

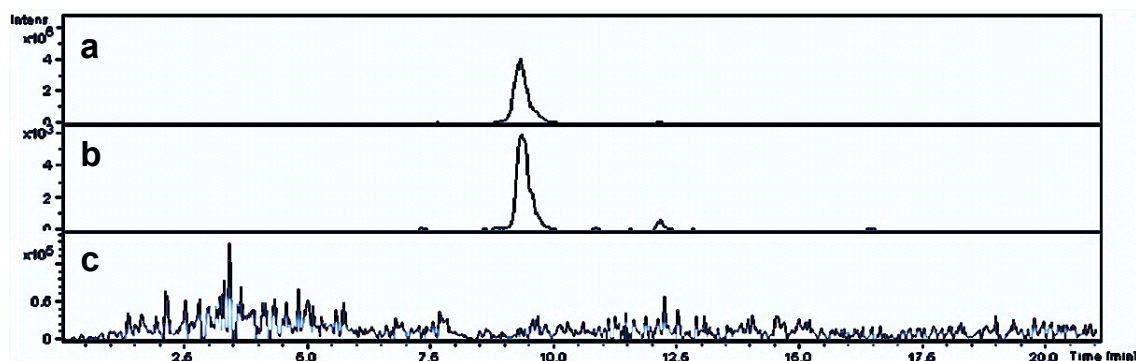
*Supporting information*

**Indole Antibiotics Production Induced by Exogenous Gene  
Derived from Sponge Metagenome**

**Yuya Takeshige<sup>a</sup>, Yoko Egami<sup>a</sup>, Toshiyuki Wakimoto,<sup>a\*</sup> Ikuro Abe<sup>a\*</sup>**

<sup>a</sup> *Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo,  
Bunkyo-ku, Tokyo 113-0033, Japan. Fax: +81 3 5841 4744; Tel: +81 3 5841 4741;  
E-mail: abei@mol.f.u-tokyo.ac.jp*

## LC-MS analysis for compound 2



**Figure S1. Extracted ion chromatogram (EIC) for m/z 364**

a) 2,2-di(3-indolyl)-3-indolone (**2**) as a standard, b) Culture solution of *E. coli* harboring ORF 25, c) Culture solution of *E. coli*'s IMPD overexpression

LC-MS analysis was performed on ODS column (COSMOSIL 5C<sub>18</sub> MS-II, 2.0 x 75 mm) with a mixture of H<sub>2</sub>O containing 0.1% acetic acid and MeCN: 0-5 min, 30% MeCN; 5-20 min, 30-100% MeCN. 0.2 mL/min. Positive ESI mode.

**Table S1. <sup>1</sup>H NMR (500 MHz) data for cpd 1 in CDCl<sub>3</sub>**

Position	$\delta$ (ppm)	Intensity	Multiplicity
1, 1', 1''-NH	7.96	3H	br. s
2, 2', 2''	6.94	3H	s
4, 4', 4''	7.36	3H	d ( $J = 8$ Hz)
5, 5', 5''	7.19	3H	t ( $J = 8$ Hz)
6, 6', 6''	7.09	3H	t ( $J = 8$ Hz)
7, 7', 7''	7.62	3H	d ( $J = 8$ Hz)

**Table S2. <sup>1</sup>H NMR (500 MHz) and <sup>13</sup>C NMR (125 MHz) data for cpd 2 in CDCl<sub>3</sub>**

Position	<sup>13</sup> C	<sup>1</sup> H	Intensity	Multiplicity
	$\delta$ (ppm)	$\delta$ (ppm)		
indoxyl				
1-NH	-	8.09	1H	br.s
2	68.2	-		
3	201.5	-		
3a	120.8	-		
4	125.5	7.73	1H	d ( $J = 8$ Hz)
5	119.4	6.89	1H	t ( $J = 8$ Hz)
6	137.6	7.51	1H	t ( $J = 8$ Hz)
7	112.9	6.91	1H	d ( $J = 8$ Hz)
7a	160.3	-		
indole				
1-NH	-	8.10	2H	br. s
2', 2''	124.1	6.94	2H	s
3', 3''	115.1			
3'a, 3''a	125.7	-		
4', 4''	120.4	7.36	2H	d ( $J = 8$ Hz)
5', 5''	119.9	7.19	2H	t ( $J = 8$ Hz)
6', 6''	122.3	7.09	2H	t ( $J = 8$ Hz)
7', 7''	111.4	7.62	2H	d ( $J = 8$ Hz)
7'a, 7''a	137.0	-		

**Table S3. Proposed ORFs in pDC112 (40.639kb)**

ORF	Size (aa)	Predicted function	Identity (%)
1	159	multi-sensor signal transduction histidine kinase	43
2	492	diguanylate cyclase [ <i>Nitrosococcus halophilus</i> Nc4]	29
3	552	multi-sensor hybrid histidine kinase	45
4	195	putative transmembrane protein [ <i>NC10 bacterium</i> ]	39
5	560	TonB-dependent receptor, plug [ <i>NC10 bacterium</i> ]	49
6	971	conserved hypothetical protein, membrane [ <i>Candidatus Poribacteria</i> sp. WGA-A3]	48
7	321	serine/threonine kinase [ <i>Streptomyces griseoaurantiacus</i> M045]	31
8	583	unnamed protein product [ <i>Desulfobacterium autotrophicum</i> HRM2]	41
9	405	amidohydrolase [ <i>Sebaldella termitidis</i> ATCC 33386]	39
10	200	conserved hypothetical protein [ <i>Oscillatoria</i> sp. PCC 6506]	52
11	464	hypothetical protein NIDE3054 [ <i>Candidatus Nitrospira defluvii</i> ]	51
12	236	proteasome subunit alpha [ <i>Candidatus Nitrospira defluvii</i> ]	59
13	226	proteasome subunit beta [ <i>Candidatus Nitrospira defluvii</i> ]	59
14	80	unnamed protein product [ <i>Streptomyces scabiei</i> 87.22]	54
15	491	putative proteasome component [ <i>NC10 bacterium</i> 'Dutch sediment']	54
16	611	ATPase [ <i>Streptomyces</i> sp. C]	58
17	261	cytochrome c-type biogenesis protein CcsB [ <i>Geobacter</i> sp. FRC-32]	37
18	420	glutamyl-tRNA reductase [ <i>Thermodesulfator indicus</i> DSM 15286]	49
19	322	hemC-porphobilinogen deaminase [ <i>endosymbiont of Riftia pachyptila</i> (vent Ph05)]	49
20	322	hemB-Porphobilinogen synthase [ <i>Thermodesulfator indicus</i> DSM 15286]	65
21	155	RsmD family RNA methyltransferase [ <i>Eubacterium infirmum</i> F0142]	35
22	355	tRNA-specific 2-thiouridylase MnmA [ <i>Geobacter</i> sp. FRC-32]	54
23	155	cytoplasmic protein[ <i>Syntrophus aciditrophicus</i> SB]	40
24	638	DNA mismatch repair protein	39
25	499	inosine-5'-monophosphate dehydrogenase[ <i>Deferribacter desulfuricans</i> SSM1]	67
26	1152	DNA polymerase III subunit alpha[ <i>Geobacter sulfurreducens</i> PCA]	52
27	236	transporter, MotA/TolQ/ExbB proton channel family [marine gamma proteobacterium]	40
28	145	Biopolymer transport protein ExbD/TolR[ <i>Moritella</i> sp. PE36]	36
29	192	N-acetyltransferase GCN5	40
30	436	group 1 glycosyl transferase	48
31	214	beta-lactamase domain protein [ <i>Arthrospira maxima</i> CS-328]	39

**Table S4. Primer list for subcloning**

name	primer sequence
Fragment 1- Fw	ATATGATCACCATGATTACGCCAAGCTAT
Fragment 1- Re	ACGTGATCAAAACTACCGCATACTCATTT
Fragment 2- Fw	ATATGATCACATCGAAATCCCGTGCGATA
Fragment 2- Re	ATATGATCATCAGAGAAAACATGCCTCGC
Fragment 3- Fw	ATATGATCATATATATGCCGACACCAGCG
Fragment 3- Re	ATATGATCAGCCGGGGTATTGAAAAACAG
Fragment 4- Fw	ATATGATCACATTGGCCACCATGGGATTT
Fragment 4- Re	ATATGATCAATACTTGGGCTCGATCTCTA
Fragment 5- Fw	ATATGATCAATAAGGAAGCGTGATGAGTG
Fragment 5- Re	ATATGATCATTTTCAGTGAGAGGTGCCTAA

**Table S5. Primer list for *E. coli* IMPD overexpression**

name	primer sequence
<i>E. coli</i> IMPD-Fw	AATTGGCCATATGCAATCGGTTACGCTCTG
<i>E. coli</i> IMPD-Re	ATAGAATTCTCAGGAGCCCAGACGGTAGTT

**Table S6. Primer list for *Pichia pastoris* expression**

name	primer sequence
<i>P. pastoris</i> -IMPD-Fw	GACTAGTATTCGGGCACAACGTTGTCA
<i>P. pastoris</i> -IMPD-Re	TATGAATTCTTAGTCCATCTGATAATTCGG

## Sequence data for inosine-5'-monophosphate dehydrogenase homolog

The sequence data were analyzed using NCBI BLAST and deposited at DDBJ/EMBL/GenBank (pDC112; accession no. LC027440).

### Nucleotide sequence

ATGTTGGAAGCAACCATACCTATGGGTCTCACGTTTGATGACGTCCTACTCG  
TTCCGGCCCGCTCGGATATCGTGCCGCGGACACGGATGTCTCCACCATTCT  
CACCCATCACATTCCGATGACCATTCCCCTCATCAGTTCCGCTATGGACACC  
GTGACTGAGAGTAGCTTGGCCATTGCCCTGGCCCAGGAGGGCGGCATCGGG  
GTGATTCACAAAACCTGTCCATCGAGGCCCAAGCTGCTGAGGTCGACAAA  
GTGAAACGCTCGGAGAGCGGCATGATTGTGACCCCATACCATGTGCGCCG  
CATCAAACCGTGCAAGACGCCCTCGACGTGATGGCGCGCTACCGCATCTCC  
GGCATTCCGATCACCGAAGGGTCCAAGCTCGTGGGCATTTTGACCAATCGC  
GATCTTCGCTTTGTCAACGACACCGCACAGCTCATTTGCGCCCTGATGACCA  
AAGATAATTTGGTCACCGTGTCCGAGGGCACGAGCCTGGACGAGGCCCAGC  
GCTTGCTCCACGCCACC GCATTGAAAACTTCTGGTGGTTCGACGAATTGTT  
TCACCTCAAAGGGCTCATCACCATCAAAGATATTGAAAAGCGGGCGCAAATA  
CCCGTTGGCCTGCAAAGACGACCTCGGGCGCTTGCGGGTGGGCGCGGCCAT  
TGGGGTGACCCCGATCGCCTCGAGCGCTTGCAAGA ACTGATGCGCTGCGG  
GGTCGATGTGGTGGTGGTTCGACAGTGCACACGGCCACTCGACCAATGTGTT  
GGCAACCATTGAAGAGGTGAAGAAAACGGCGCCGGATCTCCAAGTGATTGG  
CGGCAATGTGGCCACTGCGGGCCGCCACCCGCGACACGATTAACGCGGGGGC  
CGACGCGGTCAAAGTGGGCATTGGACCTGCCTCCATTTGTACCACTCGGGTG  
GTGGCCGGCACCGGCGTGCCGCAGCTCACGGCCATTGCGGAATGCGCCCAG  
GTGGGCGAGCAGTACGGCATTCCGATCATTGCCGATGGCGGGGTGAAGTAC  
TCCGGCGATATTGTTAAAGCGTTGGCGGGCGGGGGCGCACTCGGTAATGATT  
GGCAGTCTGTTTGCCGGGACGGAAGAGAGTCCGGGCGAGCGCGTGCTGTAT  
CAAGGCCGGAGTTACAAAGTGTACCGGGGCATGGGATCGCTCGGGGCGATG  
GCCGGGGGCCACGGCGACCGCTACTTCCAAGGCGAAGAGCGAGAGCTGAG  
CAAGCTCGTGCCGGAAGGCGTCGAGGGCCGCATTCCGTTCAAGGGGGTGCT  
GGCTGACGTGGTCTACCAGCTGGTTCGGCGGGTTGCGAGTGGGCATGGGCTA  
CTGCGGTTGCAAGAGTATCGACGCCTTGCGCCGCGAGGCGCGTTTTGTGCA

AATTACCAATGCCGGGTTCCGGGAAAGCCATGTTACACGACGTGATGGTGAC  
CCAGGAGGCCCCGAATTATCAGATGGACTAA

Amino acid sequence

MLEATIPMGLTFDDVLLVPARSDIVPRDTDVSTILTHHIPMTIPLISSAMDTVTES  
SLAIALAQEGGIGVIHKNLSIEAQAAEVDKVKRSESGMIVDPITMSPHQTVQDAL  
DVMARYRISGIPITEGSKLVGILTNRDLRFVNDTAQLISALMTKDNLVTVSEGTS  
LDEAQRLHAHRIEKLLVDELFLKGLITIKDIEKRRKYPLACKDDLGRRLRVG  
AAIGVTPDRLERLQELMRCGVDVVVVDVAHGHSTNVLATIEEVKKTAPDLQVI  
GGNVATAAATRDTINAGADAVKVGIGPASICTTRVVAGTGVPQLTAAECAQV  
GEQYGIPIIADGGVKYSGDIVKALAAGAHSVMIGSLFAGTEESPGERVLYQGRS  
YKVYRGMGSLGAMAGGHGDRYFQGEERELSKLVPEGVEGRIPFKGVLADVY  
QLVGGLRVGMGYCGCKSIDALRREARFVQITNAGFRESHVHDMVTQEAPNY  
QMD