *13*, 171–81.

(28) Pretsch, E. *Structure Determination of Organic Compounds: Tables of Spectral Data*; 4th ed.; Springer: New York, 2009.

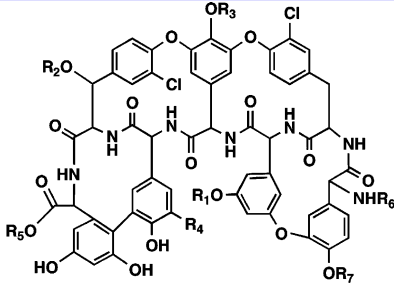


**Figure 3.** New glycopeptide congeners produced using eDNA-derived tailoring enzymes. Compounds 2–5 were produced in vivo by *S. toyocaensis* transformed with tailoring-enzyme-rich eDNA clones. Compounds 13–16 were generated in vitro using recombinantly expressed, eDNA-derived sulfotransferase AZ205.

of these enzymes using an in vitro approach. Previously, we showed that the three sulfotransferases (Teg12, -13, and -14) from the TEG gene cluster (Figure 2A, orange) could be used in vitro to produce new sulfated teicoplanin aglycon derivatives (6–12).<sup>11</sup> By coupling the in vivo approach used to generate compounds 2–5 with the in vitro approach used to generate compounds 6–12, we were able to produce an additional set of sulfated A47934 analogs.<sup>11</sup> In the presence of the cosubstrate 3'-phosphoadenosine-5'-phosphosulfate (PAPS) and the teico-

planin aglycon, recombinant AZ205 sulfotransferase expressed as a His-tagged protein in *Escherichia coli* was found to produce a monosulfated teicoplanin aglycon that is spectroscopically identical to compound 7, the sulfated product generated by Teg13. Both sulfotransferases from clone AZ205 and Teg13 therefore specifically sulfate the  $\beta$  carbon hydroxyl on the  $\beta$ -hydroxytyrosine (Bht) found at position 6 in glycopeptides. Sulfonation reactions using recombinant AZ205 sulfotransferase were subsequently carried out on A47934 as well as on each of





|    | R1                           | R2                           | R3                           | R4 | R5              | R6              | R7                           | Parent glycopeptide       | Modification method          |
|----|------------------------------|------------------------------|------------------------------|----|-----------------|-----------------|------------------------------|---------------------------|------------------------------|
| 2  | H                            | H                            | H                            | Cl | CH <sub>3</sub> | H               | SO <sub>3</sub> <sup>-</sup> | A47934 (1)                | UTA15 Cosmid                 |
| 3  | H                            | H                            | SO <sub>3</sub> <sup>-</sup> | Cl | H               | H               | SO <sub>3</sub> <sup>-</sup> | A47934 (1)                | UTD30 Cosmid                 |
| 4  | H                            | H                            | Glucose                      | Cl | H               | H               | SO <sub>3</sub> <sup>-</sup> | A47934 (1)                | CA191 Cosmid                 |
| 5  | H                            | H                            | H                            | Cl | H               | CH <sub>3</sub> | SO <sub>3</sub> <sup>-</sup> | A47934 (1)                | AZ205 Cosmid                 |
| 6  | SO <sub>3</sub> <sup>-</sup> | H                            | H                            | H  | H               | H               | H                            | Teicoplanin aglycone      | TEG12 sulfotransferase       |
| 7  | H                            | SO <sub>3</sub> <sup>-</sup> | H                            | H  | H               | H               | H                            | Teicoplanin aglycone      | TEG13 sulfotransferase       |
| 8  | H                            | H                            | SO <sub>3</sub> <sup>-</sup> | H  | H               | H               | H                            | Teicoplanin aglycone      | TEG14 sulfotransferase       |
| 9  | SO <sub>3</sub> <sup>-</sup> | SO <sub>3</sub> <sup>-</sup> | H                            | H  | H               | H               | H                            | Teicoplanin aglycone      | TEG12,13 sulfotransferases   |
| 10 | SO <sub>3</sub> <sup>-</sup> | H                            | SO <sub>3</sub> <sup>-</sup> | H  | H               | H               | H                            | Teicoplanin aglycone      | TEG12,14 sulfotransferases   |
| 11 | H                            | SO <sub>3</sub> <sup>-</sup> | SO <sub>3</sub> <sup>-</sup> | H  | H               | H               | H                            | Teicoplanin aglycone      | TEG13,14 sulfotransferases   |
| 12 | SO <sub>3</sub> <sup>-</sup> | SO <sub>3</sub> <sup>-</sup> | SO <sub>3</sub> <sup>-</sup> | H  | H               | H               | H                            | Teicoplanin aglycone      | TEG12,13,14 sulfotransferase |
| 13 | H                            | SO <sub>3</sub> <sup>-</sup> | H                            | Cl | H               | H               | SO <sub>3</sub> <sup>-</sup> | A47934 (1)                | AZ205 sulfotransferase       |
| 14 | H                            | SO <sub>3</sub> <sup>-</sup> | SO <sub>3</sub> <sup>-</sup> | Cl | H               | H               | SO <sub>3</sub> <sup>-</sup> | A47934 + UTD30 Cosmid (3) | AZ205 sulfotransferase       |
| 15 | H                            | SO <sub>3</sub> <sup>-</sup> | Glucose                      | Cl | H               | H               | SO <sub>3</sub> <sup>-</sup> | A47934 + CA191 Cosmid (4) | AZ205 sulfotransferase       |
| 16 | H                            | SO <sub>3</sub> <sup>-</sup> | H                            | Cl | H               | CH <sub>3</sub> | SO <sub>3</sub> <sup>-</sup> | A47934 + AZ205 Cosmid (5) | AZ205 sulfotransferase       |

**Figure 4.** New sulfated glycopeptide derivatives produced using combinations of in vivo and in vitro methods using eDNA-derived tailoring enzymes.

the A47934 derivatives that were generated in vivo. For compounds **1**, **3**, **4**, and **5**, we observed complete conversion of each metabolite to a new polysulfated derivative (**13**, **14**, **15**, and **16**) after 16 h. In each case, HRMS confirmed that one additional sulfate had been added to the glycopeptide starting material. As expected from the addition of a sulfate, each  $\beta$  carbon proton at position 6 is deshielded by  $\sim 0.4$  ppm compared to the chemical shift for this same proton in the starting material (Figure 3). No reaction was observed when compound **2** was used as a substrate. Methylation of the C-terminal carboxylic acid likely disrupts a key interaction that is required for catalysis or substrate recognition by the AR205 sulfotransferase.

In total, 15 new sulfated glycopeptides have been generated using the enzymes found on tailoring-enzyme-rich clones recovered from the UT, CA, and AZ eDNA megalibraries (Figure 4). This represents a significant addition to the sulfated glycopeptides that have been characterized from cultured *Actinomycetes*. Minimum inhibitory concentrations (MIC) for each compound were determined using both glycopeptide-sensitive (*Staphylococcus aureus* ATCC 6538P and *Enterococcus faecalis* ATCC 47077) and glycopeptide-resistant (methicillin resistant *S. aureus* USA300<sup>29</sup> and VanA-type-resistant *E. faecalis* EF18<sup>30</sup>) bacteria (Table 1). As has been seen with other glycopeptide congeners, the more highly functionalized derivatives tend to show higher MICs. While highly functionalized glycopeptide congeners often show increased in vitro MICs, they have proved useful as therapeutics because some exhibit

**Table 1.** MIC ( $\mu\text{g/mL}$ ) for Compounds **1–16** and the Teicoplanin Aglycon

| compd               | <i>S. aureus</i> |        | <i>E. faecalis</i> |      |
|---------------------|------------------|--------|--------------------|------|
|                     | ATCC 6538P       | USA300 | ATCC 47077         | EF18 |
| <b>1</b>            | 2                | 2      | 16                 | >64  |
| <b>2</b>            | 1                | 1      | 8                  | >64  |
| <b>3</b>            | 4                | 8      | 64                 | >64  |
| <b>4</b>            | 4                | 8      | 64                 | >64  |
| <b>5</b>            | 2                | 2      | 16                 | >64  |
| <b>6</b>            | 2                | 2      | 32                 | >64  |
| <b>7</b>            | 1                | 1      | 16                 | >64  |
| <b>8</b>            | 4                | 4      | 64                 | >64  |
| <b>9</b>            | 4                | 2      | 64                 | >64  |
| <b>10</b>           | 8                | 8      | >64                | >64  |
| <b>11</b>           | 8                | 8      | >64                | >64  |
| <b>12</b>           | 16               | 8      | >64                | >64  |
| <b>13</b>           | 2                | 4      | 16                 | >64  |
| <b>14</b>           | 16               | 16     | >64                | >64  |
| <b>15</b>           | 8                | 16     | 16                 | >64  |
| <b>16</b>           | 2                | 4      | 16                 | >64  |
| teicoplanin aglycon | 1                | 1      | 8                  | >64  |

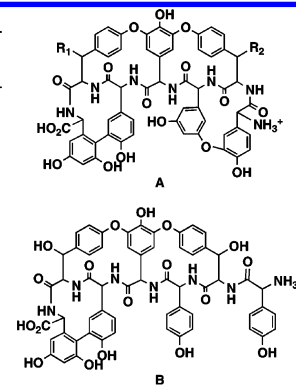
better pharmacological properties in vivo. The systematic introduction of tailoring-enzyme-rich clones to easily cultured glycopeptide producers should prove to be a simple yet effective strategy for generating collections of glycopeptides with new functionalization patterns that can be examined for improved biological activities.

**Recovery of Complete eDNA-Derived Glycopeptide Gene Clusters.** With so few sulfated glycopeptides having been characterized from cultured bacteria, it is clear that the two tailoring-enzyme-rich clones that contain sulfotransferases must have arisen from gene clusters that contain the genetic capacity to generate novel glycopeptide congeners. However, in those cases where the tailoring-enzyme-rich clones do not contain

(29) Diep, B. A.; Gill, S. R.; Chang, R. F.; Phan, T. H.; Chen, J. H.; Davidson, M. G.; Lin, F.; Lin, J.; Carleton, H. A.; Mongodin, E. F.; Sensabaugh, G. F.; Perdreau-Remington, F. *Lancet* **2006**, *367*, 731–739.

(30) Mato, R.; de Lencastre, H.; Roberts, R. B.; Tomasz, A. *Microb. Drug Resist.* **1996**, *2*, 309–17.

| Source   | Gene Cluster         | Peptide | Skeleton       |                | Transferases |          |        |          |       |      | Halogenase |
|----------|----------------------|---------|----------------|----------------|--------------|----------|--------|----------|-------|------|------------|
|          |                      |         | R <sub>1</sub> | R <sub>2</sub> | Glycosyl     | Mannosyl | Methyl | N-Methyl | Sulfo | Acyl |            |
| Cultured | Teicoplanin Molecule | A       | H or OH        | H or OH        | 2            | 1        | 0      | 0        | 0     | 1    | 1          |
|          | A47934 Molecule      | A       | H or OH        | H or OH        | 2            | 1        | 0      | 0        | 0     | 1    | 2          |
|          | A40926 Molecule      | A       | H or OH        | H or OH        | 0            | 0        | 0      | 0        | 1     | 0    | 3          |
|          | A40926 Molecule      | A       | H or OH        | H or OH        | 1            | 1        | 0      | 1        | 0     | 1    | 1          |
| eDNA     | CA37                 | A       | H or OH        | H or OH        | 1            | 1        | 0      | 1        | 0     | 1    | 2          |
|          | CA915                | A       | H or OH        | H or OH        | 4            | 1        | 3      | 0        | 0     | 0    | 1          |
|          | CA878                | A       | OH             | OH             | 4            | 2        | 2      | 0        | 0     | 0    | 1          |
|          | TEG                  | A       | OH             | OH             | 0            | 0        | 0      | 0        | 3     | 0    | 1          |
|          | VEG                  | B       | OH             | OH             | 4            | 2        | 1      | 2        | 0     | 0    | 1          |



**Figure 5.** Structures A and B correspond to the cross-linked core peptides that are predicted to arise from eDNA-derived NRPS systems. An inventory of the number and type of tailoring enzymes (or functional group transferases) found in sequenced gene clusters from cultured soil bacteria, as well as gene clusters cloned from eDNA, is shown. For gene clusters associated with known metabolites, the number of tailoring-enzyme-derived functional groups found on the encoded metabolite is also shown.

**Table 2.** Adenylation Domain Specificity Predictions for all NPRS Modules Cloned from the CA Megalibrary<sup>a</sup>

|      | CA37            |                |                  |            | CA915           |                |                  |            | CA878           |                |                  |            |
|------|-----------------|----------------|------------------|------------|-----------------|----------------|------------------|------------|-----------------|----------------|------------------|------------|
|      | observed pocket | closest pocket | closest relative | amino acid | observed pocket | closest pocket | closest relative | amino acid | observed pocket | closest pocket | closest relative | amino acid |
| Mod1 | DAFLGLL         | DAFLGLL        | Tcp-Mod1         | Hpg        | DAFLGLL         | DAFLGLL        | Tcp-Mod1         | Hpg        | DAFLGLL         | DAFLGLL        | Tcp-Mod1         | Hpg        |
| Mod2 | DASTVAAV        | DASTVAAV       | A47-Mod2         | Tyr        | DASTVAAV        | DASTVAAV       | A47-Mod2         | Tyr        | DTSKVAAI        | DTSKVAAI       | Cep-Mod2         | BHT        |
| Mod3 | DAYFSGSL        | DAYFLGTL       | Tcp-Mod3         | Dpg        | DAYFLGTL        | DAFLGLL        | A47-Mod1         | Hpg        | DAYFQGTG        | DAYFLGTL       | Tcp-Mod3         | Dpg        |
| Mod4 | DIFTLGGL        | DIFHLGGL       | Cep-Mod4         | Hpg        | DIFHLGGL        | DIFHLGGL       | Cep-Mod4         | Hpg        | DIFTLGGL        | DIFHLGGL       | Cep-Mod4         | Hpg        |
| Mod5 | DALLLGLL        | DALLLGVG       | Tcp-Mod5         | Hpg        | DALLLGVG        | DALLLGVG       | Tcp-Mod5         | Hpg        | DAVLLGGL        | DAVHLGGL       | Cep-Mod5         | Hpg        |
| Mod6 | DASTVAAV        | DASTVAAV       | A47-Mod2         | Tyr        | DASTVAAV        | DASTIAGV       | Tcp-Mod6         | Tyr        | DASTLGAI        | DASTLGAI       | Bal-Mod6         | BHT        |
| Mod7 | DPYLGGLL        | DPYLGGLL       | Cep-Mod7         | Dpg        | DPYLGGLL        | DPYLGGLL       | Cep-Mod7         | Dpg        | DPYLGGLL        | DPYLGGLL       | Cep-Mod7         | Dpg        |

<sup>a</sup> Tcp = teicoplanin, A47 = A47934, Cep, chloroeremomycin, Bal = balhimycin, Mod = module.

sulfotransferases, it is not immediately obvious from the information captured on these individual clones whether they arose from gene clusters that have the capacity to encode the production of novel congeners. A more complete picture of each CA gene cluster was therefore obtained from overlapping cosmid clones that were recovered from the CA megalibrary in additional rounds of library screening. This rescreening process was continued iteratively until end sequencing of the newly recovered overlapping clones failed to identify additional glycopeptide biosynthetic machinery. After full sequencing and annotation, each set of overlapping cosmid clones could be assembled into a single 100–120 kb contig that appears to contain a full glycopeptide biosynthetic gene cluster (Figure 2B) (GenBank accession numbers: CA37 HM486074, CA878 HM486075, CA915 HM486076). All three gene clusters recovered from the CA library are predicted to contain a complete complement of glycopeptide biosynthetic genes and show the same conserved gene cluster architecture that is seen in most previously sequenced glycopeptide gene clusters (Figure 2A).

The structure of the congener(s) encoded by a glycopeptide gene cluster is determined by the specificity of the nonribosomal peptide synthetases (NRPSs) and the tailoring enzymes that are encoded by the gene cluster. A detailed accounting of these enzymes should therefore shed light onto the structure of the molecule that each CA gene cluster has the potential to encode and thereby allow for a comparison of these predicted structures to known glycopeptide congeners (Figure 5). Each CA gene cluster contains seven NRPS modules distributed over either four or five open reading frames (Figure 2B, vertical striped green). In addition to the seven NRPS modules that are predicted to produce the heptapeptide, each CA gene cluster encodes four predicted monooxygenases that are homologues of OxyA–D

(Figure 2B, brown). OxyA–D catalyze the four oxidative couplings that are needed to generate the four macrocycles seen in many glycopeptides.<sup>18</sup>

NRPS modules are predicted to carry out a single condensation reaction that extends the growing peptide by one amino acid.<sup>31</sup> At a minimum, a module contains a condensation domain for condensing an incoming amino acid with the growing peptide, an adenylation domain for selecting and activating the incoming amino acid, and a thiolation domain that carries the growing peptide. The amino acid incorporated by a module is determined by a small number of variable residues found in the adenylation domain.<sup>32,33</sup> Table 2 shows these key amino acid specificity residues from each adenylation domain found in the three CA-derived glycopeptide gene clusters. On the basis of a comparison of these residues to those seen in adenylation domains from sequenced glycopeptide NRPS modules, all three CA gene clusters are expected to encode the production of an Hpg-Bht(Tyr)-Dpg-Hpg-Hpg-Bht(Tyr)-Dpg-like heptapeptide (Table 2).<sup>34</sup> Potential ambiguities in the predicted peptide arise from the mode of Bht incorporation, which has been reported to occur by two distinct mechanisms during glycopeptide biosynthesis. Bht can be biosynthesized and incorporated directly into the glycopeptide backbone through Bht-specific adenylation domains, or it can be produced by oxidizing tyrosine in a reaction that takes place after the selective incorporation of

(31) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. *Chem. Rev.* **1997**, *97*, 2651–2674.

(32) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6*, 493–505.

(33) Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000**, *7*, 211–24.

(34) Ansari, M. Z.; Yadav, G.; Gokhale, R. S.; Mohanty, D. *Nucleic Acids Res.* **2004**, *32*, W405–13.

tyrosine into the growing peptide.<sup>35,36</sup> Both CA37 and CA915 are predicted to generate Bht through the latter, tyrosine-oxidation strategy. Modules 2 and 6 from both gene clusters are predicted to incorporate Tyr into the growing peptide. However, as can be seen in a similar analysis of both the teicoplanin and A47934 NRPS modules, the final residues at these positions will be Bht if the tyrosines are subsequently oxidized.

For sequenced gene clusters, the number and type of tailoring enzymes found in these clusters is a reliable predictor of the features that are seen in the most highly functionalized congeners they produce (Figure 5).<sup>15–18,37</sup> The inventory of tailoring enzymes found in cryptic eDNA-derived glycopeptide gene clusters should in turn be a reasonable predictor of the features found in the congeners produced by these gene clusters. This type of analysis does not allow for precise structural predictions to be made because it does not account for differences in regioselectivity, subtle differences in substrate specificities, or the functional order of the tailoring enzymes. It can, however, provide an indication as to whether or not a cryptic gene cluster has the potential to generate a glycopeptide congener with a unique number or type of prosthetic groups.

The CA37 and CA915 gene clusters each contain genes encoding four glycosyltransferases, a mannosyltransferase, a halogenase, and three SAM-dependent methyltransferases. Both gene clusters also contain genes encoding enzymes used in the biosynthesis of a number of different rare glycopeptide-associated sugars, including vancosamine, acosamine, or ristosamine.<sup>38–40</sup> It is not possible from the DNA sequence alone to tell which of these sugars will actually be produced by the two gene clusters. Of the known glycopeptide congeners that contain five sugars, only actaplanin contains the same cross-linked heptapeptide that is predicted to arise from the CA37 and CA915 gene clusters.<sup>40</sup> These gene clusters may therefore encode the biosynthesis of actaplanin itself or more highly glycosylated and methylated actaplanin-like glycopeptides.

The CA878 gene cluster contains genes encoding four glycosyltransferases, two mannosyltransferases, a halogenase, and two methyltransferases. This gene cluster also contains homologues of the same sugar biosynthetic genes that are seen in CA37 and CA915. On the basis of the collection of tailoring-enzyme genes present in CA878, this gene cluster would appear to encode the biosynthesis of a ristocetin-like congener. Ristocetin contains the same heptapeptide as is predicted to arise from the CA878 gene cluster as well as six sugars, including a ristosamine, and two methyl substituents.<sup>41</sup> The presence of a halogenase in CA878 suggests that this cluster probably does

not produce ristocetin itself, but instead produces a halogenated ristocetin-like congener.

## Conclusions

First-generation glycopeptide antibiotics like vancomycin and teicoplanin are still used extensively in the clinic.<sup>14</sup> However, their generally poor pharmacological properties and the appearance of resistant isolates have drawn attention to the increasing need for second-generation glycopeptides with improved ADME characteristics and broader spectra of activity.<sup>25,42–44</sup> In response to this increased interest, both biosynthetic and synthetic approaches have been used to generate additional glycopeptide derivatives containing functional groups and functional group arrangements that differ from those seen in naturally occurring metabolites.<sup>27,45–47</sup> The tailoring of glycopeptide scaffolds with new eDNA-derived tailoring enzymes provides an additional means of generating collections of new glycopeptide derivatives that can be examined in future screening efforts.

With culture-independent strategies, it is possible to explore the biosynthetic potential of thousands of unique soil-associated bacterial species simultaneously. Large environmental DNA libraries are likely to be bountiful sources of tailoring-enzyme-rich clones that could serve as toolboxes for either the in vivo or the in vitro derivatization of many, if not most, different classes of bacterially derived natural products. One key advantage to using this approach, over other enzyme based strategies for generating unnatural natural product variants, is that each tailoring enzyme used in a modification reaction will have evolved to interact with the specific family of natural products being derivatized.

## Materials and Methods

**Megalibrary Construction.** The eDNA cosmid libraries screened in this study were constructed using previously described methods.<sup>48</sup> In brief, soil which had been passed through a 1/8" screen was mixed with lysis buffer (100 mM Tris-HCl, 100 mM Na EDTA, 1.5 M NaCl, 1% (w/v) cetyltrimethylammonium bromide (CTAB), 2% (w/v) SDS, pH 8.0) and incubated at 70 °C for 2 h. Soil particulates were removed by centrifugation (3500g, 30 min), and crude eDNA was precipitated from the centrifuge-clarified soil suspension with the addition of 0.7 volumes of 2-propanol. The crude eDNA extract was collected by centrifugation (3500g, 30 min), washed with 70% ethanol, and allowed to air-dry. DNA was separated from the residual mud using preparative agarose gel electrophoresis (1.0%), and after 16 h at 20 V, the high molecular weight band (>25 kb) was excised from the gel. Purified eDNA was collected by electroelution (2 h at 100 V) and concentrated using a centrifugal concentrator (30K MWCO, Vivascience). Blunt ended (End-It Kit, Epicenter) eDNA was ligated into either pWEB or pWEB-TNC, packaged into  $\lambda$  phage and transfected into *E. coli* EC100 (Epicenter). Over 2000 unique 4000–5000-membered libraries were constructed from the transfected *E. coli* for each megalibrary. For each unique library, a matching DNA miniprep (Qiagen) and glycerol stock set was archived.

- (35) Puk, O.; Bischoff, D.; Kittel, C.; Pelzer, S.; Weist, S.; Stegmann, E.; Süßmuth, R. D.; Wohlleben, W. *J. Bacteriol.* **2004**, *186*, 6093–100.  
 (36) Stinchi, S.; Carrano, L.; Lazzarini, A.; Feroggio, M.; Grigoletto, A.; Sosio, M.; Donadio, S. *FEMS Microbiol. Lett.* **2006**, *256*, 229–35.  
 (37) Goldstein, B. P.; Selva, E.; Gastaldo, L.; Berti, M.; Pallanza, R.; Ripamonti, F.; Ferrari, P.; Denaro, M.; Arioli, V.; Cassani, G. *Antimicrob. Agents Chemother.* **1987**, *31*, 1961–6.  
 (38) Chen, H.; Thomas, M. G.; Hubbard, B. K.; Losey, H. C.; Walsh, C. T.; Burkart, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11942–7.  
 (39) Heald, S. L.; Mueller, L.; Jeffs, P. W. *J. Antibiot. (Tokyo)* **1987**, *40*, 630–45.  
 (40) Debono, M.; Merkel, K. E.; Molloy, R. M.; Barnhart, M.; Presti, E.; Hunt, A. H.; Hamill, R. L. *J. Antibiot. (Tokyo)* **1984**, *37*, 85–95.  
 (41) Fehlner, J. R.; Hutchinson, R. E.; Tarbell, D. S.; Schenck, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 2420–1.

- (42) Hiramatsu, K.; Aritaka, N.; Hanaki, H.; Kawasaki, S.; Hosoda, Y.; Hori, S.; Fukuchi, Y.; Kobayashi, I. *Lancet* **1997**, *350*, 1670–3.  
 (43) Cetinkaya, Y.; Falk, P.; Mayhall, C. G. *Clin. Microbiol. Rev.* **2000**, *13*, 686–707.  
 (44) Ge, M.; Chen, Z.; Onishi, H. R.; Kohler, J.; Silver, L. L.; Kerns, R.; Fukuzawa, S.; Thompson, C.; Kahne, D. *Science* **1999**, *284*, 507–11.  
 (45) Solenberg, P. J.; Matsushima, P.; Stack, D. R.; Wilkie, S. C.; Thompson, R. C.; Baltz, R. H. *Chem. Biol.* **1997**, *4*, 195–202.  
 (46) Chen, L.; Walker, D.; Sun, B.; Hu, Y.; Walker, S.; Kahne, D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5658–63.  
 (47) Kahne, D.; Leimkuhler, C.; Lu, W.; Walsh, C. *Chem. Rev.* **2005**, *105*, 425–48.  
 (48) Brady, S. F. *Nat. Protoc.* **2007**, *2*, 1297–305.



**Clone Recovery and Sequencing.** DNA from the individual 4000–5000-membered library pools was arrayed into 8 × 8 grids. Aliquots of DNA from each pool in a grid were then combined into unique sublibraries. DNA from each sublibrary pool was screened by PCR with two different sets of *oxyC*-specific degenerate primers: PS1FWD, 5'-ATGCTSAACSCCGAGTTCACSGTVC GG-3' and PS1REV, 5'-CGAGTRRTGGAYGCCGTGCCCGAA-3'; PS2FWD, 5'-CCGCAATTCASC MARKMGAARTCSG-3' and PS2REV, 5'-TGCCCKGGCRAKGAGGTTGTC-3'. PS1 was designed based on the *oxyC* gene sequences found in five cultured glycopeptide producers: *Amycolatopsis orientalis*, which produces chloroeremomycin; *Amycolatopsis balhimyceticus*, which produces balhimycin; *S. toyocaensis*, which produces A47934; *Nonomuraea* sp. ATCC 39727, which produces A40926; and *Actinoplanes teichomyceticus*, which produces teicoplanin. PS2 was designed to encompass the five *oxyC* sequences used in the design of PS1, as well as the two *oxyC* sequences previously cloned from eDNA. Each 25  $\mu$ L PCR reaction contained 0.5  $\mu$ L of cosmid DNA, 2.5  $\mu$ L of ThermoPol Buffer (New England Biolabs), 1.25  $\mu$ L of DMSO, 0.625  $\mu$ L of each 100 mM oligonucleotide primer, 0.5  $\mu$ L of 10 mM dNTPs mix, and 1 unit of Taq DNA Polymerase. For the PS1 primer set, the cycling parameters were as follows: seven touch down cycles of 95 °C for 30 s, 70–62 °C (–1 °C/cycle) for 30 s, and 72 °C for 90 s, followed by 20 cycles of 95 °C for 30 s, 62 °C for 30 s, and 72 °C for 90 s. For the PS2 primer set the following cycling parameters were used: 30 cycles of 95 °C for 30 s, 62 °C for 30 s, 72 °C for 30 s. PCR amplicons of the correct predicted size (700–800 bp for PS1 and 380–410 bp for the PS2) were gel-purified, A/T cloned into PCR2.1 (Invitrogen), and sequenced. DNA from the individual pools that were used to create each large sublibrary that yielded an *oxyC* sequence of interest was then screened using the same set of PCR primers. Once an individual sublibrary pool containing that sequence was identified, the cosmid of interest was recovered from the corresponding glycerol stock using successive rounds of serial dilution and PCR screening. Recovered cosmid clones were end-sequenced using vector specific primers: T7 Promoter and M13 Universal Fwd –40. End-sequencing data was used to design new sets of PCR primers that could be used to recover clones containing overlapping DNA sequences. This PCR screening process was repeated iteratively as needed for each gene cluster.

All cosmid clones were 454 sequenced (Roche) by the Memorial Sloan Kettering Cancer Center Genomics Core Laboratory. Raw reads were processed with Newbler (Roche) and Velvet.<sup>49</sup> Sequence manipulation and gene identification were carried out using MacVector and BLASTX, respectively. The specificity of each NRPS adenylation domain was determined using NRPS-PKS analysis Web site.<sup>34</sup>

**Cosmid Retrofitting and Conjugation into *S. toyocaensis*.** Cosmids containing glycopeptide tailoring enzymes were each retrofitted to contain the genetic elements required for conjugation, integration, and selection in *Streptomyces*. Each cosmid was initially digested with *PsiI* (New England Biolabs) and then ligated to the 6.7 kb *DraI* fragment from pOJ436.<sup>50</sup> This *DraI* fragment contains the  $\phi$ C31 integration machinery, the RK2 *oriT* and the *aac(3)IV* apramycin resistance gene. Upon transformation into *E. coli*, retrofitted cosmids could be easily identified by selecting with both ampicillin and apramycin. Correctly retrofitted cosmids were transformed into *E. coli* S17.1 and conjugated into *S. toyocaensis* NRRL 15009 using previously described methods, with a few minor modifications.<sup>26</sup> For each conjugation, a 50 mL culture of *E. coli* S17.1 transformed with an appropriately retrofitted cosmid clone was grown to midlogarithmic phase (LB with 50  $\mu$ g/mL apramycin and 10  $\mu$ g/mL trimethoprim). Cells were then pelleted by centrifugation (3500g, 30 min) and the resulting cell pellet was washed

twice with 50 mL of antibiotic free LB. Washed *E. coli* were resuspended in 1 mL of SOC (per liter: 20 g of Bacto Tryptone, 5 g of Bacto Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 20 mM glucose). Twenty microliters of a heat-treated (50 °C for 10 min) spore slurry (OD<sub>450</sub> = 40–80) was mixed with 50  $\mu$ L of washed *E. coli*, and this cell suspension was then plated onto R2 agar. After 16 h at 30 °C, the plates were flooded with 1 mL of sterile water containing 0.1 mg of apramycin and 0.1 mg of nalidixic acid and then returned to the incubator until exconjugants appeared (3–4 days). Exconjugants were restructured on modified Bennett's agar containing 25  $\mu$ g/mL apramycin and 25  $\mu$ g/mL nalidixic acid.

**Glycopeptide Congener Production.** *S. toyocaensis* was initially grown on modified Bennett's agar for 3 days at 30 °C. Single colonies were inoculated into 50 mL of *Streptomyces* vegetative medium (SVM) and incubated at 30 °C for 3 days (200 rpm orbital shaking); 50 mL of *Streptomyces* antibiotic medium containing 1 g/L yeast extract (SAM) was inoculated with a 1:1000 dilution of the 3-days-old starter culture. After 7 days (30 °C, 200 rpm orbital shaking) the cell mass was collected by centrifugation (3500g, 10 min). The wet weight of the pellet was determined and the pellet was resuspended (vortex, 30 s) in 0.75 mL of 2.5% (v/v) NH<sub>4</sub>OH per gram of wet cell pellet. This basic cell suspension (final pH >11) was centrifuged (3500g, 10 min) and the supernatant collected. The pH of the supernatant was adjusted to 7.5 with 1 M HCl and analyzed directly by HPLC [C18 (4.6  $\mu$ m × 150 mm), 20 mM ammonium acetate:CH<sub>3</sub>CN, from 95:5 to 70:30 over 20 min, 1.5 mL/min, or C18 (10  $\mu$ m × 150 mm), 20 mM ammonium acetate:CH<sub>3</sub>CN, from 95:5 to 70:30 over 20 min, 7 mL/min].

**In Vitro Glycopeptide Production.** The sulfotransferase gene from AZ205 was amplified from cosmid DNA using the Failsafe system (Epicentre) and the following primers: AZ205Sulf-FWD(*Bam*HI), GCGCGGATCCATGAACGGAATCCGATGG and AZ205SulfREV(*Hind*III), GCGCAAGCTTTCCTAATCAGCGTACCCGTA (restriction sites added for cloning are shown in bold). PCR conditions were as follows: 30 rounds of 30 s at 98 °C, 30 s at 65 °C, and 1 min at 72 °C. The amplicon was doubly digested (*Bam*HI/*Hind*III) (New England Biolabs), ligated into similarly digested pET28a, and transformed into *E. coli* BL21(DE3) for protein expression. For protein expression, cultures were grown at 37 °C to an OD<sub>600</sub> of 0.6, induced with 0.5 mM IPTG, and then allowed to shake at 20 °C for 16 h. The induced culture was then pelleted by centrifugation (3200g for 30 min) and the cell pellet was resuspended in 40 mL of lysis buffer [50 mM Hepes, pH 7.5, 0.5 M NaCl, 5% (v/v) glycerol, 20 mM imidazole, pH 8, 10 mM  $\beta$ -mercaptoethanol, and 0.5% (v/v) Triton X-100]. The resuspended cells were lysed by sonication. The cleared cell lysate (15 000g for 30 min) was incubated with 1 mL of Ni-NTA resin for 15 min. The resulting slurry was loaded onto a column, washed with 40 mL of lysis buffer, followed by 40 mL of wash buffer [50 mM Hepes, pH 7.5, 0.5 M NaCl, 5% (v/v) glycerol, 20 mM imidazole, pH 8.0, and 10 mM  $\beta$ -mercaptoethanol], and the protein was then eluted with the addition of 15 mL of elution buffer [50 mM Hepes, pH 7.5, 0.5 M NaCl, 5% (v/v) glycerol, 125 mM imidazole, pH 8, and 10 mM  $\beta$ -mercaptoethanol]. Sulfation reactions were run at 30 °C for 16 h in a final volume of 6 mL of 15 mM HEPES, pH 7.5, 1 mM 3'-phosphoadenosine-5'-phosphosulfate (PAPS), 0.1 mM DTT. Each reaction contained 6 mg of purified recombinant protein and 0.5 mg of substrate (compound 1–5). Sulfated products (13–16) were then purified from heat-killed (99 °C, 10 min) reaction mixtures by preparative HPLC.

**Glycopeptide Congener Analysis.** High-resolution ESI-MS experiments were carried out either on a Thermo-Fisher LTQ-Obritrapp mass spectrometer interfaced with a Dionex U3000 capillary/nano-HPLC system or by direct injection onto a Waters LCT Premier XE mass spectrometer. NMR data was collected for all samples in a 3:1 mixture of D<sub>2</sub>O:CD<sub>3</sub>CN on a Bruker 600 MHz spectrometer and was processed using the MestreNova Software Package (MestreLab Research).

(49) Zerbino, D. R.; Birney, E. *Genome Res.* **2008**, *18*, 821–9.

(50) Matsushima, P.; Broughton, M. C.; Turner, J. R.; Baltz, R. H. *Gene* **1994**, *146*, 39–45.

**MIC Determination.** Stationary phase cultures of *S. aureus* and *E. faecalis*, grown in tryptic soy broth or brain heart infusion broth, respectively, were diluted 1:10<sup>6</sup>. Individual wells in 96-well microtiter plates were then filled with 100  $\mu$ L of these dilute cultures. A stock solution of an individual compound dissolved in dimethyl sulfoxide was diluted 1:100 into culture media and 100  $\mu$ L of this solution was added to the first well of a row in the filled microtiter plate. This well was then serially diluted 1:2 across the plate. Microtiter plates were incubated at 37 °C with 100 rpm orbital shaking for 18 h. MIC values are reported as the lowest concentration at which no bacterial growth was observed.

**Acknowledgment.** This work was supported by NIH GM077516, NIH MSTP GM07739 (J.W.C.), the Howard Hughes Medical Institute, and the Beckman Foundation. We thank Dr. Alexander Tomasz for providing *S. aureus* USA300 and *E. faecalis* EF18.

**Supporting Information Available:** NMR data for selected compounds and pathway ORFs. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA105825A