Volume 8 Number 23 1980 Nucleic Acids Research

Intervening sequences in a Dictyostelium gene that encodes a low abundance class mRNA

Alan R.Kimmel and Richard A.Firtel

Department of Biology, University of California at San Diego, La Jolla, CA 92093, USA

Received 19 September 1980

ABSTRACT

Using S_1 nuclease protection experiments and DNA sequencing, we have identified two intervening sequences (introns) within a Dictyostelium gene that codes for a low abundance class mRNA. The two introns are located within the protein coding region of this gene. Both are small (~100bp) and extremely (~95%) A + T rich. The splice junction sequences are similar to the splice sites in other eukaryotes. Finally, we have shown that these introns are transcribed as part of a higher molecular weight nuclear precursor.

INTRODUCTION

Many eukaryotic genes are not co-linear with their corresponding mature RNA products; instead, the DNA sequences complementary to the mature RNA are interrupted by "non-coding" intervening sequences (introns; see 1, 2). Introns have been reported within rRNA, tRNA, and mRNA genes of viral, chloroplast, mitochondrial, and nuclear genomes (see 2 for review). The genes encoding the major late adenovirus mRNAs were the first of these discontinuous genes shown to function in vivo (3-5). In the late region, introns are transcribed as part of a large precursor RNA. During mRNA maturation, introns are excised and the RNA fragments containing the mRNA sequences are ligated together. In vertebrates, nearly every gene that codes for protein is organized and processed in a manner analogous to these adenovirus genes. However, many genes in non-vertebrate metazoans and lower eukaryotes do not contain introns (2). While introns are present in some mitochondrial, tRNA, and rRNA genes of lower eukaryotes, only one protein coding nuclear gene in a lower eukaryote is reported to be discontinuous (6, 7).

In this communication we report the first example of a Dictyostelium gene that codes for mRNA and contains intervening sequences. This specific

mRNA is found in a low abundance (~0.01% of total mRNA) in vegetative cells (8). Our results support the idea that discontinuous genes are a general phenomenon in eukaryotes.

MATERIALS AND METHODS

The origin and maintenance of the M4 plasmid and the Ax-3 Dictyostelium discoideum cells used in these studies have been described previously (8).

Poly(A)+ RNA was purified from whole cells, polysomes and nuclei as described (9). Purified DNA restriction fragments were 5'-end labeled using polynucleotide kinase and 3'-end labeled using the Klenow fragment of DNA polymerase I as described previously (8, 10).

Poly(A)+ RNA preparations were fractionated by electrophoresis on 1.5% agarose gels containing methylmercuric hydroxide (11), transferred to diazotized cellulose paper (12) and hybridized in formamide buffer (8).

In vitro labeled single stranded DNA fragments (8, 10) were hybridized with a sequence excess of poly(A)+ RNA in 0.7M NaCl, 7mM TES pH 7 and 1mm EDTA at 55° to 50 times the equivalent $Rot_{\frac{1}{2}}$. Hybridization reactions were diluted 1:20 in 0.25M NaCl, 0.03M Na Ac pH 4.7 and 2mM $ZnSO_4$. S_1 nuclease was added to 7 units/ml, the reactions were incubated at 40° for 45 min., and the samples were precipitated with ethanol. The resulting fragments were sized by electrophoresis next to a DNA sequencing ladder (11, 13).

RESULTS

Organization of the M4 Gene. We have previously described the identification and characterization of the Dictyostelium M4 sequences (8, 14). The restriction map of the M4 gene is shown in Figure 1 (see, 8, 14). The small (451bp) Hind III/Eco RI fragment contains a sequence repeated ~100 times in the genome, interspersed with single-copy DNA. 1-1.5% of vegetative polysomal poly(A)+ mRNA molecules contain sequences complementary to the M4 repeat. The repeat sequences represent ~10% of the mass of these mRNAs; the remaining 90% is transcribed from single-copy DNA. The large (1.1kb) Eco RI/Hind III fragment is entirely single-copy. Using exonuclease digestion, sandwich hybridization, S_1 protection, and DNA sequencing, we have shown that transcription initiates within the small Hind III/Eco RI fragment which contains the small repeat sequence and continues through the single-copy region of the large Eco RI/Hind III fragment (8, Kimmel and Firtel, manuscript in preparation).

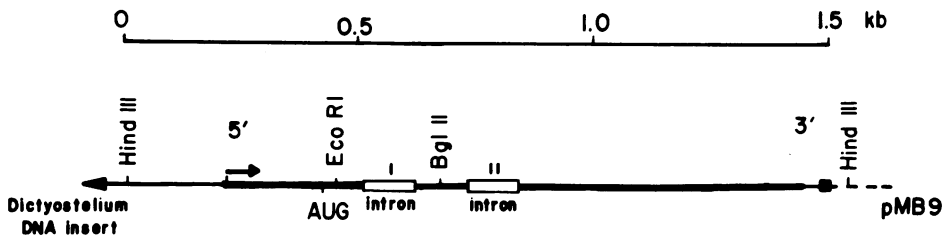

Figure 1. Organization of the M4 Gene. The small (0.45kb) Hind III/Eco RI restriction fragment contains sequences which are repeated ~100 times in the Dictyostelium genome (8, 14). The large (1.1kb) Eco RI/Hind III fragment is single-copy and contains two intervening sequences (introns). Transcription initiates within the small fragment and continues through the large fragment generating a nuclear precursor containing the intervening sequences. The translation initiation codon AUG is indicated.

The M4 mRNA is Transcribed as a Precursor. In Dictyostelium, nuclear poly(A)+ RNA is on the average ~20% larger than cytoplasmic poly(A)+ RNA (9, 15, 16). It has been suggested that this size difference reflects the processing of longer nuclear precursor RNAs into mature mRNAs (9, 15, 16). We wished to determine if the M4 mRNA were transcribed as a higher molecular weight precursor.

RNA was isolated from whole cells, polysomes, and nuclei and the poly(A)+ fractions purified. The different poly(A)+ RNA preparations were fractionated by electrophoresis on denaturing 1.5% agarose gels containing methylmercuric hydroxide (11). The RNA was transferred to diazotized paper (12) and hybridized with {^{32}P}-labeled M4 single-copy probe. RNA bands complementary to the M4 gene were identified by autoradiography (Figure 2).

As can be seen from hybridization to total cell or polysomal poly(A)+ RNA the cytoplasmic M4 mRNA is ~1.0 kb. (The size of the M4 mRNA was previously estimated to be 1.2 kb by electrophoresis using partially denaturing SDS-urea gels; 8.) Hybridization is also seen to contaminating 17s cellular (band 1) and 12s mitochondrial (band 2) rRNAs. It is clear that this is non-specific binding to rRNAs rather than hybridization to a precursor of the M4 mRNA. Hybridization to bands 1 and 2 is directly proportional to the observed amount of contaminating rRNA in the poly(A)+ RNA. These rRNA bands are prominent when stained with ethidium bromide and visualized with UV light. When poly(A)$^-$ RNA is hybridized with the M4 probe only the bands labeled 1 and 2

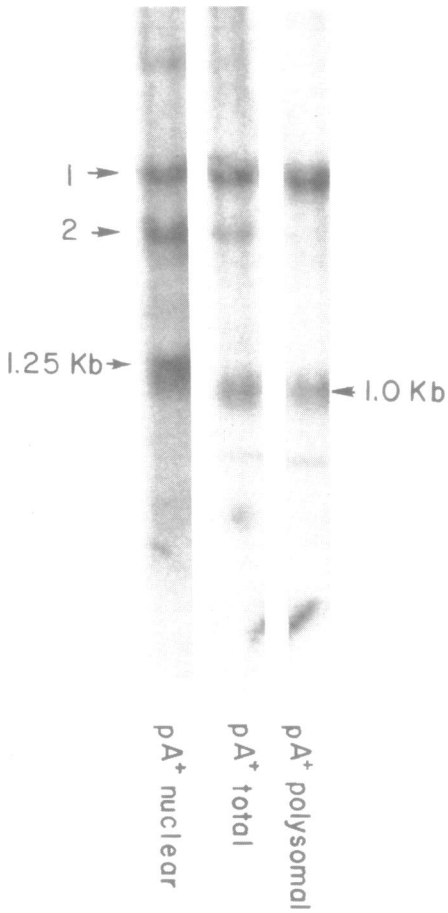

Figure 2. RNA Blot Hybridization of M4 Single-Copy Region to poly (A)+ RNA. Poly(A)+ RNA was purified from nuclei, whole cells, and polysomes (9). The RNA was electrophoresed in the presence of 10mM methylmercuric hydroxide (11) and transferred to activated cellulose paper (12). The large Eco RI/Hind III fragment (Fig. 1) was labeled in vitro by nick translation and hybridized to the RNA blots in formamide solution at 37°. The filters were washed and exposed to X-ray film (8). Molecular weights were determined by comparison with the mobilities of known Dictyostelium rRNAs and mRNAs electrophoresed in the same gel. Bands 1 and 2 are rRNAs which contaminate the poly(A)+ RNA preparations and stain prominently with ethidium bromide.

are detected. There is no hybridization to the mRNA band. In the polysomal preparation where there is no contaminating mitochondrial RNA there is also no hybridization to band 2 (see Fig. 2). When cytoplasmic poly(A)+ RNA is used, we observe M4 binding to bands 1 and 2 as well as to the 1.0 kb mRNA, but we do not detect hybridization to any bona fide M4 precursor. Finally, we have previously observed weak hybridization of the single copy M4 probe to rDNA bands using Southern hybridization (8). The rRNA contamination here represents ~20% of poly(A)+ RNA whereas the M4 mRNA sequence is only 0.01% (8). From the intensities of the hybridizations we conclude that the 1.0kb mRNA specific binding is ~1,000 time greater per RNA molecule as compared with the rRNA bands.

When the M4 single-copy probe is hybridized to nuclear poly(A)+ RNA, a 1.25kb nuclear specific RNA is observed. This suggests that the M4 mRNA is transcribed as a precursor at least 200-300 nucleotides larger than the mature mRNA (Figure 2). We have not been able to detect M4 sequences in nuclear poly(A)$^-$ RNA (data not shown). This is consistent with reports in adenovirus where poly(A) addition precedes processing of nuclear precursors (17).

Identification of Splice Sites. We were interested in identifying the regions of the nuclear RNA precursor which are removed during processing. From DNA sequencing we have identified an AUG codon 30bp upstream from the Eco RI site (see Figure 1). The 5'-end of the mRNA is 162bp 5' to this AUG, and there are no open translation reading frames within this RNA leader (Kimmel and Firtel, manuscript in preparation). The AUG is not followed closely by in-frame stop codons while both of the other translation reading frames contain stop codons (see Figure 5). We conclude that translation initiates at this AUG codon and continues into the large Eco RI Hind/III fragment. However, there is a 112 nucleotide region which is ~95% A + T rich located within the Eco RI/Bgl II fragment and is, therefore, 3' to the proposed translation initiation site. It is very unlikely that this A + T rich region codes for protein and suggests that it is contained within an intron.

To determine if the A + T rich sequence were contained within an intron, we performed a modification of the Berk and Sharp (18) S_1 nuclease mapping experiment. M4 DNA was labeled at the Bgl II site using {^{32}P}-γ-ATP and polynucleotide kinase and the labeled single-stranded Eco RI/Bgl II fragment was purified. This fragment was hybridized to mRNA and to hnRNA and the hybridized product digested with the single strand specific S_1 nuclease. The protocol for this experiment is outlined in Figure 3. If an intron existed, it would form a loop when hybridized to mRNA. This loop would be susceptible to S_1 cleavage yielding a labeled fragment shorter than the original. Similar experiments with hnRNA should yield a band which is not affected by S_1 digestion. The location of the proposed 3' intron junction can then be determined by sizing the S_1 fragments against a DNA sequencing ladder of the original Eco RI/Bgl II fragment. The results are seen in Figure 4a. Lane 4 is the uncleaved Eco RI/Bgl II fragment. Lanes 1 and 3 contain the fragment after hybridization with mRNA and cleavage with S_1. S_1 yields a series of bands which are smaller than the original fragment indicating the presence of an intron. Sollner-Webb and Reeder (19) have

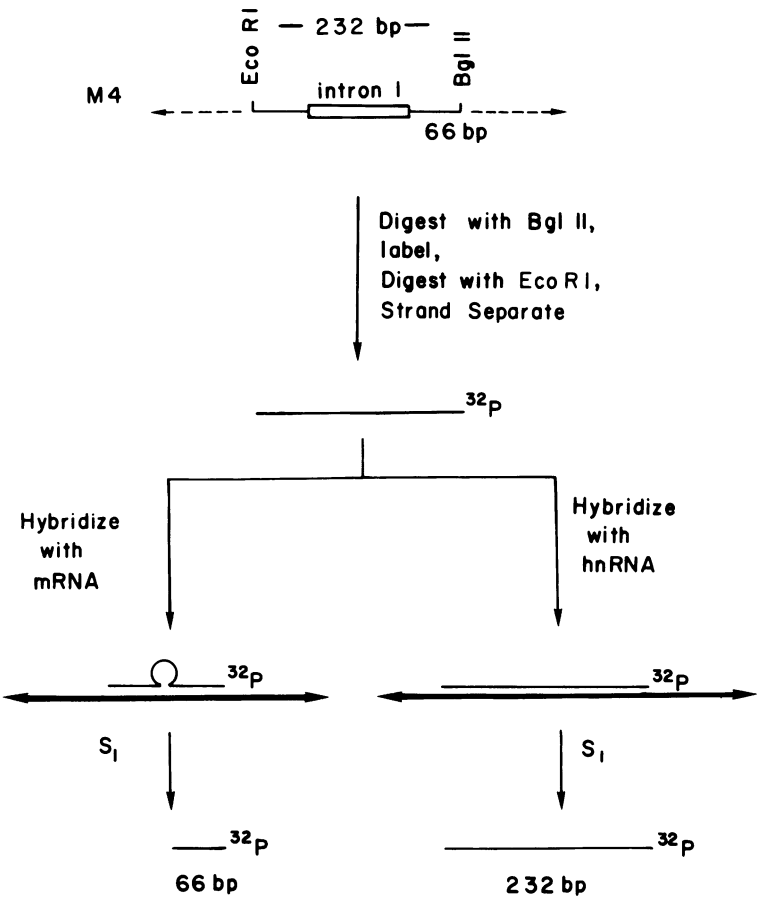

Figure 3. S_1 Mapping Procedure to Detect Intervening Sequences. The M4 plasmid was digested with the Bgl II restriction enzyme and the 5'-terminal phosphate removed with calf intestine alkaline phosphatase (10). The Bgl II site was labeled using ^{32}P-γ-ATP and the plasmid cleaved with Eco RI. The DNA was strand separated on 10% acrylamide gels and the small (262bp) labeled Eco RI/Bgl II fragment (Fig. 1) complementary to M4 RNA was purified. The labeled fragment was hybridized with cytoplasmic poly(A)+ RNA (mRNA) or nuclear poly(A)+ RNA (hnRNA) and the hybrids treated with S_1 nuclease. The mRNA does not contain introns and so S_1 would generate a fragment smaller than the original. The size of the S_1 fragment is equal to the distance from the Bgl II site to the 3'-junction of intron I. S_1 does not alter the size of the Eco RI/Bgl II fragment hybridized to hnRNA containing intron I.

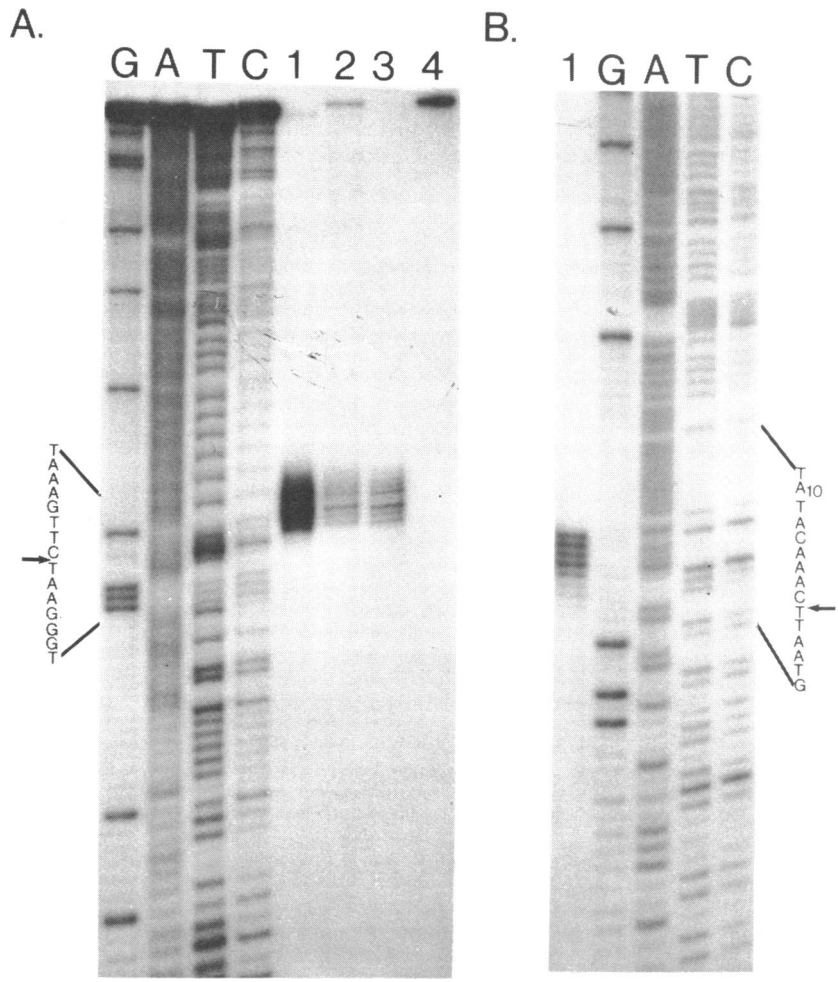

Figure 4. S_1 Mapping of Intron I.

A. 3'-splice junction. The procedure is outlined in Figure 3. Lanes G, A, T, C contain fragments generated by DNA sequencing procedure of Maxam and Gilbert (13). Lanes 1 and 3 contain the Eco RI/Bgl II fragment after hybridization with cytoplasmic poly(A)+ RNA and S_1 cleavage. Lane 4 contains uncleaved fragment. Lane 2 contains the fragment after hybridization with hnRNA and S_1 cleavage.

B. 5'-splice junction. The procedure is similar to Figure 3 except the Eco RI site was labeled at the 3'-end by the Klenow procedure (8, 11). Lanes G, A, T, C contain DNA sequencing fragments. Lane 1 contains the Eco RI/Bgl II fragment after hybridization to cytoplasmic poly(A)+ RNA and S_1 treatment.

The arrows indicate the splice junctions within the DNA sequence complementary to RNA. The sequences above the arrows are contained within the intron.

suggested a mobility correction when comparing heterogeneous S_1 generated patterns to Maxam and Gilbert (13) sequencing ladders. The DNA sequence position must be retarded 1½ nucleotides relative to the smallest S_1 generated fragment to adjust for the difference between the 3' termini generated by the S_1 nuclease or chemical sequencing procedure. We have, thus, identified the probable splice sequence at the 3'-end of the intron (see Figure 4 and Table 1). Analysis of the nucleotide sequence of the M4 mRNA with respect to open translational reading frames indicates that this must be the splice junction in that region (see Figure 5). Lane 2 shows a similar S_1 experiment using hnRNA. In this experiment approximately 20% of the labeled fragments are completely protected by nuclear RNA indicating that the intron is transcribed as part of a higher molecular weight precursor. To determine the 5'-splice junction, an analogous S_1 protection experiment was performed using the Eco RI restriction site labeled at the 3'-end using the Klenow fragment of DNA polymerase I (8, 10). The results are shown in Figure 4b and Table 1.

Additional DNA sequencing within the single-copy region of plasmid M4 shows another very A + T rich region suggesting a second intron (Figure 5). S_1 nuclease/DNA sequencing experiments have confirmed its presence (data not shown). The two M4 introns are located within the protein coding region and possess similar characteristics. Both are relatively small, ~100bp and ~95% A + T rich containing ~60% T residues. They are also flanked by sequences which are similar to other eukaryotic splice junctions (Table 1).

TABLE I

M4 RNA SPLICE JUNCTIONS

Breathnach et al.	↓GU ——— Intron ——— AG↓
M4 Intron I	U A A ↓G U U U G U - - - U U C A A G ↓A U U
M4 Intron II	A A A ↓G U A U G U - - - U A U C A G ↓G G A
Seif et al.	Pu ↓G U X X G U - - - PyPyX PyA G ↓
Lerner et al.	A G ↓G U A A G U - - - PyU X C A G ↓G

The 5'- and 3'-splice sequences of M4 introns I and II are presented. Consensus sequences of other eukaryotic splice sites derived by Breathnach et al. (28), Seif et al. (29), and Lerner et al. (30), are listed for comparison.

Py - pyrimidine Pu - purine

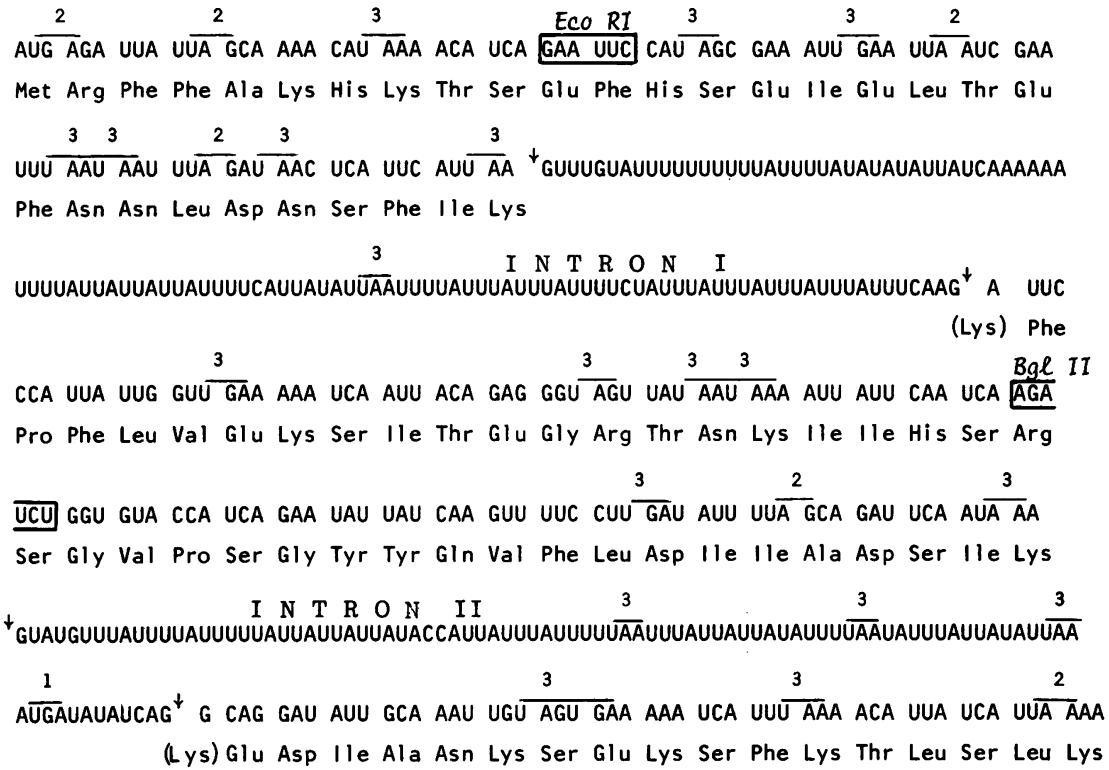

Figure 5. RNA Sequence in the Protein Coding Region Containing Introns I and II. M4 DNA (see Fig. 1) was sequenced in both directions from the Eco RI and Bgl II sites (14). The derived RNA sequence and corresponding amino acids are presented. The junction sites of introns I and II were determined by the S_1 mapping procedure. Translation stop codons are indicated by a bar above an RNA sequence with their respective reading frame, relative to the AUG initiating codon. The last amino acid shown is not the carboxy terminus. The translation termination codon is further downstream (see Fig. 1).

DISCUSSION

Intervening sequences are a common feature of many eukaryotic genes (1, 2). Except for the adenovirus polypeptide IX gene (20), all vertebrate genes studied to date which code for protein contain introns (see 1, 2). In general these genes are abundantly expressed with extreme tissue specificity (e.g., globin, ovalbumin, immunoglobin, vitellogenin) or are virally encoded (e.g., adenovirus, papovavirus). More recently, introns have been identified in genes which are likely to be expressed in many or all cell types at low to moderate levels (21-23). Furthermore, Maxwell, Maxwell and Hahn (submitted for publication) argue for the general occurence and transcription of introns in mouse genes coding for poly(A)+ RNA.

In contrast to vertebrate cells, many mRNA genes of the non-vertebrates examined thus far do not contain introns. Examples of non-spliced genes are the histone genes in sea urchin (see 2) and the (iso-)cytochrome c genes in yeast (24, 25). It is not known if the homologous genes in vertebrates contain introns. The only nuclear mRNA genes which are known to be discontinuous in non-vertebrate cells are the moderately transcribed actin genes in Drosophila (26) and yeast (6, 7) and the fibroin gene in silk moth (27).

Our results represent the first identification of intervening sequences in a gene from Dictyostelium. In Table 1, we see the 5'- and 3'-splice junction sequences of the M4 introns. Breathnach et al. have suggested that the 5'-GU and 3'-AG are conserved recognition sites for splicing (28). Seif et al. and Lerner et al. have identified several larger consensus sequences for splice points (29). The Dictyostelium splice sites agree well with these canonical sequences.

Other genes in Dictyostelium have also been examined for the presence of introns. There are 17 actin genes (10) some of which are known to be differentially regulated during development (McKeown and Firtel, unpublished observations). They do not appear to contain introns (10). The genes coding for the developmentally regulated carbohydrate binding protein, Discoidin I, have been isolated (31) and sequenced and also do not contain introns (Firtel, Poole and Rowekamp, unpublished observations). However, a duplicated gene family which is differentially expressed to a level of ~0.25-0.5% of total mRNA during early stages of the Dictyostelium developmental cycle does contain introns (Brandis and Firtel, unpublished observations). These introns are similar to those found in the M4 gene with regard to size, junction sequence, and A + T content. The high A + T content of introns in

Dictyostelium adds to the general observation in our laboratory that non-protein coding regions in the Dictyostelium genome are extremely A + T rich. These A + T rich regions also include the sequences 5' and 3' to the protein coding regions of the actin and Discoidin-I multigene families and the transcribed and non-transcribed spacers of the rDNA (10, 32; Firtel, Kimmel, McKeown, Brandis and Poole, unpublished observations).

A comparison of the transcription patterns of the actin and Discoidin-I genes with those of the spliced genes in Dictyostelium (see 8, 31, 33) yields no obvious relationship between the presence of introns and the expression of a particular gene. It has been argued biochemically and genetically that splicing is essential for the transport of mRNA into the cytoplasm for genes which contain introns (34, 35). The existence of spliced and non-spliced poly(A)+ mRNAs in Dictyostelium suggests that two transport mechanisms may exist to discriminate between these gene types.

Finally, our results raise some interesting questions. The single actin gene in yeast (6, 7) and two actin genes in Drosophila (26) are discontinuous. However, Dictyostelium actin genes are not spliced (10). This is clearly not because Dictyostelium is incapable of processing introns but may reflect the relatively rapid appearance, disappearance, or rearrangements of intervening sequences during the evolution of eukaryotes (1).

ACKNOWLEDGEMENTS

We are extremely grateful to W. Rowekamp and the Institut fur Zellforschung, Deutsches Krebsforschungs Zentrum, Heidelberg, for their hospitality during parts of this project. We also thank S. Poole, L. Goldin and M. Rosbash for their helpful suggestions on the RNA blot hybridizations, M. McKeown and J. Brandis for allowing us to cite their unpublished data, and C. Edwards, C. Lai and E. Rosen for providing useful discussions.

A. R. K. is the recipient of a Senior Fellowship for the American Cancer Society, California Division, and R. A. F. is the recipient of a Faculty Research Award from the American Cancer Society. This work was supported by grants from the National Institutes of Health and the National Science Foundation to R. A. F.

REFERENCES

1. Crick, F. (1979). Science 204, 264-271.
2. Dawid, I. B. and Wahli, W. (1979). Dev. Biol. 69, 305-328.
3. Chow, L. T., Gelinas, R. E., Broke, T. R., and Roberts, R. J. (1977), Cell 12, 1-8.

4. Berget, S. M., Moore, C. and Sharp, P. A. (1977). Proc. Natl. Acad. Sci. USA 74, 3171-3175.
5. Klessig, D. F. (1977). Cell 12, 9-21.
6. Gallwitz, D. and Sures, I (1980) Proc. Natl. Acad. Sci USA, 77, 2546-2550.
7. Ng, R. and Abelson, J. (1980). Proc. Natl. Acad. Sci. USA, 77, 3912-3916.
8. Kimmel, A. R. and Firtel, R. A. (1979). Cell 16, 787-796.
9. Firtel, R. A. and Lodish, H. F. (1973). J. Mol. Biol. 79, 295-314.
10. Firtel, R. A., Timm, R., Kimmel, A. R. and McKeown, M. (1979). Proc. Natl. Acad. Sci. USA 76, 6206-6210.
11. Bailey, J. M. and Davidson, N. (1976). Anal. Biochem. 70, 75-85.
12. Alwine, J. C. Kemp, D. J. and Stark, G. R. (1977). Proc. Natl. Acad. Sci. USA 74, 5350-5354.
13. Maxam, A. M. and Gilbert, W. (1977). Proc. Natl. Acad. Sci. USA 74, 560-564.
14. Kindle, K. H. and Firtel, R. A. (1979). Nucl. Acid Res. 6, 2403-2422.
15. Firtel, R. A., Kindle, K. and Huxley, M. P. (1976). Fed. Proc. 35, 13-22.
16. Lodish, H. F., Firtel, R. A. and Jacobson, A. (1973). Cold Spring Harbor Symp. Quant. Biol. 38, 899-914.
17. Nevin, J. R. and Darnell, J. E., Jr. (1978). Cell 15, 1477-1493.
18. Berk, A. J. and Sharp, P. A. (1978). Proc. Natl. Acad. Sci. 75, 1274-1278.
19. Sollner-Webb, B. and Reeder, R. H. (1979). Cell 18, 485-499.
20. Alestrom, P. Akusjarui, M., Perricaudet, M., Mathews, M. B., Klessig, D. F. and Pettersson, U. (1980). Cell 19, 671-681.
21. Wahl, G. M., Padgett, R. A. and Stark, G. R. (1979). J. Biol. Chem. 254, 8679-8689.
22. Schimke, R. T., Kaufman, R. J., Alt, F. W. and Kellems, R. F. (1978). Science 202, 1051-1055.
23. Nunberg, J. H., Kaufman, R. J., Chang, A. C. Y., Cohen, S. N. and Schimke, R. T. (1980). Cell 19, 355-364.
24. Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L. and Hall, B. D. (1979). Cell 16, 753-761.
25. Montgomery, D. C., Leung, D. W., Smith, B., Bhalit, P., Faye, G. and Hall, B. D. (1980). Proc. Natl. Acad. Sci. USA 77, 541-545.
26. Fyrberg, E. A., Kindle, K. L., Davidson, N. and Sodja, A. (1980). Cell 19, 365-378.
27. Tsuda, M., Oshima, Y., and Suzuki, Y. (1979). Proc. Nat. Acad. Sci. USA 76, 4872-4876.
28. Breathnach, R., Benoist, C., O'Hare, K., Gannon, G. and Chambon, P. (1978). Proc. Natl. Acad. Sci. USA 75, 4853-4857.
29. Seif, Il, Khoury, G. and Dhar, R. (1979). Nucl. Acids Res. 6, 3387-3398.
30. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. (1980). Nature 283, 220-224.
31. Rowekamp, W., Poole, S. P. and Firtel, R. A. (1980). Cell 20, 495-505.
32. Firtel, R. A., Cockburn A., Frankel, G., and Hershfield, V. (1976). J. Mol. Biol. 102, 831-852.
33. Rowekamp, W. and Firtel, R. A. (1980). Dev. Biol., in press.
34. Hamer, D. H. and Leder, P. (1979). Cell 18, 1299-1302.
35. Gruss, P., Lai, C-J., Ohan, R. and Khoury, G. (1979). Proc. Natl. Acad. Sci. USA 76, 4317-4321.