The EMBO Journal vol.4 no.12 pp.3273–3280, 1985

Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element

Wei Chen and Kevin Struhl

Department of Biological Chemistry, Harvard Medical School, Boston, MA 02115, USA

Communicated by H.Pelham

We present evidence suggesting that accurate mRNA initiation in yeast cells, unlike their higher eukaryotic counterparts, is determined primarily by specific sequences downstream from the TATA element. First, changing the distance between the *his3* TATA element and the initiation region does not affect the sites of initiation or the level of RNA. Second, reciprocal *his3-ded1* and *ded1-his3* hybrid promoters containing the upstream and TATA elements of one gene fused to the mRNA coding region of the other gene initiate transcription at sites defined by wild-type mRNA coding sequences, not by the distance from the TATA element. Third, when the *his3* or *ded1* promoter region is fused to position +2 of the *his3* gene, transcripts initiated from a position equivalent to +1 are not observed. The results also suggest that the spacing between the TATA element and initiation site is relatively flexible; distances ranging from 40 to 90 bp appear to be functionally acceptable.

Key words: yeast promoter/eukaryotic transcription/RNA polymerase II/TATA element/mRNA synthesis

Introduction

In eukaryotic cells, transcription depends on the presence of upstream and TATA promoter elements (Grosschedl and Birnstiel, 1980; Moreau *et al.*, 1981; Banerji *et al.*, 1981; Struhl, 1981, 1982; Fromm and Berg, 1982; McKnight and Kingsbury, 1982; Guarente and Mason, 1983). Upstream elements are thought to provide promoter specificity. They usually determine a promoter's particular regulatory properties, and different classes of genes contain upstream elements with different DNA sequences. When upstream regions of different promoters are swapped, the hybrid promoters are regulated according to the upstream elements. The TATA element has been regarded as a general promoter element. This sequence (consensus TATAAA) has been found upstream from the known mRNA initiation sites of all sequenced protein coding genes with the exception of a few viral genes (Baker *et al.*, 1979; Breathnach and Chambon, 1981; Cattaneo *et al.*, 1983).

The current view on the function of the TATA element is that it is involved in the mechanism directing accurate initiation of transcription by RNA polymerase II. Specifically, two types of experiments suggest that it acts as a selector for transcriptional initiation sites by measuring a fixed number of base pairs downstream (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981; Mathis and Chambon, 1981; Dierks *et al.*, 1981; Ghosh *et al.*, 1981; Kamen *et al.*, 1982; McKnight and Kingsbury, 1982). In the first type, deletions that remove sequences downstream from a TATA element including the normal initiation sites result in mRNA initiated from a new site 20–30 bp downstream from the TATA sequence. In the second type, deletions of the TATA element itself lower the level of transcription, and they also result in heterogeneous initiation at the otherwise rarely used sites in the normal promoter.

In the unicellular eukaryote *Saccharomyces cerevisiae* (baker's yeast), promoter structure is similar to that of higher eukaryotes (reviewed by Guarente, 1984; Struhl, 1985). In particular, TATA sequences are observed upstream from all sequenced protein coding genes. Moreover, the functional importance for yeast TATA elements has been demonstrated in several cases (Struhl, 1982, 1984; Guarente and Mason, 1983; Siliciano and Tatchell, 1984), which has led to the tacit acceptance of a role analogous to that proposed for higher eukaryotes. However, unlike the situation for higher eukaryotic genes, the distance between the TATA sequences and the initiation sites varies, although it should be noted that in individual yeast promoter regions there are frequently several TATA-like sequences, and it is generally not clear which, if any, of these sequences are functionally important. Nevertheless, these observations question the generally proposed role of the TATA element in transcription initiation.

In yeast, the only experiments that address the selection of initiation sites involve deletion mutants that remove the *matα1* TATA box (Siliciano and Tatchell, 1984). The wild-type gene initiates transcription from about seven different sites. In TATA deletion strains transcription is decreased, and the observed initiation comes primarily from the more downstream sites. This result is hard to interpret because there may be several TATA-like sequences only some of which are deleted. In addition, it is difficult to assess the contributions of the new sequences that are in the same position as the deleted TATA box. In more general terms, in a situation where a critical sequence has been removed with serious functional consequences, it is impossible to know whether other aspects of the promoter are behaving properly.

Here we construct mutant and hybrid promoters that retain all the important elements but alter the spacing between the TATA element and the initiation sites. To analyze their transcriptional properties *in vivo*, we introduce DNA molecules back into the yeast genome in single copy at the natural chromosomal locus. The results suggest that, in contrast to the situation in higher eukaryotes, the TATA element in yeast does not 'select' initiation sites by measuring a fixed distance from it. Instead, our evidence suggests that an 'initiator' element, distinct from the upstream and the TATA elements, contains the signal for determining where transcription begins.

Results

Varying the spacing between the his3 TATA element and initiation sites

The promoter region of the *his3* gene has been extensively studied by deletion analysis (Struhl, 1981, 1982, 1984). By functional criteria, the TATA element is localized between nucleotides −55 and −35 from the transcription initiation site. If the TATA ele-

© IRL Press Limited, Oxford, England

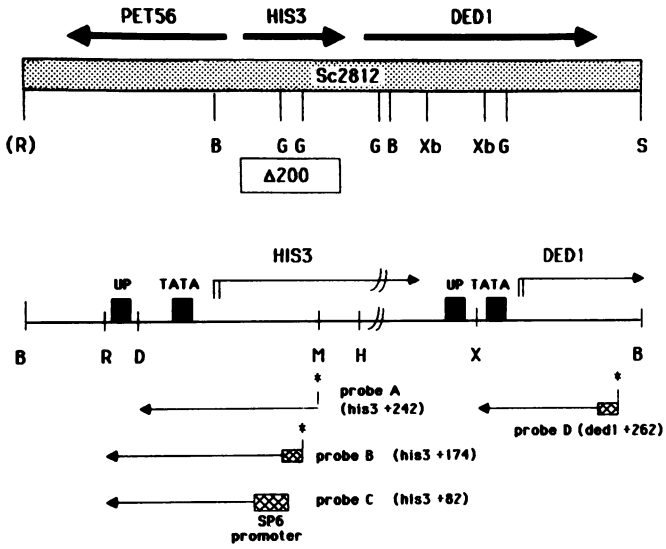

Fig. 1. Outline of the *his3* and *ded1* region. Sc2812 (shaded bar) is a derivative of the 6.1-kb *Eco*RI-*Sal*I *his3* fragment Sc2605 with the *Eco*RI site mutated. The black arrows above the hatched bar indicate the locations and orientations of the *his3*, *ded1*, and *pet56* transcripts, and the vertical lines represent the positions of restriction sites (Struhl and Davis, 1981b; unpublished data). The Δ200 allele removes the entire *his3* structural gene (nucleotides −181 to +874) but retains the *ded1* promoter (unpublished results). The expanded view of the *his3* and *ded1* promoter regions diagrammed below includes the locations of transcripts, upstream (UP) and TATA promoter elements (black boxes), and hybridization probes. 'R' is an *Eco*RI site introduced at −134 of the *his3* promoter region by *Bal*31 treatment and linker insertion. Probe A is a *Msp*I-*Dde*I DNA segment that is 5′ end-labelled and strand-separated. Probes B and D are obtained by extending an 5′ end-labelled primer starting from +174 of *his3* and +262 of *ded1* on appropriate M13 templates followed by cleavage with *Eco*RI and *Xho*I (see Materials and methods). Probe C is a SP6 RNA polymerase-generated uniformly labelled anti-message RNA probe. Abbreviation for restriction sites are R (*Eco*RI), B (*Bam*HI), G (*Bgl*II), Xb (*Xba*I), S (*Sal*I), D (*Dde*I), M (*Msp*I), H (*Hind*III), X (*Xho*I).

ment selects initiation sites by measuring a fixed distance downstream, then changing the distance between TATA element and the initiation sites should result in changes of the sites of initiation.

To evaluate the spacing effect we examined the following three alleles, all of which derive from Sc2812, a 6.1-kb fragment with *his3* centrally located (Figure 1). *His3*-Δ18 is an *Eco*RI linker insertion at position −35; thus it extends the distance between the TATA and initiation regions by 8 bp. In *his3*-Δ19 the region between −35 and −24 is replaced by an *Eco*RI linker, which results in a net deletion of 3 bp. Similarly *his3*-Δ36, which deletes sequences from −35 to −11, has a net deletion of 16 bp between the TATA and initiation region (Figure 2a). By appropriate genetic manipulations, the *his3*-Δ 200 allele of strain KY117 was replaced by the 'spacing' mutant DNAs (see Materials and methods). As shown by standard hybridization analysis (Figure 2b), the resulting strains contain one copy of the 'spacing' mutations precisely at the normal *his3* chromosomal locus.

To analyze the initiation pattern of the 'spacing' mutants, we isolated total RNA and mapped the 5′ ends by two different methods, both of which are quantitative and qualitative assays of transcription (see Materials and methods). In the first method yeast RNA was hybridized to completion with an excess of a single-stranded *his3* DNA probe that was 5′ end-labelled with ^{32}P at position +242, and then digested with S1 nuclease. In the second method RNA was hybridized to completion with an excess of complementary RNA probe (*his3* sequences +88 to −134) and then treated with RNaseA.

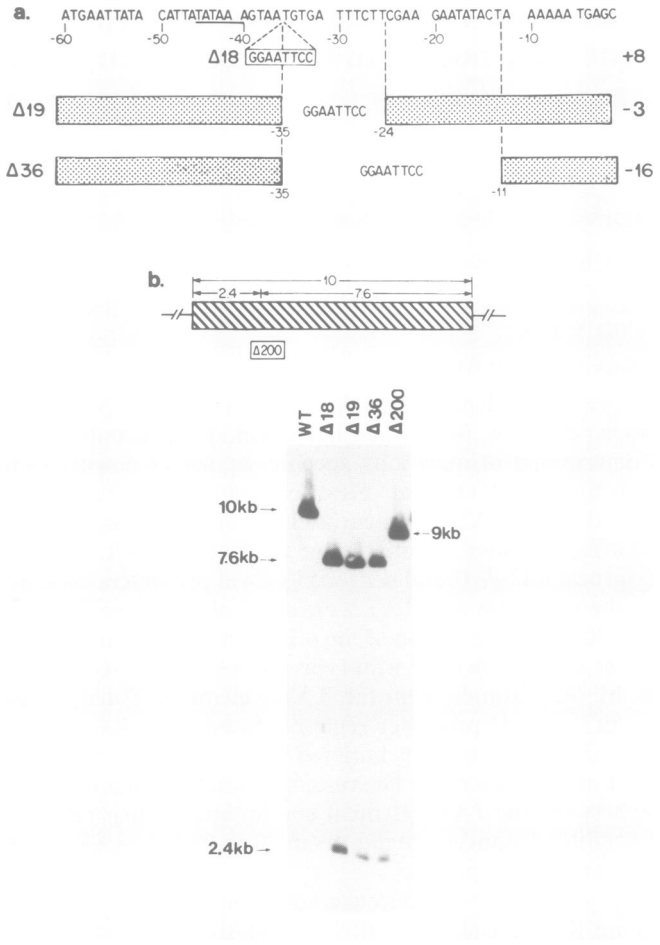

Fig. 2. Structures of the 'spacing' mutants. **Part a** shows the nucleotide sequence of the relevant part of the *his3* promoter region (Struhl and Davis, 1981a) and the structures of the 'spacing' mutants. All these insertions and deletions occur between the TATA box (underlined) and the +1 RNA initiation site; the spacing changes are listed on the right side. **Part b** shows the structure of the chromosome region after gene replacement in terms of *Eco*RI restriction fragments; the desired event should result in fragments of 7.4 kb and 2.4 kb. **Part c** shows hybridization analysis that confirms the replacement of the *his3*-Δ200 allele with the 'spacing' mutants. Genomic DNAs were prepared and digested with *Eco*RI and hybridized with ^{32}P labelled pUC8-Sc2605 DNA.

The wild-type *his3* gene has two major start sites of equal intensity mapped at +1 and +12, as well as a minor site at +22 (Struhl and Davis, 1981a; Struhl, 1984; Figure 3a,b). Figure 3a shows the results obtained from the 5′ end analysis by S1 nuclease mapping. In all three mutants, the pattern of initiation sites is indistinguishable from that of wild-type strains. Figure 3b shows an analysis of the same RNAs by the RNase method. Again, both the pattern of initiation sites and the level of RNA are the same as for the wild-type *his3* gene. Thus all the 'spacing' mutants behave as wild-type alleles in terms of transcription.

Fusions with the promoter and initiation regions 'swapped' between two different yeast genes

The transcription patterns of the 'spacing' mutants described above suggest that the TATA element does not direct the sites of transcription initiation by measuring a fixed distance from itself. To determine whether the initiation site is determined by the sequences downstream from the TATA region, we designed promoter swap experiments by constructing reciprocal fusions between the *his3* and *ded1* genes. Specifically, the upstream and TATA elements of one gene were fused to the initiation region

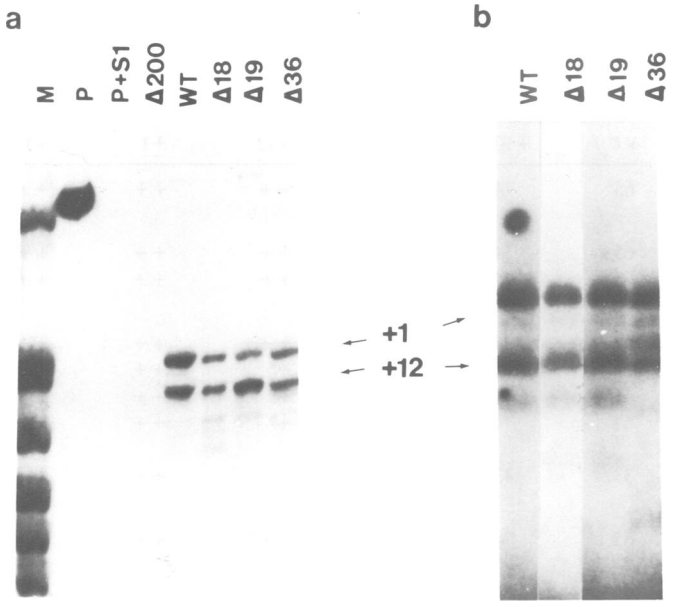

Fig. 3. Analysis of RNAs from 'spacing' mutants. In **part a**, total RNAs, from the strains indicated in the lanes were hybridized to the 5′ end-labelled DNA probe (probe A of Figure 1) and treated with S1 nuclease. Lanes representing pBR322 *Msp*I fragments as size markers (M), untreated probe (P), and probe treated with S1 nuclease (P + S1) are included as controls. In **part b**, the same RNA preparations were hybridized to the single-stranded RNA probe generated by SP6 RNA polymerase (probe C, Figure 1), and then digested with RNaseA. Equal amounts of RNA (determined by A_{260} and by independent hybridization to *ded1* probe D) were examined in each lane. The positions of +1 and +12 transcripts are shown.

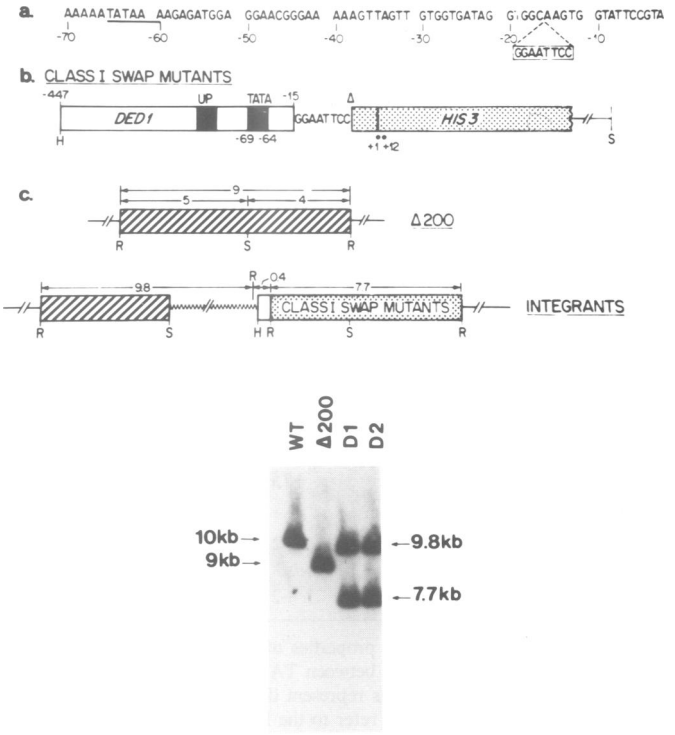

Fig. 4. Structures of the class I 'swap' mutants. **Part a** shows the *ded1* promoter sequence with the putative TATA element underlined. An *Eco*RI linker insertion is shown in the framed box at −15 with respect to the *ded1* +1 RNA initiation site. **Part b** diagrams the structures of the class I swap mutants. The *ded1* promoter region extending from −447 to −15 (open box) is fused via an *Eco*RI linker to various segments containing the *his3* initiation region and structural gene (shaded box). Δ indicates the *his3* fusion break points which are −11 (*his3*-D1), −24 (*his3*-D3), −31 (*his3*-D4) and −35 (*his3*-D2). **Part c** diagrams the genomic structures expected from single copy integration of class I swap mutants into Δ200. The hatched box represents the *Eco*RI fragment of Δ200; the wavy line represents the YIp5 vector sequence, and the open box represents the class I mutant sequences. The autoradiogram shows analysis of *Eco*RI cleaved DNAs using pUC8-Sc2605 as a hybridization probe.

of the other gene. The *ded1* gene, which is located adjacent to and downstream from *his3* gene (Figure 1), is required for cell growth (Struhl, unpubished result). Although the *ded1* TATA region has not been defined by functional analysis there is only one DNA sequence that even remotely resembles a TATA box. This sequence (TATAAA) is located between nucleotides −60 to −65 with respect to the two major initiation sites at +1 and +10 (Struhl, unpublished results; Figure 4a). In these derivatives the spacing between the TATA element and initiation site ranges between 18 and 104 bp (Table I). In this paper such distances will be measured from the upstream-most T residue of the TATAAA sequence.

The class I 'swap' fusions have the promoter region of *ded1* (nucleotides −447 to −15) fused to positions −11, −24, −31 and −35 of the *his3* gene (*his3* alleles D1, D3, D4, D2 respectively). To test whether the transcriptional pattern of the wild-type *his3* gene is retained when the *his3* initiation region is under foreign control, these DNAs were integrated into the KY117 genome in single copy at the *his3* locus (see Materials and methods and Figure 4c) and then analyzed by the SP6 and S1 nuclease mapping techniques. Figure 5a shows that the levels of *his3* RNA in the fusions are five times higher than that for the wild-type *his3* gene. These levels are comparable with the wild-type level of *ded1* RNA. This indicates that, as expected, the *ded1* promoter controls the level of *his3* transcription.

In the wild-type *his3* gene the TATA element is located 45 nucleotides away from the +1 initiation site. In *his3*-D1, where the *ded1* TATA element is 69 bp upstream from the +1 initiation site, the initiation pattern is exactly the same as for the *his3* TATA element now 93 bp upstream from the initiation site, the level of +1 initiation is normal. However, the level of +12 initiation is reduced, and there are also novel initiations from −3 to −6. The same results are observed with *his3*-D3 and *his3*-D4 in which the distances are 82 and 89 bp. This suggests that although there is no fixed 'spacing' relationship by which the TATA element directs initiation there is some limitation in the distance over which it can act properly.

To corroborate the conclusion that the initiation signal is carried by sequences downstream from the TATA region, we replaced the normal *his3-ded1* chromosomal region with the reciprocal fusion containing the *ded1* initiation region under control of the *his3* upstream sequence and TATA element (see Materials and methods and Figure 6a,b). Figure 6c shows that RNA transcribed by the class II fusion has a similar initiation pattern to that of the wild-type *ded1* gene, even though the *his3* TATA element is located only 33 bp away from the +1 initiation site compared with the 65 bp distance found in the wild-type *ded1* gene. Both normal initiation sites are observed, although the +1 site appears to be used ~2-fold less efficiently. This confirms the conclusion that the initiation region contains the signal for where transcription should begin. Interestingly, the level of RNA is higher than that observed for the wild-type *his3* gene and it is almost up to the wild-type levels of *ded1* RNA.

Mutants lacking the initiation region

If the initiation region contains the signal for accurate transcrip-

Table I. Structural and transcriptional properties of wild-type and mutant alleles of *his3* and *ded1*

Alleles	Endpoints		Spacing				Patterns of Initiation			
			his3		*ded1*		*his3*		*ded1*	
			+1	+12	+1	+10	+1	+12	+1	+10
Wild-type	–	–	45	56	65	74	++	++	++	++
Spacing Mutants										
his3-Δ18	−35	−35	53	64	65	74	++	++	++	++
his3-Δ19	−35	−24	42	53	65	74	++	++	++	++
'Swap' Mutants	*ded1*	*his3*								
his3-D1	−15	−11	69	80	65	74	++	++	++	++
his3-D3	−15	−24	82	93	65	74	++	+[a]	++	++
his3-D4	−15	−31	89	100	65	74	++	+[a]	++	++
his3-D2	−15	−35	93	104	65	74	++	+[a]	++	++
	his3	*ded1*								
ded1-H1	−35	−15	–	–	33	42	–	–	++	++
'Initiator' Mutants										
his3-Δ70	−35	+2	18[c]	29	65	74	–	++[b]	++	++
	ded1	*his3*								
his3-D5	−15	+2	58[c]	69	65	74	–	++[b]	++	++

The structural and transcriptional properties of wild-type and mutant alleles of *his3* and *ded1* determined in Figures 2−7 are summarized. Spacing mutants represent deletions and insertions between TATA element and the normal initiation sites; swap mutants represent hybrids between *his3* and *ded1* promoter and initiation regions; initiator mutants represent deletions that remove part of the initiation region. The end points are measured with respect to the normal *his3* or *ded1* +1 initiation site, and they refer to the last nucleotide that is present before the *Eco*RI linker joint. The spacing indicates the distance between the most upstream T of the relevant TATA element (see text) and the *his3* or *ded1* initiation sites. For each normal initiation site, transcriptional levels are indicated as follows: '++' indicates the level that is expected from the upstream promoter element; '+' represents detectable, but lower than expected levels; '−' indicates very low or undetectable RNA levels.

[a]There are also aberrant initiations at −3 to −6.

[b]The +22 site, normally a minor *his3* initiation site, is utilized at equal or nearly equal efficiency as the +12 site.

[c]Since the +1 nucleotide is removed in these two derivatives, these numbers refer to the distances corresponding to the equivalent +1 position (i.e., to the C residue of the *Eco*RI site).

tion, then mutants that disrupt this region should not initiate properly. Figure 7 shows the S1 analysis of two derivatives in which the *his3* +1 initiation site is deleted. In *his3*-Δ70, the region between −35 and +1 is replaced by an *Eco*RI linker; this allele is similar to the 'spacing mutants' described in the first section of the results. In *his3*-D5, the identical *his3* mRNA coding sequences have been fused to the *ded1* promoter elements as desribed for the class I swap mutants. For both of these derivatives, although the expected level of transcription is observed, there is hardly any initiation at a position that would correspond to +1. Instead, transcription is initiated at the normal +12 site as well as the +22 site which is utilized inefficiently in the wild-type *his3* gene.

Discussion

Yeast transcriptional initiation sites are determined primarily by an 'initiator' element, not by the distance from the TATA box

In principle, discrete mRNA start sites could be determined by specific 'initiator' sequences or by the distance from a required promoter element. In *Escherichia coli*, transcription is initiated at sites that are located a precise distance from the required promoter elements, the −10 and −35 sequences (reviewed by Hawley and McClure, 1983). Similarly, it has been suggested by many workers that mammalian mRNAs begin at sites that are located a fixed distance from the TATA element. In contrast, we present three lines of evidence suggesting that in yeast accurate mRNA transcription depends primarily on specific sequences located near the initiation region itself (see Table I).

(i)The *his3* spacing mutants behave indistinguishably even though the distance between the TATA element and the initiation region varies over a range of 24 bp. (ii) The promoter swap experiments indicate that accurate initiation occurs even when transcription depends on 'foreign' promoter elements that are located at different positions from the +1 initiation site. The initiation pattern of *his3*-D1 (spacing 69 nucleotides) is identical to the wild-type *his3* gene (spacing 45 nucleotides) and the initiation pattern of *ded1*-H1 (spacing 33 nucleotides) is indistinguishable from that of the wild-type *ded1* gene (spacing 65 nucleotides). Moreover, all seven derivatives containing the *his3* initiation region produce wild-type levels of the +1 transcript even though the TATA elements are located 33-93 bp away. Finally, the aberrant patterns observed in *his3*-D2, D3, and D4 are similar even though the relevant distances vary over a range of 11 bp. (iii) Both derivatives that lack the *his3* +1 site (*his3*-Δ70 and *his3*-D5) fail to initiate at a position equivalent to +1. This cannot be explained by a 'closeness' effect because in *his3*-D5, the +1 equivalent site is 58 bp away from the TATA element. Moreover, both derivatives produce the same transcriptional pattern even though the TATA elements are located at different distances from the observed +12 and +22 start sites.

Indirect support for this view comes from the observation that for the yeast genes whose DNA sequences and RNA start sites have been determined, TATA-like sequences are located at variable distances from their initiation sites. This is unlike the situation in higher eukaryotic genes in which the TATA element is located a relatively constant distance away (~25−30 bp).

Distance limitations between TATA element and initiation sites

Although there is no precise spacing relationship between the TATA element and the transcriptional initiation sites our results indicate that there are limits to the distance over which a TATA

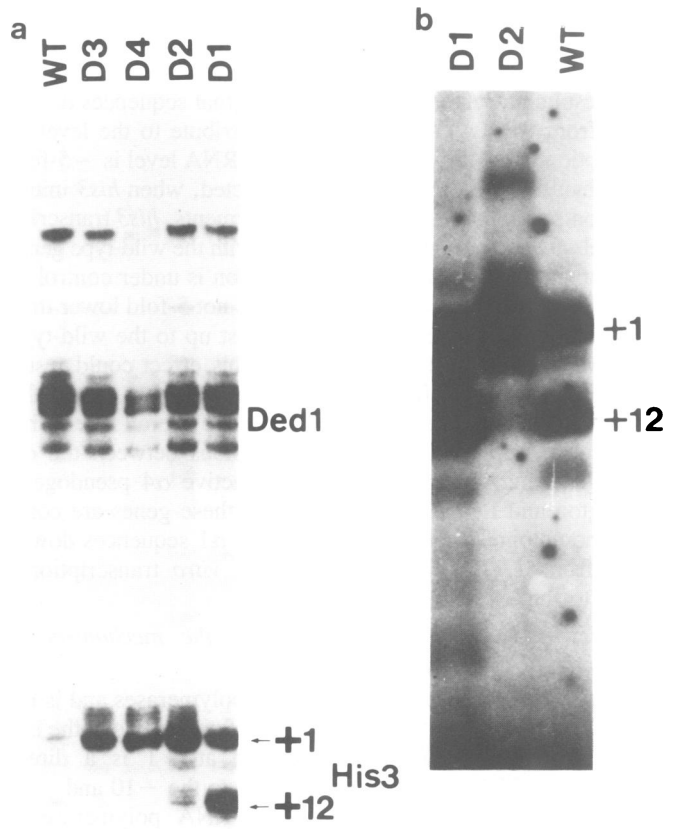

Fig. 5. Analysis of RNA from class I swap mutant. **Panel a** shows 5′ end mapping by S1 nuclease analysis using a mixture of *his3* probe B and *ded1* probe D (see Figure 1). The strains analyzed include swap mutants *his3*-D1 through *his3*-D4 and well as the wild-type control (WT). The positions of the *his3* and *ded1* transcripts are indicated. **Panel b** shows analysis of some of the same RNAs by the RNA:RNA hybridization method using probe C.

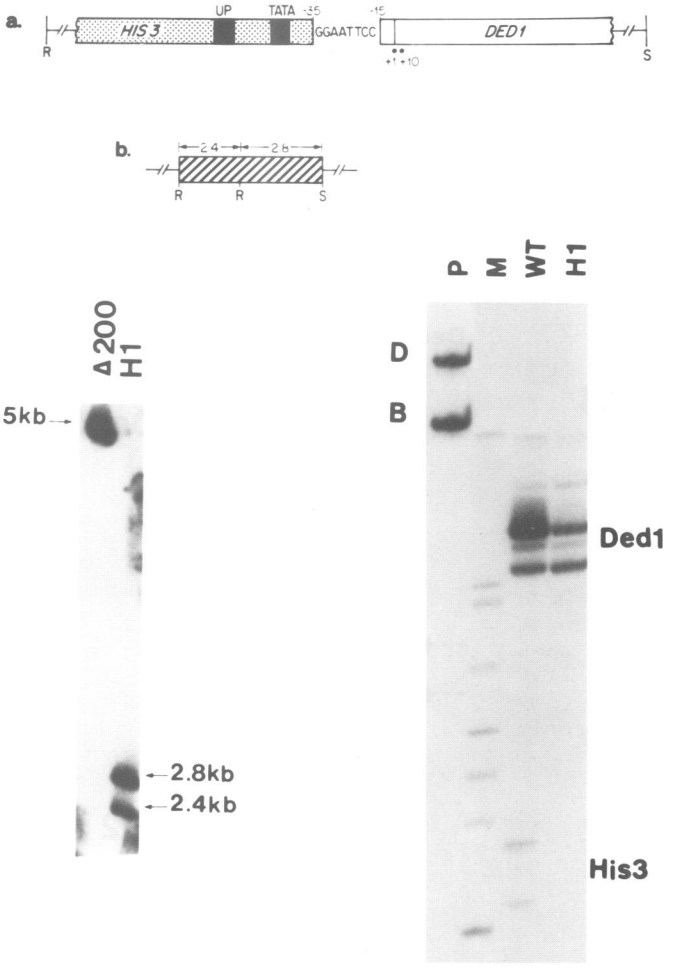

Fig. 6. Structure of the class II mutant *ded1*-H1. **Part a** diagrams *ded1*-H1 which was constructed by fusing the *his3* promoter down to −35 (shaded box) to −15 of the *ded1* gene (open box) via an *Eco*RI linker. The *his3* promoter elements are indicated as black boxes and the normal *ded1* initiation sites are indicated as closed circles. **Part b** shows the expected genomic structure of this allele following gene replacement. The autoradiogram on the left represents the hybridization of the pUC8-Sc2605 probe to genomic DNAs cleaved with *Eco*RI and *Sal*I. The autoradiogram on the right shows an analysis of RNAs produced by *ded1*-H1. The experiment was performed identically to that shown in Figure 5. The lack of *his3* hybridization in the *ded1*-H1 lane is due to the fact that this allele deletes the entire coding region.

element can act. The most informative examples are represented by several of the class I fusions (D2, D3, D4) in which initiation at +12 is reduced compared with initiation at +1, and aberrant initiation around −3 to −6 is also evident. This effect cannot be due to the *ded1* promoter because the wild-type *his3* pattern is oberved in *his3*-D1, and because the overall transcription level is not affected. Moreover, the absence of −3 to −6 transcripts in the wild-type *his3* gene cannot be explained by distance effects because these sites are located extremely close to the normal +1 site. Since the aberrant transcripts are observed only when the level of +12 transcripts are reduced, it seems likely that the two effects are related. A possible explanation is that the *ded1* TATA element is too far from the +12 site for efficient transcription. In *his3*-D2, the distance is 93 bp away from the +1 site and 104 bp from the +12 site; in *his3*-D3 the distances are 82 and 93 bp. We do not know why a distance of 93 bp can result in transcription that is efficient (+1 initiation in *his3*-D2) or inefficient (+12 initiation in *his3*-D3). Perhaps a distance of 93 bp is just at the functional boundary line such that secondary considerations influence the initiation efficiency. In any event, it appears that the maximal effective distance between the TATA element and initiation site is ~90 bp.

The minimal distance between the TATA element and initiation site is more difficult to define. In *his3*-Δ36 the distance between the +1 initiation site and the sequence TATAAA is only 29 bp, but it is not clear whether this sequence serves a functional role in this case. Perhaps it is too close to the initiation site, and the TATA function is determined by TATATA (31 bp away) or TATACA (38 bp away), sequences which lie within the region that is critical for transcription (nucleotides −35 to −55 in wild-type *his3*). Nevertheless, it appears that a separation of as few as 40 bp is probably adequate for transcription, and the minimal distance may be even less. Thus, a conservative interpretation of these results suggests that the TATA element can be located anywhere between 40 and 90 bp from the initiation site with minimal functional consequences.

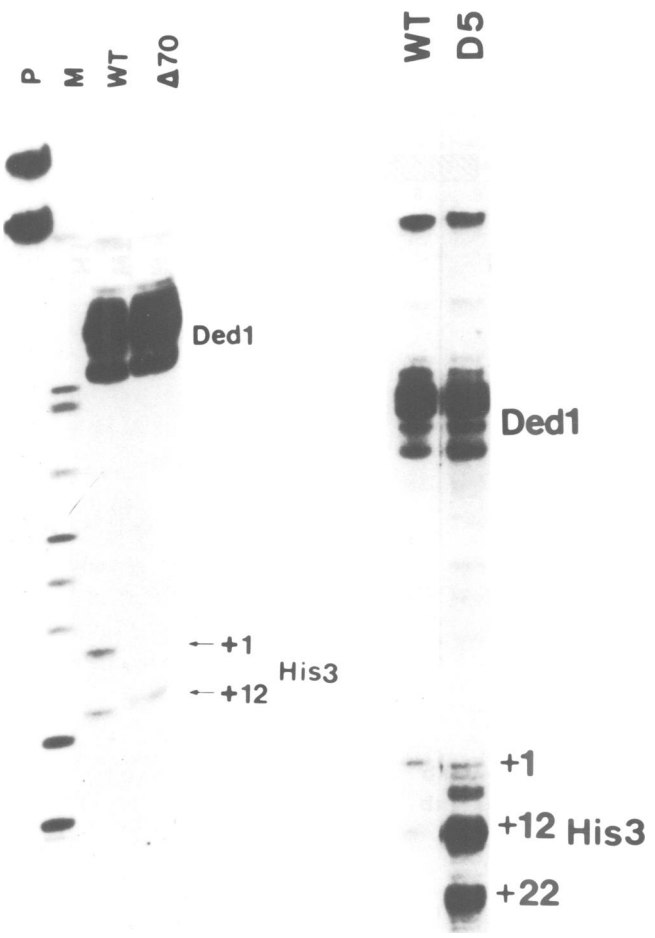

Fig. 7. Analysis of RNAs produced by derivatives lacking the *his3* +1 initiation site. S1 nuclease analysis of *his3*-Δ70 and *his3*-D5 was performed as described in Figure 5. Δ70 is a deletion between −35 and +1 of the *his3* gene with an *Eco*RI linker insertion at the break point. *His3*-D5 is a derivative that resembles the class I swap fusions except that the downstream *his3* deletion end point (+2) is identical to that of *his3*-Δ 70. Due to the structure of the probe, the faint hybridization signal at the +2 position represents all transcripts that initiate upstream from this point.

Properties of the initiator element

Although our results identify an initiator element which is distinct from the TATA box, they do not localize it precisely. In formal terms, the *his3* initiator is defined by sequences downstream of nucleotide −11, and the *ded1* initiator is defined by sequences downstream from −15. In the case of *his3*, the region between −11 and +1 is necessary for initiation in the vicinity of +1; this may be due simply to the fact that this region contains the +1 initiation site. For the *his3* +12 transcript, proper initiation does not require sequences upstream of nucleotide +2. Thus, within the limits of our deletions, sequences >10−15 bp upstream from mRNA start sites are not necessary for accurate initiation.

The main conclusion from all these cases is that the initiator element is located close to the initiation site itself. Although the DNA sequence specificity of an initiator element is not known, it cannot be defined solely by the initiating nucleotide. The TATA element appears to permit initiation anywhere within a 50-bp region, yet particular sites are chosen (probably 2 or 3 is a typical number); a single nucleotide occurs too frequently for such selectivity. At this time, the basis for site selection is obscure, although it should be noted that initiation regions frequently have short stretches of A residues in the coding strand just upstream from the RNA start site.

The results presented here also suggest that sequences downstream from the TATA element can contribute to the level of transcription. In wild-type cells, the *ded1* RNA level is ~5-fold higher than the *his3* RNA level. As expected, when *his3* initiation is controlled by the *ded1* promoter elements, *his3* transcription levels are 5-fold higher as compared with the wild-type gene. However, in *ded1*-H1 where *ded1* initiation is under control of the *his3* promoter, the level of the RNA is not 5-fold lower than the wild-type *ded1* level, but rather almost up to the wild-type *ded1* level. Although it is possible that this effect could result from sequences at the fusion point of *ded1*-H1, we prefer the idea that the initiation region can influence the overall level. This suggestion is supported by promoter fusions between the α1 globin gene and the transcriptionally inactive α4 pseudogene (Talkington and Leder, 1982). Although these genes are completely homologous in the TATA region, α1 sequences downstream from TATA elements restore *in vitro* transcriptional activity to the α4 pseudogene.

Inferences and speculations concerning the mechanism of transcriptional initiation

Since transcription is catalyzed by RNA polymerases and is initiated at discrete sites, there must be spcificity between the enzyme and DNA. In *E. coli*, initiation at +1 is a direct consequence of RNA polymerase binding to the −10 and −35 promoter sequences. What does yeast RNA polymerase II recognize to initiate transcription at the proper site?

It is extremely unlikely that RNA polymerase interacts directly with upstream promoter elements. First, these elements act at long and variable distances as well as both orientations with respect to the initiation site (Guarente and Hoar, 1984; Struhl, 1984). Second, the nucleotide sequences of upstream elements are different for individual genes. Third, the *gal* upstream element is recognized by the *gal4* protein (Bram and Kornberg, 1985; Giniger *et al.*, 1985), and the *his3* upstream regulatory sites interact with the *gcn4* protein (Hope and Struhl, 1985).

From the results presented here, we suggest that RNA polymerase does not specifically recognize the TATA element. The main evidence is that the distance between the TATA element and the initiation site can vary over a range of at least 50 bp. This observation eliminates prokaryotic models, and it indicates that if RNA polymerase binds to the TATA sequence it must 'stretch' or 'move' to the initiation site in a non-transcriptional mode. This suggestion is supported by the existence of mammalian and *Drosophila* proteins that bind to TATA sequences and act as transcription factors *in vitro* (Davison *et al.*, 1983; Parker and Topol, 1984).

Instead, we propose that specificity between RNA polymerase and promoter DNA is mediated via the initiator element. Such specificity could reflect sequence preferences of the RNA polymerase itself, or it could be due to the binding of an initiation factor(s), such as prokaryotic σ factors, which guide the polymerase to the correct start site. By either model, such interactions could explain why the initiator element can contribute to the overall level of transcription.

If the primary interaction between the RNA polymerase II and the promoter of a gene is at the initiator region, what is the role of the other promoter elements? Our suggestion is that the enzyme does not interact with DNA in the inert nucleosomal form of chromatin, but rather recognizes an 'active' structure created

by the proteins bound to the upstream and TATA elements. In this view, the enzyme recognizes a 'window' near this complex and binds just downstream from the TATA element. The size of this window would correspond to the variability in spacing between the TATA element and the initiation site. The initiator element would lie within this window, and it would represent the particular sequences preferred by RNA polymerase II or an initiation factor.

Materials and methods

Microbiological techniques

The yeast strains used were KY117 (*a ura3-52 trpl-Δ1 lys2*-801am *ade2*-101oc *his3*-Δ200) and KY114 (same as KY117 except *HIS3*$^+$) and they were propagated as described previously (Struhl, 1984). DNAs were introduced into lithium chloride-treated yeast cells as described by Ito *et al.* (1983). Hybrid DNA molecules were propagated as described previously using *E. coli* strains EQ82 (for plasmid DNAs) or JM101 (for M13 derivatives).

Construction of spacing and swap mutants

The procedures for constructing hybrid DNA molecules have been described previously (Struhl, 1983). The yeast vectors used in this study, YIp5 (Struhl *et al.*, 1979) and YRp14 (Struhl, 1983) contain the *URA3*$^+$ selectable marker; YRp14 also contains the *SUP11*° gene that permits selection for excision events by virtue of resistance to osmotic stress. Most of the DNA molecules are derived from Sc2605, a 6.1-kb DNA fragment that contains the entire *pet56, his3, ded1* region (Struhl and Davis, 1980); Sc2812 is a derivative of Sc2605 in which the *Eco*RI site has been mutated (Struhl, 1982).

To create the spacing mutants (*his3*-Δ18, *his3*-Δ19, and *his3*-Δ36), YRp14-Sc2812 DNA was partially digested with *Mbo*II (cleaves at positions −11 and −35) or with *Taq*I (cleaves at position −24), treated with *E. coli* DNA polymerase I to produce blunt ends, and then ligated to the octanucleotide *Eco*RI linker (GGAATTCC). The resulting mixtures were cleaved with *Eco*RI, ligated, and introduced into *E. coli*. DNAs from individual transformants were analyzed by restriction mapping in order to identify those containing *Eco*RI linkers at the desired positions. These structures were verified by DNA sequencing (Sanger *et al.*, 1980) of appropriate DNA fragments cloned into the mp9 vector (Messing, 1983). The spacing mutants were constructed by pairwise combinations via the *Eco*RI linkers. The resulting DNA molecules (Figure 2) cleaved with *Xba*I and introduced into KY117 by selecting for Ura$^+$ transformants (Struhl, 1984). Excision of the integrated DNA was selected by resistance to 2.5 M ethylene glycol, and the desired gene replacement events, which have lost the *his3*-Δ200 allele, were identified as ura$^-$ His$^+$ segregants.

Class I 'swap' mutants were made by inserting an *Eco*RI linker at the *Hph*I site corresponding to position −15 of the *ded1* gene by methods analogous to those just described for the spacing mutants. To create the final molecules, the *Hind*III-*Eco*RI *ded1* promoter fragment (nucleotides −447 to −15) and the appropriate *Eco*RI-*Sal*I fragment containing the *his3* initiation region and structural gene were joined to *Hind*III-*Sal*I digested YIp5 vector DNA. The *his3* segments for *his3*-D1, D2 and D4 were obtained from the spacing mutants, whereas the analogous segment for the D3 derivative was obtained by joining an *Eco*RI linker to *Bal*31 exonuclease-treated DNA (unpublished results). These class I hybrid promoters were integrated as single copies into the yeast genome of strain KY117 by selecting for Ura$^+$ transformants following *Xba*I treatment of the DNA.

The class II swap mutant, *ded*-H1 was made by ligating the *Eco*RI (−15 of *ded1*)-*Bgl*II fragment (in the structural region of *ded1*) into the *Eco*RI-*Bgl*II-cut *his3*-Δ18 (in YRp14). This fuses the *his3* promoter down to −35 with −15 of *ded1* and, as a consequence, deletes the whole *his3* structural gene. This DNA molecule was linearized with *Bgl*II and introduced into KY114 (*HIS3*$^+$) by selecting for Ura$^+$ integrants. Gene replacement events were identified by subsequent screening for red colonies (*ade2*$^-$ due to the loss of the ochre suppressing tRNA encoded by the YRp14 vector) which were ura$^-$his$^-$.

5′ End analysis of RNA

Probe A was made by labelling the 5′ end of the *Msp*I-*Dde*I fragment (Figure 1) with T4 polynucleotide kinase and separating the DNA strands on a 4% native polyacrylamide gel (Hayward, 1972). Probes B and D were synthesized by a primer extension method similar to that employed for DNA sequencing (Sanger *et al.*, 1980). Synthetic oligonucleotide primers (100 ng) were labelled at their 5′ ends using a molar excess of [^{32}Pγ]ATP (7000 Ci/mmol) and T4 polynucleotide kinase and then hybridized to 20 μg of the appropriate M13 hybrid DNAs. The *his3* 22-mer (^{32}P-GCGATTGTGTGGCCTGTTCTGC) corresponding to nucleotides +174 to +153 was annealed to mp8-Sc3233 DNA (an *Eco*RI-*Hind*III fragment that extends from −134 to +328). The *ded1* 23-mer (^{32}P-GGCCGTTAGATCTGCTGCCACCG), which corresponds to nucleotides +262 to +240, was annealed to mp9-Sc2677 (a *Hind*III-*Bam*HI fragment that extends from −451 to +352). The hybridized primers were extended with *E. coli* DNA polymerase I in the presence of all four deoxynucleotide triphosphates (0.2 mM each), cleaved with the appropriate restriction enconuclease (*Eco*RI for *his3* and *Xho*I for *ded1*), and the resulting products were separated by electrophoresis in 1.7% low-gelling agarose, 30 mM NaOH, 1 mM EDTA. The gel slices containing the desired probes were melted at 73°C, extracted with phenol, and precipitated with ethanol after the addition of 100 μg *E. coli* tRNA. It should be noted that in terms of molarity, probes B and D have equal specific activities. Probe C was synthesized by *in vitro* transcription of *Eco*RI-cleaved pSP62-Sc3293 DNA (1 μg) with SP6 RNA polymerase and [^{32}P]UTP as described by Melton *et al.* (1984). The template DNA was constructed by inserting an *Eco*RI-*Sau*3A *his3* fragment extending from nucleotides −134 to +88 into the SP6 promoter vector. The run-off transcript is a uniformly labelled RNA whose 5′ end is defined by the SP6 RNA polymerase initiation site and whose 3′ end is defined by the *Eco*RI site at −134.

Nucleic acid hybridization experiments were carried out with total RNA which was isolated as described by Struhl and Davis (1981b). 50 μg yeast RNA was used in each hybridization reaction; this corresponds to ~10 pg of *his3* and 50 pg *ded1* RNA that is complementary to the hybridization probes. For DNA probes, ~50 000 c.p.m. of each probe (~0.5 ng) was incubated for 16−24 h with RNA at 75°C in 50 mM Hepes pH 7.0, 1.0 M NaCl (total volume 30−50 μl). Following hybridization, the reaction products were treated with S1 nuclease and analyzed as described previously (Struhl and Davis, 1981a). For RNA:RNA hybridization, ~250 000 c.p.m. probe C (5 ng) was incubated with yeast RNA at 75°C for 1 h in 30 mM Hepes buffer (pH 7.0), 0.3 M NaCl, 1 mM EDTA, and then treated with 25 μl/ml of RNaseA at 0°C for 30 min. The products were immediately analyzed on an 8% native polyacrylamide gel. Both analytical methods represent qualitative and quantitative measurements of transcription because the hybridization probes are in molar excess, and because calculations and experiments indicate that the hybridization reactions have gone to completion.

Acknowledgements

We thank Alex Nussbaum for synthesizing the oligonucleotide primers, Steve Scholnick, Bruce Morgan and Dan Silverman for comments on the manuscript, and members of the laboratory for fruitful discussions during the course of the work. This research was supported by grants from the Markey Foundation (to W.C.), and from the National Institutes of Health (GM30186) and the Chicago Community Trust (Searle Scholars program) to K.S.

References

Baker,C.C., Herisse,J., Courtois,G., Galibert,F. and Ziff,E. (1979) *Cell*, **18**, 569-580.
Banerji,J., Rusconi,S. and Schaffner,W. (1981) *Cell*, **27**, 299-308.
Benoist,C. and Chambon,P. (1981) *Nature*, **290**, 304-310.
Bram,R. and Kornberg,R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 43-47.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Cattaneo,R., Will,H., Hernandez,N. and Schaller,H. (1983) *Nature*, **305**, 336-338.
Davison,B.L., Egly,J.-M., Mulvihill,E.R. and Chambon,P. (1983) *Nature*, **301**, 680-686.
Dierks,P., Ooyen,A.V., Mantel,N. and Weissmann,C. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1411-1415.
Fromm,M. and Berg,P. (1982) *J. Mol. Appl. Genet.*, **1**, 457-481.
Ghosh,P.K., Lebowitz,P., Frisque,R.J. and Gluzman,Y. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 100-104.
Giniger,E., Varnum,S.M. and Ptashne,M. (1985) *Cell*, **40**, 767-774.
Grosschedl,R. and Birnstiel,M.L. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1432-1436.
Guarente,L. (1984) *Cell*, **36**, 799-800.
Guarente,L. and Hoar,E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7860-7864.
Guarente,L. and Mason,T. (1983) *Cell*, **32**, 1279-1286.
Hawley,D.K. and McClure,W.R. (1983) *Nucleic Acids Res.*, **11**, 2237-2255.
Hayward,G.S. (1972) *Virology*, **49**, 342-344.
Hope,I.A. and Struhl,K. (1985) *Cell*, **43**, in press.
Ito,H., Fukuda,Y., Murata,K. and Kimura,A. (1983) *J. Bacteriol.*, **153**, 163-168.
Kamen,R., Jat,P., Treisman,R. and Favaloro,J. (1982) *J. Mol. Biol.*, **159**, 189-224.
Mathis,D.J. and Chambon,P. (1981) *Nature*, **290**, 310-315.
McKnight,S.L. and Kingsbury,R. (1982) *Science*, **217**, 316-324
Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R (1984) *Nucleic Acids Res.*, **12**, 7035-7056.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20-79.
Moreau,P., Hen,R., Wasylyk,B., Everett,R., Gaub,M.P. and Chambon,P. (1981) *Nucleic Acids Res.*, **9**, 6047-6068.
Parker,C.S. and Topol,J. (1984) *Cell*, **36**, 357-369.

Sanger,F., Coulson,A.R., Barell,B.G., Smith,A.J. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.
Siliciano,P.G. and Tatchell,K. (1984) *Cell*, **37**, 969-978.
Struhl,K. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 4461-4465.
Struhl,K. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7385-7389.
Struhl.K. (1983) *Gene*, **26**, 231-241.
Struhl,K. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7865-7869.
Struhl,K. (1985) in Reznikoff,I.W. and Gold,L. (eds.), *From Gene to Protein: Steps Dictating the Maximal Level of Gene Expression*, Benjamin/Cummings Publishing Co., California, in press.
Struhl,K. and Davis,R.W. (1980) *J. Mol. Biol.*, **136**, 309-332.
Struhl,K. and Davis,R.W. (1981a) *J. Mol. Biol.*, **152**, 535-552.
Struhl,K. and Davis,R.W. (1981b) *J. Mol. Biol.*, **152**, 553-568.
Struhl,K., Stinchcomb,D.T., Scherer,S. and Davis,R.W. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1035-1039.
Talkington,C.A. and Leder,P. (1982) *Nature*, **298**, 192-195.

Received on 12 August 1985