ORGANIC LETTERS

2005
Vol. 7, No. 17
3613−3616

# *N*-Acyl Derivatives of Arginine and Tryptophan Isolated from Environmental DNA Expressed in *Escherichia coli*

**Sean F. Brady and Jon Clardy***

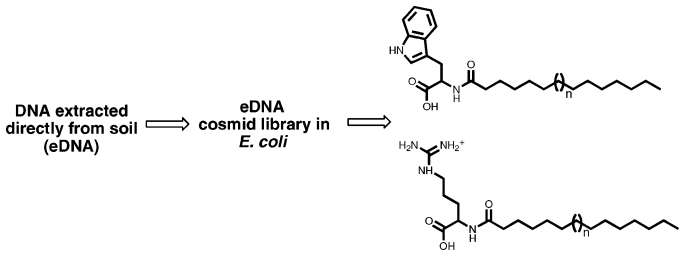
*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115*

*jon_clardy@hms.harvard.edu*

**Received April 28, 2005** *(Revised Manuscript Received June 17, 2005)*



## ABSTRACT




**Heterologous expression of microbial DNA extracted directly from environmental samples (environmental DNA, eDNA) in easily cultured hosts can provide access to natural products produced by previously uncultured bacteria. This report describes the characterization of antibacterially active long-chain *N*-acyl derivatives of tryptophan and arginine that are produced by eDNA clones hosted in *Escherichia coli*. The sequencing and subcloning of the proposed *N*-acyl amino acid synthases (NASs) for each family of natural products are also described.**

Many lines of evidence suggest that less than 1% of the microbes present in nature are readily cultured in the laboratory.[1−7] Uncultivated microorganisms are therefore a very attractive source of potentially new natural products. Although there appears to be no easy way to culture this large collection of unstudied microorganisms, it is possible to isolate large fragments of microbial DNA directly from uncultured bacteria present in environmental samples (environmental DNA, eDNA). Heterologous expression of eDNA in an easily cultured host could provide access to many of the natural products encoded by this previously inaccessible genetic material. We initially reported the isolation and characterization of a family of long-chain *N*-acyl tyrosine antibiotics produced by *Escherichia coli* transformed with eDNA.[8−11] In this report, we describe additional antibacterially active long-chain *N*-acyl amino acids and their biosynthetic enzymes that were found in a soil eDNA library expressed in *E. coli*.

An overview of the general approach that is used to clone eDNA from soil and then to screen eDNA clones for the production of antibacterial activity is shown in Figure 1. In brief, the bacteria present in an environmental sample are lysed in situ by heating in the presence of a detergent, and the freed high molecular weight DNA is collected by alcohol precipitation from the centrifuge-clarified crude lysate. Gel-purified high molecular weight eDNA is then partially digested with *Bam*H I, dephosphorylated with calf intestinal alkaline phosphatase and ligated into the *Bam*H I site of

(1) Hugenholtz, P.; Goebel, B. M.; Pace, N. R. *J. Bacteriol.* **1998**, *180*, 4765−4774.
(2) Bintrim, S. B.; Donohue, T. J.; Handelsman, J.; Roberts, G. P.; Goodman, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 277−282.
(3) Stackebrandt, E.; Liesack, W.; Goebel, B. M. *FASEB J.* **1993**, *7*, 232−236.
(4) Liesack, W.; Stackerbrandt, E. *J. Bacteriol.* **1992**, *174*, 5072−5078.
(5) Torsvik, V.; Goksoyr, J.; Daae, F. L. *Appl. Environ. Microbiol.* **1990**, *56*, 782−787.
(6) Torsvik, V.; Salte, K.; Sorheim, R.; Goksoyr, J. *Appl. Environ. Microbiol.* **1990**, *56*, 776−781.
(7) Ward, D. M.; Weller, R.; Bateson, M. M. *Nature* **1990**, *345*, 63−65.
(8) Brady, S. F.; Clardy, J. *J. Am. Chem. Soc.* **2000**, *122*, 12903−12904.
(9) Brady, S. F.; Chao, C. J.; Clardy, J. *J. Am. Chem. Soc.* **2002**, *124*, 9968−9969.
(10) Brady, S. F.; Clardy, J. *Org. Lett.* **2003**, *5*, 121−124.
(11) Brady, S. F.; Chao, C. J.; Clardy, J. *Appl. Environ. Microbiol.* **2004**, *70*, 6865−6870.

10.1021/ol0509585 CCC: $30.25 
**Published on Web 07/21/2005**

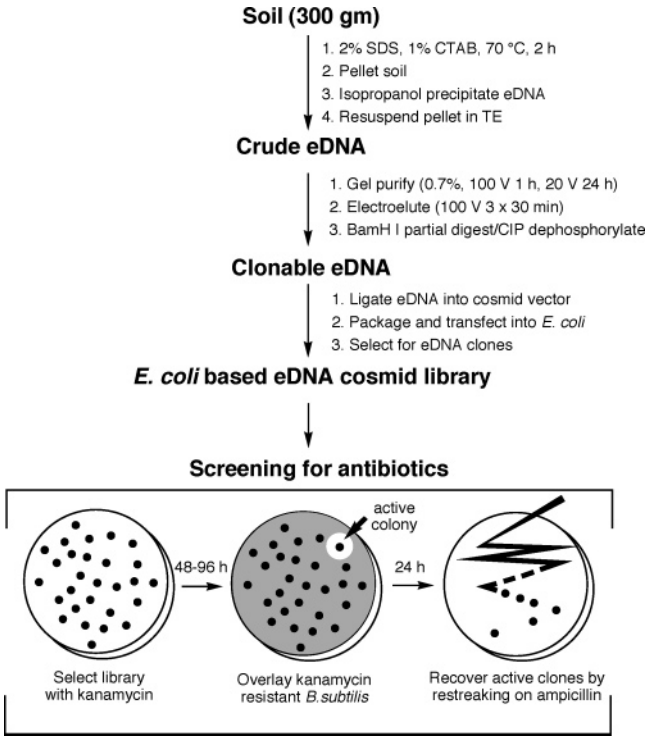



**Figure 1.** General scheme for the construction and screening of eDNA cosmid libraries from soil.

the cosmid vector pSuperCOS. The ligation reaction is then packaged into $\lambda$-phage and transfected into *E. coli*. To recover antibacterially active clones from the resulting library, eDNA clones are screened for the production of antibiosis using a top agar overlay containing *Bacillus subtilis*.

Clones that produce zones of growth inhibition in the top agar overlay are recovered from the assay plates and cleaned of the *B. subtilis* assay strain by restreaking on plates containing ampicillin. While both the eDNA clones and the *B. subtilis* assay strain are kanamycin resistant and therefore able to grow on the original selection plates, only the eDNA clones are resistant to ampicillin. Antibacterially active clones that produced antibacterially active ethyl acetate extracts when grown in liquid culture were selected for further characterization.

The antibacterially active constituents present in ethyl acetate extracts from cultures of two antibacterially active eDNA clones, CSL1 and CSL11, did not appear to result from any previously described common eDNA clone phenotypes (long-chain *N*-acyltyrosine, indigo/indirubin, aminolevulinic acid synthase (*hemA)* expression or melanin), and therefore a bioassay-guided fractionation of the active constituents present in these extracts was undertaken. Cultures of CSL1 and CSL11 grown in LB (30 $\mu$g/mL kanamycin) for 3 days at room temperature were neutralized with 2 N HCl and extracted twice with an equal volume of ethyl acetate, and the dried antibacterially active extract from each culture was partitioned by normal phase flash chromatography using a step gradient of $CHCl_3$/MeOH modified with 0.1% HOAc. Analysis by one- and two-dimensional NMR of the antibacterially active material that eluted from each normal phase flash column suggested that the active constituents were families of long-chain *N*-acyl amino acids.[12] The $\alpha$-carbon of each amino acid was apparent from the correlation of a highly deshielded methine carbon (53.4 and 55.2 ppm) to a deshielded proton doublet of doublets (4.4 and 4.72 ppm) in $^1H-^{13}C$ HMBC experiments, and the large methylene envelope of the long-chain fatty acids was easily seen in the $^1H$ spectra.

$^1H-^1H$ RelayH couplings and long-range $^1H-^{13}C$ HMBC correlations shown in Figure 2 confirmed that the antibiotics

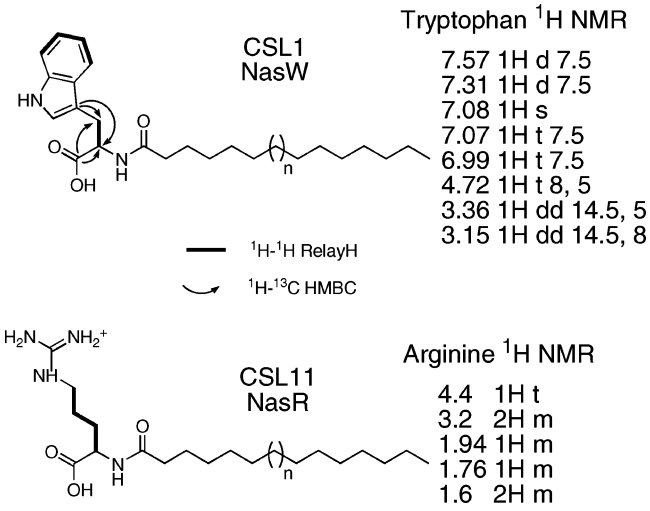



**Figure 2.** Long-chain *N*-acyl amino acids isolated from antibacterially active eDNA cosmid clones CSL1 and CSL11. In both cases, a family of compounds with different fatty acid side chains were isolated from each clone. $^1H$ NMR chemical shifts observed for each amino acid are listed.

produced by cultures of CSL1 and CSL11 were families of long-chain *N*-acyl derivatives of tryptophan and arginine, respectively. The C-3-substituted indole of the tryptophan present in the material extracted from cultures of CSL1 was apparent from the deshielded four-carbon spin system seen in the $^1H-^1H$ RelayH spectra and the deshielded proton singlet observed in the $^1H$ spectra. The two-carbon spin system that comprises the $\alpha$- and $\beta$-carbons of the tryptophan side chain was connected to the indole at one end and the free carboxylic acid at the other by long-range $^1H-^{13}C$ HMBC correlations shown in Figure 2. The arginine side chain present in the active material extracted from cultures of CSL11 was inferred from the four-carbon spin system seen in the $^1H-^1H$ RelayH spectra, the guanidinium carbon

(12) *N*-Palmiteoyl-L-tryptophan: $^{13}C$ NMR (100 MHz, $CD_3OD$) 176.2, 176.1, 138.2, 131.0, 131.0, 129.2, 124.4, 122.4, 119.8, 119.5, 112.3, 111.5, 55.2, 37.1, 33.1, 31.0, 30.5, 30.3, 30.2, 28.8, 28.3, 26.9, 23.9, 14.6. Natural mixture of long-chain *N*-acyl-L-arginines: $^{13}C$ NMR (100 MHz, $CD_3OD$) 176.6, 175.8, 158.8, 131.0, 53.4, 42.1, 37.0, 33.2, 33.1, 31.0−30.4 (m), 28.3, 27.1, 26.6, 23.9, 14.6.

at 158.8 ppm, and the three additional nitrogens predicted by HRFABMS (Table 1) to be present in the final structure.

**Table 1.** High-Resolution FABMS Data for the Major N-Acyl Amino Acids Present in Ethyl Acetate Extracts of *E. coli* Transformed with NASs from CSL1 (NasW), CSL11 (NasR), and CSL12 (NasY)

| eDNA clone compound[a] | molecular formula | (*m/z*) [M + H]$^+$ calcd | found |
|---|---|---|---|
| CSL1 | | | |
| Trp−C16:1 | $C_{27}H_{40}N_2O_3$ | 441.3117 | 441.3119 |
| Trp−C16 | $C_{27}H_{42}N_2O_3$ | 443.3274 | 443.3276 |
| Trp−C18:1 | $C_{29}H_{44}N_2O_3$ | 469.3430 | 469.3430 |
| CSL11 | | | |
| Arg−C14:1 | $C_{20}H_{38}N_4O_3$ | 383.3022 | 383.3020 |
| Arg−C14 | $C_{20}H_{43}N_4O_3$ | 385.3179 | 385.3177 |
| Arg−C15 | $C_{21}H_{42}N_4O_3$ | 399.3335 | 399.3334 |
| Arg−C16:1 | $C_{22}H_{42}N_4O_3$ | 411.3335 | 411.3335 |
| Arg−C16 | $C_{22}H_{44}N_4O_3$ | 413.3492 | 413.3494 |
| CSL12 | | | |
| Tyr−C14:1 | $C_{23}H_{35}NO_4$ | 390.2644 | 390.2645 |
| Tyr−C14 | $C_{23}H_{37}NO_4$ | 392.2801 | 392.2802 |
| Tyr−C16:1 | $C_{25}H_{39}NO_4$ | 418.2957 | 418.2956 |
| Tyr−C16 | $C_{25}H_{41}NO_4$ | 420.3114 | 420.3114 |
| Tyr−C18:1 | $C_{27}H_{43}NO_4$ | 446.3270 | 446.3271 |

[a] Compounds are listed as amino acid−fatty acid side chain.

Although the presence of the long-chain fatty acid side chain was obvious from the methyl triplet and the large methylene envelope seen in the $^1$H spectrum, the individual fatty acids present in each extract could not be resolved by NMR. The major long-chain *N*-acyl metabolites present in each extract were therefore purified by reversed-phase HPLC from the active mixture (acetonitrile and water with 0.1% triethylamine, gradient, Supelco ODP50) and then characterized using high-resolution FABMS (Table 1). On the basis of the molecular formulas predicted by HRFABMS, CSL1 produces long-chain *N*-acyltryptophans modified with both saturated and monounsaturated fatty acids that range from 16 to 18 carbons in length, and CSL11 produces long-chain *N*-acylarginines modified with both saturated and monounsaturated fatty acids that range from 14 to 16 carbons in length. Synthetic samples of *N*-palmitoleoyl-L-tryptophan and *N*-palmitoleoyl-L-arginine that were produced from the appropriate L-amino acids and the long-chain fatty acid chloride[13] are spectroscopically identical to the natural samples and, as seen with the natural samples, antibacterially active.[14]

To identify the genes responsible for the biosynthesis of each family of *N*-acyl amino acids, the eDNA cosmids isolated from CSL1 and CSL11 were transposon mutagenized with the genome priming system (GPS, New England Biolabs) and then retransformed into *E. coli*. Mutagenized cosmids that no longer conferred the production of antibiosis to *E. coli* were recovered, and the DNA flanking each transposon insertion was sequenced using transposon specific primers. Both eDNA clones were found to contain single open reading frames (ORF) that appeared to confer the production of antibacterial activity to *E. coli*. Neither the proposed long-chain *N*-acyltryptophan synthase from CSL1 (NasW−NAS tryptophan) nor the proposed long-chain *N*-acylarginine synthase from CSL11 (NasR−NAS arginine) showed any significant sequence identity (>20%) to sequences from cultured bacteria that have been deposited in GenBank or the long-chain *N*-acyltyrosine synthases (NasY−NAS tyrosine) found in other eDNA clones.[15]

To confirm that the two proposed NASs were necessary and sufficient to confer the production of long-chain *N*-acyl amino acids to *E. coli*, each ORF was subcloned as a GST fusion protein, and these constructs were then assayed for the ability to confer antibiosis to *E. coli*. NasW and NasR were amplified from the cloned eDNA using the polymerase chain reaction and cloned as GST fusion proteins in pGEX3X (Pharmacia Biotech) to give pNasWGST and pNasRGST, respectively.[16] *E. coli* cultures transformed with pNasWGST and pNasRGST were grown at 37 °C to an $OD_{600}$ of 0.7 at which point the temperature was reduced to 20 °C and the cultures induced with 10 $\mu$M IPTG. After an additional 18 h at 20 °C, the entire culture was extracted with ethyl acetate. FABMS, bioautography, and NMR analysis of these extracts confirmed that the predicted long-chain *N*-acyl amino acid synthases NasR and NasW are necessary and sufficient to confer the production of long-chain *N*-acylarginines and *N*-acyltryptophans to *E. coli*, respectively.

Long-chain *N*-acyl amino acid-producing clones have been found in every eDNA library that we have examined to date. The biosynthesis of long-chain *N*-acyl amino acids appears to be a common phenomenon among uncultured bacteria whose DNA can be readily expressed in *E. coli*. The frequency with which this family of compounds is found

(13) A 2-fold molar excess of each amino acid was stirred with 40 mg of palmitoleoyl chloride in 2 mL of 1:1 $CH_2Cl_2$/DMF with 100 $\mu$L of pyridine. After 48 h, the reactions were diluted with 10 mL of 1 N HCl and then extracted two times with 10 mL of $CH_2Cl_2$. The $CH_2Cl_2$ was removed under vacuum, and the synthetic long-chain *N*-acyl amino acids were then purified using silica gel chromatography ($CHCl_3$/MeOH + 0.1% HOAc, step gradient).

(14) In disc diffusion assays against *Staphylococus aureus*, 50 $\mu$g of Trp-C16:1 produced an 8 mm clear (14 mm hazy) zone of growth, and 5 $\mu$g of the ampicillin control produced a 14 mm clear zone of growth inhibition. Arg-C16:1 is only weakly active as an antibiotic against *S. aureus*. Growth inhibition against *S. aureus* was only observed when the compound was directly applied to the surface of a bacterial lawn. *N*-Palmiteoyl-L-tryptophan $[a]^{25}_D$ +10.8 (*c* 0.79, methanol). *N*-Palmiteoyl-L-arginine +23.6 (*c* 0.50, methanol/$CH_2Cl_2$ (1:1)).

(15) Sequence data for NasW and NasR have been deposited in GenBank under accession numbers AY214919 and AY214920, respectively.

(16) NasW and NasR were amplified from the cloned eDNA using the polymerase chain reaction and the following primer pairs: NasW 5'-CGTGGGATCCCCATGTTGATGGGCGATGAAGGGCA-3' and 5'-GCTCAATTGGCGGGATTGCTTGGCTTTGAAGCTGA-3' (*Bam*H I and *Mfe* I), NasR 5'-GGAGGGGATCCTCATGCAGCCAGAGATCTTCGCGC-3' and 5'-ATCGAATTCCTGGTCTCAGTCGCTCACATCTC-3' (*Bam*H I and *Eco*R I). The gel-purified PCR products were digested with the appropriate restriction endonucleases (as indicated above) and ligated into *Bam*H I- and *Eco*R I-digested pGEX-3X vector to give plasmids pNasWGST and pNasRGST.

using this approach suggests that *N*-acyl amino acids play an important role in the biology of uncultured bacteria. Although long-chain *N*-acyltryptophans and arginines were found using a screen for antibiotics, their weak antibacterial activity suggests that antibiosis may not be the primary role these compounds play in the native organisms that produce them.

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

**Supporting Information Available:** $^{1}H$ and $^{13}C$ NMR spectra for *N*-palmiteoyltryptophan from cultures of CSL1 and the natural mixture of *N*-acylarginines from CSL11.

OL0509585