Volume 11 Number 24 1983 Nucleic Acids Research

Nucleotide sequence of the transcriptional initiation region of the yeast *GAL7* gene

Yasuhisa Nogi and Toshio Fukasawa

Laboratory of Molecular Genetics, Keio University School of Medicine, Shinjuku-ku, Tokyo 160, Japan

Received 25 October 1983; Accepted 21 November 1983

ABSTRACT

The GAL7 gene of Saccharomyces cerevisiae encodes Gal-1-P uridylyl transferase, the second enzyme of Leloir pathway for the galactose catabolism. We have determined the sequence of 1003 base pairs surrounding and upstream of the transcriptional initiation site of the GAL7 gene. The region sequenced also encompasses the 3' end of GAL10 gene. The 5' end of GAL7 mRNA was determined on the DNA sequence by the S1 nuclease- and exonuclease VII mapping, which is located 21 to 22 base pairs upstream from the translation initiating ATG codon. The primary structure of the GAL7 5' flanking region has many features common to those of multicellular eukaryotic genes. The 3' end of GAL10 mRNA was also determined by the mapping technique with the single-strand specific nucleases to be about 600 base pairs upstream from the 5' end of GAL7 mRNA.

INTRODUCTION

The galactose utilization system in Saccharomyces cerevisiae has been extensively studied genetically as well as biochemically: Three genes designated GAL7, GAL10 and GAL1 encode galactose-1-phosphate uridylyl transferase (transferase), uridine diphosphoglucose 4-epimerase (epimerase) and galactokinase respectively, which catalize the first three steps of the galactose catabolism (1). These enzymes have been purified to homogeneity (2, 3, 4). These genes, collectively referred to as the GAL cluster genes in this paper, are tightly linked to each other near the centromere of chromosome II. The gene order is known to be centromere-GAL7-GAL10-GAL1 (5). The GAL cluster genes have been cloned into λ phage (6) and the approximate location of each gene is assigned on the cloned DNA (7). It has also been shown that the transcription of each gene starts from the respective promoter (8).

The expression of the GAL cluster genes is positively regulated by the genes GAL4 and GAL11, and negatively by the GAL80 gene in a coordinate fashion (9-13) at the transcriptional level (6, 7). This gene expression is also under the control of the catabolite repression which is presumably mediated by the GAL82 and GAL83 genes (14). By an analogy with the mechanism of the

prokaryotic regulation, all or at least some of these gene products are assumed to bind the GAL cluster genes at the respective 5' flanking regions (10, 11, 13). Recently the regulatory genes GAL4 (15-17) and GAL80 (17, 18) have been isolated, which has opened the way to purify those regulatory proteins. These knowledges prompted us to identify the signals for the controlled transcription in the GAL cluster genes. As the first step, we have determined the nucleotide sequence of the 5' flanking region of GAL7 gene. The 5' end of GAL7 mRNA as well as the 3' end of GAL10 mRNA have been assigned on the DNA sequence. These results have revealed that the 5'-flanking region of GAL7 shares many features with the corresponding region of eukaryotic genes.

MATERIALS AND METHODS

Strains, media and growth

Saccharomyces cerevisiae strain YK3 (a trp1 his3 GAL$^+$), and YK3 transformed with the plasmid pYH3003 (YK3[pYH3003]) were used as sources of polyadenylated RNA (polyA$^+$RNA). Yeast strains were grown on synthetic minimal media containing 0.67% yeast nitrogen base w/o amino acids (Difco) and either galactose (SGal) or glucose (SD) as the carbon source. Escherichia coli K12 strain JA221 (recA1 leuB6 trpΔE5 hsdR$^-$ hsdM$^+$ lacY) harboring the plasmid pYF1016 was grown on L-broth containing 25 μg/ml ampicillin. Plasmid DNA was prepared by the cleared-lysate procedure (19) and further purified through Bio-Gel A-150 m column fractionation followed by the ethidium bromide/CsCl equilibrium centrifugation.

Recombinant plasmids

The plasmid pYH3003 (17) was constructed by inserting the 2.3 kilobases (kb) SalI fragment containing the complete sequence of the GAL7 gene and a 3' region of the GAL10 gene from λgtSc481 (6) into an E. coli-yeast shuttle vector, pYC1, a gift from Dr. B. Hohn via Dr. Y. Kikuchi of our laboratory at Keio University. The plasmid pYF1016 was constructed by inserting the same 2.3 kb SalI fragment into the SalI site of pBR322.

DNA sequence analysis

The plasmid pYF1016 was digested with HinfI, electrophoresed on a 4% polyacrylamide gel and the 1.0 kb fragment was eluted from the gel. The fragment was digested with AvaII, AluI or BstNI and labeled either at the 5' end with [γ-^{32}P]ATP and T4 polynucleotide kinase (20) or at the 3' end with an appropriate [α-^{32}P]dNTP and AMV reverse transcriptase (21). The nucleotide sequence was determined by the method of Maxam and Gilbert (20). The G, A>C, C+T, C reactions were carried out and analyzed on thin (0.3 mm) 10% or 20%

polyacrylamide gel (17 cm x 42 cm) containing 8.3 M urea. One AluI fragment (-293/-139 in Fig. 2) was subcloned into the HindII site of pUR222 and sequenced according to the method described by Rüther et al. (22).

Preparation of polyA$^+$RNA

Yeast was grown to an early phase of the logarithmic growth in either SD or SGal medium. Cells were converted to spheroplasts by the use of Zymolyase 60,000 (Kirin Brewery Co.), from which bulk of nucleic acids was extracted by repeated phenol extractions. PolyA$^+$RNA was selected from the nucleic acid sample essentially according to the method of Sripati and Warner (23) by oligo (dT)-cellulose column chromatography. PolyA$^+$RNA samples from yeast cells grown on SD or on SGal are respectively referred to as glu RNA or gal RNA in the text.

Nuclease S1 and exonuclease VII mapping

The S1 nuclease mapping was performed as described in Weaver and Weissmann (24). For the 5' end mapping of GAL7 mRNA, two kinds of probe DNA were employed; 1) 1.5 kb SalI/BglII fragment 5' end-labeled at the BglII site, 2) the HhaI/HinfI fragment (-200/+61) 5' end-labeled a the HinfI site (Fig. 1). The probe DNA was hybridized with 5 µg polyA$^+$RNA in 10 µl [80% formamide, 0.4 M NaCl, 0.04 M PIPES (pH6.4), 0.001 M EDTA] for 3 hr at 42C. The 10 µl of hybridization mixture was transferred to 100 µl of S1 buffer [0.25 M NaCl, 0.03 M sodium acetate (pH6.4), 0.001 M $ZnSO_4$, 2 µg of denaturated λDNA] and digested with 400 units of S1 nuclease (Seikagaku Kogyo Co.) for 1 hr at 25C. For the exonuclease VII mapping (25) the 10 µl of hybridization mixture was transferred to 100 µl of 'exo VII buffer'[0.003 M KCl, 0.001 M Tris-HCl (pH7.4), 0.01 M EDTA] and incubated with an appropriate amount of exonuclease VII for 1 hr at 25C. The concentration of exonuclease VII was determined empirically. The nuclease-resistant materials were ethanol-precipitated and then dissolved with 20 µl of 80% formamide, 0.05 M NaOH, 0.1% xylane cyanol and 0.1% bromophenol blue. The protected DNA was loaded on 4%, 8% or 10% polyacrylamide/7 M or 8.3 M urea gel together with size markers or sequencing ladders. For the 3' end mapping of GAL10 mRNA, HinfI/HhaI fragment (-942/-427) 3' end-labeled at the HinfI site was employed as the probe. The subsequent procedures were exactly as described above in the 5' end mapping.

RESULTS

Determination of the nucleotide sequence

A yeast gal7 mutant harboring a plasmid that bears the 2.3 kb fragment from the GAL cluster region, such as pYH3003, exhibits the normal pattern of

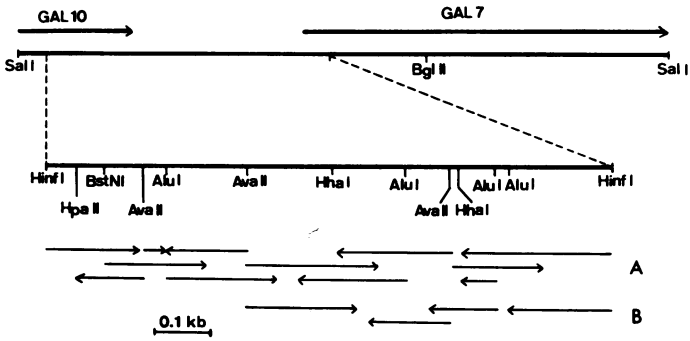

Fig. 1 Restriction map of a 2.3 kb SalI DNA fragment bearing GAL7 gene and its surrounding region and strategy for DNA sequence determination. The 2.3 kb DNA fragment between two SalI cleavage sites is drawn as the straight line in the upper part. The arrows on the fragment indicate an approximate location and the direction of the transcripts from GAL7 and GAL10. (The promoter proximal region of the GAL10 gene is deleted from this fragment; see the text.) A 1.003 bp HinfI DNA fragment is drawn under the SalI fragment in an expanded scale with the restriction endonuclease sites relevant to the present paper. The thin arrows in the lower part indicate the directions and extents of sequencing from either 5'-labeled termini (A) or 3'-labeled termini (B).

induced synthesis of transferase (17). Furthermore we constructed a GAL7-lacZ fused gene by ligating the 1.5 kb SalI/BglII fragment (the left-hand SalI to BglII in Fig. 1) and the 6.2 kb lacZ' fragment (26) in a correct frame. The expression of the GAL7-lacZ gene in yeast was under the normal control involving both GAL4 and GAL80 genes (to be published). These results indicate that the complete set of the regulatory signals is contained within the SalI/BglII fragment. Therefore, we constructed a detailed restriction endonuclease cleavage map of the 2.3 kb SalI fragment by the partial digestion technique of Smith and Birnstiel (27). In Fig. 1, shown are only the cleavage sites used for sequencing or mRNA end determination. The nucleotide sequence of 1003 bp HinfI fragment was determined according to the strategy shown in the same figure. About 90% of the total sequence was determined on both strands (either same direction or opposite direction). Each junction point was confirmed by sequencing of overlapping fragments. The N-terminal amino acid of transferase was previously shown to be threonine (3). Only one reading frame was started by Met-Thr, at the position of 61 bp upstream from the right-end HinfI site as transcription proceeds left to right on the 1.0 kb HinfI fragment. We assumed that ATG positioning at +1~+3 was the initiation codon: This assumption was justified 1) by the determination of the 5' end of GAL7 mRNA (see below) in which that ATG was located most proximal to the 5' end of the mRNA

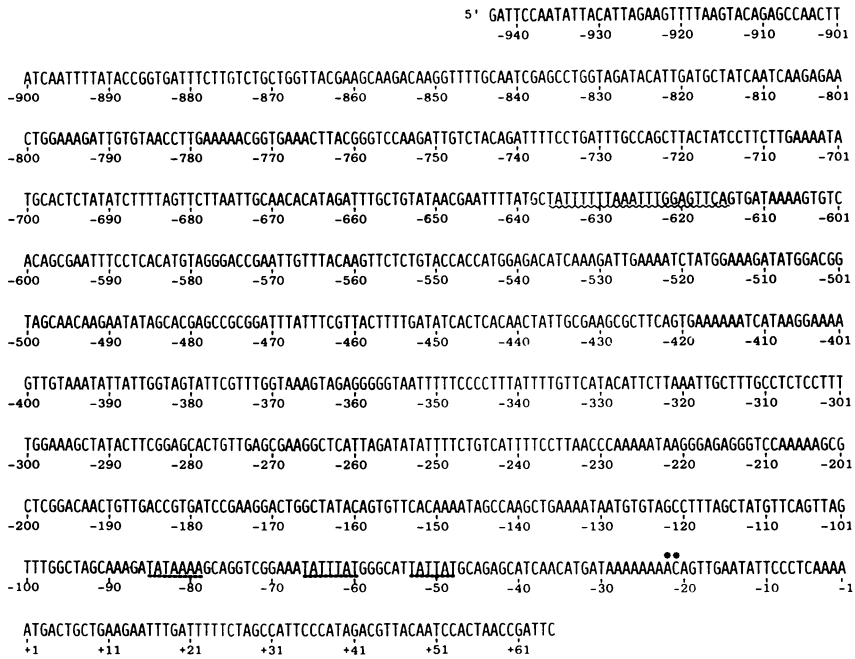

Fig. 2 DNA sequence of the region surrounding the transcription initiation site of GAL7 mRNA. The nucleotide sequence corresponds to the non-coding strand. The numbering begins at the A in the translation initiating ATG in each direction. The nucleotide sequence encompasses the 3' ends of GAL10 mRNA. The asteriks mark the 5' end of GAL7 mRNA and the wavy underline indicates 3' end of GAL10 mRNA. The putative TATA boxes are indicated by the broken underline.

as in most eukaryotic genes including the yeast genes (28, 29), and 2) by a preliminary determination of 7 N-terminal amino acid residues of transferase partially purified from YK3[pYK3003] by the manual Edman method (30), in which phenylthiohydantoin derivatives were identified by HPLC. Thr, Ala, Glu Glu, Phe, Asp, and Phe were detected in this order as major signals (unpublished), the sequence being exactly as expected from the nucleotide sequence shown in Fig. 2, taking the ATG in question as the initiation codon.

Mapping the 5' end of GAL7 mRNA

To determine an approximate location of the 5' end of GAL7 mRNA, we first used the 1.5 kb SalI/BglII fragment 5' end-labeled at the BglII site (see Fig. 1) as the probe which was hybridized to either glu RNA or gal RNA from YK3 or YK3[pYH3003]. After S1 nuclease digestion of the DNA-RNA hybrid, protected DNA was run on a 4% sequencing gel along with length markers (Fig. 3). The

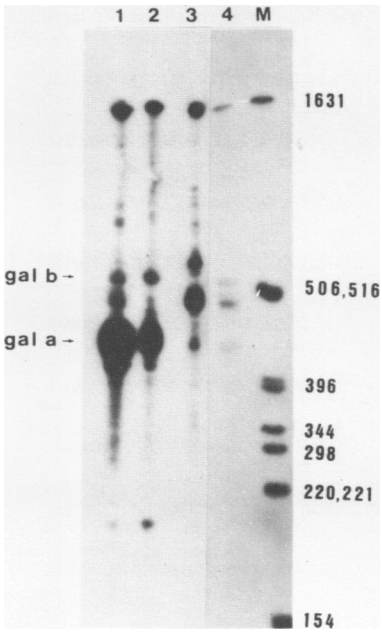

Fig. 3 S1 nuclease mapping of the 5' end of GAL7 mRNA. The SalI/BglII fragment 5' end labeled at the BglII site was hybridized to polyA$^+$RNA from the yeast grown in synthetic minimal medium containing galactose (lane 1, 2) or glucose (lane 3, 4). The DNA -RNA hybrids were digested with S1 nuclease and the nuclease resistant fragments were electrophoresed on a 4% polyacrylamide/7 M urea gel. The size markers are ^{32}P-end-labeled HinfI fragments of pBR322 (lane M). The figures on the right represent the size of the respective fragments in base pair. Lanes 1 and 3 shown the protected DNA by polyA$^+$RNA from the strain YK3 with the plasmid pYH3003 while lanes 2 and 4 the protected DNA by polyA$^+$RNA from YH3. Each lane contains the sample prepared with 5 µg polyA$^+$RNA.

major protection bands (designated gal a in Fig. 3) found in gal RNA from either strain was about 440 nucleotide long: We concluded the band to represent the 5'-end of the GAL7 mRNA for the following reasoning: 1) The protected DNA was much more abundant in gal RNA than in glu RNA in the respective strains, judged by the band intensity. 2) Similarly, that protected DNA was much more abundant in gal RNA from YK3 with the plasmid pYH3003 than in gal RNA from YK3 without the plasmid. Remember that the plasmid bears the multi-copy vector and that the transformant with the plasmid exhibits transferase activity at least five-fold higher than the normal strain (17).

The minor protection bands with a length of about 520 nucleotides in lanes 1 and 2 of Fig. 3 (designated gal b) seemed to be due to RNA transcribed from the host chromosomal gene but not to RNA from the gene on the plasmid because the abundance of the protected DNA appeared to be similar in both samples. The RNA in question may be the read-through transcript from the GAL10 promoter reported by St. John and Davis (7). Note that the GAL10 promoter is deleted in the plasmid pYH3003. We believe that the protection band (gal b) would not represent the 5'-end of that transcript but merely an artifact in the S1 mapping as discussed below. We neglected the other minor bands seen in Fig. 3, whose intensity did not increase by galactose on which yeast cells were grown for the preparation of RNA. Therefore, our attention here-

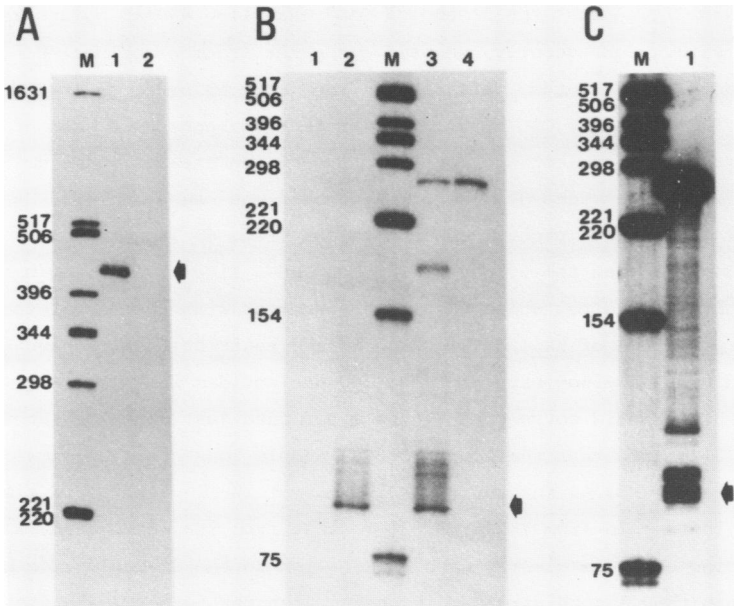

Fig. 4 S1 nuclease and exonuclease VII mapping of the 5' end of GAL7 mRNA. (A) The SalI/BglII fragment 5' end-labeled at the BglII site was hybridized to polyA$^+$RNA from YK3 with pYH3003 grown on either glucose (lane 1) or galactose (lane 2) as the carbon source. The DNA-RNA hybrids were digested with S1 nuclease and the nuclease-resistant fragments were electrophoresed on an 8% polyacrylamide/8.3 M urea sequencing gel. (B) The HhaI/HinfI fragment 5' end -labeled at the HinfI site was hybridized to polyA$^+$RNA from YK3 with pYH3003 grown on either glucose (lane 1) or galactose (lane 2). The same probe as above was hybridized to polyA$^+$RNA from YK3 grown on either galactose (lane 3) or glucose (lane 4). The DNA-RNA hybrids were treated with S1 nuclease and electrophoresed as above. The amounts of the samples loaded on lanes 1 and 2 were 1/30 of those of lanes 3 and 4. (C) The same probe to (B) was hybridized to polyA$^+$RNA from YK3 grown on galactose. The DNA-RNA hybrid was digested with exonuclease VII and protected fragments were electrophoresed as above. The protected band seen at 115 nucleotide long was assumed to be due to an incomplete digestion of the hybrid because the band was not seen when digestion was conducted at 45C (data not shown). The size standard (lane M in A to C) is the same as in Fig. 3.

after was focussed on the mojor band of about 440 nucleotides (gal a). In the experiment of Fig. 4, the amount of the sample loaded on a lane was reduced by a factor of more than 5 comparing to that in Fig. 3, such that only the major band was clearly detected in the autoradiography. The concentration of the polyacrylamide gel used was also changed to 8%. Thus, the band of about 440 nucleotide long was observed as the unique band in gal RNA while no band was obvious in glu RNA (Fig. 4A lane 1 and 2). The result indicated more

clearly than Fig. 3 that the 5' end of GAL7 mRNA lied 440±10 bp upstream from the BglII site. Subsequently, we used a shorter fragment of HhaI/HinfI (-200/+61) 5' end-labeled at the HinfI site as the probe in order to locate more precisely the 5' end of GAL7 mRNA. The protection bands were seen at a position around 85 nucleotides, irrespective of the strains, when gal RNA was used (Fig. 4B, lane 2 and 3). No promient band was seen when glu RNA from either strain was used under these conditions (Fig. 4B lane 1 and 4). The result indicated that the 5' end of the GAL7 mRNA lied about 85 bp upstream of the right HinfI site, which was in an agreement with the above experiment using the probe DNA of SalI/BglII based on the restriction map shown in Fig. 1. In addition, a minor band of about 180 nucleotides was seen (Fig. 4B lane 3), which should correspond to the minor band just described above (gal b in Fig. 3). Note that the amount of the sample loaded was 30-times as much in lane 3 (YK3) as in lane 2 (YK3[pYH3003]) in Fig. 4B. If a similar amount of the sample from YK3[pYH3003] to that from YK3 was loaded, the corresponding band was clearly seen in addition to the 85-nucleotide band, the latter exhibiting a much stronger intensity than the former (data not shown).

When the DNA/RNA hybrid prepared with the YK3 RNA was digested with exonuclease VII which cuts only a single strand end (31), instead of S1 nuclease which cuts an internally denatured region as well, two band clusters around 90- and 95-nucleotide long were observed as protected DNA (Fig. 4C). Taking into account the fact that exonuclease VII usually gives rise to a protected fragment five- to seven-nucleotide longer than that by S1 nuclease (32-34), the observed bands in Fig. 4C seemed to correspond to the band of 85 nucleotides in Fig. 3C. We thus reaffirmed that the major protection band observed in the S1 nuclease mapping represented the 5'-end of GAL7 mRNA but not an artifact which could have resulted from cleavage of DNA at an internally denatured region of the DNA-RNA hybrid by S1 nuclease. In this context, it should be pointed out that no band corresponding to the minor band in lane 3 of Fig. 4B was clearly seen in lane 1 of Fig. 4C. Therefore, we interpret the minor band seen in the S1 mapping to be an S1-digest of the probe DNA hybridized with the read-through transcript from the GAL10 promoter (7) at an internally denatured region.

In order to locate the 5' end of GAL7 mRNA at the nucleotide level, polyA$^+$ RNA from YK3[pYH3003] was hybridized to the HhaI/HinfI (-200/+61) fragment, and the S1 nuclease-resistant DNA was run on a 10% sequencing gel in parallel with the chemically cleaved products of the HhaI/HinfI fragment (Fig. 5). Comparing the position of the S1-resistant DNA fragment with the

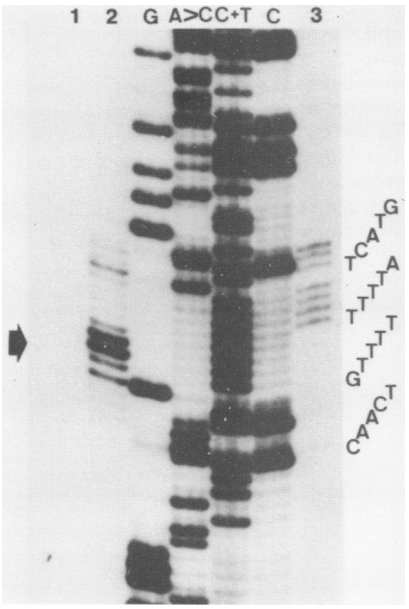

Fig. 5 Determination of the 5' end of GAL7 mRNA on nucleotide sequence. The HhaI/HinfI fragment 5' end-labeled at the HinfI site was hybridized to poly-A^+RNA from YK3 with pYH3003 grown on either glucose (lane 1) or galactose (lane 2), and polyA^+RNA from YK3 grown on galactose (lane 5). The DNA-RNA hybrids were treated with S1 nuclease (lane 1 and 2) or exonuclease VII (lane 3). Nuclease-resistant fragments were electrophoresed on a 10% polyacrylamide/8.3 M urea sequencing gel together with the sequencing ladders of the same probe (lane G, A>C, C+T, C). The partial sequence of the probe (coding strand) is shown on the right.

sequencing ladder after the correction of 1½ nucleotides (35, 36), we deduced the most probable 5' end of GAL7 mRNA to be either A at the position -22 or C at the position -21 on the non-coding strand (Fig. 5, Fig. 2). Since termini generated by S1 nuclease digestion are known to exhibit microheterogeneity of one to five nucleotides (35, 36), we were able neither to determine which of the two positions reflects the 5' end of GAL7 mRNA, nor to rule out the possibility of the 5' end of GAL7 mRNA to be double. Some additional bands around the major protection bands were thought to be due to the inherent heterogeneity of S1-resistant DNA as described above. A similar pattern of the protection bands was obtained with gal RNA from YK3 without plasmid (data not shown). When the protection fragments by exonuclease VII obtained in Fig. 4C were analyzed on the 10% sequencing gel (Fig. 5, lane 3), ladders were seen from -26 to -29 and from -32 to -34; the difference from those of lane 2 may be due to the intrinsic properties of exonuclease VII as described above. Since no precursor for the GAL7 mRNA has been detected by the Northern blot analysis (6; our unpublished data), we believe that the 5' end(s) of GAL7 mRNA described here is most likely the 5' end(s) of the primary transcript.

Mapping the 3' end of GAL10 mRNA

The HinfI/HhaI (-942/-427) fragment 3' end-labeled at the HinfI site was hybridized to glu RNA or gal RNA from YK3, digested with S1 nuclease or exo-

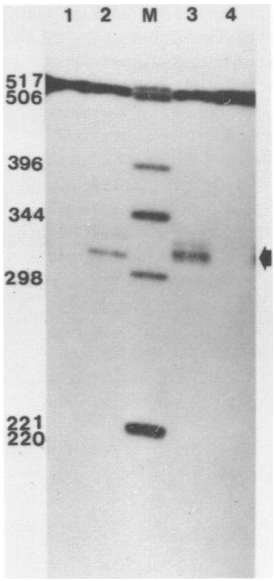

Fig. 6 S1 nuclease and exonuclease VII mapping of the 3' end of GAL10 mRNA. The HinfI/HhaI fragment 3' end-labeled at the HinfI site was hybridized to polyA$^+$RNA from YK3 grown on either glucose (lanes 1 and 4) or galactose (lanes 2 and 3). The DNA-RNA hybrids were digested with exonuclease VII (lanes 1 and 2) or S1 nuclease (lanes 3 and 4). Protected fragments were electrophoresed on a 8% polyacrylamide/ 8.3 M urea sequencing gel together with ^{32}P-labeled HinfI fragments of pBR322 as size markers (lane M).

nuclease VII and analyzed on an 8% sequencing gel along with length markers (Fig. 6). One major protection band about 315 nucleotide long was visible when gal RNA was used (Fig. 6, lanes 2 and 3) but not with glu RNA (Fig. 6, lanes 1 and 4), irrespective of the nucleases employed. Assuming the error in mapping to be within a range of 10 nucleotides, the result indicated that the 3' end of GAL10 mRNA resides around 315±10 bp downstream from left-hand HinfI site (Fig. 2). Another minor protection band about 320±10 bp was also observed (Fig. 6, lanes 2 and 3). It has been unclear yet whether this band represented the minor termination site or not.

DISCUSSION

The control signals for the transcription of eukaryotic genes by RNA polymerase II seem to be more complex than the prokaryotic counterparts. The signal for the promoter function is composed of at least three separate regions; the capping site for mRNA, the TATA box which fixes the transcriptional initiation site and the CAAT box which is required for an efficient initiation of transcription. The TATA box and the CAAT box are located generally 30 bp and about 80 bp upstream from the transcriptional initiation site respectively. Furthermore, the sequence more than 100 bp upstream from the transcriptional initiation site has been demonstrated to be indispensable for the controlled transcription in several eukaryotic genes (37-39). Com-

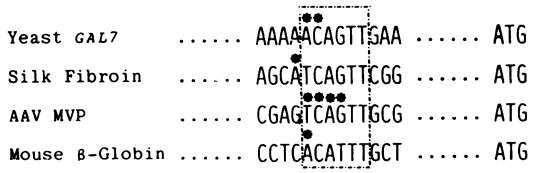

Fig. 7 The alignment of the regions surrounding the cap site or the presumed cap site of several eukaryotic genes. The sources of the nucleotide sequences are: GAL7 (this paper), Silk Fibroin (48), AAV MVP (the 2.3 kb transcript of the major viral protein gene of adenovirus-associated virus (49), Mouse major β-globin (40). The asteriks on bases indicate the 5' end of mRNA of the cap site.

parison of the DNA sequence of the 5'-flanking region of GAL7 with those of the other eukaryotic genes so far studied revealed resembling sequences at the corresponding locations: 1. The hexanucleotide ACAGTT (from -22 to -17) near the transcriptional initiation site displays a close homology with the capping site of several unrelated eukaryotic mRNAs (Fig. 7). However, a similar hexanucleotide TCAGTT is also present at -108 to -103 of GAL7 or 13 to 8 bp upstream from capping site of mouse major β-globin mRNA (40). Therefore, the hexanucleotide (A/TCAGTT) may represent only a part of the recognition signal for capping. 2. Preceeding the transcriptional initiation site, three AT rich sequences resembling the canonical TATA sequence are present; TATAAA (from -85 to -79), TATTTAT (from -66 to -60) and TATTAT (from -53 to -48). Which of those sequences is functional remains to be determined in future. The functional TATA box in yeast may not locate strictly 30 bp upstream from the transcriptional initiation site as in multi-cellular eukaryotes. In fact, the functional TATA sequence locates 36~39 bp upstream from the transcriptional initiation site in the HIS4 gene (33), while the ADH1 mRNA start sites scatter in the region about 90 to 100 bp from the possible TATA box (41). 3. The sequence GGACAACT (from -197 to -190) resembling CAAT box (GGpyCAATCT) is observed about 170 bp upstream from the mRNA start site. In yeast, the presence of the related sequence was reported only in the PGK gene (42) and the actin gene (43).

The 5' flanking region of an eukaryotic gene is known to be AT rich. The GAL7 flanking region is also AT rich, especially between -1 and -90 (70%). Two inverted repeats are observed upstream to the possible TATA boxes, (-200 to -191 and -189 to -180, and -167 to -159 and -128 to -120). Hinnebush and Fink (33) demonstrated that the 5' flanking regions of HIS1, HIS3, HIS4 and TRP5 contain the common repeated sequences and suggested these sequences to be

the recognition signal for the general amino acid control. The inverted repeats in GAL7 might be those who are recognized by the GAL4 protein and/or the GAL80 protein. In this context, we should refer to the recent work by Guarente et al. (44), in which they showed that 365-nucleotide sequence locating 133 bp upstream from the transcriptional initiation site of GAL10 gene, substituting the upstream activation sequence of the CYC1 gene, is able to activate the heterologous promoter under the normal regulation to the GAL cluster genes. They have thus suggested that the recognition signals for the GAL4 protein and possibly for GAL80 are localized in the region.

The consensus nucleotide surrounding the translation initiating ATG proposed by Kozak (28) is either purine (usually A) in position -3 or G in position +4 or both. In the GAL7 gene, A locates both in position -3 and +4, being 'semi-consistent' with Kozak's proposal. It is proposed that CT rich sequence of 5' non-coding region is related to the high efficiency of the gene expression in yeast (42) and the hexanucleotide CACACA close to the translation initiating ATG plays a role for ribosome binding (46). Neither of these sequences are present in the 5' non-coding region of the GAL7 gene. The region surrounding the 3' termini of GAL10 mRNA shown in the present work lacks the typical AATAAA sequence for polyadenylation in higher eukaryotic genes (46). However, the tripartite sequence for mRNA termination proposed by Zaret and Sherman (47) is found in the order of TAGT (from -684 to -681) . . . TAG (from -664 to -662) . . . AT rich . . . (from -644 to -642).

The comparison of the present result with the corresponding data in other co-regulated genes, such as GAL1, GAL10 or MEL1 may furnish a clue to the question as to which sequence is responsible for the recognition by the regulatory proteins. The alteration of the 5' flanking region introduced either in vivo or in vitro would allow us to identify such sequences. These experiments are now in progress in this laboratory.

ACKNOWLEDGEMENTS

We thank Dr. Haruo Ohmori and Mr. Takaaki Tamura for their advice in the DNA sequence analysis and Dr. Hisao Uchida for initiating the computer analysis for DNA sequence. Our thanks are also due to Dr. Hiroshi Handa for his generous gift of exonuclease VII and Dr. Hiroshi Matsubara for his help in the amino acid analysis. This work was supported by the MESC grant No. 58120006.

REFERENCES

1. Douglas, H.C. and Hawthorne, D.C. (1964) Genetics 49, 837-844.
2. Schell, M.A. and Wilson, D.B. (1977) J. Biol. Chem. 252, 1162-1166.

3. Segawa, T. and Fukasawa, T. (1979) J. Biol. Chem. 254, 10707-10709.
4. Fukasawa, T., Obonai, K., Segawa, T. and Nogi, Y. (1980) J. Biol. Chem. 255, 2705-2707.
5. Bassel, J. and Mortimer, R. (1971) J. Bacteriol. 108, 179-183.
6. St. John, T.P. and Davis, R.W. (1979) Cell 16, 443-452.
7. St. John, T.P. and Davis, R.W. (1981) J. Mol. Biol. 152, 285-315.
8. St. John, T.P., Scherer, S., McDonell, M.W. and Davis, R.W. (1981) J. Mol. Biol. 152, 317-334.
9. Douglas, H.C. and Pelroy, G. (1963) Biochem. Biophys. Acta 52, 155-156.
10. Douglas, H.C. and Hawthorne, D.C. (1966) Genetics 54, 911-916.
11. Perlman, D. and Hopper, J.E. (1979) Cell 16, 89-95.
12. Nogi, Y. and Fukasawa, T. (1980) Curr. Genet. 2, 115-120.
13. Matsumoto, K., Adachi, Y., Toh-e, A. and Oshima, Y. (1980) J. Bacteriol. 141, 508-527.
14. Matsumoto, K., Toh-e, A. and Oshima, Y. (1981) Mol. Cell. Biol. 1, 83-93.
15. Johnston, S.A. and Hopper, J.E. (1982) Proc. Natl. Acad. Sci. USA 79, 6971-6975.
16. Laughon, A. and Gesteland, R.F. (1982) Proc. Natl. Acad. Sci. USA 79, 6827-6831.
17. Hashimoto, H., Kikuchi, Y., Nogi, Y. and Fukasawa, T. (1983) Mol. Gen. Genet. in press.
18. Shimada, H., Nogi, Y., Matsuzaki, Y., Hashimoto, H. and Fukasawa, T. (1983) in press.
19. Clewell, D.B. and Helinski, D.R. (1969) Proc. Natl. Acad. Sci. USA 62, 1159-1166.
20. Maxam, A. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K. Eds., Vol 65, pp. 499-560, Academic Press, New York.
21. Smith, D.R. and Calvo, J.M. (1980) Nucl. Acids Res. 8, 2255-2274.
22. Rüther, U., Kaenen, M., Otto, K. and Müller-Hill, B. (1981) Nucl. Acids Res. 9, 4087-4098.
23. Sripati, C.E. and Warner, J.R. (1978) in Methods in Cell Biology, Prescott, E.M. Ed., Vol 20, pp. 61-81, Academic Press, New York.
24. Weaver, R.F. and Weissmann, C. (1979) Nucl. Acids Res. 7, 1175-1193.
25. Berk, A.J. and Sharp, P.A. (1978) Proc. Natl. Acad. Sci. USA 75, 1274-1278.
26. Casadaban, M.J., Chou, J. and Cohen, S.N. (1980) J. Bacteriol. 143, 971 -980.
27. Smith, H.O. and Birnstiel, M.L. (1976) Nucl. Acids Res. 3, 2387-2398.
28. Kozak, M. (1981) Nucl. Acids Res. 9, 5233-5252.
29. Stiles, J.I., Szostak, J.W., Young, A.T., Wu, R., Consaul, S. and Sherman, F. (1981) Cell 25, 277-284.
30. Blomback, B., Blomback, M., Edman, P. and Hessel, B. (1966) Biochim. Biophys. Acta 115, 371-396.
31. Chase, J.W. and Richardson, C.C. (1974) J. Biol. Chem. 249, 4545-4552.
32. Nasmyth, K.A., Tatchell, K., Hall, B.D., Astell, C. and Smith, M. (1981) Coldspring Harbor Symp. Quant. Biol. 45, 961-981.
33. Hinnebush, A.G. and Fink, G.R. (1983) J. Biol. Chem. 258, 5238-5247.
34. Donahue, T.F., Farabaugh, P.J. and Fink, G.R. (1982) Gene 18, 47-59.
35. Sollner-Webb, B. and Reeder, R.H. (1979) Cell 18, 485-499.
36. Hentschel, C., Irminger, J-C., Bucher, P. and Birnstiel, M.L. (1980) Nature 285, 147-151.
37. Wasylyk, B., Wasylyk, C., Augereau, P. and Chambon, P. (1983) Cell 32, 503-514.
38. Struhl, K. (1982) Proc. Natl. Acad. Sci. USA 79, 7385-7389.
39. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 7102-7106.
40. Konkel, D.A., Tilghman, S.M. and Leder, P. (1978) Cell 15, 1125-1132.
41. Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3018-3025.

42. Dobson, M.J., Tuite, M.F., Roberts, N.A., Kingsman, A.J., Kingsman, S.M., Perkins, R.E., Conroy, S.C., Dunbar, B. and Fothergill, L.A. (1982) Nucl. Acids Res. 10, 2625-2637.
43. Gallwitz, D., Perrin, F. and Seidel, R. (1981) Nucl. Acids Res. 9, 6339-6350.
44. Guarente, L., Yocum, R.R. and Gifford, P. (1982) Proc. Natl. Acad. Sci. USA 79, 7410-7414.
45. Tschumper, G. and Carbon, J. (1980) Gene 10, 157-166.
46. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
47. Zaret, K.S. and Sherman, F. (1982) Cell 28, 563-573.
48. Tsujimoto, Y. and Suzuki, Y. (1979) Cell 16, 425-436.
49. Green, M.R. and Roeder, R.G. (1980) Cell 22, 231-242.