

Each of three “TATA elements” specifies a subset of the transcription initiation sites at the *CYC-1* promoter of *Saccharomyces cerevisiae*

(yeast promoters/initiator element/iso-1-cytochrome *c* gene)

STEVEN HAHN, EILEEN T. HOAR, AND LEONARD GUARENTE

Department of Biology, 56-409, Massachusetts Institute of Technology, Cambridge, MA 02139

Communicated by S. E. Luria, August 15, 1985

ABSTRACT Transcription initiation of the yeast iso-1-cytochrome *c* gene (*CYC-1*) occurs in six major clusters at positions +1, +10, +16, +25, +34, and +43. Potential “TATA elements” lie upstream at positions –154, –106, –52, and –22. Analysis of the TATA region suggests that three of these TATA sequences are functional and contribute to initiation at *CYC-1*, with the –106 TATA promoting initiation at +1, +10, and +16; the –52 TATA, at +16, +25, +34, and +43; and the –22 TATA, at +34 and +43. Deletions changing the spacing between the TATA sequences and the region of transcription initiation do not change the location of the *CYC-1* transcription start points. This finding suggests that at least part of the information determining mRNA initiation sites is encoded within the DNA sequence at the site of transcription initiation. Analysis of 18 yeast RNA polymerase II promoters suggests that two classes of DNA sequences serve as preferred sites for transcription initiation. To test this possibility, we replaced some of the normal *CYC-1* start sites with one of these sequences, TCGA, and found that transcription initiates at this newly introduced sequence. These results are in contrast to those from higher eukaryotes, where RNA polymerase II typically initiates transcription a fixed distance downstream from the TATA element. The presence of multiple, functional TATA sequences at *CYC-1* is inconsistent with the idea that RNA polymerase or another transcription factor attaches to the template at an upstream activation site and scans for the nearest TATA element.

In higher eukaryotes, promoters transcribed by RNA polymerase II can be separated into several distinct elements that are required for maximal transcription. The first element, the “TATA” sequence, is typically located about 30 nucleotides upstream from the site of transcription initiation and is thought to play the major role in determining the location of the transcription start site (1–5). Deletion of the TATA element often leads to multiple sites of transcription initiation (1, 3, 6). In several cases, initiation at a position 30 nucleotides downstream from the TATA element is unaffected by alteration of the particular sequence at the transcription start site (1, 2, 4). A second class of elements, termed enhancers, can function up to kilobases upstream or downstream of the TATA element (6–9).

In yeast, promoters transcribed by RNA polymerase II have been separated into at least two functional elements, upstream activation sites (UAS) and TATA elements (10–18). The yeast TATA element, along with sequences at the start of transcription initiation, contains all the information necessary to specify the transcription start site. Joining either the yeast *CYC-1* or *HIS3* TATA element and initiation region to

heterologous UAS elements produces transcripts with 5' ends identical to those of wild-type *CYC-1* or *HIS3* (12–19).

Although TATA elements are required for promoters of both yeast and higher eukaryotes, several differences seem apparent. First, promoters of higher eukaryotes often display a single site of transcription initiation, whereas yeast promoters frequently have multiple initiation sites, which can be spread out over a region of up to 100 nucleotides (10). Second, unlike those of higher cells, yeast promoters often contain multiple TATA sequence motifs as far as 100 nucleotides upstream from the transcription start site, although the functional yeast TATA elements have been identified in only a few cases (11, 13, 20, 21).

A useful system in which to study initiation is the iso-1-cytochrome *c* gene (*CYC-1*), because it displays six major clusters of initiation sites spread over 50 base pairs and four potential TATA elements that lie between –154 and +1* (13, 22). The positions of the *CYC-1* transcription initiation sites do not change under either repressed or derepressed conditions (13). To clarify the relationship between the TATA sequences and mRNA start sites for *CYC-1*, and perhaps for yeast promoters in general, we have analyzed the effects of deletion, substitution, and insertion mutations in the TATA and mRNA initiation regions on the efficiency and position of *CYC-1* transcription initiation. We have deduced the specific relationship between the *CYC-1* TATA elements and initiation sites and several general rules pertaining to initiation by RNA polymerase II in yeast.

MATERIALS AND METHODS

Assays. β -Galactosidase assays were performed on strain BWG1-7a (*MATa*, *leu2-3*, *leu2-112*, *his4-519*, *ade1-100*, *ura3-52*) (14) as described (23).

TATA Deletion Mutations. Deletion mutants Δ S15, Δ M8, and Δ L8 were derived from the pLG Δ 312 derivative Δ -178-82 containing the yeast 2- μ m (“2-micron”) origin of DNA replication (13). This plasmid was opened with *Xho* I and digested with BAL-31 as described (13). The resulting linear DNA was cut with *Sac* I, at a unique site within the *lacZ* gene. The BAL-31–*Sac* I fragment was isolated and ligated into the vector pLG Δ 312s (13) cut with *Sal* I and *Sac* I. The resulting plasmids were identical to plasmid pLG Δ 312 except that they lacked the DNA between position –178 and the deletion endpoints (Fig. 1). Deletion mutant Δ H3 was derived from a fragment extending from the *Hae* III site at –138 to the *Bam*HI site and was ligated to a backbone derived from pLG Δ 312, extending from a filled-in *Xho* I site to the *Bam*HI site.

Insertion Mutations Between the TATA and Initiation Region. I-27: An *Xma* I–*Bcl* I fragment extending from –312 to

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: UAS, upstream activation site(s).

*We define +1 as the most upstream site of transcription initiation.

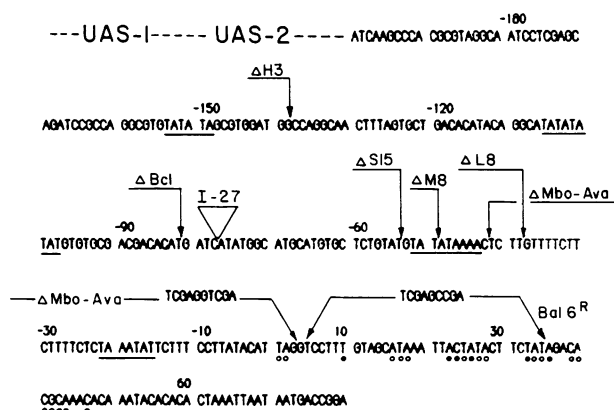


FIG. 1. Deletions and rearrangements in the *CYC-1* TATA and initiation region. Shown is the DNA sequence upstream of the *CYC-1* coding region and the deletions and rearrangements used in this work. Position +1 is defined as the most upstream start of transcription initiation (13). UAS-1 and UAS-2 lie at positions -265 and -229, respectively (14). Deletions Δ H3, Δ Bcl, Δ S15, Δ M8, and Δ L8 extend from position -178 through the respective endpoints indicated. Deletions Δ Mbo-Ava and Δ Bal6^R combine both deletions and insertions of new sequence as indicated. Δ Mbo-Ava-Bcl consists of both the Δ Mbo-Ava and Δ Bcl deletions in *cis*. I-27 is an insertion of the sequence CGCGTAGGCAATCCTCGAGTCGAGATC at position -77. The positions of wild-type *CYC-1* transcription start points are shown below the DNA sequence for both strong (●) and weak (○) initiation sites. TATA homologies are underlined.

-80 was joined to an *Mlu* I-*Bam*HI fragment extending from -190 through +80, from which the sequence between -178 and -80 had been deleted. The two fragments were ligated into an *Xma* I-*Bam*HI backbone derived from pLGΔ312. Bal6^R: An *Xho* I-*Sac* I fragment extending from +35 through part of *lacZ* and an *Xma* I-*Ava* II fragment extending from -312 to +3 were joined to an *Xma* I-*Sac* I backbone derived from pLGΔ312. Δ Mbo-Ava: An *Xho* I-*Mbo* II fragment extending from -178 to -41 and an *Xho* I (filled-in)-*Sac* I fragment extending from +3 through part of *lacZ* was ligated to an *Xho* I-*Sac* I backbone derived from pLGΔ312. Δ Mbo-Ava-Bcl: A *Bcl* I (filled-in)-*Bam*HI fragment of Δ Mbo-Ava extending from -42 to +80 was inserted into an *Xho* I (filled-in)-*Bam*HI backbone derived from pLGΔ312.

Mapping of 5' Ends of *CYC-1* mRNA. Total yeast RNA was isolated as described (24). The synthetic oligonucleotide primer used is homologous to a region of *lacI* sequence in plasmid Δ 312 located about 100 nucleotides 3' of the most downstream start of *CYC-1* transcription initiation. ³²P-labeled primer (5 ng) was incubated with total yeast RNA in 5 mM Tris Cl, pH 8/40 mM KCl/0.5 mM EDTA in a volume of 15 μ l at 48°C for 1 hr. The reaction mixture was adjusted to 100 μ M dATP, dGTP, dCTP, and dTTP; 10 mM MgCl₂; and 1 mM dithiothreitol. Actinomycin D (0.05 μ g/ml) and 8 units of reverse transcriptase (Life Sciences, St. Petersburg, FL) were added to obtain a final volume of 20 μ l, and the mixture was incubated 60 min at 37°C. The products of the reaction were analyzed in 8% acrylamide sequencing gels.

RESULTS

Effects of TATA Deletion Mutations on Expression of a *CYC-1-lacZ* Fused Gene. Previous work has shown that wild-type levels of *CYC-1* transcription depend upon sequences in the region between -60 and +3 (13), which contains two TATA-like sequences at positions -52 and -22 (numbering denotes the position of the first T in the TATA sequence). Two additional TATA motifs lie at -154 and -106. To assess the relative contributions of each of these TATA sequences in determining the levels and the position

of initiation of *CYC-1* mRNA, deletion mutants were constructed and their effects were measured by using a *CYC-1-lacZ* fusion gene borne on a plasmid with the yeast 2- μ m origin of DNA replication.

Deletion Δ H3, which extends from position -178 to -138, removing the -154 TATA sequence (Fig. 1), increases the level of *CYC-1* expression, in both glucose and derepressing lactate media, compared to the expression from the wild-type parent, Δ 312 (Table 1). This effect has been noted previously with a deletion ending at position -60 (13). Deletion Δ S15, which extends from position -178 to -54, removing both the -154 and -106 TATA sequences, likewise gives elevated levels of *CYC-1* expression. Deletion Δ M8, ending at position -49 and removing the TAT of the TATATAAAA sequence at -52, decreases *CYC-1* expression by a factor of about 3, relative to wild-type levels. This drop in expression is not due to an improper spacing between the UASs and the initiation region, since filling in the *Xho* I site of Δ M8 with DNA polymerase generates a plasmid, Δ M8s, which has spacing identical to that of deletion mutant Δ S15 but displays the same level of *CYC-1* expression as Δ M8 (Table 1). Deletion Δ L8, ending at position -38 and removing the entire -52 TATA motif, gives levels of *CYC-1* expression about one-sixth those of wild type. The effects of the Δ M8 and Δ L8 deletions are observed in both glucose and lactate media (Table 1). Finally, Δ Xho-Ava, which deletes sequences between -178 and +3, removing all four TATA motifs, reduces expression \geq 94% (13).

An internal deletion mutant, Δ Mbo-Ava, was also constructed, from which only the -22 TATA sequence had been removed. This deletion extends from position -41 to +3 and adds 8 nucleotides of linker sequences, thus changing the spacing between the normal start sites of *CYC-1* transcription and the major TATA sequence at -52 by 35 base pairs. Table 1 shows that Δ Mbo-Ava has little effect on the overall level of *CYC-1* expression in either glucose or lactate medium.

Deletions and Insertions in the TATA Region. RNA was extracted from cells grown in lactate medium and containing the various altered *CYC-1* plasmids, and the 5' ends of mRNA were mapped by extension of a primer that hybridizes to *lacI* and is thus specific for the *CYC-1* fusion message. Fig. 2 shows that mRNA from the wild-type *CYC-1* promoter (Δ 312) has very heterogeneous 5' ends, with transcription initiation in six major clusters: +1, +10, +16, +25, +34, and +43 (13, 22). Deletion Δ H3, which removes only the -154 TATA, gives a pattern of 5' ends identical to that of the wild-type

Table 1. TATA deletions and expression of the *CYC-1-lacZ* fusion gene

Plasmid	β -Galactosidase, units	
	Glucose medium	Lactate medium
Δ 312	49	420
Δ H3	112	570
Δ S15	180	600
Δ M8	13	170
Δ M8s	13	—
Δ L8	8	80
Δ Xho-Ava	3	6
Δ Mbo-Ava	34	580
Δ Mbo-Ava-Bcl	27	260
I-27	170	510
Bal6 ^R	29	490

Strain BWG1-7a, bearing the indicated plasmid, was grown in minimal medium containing either 2% (wt/vol) glucose or 2% lactate. β -Galactosidase assays were done in duplicate (individual values deviated from the average by <20%).

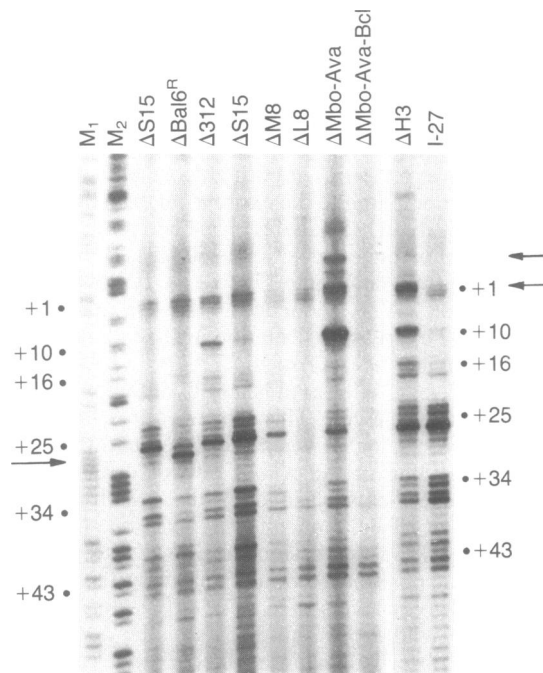


FIG. 2. Mapping of *CYC-1* transcription start sites. RNA was extracted from lactate-grown BWG1-7a containing the indicated plasmid, and the 5' ends of the *CYC-1* fusion RNAs were mapped by primer-extension. The numbers indicate position relative to the wild-type +1 transcription start site. Arrows indicate initiation events at the TCGA oligonucleotide insertions in the ΔBal6^R and $\Delta\text{Mbo-Ava}$ constructs. Reaction mixtures contained either 20 μg (ΔS12 , $\Delta\text{Mbo-Ava-Bcl}$, ΔH3 , I-27), 7.5 μg [ΔS15 (third lane)], 15 μg [ΔS15 (sixth lane), $\Delta\text{Mbo-Ava}$, ΔBal6^R], 40 μg (ΔM8), or 60 μg (ΔL8) of total RNA. Lanes M_1 and M_2 contained size standards.

promoter. However, ΔS15 , from which both the -154 and -106 TATA sequences have been removed, shows a severe reduction in initiation from the +10 site and a partial reduction in initiation from +16. This suggests that the -106 TATA sequence contributes to initiation at the +10 and +16 regions. The effects of ΔS15 on the +1 initiation site are sometimes obscured by background products of the primer-extension reaction that comigrate with the product of extension of mRNA initiating at +1. That the TATA element at -106 actually contributes to the +1 initiation site will be demonstrated below.

Deletion mutant ΔM8 , from which the TATA sequence at -52 has been partially removed, shows levels of RNA one-third to one-fifth those observed with wild-type *CYC-1*. Transcription, normally initiating at positions +10 and +16, are undetectable with this deletion. Deletion ΔL8 , which removes the entire -52 TATA sequence, shows a further reduction in *CYC-1* mRNA levels. Now the starts at +10 and +16 as well as those at +25 are absent, and the starts at +34 and +43, although reduced about 85% and 75%, respectively, are still clearly apparent. A deletion of all the TATA elements ($\Delta\text{Xho-Ava}$) and ending at +3 abolished all detectable *CYC-1* message (data not shown). These results suggest that the -52 TATA is required for initiation at +16 and +25 and contributes to initiation at both +34 and +43. The starts at +34 and +43 apparently are also promoted by the TATA element at -22 . The sequence TATA found at -7 is unlikely to contribute to initiation at +34 and +43, since the sequence TATA is shown to be insufficient to function as a TATA element in yeast in an accompanying paper (20). In summary, it would appear that the -106 TATA element specifies the start at +10 and contributes to the start at +16; the -52 TATA specifies starts at +16, +25, +34, and +43; and the -22 TATA element contributes to the +34 and +43 starts.

A possible complication to the above conclusions is the possibility that the -52 and -22 TATA elements only become functional when the upstream TATA element has been deleted. According to this view, the wild-type *CYC-1* promoter would contain but one functional TATA element, at -106 . To rule out this complication, we inserted a sequence, between the TATA elements at -106 and -52 , that displaced the -106 element 27 base pairs upstream. This insertion, I-27, greatly reduces the extent of initiation at +1, +10, and +16. Initiation at +25, +34, and +43, however, is not affected by the insertion. Thus, we conclude that in the wild-type promoter, the TATA element at -106 specifies starts at +1, +10, and +16, whereas the remaining starts are specified by the TATA elements at -52 and -22 , as described above.

Deletion Between the TATA Elements and Initiation Sites. As shown in Table 1, the internal deletion $\Delta\text{Mbo-Ava}$ removes the -22 TATA sequence but has no effect on the total levels of *CYC-1* expression. However, Fig. 2 shows that this deletion has a surprising effect on the pattern of *CYC-1* transcription. The start sites at +16, +25, +34, and +43, although their spacing from the TATA elements has been altered, are still utilized. The +10 start is actually enhanced by the deletion, and two new initiation sites appear, upstream from position +1, that will be discussed below.

We wished to test whether any of the above start sites were brought under the influence of the -106 TATA element, which was moved 35 base pairs closer to the initiation region by the $\Delta\text{Mbo-Ava}$ deletion. Thus, a deletion of DNA between -178 and -81 was constructed in *cis* to $\Delta\text{Mbo-Ava}$ ($\Delta\text{Mbo-Ava-Bcl}$). This deletion of sequences containing the -106 TATA abolishes the starts at +10, +16, +25, and +34 and reduces transcription initiating at +43. Thus, we conclude that the $\Delta\text{Mbo-Ava}$ deletion brought these start sites within range of the -106 TATA. The start at +43 apparently also receives a contribution from the -52 TATA element in the $\Delta\text{Mbo-Ava}$ construct, because initiation at +43 is not abolished by deletion of the -106 TATA element.

The experiments in the above sections indicate that the TATA element operates over a "window" between 60 and 120 nucleotides downstream. The lower limit of about 60 nucleotides derives from several consistent observations. First, in the wild-type *CYC-1* promoter, the -52 TATA element does not activate the start at +10 but does activate the cluster of starts at +16. Second, the -22 TATA element does not activate starts at +25 but does partially activate starts at +35 and +43. Third, in the $\Delta\text{Mbo-Ava}$ construction, the -52 TATA element (now moved 35 nucleotides closer to the initiation sites) does not activate starts at +35 but does activate starts at +43. The assignment of the upper limit of 120 nucleotides also rests on several consistent observations. First, in the wild-type *CYC-1* promoter, the -106 TATA element activates starts at +16 but does not activate the +25 cluster. Second, the 27-base-pair insertion I-27 places the -106 TATA element out of range of the +1 start. Third, in the $\Delta\text{Mbo-Ava}$ construction, the -106 TATA element activates starts at +43 (114 nucleotides from the element in this construction).

Consensus Sequences at Yeast Transcription Initiation Sites. If yeast TATA elements in general activate transcription over a large window, then the precise start points of transcription may be encoded within the window, close to the initiation region itself. Thus, we searched for possible consensus sequences at or close to yeast initiation sites by comparing 18 yeast promoters utilized by RNA polymerase II (Fig. 3). Our analysis revealed two different consensus sequences accounting for about 55% of all transcription initiation sites. One class, accounting for 17% of all transcription starts, consists of the sequence TC(G/A)A, with transcription usually starting at either the C and/or the adjacent nucleo-

TY16 TGAGATATAATGTGGTAATTAATATATATTTGTGGATTCATTGTTGATAAAGGCTATAATATTAGATATACAGAAATAT
ACTAGAAATTCCTCCTCAAGATATAAGAAATCCT

RP51A CATGCTGTAGCTATGCTGTGATCCTTTACTGTCATATTAGCTTAATAATATATGAAATTAAGTATTAAATCAAGCATT
ATCGACTTAATTCTAAGAAAAGTCAAGATCTCGAGACTAGCAATAAC

HIS4 TATCAGTCATTCATATAGAAAGTAAGAAAAGATATGACTATGAACAGTAGTACTGTATATAAATAGATATGGA
ACGTTATATTACCTCCGATGTGTGTTGTACATACATAAAAATATCATAGCACAACTGCGCTGTG

CYC-1 CATATATAATATGTGTGACGACACATGATCATATGTCATGCTGTATGATATAAAAATCCTTTTCTTCTT
TTCTCFAAATAATCTTTCTTATACATTAAGTCTTTGTAGCATAAATTACTATCTCTATAGACACGCAAAACACA

GAL1 CAGCGAAGCGATGATTTTGTATTAACAGATATATAAATATCAAAAATGTCATAACCCTTTAACFAATACCTTTCAA
CATTTTCGTTTTGTATTACTTTTCAATATGATAAAAAGTATCAACAAAAAATTGTA

GAL10 TAAGATAATGGGCTCTTTACATTTCCACAACATATAATAGATTAGATATGATATGATATATGATATGATATGATATGAT
GGTAATGCCATGFAAATAATGATTATTAAGCTTTGCGTCCATCCAAAAAAGTAAGAAATTTTGAATAATTC

HIS3 ATGACTCTTTTTTTTCTTAGCGATTGTCATTATCACATAATTAATTATACATTTATATAAATAAATGATGTGATTTCTTCGA
AGAAATACTAAAAAATGAGCAAGCAAGATAAACGAAAGCAAGATAAACGAAAGCAAAA

SUC2 ATCCATATAATCCTTCCTCTGAAAAGAAACAATATAAATAATATATATTCTTCAAAAACCTTCCTGTCTGTCTGCT
(1.9kb) ITTTTTTACCAATATAATTTACTTTTTTTTCTCTCAGAGAAACAGCAAAAACAAAAGCT

SUC2 TCAGAGAAACAAGCAAAAACAAAAGCTTTTCTTTCACTAACATATAATGATGCTTTTCAAGCTTTCTTTCTTTTCTTGG
(1.8kb) CTGTTTTTCAAGCCAAAATATCTGCATCAATGACA

LEU2 ACTTGAAGTTGACAATATTATTTAAAGACCTATTGTTTTTCCAATAGTGGTTAGCAATGCTCTTACTTTCTAATCTT
TCTTACCTTTTACATTTCAAGCAATATATATATATTTCAAGGATATA

ADR2 TCAATATAAATATAGATGCCAGTAGCGACTTTTTTACACTCGAAATACTTACTACTGCTCTTGTGTTTTTA
TCACITCTTTGTTCTTCTTGTAAATAGAAATATCAAGCTACAAAAA

HIS1 GGAACATATAAATAATTTTTCAGAAATATACTGCTGAAAAGTCCATTTATACGACTGAACTCAAGATATTTACGGTC
TAAACTTTGGCATTTTACTCTCTTAAAGCTTTGAACTTTGAAAATAGAAAAGAAAGATAAGTTCTAAAAAA
TTCAATAGAAAA

GAL4 GTTTTATAGCTATTTTTTCCGTACTCTTCCCAATTTTCAAGCTTCACTCCAGATTGTGTCTACGTAATGCACGCCAT
CATTTTAAGAGAGAGCAAGAAAGCAAGCCCTCTGAAAAGATG

CAR1 TAAAGATATAATGACAGCTTTCGCCCCACTTTCTATCAAGATCAAGACTGTTCTCTCTTGTCTGTATATATT
TTCTCAAAATGATGACAGAAACAACAACAACAAC

MATb1 TTGATTTTTGTCTTTGCGGAAACTGTATAAAAATCTTCCAAAAAAGAAAAATAAACAATACATCTCCTTATATC
AAAGAAAATCAAGAAAGGACAAATCATGATGATATTG

MATg2 GCAATAAAATGATCCCAACAACAAACCCAGACATCTTAATGTTTTGAAACATAAAATGTTGAACTTAAAAATAT
ATATAAAGGAGAGATGAAATCAGCTTGAAGTGGCAAGAAAAAAGAAAGATAAGCAAGAAAAAATGAA

MATg1 ACATCATATGAAACAACCTTAAACTCTTAACTACTCTTTTAACTTCACITTTTATGAAATGATCAACCATATA
TAATAACTTAATAGACGACATTAC

MATg2 AAAGAACAATAATCAAGCAATAAATAAGCTTTTCTATTTGCTGTAAATTTACCTAAGTTACCAAGAGATGATAACAACAG
AAGAAGAGTTGAATTAAGGATATATTAGAA

FIG. 3. TATA and initiation-region sequences for various yeast promoters. Published transcription initiation sites are indicated above the sequence (●) or by an overbar where heterogeneous and/or defined to within several adjacent nucleotides. Potential TATA elements are boxed. mRNA initiation sites falling into the RRYRR class are underlined and those in the TC(G/A)A class are indicated by zig-zag underlining. Data were taken from the following sources: *Ty18* (25, 26); *RP51A* (24); *HIS4* (15); *GAL1,10* (28); *HIS3* (18, 29); *SUC2* (30); *LEU2* (31); *ADR2* (32); *HIS1* (33); *GAL4* (34); *CAR1* (35); *MAT* genes (36). kb, Kilobases.

tides. The second class of sequences consists of a pyrimidine surrounded by at least two purines on either side, RRYRR. Examination of the sequences in Fig. 3 shows that the TC(G/A)A sequence is nearly always used when present more than 50 nucleotides downstream of a TATA sequence. In the rare exceptions to this rule [*SUC2* (1.9 kb), *CAR1*, and *GAL1*], the functional TATA elements have yet to be defined. The *CYC-1* promoter, which has one of the most heterogeneous patterns of transcription starts, contains neither of the two consensus sequences between the TATA regions and the most distal site of transcription initiation.

Functionality of the Sequence TCGA as a Transcription

Initiation Site. To test whether the above homologies were meaningful, we constructed an altered *CYC-1* promoter in which the DNA between +6 and 35 was deleted, removing the normal transcription start sites at +10, +16, +25, and +34, and the sequence TCGAGCCGA was placed at the deletion junction. If the sequence TCGA encoded a transcription start site, then we would expect to see a strong start at this new DNA sequence. This construction, Bal6^R, expresses nearly wild-type levels of β -galactosidase in both glucose and lactate media (Table 1). A transcription-mapping experiment (Fig. 2) shows that TCGA provides the major transcriptional start signal in this construct. Transcription

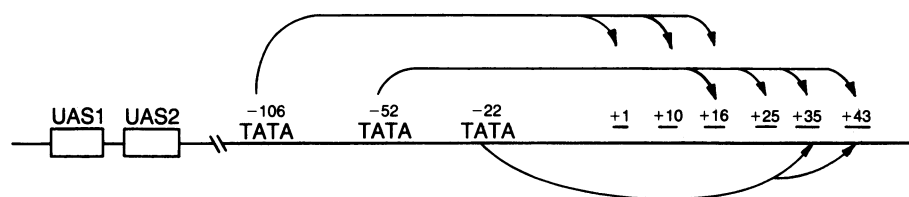


FIG. 4. Multiple TATA elements activate *CYC-1* transcription starts. mRNA initiation clusters are indicated by bars. The three functional TATA elements and the UAS elements are indicated. Proposed sites of TATA activation are indicated by arrows.

also initiates at the normal *CYC-1* +43 cluster, which has been moved to a position 74 nucleotides downstream of the -52 element. In construct Δ Mbo-Ava, two TCGA sequences were introduced on an oligonucleotide linker inserted at the deletion junction (Fig. 1). Both TCGA sequences serve as sites of initiation in this construct (Fig. 2).

DISCUSSION

Our results show that the extreme heterogeneity of transcription initiation sites for the *CYC-1* promoter can be explained, in part, by the presence of three functional TATA elements (Fig. 4). The TATA element at -106 specifies starts at +1 and +10 and contributes to the +16 cluster of starts. The TATA element at -52 specifies starts at the +16 and +23 clusters and contributes to the starts at the +34 and +43 clusters. A weaker contribution to the +34 and +43 starts is made by the TATA element at -22.

The above picture of the wild-type promoter, as well as results obtained with promoters bearing deletions, insertions, and substitutions, suggests several features that may be general for transcription initiation at RNA polymerase II promoters in yeast. First, the TATA element is capable of activating transcription only within a window 60–120 nucleotides downstream of the element. A similar conclusion was made by Nagawa and Fink (20) in studies on *HIS4* initiation. Second, the position at which initiation occurs within this window is encoded at the initiation sites themselves. Analysis of 18 promoters, transcribed by RNA polymerase II, with known start sites for transcription reveals preferred sequences for mRNA initiation. Two of these sequences, TC(G/A)A and RRYRR, can account for initiation at 31 sites (55% of the total initiation sites) in these promoters. With only a few exceptions, these sequences are always used as initiation sites unless present <50 nucleotides downstream from a TATA-like sequence. Indeed, in two separate constructs, when the sequence TCGA is inserted downstream of the *CYC-1* TATA elements, it encodes the major mRNA initiation site. Since total mRNA levels encoded by the *CYC-1* promoter are not elevated in construct Bal6^R, we conclude that the choice of an initiation site is not rate-limiting in the initiation process at *CYC-1*.

Our results show a striking difference between yeast and higher eukaryotes in the relationship between TATA elements and transcription initiation sites. In higher cells, the TATA sequence is typically located 30 nucleotides upstream from the site of transcription initiation. Although the sequence at the start of initiation has been found to affect the efficiency of initiation in a few cases (5, 37), the main determinant of initiation is thought to be the TATA sequence.

One model proposed for the mechanism of UAS elements in yeast is that the UAS acts as an RNA polymerase entry site. In one simple form of this model, RNA polymerase or some other transcription factor scans the DNA sequence downstream of the UAS until it finds a TATA sequence at which to effect initiation. Our finding that multiple TATA sites are functional at *CYC-1* appears at odds with this model. If polymerase does first interact at the UAS, it would in some way have to be accessible to all three TATA sites. One mechanism might be through a bend or loop in the DNA bringing together the UAS and TATA

elements. An alternative mechanism might be that RNA polymerase does not specifically interact with the UAS but rather with the TATA sites, perhaps via specific TATA-binding proteins analogous to those seen in higher eukaryotic cells (27). Once bound, polymerase would scan the DNA sequence 60–120 nucleotides downstream and initiate transcription at preferred sequences.

We wish to thank B. Rymond and C. Pikielny for advice on the primer-extension technique, J. Teem for a gift of *lacI* primer, and F. Nagawa and G. Fink for communication of unpublished data. This work was supported by a grant to L.G. from the National Institutes of Health and by a Damon Runyon-Walter Winchell Cancer Fund postdoctoral fellowship to S.H.

1. Benoist C. & Chambon, P. (1981) *Nature (London)* **290**, 304–310.
2. Grossveld, G. C., Shewmaker, C. K., Jat, P. & Flavell, R. A. (1981) *Cell* **25**, 215–226.
3. Tsujimoto, Y., Hirose, S., Tsuda, M. & Suzuki, Y. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4838–4842.
4. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) *Science* **209**, 1406–1414.
5. Hu, S.-L. & Manley, J. L. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 820–824.
6. Wasylyk, B., Wasylyk, C., Augereau, P. & Chambon, P. (1983) *Cell* **32**, 503–514.
7. Chandler, V., Mahler, B. & Yamamoto, K. (1983) *Cell* **33**, 489–499.
8. Banerji, J., Olsen, L. & Schaffner, W. (1983) *Cell* **33**, 729–740.
9. Queen, C. & Baltimore, D. (1983) *Cell* **33**, 741–748.
10. Guarente, L. (1984) *Cell* **36**, 799–800.
11. Struhl, K. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7385–7389.
12. Guarente, L., Yocum, R. R. & Gifford, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7410–7414.
13. Guarente, L. & Mason, T. (1983) *Cell* **32**, 1279–1286.
14. Guarente, L., Lalonde, B., Gifford, P. & Alani, E. (1984) *Cell* **36**, 503–511.
15. Donahue, T., Daves, R., Lucchini, G. & Fink, G. (1983) *Cell* **32**, 89–98.
16. Martinez-Arias, A., Yost, H. & Casadaban, M. (1984) *Nature (London)* **307**, 740–742.
17. Guarente, L. & Hoar, E. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7860–7864.
18. Struhl, K. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7865–7869.
19. Hinnenbusch, A., Lucchini, G. & Fink, G. R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 498–502.
20. Nagawa, F. & Fink, G. R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8557–8561.
21. Siliano, P. G. & Tatchell, K. (1984) *Cell* **37**, 969–978.
22. Faye, G., Leung, D., Tatchell, K., Hall, B. & Smith, M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2258–2262.
23. Guarente, L. (1983) *Methods Enzymol.* **101**, 181–191.
24. Teem, J. & Rosbash, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4403–4407.
25. Elder, R. T., Loh, E. Y. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2432–2436.
26. Gafner, J. & Philippsen, P. (1980) *Nature (London)* **286**, 414–418.
27. Parker, C. & Topol, J. (1984) *Cell* **36**, 357–369.
28. Yocum, R. R., Hanley, S., West, R. & Ptashne, M. (1984) *Mol. Cell. Biol.* **4**, 1985–1988.
29. Struhl, K. & Davis, R. (1981) *J. Mol. Biol.* **152**, 553–568.
30. Carlson, M., Taussing, R., Kustu, S. & Botstein, D. (1983) *Mol. Cell. Biol.* **3**, 439–447.
31. Andreadis, H., Usu, Y.-P., Kohlaw, G. & Schimmel, P. (1982) *Cell* **31**, 319–325.
32. Russell, D. W., Smith, M., Williamson, V. M. & Young, E. T. (1983) *J. Biol. Chem.* **258**, 2674–2682.
33. Hinnenbusch, A. G. & Fink, G. R. (1983) *J. Biol. Chem.* **258**, 5238–5247.
34. Laughon, A. & Gestland, R. F. (1984) *Mol. Cell. Biol.* **4**, 260–267.
35. Sumrada, R. A. & Cooper, T. G. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 643–647.
36. Naysmyth, K., Tatchell, K., Hall, B. D., Astell, C. & Smith, M. (1981) *Nature (London)* **289**, 244–250.
37. Concino, M., Lee, R. F., Merryweather, J. P. & Weinmann, R. (1984) *Nucleic Acids Res.* **12**, 7423–7433.