Characterization and Cloning of CelR, a Transcriptional Regulator of Cellulase Genes from Thermomonospora fusca*

(Nikolay A. Spiridonov and David B. Wilson§
From the Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853)

CelR, a protein that regulates transcription of cellulase genes in Thermomonospora fusca (Actinomycetaeae) was purified to homogeneity. A 6-kilobase NotI-SacI fragment of T. fusca DNA containing the celR gene was cloned into Escherichia coli and sequenced. The celR gene encodes a 340-residue polypeptide that is highly homologous to members of the GalR-LacI family of bacterial transcriptional regulators. CelR specifically binds to a 14-base pair inverted repeat, which has sequence similarity to the binding sites of other family members. This site is present in regions upstream of all six cellulase genes in T. fusca. The binding of CelR to the celE promoter is inhibited specifically by low concentrations of cellobiose (0.2–0.5 mM), the major end product of cellulases. The other sugars tested did not affect binding at equivalent or 50-fold higher concentrations. The results suggest that CelR may act as a repressor, and that the mechanism of induction involves a direct interaction of CelR with cellulose.

Biodegradation of cellulose by bacteria and fungi is accomplished by extracellular cellulolytic enzymes encoded by genes that are subject to transcriptional control. Several regulators of the transcription of cellulase genes in the cellulolitic filamentous fungus, Trichoderma reesei, including the activators ACE I and ACE II and the glucose repressor, Cre1, have been described (1, 2). However, little is known about the molecular mechanisms of cellulase regulation in soil bacteria.

The thermophilic actinomycete, Thermomonospora fusca, a major degrader of cellulose in plant residues, is an extensively studied cellulolytic bacterium. This species produces six different extracellular cellulases, designated E1 through E6. Three of the enzymes (E1, E2, and E5) are endocellulases, two are exocellulases (E3 and E6), and one enzyme possesses both exo- and endocellulolytic activity (E4). The major product of these enzymes is cellobiose (3–5).

Synthesis of cellulases in T. fusca is regulated by cellobiose induction and caballase repression, with any readily metabolized sugar acting as a repressor (6). The lowest level of cellulase synthesis (3 nM) was observed with xylose as a carbon source, and the highest level was found in cultures grown on microcrystalline cellulose. Endocellulases and exocellulases showed distinctly different regulation patterns, with exocellulases showing the highest level of induction (7).

The structural genes for all these enzymes, designated celA through celE (8–10) and celF, have been cloned and sequenced. All six cellulase genes in T. fusca contain the 14-bp inverted repeat sequence TGGGAGCGCTCCCA in their 5′-upstream regions. A regulatory protein that interacts with the upstream regions of cel genes was detected in T. fusca cultures grown on cellulose. The binding site was identified as the 14-bp inverted repeat by DNase I footprinting of the celE gene and by chemical footprinting of the celB gene (11, 12).

Here we report the isolation and properties of CelR, a regulatory protein from T. fusca that specifically binds to the 14-bp inverted repeat site in cellulase genes, as well as the complete nucleotide sequence of the celR gene.

MATERIALS AND METHODS

Bacterial Strains and Plasmids—T. fusca ER1 is an extracellular protease-negative strain derived from T. fusca YX (12). Escherichia coli strain DH5α (19) was used for cloning and plasmid isolation. Plasmid pE5-46 has been described previously (3). Plasmid pNS1 was constructed as described below.

Growth of Organisms—T. fusca was grown on Hagerdahl medium (20) supplemented with 0.5% Solka Floc (microcrystalline cellulose, James River Corp.). E. coli strains containing recombinant plasmids were grown in Luria broth or plated on Luria agar plates containing 0.1 mg/ml ampicillin.

Quantitation of DNA Binding Activity—Plasmid pE5-46 from E. coli strain D541, which contained the celE gene from T. fusca (3) was cut with EcoRI, SalI, and XhoI, and the 3′ ends of the fragments were labeled with [α-32P]dCTP or [α-32P]dATP, using DNA polymerase I Klenow fragment (21). The following 32P-labeled pE5-46 fragments were present in the reaction mixture: a high mobility 548-bp fragment (containing the celE promoter and regulatory region), a low motility 2656-bp fragment (containing the pE5-46 promoters), and a 1428-bp fragment (containing the coding region of the celE gene). Binding was carried out in 15–20 µl of 40 mM Tris-HCl buffer (pH 8.0), 150 mM KCl, 1 mM MgCl2, 0.1 mM EDTA, 0.1 mM DTT, and 2 ng of 32P-labeled fragments, T. fusca extract (1–6 µg of protein), 10 µg/ml poly(dG-dC), and 12.5% glycerol. After incubation at 37 °C for 15 min, samples were electro-

* This work was supported by Grant DE-FG02–84ER13233 from the Department of Energy Basic Energy Research Program. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF086819.

† Permanent address: Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Moscow Region 142292, Russia.

‡ To whom correspondence should be addressed: 458 Biotechnology Bldg., Cornell University, Ithaca, NY 14853. Tel.: 607-255-5706; Fax: 607-255-2428; E-mail: dbw3@cornell.edu.

2 The abbreviations used are: DTT, dithiothreitol, PAGE, polyacrylamide gel electrophoresis; bp, base pairs; kb, kilobase pairs; HPLC, high performance liquid chromatography.
phoresed through a 1.2% agarose gel in 90 mM Tris borate buffer (pH 8.3) containing 3 mM EDTA. Gels were dried at 80 °C under vacuum and radioautographed. The extent of binding to the celE regulatory region was estimated from dilutions that yielded 25–75% conversion of the DNA fragment to the slower mobility DNA-protein complex. One unit of DNA-protein binding activity was defined as the amount of protein that converts 50% of the DNA fragment to a DNA-protein complex under the assay conditions.

**Purification of the CelR Protein—**T. fusca strain ER1 was grown in 10 liters of Hagerdahl medium (20) containing 0.5% Solka Floc and 2 g of yeast extract at 52 °C (pH 7.2 at 40% oxygen saturation). Cells were harvested after 15 h by centrifugation, suspended in a Tris buffer containing 10 mM Tris-HCl (pH 7.8), 1 mM EDTA, 0.1 mM phenylmethylsulfonyl fluoride, and 0.1 mM DTT. All purification steps were performed at 4 °C, unless otherwise noted. Cells were lysed with a French press, and cell debris was removed by centrifugation at 10,000 × g for 25 min. Streptomyces sulfate (4%) was added to the supernatant, and the DNA precipitate was removed after 1 h by centrifugation at 10,000 × g for 25 min. Sodium chloride (0.5 M) was added, and the supernatant (700 ml) was loaded on a 200-ml phenyl-Sepharose CL-4B column (Amersham Pharmacia Biotech). The column was eluted successively with 0.5M, 20 mM NaCl in 10 mM Tris-HCl (pH 7.8) containing buffer containing 10 mM Tris-HCl (pH 7.8), 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, and 0.1 mM DTT. All purification steps were performed at 4 °C, unless otherwise noted. Cells were lysed with a French press, and cell debris was removed by centrifugation at 10,000 × g for 25 min. Streptomyces sulfate (4%) was added to the supernatant, and the DNA precipitate was removed after 1 h by centrifugation at 10,000 × g for 25 min. Sodium chloride (0.5 M) was added, and the supernatant (700 ml) was loaded on a 200-ml phenyl-Sepharose CL-4B column (Amersham Pharmacia Biotech). The column was eluted successively with 0.5M, 20 mM NaCl in 10 mM Tris-HCl (pH 7.8) containing 1 mM EDTA, 0.1 mM DTT, the same buffer without NaCl, and with distilled water. Fractions eluted with distilled water (150 ml) were applied to a 20-ml phenyl-Sepharose CL-4B column (Amersham Pharmacia Biotech) equilibrated with 10 mM Tris-HCl (pH 7.8), 0.1 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, and 0.1 mM DTT (buffer A). The column was eluted with a linear gradient of NaCl from 0.1 to 1.5 × (400 ml). Appropriate fractions (120 ml) were pooled, dialyzed against buffer A, and loaded on a 7-ml DNA-Sephrose column that contained DNA fragments with multiple copies of the 14-bp inverted repeat (TGG-CAGCCCGCTCCCA) coupled to Sepharose 4B (Amersham Pharmacia Biotech). DNA affinity chromatography was performed in buffer A at room temperature. After washing with 25 ml of buffer A, the column was successively eluted with 0.3, 0.6, and 1.0 M NaCl in buffer A. Electrophoretically pure (>95%) CelR protein was eluted with 10 ml of warm (50 °C) 1.5 M NaCl, concentrated, and desalted using Centricon-30 concentrators. Protein was determined with the BCA reagent (Pierce) using bovine serum albumin (Sigma) as a standard.

**Molecular Weight Determination—**The molecular weight of CelR was determined by 12% SDS-PAGE (22) using low and high molecular weight protein standards (Sigma). Protein was determined with the BCA reagent (Pierce) using bovine serum albumin (Sigma) as a standard.

**DNA Sequencing—**Double-stranded DNA from nps1 was used as a template for sequencing the celR structural gene and its 3′- and 5′-flanking regions. The sequences of both strands were determined by the dideoxy chain termination method (23) at the BioResource Center, Cornell University. The two degenerate 15-mer oligonucleotides were used to determine the initial nucleotide sequences. Specific primers for sequencing the celR gene were synthesized by the Bioresource Center.

**CellE Promoter Binding Activity in T. fusca Extracts—**T. fusca cultures grown on microcrystalline cellulose possessed considerable celE promoter binding activity (Fig. 1). Preliminary experiments showed that the maximum level of activity was observed after 12–24 h of cultivation, and that the levels of binding activity remained high during an additional 48 h of cultivation. Cells grown on non-inducing carbon sources had insignificant celE promoter binding activity (data not shown).

Reduced motility of the high molecular weight DNA fragment containing the E. coli plasmid promoters was also observed (Fig. 1, lanes 2–4). However, it was independent of the carbon source used to grow T. fusca. Presumably, the reduced motility of the high molecular weight fragment was due to the incomplete digestion of RNA polymerase from T. fusca. DNA from T. fusca was present in the crude extract with plasmid promoters from pE5-46. No shift was detected for the 1428-bp DNA fragment, which contained no promoter elements.

**Purification of CelR Protein—**The CelR protein was purified from 10 liters of T. fusca culture by three chromatographic steps. The process of purification was followed by the gel retardation assay, Lowry protein measurement (25), and SDS-
PAGE (22). A purification summary is presented in Table 1. Purification yielded highly active protein of greater that 95% purity (Fig. 2). The resulting CelR bound only to the DNA fragment containing the celE promoter (Fig. 1).

**Molecular Weight of CelR Protein**—The molecular mass of pure CelR protein as determined by SDS-PAGE is 41 ± 1.5 kDa. Samples of partially purified and pure CelR were separated with SDS-PAGE, and the gels were cut into 2-mm sections. Protein was eluted from each section, renatured, and tested for DNA binding activity. In each case the binding activity was associated with the 41-kDa band (data not shown). Estimation of the native molecular weight of purified CelR with HPLC gel chromatography showed that the protein aggregated under non-denaturing conditions. About 3% of CelR came off the column as a dimer (70–85 kDa), a tetramer (145–175 kDa), and an octamer (250–400 kDa). The major part of the protein formed high molecular weight aggregates that could not be resolved by HPLC. Only traces of CelR monomer were found.

**CelR Binding Affinity**—The dissociation constant was calculated as the concentration of CelR that caused 50% of the DNA to form a complex with the protein under the described conditions. The apparent \( K_d \) for the CelR-celE promoter complex in the absence of cellulose was 0.5–1 × 10^{-8} M.

**Effect of Cellulose on CelR Protein-DNA Binding**—The binding of CelR to the celE promoter region was inhibited by cellulose at concentrations of 0.2–0.5 mM and higher (Fig. 3). The apparent dissociation constant for cellulose was 5 × 10^{-4} M, as measured in gel retardation experiments under the above conditions with 75 pg/µl CelR. The effect of cellulose on the dissociation constant of the Cel-celE promoter complex is shown on Fig. 4. A 500-fold increase in cellulose concentration (from 0.1 to 50 mM) resulted in a 7-fold increase in \( K_d \) for the Cel-celE promoter complex. At the same time, equivalent or higher concentrations of other tested sugars did not affect binding. Cellotriose, sophorose, and xylose showed slight inhibition at 50 mM. Other mono- and disaccharides (glucose, galactose, mannose, xylose, arabinose, sucrose, lactose, and maltose) had no effect on binding even at 100 mM (data not shown).

**CelR Gene Cloning and Characterization**—The N terminus and an internal region of CelR were sequenced. Two degenerate oligonucleotides prepared on the basis of reverse translation of the amino acid sequence were used to select E. coli clones. About 200 E. coli transformants from a NatI-SacI library were screened. Six transformants contained plasmid DNA with a 6-kb insert that hybridized with the two different oligonucleotide probes. Plasmid DNA (pNS1) from one of the positive clones was used as a template for sequencing. The DNA insert contained a three-cistron operon bgLABC coding for two sugar permeases and a ß-glucosidase, and celIR, coding for the regulatory protein. Both strands were sequenced, and the sequence of celR is shown in Fig. 5. The celR gene has a G+C content of 68%, which is similar to the 65% G+C content of T. fusca DNA (9). A reading frame from nucleotide 109 to nucleotide 1128 encodes a 340-amino-acid moderately hydrophobic protein (125 hydrophobic amino acids). The molecular weight of CelR, inferred from its amino acid sequence, is 36,863 daltons, lower than its apparent molecular weight determined by SDS-PAGE. The reading frame that codes for CelR has a high content of G and C (85.6%) in the third positions of the codons, typical for cellulase genes from T. fusca (8–10). The N-terminal (first 37 amino acids) and an internal (amino acids 60–77) sequence of the protein read from the nucleotide sequence of the celR gene are identical to the N-terminal sequence of CelR protein and the sequence of its endo-lys-C cleavage product, as determined with the protein sequencer.

A potential ribosome binding site (GGAA) is located 5 bases upstream of the start codon. The 5’-regulatory region of celR contains an inverted repeat that may act as a transcription terminator for the bglC gene whose termination codon is located 108 nucleotides upstream of the celR start codon. A putative transcription terminator sequence for the celR gene (a 21-base palindrome) is located immediately downstream of its structural region.

It is interesting to note that the reverse strand of the celR gene contains a 1008-nucleotide open reading frame that begins with GTG start codon (nucleotides 1140–1138) and terminates with TAG (nucleotides 135–133). A putative ribosome-binding site (AGG) is located 5 nucleotides upstream from the start codon. The hypothetical 335-amino-acid protein product is 23% identical to a putative ATP/GTP-binding protein from Streptomyces coelicolor, AL031225.

**CelR Sequence Similarities**—The amino acid sequence of CelR has been scanned against the GenBank and EMBL data bases. CelR shares significant homology with a number of proteins that belong to the GalR-LacI family of bacterial transcriptional regulators. It is 49% identical to a transcriptional regulator from S. coelicolor (e1309425), 36% identical to a ribose operon repressor RbsR from E. coli (P25551), 32% identical to a transcriptional repressor CyrR from E. coli (P06964), a transcriptional regulator DegA from Bacillus subtilis (e1173517), and a ribose operon repressor RbsR from Haemophilus influenzae (P44329), 30% identical to a galactose operon repressor GalR (P03024), a maltose repressor MalI (M60722), and a sucrose operon repressor CscR (P40715) from E. coli, 29% identical to the lactose repressor from E. coli (P03023) and to many other members of the family.

The N-terminal helix-turn-helix motif, located between amino acids 8 and 29, is the most conserved part of the CelR protein. It is 65% identical and 80% similar to the GalR-LacI consensus (Fig. 6). An unusual feature of CelR is a cluster of four consecutive arginine residues near the N terminus that is followed by proline and threonine (amino acids 8–9). This may represent a phosphorylation site for a cAMP- and cGMP-de-

---

4 N. A. Spiridonov, unpublished results.
pendent protein kinase (26). Location of this site at the boundary of the DNA-binding motif may indicate its involvement in regulation of DNA-protein interactions. The amino acid sequence (residues 205–210, 256–260, and 282–285) of CelR also shares some homology with elements of the sugar binding sites of the GalR/LacI proteins (27, 28).

### DISCUSSION

A functionally active DNA-binding protein that interacts specifically with the 5’-regulatory region of the celE gene was isolated from *T. fusca* grown on microcrystalline cellulose. CelR did not bind to the *E. coli* promoters in pE5-46 or to the coding region of the celE gene; it bound tightly to the celE promoter.

Sequence similarities showed that CelR belongs to the GalR-LacI family of transcriptional regulators. The GalR-LacI regulators bind to their target DNA sites as homodimers, and their operator sequences are inverted repeats. Each monomer of a homodimer interacts with its half of an operator sequence with an N-terminal DNA-binding domain that contains a characteristic helix-turn-helix motif (27). The CelR protein shares a number of common features with other members of this family.

CelR possesses a typical DNA binding domain, based on sequence comparison, and its binding site is similar to operator sites for the lactose repressor LacI from *E. coli*, the amylase repressor CcpA from *Bacillus subtilis*, and some other members of the GalR-LacI family (Table II). Unlike other GalR-LacI proteins that form dimers and tetramers, CelR, because of its hydrophobicity, shows strong aggregation under non-denaturing conditions.

The 5’ regulatory regions of all known cellulase genes in *T. fusca* contain from one to three copies of this sequence (7). Presumably, all six cel genes in *T. fusca* are under transcriptional control of the CelR protein. Apparently, cellulase regulation in *T. fusca* follows the general design of transcriptional control of genes encoding enzymes for carbohydrate catabolism in other eubacteria. In particular, it resembles the regulatory system of the amylase and chitinase genes in *S. lividans* that are controlled by Reg1, a member of the GalR/LacI family, that acts as a positive regulator of chitinase genes under inducing conditions, or as a negative regulator of *α*-amylase and chitinase production in the presence of glucose (29). Similar to *reg1*, which is not adjacent to the *α*-amylase genes in *S. lividans* controlled by *reg1*, celR is not adjacent to any of the *T. fusca* cellulase genes. Cellulase regulation in *T. fusca* is different from regulation of cellulase genes in the fungus *T. reesei*, although the two evolutionally distant species possess similar cellulolytic enzyme systems (30). In *T. reesei*, the cellulase genes are controlled by the transcriptional activators ACE I and ACE II and a glucose repressor Cre1 belonging to the family of Cys₆-His₃ zinc finger proteins (1, 2).

The vast majority of proteins of the GalR/LacI family bind carbohydrate or nucleoside effectors (27). Our results show that

### Table I

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Volume ml</th>
<th>Total protein mg</th>
<th>Total activity unit × 10⁶</th>
<th>Yield %</th>
<th>Specific activity unit/mg × 10⁶</th>
<th>Purification fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude extract</td>
<td>650</td>
<td>3,120</td>
<td>13,000</td>
<td>100</td>
<td>4.2</td>
<td>1</td>
</tr>
<tr>
<td>Phenyl-Sepharose</td>
<td>150</td>
<td>117</td>
<td>6,000</td>
<td>46</td>
<td>51.3</td>
<td>12</td>
</tr>
<tr>
<td>Heparin-Sepharose</td>
<td>120</td>
<td>23</td>
<td>4,800</td>
<td>37</td>
<td>210</td>
<td>50</td>
</tr>
<tr>
<td>DNA affinity-Sepharose</td>
<td>12</td>
<td>0.105</td>
<td>240</td>
<td>1.8</td>
<td>2,290</td>
<td>547</td>
</tr>
</tbody>
</table>

**FIG. 2.** Electrophoregrams of CelR purification fractions separated by 12% SDS-PAGE and stained with Coomassie Brilliant Blue R-250. Lane 1, protein molecular weight markers; lane 2, crude extract of *T. fusca*; lane 3, phenyl-Sepharose; lane 4, heparin-Sepharose; lane 5, affinity DNA-Sepharose.

**FIG. 3.** Inhibition of CelR protein binding to the celE promoter by cellobiose. A–D, as in Fig. 1. Lane 1, no protein; lane 2, 1.2 ng of CelR protein; lanes 3–8, 1.2 ng of CelR protein and cellobiose (0.2, 0.5, 1, 5, 10, or 50 mM).

**FIG. 4.** The effect of cellobiose on the dissociation constant of CelR-celE promoter complex.
the DNA binding activity of CelR is modulated by cellobiose. Results of in vitro experiments explain the literature data on the induction of cellulase production in *T. fusca* cultures by cellobiose (6, 7). The fact that low physiological levels of cellobiose (0.2–0.5 mM) are sufficient for dissociation of the CelR-celE promoter complex in vitro is evidence that cellobiose is the true inducer of the *cel* genes, and that the mechanism of induction involves a direct interaction of CelR with cellobiose. In the absence of cellobiose, CelR presumably forms complexes with the 14-bp inverted repeats next to the *cel* genes and blocks their transcription. However, a very low constitutive level of cellulase synthesis (about 3 nM) was observed even when cells were grown on non-inducing sugars (7). When cellulose is present in the environment, cellobiose resulting from its digestion enters cells and forms complexes with CelR, allowing transcription to proceed. As cellobiose in the cells is exhausted, CelR likely represses transcription. This simple mechanism may allow quick adaptation of cells to changing environments, improve the efficiency of the *cel* regulon, and help avoid the unnecessary production of extracellular enzymes. The occurrence of the 14-bp inverted repeat in the upstream regions of *cel* genes from different *Streptomyces* species (13–16) suggests that this regulatory mechanism may be also shared by other cellulase genes. One result that is not explained by this model is the absence of active CelR protein in cells grown without cellulose or cellobiose. The mechanism of cellulase regulation in *T. fusca* under non-inducing conditions requires further investigation.

**Acknowledgments**—We gratefully thank Diana Irwin, Sheng Zhang, Joseph Calvo and Theodore Thannhauser for advice and discussions, Guifang Lao for preparation of the DNA affinity column, and William Enslow for expert protein sequencing and HPLC gel chromatography.

**REFERENCES**


---

**(a)** N = any base, K = G/T.
CelR, a Transcriptional Regulator of Cellulase Genes
