Multiple copies of a G-rich element upstream of a cAMP-inducible *Dictyostelium* gene are necessary but not sufficient for efficient gene expression

Catherine J. Pears⁺ and Jeffrey G. Williams*

Imperial Cancer Research Fund, Clare Hall Laboratories, Blanche Lane, South Mimms, Herts. EN6 3LD, UK

Received April 29, 1988; Revised and Accepted August 2, 1988

ABSTRACT

The cysteine proteinase 1 (CP1) and cysteine proteinase 2 (CP2) genes of *Dictyostelium discoideum* encode co-ordinately expressed mRNA sequences which are inducible by extracellular cAMP. There are short, G-rich sequence elements upstream of both genes and we have previously shown that deletion of these elements from the CP2 gene abolishes cAMP-inducibility. We show here that the G-rich element from the CP1 gene is functionally homologous to that in the CP2 gene by reconstituting cAMP-inducibility in a deletion mutant of the CP2 gene using CP1-derived sequences. Both the CP1 and CP2 genes contain multiple G-rich elements. We show that efficient induction requires at least two copies of the CP1 element and that their relative orientation is unimportant. Two copies of an inverted relative orientation are, however, inactive when moved upstream of their normal position and are incapable of conferring cAMP-inducibility on a heterologous gene. These observations suggest that these sequences are either essential promoter elements, not themselves interacting with the inducer, or that their interaction with a separate class of control sequences is necessary for inducible expression.

INTRODUCTION

The cysteine proteinase 1 (CP1) and cysteine proteinase 2 (CP2) genes of *Dictyostelium discoideum* encode sulphhydryl proteinases with a possible role in the autodigestion which accompanies differentiation (1, 2, 3). The two mRNA sequences are perfectly co-regulated, both are first expressed late during cellular aggregation and accumulate prematurely in response to exogenous cAMP (1, 2, 3, 4, 5). This suggests that the rise in extracellular cAMP levels during aggregation may be responsible for their induction. The precise co-regulation of these genes implies the existence of common control elements activating their transcription.

The CP1 gene is situated 910 nucleotides away from a divergently transcribed gene of unknown function which is

co-ordinately activated with it and which is also cAMP-inducible (6). We show here that the CP2 gene also forms part of a divergently transcribed gene pair but that in this case the other gene is expressed during vegetative growth and development. The regions between the CP1 and CP2 gene pairs are extremely rich in A and T residues as is typical of Dictyostelium non-coding DNA (7). The DNA upstream of the CP1 gene contains several G-rich regions (6) and G-rich elements upstream of the CP2 gene are known to be necessary for efficient expression and cAMP-inducibility (4, 5). Deletions and point mutations in one member of the pair of closely spaced, homologous G-rich elements upstream of the CP2 gene drastically reduce its expression (4, 5). Here we show, by inserting synthetic oligonucleotides containing sequences from the CP1 gene into a deletion mutant of the CP2 gene, that the G-rich elements from the two genes are functionally homologous and that multiple copies are required for optimal expression. Having defined the minimal element which will reconstitute efficient expression, we have investigated its positional dependence and ability to activate a heterologous gene.

MATERIALS AND METHODS

Dictyostelium Culture and Development

Dictyostelium cells (strain Ax2 - provided by J. Ashworth) were grown at 22°C in shaking suspension in HL5 medium (8). The cells were developed in shaken suspension as described previously (5) and induced with cAMP at 1mM final concentration after 1.5 hours.

Transformation of Ax2 Cells and Analysis of Expression from Transformed Cells

Dictyostelium Ax2 cells were transformed by a calcium phosphate co-precipitation method (9,10) using modified conditions of selection with G418 (11). At least two independently-derived populations of transformed cells were analysed for each construct. Pooled populations of cells were selected by growth in shaken suspension in 10µg/ml of G418 and derived from an average of 20-30 original, independent clones. Total nucleic acid was isolated from transformed cells, the average copy number of each pooled population was determined by Southern transfer and the expression from the transformed

sequences determined by Northern transfer as described previously (5). Total cellular RNA isolated from cells transformed with "mini-gene" constructs of the CP2 gene was analysed using as probe a mixture of the 5' and 3' proximal KpnI fragments from the cysteine proteinase 2 cDNA clone pDd8 (5). Total cellular RNA isolated from cells transformed with constructs derived from Actin15-CP, a derivative of the actin 15 gene marked with an internal fragment of the CP1 gene, were probed with the CP1 cDNA clone pDd11.7.10 (12). Autoradiograms were analysed using an LKB ultrascan XL laser scanning densitometer.

Site-directed Mutagenesis

A BamHI site was created 190 nucleotides upstream of the initiation codon of the fusion gene present on Actin15-CP using the two primer method of site-directed mutagenesis (13). Transformed colonies of *E. coli* containing the mutated plasmid were identified by hybridization to the mutating oligonucleotide primer (14).

Insertion of Double-stranded Synthetic Oligonucleotides

Two complementary synthetic oligonucleotides were annealed by incubating at 60°C for one hour in 50mM Tris pH 7.5, 10mM MgCl₂, 10mM DTT and 0.5M ATP. The vector was cleaved with the relevant restriction enzyme and ligated to a 100-fold molar excess of double-stranded oligonucleotide, which was unphosphorylated to prevent multiple insertions. After transformation into *E.coli*, colonies containing plasmid DNA which had incorporated the insert were identified by screening with one of the two oligonucleotides. The orientation of the insert was determined by sequence analysis. The sequence of oligonucleotides inserted at -206 in the Δ30 mutant of the CP2 mini-gene was determined by the dideoxy chain termination method (15) using a synthetic oligonucleotide primer which was complementary to nucleotides -315 to -299 of the CP2 gene. The sequence of all other oligonucleotide insertions was determined by the chemical method (16) after cleavage and labelling at the restriction site regenerated at one end of the synthetic oligonucleotide insert.

RESULTS

1. A Comparison of the Sequence Organization Upstream of the CP1 and CP2 Genes

A schematic representation of the organization of the 5'

Figure 1A. Schematic representation of the organization of DNA found upstream of the CP1 and CP2 genes

The organization of the DNA upstream of the CP1 and CP2 genes is shown and the positions of the divergently transcribed genes found upstream of them are indicated. GC-rich elements found in the 5' flanking DNA are represented by arrows which indicate their relative orientation. The symbol → represents a G-rich element, and the symbol ← a C-rich element, relative to the sense strand of the relevant cysteine proteinase gene.

Figure 1B. The nucleotide sequence of the CP1 and CP2 G-rich elements

The nucleotide sequence of the G-rich elements found upstream of the CP1 and CP2 genes are shown and their orientation relative to the start site of transcription of the relevant gene is indicated by an arrow. Again an arrow pointing left to right indicates that the sequence is presented in the sense strand of the CP1 or CP2 gene.

proximal regions of the CP1 and 2 genes is shown in Figure 1A. The cap site of the CP1 gene lies 910 nucleotides away from the cap site of a divergently transcribed gene of unknown function

Figure 2. Analysis of the expression of the gene upstream of the CP2 gene

D. discoideum NC4 cells were allowed to develop on buffered agar and total RNA was extracted at intervals as indicated. A Northern transfer was probed with a nick-translated HindIII-TaqI fragment of 1.7Kb purified from pDdG4, the genomic clone containing the gene upstream of the cysteine proteinase 2 gene (2). Poly(A)+RNA, isolated from cells in the fourth hour of development, was also included (4+). This analysis reveals two transcripts, both of which hybridize to the probe at a high stringency of washing (0.5xSSC, 0.2%SDS, 65°C). The stages of development were 12hr tipped aggregates, 16hr mexican hats, 20hr midculminants and 24hr culminants. The sample labelled 4+ was isolated from cells shaken in suspension at 120rpm and then exposed to cAMP for a further 3hr.

(6). The two genes are co-regulated, first being expressed during late aggregation and they are inducible in response to extracellular cAMP. The intergenic DNA is extremely AT-rich apart from five GC-rich regions (Figure 1A and B) and is capable of directing correctly regulated, bidirectional transcription (6). The region upstream of the cysteine proteinase 2 gene was analysed to determine its organization. When Northern transfers were probed with a restriction fragment containing the DNA sequences from -921 to approximately -2600 upstream of the cysteine proteinase 2 gene, two transcripts were detected. One of these transcripts, of approximately 0.7kb in length, was expressed in vegetative growth and throughout early development but with a marked decrease in level after aggregation. The second transcript, of about 2kb, was also present throughout development and decreased in concentration later during

A CP2

Figure 3A. The nucleotide sequence the of 5' flanking region of the CP2 gene showing the endpoints of the deletion in the Δ30 mutant and showing the oligonucleotides used to investigate promoter activity

The sequences of the G-rich areas found upstream of the CP1 and CP2 genes are shown with the start sites of transcription designated as +1 and the centres of dyad symmetry indicated by arrows. Residues which are matched in the two halves of the dyad are starred. In the case of CP2 the extent of the internal deletion in the Δ30 construct is indicated below and the short direct G-rich repeats are underlined. For CP1 the short G-rich element is boxed and the extent of the 68 base (oligonucleotide L) and 28 base (oligonucleotide S) synthetic oligonucleotides used to reconstitute promoter activity in Δ30 are shown below. The sequence at the termini of the two inserts is indicated to show the base alterations necessary to generate BamHI compatible cohesive ends and to regenerate a BamHI site at only one end of the insert.

Figure 3B. Restriction map of the CP2 mini-gene construct $\Delta 30$

The CP2 gene, with 921bp of 5' flanking DNA, was inserted into the *Dictyostelium* transformation vector pB10 (9) and a mini-gene created by the deletion of an internal KpnI fragment from the coding region of the CP2 gene. The polyadenylation signal from CP1 was fused to the 3' end of the mini-gene and that from another gene, gene Y (11), onto the neomycin resistance gene (5). The deletion mutant $\Delta 30$ contains a 30bp deletion in the 5' flanking DNA of the mini-gene and there is a unique BamHI site at the position of the deletion (5).

aggregation (Figure 2). Southern transfer showed these DNA sequences to be present as a single copy in the genome (data not shown). S1 nuclease analysis showed a start site of transcription for one, or possibly both, of these RNA species about 1.3kb upstream of the cap site of the cysteine proteinase 2 gene and indicated the gene to be divergently transcribed with respect to CP2 (data not shown).

2. Insertion of 5' Flanking Sequences from CP1 into a Deleted CP2 Promoter and Analysis of Expression from the Fusion Promoter

There are two short G-rich elements present as a direct repeat 200 nucleotides upstream of the cap site of the CP2 gene (Figure 1A and 1B, Figure 3A; 21). In order to determine whether the G-rich elements found upstream of the two cysteine proteinase genes are functionally equivalent, sequences containing the G-rich region from the CP1 gene were inserted into a deletion mutant of the CP2 gene. The constructs used contained a "mini-gene", created by an in-frame fusion of the 5' and 3' regions of the CP2 gene, inserted into a *Dictyostelium* transformation vector (11). The mini-gene (Figure 3B) contains 3' non-coding sequences and the polyadenylation signal from the cysteine proteinase 1 gene and, after transformation into *Dictyostelium* Ax2 cells, the gene generates a discrete transcript which is induced to accumulate in response to extracellular cAMP (5). The mutant $\Delta 30$ is a derivative of this mini-gene, containing a 30bp deletion which removes both copies of the G-rich element at -202 and which displays an approximate 50-fold reduction in the level of regulated transcription relative to the wild-type mini-gene (5).

Two 68 residue complementary oligonucleotides, with BamHI cohesive ends, were synthesized. They contain almost an entire

Figure 4. Analysis of expression from $\Delta 30$ after insertion of CP1 upstream sequences

Transformed populations of *Dictyostelium* Ax2 cells, selected by growth in 10 μ g/ml of G418, were washed, resuspended at 10⁷ cells/ml in 17mM potassium phosphate buffer and shaken at 120rpm. cAMP was added to 1mM to half of the cells after 1.5 hours of development. Cells were harvested 3 hours later and total RNA extracted and analysed by Northern transfer, using the 5' and 3' proximal KpnI fragments from the CP2 cDNA clone, pDd8, as a probe (5). Lanes 1-3 contain RNA extracted from these cells either before development or after 4.5 hours in the presence or absence of cAMP. Total cellular RNA extracted from cells transformed with the wild-type mini-gene, which had been developed for 4.5 hours in the presence of cAMP added after 1.5 hours, was included on the Northern transfer as a control for the level of mini-gene expression (Δ Kpn) expected from an efficient promoter (Lane 4). The cells, transformed with the reconstituted promoter, have an average copy number of 30-40 per cell while the original mini-gene is present at 40-50 copies per cell as estimated by Southern transfer.

copy of the proximal direct repeat from the CP1 upstream region (6), extending from position -241 to -181 relative to the cap site of the CP1 gene (Figure 3A). The two oligonucleotides were annealed (to give "oligonucleotide L"), inserted into the BamHI site of the $\Delta 30$ construct at residue -206 which is the position normally occupied by the two G-rich elements in the wild type CP2 gene. A recombinant was chosen in which the orientation of the insert relative to the coding strand of the adjacent promoter was the same as is found in the CP1 gene. This construct was transformed into *Dictyostelium* Ax2 cells and two pooled populations of cells were analysed for expression of the fusion promoter in response to cAMP (Figure 4). Oligonucleotide L is

Figure 5. Analysis of the expression from $\Delta 30$ containing CP1 upstream sequences displaced from their normal position

The CP1 upstream sequences from -241 to -181 were inserted into the HindIII site at -891 of $\Delta 30$ (see Fig. 3B). The construct shown in diagram A contains the insert in the orientation, relative to the coding strand of the cysteine proteinase gene, found in the genome whereas that shown in diagram B contains the insert in the opposite orientation. The inserted oligonucleotide contains the same sequence as oligonucleotide L (Figure 3A) except that the termini were altered to generate a molecule with HindIII-compatible cohesive ends. Pooled populations of cells transformed with these two constructs were subjected to development and expression of the mini-gene was analysed as in the legend to Figure 4. Lanes 1-3 contain RNA extracted from cells transformed with DNA containing the CP1 sequences in the correct orientation, lanes 4-6 contain RNA extracted from cells transformed with DNA containing the insert in the opposite orientation and lane 7 contains RNA derived from cells transformed with the wild-type mini-gene developed with the addition of cAMP for 4.5 hours. The average copy number for all the pooled populations was 40-50 per cell.

capable of restoring efficient, correctly regulated, promoter activity to $\Delta 30$. The level of the mini-gene transcript is not significantly different from that seen from the original, wild-type mini-gene promoter, when adjusted for the average copy number found in the pooled population and normalised to the amount of transcript from the endogenous CP2 gene. The ability of sequences from the upstream region of the CP1 gene to functionally replace the G-rich elements from the CP2 gene indicates that the sequences from the two genes are functionally homologous.

3. The G-rich Elements are Inactive when Displaced Upstream of their Normal Position and are Unable to Confer cAMP-Inducibility on a Heterologous Gene

Eukaryotic enhancers have been shown to function in a distance and orientation-independent manner, whereas promoter elements are much more dependent upon position (reviewed in 17). In order to determine the positional dependence of the G-rich elements the CP1 sequence was inserted into the $\Delta 30$ mutant at a position far upstream of -206, the position where we have shown it to be effective in reconstituting expression. An oligonucleotide of 68 residues in length, equivalent to oligonucleotide L but with HindIII cohesive ends, was inserted into the HindIII site of $\Delta 30$ at a position 891 nucleotides upstream of the cap site (see Figure 3B). Recombinants were isolated which contained the insert in both orientations relative to the mini-gene coding sequences (Figure 5) and transformed into *Dictyostelium* Ax2 cells. The response to cAMP of two independently-derived pooled populations of G418-resistant cells

Figure 6. Schematic representation of Actin15-CP and analysis of expression from the actin 15 promoter containing CP1 upstream sequences

The Actin15-CP construct contains the actin 15 coding sequences, along with 270bp of 5' and 500bp of 3' flanking DNA, subcloned into the *Dictyostelium* transformation vector pB10TP1 11). The actin 15 gene was marked with a 340bp BglIII-HindIII fragment from the CP1 cDNA clone pDd11.7.10 (1) inserted at +20 of the actin coding sequences via a short multilinker (E. Henderson, personal communication). Oligonucleotide L, containing the CP1 upstream sequences, was inserted into a BamHI site, created by site-directed mutagenesis, 190bp upstream of the initiation codon. These constructs were transformed into *Dictyostelium* Ax2 cells and pooled populations of transformed cells, with average copy numbers of 40-70 per cell, were developed in shaking suspension in the presence and absence of 1mM cAMP, as in the legend to Figure 4. Total RNA was isolated and analysed by Northern transfer using as a probe the CP1 cDNA clone, pDd11.7.10. This hybridizes both to the endogenous CP1 transcript of 1200 nucleotides in length and to the fusion transcript from Actin15-CP of 1700 nucleotides in length. Lanes 1-3 contain RNA isolated from cells transformed with Actin15-CP, lanes 4-6 with Actin15-CP containing the three point mutations used to generate the BamHI site, and lanes 7-9 with the Actin15-CP containing oligonucleotide L.

were analysed for each of the two constructs (Figure 5). When normalised to the concentration of endogenous CP2 transcript and to the average copy number of the plasmid sequences per cell as determined by Southern transfer, neither construct showed any expression above the low, background level of $\Delta 30$.

Oligonucleotide L was also inserted upstream of a heterologous *Dictyostelium* promoter to see if it was able to confer cAMP-inducibility on a gene whose expression is not

usually modulated by a rise in extracellular cAMP levels. The gene chosen for this analysis was actin 15. A fragment containing 270bp of 5' flanking DNA from this gene is known to confer correct developmental regulation on a fusion gene when transformed into Dictyostelium cells (18). The construct Actin15-CP (Figure 6) contains the entire actin 15 coding region, along with 270bp of 5' and 500bp of 3' flanking DNA, subcloned into the Dictyostelium transformation vector pB10TP1 (11). This was marked, in order to distinguish its transcripts from those derived from the endogenous actin 15 gene, by the insertion of a 0.34kb BglIII-HindIII fragment from the CP1 cDNA clone pDd11.7.10 (1). This was inserted into the HindIII site, which lies 20 nucleotides downstream of the initiation codon of the actin 15 gene, via a short multilinker.

In order to insert the CP1 sequences into this promoter, a restriction site was created in the 5' flanking DNA of the actin 15 gene by site-directed mutagenesis. A BamHI site was created at -190 relative to the initiation codon of the actin 15 gene in a process which involved three single base alterations in one of the four GC-rich elements found upstream of this gene. Previous analysis has shown transcription from the actin 15 promoter to be unaffected by deletion of this particular element (18). The construct was then cleaved with BamHI and oligonucleotide L was inserted (Figure 6). The original Actin15-CP construct, the derivative containing the BamHI site and the derivative containing oligonucleotide L were transformed into Dictyostelium cells and the expression of the fusion gene analysed in the presence and absence of exogenous cAMP. Total RNA extracted from developing cells was analysed by Northern transfer using a CP1 cDNA fragment as a probe. The endogenous CP1 transcript was detectable after induction with cAMP, proving that the induction had been successful. The Actin15-CP fusion gene transcript of 1700 nucleotides in length was readily distinguishable from the CP1 transcript of 1200 nucleotides and showed the correct developmental regulation for an actin 15 promoter (Figure 6). When normalised relative to the average copy number per cell, the three point mutations introduced to create the BamHI site caused a very slight reduction in the level of fusion gene transcript (to about 80% of the level from Actin15-CP). The insertion of the

Figure 7. Analysis of expression from $\Delta 30$ containing two copies of the CP1 G-rich element

Dictyostelium cells transformed with $\Delta 30$ containing two copies of the CP1 G-rich element (oligonucleotide S), either as a direct ($\rightarrow\rightarrow$) or an inverted ($\rightarrow\leftarrow$) repeat and containing an average of 30-40 copies per cell, were developed and expression from the mini-gene was analysed by Northern transfer as in the legend to Figure 4. Lanes 1-3 contain RNA isolated from cells transformed with the direct repeat and lanes 4-6 contain RNA from cells transformed with the inverted repeat. Lanes 1 and 2 contain RNA isolated from vegetative cells. Lanes 2 and 5 contain RNA from cells induced with cAMP at 1.5hr and harvested at 4.5hr and Lanes 3 and 6 contain RNA from cells developed for 4.5hr in the absence of cAMP

CP1 upstream sequences caused a greater reduction in the level of transcription from this gene (down to 50% of that from Actin15-CP) but had no effect on the developmental regulation of the gene, with no increase in the concentration of transcript detectable after cAMP-induction. We have also analysed a construct in which the actin 15 upstream sequences between -190 and -135 relative to the initiation codon, were deleted and replaced with multiple copies of an oligonucleotide equivalent to oligonucleotide L (Driscoll, D.M., Duffy, K.T. and Pears, C.J, unpublished data). The deletion mutant was totally inactive and insertion of the oligonucleotides as single or multiple copies, did not re-activate expression prior to, or during, aggregation.

Table 1

Transforming plasmid	No. of populations analysed	Average expression of mini-gene as a % of that seen from the control per copy of the mini-gene	Expression
mini-gene	4	100	+++
Δ 30	3	15	-/+
single copy + orientation →	4	30	+
Single copy - orientation ←	3	15	-/+
Direct repeat (+) → →	3	130	++++
Inverted repeat → ←	4	170	++++

Expression from mini-gene promoters containing one or two copies of the CP1 G-rich element inserted into Δ 30

For each of the constructs whose expression was examined after transformation into Dictyostelium Ax2 cells, the number of separate pooled populations analysed is shown. In each case Northern transfers of cAMP-induction experiments were analysed by scanning densitometry and for each of these populations at least two Northern transfers were quantified. The expression from each mini-gene was normalised relative to the expression from the endogenous CP2 gene and to the average copy number per cell. The level of expression from the mutated and reconstituted promoters is shown as a percentage of that seen from the original wild-type minigene.

4. G-rich Elements Function Optimally when Multiple Copies are Present

The CP1 sequences, which were active in restoring cAMP-inducibility to Δ 30, form an inverted repeat with a G-rich element at the centre of each half of the dyad (Figure 3A). It seemed possible, therefore, that the synthetic oligonucleotide functions because it contains two copies of the G-rich sequence element in opposite orientations. In order to investigate this possibility shorter, complementary synthetic oligonucleotides were synthesized. "Oligonucleotide S" is a 28 residue, double-stranded oligonucleotide containing a single copy of the G element from within the proximal cysteine proteinase 1 direct repeat (see Figure 3A). The two complementary oligonucleotides

were designed such that, after insertion into a BamHI site, the restriction site is regenerated at one end of the insert. This allows the number and orientation of the elements present to be controlled. A single copy of oligonucleotide S was inserted into the BamHI site at -206 in $\Delta 30$ and recombinants were isolated that contained the insert in both orientations. A second copy was then inserted into the BamHI site regenerated in these recombinants, and constructs containing two copies of oligonucleotide S, either as a direct repeat or an inverted repeat, were identified by sequence analysis. The direct repeat contained the elements in the same orientation as found upstream of the CP1 gene, with the G residues in the sense strand of the mini-gene. The inverted repeat contained a proximal "C element" and a distal "G element."

The mini-gene constructs containing two copies of oligonucleotide S were transformed into Dictyostelium Ax2 cells and the ability of the promoter to increase transcription in response to extracellular cAMP during early development was determined (Figure 7). Two copies of the G-rich element from cysteine proteinase 1 were capable of restoring efficient cAMP-inducible expression in $\Delta 30$, whether present as a direct or an inverted repeat. However a single copy of the G-rich element in the correct orientation, that is with G residues in the coding strand, was capable of causing only a slight increase in promoter activity in response to extracellular cAMP, and one copy in the opposite orientation gave no stimulation (Table 1). As the increase in regulated transcription above the background level in $\Delta 30$ caused by one copy of the G element was small, the results were confirmed by repeating the analysis on several pooled populations of cells for each construct (Table 1). In each case the level of mini-gene expression was normalised to the average copy number per cell and to the concentration of endogenous cysteine proteinase 2 transcript, and the promoter activity expressed as a percentage of that seen from the original mini-gene under the same conditions. This analysis confirmed that two copies of the CP1 G-rich element were capable of restoring efficient regulated expression, whether present as a direct or an inverted repeat. One copy was capable of a slight (2-fold) increase in the level of transcription in response to cAMP when present in the correct orientation, but not when inverted.

The demonstration that a single copy of the G-rich element,

in the correct orientation, is able partially to confer cAMP-inducibility on a deletion mutant of the CP2 gene, coupled with previous evidence showing that mutations in just one of the two copies of the CP2 element greatly reduce gene expression, suggests that a single copy of the G-rich element acts as a discrete unit. While it is not yet possible to define a strict consensus (see Figure 1B) it seems reasonable to give the elements from both genes a common name, and we propose the term GSE (G-rich Sequences Element).

DISCUSSION

The control regions of genes in higher eukaryotes can now be understood as a set of short, discrete sequence elements necessary for efficient transcriptional initiation in the cell type in which the gene is expressed. If we are to understand gene regulation in Dictyostelium, it is necessary to identify both the basal promoter elements and the regulatory sequences necessary for inducibility at particular developmental stages and in the two cell types. The GSEs upstream of the CP2 gene are by far the best characterized Dictyostelium elements, as the effect of deletion and point mutations in the region have been determined in two separate studies (4, 5). The aim of this work was to further define the elements and to attempt to determine their role in gene induction.

As a first step we showed that the elements upstream of the perfectly co-regulated, but structurally highly diverged, CP1 and CP2 genes are functionally similar. This is by no means a completely predictable result because the sequences of the GSEs from the two genes are not highly conserved. The common, striking sequence feature is the presence of a very high proportion of G residues in the coding strand of the gene, but it is not possible to define a strict consensus sequence (Figure 1B). Our observation of replacement of function does not necessarily imply that a single protein, with a high degree of flexibility in its cognate recognition sequence, binds to both the CP1 and CP2 sequences. Indeed, as relatively minor changes in the CP2 sequence elements produce a major reduction in gene expression (5), a high degree of specificity in the interaction seems likely. It remains possible, therefore, that different proteins recognize the GSEs upstream of the two cysteine proteinase genes.

The major conclusion we can draw from this experiment is that the GSE from the CP1 gene is a discrete unit which is functional when moved out of its normal context.

A feature common to the CP1 and CP2 genes is the presence of multiple copies of the GSE. In the intergenic region upstream of the CP1 gene there are no fewer than five such elements in varying orientations relative to the direction of transcription. This may reflect the fact that the intergenic region mediates bidirectional transcription of two co-regulated genes. The presence of only two GSEs upstream of the CP2 gene is consistent with this interpretation since the adjoining, divergently transcribed, gene is not co-regulated with CP2. We have previously shown that deletion of the distal of the two GSEs upstream of the CP2 gene resulted in a major reduction in expression (5). This led us to suggest either that the proximal element is redundant or that a single copy of the element is insufficient for efficient expression. Our observation that a single GSE in one orientation gave very little stimulation of expression and in the opposite orientation gave no stimulation, suggests the latter explanation to be correct. Two G-rich elements were able completely to restore efficient transcription. This restoration of function occurred irrespective of the relative orientation of the two GSEs. This flexibility of relative orientation appears to be reflected in the intact genes, since the two pairs of GSEs upstream of the CP1 gene are present in an inverted orientation while the CP2 elements are present as direct repeats.

In order to determine whether the presence of GSEs is the only requirement for efficient inducible expression we attempted to activate a heterologous gene by inserting GSEs into its upstream sequence. We chose the actin 15 gene because the upstream region has been subjected to deletion analysis and an essential sequence element has been approximately defined (18). This element is itself a G-rich sequence. We were unable to confer cAMP-inducibility on a minimally disrupted, transcriptionally active actin 15 gene or on one which had had its control elements deleted and replaced by CP1 upstream sequences. We are to some extent constrained in interpreting these results by the lack of information concerning the precise sequence boundaries of the actin 15 elements, their role in

transcription and relationship with other sequence elements in this gene. However, the inability of the GSEs to stimulate transcription either in the manner of an actin 15 gene or in response to extracellular cAMP, suggest that the GSEs must act in concert with other sequences upstream of the cysteine proteinase genes.

By analogy with other eukaryotic transcriptional control sequences (reviewed in 17) there appear to be two obvious possibilities as to the function of the GSEs. The GSE could be a basal promoter element, required for efficient expression but not directly interacting with the regulatory protein, or proteins, we presume to be responsible for gene induction. Alternatively, the GSE could be a regulatory element requiring interaction with a specific promoter element not present upstream of the actin 15 gene. One characteristic feature of regulatory elements is that they tend to behave like enhancer elements, being relatively flexible in their position relative to the cap site, while promoter elements, are much more position-dependent. When the duplicate GSE element, which was able to restore complete cAMP-inducibility when positioned at nucleotide -206 relative to the cap site of the CP2 mini-gene, was inserted at -891 it was totally inactive. Again, a negative result must be interpreted with caution but two further pieces of inferential evidence are consistent with the GSE being an essential promoter element rather than a regulatory, enhancer-like sequence.

When both GSEs upstream of the CP2 gene are deleted there is a low level of residual expression with transcription being correctly regulated and cAMP-inducible (4, 5). There are other possible interpretations but the most straightforward is that the regulatory sequence element lies elsewhere and that the GSE is necessary for optimal expression. In this respect the GSE would play a similar role to the PUE (positive upstream element) shown to reside upstream of the actin 6 gene of Dictyostelium (19). Another circumstantial piece of evidence for such a role comes from the ubiquity of GSE-like sequences. They are found upstream of vegetatively transcribed actin genes (20), the cAMP-repressible discoidin I genes (21, 22) and the DIF-inducible pDd56 gene (23, 24). It will obviously be necessary to prove the functional identity of these sequences but, assuming this to be the case, then these elements would appear to be playing a role

in Dictyostelium gene transcription analagous to that played by promoter-elements of higher eukaryotes such as "CCAAT" box or the SP1 binding site. The GSE shows some similarity to the SP1 binding site in being very G-rich in functioning bi-directionally and in its dependence upon the presence of multiple copies for optimal activity (25, 26, 27, 28, 29). There is now one striking example of conservation of both sequence, and apparent function, between a yeast and a vertebrate regulatory protein (30). It will be of interest, therefore, to compare the SP1 binding protein and the GSE-binding protein once it has been purified.

*To whom correspondence should be addressed

+Present address: Ludwig Institute, 91 Riding House Street, London, W1P 8T, UK

REFERENCES

1. Williams, J.G., North, M.J. and Mahbubani, H. (1985) *EMBO J.* 4, 999-1006.
2. Pears, C.J., Mahbubani, H. and Williams, J.G. (1985) *Nucl. Acids. Res.* 13, 8853-8866.
3. Datta, S., Gomer, R.H. and Firtel, R.A. (1986) *Mol. Cell Biol.* 6, 811-820.
4. Datta, S. and Firtel, R.A. (1987) *Mol. Cell. Biol.* 7, 149-159.
5. Pears, C.J. and Williams, J.G. (1987) *EMBO J.* 6, 195-200.
6. Driscoll, D.M. and Williams, J.G. (1977) *Mol. Cell Biol.* 7, 4482-4489.
7. Kimmel, A.R. and Firtel, R.A. (1982) in the Development of Dictyostelium discoideum, Loomis, W.F. ed. pp234-316, Academic Press, New York.
8. Watts, D.J. and Ashworth, J.M. (1970) *Biochem. J.* 119, 171-174.
9. Nellen, W., Silan, C. and Firtel, R.A. (1984) *Mol. Cell. Biol.* 4, 2890-2898.
10. Nellen, W. and Firtel, R.A. (1985) *Gene* 39, 155-163.
11. Early, A. and Williams, J.G. (1987) *Gene* 59, 99-105.
12. Williams, J.G. and Lloyd, M.M. (1979) *J. Mol. Biol.* 129, 19-35.
13. Norris, K., Norris, F., Christianson, L and Fiil, N. (1983) *Nucl. Acids Res.* 11, 5103-5112.
14. Zoller, M.J. and Smith, M. (1982) *Nucl. Acids. Res.* 10, 6487-6500.
15. Sanger, F., Nicklen, S. and Coulson, A.R (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463-5467.
16. Maxam, A. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
17. Maniatis, T., Goodburn, S. and Fischer, J.A. (1987) *Science* 236, 1237-1245.
18. Cohen, S.M., Knecht, D., Lodish, H.F. and Loomis, W.F. (1986) *EMBO J.* 5, 3361-3366.
19. Nellen, W., Silan, C., Saur, V and Firtel, R.A. (1986) *EMBO J.* 5, 3367-3372.

20. Romans, P. and Firtel, R.A. (1985) *J. Mol. Biol.* **183**, 311-326.
21. Williams, J.G., Tsang, A.S. and Mahbubani, H. (1980) *Proc. Natl. Acad. Sci. U.S.A* **77**, 7171-71715.
22. Poole, S.J. and Firtel, R.A. (1984) *J. Mol. Biol.* **172**, 203-220.
23. Jermyn, K.A., Berks, M., Kay, R.R. and Williams, J.G. (1987) *Development* **100**, 745-756.
24. Ayres, K., Newman, W., Rowekamp, W.G. and Chung, S. (1987) *Mol. Cell. Biol.* **7**, 1823-1829.
25. Dynan, W.S. and Tjian, R. (1983) *Cell* **32**, 669-680.
26. Everett, R.D., Baty, D. and Chambon, P. (1983) *Nucl. Acids Res.* **11**, 2447-2464.
27. Gidoni, D., Kadonaga, J.T., Barrera-Saldna, H., Takahashi, K., Chambon, P. and Tjian, R. (1985) *Science* **230**, 511-517.
28. Jones, K.A., Yamamoto, K.R. and Tjian, R. (1985) *Cell*, **42**, 559-572.
29. Melton, D.W., McEwan, C., McKie, A.B. and Reid, A.M. (1986) *Cell* **44**, 319-328.
30. Struhl, K. (1987) *Cell* **50**, 841-846.