ORGANIC LETTERS

2001
Vol. 3, No. 13
1981−1984

Cloning and Heterologous Expression of a Natural Product Biosynthetic Gene Cluster from eDNA

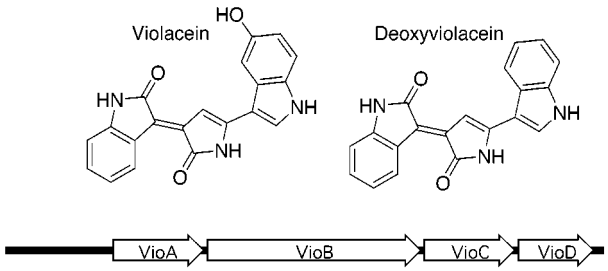
Sean F. Brady,[†] Carol J. Chao,[†] Jo Handelsman,[‡] and Jon Clardy*,[†]

Department of Chemistry and Chemical Biology, Cornell University, Baker Laboratory, Ithaca, New York 14853-1301, and Department of Plant Pathology, University of Wisconsin—Madison, 1630 Linden Drive, Madison, Wisconsin 53706

jcc12@cornell.edu

Received April 5, 2001

ABSTRACT



To study the natural products produced by uncultured microorganisms, an environmental DNA (eDNA) cosmid library was constructed and screened for the heterologous production of small molecules. A blue clone, CSL51, found in the eDNA library produces deoxyviolacein and the broad spectrum antibiotic violacein. The full sequence of the 6.7 kb eDNA violacein gene cluster and the characterization of violacein and deoxyviolacein from an eDNA clone are reported here.

Although cultured microorganisms have produced many of the most biologically active and medically useful small molecules,[1] the continued screening of easily cultured microbes inevitably faces diminishing returns. Since uncultured microbes vastly outnumber their cultured counterparts by at least 100:1,[2] screening uncultured microorganisms could avoid this liability. Of course, uncultured soil microbes have their own liability in our inability to culture them and then characterize the resulting natural products found in their culture broth. The DNA of the entire microbial community can be extracted directly from the environment, eDNA, and heterologous expression of eDNA in easily cultured hosts could provide access to many of the natural products encoded by the extracted DNA.[3−6] As many studies on biosynthetic pathways of known natural products have shown, secondary metabolite biosynthetic pathways are tightly clustered on bacterial chromosomes, and as a result the cloning and heterologous expression of secondary metabolites from single continuous fragments of genomic DNA has been possible.[7] Here we report the characterization of CSL51, a blue clone found in an eDNA cosmid library constructed in *Escherichia*

[†] Cornell University.
[‡] University of Wisconsin—Madison.

(1) Newman, D. J.; Cragg, G. M.; Snader, K. M. *Nat. Prod. Rep*. **2000**, *17*, 215. Cragg, G. M.; Newman, D. J.; Snader, K. M. *J. Nat. Prod*. **1997**, *60*, 52.
(2) 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. Torsvik, V.; Salte, K.; Sorheim, R.; Goksoyr, J. *Appl. Environ. Microbiol*. **1990**, *56*, 776. Ward, D. M.; Weller, R.; Bateson, M. M. *Nature* **1990**, *345*, 63.
(3) Brady, S. F.; Clardy, J. *J. Am. Chem. Soc.* **2000**, *122*, 12903.
(4) Handelsman, J.; Rondon, M. R.; Brady, S. F.; Clardy, J.; Goodman, R. M. *Chem. Biol.* **1998**, *5*, R245.
(5) Rondon, M. R.; August, P.; Bettermann, A. D.; Brady, S. F.; Grossman, T. H.; Liles, M. R.; Loiacono, K.; Lynch, B.; 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.
(6) 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.

10.1021/ol015949k CCC: $20.00
Published on Web 05/26/2001

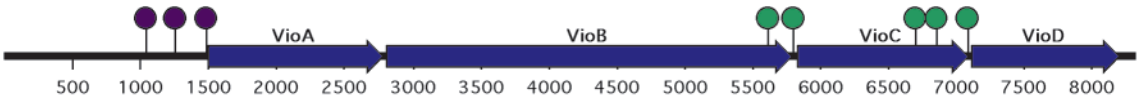


Figure 1. The violacein gene cluster sequenced from CSL51, a blue eDNA clone. The positions of transposon insertions that lead to darker blue colonies and green colonies are shown as dark blue or green flags, respectively.

coli. The eDNA captured in CSL51 contains a four-gene biosynthetic cluster and successfully confers the production of violacein (**1**) and deoxyviolacein (**2**) to the *E. coli* host.

An eDNA cosmid library was constructed from soil collected in the Ithaca (NY) area and screened for colored clones. High molecular weight eDNA was isolated directly from soil samples by heating with detergent, extracting with chloroform, and finally precipitating with 2-propanol from the centrifuge-clarified aqueous phase. Large fragments of partially digested and dephosphorylated gel-purified eDNA were then used to construct a SuperCos I[8] based cosmid library in *E. coli* using standard methodologies.[9]

The presence of color in microbial cultures is often an indication of small molecule biosynthesis. Therefore, color can be used as an initial screen to identify eDNA cosmid clones containing biosynthetic gene clusters for natural products. A blue clone, CSL51, found in the eDNA library was selected for further characterization. *E. coli* retransformed with pCSL51, the cosmid isolated from CSL51, was also blue, indicating that the cloned eDNA is responsible for the blue phenotype. Random transposon mutagenesis of pCSL51 with the Genome Priming System (GPS)[10] produced clones with two distinct phenotypes, darker blue colonies and green colonies. We focused on these two groups of transposon mutants with obvious color differences because transposon-induced knockouts that no longer produced a blue color could not be distinguished easily from pCSL51-transformed *E. coli* cultures with low levels of color production.

The DNA sequence derived from the transposon mutants indicated that the eDNA responsible for the colored phenotypes resembled that of the violacein gene cluster sequenced from the cultured bacterium *Chromobacterium violaceum*.[11,12] In addition, the green phenotype has been observed in cultures of *E. coli* transformed with a transposon-mutagenized violacein gene cluster from *C. violaceum*.[11,13] Primer walking was used to fill the gaps between the sequence obtained from these transposon insertions. CSL51 eDNA contains a four-gene natural product biosynthetic gene cluster, vioA, vioB, vioC, and vioD, that spans 6.7 kb (Figure 1).[14,15] Transposon insertions in pCSL51 that result in darker blue colonies are located upstream of vioA while transposon insertions that result in green colonies are found in vioC and the very C-terminal end of vioB (Figure 1).

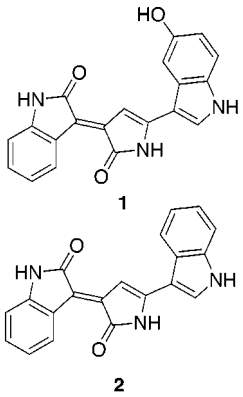


Figure 2. Violacein (**1**) and deoxyviolacein (**2**), blue compounds isolated from *E. coli* cultures containing eDNA.

The blue material produced by *E. coli* transformed with eDNA from pCSL51 was characterized from cultures of a GPS transposon mutant, CSL51-GPS13, and a *Not* I partial digest subclone of pCSL51, CSL51-N3, both of which produce darker blue colonies. Ethyl acetate extracts obtained from cultures grown in LB broth at 24 °C were resuspended in 80% methanol and extracted 3× with hexanes. The 80% methanol fraction was then diluted with water to approximately 50% methanol and re-extracted 3× with $CHCl_3$. The $CHCl_3$ fraction contained most of the blue color and was partitioned by normal phase flash chromatography using a hexanes:ethyl acetate step gradient modified with 0.1% HOAc. The blue material that eluted with 100% ethyl acetate was then purified by reversed phase HPLC.[16] The major colored metabolite isolated from cultures of CSL51-N3 is spectroscopically identical to violacein (**1**) while the major colored metabolite isolated from cultures of CSL51-GPS13 is spectroscopically identical to deoxyviolacein (**2**) (Figure 2).[17,18] The entire violacein biosynthetic gene cluster was therefore captured on a single fragment of eDNA and successfully expressed in a heterologous host.

Violacein and deoxyviolacein are amino acid dimers that are believed to arise from the dimerization, decarboxylation,

(7) Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien, B. *Science* **2000**, *287*, 640. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Betlach, M.; Ashley, G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1846. Bennett, J. W. *Can. J. Bot.* **1995**, *73*, S917 and references therein. Stone, M. J.; Williams, D. H. *Mol. Microbiol.* **1992**, *6*, 29 and references therein.

(8) Stratagene (CA).

(9) A detailed description of the methods used to construct the eDNA library will be described elsewhere.

(10) New England BioLabs (MA).

(11) August, P. R.; Grossman, T. H.; Minor, C.; Draper, M. P.; MacNeil, I. A.; Pemberton, J. M.; Call, K. M.; Holt, D.; Osburne, M. S. *J. Mol. Microbiol. Biotechnol.* **2000**, *2*, 513.

(12) Violacein gene clusters from *C. violaceum* have been deposited in Genbank under accession numbers AF172851 and AB032799.

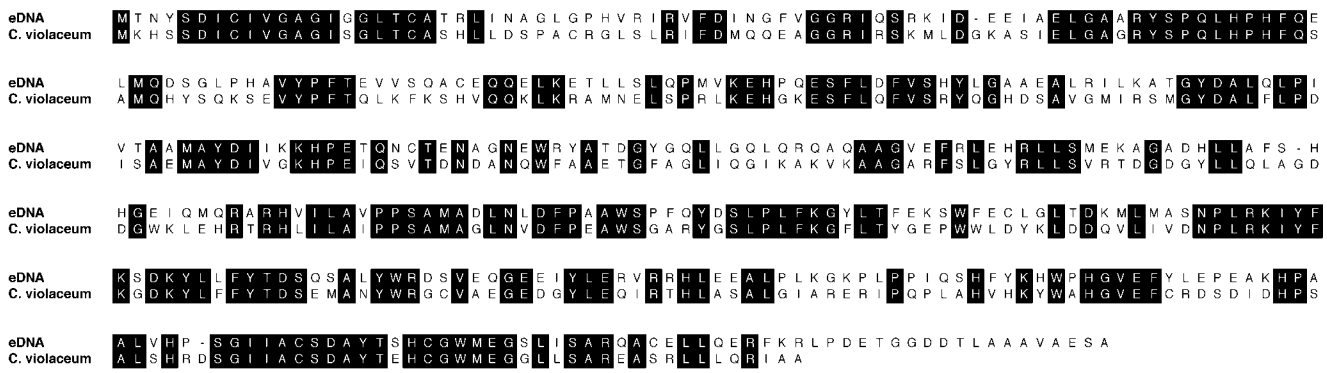


Figure 3. ClustalW alignment of VioA from the eDNA cloned in CSL51 (434 amino acids) and VioA from the cultured bacterium *C. violaceum* (418 amino acids). Identical residues are highlighted with dark boxes.

and oxidation of two molecules of tryptophan.[18,19] VioA, a predicted L-amino acid oxidase, and VioB are thought to be involved in the early steps of violacein biosynthesis. The disruption of VioC, a predicted FAD monooxygenase, results in the production of a green pigment seen in some of the transposon mutants. VioD, a predicted hydroxylase, is required for the aromatic hydroxylation seen in violacein.[11] Cultures of the dark blue GPS transposon mutant CSL51-GPS13 produce only deoxyviolacein, suggesting that VioD may not efficiently carry out the oxidation when the violacein biosynthetic gene cluster is significantly up regulated.

It is now theoretically possible to construct eDNA cosmid libraries with more than $0.1-1.0 \times 10^9$ members, the overwhelming majority of which will not produce natural products. If eDNA libraries are to be a viable source of new small molecules, simple methods for quickly identifying the tiny minority of eDNA clones that may produce natural products must be developed. The rapid identification of clones that exhibit "natural product-like phenotypes", including antibacterial activities and color production, has enabled us to find small molecule-producing clones in very large eDNA cosmid libraries. We recently described a high throughput double antibiotic selection assay that can screen eDNA libraries with $>1 \times 10^6$ members for cytotoxic activity.[3] The identification of colored clones in eDNA libraries also appears to be a very simple filter to identify potential natural product-producing clones.

Violacein, which we characterized because of its color, is a broad spectrum antibiotic with activity against *Staphylococcus aureus*, *Bacillus* sp., and *Streptococcus* sp.[20] and has been reported to induce apoptosis in fibroblast cells.[21] Natural products have evolved to play a vital role, primarily defensive in nature, for the producing organisms. The use of simple high throughput assays as an initial filter to identify natural product-producing clones, in lieu of more cumbersome bioassays or analysis of the culture broth from individual clones, is likely to be the most efficient means of finding biologically active small molecules in very large eDNA libraries.

In subsequent screens of eDNA clones produced from a soil sample collected in a second Ithaca (NY) site, an additional blue clone, CSL133, was found. This clone was not extensively characterized; however, the vioA gene was completely sequenced from pCSL133, the cosmid isolated from CSL133, and found to be 100% identical to the vioA gene sequenced from pCSL51. While the two eDNA vioA genes are 100% identical, the DNA sequences for vioA, vioB, vioC, and vioD from CSL51 are only 59, 69, 74, and 72%, identical to the corresponding sequences from *C. violaceum*. The predicted amino acid sequences for VioA, VioB, VioC, and VioD from CSL51 are 48, 62, 71, and 69%, identical to the predicted sequences from *C. violaceum*. A representative ClustalW[22] alignment of the most divergent protein sequences, VioA from pCSL51 (434 amino acids) and *C. violaceum* (418 amino acids), is shown in Figure 3.

The DNA sequences from pCSL51 and *C. violaceum* that code for violacein biosynthesis genes and the region upstream of vioA in *C. violaceum* have very high G + C content, 62.8, 67.5, and 64.9%, respectively. The G + C content in

(13) Pemberton, J. M.; Vincent, K. M.; Penfold, R. J. *Curr. Microbiol.* **1991**, *22*, 355.

(14) The sequence of the CSL51 violacein gene cluster has been deposited with Genbank under accession number AF367409.

(15) Nucleic acids analysis was done using *MacVector*, version 6.5.3; Oxford Molecular Ltd: Oxford, U.K., 1999.

(16) HPLC conditions: deoxyviolacein, C18 column 10 mm × 25 cm, 5 mL/min, 6:4 water:acetonitrile with 0.1% HOAc (t_R = 10.6 min); violacein, C18 column 10 mm × 25 cm, 1.5 mL/min, 6:4 water:acetonitrile with 0.1% HOAc (t_R = 13.4 min).

(17) Violacein (**1**): HREIMS *m*/*z* M^+ calcd for $C_{20}H_{13}N_3O_3$, 343.0957; found 343.0959; ^{13}C NMR d_6-DMSO 171.7, 170.2, 152.9, 147.6, 141.8, 137.0, 131.6, 129.6, 129.4, 126.4, 125.6, 122.4, 120.8, 118.7, 113.4, 113.1, 109.0, 105.8, 104.6, 96.9; ^1H NMR d_6-DMSO 11.89 (d, 2.5), 10.74 (d, 1.5), 10.62 (s), 9.34 (s), 8.93 (d, 8.0), 8.07 (d, 2.5), 7.55 (d, 2), 7.35 (d, 8.5), 7.24 (d, 2), 7.20 (td, 7.5, 1.0), 6.95 (td, 7.5, 1.0), 6.82 (d, 8.0), 6.79 (dd, 8.5, 2.5). Deoxyviolacein (**2**): HREIMS *m*/*z* M^+ calcd for $C_{20}H_{13}N_3O_2$, 327.1008; found 327.1008; ^{13}C NMR d_6-DMSO 171.5, 170.2, 147.1, 142.0, 137.5, 136.8, 129.8, 129.5, 126.5, 124.5, 123.1, 122.3, 121.6, 120.9, 119.8, 119.6, 113.0, 109.1, 106.3, 97.4; ^1H NMR d_6-DMSO 12.11 (s), 10.82 (d, 1.5), 10.65 (s), 8.93 (d, 8.0), 8.18 (s), 7.85 (m), 7.65 (d, 2), 7.56 (m), 7.31 (2H, m), 7.22 (td, 7.5, 1.0), 6.96 (td, 8.0, 1.0), 6.83 (d, 8.0).

(18) Hoshino, T.; Kondo, T.; Uchiyama, T.; Ogasawara, N. *Agric. Biol. Chem.* **1987**, *51*, 965.

(19) Momen, A. Z. M. R.; Hoshino, T. *Biosci. Biotechnol. Biochem.* **2000**, *64*, 539.

(20) Lichstein, H. C.; van de Sand, V. F. *J. Infect. Dis.* **1945**, *76*, 47.

(21) Da Silva, M. S. S. M. P.; De Campos, V. B.; Haun, M.; Duran, N. *In Vitro Cell. Dev. Biol. Animal* **2000**, *36*, 539.

(22) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucl. Acids Res.* **1994**, *22*, 4673.

the 1.5 kb immediately upstream of vioA in pCSL51 drops to 55.8%, and this sequence is not related (40% identity with multiple gaps) to the corresponding sequence from *C. violaceum*. Taken together, these observations suggest that although the coding regions for these two violacein gene clusters show significant similarity, the similarity quickly disappears outside the coding regions. Therefore, the two violacein pathways likely originated from different bacteria.

This study shows that cloning large fragments of eDNA in *E. coli*-based cosmid libraries is a feasible means to access complex natural products from uncultured microorganisms. It also suggests that simple surrogate assays, which rapidly identify clones with natural product-like phenotypes, can be used to find eDNA clones that produce bioactive small molecules.

Acknowledgment. This work was supported by the NIH (CA24487) and the David and Lucile Packard Foundation. Mass spectra were obtained by the University of Illinois mass spectrometry facility and the Cornell University (Department of Chemistry and Chemical Biology) mass spectrometry facility.

OL015949K