Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary

Jacob J. Banik and Sean F. Brady[1]

Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, 1230 York Avenue, New York, NY 10065

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved September 23, 2008 (received for review August 1, 2008)

Glycopeptide antibiotics have long served as drugs of last resort for the treatment of antibiotic-resistant Gram-positive bacterial infections. Resistance to the clinically relevant glycopeptides, vancomycin and teicoplanin, threatens to undermine the usefulness of this important class of antibiotics. DNA extracted from a geographically diverse collection of soil samples was screened by PCR for the presence of sequences related to OxyC, an oxidative coupling enzyme found in glycopeptide biosynthetic gene clusters. Every soil sample examined contained at least 1 unique OxyC gene sequence. In an attempt to access the biosynthetic gene clusters associated with these OxyC sequences, a 10,000,000-membered environmental DNA (eDNA) megalibrary was created from a single soil sample. Two unique glycopeptide gene clusters were recovered from this eDNA megalibrary. Using the teicoplanin aglycone and the 3 sulfotransferases found in one of these gene clusters, mono-, di-, and trisulfated glycopeptide congeners were produced. The high frequency with which OxyC genes were found in environmental samples indicates that soil eDNA libraries are likely to be a rewarding source of glycopeptide gene clusters. Enzymes found in these gene clusters should be useful for generating new glycopeptides analogs. Environmental DNA megalibraries, like the one constructed for this study, can provide access to many of the natural product biosynthetic gene clusters that are predicted to be present in soil microbiomes.

antibiotics | eDNA | metagenomics | natural products | uncultured bacteria

Vancomycin and teicoplanin are clinically useful heptapeptide antibiotics that inhibit cell wall biosynthesis by noncovalently binding the D-ala-D-ala terminus of growing petidoglycans (1–3). For almost 50 years, this class of glycopeptide antibiotics has served as the last line of defense against methicillin-resistant *Staphylococcus aureus* (MRSA). With the appearance of vancomycin-resistant Enterococci in the late 1980s and resistant Staphylococci in the 1990s, there has been a renewed interest in the discovery of novel glycopeptide congeners that might reinvigorate this important class of clinically useful antibiotics (4–8). Although the screening of bacterial culture broths for the presence of glycopeptide antibiotics initially yielded a large number of novel congeners, few additional naturally occurring members of this important class of antibiotics have been reported in recent years.

A single gram of soil is predicted to contain >10,000 unique bacterial species (9–11). DNA extracted directly from soil (environmental DNA, eDNA) should contain a very diverse collection of bacterial natural product biosynthetic gene clusters. Only a small fraction of these gene clusters is likely to have been functionally accessed to date, because the majority of bacteria are not readily cultured in the laboratory, and of those that are cultured, only a fraction of their secondary metabolite biosynthetic gene clusters is typically activated in laboratory fermentations (9–13). Systematically screening large eDNA libraries for conserved sequences associated with the biosynthesis of pharmacologically relevant natural products could prove to be a rewarding strategy for the discovery of gene clusters that encode the biosynthesis of new natural analogs. Although the heterologous expression of intact eDNA-derived secondary metabolite gene clusters remains a significant challenge, enzymes found in these gene clusters will likely be useful for functionalizing many natural products in novel ways.

In a PCR-based survey of eDNA extracted from geographically diverse soil samples, every sample we examined yielded OxyC sequences that closely resemble those found in glycopeptide gene clusters. Here, we report the cloning of 2 glycopeptide biosynthetic gene clusters from a 10,000,000-membered soil eDNA megalibrary, both of which are predicted to encode the biosynthesis of highly functionalized glycopeptide congeners. By using the sulfotransferases found in 1 of the 2 gene clusters, a unique family of sulfated teicoplanin analogs was generated in vitro from the teicoplanin aglycone. The work presented here suggests that environmental DNA megalibraries are likely to be a rich source of new glycopeptide biosynthetic gene clusters, and that the enzymes found in these gene clusters should be useful for generating a variety of new glycopeptide analogs.

Results and Discussion

Our search for glycopeptide biosynthetic gene clusters began by probing a geographically diverse collection of crude eDNA extracts for the presence of OxyC-like gene sequences. OxyC catalyzes the formation of the C–C bond between the hydroxyphenylglycine at position 5 and the dihydroxyphenylglycine at position 7 in both vancomycin- and teicoplanin-like glycopeptides (Fig. 1) (14). This enzyme is highly conserved in sequenced glycopeptide gene clusters and easily distinguished from related oxidative enzymes that appear in unrelated biosynthetic gene clusters, making it an ideal probe to use in the search for unique glycopeptide gene clusters. By using a set of nested OxyC-based degenerate primers, we were able to amplify OxyC-like sequences from every eDNA sample that we examined (Fig. 1 and *SI Text*). The soils used for this study were collected in North America, Central America, and Africa (Fig. 1). This PCR-based survey strongly suggested that eDNA would be a rich source of glycopeptide gene clusters; however, it did not provide functional access to these clusters.

Although soils are thought to be the most microbially diverse ecosystems on earth and therefore represent attractive sources of

Author contributions: J.J.B. and S.F.B. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in GenBank database [accession nos. EU874252 (VEG pathway) and EU874253 (TEG pathway)].

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: sbrady@rockefeller.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0807564105/DCSupplemental.

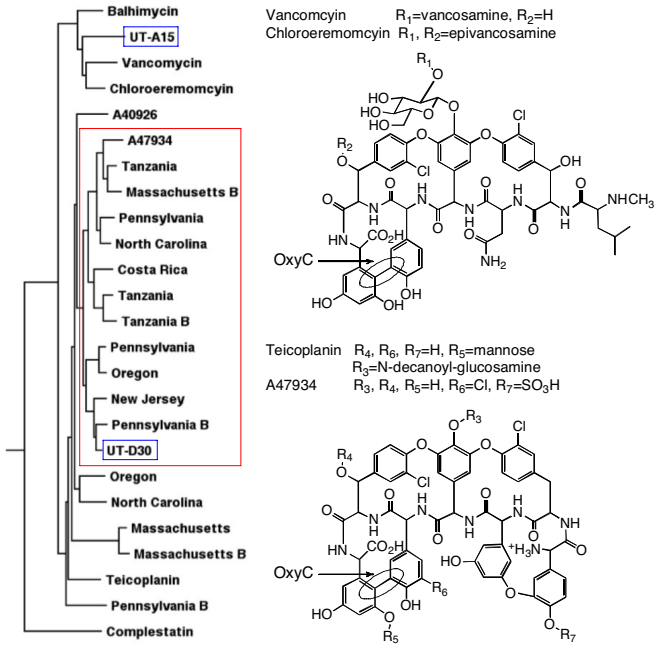


Fig. 1. The phylogenetic tree contains OxyC sequences from cultured bacteria and a subset of the eDNA OxyC sequences amplified from a geographically diverse collection of soils. Blue boxes show OxyC sequences from the Utah library and the red box highlights sequences that are closely related to those present in glycopeptide gene clusters known to contain sulfotransferase genes.

genetic diversity for the discovery of new natural product biosynthetic gene clusters, the immense size of a soil metagenome has made cloning natural product gene clusters directly from soil microbiomes very difficult. In an attempt to access glycopeptide gene clusters associated with OxyC-like sequences that could be amplified from crude eDNA, we constructed a 10,000,000-membered cosmid-based library by using DNA extracted from desert soil collected in Utah. This eDNA megalibrary contains the equivalent of 100,000 4-Mb bacterial genomes, and although even 10,000,000 clones may not capture all of the genetic information present in a soil sample, a library this size should allow for the recovery and reconstruction of many previously unstudied gene clusters present in soil metagenomes.

Aliquots of DNA from each of the ≈2,500 unique sublibraries that make up the eDNA megalibrary were used as amplification templates in PCRs with OxyC-based degenerate primers. Using these primers, 2 unique OxyC-like sequences were found in the megalibrary (Fig. 1; UT-A15 and UT-D30). One of the OxyC sequences is more closely related to those found in sequenced vancomycin-like gene clusters (vancomycin-like eDNA derived gene cluster, VEG cluster), whereas the other resembles those found in sequenced teicoplanin-like gene clusters (teicoplanin-like eDNA derived gene cluster, TEG cluster). Only teicoplanin-like sequences were found in our initial screen of crude eDNA samples, suggesting that these sequences may be preferentially amplified from OxyC mixtures, and that environmental samples are likely to contain an even more diverse collection of glycopeptide gene clusters than is suggested by Fig. 1.

Cosmid clones containing each of the Utah eDNA OxyC sequences were recovered from the appropriate sublibraries, and then subsequent rounds of library screening were carried out to identify and recover overlapping clones that could be used to reconstruct 2 independent glycopeptide biosynthetic gene clusters (Fig. 2). Cosmids containing the more common VEG pathway appeared in approximately 1 of every 2 million clones, whereas clones containing the less common TEG pathway were found in 1 of every 5–10 million clones. For the common pathway, >100 kb of continuous sequence were reconstructed from 6 overlapping cosmid clones. Almost 75 kb of this sequence are predicted to be part of a glycopeptide antibiotic biosynthetic gene cluster; >50 kb of continuous sequence from the second gene cluster were reconstructed from 3 overlapping cosmid clones. Although the TEG gene cluster is truncated at one end, a comparison of the 2 gene clusters, which are closely related at both ends but differ significantly in the central finishing enzyme regions, suggests that the missing region is likely to contain conserved biosynthetic machinery that is common to most glycopeptide gene clusters.

A detailed analysis of these 2 eDNA-derived gene clusters indicates that each is predicted to encode the biosynthesis of a previously unknown glycopeptide congener (Tables S1–S3). As expected, both clusters contain large NRPS systems that are predicted to encode heptapeptides (Fig. 2, green). The NRPS

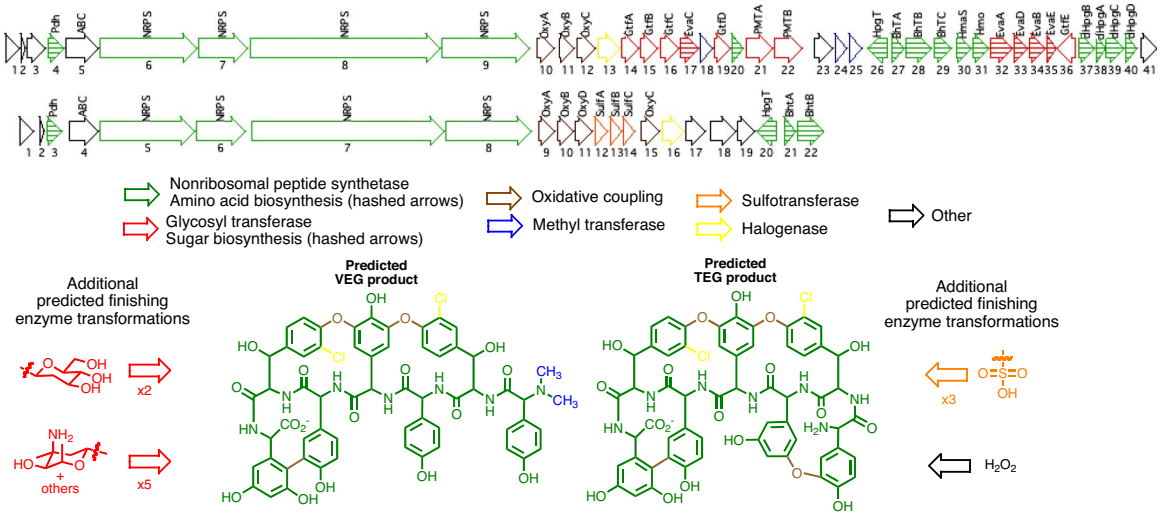


Fig. 2. The VEG and TEG gene clusters have been color coded to denote the general biosynthetic transformation of each ORF. In those cases where the predicted function of the ORF has a standard abbreviation in glycopeptide biosynthesis, the ORF has also been labeled with this abbreviation. The predicted core structure of each glycopeptide congener is shown.

system in the VEG pathway is predicted to encode the biosynthesis of a type II heptapeptide [hydroxyphenylglycine (HPG)-β-hydroxytyrosine (Bht)-Hpg-Hpg-Hpg-Bht-dihydroxyphenylglycine (Dpg)], and the TEG pathway is predicted to encode the biosynthesis of a type III heptapeptide (Hpg-Bht-Dpg-Hpg-Hpg-Bht-Dpg) (3). These predictions are supported by the presence of the 3 conserved oxidative enzymes (Fig. 2, brown, Veg10–12) that are required for the biosynthesis of an oxidatively cross-linked type II glycopeptide in the first gene cluster and the 4 conserved oxidative enzymes (Fig. 2, brown, Teg9, 10, 11, and 15) that are required for the biosynthesis of an oxidatively cross-linked type III glycopeptide in the second gene cluster.

The VEG cluster contains a complete complement of the genes predicted to be necessary for the biosynthesis of an oxidatively cross-linked heptapeptide. This includes all of the biosynthetic machinery required for the production of the 3 nonproteinogenic amino acids (Hpg, Bht, and Dpg) found in the peptide core (Fig. 2, hashed green), a halogenase for the production of chloro-Bht (Fig. 2, yellow), and, as mentioned, the 3 oxidative enzymes (Fig. 2, brown) that oxidatively cross-link the heptapeptide core (1). In addition to the core biosynthetic machinery, the VEG cluster contains 7 glycosyltransferases (Fig. 2, red) and 3 methyltransferases (Fig. 2, blue). Two of the 3 methyltransferases (Veg24, 25) are predicted *N*-methyltransferases suggesting that the product of the VEG pathway is doubly *N*-methylated at the N terminus. The third methyltransferase (Veg18) is related to putative rRNA-methyltransferases. rRNA-methyltransferases are commonly used for self-resistance against antibiotics that target the ribosome (15). The role this enzyme would play in a glycopeptide gene cluster is not immediately obvious.

Four of the glycosyltransferases in the VEG gene cluster (Veg14, 15, 16, and 19) are related (75–88% identity) to glycosyltransferases from chloroeremomycin biosynthesis, two (Veg21, 22) are related to putative mannosyltransferases (PMT, 80–82% identity) from teicoplanin-like biosynthetic gene clusters, and the seventh (Veg36) is related to the glycosyltransferase CalG1 (47% identity) from calicheamycin biosynthesis. Two of the related glycosyltransferases from chloroeremomycin biosynthesis (GtfA and C) use derivatives of vancosamine as substrates. Homologs of the 5 enzymes (EvaA–E) required for the biosynthesis of a vancosamine derivative are also found in the VEG gene cluster (Fig. 2, hashed red) (16, 17). To the best of our knowledge, no glycopeptide congener containing the unique assortment of functionality encoded by the VEG gene cluster (doubly *N*-methylated, vancosamine functionalized, type II glycopeptide) has been characterized from cultured bacteria.

The VEG gene cluster is rich in methyl- and glycosyltransferases, whereas the TEG gene cluster is rich in sulfotransferases (Fig. 2, orange). In addition to containing biosynthetic machinery that is required for the production of an oxidatively cross-linked heptapeptide core, the TEG gene cluster contains 3 predicted sulfotransferases (Teg12, 13, and 14), a putative dyp-type peroxidase (Teg17), and 2 hypothetical proteins of unknown function (Teg18, 19). Homologs of the peroxidase and the 2 hypothetical proteins of unknown function do not exist in any sequenced glycopeptide gene cluster from cultured bacteria. They may therefore provide a unique functionality to this family of antibiotics. Although a large number of glycosylated, halogenated, and alkylated glycopeptide congeners have been identified from studying cultured bacteria, to the best of our knowledge only 2 sulfated, and no polysulfated, glycopeptides have been reported (3, 18). The presence of 3 sulfotransferases in the TEG cluster suggested that it was likely to encode the biosynthesis of the first polysulfated glycopeptides. Polysulfated glycopeptides would have very different surface-charge distributions than any reported glycopeptide congeners. We therefore sought to generate such compounds by using the unique genetic information captured in the TEG gene cluster.

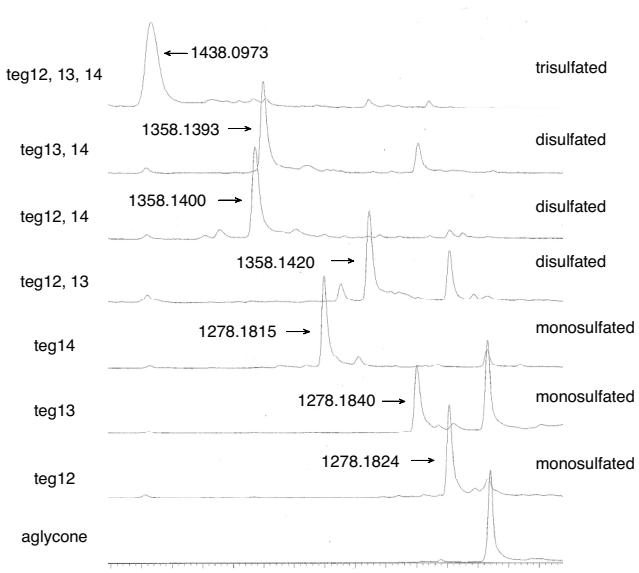
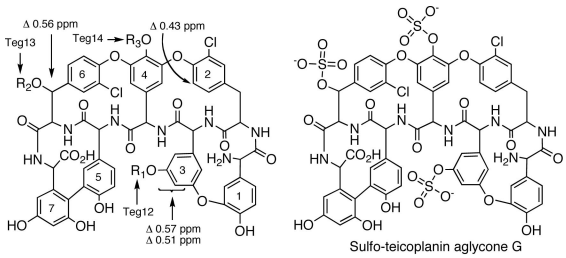


Fig. 3. HPLC traces and observed ESI-HRMS *m/z* data for the compounds that are produced by all possible combinations of the TEG sulfotransferases (*m/z* $[M]^+$ calcd for (mono-) $C_{58}H_{46}Cl_2N_7O_{21}S$, 1278.1839, $[M]^+$ calcd for (di-) $C_{58}H_{46}Cl_2N_7O_{24}S_2$, 1358.1407, $[M]^+$ calcd for (tri-) $C_{58}H_{46}Cl_2N_7O_{27}S_3$, 1438.0975).

It is possible to envision using either in vivo heterologous expression studies or in vitro reconstitution studies to access natural products from cryptic gene clusters. The TEG gene cluster encodes the biosynthesis of a putative heptapeptide core that only differs from the teicoplanin aglycone by the substitution of a β-hydroxytyrosine for the tyrosine at position 2. The teicoplanin aglycone is readily available and therefore we elected to use this closely related substrate in in vitro studies with the TEG sulfotransferases. The 3 sulfotransferases (Teg12, 13, and 14) in the TEG cluster are homologs of StaL, a 3′-phosphoadenosine 5′-phosphosulfate (PAPS)-dependent sulfotransferase found in the A47934 glycopeptide gene cluster (19). StaL has been used to generate monosulfated glycopeptides in vitro (20). For our in vitro glycopeptide sulfation studies Teg12, 13, and 14 were PCR amplified and cloned into the pET28a expression vector. Each recombinant 6-His-tagged sulfotransferase was then Ni-affinity-purified from cultures of *Escherichia coli* BL21(DE3) (Fig. S1). In the presence of PAPS and the teicoplanin aglycone each predicted sulfotransferase produces a unique monosulfated glycopeptide derivative (Fig. 3). In reactions with 2 sulfotransferases, the 3 possible disulfated derivatives are formed, and in a reaction with all 3 sulfotransferases a trisubstituted derivative is produced (Fig. 3).

Mass spectrometry and 1D and 2D NMR were used to identify the sulfation site in each monosulfated product. The sulfation patterns seen in the di- and trisulfated teicoplanin aglycone analogs were then inferred from the sulfation specificities of the sulfotransferases used to synthesize these derivatives. On fragmentation by negative ion ESI-MS/MS, each of the monosulfated sulfated aglycone derivatives produces a daughter ion with an $m/z = 906$ (Fig. S2). This fragment, which is not produced by the teicoplainin aglycone, corresponds to a sulfated product that has lost the macrocycle formed by amino acids 5 and 7. The presence of this sulfated fragment eliminates 3 of the 7 hydroxyls on the teicoplainin aglycone as candidates for sulfation. To determine which of the 4 remaining hydroxyls are sulfated,



	R_1	R_2	R_3	MIC USA300	MIC JH2
Sulfo-teicoplanin aglycone A	SO_3H	H	H	2.5	5.0
Sulfo-teicoplanin aglycone B	H	SO_3H	H	1.0	2.0
Sulfo-teicoplanin aglycone C	H	H	SO_3H	3.5	3.5
Sulfo-teicoplanin aglycone D	SO_3H	SO_3H	H	1.3	2.5
Sulfo-teicoplanin aglycone E	SO_3H	H	SO_3H	5.0	10
Sulfo-teicoplanin aglycone F	H	SO_3H	SO_3H	>4.4	>4.4
Sulfo-teicoplanin aglycone G	SO_3H	SO_3H	SO_3H	5.0	10
Teicoplanin				1.3	5
Vancomycin				1.3	2.5

Fig. 4. The sulfo-teicoplanin aglycones are produced by sulfation of the teicoplanin aglycone with combinations of Teg12, 13, and 14. Minimum inhibitory concentrations (MIC, μg/mL) were determined against *S. aureus* USA 300 (methicillin-resistant) and JH2 (vancomycin-intermediate-resistant).

HPLC-purified sulfo-teicoplanin aglycones, derived from milligram-scale sulfation reactions, were analyzed by 1D and 2D NMR, and these spectra were compared with those obtained with the aglycone (21, 22) (Figs. S3 and S4). Compared with their hydroxylated counterparts, sulfated compounds show significant deshielding (≈0.5 ppm) of aromatic protons ortho to the site of sulfation and deshielding of aliphatic protons that are attached to a sulfated carbon (Tables S4 and S5) (18, 23, 24). Sulfation of the hydroxyls on the amino acids at positions 3, 6, and 4 by Teg12, 13, and 14, respectively, could be inferred from the sulfate-induced deshielding of protons present in the ^{1}H NMR spectra of sulfo-teicoplanin aglycones A–C (Fig. 4). In sulfo-teicoplanin aglycone A (the Teg12 product), the 2 protons ortho to the hydroxyl on the resorcinol side chain of amino acid 3 are deshielded by 0.51 and 0.57 ppm relative to their chemical shifts in the aglycone. In sulfo-teicoplanin aglycone B (the Teg13 product), the β-carbon proton of amino acid 6 is deshielded by 0.59 ppm relative to the aglycone. In sulfo-teicoplanin aglycone C (the Teg14 product), no protons ortho to a potential sulfation site are significantly deshielded compared with the aglycone. The hydroxyphenylglycine at position 4 is the only potential sulfation site that is not predicted to undergo significant proton chemical shifts changes on sulfation. The cup-shaped conformation of glycopeptides places a sulfate at position 4 in close proximity to the protons on amino acid 2, which could explain the 0.43-ppm deshielding of the proton meta to the chlorine on amino acid 2 in sulfo-teicoplanin aglycone C. To the best of our knowledge, all 3 sites sulfated by the TEG sulfotransferases differ from those sulfated in known glycopeptide congeners.

In antibacterial assays run against methicillin-resistant (USA300) and vancomycin-intermediate-resistant (VISA) (JH2) strains of *Staphylococcus aureus*, the sulfated teicoplanin aglycone analogs show similar activity to vancomycin and teicoplanin (25, 26). The general trend observed within this family of metabolites is that each successive sulfate addition increases the MIC slightly. Although sulfates have rarely been seen in naturally occurring glycopeptide antibiotics, a related negatively charged substituent, a phosphonic acid, appears on the dihydroxyphenylglycine at position 7 in Telavancin, a second-generation semisynthetic glycopeptide antibiotic (6). Whether the negatively charged sulfates play a role beyond increasing glycopeptide solubility remains to be seen.

Conclusions

Both of the glycopeptide gene clusters recovered from the 10,000,000-membered eDNA megalibrary constructed for this study are predicted to encode the biosynthesis of previously unknown glycopeptide congeners. The sulfotransferases found in the TEG gene cluster introduce multiple examples of a rarely seen functionality onto a glycopeptide skeleton. Using degenerate primers based on OxyC sequences found in known glycopeptide gene clusters, we were able to amplify OxyC-like sequences from every eDNA sample that we examined. OxyC sequences related to those found in gene clusters that encode the biosynthesis of sulfated glycopeptides were found in every eDNA sample we examined (Fig. 1, red box). Although only 2 naturally occurring sulfated glycopeptides have been characterized to date, sulfated glycopeptide congeners may actually be very common in nature. Environmental DNA megalibraries, like the one constructed for this study, should be generally useful resources for cloning natural product biosynthetic gene clusters from soil metagenomes. Although functionally accessing the molecules encoded by these gene clusters remains a significant challenge, this work shows that the finishing enzymes found in these clusters should be useful for functionalizing known natural products in new ways.

Methods

PCR Screening of eDNA for OxyC Sequences. Crude eDNA samples were prepared from soil samples collected in Pennsylvania (samples A and B), New Jersey, Massachusetts (samples A and B), Utah, Oregon, North Carolina, Tanzania (samples A and B), and Costa Rica by using standard eDNA isolation methodology (27). In brief, a 1:1 mixture (wt/vol) of soil and lysis buffer (100 mM Tris·HCl, 100 mM Na EDTA, 1.5 M NaCl, 1% (wt/vol) CTAB, 2% (wt/vol) SDS, pH 8.0) was heated for 2 h at 70 °C, and then, the soil was removed by centrifugation (4,000 × *g*, 30 min). Crude environmental DNA was isopropyl alcohol precipitated from the supernatant with the addition of 0.6 volumes of isopropyl alcohol, collected by centrifugation (4,000 × *g*, 30 min), washed with 70% ethanol, and then resuspended in TE (10 mM Tris, 1 mM EDTA, pH 8.0). The remaining dirt present in the crude extract was removed by large-scale gel purification on a 1% agarose gel (16 h at 20 V), and purified high-molecular-weight eDNA was electroeluted from the gel. PCR-ready eDNA was prepared from the gel-purified eDNA by using the QIAamp DNA Stool Mini Kit (Qiagen). OxyC sequences were amplified from these eDNA samples by using the following nested PCR primers: 1st round forward: ATGCTSACSCCSGAGTTCACSGTVCGG, 1st round reverse: GCAGTRRTGGAYGCCGTGCCCGAA, 2nd round forward CTGTGYGARCTGCTCGGSRTCC, 2nd round reverse CGACRCCRCCSAGGAKCAGC. eDNA (100 ng) was used as a template in all first-round amplification reactions (cycling parameters: 30 rounds of PCR, denatured 95 °C for 30 s, 68 °C for 30 s, 72 °C for 90 s, 25-μL reaction conditions: 2.5 μL of Thermo Pol Buffer (New England Biolabs), 1.25 μL of DMSO, 0.625 μL of each 100 mM oligonucleotide primer, 0.5 μL of 10 mM dNTPs mix, 1 unit of TaqDNA Polymerase, and water as needed). In second-round amplification reactions 1 μL of the first-round PCR was used as a template (cycling parameters: 30 rounds of 95 °C for 30 s, 65 °C for 30 s, and 72 °C for 30 s; reaction conditions were identical to those used in the first-round amplification reactions). Amplicons of the correct predicted size (245 bp) were gel-purified and topocloned into TOPO 2.1 (Invitrogen). Eight to 10 unique clones obtained from each soil sample were sequenced.

eDNA Isolation, Library Construction, and Screening. High-molecular-weight DNA electroeluted from the large-scale gel purification of the crude eDNA extract derived from soil collected in Utah was used to prepare 10,000,000 cosmid clones by using standard cosmid-cloning methods. Environmental DNA was blunt ended (Epicentre, End-It), ligated into the SmaI site of pWEB, packaged into lambda phage, and transfected into *E. coli* EC100. The megalibrary was prepared as unique 4,000- to 15,000-membered sublibraries that were arrayed into 8 × 8 grids of matching minipreps and glycerol stocks. Miniprep samples were used for PCR-based screening, and the matching glycerol stocks were used for clone recovery. Aliquots from each row and column of the 8 × 8 miniprep grids were pooled to facilitate PCR screening.

Library Screening and Clone Recovery Protocol. To identify OxyC sequences present in the arrayed eDNA library, each set of pooled rows and columns was screened by PCR with the first-round OxyC primers described above (cycling parameters: 7 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 standard cycles of 95 °C for 30 s, 62 °C for 30 s, 72 °C for 90 s). Amplicons of the correct predicted size (739 bp) were

gel-purified and sequenced to identify the specific library pools containing OxyC sequences. Individual clones of interest were recovered from the appropriate sublibraries by successive rounds of library dilution and PCR screening. In brief, the glycerol stock corresponding to a miniprep known to contain an OxyC sequence was diluted to an OD_{600} ≈0.2 and then 12.5 μL of this solution (≈100–200 clones) were plated into 96-well microtiter plates prefilled with 100 μL of LB-agar containing 30 μg/mL kanamycin. After incubation overnight at 37 °C the cells were resuspended in 75 μL of LB and then each resuspended well was screened by whole-cell PCR for the presence of a clone containing the desired sequence. This screening process was then repeated with the positive pools plated at successively lower titers (OD_{600} ≈0.02 and ≈0.002). In the final round of screening, individual colonies were assayed by PCR. Each recovered clone was end sequenced, and the primers designed to this sequence were used in subsequent rounds of screening to recover overlapping cosmid clones from the library. Six overlapping clones (cosmid clones: B128, A15, A6, Y69, X150, and W186) containing portions of the VEG pathway were recovered from the library. Three overlapping clones (cosmid clones: D30, J44, and B95) that contain the TEG pathway were recovered from the library. All cosmid clones were sequenced by pyrosequencing (454 Life Sciences).

Cloning, Expression, and Purification of Teg12, 13, and 14 Sulfotransferase Genes. Teg12 (Sulf 1), Teg13 (Sulf 2), and Teg14 (Sulf 3) were amplified (30 cycles of 95 °C for 30s, 60 °C for 30s, and 72 °C for 90s; FailSafe system from Epicentre) from clone D30 by using the following primers: Sulf1For(BclI): **GCGCTGATCA**ATGAACGGAATTCGATGG, Sulf1rev(HindIII): **GCGCAAGCTT**TCCTTAACCGGCATACCCGTA, Sulf2For(BclI): **GCGCTGATCA**ATGAACGGCATTCGATGGATC, Sulf2rev(HindIII): **GCGCAAGCTT**ATCTCTCCTCCCTCAGCCGGC, Sulf3For (BclI): **GCGCTGATCA**ATGAACGGTATTCGATGGATC, Sulf3Rev (HindIII): **GCGCAAGCTT**ACAATCCGCCCGTTAGCCGGC. The restriction sites added for cloning purposes are shown in bold. Amplicons were doubly digested with BclI/HindIII, ligated into the BamHI/HindIII-digested pET28a, and then transformed into *E. coli* BL21 (DE3).

Fusion protein purification and sulfation reaction protocols were modified from published procedures used for StaL (20). Overnight growths were used to inoculate (1:1,000 dilution) 1-L cultures of LB, which were grown at 37 °C until the OD_{600} reached 0.6. The temperature was then reduced to 20 °C, and after 1 h, the cultures were induced with IPTG (0.5 mM). After 14–16 h at 20 °C, the cultures were harvested by centrifugation (3,200 × *g* for 30 min). The cell pellet was resuspended in 40 mL of lysis buffer [50 mM Hepes, pH 7.5, 0.5 M NaCl, 5% (vol/vol) glycerol, 20 mM imidazole, pH 8, 10 mM β-mercaptoethanol, and 0.5% (vol/vol) Triton X-100], and the cells were lysed by sonication. Crude cell lysates were centrifuged at 25,000 × *g* for 30 min, and the supernatants were then incubated for 15 min (24 °C) with 1 mL of Ni-NTA resin. After 15 min, this 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% (vol/vol) glycerol, 20 mM imidazole, pH 8, and 10 mM β-mercaptoethanol], and then the protein was eluted from the resin with 15 mL of elution buffer [50 mM Hepes, pH 7.5, 0.5 M NaCl, 5% (vol/vol) glycerol, 125 mM imidazole, pH 8, and 10 mM β-mercaptoethanol]. The protein concentration was determined with the Bradford assay (yields: 74 mg/L TEG12, 31 mg/L TEG13, and 88 mg/L TEG14) and these samples were used in sulfation reactions without further purification. Two-milliliter sulfation reactions containing 1 mg of each sulfotransferase, 1 mg of teicoplanin aglycone, and 0.5 mg of 3′-phosphoadenosine 5′-phosphosulfate (PAPS) were set up in reaction buffer (250 mM Hepes, pH 7.5, 0.1 mM DTT) and incubated overnight at 30 °C. Excess sulfotransferase was added to overcome the known inhibition of PAPS-dependent sulfotransferases by the sulfotransferase by-product 3′-phosphoadenosine 5′ phosphate (PAP) (28). The reactions were then placed in a boiling water bath for 10 min, placed on ice for 10 min, and centrifuged (21,000 × *g* at 4 °C) for 10 min. The supernatant (and, if necessary, a water wash of the pellet) was evaporated to dryness and resuspended in 400 μL of a 50:50 DMSO/H_2O. The analytical HPLC conditions used to generate Fig. 2 were as follows: C18 (4.6 × 150 mm), 20 mM ammonium acetate/acetonitrile, 95:5 to 70:30 over 20 min, 1.5 mL/min. Preparative HPLC was carried out by using the same gradient conditions on a 10 × 150 mm column at a flow rate of 7 mL/min.

Supporting Information. For further information regarding OxyC amplicons, please see supporting information (SI) *Text*, Figs. S1–S4, and Tables S1–S5.

ACKNOWLEDGMENTS. We thank The Instituto National de Bioversidad and Serengeti Genesis for their assistance with collecting environmental samples. This work was supported by National Institutes of Health GM077516, The Beckman Foundation, and The Searle Foundation. NMR data were collected at the New York Structural Biology Center, which is supported in part by National Institutes of Health Grant GM66354.

1. Hubbard BK, Walsh CT (2003) Vancomycin assembly: Nature's way. *Angew Chem Int Ed* 42(7):730–765.
2. Kahne D, Leimkuhler C, Lu W, Walsh C (2005) Glycopeptide and lipoglycopeptide antibiotics. *Chem Rev* 105(2):425–448.
3. Nicolaou KC, Boddy CNC, Brase S, Winssinger N (1999) Chemistry, biology, and medicine of the glycopeptide antibiotics. *Angew Chem Int Ed* 38(15):2096–2152.
4. Uttley AH, Collins CH, Naidoo J, George RC (1988) Vancomycin-resistant enterococci. *Lancet* 1(8575-6):57–58.
5. Hiramatsu K, *et al.* (1997) Dissemination in Japanese hospitals of strains of Staphylococcus aureus heterogeneously resistant to vancomycin. *Lancet* 350(9092):1670–1673.
6. Higgins DL, *et al.* (2005) Telavancin, a multifunctional lipoglycopeptide, disrupts both cell wall synthesis and cell membrane integrity in methicillin-resistant Staphylococcus aureus. *Antimicrob Agents Chemother* 49(3):1127–1134.
7. Ge M, *et al.* (1999) Vancomycin derivatives that inhibit peptidoglycan biosynthesis without binding D-Ala-D-Ala. *Science* 284(5413):507–511.
8. Wolter F, Schoof S, Süssmuth R (2007) Synopsis of structural, biosynthetic, and chemical aspects of glycopeptide antibiotics. *Glycopeptides and Glycoproteins* (Springer, Berlin), pp 143–185.
9. Torsvik V, Goksoyr J, Daae FL (1990) High diversity in DNA of soil bacteria. *Appl Environ Microbiol* 56(3):782–787.
10. Curtis TP, Sloan WT, Scannell JW (2002) Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci USA* 99:10494–10499.
11. Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. *Annu Rev Microbiol* 57:369–394.
12. Bentley SD, *et al.* (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* 417(6885):141–147.
13. Ikeda H, *et al.* (2003) Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nat Biotechnol* 21(5):526–531.
14. Pylypenko O, Vitali F, Zerbe K, Robinson JA, Schlichting I (2003) Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C-C coupling reaction during vancomycin biosynthesis. *J Biol Chem* 278(47):46727–46733.
15. Mosbacher TG, Bechthold A, Schulz GE (2003) Crystal structure of the avilamycin resistance-conferring methyltransferase AviRa from Streptomyces viridochromogenes. *J Mol Biol* 329(1):147–157.
16. Lu W, *et al.* (2004) Characterization of a regiospecific epivancosaminyl transferase GtfA and enzymatic reconstitution of the antibiotic chloroeremomycin. *Proc Natl Acad Sci USA* 101:4390–4395.
17. Chen H, *et al.* (2000) Deoxysugars in glycopeptide antibiotics: Enzymatic synthesis of TDP-L-epivancosamine in chloroeremomycin biosynthesis. *Proc Natl Acad Sci USA* 97:11942–11947.
18. Boeck LD, Mertz FP (1986) A47934, a novel glycopeptide-aglycone antibiotic produced by a strain of Streptomyces toyocaensis taxonomy and fermentation studies. *J Antibiot* 39(11):1533–1540.
19. Shi R, *et al.* (2007) Crystal structure of StaL, a glycopeptide antibiotic sulfotransferase from Streptomyces toyocaensis. *J Biol Chem* 282(17):13073–13086.
20. Lamb SS, Patel T, Koteva KP, Wright GD (2006) Biosynthesis of sulfated glycopeptide antibiotics by using the sulfotransferase StaL. *Chem Biol* 13(2):171–181.
21. Boger DL, *et al.* (2000) Thermal atropisomerism of teicoplanin aglycon derivatives: Preparation of the P,P,P and M,P,P atropisomers of the teicoplanin aglycon via selective equilibration of the DE ring system. *J Am Chem Soc* 122(41):10047–10055.
22. Malabarba A, Ferrari P, Gallo GG, Kettenring J, Cavalleri B (1986) Teicoplanin, antibiotics from Actinoplanes teichomyceticus nov. sp. VII. Preparation and NMR characteristics of the aglycone of teicoplanin. *J Antibiot* 39(10):1430–1442.
23. Yates EA, *et al.* (2000) Effect of substitution pattern on 1H, 13C NMR chemical shifts and 1J(CH) coupling constants in heparin derivatives. *Carbohydr Res* 329(1):239–247.
24. Pretsch E, Bühlmann P, Affolter C (2000) *Structure Determination of Organic Compounds : Tables of Spectral Data* (Springer, Berlin), 3rd completely rev. and enl. English ed, pp xv and 421.
25. Diep BA, *et al.* (2006) Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant Staphylococcus aureus. *Lancet* 367(9512):731–739.
26. Sieradzki K, Leski T, Dick J, Borio L, Tomasz A (2003) Evolution of a vancomycin-intermediate Staphylococcus aureus strain in vivo: Multiple changes in the antibiotic resistance phenotypes of a single lineage of methicillin-resistant S. aureus under the impact of antibiotics administered for chemotherapy. *J Clin Microbiol* 41(4):1687–1693.
27. Brady SF (2007) Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat Protocols* 2:1297–1305.
28. Chapman E, Best MD, Hanson SR, Wong CH (2004) Sulfotransferases: Structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew Chem Int Ed* 43(27):3526–3548.