J|A|C|S
COMMUNICATIONS

Published on Web 08/03/2002

New Natural Product Families from an Environmental DNA (eDNA) Gene Cluster

Sean F. Brady, Carol J. Chao, and Jon Clardy*

Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301

Received May 14, 2002

Over 50 years ago Selman Waksman isolated the small-molecule antibiotics actinomycin and streptomycin from cultured soil bacteria found in a New Jersey swamp, and ever since cultured soil microbes have played a central role in drug discovery. Many lines of evidence, from light microscopy to nucleic acid-based analyses, indicate that existing culture techniques uncover only a minuscule (less than 1%) and statistically unrepresentative fraction of soil microbes.[1] The vast majority of uncultured soil microbes constitute a potentially rich source of biologically active small molecules that are inaccessible using current isolation paradigms. We have been exploring a method to access the natural products of uncultured microbes that involves isolating DNA directly from soil (environmental DNA, eDNA), making eDNA cosmid libraries in *E. coli*, and screening the cosmid libraries for clones with the ability to biosynthesize biologically active natural products.[2,3] In an earlier publication we described the general approach and the isolation of long-chain *N*-acyltyrosine antibiotics that are derived from a single eDNA open reading frame (ORF).[3] We now describe a biosynthetic gene cluster from an antibiotic-producing clone (CSLC-2) that produces two additional families of natural products derived from long-chain *N*-acyltyrosines.

Clones that produce long-chain *N*-acyl aromatic amino acids have been found in every eDNA library that we have screened. To explore this common family of new compounds we thoroughly investigated a number of these clones. During the characterization of CSLC-2, an eDNA clone that produces long-chain *N*-acyltyrosine antibiotics, it became clear that extracts from this clone contained two additional families of clone-specific small molecules. Low-resolution FABMS analysis of ethyl acetate extracts from CSLC-2 showed major peaks at $m/z = 336$, 364, 392, 418, and 420, identical to those observed in extracts from eDNA clones known to produce long-chain *N*-acyltyrosines.[3] This family of long-chain *N*-acyltyrosines (Figure 1, family **1**) includes saturated and monounsaturated fatty acids ranging from C_{10} to C_{18}. DNA sequencing of transposon-induced antibacterial knockouts suggested that a single ORF was responsible for the production of long-chain *N*-acyltyrosines. The predicted long-chain *N*-acyl amino acid synthase (NAS) from CSLC-2 is 23% identical to that previously characterized from CSL-12, an eDNA clone also shown to produce long-chain *N*-acyltyrosines.[2]

The two additional families of clone-specific small molecules not previously seen in extracts from other long-chain *N*-acyl amino acid-producing clones were characterized by HRFABMS and NMR, and a representative member from each family of predicted structures was confirmed by total synthesis. HRFABMS indicated that one family (**2**) of compounds differed from the long-chain *N*-acyltyrosines by 46 mass units (H-COOH), suggesting that they were oxidatively decarboxylated long-chain *N*-acyltyrosines. The

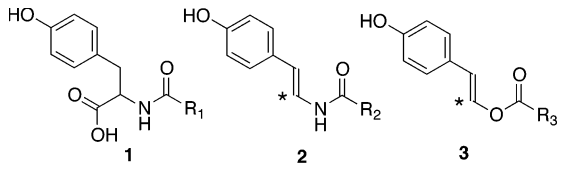

Figure 1. Three families of long-chain acyl phenols (families **1**, **2**, and **3**) produced by CSLC-2 are functionalized with both fully saturated and monounsaturated long-chain fatty acids, C_{10} to C_{18}.

large methylene envelope in the ^{1}H spectrum confirmed the presence of the long-chain fatty acids, while the para-substituted phenol of the tyrosine was seen as two highly deshielded proton doublets each integrating for two protons. A pair of highly deshielded doublets, identified as a trans ($J = 14.5$ Hz) double bond in the ^{1}H spectrum and ^{1}H−^{1}H RelayH experiments, replaced the H_α and H_β typically seen in the ^{1}H spectrum of long-chain *N*-aryltyrosines. Long-range ^{1}H−^{13}C HMBC correlations from this double bond to the aromatic ring confirmed the presence of the oxidatively decarboxylated tyrosine headgroup. Each of the fatty acids connected to the decarboxylated tyrosine was ultimately deduced from HRFABMS data. Members of the second family of clone-specific compounds (**2**) are therefore oxidatively decarboxylated tyrosines N-acylated with both saturated and unsaturated fatty acids ranging from C_{12} to C_{16}. The structure of the N-palmitoylated eneamide was independently confirmed by comparison to a synthetic sample.

HRFABMS data from the third, even more hydrophobic, family of metabolites indicated that they differed from the family of eneamides (**2**) by one mass unit, and the molecular formula predicted by HRFABMS suggested that this difference resulted from the substitution of an oxygen for the amide NH. The only significant differences observed in the ^{1}H and ^{13}C NMR spectra between these two families of compounds were the chemical shifts of C* (Figure 1). Deshielding of this carbon (+14.3 ppm) and proton (+0.5 ppm) supported the presence of an ester. The proposed structure, a long-chain fatty acid enol ester, was confirmed by comparison of the major metabolite in this family, the palmitoyl derivative, to a synthetic sample.[4] Each of the fatty acids conjugated onto the enol ester headgroup was ultimately determined on the basis of the molecular formula predicted from HRFABMS experiments. As with the tyrosine (**1**) and the oxidatively decarboxylated tyrosine (**2**) families, a range of saturated and monounsaturated fatty acids from C_{14} to C_{16} are attached to the enol ester headgroup (**3**).

The DNA surrounding the long-chain *N*-acyltyrosine synthase from CSLC-2 was sequenced and found to be part of a 13 ORF secondary metabolite gene cluster.[5] The 13 ORFs, *feeA*−*M* (*f*atty acid *e*nol *e*ster) are organized into a right- and a left-hand operon (Figure 2). The entire gene cluster was saturated with transposons, and ethyl acetate extracts from clones containing transposon

* To whom correspondence should be addressed. E-mail: jcc12@cornell.edu.

10.1021/ja0268985 CCC: $22.00

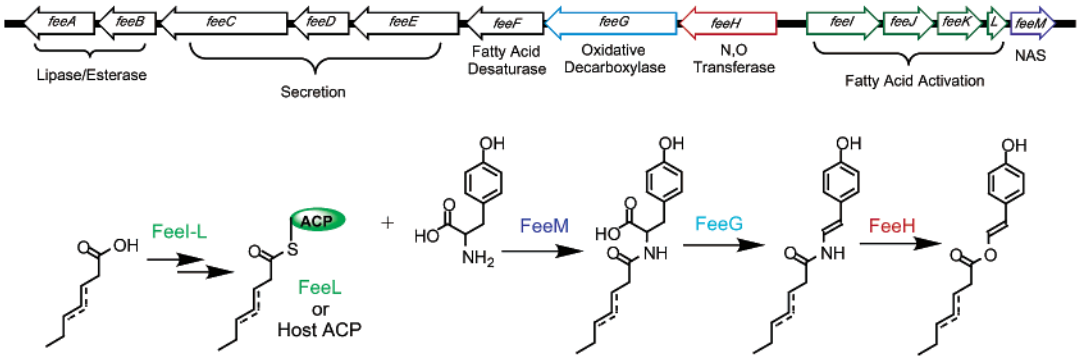

Figure 2. The *fee* gene cluster (*f*atty acid *e*nol *e*ster) cloned from eDNA. A biosynthetic scheme inferred from structural studies, transposon mutagenesis experiments, and sequencing results is shown below the gene cluster.

insertions in each ORF were analyzed by HPLC. Transposon insertions in *feeM* knock out the production of all three families of compounds. Insertions in *feeG* knock out the production of both the eneamides (**2**) and enol esters (**3**), while insertions in *feeH* exclusively inhibit the production of the enol esters (**3**). A proposed biosynthetic scheme, which has been inferred from the transposon mutagenesis results, is given in Figure 2. FeeM, -G, and -H are presumptively identified as a long chain *N*-acyl amino acid synthase, decarboxylase, and *N*,*O*-acyltransferase, respectively.

In a BLAST search against sequences deposited in GenBank, the predicted decarboxylase (FeeG) shows significant similarity (23−26%) to a family of flavoprotein oxidases.[6] All of these flavoprotein oxidases are believed to involve a quinone methide intermediate, and this putative intermediate suggests a novel tyrosine decarboxylation mechanism where the *p*-quinone methide eliminates CO_2 to produce a decarboxylated and rearomatized long-chain *N*-acyltyrosine.

The predicted *N*,*O*-acyltransferase, FeeH, shows limited sequence similarity to PseA and genes of unknown function from lipopolysaccharide gene clusters that make *N*-acyl and *N*-acetamidino sugars.[7] Only PseA has been analyzed functionally, and it is thought to play a role in the biosynthesis of an *N*-acetamidino sugar.[7] This group of enzymes may therefore be part of a yet unrecognized family of biosynthetic enzymes that modify N-acetylated natural products.

The role of each of the remaining ORFs in the *fee* gene cluster was inferred from BLAST searches. FeeJ, -K, -L, and -M show sequence similarity to acyl acyl carrier proteins (AACP), phosphopantethienyl transferases (PPT), acyl carrier proteins (ACP), and *N*-acyl amino acid synthases (NAS), respectively. A plausible biosynthetic scheme based on the biosynthesis of long chain *N*-acylhomoserine lactones produced by cultured bacteria can be formulated.[8] The NAS (FeeM) would accept fatty acids from an ACP (FeeL) for the assembly of long-chain *N*-acyltyrosines from tyrosine. The PPT (FeeK), which posttranslationally modifies ACPs (FeeL) to their catalytically active holo-forms and the AACP (FeeJ), which catalyzes the ligation of activated fatty acids onto an ACP, would be required to create the acyl-ACP used in the process. FeeI is related to proteins that adenylate carboxylic acids and may therefore be used to activate the free fatty acids used in this process.

Transposon insertions in *feeL*, the predicted ACP, result in a significant decrease in the quantity of natural products produced by CSLC-2 but do not completely abrogate their production. The ACP from de novo fatty acid biosynthesis in *E. coli* may therefore also function as a source of fatty acids during log phase growth. However if these compounds are needed during stationary phase, when de novo fatty acid biosynthesis is almost completely shut down, a system for providing ACP-linked fatty acids (FeeI-L) would be necessary. FeeA, -B, and -C, which show sequence similarity to lipases, esterases, and fatty acid desaturases, respectively, may provide the pool of saturated and unsaturated fatty acids that is loaded onto the ACP (FeeL). The remaining enzymes in the gene cluster, FeeC, -D, and -E, show sequence similarity to ABC transporter proteins (FeeD and -E) and membrane fusion proteins (FeeC) and may therefore be used to secrete these compounds from the cell.

The eDNA-based approach provides a direct link between bacterial metabolites and the biosynthetic machinery that produces them. These results confirm that culture-independent methods of examining the metabolites of soil bacteria will reveal both new molecular families and biosynthetic activities.

Acknowledgment. This work was supported by a grant from the Ellison Medical Foundation and CA24487.

Supporting Information Available: Isolation procedures and HRFABMS for families **2** and **3**, NMR data for a representative compound from families **2** and **3**, and synthetic methods for a representative compound from family **2** (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Torsvik, V.; Sorheim, R.; Goksoyr, J. *J. Ind. Microbiol.* **1996**, *17*, 170. Torsvik, V.; Goksoyr, J.; Daae, F. L. *Appl. Environ. Microbiol.* **1990**, *56*, 782. Hugenholtz, P.; Goebel, B. M.; Pace, N. R. *J. Bacteriol.* **1998**, *180*, 4765. Stackebrandt, E.; Liesack, W.; Goebel, B. M. *FASEB J.* **1993**, *7*, 232. Ward, D. M.; Weller, R.; Bateson, M. M. *Nature* **1990**, *345*, 63.
(2) Brady, S. F.; Chao, C. J.; Handelsman, J.; Clardy, J. *Org. Lett.* **2001**, *3*, 1981.
(3) Brady, S. F.; Clardy, J. *J. Am. Chem. Soc.* **2000**, *122*, 12903.
(4) Brady, S. F.; Clardy, J. In preparation.
(5) The *fee* gene cluster has been deposited with GenBank under accession No. AY128669.
(6) Fraaije, M. W.; van Berkel, W. J. H. *J. Biol. Chem.* **1997**, *272*, 18111. Fraaije, M. W.; van den Heuvel, R. H. H.; Roelofs, J. C. A. A.; van Berkel, W. J. H. *Eur. J. Biochem.* **1998**, *253*, 712. McIntire, W. S.; Everhart, E. T.; Craig, J. C.; Kuusk, V. *J. Am. Chem. Soc.* **1999**, *121*, 5865. Priefert, H.; Overhage, J.; Steinbuchel, A. *Arch. Microbiol.* **1999**, *172*, 354.
(7) Thibault, P.; Logan, S. M.; Kelly, J. F.; Brisson, J.-R.; Ewing, C. P.; Trust, T. J.; Guerry, P. *J. Biol. Chem.* **2001**, *276*, 34862.
(8) More, M. I.; Finger, L. D.; Stryker, J. L.; Fuqua, C.; Eberhard, A.; Winans, S. C. *Science* **1996**, *272*, 1655.

JA0268985