Dual regulation of a polyethylene glycol degradative operon by AraC-type and GalR-type regulators in *Sphingopyxis macrogoltabida* strain 103

Jittima Charoenpanich, Akio Tani, Naoko Moriwaki, Kazuhide Kimbara and Fusako Kawai

The genes for polyethylene glycol (PEG) catabolism (*pegB*, *C*, *D*, *A* and *E*) in *Sphingopyxis macrogoltabida* strain 103 were shown to form a PEG-inducible operon. The *pegR* gene, encoding an AraC-type regulator in the downstream area of the operon, is transcribed in the reverse direction. The transcription start sites of the operon were mapped, and three putative σ70-type promoter sites were identified in the *pegB*, *pegA* and *pegR* promoters. A promoter activity assay showed that the *pegB* promoter was induced by PEG and oligomeric ethylene glycols, whereas the *pegA* and *pegR* promoters were induced by PEG. Deletion analysis of the *pegB* promoter indicated that the region containing the activator-binding motif of an AraC/XylS-type regulator was required for transcription of the *pegBCDAE* operon. Gel retardation assays demonstrated the specific binding of PegR to the *pegB* promoter. Transcriptional fusion studies of *pegR* with *pegA* and *pegB* promoters suggested that PegR regulates the expression of the *pegBCDAE* operon positively through its binding to the *pegB* promoter, but PegR does not bind to the *pegA* promoter. Two specific binding proteins for the *pegA* promoter were purified and identified as a GalR-type regulator and an H2A histone fragment (histone-like protein, HU). The binding motif of a GalR/LacI-type regulator was found in the *pegA* and *pegR* promoters. These results suggested the dual regulation of the *pegBCDAE* operon through the *pegB* promoter by an AraC-type regulator, PegR (PEG-independent), and through the *pegA* and *pegR* promoters by a GalR/LacI-type regulator together with HU (PEG-dependent).

**INTRODUCTION**

Polyethylene glycols (PEGs) are man-made polymers that have been widely used as commodity chemicals in industrial products such as pharmaceuticals, cosmetics and lubricants and as raw materials in the synthesis of surfactants and polyurethanes. Since they are water-soluble, they finally show up in natural streams or wastewater systems as dilute solution, which can neither be recycled nor incinerated. Because of their random coil formation (Cox, 1978), the alcohol groups at the termini are randomly distributed in the space taken up by the macromolecule, making it difficult to attack the termini with enzymes. Moreover, aliphatic ether bonds are chemically very stable; therefore, the compounds have long persistence (Bailey & Koleske, 1976). Hence, microbiological degradation appears to be the only means to decompose this group of polymers.

Sphingomonads include *Sphingomonas*, *Sphingopyxis*, *Sphingobium* and *Novosphingobium* (Pal et al., 2006), which are known to utilize several types of xenobiotic polymers as well as PEG (Kawai, 1999). PEG metabolism has been reported in both axenic, e.g. *Sphingomonas macrogoltabidus* (Kawai & Enokibara, 1996), and mixed culture, e.g. *Sphingomonas terrae* with a *Rhizobium* species (Kawai & Yamanaka, 1986). These siphingomonads were reclassified as *Sphingopyxis macrogoltabida* and *Sphingopyxis terrae*, respectively (Takeuchi et al., 2001). PEG is aerobically degraded by successive oxidation of a terminal alcohol group to an aldehyde and a carboxyl group. An ether bond adjacent to a carboxymethyl group is split to yield glyoxylic acid, resulting in depolymerization by one glycol unit (Enokibara & Kawai, 1997; Kawai, 2002). Two enzymes...
involved in the first and second steps of PEG degradation were characterized as PEG dehydrogenase (PEG-DH) (Sugimoto et al., 2001) and PEG-aldehyde dehydrogenase (PEGAL-DH) (Ohta et al., 2005) from *S. terrae*. PEG-DH is a novel flavoprotein alcohol dehydrogenase containing FAD as a cofactor. PEGAL-DH is a Ca$^{2+}$-activated nictotinoprotein aldehyde dehydrogenase, an enzyme type not previously reported. An ether-bond splitting enzyme was suggested to be hydroxyacid dehydrogenase in *S. terrae* (Enokibara & Kawai, 1997) or superoxide dismutase in the Gram-positive PEG degrader *Pseudonocardia* sp. strain K-1 (Yamashita et al., 2004). The gene encoding the PEG-DH, pegA, was cloned from *S. terrae* (Sugimoto et al., 2001). The flanking regions of pegA were also sequenced from three sphingomonads, *S. terrae*, *S. macrogoltabida* strain 103 and strain 203 in our laboratory (accession numbers AB239603, AB196775 and AB239080, respectively).

In this paper we demonstrate the operonic nature of the genes involved in the flanking region of *pegA* in *S. macrogoltabida* strain 103, the regulation of the operon by *pegR*, and involvement of another transcriptional regulator specific for *pegA* and *pegR*.

**METHODS**

**Bacterial strains, plasmids, media and chemicals.** The bacterial strains and plasmids used in this study are listed as supplementary material with the online version of this paper (Table S1). The wild-type *S. macrogoltabida* strain 103 (Takeuchi et al., 2001) was routinely subcultured on nutrient broth (NB) (Atlas, 1995) and cultivated at 28 °C in glucose medium, PEG (molecular mass 4000, PEG 4000) medium (Kawai et al., 1985) or W minimal salts medium (WM) (Kimbara et al., 1989) containing an appropriate carbon source at the final concentration of 0-5% (v/v) or (v/v). For routine cloning purposes, *Escherichia coli* DH5α was grown in Luria–Bertani (LB) medium (Sambrook et al., 1989) at 37 °C for 12–15 h as a host. *E. coli* K-12 MC4100 was used in the promoter–lacZ fusion analysis experiment. For promoter activity assays, *E. coli* K-12 MC4100 was cultivated in LB diluted 1/5, with and without 0-5% (v/v) or (v/v) triethylene glycol (TEG) or PEG 4000. When necessary, cultures were supplemented with 50 μg ampicillin ml$^{-1}$, 1 mM IPTG and 1 mM X-Gal.

**DNA manipulations, PCR and nucleotide sequence analysis.** Transformations of *E. coli*, restrictions and ligation were carried out using standard procedures (Sambrook et al., 1989). Recombinant plasmids were introduced into *S. macrogoltabida* strain 103 by electroporation using a modification of the method of Diver et al. (1990). The Gene Pulser (Bio-Rad) was used under the following conditions: 0.1 cm cuvette, 500 μF, 1.25 kV, and a pulse time of 4–7–5 ms. A 1 ml aliquot of NB was added immediately after the electric pulse. The cells were incubated at 28 °C for 1 h prior to plating onto NB diluted 1/3 and containing ampicillin. Total DNA of *S. macrogoltabida* strain 103 was extracted by the method of Marmur (1961). Restriction enzymes and other DNA-modifying enzymes were purchased from TOYOBO or Takara bio and used as specified by the manufacturers. Plasmid DNA from *E. coli* was prepared with a Wizard Plus SV miniprep DNA purification system (Promega) whereas the method of Kado & Liu (1981) was used for *S. macrogoltabida* strain 103. DNA fragments were purified from agarose by using a Mag-Extractor DNA purification kit (TOYOBO) or a GenElute minus EtBr spin column (Sigma). PCR mixtures (25 μl) contained 0-4 pmol of each primer μl$^{-1}$, 2 mM MgCl$_2$, 0-25 mM of each deoxynucleotide, 1 × Ex Tag DNA polymerase buffer, and 0-5 U Ex Taq DNA polymerase (Takara). All PCR fragments were cloned into pCR-2.1 TOPO vector (Invitrogen) and their sequences were checked by sequencing with an ABI3100 Genetic Analyser and a BigDye Cycle Sequencing kit version 1.1 (Applied Biosystems). Primers used for PCRs and sequencing reactions are listed as supplementary material with the online version of this paper (Table S2).

**Preparation of total RNA and RT-PCR.** *S. macrogoltabida* strain 103 was grown at 28 °C on WM medium containing ethylene glycol (EG) or PEG 4000 to an OD$_{600}$ of approximately 0-3. Total RNA was isolated using ISOGEN (Nippon Gene). The RNA samples were treated with DNase I (Invitrogen). RT-PCR was performed with a One-Step RT-PCR kit (Qiagen) according to the supplier’s instructions. Negative control experiments were done by omitting the reverse transcription step. Positive control experiments were done with genomic DNA as a template. The sequences and positions of primers used for each PCR amplification are indicated in Table S2 and Fig. 1.

**RNA dot-blot hybridization.** Three micrograms of total RNA from each sample was spotted and heat-fixed on nylon membranes (Hybond-N+, Amersham Biosciences) at 80 °C for 120 min. Probes were generated by amplifying the DNA fragments of the ORFs of each gene with appropriate primers, listed in Table S2 and Fig. 1. The probes were labelled as described in the manual for the DIG system (Roche). The conditions used for pre-hybridization, hybridization, washing and detection followed the instructions recommended by the manufacturer.

**Construction of promoter–lacZ fusion plasmids and deletion analysis of the pegB promoters.** To generate the promoter–lacZ fusion plasmids, the upstream regions of the *pegA*, pegB and pegR genes were amplified by PCR using the primers illustrated in Fig. 1 and listed in Table S2. The PCR products were digested with *Hind*III and cloned in the *Kpn*I–*Hind*III site upstream of the promoterless lacZ gene in the vector pQF50 (Farinha & Kropinski, 1990). Plasmids pQF50pB1, pQF50pA and pQF50pR were constructed in this way. Insertion was confirmed by sequencing. Deletion analysis of pQF50pB1 was done by the same procedure as described above by fixing either primer Act-B1 or Act-B2. Primer names and positions are illustrated in Fig. 2 with the constructed vector pQF50pBX (X = 1–10). To test the possible interaction between PegR and the promoter of pegB or pegA in vivo, pegR was introduced into pQF50pB10 and pQF50pA. The primer R1, which contains the promoter region of *pegR*, and the primer R2, which contains the trpA transcription terminator sequence and the 3′-region of *pegR* were used to amplify *pegR*. The resulting fragment contained the trpA terminator downstream of *pegR*, to prevent readthrough from *pegR* to the reporter gene. The fragment was digested with *Bgl*II and *Xba*I and ligated into a *Bgl*II–*Xba*I site, upstream of the promoter–lacZ fragment in pQF50pB10 or pQF50pA (Table S1). The resulting plasmids were transferred into either *E. coli* K-12 MC4100 or *S. macrogoltabida* strain 103 as host for promoter activity assay.

**β-Galactosidase activity assay.** Cultures harbouring fusion plasmids were grown to exponential phase in different media. β-Galactosidase activity was measured as described by Miller (1972); enzyme activity was expressed in Miller units. The data presented were obtained from three independent experiments, each of which was performed in duplicate.

**Identification of transcription initiation points of pegB, pegA and pegR.** We employed 5′-rapid amplification of cDNA ends (5′-RACE) (Frohman et al., 1988) to determine the transcription initiation
The transcription initiation point of **pegA** was also determined by primer extension analysis. A reaction mixture (20 μl) containing 2 pmol of an FITC-labelled primer, FITC-A, μl⁻¹ (Table S2), 2 μg RNA obtained from PEG 4000-grown cells, 10 mM each dATP, dCTP, dGTP, dTTP (Roche) and 7-deaza-dGTP (New England Biolabs) was incubated at 75 °C for 5 min and was allowed to cool on ice. The reverse transcription reaction was started by addition of 8 μl of a mixture containing reaction buffer (50 mM Tris/HCl pH 8.3, 75 mM KCl and 3 mM MgCl₂), 10 mM DTT, 20 U RNase inhibitor (RNaseOUT, Invitrogen) and 100 U Superscript II reverse transcriptase (Invitrogen). The reaction mixture was incubated at 42 °C for 60 min and 70 °C for 15 min. The excess RNA was digested with 2 μl 10 mg RNase 1 ml⁻¹ at 37 °C for 30 min. The cDNA was recovered by ethanol precipitation and applied to a standard sequencing gel of a DSQ-2000L DNA sequencer (Shimadzu). An RNA sample obtained from glucose-grown cells was used as a negative control. Sequence ladders using the same primer and pCR-PpegA (described below) as a template were prepared with the Thermo Sequenase Fluorescent-Labelled Primer Cycle Sequencing kit (Amersham Pharmacia). pCR-PpegA was constructed by amplification of the **pegA** region with primers 9-F and C-pegA (Tables S1 and S2) and cloning into the pCR-2.1 TOPO vector and sequenced as described above.

The transcription initiation point of **pegA** was also determined by primer extension analysis. A reaction mixture (20 μl) containing 2 pmol of an FITC-labelled primer, FITC-A, μl⁻¹ (Table S2), 2 μg RNA obtained from PEG 4000-grown cells, 10 mM each dATP, dCTP, dGTP, dTTP (Roche) and 7-deaza-dGTP (New England Biolabs) was incubated at 75 °C for 5 min and was allowed to cool on ice. The reverse transcription reaction was started by addition of 8 μl of a mixture containing reaction buffer (50 mM Tris/HCl pH 8.3, 75 mM KCl and 3 mM MgCl₂), 10 mM DTT, 20 U RNase inhibitor (RNaseOUT, Invitrogen) and 100 U Superscript II reverse transcriptase (Invitrogen). The reaction mixture was incubated at 42 °C for 60 min and 70 °C for 15 min. The excess RNA was digested with 2 μl 10 mg RNase 1 ml⁻¹ at 37 °C for 30 min. The cDNA was recovered by ethanol precipitation and applied to a standard sequencing gel of a DSQ-2000L DNA sequencer (Shimadzu). An RNA sample obtained from glucose-grown cells was used as a negative control. Sequence ladders using the same primer and pCR-PpegA (described below) as a template were prepared with the Thermo Sequenase Fluorescent-Labelled Primer Cycle Sequencing kit (Amersham Pharmacia). pCR-PpegA was constructed by amplification of the **pegA** region with primers 9-F and C-pegA (Tables S1 and S2) and cloning into the pCR-2.1 TOPO vector and sequenced as described above.

**Cloning, expression and purification of the recombinant PegR.** A 1.1 kb DNA fragment of the **pegR** ORF was generated by PCR using primers R3 and R4 (Table S2). The PCR product was cloned into the EcoRI-HindIII site of pCold-1 (Takara), generating plasmid pCold-pegR. E. coli DH5α carrying pCold-pegR was grown in 100 ml LB medium containing ampicillin at 37 °C to an OD₆₀₀ of approximately 0.3. Then the culture temperature was shifted to 15 °C, and IPTG was added. After 24 h induction at 15 °C, the cells were harvested, resuspended in 5 ml cold lysis buffer (50 mM
NaH₂PO₄·7H₂O, 300 mM NaCl, 5 mM imidazole and 0.15 mM PMSF at pH 8.0) and then broken by sonication (ultrasonic disruptor UD-200, Tomy Seiko) on ice with 5 kHz, 10 s pulse on/off for 10 min. The sample was centrifuged at 10 000 g for 15 min at 4°C and the resulting supernatant was designated cell-free extract. The recombinant 6xHis-tagged PegR was purified with Ni²⁺-NTA-activated resin (Qiagen) as recommended by the manufacturer. Purity of the PegR was confirmed by SDS-PAGE (Laemmli, 1970). The purified PegR was dialysed against lysis buffer before the gel retardation assay. Protein concentration was measured with the Bio-Rad protein assay kit using BSA as a standard.

Gel retardation assay. A 268 bp DNA fragment of the pegB promoter in pQF50pB10 was recovered by digesting the plasmid with KpnI and HindIII (Table S1) and was labelled with DIG as described in the manual for DIG-gel shift kit (Roche). The labelled pegB promoter (0-05 pmol) was incubated with various amounts of recombinant PegR (0-25 pmol) in a reaction mixture as recommended by the manufacturer. After 15 min incubation at 28°C, 4 μl loading dye (0-1%, w/v, xylene cyanol in 40%, v/v, glycerol, 22-25 mM Tris, 20 mM boric acid and 0.5 mM EDTA, pH 8-0) was added to stop the reactions and the samples were loaded on a 30 min pre-run 5% (w/v) polyacrylamide/0-1% (w/v) bisacrylamide gel under non-denaturing conditions. Negative control reactions were run with 0-05 pmol DIG-labelled 142 bp pegD intergenic region (described below) incubated with or without 0-25 pmol recombinant PegR. The pegD intergenic region was amplified by PCR with primers pD1 and pD2 (Table S2). The electrophoresis was carried out at 4°C in 0-5 x TBE buffer (44-5 mM Tris, 44-5 mM boric acid and 1 mM EDTA, pH 8-0) at a constant 5 V cm⁻1 for 45 min. After electrophoresis, nucleotides were transferred to a positively charged nylon membrane (Nippon Gene) by electroblotting at 300 mA for 30 min at 4°C. Nucleotides were fixed onto the membrane by heating at 120°C for 30 min. A chemiluminescent signal system was used for washing and detection following the instructions recommended by the DIG system manufacturer (Roche).

Purification of the pegA promoter-binding protein. S. macrogoltabida strain 103 cells grown in PEG 4000 medium (OD₆₀₀ 0.6-0.8) were harvested by centrifugation (5000 g, 10 min, 4°C), washed with 0.9% (w/v) NaCl, and resuspended in 25 mM phosphate buffer (pH 8.0) containing 1 mM PMSF and 1 mM DTT. The cells were then disrupted by sonication as described above. Cell debris was removed by centrifugation at 10 000 g for 30 min at 4°C. Cell-free extract was collected by ultracentrifugation at 40 000 g for 45 min at 4°C. The concanameric oligonucleotides containing the

Fig. 2. Deletion analysis of the pegB promoter region. A series of lacZ reporter plasmids containing various deletions of the pegB promoter were constructed. Numbers adjacent to the constructs represent nucleotide positions relative to the transcription initiation site of the pegB promoter. Thin arrows indicate the primer positions used for this study. β-Galactosidase activities were measured with S. macrogoltabida strain 103 grown on NB medium and E. coli K-12 MC4100 grown on 0.2 x LB medium. Both strains carried the plasmids indicated in the figure.
The six genes were expressed in PEG medium, but not in EG medium, showing that they were inducible by PEG 4000 but not by EG (Fig. 1a). The transcripts of each gene were also detected by RT-PCR using gene-specific primers (Fig. 1b). RT-PCRs gave amplified products with the same size as the normal PCR products using genomic DNA as a template. Negative controls prepared by omitting the reverse-transcription step gave no detectable band, showing no detectable contamination of DNA in the RNA samples. The RT-PCR products were only found in the RNA samples from PEG 4000-grown cells. Next, we designed primers to amplify the intergenic regions of the adjacent genes, using the RT-PCR technique. Four intergenic regions between adjacent genes (pegB, C, D, A and E) were amplified (Fig. 1c) from the RNA samples of the PEG 4000-grown cells. These results suggest that five of the genes (pegB, C, D, A and E) form an operon. Comparison of cells grown on EG and PEG 4000 showed that expression of the pegBCDAE operon and pegR was induced by PEG.

**Transcriptional analysis of promoter regions for the pegBCDAE operon**

In order to study the promoter activity of the pegBCDAE operon, we cloned the upstream regions of pegB, pegA and pegR into the broad-host-range vector pQF50. In all media tested, β-galactosidase activity was detected in both *S. macrogoltabida* strain 103 and *E. coli* K-12 MC4100 without pQF50. Basal expression levels of a promoterless pQF50 showed 7–15 Miller units in *E. coli* and less than 84 Miller units in *S. macrogoltabida* strain 103 (Table 1). More than 1.9-fold levels of β-galactosidase activity were detected in the cells harbouring pQF50pB1 grown on all the media except EG. This result suggests that the transcription of the pegBCDAE operon starts from the pegB promoter, induced by oligomers of PEG 4000. Higher activity was also detected in EG and PEG 4000 compared to LB medium.

### RESULTS

**Inducibility of the peg operon**

Six DNA fragments of the genes in the cluster (pegB, C, D, A, E and R) were used as probes for RNA dot-blot hybridization analysis. The six genes were expressed in PEG medium, but not in EG medium, showing that they were inducible by PEG 4000 but not by EG (Fig. 1a). The transcripts of each gene were also detected by RT-PCR using gene-specific primers (Fig. 1b). RT-PCRs gave amplified products with the same size as the normal PCR products using genomic DNA as a template. Negative controls prepared by omitting the reverse-transcription step gave no detectable band, showing no detectable contamination of DNA in the RNA samples. The RT-PCR products were only found in the RNA samples from PEG 4000-grown cells. Next, we designed primers to amplify the intergenic regions of the adjacent genes, using the RT-PCR technique. Four intergenic regions between adjacent genes (pegB, C, D, A and E) were amplified (Fig. 1c) from the RNA samples of the PEG 4000-grown cells. These results suggest that five of the genes (pegB, C, D, A and E) form an operon. Comparison of cells grown on EG and PEG 4000 showed that expression of the pegBCDAE operon and pegR was induced by PEG.

**Table 1. Promoter activity in *S. macrogoltabida* strain 103 and *E. coli* K-12 MC4100 in different growth media**

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>WM + EG</th>
<th>WM + DEG</th>
<th>WM + TEG</th>
<th>WM + TTG</th>
<th>WM + PEG 4000</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. macrogoltabida</em> 103</td>
<td>2.5±0.2</td>
<td>3.5±0.3</td>
<td>1.6±0.3</td>
<td>1.7±0.2</td>
<td>2.7±0.1</td>
</tr>
<tr>
<td>pQF50</td>
<td>84±5</td>
<td>71±5</td>
<td>71±3</td>
<td>74±5</td>
<td>81±6</td>
</tr>
<tr>
<td>pQF50pB1</td>
<td>90±6</td>
<td>151±4</td>
<td>181±4</td>
<td>182±2</td>
<td>184±4</td>
</tr>
<tr>
<td>pQF50pA</td>
<td>88±6</td>
<td>95±7</td>
<td>92±5</td>
<td>96±7</td>
<td>142±8</td>
</tr>
<tr>
<td>pQF50pR</td>
<td>76±3</td>
<td>71±6</td>
<td>88±6</td>
<td>89±7</td>
<td>100±4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>WM + EG</th>
<th>WM + DEG</th>
<th>WM + TEG</th>
<th>WM + TTG</th>
<th>WM + PEG 4000</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> K-12 MC4100</td>
<td>1.2±0.1</td>
<td>1.8±0.1</td>
<td>3.7±0.2</td>
<td>3.7±0.2</td>
<td>3.7±0.2</td>
</tr>
<tr>
<td>pQF50</td>
<td>15±1</td>
<td>7.4±0.3</td>
<td>7.6±0.1</td>
<td>7.6±0.1</td>
<td>7.6±0.1</td>
</tr>
<tr>
<td>pQF50pB1</td>
<td>132±3</td>
<td>118±2</td>
<td>118±1</td>
<td>118±1</td>
<td>118±1</td>
</tr>
<tr>
<td>pQF50pB10</td>
<td>135±3</td>
<td>110±2</td>
<td>139±2</td>
<td>139±2</td>
<td>139±2</td>
</tr>
<tr>
<td>pQF50pB10R</td>
<td>994±11</td>
<td>776±19</td>
<td>687±19</td>
<td>687±19</td>
<td>687±19</td>
</tr>
<tr>
<td>pQF50pA</td>
<td>18±2</td>
<td>6.7±0.4</td>
<td>11±1</td>
<td>11±1</td>
<td>11±1</td>
</tr>
<tr>
<td>pQF50pAR</td>
<td>20±2</td>
<td>6.9±0.2</td>
<td>12±1</td>
<td>12±1</td>
<td>12±1</td>
</tr>
</tbody>
</table>

*0-2× LB means LB medium diluted 1/5.*
in E. coli harbouring pQF50pB1, implying a highly functional promoter even in E. coli. In strain 103, expression of the pegA and pegR promoters carried on pQF50pA and pQF50pR was induced by PEG 4000, suggesting that these PEG-responsive promoters might be regulated by PEG-responsive regulators. In E. coli harbouring pQF50pA, β-galactosidase activity could not be detected even in the presence of PEG 4000. This result suggests that a PEG-responsive regulator for pegA must exist in strain 103 but not in E. coli.

Identification of the transcription initiation sites of the pegBCDAE operon

Analysis of promoter activity showed high β-galactosidase activities with the pegB, pegA and pegR promoters, as described above. These results led us to seek transcription start sites located in these regions. The nucleotide sequence of the 5′-RACE PCR product for pegB showed that transcription starts at the G nucleotide located 90 bp upstream of the pegB translation initiation codon (see Fig. S1a, available as supplementary data with the online version of this paper). A putative −10 sequence (AGACTT) (deHaseth et al., 1998), which is separated by 13 bp from a putative −35 sequence (TTGCGG), existed upstream of this transcription start site. Furthermore, a putative RBS (GGAG) (Shine & Dalgarno, 1975) was found 14 bp upstream of the pegB translation initiation codon. DNA-binding motifs similar to those of the regulatory proteins in the AraC/XylS family (Hendrickson & Schleif, 1985; Gallegos et al., 1997) were found upstream of the pegB promoter (Fig. S1a). Primer extension analysis showed that transcription starts at 335 bp (t1) and 109 bp (t2) upstream of the pegA translation start codon, ATG (Fig. S1b, d). The major transcriptional start site (the stronger signal: t2) was positioned 225 bp downstream of the minor site (t1). Putative −10 (TGATAC) and −35 (TTCTCA) regions, separated by a 20 bp spacer, were found upstream of the minor transcriptional initiation site (t1). Upstream of the major site (t2) is a putative −10 (ATTTT) region, separated by 18 bp from a −35 (GTGACC) region. Putative RBs (GAA for pegA1 and AGGA for pegA2) (Shine & Dalgarno, 1975) were found 5 bp and 8 bp upstream of the translation initiation codon (pegA1 and pegA2, respectively). These data suggest that the transcription of pegA starts at two distal nucleotides. Moreover, upstream regions of both transcriptional start sites contain the motif sequences for binding sites of LacI family regulators (Sadler et al., 1983) and O6 and O8 of the GalR family (Majumdar & Adhya, 1987) as well as a histone-like protein (HU) binding site, hbs (ATTTTATAT) (Fig. S1b). In the pegR promoter region (Fig. S1c) transcription starts at the A nucleotide 137 bp upstream of the pegR translation initiation codon. A putative −10 sequence (CAACCTT) (deHaseth et al., 1998) is separated by 18 bp from a putative −35 sequence (GGCTCA). A putative RBS (GGA) (Shine & Dalgarno, 1975) was found 11 bp upstream of the pegR translation initiation codon. DNA-binding sites of a GalR/LacI family regulator (Sadler et al., 1983; Majumdar & Adhya, 1987) were also found upstream of the pegR promoter.

Deletion analysis of the pegB promoter

To identify the important region of the pegB promoter, deletion analysis was performed. The reporter plasmid pQF50pB1, which contained a 416 bp fragment covering the full length of the upstream region of the pegB promoter, showed β-galactosidase activity of about 398 Miller units in S. macrogoltabida strain 103 and 273 Miller units in E. coli K-12 MC4100 (Fig. 2). Deletion of the putative RBS in pQF50pB2 reduced the activity by 60% in both strains. The strains carrying pQF50pB3 and pQF50pB4 showed the same level of β-galactosidase activity as that in pQF50pB2. The activity in pQF50pB5 decreased to nearly the vector background level, indicating that the promoter activity appeared to require a region approximately 108–217 bp upstream of the pegB transcription start site. These results were confirmed by the activity in pQF50pB10. In pQF50pB6, which contains the putative RBS, full activity was detected at 374 Miller units in strain 103 and 233 Miller units in E. coli. Strain 103 and E. coli carrying pQF50pB7 had activity reduced to approximately 84% and 67%, respectively, of the level with pQF50pB6. Basal-level activities were found in strains carrying pQF50pB8 and pQF50pB9, although pQF50pB8 contains the transcriptional start site and putative −10 and −35 sequences of pegB. This might suggest that a regulator-binding site is necessary for transcription of pegB. From these results, we can conclude that the sequences between positions 108–217 bp upstream and 50–90 bp downstream of the pegB transcription initiation site are required for transcription of the pegB promoter and the pegBCDAE operon.

Interaction assay of PegR with the pegA and pegB promoters in E. coli

To assess whether the PegR protein acts as a regulator for the pegBCDAE operon, we fused pegR into the reporter plasmids pQF50pA and pQF50pB10. The introduction of pegR resulted in more than fivefold higher activity in pQF50pB10, but no obvious change in pQF50pA in all growth conditions (Table 1). These results suggest that PegR interacts with the pegB promoter but not with the pegA promoter. PegR enhances transcription of the pegB promoter, resulting in the positive expression of the pegBCDAE operon.

Purification of recombinant PegR, and gel-retardation assay

SDS-PAGE analysis of purified recombinant PegR gave a subunit molecular mass of 41 kDa (Fig. 3a). After dialysis, the purified PegR and the 268 bp DNA fragment (position −217 to +50) in pQF50pB10 were used for the gel retardation assay. As shown in Fig. 3(b), the mobility of the labelled pegB promoter was retarded by the recombinant PegR added at an amount more than equivalent to the pegB
promoter, but there was no obvious shift in the labelled pegD promoter. This result suggested that the specific binding region for PegR is within the 268 bp pegB fragment.

Purification of the specific DNA-binding-protein for the pegA promoter

The specific binding protein for the pegA promoter was purified from cell-free extracts prepared from cells grown on PEG 4000. Two purified proteins showed molecular masses of approximately 20 and 40 kDa (Fig. 3c, d). The purified proteins were digested with trypsin and analysed by MALDI-TOF MS. The peptide peak of \( m/z \) 1479 ± 0.005 from the 20 kDa protein was assigned to AKAKTNSCKVGSGTK. Seven amino acid residues (AKAKTNS) showed 85% homology with the H2A histone fragment from Rattus norvegicus (accession no. Q9Z2R1 in the MS database). The peptide of \( m/z \) 2216 ± 0.022 from the 40 kDa protein corresponded to WDYVSENTRATNPVFMMR, where eight amino acid residues (WDYVSENTR) showed 87% homology with the GalR/LacI family transcription regulator from Bacillus clausii KSM-K16 (accession no. Q5WBX5 in the MS database).

DISCUSSION

Research to date on PEG degradation in micro-organisms has mainly concentrated on characterization of the relevant enzymes (Yamanaka & Kawai, 1989; Schramn & Schink, 1991; Frings et al., 1992; Sugimoto et al., 2001; Yamashita et al., 2004; Ohta et al., 2005). The structural and regulatory features of the degradative genes have not been reported yet. Recently our laboratory has sequenced the flanking regions of pegA from three PEG-utilizing sphingomonads (accession numbers AB239603, AB196775 and AB239080). The peg catabolic cluster encoded two major enzymes (PEG-DH and PEGAL-DH) required for PEG degradation, a TonB-dependent receptor, a permease and an acyl-CoA ligase (unpublished results). Although the roles of the latter three genes have not been clarified yet, they seem to have relevance to the degradation of PEG. The primary structures of the region (pegBCDAE) from the three strains shared more than 99% identity, except that a transposase gene named pegF was inserted between pegD and pegA only in strain 203 (unpublished results). pegBCDAE encodes a putative TonB-dependent receptor, PEGAL-DH, a putative permease, PEG-DH and a putative acyl-CoA ligase,
respectively (Fig. 1). A pegR gene located downstream of pegBCDAE showed similarity to AraC-type regulators from various microorganisms. Understanding the regulatory system at a molecular level is a prerequisite for elucidation of PEG degradation in sphingomonads. In this study, we investigated the transcriptional organization and regulation of the pegBCDAE operon in *S. macrogoltabida* strain 103.

We have demonstrated the existence of a putative GalR/LacI-type regulator and a histone-like protein specific to the pegA promoter. Analysis of the intergenic sequence in the pegA promoter showed the sequence corresponding to the consensus DNA-binding motifs of these proteins (Majumdar & Adhya, 1987; Sadler et al., 1983). The GalR/LacI family regulators are well known as negative regulators of transcription of inducible genes in the coordination of catabolic, metabolic and transport operons (Weickert & Adhya, 1992). Repression requires an interaction between two operator (O_E and O_I)-bound repressors, and formation of a loop of the intervening 113 bp DNA encompassing the two promoters (Irani et al., 1983; Majumdar & Adhya, 1984). One of the sites (O_E) is located between –75 and –50 upstream of the transcription start site t_1. The other (O_I) is located between +43 and +67 downstream. The two operators are separated by more than 90 bp (Majumdar & Adhya, 1984). The interaction is facilitated by the binding of a histone-like protein HU at a specific region between the two operators, hbs (Aki & Adhya, 1997). HU has been shown to bind to 9 bp segments of AT-rich hbs containing three overlapping weak dyad symmetries, which are centred at position +6.5 (Aki & Adhya, 1997). HU is a small, basic DNA-binding protein capable of wrapping DNA, and its primary structure is highly conserved among bacterial species (Drlica & Rouviere-Yaniv, 1987). It shows similarity to eukaryotic histones and H2A fragment (Rouviere-Yaniv & Gros, 1975). HU is a heterotypic dimer having an apparent molecular mass of 20 000 as determined by gel electrophoresis in the presence of ionic detergents (Rouviere-Yaniv & Gros, 1975), which agreed well with the molecular mass of the purified pegA-binding protein (Fig. 3c). It has been reported that full derepression by the GalR regulator occurs by the binding of an inducer to the operator-bound repressor, and the inhibitory effect of the repressor on RNA polymerase is found without dissociating the repressor from DNA (Chatterjee et al., 1997). Although the putative gene has not been found around the peg cluster, the properties and positions of these consensus repression sites as well as those of hbs agreed well with the expression of pegA induced with PEG. Moreover, the consensus DNA-binding motif for GalR/LacI family regulators (Sadler et al., 1983; Majumdar & Adhya, 1987) was also found in the upstream region of pegR. The hbs was not found in this promoter region. It is reported that HU can interact with GalR and does not necessarily use its DNA-binding property (Aki et al., 1996). Thus, we can conclude that transcription of the pegA and pegR promoters was repressed by GalR/LacI family transcriptional regulators with the help of HU, and possibly derepressed by PEG 4000 as an inducer (Fig. 4).

In the pegB promoter, many putative DNA-binding sites for regulatory proteins of the AraC/XylS family such as AraC, XylS and SoxS were found. For positive regulation by AraC, the target sequences in the cognate promoters have been located adjacent to or overlapping the –35 region of the promoter (Busby & Ebright, 1994; Collado-Vides et al., 1991; Harrison, 1991). The putative AraC operator site, araI, in the pegB promoter was remote from the –35 region, but included the similar overlapping sequence, TTCCAA, with –35 regions. On the other hand, the XylS-binding site is probably represented by the motif T(C/A)CAN4TGCA, such that the exact location of the RNA polymerase binding site distal motif was between –67 and –78 (Gallegos et al., 1996). The motif corresponded to TAGAATGTGGCA (mismatched nucleotides are shown in bold) found in the

![Fig. 4. Regulatory circuit model of the peg operon. Details are given in the text.](image-url)
Transcriptional regulation of the peg operon

The Sox-box consensus sequence, AN₃GCA(C/T)N(A/G)AN₃(A/G)N₂AA(A/G)N, was also found in the pegB promoter. A putative Sox-box for upstream from the −35 element of the pegB promoter has similarity with the SoxS-binding site of ribA, encoding GTP cyclohydrolase II, in E. coli, which is located from −146 to −118 (Koh et al., 1996). In many micro-organisms, the genes encoding the regulator are near the structural genes, and the regulator protein activates the transcription (van der Meer et al., 1992). Deletion analysis demonstrated that the promoter sequences from −108 to −217 upstream and +50 to +90 downstream are crucial for the function of the pegB promoter (Fig. 2). This region contains sequences similar to those of binding motifs for SoxS and araL and araO of AraC-type regulators, but not the sequence of the binding motif for the XylS regulator. These data strongly suggest that the pegB promoter contains the operator sites for the AraC-type regulator, PegR.

From DNA sequence analysis we predicted that the pegR gene encoded a protein of 41 kDa, which was confirmed by expression of the gene. Members of the AraC regulatory family have several common features. All have a helix–turn–helix motif in the C-terminal region implicated in DNA-binding sequences (Ramos et al., 1990). These regulatory proteins are often translated in the reverse direction from the operon they regulate (van der Meer et al., 1992). Most of the characterized proteins of this family are positive transcriptional regulators (Gallegos et al., 1993). The finding that gene regulation by this family appears in a broad spectrum of micro-organisms implies that this family of regulatory proteins regulates many genes in general. In E. coli, high β-galactosidase activity in pQF50pB10R indicated that the PegR plays an essential role in activation of the pegB gene (Table 1). It is known that the single conserved domain of the AraC protein, araL, can bind to DNA and activate transcription from cognate promoters (Reeder & Schleif, 1993). Moreover, the domain does not appear to bind effector molecules. All the data supported the assumption that PegR belongs to the AraC/XylS family of transcriptional activators and binds to the pegB promoter.

Gel retardation studies using PegR showed that the target for DNA binding is within the promoter-control region (Fig. 3b), as found with other AraC members. Although the exact DNA sequence for PegR binding has not been determined, it must be localized in the 268 bp region including the consensus AraC/XylS family DNA-binding sites. Experiments both in vitro by gel-retardation assay and in vivo by PegR fusion suggested that PegR regulates the pegBCDAE operon.

The data reported in this paper allow us to propose a model for transcription of the pegBCDAE operon (Fig. 4). Generally, a putative GalR/LacI-type regulator inhibits transcription from the pegA promoter when it binds to the cognate two operators, O₇ and O₈, with the help of HU protein. This would be the same with the pegR promoter. Derepression of pegA promoter occurs by the interaction of operator-bound GalR with inducer PEG 4000, which stimulates transcription of pegA, encoding the first-step enzyme (PEG-DH) in PEG degradation. Similarly, PegR is produced by derepression. PegR binds to the pegB promoter at the araI half-site and activates the transcription of the pegBCDAE operon. Altogether, PEG degradation is triggered by PEG. This is believed to be the first report on the transcriptional regulation of a PEG-degradative operon in sphingomonads.

ACKNOWLEDGEMENTS

This work was financially supported in part by Grant-in-Aid for Scientific Research (C) 14560068 from the Japan Society for the Promotion of Science. We are grateful for grants to F. K. from Wesco Scientific Promotion Foundation and the Yamada Foundation for Environmental Science and Power Research and Encouragement of Studies to J. C. from the Graduate School of Natural Science and Technology, Okayama University. We thank Dr J. Kato (Hiroshima University) for the kind gifts of pQF50 and E. coli K-12 MC4100 and Dr T. T. Kaitani and Dr H. H. Yurimoto (Kyoto University) for their kind help in the primer extension analysis.

REFERENCES

Probing the structure of Gal
Cleavage of structural proteins during the
Kimbara, K., Hashimoto, T., Fukuda, M., Koana, T., Takagi, M., Oishi, M.,
Microbial degradation of polyethers.
A structural taxonomy of DNA-binding domains.