Volume 17 Number 23 1989 Nucleic Acids Research

Primary structure and regulation of vegetative specific genes of *Dictyostelium discoideum*

Charles K.Singleton*, Suzanne S.Manning and Ruey Ken

Department of Molecular Biology, Vanderbilt University, Box 1820, Station B, Nashville, TN 37235, USA

Received August 30, 1989; Revised and Accepted October 20, 1989 EMBL accession nos+

ABSTRACT

We have examined the expression and structure of several genes belonging to two classes of vegetative specific genes of the simple eukaryote, *Dictyostelium discoideum* . In amebae grown on bacteria, deactivation of all vegetative specific genes occurred at the onset of development and very little mRNA exists by 8 to 10 hours. In contrast, when cells were grown in axenic broth, the mRNA levels remained constant until a dramatic drop occurred around 10 to 12 hours. Thus, regulation of both classes of genes during the first several hours of development is dependent upon the prior growth conditions. Analysis of genomic clones has resulted in the identification of two V genes, V1 and V18, as ribosomal protein genes. Several other V genes were not found to be ribosomal protein genes, suggesting that in Dictyostelium non-ribosomal protein genes may be coordinately regulated with the ribosomal protein genes. Finally, using deletion analysis we show that the promoters of two of the V genes are composed of a constitutive positive element(s) located upstream of sequences involved in the regulated expression of these genes and within the first 545 upstream bp for V18 and 850 bp for V14. The regions involved in regulated expression were localized between −7 and −222 for V18 and −70 and −368 for V14. The sequences conferring protein synthesis sensitivity were shown to reside between −502 and −61 of the H4 promoter.

INTRODUCTION

Upon the initiation of the developmental program of *Dictyostelium discoideum* a series of morphogenetic and biochemical changes are set in motion which lead to the formation of at least two cell types from an initial pool of identical cells (1,2). The cell types become associated with one another in a defined manner resulting in a multicellular phase for this organism and culminating in the formation of a fruiting body. The morphogenetic and biochemical changes which occur are the observable manifestations of a complex program of differential gene expression (2,3). The details of the control of gene expression leading to these changes in biochemistry and morphology are being vigorously pursued from a number of different directions (reviewed in 3; and in the papers found in *Developmental Genetics* 9, no. 45). The results of these studies indicate that the control of gene expression during development of *Dictyostelium* is quite complex and multifaceted.

Because of the independence of the developmental program and the growth phase as well as for other obvious reasons, many genes whose products are required primarily during growth show reduced levels of expression during development. We have identified and characterized a number of genes which are vegetative specific; that is, their expression is deactivated upon the initiation of the developmental program (4−6). Our prior work characterized the deactivation of these vegetative specific genes and resulted in their division into two major classes with respect to the general mechanism of their regulation (4−6).

The deactivation at the onset of development for the majority of vegetative specific genes (V genes) is independent of ongoing protein synthesis. In addition, if protein synthesis is inhibited during growth, the expression of the V genes slowly declines with time. A simple model has been proposed to account for the regulated expression of these genes (6). We have postulated that a positive factor is present in growing cells, and this factor is responsible for allowing the expression of the V genes. Upon initiation of the developmental program, the factor is lost or inactivated resulting in the deactivation of the expression of these genes. At least 2 subclasses of V genes exist suggesting a different positive factor for each subclass. In contrast, the H genes constitute about 5% of vegetative specific genes and this class is regulated by a distinctly different mechanism. Inhibition of protein synthesis at the onset of development abolishes the deactivation of the H genes and in fact brings about a superinduction for most of the members of this class. The proper regulation of the H genes at all times is dependent upon continued protein synthesis. Our model for the regulation of this class of genes is based upon a balance of positive and negative factors, where the negative factor has a short half life and thus must be continuously made (6). Upon the onset of development, we postulate that the negative factor(s) increases in activity and override the ability of the positive factor(s) to bring about transcription. The positive factor(s) is still present, however, since induction can occur if protein synthesis is inhibited well after the normal deactivation event.

The work presented herein extends our previous investigations of the two classes of vegetative specific genes in several respects. Genomic clones of 2 V genes and one H gene are described and reveal in greater detail their structure and organization. Sequence analysis of the cDNA and genomic clones has identified two of the V genes as encoding ribosomal proteins. Finally, the promoter regions for the three genes have been characterized using deletion analysis. The results are discussed in relation to the models described above.

METHODS

D. discoideum strain KAx3 (7) was grown and cultured as described previously (4). Development was initiated by removing bacteria or nutrient broth, followed by plating the cells in starvation buffer as described (4). In some experiments, protein synthesis was inhibited by the use of cycloheximide at 400 μg/ml (5).

RNA Isolation And Analysis

Total cellular RNA from growing or developing cells was isolated by the quick-prep procedure (5). 1.5×10^7 cells were used per time point. RNA was quantitated by optical density measurements. Probe generation, northern blots, and hybridizations were performed as described previously (4).

Genomic Clones

Libraries of sucrose fractionated genomic DNA (8) digested with the appropriate restriction endonuclease were constructed by ligating fractions containing the genomic fragment corresponding to one of the cDNA clones into pGEM-1 (Promega). The appropriate size fraction and endonuclease to be used had been previously determined by Southern analysis (9). The resulting colonies were screened (10) using probes made by random primer labelling of the cDNA inserts (11). Restriction mapping was carried out on positive colonies using the partial digestion technique (12). The location of the transcribed region for each isolated genomic clone and the corresponding position of the cDNA insert were determined by digestion with the appropriate restriction endonucleases, fractionation by gel electrophoresis, and blotting to nylon. The filters were then hybridized using either a radioactive cDNA

insert or in vitro end labelled poly(A)$^+$ RNA (obtained with polynucleotide kinase and [γ ^{32}P]-ATP) from cells expressing the gene of interest. In several cases, comparison of the results of the RNA mapping with those of the cDNA mapping allowed for the determination of the orientation of the gene.

Appropriate smaller fragments containing the gene of interest and/or the 5′ upstream sequences were subcloned into pGEM-1. The smaller fragments or exonuclease III deletion products derived from them (13) were used for sequence analysis. Sequencing was carried out by employing the dideoxy procedure (14) on double stranded templates with Sequenase (USB Corp.). The manufacturer's suggested buffers were adjusted to account for the very A/T rich *Dictyostelium* genome. The seqences obtained have been entered into the EMBL/Genbank/DDBJ nucleotide sequence databases under the following accession numbers: X15382 (V18), X15383 (V14), and X15388 (H4).

Transcription start sites and exon/intron boundaries were mapped to the single nucleotide level by using either S1 nuclease (15) or RNase (16) protection assays. Singly end labelled restriction fragments spanning the suspected start site were used for the Sl protection assays. For RNase protection, radioactively labelled RNAs complementary to a particular mRNA were generated by using SP6 or T7 polymerase and the appropriate linearized subclone as template. The cRNAs were hybridized as described (15,16) to total RNA isolated from amebae expressing the gene under investigation or to tRNA as a negative control. After nuclease digestions and workup, the samples were electrophoresed under denaturing conditions along with sequencing reactions as size standards.

Promoter Constructs

The promoter constructs for V18 were made by digesting various exonuclease III deletion clones (deleted from the upstream direction) with HindIII (from the polylinker of the pGEM-1 parent vector) and BglII (located at +303 within the coding region of the V18 gene). The fragment in each case was isolated from a low melting point agarose gel and ligated to pPAV (17) which had been digested with HindIII and BamHI. This resulted in the insertion of the V18 promoter region of various lengths plus the first 303 bp of the transcribed region in front of the chloramphenicol acetyl transferase gene in the appropriate orientation for possible expression. The V14 constructs were generated in a similar manner by employing the BamHI site located at +425 within the coding region of the V14 gene. Finally, the HindIII-EcoRI fragment corresponding to −629 to −61 from the H4 gene was inserted in front of the CAT gene, generating the plasmid designated pH4b-2. An exonuclease III deletion derivative was made from this construct in which 128 bp from the 5′ most side were removed. All constructs were confirmed by restriction mapping and sequence analysis.

Transformations were carried out essentially as described (18). Equal amounts of a given promoter construct was cotransformed with the selectable marker-containing vector, Neo 1ΔX TBR (19). Transformed populations were grown in axenic media under G418 selection. Copy number was determined by isolating DNA from 10−20 ml of culture. Cells were collected and lysed as described above. Nuclei were pelleted and lysed in 50 mM Tris·HCl, pH 8, 10 mM EDTA, 4% sarcosyl by heating at 55°C for 15 minutes. The samples were digested with pancreatic ribonuclease followed by a digestion with proteinase K. A phenol extraction and two chloroform extractions were performed followed by an isopropanol precipitation. The resulting pellet was dissolved in water, and the DNA was quantitated by optical density measurements. Samples were digested with HindIII which recognizes a single site in the promoter construct vectors. This resulted in the release of

linear plasmids from the tandemly integrated copies. Southern blots of these samples were hybridized with a CAT specific probe. Copy number was determined by comparing band intensities to those of samples of added linearized plasmid corresponding to 1, 5, or 50 copies per genome equivalent. In all cases, aproximately 50% of all copies of integrated plasmid were the result of recombination events between the 2 cotransforming vectors (occurring homologously between the actin 15 terminator region contained in both plasmids).

Other Procedures

Restriction endonuclease digests were performed under the manufacturer's recommended conditions. The reaction conditions for utilizing other nucleic acid metabolizing enzymes were similarly those suggested by the supplier. Generation of radioactive cRNA transcripts using the pGEM-1 vector and either SP6 or T7 polymerase was accomplished by the procedures outlined in the technical bulletin supplied by Promega.

RESULTS

Primary Structure of the Genes and the Flanking Regions

Genomic clones for several of the V and H genes have been isolated. From Southern analysis, appropriately sized restriction fragments were identified which hybridized to a particular cDNA clone. Genomic DNA was digested with the appropriate restriction endonuclease, fractionated on a sucrose gradient, and the DNA in the fraction corresponding to the desired length was used to generate a plasmid based library. These were then screened using the cDNA inserts as probes. Maps of the genomic clones for V14, V18, and H4 as well as other pertinent information is shown in figure 1. Previous Southern analysis has shown that each of these three genes are present in a single copy in the *Dictyostelium* genome. All or part of the two V genes reside on *EcoR*I fragments: V14 on a 3.2 kb fragment and V18 on a 4.3 kb fragment. The H4 gene is located within a 6.4 kb *Hind*III fragment. By probing restriction digests of each clone with the cDNA inserts, the approximate location of each gene within the genomic fragments was determined. Sequence analysis of the cDNA inserts and the genomic clones established the directionality and confirmed the postion of each gene. The transcription start site and exon-intron junctions for each gene were mapped; several junctions were also confirmed by comparison of the cDNA and genomic sequences. There are at least three introns which have not been precisely localized and thus are not shown in the figure for H4. Also, since the 3′ end of the H4 cDNA obviously does not correspond to the poly(A) tail, the exact 3′ terminus of this gene is unknown. The nucleotide sequence of the transcribed regions as well as upstream sequences are given in figures 2−4.

Also in the figures, the predicted amino acid sequence for the coding region of each gene (or portion of the gene) is shown in addition to the nucleotide sequence. About 700 bp of coding sequence within the 3′ half of the H4 gene (see figure 1) was determined using the genomic and cDNA clones but is not shown. Data base searches revealed no significant similarities to existing sequenced proteins for V14 and H4. No obvious similarities with known proteins were found for two other V genes (V4, V12) and 3 additional H genes (H5,H6, H7) (CKS, SSM unpublished). In contrast, the V18 predicted amino acid sequence possessed 79.7% identity with the ribosomal protein L16 of yeast (39.5% with L5 of *E. coli*). Similarly, as detailed elsewhere (20), Vl also encodes a ribosomal protein corresponding to the L2 ribosomal protein of *E. coli*.

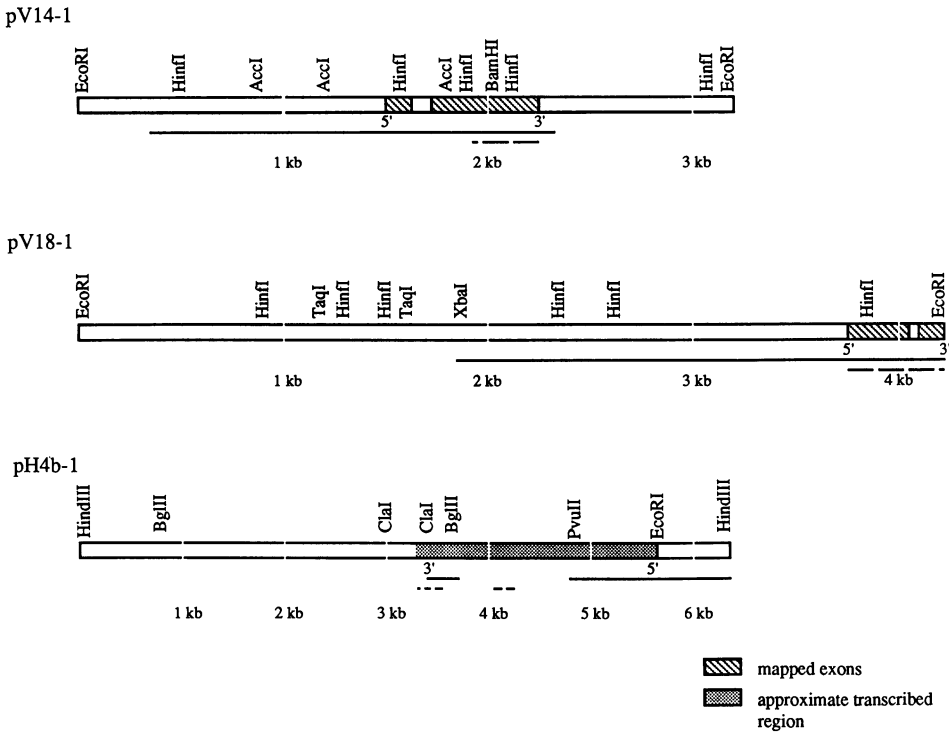

Figure 1. Genomic clones of V and H genes. The isolated genomic fragments containing the V14, V18, and H4 genes are shown schematically. The postions of the genes, both exons and introns, and the direction of each is illustrated. The solid underline represents the regions of the genomic clone that were sequenced. The dashed underline corresponds to the regions of the cDNA inserts that were sequenced. The exact location of the 3′ terminus of the H4 gene is not known as well as the exact position of at least three introns within the 3′ half of the gene.

Analysis of Promoter Regions

We have begun an analysis of the sequences required for the expression and developmental control of the V and H genes by examining a limited deletion series for several of the genes. Constructs were made in which various amounts of upstream sequences and a portion of the mRNA were inserted upstream of the chloramphenicol acetyl transferase (CAT) gene in the plamsid pPAV (17). The CAT gene in this plasmid lacks a promoter region and is flanked at its 3′ end by a functional *Dictyostelium* transcriptional terminator. Such constructs for V14, V18, and H4 were cotransformed into *Dictyostelium* amebae along with the selectable plasmid Neo 1ΔX TBR (19) which confers resistance to the antibiotic G-418. Populations of transformed cells for each construct were analysed to determine the average copy number of the inserted plasmid and for the expression of CAT specific mRNA under various conditions.

Figure 5 represents an autoradiogram illustrating the CAT specific mRNA levels found in several strains obtained by transforming with different V14 and V18 constructs. Table 1 summarizes the results obtained for the V18 and V14 constructs and indicates the region

-1148 TAGGTAAAAATAAAAAAAAAAAAAAAAAAA

-1120 AA AAAAA GTTATTATTTT GATTTAAAATTT ATTTT GATTTAAAATT TAAAAAAAAA TAAAAATAATCAGA

-1050 ATCTAAAAAAAAAAAAAATTAATTAATTAAA CAAAAAAAAAAAT AAAAGTAATCAAACCCAAAAAGTTAA

-980 TTTTCCATTAAAAAAAAAAAAAATTCGTGAAA CAATATCTAT AAAAAAAAAAATAC AAAAATTAGCGTTG

-910 GCCACAAAAAAA TTTTACATTGGGCTATGTAAACCTATGTATAAAA TTGGCAAAAT TACCC CTCTCGCAA

-840 CAA TTTTTT TTTTTTT GTGTTTT TATG CGT CTATCAA AA TTTT TTT CAATT TTTTT TGAATT TTTTTTTT

-770 TACT TCTAAGTTGA ACGAGTTT TTGGTATG TGTATACAC ATTATT TGTT TACGA TTT TTTTTT TTTTTTT

-700 TTTTT TTTTGTAAGGCAAAAATAAAATAAAA AAAAAA AAAATAAAATAAAAATAAAAAGATTTTAAAGAC

-630 AGAACACAAGTACACATATGATTATAAAGATAGCAAAAAAAAAAAAAAAATAATAATAATAAAAAGCAAT

-560 AAAATAAAAAAAAAAAAAAATAAAAAAAGAAAAAAAAAAAATAAAAAAATAATAATAATAATAAAAATA

-490 TGAT TTTTA AATGGTAGA AATAGTT TGGTATTTAGAAATTTACAATCTGTTGTAGAGAAATTTTAAAAATC

-420 AACTACAATT TTTAT TATTCAAGT TAAACCTT TATTTT TATTTGTT TAATTTATCC TTGAC TAGGATATA

-350 TTAA ACAGTAGACATGGAGAA TAGACTATGTGAAA ATACAGATAACAAATTTATT GGTTGCCAATTTTAA

-280 AAATTGATTAAAAATAAAA ATACATCACCATATAAAAATCTT TTGAAAAAAATACACAATAATAATAAAA

-210 AAAATAATACAACACAATAATAAATAAAATAATAATACAACA GGGCAAATAAAAAAA TAAAATAACAAA

-140 CCATCAGTCTTTTA TCATCACTTTTT TTTTTT CCGT AAAAA TTTTCGT TTTTTTT AGTTACGG TTTTTTT

-70 TTCGATT CCAAA TTTTTTTTTTTT AAAACACAAT TAAAAATT AGTCACTAACCTTT CA TTATCTCTATT

+1 ATTTAGACAATCCTTATAAAATCTAAAACAACAAAATGGTCTCTCTTAAATTACAAAAGAGATTAGC TGC
M V S L K L Q K R L A A

+71 CTCTATTCTCAAATGCGGCAAAGGTAGAGTTT GGATCGATCCAAATGAAATTGCTGATGTTGCCATGGCC
S I L K C G K G R V W I D P N E I A D V A M A

+141 AACTCAAGTAAGTATTATATTTTTTTAAAAAATT CAAATCAAT CAAAATCAAATAATTAGTAGTAAAAAA
N S R

+211 AACATAAAACTGATAAT GAAATTATAAATTGAAAATTTATAGGAGATAACGTCCGTAGACTCATCGCTAC
D N V R R L I A T

+281 CGGTTTCATCATGCGTAAACCAGTTGTTGTCCACTCTCGTTCACGTGCTAGAGAACACAACGCTGCCAAA
G F I M R K P V V V H S R S R A R E H N A A K

+351 CGTTTAGGTCGTCACAGAGGTGCTGGTAACAGATTAGGTACCCGTGAAGCTCGTCTCCCATCCAAGATTC
R L G R H R G A G N R L G T R E A R L P S K I L

+421 TTTGGATCCGTAGAATCAGAGTTTTAAGAAGACTCTTAAAGAAATACCGTGAAGCCAAGAAAATCGATAA
W I R R I R V L R R L L K K Y R E A K K I D K

+491 ACACTCTTACAGAGAGTTATACCTCAAAGCCAAAGGTAATGTCTTCAAGAACAAACGTACCTTAATTGAA
H S Y R E L Y L K A K G N V F K N K R T L I E

+561 TACATCGTCAAATCTAAGACTGAAAAATTAAGAGAAAAACTCATCCAAGATCAAGCTCAAGCCCGTAGAA
Y I V K S K T E K L R E K L I Q D Q A Q A R R N

+631 ACAAGAACAAATCAATCAAAGACAGACGTGCTGCTAAATCTGCCGCCAAAGAAATTGTTTCATCAAACTA
K N K S I K D R R A A K S A A K E I V S S N *

+701 AAA TAAATAATA CAAT CTTTTTTAA TATTTTATTTA GATTTAAA TTTATTT CCATTAAT TATATCTTATC

+771 ATTCTATTTATTCTTTTTTTTA

of each gene and upstream region contained in each construct. 222 base pairs of upstream sequence was sufficient to give expression for V18 as seen in fig. 5a for V18P2. Adding sequences out to −545 (V18P3) resulted in at least a 10 fold increase in CAT mRNA levels during growth (panel A). Surprisingly, V18P4, containing out to −740 bp, gave a much lower level of expression than V18P3, suggesting the presence or artefactual construction of a negative element within the region between −545 and −740. Although differing in the level of expression, both V18P3 and V18P2 demonstrated proper developmental regulation; i.e., deactivation when development was initiated (fig.5b). Thus, the sequences necessary for regulation lie within the first 222 bp upstream of the transcriptional start site. Except for the negative effect revealed by V18P4, a similar promoter arrangement was found for V14. Sequences between −859 and −368 were responsible for at least a 10 fold increase in V14-CAT mRNA levels when compared to a construct containing only the first 368 upstream base pairs (V14P4 verses V14P3). Again, both constructs resulted in the deactivation of the V14-CAT mRNA upon the onset of development with the loss of the CAT mRNA being faster than the endogenous V14 mRNA (fig. 5b). Deleting all sequences past −70 (V14P1) abolished expression and thus regulation as well. In each of the transformants analysed, the endogenous V18 or V14 mRNA levels were also determined as a control; both endogenous genes showed normal expression and regulation in all transformants.

A similar but more limited analysis was performed to characterize the H4 promoter. The region spanning from −629 to −61 for the H4 gene was inserted in front of the CAT gene and transformants containing this construct (pH4b-2) were obtained. RNA was isolated from transformants growing axenically before or 2 hours after the addition of cycloheximide. These samples were fractionated and probed for the CAT specific mRNA. As shown in figure 5c, this region was sufficient to confer cycloheximide inducibility to the CAT gene. Very little or no CAT mRNA was detected in cells which were not treated with cycloheximide. However the level of expression for the endogenous H4 gene is very low and sometimes hard to detect, thus making it difficult to discern if the upstream region contained in this promoter construct resulted in basal level expression. Deleting the region corresponding to −629 to −503 from pH4b-2 resulted in a construct from which the CAT gene was also inducible by cycloheximide addition (not shown), thus further delimiting the sequences necessary for conferring sensitivity of expression of H4 to the inhibition of protein synthesis. Results presented in figure 5c also illustrate the lack of CAT specific mRNA when no promoter is inserted into the pPAV vector, and it is transformed into amebae (negative control).

Effect of Growth Conditions on Developmental Regulation

As detailed above, two of the V genes, V1 and V18, have been identified as ribosomal protein genes. Our previous findings of a rapid and continuous loss of the corresponding mRNAs for these genes (4) and that the loss is due to a concomitant decrease in transcription

Figure 2. The sequence of the coding strand of the V14 gene and surrounding sequences. +1 corresponds to the mapped transcriptional start site. The predicted amino acid sequence of the single, large open reading frame is also presented. The underlined G at position 148 and 252 represent the first and last nucleotide, respectively, of the mapped intron. The AA sequence at position 767 is the poly(A) addition site. A TATA like sequence is found centered at −30. The other underlined sequences are described in the discussion section.

-1904 TCTAGAGTTTAAAT

-1890 TTCT CAGTT CCAACTTTTTTA GAATC TTCCTC ATTAAAATTTT AATATAT CAAAT AATTCATTCA CCTCAT

-1820 TATCCT CTTCCT CTGGCT CTTGA TTTTT ATTT AAATT CTCATT CTTT TTAT TTTTA CTTTTAT CTTCTTC

-1750 ATCTATATTTT CCTTAT CTTCTTCA GTTCGAT CGATT GGGT GTAATAACT CTAAACATTT CTAAAAAAAA

-1680 ATAAAAAT AAAAAAAAT TATAGTAAAT ATAAAACATTAG TATATTTAATT ATTATGATG ATAA CTATGAT

-1610 TATGATGAATAAT CAAGAT GATGATGATG ATGATGATGAT GATGATGAT GATGATGATGATG ATGATGAT

-1540 GATGATGAT GATGATGAT GATGATGATGAT GATCATGATGAT GATGATGATGAT GATGATGATC ATGATG

-1470 ATGATGATGAT GATGATGATCAT GATGATGATGAT GATGATGATGA TGATCAT GATGATGAT GATGATGA

-1400 TGATGAT GATTAAACAATAACTT CAAAGG TTCTAA CCCTTTTACAAACT GTTGGTG CTTCTCTTT GAATT

-1330 AACCAATAT GATTCATTTT GCTATTTAA TATTAAATAAT AATAATAATAATAAAAG ATAATAAAAT AATA

-1260 ATAATAA CTGTTAG TTATATTATATTA TATTATATTATA TTATATT AAATT TTTTAAAT AAAAAA AAAAA

-1190 TTTTTTAATAATA CCATCTG TAATAACT CCATTGAAAC CAATTCTTGATC TTCAAA TCCATC CATTTTTT

-1120 AATT ATAAATACAAA CTAAA GCTATTTA GTTT AATTAT CTTTTT ATTTATTG ATTTAATTTAA CTTTTAG

-1050 TGTATTGACAAAAAAAAAAAAAAAATAATTATAAAAGAAAAAAAAATAAAAAATT GAAAAAAAAAAAAAAA

-980 AAAAAAAAAATTAAAAAAAAAATTAAAAAAAAAAAATTA AAATAATTAAAAATATA CATAAATGGAATCG

-910 TTTACCAGGCT GTATTATAT CAAATTATAAAAT CAAAATAATAT ATTAAAATATCTAAAC AATAATTTCA

-840 AAAAAA CTATTTGAA GGGGG ATTT GATTTA TATAAAT CAATTAAA TTTTT TTTTTT TTTTTTTTTTTT

-770 TTTTTTTGTTTTTT CCAAAGGTGATAT CTTATTTGCAAAATAATA AATAAACCACAAA CCAAAAATGAAT

-700 AATTTCCTTT GCATTTAAATT AAAAAAAAAAAAAA GATTTTTCTT GAAAATAATT GTTTATAAAAAATAG

-630 TGC TTTTT TTTTT TTTTTT TTTT ATTT AT TTTT ATTTA TTA TTA TTTT TTTTT TTTTTC TTTATT CATT

-560 TATTTTT TTTTTTTA TTTTTTT AATAA TTCCCC TTTTT AA CTTTATTTT AATCT CAAAAA TTTGT CATCT

-490 AAAAAAAATAAATAAATAAATT AAAAAAAAAAAAAAATTTCTCGCCCTAATCCTTTTTC AAACAAAAAAAA

-420 AAAAAAAAAAA CGTTA GGCTTTAGG TTTTTCAA TTTTTTTTTTT CAATTTTTA AAAAAAAAA AATTTATT

-350 GGTGTTGAG AAAATCCTCA CAATTTTT CAATTTTTACG TTTTATCTTCCAT CCA CAAAAT GGTATGTATT

-280 ATGCGAT GTCTATATATAT CATTTTATAT GTATGAAAATATA GGTGGGT GTTTGG TTAAT GTTTGGACTA

-210 TAGGTTGAAAT TAAATAGTT GATGTGGATTAATTAAA GATTATTAAAGATTAA CAAGGATT GATAATGAC

-140 AGTATGATAACGATAATGGATATACAATAT GAAATAGATAGT CAAAAAGTTTAAAC ATTAAAAAAAACCT

-70 ATCACCCT CCTTATATTT CACAATTACACA GAAAGTAAAAAA GAAATTAACATTTTATTTAAT GACAGTC

+1 TGCCAAAGCTGCCACCAAAAACGCTACTAAGGTTGCCGTCAAAGCACCAGAAGCAACTACACCAGTTGAA

+71 ACCAAAAAATCAAAGAAAGACAATGTCATGAGAGGT CTCAGAATTGAAAAG TTAGTTCTCAACATTTGTG
M R G L R I E K L V L N I C V

+141 TTGGTGAATCTGGTGATAGATTA GTTCGTGCTGCTAAAGTACTTG AACAATTAACTGGTCAAA CCCCAGT
G E S G D R L V R A A K V L E Q L T G Q T P V

+211 CTACTCTAAAGGTAAATTATTTAAATAATAATAA CTAAAAATAATACCATAGACTAAC CATAATATTAAA
Y S K A

+281 AATTATAGCTCGTTACACTGTCA GATCTTTCAACATTCGTCGTAATGAA CAAATCGCTGCT CACGTCACT
R Y T V R S F N I R R N E Q I A A II V T

+351 GTCCGTGGTGAAAAAGCTGCTGAAATCTTAGAAATAGGTCTCAATGTTAGAA
V R G E K A A E I L E I G L N V R

rate of the genes (5) is in contrast to work by others (21) in which the levels for the mRNAs corresponding to 5 ribosomal protein genes did not decrease upon the onset of development but instead remained at the vegetative levels for up to 10 hours into development. Only then did a rapid and large decrease occur. These conflicting results were initially puzzling since it is assumed that all ribosomal protein genes are similarly regulated. One procedural difference between our analyses and those of Steel and Jacobson was the method of growing the cells. In all of our studies cells feeding on bacteria were used; in contrast, Steel and Jacobson used cells growing in axenic media. Thus, we examined the expression of V14, V18, and one of Steel and Jacobson's genes, rp1024, in developing cells grown under both conditions.

Cells were grown on plates in the presence of bacteria or grown axenically in shaking cultures. Both cultures were harvested, the nurients or bacteria were removed, and the amebae were plated for development. Morphologically, development proceded with the same timing for both samples: ripples formed by 7 hours and tipped aggregates by 12 hours. RNA was isolated from each population at various times before and after the initiation of development and subjected to northern analysis. Figure 6 shows the resulting autoradiogram after probing for V18 mRNA. Both patterns of expression originally reported by the two groups were found but with a dependence on the growth conditions. Similar results were obtained for V1 and rp1024, indicating that the manner in which ribosomal protein genes are regulated during the early stages of development is dependent on the method used to grow the cells prior to development. Surprisingly, this phenomenon is not limited to ribosomal protein genes. In fact, the same dual pattern of regulation depending upon the growth conditions was found for all of the V and H genes tested except V4.

DISCUSSION

We have examined the structure and expression of several genes which are expressed during growth and whose expression is deactivated upon the onset of the developmental program of *Dictyostelium discoideum* . We have isolated genomic clones corresponding to several of the genes in both classes. Mapping experiments and sequence analysis defined the organization and structure of these genes. By comparing the predicted amino acid sequence of each gene to that of known proteins, two of the V genes were identified as encoding ribosomal proteins. We have only determined the sequence for the amino terminal half of the predicted V18 protein. The amino acid sequence for this portion of the protein is 79.7% identical to the corresponding portion of the L16 ribosomal protein of *Saccharomyces ceriveciae* and 39.5% identical to the L5 ribosomal protein of *E. coli* . The first 210 amino acids for the predicted V1 protein are 28.1% identical to the L2 ribosomal protein of *E. coli*, while the carboxy terminal 27 amino acids are divergent (20). Although not identified by data base searches as being a ribosomal protein gene, the predicted V14 gene product is similar overall to ribosomal proteins by being small (186 residues) and basic (13% lys, 15% arg). The remaining V genes and the H genes which have been analysed show no

Figure 3. The sequence of the coding strand of the V18 gene and surrounding sequences. +1 corresponds to the mapped transcriptional start site. An alternate, mapped start site is found at −2. The predicted amino acid sequence of the single, large open reading frame is also presented. Those amino acids that are different from the yeast homolog of the encoded ribosomal gene are underlined; all other residues are identical between the genes of the two species. The underlined G at position 222 and 288 represent the first and last nucleotide, respectively, of the mapped intron. A TATA like sequence is found centered at −30. The other underlined sequences are described in the discussion section.

-629 TTAAT GTA TTTTTA TTTTT ATTA TTTTT TTTTTA TTTT TTTA TTTTT TTTGG TATTAC TTTATT GTGG TT

-560 TTCCCAATA TAATATATAGTAAATTT GGAAATT GAATAATTTTT AATACATTATC TGGAAA TAAAAAAAA

-490 ATAAAAAAAA AAACCTTT CAAAAAT GAAAAT GTTTTTTATTGGTT CATTTTTTGTAACAAAT CAAAAAAA

-420 AAAAAAAAAAAAGAAGTTTTTA CAAAAAAAGTTAA TTATTTAT AAAATTTTAT AAAATAAC TAAGGAAAA

-350 AAAAAAAAAAATTGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGATAAAAAAAGAAATAAAAATA

-280 AATTAAAAAAATGTT CTTT CGTTTT AACCACCAA TAGTTCA CAAA CATTC CGATTTTTTTTAT TTTTATT

-210 TTTTT CTC TAAA TTTT TTT TTTT TT TTTT CTC AATTT TTTTT T TTTT TTT TT TTT TTTTTA TTTTA TTTT

-140 TCATT TTTTT TTATT CTA TTTT ATA ATTTT AATTT CA TTTT TTT CAATA TTATCA TTATTA TT TTTA TT

- 70 TTAGAATTCAAAGAAAAAGAAGGGAAA TAAAAAAAA TAAAAAAAATA GAAATATAGAAAA TTAAATAAA

+1 AAAAAATAATAAAAATGTCAACAAAACCAACAA TTAATAAAGATG AATTAGTTACATTATTCTCACAAATT
M S T K P T I N K D E L V T L F S Q I

+71 GGTTTAGATTCATCAAAAGCTAAAGA GACAACCAATAAT GCAACTTT ATCAAGCAATTTA CAAGAGATCA
G L D S S K A K E T T N N A T L S S N L Q E I I

+141 TTAAAGAAGCA GGTGCT GAAAGTGGTTGT GAAAAATCTGTTGG TTTATTATTATACA CATTAGCAACTAA
K E A G A E S G C E K S V G L L L Y T L A T K

+211 ATATCCAGCCAATGCAATGAAACATCGTGCTACACTCGTT GACTATATT GCCAATAAGAAA TCAGTAAAC
Y P A N A M K H R A T L V D Y I A N K K S V N

+281 TCAATCAATTTGCAAGCATG TTTAGATTA CCTCCGTCGTACT GCCAATGAGGAATTAAA CGTAGCAGAAT
S I N L Q A C L D Y L R R T A N E E L N V A E F

+351 TTGAGCAAT CATGTGGTGTT GGTGTTGTAATTACAA GAGAACAAGTCGCACAA GCCGTATCAGATTATAT
E Q S C G V G V V I T R E Q V A Q A V S D Y I

+421 TAACAAAAATAAGAGTGACC TCTTGAGAA CGTTATCAATT CAATATTGGTGGTATTTAAT GGAGATTAAA
N K N K S D L L R T L S I Q Y W W Y L M E I K

+491 AATAGTCTCAAATGGGCAAACGCAAAAGATATCAAAGAA GAAGTCGAT GCTGCAATTC TTTCACTCTTAG
N S L K W A N A K D I K E E V D A A I L S L L G

+561 GTCCAAAAACCGATGCTGATAAAGCACCACCAGCAAAACCAGTTAAA CCAACCACTCCAACTGCTGTTGC
P K T D A D K A P P A K P V K P T T P T A V A

+631 CACCACCACTGCTGCCACCACCACCACTGGTGATTTCTCA CCAATTATTCCAGCTGAA TTAAAACCAGCC
T T T A A T T T T G D F S P I I P A E L K P A

+701 AAAGAAGAAATCAAATTCCCAGATCCAA GTGATAACATTCAAAACACA CCAAAACTCTTGGCTGATCATT
K E E I K F P D P S D N I Q N T P K L L A D H L

+771 TGAAAACCACAGGTGGTAAAATCGTTACTCGTTTCCCACCA GAACCAAATGGTTATCTCCA CA TTGGTCA
K T T G G K I V T R F P P E P N G Y L H I G H

+841 TGCTAAAG
A K

Figure 4. The sequence of the coding strand of the H4 gene and surrounding sequences. +1 corresponds to the mapped transcriptional start site. (This corresponds to the only protected band present; however a number of bands arising aparently from intrastrand base pairing within the probes used in the mapping experiments were also present in this sample as well as in the negative control). The predicted amino acid sequence of the single, large open reading frame is also presented. A TATA like sequence is found centered at −30.

extensive identity to known proteins. The other V genes show essentially identical expression behavior to that of the ribosomal protein encoding V genes as determined under a wide range of conditions and in various mutants, thus suggesting that in *Dictyostelium*

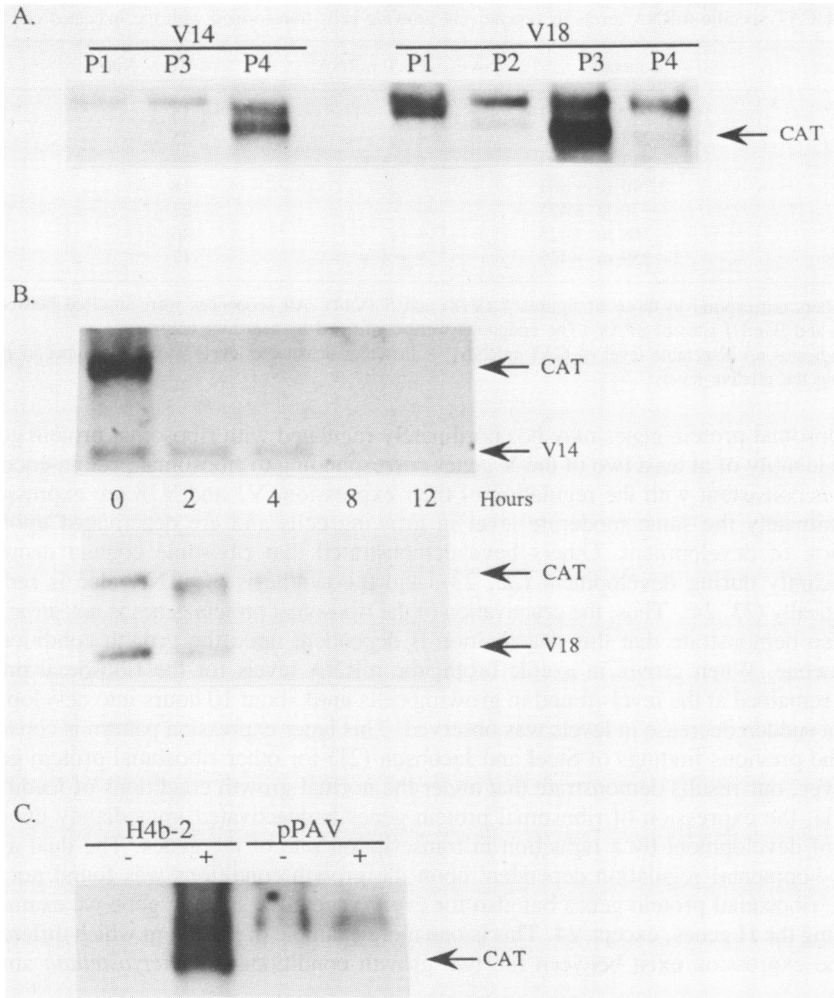

Figure 5. Expression of the CAT specific mRNA in strains transformed with promoter constructs. Panel A: Northern analysis was performed on total cellular RNA isolated from axenically growing cells of strains transformed with the promoter constructs V14P1, V14P3, V14P4, V18P1, V18P2, V18P3, and V18P4. After blotting, the filter was hybridized with a probe specific for CAT sequences. The resulting autoradiogram following exposure to the filter is shown. The arrow indicates the CAT specific chimeric mRNA species. The probe gives substantial hybridization to the 17S rRNA (band not marked) which can be reduced by more stringent wash conditions as seen in panel B. Panel B: Northern analysis was performed on total RNA isolated from strains transformed with either the V14P4 (upper) or V18P3 (lower) construct. The amebae were either growing in the presence of bacteria (0 hour time point) or had developed for the indicated times in hours. The resulting filters were hybridized with probes specific for the endogenous V14 or V18 mRNAs and CAT sequences. Panel C: Northern analysis was performed on total RNA isolated from strains transformed with either pH4b-2 or the parental plasmid for all of the promoter constructs, pPAV. Amebae were growing axenically before (−) and 2 hours after (+) the addition of cycloheximide at the time of cell lysis. The resulting filter was hybridized with the CAT specific probe. As in panel A, the probe hybridized to the 17S rRNA species.

Table 1 CAT specific mRNA levels in vegetatively growing cells transformed with the indicated construct.

Construct	Sequence[a]	CAT mRNA[b]	Copy Number
V18P1	−7 to +303	−	15
V18P2	−222 to +303	+	25
V18P3	−545 to +303	++++	25
V18P4	−740 to +303	−	18
V14P1	−70 to +425	−	21
V14P3	−368 to +425	+	46
V14P4	−859 to +425	++++	17

[a] Numbers correspond to those of figures 4 (V18) and 5 (V14). All sequences were inserted between the HindIII and BamHI sites of pPAV. The endpoints were confirmed by sequence analysis.
[b] − indicates no detectable level of CAT mRNA; + indicates detectable levels with the number of plusses reflecting the relative levels.

nonribosomal protein genes may be coordinately regulated with ribosomal protein genes.

The identity of at least two of the V genes corresponding to ribosomal protein-encoding genes is consistent with the regulation of their expression. V1 and V18 are expressed at approximately the same moderate level in growing cells and are deactivated upon the initiation of development. Others have demonstrated that ribosome content decreases significantly during development (22, 23), and the synthesis of rRNA also is reduced dramatically (23, 24). Thus, the deactivation of the ribosomal protein genes is not surprising. We also demonstrate that this deactivation is dependent upon the growth conditions of the amebae. When grown in axenic broth, the mRNA levels for the ribosomal protein genes remained at the levels found in growing cells until about 10 hours into development when a sudden decrease in levels was observed. This latter expression pattern is consistent with the previous findings of Steel and Jacobson (21) for other ribosomal protein genes. However, our results demonstrate that under the normal growth conditions of feeding on bacteria, the expression of ribosomal protein genes is deactivated immediately upon the onset of development by a reduction in transcription rate of the genes. The dual aspect of developmental regulation dependent upon the growth conditions was found not only for the ribosomal protein genes but also for every vegetative specific gene we examined, including the H genes, except V4. This is one more example of several in which differences in gene expression exist between the two growth conditions for *Dictyostelium* amebae

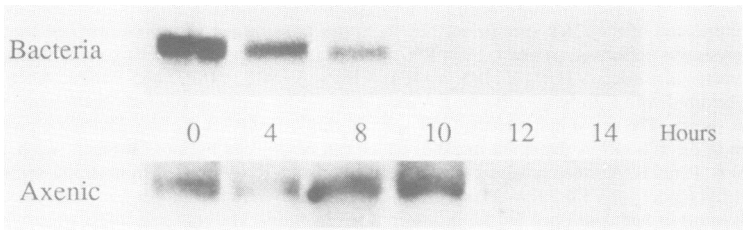

Figure 6. Dependence of V18 gene expression during development on the prior growth conditions of the amebae. Cells grown in the presence of bacteria or in axenic broth were plated for development after removal of nutrients. Total cellular RNA was isolated before nutrient removal (0 hr) and at various times after development was initiated. Equal amounts from each sample were separated eletrophoretically, transferred to nylon filters, and hybridized with a V18 specific radioactive probe. The resulting autoradiogram following exposure to the filter is shown.

(discussed in 25). The interesting aspect and significance of the dual regulation of the V and H genes, however, lies in the fact that these genes are expressed at identical levels in cells growing on bacteria or growing in axenic media. It is their regulation once development has been initiated and growth terminated that is dependent on prior growth conditions. Other genes show distinct differences in expression behavior under the different growth conditions themselves.

To begin to delineate the sequences necessary for the proper regulation and expression of the V and H genes, we have attached various upstream regions derived from several of these genes to a reporter gene (CAT gene). Our initial results indicate that proper developmental control of V18 requires at most the first 222 bp upstream of the start site of transcription while that for V14 occurs with only the first 366 bp. Deletion of all but the first 70 upstream nucleotides for V14 abolishes regulation and detectable expression. Further upstream of these regions for each gene, we have tentatively identified a positive element not involved in regulation but instead influencing the level of transcription for each gene by at least 10 fold. Although our studies to date are by no means extensive, these findings are similar to those others have found for several genes in *Dictyostelium* which show different patterns of regulation. A positive element affecting the levels of expression of the actin 6 gene has been demonstrated (26) and is located in a similar position to the V18 and V14 elements. The sequences responsibe for regulation and expression for several genes have been delineated to within the first 300 bp upstream of the transcriptional start site, and in each case, it has not been possible to disengage regulation and expression (19, 26–29). Our new findings give rise to a refinement of the model we have postulated for V gene expression. In addition to the postulated regulatable element, we find a constitutive positive element involved in increasing the levels of expresion and having no effect on regulation.

Although the promoter regions for V18 and V14 were found to have a similar design and previous work has shown these genes to be expressed in a similar manner under a variety of conditions, obvious sequence similarities in the upstream regions of the two genes were cryptic. The major complication in such a comparison is the very high A/T composition of these regions and other intergenic regions in the *Dictyostelium* genome. Determining which sequence similarities stand out or are significant is virtually impossible since many A/T rich sequences are repeated many times. Nonetheless, both genes possess a small inverted repeat around −300 and a large direct repeat around −170. In each case, the sequences involved are similar but different. Also, the sequence AAATTTATTGG is found centered at −300 in V14 and at −350 in V18. Both genes possess a TATA like sequence, TAAAAA, centered at −30. Finally, these genes as well as two other V genes (CKS, unpublished) possess blocks of C/A residues between the TATA box and −80, a feature not found within this region for H genes. The significance of these and other less salient similarities will require further analysis. One interesting feature of the V18 surrounding sequences which are apparently not involved in expression is the tandem repeat of the sequence GAT centered around −1500. This triplet is repeated 68 times with every tenth or so repeat being CAT. Also, flanking the ends of the repeat are imperfect repeats of this sequence or repeats of TAA and TTATA.

Finally, our analysis of the promoter structure and organization for H genes is less well developed. Sequences conferring sensitivity to protein synthesis inhibition for the H4 gene were found to be located between −502 and −61. The upstream sequences of H4 are superficially similar to those of the two V genes discussed above primarily due to the extreme A/T richness of these regions (89% for H4). As with the two V genes, a TATA box,

TAAAAA, is found centered at −30. Even though this region is missing in H4b-2, cycloheximide inducibility does occur. Due to the normally low level of expression for the H4 gene, it was not possible to determine if this promoter construct also resulted in basal level expression in growing cells. Other constucts are being made to address this and other issues concerning the regulation of this gene. We recently obtained genomic clones for the H6 gene and two of the three H5 genes. Comparative studies on these genes along with further analysis of the H4 promoter region should facilitate an understanding of how this class of genes is regulated during growth and development.

ACKNOWLEDGEMENTS

We thank Dr. Laura Steel for providing the rp1024 clone, and Russ Delude for work on mapping the transcription start site for the H4 gene. This work was supported by grant DCB-8506117 and DCB-8809026 from the National Science Foundation

*To whom correspondence should be addressed

+EMBL accession nos X15382, X15383, X15388

REFERENCES

1. Loomis, W.F. (1975) 'Dictyostelium discoideum a developmental system.' New York, Academic Press.
2. Loomis, W.F. (ed). (1982). 'The development of Dictyostelium discoideum.' Academic Press, New York.
3. Gerisch, G. (1987) *Ann. Rev. Biochem.* 56, 833−879.
4. Singleton, C.K.,. Delude, R. L, and McPherson, C. E. (1987) *Dev. Biol.* **119**, 433−441.
5. Singleton, C.K., Manning, S.S., and Feng, Y. (1988a) *Mol. Cell. Biol.* 8,10−16.
6. Singleton, C.K., McPherson, C.E., and Manning, S.S. (1988b) *Dev. Genetics* 9, 327−335.
7. Poole, S.J. and Firtel, R.A. (1984) *Mol. Cell. Biol.* 4, 671−680.
8. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Simn, D.K., and Efsratiadis, A. (1978) *Cell* 15, 687−698.
9. Southern, E. (1975) *J. Mol. Biol.* 98, 503−517.
10. Grunstein, M. and Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3961−3965.
11. Feinberg, A. and Vogelstein, B. (1984) *Anal. Biochem.* 137, 266−267.
12. Smith, H.O. and Bernsteil, M.L. (1976) *Nucleic Acids Res.* 3, 2387−2394.
13. Henikoff, S. (1984)*Gene* **28**, 351−359.
14. Sanger, F., Nicklen, S., and Coulson,A.R. (1977)*Proc. Natl. Acad. Sci. USA* 74, 5463−5467.
15. Berk, A. J. and Sharp, P.A. (1977) *Cell* **12**, 721−728.
16. Zinn, K., DiMaio, D., and Maniatis, T. (1983) *Cell* 34, 865−879.
17. Singleton, C.K. (1987) *Nucleic Acids Res.* 15, 9610.
18. De Lozanne, A. and Spudich, J.A. (1987)*Science* 236, 1086−1091.
19. Knecht, D.A., Cohen, S.M., Loomis, W.F., and Lodish, H.F. (1986) *Mol. Cell. Biol.* **6**, 3973−3983.
20. Singleton, C.K. (1989). Nucleic Acids Res., in press.
21. Steel, L.F. and Jacobson, A. (1987a) *Mol. Cell. Biol.* 7, 965−973.
22. Hames, B.D. and Ashworth, J.M. (1974) *Biochem. J.* 142, 301−315.
23. Cocucci, S. and Sussman M. (1970) *Cell Biol.* 45, 399−407.
24. Mangiarotti, G., Altruda, F. , and Lodish,H.F. (1981). *Mol. Cell. Biol.* 1, 35−42.
25. Oyama, M. and Blumberg, D.D. (1986) *Dev. Biol.* 117, 550−556.
26. Nellen, W., Silan, C., Saur, U., and Firtel, R.A. (1986) *EMBO J.* 5, 3367−3372.
27. Pavlovic, J., Haribabu, B., and Dottin, R.P. (1988)*Dev. Genetics* 9, 371−382.
28. Pears, C.J. and Williams, J.G. (1988). *Nucleic Acids Res.* 17, 8467−8486.
29. Datta, S. and Firtel, R.A. (1988) *Genes & Dev.* 2, 294−304.