The EMBO Journal vol.11 no.12 pp.4583-4590, 1992

# Sequence-specific initiator elements focus initiation of transcription to distinct sites in the yeast *TRP4* promoter

**Hans-Ulrich Mösch, Roney Graf and Gerhard H.Braus**

Institute of Microbiology, Swiss Federal Institute of Technology, (ETH), Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland

Communicated by W.Schaffner

**Transcription from the yeast *TRP4* promoter initiates at two basal (i127 and i76) and three GCN4 dependent (i31, i25 and i12) initiator elements. All of these elements contain not more than one deviation from the earlier proposed initiator consensus sequence PuPuPyPuPu, a pyrimidine nucleotide flanked on either side by two purine nucleotides. A point mutation analysis of these elements in various combinations was performed and revealed that the central pyrimidine nucleotide and at least one of the 3′ flanking purine nucleotides of the PuPuPyPuPu consensus sequence are essential but alone not sufficient to define a functional initiator element. Multiple cryptic transcription start sites, which function independently whether they are located on the coding or the non-coding strand, can replace the function of mutated initiator elements and therefore the overall level of transcription initiation is not affected. The sequence specificity is identical for basal and GCN4 dependent initiator elements demonstrating that they are functionally homologous. These findings imply that the role of initiator elements is to 'focus' the start point(s) of transcription to distinct sites located in the region between the site(s) of the assembly of the transcriptional complex and the start codon of translation.**

*Key words:* amino acid biosynthesis/initiator element/promoter/*Saccharomyces cerevisiae*/transcription initiation

## Introduction

In yeast, transcriptional regulation of RNA polymerase II promoters requires three different types of DNA sequence elements, namely upstream (UAS), TATA and initiator (I) elements (reviewed by Guarente, 1987, 1988; Struhl, 1987, 1989).

Upstream elements are short DNA sequences, typically 10–30 bp in length, which work in a distance and orientation independent manner ~100–600 bp upstream from the transcription initiation site; they act as targets for specific DNA binding proteins that activate or repress transcription. Most eukaryotic RNA polymerase II promoters also contain a TATA element, consensus TATAAA, which is necessary but not sufficient for accurate initiation of transcription *in vivo* and *in vitro* (Dynan and Tjian, 1985; Struhl, 1987). Functional TATA elements are located between UAS and mRNA initiation site(s) and mediate the first step in the pathway of transcription initiation by binding the general transcription factor TFIID (Buratowski *et al.*, 1988, 1989; Van Dyke *et al.*, 1988).

The transcription initiator (I) element is the primary determinant of where transcription begins in yeast (for review see Struhl, 1989). Yeast mRNA initiation sites are determined primarily by specific initiator sequences, and not, as in many genes of higher eukaryotes, by the distance (25–30 bp) from the TATA element (Chen and Struhl, 1985). The spacing between yeast initiation sites and TATA elements is much more flexible and somewhat larger, although there are limits to the distance over which a TATA element can act (~40–120 bp). The DNA sequence requirements for yeast initiator elements are poorly understood, although it is clear that many different sequences can carry out the function. Two possible initiation site consensus sequences, TCG/AA and PuPuPyPuPu, a pyrimidine flanked by at least two purines on each side, have been shown to function as an initiation site when inserted at one location upstream of the yeast *CYC7* gene (Hahn *et al.*, 1985). Using random oligonucleotide mutagenesis McNeil and Smith (1986) found a preference for 5′ ends at A residues preceded by a short track of pyrimidine residues. Healy *et al.* (1987) surveyed sequences within a 2–3 bp range of 99 published yeast initiation sites and argue that the sequence PyAAPu is a functional start site element.

In many genes from *Drosophila melanogaster* the highly conserved sequence 5′PyATCAG/TTPy3′ has been found as a transcription initiator element (Hultmark *et al.*, 1986), whereas in mammals the transcription initiation sequence 5′PyPyCAPyPyPyPyPy3′ has been identified in multiple genes carrying a TATA box (Corden *et al.*, 1980). An 'initiator' (Inr) as a transcription control element was recently described for the lymphocyte-specific terminal deoxynucleotidyl transferase gene (Smale and Baltimore, 1989), where TATA and Inr elements were shown to be functionally similar (Smale *et al.*, 1990). Additional subunit(s) or domain(s) of the mammalian transcription factor TFIID were suggested to be required for mediating transcriptional activation via Inr elements (Nakatani *et al.*, 1990; Smale *et al.*, 1990).

The 'general control' system of yeast connects transcriptional derepression of >30 unlinked amino acid biosynthetic genes to the physiological signal 'amino acid starvation' (reviewed by Hinnebusch, 1988). Typical amino acid biosynthetic promoters are dual promoters that can be regulated by two control systems, namely 'general' and basal (Struhl, 1986; Arndt *et al.*, 1987). Different upstream, TATA and initiator elements are required for controlling the basal or 'general' control response. The upstream activator protein GCN4 is the key regulator in the 'general control' system of yeast (Hinnebusch and Fink, 1983; Thireos *et al.*, 1984).

The *TRP4* gene of *Saccharomyces cerevisiae* encodes for the enzyme anthranilate phosphoribosyl transferase (EC

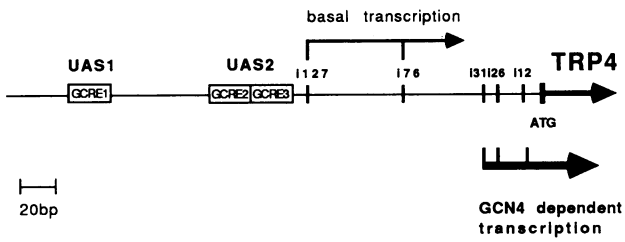



**Fig. 1.** *TRP4* gene promoter region. UAS elements for GCN4 dependent transcription are indicated and GCN4 responsive elements (GCRE1, GCRE2 and GCRE3) are represented as boxes (Mösch *et al.*, 1990). Initiator elements for basal transcription (i127 and i76) and for GCN4 dependent transcription (i31, i25 and i12) are indicated.

2.4.2.18) and is subject to the 'general control' system in yeast (Furter *et al.*, 1986, 1988; reviewed by Braus, 1991). *TRP4* expression is regulated by GCN4 via two UAS elements that act together synergistically to derepress the *TRP4* gene (Braus *et al.*, 1989; Mösch *et al.*, 1990). The *TRP4* promoter contains completely separate initiator elements for basal (i127 and i76) and GCN4 dependent (i31, i25 and i12) transcription (Figure 1; Furter *et al.*, 1988; Mösch *et al.*, 1990). This is in contrast to other described GCN4 dependent promoters (e.g. the yeast *HIS3* and *HIS4* genes), where basal and GCN4 dependent transcription initiate from identical initiator elements or where part of the basal initiator elements are used by the GCN4 dependent promoter (Chen and Struhl, 1985; Nagawa and Fink, 1985).

| | i127 i121 i113 i110 | i76 | i31 i25 i12 |
|---|---|---|---|
| wt | ..TCAGTAAACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 1 | ..TCAGCAAACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 2 | ..TCAGAAAACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 3 | ..TCAGGAAACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 4 | ..TCTCTAAACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 5 | ..TCAGTTCACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 6 | ..TCTTACTACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 7 | ..TCTTGCTACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 8 | ..TCTTTCTACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 9 | ..TCTTCCTACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 10 | ..TCAGGAAACAGAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 11 | ..TCAGGAAACAGAAACTCTCTTCCA.. | ..AAAGCATCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 12 | ..TCAGTAAACATAAACTCTGTTGCA.. | ..AAAGCCTCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 13 | ..TCAGTAAACATAAACTCTGTTGCA.. | ..AAAGCATCTGAT.. | ..GCAAGAATCAGAAGAAAACTTAGGATTCC.... |
| 14 | ..TCAGTAAACATAAACTCTGTTGCA.. | ..AAAGCCTCTGAT.. | ..GCAAGAATCAGAAGAAAACTTAGGATTCC.... |
| 15 | ..TCAGGAAACAGAAACTCTCTTCCA.. | ..AAAGCATCTGAT.. | ..GCAAGAATCAGAAGAAAACTTAGGATTCC.... |
| 16 | ..TCAGGAAACAGAAACTCTCTTCCA.. | ..AAAGCCTCTGAT.. | ..GCAATAATCACAAGAAAACTTAGTATTCC.... |
| 17 | ..TCAGGAAACAGAAACTCTCTTCCA.. | ..AAAGCCTCTGAT.. | ..GCAAGAATCAGAAGAAAACTTAGGATTCC.... |

**Fig. 2.** Nucleotide sequences of wild-type (wt) and mutant (1–17) *TRP4* promoter alleles. Major (i127, i76, i31, i25 and i12) and cryptic (i121, i113, i110) initiation sites are bold and their positions (relative to translational start site at +1) are indicated by arrows. Mutated nucleotides were obtained by site-directed mutagenesis and are underlined and marked by asterisks. Nucleotides corresponding to the PuPuPyPuPu consensus sequence (Hahn *et al.*, 1985) on coding (—) or non-coding (· · · · · ·) strands, respectively, are underlined.

In this study we performed a fine analysis of the basal as well as the GCN4 dependent initiator elements of the *TRP4* gene. We find that: (i) basal and GCN4 mediated transcription select functionally homologous initiator elements which are highly sequence specific and which work in either orientation; (ii) multiple cryptic transcription start sites can replace the function of mutated initiator elements; (iii) the initiator elements 'focus' basal and GCN4 mediated transcription to distinct start sites; and (iv) initiator elements do not significantly affect the overall level of transcription initiation.

## Results

### *Basal and GCN4 dependent initiator elements are functionally homologous, highly sequence specific and function in both orientations*

Our aim was to obtain a better understanding of the DNA sequence requirements of initiator elements in the yeast *S.cerevisiae*. Therefore we performed site-directed mutagenesis of all basal (i127 and i76) as well as GCN4 dependent (i31, i25 and i12) initiator elements of the *TRP4* gene. Whilst the two most upstream located basal and GCN4 dependent initiator elements (i127 and i31, respectively) perfectly match the proposed PuPuPyPuPu consensus sequence (Hahn *et al.*, 1985) the other elements i76, i25 and i12 consist of shortened forms of either a PuPuPyPu (i76 and i12) or a PuPyPuPu (i25) sequence (Figure 2 and Table I). Interestingly, for i127, i31, i25 and i12 the consensus sequence occurs on the coding strand and for i76 on the non-coding strand, respectively (Figure 2 and Table I). In all initiator elements the central pyrimidine nucleotide was the start point for transcription (Figures 3 and 4, all wt lanes).

We first created different point mutations of the major basal initiator element i127 (Figure 2, mutant promoters Nos 1−9) to test which nucleotides of the PuPuPyPuPu consensus are required for proper initiator function. The different mutant promoter alleles were then substituted for the wild-type *TRP4* promoter by gene replacement in the genomic *TRP4* locus and transcripts of the various mutated *TRP4* promoters were analysed by primer extension (Figure 3). A transition of the central T pyrimidine nucleotide of i127 to a C pyrimidine nucleotide had no influence on the selection of i127 as a major basal initiator element, as the transcription initiation pattern of mutant promoter No. 1 is indistinguishable from a wild-type situation (Figure 3, lane 1). However, transversion of the same T pyrimidine nucleotide to either an A or a G purine nucleotide resulted in a quantitative loss of transcription initiating from i127 (Figure 3, lanes 2 and 3). Surprisingly, transcription initiated more downstream at multiple cryptic start sites instead. A sequence analysis revealed that the sequences of these basal cryptic sites also correspond to the PuPuPyPuPu consensus sequence and are located either on the coding or non-coding strand (Figure 2 and Table I).

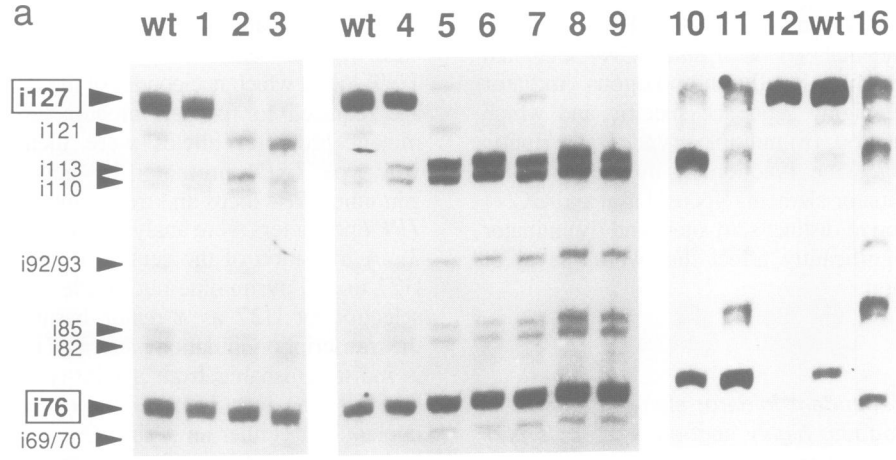
Transversion of the 3′ flanking purine nucleotides of i127 to pyrimidine nucleotides also completely abolished transcription initiation from i127 and again transcription initiated from the cryptic start sites instead (Figure 3, lane 5). Substitutions of the 5′ flanking purines for pyrimidine nucleotides in i127, however, still rendered an almost normal wild-type initiation pattern (Figure 3, lane 4). These results show that the central pyrimidine nucleotide and at least one of the 3′ flanking purine nucleotides are essential parts of

**Table I.** DNA sequences corresponding to the PuPuPyPuPu consensus initiator sequence in the *TRP4* promoter region

| PuPuPyPuPu - sequences | | PuPuPyPu - sequences | | PuPyPuPu - sequences | |
|---|---|---|---|---|---|
| **position** | **functional** | **position** | **functional** | **position** | **functional** |
| coding | | coding | | coding | |
| **-127** AGTAA | +++ main$^{basal}$ | -123 AACA | - | -121 ATAA | + cryptic$^{basal}$ |
| -82 AATAA | + cryptic$^{basal}$ | -77 AGCA | - | -85 ATAA | + cryptic$^{basal}$ |
| -55 GACAA | - | -70 GATA | + cryptic$^{basal}$ | -65 ACAG | - |
| -47 AGTGA | - | **-12** AGTA | +++ main$^{GCN4}$ | -34 GCAA | - |
| **-31** AATAA | +++ main$^{GCN4}$ | | | **-25** ACAA | +++ main$^{GCN4}$ |
| non-coding | | non-coding | | non-coding | |
| -113 AACAG | + cryptic$^{basal}$ | -108 AATG | - | -110 GCAA | + cryptic$^{basal}$ |
| -104 AATGA | - | **-76** GATG | +++ main$^{basal}$ | -86 ATAA | - |
| -60 AGTAG | - | -69 AATA | + cryptic$^{basal}$ | -66 GTAA | - |
| -3 GATAA | - | -11 AATA | - | -35 GCAG | - |
| | | | | -26 GTGA | - |

All DNA sequences in the *TRP4* promoter, between position −127 (i127) and the translational start site at +1, corresponding to either the PuPuPyPuPu consensus sequence or the shortened forms PuPuPyPu and PuPyPuPu on the coding or non-coding strand are shown, as determined by a computer search. Numbers indicate the position relative to the translational start site at +1. Major initiation start sites are bold and underlined. Functions are indicated by +++ (major initiators), + (cryptic initiators) or − (non-functional as initiators).

a




b

| | wt | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **i127** | 66 | m64 | m12 | m9 | m54 | m3 | m- | m3 | m- | m- | m9 | m9 | 64 | m13 |
| i121 | 4 | 4 | 26 | 37 | 4 | 5 | 1 | - | - | - | m4 | m5 | 4 | m5 |
| i113 | 4 | 4 | 14 | 9 | 6 | 26 | 29 | 28 | 27 | 28 | 37 | m7 | 4 | m11 |
| i110 | 4 | 4 | 5 | 5 | 6 | 16 | 17 | 16 | 17 | 18 | 4 | m4 | 4 | m5 |
| i92/93 | - | - | - | - | - | 3 | 4 | 4 | 5 | 4 | - | 3 | - | 6 |
| i85 | - | - | - | - | - | 4 | 3 | 3 | 4 | 3 | - | 13 | - | 19 |
| i82 | - | - | - | - | - | 4 | 3 | 3 | 3 | 3 | - | 10 | - | 10 |
| **i76** | 22 | 23 | 38 | 38 | 21 | 28 | 29 | 32 | 31 | 32 | 31 | 36 | m1 | m1 |
| i69/70 | - | - | - | - | - | - | - | - | - | - | - | - | 5 | 14 |
| % wt | 100a | 99a | 95a | 98a | 91a | 89a | 86a | 89a | 87a | 88a | 85a | 87a | 82a | 84a |

**Fig. 3.** Transcription start site selection patterns of basal *TRP4* gene transcripts under control of wild-type and mutant promoters. (**a**) Primer extension analysis of basal wild-type (wt) and mutant (1−12 and 16, for numbering see Figure 2) *TRP4* gene transcripts. 100 μg of poly(A)$^+$ RNA from identical preparations to those used for quantitative Northern hybridizations in Figure 5 were hybridized to completion with an excess of a $\gamma$-$^{32}$P-labelled *TRP4* primer and subsequently elongated using AMV reverse transcriptase. Initiation sites of *TRP4* transcription are designated as i (numbers indicate their position relative to translational start at +1). Major basal initiator elements i127 and i76 are boxed. (**b**) Quantification of basal wild-type and mutant *TRP4* transcription selection patterns. Individual lanes of primer extension analysis from (a) were scanned using laser densitometry. The numbers in columns 1−12 and 16 represent the usage of each initiation site in comparison with a wild-type situation (wt) and are percentage values of the overall basal *TRP4* transcription of a given promoter (summarized in row % wt). Mutation initiation sites are underlined and marked by a superscribed m. Overall basal *TRP4* transcript levels of different mutant promoter alleles (see Figure 5) are shown in the last row (% wt) and are indicated by a superscribed a.

the earlier proposed PuPuPyPuPu consensus sequence, whereas the 5′ flanking purines do not seem to be required.

In mutant promoters Nos 6, 7, 8 and 9 the PuPuPyPuPu sequence of i127 was inverted from a 5′AGTAA3′ sequence to the four possible sequences 5′TT(A/C/G/T)CT3′, to test whether this part of i127 alone is sufficient for proper initiator function. Inversion of the PuPuPyPuPu sequence of i127 led to a non-functional i127 element, independently of the type of the central nucleotide, and transcription from the cryptic start sites was found again (Figure 3, lanes 6, 7, 8 and 9). This shows that the PuPuPyPuPu sequence is essential but alone not sufficient to define an initiator element and suggests that additional nucleotides upstream or downstream are required. This is in agreement with the finding that the half-sites upstream and downstream of the transcription start are not identical, therefore suggesting an initiator sequence to be asymmetric.

We further created point mutations of the major basal initiator element i76 and the major GCN4 dependent initiators i31, i25 and i12 *in vitro* by substituting their PuPuPyPuPu sequences for PuPuPuPuPu sequences instead. These mutant promoter alleles (Figure 2, constructs Nos 12 and 13) were also targeted to the chromosomal *TRP4* locus and *TRP4* transcripts were examined from yeast cells lacking any GCN4 protein (Figure 3) or containing high amounts of GCN4 (Figure 4). A conversion of the central pyrimidine nucleotide in both the major basal initiator element and i76 (Figure 3, lane 12) as well as the major GCN4 dependent

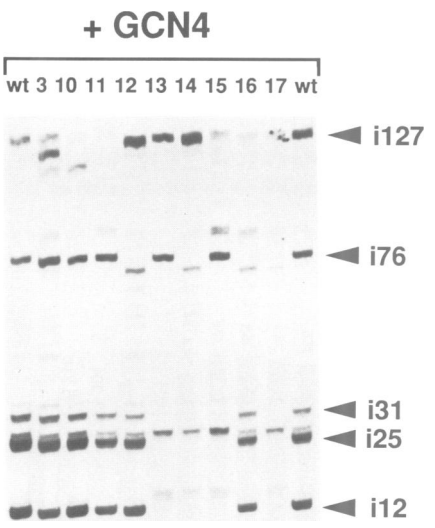



**Fig. 4.** Transcription start site selection patterns of *TRP4* gene transcripts under control of GCN4. Primer extension analysis of wild-type (wt) and mutant (3 and 10–17; numbering according to Figure 2) *TRP4* gene transcripts from yeast cells containing high amounts of GCN4 protein (+GCN4). Primer elongation reactions were performed as described in Figure 3a and initiation sites of *TRP4* transcription are designated as i (numbers indicate their position relative to translational start at +1).

initiator elements i31, i25 and i12 (Figure 4, lane 13) resulted in a quantitative loss of transcription initiating from these mutated elements. Again transcription initiated further downstream at multiple cryptic start sites instead.

These results demonstrate that (i) initiator elements in yeast are elements with high sequence specificity as their usage by the transcription machinery can be completely blocked by single point mutations in their central pyrimidine nucleotide normally used as the start site of transcription; (ii) the central pyrimidine nucleotide and at least one of the 3′ flanking purine nucleotides of the PuPuPyPuPu consensus are essential but alone not sufficient to define a functional initiator element; (iii) inversion of the core sequence results in a non-functional initiator element; and (iv) major basal and GCN4 dependent transcription select initiator elements that are completely different in respect of their location but have the same DNA sequence requirements.

### *The TRP4 promoter carries redundant cryptic initiator sites which replace the mutated initiator elements*

Basal transcription of the *TRP4* gene mainly initiates at the two major start sites i127 and i76. In addition, several minor start sites for basal *TRP4* transcription could be detected when autoradiograms of primer extension analyses of basal *TRP4* transcripts were carefully analysed (Figure 3). Besides the major initiator elements i127 and i76, three minor start sites could be mapped at positions −121 (i121), −113 (i113) and −110 (i110) relative to the translational start site at +1 (Figure 3a). Whereas the minor initiator element i113 perfectly matches the PuPuPyPuPu consensus sequence, the elements i121 and i110 each consist of the shortened PuPyPuPu sequence comparable with the i76 element. For i121 the consensus sequence lies on the coding strand and for i113 and i110 on the non-coding strand, respectively (Figure 2 and Table I). A quantification of all of these basal start sites was subsequently performed using laser densitometry (Figure 3b). The major i127 and i76 elements were selected by the basal transcription machinery at a level of 66 and 22%, respectively, whereas the minor start sites were each used to an extent of not more than 4% of the total transcripts.

A substitution of the central pyrimidine for a purine nucleotide in the major i127 element or the inversion of i127 (Figure 2, constructs Nos 2, 3 and 6) led to a 5- to 20-fold decrease in the usage of this element for transcription initiation and to a 2- to 7-fold increase in the transcription machinery initiating at each of the two minor elements i121 and i113 (Figure 3a and b, lanes 2, 3 and 6). We therefore constructed further *TRP4* promoter mutant alleles *in vitro*, again by substituting the central pyrimidine nucleotide of the PuPuPyPuPu sequences of the minor initiator elements i121, i113 and i110 for a purine nucleotide to result in PuPuPuPuPu sequences instead (Figure 2, constructs Nos 10, 11 and 16). These mutant promoters were also targeted to the chromosomal *TRP4* locus, substituting the wild-type *TRP4* promoter. A primer extension analysis of *TRP4* transcripts generated *in vivo* under the control of these mutant promoters revealed that simultaneous mutations in elements i127 and i121 (mutant promoter No. 10) still resulted in a 7-fold decrease in the usage of i127, but this was replaced by a 9-fold increase of transcription initiation solely from the i113 element (Figure 3a and b, lanes 10). When the four elements i127, i121, i113 and i110 were simultaneously mutated (mutant promoter No. 11), usage of i127 still dropped 7-fold and transcription at i121, i113 and i110 initiated at wild-type levels, but now basal transcription initiating from two additional cryptic start sites at positions −85 (i85) and −82 (i82) could be found (Figure 3a and b, lanes 11). The DNA sequences of these cryptic initiator elements i85 (−86 ATAA −83) and i82 (−84 AATAA −80) also match to the PuPuPyPuPu consensus sequence with not more than one terminal deviation. Finally, simultaneous mutations in the five elements i127, i121, i113, i110 and i76 (mutant promoter No. 16) led to a random distribution of transcription initiating at the various mutated major and minor initiator elements, each at levels between 1 and 14% of the total basal transcription (Figure 3a and b, lanes 16). Two additional initiator elements were also used in mutant No. 16 at positions −92/−93 (i92/93) and −69/−70) (i69/70) to an extent of 6 and 14%, respectively. The element i69/70 contains a PuPuPyPu sequence (−72 GATA −69), whereas the weak element i92/93 does not correspond to a PuPuPyPuPu consensus sequence. Although in mutant promoter No. 16 the two wild-type initiation sites as well as three cryptic start sites were mutated, the transcription machinery could not be urged to initiate at the more downstream GCN4 dependent initiator elements, suggesting that the basal promoter can only initiate within a window that has its downstream border in this region.

Transcription mediated by GCN4 also initiated more downstream from cryptic start sites when the GCN4 dependent initiator elements i31, i25 and i12 were mutated (Figure 4, mutant promoters Nos 13, 14, 15 and 17). However, the DNA sequences of these cryptic GCN4 dependent start sites at positions −28/−29 and −15/−16 do not correspond to the PuPuPyPuPu consensus sequence.

Taken together, these results show that: (i) the *TRP4*

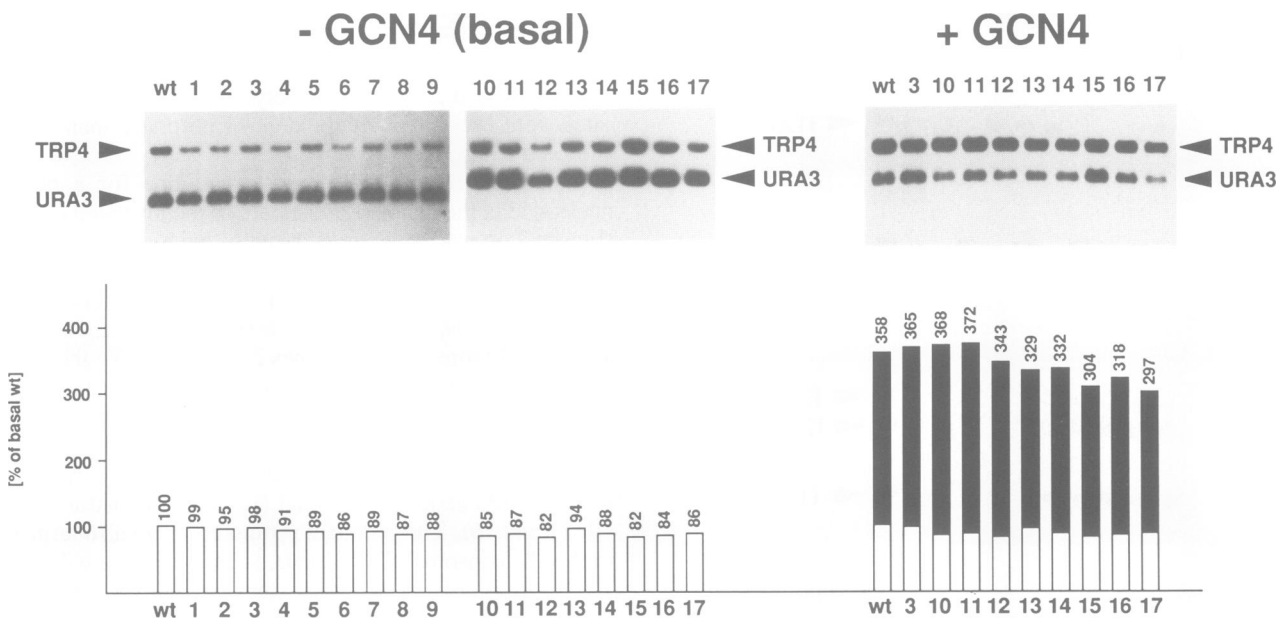



**Fig. 5.** Expression of *TRP4* gene under control of wild-type and mutant promoters. Quantification of *TRP4* transcript levels are shown in the lower part. The values given (% of basal wt) are the means of at least four independent measurements obtained by performed laser densitometry on autoradiograms of at least four different Northern analyses using two independent poly(A)$^+$ RNA isolations of each wild-type (wt) and mutant (1−17; numbering according to Figure 2) strain (standard deviations did not exceed 20%). Values are *TRP4* transcript levels in percentage of basal wild-type expression and were obtained by using *URA3* transcript levels as internal standards. Basal expression is indicated by white bars; GCN4 dependent expression (black bars) was obtained by subtracting basal levels (white bars) from total expression levels occurring in the presence of high amounts of GCN4 protein in the cells (+GCN4). The upper part represents one example of an autoradiogram of a Northern analysis that was scanned by laser densitometry as described above. Poly(A)$^+$ RNA from yeast strains containing no GCN4 protein (−GCN4) or a high amount of GCN4 protein (+GCN4) and carrying wild-type (wt) and mutant *TRP4* promoter alleles (1−17; numbering according to Figure 2) at the chromosomal *TRP4* locus was isolated and subsequently cohybridized against radiolabelled *TRP4* and *URA3* probes. The *URA3* transcript was chosen as an internal standard. Transcript sizes are: *TRP4*, 1.4 kb; *URA3*, 0.9 kb.

promoter region contains multiple cryptic initiator elements; (ii) the transcription machinery is urged to initiate further downstream from these cryptic initiator elements when upstream initiation sites are increasingly mutated; and (iii) basal transcription cannot be urged to initiate at the GCN4 dependent initiator elements.

### *Eviction of initiator elements results in defocused mRNA 5′ ends but does not substantially decrease the total level of transcription initiation*

A quantitative *TRP4* transcript analysis was performed to determine the overall basal and GCN4 mediated expression of the *TRP4* gene under control of the different mutant promoter alleles (Figure 5). The overall basal or GCN4 mediated transcription did not drop below 80% of wild-type transcription for any of the mutant promoters, even when all the basal i127 and i76 elements, the basal cryptic i121, i113 and i110 elements as well as the GCN4 dependent i31, i25 and i12 elements were simultaneously mutated (Figure 5, mutant promoter No. 17). Although such a mutant promoter still revealed ~80% of normal wild-type *TRP4* transcription, the 5′ ends of these transcripts could no longer be mapped to the five distinct positions i127, i76, i31, i25 and i12 of a wild-type *TRP4* promoter, but transcription initiation from such a construct seemed to be 'defocused' and starting randomly from multiple weak cryptic start sites distributed over the *TRP4* promoter region.

These results show that: (i) though mutations of the central pyrimidine nucleotide in a given initiator element lead to a quantitative loss of transcription initiating at such a mutated site the overall transcription level is unaffected due to the usage of multiple cryptic initiator elements in the *TRP4* promoter region; (ii) the presence of an initiator element is required for a 'focused' pattern of basal as well as for GCN4 mediated transcription initiation.

## Discussion

The current theory of yeast mRNA initiation site selection suggests that the sequence at the site of RNA initiation dictates the selection of the RNA initiation site. Comparison of the sites of RNA initiation of many yeast genes has led to the identification of several sequences which are found preferentially at the site of RNA initiation: PuPuPyPuPu, TC(A/G)A or PyAAPu (Hahn *et al.*, 1985; Healy *et al.*, 1987). However, none of these suggested initiation site consensus sequences has been investigated by a site-directed mutational analysis so far.

Here, we performed a point mutation analysis of the proposed PuPuPyPuPu consensus sequence of the basal as well as the GCN4 dependent initiation elements in the *TRP4* gene promoter in various combinations. We made use of the fact that the two major basal (i127 and i76), three cryptic basal (i121, i113 and i110) and the three GCN4 dependent (i31, i25 and i12) *TRP4* initiator elements all contain not more than one deviation from the PuPuPyPuPu consensus sequence and could be mutated by substituting the central pyrimidine for a purine. We find that transcription starts at the central pyrimidine nucleotide of the PuPuPyPuPu sequences in all of the *TRP4* initiation elements, independent

of whether they lie on the coding or non-coding strand. This suggests that a double stranded DNA sequence rather than a single stranded sequence is recognized as a transcription initiation site by the transcription machinery, and therefore transcription initiation would start before the process of melting of the double stranded DNA.

In other studies initiator elements were often only defined by performing deletions in the tested promoter sequences (Chen and Struhl, 1985; Hahn *et al.*, 1985; Healy and Zitomer, 1990; Maicas and Friesen, 1990; Rathjen and Mellor, 1990). It is therefore important to mention that here single point mutations in the start point of transcription lead to the loss of a given initiator element to be selected for transcription initiation, and that no spacing effects or deletion of additional promoter sequences can be responsible. Our findings, that transversion of the central pyrimidine nucleotide to a purine nucleotide always resulted in a drastic loss of transcription initiation from such a mutated site, therefore demonstrate that it is indeed the sequences at the start point of transcription that are essential for the selection of an initiator element by the transcription machinery. Simultaneously the initiation level from other intact cryptic elements was found to be increased. Early studies showed that in mammals the TATA box is the dominant element in the selection of initiation points. For example, when the cap site is eliminated, other kinds of 'cryptic' initiators ~25 bp downstream of the TATA box were recruited as new initiation points (Grosschedl and Birnstiel, 1980; Breathnach and Chambon, 1981). Recent studies show that in mammals the initiator itself also plays an important role and is sufficient to serve as a precise, albeit not very efficient initiation site even in a promoter lacking a TATA box (Smale and Baltimore, 1989; Smale *et al.*, 1990). Here we show that though in yeast the initiator was also recognized as an important promoter element, cryptic sites within a certain window, but not at a defined distance, are redundantly present in a promoter and are activated *in vivo* when the original site is evicted. Therefore the importance of the initiator elements in yeast and mammals seems to be more similar than previously assumed. It seems that in mammals there is a strict spacing and a looser consensus sequence, whereas in yeast the spacing is more relaxed (initiation window) but the consensus sequence is stricter.

Table I summarizes all DNA sequences corresponding either to the PuPuPyPuPu consensus sequence or to the shortened forms PuPuPyPu and PuPyPuPu, that are found in the *TRP4* promoter region between position −127 (i127) and the translational start site. Interestingly, only 12 out of the 27 possible consensus sequences are used *in vivo* either as major initiator elements or as cryptic start points. All five major initiator elements belong to the class of PuPuPyPuPu consensus sequences as well as seven out of the 10 cryptic initiator elements. A comparison of all found major and cryptic initiators of the PuPuPyPuPu class reveals that in 11 out of these 12 sequences the central nucleotide is a pyrimidine followed by an A purine nucleotide. We therefore propose the sequence 5'PyA3' to be an essential part of functional initiator sequences, as mutations of these nucleotides or inversion of the core sequence lead to a non-functional initiator element. There are three exceptions: the cryptic basal element i92/93 and the cryptic GCN4 dependent elements i28/29 and i15/16. Therefore not only the sequence but also the position of an initiator element in the promoter region, e.g. the distance from the TATA box or UAS elements, are important for its usage. It is furthermore possible that additional auxiliary 'determinator' or 'locator' sequences upstream or downstream of the basal *TRP4* initiation elements are involved in the selection of the start site of transcription, as found for the *PGK* and *TCM1* genes of *S.cerevisiae* (Maicas and Friesen, 1990; Rathjen and Mellor, 1990).

While transversion of the start point of transcription had a drastic effect on selection of an initiator element, the same mutations did not affect the overall level of basal *TRP4* transcription, even when multiple major and minor initiator elements were mutated. Thus, the basal *TRP4* initiator elements seem only to direct the RNA polymerase II to initiate transcription at specific sites, but are not responsible for the overall amount of transcription initiation. This suggests a model where upstream (UAS) and TATA elements attract the transcription machinery to a specific promoter, and the initiator element determines the start site of transcription during or after assembly of the transcriptional complex. In such a model upstream activator proteins and TATA box binding proteins would open a 'window' for transcription initiation and transcription would then start at initiator elements that are abundantly present within this 'window' of initiation. Such a model is supported by our finding that mutation of both major basal (i127 and i76) and the cryptic basal initiator elements (i121, i113 and i110) did not lead to basal transcription initiation at points more downstream than position −69. A similar 'window' of transcription initiation has been proposed in the case of the *ADH1* gene from *Schizosaccharomyces pombe* (Furter-Graves and Hall, 1990).

Studies on DNA sequence requirements for yeast TATA element function have revealed that besides the optimal consensus sequence TATAAA the sequences TATATA and TATCTA (Chen and Struhl, 1988) or TATTTA (Harbury and Struhl, 1989) are also functional, albeit to a lesser extent, *in vivo*. In addition, an *in vitro* study defined, besides TATAAA, the sequences AATAAA, CATAAA, TTTAAA, TATATA or TATAAG as also being able to mediate TFIID dependent transcription (Wobbe and Struhl, 1990). In higher eukaryotes an 'initiator' (Inr) as a transcription control element has been described for the terminal deoxynucleotidyl transferase (*TdT*) gene (Smale and Baltimore, 1989). A further *in vitro* analysis of *TdT* gene transcription revealed that TATA and Inr elements are functionally similar (Smale *et al.*, 1990). In this context, it is interesting that two cryptic *TRP4* initiator elements perfectly match functional TATA sequences, namely CATAAA (i121) and AATAAA (i82) (Figure 2c). However, in yeast it has never been shown that a TATA box and an initiator element are functionally interchangeable and it therefore remains to be tested whether these cryptic *TRP4* initiator elements can functionally replace a TATA box.

Furthermore it remains to be elucidated which proteins might recognize the initiator element and whether such proteins are part of the RNA polymerase II complex or belong to the general transcription factors or whether there exist specific initiator factors for specific promoters.

Our results indicate that the presence of multiple initiator elements with degenerate sequences guarantees that transcription starts upstream of the translational start site ATG and therefore all transcribed mRNA molecules can

be properly translated. This renders the promoter less sensitive to single point mutations. Therefore, it might be advantageous for the cell that the sequence of initiator elements is degenerate, similar to other degenerate sequence motifs in nature, such as 'acid blobs' in transcriptional activation or target sequences of proteins that are transported into mitochondria. As such a 'focusing' mechanism has been selected during evolution it would seem to have important advantages when compared with a 'defocused' and stoichastic start mechanism of transcription initiation.

## Materials and methods

### *Construction of yeast strains carrying TRP4 mutant alleles*

All yeast strains carrying *TRP4* mutant alleles were constructed using the gene replacement technique (Rudolph *et al.*, 1985). The complete *TRP4* promoter was first evicted (from positions −279 to +196, relative to the translational start site at position +1) and then substituted by the *URA3* gene. *TRP4* promoter mutant alleles obtained by site-directed point mutagenesis were then re-introduced, replacing the *URA3* gene. Linear fragment yeast transformations were performed using the Li-acetate treatment method (Ito *et al.*, 1983). The integration of the mutant alleles at the original *TRP4* locus on the chromosome was confirmed using the Southern blot technique (Southern, 1975). *TRP4* gene replacements were performed in two isogenic derivatives of the *S.cerevisiae* laboratory strain S288C: RH1408 (*MAT***a** *ura3-52 gcn4-103*) and RH1378 (*MAT***a** *Δura3 gcd2-1*). The *gcn4-103* and the *gcd2-1* mutations have been described earlier (Hinnebusch, 1985; Niederberger *et al.*, 1986). The *gcn4-103* mutation contains a large deletion of the *GCN4* gene whereas the *gcd2-1* mutation causes constitutively high amounts of GCN4 protein in the cell.

### *Site-directed mutagenesis*

Point mutations were generated using the Muta-Gene *in vitro* mutagenesis kit from Bio-Rad (Richmond, CA, USA), based on a method described by Kunkel (1985). The promoter regions of all *TRP4* mutant alleles obtained by this procedure were sequenced using the dideoxy method (Sanger *et al.*, 1977) thereby ruling out possible second site mutations.

### *Northern analysis*

Poly(A)$^+$ RNA was isolated as described earlier (Furter *et al.*, 1986). For Nothern hybridization poly(A)$^+$ RNA was separated on a formaldehyde–agarose gel, electroblotted onto a nylon membrane and hybridized with DNA fragments labelled according to the 'oligo-labelling' technique described by Feinberg and Vogelstein (1984). Band intensity on autoradiograms was quantified using a Molecular Dynamics Computing Densitometer model 300A (Sunnyvale, CA, USA).

### *Primer extension analysis*

Primer extension analysis was performed according to Kassavetis and Geiduschek (1982). 100 μg of poly(A)$^+$ RNA and 5 × $10^6$ c.p.m. of either a 5′-end-labelled 50 bp primer (from positions +29 to +79 relative to the translation start site of the *TRP4* gene) or a 5′-end-labelled 37 bp primer (from positions −108 to +144) were hybridized in 20 mM Tris (pH 8.3), 200 mM NaCl and 0.1 mM EDTA in a 30 ml assay at 60°C for 2 h. Annealed primers were then elongated using AMV reverse transcriptase in 100 mM Tris (pH 8.3), 10 mM $MgCl_2$, 30 mM NaCl, 10 mM DTT and in the presence of all four DNA constituting deoxyribonucleotide triphosphates in a 100 ml reaction at 42°C for 1 h. The synthesized cDNAs were separated on a 6% polyacrylamide–7 M urea standard sequencing gel, together with the four dideoxy sequencing reactions (Sanger *et al.*, 1977), obtained by using the identical 50 bp or 37 bp primers as for poly(A)$^+$ RNA elongation reactions. Band intensities on autoradiograms were quantified using a Molecular Dynamics Computing Densitometer model 300A (Sunnyvale, CA, USA).

### *Media for yeast strain cultivation*

Yeast strains were cultivated in YEPD complete medium or MV minimal medium supplemented with uracil (40 mg/l) and arginine (40 mg/l) as described earlier (Miozzari *et al.*, 1978).

## Acknowledgements

We thank R.Hütter for generous support. We are grateful to D.Jones and M.Künzler for critical reading of the manuscript and S.Irniger for helpful discussions. We also appreciate the synthesis of oligonucleotides by P.Kast. This work was supported by the Swiss National Science Foundation Grant No. 31.29926.90, by grants of the Swiss Federal Institute of Technology and by a grant from the 'Huber-Kudlich' Foundation.

## References

Arndt,K.T., Styles,C.A. and Fink,G.R. (1987) *Science*, **237**, 874–880.
Braus,G.H. (1991) *Microbiol. Rev.*, **55**, 349–370.
Braus,G., Mösch,H.-U., Vogel,K., Hinnen,A. and Hütter,R. (1989) *EMBO J.*, **8**, 939–945.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.
Buratowski,S., Hahn,S., Sharp,P.A. and Guarente,L. (1988) *Nature*, **334**, 37–42.
Buratowski,S., Hahn,S., Guarente,L. and Sharp,P.A. (1989) *Cell*, **56**, 549–561.
Chen,W. and Struhl,K. (1985) *EMBO J.*, **4**, 3273–3280.
Chen,W. and Struhl,K. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2691–2695.
Corden,J., Wasylyk,B., Buchwalder,A., Sassone-Corsi,P., Kedinger,C. and Chambon,P. (1980) *Science*, **209**, 1405–1414.
Dynan,W.S. and Tjian,R. (1985) *Nature*, **316**, 774–777.
Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266–267.
Furter,R., Paravicini,G., Aebi,M., Braus,G., Prantl,F., Niederberger,P. and Hütter,P. (1986) *Nucleic Acids Res.*, **14**, 6357–6373.
Furter,R., Braus,G., Paravicini,G., Mösch,H.-U., Niederberger,P. and Hütter,R. (1988) *Mol. Gen. Genet.*, **211**, 168–175.
Furter-Graves,E.M. and Hall,B.D. (1990) *Mol. Gen. Genet.*, **223**, 407–416.
Grosschedl,R. and Birnstiel,M.L. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1432–1436.
Guarente,L. (1987) *Annu. Rev. Genet.*, **21**, 425–452.
Guarente,L. (1988) *Cell*, **52**, 303–305.
Hahn,S., Hoar,E.T. and Guarente,L. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8562–8566.
Harbury,P.A.B. and Struhl,K. (1989) *Mol. Cell. Biol.*, **9**, 5298–5304.
Healy,A.M. and Zitomer,R.S. (1990) *Curr. Genet.*, **18**, 105–109.
Healy,A.M., Helser,T.L. and Zitomer,R.S. (1987) *Mol. Cell. Biol.*, **7**, 3785–3791.
Hinnebusch,A.G. (1985) *Mol. Cell. Biol.*, **5**, 2349–2360.
Hinnebusch,A.G. (1988) *Microbiol. Rev.*, **52**, 248–273.
Hinnebusch,A.G. and Fink,G.R. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5374–5378.
Hultmark,D.R., Klemenz,R. and Gehring,J.W. (1986) *Cell*, **44**, 429–438.
Ito,H., Jukuda,Y., Murata,K. and Kimura,A. (1983) *J. Bacteriol.*, **153**, 163–168.
Kassavetis,G.A. and Geiduschek,E.P. (1982) *EMBO J.*, **1**, 107–114.
Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488–492.
Maicas,E. and Friesen,J.D. (1990) *Nucleic Acids Res.*, **18**, 3387–3393.
McNeil,J.B. and Smith,M. (1986) *J. Mol. Biol.*, **187**, 363–378.
Miozzari,G., Niederberger,P. and Hütter,R. (1978) *J. Bacteriol.*, **134**, 48–59.
Mösch,H.-U., Graf,R., Schmidheini,T. and Braus,G. (1990) *EMBO J.*, **9**, 2951–2957.
Nagawa,F. and Fink,G.R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8557–8561.
Nakatani,Y., Horikoshi,M., Brenner,M., Yamamoto,T., Besnard,F., Roeder,R.G. and Freese,E. (1990) *Nature*, **348**, 86–88.
Niederberger,P., Aebi,M. and Hütter,R. (1986) *Curr. Genet.*, **10**, 657–664.
Rathjen,J. and Mellor,J. (1990) *Nucleic Acids Res.*, **18**, 3219–3225.
Rudolph,H., Koenig-Rauseo,I. and Hinnen,A. (1985) *Gene*, **36**, 87–95.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Smale,T.S. and Baltimore,D. (1989) *Cell*, **57**, 103–113.
Smale,S.T., Schmidt,M.C., Berk,A.J. and Baltimore,D. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4509–4513.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
Struhl,K. (1986) *Mol. Cell. Biol.*, **6**, 3847–3853.
Struhl,K. (1987) *Cell*, **49**, 295–297.
Struhl,K. (1989) *Annu. Rev. Biochem.*, **58**, 1051–1077.
Thireos,G., Driscoll-Penn,M. and Greer,H. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5096–5100.
Van Dyke,M.W., Rhoeder,R.G. and Sawadogo,M. (1988) *Science*, **241**, 1335–1338.
Wobbe,C.R. and Struhl,K. (1990) *Mol. Cell. Biol.*, **10**, 3859–3867.

*Received on March 2, 1992; revised on August 6, 1992*