Characterization of the Dictyostelium discoideum Cellulose-binding Protein CelB and Regulation of Gene Expression*

(Received for publication, March 20, 1997, and in revised form, July 2, 1997)

Ramachandran Ramalingam‡, and Herbert L. Ennis§
From the Roche Institute of Molecular Biology, Roche Research Center, Nutley, New Jersey 07110

Similar to other stages of Dictyostelium development, spore germination is a particularly suitable model for studying regulation of gene expression. The transition from spore to amoeba is accompanied by developmentally regulated changes in both protein and mRNA synthesis. A number of spore germination-specific cDNAs have been isolated previously. Among these are two members of the 270 gene family, a group of four genes defined by the presence of a common tetrapeptide repeat of Thr-Glu-Thr-Pro. celA (formerly called 270–6) and celB (formerly 270–11) are expressed solely and coordinately during spore germination, the levels of the respective mRNAs being low in dormant spores, rising during germination to a maximum level at about 2 h, and then rapidly declining as amoebae are released from spores. The mRNAs are not found in growing cells or during multicellular development. The rapidity with which these transcripts accumulate and then disappear during germination implies that the respective products may be important for the process. We reported previously that the purified CelA protein is a cellulase (endo-1,4-β-glucanase (PC 270–6)). In the present investigation, properties of the CelB protein, a glycosylated protein of 532 amino acids, 36% of which are serine or threonine, were examined, and the upstream sequences involved in the developmental regulation of the expression of the gene have been determined. The CelB protein does not demonstrate cellulase activity, but it has a cellulose-binding domain. Its role, if any, in degradation of the cellulose-containing spore wall is unknown. To identify cis-acting elements in the celB promoter, unidirectional 5′ deletions of the celB upstream noncoding region were constructed and used to transform amoebae. Analysis of promoter activity during different stages of development shows that a short, very A/T-rich sequence of approximately 81 base pairs is sufficient for spore-specific celB transcription. Contained in this sequence is the Myb oncogene protein binding site, TAACTG, which was shown previously to be a negative regulator of celA transcription. Dictyostelium and mouse Myb proteins bind to this region of the promoter, suggesting that Myb might regulate celB gene expression negatively as it does in celA.

Dictyostelium discoideum is a favorable organism for study-

* 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 1734 solely to indicate this fact.

‡ Present address: Pulmonary Research Laboratory, Cornell University Medical College, 520 East 70th St., New York, NY 10021.
§ Present address and to whom correspondence should be addressed: Dept. of Anatomy and Cell Biology, College of Physicians and Surgeons of Columbia University, 630 West 168th St., New York, NY 10032. Tel.: 212-305-1478; Fax: 212-305-3970; E-mail: he28@columbia.edu.


ging macromolecular events coincident with and necessary for eukaryotic development. Depletion of the food supply sets in motion an orderly succession of developmental events (1). New programs of gene expression allow individual vegetative cells to enter the multicellular mode and subsequently form the fruiting body which contains two morphologically distinct cells, those in the stalk and in the spores. Similar to other stages of Dictyostelium development, spore germination is a particularly suitable model for studying regulation of gene expression. The synchronous transition from spore to amoeba is accompanied by developmentally regulated changes in both protein and mRNA synthesis and requires ongoing protein synthesis (1–5).

A number of spore germination-specific cDNAs have been isolated previously (6). One of these, the 270 gene family, is a group of four genes defined by the presence of a common tetrapeptide repeat of Thr-Glu-Thr-Pro (7). Two of the members of the family, celA (formerly called 270–6) and celB (formerly 270–11) are expressed solely and coordinately during spore germination, the levels of the respective mRNA being low in dormant spores, rising during germination to a maximum level at about 1.5–2 h, and then rapidly declining as amoebae are released from spores (7). The mRNAs are not found in growing cells or during multicellular development. The rapidity with which these transcripts accumulate and then disappear during germination implies that the respective products may be important for the process.

We reported previously that the purified CelA protein is a cellulase and is composed of three separable functional domains, catalytic, spacer containing the amino acid repeat and a cellulose-binding domain (8). We have also defined a 5′ upstream region of the gene that is important for its developmental regulation during spore germination (9). In the present investigation, properties of the CelB protein, a glycosylated protein of 532 amino acids 36% of which are serine and threonine, were examined, and the upstream sequences involved in the developmental regulation of the expression of the gene have been determined.

EXPERIMENTAL PROCEDURES

General Methods—Methods for growing D. discoideum, preparation of spores, germination of spores and isolation of DNA and RNA have been described earlier (10). Stable transformants of amoebae were made as described previously (11). RNA blot analysis was performed as described (10).

Construction of celB Expression Vectors—The D. discoideum transformation vector pDN3367 (12), which contains an actin 6 promoter that is turned on during vegetative growth, was provided by Dr. C. Klein, St. Louis University. The desired portion of celB coding region (GeneBank™/EBI Data Bank accession number J02916) was constructed using PCR and subcloned downstream of the actin promoter using BamHI and XhoI sites (underlined in the PCR primers). For both the full-length (1596 bp) and truncated (1011 bp) celB coding sequence, the 5′ primer was 5′-GGGCGGGATCCGAAATTATAATATATGTTATTC-3′.

The abbreviations used are: bp, base pair(s); CAT, chloramphenicol acetyl transferase; ConA, concanavalin A; EMSA, electrophoretic mo-
Dictyostelium discoideum Cellulose-binding Protein

**Fig. 1. Construction of Dictyostelium celB transformants.** Schematic representation of the vegetative expression constructs pDNeo67.CelB and pDNeo67.CelB-T. PCR primers containing BamHI and XhoI sites were used to produce celB-coding sequence for insertion into the Dictyostelium integrating transformation vector pDNeo67. In these constructs, the act 6 promoter drives expression of the inserted celB coding sequence. For pDNeo67.CelB, the coding sequence extends from nt 1 to 1596 and for pDNeo67.CelB-T from nt 1 to 1011.

**Fig. 2. RNA blot analysis of celB transformants.** Ten μg of total RNA isolated from vegetatively growing amoebae transformed with each of the indicated constructs was size-fractionated on an agarose-formaldehyde gel, transferred to nitrocellulose, and hybridized to a random primer-labeled full-length celB cDNA clone. The size markers are ribosomal RNA.

transformed amoebae. Transcripts of 1600 and 1250 nt for celB and celB-T, respectively, were found, consistent with the sizes of the cDNAs that were cloned (1596 and 1011 bp, respectively). No celB-specific transcript was present in either untransformed amoebae or in amoebae transformed with only the pDNeo67 vector (data not shown).

**Synthesis and Secretion of CelB and CelB-T Protein by Transformed Amoebae—**The deduced amino acid sequence of the CelB N terminus suggested the presence of a signal peptide sequence (7). Consequently, the media of vegetatively growing amoebae stably transformed with celB expression constructs was examined by SDS-PAGE for the presence of secreted proteins (Figs. 3 and 4). Amoebae containing constructs pDNeo67.CelB and pDNeo67.CelB-T secreted single predominant polypeptides of apparent molecular weights of 82,000 and 47,500, respectively. N-terminal amino acid analysis of the CelB protein gave the protein sequence NNAFI. The N-terminal sequence of the deduced CelB protein was MKNIYSLFLLFLASATFANNAFI etc. (7). Consequently, the first 19 amino acids were not present in the secreted protein. No CelB was associated with amoebae (data not shown). The molecular weights of the deduced CelB and CelB-T proteins on the basis of amino acid content were 55,205 and 34,611, respectively, and the respective proteins minus the first 19 amino acids were 53,059 and 32,465. Thus, the secreted proteins were larger than their predicted molecular masses.

**CelB and CelB-T Are Glycosylated—**CelA protein is larger than its predicted molecular weight because it is glycosylated (8, 14). To investigate whether CelB proteins were glycosylated, three different methods were used. First, concentrated supernatants from growing transformed amoebae were passed over ConA-agarose columns, and bound proteins were then eluted with methyl-D-mannopyranoside. SDS-PAGE analysis of the fractions showed that CelB bound to ConA, whereas CelB-T did not (Fig. 3). Compare lanes 5 and 6 (CelB unbound and bound fractions, respectively) with lanes 8 and 9 (CelB-T unbound and bound fractions, respectively). CelA protein, which has been shown previously (8, 14) to bind to ConA, was included as a control (lanes 2 and 3). These results immediately indicated that CelB was glycosylated and that it contained sugars necessary for ConA binding. However, lack of binding to ConA of CelB-T did not mean that the protein was not glycosylated because the glycosyl moiety could contain sugars not recognized by ConA.

The next method used was reaction of proteins with periodic acid-Schiff reagent (15). The results of this experiment showed that both CelB and CelB-T reacted with this reagent, indicating that both proteins were glycosylated (Fig. 4). It is notewor-
protein to bind to Avicel was determined. It is clear from the results shown in Fig. 7 that CelB but not CelB-T bound to cellulose. However, we have been unable to find an enzymatic activity for CelB. The protein did not hydrolyze a large number of carbohydrates tested, including carboxymethyl cellulose, crystalline cellulose, microcrystalline cellulose, xylose, barley glucan, amorphous avicel, laminarin, and cellulose purified from D. discoideum spores (data not shown).

Transcription Start Site—The 5'-untranslated region of celB was not known. To understand the transcriptional regulation of CelB expression, we had to determine this sequence. Consequently, a genomic clone containing the 1.2-kbp upstream sequence was isolated from a genomic library and characterized (Fig. 8). As a first step, the transcription start site was determined. Primer extension analysis of RNA isolated from 1.5-h germinating spores showed that transcription was initiated from nine adjacent sites located 215 bases upstream of the translation start site (Fig. 9). A consensus TATA box sequence was located upstream of the transcriptional start sites at −29 to −24.
Identification of Regulatory Sequences in the celB Promoter—To characterize regulatory sequences that determine spore germination-specific expression of celB, the 1.2-kbp upstream region was cloned into the reporter construct pAVCATII (−984) such that the cloned celB upstream region drives expression of the plasmid-borne CAT coding sequence. Amoebae transformed with this promoter construct were allowed to develop and form spores, spores were collected and induced to germinate, and CAT activity in 1.5-h germinating spores was measured. As shown in Fig. 10, the −984 construct possessed spore germination-specific promoter activity.

To identify cis-acting elements in the celB promoter, unidirectional 5’ deletions of construct −984 were made. Spore germination-specific promoter activity was measured in the transformants of the deletion promoter constructs as described above. Constructs −684 and −494 were about 24- and 4-fold more active, respectively, than the −984 construct (Fig. 10).

Since this vector inserts multiple copies in the Dictyostelium genome, the copy number in all transformants was determined and found to be about equal in all (data not shown). Therefore, the difference in activity is not due to different copy number of the integrated construct. Construct −420 showed the same activity as −984. When the deletion extended to −242, spore germination-specific promoter activity was lost. These data suggested that a critical sequence that was required for celB transcription was present in the 178 bp located between −242 and −242. In support of this proposal was the observation that an internal deletion of this region Δ −240/−242 rendered the construct inactive.

A data base search for the presence of known transcription factor binding sites in the celB promoter showed a Myb binding site TAACTG (19) in reverse orientation just upstream of −420. The Myb gene has been shown to be present and expressed in Dictyostelium (19). An internal deletion of sequences containing the Myb binding site, in addition to another sequence (Δ −494/−242), was prepared and analyzed. The transformant containing this construct was as active as the −984 construct. An EMSA was performed to determine whether Myb protein bound to this sequence. The probe was a 32P-labeled oligonucleotide corresponding to the sequence −436 to −416 of the celB promoter (see Fig. 8). Bacterially synthesized D. discoideum Myb and mouse Myb were used. The Dictyostelium protein was the entire Myb sequence while the mouse protein was a truncated version lacking the C terminus but containing the DNA binding and transactivating domains (9). As shown in Fig. 11, both Dictyostelium and mouse Myb bound to the celB promoter containing the MRE (lanes 6 and 2, respectively), and excess unlabeled probe competed for binding to the complex (lanes 7 and 3, respectively). However, excess unlabeled MRE mutant sequence (see “Experimental Procedures” for sequence), which was previously shown not to bind Myb, did not compete (lanes 4 and 3, respectively), demonstrating that the complex formed was specific. An extract of bacteria that did not

**Fig. 7. Binding of CelB proteins to cellulose.** Binding assays were performed as described previously using Avicel (8). The unbound and bound proteins were concentrated by filtration, fractionated on an SDS-polyacrylamide gel, and stained with Coomassie Blue. Lane 7, CelB concentrate; lane 3, CelB-T concentrate; lane 6, CelB-unbound fraction; lane 5, CelB bound fraction; lane 2, CelB-T-unbound fraction; and lane 1, CelB-T-bound fraction. Protein standards are shown in lane 4 in thousands.

**Fig. 8. Nucleotide sequence of D. discoideum celB promoter.** Transcription start sites are boxed. The most upstream transcription start site is designated +1. The putative TATA box is underlined. The sequence from +45 to ATG codon has been published previously (7). The MRE sequence is bold. Down arrows indicate 5’ ends of the deletions. Up arrows indicate the end points of internal deletions. (See also Fig. 10.)
contain Myb did not form a complex (lane 5).

The promoter construct −420 was tested for promoter activity during different developmental stages to verify that it contained sequences necessary for the proper temporal regulation of celB expression. The data presented in Fig. 12 show that the gene was not transcribed during vegetative growth or 5- and 16-h multicellular development but was expressed during spore germination in a similar manner as the normal gene. The other constructs that displayed activity were also shown to be active during germination but not in growing amoebae (data not shown).

DISCUSSION

The present study encompasses two aspects of CelB expression, the characterization of the protein and the definition of the gene promoter. The work may have special significance because the narrow window of gene expression intimates importance of the protein during this restricted portion of the organism life cycle. Consequently, an understanding of the role the protein plays in development may be a useful model of how other genes, similarly expressed during other stages of organism development, regulate their expression for use exclusively at a particular time.

CelB protein has an uncommon sequence organization (18). Its 532 amino acids can be arranged into five distinct portions of approximately similar size. The N-terminal 110 amino acids is followed by a 118-amino acid segment of mostly threonine and serine, after which is 104 amino acids with homology to bacterial cellulose-binding domains, connected to a 104-amino acid region containing the Thr-Glu-Thr-Pro repeat, and finally the C-terminal 96 amino acids that also show good homology to bacterial cellulose-binding domains. The two proline-, threonine-, glutamic acid-, and serine-rich segments correspond to the linker sequences connecting domains in β-1,4-glucanases (20).

To facilitate the isolation of large amounts of CelB protein, expression vectors were constructed that allowed the synthesis and secretion of full-length (532 amino acids) and truncated protein lacking the 195 C-terminal amino acids during growth of transformed amoebae, a stage in the life cycle which does not normally synthesize CelB. The N terminus of the secreted protein was shown to lack the first 19 amino acids of the deduced protein, defining the signal sequence.

Both proteins were larger than their predicted molecular weights based on their amino acid content because of O-linked glycosylation in different portions of the protein. The full-length protein was glycosylated in the Thr-Glu-Thr-Pro repeat region as well as in an undefined more N-terminal region. The truncated protein, which lacks the amino acid repeat region was also glycosylated. The two regions obviously contained different glycosylations since the full-length protein bound to ConA, whereas the truncated protein did not. Both glycosylations

---

![Image](http://example.com/image1)

**Fig. 9. Identification of celB transcription start sites.** Primer extension analysis was carried out using a 32P-labeled synthetic oligonucleotide primer corresponding to the +55 to +81 5'-untranslated sequence of the celB gene (5'-ATTAAATTACTACAAAAAGAAATAAT-3'), and mRNA isolated from 1.5-h germinating spores sequencing reactions generated with the same primer using a −984 celB promoter construct are shown as A, T, G, and C. Primer extension product is marked P.

![Image](http://example.com/image2)

**Fig. 10. Analysis of celB promoter constructs.** A, schematic representation of celB promoter deletion constructs. Constructs were made as described under “Experimental Procedures.” The thin line represents the celB 5'-untranslated region, the arrow represents the transcription start site, and the stippled box represents the CAT coding sequence. Numbers to the left indicate deletion end points. Internal deletions are shown by broken lines. Numbers to the right are promoter activity (% of [14C]chloramphenicol converted). B, the ability of these constructs to drive CAT gene expression in 1.5-h germinating spores was assayed as described previously (9). An equal amount of protein was used for each sample in the assay.
contained the MUD50 epitope, which is an O-linked oligosaccharide (16).

The C-terminal 195 amino acid region contained a cellulose-binding domain since the full-length protein bound avidly to crystalline cellulose whereas the truncated protein did not. Meinke, et al. (18), predicted two cellulose-binding domains in CelB, on the basis of sequence homology to other cellulose-binding domains, one at the C terminus at around amino acid 460 to 503, and the other internal at about 270 to 302. These regions showed good homology to bacterial cellulose-binding domains. Our data showed that the C-terminal region did, in fact, contain a binding domain, but the other predicted sequence did not, under the conditions of our experiment. (Fig. 7).

We were unable to demonstrate enzymatic activity for the CelB protein even though many different substrates were used. The most likely portion of the protein that could contain catalytic activity would be the N-terminal 110 amino acids. However, it is unlikely that this region has activity since the smallest cellulase catalytic domain described thus far is about 280 amino acids (21). Cellulose-binding proteins lacking enzymatic activity have been found in other organisms, but they are usually a component of a multi-subunit complex (22). Although the cellulose-binding protein has no apparent enzymatic activity, it is nevertheless necessary for converting crystalline cellulose to a form that can be hydrolyzed by catalytic subunits in the cellulase complex. Other studies (23, 24) have shown that the cellulose-binding domain of a bacterial cellulase bound to crystalline cellulose, disrupted the structure of cellulose fibers and released small particles. There is synergy between the binding protein and the catalytic domain, which might explain why most cellulases contain both catalytic and cellulose-binding domains on the same protein. A conclusion from this work was that the cellulose-binding domain disorders the substrate to allow the catalytic domain easier access, and in this way allows for better hydrolysis of cellulose. Perhaps CelB protein binds to cellulose and makes the substrate more accessible to CelA cellulase. Because CelA and CelB are expressed coordinately during spore germination, it is tempting to suggest that these two proteins cooperate to digest the spore wall to enable the amoeba to emerge from the spore.

Because celB is expressed during a narrow window of the D. discoideum life cycle, study of the mechanism by which promoter sequences can regulate the gene transcription should be important for understanding how specific sequences can regulate both the temporal and transcriptional control of the gene. We therefore set out to clone, sequence, and characterize the celB promoter. The celB promoter sequence was unusually AT-rich with several homoplastic tracts of A or T. Transcription was initiated from eight A residues interspersed with one T residue. All the start sites seemed to be utilized at about equal frequency, as assessed by the intensity of the primer extension bands.

The parental promoter construct −984 possessed regulatory sequences essential for spore germination-specific expression of celB. Deletion of sequences between −984 and −684 resulted in a 24-fold increase in celB promoter activity, indicating that there was a negatively acting sequence located between −984 and −684 (90% AT). Close inspection of this region revealed the presence of an A-rich (96% A) sequence located between −858 and −804, including a homoplastic run of 24 A residues. Another interesting feature of this sequence was the presence of a T-rich (91% T) sequence located between −757 and −687. The almost homoplastic tract of T residues was punctuated by A residues occurring once in nine or ten residues. This arrangement of sequence might place the A residue on the helical face of the DNA. Similar homoplastic tracts of A and T residues were present several times throughout the celB promoter sequence. Perhaps this AT-rich sequence can act negatively or positively by placing structural constraints on the chromatin. Deletion of sequences up to −494 and −420 gradually reduced celB promoter activity compared with the −684 construct. Finally, deleting the 178-bp sequence between −420 and −242 resulted in loss of celB promoter activity during spore germination. This was further confirmed by the observation that an internal deletion of this 178-bp sequence also led to loss of celB promoter activity (Δ−420/−242; Fig. 10).

An internal deletion of sequences from −494 to −420 restored promoter activity to the −984 level, suggesting that another negatively acting sequence element was located between −494 and −420. This sequence included a perfect MRE in a reverse orientation. In celA, a gene that is coordinately expressed with celB, the MRE played a dual role in regulation of gene expression. In the absence of an adjacent AT-rich sequence, the MRE was required for correct gene expression while deletion of the MRE alone stimulated celA promoter activity (9). Since Dictyostelium Myb protein binds to the celB MRE (Fig. 11), it is a reasonable suggestion that Myb might regulate celB gene expression negatively as was observed in celA regulation (9).

There was a high homology at the nucleotide level (57%) between the promoter sequences of these two genes. The highly conserved sequences included a tract of 70 T-rich residues, several runs of A residues, and the presence of the MRE. The high sequence homology and presence of similar regulatory sequences in both celA and celB promoters suggested that expression of celB, which is expressed at the same time as celA,
is possibly regulated by a similar mechanism. Since celB is thought, but not yet proved, to perform an accessory function for breakage of the spore wall during spore germination along with celA, it is conceivable that this family of genes is coordinately regulated by a common set of transactivating factors.

Acknowledgments—We thank Dr. Stephen Alexander, for help in performing the experiment described in Fig. 6, and Drs. Kalyan Ganguly and Joseph S. Lipsick, for Myb samples.

REFERENCES
1. Loomis, W. F. (1982) in The Development of Dictyostelium discoideum, pp. 3–10, Academic Press, New York, NY
2. Cotter, D. A., and Raper, K. B. (1966) Proc. Natl. Acad. Sci. U. S. A. 69, 1402–1412
3. Giri, J. G., and Ennis, H. L. (1977) Biochem. Biophys. Res. Commun. 77, 282–289
4. Giri, J. G., and Ennis, H. L. (1978) Dev. Biol. 67, 189–201
5. Dowbenko, D. J., and Ennis, H. L. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 1791–1795
6. Kelly, L. J., Kelly, R., and Ennis, H. L. (1983) Mol. Cell. Biol. 3, 1943–1948
7. Giorda, R., Ohmachi, T., Shaw, D. R., and Ennis, H. L. (1990) Biochemistry 29, 7264–7269
8. Ramalingam, R., Blume, J. E., and Ennis, H. L. (1992) J. Bacteriol. 174, 7834–7837
9. Ramalingam, R., and Ennis, H. L. (1995) Nucleic Acids Res. 23, 3018–3025
10. Giorda, R., and Ennis, H. L. (1987) Mol. Cell. Biol. 6, 2097–2103
11. Nellen, W., Dutta, S., Reymond, C., Siversten, A., Mann, S., Crowley, T., and Firtel, R. A. (1987) Methods Cell Biol. 28, 67–100
12. Da Silva, A. M., and Klein, C. (1990) Dev. Biol. 140, 139–148
13. May, T., Kern, H., Muller-Taubenberger, A., and Nellen, W. (1989) Mol. Cell. Biol. 9, 4653–4659
14. Blume, J. E., and Ennis, H. L. (1991) J. Biol. Chem. 266, 15432–15437
15. Segrest, J., and Jackson, R. (1972) Methods Enzymol. 28, 54–63
16. Alexander, S., Smith, E., Davis, L., Gooley, A., Por, S. B., Browne, L., and Williams, K. L. (1988) Differentiation 38, 82–90
17. Alexander, A., and Elder, J. H. (1982) Methods Enzymol. 170, 505–518
18. Meinke, A., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr., and Warren, R. A. J. (1991) Protein Sequences Data Anal. 4, 349–353
19. Stober-Grasser, U., Brydolf, B., Bin, X., Grasser, F., Firtel, R. A., and Lipsick, J. S. (1992) Oncogene 7, 589–596
20. Gilkes, N. R., Henriassat, B., Kilburn, D. G., Miller, R. C., Jr., and Warren, R. A. J. (1991) Microbiol. Rev. 55, 303–315
21. Wong, W. K. R., Gerhard, B., Guo, Z. M., Kilburn, D. G., Warren, R. A. J., and Miller, R. C., Jr. (1986) Gene (Amst.) 44, 315–324
22. Shoseyov, O., and Doi, R. H. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 2192–2195
23. Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Jr., Warren, R. A. J., and Kilburn, D. (1991) Bio/Technology 9, 1096–1099
24. Din, N., Damude, H. G., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. J., and Kilburn, D. G. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 11383–11387