Cloning and Heterologous Expression of the Vibrioferrin Biosynthetic Gene Cluster from a Marine Metagenomic Library

Masaki J. Fujita, Nobutada Kimura, Atsushi Sakai, Yoichi Ichikawa, Tomohiro Hanuy and Masami Otsuka

1 Priority Organization for Innovation and Excellence, Kumamoto University, 5-1 Oe-Honmachi, Kumamoto 862-0973, Japan
2 Bioproduction Research Institute, Advanced Industrial Science and Technology, 1-1-1 Higashi, Tsukuba 305-8568, Japan
3 Graduate School of Pharmaceutical Sciences, Kumamoto University, 5-1 Oe-Honmachi, Kumamoto 862-0973, Japan
4 Priority Organization for Innovation and Excellence, Kumamoto University, 5-1 Oe-Honmachi, Kumamoto 862-0973, Japan

Received May 13, 2011; Accepted September 3, 2011; Online Publication, December 7, 2011

A biosynthetic gene cluster of siderophore consisting of five open reading frames (ORFs) was cloned by functional screening of a metagenomic library constructed from tidal-flat sediment. Expression of the cloned biosynthetic genes in Escherichia coli led to the production of vibrioferrin, a siderophore originally reported for the marine bacterium Vibrio parahaemolyticus. To the best of our knowledge, this is the first example of heterologous production of a siderophore by biosynthetic genes cloned from a metagenomic library. The cloned cluster was one of the largest of the clusters obtained by functional screening. In this study, we demonstrated and extended the possibility of function-based metagenomic research.

Key words: metagenome; siderophore; vibrioferrin; biosynthesis

Marine organisms, including marine bacteria, are rich sources of bioactive compounds, and many of these are promising potential pharmaceuticals. To date, more than 20,000 marine natural products have been reported, and some are in medical use or clinical trials, but the number of compounds of marine origin used in clinical treatment is still quite limited. The major barrier is limited supply, often caused by difficulty in collecting or cultivating the source organism, coupled with a degree of structural complexity that discourages synthetic approaches. Recent studies suggest that a large fraction of the natural products isolated from marine invertebrates are produced by symbiotic bacteria, but it is also known that most environmental bacteria, especially marine invertebrate symbiotic microorganisms, are difficult to culture under artificial conditions.

In view of these facts, we attempted to construct a practical supply system for bacterially derived marine natural products by heterologous expression of biosynthetic genes in easily cultured hosts. Using a metagenomic approach, biosynthetic genes were expected to be able to obtain from marine environmental samples containing marine invertebrates. This approach has been used to obtain genomic DNA directly from environmental samples without microbial isolation or cultivation. In principle, it opens up the genetic resources of uncultured environmental bacteria. Expression of the biosynthetic genes cloned from uncultured species is expected to lead to the discovery of novel compounds and also to the production of known valuable compounds. The production of several compounds such as pigments and antibiotics from a soil metagenomic library has already been reported.

In our application of a metagenomic approach to marine natural products chemistry, we chose a tidal-flat sediment collected from the Ariake Sea as genomic source. Tidal-flats are characterized by high bacterial diversity, because they provide a nutrient- and oxygen-rich environment. According to a 16S rDNA analysis of metagenomic DNA extracted from tidal-flat sediment in the temperate zone, it consisting of 13 lineages of the domain Bacteria, including uncultured candidates.

Developing assay methods to screen large libraries for desired genes is the most important and difficult step in the functional metagenomic research. Especially, methods of detecting small molecule production are quite limited, and this prevents the application of metagenomics to natural products chemistry. In this study, we employed a metal-binding compound detection assay utilizing the metal ion indicator, chrome azurol S (CAS), for screening purposes. Because the marine environment contains very little bio-available iron, an essential element for almost all organisms, most marine bacteria produce siderophores, iron-binding molecules, to capture and import iron. Additionally, metal-binding molecules, such as siderophores, serve as useful leads in the development of inhibitors against metalloenzymes including metalloproteinases which are involved in cancer metastasis. Siderophores are thus potential anti-metastatic agents. Here we report the construction of a marine metagenomic library from...
tidal-flat sediment, the cloning of a siderophore biosynthetic gene cluster, and the heterologous production and identification of vibrioferrin (Fig. 1).

**Results and Discussion**

*Marine metagenomic library construction and siderophore screening*

Tidal-flat sediment collected at the Ariake Sea (Kumamoto, Japan, August 2009) at low tide was suspended in sterilized sea water and shaken vigorously overnight to separate bacterial cells from mud particles. The inorganic fraction was removed by low-speed centrifugation, and cells were pelleted by subsequent high-speed centrifugation of the supernatant. The cell pellet was then washed with TES buffer. Pure high molecular weight DNA (over 30 kb) was obtained using a modified soil metagenomic library production protocol. Crude metagenomic DNA was extracted from the cell pellet by treating it with lysozyme, RNase, proteinase K, cetyltrimethylammonium bromide (CTAB), and phenol-chloroform. The crude DNA was then fractionated by agarose gel electrophoresis to prepare DNA pure enough for further cloning procedures. A metagenomic library was constructed using a CopyControl Fosmid Library Production Kit (Epicentre, Madison, WI) following the manufacturer’s protocol. Purified metagenomic DNA was enzymatically blunt-ended, ligated into the fosmid vector, and transfected into E. coli to construct a subclone library. Several active subclones were found on the second screening, and the one with the minimum insert size, MG93-1, was selected for full DNA sequencing. A nucleotide sequence of 7,903 bp was determined, and it revealed the presence of five continuous ORFs showing the same transcriptional direction. Judging from the results of a homology search, these ORFs were quite similar to pvsA-E, a biosynthetic gene cluster for vibrioferrin, a siderophore originally reported from the pathogenic marine bacteria *Vibrio parahaemolyticus* and *Vibrio alginolyticus*.

Based on the amino acid sequence, all five predicted proteins exhibited 98% identity with PvsA, B, C, and E from *V. parahaemolyticus* and PvsD from *V. alginolyticus* (Fig. 2).

Siderophore activity was detected in the culture broth of the subclone. The active molecule was concentrated by C18 resin solid phase extraction after acidification of the medium to pH 2 with trifluoroacetic acid. The methanol eluate from the resin was fractionated by C18 reversed phase column chromatography, gel filtration chromatography with Sephadex G-10, and C18 HPLC to afford the pure active compound as a clear viscous oil (296.3 mg from 3.2 L of culture, 92.6 mg/L based on isolation yield).

Positive and negative mode FAB mass spectra, including negative mode high resolution data, suggested that the molecular formula of the active compound matched that of vibrioferrin (1), an expected compound based on enzyme homology. Judging from the 1H and 13C NMR data shown in Table 1, it was obvious that two isomers were present at a ratio of 55:45. These data also indicated that the product was related to compound 1, which is known to form two diastereomeric aminal isomers at the C-5 position. Additionally, interpretation of the 1D and 2D NMR data (Fig. 3, and supplemental data available at the Biosci. Biotechnol. Biochem. Web site) confirmed that the gross structure of the product was the same as that of compound 1. Finally, the positive optical rotation value also confirmed its identity, including absolute stereochemistry. Here we concluded that the active molecule produced was intact vibrioferrin (1).

Vibrioferrin (1) consists of one each of 2-ketogluutaric acid, t-alanine, ethanolamine, and citric acid, all of which was about 40 kbp, was chosen for further study due to its prominent activity. In order to minimize the sequencing effort, fosmid DNA was randomly digested up to 10 kbp using Sau3AI. The 5–10 kbp DNA fragments recovered were transformed into *E. coli* to construct a subclone library. Several active subclones were found on the second screening, and the one with the minimum insert size, MG93-1, was selected for full DNA sequencing. A nucleotide sequence of 7,903 bp was determined, and it revealed the presence of five continuous ORFs showing the same transcriptional direction. Judging from the results of a homology search, these ORFs were quite similar to pvsA-E, a biosynthetic gene cluster for vibrioferrin, a siderophore originally reported from the pathogenic marine bacteria *Vibrio parahaemolyticus* and *Vibrio alginolyticus*.

Based on the amino acid sequence, all five predicted proteins exhibited 98% identity with PvsA, B, C, and E from *V. parahaemolyticus* and PvsD from *V. alginolyticus* (Fig. 2).

Siderophore activity was detected in the culture broth of the subclone. The active molecule was concentrated by C18 resin solid phase extraction after acidification of the medium to pH 2 with trifluoroacetic acid. The methanol eluate from the resin was fractionated by C18 reversed phase column chromatography, gel filtration chromatography with Sephadex G-10, and C18 HPLC to afford the pure active compound as a clear viscous oil (296.3 mg from 3.2 L of culture, 92.6 mg/L based on isolation yield).

Positive and negative mode FAB mass spectra, including negative mode high resolution data, suggested that the molecular formula of the active compound matched that of vibrioferrin (1), an expected compound based on enzyme homology. Judging from the 1H and 13C NMR data shown in Table 1, it was obvious that two isomers were present at a ratio of 55:45. These data also indicated that the product was related to compound 1, which is known to form two diastereomeric aminal isomers at the C-5 position. Additionally, interpretation of the 1D and 2D NMR data (Fig. 3, and supplemental data available at the Biosci. Biotechnol. Biochem. Web site) confirmed that the gross structure of the product was the same as that of compound 1. Finally, the positive optical rotation value also confirmed its identity, including absolute stereochemistry. Here we concluded that the active molecule produced was intact vibrioferrin (1).
which are primary metabolites present in *E. coli* (Fig. 4). There are five genes, *pvsA* to *pvsE*, in the cluster, and based on sequence homology, PvsC is most likely the transporter involved in secreting the product from the cell. Two proteins, PvsB and PvsD, are annotated as enzymes for amide bond formation between 2-ketoglutarate and L-alanine, or L-alanine and ethanolamine. One of the two remaining enzymes, PvsE, has high homology with 2,6-diaminopimelate decarboxylase, indicating that it catalyzes decarboxylation from an amino acid to ethanolamine or its precursor. The function of the last enzyme, PvsA, is thought to form an ester bond between citrate and ethanolamine. Considering that the active molecule was found in the culture broth, the final product was secreted by transporter PvsC. These data indicate that all five genes are functionally expressed in the heterologous host.

It is quite rare that such a long gene cluster consisting of five ORFs with 7,245 bp is cloned by the functional metagenomic method and works efficiently in an alternative host. This encouraged us to discover new biosynthetic gene clusters as well as to produce unknown molecules from the randomly generated metagenomic library. It has been found that the genome mining technique utilizing sequence homology data is a powerful method, especially for well-studied polyketide synthases and non-ribosomal peptide synthases.20) However, there are many marine natural products with small and unique structures whose biosynthetic genes are unpredictable from sequence homology. Furthermore, in many cases, not only biosynthetic genes but even actual producers remain unknown. For these compounds, random metagenomic screening and heterologous expression is believed to be one of the most efficient methods of revealing the mechanism of biosynthesis.

![Fig. 3. Key 2D NMR Correlations.](image)

![Fig. 4. Hypothetical Biosynthetic Pathway of Compound 1.](image)

Table 1. 1H and 13C NMR Data for Compound 1 Produced by Metagenomic Clone MG93-1 in D2O + CD3OD (6:1) (500 MHz for 1H and 125 MHz for 13C)

<table>
<thead>
<tr>
<th></th>
<th>δH (mult. J (Hz))</th>
<th>δC</th>
<th></th>
<th>δH (mult. J (Hz))</th>
<th>δC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.77 (d, 13.5)</td>
<td>4.9</td>
<td>4.49</td>
<td>44.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.94 (d, 13.5)</td>
<td>4.44</td>
<td>4.42</td>
<td>44.2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.74 (d, 13.5)</td>
<td>4.24</td>
<td>4.24</td>
<td>44.2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.94 (d, 13.5)</td>
<td>4.24</td>
<td>4.24</td>
<td>44.2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>171.8</td>
<td>171.8</td>
<td>174.0</td>
<td>174.0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>177.2</td>
<td>177.2</td>
<td>177.2</td>
<td>177.2</td>
<td></td>
</tr>
<tr>
<td>1'</td>
<td>4.07 (m)</td>
<td>64.4</td>
<td>64.4</td>
<td>64.4</td>
<td></td>
</tr>
<tr>
<td>2'</td>
<td>3.37 (m)</td>
<td>39.0</td>
<td>39.0</td>
<td>39.0</td>
<td></td>
</tr>
<tr>
<td>3'</td>
<td>3.97 (q, 7.0)</td>
<td>53.5</td>
<td>53.5</td>
<td>53.5</td>
<td></td>
</tr>
<tr>
<td>4'</td>
<td>1.31 (d, 7.0)</td>
<td>14.7</td>
<td>14.7</td>
<td>14.7</td>
<td></td>
</tr>
<tr>
<td>5'</td>
<td>2.48 (m)</td>
<td>29.6</td>
<td>29.6</td>
<td>29.6</td>
<td></td>
</tr>
<tr>
<td>6'</td>
<td>2.09 (m)</td>
<td>32.8</td>
<td>32.8</td>
<td>32.8</td>
<td></td>
</tr>
<tr>
<td>7'</td>
<td>2.50 (m)</td>
<td>29.9</td>
<td>29.9</td>
<td>29.9</td>
<td></td>
</tr>
</tbody>
</table>

Vibrio ferrin from a Marine Metagenome 2285
Generally, the expression of siderophore biosynthetic genes is strictly regulated by the concentration of ferric ion in the environment. However, in this clone, the putative promoter sequence, and the ferric uptake repressor (fur) binding site (Fur box), which regulate the expression of vibrioferrin biosynthetic genes, were completely lost. These observations suggested that the cloned cluster was expressed by the promoter in the cloning vector, and that enabled constitutive and large-volume production of vibrioferrin in a regular medium without any iron removal.

The isolation yield of compound 1 (92.6 mg/L) was high compared to the known producing strain, V. parahaemolyticus (2.5 mg/L), which requires strictly iron-regulated media. High production made it possible to provide enough material for a variety of bioassays to discover unrecognized bioactivities of the compound. Several bioassays are now in progress.

A recent study suggested that siderophores are key molecules within environmental bacterial communities promoting survival and expressing various bio-phenomena, and it is also known that some siderophores are utilized by species other than those that produced them. It is expected that there are more unknown metal-binding molecules in nature produced by uncultured bacteria. Functional screening of metagenomic libraries utilizing metal indicators including noble and heavy metals is potentially method of discovering not only clinical resources but also molecules valuable in a variety of applications.

Conclusions

We cloned several biosynthetic gene clusters for siderophores from a marine metagenomic library constructed from tidal-flat sediment, and successfully produced vibrioferrin (1) by heterologous expression using one of the cloned clusters. To date, several secondary metabolites have been produced by similar methods. However, most of them were pigments or antibiotics. To the best of our knowledge, this is the first report of heterologous production of siderophores from a metagenomic clone. The product was one of the most complex molecules found by functional metagenomic screening, which required five genes for production and secretion. These results indicate the potential of this methodology to enable identification of relatively large biosynthetic gene clusters and to produce complex compounds, and extend the power of metagenomic approach to natural products chemistry. We also prepare an efficient supply system for a valuable natural product.

In this study, the compound obtained was known previously, but the result is a step toward discovering novel compounds produced by uncultured marine bacteria, and also to create a practical supply of limited marine natural products for clinical use.

Experimental

General experimental procedures. Negative mode FAB mass spectra were measured on a JEOL JMS-700 MSStation mass spectrometer using 3-nitrobenzyl alcohol (NBA) as matrix. Optical rotation was determined with a Jasco DIP-1000 digital polarimeter in MeOH. UV spectra were recorded on a Jasco V-530 UV/Vis spectrophotometer in MeOH. NMR spectra were recorded on a JEOL JMN-A500 NMR spectrometer at 500 MHz for 1H and at 125 MHz for 13C in D2O + CD3OD (6:1) at 20 °C. Chemical shifts in the 1H and 13C NMR spectra were referenced to the solvent peaks: δH 3.30 and δC 49.0 for CD3OD.

Extraction of metagenomic DNA from tidal-flat sediment. One liter of tidal-flat sediment collected at Ariake Sea (Amakusa Islands and Uto peninsula, August 2009) was vigorously shaken with an equal volume of sterilized sea water. The flasks were kept for 30 min to settle the sediment, and then the supernatant was transferred to centrifuge bottles. After slow centrifugation (1,000 rpm for 5 min) to remove sand, the supernatant was re-centrifuged at 5,000 rpm for 5 min to afford cell pellet. This was suspended in total 20 mL of TES buffer (20 mM Tris–HCl pH 8.0, 100 mM EDTA pH 8.0, 50 mM NaCl, and 25% sucrose), and then centrifuged at 8,000 rpm for 10 min, followed by re-suspended in 18 mL of TES buffer. A 900 μL of 10% SDS and 90 μL of protease K (100 mg/mL) were added to the suspension, and this was incubated at 37 °C for 1 h. After the addition of 3 mL of 5 M NaCl and 2.4 mL of 10% cetyltrimethylammonium bromide (CTAB) in 0.7 M NaCl, the cell lysate was heated at 65 °C for 10 min. An equal volume of a combined solvent of phenol, chloroform, and isooamyl alcohol (25:24:1) was added to the cell lysate. This was mixed gently, and then centrifuged at 8,000 rpm for 10 min. The recovered upper layer was washed with chloroform and isooamyl alcohol (24:1). Followed by the crude metagenomic DNA was concentrated by precipitation from the aqueous layer with the addition of 0.6 volume of isopropanol.

Purification of metagenomic DNA and library construction. Crude metagenomic DNA dissolved in TE buffer was size-fractionated by agarose gel electrophoresis (1% low melting point agarose gel, 30 V for 15 h). The agarose gel containing DNA above 23 kbp was cut off, digested by Sau3AI to 5–10 kbp, gel purified, ligated into pHY300PLK vector, and that enabled constitutive and large-volume production of vibrioferrin in a regular medium containing 20 mg/mL of chloramphenicol. The amplified DNA was ligated into pEPI300 fosmid vector (Epicentre) and then transfected into Escherichia coli EPI300-T1R (Epicentre) and plated on LB agar plates containing chloramphenicol as selection marker following the protocol provided by the manufacturer.

Siderophore assay. Sterilized chrome azurol S (CAS) stock solution (0.1 mM Fe(C1), 1.0 mM HCl, 1.0 mM CAS, 2.0 mM CTAB) was mixed with melted LB agar (1:9) to make CAS assay plates. Metagenomic clones or paper discs (6 mm in diameter) absorbing the test samples were put on CAS assay plates, and kept overnight at 37 °C and room temperature respectively.

Cloning of the vibrioferrin biosynthetic gene cluster. Fosmid DNA prepared from siderophore producing clone MG93-1 was partially digested by Sau3AI to 5–10 kb, gel purified, ligated into pH300PK plasmid vector (Takara, Shiga, Japan), and then transformed into NEB10β competent cell (New England Biolabs, Ipswich, MA). Siderophore producing subclone MG93-1 containing an approximately 8-kbp insert of DNA was chosen for DNA sequencing by the shotgun and primer walking methods (DNA sequencing was performed by Hokkaido System Science (Sapporo, Japan), and oligo DNAs were purchased from Invitrogen (Carlsbad, CA)). DNA analysis, such as contig construction and open reading frame analysis, was done by Vector NTI (Invitrogen). A homology search was performed by NCBI BLAST.

DNA sequence of vibrioferrin biosynthetic gene cluster. Accession Code: DNA Data Bank of Japan (DDBJ); AB585998, vibrioferrin biosynthetic gene cluster from the marine environmental sample.

Production and isolation of vibrioferrin. Siderophore-producing clone MG93-1 was cultured in LB medium containing 20 μg/mL of tetracycline (3.2 L) at 30 °C for 2 d. The culture-clarified culture broth was acidified to pH 2.0 by the addition of trifluoroacetic acid, and then subjected to the solid phase extraction with C18 resin. The active compound, eluted by 30% aqueous MeOH from the resin, was further fractionated by C18 reversed phase column chromatography by
a gradient elution system from pure water to 30% aqueous MeOH. The obtained active fraction was subjected to size exclusion chromatography using Sephadex G-10 as a solid support and water as a mobile phase. The active substance was eluted around the middle of the fractions, and it was purified by reversed phase HPLC by Develosil C30 (Nomura Chemical, Aichi, Japan) with gradient elution from 0% to 30% aqueous MeOH containing 0.05% TFA to afford pure vibrioferin (296.3 mg).

Clear oil; $\frac{1}{2}$C$_{11}$C$_{13}$8$^2\theta + 2.43$ (c 1.0, MeOH); UV (MeOH) No $\lambda_{\text{max}}$ above 220 nm; $^1$H and $^{13}$C NMR, see Table 1; negative mode HR-FABMS (NBA) $m/z$: Calcd. for C$_{16}$H$_{21}$N$_2$O$_{12}$: 433.1094, Found: 433.1100.

Acknowledgment

This project was supported by JSPS Grants-in-Aid for Scientific Research (20810022) and JST Special Coordination Funds for Promoting Science and Technology.

References