Curr Genet (1990) 18:105–109

Current Genetics

A sequence that directs transcriptional initiation in yeast

A.M. Healy* and R.S. Zitomer

Department of Biological Sciences State University of New York at Albany, Albany, NY 12222, USA

Received April 9, 1990

Summary. While RNA polymerase II of the yeast *Saccharomyces cerevisiae* initiates transcription at discrete sites, these sites are located over a wide range of distances from the TATA box for different genes. This variability has led to a number of proposals for consensus sequences located at the initiation site which, in conjunction with the TATA box, would direct initiation. We tested this hypothesis via oligonucleotide-directed mutagenesis, by placing the sequence CAAG, a member of one of these consensus sequences, upstream of the coding sequence of the *CYC7* gene at a site at which initiation does not occur. The distance between the TATA sequence and this putative initiation site was varied by inserting it into the wild-type gene and three deletion mutants. The results demonstrated that this sequence can serve as an initiation site when located 49, 77, or 106 bp from the TATA sequence, but not when located 30 bp away.

Key words: *CYC7* gene – Cytochrome c – TATA box

Introduction

In yeast, as in all eukaryotes, the TATA box directs RNA polymerase II to the transcription initiation site. However, unlike the enzyme in other eukaryotes, which initiates transcription at a site 30 bp downstream from this element (reviewed in Breathnach and Chambon 1981), RNA polymerase II of *Saccharomyces cerevisiae* initiates transcription at discrete sites anywhere from 40 to 120 bp downstream from the TATA box (Chen and Struhl 1985; Hahn et al. 1985; Nagawa and Fink 1985; McNeil and Smith 1986; Healy et al. 1987; Rudolph and Hinnen 1987). Thus, while the TATA box appears to provide sufficient information for initiation in other systems, in yeast the conclusion seems inescapable that an additional sequence is required, one that signals the actual start of transcription. Whether this difference between initiation in yeast and higher eukaryotes represents a fundamental difference in the mechanism of initiation, or perhaps more likely a small difference in otherwise similar processes, a comparison of the two systems should provide some interesting insights into the evolution of the RNA polymerase II enzyme and its mode of action.

A number of different sequences have been proposed to serve as this postulated initiation site, based upon comparisons of a number of genes (Dobson et al. 1982; Burke et al. 1983; Hahn et al. 1985). In an earlier study (Healy et al. 1987), in which the distance between the TATA box and the coding sequence of the *CYC7* gene encoding iso-2-cytochrome c was varied, we found that there were eleven potential initiation sites in the region immediately upstream from the coding sequence, and all conformed to the initiation sequence proposed by Dobson et al. (1982) and Burke et al. (1983). We also reported that a survey revealed that 80% of 99 initiation sites in 32 genes contained this sequence. In the present study, to directly test the ability of this sequence to serve as an initiation site, we inserted it via oligonucleotide mutagenesis into a position immediately upstream from the coding sequence where no initiation occurs. Our results clearly demonstrated that this sequence, PyAAPu, directs transcriptional initiation in a manner dependent upon its distance from the TATA box and the presence or absence of other initiation sites.

Materials and methods

Strains and transformations. The *Saccharomyces cerevisiae* strain AH10 (*MATα, leu2-3, leu2-112, trp1-1, his4-519, ura3-52, cyc1-1, cyc7:LEU2, cyp1-16*, gal-) has been previously described (Healy et al. 1987). Plasmids were transformed into this strain by the frozen-cell method (Klebe et al. 1983). The respiratory competence of mutants was assayed on glycerol and lactic acid plates (Sherman et al. 1974).

* *Present address:* Department of Biology, The University of Michigan, Ann Arbor, MI 48109, USA

Offprint requests to: R.S. Zitomer

The *Escherichia coli* strain HB101 (Boyer and Roulland-Dussoix 1969) was used for most plasmid constructions and was transformed using the method of Hanahan (1983). The *E. coli* strains JM101 (Yanisch-Perron et al. 1985) and BW313 (Kunkel 1985) were used for the propagation of M13 vectors.

Plasmids. The centromeric *TRP1-ARS1* plasmid YCp*CYC7*(2)r, which contains a 2 kb fragment from the *CYC7* locus, has been described previously (Wright and Zitomer 1984). The plasmids X108, X80, and X61 are derivatives of YCp*CYC7*(2)r, which contain deletions extending from the *Xho*I site (–142) to the base pair indicated (–108, –80, or –61), have also been described previously (Healy et al. 1987), and are depicted in Fig. 1. An M13mp8 vector (Messing and Vieira et al. 1982) containing the *Bam*HI-*Kpn*I (–715 to +264) fragment from YCp*CYC7*(2)r was used to generate a sequence ladder in primer extension analyses.

Plasmid constructions. The *Xho*I-*Kpn*I fragments from the plasmids YCp*CYC7*(2)r, X108, X80, and X61 were gel-purified (Wright and Zitomer 1984) and ligated into the *Xho*I-*Kpn*I sites of the M13 vector um20 (a derivative of mp18 purchased from International Biotechnologies Inc, New Haven, Conn). These um20 derivatives were subjected to oligonucleotide mutagenesis by the method of Zoller and Smith (1982), using the oligodeoxyribonucleotide 5'-GCATTTATTCAAGCTTACTTTAAG-3'. After mutagenesis and transfection, potential mutants were screened by plaque hybridization (Benton and Davis 1977), using the same oligonucleotide as a probe (Wallace et al. 1981). The presence of the desired mutations was confirmed by sequence analyses (Sanger et al. 1977). The *Xho*I-*Kpn*I fragments from the derivatives of um20 containing the new mutations were purified and ligated into *Xho*I-*Kpn*I-digested YCp-*CYC7*(2)r, creating the plasmids YCp*CYC7*(2)r-i51, X108-i51, X80-i51, and X61-i51. Plasmid X80-i51m was created in a similar manner. Template from the um20 derivative containing the X80-i52 mutation was prepared in BW313, as described by Kunkel (1985), and subjected to mutagenesis with the oligodeoxyribonucleotide 5'-GCATTTATCCAAGCTTAC-3', as described in Biorad technical bulletin 1311 using two units of T4 DNA polymerase and 25 μg/ml of T4 gene 32 protein.

RNA extractions. Cells were grown on raffinose medium as described by Lowry et al. (1983). Total cellular RNA was prepared as described by Zitomer and Hall (1976). Poly(A)$^+$ RNA was selected by chromatography on oligo(dT)-cellulose (Maniatis et al. 1982).

Mapping of 5' ends of CYC7 mRNA. Primer extension analysis was used to map the 5' ends of *CYC7* mRNAs as described by McNeil and Smith (1986) and modified by Healy et al. (1987) using an oligonucleotide complementary to residues +17 to +36 of the mRNA.

Nomenclature. The base pairs of the plasmids are numbered with respect to the *CYC7* coding sequence. The A of the initiation codon is numbered 1; bases 3' are numbered in positive integers and bases 5' are numbered in negative integers.

Materials. All enzymes were purchased from Boehringer Mannheim Biochemicals (Indianapolis, Ind) and New England Biolabs, Inc (Beverly, Mass) with the exception of avian myeloblastosis virus reverse transcriptase which was purchased from Life Sciences Inc (St Petersburg, Fla). T4 Gene 32 protein was purchased from Biorad Laboratories (Richmond, Calif). Enzyme reactions were carried out in accordance with the instructions of the vendors. The oligonucleotides were purchased from Nadrian Seeman of this department.

Results

Previous analysis of deletion mutations in the upstream region of the *CYC7* gene led to the identification of the TATA box and potential initiation sequences, and to the characterization of the spatial relationship between them (Healy et al. 1987). The choice of start sites depended in part on the distance from the TATA box, and in part on the sequence at the initiation site. Although reducing the distance between the TATA box and the start site resulted in the displacement of initiation sites 3'-wards, the initiation sites were not displaced in exact proportion to the amount of DNA deleted, but were located at discrete sites. For the wild-type and mutant *CYC7* genes, a total of eleven start sites were found over a 90 bp region (Table 1). We also found that a consensus sequence, PyAAPu, previously proposed to play a role in initiation (Dobson et al. 1982; Burke et al. 1983), was present at the start sites of the *CYC7* gene (Fig. 1) and many other *S. cerevisiae* genes (Healy et al. 1987). Furthermore, we proposed that initiation sites within 45 bp of the TATA box could not be used. In the experiments described here, we tested these hypotheses by placing the consensus sequence, PyAAPu, in the upstream region of the *CYC7* gene at –51 by oligonucleotide-directed mutagenesis. This region was targeted because it represented the middle of a 20 bp region (–40 to –60) which contained no consensus sequence or initiation sites (Fig. 1) and, therefore, it would be easy to discern any new initiations at this site. The consensus sequence was created by replacing the

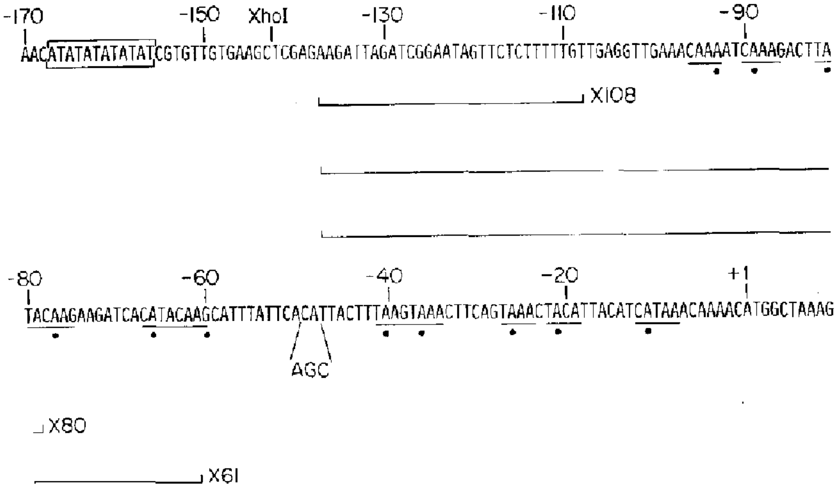

-170
-150
XhoI
-130
-110
-90
AACATATATATATATATCGTGTTGTGAAGCTCGAGAAGATTAGATCGGAATAGTTCTCTTTTTGTTGAGGTTGAAACAAAATCAAAGACTTA
X108
-80
-60
-40
-20
+1
TACAAGAAGATCACATACAAGCATTTATTCACATTACTTTAAGTAAACTTCAGTAAACTACATTACATCATAAACAAAACATGGCTAAAG
AGC
X80
X61

Fig. 1. The sequence of the coding strand for part of the upstream region of the CYC7 gene. The TATA sequence is *boxed.* Initiation sites previously determined (Healy et al. 1987 and Table 1) are indicated by a *filled circle* under the base, and the corresponding consensus sequence, discussed in the text, is *underlined.* Deletion used in this study are represented by a *line below the sequence.* The changes made via oligonucleotide-directed mutagenesis to construct the consensus sequence at –51 are also indicated

Table 1. Summary of plasmids used to construct putative initiation sites

Plasmid	Base pairs deleted	Initiation sites[a]	Distance between TATA and −51 (bp)[b]
YCpCYC7(2)r	–	−93, −89, −81	106
X108	−136 to −108	−89, −81, −77, −66, −60, −40, −36	77
X80	−136 to −80	−40, −36, −26, −21, −11	49
X61	−136 to −61	−40, −36, −26, −21, −11	30

[a] Initiation sites were determined previously (Healy et al. 1987). Major start sites are underlined

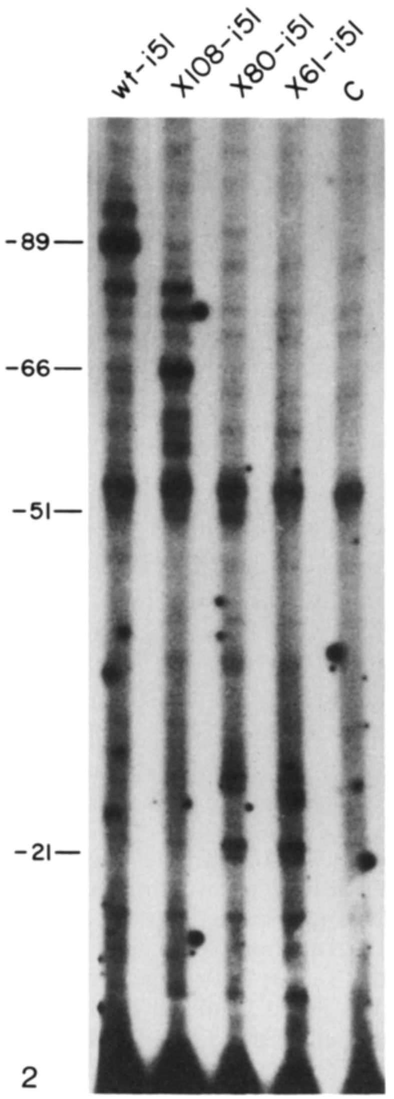
[b] Distances were measured from the 3′ most base pair of the TATA box

2 bp sequence CA at −49 and −48 with the 3 bp sequence AGC (Fig. 1), resulting in a net gain of one base pair and fortuitously creating a *Hin*dIII site (5′-AAGCTT-3′) which aided in the screening for mutants. This putative start site at −51 could be located at varying distances from the TATA sequence by mutagenizing the wild-type gene and the three deletion mutants, shown in Fig. 1, individually. The distance between the TATA sequence and the newly introduced consensus initiation site is given in Table 1 for each construction.

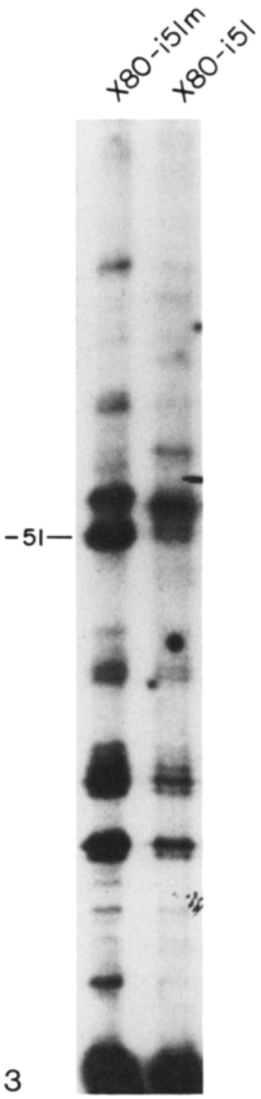
The plasmids containing this putative initiation site [YCp*CYC7*(2)r-i51, X108-i51, X80-i51, and X61-i51] were transformed into the *S. cerevisiae* strain AH10 which contains deletions of both cytochrome c genes, *CYC1* and *CYC7*, and is dependent on the plasmid-borne *CYC7* gene for growth on non-fermentable energy sources. As previously described, no changes in the level of *CYC7* expression were evident in transformants carrying deletions extending into the initiation region (Healy et al. 1987). Similarly, mutants carrying the new sequence exhibited wild-type levels of *CYC7* expression when streaked on glycerol and lactic acid plates, an assay sensitive to changes in levels of cytochrome c. Also, Northern blot analysis demonstrated that wild-type levels of *CYC7* mRNA were present in the mutants (data not shown).

Transcriptional initiation sites were determined by mapping the 5′ end of the *CYC7* gene in a primer extension assay. A synthetic primer complementary to *CYC7* mRNA was annealed to and extended, using reverse transcriptase to the 5′ end of the message. The 5′ ends of the mRNAs were determined by comparing the length of the resulting cDNAs with a sequencing ladder generated with the same primer annealed to the gene.

As shown in Fig. 2, the new consensus sequence at −51 was capable of serving as a site for the start of transcription in three of the constructions. In the wild-type gene [YCp*CYC7*(2)r-i51] a single start site was evident at −51, 106 bp from the TATA sequence. Two start sites at −51 and −50 were used in X108-i51 and X80-i51, which were located 77 and 49 bp from the TATA sequence, respectively. Finally, no initiation of transcription was evident at this site in X61-i51, which was located 30 bp from the TATA sequence. (There is an additional band visible in this region which did not arise from extension of the primer hybridized to *CYC7* transcripts; it can also be seen in the control sample which was generated with RNA from untransformed cells containing a deletion of the *CYC7* gene.)



2



3

Fig. 2. Primer analysis of CYC7 initiation sites from cells transformed with plasmids containing the putative initiation site at −51. Poly(A)$^+$ RNA was prepared from AH10 cells carrying the plasmid indicated above the lane. The sample in *lane C* was derived from untransformed cells and served as a control. The location of some of the initiation sites, including that at −51 are indicated and were determined by using a sequence ladder for size markers (data not shown)

Fig. 3. Primer extension analysis of CYC7 initiation sites from cells transformed with plasmids containing the putative initiation site at −51. Poly(A)$^+$ RNA was prepared from AH10 cells carrying the plasmid indicated above the lane. The location of the initiation site at −51 is also indicated

The newly introduced initiation site in these mutants did overlap with a different proposed start site, TCPuA (Hahn et al. 1985). To determine whether the introduc-

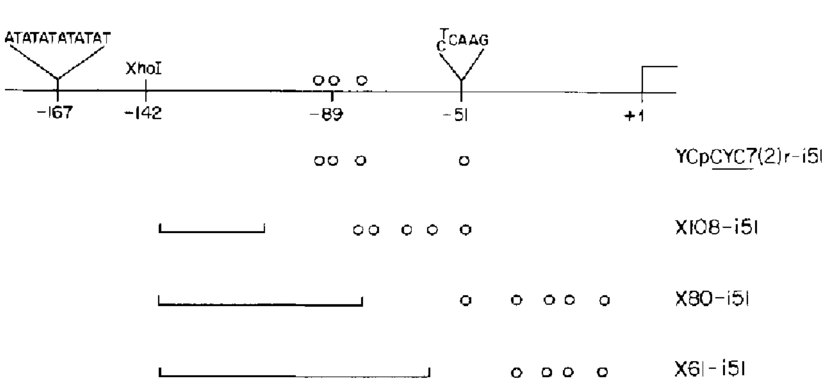


Fig. 4. Summary of plasmid constructions and their corresponding initiation sites. The wild-type gene is represented on top and the *lines below* represent the region deleted in the plasmid indicated. Initiation sites are represented by *open circles*

tion of the CAAG was the determinant factor or whether the T 5′ to the C played an essential role, we constructed a derivative of X80-i51 (X80-i51m) in which the T at –53 was replaced by a C. As can be seen in Fig. 3, the single base pair change did not affect the use of this sequence as an initiation site.

Discussion

The results presented here demonstrate unequivocally that the sequence CAAG directs initiation of transcription by yeast RNA polymerase II when placed an appropriate distance downstream from a TATA box. The necessity for such a sequence was noted by a number of workers who found that, unlike the situation in higher eukaryotes where the distance between the TATA box and the initiation site is fixed at about 30 bp, in yeasts the spacing between these two sites is both larger and variable, ranging over 40 to 120 bp (Dobson et al. 1982; Hahn et al. 1985; Nagawa and Fink 1985; McNeil and Smith 1986; Healy et al. 1987). The sequence demonstrated to function in conjunction with the TATA box in yeast is a representative of the consensus sequence PyAAPu proposed by Dobson et al. (1982) and Burke et al. (1983), based upon a survey of the initiation sites known at the time. In a more recent study (Healy et al. 1987), we found that this sequence was present at the eleven start sites identified for the wild-type and mutant *CYC7* genes and at 80 of 99 initiation sites in 32 class II genes with zero or one mismatch, with a strong preference for initiation at the second position, which agrees with the results of this present study. The pervasive presence of this DNA sequence at the start sites of numerous genes strongly implicates it as the major transcriptional initiation site in yeast.

We set out to test two features of transcription initiation in yeast; the spatial relationship between the TATA sequence and the initiation site and the sequence specificity at the start site. Our results (summarized in Fig. 4) demonstrated that CAAG can serve as an initiation site and confirmed the minimum distance requirements previously proposed. As we anticipated, initiation occurred at the new site when it was placed 49 bp from the TATA box in the deletion plasmid X80-i51. In this case, this site was the first such site greater than 45 bp from the TATA box. Also, we predicted that placing an initiation site 30 bp from the TATA sequence, as in X61-i51, would be too close, and such was the case. On the other hand, we were surprised to find that initiation occurred at the –51 site when located 77 or 106 bp from the TATA sequence, as in X108-i51 and YCp*CYC7*(2)r-i51, even though there were major start sites located upstream. In the wild-type construction this is the only initiation site downstream of the major sites which is used, suggesting it is a strong site. The use of this site in YCp*YCY7*(2)r-i51 implies that RNA polymerase II can identify strong start sites located downstream, which is not completely consistent with a model for start site selection where RNA polymerase scans 5′ to 3′ in search of preferred initiation sites.

Other consensus sequences have been suggested for the initiation site. In an analysis of 18 class II promoters, two sequences, PuPuPyPuPu and TCPuA, were found to account for 55% of the initiation sites (Hahn et al. 1985). We evaluated these sequences, as well as PyAAPu, with respect to the initiation sites of 32 class II genes (Healy et al. 1987). PuPuPyPuPu was present around 80% of the initiation sites with zero or one mismatch but without any specificity for the residue at which initiation occurred. This appearance was somewhat higher than chance, but could be due to its overlap with the PyAAPu sequence. The sequence PuPuPyPuPu was not present at the start site we created. The other proposed consensus sequence TCPuA has been found at start sites in a few (17%) cases. This consensus also overlaps the sequence PyAAPu which may explain its appearance at start sites. This sequence was present at the start site we constructed. To differentiate between the two sequences present, we altered the TCAAG sequence to CCAAG. This change had no effect on initiation at this site, indicating that CAAG was the determinative sequence.

In summary, we have demonstrated that the consensus, PyAAPu, for the initiation site, is capable of signaling the start of transcription when located an appropriate distance from the TATA sequence. For many yeast genes, including *CYC7*, there are multiple initiation sites which may vary in strength. The characteristics of a strong site are not obvious through inspection of the sequence and its location with respect to the TATA sequence. Also, a few start sites do not fit the consensus sequence demonstrated here. These observations imply that there is additional information involved in the selection of a start site. This information may be in the form of other specific sequences not yet uncovered, may involve the conformation of DNA, or may rely on the accessibility of certain sites in the initiation region to the bound RNA polymerase molecule.

Acknowledgements. This study was supported by a grant from the National Institutes of Health.

References

Benton WD, Davis RW (1977) Science 196:180–182

Boyer HW, Ruolland-Dussoix D (1969) J Mol Biol 4:459–472

Breathnach R, Chambon P (1981) Annu Rev Biochem 50:349–383

Burke RL, Tekamp-Olson P, Najarian R (1983) J Biol Chem 258:2193 2201

Chen W, Struhl K (1985) EMBO J 4:3273–3280

Dobson MJ, Tuite MF, Roberts NA, Kingsman AJ, Kingsman SM (1982) Nucleic Acids Res 10:2625–2637

Hahn S, Hoar ET, Guarente L (1985) Proc Natl Acad Sci USA 82:8562–8566

Hanahan II (1983) J Mol Biol 166:557 580

Healy AM, Helser TL, Zitomer RS (1987) Mol Cell Biol 7:3785–3791

Klebe RJ, Harriss JV, Sharp ZD, Douglas MG (1983) Gene 25:333–341

Kunkel TA (1985) Proc Natl Acad Sci USA 82:488–492

Lowry CV, Weiss JL, Walthall DA, Zitomer RS (1983) Proc Natl Acad Sci USA 80:151–155

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Messing J, Vieira J (1982) Gene 19:269–276

McNeil JB, Smith M (1986) J Mol Biol 187:363–378

Nagawa F, Fink G (1985) Proc Natl Acad Sci USA 82:8557–8561

Rudolph H, Hinnen A (1987) Proc Natl Acad Sci USA 84:1340–1344

Sanger F, Nicklen S, Coulson SA (1977) Proc Natl Acad Sci USA 74:5463–5467

Sherman F, Stewart JW, Jackson M, Gilmore RA, Parker JH (1974) Genetics 77:255–284

Wallace RB, Johnson MJ, Hirose T, Miyake T, Kawashima EH, Itakura K (1981) Nucleic Acids Res 9:879–894

Wright CF, Zitomer RS (1984) Mol Cell Biol 4:2023–2030

Yanisch-Perron C, Vieira J, Messing J (1985) Gene 33:103 119

Zitomer RS, Hall BD (1976) J Biol Chem 251:6320–6326

Zoller MJ, Smith M (1982) Nucleic Acids Res 10:6487–6500

Communicated by C. P. Hollenberg