J. Mol. Biol. (1981) **152**, 553–568

Promoter Mutants of the Yeast *his3* Gene

KEVIN STRUHL†

Medical Research Council Laboratory of Molecular Biology
Hills Road, Cambridge, CB2 2QH, England

AND

RONALD W. DAVIS

Department of Biochemistry
Stanford Medical School, Stanford, Calif. 94305, U.S.A.

(Received 27 February 1981)

The DNA sequences, transcription patterns and genetic behaviours of the wild-type *his3* gene from *Saccharomyces cerevisiae* (yeast) and of seven mutations that delete sequences flanking the 5′-end of the messenger RNA coding region are described. The deleted derivatives behave as promoter mutants because they are *cis*-dominant and fail to make detectable levels of *his3* mRNA. The results indicate that sequences further than 44 base-pairs upstream from the mRNA coding region are necessary for proper *his3* expression. The seven derivatives that inactivate *his3* promoter function in yeast cells also delete the promoter sequence necessary for expression in *Escherichia coli*.

1. Introduction

Concepts of gene regulation, protein–DNA interactions, and DNA sequence recognition have been developed in prokaryotic systems. A variety of genetic signals that influence the expression and the regulation of specific structural genes have been analysed in great detail by correlating physical DNA structures of wild-type and mutant genes with their biochemical and physiological functions (reviewed by Rosenberg & Court, 1979). In eukaryotic organisms, mechanisms of gene expression and of gene regulation, though under heavy experimental attack, remain to be elucidated.

Here, our inquiries are confined to the imidazoleglycerolphosphate dehydratase (*his3*) gene of the simple eukaryote *Saccharomyces cerevisiae* (bakers' yeast). A DNA fragment containing *his3* has been cloned (Struhl *et al.*, 1976; Struhl & Davis, 1977) and structurally, transcriptionally and genetically mapped (Struhl *et al.*, 1980; Struhl & Davis, 1980,1981). Numerous derivatives that successively delete into *his3* DNA have been isolated and physically characterized (Struhl & Davis, 1980).

In a variety of ways, it is possible to introduce cloned yeast DNA back into genetically defined strains (Hinnen *et al.*, 1978; Beggs, 1978; Struhl *et al.*, 1979*b*).

† Present address: Department of Biological Chemistry, Harvard Medical School, 25 Shattuck St., Boston, Mass. 02115, U.S.A.

0022-2836/81/310553-16 $02.00/0

Such hybrid DNA molecules are analogous to specialized transducing phages or conjugative plasmids. Thus, the sophistication of yeast molecular genetics equals that of *Salmonella typhimurium* and *Escherichia coli*. In particular, it is possible to determine accurate *in vivo* phenotypes of cloned derivatives of the wild-type gene.

In this paper, seven *cis*-dominant mutations that alter *his3* expression are described. The potential locations of *his3* promoter and regulatory elements with respect to the DNA sequence and to the messenger RNA transcript are considered.

2. Materials and Methods

(a) *Plasmid, phage, bacterial and yeast strains*

Plasmid DNAs YIp5, YRp7 (Struhl *et al.*, 1979*b*), pGT1-Sc2602, pGT2-Sc2605, pBR322-Sc2676, pBR322-Sc2710, pMB9-Sc2731 (Struhl & Davis, 1980), and pMB9-Sc2683 (Brennan & Struhl, 1980) have been described. All bacteriophage λgt-*his3* hybrid phages were characterized by Struhl & Davis (1980). The bacterial strain lacking imidazoleglycerol-phosphate dehydratase activity was *hisB*463 (Struhl *et al.*, 1976). The yeast strain used in all the experiments, SC3 (λ *trp1*-289 *ura3*-52 *his3*-Δ1 *gal2 gal10*), was constructed by and obtained from Stewart Scherer (Scherer & Davis, 1979). These organisms were propagated as described by Struhl & Davis (1980) and Struhl *et al.* (1979*b*).

(b) *DNA biochemistry*

The technical details for isolation of phage, plasmid and yeast DNAs, and for construction of hybrid DNA molecules were described by Struhl & Davis (1980). Hybrid DNAs were constructed by cleavage of the appropriate starting DNAs with a restriction endonuclease(s) followed by covalent joining with T4 DNA ligase. They were introduced and propagated in *E. coli* cells by transformation of the strain *hisB*463 to ampicillin or tetracycline resistance. Plasmid DNAs from transformants with the desired phenotype were prepared and analysed by cleavage with the appropriate restriction endonucleases.

DNA from each *his3* derivative was subcloned into an appropriate yeast vector in order to determine its phenotype *in vivo*. The vectors used were YRp7 and YIp5 (Struhl *et al.*, 1979*b*). YRp7 contains the yeast *trp1* gene as well as *ars1*, a chromosomal sequence that allows hybrid molecules to replicate autonomously in the absence of significant genomic integration (Struhl *et al.*, 1979*b*; Stinchcomb *et al.*, 1979). YIp5 contains the yeast *ura3* gene but is incapable of autonomous replication. YIp5-*his3* hybrid DNA molecules transform yeast strains harbouring the *ura3*-52 mutation to Ura^+ only by chromosomal integration at the *his3* locus (Scherer & Davis, 1979). Specific *his3* hybrid DNA molecules were constructed as described below.

YRp7 hybrids containing DNA fragments Sc2715, Sc2716, Sc2734, Sc2735, Sc2737 and Sc2738 were constructed by cleavage of YRp7 DNA and of λgt-Sc2639, λgt-Sc2671, λgt-Sc2666, λgt-Sc2667, λgt-Sc2669 and λgt-Sc2670 with *Bam*HI endonuclease followed by covalent joining with T4 DNA ligase. Transformants resistant to ampicillin were selected and screened for sensitivity to tetracycline and His^+. DNAs used in the yeast transformation experiments all have the *his3* fragment in the orientation shown in Fig. 2.

YRp7-Sc2713 and YRp7-Sc2714 were constructed by cleavage of YRp7 DNA and of either λgt-Sc2694 or λgt-Sc2695 DNAs with *Sal*I endonuclease followed by ligation. The transformants were analysed similarly to those YRp7 hybrids described above.

YRp7-Sc2710 was constructed by cleavage of YRp7 and pBR322-Sc2710 DNAs with *Eco*RI endonuclease followed by ligation. Tetracycline-resistant transformants were screened for sensitivity to ampicillin and for Trp^+ (by complementing the *E. coli* mutation *trpC*9830).

YRp7-Sc2731 was constructed by complete *Eco*RI cleavage of pMB9-Sc2731 DNA and partial *Eco*RI cleavage of YRp7 DNA followed by ligation. Ampicillin-resistant

transformants were screened by the colony filter hybridization method of Grunstein & Hogness (1975) for those capable of hybridizing ^{32}P-labelled pGT2-Sc2605 DNA.

YRp7-Sc2742 was constructed by partial *Eco*RI and total *Bam*HI cleavage of YRp7 DNA and total *Eco*RI and *Bam*HI cleavage of pGT2-Sc2605 DNA followed by ligation. Ampicillin-resistant transformants that were tetracycline-sensitive and His$^+$ were selected.

YIp5-Sc2683 was constructed by *Eco*RI cleavage of YIp5 and pMB9-Sc2683 DNAs with *Eco*RI endonuclease followed by ligation. Ampicillin-resistant transformants that were His$^+$ were selected.

YIp5-Sc2732 was constructed by *Eco*RI and *Bam*HI cleavage of YIp5 and pGT1-Sc2602 DNAs followed by ligation. Ampicillin-resistant, tetracycline-sensitive transformants were selected.

YIp5 hybrids containing DNA fragments Sc2739, Sc2744, Sc2745, Sc2746 and Sc2747 were constructed by *Sal*I endonuclease cleavage of YIp5-Sc2732 DNA and of λgt-Sc2694, λgt-Sc2639, λgt-Sc2766, λgt-Sc2666, and λgt-Sc2670 DNAs. Ampicillin-resistant transformants that were His$^+$ were selected. Those used have the subcloned *his3* DNA fragment inserted in the orientation shown in Fig. 2.

The structures of all the DNAs constructed in the above paragraphs were verified by restriction endonuclease cleavage (data not shown). The protocols for nick-translation and for hybridization of ^{32}P-labelled probes to nitrocellulose filters have been described previously (Struhl & Davis, 1980). DNA was purified from acrylamide gels following electrophoresis by placing the gel slice in a dialysis bag in 5 mM-Tris-borate (pH 8·3), 0·3 mM-EDTA. The bag was immersed in a tank containing the same solution and subjected to electrophoresis. The clamped ends of the dialysis bag were oriented perpendicular to the electric field. The DNA fragment was concentrated by precipitation with ethanol; greater than 90% of the DNA was recovered by this procedure.

(c) *RNA biochemistry*

Yeast strains were grown to a density of 2×10^7 cells/ml in glucose minimal medium at 30°C containing histidine and uracil, conditions that maintained selection for the transforming DNA. RNA was extracted, denatured in glyoxal, electrophoretically separated in 1·7% (w/v) agarose, transferred to a strip of diazotized paper, and challenged for hybridization with a ^{32}P-labelled probe of Sc2676 DNA. The technical details for these procedures can be found in the previous paper (Struhl & Davis, 1981).

(d) *Mapping with S_1 nuclease*

The *Bam*HI-*Hin*dIII fragment from Sc2671 containing the 5′-end of the *his3* gene and 240 base-pairs of λ DNA that had been purified by velocity sedimentation in a sucrose gradient by J. A. Jaehning was lightly labelled (to a spec. act. around 3×10^7 disints/min of ^{32}P/μg DNA) by nick-translation, cleaved with *Hin*fI endonuclease and electrophoretically separated in a 5% (w/v) acrylamide gel. Because Sc2671 deletes the rightmost *Hin*fI site indicated in Fig. 1, the 183 base-pair fragment encoding the 5′-end of the mRNA is easily separated from all other fragments. This base-pair DNA fragment was purified, denatured in alkali, and subjected to gel electrophoresis under conditions that permit strand separation (Hayward, 1972). The separated strands were electrophoretically eluted from the gel; the fast migrating strand is purine-rich, while the slow migrating strand is pyrimidine-rich. A portion of each labelled probe (roughly 10,000 cts/min or 0·3 ng) was incubated with 30 ng of *Hha*I and *Hae*III-cleaved plasmid DNA that had been denatured in alkali or 10 μg of poly(A)-containing RNA for 60 min in 40 mM-PIPES (pH 7·3), 0·8 mM-EDTA, 1 M-NaCl at 60°C. The DNA was cleaved to decrease its rate of self-annealing. Following hybridization, the reaction mixture was diluted into buffer containing S_1 nuclease and treated as described by Struhl & Davis (1980). The resulting products were electrophoretically separated under the same conditions used for DNA sequencing (Maxam & Gilbert, 1977). The sizes of these products were calibrated using restriction endonuclease-generated fragments of pBR322 DNA of known length (Sutcliffe, 1978).

(e) *DNA sequencing*

pBR322-Sc2676 DNA was cleaved with *Taq*I endonuclease, treated with calf intestinal alkaline phosphatase to remove the 5'-terminal phosphates and then incubated with T4 polynucleotide kinase and ATP labelled in the gamma position in order to complete the phosphate exchange. The 2 appropriate DNA fragments were purified after separation in a 5% acrylamide gel and then cleaved with *Hin*dIII or with *Hha*I endonuclease, and rerun in 5% acrylamide. The 2 DNA fragments each labelled only at the *Taq*I recognition site at nucleotide −63 were isolated and their DNA sequences determined by the chemical, base-specific, partial cleavage method of Maxam & Gilbert (1977). To confirm most of this sequence, the 180 base-pair *Hin*fI fragment that includes the *Taq*I site was labelled at both ends and then cleaved with *Taq*I. The sequences of the appropriate DNA fragments were determined.

3. Results

(a) his3 *mutants that delete sequences adjacent to but not within the structural gene*

The yeast imidazoleglycerolphosphate dehydratase gene (*his3*), when introduced into *E. coli hisB* auxotrophs lacking the bacterial imidazoleglycerolphosphate dehydratase, allows the resulting cells to grow in the absence of histidine (Struhl *et al.*, 1976). The wild-type yeast *his3* gene is functionally expressed in *E. coli* to produce an imidazoleglycerolphosphate dehydratase activity strongly resembling that found in wild-type yeast cells (Struhl & Davis, 1977). Cloned mutant *his3* genes are non-functional (Struhl & Davis, 1977) and a *his3* amber lesion can be suppressed by the appropriate *E. coli* transfer RNA species (Struhl *et al.*, 1979*a*). Therefore, yeast *his3* expression in *E. coli* assays cloned DNA fragments for the presence of the intact *his3* structural gene.

The *his3* structural gene has been localized by correlating the expression and the physical structures of numerous derivatives of the original cloned DNA fragment (Struhl & Davis, 1980). Most of these derivatives were isolated as deletion mutants

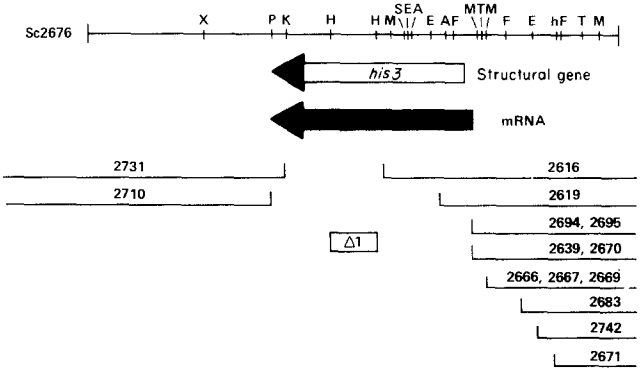

FIG. 1. Physical and genetic map of the *his3* gene. The structure of the *his3*-containing *Bam*HI DNA fragment (Sc2676) is shown by the top horizontal line. Restriction endonuclease cleavage sites (vertical lines) to the rightmost *Hin*dIII (H) site are indicated for *Mbo*II (M), *Taq*I (T), *Hin*fI (F), *Hha*I (h), *Hae*III (E), *Ava*II (A) and *Msp*I (S). *Kpn*I (K), *Pst*I (P), and *Xho*I (X) sites are also shown. The location of the structural gene (presumed translation sequences) and transcribed region (determined in this work) are shown. Mapping positions of deletion mutations were determined by Struhl (1981*a*) with the exception of *Δ*1 (Scherer & Davis, 1979), 2683 (Brennan & Struhl, 1980), and Sc2742 (this work).

of a λ*his3* hybrid phage. Functional expression of individual λ*his3* deletion mutants was determined under conditions in which transcription is initiated from the λ promoter P_L; such an assay requires only the presence of the intact structural gene (Struhl *et al.*, 1980). By this analysis, the *his3* structural gene maps within a 700 base-pair region (see Fig. 1). The 5′-end lies between the deletion break points of λgt-Sc2694 and λgt-Sc2619 and the 3′-end lies between the *Pst*I and the *Kpn*I sites (fragments Sc2710 and Sc2731). In this paper, seven *his3* deletion mutations generated in *E. coli* by illegitimate recombination between *his3* and λ sequences are described in detail (Table 1). Their mapping positions were determined by a heteroduplex-S_1 nuclease technique or by direct DNA sequencing (Struhl, 1981*a*). Most importantly, all seven deletion mutants express *his3* in *E. coli*; thus, they contain the intact structural gene.

(b) *Phenotypes of the mutants*

The cloned *his3* derivatives were isolated and mapped by methods that did not involve their propagation in yeast cells. To establish their *in vivo* phenotype, DNA from each derivative was subcloned into two different kinds of vectors (Fig. 2). The YRp7 vector contains the yeast *trp1* gene as well as *ars1*, a chromosomal sequence that allows hybrid molecules to replicate autonomously in the absence of significant genomic integration (Struhl *et al.*, 1979*b*; Stinchcomb *et al.*, 1979). YIp5 is a vector containing the yeast *ura3* gene; when used in conjunction with strains

TABLE 1

Properties of his3*-containing DNA fragments*

Cloned *his3* DNA fragments			Deletion endpoints		*his3*
Original (λ)	YRp7	YIp5-Sc2732	Yeast	λ	phenotypes
Sc2616	Sc2711	ND	+325±25	*att*	−
Sc2619	Sc2712	ND	+55±15	*att*	−
Sc2694	Sc2713	Sc2739	−45± 2	0·633	−
Sc2695	Sc2714	ND	−45± 2	0·633	−
Sc2639	Sc2715	Sc2744	−49	*att*	−
Sc2670	Sc2738	Sc2747	−49	*att*	−
Sc2667	Sc2735	ND	−76	*att*	−
Sc2666	Sc2734	Sc2745	−80	*att*	−
Sc2669	Sc2737	Sc2746	−84	*att*	−
Sc2683		Sc2683	−170±40	None	+
	Sc2742		−200±20	None	+
Sc2671	Sc2716	ND	−300± 5	*att*	+

The isolation numbers of the original deletion mutants of λgt-Sc2601 are indicated (see Struhl & Davis, 1980). The *his3*-containing DNA fragments were cloned either into YRp7 or YIp5-Sc2732. Fragments cloned into YRp7 were produced by *Bam*HI endonuclease with the exception of Sc2713 and Sc2714, which were generated by *Sal*I endonuclease. Fragments cloned into YIp5-Sc2732 were generated by *Sal*I. ND, indicates that the particular hybrid DNA molecule was not constructed. The structure of a prototype of these molecules is shown in Fig. 1. The deletion endpoints within the *his3* sequence and within the λ sequence are indicated. The co-ordinate scale is defined in Fig. 5; the mapping positions with respect to this scale have been determined here or elsewhere. The λ endpoints are defined on the normal λ map (*att* indicates the core of the λ attachment site). The *his3* phenotypes are indicated in the last column. In all cases, the YRp7 and YIp5 hybrids of a given allele showed identical phenotypes.

harbouring the *ura3*-52 deletion mutation, transformation occurs by integration at the *his3* locus (Scherer & Davis, 1979). Hybrid DNA molecules were introduced into yeast strain SC3 (relevant genotype *ura3*-52 *trp1*-289 *his3*-Δ1) by selecting for Ura$^+$ or Trp$^+$ transformants. These were tested for the *his3* character by growth in the absence of histidine. The results are shown in Table 1.

his3 derivatives that retain more than 170 base-pairs before the 5′-end of the structural gene (such as Sc2683, Sc2742 and Sc2671, see Fig. 1) grow at the wild-type rate in the absence of histidine. Furthermore, these transformants grow in the presence of aminotriazole (a competitive inhibitor of imidazoleglycerolphosphate dehydratase), thus indicating that these derivatives express *his3* at the wild-type level. Derivatives lacking the 5′-end of the structural gene (e.g. Sc2619 and Sc2616)

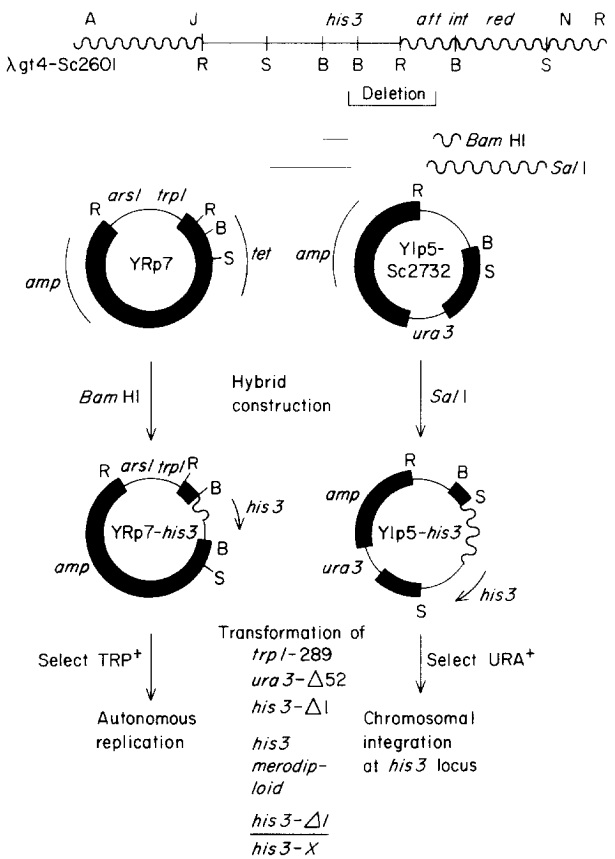

FIG. 2. Phenotype *in vivo* in yeast cells of *his3* derivatives. The structure of a typical *int*-generated deletion mutant of a λ*his3* hybrid is shown. λ sequences are shown as a wavy line and *his3* sequences are shown as a solid line. The *his3*-containing *Bam*HI or *Sal*I DNA fragments were cloned, respectively, into YRp7 or into YIp5-Sc2732 (containing the rightmost *Bam*HI-*Eco*RI DNA fragment of Sc2601 DNA cloned in YIp5). Hybrid molecules were introduced into SC3 by selecting for vector-encoded genes (*trp1* or *ura3*). YRp7 hybrids replicate autonomously in yeast cells while YIp5 hybrids are propagated only after chromosomal integration at the *his3* locus. The resulting transformants are merodiploid for the *his3* locus. One copy is the original chromosomal lesion (Δ1), which is deleted for 150 base-pairs internally within the gene; the other copy comes from the transforming DNA.

are non-functional. For derivatives with endpoints near the 3′-end, Sc2710 is phenotypically His^+ in both *E. coli* and yeast, while Sc2731 is His^- in both organisms. These results indicate that the structural gene defined in *E. coli* is similar, if not identical, to the structural gene defined in yeast.

However, seven of the deletion mutants listed in Table 1 all fail to grow in the absence of histidine, even though they contain the intact structural gene. Therefore, these deletion alleles inactivate a non-structural region necessary for *his3* expression in yeast cells.

It is important to determine whether these seven mutations are genetically dominant or recessive and whether they act in *cis* or in *trans*. The transformation of haploid yeast cells by exogenous DNA capable of autonomous replication, analogous to F′-containing *E. coli* strains, ensures that the resulting transformants are merodiploid for yeast sequences present on the transforming DNA (Fig. 2). In the case at hand, the YRp7 hybrid DNAs of the deletion mutants were introduced into a strain carrying the *his3-Δ1* allele. This allele was constructed such that a 150 base-pair *Hin*dIII fragment located within the *his3* structural gene is removed and as expected, it does not revert (Scherer & Davis, 1979). *his3-Δ1* is transcribed to produce normal amounts of a shortened mRNA species (0·55 kb† in length) easily distinguishable from the 0·7 kb wild-type species (Fig. 3, lanes (1) and (2)). Transformation of SC3 to Trp^+ by a YRp7 hybrid containing an intact functional *his3* gene (such as YRp7-Sc2716) generates a His^+ strain in which both mRNA species are seen (Fig. 3, lane (5)). That the normal (0·7 kb) species is present at higher levels than the deleted species might be due to multiple copies of the YRp7 hybrid molecules in these cells. However, the seven derivatives listed in Table 1 as promoter mutants (two of which are shown in lanes (3) and (4) of Fig. 3) accumulate normal amounts of the shortened mRNA, but less than 10% (if any) of the normal sized mRNA. Because the wild-type level of *his3* mRNA is approximately one molecule per cell (Struhl & Davis, 1981), this indicates that most cells do not contain any of the mRNA species encoded by the transforming DNA.

Therefore, the *his3* deletion mutations tested confer their properties (inability to make RNA) in a *cis*-dominant manner. *cis*-dominant mutants that produce levels of mRNA below the basal level are analogous to promoter mutants described in many prokaryotic systems.

(c) *Mapping the 5′-end of the* his3 *mRNA*

Promoter mutants have the property that they alter transcription without altering the mRNA coding region. Thus, it is important to determine whether the *his3* derivatives described above delete any transcribed sequences. In the preceding paper, the direction of transcription and the approximate location of the 5′ and 3′-ends of the stable RNA species were determined (Struhl & Davis, 1981). The 5′-end of the RNA maps in the 183 base-pair *Hin*fI DNA fragment that includes the start of the structural gene as defined by expression *in vivo* of *his3* derivatives in yeast

† Abbreviation used: kb, 10^3 bases or base-pairs as appropriate.

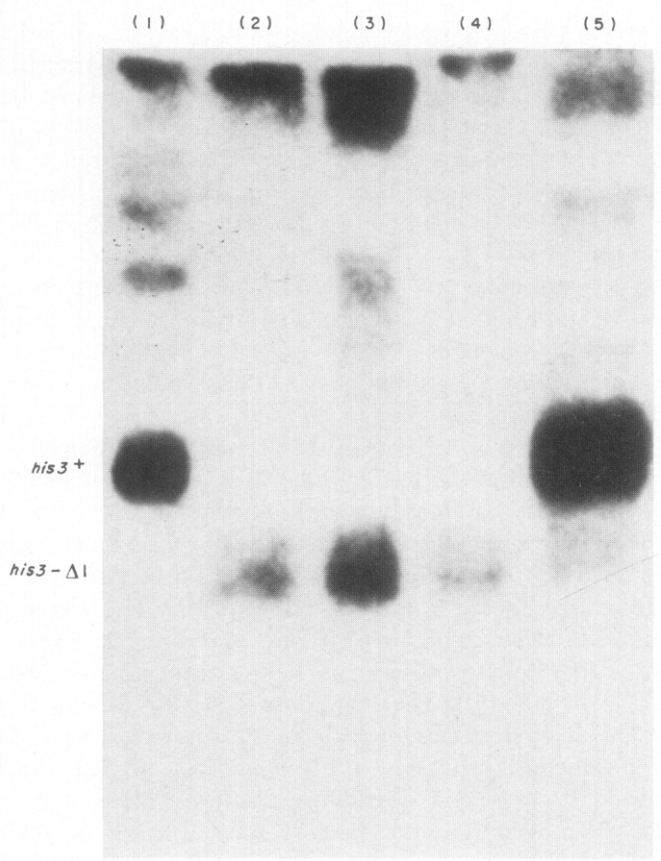

FIG. 3. *cis-trans* test for *his3* deletion mutations. RNA from (1) S288C; (2) SC3; (3) SC3 (YRp7-*Sal*I fragment from Sc2694); (4) SC3 (YRp7-*Bam*HI fragment from Sc2667); and (5) SC3 (YRp7-*Bam*HI fragment from Sc2671) was isolated, denatured in glyoxal, electrophoretically separated in a 1·7% (w/v) agarose gel, transferred to a strip of diazotized paper, and challenged for hybridization with a ^{32}P-labelled probe of Sc2676 DNA. The mobilities of mRNA species produced by *his3*$^+$ (0·7 kb) and *his3-Δ1* (0·55 kb) are indicated. The band at the top of the gel probably results from hybridization of ^{32}P-labelled DNA to contaminating cellular DNA.

and in *E. coli* (see Fig. 1). In this section, the 5′-endpoint of the wild-type *his3* mRNA is mapped accurately by measuring the length of the appropriate ^{32}P-labelled, single-stranded *his3* DNA fragment protected from S_1 nuclease digestion by hybridization with *his3* mRNA (see Materials and Methods).

As presented in Figure 4, total poly(A)-containing RNA from yeast hybridizes only to the slow (pyrimidine-rich) strand. The predominant protected fragment is 92 ± 1 nucleotides in length; therefore, the *his3* message starts this far from the *Hin*fI cleavage site in the structural gene. Two minor bands resulting from shortened RNA–DNA hybrid structures are also present.

The position of the 5′-end of the *his3* mRNA was also determined relative to the endpoints of deletion mutants described in the previous section. The results (Fig. 4) indicate that DNAs from deletions with known map positions protect the expected

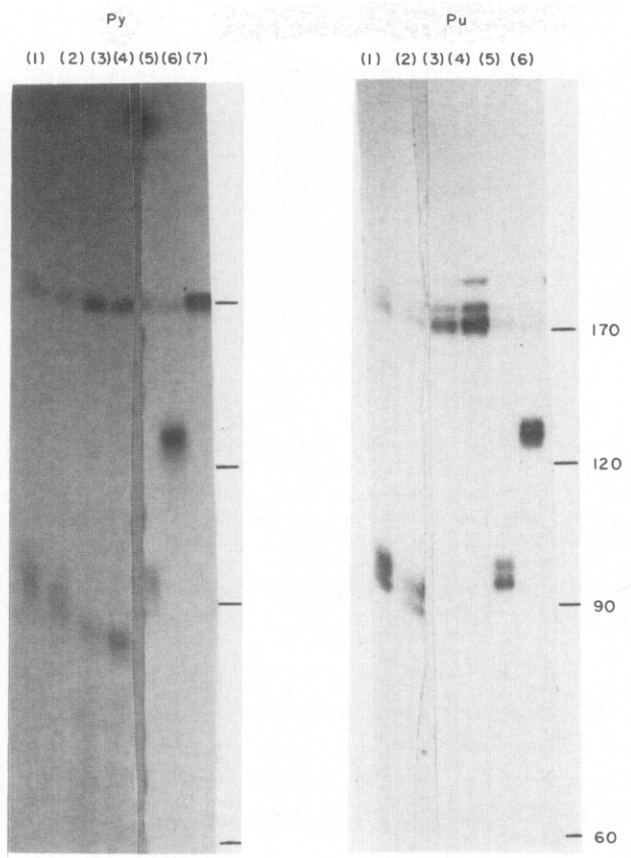

FIG. 4. High resolution mapping of the 5′-end of the *his3* mRNA. The nucleic acids tested for hybridization are as follows. (1) YRp7-Sc2713 (from Sc2694); (2) YRp7-Sc2714 (from Sc2695); (3) and (4) poly(A)-containing RNA from yeast cells; (5) YRp7-Sc2715 (from Sc2669); (6) YRp7-Sc2737 (from Sc2669); and (7) YRp7-Sc2716 (from Sc2671). The probes used were separated single strands of the 180 base-pair *Hin*fI fragment (Pu indicates purine-rich strand and Py indicates pyrimidine-rich strand). Size standards are indicated in base-pairs.

length of each strand of the labelled fragment (the endpoints are probably accurate within a range of 2 to 3 nucleotides). Deletions Sc2694 and Sc2695 have not been mapped previously because the deletion endpoints in λ DNA are not at the λ attachment site (both map between *att* and *red* roughly at co-ordinate 0·633 on the λ map; Struhl & Davis, 1980). Because their endpoints appear identical, these two deletions probably represent two independent isolates of the same recombination event. Because the mRNA endpoint and known deletion endpoints are determined by hybridization protection of the identical ^{32}P-labelled DNA probe, their relative mapping positions are directly established.

All seven deletion mutant DNAs protect a larger segment of labelled probe than the *his3* mRNA. This indicates that they contain the entire mRNA coding region for the *his3* gene.

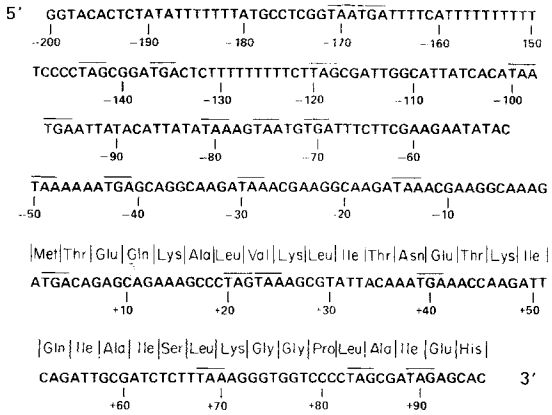

FIG. 5. DNA sequence of the *his3* gene region. The nucleotide sequence of the "sense" strand of a 300 base-pair region at the 5'-end of the *his3* gene is shown. The sequence is oriented such that the 5'-3' direction proceeds from left to right (this is the opposite direction from that defined in Fig. 1). The nucleotides are numbered such that +1 is the presumptive site of translational initiation and that −1 is the first nucleotide preceding it. The presumptive amino acid sequence for the initial 32 codons of the *his3* translation product is indicated above the relevant codons. Termination codons are indicated by horizontal bars above the nucleotides. Hyphens have been omitted for clarity.

(d) *DNA sequence of the* his3 *promoter region*

The nucleotide sequence of a 300 base-pair region at the 5'-end of the *his3* gene was determined by the chemical, base-specific, partial cleavage method of Maxam & Gilbert (1977) (see Fig. 5). The endpoints of the deletion mutants and the position of the 5'-end of the *his3* mRNA are included in the schematic view of the *his3* regulatory region (Fig. 6).

4. Discussion

(a) *Structural gene sequences*

These are defined herein as those translated to produce the *his3* gene product (imidazoleglycerolphosphate dehydratase). Because eukaryotic organisms initiate translation at AUG codons, the start of the *his3* structural gene can be positioned uniquely to the codon defined by nucleotides 1 to 3 in Figure 5. Translation of the genetic code beginning with this AUG proceeds through a reading frame that is "open" (without termination codons) for at least 150 nucleotides. Two UAA termination codons (at positions −13 to −15 and −28 to −30) are found in this reading frame just prior to the AUG at nucleotides 1 to 3. The only other AUG codon in the sequenced region (at position −44 to −42) is not transcribed and it specifies a reading frame filled with termination codons (the first one being the UGA at positions 2 to 4). Both reading frames other than the one defined by the AUG at nucleotides 1 to 3 contain termination codons. This is particularly significant in the light of previous results (Struhl & Davis, 1981), suggesting that the *his3* mRNA is not spliced in a manner described for many eukaryotic messages (Berget *et al.*, 1977; Chow *et al.*, 1977; Jeffreys & Flavell, 1977). The possibility that

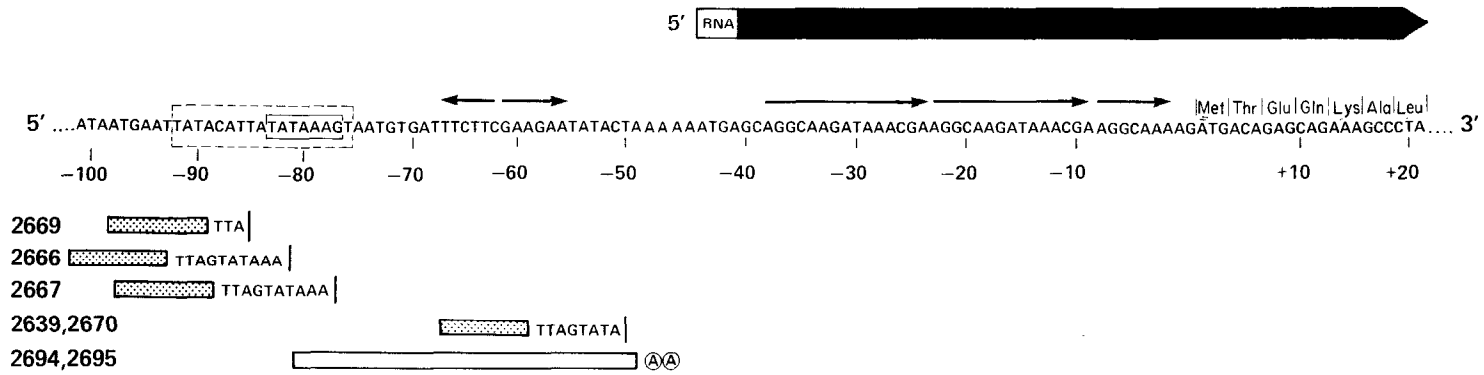

FIG. 6. The *his3* regulatory region. The nucleotide sequence of the "sense" strand of a 100 base-pair region immediately preceding the start of the *his3* structural gene is shown. The 5′-end and the direction of the *his3* transcript are indicated by the thick arrow above the sequence. The uncertainty in the exact position of the 5′-end is indicated by the letters RNA in the open box. The thin arrows above the sequence indicate the 36 base-pair direct repeat and the 12 base-pair inverted repeat (see text). The 5′-proximal amino acid residues of the *his3* translation product are shown above the relevant codons. The sequences and mapping positions of the *his3* promoter deletion alleles (determined by Struhl, 1981*a*) are indicated below the wild-type sequence. The endpoints within the yeast DNA sequence are indicated by vertical lines; the deletion mutations contain all the yeast nucleotides to the right of this line. The nucleotides to the left of this line come from the core of the λ attachment site; different deletion mutations have different endpoints within the core (Struhl, 1981*a*). The shaded box indicates nucleotides corresponding to the P′ region of the bacteriophage λ attachment site (represented in the opposite orientation as defined normally for λ). The open box shown for mutations Sc2694 and Sc2695 indicates λ DNA that has not been sequenced. The circled nucleotides shown for these 2 mutations indicate that they may or may not be present in the alleles. The Pribnow box proposed to act as a promoter for expression in *E. coli* is indicated by the solid box around the 7 base-pair region (−77 to −83). The conserved sequence that is homologous to 5′-ends of other eukaryotic genes (nucleotides −76 to −92) is indicated by the broken box. Hyphens have been omitted for clarity.

GUG and UUG occasionally initiate prokaryotic translation (Miller, 1974) is ruled out by their absence in the correct reading frame within this region. The location of the 5′-end of the structural gene determined by DNA sequence analysis agrees with previous genetic studies depending upon functional *his3* expression in *E. coli*, which indicated that it mapped between the deletion endpoints defined by λgt-Sc2619 and λgt-Sc2694 (nucleotides −44 to +70; Struhl & Davis, 1980).

(b) *Transcribed sequences*

The predominant *his3* mRNA contains a 41 ± 1 nucleotide leader before the 5′-end of the structural gene (Fig. 4). Two minor RNA species are also apparent. These might represent either (1) products of RNA processing such as by the removal of intervening sequences, (2) specific degradation of *his3* mRNA, or (3) mRNA species with different points of transcriptional initiation. The possibility that *his3* contains an intervening sequence(s) that must be spliced out of the initial transcript is unlikely from results of the previous paper (Struhl & Davis, 1981), because the molar amounts of the shorter RNA species are far below the molar amount of the predominant species, and because no protected fragment larger than the predominant species is observed. Although there is no evidence concerning the possibility of specific mRNA degradation, we feel that the minor bands represent *his3* mRNAs with different 5′-ends. The cytochrome *c* gene (*cyc1*) of yeast appears to code for a large number of transcripts each with a distinct 5′-terminal sequence (B. D. Hall, personal communication). Therefore, it is most likely that the DNA encoding the 5′-end of the RNA consists of a single continuous block of nucleotides having no intervening sequences.

In eukaryotic organisms, the absence of polycistronic operons and the unlikelihood of translational re-initiation has led to the suggestion that eukaryotic ribosomes recognize the 5′-proximal AUG in the mRNA (Sherman & Stewart, 1975; Kozak & Shatkin, 1978); the *his3* gene fits this pattern.

In addition, two striking structural features are noted. First, most of the leader (36 out of 41 nucleotides) consists of 2·4 tandem, directly repeated copies of a 15 base-pair sequence. Second, the leader is extremely rich in purine residues. This strong strand bias with respect to purine or pyrimidine bases continues for 30 more nucleotides upstream from the mRNA coding sequences. Such asymmetry has been found immediately adjacent to the 5′-untranslated sequences of the yeast *cyc1* and *ura3* genes. The significances of these situations remain to be seen.

(c) *Non-transcribed, untranslated sequences necessary for* his3 *expression*

Non-transcribed sequences upstream from the 5′-end of the gene are essential to express *his3* because deletion mutations that retain the intact mRNA coding region are not functional. Like prokaryotic genes, eukaryotic genes transcribed by RNA polymerase II require flanking sequences for proper expression (Struhl, 1979; Grosschedl & Birnstiel, 1980). Such is not the case for the frog *5S* genes, which are transcribed by RNA polymerase III. Deletion of the entire 5′ or 3′ sequences flanking the transcribed region does not prevent expression (Bogenhagen *et al.*, 1980; Sakonju *et al.*, 1980).

Two subsets of the sequenced region have properties suggestive of regulatory functions. First, a six base-pair sequence (T-T-C-T-T-C) is tandemly repeated in the inverted orientation. This 12 base-pair palindrome (nucleotides −56 to −67) is located 15 to 26 base-pairs before the mRNA start, thereby suggesting analogies to negative control such as that in the *E. coli* lactose operon. Second, the region located between 35 to 52 base-pairs before the mRNA start (nucleotides −76 to −93) shows significant homology with regions of DNA adjacent to other eukaryotic genes (first described by Goldberg, 1979). The homologous sequences before these eukaryotic structural genes are all extremely A + T-rich; the canonical sequence now depends upon definitions of individual investigators. The involvement of these (or any other sequence in regulation and/or expression of *his3*) must be demonstrated by the phenotype of mutations within it.

(d) *Comments regarding* his3 *expression and regulation*

A promoter is defined classically by mutations that are (1) closely linked but outside the structural gene, (2) *cis*-dominant, and (3) altered in the expression level of all genes in the transcription unit (Scaife & Beckwith, 1966). Seven deletion mutations described here behave as "promoter down" mutations by these criteria. Defined in a more physiological sense, promoter mutations alter gene expression at the level of RNA transcription without inactivating any transcribed sequences. The promoter deletions contain all mRNA coding sequences and produce significantly less (if any) *his3* mRNA than wild-type controls.

Analysis of deletion mutations is subject to several important caveats. First, deletions cannot be viewed solely as the absence of a particular sequence. All deletion alleles fuse new genetic material to the region of interest at the deletion break point. In five of the mutants described here, DNA from the core of the λ attachment site is directly connected to *his3* sequences; in these cases it is unlikely that the flanking sequences have differential effects. Second, the nucleotide sequences at the novel joints of the deletion mutations are sometimes unusual; the fusion sequence can appear as a point mutation(s) of the original wild-type sequence (Struhl, 1981*a*). Third, the hybrid DNA molecules used in these experiments contain prokaryotic DNA interspersed with eukaryotic DNA and the yeast DNA segments originated on different chromosomes in the normal yeast genome. For these reasons, the phenotype of any particular *his3* deletion allele could be affected by the sequences fused to it, by the sequence at the novel joint, or by the unusual structure of the transforming DNA. Derivatives in which sequences adjacent to the deletion endpoint promote or inhibit gene expression have been described (Struhl, 1979,1981*b*). However, our analysis of the phenotypes of cloned *his3* derivatives assumes that any of these effects will represent the exception rather than the rule.

Bearing the above caveats in mind, it seems highly probable that yeast RNA polymerase II recognizes a subset of the DNA sequences missing in the deletion mutants. The phenotype of Sc2669 indicates that sequences more distal to the structural gene than nucleotide −84 are important for gene expression. The phenotypes of derivatives Sc2683 and Sc2741 suggest that all the elements necessary for a functional promoter are located less than 170 base-pairs from the

structural gene. However, these derivatives were isolated in quite different ways from the seven promoter deletions. In particular, Sc2683 was isolated by virtue of increased *his3* expression in *E. coli* (Brennan & Struhl, 1980). Its non-*his3* endpoint maps at the 5′-end of the gene encoded by the 1·6 kb mRNA described in the previous paper (Struhl & Davis, 1981). The derivative that is the best control for the promoter deletions is λgt-Sc2671, which has a deletion break point approximately 300 base-pairs from the mRNA coding region. Determination of the minimum contiguous sequence necessary for wild-type *his3* promoter function will depend upon more deletion mutants.

In defining the key features of a eukaryotic promoter, much of the emphasis has been placed upon the conserved sequence described by Goldberg (1979) located between 20 and 40 base-pairs from the mRNA start site. This is because the conserved region (canonical sequence T-A-T-A-A-A-T-A) strongly resembles the Pribnow box (T-A-T-A-A-T), an important element for RNA polymerase binding and promoter function in prokaryotes (Pribnow, 1975; Rosenberg & Court, 1979). Experiments designed to determine the importance of this conserved sequence have given contradictory results. A 20 base-pair adenovirus DNA region containing it seems to be necessary and sufficient for correct transcription initiation *in vitro* (Corden *at al.*, 1980). Complete deletion of the analogous sequences in simian virus 40 DNA was alleged to have no effect *in vivo* (Benoist & Chambon, 1980), but the analysis was not quantitative and the effects of DNA replication and gene regulation were not considered experimentally. Deletion of a 54 base-pair DNA fragment of a sea urchin histone gene that included the conserved sequence removed the specificity of initiation but not the transcription in frog oocytes (Grosschedl & Birnstiel, 1980). However, it is not clear if this result is due (1) to the absence of the conserved sequence (as opposed to the other 45 base-pairs), (2) to the novel joint sequences generated by the deletion, or (3) to the heterologous nature of the assay.

The experiments described here clearly indicate that the conserved sequence is not sufficient for *his3* expression. Sc2669 retains what is likely to be the wild-type conserved sequence (T-A-T-A-A-A-G-T) and Sc2667, by virtue of the recombination event generates a "perfect fit" (Struhl, 1981*a*), yet both derivatives fail to make mRNA. However, the results do not bear on whether the conserved region is necessary for expression.

The *his3* gene is regulated at the transcriptional level in response to amino acid starvation (Struhl & Davis, 1981). From analysis of mutations in unlinked genes that affect *his3* expression as well as many other amino acid biosynthetic genes, it was suggested that control occurred in both positive and negative fashion (Wolfner *et al.*, 1975). However, because several genes seem to be involved, it seems unlikely that all their products interact directly with the *his3* gene. Though the location and nature of the 12 base-pair palindrome suggest analogies to negative control, such a palindrome is not present in the *his4* gene (which is regulated similarly to *his3*; Farabaugh & Fink, personal communication). It is unlikely that *his3* is regulated in a manner similar to the *S. typhimurium* histidine operon (DiNocera *et al.*, 1978; Barnes, 1978). Attenuation of an initial transcript is unlikely because no peptide leader is expected from the nucleotide sequence and because there are no histidine

codons prior to the structural gene. The question of positive or negative control of *his3* remains to be elucidated.

(e) *Comparison of* his3 *expression in yeast and in* E. coli

A promoter region for *his3* gene expression in yeast cells has been initially defined by mutations that inactivate it. An analogous region has been defined for *his3* expression in *E. coli* (Struhl *et al.*, 1980; Struhl & Davis, 1980; Struhl, 1981*a*). The promoter mutations defined in *E. coli* behave as promoter mutations in yeast. As yet, the exact position of the *E. coli* promoter is not known. However, there are three regions (at positions −83 to −77, −88 to −82, and −102 to −96) that have most if not all the highly conserved residues of a canonical Pribnow box (reviewed by Rosenberg & Court, 1979). From this as well as the *E. coli* RNA polymerase binding and transcription studies *in vitro* of J. A. Jaehning (unpublished results), it seems likely that one of these Pribnow boxes is the promoter recognized by *E. coli*. Perhaps not coincidentally, these Pribnow boxes and the conserved sequence in front of eukaryotic genes share common nucleotides. Because (1) *E. coli* transcripts are initiated less than ten base-pairs from the Pribnow box and (2) the yeast transcript maps more than 35 base-pairs from the analogous sequence, it is likely that the mechanism of prokaryotic and eukaryotic RNA polymerases differs in this basic step. Nevertheless, recognition of promoter sequences of the yeast imidazoleglycerolphosphate dehydratase structural gene in prokaryotic and eukaryotic organisms may literally share similar features.

We thank Phil Anderson and Andrew Travers for comments on the manuscript, and Judith Jaehning for permitting us to quote her unpublished results.

REFERENCES

Barnes, W. M. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 4276–4280.
Beggs, J. (1978). *Nature (London)*, **275**, 104–109.
Benoist, C. & Chambon, P. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 3865–3869.
Berget, S. M., Moore, C. & Sharp, P. A. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 3171–3175.
Bogenhagen, D. F., Sakonju, S. & Brown, D. D. (1980). *Cell*, **19**, 27–40.
Brennan, M. B. & Struhl, K. (1980). *J. Mol. Biol.* **136**, 333–338.
Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. (1977). *Cell*, **12**, 1–8.
Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980). *Science*, **209**, 1406–1414.
DiNocera, P. P., Blasi, F., DiLauro, R., Frunzio, R. & Bruni, C. B. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 4276–4280.
Goldberg, M. L. (1979). Ph.D. thesis, Stanford University.
Grosschedl, R. & Birnstiel, M. L. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 1432–1436.
Grunstein, M. & Hogness, D. S. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 3961–3965.
Hayward, G. S. (1972). *Virology*, **49**, 342–344.
Hinnen, A., Hicks, J. B. & Fink, G. R. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 1929–1933.
Jeffreys, A. J. & Flavell, R. A. (1977). *Cell*, **12**, 1097–1108.
Kozak, M. & Shatkin, A. J. (1978). *Cell*, **13**, 201–212.
Maxam, A. M. & Gilbert, W. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 560–564.
Miller, J. H. (1974). *Cell*, **1**, 73–76.
Pribnow, D. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 784–788.
Rosenberg, M. & Court, D. (1979). *Annu. Rev. Genet.* **13**, 319–353.

Sakonju, S., Bogenhagen, D. F. & Brown, D. D. (1980). *Cell*, **19**, 13–27.
Scaife, J. & Beckwith, J. R. (1966). *Cold Spring Harbor Symp. Quant. Biol.* **31**, 403–408.
Scherer, S. & Davis, R. W. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 4951–4955.
Sherman, F. & Stewart, J. W. (1975). In *Proc. 10th FEBS Meeting* (Bernardi, G. & Gros, F., eds), vol. 38, pp. 175–191, American Elsevier, New York.
Stinchcomb, D. T., Struhl, K. & Davis, R. W. (1979). *Nature (London)*, **282**, 39–43.
Struhl, K. (1979). Ph.D. thesis, Stanford University.
Struhl, K. (1981*a*). *J. Mol. Biol.* **152**, 517–533.
Struhl, K. (1981*b*). *J. Mol. Biol.* **152**, 569–575.
Struhl, K. & Davis, R. W. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5255–5259.
Struhl, K. & Davis, R. W. (1980). *J. Mol. Biol.* **136**, 309–332.
Struhl, K. & Davis, R. W. (1981). *J. Mol. Biol.* **152**, 535–552.
Struhl, K., Cameron, J. R. & Davis, R. W. (1976). *Proc. Nat. Acad. Sci., U.S.A.* **73**, 1471–1475.
Struhl, K., Davis, R. W. & Fink, G. R. (1979*a*). *Nature (London)*, **279**, 78–79.
Struhl, K., Stinchcomb, D. T., Scherer, S. & Davis, R. W. (1979*b*). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 1035–1039.
Struhl, K., Stinchcomb, D. T. & Davis, R. W. (1980). *J. Mol. Biol.* **136**, 291–307.
Sutcliffe, J. G. (1978). *Nucl. Acids Res.* **5**, 2721–2728.
Wolfner, M., Yep, D., Messenguy, F. & Fink, G. (1975). *J. Mol. Biol.* **96**, 273–290.

Edited by S. Brenner