Volume 14 Number 16 1986 Nucleic Acids Research

The *TRP4* gene of *Saccharomyces cerevisiae*: isolation and structural analysis

Rolf Furter*, Gerhard Paravicini, Markus Aebi, Gerhard Braus, Franziska Prantl, Peter Niederberger and Ralf Hütter

Mikrobiologisches Institut, Eidgenössische Technische Hochschule, ETH-Zentrum, CH-8092 Zürich, Switzerland

Received 3 July 1986; Accepted 29 July 1986

ABSTRACT

The TRP4 gene of Saccharomyces cerevisiae, encoding the anthranilate phosphoribosyl transferase, was isolated and subcloned by functional complementation in yeast. A 2 kb fragment containing information for a polypeptide of 380 amino acids and the 5'- and 3'-flanking regions was sequenced. The TRP4 transcript was identified and mapped with S1 nuclease. Homologies to two prokaryotic genes encoding the same function, and sequences potentially involved in transcription start and termination and in regulation of TRP4 gene expression are discussed.

INTRODUCTION

In Saccharomyces cerevisiae five unlinked genes encode five tryptophan biosynthetic enzymes (1). The expression of four of these genes is regulated by the general control of amino acid biosynthesis (2,3).

Genes TRP1 and TRP5 have been isolated by complementation of equivalent mutations in Escherichia coli (4,5), the other three genes, TRP2, TRP3, and TRP4 by functional complementation in yeast (6; this work). The nucleotide sequence of TRP1, TRP2, TRP3, and TRP5 have been established (7,8,9,10).

The TRP4 gene of S.cerevisiae encodes the monofunctional anthranilate phosphoribosyl transferase (PRtransferase EC 2.4.2.18), catalyzing the second step in the tryptophan biosynthesis, and is subject to the general control of amino acid biosynthesis (2,3). In some prokaryotic organisms such as E.coli and Salmonella typhimurium, the PRtransferase is fused with the carboxylterminus of the anthranilate synthase subunit II (11).

In this report we describe the isolation of a yeast DNA fragment carrying the TRP4 gene. We present studies on the nucleotide sequence and transcription of this gene. We show that the cloned and sequenced fragment contains the complete coding sequence, and the 5'- and 3'-flanking sequences, necessary for expression and regulation under the general control.

MATERIALS AND METHODS

Strains and plasmids

All mutant strains used in this study are derived from the two haploid wild type strains X2180-1A MATa or X2180-1B MATα: RH375 MATa gcn1-1, RH993 MATa trp4-21 leu2-2 met8-1, RH1191 MATα ade6 ura3, RH1227 MATa ura3 trp3B-4 trp4-21. RH1244 MATa ura3-251 ura3-328 ura3-373 was obtained from F. Lacroute, Strasbourg. Escherichia coli strains HB101 r_K^- m_K^- leuB6 proA2 recA13 and GM242 dam$^-$ were used for propagation of plasmid DNA; strain JA196 trpC1117 leuB6 thi hsr_K was used for the preparation of the gene library. Plasmid pJDB207 was obtained from J. Beggs, London, and plasmid YEp24 from P. Philippsen, Basel. Plasmid pME512 was constructed by replacing the "mini" pBR322 sequence of pJDB207 with the whole pBR322 sequence.

Media

MV-medium for yeast was described previously (2). E.coli media were prepared according to Vogel and Bonner (12).

Cloning procedures

The methods for plasmid isolation, endonuclease digestion, plasmid pool construction and ligation were described earlier by Aebi et al. (6). Deletions were constructed by treating the linearized plasmids with exonuclease III and nuclease SI (Boehringer, Mannheim) and ligating in the presence of linkers according to Roberts and Lauer (13).

Transformation of yeast and *E.coli*

The methods used were previously described (6).

Southern analysis

The modification of Smith and Summers (14) of the method of Southern (15) was used.

DNA sequencing

The method of Maxam and Gilbert (16) was used with the following modification: The A+G reaction was performed in 80 % formic acid (instead of piperidine) at room temperature for 6 min. Fragments were isolated from 5 % polyacrylamide gels by electroelution of ethidiumbromide-stained DNA fragments onto NA45 anion-exchange paper (Schleicher and Schuell, Feldbach, Switzerland). DNA was eluted with 2M NaCl, 5fold concentrated TAE buffer (17) at 68°C for 15 min, precipitated with 2 volumes of ethanol, resuspended in 50 µl of deion. water, and reprecipitated with 1 ml ethanol, in an ethanol/dry ice bath for 10 min. (^{32}P)-labelled nucleotides were obtained from Amersham.

Poly(A)$^+$-RNA isolation

Exponentially growing cells were harvested, disrupted with acid-washed glassbeads (Ø 0.45 mm) by vigorous shaking for 1 min in the presence of extraction buffer [100 mM PIPES, pH 7.5, 100 mM $LiCl_2$, 1 mM EDTA, 1 % (w/v) SDS, 1 % (v/v) diethylpyrocarbonate and 0.2 volume of phenol/chloroform (1:1, v/v)]. The aqueous phase was re-extracted twice with phenol/chloroform, precipitated twice with 1.5 volumes of isopropanol for 20 min on ice. Total RNA was enriched on oligo(dT)-cellulose as described by Aviv and Leder (18) and Bantle et al. (19). The eluted poly(A)$^+$-RNA was extracted with phenol/chloroform and stored precipitated at -20°C.

S1 nuclease mapping

50 µg poly(A)$^+$-RNA were hybridized with endlabelled fragments (20'000 cpm) in sealed glass capillaries (20,21). The temperature was decreased from 90°C to the hybridization temperature of 46°C. S1 nuclease digestion (Boehringer, Mannheim) was done with 20 to 80 enzyme units. These variations yielded identical protection patterns. The S1-resistant hybrids were analyzed on 6 % polyacrylamide/7M urea sequencing gels. Gels were dried and autoradiographed on Fuji X-ray films.

PRtransferase assay

PRtransferase was assayed *in situ* according to Miozzari et al. (2) in Triton X-100 permeabilized cells (22). As a source for PRAisomerase, strain RH218 carrying the TRP1/EcoRI circle (23) was used instead of the RH28 *trp4* mutant.

RESULTS

Cloning of the *TRP4* gene

A population of BglII fragments from total genomic DNA of strain RH375 gcn1-1 was inserted into the BamHI site of the yeast 2 µm DNA-based plasmid pJDB207. This gene library was used to transform strain RH993 *trp4 leu2 met8*, and Trp$^+$Leu$^+$ colonies were selected. Recombinant plasmids were recovered from such transformants after transformation of *E.coli* strain HB101. A representative isolate, designated pME506 was found to carry a 6 kb insert and its restriction map was determined (Fig. 1a). The location of the TRP4 gene on the BglII fragment was determined by insertion of various subfragments into plasmids pJDB207 or pME512. Their function was tested by the ability to transform strain RH993 to Trp$^+$Leu$^+$ at high frequency. The location of these frag-

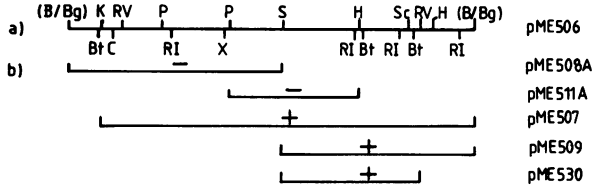

Figure 1. Mapping of the TRP4 region on a cloned 6 kb BglII restriction fragment. a) map of restriction sites for BamHI (B), BglII (Bg), BstEII (Bt), ClaI (C), HindIII (H), KpnI (K), EcoRI (RI), EcoRV (RV), SalI (S), SacI (Sc) and XhoI (X). b) Subclones were tested for their complementing ability in the trp4 strain RH993 as pJDB207 derivatives. Symbols (+/-) give the qualitative results of Trp^+ transformation for each subclone.

ments within the BglII insert and the qualitative behaviour in the transformation experiment are depicted in Fig. 1B. The SalI-EcoRV fragment carried on the 2 µm DNA-based multicopy plasmid pME530 yielded about 2000 Trp^+Leu^+ transformants per µg plasmid DNA.

It was shown by Southern hybridization that the cloned DNA-fragment contains a unique sequence in the yeast genome. For this purpose the SalI-EcoRV fragment was nick-translated, and hybridized against a BglII and a SalI-EcoRV digest of total yeast DNA. As expected from the restriction map (Fig. 1a) a single 6 kb and a single 2 kb fragment was generated in the first and the second case, respectively (data not shown).

It was verified further by a "gene disruption" experiment (24) that the cloned SalI-EcoRV fragment indeed carries the TRP4 gene. For that purpose the URA3 gene (derived as a HindIII fragment from vector YEp24; 25) was inserted into the single HindIII site in the presumptive TRP4 gene (see Fig. 1). 30 µg of the linear SalI-SacI DNA fragment were used to transform the ura3 mutant strain RH1244 to Ura^+. Of 45 Ura^+ transformants, 23 were Trp^-, indicating an integration event at the TRP4 gene locus. One of the Ura^+Trp^- transformants was then crossed to a second ura3 mutant strain RH1191. The meiotic segregants were obtained by the "random spore" technique and analysed for their phenotypes (Table I). The absence of Ura^-Trp^- segregants showed the genetic linkage between the Trp^- phenotype, generated by the integration event of the URA3-disrupted trp4 fragment, and the TRP4 gene. The unexpectedly high rate of Trp^+ Ura^+ segregants probably stems from diploid or aneuploid fusion products which arose in the preceeding PEG-mediated yeast transformation experiment.

Integration of the URA3 gene at the TRP4 locus was visualized further by

Table I. Random spore analysis of the cross between a Ura^+Trp^- transformant and strain RH1191 *ura3 ade6*

Phenotype	Number of segregants
Trp^+Ura^-	69
Trp^-Ura^+	61
Trp^+Ura^+	12
Trp^-Ura^-	0

a Southern hybridization analysis. The SalI-EcoRV fragment was used as probe. The Southern patterns of EcoRV-digested chromosomal DNA, isolated from two independant Ura^+Trp^- transformants showed the expected change, as compared with the original untransformed strains RH1244 (Fig. 2).

A final proof for the isolation of the genuine TRP4 gene was the high activity of PRtransferase in strain RH993, transformed with the 2 µm DNA-based multicopy plasmid pME530. Such cells derived from a liquid culture of a transformant showed a 120-fold increase in enzyme activity as compared to the wild strain X2180-1A (140 nmoles per min and mg protein as compared to 1.2 in the

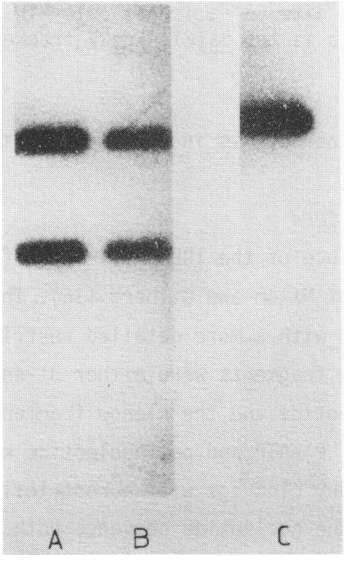

Figure 2. Southern analysis of integrative transformants. Strain RH1244 ura3 was transformed to Ura^+ prototrophy with a linear DNA fragment, carrying the URA3-disrupted trp4 gene. Total DNA from two independent Ura^+Trp^- transformants (A,B) and the non-transformed strain RH1244 (C) was digested with EcoRV and probed with the nick-translated SalI-EcoRV TRP4 fragment.

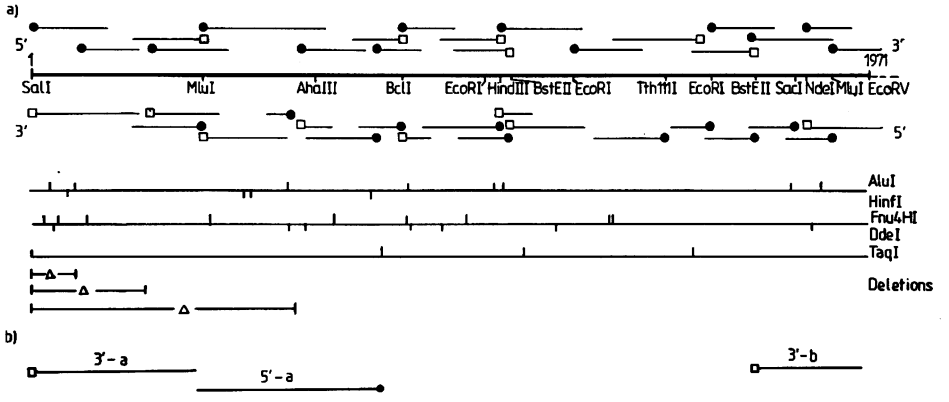

Figure 3. Restriction endonuclease sites and sequencing strategy for the TRP4 gene. a) Sequence strategy for both strands of the 1971 bp SalI-EcoRV fragment. Not all sites indicated were used for end-labelling. In 3 cases, deletion plasmids were used for sequencing. In these constructions RNA was deleted starting from the SalI site which was restored with synthetic linkers 5'-end-labelling with (γ-^{32}P)-ATP and polynucleotide kinase (●); 3'-end-labelling with an (α-^{32}P)-deoxynucleotide and the Klenow fragment of DNA polymerase I (□). The lines indicate the extension of readable sequence information. b) Probes used in S1 nuclease mapping. 5'-a is the TaqI-MluI-fragment 5'-end-labelled (●) at the TaqI site. 3'-a is the SalI-MluI-fragment 3'-end-labelled at the SalI site, 3'-b is the BstEII-EcoRV fragment 3'-end-labelled at the BstEII site.

wild type strain). This increase is in the range of the expected copy number for this vector (26).

DNA-sequence analysis of *TRP4*

The nucleotide sequence of the TRP4 SalI-EcoRV fragment was determined according to the method of Maxam and Gilbert (16). The sequencing strategy is shown in Fig. 3a together with a more detailed restriction map derived from the sequence. Overlapping fragments were either 3'-end-labelled with an appropriate (α-^{32}P)-deoxynucleotide and the Klenow fragment of DNA-polymerase I or 5'-end-labelled with (γ-^{32}P)-ATP and polynucleotide kinase. They were prepared for sequencing by secondary cleavage with a restriction enzyme or by strand separation. For 98 % of the nucleotide sequence both strands were sequenced. Each restriction site, used for end-labelling was oversequenced at least once.

Deduced amino acid sequence of *TRP4* protein

The 1971 bp SalI-EcoRV DNA fragment encoding the PRtransferase activity contains a single long open reading frame. The deduced amino acid sequence is

presented in Fig. 4. It starts with the ATG codon (position +1) and extends to the first TAG stop codon at position +1141. The calculated molecular weight (M_r) for this peptide chain of 380 amino acids is 41'370 d. The deduced amino acid sequences of the corresponding genes from E.coli trpD (27), and B.subtilis (28) were aligned with that of the yeast TRP4 gene. Homology between the sequences was determined in blocks of 20 amino acids (Fig. 5; deletions or insertions of amino acids are indicated but not considered). Within two regions A and B a simultaneous beyond-average homology between E.coli and B.subtilis (> 34%) and between E.coli and S.cerevisiae (>27%) was found. It seems likely that these more strongly conserved regions define binding sites for the substrate anthranilate on the one hand and phosphoribosyl pyrophosphate on the other.

Codon usage in the TRP4 sequence was compiled in Table II. The codon bias index, defined according to Bennetzen and Hall (29), is 0.21. This number lies in the same range as it was found for iso-2-cytochrome C (0.15), which constitutes 0.003 % of total mRNA (29). The proportion of PRtransferase compared to the total cell protein can be estimated to be around 0.01 %. This was judged from a protein band appearing on SDS polyacrylamide gels from different yeast transformants overexpressing the enzyme 130- and 400-fold (Fig. 6, tracks 4 and 5).

Transcription of the *TRP4* gene and 5'-flanking sequences

Transcription of the TRP4 gene from the SalI-EcoRV fragment was studied by Northern analysis. Poly(A)$^+$-RNA isolated from the wild type strain X2180-1A was separated on a 1.5 % agarose gel under denaturing conditions, transferred to nitrocellulose filters, and hybridized against the nick-translated ^{32}P-labelled SalI-EcoRV fragment of plasmid pME530. Two transcripts of 1.4 kb and 1.25 kb were detected (data not shown). To identify the TRP4 transcript two experiments were done: In the first one, the poly(A)$^+$-RNA filter was probed with different subfragments of the SalI-EcoRV fragment. This procedure showed that the SalI-MluI fragment (-668 to -279; see Fig. 3) hybridizes only with the shorter transcript, whereas all fragments in the MluI-EcoRV segment (-279 to +1303) hybridize to the 1.4 kb transcript. This finding excludes the possibility that the shorter transcript is a degradation product of the longer one and shows that the TRP4 gene encodes the 1.4 kb transcript. In the second experiment, the poly(A)$^+$-RNA filter was hybridized with two RNA probes derived from the SalI-SacI fragment of plasmid pME530, and inserted in both di-

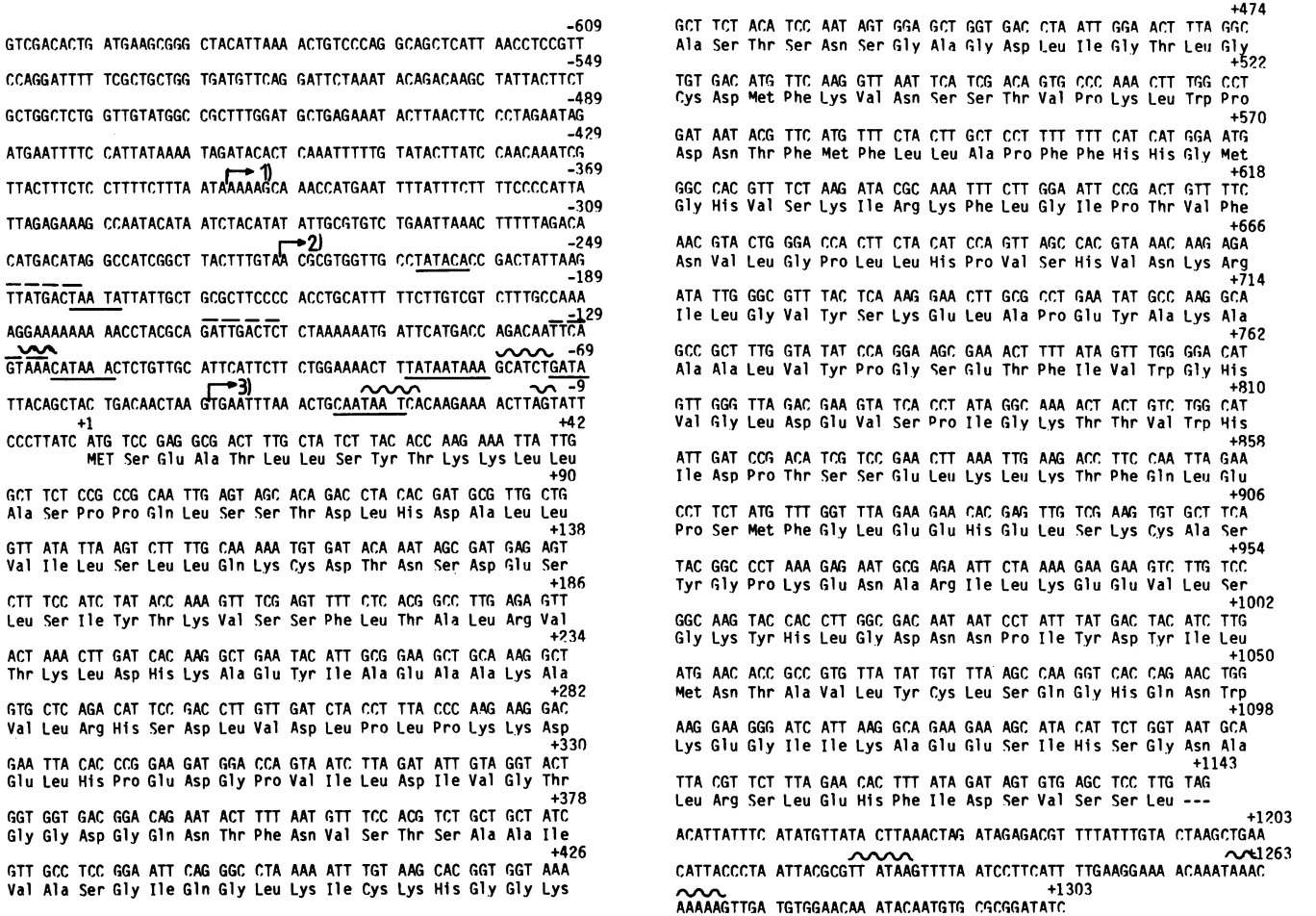
-609
GTCGACACTG ATGAAGCGGG CTACATTAAA ACTGTCCCAG GCAGCTCATT AACCTCCGTT
-549
CCAGGATTTT TCGCTGCTGG TGATGTTCAG GATTCTAAAT ACAGACAAGC TATTACTTCT
-489
GCTGGCTCTG GTTGTATGGC CGCTTTGGAT GCTGAGAAAT ACTTAACTTC CCTAGAATAG
-429
ATGAATTTTC CATTATAAAA TAGATACACT CAAATTTTTG TATACTTATC CAACAAATCG
-369
TTACTTTCTC CTTTTCTTTA ATAAAAAGCA AACCATGAAT TTTATTTCTT TTCCCCATTA
-309
TTAGAGAAAG CCAATACATA ATCTACATAT ATTGCGTGTC TGAATTAAAC TTTTTAGACA
-249
CATGACATAG GCCATCGGCT TACTTTGTAA CGCGTGGTTG CCTATACACC GACTATTAAG
-189
TTATGACTAA TATTATTGCT GCGCTTCCCC ACCTGCATTT TTCTTGTCGT CTTTGCCAAA
-129
AGGAAAAAAA AACCTACGCA GATTGACTCT CTAAAAAATG ATTCATGACC AGACAATTCA
-69
GTAAACATAA ACTCTGTTGC ATTCATTCTT CTGGAAAACT TTATAATAAA GCATCTGATA
-9
TTACAGCTAC TGACAACTAA GTGAATTTAA ACTGCAATAA TCACAAGAAA ACTTAGTATT
+1
+42
CCCTTATC ATG TCC GAG GCG ACT TTG CTA TCT TAC ACC AAG AAA TTA TTG
MET Ser Glu Ala Thr Leu Leu Ser Tyr Thr Lys Lys Leu Leu
+90
GCT TCT CCG CCG CAA TTG AGT AGC ACA GAC CTA CAC GAT GCG TTG CTG
Ala Ser Pro Pro Gln Leu Ser Ser Thr Asp Leu His Asp Ala Leu Leu
+138
GTT ATA TTA AGT CTT TTG CAA AAA TGT GAT ACA AAT AGC GAT GAG AGT
Val Ile Leu Ser Leu Leu Gln Lys Cys Asp Thr Asn Ser Asp Glu Ser
+186
CTT TCC ATC TAT ACC AAA GTT TCG AGT TTT CTC ACG GCC TTG AGA GTT
Leu Ser Ile Tyr Thr Lys Val Ser Ser Phe Leu Thr Ala Leu Arg Val
+234
ACT AAA CTT GAT CAC AAG GCT GAA TAC ATT GCG GAA GCT GCA AAG GCT
Thr Lys Leu Asp His Lys Ala Glu Tyr Ile Ala Glu Ala Ala Lys Ala
+282
GTG CTC AGA CAT TCC GAC CTT GTT GAT CTA CCT TTA CCC AAG AAG GAC
Val Leu Arg His Ser Asp Leu Val Asp Leu Pro Leu Pro Lys Lys Asp
+330
GAA TTA CAC CCG GAA GAT GGA CCA GTA ATC TTA GAT ATT GTA GGT ACT
Glu Leu His Pro Glu Asp Gly Pro Val Ile Leu Asp Ile Val Gly Thr
+378
GGT GGT GAC GGA CAG AAT ACT TTT AAT GTT TCC ACG TCT GCT GCT ATC
Gly Gly Asp Gly Gln Asn Thr Phe Asn Val Ser Thr Ser Ala Ala Ile
+426
GTT GCC TCC GGA ATT CAG GGC CTA AAA ATT TGT AAG CAC GGT GGT AAA
Val Ala Ser Gly Ile Gln Gly Leu Lys Ile Cys Lys His Gly Gly Lys
+474
GCT TCT ACA TCC AAT AGT GGA GCT GGT GAC CTA ATT GGA ACT TTA GGC
Ala Ser Thr Ser Asn Ser Gly Ala Gly Asp Leu Ile Gly Thr Leu Gly
+522
TGT GAC ATG TTC AAG GTT AAT TCA TCG ACA GTG CCC AAA CTT TGG CCT
Cys Asp Met Phe Lys Val Asn Ser Ser Thr Val Pro Lys Leu Trp Pro
+570
GAT AAT ACG TTC ATG TTT CTA CTT GCT CCT TTT TTT CAT CAT GGA ATG
Asp Asn Thr Phe Met Phe Leu Leu Ala Pro Phe Phe His His Gly Met
+618
GGC CAC GTT TCT AAG ATA CGC AAA TTT CTT GGA ATT CCG ACT GTT TTC
Gly His Val Ser Lys Ile Arg Lys Phe Leu Gly Ile Pro Thr Val Phe
+666
AAC GTA CTG GGA CCA CTT CTA CAT CCA GTT AGC CAC GTA AAC AAG AGA
Asn Val Leu Gly Pro Leu Leu His Pro Val Ser His Val Asn Lys Arg
+714
ATA TTG GGC GTT TAC TCA AAG GAA CTT GCG CCT GAA TAT GCC AAG GCA
Ile Leu Gly Val Tyr Ser Lys Glu Leu Ala Pro Glu Tyr Ala Lys Ala
+762
GCC GCT TTG GTA TAT CCA GGA AGC GAA ACT TTT ATA GTT TGG GGA CAT
Ala Ala Leu Val Tyr Pro Gly Ser Glu Thr Phe Ile Val Trp Gly His
+810
GTT GGG TTA GAC GAA GTA TCA CCT ATA GGC AAA ACT ACT GTC TGG CAT
Val Gly Leu Asp Glu Val Ser Pro Ile Gly Lys Thr Thr Val Trp His
+858
ATT GAT CCG ACA TCG TCC GAA CTT AAA TTG AAG ACC TTC CAA TTA GAA
Ile Asp Pro Thr Ser Ser Glu Leu Lys Leu Lys Thr Phe Gln Leu Glu
+906
CCT TCT ATG TTT GGT TTA GAA GAA CAC GAG TTG TCG AAG TGT GCT TCA
Pro Ser Met Phe Gly Leu Glu Glu His Glu Leu Ser Lys Cys Ala Ser
+954
TAC GGC CCT AAA GAG AAT GCG AGA ATT CTA AAA GAA GAA GTC TTG TCC
Tyr Gly Pro Lys Glu Asn Ala Arg Ile Leu Lys Glu Glu Val Leu Ser
+1002
GGC AAG TAC CAC CTT GGC GAC AAT AAT CCT ATT TAT GAC TAC ATC TTG
Gly Lys Tyr His Leu Gly Asp Asn Asn Pro Ile Tyr Asp Tyr Ile Leu
+1050
ATG AAC ACC GCC GTG TTA TAT TGT TTA AGC CAA GGT CAC CAG AAC TGG
Met Asn Thr Ala Val Leu Tyr Cys Leu Ser Gln Gly His Gln Asn Trp
+1098
AAG GAA GGG ATC ATT AAG GCA GAA GAA AGC ATA CAT TCT GGT AAT GCA
Lys Glu Gly Ile Ile Lys Ala Glu Glu Ser Ile His Ser Gly Asn Ala
+1143
TTA CGT TCT TTA GAA CAC TTT ATA GAT AGT GTG AGC TCC TTG TAG
Leu Arg Ser Leu Glu His Phe Ile Asp Ser Val Ser Ser Leu ---
+1203
ACATTATTTC ATATGTTATA CTTAAACTAG ATAGAGACGT TTTATTTGTA CTAAGCTGAA
+1263
CATTACCCTA ATTACGCGTT ATAAGTTTTA ATCCTTCATT TTGAAGGAAA ACAAATAAAC
+1303
AAAAAGTTGA TGTGGAACAA ATACAATGTG CGCGGATATC

Table II. Codon usage in the coding region of the *S.cerevisiae TRP4* gene.

TTT	Phe	9	TCT	Ser	8	TAT	Tyr	5	TGT	Cys	5
TTC	Phe	4	TCC	Ser	9	TAC	Tyr	6	TGC	Cys	0
TTA	Leu	13	TCA	Ser	4	TAA	---	0	TGA	---	0
TTG	Leu	13	TCG	Ser	4	TAG	---	0	TGG	Trp	4
CTT	Leu	11	CCT	Pro	8	CAT	His	7	CTG	Arg	1
CTC	Leu	2	CCC	Pro	2	CAC	His	10	CGC	Arg	1
CTA	Leu	8	CCA	Pro	4	CAA	Gln	4	CGA	Arg	0
CTG	Leu	2	CCG	Pro	5	CAG	Gln	3	CGG	Arg	0
ATT	Ile	10	ACT	Thr	9	AAT	Asn	10	AGT	Ser	6
ATC	Ile	5	ACC	Thr	4	AAC	Asn	4	AGC	Ser	7
ATA	Ile	7	ACA	Thr	5	AAA	Lys	12	AGA	Arg	4
ATG	Met	6	ACG	Thr	3	AAG	Lys	16	AGG	Arg	0
GTT	Val	13	GCT	Ala	11	GAT	Asp	10	GGT	Gly	9
GTC	Val	2	GCC	Ala	5	GAC	Asp	9	GGC	Gly	8
GTA	Val	6	GCA	Ala	4	GAA	Glu	18	GGA	Gly	10
GTG	Val	4	GCG	Ala	5	GAG	Glu	4	GGG	Gly	2

The number of codon triplets used in the coding region of the TRP4 gene is shown. The preferred codons according to Bennetzen and Hall (29) are underlined.

rections in the two SP6 transcription vectors SP64 and SP65 (Fig. 7). Only the RNA transcribed from SP65 (plasmid pME562) yielded signals, again as in the Northern experiment of the sizes 1.25 kb and 1.4 kb. This shows that both transcripts move in the same direction from the SalI towards the SacI site.

Mapping of the 5'- and 3'-ends of the *TRP4* transcript

The S1-mapping technique of Berk and Sharp (20), as modified by Weaver and Weissmann (21), was used to map the 5'- and 3'-ends of the 1.4 kb TRP4 transcript and the 3'-end of the 1.25 kb transcript (of unidentified nature). The fragments used as probes are shown in Fig. 3b and designated 5'-a, 3'-b, and 3'-a.

For the mapping of the TRP4 transcription start sites the TaqI-SalI fragment (+159 to -668) was labelled with polynucleotide kinase in the presence of (γ-^{32}-P)-ATP and recut with MluI (-279). The TaqI-MluI-fragment (5'-a; +159 to -279) was isolated and hybridized against wild type poly(A)$^+$-RNA. The

Figure 4. Nucleotide sequence and deduced amino acid sequence of S.cerevisiae TRP4 gene. The nucleotide sequence is shown in the 5'- to 3'-direction and numbered from the initiation codon ATG (+1). Initiation and termination sites for transcription in the 5'- and 3'-sequences respectively, are marked with wavy lines. Postulated TATA elements are underlined, consensus sequences according to Donahue et al. (32) are overlined with dashed lines.

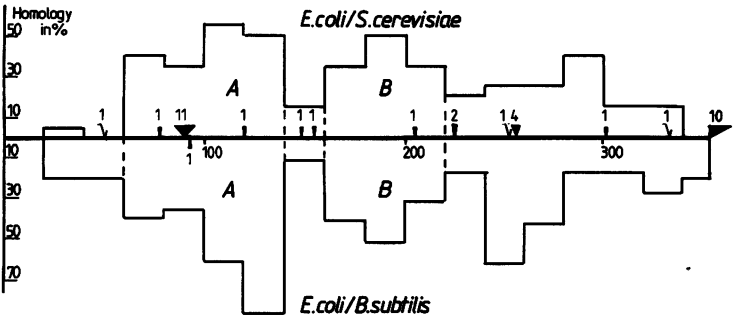

Figure 5. Homology between different PRtransferase sequences. The amino acid sequences of the PRtransferase of E.coli and S.cerevisiae (upper part) and E.coli and B.subtilis (lower part) were compared. Position and number of amino acids inserted with the yeast sequence (▼) compared to the E.coli sequence and within the E.coli sequence (√) compared to the yeast sequence are indicated but not counted for the comparison. Only in block A (amino acids 60-120) and block B (amino acids 140-220) a simultaneous beyond-average homology was observed for the three organisms.

S1-resistant probes were separated, together with samples of the four Maxam and Gilbert reactions of the same fragment, on a 6 % polyacrylamide/7M urea DNA sequencing gel. The autoradiogram of this analysis is shown in Fig. 8. One major start site between -127 and -123, and minor start sites around -76, -31 to -26 and -14 to -12 were found.

For the mapping of the 3'-end of the TRP4 transcript, the double stranded BstEII-EcoRV fragment (3'-b; +1036 to +1303) was labelled with polymerase I (Klenow fragment) in the presence of (α-^{32}P)-dGTP at the 3'-end of the BstEII site (+1036). This probe was hybridized against wild type poly(A)$^+$-RNA. The S1-resistant DNA pieces were separated together with the Maxam and Gilbert reactions of the SalI-MluI fragment on a DNA-sequencing gel. Two major transcription termination sites for the TRP4 gene around positions +1225 and +1265 (Fig. 9) were found.

Deduced from the S1-mapping data, the TRP4 message has a length of between 1350 to 1390 bp, which agrees with the Northern data.

The 3'-end of the unidentified 1.25 kb transcript was mapped by labelling the SalI-MluI fragment (3'-a; -668 to -279) with DNA polymerase I (Klenow fragment) in the presence of (α-^{32}P)-dTTP at the 3'-end of the SalI site (-668). This probe was hybridized against poly(A)$^+$-RNA from pME506-transformed cells of strain RH1227. The S1-resistant DNA pieces were separated on a

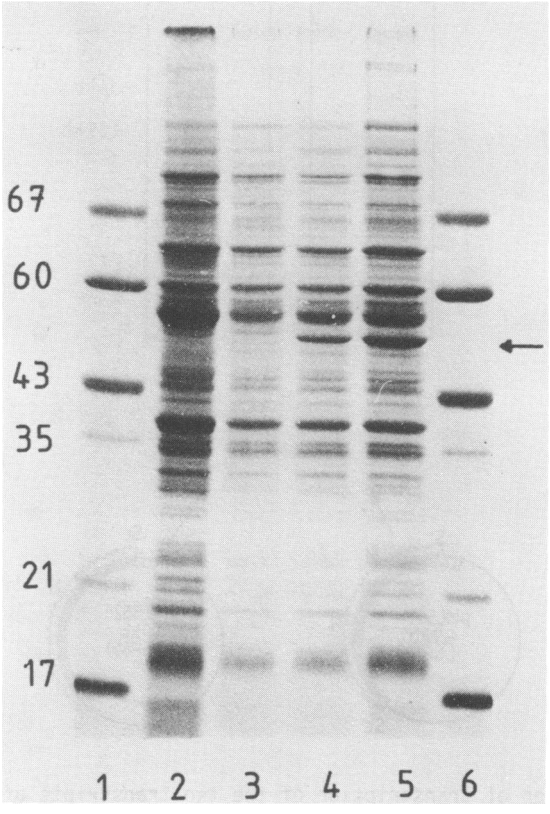

Figure 6. Identification of the PRtransferase on a SDS polyacrylamide gel. Crude extracts of different yeast strains were separated on a 10 % SDS polyacrylamide gel. Track 1,6: marker proteins with the indicated M_r; 2: crude extract from strain X2180-1A; 3: strain RH995 (pME512); 4: strain RH993 (pME530), overexpression of PRtransferase 130-fold; 5: strain RH995 (pME530), overexpression 400-fold. The arrow indicates the PRtransferase, estimated to constitute 1%-2% of total protein.

DNA-sequencing gel as described above. Two major transcription termination sites for the 1.25 kb transcript at positions -427 to -401 and -390 to -375 were found (Fig. 9).

DISCUSSION

The *TRP4* coding sequence

A unique DNA sequence was isolated from a yeast gene library, which (i) complements a trp4 mutation, (ii) leads to a copy-number effect of the TRP4

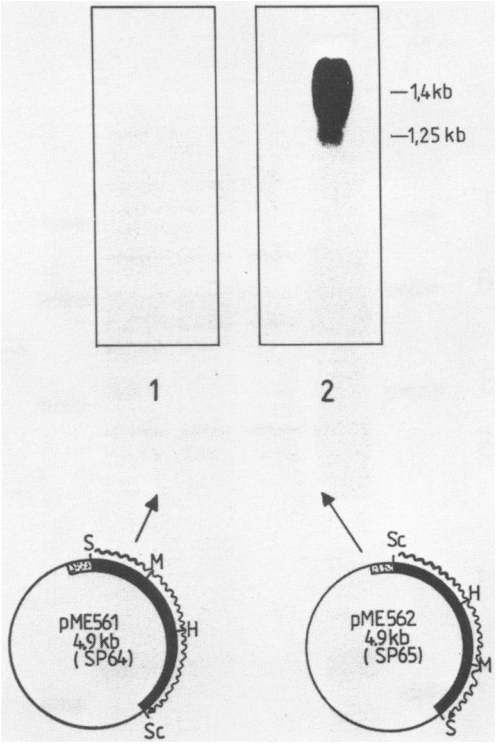

Figure 7. Direction of transcription of the two transcripts of the SalI-EcoRV fragment. Poly(A)$^+$-RNA, isolated from strain X2180-1A was probed with single strand specific, ^{32}P-labelled RNA, synthesized either from pME561 linearized with SacI (track 1) or from pME562, linearized with SalI (track 2). The 1.45 kb as well as the 1.25 kb transcript hybridized only with the in vitro RNA transcribed from plasmid pME562. ⁘ SP6 promoter; ⇝ direction of transcription; HindIII (H), MluI (M), SalI (S), SacI (Sc).

gene product, when put on a multicopy plasmid, (iii) integrates at the TRP4 locus in a gene disruption experiment and (iv) shows sequence homology to the homologous genes from E.coli and B.subtilis.

The amino acid sequence homology was relatively low (27 %) between the corresponding S.cerevisiae TRP4 and the E.coli trpD sequences as compared to other sequences of homologous TRP genes, namely 35 % for InGPsynthase (9), 40 % for tryptophan synthase (8,10). The relatively low overall homology of PRtransferase may be due to monofunctionality of this enzyme. There are two domains on the peptide however (see Fig. 5) with homologies higher than 30 %,

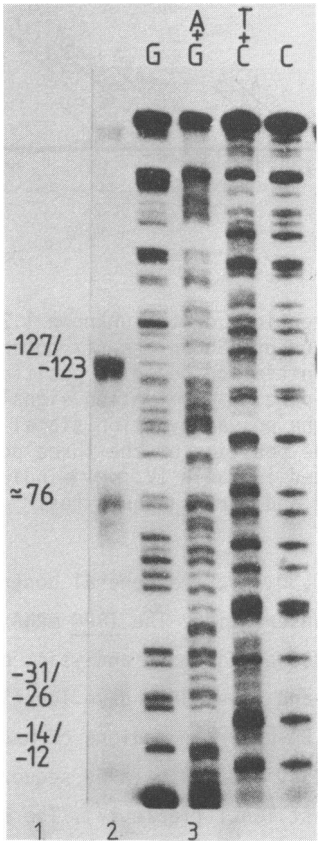

Figure 8. Transcript mapping of the SalI-EcoRV fragment. Start sites of TRP4 gene transcription, mapped on a 6 % polyacrylamide/7M urea sequencing gel. Track 1: position of the start points (ATG = +1). Track 2: S1 nuclease-resistant DNA fragments of the coding strand (60 units S1 endonuclease, hybridisation temperature 46°C). The poly(A)$^+$-RNA used in this protection experiment was isolated from the wild type strain X2180-1A. Track 3: Maxam Gilbert sequence ladder of the used 5'-a fragment (see Fig. 3b).

which even applies if the sequences from S.cerevisiae, E.coli and B.subtilis are compared simultaneously. One may speculate that these two domains are the acceptor sites for the substrate anthranilate and the co-substrate phosphoribosyl pyrophosphate.

The *TRP4* flanking sequences

The TRP4 gene is regulated in the frame of the general control together with TRP2, TRP3, TRP5 and genes of other amino acid biosyntheses (2). In all

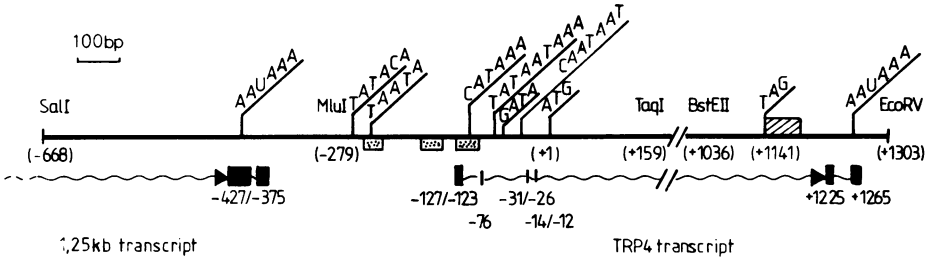

Figure 9. Positions and directions of the unknown 1.25 kb transcript and the TRP4-specific, 1.4 kb transcript. Shown are six potential TATA elements, listed in Table III, the ATG initiation codon at position +1, the stop codon TAG at position +1141, the tri-partite termination signal as proposed by Zaret and Sherman (36). ▨, and the polyadenylation signal AAUAAA, as proposed by Nevins (37). ⊡ indicates the position of the three potential "general control" consensus sequences as listed in Table IV. ∿▶ indicates direction of mRNA synthesis, ▮—▮ mRNA start and termination sites.

cases examined so far (6,10,30,31), the general control has been shown to operate at the level of transcription. The TRP4 mRNA is also increased under derepressed conditions (based on Northern analysis; data not shown). Donahue et al. (32) and Hinnebusch and Fink (30), have identified a consensus sequence 5'-A$^{A}_{T}$GTGACTC-3', in the 5'-upstream regions of four genes subject to general control: HIS1, HIS3, HIS4, and TRP5. Three sequences in the TRP4-promoter show conservation of at least TGACT (Table IV). The sequence at position -124/-132 is inverted but may, according to Hinnebusch (33), still act as an activation site. The best candidate for a functional consensus sequence is the element at position -168/-160, possibly in connection with the other two elements.

The 5'-ends for TRP4 mRNA were determined by S1 nuclease mapping. In con-

Table III. Postulated "TATA elements" in the 5'-flanking region of the *TRP4* gene.

Position	Sequence
-266/-261	TATACA
-241/-237	TAATA
-123/-118	CATAAA
-87/-79	TATAATAAA
-72/-69	GATA
-34/-28	CAATAAT

All AT-rich elements, with at least a conserved ATA, up to position -280 are listed.

Table IV. General control consensus sequences in the 5'-flanking region of the *TRP4* gene.

Position	Sequence	Orientation
-248/-240	5' TTATGACTA 3'	normal
-168/-160	5' GATTGACTC 3'	normal
-124/-132	3' AAATGACTT 5'	inverted
consensus sequence (31)	5' A(A/T)GTGACTC 5'	normal

All elements in the 5'-flanking region up to position -668 with at least a conserved TGACT (in both orientations) are listed. Homologies to the "ideal" consensus sequence are underlined.

trast to the TRP2 (8), TRP3 (8,9), and TRP5 genes (10), where multiple 5'-ends of similar intensities were detected, most of the TRP4 mRNA is initiated at one major point around position -127/-123 and at a minor point around -76.

In most yeast gene promoters, AT-rich sequences with similarity to the "Goldberg-Hogness box" were found within 35 to 180 bp upstream of the first mRNA start site (34). Furthermore it was shown for the HIS3 promoter region, that the spacing between the presumptive TATA box and the initiation sites can be varied over a certain range without altering the sites of initiation (35). This is an indication that the sequences at the initiation site show a certain specificity themselves. In Table III such AT-rich sequences in the TRP4 promoter with at least a conserved ATA are listed. Out of six elements four can not be functional since they are located within the transcribed region when starting from the mainly used -127/-123 region (Fig. 9). If the functional arrangement of the promoter elements is 5'end/consensus sequence /TATA box / site for transcription initiation/3'end, then the TATA elements at position -266/-261 and -241/-237 can not be functional either, as they are upstream of the regulatory consensus sequences. So it seems that none of the postulated TATA sequences of Table III is actually used in vivo.

The only good candidate for a TATA-like element, in relation to the sequences between -127/-123 (start site of transcription) and -168/-160 (consensus sequence), is the sequence 5'-TAAAAAATGATT-3' (-157/-146). Further experiments are required to evaluate the postulated role of this sequence.

S1 nuclease mapping provided evidence for heterogeneity at the 3'-end of TRP4 mRNA. Since the isolated mRNA was polyadenylated, 3'-heterogeneity can not be due to partial degradation in vivo or during isolation. The sequence

TAG...TATGT...TTT suggested by Zaret and Sherman (36) as a component of a termination/polyadenylation signal in yeast is found in TRP4 between nucleotides +1141 and +1183 and precedes the 3'-mRNA ends determined by S1 nuclease mapping around +1225 and +1265 (Fig. 9). A second potential TRP4 polyadenylation sequence, AAUAAA (37), is present at nucleotides +1257 to +1265 (Fig. 9). This sequence could potentially serve as a polyadenylation signal only for the longer class of transcripts encoding around +1265, but not for the shorter ones that were also isolated as poly(A)$^+$-transcripts. Additionally, a 3'-end-deleted, plasmid-borne TRP4 gene, lacking the sequences downstream of +1217 seems to be expressed still normally (data not shown). This finding argues against a functional role of the AATAAA sequence in the TRP4 terminator, since non-polyadenylated mRNAs are very unstable and would likely not allow for proper expression of PRtransferase.

ACKNOWLEDGMENT. This work was supported by the Swiss National Science Foundation, Grant No. 3.606-0.84.

*Present address: Department of Genetics, University of Washington, Seattle, WA 98195, USA

REFERENCES

1. Mortimer, R.K. and Hawthorne, D.C. (1966) Genetics 53, 165-173.
2. Miozzari, G., Niederberger, P. and Hütter, R. (1978) J.Bacteriol. 134, 48-59.
3. Niederberger, P., Miozzari, G. and Hütter, R. (1981) Mol. Cell. Biol. 1, 584-593.
4. Walz, A., Ratzkin, B. and Carbon, J. (1978) Proc. Natl. Acad. Sci. USA 75, 6172-6176.
5. Struhl, K., Stinchcomb, D.T., Scherrer, S. and Davis, R.W. (1979) Proc. Natl. Acad. Sci. USA 76, 1035-1039.
6. Aebi, M., Niederberger, P. and Hütter, R. (1982) Curr. Genet. 5, 39-46.
7. Tschumper, G. and Carbon, J. (1980) Gene 10, 157-166.
8. Zalkin, H., Paluh, J.L., van Cleemput, M., Moye, W.S. and Yanofsky, C. (1984) J. Biol. Chem. 259, 3985-3992.
9. Aebi, M., Furter, R., Prantl, P., Niederberger, P. and Hütter, R. (1984) Curr. Genet. 8, 165-172.
10. Zalkin, H. and Yanofsky, C. (1982) J. Biol. Chem. 257, 1491-1500.
11. Zalkin, H., (1973) Adv. Enzymol. 38, 1-39.
12. Vogel, H. and Bonner, J. (1956) J. Biol. Chem. 218, 97-106.
13. Roberts, T.M. and Lauer, G.D. (1979) Methods Enzymol. 68, 473-482.
14. Smith, G.E. and Summers, M.D. (1980) Anal. Biochem. 109, 123-129.
15. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
16. Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
17. Dretzen, G., Bellard, M., Sassoue-Corsi, P. and Chambon, P. (1981) Anal. Biochem. 112, 295-299.

18. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
19. Bantle, J.A., Maxwell, J.H. and Hahn, W.E. (1976) Anal. Biochem. 72, 413-427.
20. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
21. Weaver, R.F. and Weissmann, C. (1979) Nucl. Acid. Res. 7, 1175-1193.
22. Miozzari, G., Niederberger, P. and Hütter, R. (1978) Anal. Biochem. 90, 220-223.
23. Zakian, V.A., and Scott, J.F. (1982) Mol. Cell. Biol. 2, 221-232.
24. Rothstein, R.J. (1983) Methods Enzymol. 101, 202-211.
25. Botstein, D., Falco, S.C., Stewart, S.E., Scherer, S., Stinchcomb, M.T., Struhl, K. and Davis, R.W. (1979) Gene 8, 17-24.
26. Erhart, E. and Hollenberg, C.P. (1981) Curr. Genet. 3, 83-89.
27. Yanofsky, C. (1981) Nature 289, 751-758.
28. Band, L., Shimotu, H. and Henner, D.J. (1984) Gene 27, 55-65.
29. Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3026-3031.
30. Hinnebusch, A.G. and Fink, G. (1983) J. Biol. Chem. 258, 5238-5247.
31. Struhl, K. and Davis, R.W. (1981) J. Mol. Biol. 152, 535-552.
32. Donahue, T.F., Daves, R.S., Lucchini, G. and Fink, G.R. (1983) Cell 33, 89-98.
33. Hinnebusch, A.G. (1985) Mol. Cell. Biol. 5, 2349-2360.
34. Sentenac, A. and Hall, B.D. (1982) In The Molecular Biology of the Yeast Saccharomyces cerevisiae, Metabolism and Gene Expression, Strathern, J.N., Jones, E.W. and Broach, J.R. Eds., pp. 561-606, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
35. Chen, W. and Struhl, K. (1985) EMBO J. 4, 3273-3280.
36. Zaret, K.S. and Sherman, F. (1982) Cell 28, 563-573.
37. Nevins, J.R. (1983) Ann. Rev. Biochem. 52, 441-466.