Long-Chain *N*-Acyl Amino Acid Antibiotics Isolated from Heterologously Expressed Environmental DNA

Sean F. Brady and Jon Clardy*

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

Received August 11, 2000

The tiny minority of soil microorganisms that can easily be cultured with standard techniques, roughly 0.1 to 1.0%,[1] produce a spectacular array of biologically active natural products. The uncultured majority likely produces natural products with chemical diversity and biological activity similar to that of cultured microorganisms. To access the natural products produced by uncultured microorganisms,[2] a cosmid library of DNA extracted directly from soil samples (environmental DNA, eDNA) was constructed and screened for the production of biologically active small molecules.[3] One of the active cosmid clones, CSL12, produces a series of long-chain *N*-acyl-L-tyrosine antibiotics. Long-chain *N*-acyl amino acids are a growing family of bacterial natural products for which no biosynthesis genes have yet been identified. Analysis of the eDNA cloned in CSL12 indicated that a single open reading frame (ORF) was responsible for the production of these antibiotics and thus led to the identification of what we believe to be the first long-chain *N*-acyl amino acid biosynthesis gene. In this communication we report the characterization of these new natural product antibiotics and the sequence for a long-chain *N*-acyl amino acid synthase.

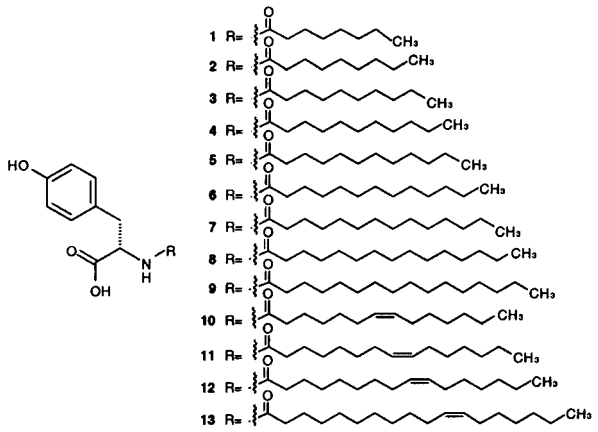
A sequential two antibiotic selection scheme was used to identify and then recover cosmid clones that produce antibacterial activities. The cosmid library of eDNA was originally selected on LB plates containing kanamycin and allowed to incubate at 30 °C for 2−4 days. The mature colonies were then overlayed with top agar containing kanamycin resistant *Bacillus subtilis*. After an additional 24 h of incubation at 30 °C, colonies that produced zones of growth inhibition in the *B. subtilis* lawn, indicating the production of antibacterial activity, were recovered by streaking bacteria picked from the active colonies onto LB plates containing ampicillin (50 μg/mL). The second selection, on ampicillin, removes the *B. subtilis* assay strain and allows for the recovery of antibacterial hits directly from the assay plates.

Of the approximately 700 000 clones screened, 65 antibacterial active colonies were found. Since we were most interested in small molecule antibiotics, ethyl acetate extracts from small-scale cultures of the active clones were assayed for antibacterial activity. One of the clones that produced a very active organic extract, CSL12, was chosen for further characterization. When the purified cosmid from CSL12 was retransformed into *E. coli*, it continued to confer antibacterial activity indicating that the cloned eDNA was responsible for the observed activity.

Table 1. HRMS Data for CSL12-A through CSL12-M (**1**−**13**)

CSL12-	molecular formula	(*m*/*z*) [M + H]$^+$ calcd	found	acyl side chain
A (**1**)	$C_{17}H_{25}N_1O_4$	308.1862	308.1864	$C_8H_{15}O$
B (**2**)	$C_{18}H_{27}N_1O_4$	322.2018	322.2011	$C_9H_{17}O$
C (**3**)	$C_{19}H_{29}N_1O_4$	336.2175	336.2176	$C_{10}H_{19}O$
D (**4**)[a]	$C_{20}H_{31}N_1O_4$	350.2331	350.2336	$C_{11}H_{21}O$
E (**5**)	$C_{21}H_{33}N_1O_4$	364.2488	364.2471	$C_{12}H_{23}O$
F (**6**)	$C_{22}H_{35}N_1O_4$	378.2644	378.2638	$C_{13}H_{25}O$
G (**7**)	$C_{23}H_{37}N_1O_4$	392.2801	392.2797	$C_{14}H_{27}O$
H (**8**)	$C_{24}H_{39}N_1O_4$	406.2957	406.2955	$C_{15}H_{29}O$
I (**9**)	$C_{25}H_{41}N_1O_4$	420.3114	420.3131	$C_{16}H_{31}O$
J (**10**)[a]	$C_{21}H_{31}N_1O_4$	362.2331	362.2329	$C_{12}H_{21}O$
K (**11**)	$C_{23}H_{35}N_1O_4$	390.2644	390.2652	$C_{14}H_{25}O$
L (**12**)	$C_{25}H_{39}N_1O_4$	418.2957	418.2967	$C_{16}H_{29}O$
M (**13**)	$C_{27}H_{43}N_1O_4$	446.3270	446.3260	$C_{18}H_{33}O$

[a] **4** and **10** were isolated by RPHPLC as a mixture.



The active constituents in the ethyl acetate extract from cultures of CSL12 were isolated using a bioassay-guided fractionation against *B. subtilis*. Crude ethyl acetate extracts were obtained from neutralized cultures grown in LB (30 μg/mL of kanamycin) at 30 °C for 60 h. The ethyl acetate extract was partitioned by normal phase flash chromatography ($CHCl_3$:MeOH step gradient with 0.1% HOAc) and the active material that eluted from the silica column was then further partitioned by reversed-phase LC-CN flash chromatography (CH_3CN:H_2O step gradient with 0.1% triethylamine). Thirteen related compounds trivially named CSL12-A through CSL12-M were isolated by reversed-phase HPLC from the antibacterial active material that eluted from the second flash column.[4] The active material from the ethyl acetate extract of a 3.5 kb BamH I subclone of CSL12, CSL12.1, produced an identical reversed-phase HPLC trace, and this subclone was used for the remainder of our characterization studies.

Mass spectral analysis of CSL12-A through CSL12-M suggested that they were a family of long-chain saturated and unsaturated acyl derivatives of tyrosine, **1**−**13**. Tyrosine was apparent from the fragments observed in the LRESI MS spectra (*m*/*z* 182 and 136) and the acyl side chains were deduced from the high-resolution mass spectra (Table 1). Compounds **1**−**13** were all negative in a ninhydrin assay suggesting the acyl substituents were attached through an amide bond to the nitrogen

(1) Hugenholtz, P.; Goebel B. M.; Pace, N. R. *J. Bacteriol.* **1998**, *180*, 4765. Bintrim, S. B.; Donohue, T. J.; Handelsman, J.; Roberts, G. P.; Goodman, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 277. Stackebrandt, E.; Liesack, W.; Goebel, B. M. *FASEB J.* **1993**, *7*, 232. Liesack, W.; Stackebrandt, E. *J. Bacteriol.* **1992**, *174*, 5072. Torsvik, V.; Goksoyr, J.; Daae, F. L. *Appl. Environ. Microbiol.* **1990**, *56*, 782. Ward, D. M.; Weller, R.; Bateson, M. M. *Nature* **1990**, *345*, 63.

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

(3) Construction of the cosmid library: A 1:1 (w:v) mixture of soil and lysis buffer (Zhou, J.; Bruns, M. A.; Tiedge, J. M. *Appl. Environ. Microbiol.* **1996**, *62*, 316−322.) was heated at 70 °C for 2 h and extracted once with an equal volume of chloroform, and the eDNA was then precipitated with 2-propanol (0.6 vol) from the centrifuge clarified aqueous phase. Crude eDNA was gel purified on preparative 0.7% agarose gels (100 V for 1 h then 20 V for 24 h) and the high molecular weight (HMW) DNA collected by electroelution (100 V, 3 × 30 min). Purified eDNA was partially digested with BamH I and dephosphorylated with calf intestinal alkaline phosphatase prior to ligating into the SuperCos I vector (Stratagene). DNA ligated into the BamH I site in the SuperCos I vector was subsequently packaged into Gigapack III Gold packaging extracts (Stratagene) and transformed into *E. coli* XL1 blue MR cells.

(4) HPLC conditions: ODP50, 10 mm × 25 cm, 4 mL/min, gradient CH_3CN:H_2O 10:90 to 40:60 over 30 min with 0.1% triethylamine.

10.1021/ja002990u CCC: $19.00
Published on Web 12/27/2000

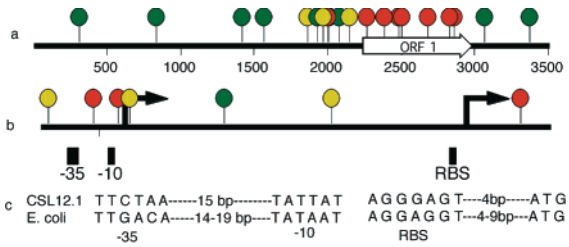


Figure 1. Genome priming systems (GPS) transposon mutagenesis of CSL12.1. Transposon insertions that knock out, reduce, or do not affect the antibacterial activity produced by CSL12.1 are marked with red, yellow, or green flags, respectively: (a) 3.5 kb insert from CSL12.1, (b) expanded region containing the predicted promoter for ORF1, and (c) proposed eDNA ribosome binding site (RBS), −35 and −10 sequences aligned with the corresponding consensus sequences from *E. coli*.

on tyrosine. The *N*-acyl-tyrosine structure and the configuration of the tyrosine in this series of compounds was ultimately confirmed by the total synthesis of two representative examples: CSL12-C (**3**), the most abundant compound isolated from CSL12.1, and CSL12-G (**7**), one of the most active compounds isolated from CSL12.1.[5] Both **3** and **7** are spectroscopically identical to their synthetic counterparts *N*-decanoyl-L-tyrosine and *N*-myristoyl-L-tyrosine, respectively.

CSL12-A to CSL12-M (**1**−**13**) show varying degrees of antibacterial activity, with the length of the aliphatic substitution on tyrosine markedly affecting the antibacterial activity of these natural products. In assays against *B. subtilis*, C_{13} to C_{16} saturated and unsaturated *N*-acyl substituted L-tyrosines were the most active. Tyrosine derivatives with side chains one and two carbons longer or shorter were markedly less active and derivatives with even shorter side chains were essentially inactive.

The genome priming system transposon mutagenesis (New England BioLabs, MA) of the CSL12.1 insert indicated that a single open reading frame (ORF1) produced the observed antibacterial activity and permitted the complete sequencing of the CSL12.1 insert.[6] With two exceptions, all of the transposon insertions that knock out the production of antibacterial activity were found in ORF1 (Figure 1a). The remaining two knock-out insertions were found in what is predicted to be the endogenous promoter in the eDNA for ORF1 (Figure 1b,c). In addition to transposons which completely knocked out the antibacterial activity, we also found a group of insertions both upstream and downstream of the proposed promoter but not in ORF1 that reduced the antibacterial activity observed in agar plate assays (Figure 1b). The transposon insertion studies suggest that ORF1 is responsible for the antibacterial activity and that the eDNA contains its own promoter that is successfully used by *E. coli*. In vitro biochemical studies are currently underway to confirm that ORF1 is an *N*-acyl transferase and not functioning indirectly to cause the production of *N*-acyl-tyrosine compounds in the *E. coli* host.

Three additional hypothetical ORFs (>150 amino acids), which do not appear to be organized as a secondary metabolite gene cluster, are present in CSL12.1. A BLAST[7] search did not identify any deposited sequences with sequence similarity to these ORFs. A BLAST search against deposited sequences found one sequence, a hypothetical protein (MJ1207) from *Methanococcus jannaschii*,[8] similar to ORF1. MJ1207 is a member of a small superfamily of hypothetical proteins that are predicted *N*-acyl transferases. Although ORF1 is only distantly related to this group of hypothetical *N*-acyl transferases by sequence homology, its functional similarity, the production of *N*-acyl products, argues that the sequence homology could be significant.

Long-chain *N*-acyl aminoacylases have been characterized from bacteria and appear to be widespread among microorganisms.[9] However, no enzymes that catalyze the formation of these compounds have been identified and the role of this growing family of natural products remains unclear.[10] With the observation that these long-chain *N*-acyl amino acids possess antibacterial activities, it is likely that long-chain *N*-acyl amino acids represent a general class of microbial antibiotics. The identification of ORF1 as an *N*-acyl amino acid biosynthesis gene should aid in determining the true role of this family of natural products as well as help in the assignment of function to other hypothetical *N*-acyl transferases from bacterial and archaeal sequencing projects.

The heterologous expression of eDNA accesses the chemical diversity of uncultured microorganisms. CSL12-A through CSL12-M (**1**−**13**) are the first new biologically active natural products that have been characterized from the heterologous expression of environmental DNA.[11] This heterologous expression approach automatically couples biologically active natural products with their biosynthetic genes. In this case, the characterization of the eDNA cloned in CLS12 identified the first biosynthesis gene for long-chain *N*-acyl amino acids. The straightforward strategy for constructing and screening large eDNA libraries that we have described should provide ready access to many of the natural products produced by previously inaccessible microorganisms. This method can also be used for heterologous expression in alternative host organisms using cosmid shuttle vectors and such studies are currently underway.

Acknowledgment. We thank Jo Handelsman and Robert Goodman for helpful discussions. Mass spectra were obtained in the Mass Spectrometry Facility, Department of Chemistry and Chemical Biology, Cornell University. This work was supported by the NIH (CA24487) and the David and Lucile Packard Foundation.

Supporting Information Available: NMR spectral data for **3** and characterization of the monounsaturated sides in **11**, **12**, and **13** (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA002990U

(5) A slight excess of the appropriate acid chloride was stirred with 20 mg of L-tyrosine in 1 mL of DMF at room temperature. After 3 h the reaction was diluted with 10 mL of 1 N HCl, extracted 3× with ethyl acetate, and the synthetic *N*-acyl-tyrosines were purified by reversed-phase HPLC.

(6) The nucleotide sequence of CSL12.1 has been deposited in GenBank under Accession Number AF324335.

(7) WU-BLAST (Washington University) 2.0.
(8) Bult, C. J.; et al. *Science* **1996**, *272*, 1058.
(9) Fukada, H.; Iwade, S.; Kimura, A. *J. Biochem.* **1982**, *91*, 1734. Shintani, Y.; Fukuda, H.; Okamoto, N.; Murata, K.; Kumura, A. *J. Biochem.* **1984**, *96*, 6343. Matsumoto, J.; Nagai, S. *J. Biochem.* **1972**, *72*, 269.
(10) Pohnert, G.; Jung, V.; Haukioja, E.; Lempa, K.; Boland, W. *Tetrahedron* **1999**, *55*, 11275. Alborn, H. T.; Turlings, T. C. J.; Jones, T. H.; Stenhagen, G.; Loughrin, J. H.; Tumlinson, J. H. *Science* **1997**, *276*, 945. Yagi, H.; Corzo, G.; Nakahara, T. *Biochim. Biophys. Acta* **1997**, *1336*, 28 and references therein. Yagi, H.; Corzo, G.; Yochochi, T.; Kamisaka, Y.; Yamaoka, M.; Nakahara, J. *Jpn. Oil Chem. Soc.* **1995**, *44*, 635. Asselineau, J. *Prog. Chem. Org. Nat. Prod.* **1991**, *56*, 1.
(11) Wang, G.; Graziani, E.; Waters, B.; Pan, W.; Li, X.; McDermott, J.; Meurer, G.; Saxena, G.; Anderson, R.; Davies, J. *Org. Lett.* **2000**, *2*, 2401.