Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

# A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation

**Victoria L. Singer, C. Richard Wobbe, and Kevin Struhl**

Department Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115 USA

**We created a library of DNA molecules in which the required TATA element of a yeast *gal–his3* promoter is replaced by random-sequence oligomers averaging 16 bp in length. Surprisingly, 1% of such random sequences functionally replace the native yeast TATA element. In many cases, sequences completely unrelated to the consensus TATA element (TATAAA) stimulate transcription with equal or increased efficiency. Transcription mediated by these synthetic elements requires GAL4 and is initiated from normal *his3* initiation sites, suggesting that it occurs by a mechanism indistinguishable from that involving wild-type TATA elements. Many, but not all, of these elements act as substrates for yeast TFIID in reconstituted transcription reactions in vitro. These observations indicate that yeast TFIID can stimulate transcription from TATA elements whose sequences differ from the consensus, and they suggest the possibility of alternative factors that may provide a related function for transcriptional activation.**

[*Key Words*: Eukaryotic promoters; RNA polymerase II; TATA-binding protein; transcriptional initiation; *Saccharomyces cerevisiae*]

Received November 27, 1989; revised version accepted January 25, 1990.

Yeast RNA polymerase II promoters consist of three main functional elements, with distinct DNA sequence requirements (Guarente 1988; Struhl 1989). Upstream activating sequences (UASs) function at long and variable distances from the mRNA initiation site and generally act by binding gene-specific regulatory proteins such as GCN4 or GAL4. Although upstream elements in unrelated promoters have very different DNA sequences and interact with different proteins, individual UASs have relatively stringent DNA sequence requirements because of the constraints of specific protein–DNA interactions. TATA elements, located 40–100 bp upstream from the mRNA initiation site, strongly influence the level of transcription and the selection of start sites; they typically contain sequences related to TATAAA. Initiator elements, which are located very near the mRNA start site, are responsible for determining precisely where transcription begins. Initiator elements appear to have nonstringent sequence requirements, and it has been suggested that they may reflect local sequence preferences for initiation by RNA polymerase II (Chen and Struhl 1989).

The presence of the conserved TATAAA sequence in most eukaryotic promoters has led to the idea that TATA elements function by binding a common factor, TFIID, which is a component of the general transcriptional machinery. Thus, TATA elements have been commonly defined by the following criteria: (1) their downstream location in promoters, which strongly influences the location of transcriptional start sites; (2) their similarity to the consensus sequence TATAAA; (3) their ability to bind the factor TFIID, which is required early in the transcription process in vitro; and (4) their ability to stimulate transcription in combination with upstream activator proteins such as GCN4 and GAL4. Because the mechanism of transcriptional initiation by RNA polymerase II is not well understood, however, it is unclear whether all of these criteria should be grouped together. By analogy, before the discovery of distinct σ-factors, it was thought that prokaryotic promoters used a single factor that recognized the conserved −10 and −35 promoter elements.

The yeast *his3* promoter contains two functionally distinct TATA elements: $T_C$, which is required for constitutive transcription, and $T_R$, which is required for expression induced by the activator protein GCN4 (Struhl 1986). Of these elements, $T_R$ contains the sequence TATATAAA, which includes the TATA consensus sequence. Extensive point mutagenesis of $T_R$ in the context of a *gal–his3* hybrid promoter indicates that oligonucleotides containing the sequences TATAAA or TATATA provide $T_R$ function (Chen and Struhl 1988). In contrast, all other possible single point mutations of TATAAA are functionally defective for transcriptional activation. In addition, of 12 single point mutations of TATATA, only one, TATTTA, functions as a $T_R$ element (Harbury and Struhl 1989).

From these apparently stringent sequence require-

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press



ments, it has been suggested that $T_R$ sites are bound by a specific protein that is relatively intolerant of sequence deviation in its binding site. However, the yeast *his3* $T_C$ sequence and a number of other yeast (and higher eukaryotic) promoters do not contain sequences similar to TATAAA, TATATA, or TATTTA. In addition, the $T_R$ element can be replaced by a GCN4-binding site to permit transcriptional activation of a *gal–his3* promoter (Chen and Struhl 1989). In this case, GCN4 activates transcription when bound at the position of a TATA element but apparently in the absence of the conventional TATA-binding factor. Therefore, it seems likely that functional TATA elements containing sequences unrelated to canonical TATA sites must exist.

To test this hypothesis, we used the method of random selection (Oliphant and Struhl 1987) to obtain a novel class of functional $T_R$ elements. Specifically, we generated a library of DNA molecules, in which the wild-type *his3* $T_R$ element in a *gal–his3* fusion promoter was replaced with synthetic random-sequence oligonucleotides, and assayed the resulting clones for $T_R$ function in yeast. We find that many sequences are capable of serving as fully functional $T_R$ elements in this promoter and that a large percentage of these sequences bear no similarity to known TATA sites. We then examined some of these functionally selected DNA sequences for their ability to respond to purified yeast TFIID in reconstituted transcription reactions in vitro. Interestingly, several of these synthetic elements supported TFIID-dependent transcription, whereas others did not. We discuss the implications of these results with respect to the properties of TFIID and to the possibility of other factors that may perform a similar function.

## Results

### *A hybrid promoter is an effective probe for $T_R$ function*

A plasmid was constructed that would enable the identification of DNA fragments that could act as $T_R$ elements (Fig. 1). This plasmid, pD–*lacZ*, contains a *gal–his3* promoter identical to that used in previous studies for site-directed mutagenesis of the $T_R$ element (Chen and Struhl 1988). Specifically, the 365-bp *GAL1,10* fragment, containing four GAL4-binding sites, was fused upstream of *Eco*RI and *Sac*I restriction endonuclease sites, between which test oligonucleotides could be inserted. The wild-type *his3* initiation region and the amino-terminal half of the *his3* structural gene were present downstream of these sites, and a functional *Escherichia coli lacZ* gene was fused in frame to *his3*. These studies differed from previous work, in that the plasmids were propagated as circular minichromosomes at one copy per cell rather than being integrated into the yeast chromosome by gene replacement.

To test the probe plasmid, three DNA fragments characterized previously were inserted into the vector and examined for their ability to express the *his3–lacZ* fusion protein and hence give rise to blue colonies on plates containing the indicator dye X-gal (Fig. 2). As ex-

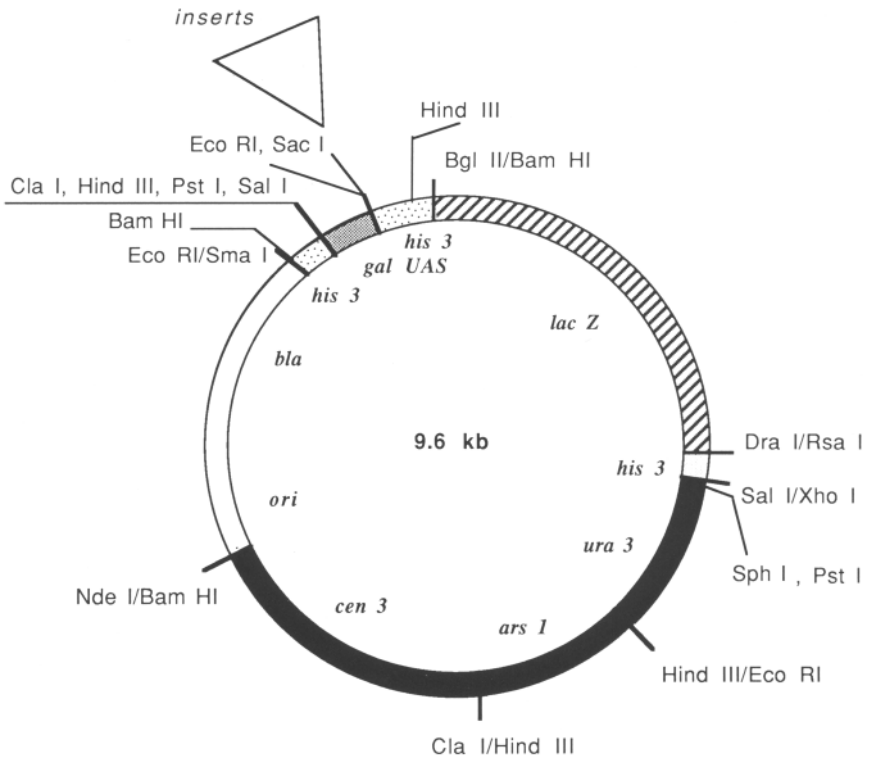


**Figure 1.** Restriction and genetic map of the $T_R$ probe plasmid. All oligonucleotide inserts were subcloned between the *Eco*RI and *Sac*I restriction endonuclease sites.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press



| *his3* allele pattern | DNA | insert sequence | phenotype | |
|---|---|---|---|---|
| G17,D | 3800 | (2.2 kb *Drosophila* DNA insert) | - | |
| G17 | 3801 | CTATAAAGTAATGTG | ++ | A |
| wild type | 3802 | (wild-type *his3* promoter) | - | A,B |
| G17,206 | 3803 | CTTTAAAGTAATGTG | - | |
| G17,R6 | 3821 | GACCGCATTTAATCC | +++ | D |
| G17,R17 | 3822 | CTATAACGTAATGTG | +++ | |
| G17,R35 | 3823 | ATTATCATTTAATTAC | +++ | D |
| G17,R37 | 3824 | TATATAAATTTTCCT | +++ | A,B |
| G17,R45 | 3825 | TATATAAGAAACATTC | +++ | B |
| G17,R40 | 3826 | Triple Insertion | ++ | |
| G17,R2 | 3827 | CGCCGCCTATTTAATCC | ++ | C |
| G17,R3 | 3828 | ACTCTTGTGCCGATA | ++ | E |
| G17,R7 | 3829 | CTACTATAAAAACCCA | ++ | A,F |
| G17,R60 | 3830 | CATTACTATAAACAGAC | ++ | A |
| G17,R72 | 3831 | CGGCCTTTAATATTAA | ++ | |
| G17,R80 | 3832 | AACTTAATAAACCAATC | ++ | |
| G17,R30 | 3833 | GCGACCGATAACCCTC | + | |
| G17,R4 | 3834 | GCAATGTAGATTTTTT | + | |
| G17,R16 | 3835 | TATAAAACAAAACCCC | + | A,F |
| G17,R48 | 3836 | CAATAAATATTCTAACCT | + | |
| G17,R74 | 3837 | ATACATAAAATGACTC | + | |
| G17,R81 | 3838 | TATTATACCCGACCT | + | |
| G17,R225 | 3839 | TATATATACGCATACAT | + | B |
| G17,R233 | 3840 | CACTCTATTTAACCAA | + | C |
| G17,R1 | 3841 | GCCCGTTTCCAACC | +/- | |
| G17,R9 | 3842 | CTCGATCCTCGGCTCCA | +/- | |
| G17,R12 | 3843 | CCCCCCTTCAAATATCA | +/- | |
| G17,R15 | 3844 | CACCGCTATTTAAATCCC | +/- | C |
| G17,R26 | 3845 | ACTCTGTGCCGCTA | +/- | E |
| G17,R33 | 3846 | CTACTACTATTAAAACCCA | +/- | F |
| G17,R44 | 3847 | AGTCTTATACATTTAAC | +/- | |
| G17,R47 | 3848 | CATCCTTCACTAATCT | +/- | |
| G17,R51 | 3849 | TTAAAAGCGTCCCATTTC | +/- | |
| G17,R234 | 3850 | AAACATACACACCACA | +/- | |
| G17,R238 | 3851 | CACTTTAGATCCTCTA | +/- | |

*Key to patterns:

A = TATAAA
B = TATATA
C = TATTTAA
D = CATTTAAT
E = TGTGCCG
F = AAAACCC

**Figure 2.** Insert sequences of the $T_R$ elements. The sequences and phenotypes of the $T_R$ elements identified in these studies are shown. With the exception of 3800, 3801, 3802, and 3803, all oligonucleotide sequences are flanked by the sequence GGAATTC on the 5′ end and CGAGCTC on the 3′ end; 3801 and 3803 are flanked by the sequences 5′-TGAATTC and 3′-GAGCTCC. The phenotypes are based on intensity of color on galactose/X-gal indicator plates, using the following standards: 3800 (−), 3841 (+/−), 3834 (+), 3801 (++), and 3821 (+++). The occurrence of representative sequence patterns is shown in the last column, with the key to the patterns listed below; the list is not meant to be exhaustive.

pected, insertion of a functional $T_R$ element (TATAAA) resulted in blue colonies on plates containing galactose but not glucose, whereas cells containing either a point mutation of $T_R$ (TTTAAA) or *Drosophila* sequences lacking a TATA site were white. Therefore, the probe plasmid efficiently distinguishes functional $T_R$ elements from other DNA sequences.

### *Selection of functional $T_R$ elements from random-sequence oligonucleotides*

To identify new $T_R$ elements, a library containing ~3000 unique clones was constructed by insertion of synthetic random-sequence oligonucleotides, ~16 bp in length, into the probe vector. The sequences of 10 clones chosen at random from the library indicated that the frequency at which individual nucleotides arose was not entirely random: G (19%), A (25%), T (23%), and C (33%). On the basis of these observed frequencies, however, it was possible to estimate the expected frequency of occurrence of particular sequence patterns in our library.

To assay for $T_R$ function, the library was introduced into yeast cells and transformants were patched onto X-gal indicator plates containing either glucose or galactose as the carbon source. Approximately 1% of the transformants yielded a blue patch after 4 days when grown on plates containing galactose. None of the transformants produced blue colonies on glucose plates.

The intensity of the blue color varied greatly among individual transformants; therefore, four representative colonies were chosen to indicate relative levels of $T_R$ activity. All other colonies were classified according to similarity with those standards. Assays of β-galactosidase activities present in these colonies confirmed the original groupings and showed that weak elements detectable on plates were essentially not detectable in liquid assays (Fig. 2, Table 1). For example, alleles R5 and R33, which were grouped as +/−, gave rise to 5–20 units of β-galactosidase activity (nearly uninduced levels). In contrast, alleles grouped as + (R4, R30, and R74) gave rise to 20–85 units, alleles grouped as ++ (R2, R3, R40, and R80) gave 90–230 units, and alleles classified as +++ (R6, R17, R35, and R37) yielded 300–500 units. Note that G17, which contains the sequence TATAAA and is the wild-type element used in previous studies, is grouped as ++ and gives rise to only 100 units. Therefore, many of the newly identified $T_R$ elements result in higher levels of gene expression than

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press



**Table 1.** *Expression conferred by randomly selected $T_R$ elements*

| Allele | Insert sequence | Glucose | Galactose |
|---|---|---|---|
| D | 2.2-kb *Drosophila* DNA insert | 1 | 0 |
| $T_R$ | CTATAAAGTAATGTG | 0 | 100 |
| R2 | CGTCCGCCTATTTAATCC | 2 | 140 |
| R3 | ACTCTTGTGCCGATA | 7 | 100 |
| R4 | GCAATGTAGATTTT | 8 | 60 |
| R6 | GACCGCATTTAATCC | 9 | 430 |
| R7 | CTACTATAAAAACCCA | 15 | 120 |
| R16 | TATAAAACAAAACCCC | 12 | 150 |
| R17 | CTATAACGTAATGTG | 2 | 450 |
| R30 | GCGACCGATAACCCTC | 0 | 60 |
| R33 | CTACTACTATTAAAACCCA | 3 | 20 |
| R35 | ATTATCATTTAATTAC | 6 | 450 |
| R37 | TATATAAATTTTCCT | 7 | 450 |
| R40 | triple insertion | 6 | 230 |
| R45 | TATATAAGAAACATTC | 15 | 450 |
| R72 | CGGCCTTTAATATTAA | 9 | 190 |
| R74 | ATACATAAAATGACTC | 10 | 80 |
| R80 | AACTTAATAAACCAATC | 6 | 220 |

Yeast cells containing the indicated plasmids were grown in minimal media containing the indicated carbon source(s). Approximately $10^7$ cells from each culture were used for each assay. Numbers are units of β-galactosidase (Miller 1972) that represent an average of three independent assays. Because of inherent errors involved in measurements of low β-galactosidase activities, apparent differences in the glucose/galactose induction ratios are probably not significant.

the wild-type TATA element. As it was likely that both the transcript and translated products were identical in all cases, the relative intensities of blueness and β-galactosidase levels were taken as an initial indication of transcriptional activity. In fact, this assumption was confirmed by S1 nuclease analysis, as described below.

### *Some efficient $T_R$ elements contain sequences unrelated to canonical TATA elements*

Plasmids were rescued from yeast as described (Materials and methods), and inserts were sequenced to identify the $T_R$ elements. Insert sequences and relative promoter strengths of 33 of these clones are shown (Fig. 2). As expected, the sequences TATAAA, TATATA, and TATTTA arose in this collection, appearing four times, four times, and three times, respectively. In addition, many clones contained inserts completely unrelated to known $T_R$ sites. For example, R3 does not even contain an AT-rich insert and is still of similar efficiency to T, which contains the sequence TATAAA (Fig. 2).

To look for sequence elements that may represent binding sites for proteins other than the normal $T_R$ binding protein, we analyzed inserts for repeated sequence motifs. This was accomplished with a computer program designed to identify sequence 'words' that arose at a greater frequency than that predicted by random occurrence (Oliphant 1989). In addition to locating the sequences TATAAA and TATATA, several other recurring words were identified. The most prevalent of these, TATTTAA, CATTTAAT, TGTGCCG, and AAAACCC, are indicated where they occur in the $T_R$ inserts (Fig. 2). Each of these sequences was found far more frequently than can be accounted for by random occurrence. For example, CATTTAAT arose twice among the 33 sequenced inserts (R6 and R35) but would be predicted to arise once at a frequency of $8 \times 10^{-3}$ or twice at a frequency of $7 \times 10^{-5}$; interestingly, R6 and R35 are among the strongest elements.

Note that TATTTAA and CATTTAAT are actually similar sequences and that these and several related sequences (R44, CATTTAAC; R15, TATTTAAA; R2–33, TATTTAAC) can be explained by considering them all to be representatives of an element with consensus sequence (C/T)ATTTAAt. If this explanation is correct, the exceptional double mutation identified previously (TATTTA; Harbury and Struhl 1989) could be viewed as a single point mutation of this new element.

### *Novel $T_R$ elements behave indistinguishably from the* his3 *$T_R$ element*

To determine whether transcription mediated by these novel $T_R$ elements occurs by essentially the same mechanism as the wild-type *his3* element, 5′ ends of the transcripts were mapped by S1 analysis. In all cases, transcription was initiated from the normal *his3* RNA start sites, designated +1, +12, and +22 (Fig. 3). More importantly, in all but one case, the vast majority of transcripts were initiated from the +12 site, as is observed for the wild-type *his3* $T_R$ element. The sole exception to this observation, R40, gives rise to transcripts that initiate primarily at +1, but this derivative contains three oligonucleotide insertions, including a TATA-like sequence located farther upstream. In accord with previous observations (Struhl 1986), it is extremely likely that the unusual initiation pattern seen in R40 merely reflects an increased distance between the putative $T_R$

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press



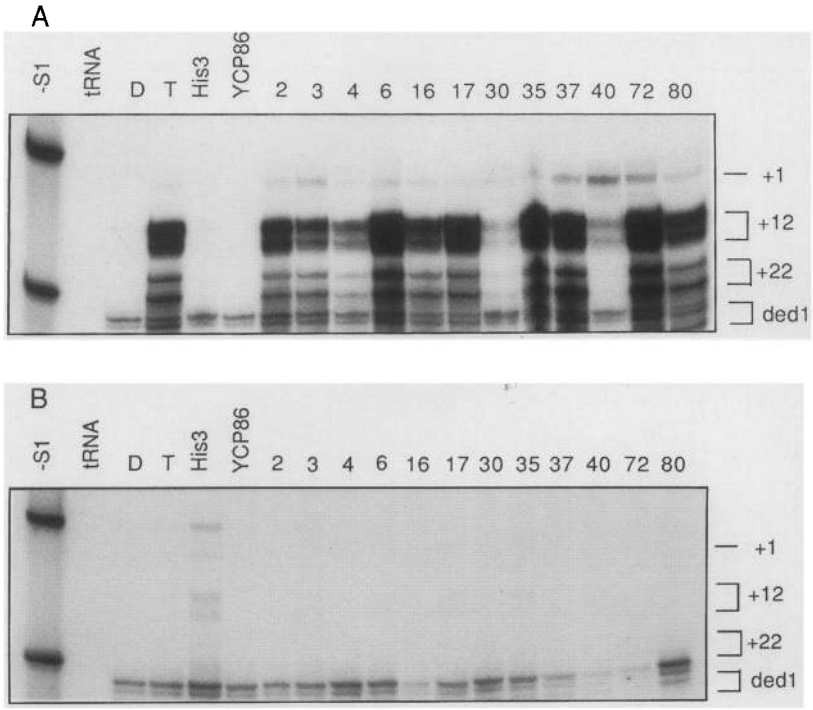


**Figure 3.** RNA analysis. RNA was prepared from yeast KY320 cells containing the indicated plasmids that were grown in YP media containing either (*A*) galactose or (*B*) glucose as a carbon source. For each sample, 50 μg of RNA was hybridized to an equimolar mixture of *his3* and *ded1* 5′-end-labeled oligonucleotide probes and then digested with S1 nuclease: (Lane −S1) probe alone; lane *tRNA*; (lane *D*) 3800; (lane *T*) 3801; (lane *His3*) wild-type *his3* gene; (lane *YCp86*) vector YCp86; (lane *2*) 3827; (lane *3*) 3828; (lane *4*) 3834; (lane *6*) 3821; (lane *16*) 3835; (lane *17*) 3822; (lane *30*) 3833; (lane *35*) 3823; (lane *37*) 3824; (lane *40*) 3826; (lane *72*) 3831; and (lane *80*) 3832. The locations of DNA fragments corresponding to transcripts initiating at the *his3* and *ded1* initiation sites are shown. The primary transcript initiated by the fully induced wild-type $T_R$ element is +12, although transcription is also induced from +1 and +22 sites.

element and the *his3* initiation region. So, as in the wild-type *his3* promoter, the novel $T_R$ elements direct transcription from the wild-type initiation sites, and induction occurs at the site where it occurs in promoters containing wild-type $T_R$ elements.

Two experiments indicate that like the wild-type $T_R$ element, transcription mediated by the novel $T_R$ elements requires the presence and activity of GAL4 protein. First, β-galactosidase levels and relative blueness of patches on indicator plates were measured when cells were grown in the presence of galactose, raffinose, or glucose as the carbon source. Second, the same assay methods were used to examine gene expression conferred by six of the novel elements in an isogenic strain containing a deletion of the *gal4* structural gene. For both the wild-type and novel $T_R$ elements, expression was dependent on the presence of galactose in the media (Fig. 2; Table 1) and on the presence of an intact copy of the *GAL4* structural gene (Table 2). Thus, as far as can be determined by these assays, the mechanism of activation by the new $T_R$ elements is indistinguishable from that used by 'normal' $T_R$ elements.

### *Some, but not all, novel $T_R$ elements support TFIID-dependent transcriptional initiation in vitro*

Mammalian cell-free extracts contain a variety of chromatographically separable factors that can reconstitute accurate transcription in vitro when combined. Interestingly, a yeast TATA-binding protein can functionally substitute for the mammalian TATA-binding factor TFIID in such reconstituted transcription reactions (Buratowski et al. 1988, 1989; Cavallini et al. 1988). Using this assay, we have purified yeast TFIID extensively and tested its ability to activate transcription from promoters containing some of the novel $T_R$ elements described here.

The templates for in vitro transcription are derived from pML($C_2$AT)19, which contains the TATA element and initiation region of the adenovirus major late promoter fused to RNA-coding sequences lacking any G residues, the so-called 'G-less cassette' (Sawadogo and Roeder 1985). Transcription from this promoter requires TFIID and the other general factors but does not depend on upstream activator proteins such as GAL4. To study our collection of mutated and randomly selected TATA elements, pML($C_2$AT)19 DNA was modified by introducing *Eco*RI and *Sac*I sites at positions that flank the TATA sequence of the adenovirus promoter (Fig. 4; C.R. Wobbe and K. Struhl, in prep.). *Eco*RI–*Sac*I fragments containing representatives of the novel $T_R$ elements were then cloned in place of the adenovirus TATA sequence and examined for their ability to support TFIID-dependent transcriptional initiation in vitro.

Five of the novel $T_R$ elements tested (R2, R35, R37, R72, and R80) clearly result in TFIID-dependent transcription (Fig. 4). All of these elements contain a stretch of eight or more consecutive A or T residues, but only R37 contains a TATAAA or TATATA sequence. Quantitative measurements (Table 3) indicate that, in comparison to transcription mediated by the wild-type *his3* $T_R$ element in G17, the R35, R37 and, possibly, R2 elements are somewhat more efficient, R80 is similarly efficient, and R72 may be somewhat less efficient. These relative efficiencies for basal TFIID-dependent transcription in vitro are very similar to the relative efficiencies

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press



**Table 2.** *GAL4 regulation of randomly selected $T_R$ elements*

| Allele | Wild type | | | Gal4 deletion | | |
|---|---|---|---|---|---|---|
| | raffinose | raffinose/glucose | raffinose/galactose | raffinose | raffinose/glucose | raffinose/galactose |
| *Drosophila* | 1 | 1 | 4 | 0 | 0 | 4 |
| *his3* $T_R$ | 0 | 1 | 80 | 2 | 0 | 0 |
| None[1] | 0 | 1 | 2 | 3 | 1 | 1 |
| R2 | 1 | 4 | 110 | 3 | 2 | 1 |
| R3 | 1 | 3 | 120 | 4 | 2 | 2 |
| R17 | 2 | 4 | 270 | 4 | 3 | 2 |
| R35 | 2 | 9 | 430 | 4 | 3 | 4 |
| R40 | 1 | 6 | 230 | 4 | 4 | 6 |
| R80 | 2 | 6 | 230 | 5 | 4 | 4 |

Yeast cells containing plasmids with the indicated $T_R$ derivative were grown in minimal medium containing the indicated carbon source(s), as described (Materials and methods). Approximately $10^7$ cells from each culture were used for each assay. Numbers are units of β-galactosidase (Miller 1972) that represent an average of three independent assays.
[1]This allele contains a wild-type his3 promoter with no GAL4-binding sites.

for GAL4-activated transcription in vivo, with the possible exception of R72, which may be somewhat weaker in vitro than expected. Thus, TFIID can interact functionally with sequences that differ from the canonical TATA element and in some cases, with greater efficiency.

Interestingly, two of the novel $T_R$ elements, R3 and R4, do not appear to support TFIID-dependent transcription in vitro. The R3 element does not contain any AT-rich sequence, whereas the R4 element is relatively AT-rich. The failure of R3 and R4 to permit TFIID-dependent transcription in vitro is unlikely to be caused by an abnormal spacing between potential TFIID target sites and the mRNA initiation site. In the functional $T_R$ elements assayed in Figure 4, the likely TFIID interaction sites are located at various distances from the initiation site, yet the levels of transcription in vitro are in accord with those achieved during GAL4 induction in vivo.

## Discussion

### *A wide variety of unrelated sequences act as TATA elements in yeast*

Previous analysis indicated that the *his3* $T_R$ element has relatively stringent DNA sequence requirements; that is, almost all single mutations of TATAAA or TATATA severely reduced the level of transcriptional activation by GAL4 or GCN4 protein (Chen and Struhl 1988; Harbury and Struhl 1989). In striking contrast to these observations, the results here indicate that a wide variety of sequences, many having no similarity to known TATA elements, can function efficiently as TATA sites in yeast. In the identical context of a *gal–his3* promoter, these novel $T_R$ elements stimulate transcriptional initiation from the same sites as the *his3* $T_R$ element and are responsive to induction by GAL4. Many of the novel sites stimulate transcription at least as efficiently as the wild-type $T_R$ element. Therefore, even though many of these sequences are dissimilar to the canonical TATA sequence, they appear to function by a mechanism indistinguishable from that used by wild-type elements.

### *TFIID stimulates transcription from sequences that differ from the TATA consensus*

One explanation for our observations is that the known TATA-binding factor, TFIID, can functionally interact with sequences other than TATAAA or TATATA, the canonical TATA element. By analogy, yeast HAP1 (Pfeifer et al. 1987), steroid receptors (Sakai et al. 1988), C/EBP (Landschulz et al. 1988), TEF (Davidson et al. 1988), and OTF-2 (Poellinger et al. 1989) have been proposed to bind dissimilar sequences. The results in Figure 4 indicate that many derivatives lacking a canonical TATA sequence (R2, R35, R80, and R72) are transcribed in vitro in a yeast TFIID-dependent manner. Moreover, when these randomly selected elements are compared to the wild-type *his3* $T_R$ element, their relative levels of 'basal-TATA' transcription in vitro are very similar to their relative levels of GAL4-enhanced transcription in vivo (with the possible exception of R72, which may be somewhat less effective in vitro). Thus, it is very likely that many of the novel $T_R$ elements represent TFIID binding sites that differ in sequence from the TATA consensus. During the preparation of this manuscript, it was shown that some noncanonical sequences from natural promoters could respond to TFIID in vitro (Hahn et al. 1989).

The fact that TFIID interacts functionally with noncanonical TATA sequences does not conflict with our previous mutational analyses indicating that the TATA factor has a high sequence specificity (Chen and Struhl 1988; Harbury and Struhl 1989). These apparently contradictory observations can be reconciled by proposing that sequences flanking the 6-bp core of the TATA element can influence the TFIID interaction. Indeed, point mutations downstream from the core can influence transcriptional activity significantly, with A and T residues being preferred over G and C residues (Harbury and Struhl 1989). In this regard, R2, R35, R80, and R72 all contain at least eight consecutive A or T residues, even though they lack a TATAAA or TATATA sequence. Perhaps, flanking A or T residues can compensate for a nonconsensus hexanucleotide core. However,

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press



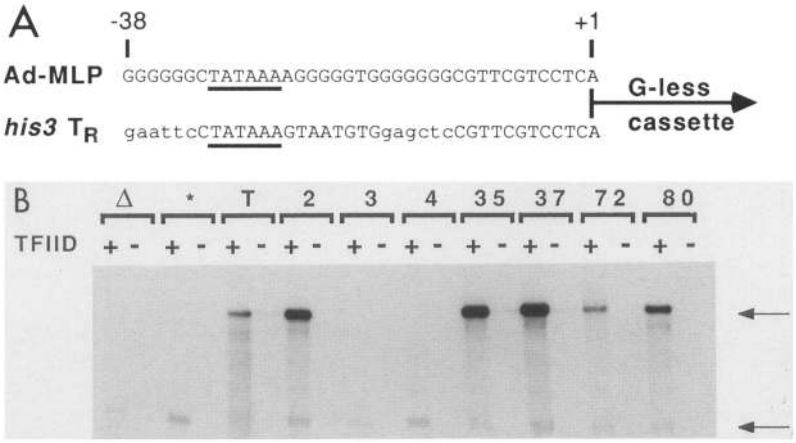


**Figure 4.** Transcription in vitro. (*A*) Sequences of the TATA element region (nucleotides −38 to +1, with TATA sequence underlined) of the adenovirus major late promoter (Ad-MLP) and of the wild-type *his3* promoter flanked by *Eco*RI and *Sac*I sites (lowercase letters) that are fused to the G-less cassette (Sawadogo and Roeder 1985). Randomly selected $T_R$ elements were cloned as *Eco*RI–*Sac*I oligonucleotides in place of the *his3* $T_R$ element. (*B*) An autoradiograph in which DNA templates containing the indicated $T_R$ derivatives were transcribed by RNA polymerase II and HeLa cell factors A, B, and E + F in the presence (+) or absence (−) of yeast TFIID. The derivative indicated as Δ lacks any TATA oligonucleotide, and the derivative indicated by an asterisk (*) contains the *his3*-214 mutation in $T_R$, TATAGA, that abolishes function in vivo (Chen and Struhl 1988). The top arrow indicates the position of the transcript initiated from the correct adenovirus mRNA start site; the bottom arrow indicates the position of a transcript initiated 25–30 bp downstream from a TATA-like sequence within the G-less cassette. The shorter transcript in a given lane serves as an internal control for quantitating the level of the correctly initiated transcript in that lane (C.R. Wobbe and K. Struhl, in prep.).

it is very unlikely that TFIID can recognize all sequences containing eight consecutive A or T residues because in that case, the expected frequency of obtaining functional $T_R$ elements in our experiment would be ~4%.

### *Are there multiple factors that perform the TATA function?*

At least two of the synthetic elements described here, R3 and R4, fail to support TFIID-dependent transcription in vitro even though they functionally replace the $T_R$ element required for GAL4 activation in vivo. At face value, this indicates that TFIID does not interact with the R3 and R4 elements and, hence, that GAL4 activation can occur without TFIID. However, we cannot exclude the possibility that GAL4 increases either the affinity of TFIID for the R3 and R4 elements or the activity of TFIID bound to these elements. Such effects, if they occur, may reflect allosteric interactions between GAL4 and TFIID or GAL4-induced alterations in DNA structure that facilitates TFIID binding to certain sequences. A strong, but not compelling, argument against these possibilities is that similar analysis of many of the TATA mutations described previously (Chen and Struhl 1988; Harbury and Struhl 1989) indicates an extremely good correlation between the relative levels of TFIID-dependent transcription in vitro and GAL4-activated transcription in vivo (C.R. Wobbe and K. Struhl, in prep.). In addition, unlike the wild-type $T_R$ element and several of the derivatives described here, R3 and R4 do not support activation by a GAL4 derivative in yeast nuclear extracts, conditions that require a functional TATA element and presumably TFIID for transcription (A.S. Ponticelli, V.L. Singer, and K. Struhl, unpubl.).

If R3 and R4 do not interact with TFIID, the simplest explanation for their ability to functionally replace the $T_R$ element is that they interact with distinct proteins that carry out the TFIID function. In this model, yeast cells would contain a set of specificity factors that each interact with different downstream promoter sequences, as is the case for bacterial σ-factors (Losick et al. 1986; Helmann and Chamberlin 1988). This hypothesis does not imply that TATA factors function in any mechanistic sense like σ-factors; rather, it implies that a TATA factor may be required for transcription of all promoters, but the particular factor needed for a particular promoter would depend on the DNA sequences of the downstream element. In this view, TFIID is the major, but not unique, TATA factor, much like $\sigma^{70}$ is the major, but not sole σ-factor in *E. coli*. Attempts to biochemically identify such an alternative factor by assaying all chromatographic fractions from the initial heparin–agarose column used to purify TFIID for their activities on promoters containing R3 and R4 in the reconstituted in vitro system have not yet been successful.

**Table 3.** *In vitro activity of randomly selected $T_R$ elements*

| Allele | Sequence | Relative activity |
|---|---|---|
| *his3* $T_R$ | CTATAAAGTAATGTG | 100 |
| G3 | deletion | <1 |
| 214 | CTTTAAAGTAATGTG | <1 |
| R2 | CGCCGCCTATTTAATCC | 290 |
| R3 | ACTCTTGTGCCGATA | <1 |
| R4 | GCAATGTAGATTTTTT | <1 |
| R35 | ATTATCATTTAATTAC | 300 |
| R37 | TATATAAATTTTCCT | 300 |
| R72 | CGGCCTTTAATATTAA | 60 |
| R80 | AACTTAATAAACCAATC | 130 |

Relative levels of TFIID-dependent transcription in vitro were determined as described (Materials and methods), with the activity of the wild-type *his3* $T_R$ element being defined as 100. The underlined sequences represent motifs described in Fig. 2, and the numbers represent the average of two different assays.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press



An alternative explanation for derivatives such as R3 and R4 is that they may give rise to novel DNA structures that bypass the requirement for TFIID and stimulate transcription in its absence. If TATA sites represent melting nucleation sites for the formation of an initiation bubble, proteins such as TFIID may then be required to mediate the melting event in some cases, in other cases, DNA structure may be sufficient. However, a major difficulty with this hypothesis is that it fails to explain the requirement for GAL4 (or GCN4) for transcriptional activation through the $T_R$ elements. Another possibility is that R3 and R4 are binding sites for transcriptional activators that normally do not act as TATA factors but can substitute for such proteins. Indeed, GCN4, normally an upstream activator protein, can stimulate transcription when its binding site is present in exactly the same context as the novel $T_R$ elements (Chen and Struhl 1989). Unlike all of the novel $T_R$ elements described here, however, GCN4 activation from the equivalent downstream location occurs in glucose medium, is not stimulated by GAL4, and does not occur with the normal $T_R$-dependent initiation pattern; thus, we believe that this explanation is not very likely.

### *Implications for eukaryotic promoters*

The observation that sequences very different from TATAAA function indistinguishably from TATA elements in yeast has implications for the structure and function of eukaryotic promoters. For promoters without readily identifiable TATA sequences, we suggest that there are functionally equivalent elements that are simply not easy to identify given our current knowledge of TATA sequences. Furthermore, our studies show that it is difficult to predict the location and the functional activity of a TATA element simply by inspecting the DNA sequence. Indeed, even for the two well-characterized motifs, TATAAA and TATATA, the sequence of the four downstream nucleotides exerts a significant effect on function (Harbury and Struhl 1989).

It is important to point out that we did not do an exhaustive search for $T_R$ elements, and we probably identified only a subset of sequences capable of functioning as $T_R$ elements. Given ~3000 inserts averaging 16 nucleotides in length, it can be predicted that an element whose specificity is equivalent to six exact and contiguous nucleotides would arise approximately eight times or that it almost certainly would have been detected. For an element specified by 7, 8, 9, or 10, nucleotides, however, the probability of detection would be ~75%, 30%, 8%, and 2%, respectively. Thus, it is not surprising that many of the novel $T_R$-elements cannot be classified into a pattern because they may represent individual isolates.

Finally, the frequency of functional $T_R$ elements, 1% under the specific experimental conditions here, provides some information about the number and specificity of the interacting factors. For example, the observed result is compatible with a single factor whose specificity is equivalent to 5 bp or with 20 factors whose specificities are equivalent to 7 bp. These calculations should be considered only as rough estimates because they are influenced by the specific promoter, by the genetic screen, and by the nature of the element (i.e., localized to contiguous base pairs or spread out over a larger number of positions). In any event, the high frequency of sequences that perform the TATA function predicts that such elements should often be found downstream of enhancers (expected frequency about once per 1000 bp) and, hence, activate transcription from undesired initiation sites. Several explanations could be invoked for why this does not appear to occur; certain elements may be responsive only to particular UAS elements, evolution may select against functional TATA elements at undesired positions in promoters, or initiator elements downstream of the TATA elements may be necessary for efficient transcriptional initiation.

## Materials and methods

### *Strains and plasmids*

The bacterial strains used in these experiments were *E. coli* SCS1 (*endA1, hsdR17* ($r_{K-}$, $m_{K+}$), *supE44, thi-1, recA1, gyrA96, relA1*) for construction and propagation of the $T_R$ library and *E. coli* HB101 [*hsdS20* ($r_{K-}$, $m_{K-}$), *recA13, ara-14, proA2, lacY1, galK2, rpsL20* ($Sm^r$), *xyl-5, mtl-1, supE44*] for rescue of plasmids into bacteria after screening in yeast. Yeast strains were KY320 (*trp1-Δ1, ura3-52, lys2, ade1, leu2-P1, his3-Δ200*) (Chen and Struhl 1988) and an isogenic strain containing a deletion of the *gal4* structural gene.

Plasmids used in these experiments are all derivatives of YCp86, YCp86-Sc3416, and YIp55-Sc3640. YCp86 contains the 2-kb *Eco*RI–*Nde*I fragment of pUC18 that includes the *bla* gene and the origin of replication, a 1.9-kb *Bam*HI–*Eco*RI fragment of yeast DNA containing *cen3* and *ars1*, a 1.1-kb fragment of DNA containing the *ura3* gene, and a polylinker. Sc3416 (constructed by Joan Sellers) consists of a 2.9-kb *Eco*RI–*Bgl*II fragment of the *his3* locus (coordinates −2500 to +419) fused to the 3.4-kb *Bam*HI–*Dra*I fragment of the *E. coli lacZ* gene (includes all coding sequences downstream of codon 8 but contains mutations in the *Eco*RI and *Sac*I sites), followed by a 260-bp *Rsa*I–*Xho*I fragment that includes sequences involved in transcriptional termination of *his3*. YCp86-Sc3802 was constructed by mutating the *Eco*RI and *Hin*dIII sites at the boundary of the polylinker in YCp86-3416 and filling in the restriction sites with Klenow enzyme. The *Bam*HI–*Hin*dIII fragment of YIp55-3640 (Chen and Struhl 1988) containing the *gal1,10* UAS region was then inserted into the resulting plasmid, in place of *his3* promoter sequences, giving rise to plasmid Sc3801, which contains a wild-type TATA site. pD–*lacZ* (designated Sc3800) was then constructed by subcloning a 2.2-kb fragment of *Drosophila* DNA into the *Eco*RI and *Sac*I sites of Sc3801, in place of the TATA sequences.

### *Construction and propagation of the $T_R$ library*

Single-stranded oligonucleotides containing random DNA sequences averaging 16 bp in length were converted to double-stranded DNA by mutually primed synthesis (Oliphant et al. 1986). A library of ~3000 clones was prepared in *E. coli* by cloning these random-sequence oligonucleotides into the *Eco*RI and *Sac*I restriction sites of pD–*lacZ*. DNA was prepared from the library and then transformed into yeast strain KY320. 7900

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press



yeast transformant colonies were patched onto X-gal indicator plates containing either glucose or galactose, and those plates were incubated 3–5 days until blue color developed in the control strains. Of these transformants, 81 produced reproducibly blue patches. Plasmids were rescued from those transformants, as described (Hoffman and Winston 1987), and DNA inserts were sequenced.

*RNA analysis*

Procedures for isolation of total yeast RNA, preparation of oligonucleotide DNA probes 5′-end labeled with $^{32}P$, DNA–RNA hybridization, S1 nuclease digestion, and analysis of products by gel electrophoresis were performed essentially as described previously (Chen et al. 1987; Chen and Struhl 1988). For all cases, cells were grown in YP (1% yeast extract, 2% peptone) media containing either 2% glucose or 2% galactose.

*Phenotypic analysis*

Plates used for β-galactosidase assays contained 0.17% Difco yeast nitrogen base, 0.2% ammonium sulfate, 100 mM potassium phosphate (pH 7.0), 2% glucose, 1.5% Difco Bacto Agar, 40 μg/ml X-gal, and amino acids at appropriate concentrations. Single colonies of yeast cells containing the desired plasmids were patched onto indicator plates and examined visually for the intensity of blue color after 2–3 days . All colonies were compared to cells containing G17, D (−); G17, R9 (+/−); G17, R4 (+); G17 (++); and G17, R6 (+++). Liquid assays for β-galactosidase were done essentially as described (Guarente 1983) on cells grown in YP media containing either 2% glucose, 2% galactose, or 2% raffinose in addition to 2% glucose or galactose) to mid-exponential growth phase ($OD_{600} \cong 1–2$). Enzyme activities were determined in triplicate and normalized to the $OD_{600}$ of the cultures; the values are accurate to ±25%.

*Computer analysis*

The computer program ALIGN (Oliphant 1989) was used to analyze our sequence data. This program treats entered sequences as strings of continuous information, and searches for the occurrence of two to seven nucleotide patterns or words at a frequency higher than that predicted by random occurrence.

*In vitro transcription assays*

The transcription assay substrates were created by fusing the 0.4-kb *Bst*UI–*Sma*I fragment containing the 'G-less cassette' of pML($C_2$AT)19 (Sawadogo and Roeder 1985) to a synthetic oligonucleotide containing the $T_R$ element of the yeast *his3* gene flanked *Eco*RI and *Sac*I restriction sites (Chen and Struhl 1988) and ligating the resulting fragment into pUC19 as described elsewhere (C.R. Wobbe and K. Struhl, in prep.). The resulting molecule is equivalent to pML($C_2$AT)19, except that the *his3* $T_R$ element is at the precise position normally occupied by the TATA sequence of the adenovirus major late promoter. Oligonucleotides corresponding to some randomly selected $T_R$ oligonucleotides described above were then inserted between the *Eco*RI and *Sac*I sites.

HeLa cell transcription factors were partially purified as described in detail in C.R. Wobbe and K. Struhl (in prep.). Briefly, HeLa cell crude nuclear extract was fractionated initially by step-gradient chromatography over phosphocellulose to generate fractions containing TFIIA, TFIIB+E+F, and TFIID, as described previously (Reinberg and Roeder 1987). TFIIA was purified further by chromatography on DEAE–cellulose and Sephacryl S-200; TFIIB, by chromatography on DEAE–cellulose and concentration on phosphocellulose; and TFIIE+F, by chromatography on DEAE–cellulose. Yeast TFIID-complementing activity was purified by a modification of the protocol of Buratowski et al (1988) using chromatography on heparin–agarose, DEAE–cellulose, Mono S, and Sephacryl S-200 columns. The latter column yielded a single species of TFIID complementing activity eluting at a position expected for a globular protein with $M_r$ 26,000.

Transcription assays (20 μl) were carried out as described previously (Sawadogo and Roeder 1985), except that 25 μM [α-$^{32}P$]UTP was used as the label. Reactions contained 400 ng of DNA substrate, 2.3 μg of TFIIA, 0.18 μg of TFIIB, 1.7 μg of TFIIE/F, 5 units of calf thymus RNA polymerase II (for unit definition, see Reinberg and Roeder 1987), and 30 ng of yeast TFIID. Following a 60-min incubation at 30°C, reactions were terminated and analyzed by denaturing polyacrylamide gel electrophoresis as described elsewhere (C.R. Wobbe and K. Struhl, in prep.). The level of TFIID-dependent transcription initiated from the correct site was quantitated by densitometry, using the level of transcription initiated from a site within the G-less cassette as an internal control. Transcription from this internal site depends on TFIID, occurs 25–30 bp downstream from a TATA-like sequence, and is not affected by the amount of transcription from the correct site (C.R. Wobbe and K. Struhl, in prep.). It is relatively inefficient, however, probably because the sequence of the putative TATA element differs from the consensus.

## Acknowledgments

We thank members of the laboratory for useful comments throughout the work and on the manuscript. This research was supported by an National Institutes of Health (NIH) postdoctoral fellowship to V.S., a Merck postdoctoral fellowship to C.R.W., and by grants to K.S. from the NIH (GM-30186) and from the Lucille Markey trust.

## References

Buratowski, S., S. Hahn, P.A. Sharp, and L. Guarente. 1988. Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* **334:** 37–42.

Buratowski, S., S. Hahn, L. Guarente, and P.A. Sharp. 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56:** 549–561.

Cavallini, B., J. Huet, J.L. Plassat, A. Sentenac, J.M. Egly, and P. Chambon. 1988. A yeast activity can substitute for the HeLa TATA box factor. *Nature* **334:** 77–80.

Chen, W. and K. Struhl. 1988. Saturation mutagenesis of a yeast *his3* TATA element: Genetic evidence for a specific TATA-binding protein. *Proc. Natl. Acad. Sci.* **85:** 2691–2695.

———. 1989. Yeast upstream activator protein GCN4 can stimulate transcription when its binding site replaces the TATA element. *EMBO J.* **8:** 261–268.

Chen, W., S. Tabor, and K. Struhl. 1987. Distinguishing between mechanisms of eukaryotic transcriptional activation with bacteriophage T7 RNA polymerase. *Cell* **50:** 1047–1055.

Davidson, I., J.H. Xiao, R. Rosales, A. Staub, and P. Chambon. 1988. The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. *Cell* **54:** 931–942.

Guarente, L. 1983. Yeast promoters and *lacZ* fusions designed to study expression of cloned genes in yeast. *Methods Enzymol.* **101:** 181–191.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press



———. 1988. UASs and enhancers: Common mechanism of transcriptional activation in yeast and mammals. *Cell* **52:** 303–305.

Hahn, S., S. Buratowski, P.A. Sharp, and L. Guarente. 1989. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc. Natl. Acad. Sci.* **86:** 5718–5722.

Harbury, P.A.B. and K. Struhl. 1989. Functional distinctions between yeast TATA elements. *Mol. Cell. Biol.* **9:** 5298–5304.

Helmann, J.D. and M.J. Chamberlin. 1988. Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* **57:** 839–872.

Hoffman, C.S. and F. Winston. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* **57:** 267–272.

Landschulz, W.H., P.F. Johnson, and S.L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240:** 1759–1764.

Losick, R., P. Youngman, and P.J. Piggot. 1986. *Annu. Rev. Genet.* **20:** 625–629.

Miller, J.H. 1972. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Oliphant, A. 1989. Defining genetic elements and protein structure by selecting functional sequences from random DNA. Ph.D. thesis, Harvard University, Cambridge.

Oliphant, A. R. and K. Struhl. 1987. The use of random-sequence oligonucleotides for determining consensus sequences. *Methods Enzymol.* **155:** 568–582.

Oliphant, A. R., A. L. Nussbaum, and K. Struhl. 1986. Cloning of random-sequence oligodeoxynucleotides. *Gene* **44:** 177–183.

Pfeifer, K., T. Prezant, and L. Guarente. 1987. Yeast HAP1 activator binds to two upstream activation sites of different sequence. *Cell* **49:** 19–27.

Poellinger, L., B.K. Yoza, and R.G. Roeder. 1989. Functional cooperativity between protein molecules bound at two distinct sequence elements of the immunoglobulin heavy-chain promoter. *Nature* **337:** 573–576.

Reinberg, D. and R.G. Roeder. 1987. Factors involved in specific transcription by mammalian RNA polymerase II: Purification and functional analysis of initiation factors IIB and IIE. *J. Biol. Chem.* **262:** 3310–3321.

Sakai, D.D., S. Helms, J. Carlstedt-Dele, J.-A. Gustaffson, F.M. Rottman, and K.R. Yamamoto. 1988. Hormone mediated repression: a negative glucocorticoid response element from the bovine prolactin gene. *Genes Dev.* **2:** 1144–1154.

Sawadogo, M. and R.G. Roeder. 1985. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proc. Natl. Acad. Sci.* **82:** 4394–4398.

Struhl, K. 1986. Constitutive and inducible *Saccharomyces cerevisiae* promoters: Evidence for two distinct molecular mechanisms. *Mol. Cell. Biol.* **6:** 3847–3853.

———. 1989. Molecular mechanisms of transcriptional regulation in yeast. *Annu. Rev. Biochem.* **58:** 1051–1077.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

# A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation.

V L Singer, C R Wobbe and K Struhl

*Genes Dev.* 1990, **4:**
Access the most recent version at doi:10.1101/gad.4.4.636

**References** This article cites 24 articles, 8 of which can be accessed free at:
**http://genesdev.cshlp.org/content/4/4/636.full.html#ref-list-1**

**License**

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**


Copyright © Cold Spring Harbor Laboratory Press