Characterisation of sequences required for RNA initiation from the *PGK* promoter of *Saccharomyces cerevisiae*

J.Rathjen and J.Mellor*
Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received March 9, 1990; Revised and Accepted May 11, 1990 EMBL accession nos K00553, X05969

ABSTRACT

In the phosphoglycerate kinase (*PGK*) gene of yeast, as in other highly expressed yeast genes, the sequences surrounding the site of RNA initiation have a loosely conserved structure of a CT rich stretch followed by the tetranucleotide CAAG. Using internal deletions and insertions we have identified the elements in the *PGK* promoter which are required for correct RNA initiation at the CAAG sequence at −39. The results indicate that two different components of the *PGK* promoter contribute to correct RNA initiation, the TATA homologies, located at −152 and −113, and the sequences at the site of initiation. Both TATA elements can function in RNA initiation. Deletion of the upstream TATA element, TATAI, results in slightly heterogeneous RNA initiation, but the majority of the RNA initiates correctly. Deletion of both the *PGK* TATA elements results in the majority of the RNA initiating at sites downstream from the wild-type I site, within the structural gene between +40 to +80. The CT rich box is not essential for correct mRNA initiation as shown by deletion analysis. The site of RNA initiation in the *PGK* promoter appears to be determined by sequences located immediately 5′ of the CAAG sequence motif. This short sequence, ACAGATC, when located the correct distance from the TATA elements may be sufficient to determine a discrete initiation site.

INTRODUCTION

In many of the yeast promoters analysed to date three elements have been identified as necessary for the accuracy, amount and regulation of mRNA transcription (reviewed in 1). Upstream elements function over large distances to stimulate or depress transcription from a promoter. These sequences are characteristic of a gene or co-regulated family of genes and usually determine the regulatory properties of a promoter. The second element which is considered a general promoter element is the TATA sequence, consensus TATAAA. This element is found in both prokaryotic and eukaryotic genes. In higher eukaryotic promoters the TATA element functions to set the RNA initiation site a fixed distance, usually 25 to 30bps, downstream of itself (2). The site of RNA initiation is determined by the TATA element and is unaffected by alterations of the sequences at the site. In yeast promoters, however, RNA initiation occurs within a window, the initiation window, 40−120bp 3′ of the TATA element (3). Sequences at the site of RNA initiation, I site sequences, appear to be important in the process of start site selection within the initiation window (4,5). I site sequences can be moved within the initiation window and still determine the site of RNA initiation (4,5,6). In addition, RNA initiation sequences can be inserted into the RNA initiation window of a heterologous promoter and function to determine an RNA initiation site (4). A survey of yeast RNA initiation sites has suggested a number of sequences which are found preferentially at the I site, particularly when the distance between the TATA box and I site is greater than 50bps (5,6). Two common motifs are RRYRR and YAAR (4,5; R= purine, Y= pyrimidine).

The *PGK* gene in yeast encodes the glycolytic enzyme, phosphoglycerate kinase. Like many other highly expressed genes, sequences around the I site of the *PGK* promoter conform to a general consensus of a CT rich region, located in *PGK* between −70 and −50 followed a short distance downstream by a tetranucleotide CAAG (7). mRNA initiation from the *PGK* promoter occurs at −39, within the CAAG sequence (7). The upstream activator sequence (UAS_{PGK}) has been localised to a region between −402 and −473 of the *PGK* promoter (8). This region functions bidirectionally to activate transcription from both the *PGK* promoter and heterologous promoters (8). Deletion of the UAS results in a significant drop in mRNA levels (8). Potential TATA elements have been identified within the promoter at −152 and −113 (7,8).

Despite the frequency with which the CT box..CAAG motif appears in yeast promoters, no function for this sequence configuration has been described. Using a series of small internal deletions and insertions we have investigated the role of sequences located downstream of −217, including the TATA homologies and the CT...CAAG, in determining correct mRNA initiation in the *PGK* promoter. In this paper we show that TATA elements are required for correct start site usage, that the CT rich region has no apparent function and that a short sequence ACAGATC between the CT box and the CAAG may be sufficient to determine the RNA initiation site.

MATERIALS AND METHODS

Strains, media and growth conditions

Bacterial transformation and large scale propagation of plasmid DNA were performed in *Escherichia coli* MC1061 (9). *E. coli*

* To whom correspondence should be addressed

cultures were grown at 37°C in Luria broth with ampicillin (Sigma) at 25μg/ml when necessary.

Saccharomyces cerevisiae MD40/4c (α, *ura2 trp1 leu2.3 leu2.112 his3.11 his3.15*) was used. Yeast cultures were grown at 30°C in YEPD (1% yeast extract, 2% peptone and 2% glucose) or in a defined minimal media (6.7% yeast nitrogen base w/o amino acids, 2% glucose, amino acids as required). Cultures were harvested at 6 to 8×10^6 cells/ml.

DNA manipulations and plasmid constructions

Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase and Klenow fragment were obtained from BRL, Inc.. Radioactively labelled isotopes, [α-^{32}P] dTTP (>3000Ci/mmol) and [γ-^{32}P] ATP (>3000Ci/mmol) were from Amersham International. All other reagents required were analytical grade. DNA manipulations followed standard procedures as outlined in Molecular Cloning; A Laboratory Manual (10), enzymes were used according to the manufacturers recommendations.

Plasmids pMA27, and pMA773−777 were obtained from J. Ogden and C. Stanway (8). pMA118 was constructed by C. Stanway by ligating the PvuII/BamHI fragment from a molecule containing a PGK promoter deletion with a 3′ deletion endpoint at −49 ligated to BamHI linker (5′ CCGGATCCGG 3′, BRL) with a BamHI/PvuII fragment from an molecule containing a PGK promoter derivative with a 5′ endpoint at −43, again ligated to a BamHI linker, in a manner similar to the construction of pMA777 (8). In this way a molecule containing the PGK gene driven by a promoter with a BamHI linker at −43, and a duplication of sequences −49 to −43 flanking the linker was constructed.

To construct the molecules pMA118S1 and pMA118S2, pMA118 was cut at the BamHI site and a synthetic linker (5′ GATCGGATCC 3′; Celltech) ligated into the molecule. This resulted in the loss of the original BamHI site and the insertion of a new site with the linker, creating pMA118S1. pMA118S2 was created in the same manner from pMA118S1. The molecules result in a net increase of DNA of 10 and 20bp respectively when compared to pMA118.

Construction of pMA777\1, pMA777\2, pMA77791, pMA118\1, pMA118\2 and pMA11891 was achieved by digesting pMA118 (figure 3A) and pMA777 (8, figure 2A) with BamHI and inserting into these sites purified DNA fragments from lambda (\1 and \2) or pBR322 (91). The lambda fragments used were the 60bp BglII fragment (\2) and the 79bp BamHI/BglII fragment (\1). A 91bp Sau3A1 fragment from pBR322 (91) was also used.

Deletion of the CT box was achieved by ligating 3′ deletion endpoints of the PGK promoter which progressively delete the CT box to a 5′ endpoint at −85. Both endpoints were ligated to a BamHI site, and construction of the molecules was by a method analogous to that used by Ogden *et al.* (8) for the construction of pMA777. The 3′ endpoints used were at −73, −65 and −49.

The deletion of sequences between the CT box and the I site, −49 to −39, was achieved be cleaving pMA118 (figure 3A) at the BamHI site and digesting the linearised molecule with the exonuclease III and Mung Bean Nuclease (BRL). Linearised DNA was resuspended to a final concentration of 200 ng of DNA per microgram in BRL 'React 1' (50mM Tris-HCl, pH 8.0; 10 mM $MgCl_2$) and chilled on ice. To the chilled DNA solution 100 units of Exonuclease III (Pharmacia) was added, and the reaction allowed to proceed for 15 seconds. The reaction was stopped by the addition of phenol, and the DNA recovered from the aqueous phase by ethanol precipitation. The DNA was treated with Mung Bean Nuclease (Pharmacia) to make flush ended. The deleted DNA was religated and the endpoints characterised using M13 dideoxy sequencing (11).

Yeast Transformation

Yeast transformations were performed following the procedure of Hinnen *et al.* (12). Transformed yeast strains were designated T, e.g. T777 is a strain containing pMA777.

RNA Analysis

RNA was isolated from yeast transformants, grown in defined minimal media as described above, according to the method of Dobson (13). RNA was quantitated using Northern blots (14). Filters were hybridised according to Thomas (14) using nick-translated probes (15) from the yeast PGK gene coding sequence and a yeast ribosomal DNA fragment (16).

RNA start site mapping was performed using primer extension analysis. 7ng of an endlabelled oligonucleotide was mixed with 5μg of total RNA and 3μl of 10× reverse transcriptase buffer (0.5M Tris pH 7.4, 0.75M KCl and 0.03M $MgCl_2$) in a total of 25.5μl and the mixture heated to 90°C for 5 minutes followed by a slow cooling to room temperature to allow annealing of the oligonucleotide and RNA. The annealed RNA mix was made upto 30μl with the addition of 3μl 0.1M DTT, 0.6μl of 50× dNTP mix (25mM of each of the dNTPs) and 200u of reverse transcriptase (Moloney Murine Leukaemia Virus Reverse Transcriptase, BRL) are added and the reaction incubated at 37°C for 1 hour. Following this, the RNA is degraded with the addition of 1/10th volume of 1m NaOH and heating to 70°C for 10 minutes. The mixture is neutralised with the addition of 1/10th volume of 1M Tris, pH 7.4, and the transcription products ethanol precipitated. Products were resolved on a standard 8M urea, 7.5% polyacrylamide sequencing gel and the size of products ascertained by comparison with a dideoxy sequencing reaction or comparison with radioactively labelled markers.

Determination of plasmid copy number

Total DNA was isolated from transformants using the method of Holm (17), digested with EcoRI, electrophoresed on 1% agarose gels, and transferred to nitrocellulose filters as described by Southern (18). Filters were hybridised with pBR322 and rDNA specific nick translated probes (15), and comparison of the two bands an estimate of plasmid copy number. In all transformants used a copy number of approximately 100 was seen.

RESULTS

The sequence of the *PGK* promoter from −420 to +119 is shown in figure 1. The effect of deletions on *PGK* promoter function were determined by using Northern blot to examine the steady-state *PGK* RNA levels of yeast transformants, and by using primer extension analysis of *PGK* RNA to determine the sites of RNA initiation used in the transformants. Finally, the levels of PGK protein were examined on Coomassie blue stained SDS polyacrylamide gels.

Sequences upstream from the I site are essential for correct mRNA initiation

In the majority of yeast promoters analysed to date TATA element(s) have been found to have an essential function in correct mRNA initiation (5,19,20,21,22,). The sequence of the *PGK*

-420
CTCACACAAC AAGGTCCTAG CGACGGCTCA CAGGTTTTGT AACAAGCAAT

-370
CGAAGGTTCT GGAATGGCGG GAAAGGGTTT AGTACCACAT GCTATGATGC

-320
CCACTGTGAT CTCCAGAGCA AAGTTCGTTC GATCGTACTG TTACTCTCTC

-270
TCTTTCAAAC AGAATTGTCC GAATCGTGTG ACAACAACAG CCTGTTCTCA

-220
CACACTCTTT TCTTCTAACC AAGGGGGTGG TTTAGTTTAG TAGAACCTCG

-170
TGAAACTTAC ATTTACATAT **ATATAAACTT** GCATAATTG GTCAATGCAA

-120
GAAATACATA TTTGGTCTTT TCTAATTCGT AGTTTTTCAA GTTCTTAGAT

-70
GCTTTCTTTT TCTCTTTTTT ACAGATCATC AAGGAAGTAA TTATCTACTT

-20 +1
TTTACAACAA ATATAAAACA ATGTCTTTAT CTTCAAAGTT GTCTGTCCAA

+30
GATTTGGACT TGAAGGACAA GCGTGTCTTC ATCAGAGTTG ACTTCAACGT

+70
CCCATTGGAC GGTAAGAAGA TCACTTCTAA CCAAAGAATT GTTGCTTGCTT

Figure 1. The nucleotide sequence of the *PGK* promoter between −420 and +119. The *PGK* TATA homologies (bold type), the CT...CAAG and the ATG (underlined) are all indicated. The site of RNA initiation within the CAAG sequence is indicated by bold type. The numbering of the *PGK* promoter is relative to the ATG at +1.

promoter reveals two TATA consensus sequences, one perfect (TATA1 at −152) and one imperfect (TATA2 at −113; figure 1). *PGK* promoter derivatives deleted of sequences including one or both of the TATA homologies, (8, figure 2A), were analysed to assess the role of these sequences in RNA initiation. Transformants containing the deleted derivatives of the *PGK* promoter (figure 2A) and the wild-type *PGK* promoter (pMA27, 8, figure 2A) on high copy number 2μ based plasmids are represented by T followed by the number given to the deletion. The data in figures 2B and 2C shows the effect of deleting a region containing TATA1, TATA2, or both TATA1 and TATA2 on steady state RNA and protein levels compared to the wild type gene on a high copy number plasmid (T27) and the chromosomal gene (MD40/4c). Deletion of TATA2 alone (T777) or a small region between the TATA elements (T776) had no effect on either the steady state RNA levels or the amount of PGK protein produced. Deletion of TATA1 (T773 and 774) resulted in mRNA levels similar to the intact promoter but a reduction in the levels of PGK protein compared to T27. Deletion of a region containing both TATA elements (T775) had a small effect on the steady state RNA levels but the levels of PGK protein were reduced dramatically. These results indicated that although high mRNA levels were seen with all deletions, in T775 the *PGK* RNA does not appear to be translated into PGK protein.

The RNA initiation sites used in the deleted promoters were determined by primer extension using an oligonucleotide which hybridised to sequences between +199 to +216 of the *PGK* RNA (figure 3D). Primer extension of wild-type *PGK* RNA from this oligonucleotide generates a single product of 256 bases. T27, T776 and T777, containing either both TATA1 and TATA2 (T27 and T776) or TATA1 alone (T777) produce high levels of RNA which is initiated at the wild-type I site. T773 and T774, containing deletions of the region containing TATA1, produce slightly reduced levels of the 256 base primer extension product

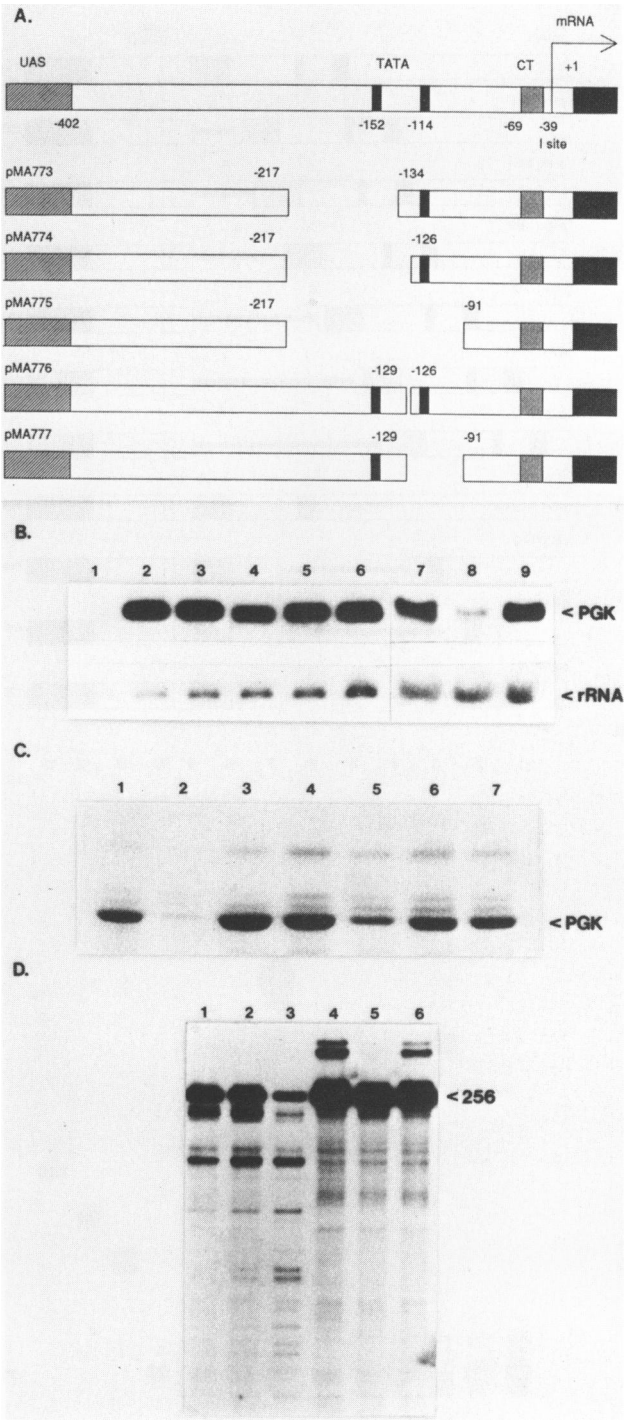

Figure 2. Analysis of the role of the TATA homologies within the *PGK* promoter. **2A.** Schematic representation of the wild-type *PGK* promoter and promoter derivatives pMA773, pMA774, pMA775, pMA776 and pMA777 (8) which have been deleted of one or both of the *PGK* TATA homologies (solid boxes). **2B.** Northern blot analysis of the steady state *PGK* RNA levels of MD40 \ 4c (lanes 1 and 8) and MD40 \ 4c transformed with pMA773 (lane 2), pMA774 (lane 3), pMA775 (lane 4), pMA776 (lane 5), pMA777 (lanes 6 and 7) and pMA27 (lane 9). Filters were hybridised with a *PGK* specific probe, and a rRNA probe. The rRNA signal was used as an internal loading control. **2C.** Total protein extracts of the haploid yeast, MD40 \ 4c, (lane 2) and MD40 \ 4c transformed with pMA27 (lane 1), pMA773 (lane 3), pMA774 (lane 4), pMA775 (lane 5), pMA776 (lane 6) and pMA777 (lane 7). PGK protein is indicated by the arrow. **2D.** Primer extension of RNA extracted from T773 (lane 1), T774 (lane 2), T775 (lane 3), T776 (lane 4), T777 (lane 5) and T27 (lane 6) from a primer which hybridised to sequences between +199 and +219 of the *PGK* coding sequence. The wild-type *PGK* RNA yields a product of 256 bases, as indicated.

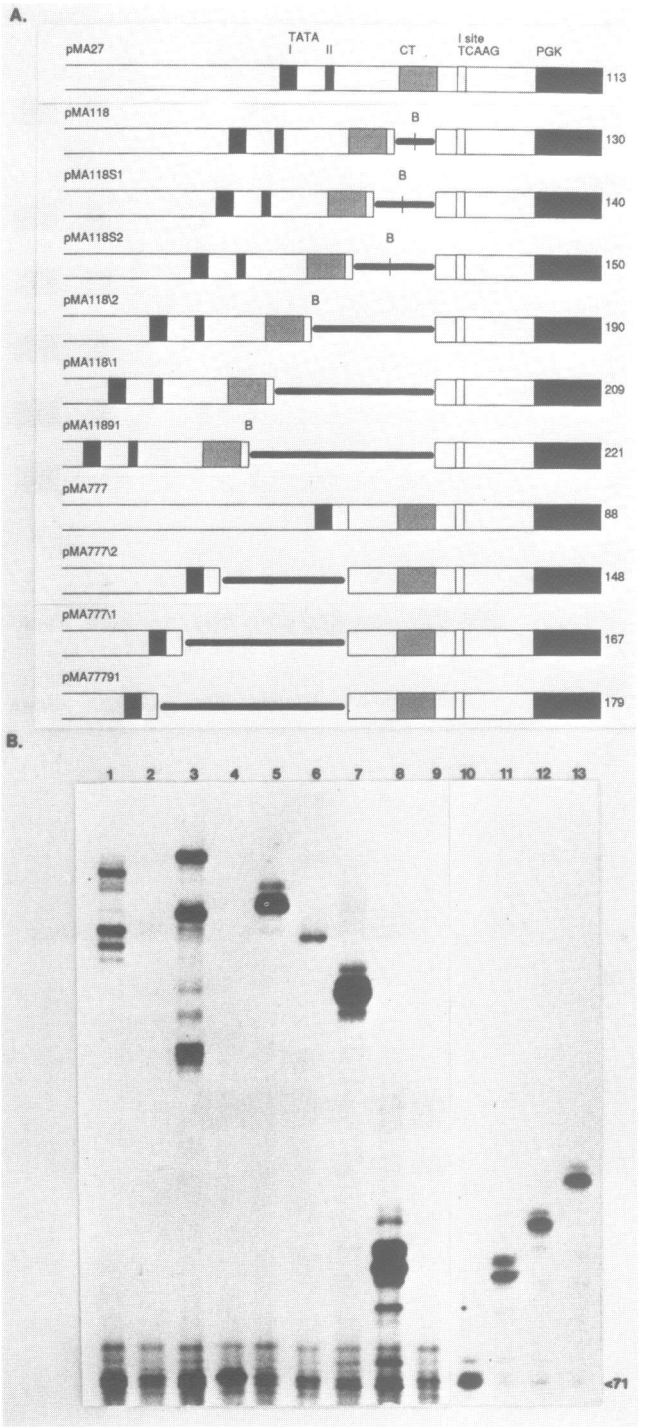

Figure 3. Analysis of *PGK* promoter derivatives spaced between TATAI and the I site. **3A.** A diagrammatic representation of the wild-type *PGK* promoter, as contained on pMA27, and the derivative promoters which contain spacing DNA (B, BamHI site; CT, CTbox; solid region, the *PGK* coding region). To space TATAI and the I site heterologous DNA derived form bacteriophage lambda and pBR322 was inserted into the BamHI sites of pMA118 and pMA777, both of which are shown. The construction of the promoters is outlined in materials and methods. The numbers to the right of each promoter indicate the distance in base pairs between TATAI and the wild type I site. **3B.** Primer extension of RNA extracted from MD40 \ 4c (lane 9), T27 (lane 10), T118 (lanes 8 and 11), T118S1 (lane 12), T118S2 (lane 13), T118 \ 2 (lane 7), T118 \ 1 (lane 6), T11891 (lane 5), T777 (lane 4), T777 \ 2 (lane 3), T777 \ 1 (lane 2) and T77791 (lane 1) from a *PGK* specific probe which hybridises to sequences between +10 and +31 of the *PGK* coding region. The extension product of 71 bases generated from wild-type *PGK* RNA is indicated. The large extension products in lanes 5, 6 and 7 have been sized at approximately 180, 167 and 150 bases respectively.

and, in addition, a number of smaller extension products. Primer extension of RNA extracted from T775 shows that the majority of the RNA produced from this plasmid initiates at a number of sites 3′ to the wild-type I site. These new RNA initiation sites mapped to sequences between +40 and +85 of the *PGK* coding region (23). The amount of the 256 base extension product for each transformant reflects the steady-state levels of PGK protein produced, and suggests that the new initiation sites used in T775 produce RNA which is not translated into full length PGK protein. These data suggest that the TATA homologies found within the *PGK* promoter are required for RNA initiation at −39. TATA1 can function alone to determine a discrete initiation site, but in its absence TATA2 is partly functional.

The RNA initiation site is not determined by the distance from the TATA element nor by the I site sequence (CAAG)

In order to investigate the spatial relationship between the region of the *PGK* promoter containing the TATA elements and the I site, the TATA elements were separated from the I site by insertion of heterologous DNA fragments at two sites within the promoter (Figure 4A). In the first experiment linker sequences or fragments from bacteriophage lambda (60bps or 79bps) and pBR322 (91bps) with GATC ends were inserted into a BamHI linker in the *PGK* promoter between the CT box and the CAAG. This resulted in a series of promoter derivatives in which the TATA1:I site distance was increased from 113bp in the wild-type promoter to between 130 and 221bp. In all of the promoters constructed the wild-type *PGK* I site, the CAAG sequence, was moved beyond the generally accepted outer limits of the initiation window (3). Plasmids containing these promoter derivatives were transformed into yeast on a 2μm based high copy number plasmid and the pattern of RNA initiation assessed by primer extension from an oligonucleotide complementary to sequences between +10 to +31 of the *PGK* RNA (Figure 4B). RNA initiation in all promoters containing inserted DNA occurred at a single, major site (figure 4B). This site mapped to a position approximately 12bp 3′ of the CT box in a position analogous to the wild-type I site, regardless of the size of the inserted DNA. This suggests that sequences 5′ to the site of RNA initiation in the *PGK* promoter contain the information necessary to determine discrete RNA initiation at that position.

Sequences 5′ to the *PGK* RNA initiation site include the TATA elements. In mammalian promoters, the site of RNA initiation is determined by the position of the TATA element (24,25). To investigate the possibility that a similar situation occurs in the *PGK* promoter the DNA fragments from bacteriophage lambda and pBR322 were inserted into a second site in the *PGK* promoter, between TATA1 and the CT box...CAAG (figure 4A). pMA777 (8, figure 2A) has a deletion between −129 and −91 and a BamHI linker in place of these sequences which results in a net deletion of 25bps of the *PGK* promoter. This deletion has no effect on the steady-state RNA levels or the site of RNA initiation (see figure 2A). Insertion of DNA at this site resulted in the distance between TATA1 and the RNA initiation site being increased from 113bps in the wild type promoter to 148bps (pMA777 \ 2), 167bps (pMA777 \ 1), and 179bps (pMA77791). RNA from these promoters was initiated at a number of sites within the inserted DNA and not at a single site 113bp 3′ of TATA1. These data together with data from *PGK* promoter derivatives deleted of sequence between the TATA homologies and the I site (8, figure 2) suggest that the RNA initiation site is not determined a set distance downstream from the TATA element.

The data from these spacing experiments and from pMA777, suggests that the site of RNA initiation in the *PGK* promoter is determined by sequences located 5′ to the CAAG motif at −41. The spacing experiments suggest that the information determining the site of RNA initiation is contained between −91 and −43. This region of sequence includes the CT box and also sequences between the CT box and the CAAG motif.

The CT box is not essential for transcription initiation in the *PGK* promoter

A purine rich region, the CT box, which precedes the site of RNA initiation is a feature common to many highly expressed yeast genes (7). However, no function has been ascribed to this region. A series of small deletions of the *PGK* promoter which left the sequences 5′ to −85 intact and removed all or part of the CT box were created (figure 3A). Neither the levels of *PGK* RNA produced or the site of RNA initiation used was affected by these promoter modifications (figure 3B). This data suggests that the CT box is not involved in RNA initiation site selection.

The short sequence between the CT box and the TCAAG functions in mRNA start site determination

As deletion of the CT box did not influence start site determination in the *PGK* promoter, we looked more closely at a possible role for the DNA located between the CT box and the CAAG in start site determination. A series of small internal deletions were created in this region and the end points religated in the absence of linker sequences (figure 5A). Deletion of 6, 7 or 9bps from this region resulted in a reduction in the amount of *PGK* RNA which was initiated at the wild-type I site, and the use of a number of sites 3′ to the wild-type I site for RNA initiation. RNA from deletion 7, which removes 6bps but leaves the (T)CAAG sequence at the I site intact, was initiated at the wild-type I site and a number of sites 3′ to this (figure 5B). Further deletions which both removed part of the TCAAG sequence, disrupted initiation at the wild-type I site and resulted in the use of the 3′ I sites (figure 5B). From this we conclude that the short sequence found between the CT box and the CAAG, 5′.ACAGATCA.3′, functions to determine the single site of RNA initiation used in the *PGK* promoter.

DISCUSSION

We have undertaken to define *cis*-acting sequences required for the correct initiation of RNA from the *PGK* promoter. Using deletion and insertion analysis of the promoter we have defined two *cis*-acting sequences which function in RNA initiation, the TATA element at −152 (TATAI) and the sequences immediately 5′ to the CAAG sequence motif at the site of RNA initiation. In addition, we examined the CT box by deletion analysis but were unable to ascribe a function to this sequence.

The role of TATAI and TATAII in the initiation of RNA from the *PGK* promoter

The DNA sequence of the *PGK* promoter shows the presence of TATA homologies at −152 (TATAI) and −113

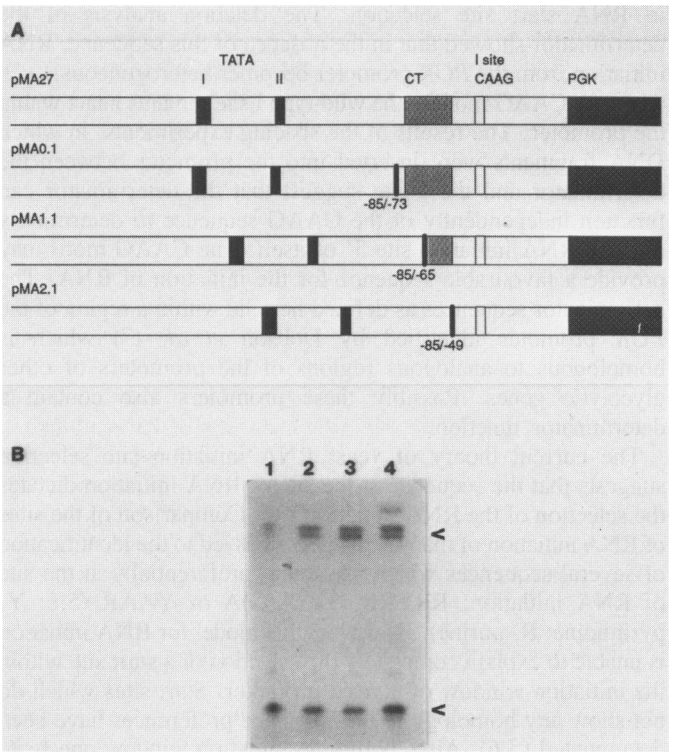

Figure 4. Analysis of *PGK* promoter derivatives deleted of the purine rich region 5′ of the I site, the CT box. **4A.** Schematic representation of the wild-type promoter, pMA27, and the deletions pMA0.1, pMA1.1 and pMA2.1. pMA0.1 contains a deletion between −85 and −73, 5′ to the CT box. pMA1.1 is deleted of sequences between −85 and −65, which includes about half of the CT box. pMA2.1 is deleted of the entire CT box. **4B.** Primer extension of RNA extracted from T27 (lane 1), T0.1 (lane 2), T1.1 (lane 3) and T2.1 (lane 4) using *PGK* and *PYK* specific primers. The upper arrow indicates the extension product generated from wild type *PGK* RNA. The lower arrow shows the position of the *PYK* product which was used as an internal loading control.

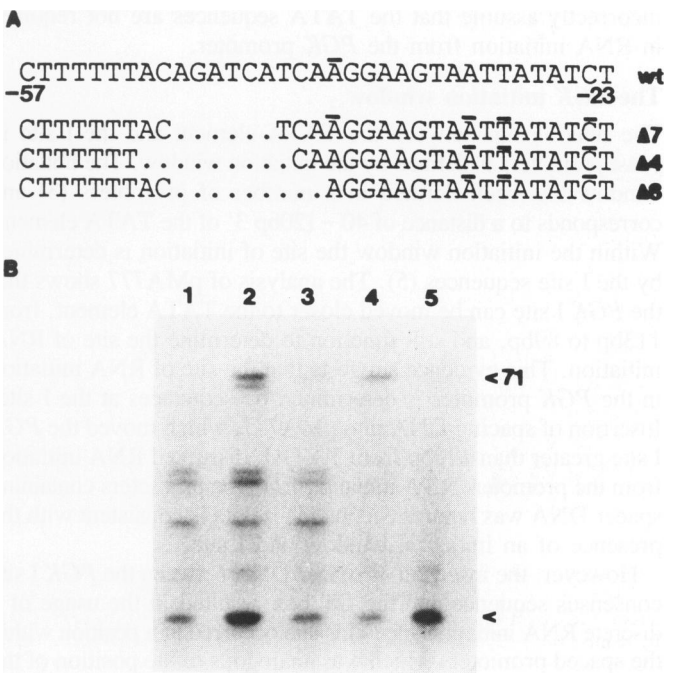

Figure 5. Analysis of *PGK* promoters deleted of sequences between the CT box and the I site. **5A.** Sequence across the site of the deletions between the CT box and the CAAG. The sequences of pMA118Δ7 (Δ7), pMA118Δ4 (Δ4) and pMA118Δ8 (Δ8) are shown compared to the wild-type (wt) sequence. **5B.** Primer extension analysis of RNA extracted from MD40 \ 4c (lane 5), T27 (lane 4), T118Δ8 (lane 1), T118Δ7 (lane 2) and T118Δ4 (lane 3). RNA was extended with *PGK* and *PYK* specific probes. The *PGK* primer extension product of 71 bases is indicated by the labelled arrow. The lower arrow indicates the *PYK* specific extension product.

(TATAII)(figure 1). Analysis of *PGK* promoters deleted of either TATAI or TATAII suggest that both of these sequences are capable of functioning in RNA initiation. Promoters containing TATAI alone initiate RNA efficiently at the wild-type I site. The pattern of RNA initiation from promoters containing TATAII alone is slightly heterogeneous, but the majority of RNA initiates correctly. A promoter deleted of both TATAI and TATAII produces high levels of RNA but primer extension analysis of the RNA shows that the majority of the RNA initiates 3′ of the wild-type I site, within the coding region. The high levels of transcription from pMA775, and the position of the sites used to initiate the aberrantly initiated RNA from pMA775 are consistent with the creation of a novel promoter in the deletion. This promoter may consist of the *PGK* UAS activating transcription from a TATA homology found at −10, within the 5′ untranslated region. In the absence of an I site this novel promoter initiates RNA heterogeneously within a window approximately 50 to 100bps 3′ of the −10 TATA sequence.

The analysis of these deleted promoter derivatives suggests that RNA initiation from the *PGK* promoter requires TATAI. The results from a previous analysis of the promoter deletions used in this study (8) suggested that the TATA homologies within the *PGK* promoter were not required for correct RNA initiation. One explanation for the difference between our results and those of Ogden *et al.* (8) is that the primer extensions presented here were performed using an oligonucleotide which hybridised to sequences between +199 and +216 and would detect the aberrantly initiated RNA species. In contrast, the riboprobe analysis presented by Ogden *et al.* (8) used a probe which hybridised to sequences 5′ of +37 and which would not have detected the incorrectly initiated RNA species. This may have led the authors to incorrectly assume that the TATA sequences are not required in RNA initiation from the *PGK* promoter.

The *PGK* initiation window

The relationship between the TATA element and the I site in yeast promoters is defined by the initiation window. The initiation window has been defined in a number of promoters (3) and corresponds to a distance of 40−120bp 3′ of the TATA element. Within the initiation window the site of initiation is determined by the I site sequences (5). The analysis of pMA777 shows that the *PGK* I site can be moved closer to the TATA element, from 113bp to 89bp, and still function to determine the site of RNA initiation. This evidence suggests that the site of RNA initiation in the *PGK* promoter is determined by sequences at the I site. Insertion of spacing DNA into pMA777, which moved the *PGK* I site greater than 120bp from TATAI, disrupted RNA initiation from the promoter. RNA initiation from the promoters containing spacer DNA was heterogeneous. This data is consistent with the presence of an initiation window and I site.

However, the insertion of spacer DNA between the *PGK* I site consensus sequence and the CT box resulted in the usage of a discrete RNA initiation site. This site occurred at a position within the spaced promoters which was analogous to the position of the wild-type I site, relative to TATAI. In all of the promoters based on pMA118 this site occurred on a BamHI linker sequence used for the insertion of the spacing DNA fragments. However, analysis of *PGK* promoter deletion pMA777 (8, figure 2), in which a BamHI site has been inserted into the *PGK* initiation window, suggest that the BamHI linker sequence is not capable of functioning as an initiation site. In these experiments it appeared that the I site determinant had not been removed from the initiation window, suggesting that a sequence distinct form the I site consensus sequence CAAG was involved in the determination of the RNA start site.

The role of sequences between −91 and −43 in RNA initiation from the *PGK* promoter

The data from the *PGK* promoters containing spacing DNA suggested that a sequence distinct from the CAAG motif was involved in the determination of the site of RNA initiation. This sequence mapped to between −91 and −43. A deletion analysis of sequences within this region identified sequences, between −43 and −49, as functional in the determination of an RNA initiation site. Deletion of this sequence results in heterogeneous RNA initiation.

It is interesting to note that the *cis*-acting sequence requirements of RNA initiation in *PGK* do not include the CT box. The deletion of the CT box showed that this sequence is not required for the accurate initiation of RNA or for the maintenance of high levels of transcription. This is a surprising observation given the high level of conservation of this sequence between the promoters of highly expressed yeast genes (7).

Sequence requirements at the site of RNA initiation of the *PGK* promoter

Analysis of the *cis*-acting sequence requirements at the *PGK* I site suggested that the I site was defined by a sequence immediately 5′ to the site of RNA initiation the CAAG. These sequences, which we call the determinator, appear to function in RNA start site selection. The deletion analysis of the determinator showed that in the absence of this sequence, RNA initiation from the *PGK* promoter becomes heterogeneous, even when the CAAG motif at the wild-type I site remains intact within the promoter. The results of the spacing experiments, in which DNA fragments were inserted into the promoter between the determinator and the I site suggest that the determinator can function independently of the CAAG sequence to determine a discrete RNA initiation site 3′ of itself. The CAAG motif may provide a favourable sequence for the initiation of RNA. The determinator sequences as defined here lie within a region of the *PGK* promoter identified by Dobson *et al.* (7) which is homologous to analogous regions of the promoters of other glycolytic genes. Possibly these promoters also contain a determinator function.

The current theory of yeast RNA initiation site selection suggests that the sequence at the site of RNA initiation dictates the selection of the RNA initiation site. Comparison of the sites of RNA initiation of many yeast genes has led to the identification of several sequences which are found preferentially at the site of RNA initiation; RRYRR, TC(A/G)A or YAAR (5,6; Y, pyrimidine; R, purine). However, this model for RNA initiation is unable to explain completely the selection of a start site within the initiation window of a yeast promoter. Start sites which do not show any homology to the sequence preferences have been documented (5,6). Also, within an initiation window one I site consensus may be used in preference to other, apparently equivalent, I site consensus sequences (6,26). From the data presented here the *PGK* I site has been defined as a *cis*-acting sequence element, the determinator, followed by an I site consensus sequence. The determinator is functionally separable form the I site and appears to play a role in start site selection. The proposed function of the determinator in the *PGK* promoter resolves some of the inconsistencies inherent in the previous

model of start site selection. In the *PGK* promoter the RNA initiation site is not defined by an I site consensus sequence, but by the determinator.

Accurate initiation from the *PGK* promoter requires TATAI and the determinator. Although a determinator function has not been identified in the promoters of other yeast genes the sequence requirements for yeast RNA initiation in the majority of promoters have not been determined experimentally. Possibly a determinator function is present in the promoters in which RNA initiation occurs at a discrete site.

ACKNOWLEDGMENTS

We are grateful to C. Stanway, J. Ogden, A.J. Kingsman and S.M. Kingsman for the generous gift of plasmids, and K. Johnson for the photography. This work was supported by a Medical Research Council grant to J.Mellor.

REFERENCES

1. Mellor,J. (1989) in Walton,E.F. and Yarranton, G.T., (eds) 'Molecular Biology of Yeasts'. Blackie, Glasgow and London. 1–42.
2. Breathnach,R. and Chambon,P. (1981) *Ann. Rev. Biochem.* **50**, 349–383.
3. Struhl,K. (1987) *Cell* **49**, 295–297.
4. Chen,W. and Struhl,K. (1985) *EMBO. J.* **4**, 3273–3280.
5. Hahn,S., Hoar,E.T. and Guarente,L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8562–8566.
6. Healy,A.M., Hesler,T.L. and Zitomer,R.S. (1987) *Mol. Cell. Biol.* **7**, 3785–3791.
7. Dobson,M.J., Tuite,M.F., Roberts,N.A., Kingsman,A.J. and Kingsman,S.M. (1982) *Nucl. Acids Res.* **10**, 2625–2637.
8. Ogden,J.E., Stanway,C., Kim,S., Mellor,J., Kingsman,A.J. and Kingsman,S.M. (1986) *Mol. Cell. Biol.* **6**, 4335–4343.
9. Casadaban,M.J. and Cohen,S.N. (1980) *J. Mol. Biol.* **138**, 179–207.
10. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
11. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
12. Hinnen,A., Hicks,J.B. and Fink,G.R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1929–1933.
13. Dobson,M.J., Mellor,J., Fulton,A.M., Roberts,N.A., Bowen,B.A., Kingsman,S.M. and Kingsman,A.J. (1984) *EMBO J.* **3**, 1115–1119.
14. Thomas,P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
15. Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.* **113**, 237–251.
16. Petes,T.D, Hereford,L.M. and Konstantin,K.G. (1978) *J.Bacteriol.* **134**, 295–305.
17. Holm,C., Meeks-Wagner,D.W., Fangman,W.L. and Botstein,D. (1986) *Gene* **42**, 169–173.
18. Southern,E.M. (1975) *J. Mol. Biol.* **98**, 503–517.
19. Struhl,K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4451–4465.
20. Faye,G., Leung,D., Tatchell,K., Hall,B.D. and Smith,M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2258–2262.
21. Nagawa,F. and Fink,G.R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8557–8561.
22. Siliciano,P.G. and Tatchell,K. (1984) *Cell* **37**, 969–978.
23. Rathjen,J. (1989) D. Phil. Thesis, Oxford.
24. Benoist,C. and Chambon,P. (1981) *Nature* **290**, 305–310.
25. Ghosh,P.K., Lebowitz,P., Frisque,R.J. and Gluzman,Y. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 100–104.
26. Rudolph,H. and Hinnen,A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1340–1344.