Characterization of LgnR, an IclR family transcriptional regulator involved in the regulation of L-gluconate catabolic genes in Paracoccus sp. 43P

Tetsu Shimizu and Akira Nakamura

Five genes encoding enzymes required for L-gluconate catabolism, together with genes encoding components of putative ABC transporters, are located in a cluster in the genome of Paracoccus sp. 43P. A gene encoding a transcriptional regulator in the IclR family, lgnR, is located in front of the cluster in the opposite direction. Reverse transcription PCR analysis indicated that the cluster was transcribed as an operon, termed the lgn operon. Two promoters, P_{lgnA} and P_{lgnR}, are divergently located in the intergenic region, and transcription from these promoters was induced by addition of L-gluconate or D-idonate, a catabolite of L-gluconate. Deletion of lgnR resulted in constitutive expression of lgnA, lgnH and lgnR, indicating that LgnR encodes a repressor protein for the expression of the lgn operon and lgnR itself. Electrophoretic mobility shift assay and DNase I footprinting analyses revealed that recombinant LgnR binds to both P_{lgnA} and P_{lgnR}, indicating that LgnR represses transcription from these promoters by competing with RNA polymerase for binding to these sequences. D-idonate was identified as a candidate effector molecule for dissociation of LgnR from these promoters. Phylogenetic analysis revealed that LgnR formed a cluster with putative proteins from other genome sequences, which is distinct from those proteins of known regulatory functions, in the IclR family of transcriptional regulators. Additionally, the phylogeny suggests an evolutionary linkage between the L-gluconate catabolic pathway and D-galactonate catabolic pathways distributed in Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria and Actinobacteria.

INTRODUCTION

L-Gluconate is an enantiomer of D-gluconate belonging to the C6 aldonic acid family of saccharides. While D-gluconate is commonly found as a catabolic intermediate of the Entner–Doudoroff pathway, L-gluconate has only recently been found in nature (Shimizu et al., 2012). An L-gluconate catabolic pathway and a gene cluster encoding enzymes required for this catabolism, termed the lgn gene cluster, were identified in Paracoccus sp. strain 43P through an analysis of its l-glucose catabolic pathway (Shimizu et al., 2012). This catabolic pathway begins with the C-5 epimerization of L-gluconate to form D-idonate, and is catalysed by L-gluconate dehydrogenase (L-GnDH; LgnH) and 5-keto-L-gluconate reductase (LgnI). Dehydration of D-idonate to form D-idonate, and is catalysed by D-idonate dehydratase (LgnE). Finally, the conversion of LgnE produces 2-keto-3-deoxy-D-galactonate (KDGal) and is catalysed by KDGal aldolase (LgnG). Although genes and biochemical functions involved in this L-gluconate catabolic pathway have been characterized, the transcriptional organization and regulation of these genes are still poorly understood. Based upon nucleotide sequence data, lgnR, which is located upstream of the cluster in the opposite direction, has amino acid sequence similarity to IclR family transcriptional regulators, such as YiaJ (Ibañez et al., 2000) and KdgR (Pouyssegur & Stoebber, 1974). This suggests that lgnR is responsible for transcriptional regulation of the lgn gene cluster.

IclR was first characterized as a repressor for the Escherichia coli aceBAK operon responsible for acetate utilization (Maloy & Nunn, 1982; Sunnarborg et al., 1990). Further studies revealed that IclR autologously represses transcription of iclR itself (Gui et al., 1996), and two different modes of repression have been proposed for repression of aceBAK transcription by IclR (Yamamoto & Ishihama, 2003). Recent expansion of genome databases revealed that iclR orthologues are distributed in a wide range of prokaryotes, including Archaea (Krell et al., 2006), and the proteins are characterized by a DNA-binding helix–turn–helix motif in pyruvate is catalysed by KDGal kinase (LgnF) and 2-keto-3-deoxy-6-phospho-D-galactonate (KDPGal) aldolase (LgnG).
the N-terminal region and an effector binding site in the C-terminal region (Molina-Henares et al., 2006; Zhou et al., 2012). Members of this family are involved in diverse physiological functions, such as quorum sensing (Kim et al., 2007), multidrug resistance (Rojas et al., 2003), secondary metabolism (Yamazaki et al., 2003), degradation of aromatic compounds (Gerischer et al., 1998) and pathogenicity in plants (Lu et al., 2011). Unlike the LysR family of transcriptional regulators that possess a conserved operator sequence motif (TN_{11}A; Maddocks & Oyston, 2008), operator sequences of genes regulated by the IclR family proteins are more variable in their base composition and secondary structures, which may include palindromes, pseudopalindromes and direct repeats. Moreover, the regulatory mechanisms through which IclR proteins regulate transcription are also diverse, as they act as either repressors or activators, and some family members possess dual functions including autoregulation (Chao & Zhou, 2013).

It is therefore challenging to predict the operator sequences, as well as regulatory mechanisms, important for the function of uncharacterized proteins in the IclR family of transcriptional regulators. Therefore, we undertook a biochemical characterization of a newly discovered member of this family involved in a novel physiological process in order to explore the physiological roles and functional diversity present within the IclR family.

In this report, we investigated the transcriptional organization of the lgn gene cluster as well as its regulation by LgnR, which belongs to the IclR family of transcriptional regulators. Moreover, we discuss the evolutionary origin of L-gluconate catabolic genes.

**METHODS**

**Bacterial strains, plasmids, media and chemicals.** Bacterial strains and plasmids, and primers, used in this study are listed in Tables 1 and 2, respectively. E. coli strains DH10B, BL21(DE3) and S17-1pir (Simon et al., 1983) were used for subcloning, production of His$_6$-tagged LgnR and conjugation, respectively. E. coli strains were grown in Luria–Bertani (LB) medium. *P. aeruginosa* sp. L. Paracoccus sp. 43P and its mutant strain were grown in LB medium or a synthetic minimal medium [20 mM NH$_4$Cl, 20 mM potassium phosphate (pH 7.0), 5 mM MgSO$_4$, 10 mM KCl, 0.2 % Hutner’s trace element solution (Hutner et al., 1950)] supplemented with the carbon sources described below. L-Glucuronate was synthesized from L-glucose (Tokyo Chemical Industry) using hypoxoide-methanol oxidation as described elsewhere (Moore & Link, 1940). D-Idonate was synthesized by base hydrolysis of D-ido-mono-1,4-lactone (Carbosynth). L-5-Keto-glucuronate, KDGal and KDPGal were prepared as previously described (Shimizu et al., 2012).

**Construction of the ΔlgnR strain.** To obtain a mobile suicide plasmid for disruption of *lgnR*, PCR was performed using genomic DNA isolated from strain 43P and the LgnR$_{SF}$/LgnR$_{SR}$ primer set. The amplified 2665 bp fragment was cloned into the SpⅢ-BamHI site of pUC-mob (Shimizu et al., 2012) to obtain pUClgnRM. The resultant plasmid was digested with *Hind*II, for which there is a restriction site located within the *lgnR* ORF, and a kanamycin-resistance gene cassette, amplified by PCR using the kanF/kanR primer pair and pBBR1CMCS2 (Kovach et al., 1995) as a template, was inserted into the restriction site, generating pUClgnRMK.

**L-GnDH activity assay.** Strain 43P and its ΔlgnR mutant were cultured in minimal medium supplemented with 15 mM L-glucuronate, 15 mM D-Ido-ionate or 20 mM succinate until the O.D.$_{600}$ of the culture reached 0.5–0.7. Cells were harvested by centrifugation and disrupted by sonication with a Sonifier 250 (Branson) in 100 mM Tris/HCl, 10 % (v/v) glycerol, 1 mM DTT pH 8.0. After the cell debris was removed by centrifugation at 20 400 g for 15 min, the supernatant was used as a cell-free extract. L-GnDH activity was assayed using a DU 800 spectrophotometer (Beckman-Coulter) by measuring the reduction of NAD$^+ $ at 25 °C in 200 μl reaction mixture composed of 1 mM potassium L-glucuronate, 1 mM NAD$^+ $, 2.5 % (v/v) cell-free extract, 100 mM Tris/HCl pH 9.0. One unit of enzyme activity was defined as the amount of enzyme that reduced 1 μmol NAD$^+$ min$^{-1}$ under the defined assay conditions.

**RNA techniques.** Strain 43P and its ΔlgnR mutant were cultured as described above until the O.D.$_{600}$ of the cultures reached 0.5–0.8. Total RNA was prepared using a hot SDS/phenol method (Jahn et al., 2008), and cDNA was prepared from total RNA using a QuantiTect reverse transcription kit (Qiagen) according to manufacturer’s instructions. Reverse transcription PCR (RT-PCR) was performed with 20 ng total RNA as a template with or without reverse transcription using the primer sets listed in Table 2. Rapid amplification of cDNA ends (RACE) analysis was performed using a 5’ RACE system (Invitrogen) and the primers in Table 2. Quantitative RT-PCR (Q-RT-PCR) (Q-RT-PCR) was carried out with SYBR Green reaction mix (Bio-Rad) using the primer sets in Table 2 and a reverse transcription product from 5 ng total RNA as a template. Amplification of PCR products was monitored using a MiniOpticon real-time PCR system (Bio-Rad). For each sample, expression of *lgn* genes was standardized to 16S rRNA as an internal control. Q-RT-PCR experiments were conducted as three technical replicates from each of two independent cultures for each gene analysed. The relative expression level of each gene for each culture condition was compared using the 2$^{-ACT}$ method (Livak & Schmittgen, 2001).

**Expression and purification of His$_6$-LgnR.** To construct an *IclR*-expression plasmid, the ORF was PCR amplified with primers LgnR-NdeF and LgnR-EcoRI from 43P genomic DNA, and the resultant fragment was cloned into the NdeI–EcoRI site of pET28a (+). The resultant plasmid was designated pET-IgnR. *E. coli* BL21(DE3) harbouring pET-IgnR was cultured in LB medium containing kanamycin at 37 °C until the O.D.$_{600}$ of the cultures reached 0.5. IPTG was added to the cultures at a final concentration of 0.2 mM, and bacteria were grown for a further 3 h at 28 °C. Cells were harvested by centrifugation and disrupted by sonication in 50 mM Tris/HCl (pH 8.0), 10 % (v/v) glycerol, 0.5 M NaCl, and 40 mM imidazole. After the cell debris was removed by centrifugation at 24 000 g for 30 min, the supernatant was passed through a 0.45 μm filter. His$_6$-LgnR was purified using a HisTrap FF column (1 ml) (GE Healthcare) by measuring the reduction of NAD$^+$ at 25 °C until further use.

**Size exclusion chromatography.** Protein aliquots (25 μl) containing 14 mg purified His$_6$-LgnR ml$^{-1}$ were loaded onto a Superose 12 10/300 GL (GE Healthcare) column installed on an AKTA system (GE Healthcare) pre-equilibrated with buffer containing 50 mM HEPES-KOH, 10 % glycerol, 250 mM NaCl, 1 mM DTT (pH 7.2). Gel filtration was performed at a flow rate of 0.5 ml min$^{-1}$ at room temperature with a flow rate of 0.5 ml min$^{-1}$ at room temperature.
Characterization of LgnR

Table 1. Bacterial strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Genotype or description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH10B</td>
<td>F− mcrAΔ (mrr-hsdRMSS-mcrB) p80dlacZAM15 ΔlacX74 deoR recA1 araD139 Δ(ara leu)7697 galU galK λ− rpsL endA1 supG</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>BL21(DE3)</td>
<td>F− ompT hsdS(rB− mB−) Δ(wcaD) DE3 galK λ dcm 857 indI 2 sam7 nin5 lacUV5−T7gene1</td>
<td>Novagen</td>
</tr>
<tr>
<td>S17-1pir</td>
<td>RP4-2-Tc::Mu-Km::Tn7 λpir; host for conjugative transfer</td>
<td>Simon et al. (1983)</td>
</tr>
<tr>
<td>Paracoccus sp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43P</td>
<td>DlgvR− 157 bp upstream of the lgnR site</td>
<td>Shimizu et al. (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUC-mob</td>
<td>Ap5 mob+</td>
<td>Shimizu et al. (2012)</td>
</tr>
<tr>
<td>pUClgnRM</td>
<td>2.6 kb fragment from strain 43P DNA cloned into pUC19-mob</td>
<td>This study</td>
</tr>
<tr>
<td>pUClgnRMK</td>
<td>Km− cassette from pBBR1MCS2 cloned into pUClgnRM</td>
<td>This study</td>
</tr>
<tr>
<td>pET28a(+)</td>
<td>Km−</td>
<td>Novagen</td>
</tr>
<tr>
<td>pBBR1MCS2</td>
<td>Km− mob+</td>
<td>Kovach et al. (1995)</td>
</tr>
<tr>
<td>pET-LgnR</td>
<td>Km− PT7lac-lgnR</td>
<td>This study</td>
</tr>
</tbody>
</table>

Phylogenetic analysis of LgnR. Using the complete amino acid sequence of LgnR as the query sequence, a BLASTP search was performed in the online interface module of the KEGG genome database (http://www.genome.jp/kegg/). Amino acid sequences showing scores higher than 170 bits in the BLASTP search (89 sequences), in addition to YiaJ and KdgR from E. coli as representative proteins in the IclR family of transcriptional regulators, were multiply aligned using CLUSTAL_X (Chenna et al., 2003). A phylogenetic tree was generated with the MEGA4 package (Tamura et al., 2007) using the neighbour-joining method and 1000 replicates. Comparisons of gene organization surrounding lgnR orthologues were done by manual validation using the genome map tool in the KEGG database.

RESULTS AND DISCUSSION

Transcriptional mapping of the lgn gene cluster

A schematic of the gene organization of the lgn cluster, as well as the positions of the primer sets used for the RT-PCR analysis, are shown in Fig. 1(a). Our RT-PCR analysis showed that the nine genes expressed from the same strand (lgnA−I) were co-transcribed (Fig. 1b), indicating that these ORFs constitute an operon. However, as suggested by its reverse orientation, the lgnR gene, which encodes a putative IclR family transcriptional regulator, was transcribed separately, because no RT-PCR product was observed with the primer set designed to amplify the 2650 bp fragment containing the intergenic region between lgnA and lgnR. Based on these data, we focused on the intergenic region between lgnA and lgnR, which is likely to contain promoter regions for both the lgnR gene as well as the lgn operon.

5′-RACE analysis revealed that the transcriptional start sites of lgnR and the lgn operon were located at 22 and 157 bp upstream of the lgnR and lgnA ORFs, respectively (Fig. 1c). We identified sequences similar to the consensus sequence of the E. coli EoT76 promoter (Harley & Reynolds, 1987) further upstream of these transcriptional start sites:
### Table 2. Oligonucleotides used in this study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'→3')*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primers used for construction of expression plasmids</strong></td>
<td></td>
</tr>
<tr>
<td>LgnR_NdeF</td>
<td>GGCCCATATGGAGAACGATGCCAACAAGGC</td>
</tr>
<tr>
<td>LgnR_EcoR</td>
<td>GGCCGGAATTCCTTGAAGGTTGAGACAGCTA</td>
</tr>
<tr>
<td><strong>Primers used for construction of pUClgnRMK</strong></td>
<td></td>
</tr>
<tr>
<td>LgnR_SF</td>
<td>GATTTTCGCCAAAGGTTCG</td>
</tr>
<tr>
<td>LgnR_SR</td>
<td>ATTTCCCCGAAGTAAGCGCTT</td>
</tr>
<tr>
<td>kanF</td>
<td>ATGTCAGCTACTGGGCTATCTGGA</td>
</tr>
<tr>
<td>kanR</td>
<td>TGGTTCGTCATTTTCGAAAAC</td>
</tr>
<tr>
<td><strong>Primers used for RT-PCR</strong></td>
<td></td>
</tr>
<tr>
<td>1F</td>
<td>TTGGGAAACGATGCCAACAAGGC</td>
</tr>
<tr>
<td>1R</td>
<td>TTAGGGGAGTCACAGCAGAATAC</td>
</tr>
<tr>
<td>2F</td>
<td>CCGAACATCATGATCCTGGCA</td>
</tr>
<tr>
<td>2R</td>
<td>GACTTCGTCACCAACAGCATGAGT</td>
</tr>
<tr>
<td>3F</td>
<td>AGCCGCTTTTCTCCTGATGCT</td>
</tr>
<tr>
<td>3R</td>
<td>ATTTGCGCAGTAAGCGCTT</td>
</tr>
<tr>
<td>4F</td>
<td>CTGTAAAGGACGACCAGCTAC</td>
</tr>
<tr>
<td>4R</td>
<td>GAAATTTCGCCATTTTCGAAAAC</td>
</tr>
<tr>
<td>5F</td>
<td>AGATTCACGTCGTCATTTTCG</td>
</tr>
<tr>
<td>5R</td>
<td>GACCTGCTGATGTAAGCGATTC</td>
</tr>
<tr>
<td>6F</td>
<td>ATGGAAGCTAAAACAAACTCC</td>
</tr>
<tr>
<td>6R</td>
<td>TCAGTCAGCTACCCGCAAGCT</td>
</tr>
<tr>
<td>7F</td>
<td>CAAGCCTCAGAAGCGAGTT</td>
</tr>
<tr>
<td>7R</td>
<td>GAAACCCTCGGAGATGCCG</td>
</tr>
<tr>
<td>8F</td>
<td>ATGAAAGCGCTTTATCATCGA</td>
</tr>
<tr>
<td>8R</td>
<td>TTAGAACCTCTCGGAGTCCC</td>
</tr>
<tr>
<td><strong>Primers used for RACE</strong></td>
<td></td>
</tr>
<tr>
<td>RC_lgnR_midR</td>
<td>TGATCCCATCCCAAGGAAATC</td>
</tr>
<tr>
<td>RC_lgnR_midRII</td>
<td>GTCTACGCACTTCCGAGTTC</td>
</tr>
<tr>
<td>RC_lgnR_midRIII</td>
<td>GATTCAGGACATGAGAATAC</td>
</tr>
<tr>
<td>RC_lgnA_midR</td>
<td>GCAGAAGCTACCCGCAAGAG</td>
</tr>
<tr>
<td><strong>Primers used for Q-RT-PCR</strong></td>
<td></td>
</tr>
<tr>
<td>lgnA_RT_F</td>
<td>AGCGGCTTTTCTCCTGATGTC</td>
</tr>
<tr>
<td>lgnA_RT_R</td>
<td>AATCTCGAAGGCAGCAAGG</td>
</tr>
<tr>
<td>lgnH_RT_F</td>
<td>TGGTGCGCAGTCTGATGTC</td>
</tr>
<tr>
<td>lgnH_RT_R</td>
<td>CGCCATCTCCGATGTTTC</td>
</tr>
<tr>
<td>lgnR_RT_F</td>
<td>CCAACAAGGACGAGCAAGAC</td>
</tr>
<tr>
<td>lgnR_RT_R</td>
<td>AACCCTCGATCATGATAG</td>
</tr>
<tr>
<td>43P_16S_RT_F</td>
<td>GAATAGGCTCAGAAGGAG</td>
</tr>
<tr>
<td>43P_16S_RT_R</td>
<td>GCCATGGCCTAGATCAGGTT</td>
</tr>
<tr>
<td><strong>Primers and oligonucleotides used for preparation of DNA fragments used in EMSA</strong></td>
<td></td>
</tr>
<tr>
<td>lgnAR_F_B</td>
<td>Biotin-AGCTGCTCTCCCTCTCATAT</td>
</tr>
<tr>
<td>lgnAR_R_B</td>
<td>Biotin-AGCTGCTCTCCCTCTCATAT</td>
</tr>
<tr>
<td>lgnAR_R</td>
<td>GCCATCTCCGATGTTTC</td>
</tr>
<tr>
<td>lgnAR_-35F</td>
<td>CTATATGGGAAATTTGGTTCTAGTAGCAGATG</td>
</tr>
<tr>
<td>lgnAR_-35R</td>
<td>ATGTGACATGGAAACATACAAATATTTTGAATG</td>
</tr>
<tr>
<td>lgnAR_midF</td>
<td>GCAATTACGTTTACGGGAG</td>
</tr>
<tr>
<td>lgnAR_midR</td>
<td>CCTCCCTAATCTGATATGGA</td>
</tr>
<tr>
<td><strong>Primers used for preparation of DNA fragments in DNase I footprinting</strong></td>
<td></td>
</tr>
<tr>
<td>lgnR_CD_IRD800label</td>
<td>IRD800-GCAACAAATTTGCTCCCTATAT</td>
</tr>
<tr>
<td>lgnR_CD_nonlabel</td>
<td>GCCAAGAAATTCTCCCTATAT</td>
</tr>
<tr>
<td>lgnA_CD_IRD800label</td>
<td>IRD800-GAATCTGGCTTTGCGAAG</td>
</tr>
<tr>
<td>lgnA_CD_nonlabel</td>
<td>GAATCTGGCTTTGCGAAG</td>
</tr>
</tbody>
</table>

*Introduced restriction sites are underlined.
2 and 10 regions for the lgnR promoter, 5’-TTCATA-3’ and 5’-CATAAT-3’, and those for the promoter of the lgn operon, 5’-TTGACA-3’ and 5’-CATAAT-3’. This result is consistent with the early observation that the genus Paracoccus possesses typical σ70-dependent transcription (Baker et al., 1998). Taken together, we concluded that lgnR and the lgn operon are transcribed under the direction of separate promoters, P_{lgnR} and P_{lgnA}, located in this intergenic region. Genes encoding LgnABCD, putative ABC transporter components, were co-transcribed with lgn catabolic genes, suggesting that they might be involved in incorporation of L-gluconate.

**Effect of lgnR disruption on the expression of lgn genes**

To determine whether LgnR functions as a transcriptional regulator of the lgn operon, we constructed a ΔlgnR strain and examined the effect of deleting this gene on biochemical production of L-GnDH as well as expression of genes in the lgn operon. Insertion of the Km r cassette to the lgnR ORF was confirmed by PCR using primer set LgnR_SF/LgnR_SR, as the fragment amplified from the genomic DNA of the ΔlgnR strain was shown to be 1000 bp larger than that from the wild-type strain (data not shown).

---

**Fig. 1.** RT-PCR analysis of lgnR and the lgn gene cluster. (a) Schematic representation of the lgn gene cluster and predicted lengths of RT-PCR products amplified with the primer pairs used in this study. Genes encoding LgnR and enzymes for L-gluconate catabolism are shaded in dark and light grey, respectively. The predicted positions of RT-PCR products are shown with lines and their lengths are also indicated. (b) Agarose gel (1%) showing RT-PCR products obtained with template RNA isolated from Paracoccus sp. 43P cultured with minimal medium containing 15 mM L-gluconate. Lanes 1 to 8 correspond to the indicated numbers on the schematic represented in (a), and lane M denotes molecular mass markers. (c) Nucleotide sequence of the intergenic region between lgnR and lgnA. Initiation codons of lgnR and lgnA ORFs are boxed. Transcriptional start sites (+1) determined by 5’ RACE analysis are shown by arrows, and putative −10 and −35 sequences are indicated.
When the wild-type strain was cultured in minimal medium supplemented with L-gluconate, D-idonate or succinate as a control, we detected L-GnDH activity catalysed by LgnH at 0.272, 0.322 and 0.002 unit (mg protein)\(^{-1}\), respectively (Fig. 2a). Q-RT-PCR analysis showed that the expression of \(lgnA\) and \(lgnH\) in the \(lgn\) operon was induced approximately 60-fold by L-gluconate and 70-fold by D-idonate when compared with expression when cells were grown in succinate (Fig. 2b). This result was in good correlation with the LgnH activity measured (136- and 161-fold with L-gluconate and D-idonate, respectively, compared with that of succinate); although, the induction level was about twice in the enzyme activity measurement. In contrast, the \(\Delta lgnR\) strain exhibited similar L-GnDH activity irrespective of the carbon source used for culturing the cells, the activity of which was almost equal or slightly higher than that of the wild-type cultured with L-gluconate or D-idonate (Fig. 2a). Additionally, the transcriptional levels of \(lgnA\) and \(lgnH\) in the \(\Delta lgnR\) strain cultured with succinate were significantly higher than those in the wild-type cultured under the same conditions (170- and 340-fold for \(lgnA\) and \(lgnH\), respectively), and reached the levels of those in the \(\Delta lgnR\) and the wild-type strains cultured with L-gluconate or D-idonate (Fig. 2b). These results suggest that the expression of the \(lgn\) operon is induced by either L-gluconate or D-idonate, and LgnR functions as a repressor for its expression.

As the \(\Delta lgnR\) strain was constructed by inserting a Km\(^r\) cassette into the \(lgnR\) ORF, we measured \(lgnR\) expression from \(P_{lgnR}\) in the \(\Delta lgnR\) strain by using a primer set (\(lgnR\_RT\_F\) and \(lgnR\_RT\_R\)) located in a region upstream of the Km\(^r\) cassette. Interestingly, expression of \(lgnR\) was regulated by LgnR itself. Expression of the gene was induced by L-gluconate or D-idonate in the wild-type strain, and deletion of \(lgnR\) resulted in higher levels of

![Graph](https://via.placeholder.com/150)

**Fig. 2.** Effects of \(lgnR\) disruption on production of L-GnDH, and expression of \(lgnR\) and the \(lgn\) operon. (a) L-GnDH activity in cell-free extracts prepared from wild-type (white bars) and \(\Delta lgnR\) (grey bars) strains cultured with succinate, L-gluconate or D-idonate. Cell-free extracts were prepared from cells in late-exponential growth phase (\(OD_{600} 0.5–0.8\)). Mean values ± SD of three independent experiments are shown. Note that the activity in the wild-type cultured with succinate was 0.002 unit (mg protein)\(^{-1}\). (b) Relative expression of \(lgn\) genes in wild-type (white bars) and \(\Delta lgnR\) strain (grey bars). Values shown are fold-increases in relative mRNA levels in comparison with the wild-type strain cultured with 20 mM succinate.
Fig. 3. Binding of His6-LgnR to the lgnR–lgnA intergenic region. (a) Gel filtration analysis of the purified His6-LgnR. Open circles indicate elution points of molecular marker proteins (aldolase, 158 kDa; conalbumin, 75 kDa; carbonic anhydrase, 29 kDa; RNase A, 13.7 kDa; aprotinin, 6.5 kDa), and the filled circle indicates that of His6-LgnR. (b) Schematic representation of the lgnA–lgnR intergenic region and positions of the biotin-labelled DNA fragments used in EMSA. The locations of labelled DNA fragments are shown as lines with numbers. The length of each fragment is shown on the right. Grey bars indicate the positions of −10 and −35 boxes of P_{lgnA} and P_{lgnR}. (c) EMSA with 1 nM biotin-labelled DNA fragments in the presence (+) or the absence (−) of 500 nM His6-LgnR. The fragments used are indicated with numbers corresponding to those shown in (b). (d) EMSA with 1 nM biotin-labelled DNA fragments in the presence of different amounts of His6-LgnR. Experiments were conducted in triplicate, and representative results are shown. F, Unbound DNA; C1, complex 1; C2, complex 2. (e) Hill plots of the data obtained in (d). Circles, triangles, and squares indicate the results of three independent experiments. Note that for fragment 1 >99.8% of the fragment was shifted at 400 nM LgnR, and therefore the data for 600 nM LgnR were omitted from the plot.
expression (Fig. 2b). These observations suggest that LgnR functions as a repressor for the lgn operon and lgnR itself, and that L-gluconate, D-idonate, or related compounds may act as an inducer for the expression of the lgn operon and lgnR.

**LgnR binds to both the P\textsuperscript{lgnR} and P\textsuperscript{lgnA} regions**

To further investigate the biochemical function of LgnR, we produced and purified a recombinant His\textsubscript{6}-LgnR, using the pET-system in E. coli. The purified protein was detected at a molecular mass of 27 kDa by SDS-PAGE (data not shown) and 48.6 kDa by gel filtration (Fig. 3a), suggesting that His\textsubscript{6}-LgnR may exist as a homodimer in vivo. To determine whether LgnR binds to the P\textsuperscript{lgnR} and P\textsuperscript{lgnA} regions, EMSA was performed using His\textsubscript{6}-LgnR and several biotin-labelled DNA fragments containing the lgnA–lgnR intergenic region (Fig. 3b). Two band-shifts were observed using fragments containing both P\textsuperscript{lgnA} and P\textsuperscript{lgnR} in the presence of His\textsubscript{6}-LgnR (Fig. 3c, fragment 1 and 2), while no band-shift was observed using a fragment containing neither P\textsuperscript{lgnA} nor P\textsuperscript{lgnR} regions (Fig. 3c, fragment 3). Moreover, a single shifted band was observed using fragments lacking either of the −10 regions in P\textsuperscript{lgnA} and P\textsuperscript{lgnR} (Fig. 3c, fragment 4 and 5). These results indicate that there are two LgnR binding sites within the lgnA–lgnR intergenic region, and the hyper-shifted bands observed for fragments 1 and 2 may represent formation of higher-order complexes as a result of LgnR binding to two sites in the same DNA fragment. In addition, the −10 regions identified within both P\textsuperscript{lgnA} and P\textsuperscript{lgnR} seem to play a critical role in binding of LgnR.

The binding affinities of His\textsubscript{6}-LgnR to each fragment were compared by performing EMSA in the presence of various concentration of His\textsubscript{6}-LgnR (Fig. 3d). After quantification of the bound and unbound EMSA, we plotted the data on Hill plot (Fig. 3e). The Hill coefficients for fragments 1, 4 and 5 were 5.2±0.8, 4.1±0.1 and 2.8±0.2, respectively, indicating that binding of LgnR to these fragments is cooperative. LgnR exists as a dimer in solution, and its binding to fragments 4 or 5 produces a single shifted band. If we assume that fragments 4 and 5 contain only one binding site in each fragment, this result may indicate that two or more LgnR dimers cooperatively bind to a single site. In the IcR family of transcriptional regulators, BlcR is known to bind its binding site as tetramer; although, it exists as dimer in solution (Pan et al., 2011). It is thus possible that LgnR also forms a homotetramer upon binding to its binding site.

Apparent dissociation constants of His\textsubscript{6}-LgnR to fragments 1, 4 and 5 were calculated from the plots as monomer to be 118±11, 136±13 and 248±21 nM, respectively. His\textsubscript{6}-LgnR showed 1.8-fold higher binding affinity for fragment 4, containing the P\textsuperscript{lgnA} region, than for fragment 5, containing the P\textsuperscript{lgnR} region. This suggests that transcriptional repression mediated by LgnR on the lgn operon is stronger than the repression of lgnR. Also, there was only slight difference between binding affinity of His\textsubscript{6}-LgnR to fragment 4 and fragment 1, which contains both P\textsuperscript{lgnA} and P\textsuperscript{lgnR} regions, suggesting that binding of LgnR to P\textsuperscript{lgnA} and P\textsuperscript{lgnR} are independent of each other.

**Effector molecule of LgnR**

To identify potential effector molecule(s) that may be required for dissociation of LgnR from its binding sites, L-gluconate and catabolic intermediates of this pathway were added as additional components in further EMSA assays. Addition of 100 μM D-idonate effectively dissociated LgnR from the DNA fragment containing both promoter regions (Fig. 4a). Other compounds, including L-gluconate, had no effect on this dissociation. Our EMSA results with D-idonate agreed with those of the in vivo expression analyses (Fig. 2). However, we previously described that addition of L-gluconate or D-idonate to culture media induced the expression of both lgnR and the lgn operon. As L-gluconate is converted to D-idonate by LgnH–LgnI reactions in this pathway, it is possible that supplementary L-gluconate acts as an inducer only after its conversion to D-idonate.

**Fig. 4.** Effect of D-idonate addition on the binding of His\textsubscript{6}-LgnR to P\textsuperscript{lgnA} and P\textsuperscript{lgnR}. (a) Effect of catabolic intermediates in the L-gluconate pathway on binding of His\textsubscript{6}-LgnR to the lgnA–lgnR intergenic region. EMSA was performed with 400 nM His\textsubscript{6}-LgnR and 1 nM biotin-labelled fragment 1 in the presence of 100 μM of each catabolic intermediate. F, Unbound DNA; C1, complex 1; C2, complex 2; 5-KLGlu, 5-keto-L-gluconate. (b) EMSA with 400 nM His\textsubscript{6}-LgnR and 1 nM fragment 4 and 5 in the presence of various concentrations of D-idonate. D-idonate was added at (from left to right) 0, 50, 100, 200 and 400 μM, respectively.
To further demonstrate that D-idonate acts as an effector on the dissociation of LgnR from both of its binding sites, we performed EMSA using fragments containing \( P_{lgnA} \) or \( P_{lgnR} \) (4 or 5) in the presence of various concentration of D-idonate. As expected, dissociation of LgnR from both fragments was observed in a manner dependent on the concentration of D-idonate (Fig. 4b), suggesting that D-idonate is a genuine inducer molecule mediating expression from both \( P_{lgnA} \) and \( P_{lgnR} \).

**DNase I footprinting analysis**

To identify LgnR binding sequences within the two promoters, DNase I footprinting analysis was performed on a 345 bp fragment extending from the +120 region of \( P_{lgnA} \) to the +136 region of \( P_{lgnR} \). This DNA sequence was labelled with IRD-800 dye at each end individually. As shown in Fig. 5, protection was observed at the two promoter regions within both the \( lgnA \)- and \( lgnR \)-coding strands in the presence of 400 nM LgnR. In the \( P_{lgnR} \) region, sequence extending from +5 to −35, containing the transcriptional start site as well as the −10 and −35 boxes, was protected by the presence of LgnR. In the \( P_{lgnA} \) region, sequence extending from −10 to −36, containing the −10 and −35 boxes, was protected. These data indicate that LgnR represses transcription from these two promoters by binding to the promoter regions and thereby preventing the binding of RNA polymerase from the promoters with steric hindrance, as has been described for other bacterial repressor proteins (Browning & Busby, 2004). Both of these protected regions contain an inverted repeat (IR) sequence (5’-CATAAATAG-3’) at their −10 boxes, suggesting that these IR sequences are important for the binding of LgnR to \( P_{lgnA} \) and \( P_{lgnR} \). This was supported by the EMSA results that deletion of either of the −10 regions produced loss of the hyper-shifted bands.

In the presence of D-idonate, protection by LgnR at the −35 regions of the two promoters was only slightly reduced.

---

**Fig. 5.** DNase I footprinting analysis using His6-LgnR and the \( lgnA-lgnR \) intergenic region. DNase I footprinting was conducted with 100 ng DNA fragment labelled with IRD-800 at the 5’ end and 400 nM His6-LgnR in the presence or the absence of the indicated concentration of D-idonate. Positions of \( P_{lgnA} \) and \( P_{lgnR} \) are shown on the left. ‘ACGT’ indicates the sequencing ladders generated by the chain-terminating sequencing reactions. ‘N’ denotes DNase I reaction without LgnR addition. Black bars indicate regions protected by LgnR within the \( lgnA \)- and \( lgnR \)-coding strand, and the sequences of protected regions are shown on the right. Inverted repeat sequences found in both protected regions are indicated by arrows on the sequences. Grey bars indicate the regions where LgnR protection was slightly reduced by D-idonate addition.
reduced (Fig. 5, grey bars), while the −10 regions continued to be protected to the same degree. The surprisingly modest influence of D-idonate on the dissociation of LgnR from the operator sites may be due to the high concentration of LgnR used in our assays. A second possibility is that D-idonate requires a synergistic effector compound for complete dissociation of LgnR from its target DNA. A comparable mechanism has been described...
for the regulation of MhpR activator, where phenylpro- 
pionate itself is not an effector for MhpR binding to DNA, 
but rather phenylpropionate enhances the activity of its 
true effectors, 3-hydroxy-phenylpropionate and 3-(2,3-
dihydroxyphenyl)propionate (Manso et al., 2009). To 
examine this second possibility, we performed the DNase I 
protection assay in the presence of D-idonate and 
addition compounds, including all intermediates found 
in the L-gluconate catabolic pathway. No synergistic effects 
were observed with the compounds tested (data not 
shown). Nevertheless, based on the results from our in 
vivo expression and EMSA analyses, we propose that D-
idonate is a true effector for LgnR. However, we cannot 
xclude the possibility that an unknown compound(s) may 
also be involved in the induction of P\(_{lgnA}\) and P\(_{lgnR}\) either 
acting directly or synergistically.

**Phylogenetic analysis of LgnR and evolutionary 
origin of the lgn operon**

In the KEGG genome database, LgnR orthologues are 
distributed in the classes Alphaproteobacteria, Betaproteobacteria 
and Gammaproteobacteria, and in the phylum Actinobacteria. 
Phylogenetic analysis based on the amino acid sequences of 
LgnR orthologues, along with known members of the IclR 
family of transcriptional regulators, revealed that LgnR 
orthologues form a cluster with putative proteins from other 
genome sequences, which is deeply branched from a cluster 
of proteins of known function, YiaJ and KdgR, which are 
responsible for the catabolism of acidic sugars like L- 
gluconate. KdgR controls both the utilization of 2-keto-3-
deoxy-D-gluconate in E. coli (Nasser et al., 1992), as well as 
pectin degradation by plant pathogens (Reverchon et al., 
1991), and YiaJ regulates the expression of L-ascorbate 
catabolic genes in E. coli (Campos et al., 2008). The amino 
acid sequence of LgnR shares only 28 and 32 % identity with 
the sequences of E. coli KdgR and YiaJ, respectively.

In the cluster containing LgnR, two subgroups (groups I 
and II) are apparent and LgnR is present in group I (Fig. 
6a). Genes encoding group II proteins are often found in 
the respective bacterial genomes in clusters with rhamD 
orthologues encoding putative L-rhamnionate dehydratases 
(Rakus et al., 2008). All of the genes encoding group I 
proteins, with the exception of two Pseudomonas genes 
(PFA506_3986, PFLU4681), are located in gene clusters 
with lgnF and/or lgnG orthologue(s) encoding putative 
KDGal kinases and KDPGal aldolases, respectively (Fig. 
6b). This may reflect the co-evolutionary history of lgnF/ 
lgnG orthologues with group I lgnR orthologues. dgoD 
orthologues encoding putative D-galactonate dehydratases 
(Babbitt et al., 1995) are present in some of these gene 
clusters, suggesting that these gene clusters are responsible 
for D-galactonate utilization. Notably, proteins placed in 
the downstream portion of the L-gluconate catabolic pathway 
in Paracoccus sp. 43P and those of the D-galactonate catabolic 
pathway reported in E. coli (Deacon & Cooper, 1977) and 
non-pathogenic Mycobacteria (Szumilo, 1981) share identical 
catabolic processes from KDGal, which are mediated by 
LgnF/LgnG orthologues. Based on our previous phylogenetic 
analyses, lgnE, lgnH and lgnI genes in the lgn operon form 
respective clusters with sequences from other bacterial 
affiliations, and therefore are assumed to be derived from 
other bacterial origins (Shimizu et al., 2012). The catabolic 
and genetic linkages between the L-gluconate catabolic 
pathway in Paracoccus sp. 43P and the D-galactonate catabolic 
pathways in other bacteria may indicate bottom-up evolution 
of these catabolic pathways. The KDGal catabolic pathway, 
which is regulated by an lgnR orthologue, was present in the 
past as an ancestor, and this pathway may have been 
subsequently adapted to D-galactonate utilization by acquiring 
dgoD orthologues, or to L-gluconate utilization by acquiring 
lgnE, lgnH and lgnI genes.

**Conclusions**

In this report, we determined the transcriptional organi-
zation of genes involved in L-gluconate catabolism in 
Paracoccus sp. 43P. We found that the lgn gene cluster is 
co-transcribed as an operon. A transcriptional regulator in 
the IclR family, encoded by lgnR, is located in the upstream 
region of lgnA and negatively regulates transcription of the 
Lgn operon and lgnR itself. EMSA and DNase I footprinting 
analysis revealed that LgnR binds to both P\(_{lgnA}\) and P\(_{lgnR}\) 
indicating that LgnR inhibits transcription from these 
promers by competing with RNA polymerase. D-Idonate, an 
intermediate of the L-gluconate catabolic pathway, is a 
candidate inducer molecule for expression from these 
promers. Phylogenetic analysis revealed that LgnR belongs 
to a cluster of putative proteins from genome sequences in 
the IclR family of transcriptional regulators. Further gene 
cluster analysis implied an evolutionary linkage between the 
L-gluconate catabolic pathway and D-galactonate catabolic 
pathways distributed in Alphaproteobacteria, Betaproteo-
bacteria, Gammaproteobacteria and Actinobacteria.

**ACKNOWLEDGEMENTS**

This work was supported by JSPS KAKENHI grant no. 21380054.

**REFERENCES**

Babbitt, P. C., Mrachko, G. T., Hasson, M. S., Huisman, G. W., Kolter, 
functionally diverse enzyme superfaimly that abstracts the alpha 

Baker, S. C., Ferguson, S. J., Ludwig, B., Page, M. D., Richter, O. M. 
Paracoccus: metabolically versatile bacteria with bioenergetic flexibili-


Campos, E., de la Riva, L., Garces, F., Giménez, R., Aguilar, 
J., Baldoma, L. & Badia, J. (2008). The yiaKLX1Z2PQRS and 
ulaABCDEF gene systems are required for the aerobic utilization
of l-ascorbate in Klebsiella pneumoniae strain 13882 with l-ascorbate-6-phosphate as the inducer. *J Bacteriol* 190, 6615–6624.


Edited by: F. Sargent