

The molecular basis for alternative splicing of the CABP1 transcripts in *Dictyostelium discoideum*[§]

Caroline E. Grant⁺, Gerard Bain and Adrian Tsang^{*}

Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada

Received June 13, 1990; Accepted July 30, 1990

EMBL accession no. X52688

ABSTRACT

We have determined the nucleotide sequence of the CABP1 gene from *Dictyostelium discoideum*. Together with previous data on cDNA sequences, we establish that alternative splicing of transcripts derived from this gene is responsible for the production of the two CABP1 subunits. RNA blot analysis suggested that alternative splicing of the CABP1 transcripts occurs during growth and throughout development. In addition, we have compiled the intron sequences of *Dictyostelium* pre-mRNAs and observed that the GUAAGU hexanucleotide at the 5' splice site is highly conserved. The 5' splice site of CABP1 deviates from the consensus hexanucleotide in having a sequence of GUAUA. To assess the role of the modified 5' splice on differential splicing, we have constructed an actin-CABP1 fusion gene and transformed it into *Dictyostelium* cells. Analysis by immunoprecipitation with anti-CABP1 antibody and amplification of specific cDNAs by polymerase chain reaction show that the transcripts generated by the fusion gene are alternatively spliced. When the 5' splice site of the fusion gene is mutated to conform to the consensus sequence, the resulting transcripts are constitutively spliced. These observations suggest that changes in positions 5 and 6 of the donor splice site are involved in the alternative splicing of the CABP1 transcripts.

INTRODUCTION

Most eukaryotic pre-mRNAs contain intervening sequences (introns) which do not form part of the mature mRNAs. The junctions of the introns are demarcated by consensus sequences which are required for intron recognition. Removal of the introns constitutes an essential step in the regulation of gene expression (reviewed in Ref. 1). In some genes, however, an intron may be retained in a subpopulation of mature mRNA. The alternative splicing of a pre-mRNA generates mRNAs with different primary sequences (reviewed in Refs. 2 and 3). Many differentially spliced mRNAs involve the utilization of different transcription start sites or different polyadenylation sites. In a few cases, an internal

translatable sequence is alternatively spliced to yield protein isoforms (3–5).

Despite the identification of a large number of protein coding genes, introns in *Dictyostelium discoideum* remain poorly characterized. We have previously shown that the cAMP binding protein CABP1 in *D. discoideum* is composed of two subunits, CABP1A and CABP1B (6). In strains NC4 and AX2, the molecular weights of CABP1A and CABP1B are 43,000 and 38,000 respectively (7). The cDNAs which encode the two subunits have identical nucleotide sequences except that the CABP1A cDNA possesses an additional 111 bp. Genomic analysis demonstrated that the two subunits are encoded by a single gene (8). These data suggest that the CABP1 pre-mRNA is alternatively spliced. Without knowledge on the sequence organization of the CABP1 gene and the characteristics of *Dictyostelium* introns, we were unable to assess the pattern of differential splicing. In the present report, we have isolated and determined the nucleotide sequence of the CABP1 gene and its flanking regions. By comparing the CABP1 intron to the known *Dictyostelium* introns we show that the 5' splice site of CABP1 deviates from the consensus sequence. Moreover, we demonstrate that differences in the 5' splice site are responsible for the alternative splicing of the CABP1 transcripts.

MATERIALS AND METHODS

Culture conditions for *Dictyostelium discoideum*

The wild-type strain NC4 and its axenic derivative, AX2, were used throughout this study. Amoebae of AX2 were grown in suspension of HL5 medium (9), and cells of NC4 and transformants were cultured on lawns of *Enterobacter aerogenes* on SM plates (10).

Construction and screening of genomic DNA library

Genomic DNA was isolated from AX2 cells, digested partially with the restriction endonuclease *Sau3A*, and then size-fractionated on sucrose density gradients (11). Fragments with an average size of 10kb were pooled and ligated into the *Bam*HI site of the phage vector L47.1 (12). The recombinant phage were screened with radiolabelled CABP1 cDNA as a probe (8).

* To whom correspondence should be addressed

⁺ Present address: Cancer Research Laboratories, Botterell Hall, Queen's University, Kingston, Ontario K7L 3N6, Canada

[§] The first two authors contributed equally to this work

Restriction fragments of *Dictyostelium* DNA from the positive recombinants were subcloned into the Bluescript vectors (Stratagene) for sequence analysis.

Construction of actin-CABP1 fusions

To express CABP1 under the control of the actin 15 promoter in *Dictyostelium* cells, a cDNA clone encoding CABP1A (8) was digested with *Sma*I and *Eco*RV to release a fragment containing the entire open reading frame present in the cDNA plus a small amount of sequence derived from the polylinker of the plasmid vector. After the addition of *Hind*III linkers, this fragment was inserted into the *Hind*III site of the expression vector pB10act15BKH (a gift from Dr. J.G. Williams, Ref.13). Constructs containing the insert in the correct orientation for expression were identified by restriction endonuclease analysis.

To construct a plasmid expressing a mutated CABP1 under the control of the actin 15 promoter, oligonucleotide-directed mutagenesis (14) was employed to alter the 5' splice site present in the CABP1A cDNA from GTAATA to GTAAGT. The presence of the desired mutation was confirmed by DNA sequence analysis. This construct is identical to the plasmid expressing the wild-type CABP1A cDNA except for the two-nucleotide change introduced by mutagenesis.

Transformation of *Dictyostelium*

Plasmid DNA was introduced into strain AX2 by the calcium phosphate coprecipitation technique as described (15). Stable transformants were selected in HL5 medium (9) containing the antibiotic G418 (Gibco) at a concentration of 20 μ g/ml.

Determination of nucleotide sequence

All sequencing was done using a modification (16) of the chain termination method (17). Staggered deletions were generated with the DNase I procedure (18).

Metabolic labelling and immunoprecipitation

Transformants were grown on bacterial lawns, washed free of bacteria by centrifugation with 20 mM potassium phosphate, pH 6.2 (KKP), and allowed to develop on filters saturated with KKP (10). Each filter (47 mm diameter, Nuclepore) contained 5×10^7 cells. Following a 3h incubation at 22°C, the filters were transferred to a clean petri dish and then 100 μ l of KKP containing 20 μ Ci of [³⁵S]methionine (Amersham) were added directly to them. After a 3h labelling period, the cells were dislodged from the filters by vigorous shaking in 10 ml of KKP and collected by centrifugation at 450 \times g. The cells were lysed and prepared for immunoprecipitation with anti-CABP1 monoclonal antibody B9 as described (6). The resulting polypeptides were analyzed by SDS polyacrylamide gel electrophoresis and detected by fluorography (19).

Amplification of specific DNA fragments by polymerase chain reaction (PCR)

Sequences complementary to the actin-CABP1 fusion transcripts were amplified as follows. Cytoplasmic RNA was isolated from cells which had developed for 2–4h. First strand cDNA was synthesized from 2 μ g of total RNA using the P2 oligonucleotide primer (5'GGAAGAGCGCCTTGATACC 3') which is complementary to a segment near the 3' end of the CABP1 transcripts. The reaction was carried out with 5 units of AMV reverse transcriptase (Promega) at 42°C for 1h. The enzyme was then inactivated by a 5 min incubation at 95°C, and the reaction

mixture was diluted 5-fold with 50mM KCl, 10mM Tris-HCl pH 8.3, 1.5mM MgCl₂, and 0.01% gelatin. A second oligonucleotide primer, P1 (5'CAAGCTTGGATCGAATTC-CGGTATA 3'), corresponding to the 5' end of the actin 15-CABP1 fusion gene was added along with 2 units of *Taq* DNA polymerase (Pharmacia). The reaction was first incubated at 94°C for 1 min, then 1 min at 58°C, and finally for 2 min at 72°C. A total of 35 cycles of amplification were performed. To amplify the actin-CABP1 fusion DNA in the transformants, 5 ng of genomic DNA were used as template. The conditions for amplification were the same as described above except that the steps involving the synthesis of first strand cDNA were omitted. The amplification products were analyzed by electrophoresis through 1.2% agarose gels followed by DNA blot analysis.

Nucleic acids blot analyses

DNA and RNA blot analyses were conducted as described (8).

RESULTS

Nucleotide sequence of the CABP1 gene

We have found that the two subunits of CABP1 are encoded by a single gene (8). To examine the mechanism by which this is accomplished, we have isolated by molecular cloning the CABP1 gene. Two genomic clones were obtained. Analysis by restriction mapping demonstrated that the 6 kb genomic clone represents an internal fragment of the 8 kb clone (Fig. 1). The restriction map of the genomic clones corresponds well with the pattern of hybridizing bands observed on a genomic DNA blot (data not shown). We have subcloned two overlapping restriction fragments, a 4 kb *Eco*RI fragment and a 5.8 kb *Xba*I fragment, into Bluescript vectors for sequence analysis. The entire CABP1 gene is located within the 2.2 kb *Xba*I-*Eco*RI fragment, and the nucleotide sequence of this fragment is shown in Fig. 2.

A single open reading frame (ORF) is present within the *Xba*I-*Eco*RI fragment. In most features, the CABP1 gene resembles a typical *D. discoideum* gene (20,21). The sequences flanking the CABP1 gene are rich in A+T, 88% at the 5' side and 84% in the 3' side. Primer extension experiments showed that

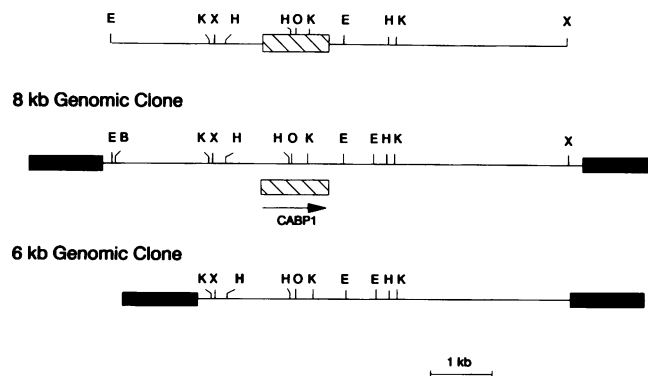


Figure 1. Restriction enzyme cleavage maps of CABP1 genomic clones. The genomic map was based on DNA blot analysis with the 0.67 kb CABP1 cDNA as probe (8). The maps of the 2 genomic clones were constructed by determining the sizes of various restriction enzyme fragments as well as by DNA blot analysis. The solid boxes represent the vector sequences. Stippled boxes delineate the CABP1 gene. The arrow indicates the direction of transcription. The restriction enzyme sites are: B, *Bam*HI; E, *Eco*RI; H, *Hin*fi; K, *Kpn*I; X, *Xba*I; and, O, *Xho*II. The maps are drawn only approximately to scale.

transcription starts at about 35 nucleotides upstream of the ATG codon (data not shown). Approximately 400 bp upstream of the translation initiation codon are two stretches of G-rich sequences. Similar G-rich sequences have been reported to be essential for promoter function (22–24). The ORF ends in three in-frame TAA stop codons. Fifty and 62 nucleotides downstream from the first translation stop codon are two consensus AATAAA polyadenylation signals. Comparison with cDNA sequences suggests that the poly(A) tails are added about 35 nucleotides after the first polyadenylation signal. Thus, the 3' noncoding region is approximately 85 nucleotides in length.

Previous results suggest that the two CABP1 subunits are encoded by two mRNA populations. The larger transcript codes for CABP1A while the smaller transcript codes for CABP1B (8). The present analysis shows that the sequence of the ORF in the *Xba*I-*Eco*RI fragment is identical and colinear with the CABP1A cDNA, whereas the CABP1B cDNA lacks a 111 nucleotide

```

TCTAGAAAAAAAAATATAATTTATTTTACAAAAAGTTGGTGATTAATAAATCATATA 63
ATTCAAAAAATAAAAAAAAAAAAAAAAAAATAACAATTACAATAAATCATTTCAGTGCAAT 126
AAATAACTCTTTAATAAATTAAGGAAAAATGATTTTTTGTGGTTTAAAAATCTTTTTTT 189
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAATTTTCAGACTCTGGGTGTAGACAAGA 252
AATAAAAAATTTAAAAAATAAATAAATAAATGATCATTACCAAGTTTTTTTTAAAAATAATTT 315
TATTGTGTTGATTAATAATTTCAATAAATTAATAAATAAATAAATAAATAAATAAATAA 378
AAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAATTTCAATGAATATATTATAATAAATTCAC 441
TCGTATTGGTGGCATTAAATAAATAAAGCAGCATTGATCATTTTTTTGTGAGGCTGACCTTAG 504
AATTTTAGACTTTTGTGAGATATCAAGCAAAAAATTTTATGAAAAAATAAATAAATAAATAA 567
AAACAATAAAAAAAAAAATAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 630
CATACACAGTGTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 693
TTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTTTAAATTTTAAATTTTATTTTT 756
TATTTATTTTTTCTTTTATTTTTTTTTTTTTTTTTTAAATTTCAAATTTAATTTTCATTTTT 819
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTAATAGAAATATAAATTTATATA 882
TATAAAAAATAAAAAATGTATAACCCACCACCACCATCTGGTTCACAGGTAATAAATAATT 945
      M Y N P P P P S G S Q G N N N Y
ATAGACAAGCATCATCCAGCCGGTGTATCAAAACCCCAAGCTAATCAATTTTTTAC 1008
Y R Q P S S T P G V S N P N P O A N Q F L
CAGCACAACCATCTAATACTACAGCAACCCAGGAGTATCCACCACAACAACAAGAGAC 1071
P Q P S N T T Q T P G G Y P P Q Q Q Q P
CACCACCTGGAGCACCACAACAACAGGAGGTTATCCAACTGCACCACCAGGAGGACCAG 1134
P P T G A P Q Q P G G Y P T P P P P G A P
GAGGTTATCCACCACAACAACAACCCGGTCAATATGGAGCACCACCACAACAACAACAGC 1197
G G Y P P Q Q Q P A G Q Y G A P P Q Q Q Q P
CCGGTCAATACGGAGCACCACAACAACAGGAGGTTATCCAACTGCACCACCAGGAGGACAAT 1260
A G Q Y G A P Q P A G Q Y G A P Q P A G Q
ATGGTGACCACAACCCAGGAGGACAATATGGGGCACCACCACCAGGAGGAGGTCAGGTA 1323
P L P P P P A G Q Y G A P P P P P G G A G
TTTCATTAGTAAAGCAATCAAAATTTCAATTAACAAAGAGATCCAACTCTTAGAAAAATAA 1386
I S L V K N Q Q I S L T K E D P T L R K L
CAATTTGGTTAGTTGGGATGTAACACCACCACCACTGCACCATTGGATTCAGTTCAGTTG 1449
T I G L G W D V N T T P T A P F D L D A V
TTTTCATGTTGAATGCACAAGGTAGAGTTAGAACCCTACAAGATTTTCTACATAAACA 1512
V F M L N A Q G R V R T S Q D F I F Y N N
AGGTATCAAGAGATACTCTGTTCTCATCAAGGTGATAATTAACAGGTCAAGGTGAAGGTC 1575
K V S R D N S V S H Q G D N L T G Q G E G
ATGATGAAGTTGCTCCTGTAACCTACAAGCAGTTTCAACAGCAGTCACTCGTTTGGTTTTCG 1638
D D E V V L V N L Q A V S P D V T R L V F
CTGTACCACTCATTAGCCGATGAAAGAGACAAAACCTTACAATGGTACCAAGAGCTTTCA 1701
A V T I H L A D E R R Q N F T M V P R A F
TTCGGTTGCCAATCAAGAACTGGTAGAAATATCTGTCGTTACGATCTCTCAAGAAGGTC 1764
I R V A N Q E T G R N I C R Y D L S Q E G
GTCCAAACACTGCCCTCATTGCTGTTAAGTTTATCGTGATCCATCAAAATCCAAACAAATGGT 1827
G P N T A L I A G E V Y R D P S N P N N W
CTTTGTTGCTGTGGTAAAGGTATGCAAGGCGTTCTCCAGGTTTACTCCAAATCTTTGGTT 1890
S L L L V K V C K A F L P G L Q I F G
GTCAATAAATTAATATTATTAAATTTAAAAAATAAATAAATAAATAAATAAATAAATAA 1953
C Q *
CAATAAATAACTATTGTAATCTTGAATAAATTTTATTTTACTTTATATTTTTTAAAAATC 2016
CTTTTTTTTTTTTTTTTTTTTTTAAAAATTAATTTGTAATTTGATTTTTTTTTTTTTTTTTTT 2079
TTTTTTTAAATTTTAAATTCGCACATCAGCAATGATTAATTTGAAAGTATCAGATCTT 2142
TTAGGCAAATTAAGGTTCTGTAAGGTTGAATATCCAACTTACACCATTACATATACA 2205
GAAGAGATTGAATTC 2220
    
```

Figure 2. Nucleotide and deduced amino acid sequences of the CABP1 gene. The sequence of the 2.2 kb *Xba*I-*Eco*RI fragment which contains the CABP1 gene along with approximately 900 nucleotides of 5' flanking sequence and 300 nucleotides of 3' flanking sequence is shown. The region which has been mapped as the transcription start site by primer extension is indicated by a double underline. In the coding region, the nucleotide and amino acid sequences that are absent from CABP1B cDNAs are underlined. Two potential polyadenylation signals located at the 3' end of the gene are also underlined.

segment located just downstream of the translation initiation codon (Fig.3, Ref.8). Hence, it can be concluded that the CABP1B transcript is generated by the splicing of 111 nucleotides

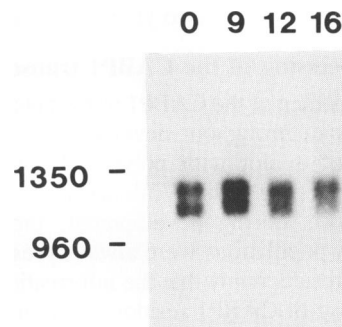


Figure 3. Expression of the CABP1 mRNAs during growth and development. Poly(A)+RNA was denatured and resolved by electrophoresis through a 1.5% formaldehyde gel. Each lane contains 2.5 µg of poly(A)+RNA. After transfer to a nitrocellulose membrane, the blot was probed with the 0.67 kb CABP1 cDNA (8). RNA isolated from vegetative cells, 0, and from cells developed for 9h, 12h and 16h are shown. Positions of size markers are indicated.

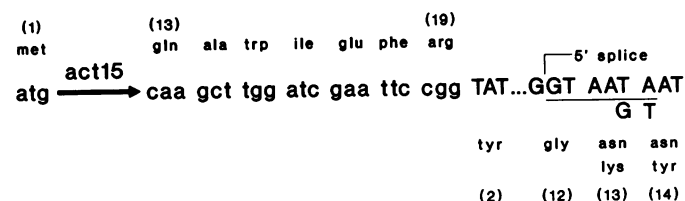


Figure 4. Actin 15-CABP1 fusion constructs. The structure of the amino terminus of the fusion polypeptides is shown. The nucleotide sequences provided by the vector are shown in small letters while those present in the CABP1A cDNA are in capitals. The numbers above the sequence represent amino acid position in the fusion while those below the sequence indicate the codon position in the CABP1A cDNA. The 5' splice site is underlined. The mutated actin 15-CABP1 fusion construct is identical to this except for the two nucleotide change introduced to convert the 5' splice site to the *Dictyostelium* consensus. These mutations are indicated underneath the underlined sequence, as are the resulting amino acid changes.

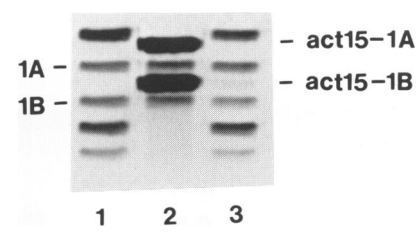


Figure 5. Immunoprecipitation analysis of transformants. After introducing the wild-type and mutant actin 15-CABP1 fusion constructs into *Dictyostelium* cells, the transformants were analyzed by metabolic labelling with [³⁵S]methionine followed by immunoprecipitation with an anti-CABP1 monoclonal antibody. The resulting polypeptides were resolved by electrophoresis through SDS polyacrylamide gels and detected by fluorography. As a control, cells transformed with the expression vector alone were treated similarly. The endogenous CABP1A and CABP1B subunits are labelled, as are the fusion polypeptides. Lane 1, immunoprecipitated products obtained from cells transformed with the expression vector pB10act15BKH; lane 2, immunoprecipitated products obtained from cells transformed with the wild-type actin 15-CABP1 construct; lane 3, immunoprecipitated products obtained from cells transformed with the mutated actin 15-CABP1 expression construct. The anti-CABP1 monoclonal antibody crossreacts with several polypeptides (6). These crossreactive polypeptides were not detected in cells transformed with the wild-type fusion gene. Presumably, the vast amounts of the fusion CABP1 polypeptides had exhausted the available antibodies, hence no crossreactive polypeptides were precipitated.

from the primary transcript. This conclusion is supported by the observation that the segment which is absent from the CABP1B cDNA is flanked precisely by the GT/AG splice consensus sequences (8). Furthermore, it appears that about half of the primary transcript is unspliced to yield the CABP1A transcript.

Alternative processing of the CABP1 transcripts

The alternative splicing of the CABP1 transcripts appears to occur during growth and throughout development. Fig. 3 show an autoradiograph of a blot with poly(A)⁺RNAs isolated from growing and developing cells. While there were fluctuations in their overall levels during development, the CABP1A and CABP1B mRNA populations were always present in a roughly 1:1 ratio. These results imply that the information regulating the alternative splicing of CABP1 resides in the primary transcript and may not involve control by *trans*-acting factors.

To test the ability of the CABP1A transcript to undergo alternative splicing, we have fused the CABP1A cDNA with the actin gene of the transformation vector. The resulting fusion

polypeptide is expected to be 20 amino acids longer than the native CABP1A (Fig. 4). The recombinant DNA was transformed into AX2 cells. Approximately 200 transformants were pooled and labelled metabolically with [³⁵S]methionine. We used immunoprecipitation for analysis because it allowed us to distinguish the fusion products from the endogenous products. Besides the endogenous CABP1A and CABP1B polypeptides, the transformants expressed large amounts of two immunoreactive polypeptides of slightly higher molecular sizes (Fig. 5, Lane 2). The two new products most likely represent the plasmid-encoded CABP1A and CABP1B polypeptides. Analysis of the transformants by PCR (Fig. 6) suggested that two distinct mRNAs are derived from the actin-CABP1 fusion gene. These observations suggest that the fusion CABP1A transcript is alternatively spliced *in vivo*.

Characteristics of *Dictyostelium* pre-mRNA introns

Results from the above experiments suggest that the CABP1A transcript has the necessary information to modulate the splicing

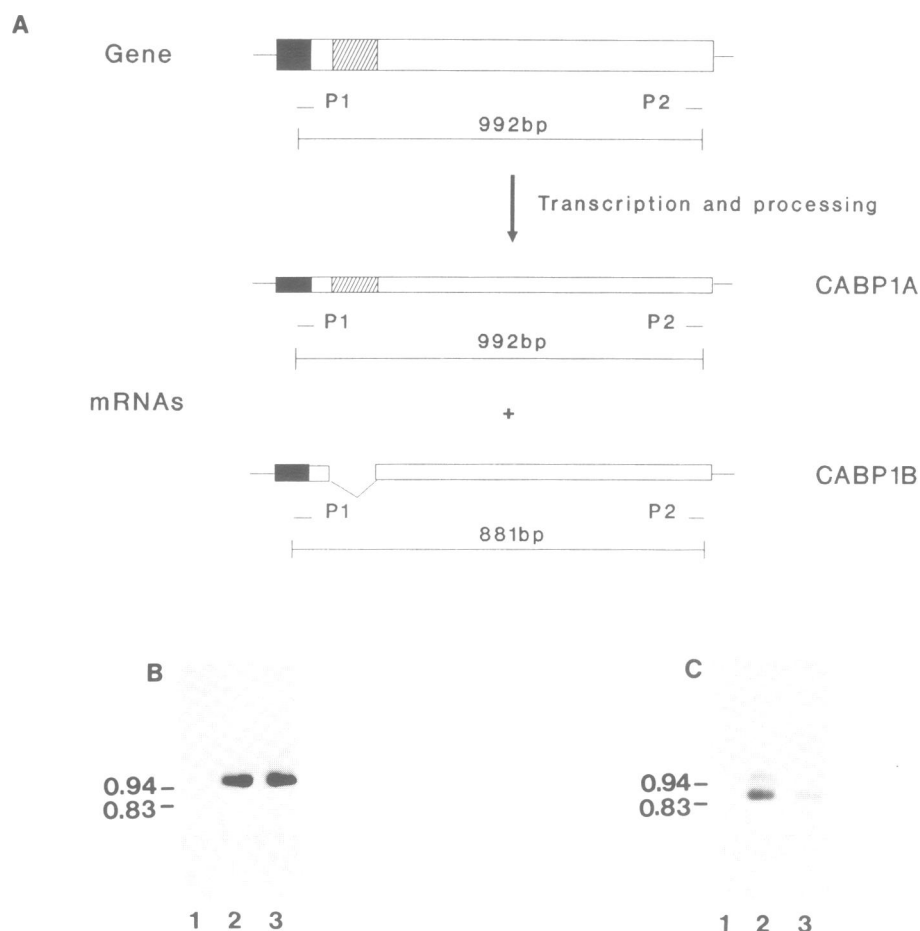


Figure 6. Analysis by PCR of actin-CABP1 fusion genes and transcripts derived from them. *A.* The strategy for analysis by PCR. Untranslated sequences are represented by a line while coding sequence is depicted by a rectangle. The sequences provided by actin 15 are represented by the solid zone. The region which is alternatively spliced to generate the CABP1B subunit is indicated by diagonal lines. The locations of the two primers used for PCR, P1 and P2, are shown. The predicted sizes of the PCR products are indicated. *B.* PCR analysis of genomic DNA isolated from various transformants. 0.01% of the PCR reactions were electrophoresed through 1.2% agarose, transferred to a nylon membrane, and probed with a CABP1 cDNA fragment. The positions of the 0.83 kb and 0.94 kb size markers obtained by digesting lambda DNA with *Eco*RI and *Hind*III are indicated. The PCR products obtained from genomic DNA isolated from cells transformed with the expression vector pB10act15BKH (lane 1), the wild-type actin 15-CABP1 fusion construct (lane 2), and the mutated actin 15-CABP1 fusion construct (lane 3) are shown. *C.* PCR analysis of RNA transcripts isolated from various transformants. The PCR reactions were analyzed as above. PCR products obtained from cytoplasmic RNA isolated from cells transformed with the expression vector pB10act15BKH (lane 1), the wild-type actin 15-CABP1 fusion construct (lane 2), and the mutated actin 15-CABP1 expression construct (lane 3) are shown.

event. Thus, the basis for alternative splicing may lie in the differences between the spliced region of CABP1 and other *Dictyostelium* introns. Until recently few intron sequences of *Dictyostelium* have been reported, and the requirements for intron recognition are not fully understood. We have, therefore, compiled the 37 *D. discoideum* intron sequences published to date. Sequences in the 5' and 3' splice sites are summarized in Table 1. From these introns a consensus sequence of /GUAAGU can be deduced for the 5' splice site. For this consensus sequence the nucleotides in positions 4 and 5 are most variable. In position 4, A is the preferred nucleotide followed by U. In position 5, G is the preferred nucleotide followed by A. The dinucleotide AG/ appear to be the only requirement for the 3' splice site. However, adjacent to the 3' splice site U appears to be the preferred nucleotide. Thus, the 5' and 3' splice sites of *Dictyostelium* introns are very similar to those of other organisms. The majority of *Dictyostelium* introns is about 100 nucleotides long. Only 5 out of 37 introns have lengths longer than 300 nucleotides. *Dictyostelium* introns are AU-rich with an average of 90%, compared to an average of 63%AU for the protein coding sequences. (The A+U content of protein coding sequences

is estimated from a list of 58 *Dictyostelium* genes cited in Ref. 41.) In addition, no polypyrimidine tract adjacent to the 3' splice site can be discerned. It is also unclear if consensus branch sites, similar to those found in vertebrate and yeast introns, exist in *Dictyostelium* pre-mRNA introns.

A modified 5' splice site is responsible for alternative processing

The spliced region of the CABP1 gene deviates from the typical *Dictyostelium* intron in two aspects. The GU dinucleotide in positions 5 and 6 of a typical 5' splice site is substituted with UA in CABP1 (Table 1, Fig. 2). The A+U content of the CABP1 intron is 63% (Fig. 2), a frequency which is significantly lower than other *Dictyostelium* introns.

Mutations in the first 6 bases of the 5' splice sites have been shown to affect the splicing efficiency of mammalian introns (42). We therefore examined the role of the modified 5' splice site in the alternative splicing of the CABP1 transcripts. The 5' splice site of the actin-CABP1 gene fusion was changed by oligonucleotide-mediated mutagenesis into the consensus GUAAGU sequence (Fig. 4). The resulting construct was

Table 1. Pre-mRNA introns of cloned *D. discoideum* genes

Gene	Number of introns	Length of intron (nucleotides)	Percentage A + U	Sequence of 5' splice site	Sequence of 3' splice site	Reference
CP2	4	102	91	GUA/GUAUGU	UAG/G	23
		81	94	UAA/GUAAGU	AAG/U	23
		103	91	UUU/GUAAGU	UAG/U	23
		127	86	UAA/GUAAGU	AAG/U	23
M4	2	112	94	UAA/GUUUGU	AAG/A	25
		93	92	AAA/GUAUGU	CAG/U	25
M3L	2	100	94	AUU/GUAAGC	AAG/U	26
		144	90	CGT/GUAUGU	CAG/G	26
M3R	2	87	93	AUU/GUAAGU	AAG/T	26
		115	86	CGU/GUAAGU	CAG/G	26
D2	1	100	85	ACU/GUAAGU	AAG/G	26
UDPGP	3	178	92	GAA/GUAAGU	UAG/C	27
		74	91	UUC/GUAAGU	UAG/A	27
		113	96	CCA/GUAAGU	AAG/C	27
29C		162	90	UUG/GUAAGU	UAG/A	28
Ddras	3	103	96	CAA/GUAUGU	UAG/A	29
		100	97	UUG/GUAAAU	UAG/A	29
		92	92	UUG/GUUUGU	UAG/A	29
		351	75	AAC/GUAGGU	CAG/U	30
PR1024	1	351	75	AAC/GUAGGU	CAG/U	30
PDE	3	488	82	AAG/GUAUGU	AAG/A	31
		1681	85	AAU/GUUUGU	AAG/A	31
		150	89	AUU/GUAAGU	UAG/A	31
SP60	1	121	91	UAG/GUAAGU	UAG/A	32
SP70	1	300	95	UUA/GUAAGU	UAG/A	32
BP74	2	308	88	GUG/GUAUGU	UAG/A	33
		116	94	GAG/GUAAAU	UAG/U	33
ambA	1	170	93	UAU/GUAUGU	AAG/A	34
A11H2	3	123	98	UGG/GUAAAU	UAG/U	35
		79	95	UUG/GUAAAU	UAG/A	35
		97	95	AUG/GUAUGU	UAG/A	35
		90	93	CUU/GUAAGU	UAG/C	36
Thy1	1	90	93	CUU/GUAAGU	UAG/C	36
		115	97	AGC/GUAAAA	UAG/A	37,38
α -actinin	2	101	86	CUU/GUAUGU	AAG/C	37,38
		101	85	AAG/GUAAUU	UAG/G	39
gp24A	1	101	88	AAG/GUAAUU	UAG/G	39
gp24B	1	101	88	AAG/GUAAUU	UAG/G	39
V14	1	104	87	CAA/GUAAGU	UAG/G	40
V18	1	66	86	AAG/GUAAAU	UAG/C	40
<i>D. discoideum</i> consensus				/GUAAGU	AG/	
Vertebrate consensus				AG/GUAAGU	AG/	1
Plant consensus				AG/GUAAGU	AG/	1
Yeast consensus				/GUAUGU	AG/	1

transformed into AX2 cells. Over 150 independent transformants were pooled and then analyzed by immunoprecipitation with an anti-CABP1 monoclonal antibody. In five separate experiments we could not detect the fusion CABP1A polypeptide. In all cases, a small amount of the plasmid-encoded CABP1B polypeptide was immunoprecipitated. The result of a typical experiment is shown in Fig. 5 (lane 3). This finding implies that all the transcripts generated by the mutated actin-CABP1 gene fusion are spliced.

It is possible, for a variety of undetermined causes, that the presence of the fusion CABP1A generated by the mutated plasmid cannot be detected by immunoprecipitation. To rule out this possibility, we analyzed by PCR the actin-CABP1 genes and the transcripts derived from them. Fig. 6B shows that the products amplified by PCR when genomic DNAs from the transformants were used as templates. Cells transformed with the wild-type actin-CABP1 yielded a single amplified product (Fig. 6B, lane 2) which was identical in size to the fragment amplified from cells transformed with the mutant gene (Fig. 6B, lane 3). This datum suggests that no gross rearrangement had occurred to either the wild-type or the mutant actin-CABP1 genes during transformation. The actin-CABP1 transcripts were then analyzed by PCR amplification of the cDNAs derived from them (Fig. 6C). Cells transformed with the vector alone did not yield amplified products (Fig. 6C, lane 1) indicating that the primers used were highly specific for the actin-CABP1 fusion genes. Cells transformed with the actin-CABP1 fusion gene produced two fragments with sizes corresponding to the fusion CABP1A and CABP1B (Fig. 6C, lane 2). Only one fragment corresponding in size to the fusion CABP1B was detected in cells transformed with the mutated fusion gene (Fig. 6C, lane 3). Even after extensive overexposure of the autoradiogram, a band corresponding to the unspliced actin-CABP1 transcript could not be detected (not shown). These results support the earlier conclusion that the CABP1 transcripts derived from the mutated fusion gene are constitutively spliced.

DISCUSSION

The role of conserved sequences at the 5' splice site in intron splicing has been characterized in detail in yeast and vertebrates. Deletion analysis showed that the hexanucleotide at the 5' splice site is important for intron recognition (42). Mutations at position 1 or 2 in the invariant GU dinucleotide abolish splicing (1,43,44). Some mutations at either of these positions allow precise cleavage at the 5' splice site and lariat formation, but prevent cleavage at the 3' splice site and exon ligation (44–46). Furthermore, mutations at position 5 can result in inaccurate splicing and mutations in position 6 can affect splicing efficiency (46,47). Based on a comparison of 37 introns present in various *Dictyostelium* genes we show that the GUAAGU hexanucleotide at the 5' splice site is highly conserved. This consensus sequence is identical to that found in most other organisms (1). The GU dinucleotide at the 5' splice site is absolutely conserved. The recent observation that a mutation at position 1 of the 5' splice site of the α -actinin gene drastically reduces splicing efficiency (37) confirms the idea that this nucleotide is an essential landmark of the donor splice site. Prior to this study, it was unclear if the 5' splice sites of *Dictyostelium* introns extend beyond the GU dinucleotide (25,26). Table 1 shows that the conserved region at the 5' splice site encompasses at least the first 6 nucleotides. Our demonstration that significant deviation from the consensus GU dinucleotide in both positions 5 and 6 leads to alternative

splicing suggests that this dinucleotide plays a role in modulating splicing efficiency.

The data in Table 1 reveals that *Dictyostelium* introns are much more similar to plant introns than to their vertebrate counterparts. Plant introns lack a polypyrimidine tract at the 3' end (48) which forms part of the recognition system in vertebrate introns (1). Moreover, the frequency of A + U of plant introns is about 20% higher than their flanking exons (49). In general, plant cells fail to splice vertebrate introns (48,50). Using model introns as substrates it was found that AU-rich sequences, and not the polypyrimidine tract, are required for intron recognition by plant cells (49). Almost all *Dictyostelium* introns have structures similar to that of plant introns (Table 1). *Dictyostelium* introns have no obvious polypyrimidine tract at the 3' end. The introns average 90% AU, compared to an average of 63% AU for the protein coding sequences. The CABP1 intron appears to be an exception to the latter rule. The frequency of A + U in the CABP1 intron is 63%, similar to an average *Dictyostelium* exon but slightly higher than the 54% AU found in the remaining translatable region of CABP1. These observations suggest that the splicing machinery in *Dictyostelium* cells, unlike that of plant cells, can recognize introns of different structures.

As in all known introns the AG dinucleotide at the 3' splice sites of *Dictyostelium* introns is absolutely conserved. In the CABP1 intron, there are two other AG dinucleotides which would potentially yield introns of 19 and 63 nucleotides in length (Fig. 2). Introns appear to have a minimum-size requirement (51). The smallest *Dictyostelium* intron observed so far is 66 bases (Table 1, Ref. 40). The minimum-size requirement may eliminate the first, but probably not the second, AG dinucleotide as the 3' splice site. We suspect, at least for CABP1, additional information encoded in the transcript is required for intron recognition. One possibility is that the CABP1 transcript requires a branchpoint sequence near the 3' splice site as in the case of yeast and vertebrate introns (1).

Cells transformed with the actin-CABP1 fusion gene express large quantities of plasmid-encoded CABP1 polypeptides, whereas cells transformed with the mutated gene yield only a small amount of the fusion CABP1B polypeptide (Fig. 5). The plasmid-encoded CABP1B transcripts produced by these two populations of transformed cells are roughly the same as estimated by PCR analysis (Fig. 6). These results can be taken to suggest that the stability of the CABP1B polypeptide is normally maintained by its association with the CABP1A polypeptide. We have immunocytochemical data which suggest that the plasmid-encoded molecules are targeted to a different cellular compartment than the endogenous polypeptides (unpublished results). Thus, the fusion CABP1B is probably unable to interact with the native CABP1A in the transformants thereby leading to its rapid turnover. If the formation of heterodimers or multimers is required for stability of the protein, then alternative splicing of the CABP1 transcripts constitutes an essential step in the regulation of CABP1 expression.

ACKNOWLEDGEMENT

This work was supported by grants from NSERC and NCI of Canada.

REFERENCES

1. Krainer, A.R. and Maniatis, T. (1988) In Hames, B.D. and Glover, D.M. (eds.). *Transcription and Splicing*. IRL Press, Oxford, pp. 131–206.

2. Leff, S.E. and Rosenfeld, M.G. and Evans, R.M. (1986) *Annu. Rev. Biochem.* **55**, 1091–1117.
3. Breitbart, R.E. Andreadis, A. and Nadal-Ginard, B. (1987) *Annu. Rev. Biochem.* **56**, 467–495.
4. Santoro, C., Mermod, N., Andrews, P.C. and Tijan, R. (1988) *Nature* **334**, 218–224.
5. Johnson, K.R., Lehn, D.A. and Reeves, R. (1989) *Mol. Cell. Biol.* **9**, 2114–2123.
6. Tsang, A.S. and Tasaka, M. (1986) *J. Biol. Chem.* **261**, 10753–10759.
7. Tsang, A.S., Kay, C.A. and Tasaka, M. (1987) *Dev. Biol.* **120**, 294–298.
8. Grant, C.E. and Tsang, A. *Gene* (in press).
9. Watts, D.J. and Ashworth, J.M. (1970) *Biochem. J.* **119**, 171–174.
10. Sussman, M. (1966) In Prescott, D. (ed.), *Methods in cell physiology*. Academic Press, N.Y., Vol. 2, pp. 397–410.
11. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
12. Loenen, W.A. and Brammar, W.J. (1980) *Gene* **10**, 249–260.
13. Chang, A.C.M., Williams, K.L., Williams, J.G. and Ceccarelli, A. (1989) *Nucleic Acids Res.* **17**, 3655–3661.
14. Zoller, M.J. and Smith, M. (1987) *Methods Enzymol.* **154**, 329–350.
15. Early, A.E. and Williams, J.G. (1987) *Gene* **59**, 99–106.
16. Stambaugh, K. and Blakesley, R. (1988) *Focus* **10**, 29–31.
17. Sanger, F., Nicklen, G. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Laughon, A. and Scott, M.P. (1984) *Nature* **310**, 25–31.
19. Skinner, M.K. and Griswold, M.D. (1983) *Biochem. J.* **209**, 281–284.
20. Kimmel, A.R. and Firtel, R. (1982) In Loomis, W.F. (ed.), *The development of Dictyostelium discoideum*. Academic Press, N.Y., pp. 233–324.
21. Kimmel, A.R. and Firtel, R. (1983) *Nucleic Acids Res.* **11**, 541–552.
22. Pears, C.J. and Williams, J.G. (1988) *Nucleic Acids Res.* **16**, 8467–8486.
23. Datta, S. and Firtel, R.A. (1987) *Mol. Cell. Biol.* **7**, 149–159.
24. Datta, S. and Firtel, R.A. (1988) *Genes Dev.* **2**, 294–304.
25. Kimmel, A.R. and Firtel, R. (1980) *Nucleic Acids Res.* **8**, 5599–5610.
26. Mann, S.K.O. and Firtel, R.A. (1987) *Mol. Cell. Biol.* **7**, 458–469.
27. Ragheb, J.A. and Dottin, R.P. (1987) *Nucleic Acids Res.* **15**, 3891–3906.
28. Ayres, K., Neuman, W., Rowekamp, W.G. and Chung, S. (1987) *Mol. Cell. Biol.* **7**, 1823–1829.
29. Reymond, C.D., Gomer, R.H., Mehdy, M.C. and Firtel, R.A. (1984) *Cell* **39**, 141–148.
30. Steel, L.F., Smyth, A. and Jacobson, A. (1987) *Nucleic Acids Res.* **15**, 10285–10298.
31. Podgorski, G.J., Franke, J., Faures, M. and Kessin, R.H. (1989) *Mol. Cell. Biol.* **9**, 3938–3950.
32. Fosnaugh, K.L. and Loomis, W.F. (1989) *Mol. Cell. Biol.* **9**, 5215–5218.
33. Hopkinson, S.B., Pollenz, R.S., Drummond, I. and Chisholm, R.L. (1989) *Mol. Cell. Biol.* **9**, 4170–4178.
34. Titus, M.A., Warrick, H.M. and Spudich, J.A. (1989) *Cell Reg.* **1**, 55–63.
35. Muller-Taubenberger, A., Westphal, M., Noegel, A. and Gerisch, G. (1989) *FEBS Lett.* **246**, 185–192.
36. Dynes, J.L. and Firtel, R.A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7966–7970.
37. Witke, W. and Noegel, A.A. (1990) *J. Biol. Chem.* **265**, 34–39.
38. Noegel, A., Witke, W. and Schleicher, M. (1987) *FEBS Lett.* **221**, 391–396.
39. Loomis, W.F. and Fuller, D.L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 886–890.
40. Singleton, C.K., Manning, S.S. and Ken, R. (1989) *Nucleic Acids Res.* **17**, 9679–9692.
41. Sharp, P.M. and Devine, K.M. (1989) *Nucleic Acids Res.* **17**, 5029–5039.
42. Wieringa, B., Meyer, F., Reiser, J. and Weissman, C. (1983) *Nature* **301**, 38–43.
43. Newman, A.J., Lin, R.-J., Cheng, S.-C. and Abelson, J. (1985) *Cell* **42**, 335–344.
44. Fouser, J.A. and Friesen, J.D. (1986) *Cell* **45**, 81–93.
45. Aebi, M., Hornig, H., Padgett, R.A., Reiser, J. and Weissman, C. (1986) *Cell* **47**, 555–565.
46. Jacquier, A., Rodriguez, J.R. and Rosbash, A. (1985) *Cell* **43**, 423–430.
47. Parker, R. and Guthrie, C. (1985) *Cell* **41**, 107–118.
48. Wiebauer, K., Herrero, J.-J. and Filipowicz, W. (1988) *Mol. Cell. Biol.* **8**, 2042–2051.
49. Goodall, G.J. and Filipowicz (1989) *Cell* **58**, 473–483.
50. van Santen, V.L. and Spritz, R.A. (1987) *Gene* **56**, 253–265.
51. Wieringa, B., Hofer, E. and Weissman, C. (1984) *Cell* **37**, 915–925.