J. Mol. Biol. (1986) 191, 221–229

Yeast *HIS3* Expression in *Escherichia coli* Depends upon Fortuitous Homology between Eukaryotic and Prokaryotic Promoter Elements

Kevin Struhl

Department of Biological Chemistry
Harvard Medical School
Boston, Mass. 02115, U.S.A.

(Received 5 December 1985, and in revised form 23 April 1986)

The yeast imidazoleglycerolphosphate dehydratase gene *HIS3*, when introduced into *Escherichia coli*, is transcribed and translated with sufficient fidelity to produce functional enzyme. The following lines of evidence indicate that *E. coli* RNA polymerase recognizes a particular region of *HIS3* DNA as a promoter sequence. First, this promoter contains nucleotide sequences that resemble the canonical prokaryotic promoter elements, the −10 and −35 regions. Second, *HIS3* transcription *in vitro* by *E. coli* RNA polymerase is initiated at the predicted site downstream from the conserved sequences. Third, deletion mutations that successively encroach upon the 5′ end of the *HIS3* gene indicate that the promoter is necessary and sufficient for expression in *E. coli*. Fourth, a single base-pair change that behaves as an "up-promoter" mutation alters the −35 region such that it becomes identical with the consensus sequence.

Because the −10 region of this promoter coincides with the TATA promoter element that is necessary for expression in yeast cells, it is possible directly to compare prokaryotic and eukaryotic promoter function. Analysis of 51 deletion and substitution mutations indicates that the patterns of mutant phenotypes are quite different for each organism. Therefore, although prokaryotic −10 regions are similar in sequence to eukaryotic TATA elements and although the same *his3* region serves both functions, it appears that this represents an evolutionary coincidence whose current functional basis is minimal. The evolutionary significance of the homology between prokaryotic and eukaryotic promoter elements is discussed.

1. Introduction

Eukaryotic proteins can be synthesized in *Escherichia coli*, usually by converting a eukaryotic gene into a bacterial gene. Specifically, eukaryotic protein-coding sequences are fused to a prokaryotic regulatory region that specifies signals for efficient transcription and translation in *E. coli*. In many cases, the eukaryotic gene product is made as a hybrid protein in which the N-terminal amino acid residues derive from a bacterial gene.

Such genetic engineering is usually necessary. Most eukaryotic protein-coding regions are split by intervening sequences. These must be removed by RNA splicing in order to produce a functional message, a feat *E. coli* is unable to manage. In addition, the identity of the genetic code and the homology between the prokaryotic −10 region (consensus sequence TATAAT†; for reviews, see Rosenberg & Court, 1979; Hawley & McClure, 1983) and the eukaryotic TATA promoter element (consensus sequence TATAAA; for a review, see Breathnach & Chambon, 1981) is frequently insufficient for expression of foreign genes in *E. coli*. For example, lack of expression can be due to mRNA instability, inefficient translational initiation, or protein instability.

Nevertheless, functional expression of eukaryotic DNA in *E. coli* was initially demonstrated with native yeast DNA, specifically a segment encoding the imidazoleglycerolphosphate dehydratase gene *HIS3*. When introduced into *E. coli*, this segment permits bacterial mutants that lack the analogous bacterial gene (*hisB*) to grow in the absence of histidine (Struhl *et al.*, 1976); furthermore, it directs the synthesis of the yeast enzyme (Struhl & Davis, 1977). The protein synthesized in *E. coli* is almost certainly identical with that produced in yeast

† Sequence hyphens have been omitted throughout for clarity.

0022-2836/86/180221-09 $03.00/0

because (1) the *HIS3* gene contains no intervening sequences and (2) a particular AUG codon is the only reasonable candidate for translation initiation in either organism (Struhl & Davis, 1981).

Because this expression was originally observed in the absence of any known prokaryotic promoter sequence, it was concluded that the segment of yeast DNA was responsible for transcriptional initiation (Struhl *et al.*, 1976). The existence of such a promoter was demonstrated by deletion mutations that eliminated it (Struhl *et al.*, 1980), and crude deletion analysis suggested that it was located close to the structural gene (Struhl & Davis, 1980). However, precise localization was complicated because many of the deletion mutants examined conferred *HIS3* function as a consequence of the promoter at the bacteriophage λ attachment site (Struhl, 1981*a*).

Nevertheless, a probable site for the promoter that functions in *E. coli* can be inferred from the *HIS3* DNA sequence (Struhl, 1985; Fig. 1). The best fit to a typical *E. coli* promoter is a presumptive −35 region (TTGGCA located 94 to 99 nucleotides upstream from protein-coding sequences) and a potential −10 region (TACATT located between nucleotides −75 and −70). Although other DNA sequences in the general vicinity are more homologous to the canonical −10 region, they are not associated with sequences that even remotely resemble a classical −35 region.

Although eukaryotic promoters differ in many ways from their prokaryotic counterparts, one similarity has been mentioned frequently. The prokaryotic −10 and eukaryotic TATA elements are homologous in DNA sequence, and both are critical for achieving the maximal level of transcription and for specifying the actual start point (Rosenberg & Court, 1979; Mathis & Chambon, 1981; McKnight & Kingsbury, 1982; Grosveld *et al.*, 1981; Struhl, 1982*a*). In this light, the proposed promoter that mediates expression in *E. coli* is of interest because its −10 region coincides with the *HIS3* TATA element. Thus, it is possible directly to compare prokaryotic and eukaryotic promoter function by assessing the phenotypes of *his3* promoter mutants in yeast and in *E. coli*. In addition, this situation permits an unusual method for creating eukaryotic promoter mutations. By selecting for point mutations that cause either increased or decreased levels of expression in *E. coli*, alterations of the *HIS3* TATA element may arise; these could then be analyzed in yeast cells.

This first part of this paper demonstrates that the promoter necessary for *HIS3* expression in *E. coli* is indeed determined by the sequences that best fit the *E. coli* consensus. This is accomplished by transcription of *HIS3* DNA by *E. coli* RNA polymerase *in vitro*, and by DNA sequence analysis of "up-promoter" mutations. The second part of this paper compares the phenotypes of numerous *his3* derivatives for their ability to support expression in yeast and in *E. coli*. The main conclusion is that, although the *HIS3* TATA element and −10 region play important roles in the relevant host, their coincidence is fortuitous.

2. Materials and Methods

(a) *Phenotypic analyses*

The *λhis3* hybrid phages that successively delete *HIS3* 5′-flanking sequences were isolated and characterized in previous work (Struhl, 1981*b*). These deletion mutants were tested for *HIS3* expression in *E. coli* by the double lysogen method (Struhl *et al.*, 1976, 1980). The phages were infected into wild-type λ lysogens of *hisB*463, a strain lacking IGP† dehydratase activity. As the *λHIS3* phages are defective in *int* and *att* function, they are presumably integrated into the *E. coli* genome by homologous recombination with the prophage. The infected mixtures were plated in glucose minimal medium in the absence of histidine; derivatives capable of *HIS3* expression produce His^+ colonies. This method directly tests for the promoter, because all the derivatives contain the intact structural gene, and because all relevant λ promoters are repressed (Struhl *et al.*, 1980).

Internal deletion and M13 substitution mutations of the *HIS3* promoter region have been described (Struhl, 1982*a*,*b*). All the molecules contain a 6·1 kb *his3* DNA segment generated by *Eco*RI and *Sal*I cleavage that is cloned into the YRp14 vector. The mutant DNAs were introduced into the *hisB*463 strain by selecting for ampicillin-resistant transformants. The resulting strains were tested both for their ability to grow in the absence of histidine, and for their ability to grow in the presence of 10 mM or 20 mM-aminotriazole. As the yeast *HIS3* gene is not regulated in *E. coli* as a function of histidine starvation (Struhl *et al.*, 1980), such growth tests constitute a direct assay *in vivo* for the level of *HIS3* expression in *E. coli*.

(b) HIS3 *transcription* in vitro

Transcription of 1 μg of pUC8-Sc2605 (closed circular form) was performed in 5-μl reactions containing 40 mM-Tris·HCl (pH 7·9), 10 mM-$MgCl_2$, 0·1 mM-dithiothreitol, 250 μM each ATP, GTP, CTP and UTP, 10 units of RNase inhibitor (Promega Biotec), KCl (at 0, 50 or 150 mM, depending on the reaction) and 1 unit of *E. coli* RNA polymerase holoenzyme (New England Biolabs). pUC8-Sc2605 contains the 6·1 kb *Eco*RI–*Sal*I *HIS3* fragment (Struhl & Davis, 1980) cloned into the pUC8 vector of Vieira & Messing (1982). The reactions were incubated for 15 min at 37°C and terminated by the addition of 15 μl of a hybridization solution containing approx. 5 ng of ^{32}P-labeled probed (10^5 cts/min) and 40 μg of tRNA, such that the final concentrations were 40% (v/v) formamide, 0·1 M-Tris·HCl (pH 7·5), 0·4 M-NaCl, 5 mM-EDTA. The hybridization probe was prepared by labeling the 5′-end of the relevant 169 base-pair *Hin*fI segment with ^{32}P, and purifying the non-coding strand after denaturation with alkali and electrophoresis in 5% (w/v) polyacrylamide (Struhl & Davis, 1981). After hybridization for 8 h at 30°C, the products were treated with S_1 nuclease and then electrophoretically separated in 6% polyacrylamide gels containing 7 M-urea. The lengths of the protected DNA fragments were determined by comparing the mobility

† Abbreviations used: IGP, imidazoleglycerol-phosphate; kb. 10^3 bases or base-pairs.

with a series of standards produced by dideoxy sequencing reactions (data not shown).

3. Results

(a) *Sequential deletion analysis*

Previous experiments designed to localize the *HIS3* sequences that function as a promoter in *E. coli* were difficult to interpret. Many of the deletion mutations resulted in the fusion of *HIS3* structural sequences to the bacteriophage λ attachment site. Surprisingly, the core of these hybrid attachment sites formed part of a functional promoter, thereby making it impossible to determine when the promoter had actually been eliminated (Struhl, 1981*a*). This problem is circumvented by using a different set of deletion mutations that successively remove sequences upstream from the structural gene. In these derivatives, *HIS3* sequences are fused to different positions of DNA, all of which are unlikely to contain bacteriophage promoters that function during conditions of lysogeny (Struhl, 1981*b*; see Fig. 1).

Fifteen *λhis3* hybrid phages were introduced into the *E. coli* genome, as described in Materials and Methods. All the derivatives that contain at least 119 base-pairs upstream from the IGP dehydratase coding region express *HIS3* in *E. coli* (Table 1). This indicates that the promoter is located no more than 119 base-pairs from the structural gene, a region that contains the putative −10 and −35 elements mentioned in the Introduction. In contrast, most of the derivatives that contain less than 98 base-pairs adjacent to the structural gene fail to express the gene. All of these derivatives remove the putative −35 region and some of them remove the putative −10 region as well. As indicated in Table 1, Sc2782 confers a His^+ phenotype, even though it presumably deletes the putative −35 element. Possible explanations for this exceptional case are a fortuitous −35 element located in the flanking DNA or readthrough transcription from the flanking sequences. In this regard, it should be noted that Sc2782 is unusual, in that the *HIS3* sequences ending at position −86 are not fused to λ sequences, but rather to IS*1* sequences (Oettinger & Struhl, 1985).

```
                2778  2765                                2784          2783
-290  CTAGGAGTCA CTGCCAGGTA TCGTTTGAAC ACGGCATTAG TCAGGGAAGT CATAACACAG

                          2781                                2757
-230  TCCTTTCCCG CAATTTTCTT TTTCTATTAC TCTTGGCCTC CTCTAGTACA CTCTATATTT

                                    2771                2786/2755
-170  TTTTATGCCT CGGTAATGAT TTTCATTTTT TTTTTTCCAC CTAGCGGATG ACTCTTTTTT

                2767        2782                                2763
-110  TTTCTTAGCG ATTGGCATTA TCACATAATG AATTATACAT TATATAAAGT AATGTGATTT

                                  2779 2773                  MetSerGlu
 -50  CTTCGAAGAA TATACTAAAA AATGAGCAGG CAAGATAAAC GAAGGCAAAG ATGACAGAGC
```

Figure 1. Nucleotide sequence of the *HIS3* promoter region with respect to deletion end points. The DNA sequence of the *HIS3* coding strand from nucleotides −290 to +10 (with respect to the protein-coding region) is shown. Putative −10 and −35 regions are underlined. Approximate *HIS3* end points for sequential 5′ deletions are shown above the DNA sequence (see Table 1).

(b) HIS3 *transcription* in vitro *using* E. coli *RNA polymerase holoenzyme*

Judith Jaehning (unpublished results) has shown that *E. coli* RNA polymerase holoenzyme binds specifically to the promoter region and initiates transcription *in vitro*. To confirm this observation, supercoiled pUC8-Sc2605 DNA (containing a 6·1 kb fragment of *HIS3* DNA) was transcribed with *E. coli* RNA polymerase. After the synthesis reaction, the products were immediately hybridized to single-stranded probe DNA and then treated with S_1 nuclease. The hybridization probe was end-labeled with ^{32}P at the *Hin*fI site located 48 bases from the AUG initiation codon.

The results of this experiment (Fig. 2) indicate that *E. coli* RNA polymerase initiates transcription 112±2 bases away from the *Hin*fI site, which corresponds to nucleotide −64 with respect to the *HIS3* structural gene. By analogy with the properties of a large number of *E. coli* promoters, this result strongly suggests that the sequence TACATT located between nucleotides −75 and −70 is recognized as a −10 region by *E. coli* RNA polymerase holoenzyme. From Figure 2, a faint band 103 bases in length and two strong bands 162 and 155 bases in length are also apparent. The faint band corresponds to a a minor transcript initiating at nucleotide −55, and it is probably due to a −10 element at nucleotides −68 to −63 (TATAAA) or −70 to −65 (TATATA) and a −35 element at −91 to −86 (ATCACA) or −93 to −88 (TTATCA). These presumptive sequences are less homologous to the canonical *E. coli* promoter elements (Hawley & McClure, 1983) than to the sequences that specify

Table 1
Phenotypes of sequential 5′ deletions

Phage	*HIS3* end point	λ end point	Phenotype
Sc2778	−282±5	0·560	+
Sc2765	−277±5	0·558	+
Sc2784	−247±5	0·560	+
Sc2783	−232±5	0·558	+
Sc2781	−207±5	0·550	+
Sc2757	−182±3	0·544	+
Sc2771	−140±3	0·564	+
Sc2786	−118±2	0·592	+
Sc2755	−119	0·566	+
Sc2767	−98±2	0·568	−
Sc2782	−89	IS*1*-0·675	+
Sc2763	−55±2	0·566	−
Sc2779	−24±2	0·560	−
Sc2787	−22±2	0·550	−
Sc2773	−20±2	0·566	−

The *λhis3* deletion phages listed in the left column are derived from λgt9-Sc2601, and the *HIS3* and λ deletion end points have been determined (Struhl, 1981*b*; Oettinger & Struhl, 1985). In this paper, the *HIS3* end points are defined with respect to the AUG initiation codon (see Fig. 1). In most cases, these were determined by S_1 nuclease mapping and hence are not localized to a precise nucleotide. Sc2782 is unusual, in that the *E. coli* IS*1* element is inserted between the *HIS3* and λ deletion break points. The *HIS3* phenotypes were determined and are indicated by their ability (+) or inability (−) to grow in the absence of histidine.

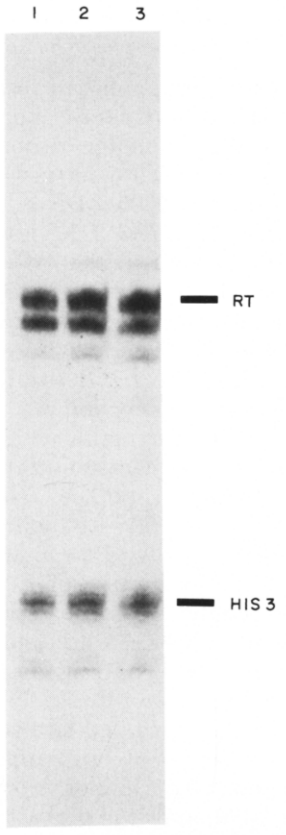


Figure 2. *HIS3* transcription *in vitro* by *E. coli* RNA polymerase holoenzyme. Positions corresponding to the major *HIS3* transcript at +64 (HIS3) and readthrough transcripts (RT) are indicated. Transcription reactions were performed as described in Materials and Methods, and differ solely by the concentration of KCl (lane 1, 0 mM: lane 2, 50 mM: lane 3, 150 mM).

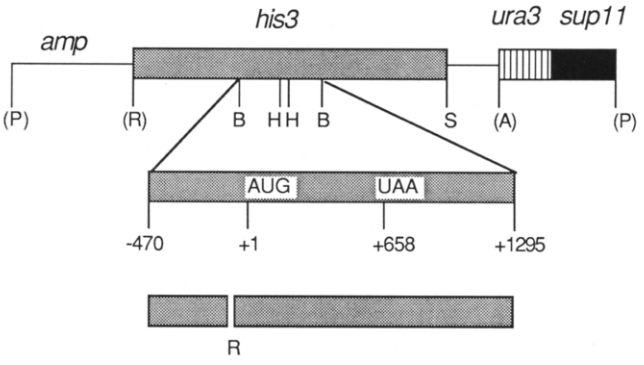


Figure 3. Structure of the YRp14-*HIS3* DNA molecules. The top part of the Figure illustrates the structure of the YRp14-*his3* molecules (drawn to scale and depicted in the linear form). The 6·1 kb *HIS3* gene region is represented as a filled gray bar. The YRp14 vector is composed of pBR322 sequences containing the ampicillin-resistance gene (*amp*) and the origin of replication (line), the yeast *ura3* gene (striped box) and the ochre allele of the yeast *sup11* gene (black box). Restriction endonuclease cleavage sites are indicated as follows: P, *Pvu*II; R, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; S, *Sal*I; A, *Ava*I; other *Pvu*II and *Ava*I sites are not shown. Parentheses around the letters indicates mutated restriction sites. The 2nd line shows an expanded view of the *HIS3* region, including the co-ordinates of the *Bam*HI sites, the AUG initiation codon, and the UAA termination codon. The bottom line indicates the structure of the *his3* internal deletions. Many of these contain *Eco*RI linkers at the deletion break points.

the major transcript. The strong 162-base band is due to complete protection of the hybridization probe from nuclease S_1, and is indicative of transcripts that initiate further upstream than nucleotide −140. Since these experiments *in vitro* are performed in the absence of the *rho* termination factor, it is not clear if these readthrough transcripts are actually synthesized *in vivo*. The strong band that is approximately 155 bases in length appears to be produced by artifactual S_1 nuclease digestion of the nine base-pair dA-dT homopolymer stretch; hence it probably also corresponds to readthrough transcription.

(c) his3 *mutations that increase expression in* E. coli

When integrated into the *E. coli* chromosome in single copy at the bacteriophage λ attachment site, the *HIS3* gene allows *E. coli* hisB mutants lacking IGP dehydratase activity to grow in the absence of histidine (Struhl *et al.*, 1976). However, the level of expression in such strains is insufficient to allow growth in the presence of aminotriazole, a competitive inhibitor of yeast IGP dehydratase (Struhl *et al.*, 1980). Spontaneously derived mutants selected for aminotriazole resistance result in strains that overproduce yeast IGP dehydratase (Brennan & Struhl, 1980). These were due to *E. coli* chromosomal mutations, or to alterations of the *HIS3* gene, which included a deletion, an IS2 insertion, and a point mutation that destroyed a terminator for readthrough transcription (Brennan & Struhl, 1980).

To enrich for point mutations of the *HIS3* promoter region, eight independently derived, high titer stocks of λgt4-Sc2601 (Struhl *et al.*, 1980) were prepared following growth in NK5154, a strain containing the *mutD* mutation described by Degnan & Cox (1974). The resulting phage were infected into an *E. coli* strain containing the *hisB*463 mutation, and aminotriazole-resistant colonies were selected. The frequency of aminotriazole-resistant colonies was three to four orders of magnitude higher when compared with phage that were passaged on normal strains. The prophages from eight colonies (one from each original phage stock) were induced, and the phage arising conferred aminotriazole resistance when reinfected into fresh *hisB*463 cells. This proves that the mutations causing aminotriazole resistance were carried on the original infecting phage.

To determine the DNA sequence changes in these eight presumptive up promoter mutations, the relevant *Bam*HI–*Hin*dIII fragments were cloned

into mp8 (Vieira & Messing, 1982). Single-stranded DNAs from these hybrid phages were prepared and annealed to a 17-base oligonucleotide primer that corresponds to nucleotides between +3 and +19 of the anti-sense strand. These hybridization mixtures were then subjected to the dideoxy sequencing method described by Sanger *et al.* (1980). All eight phages contain the same DNA sequence alteration, in that the G residue at position −96 has been changed to an A residue. This location coincides with the −35 promoter element predicted in the previous sections. Moreover, the new sequence, TTGACA, is identical with the canonical −35 sequence for *E. coli* promoters.

(d) *Expression of* his3 *promoter mutants in* E. coli

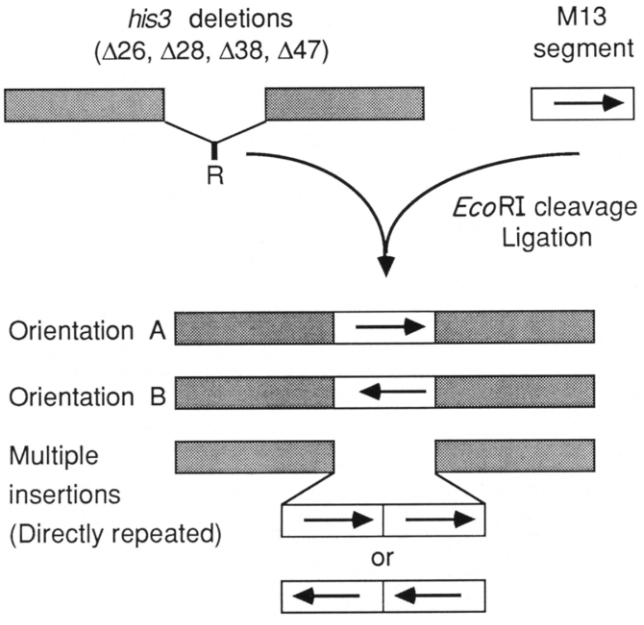
In previous work, I generated two kinds of *his3* derivatives to analyze the TATA promoter element in yeast (Struhl, 1982*a*). In one class, the promoter region was divided into upstream and downstream "halves" by *Eco*RI linker insertions, and then recreated by pairwise combination; this results in a matrix of deletion mutations (Fig. 3). In the other type, a 31-base-pair sequence from phage M13 replaced *HIS3* sequences deleted in four representatives of the matrix. The M13 sequence was inserted in both possible orientations and in single or multiple, directly repeated copies (Fig. 4). All the molecules used in these experiments contained a 6·1 kb *Eco*RI–*Sal*I *HIS3* DNA fragment that was cloned into YRp14 (Struhl, 1982*a,b*; Fig. 3). Upon introduction into the *his*B463 strain, the resulting strains were tested both for their ability to grow in the absence of histidine, and for their ability to grow in the presence of 10 mM or 20 mM-aminotriazole. These growth tests constitute a direct assay *in vivo* for the level of *HIS3* expression in *E. coli* (Struhl *et al.*, 1980; Brennan & Struhl, 1980).

Strains containing the wild-type *HIS3* derivative YRp14-Sc2605 grow at wild-type rates in the absence of histidine. They also grow, although a bit

Table 2
Phenotypes of internal deletions

DNA fragment	*his3* allele	End points			Phenotypes	
		Upstream	Junction	Downstream	*E. coli*	*S. cerevisiae*
Wild type					+	+*
Sc2857	Δ19	−58	GGAATTCC	−58	+	+*
Sc2854	Δ18	−58	GGAATTCC	−47	+	+*
Sc2855	Δ36	−58	GGAATTCC	−34	++	+*
Sc2883	Δ39	−106	GGAATTCC	−34	−	−
Sc2889	Δ40	−132	GGAATTCC	−34	±	−
Sc2886	Δ41	−139	GGAATTCC	−34	−	−
Sc2895	Δ42	−146	GGAATTCC	−34	±	−
Sc2882	Δ43	−106	GGAATTCC	−47	±	−
Sc2888	Δ44	−132	GGAATTCC	−47	±	−
Sc2885	Δ45	−139	GGAATTCC	−47	+	−
Sc2894	Δ46	−146	GGAATTCC	−47	±	−
Sc2884	Δ38	−106	GGAATTCC	−58	±	−
Sc2890	Δ47	−132	GGAATTCC	−58	±	−
Sc2887	Δ48	−139	GGAATTCC	−58	±	−
Sc2896	Δ49	−146	GGAATTCC	−58	±	−
Sc3121	Δ26	−106	GGAATTCC	−78	±	+*
Sc3125	Δ28	−132	GGAATTCC	−78	++	+
Sc3129	Δ30	−146	GGAATTCC	−78	+	+
Sc3122	Δ27	−106	GGAATTCC	−98	++	+*
Sc3126	Δ29	−132	GGAATTCC	−98	++	+
Sc3130	Δ37	−146	GGAATTCC	−98	++	+
Sc3101	Δ20	−69		−57	+	+
Sc3102	Δ21	−82		−32	±	+
Sc3110	Δ22	−76		−57	+	+
Sc3111	Δ23	−73		−50	±	+
Sc3112	Δ24	−67	CC	−58	±	+
Sc3113	Δ25	−84	TCCC	−58	−	+
Sc3138	Δ31	−132	GGAATTCC	−135	±	+*
Sc3159	Δ32	−132		−128	+	+*
Sc3160	Δ33	−132		−126	+	+*
Sc3161	Δ34	−144		−133	+	+*
Sc3165	Δ35	−162		−127	+	+

The structures of these *his3* deletion mutants are shown in Fig. 3 and their DNA sequences have been determined (Struhl, 1982*a,b*). The phenotypes produced in *E. coli* are listed as follows: ++ indicates wild-type growth rates in 10 mM and 20 mM-aminotriazole; + indicates growth in 10 mM but not 20 mM-aminotriazole; ± indicates no growth in 10 mM-aminotriazole and reduced growth rates in the absence of this inhibitor; − indicates extremely poor, or no, growth in the absence of histidine. The phenotypes produced in yeast have been determined (Struhl, 1982*a,b*) and are listed as + for growth in the absence of histidine or − for no growth. Derivatives that result in induced *HIS3* expression under conditions of amino acid starvation are indicated by an asterisk.



(A) 5'-AATCCGGACGTTGTAAAACGACGGCCAGTG-3'

(B) 5'-AATCACTGGCCGTCGTTTTACAACGTCCGG-3'

Figure 4. Structure of the M13 substitition mutants. The top of the Figure shows the structure of the original *his3* deletion mutants as grey bars (wild-type sequence), empty space (deleted DNA), and the *Eco*RI site. Ligation of the *Eco*RI-cleaved *his3* mutant DNA and the 31 base-pair M13 *Eco*RI segment (an open box with an arrow pointing rightward for orientation A and leftward for orientation B) produces the structures shown below. Many derivatives have multiple tandem insertions, all oriented in the same direction. The sequences of the coding strands of orientation A and B are shown, and a possible −35 region is underlined.

slowly, when 10 mM-aminotriazole is added to the medium; but they fail to grow when the drug concentration is raised to 20 mM. This aminotriazole resistance conferred by YRp14-Sc2605, which is not observed with λgt4-Sc2601, is due to the multiple copies of the plasmid/cell (Struhl *et al.*, 1980; Brennan & Struhl, 1980). The phenotypes conferred by 51 mutant derivatives are listed in Tables 2 and 3. A + entry in the Table indicates that the derivative of interest confers the same growth properties as the wild-type gene. An entry of ++ indicates that the strain grows at wild-type rates in the presence of 10 mM and 20 mM-aminotriazole, and hence an increased level of *HIS3* expression. A ± phenotype indicates that the strains fail to grow in 10 mM-aminotriazole, although they do grow at reduced rates in the absence of this inhibitor; this corresponds to reduced levels of *HIS3* expression in comparison with the wild type gene. A listing of − means that the strains grow poorly, if at all, in the absence of histidine, indicating a severe defect in *HIS3* expression.

A number of phenotypic patterns emerge from this experiment. First, small alterations downstream from the putative −10 region (Δ18, Δ19) do not affect *HIS3* expression, whereas a more extensive deletion (Δ36) that removes sequences downstream of −47 (relative to the ATG initiation codon) results in increased expression. Second, deletion mutants that remove the putative −10 region (Δ21, Δ23 and Δ25) generally reduce *HIS3* expression, although Δ22 apears to be an exception. Third, with the possible exception of Δ45, deletions that remove both the −10 and −35 regions (Δ38 to Δ44 and Δ47 to Δ49) reduce *HIS3* expression; in some cases (Δ39 and Δ41), the phenotype is especially severe. Fourth, small alterations upstream from the putative −35 region (Δ31 to Δ35) do not affect the level of *HIS3* expression. Fifth, derivatives with a downstream end point at −98 (Δ27, Δ29 and Δ37) all have increased levels of *HIS3*, whereas those with a downstream end point at −78 (Δ26, Δ28 and Δ30) have variable levels depending on the fused sequences. Sixth, all 11 derivatives containing the M13 segment in orientation A show reduced *HIS3* expression, and many of these are severely defective. Seventh, some derivatives containing the M13 segment in orientation B (Δ53, Δ56 and Δ63) show unusually high levels of *HIS3* expression, whereas others (Δ58, Δ60 and Δ67) confer extremely poor expression.

4. Discussion

(a) *Promoter sequences that mediate* HIS3 *expression in* E. coli

As described in the Introduction, previous genetic experiments indicated that sequences upstream from the yeast *HIS3* structural gene act as a promoter for expression in *E. coli* (Struhl *et al.*, 1976; Struhl & Davis, 1980). Several lines of evidence suggest that *HIS3* sequences between coordinates −70 and −75 and between −94 and −99 act as −10 and −35 elements, respectively, to constitute a functional *E. coli* promoter. First, these sequences are similar to the consensus promoter sequences, and they are separated by 18 base-pairs, a functionally acceptable distance (Hawley & McClure, 1983). The TACATT sequence between −70 and −75 represents a four out of six fit with the canonical −10 element, including the three most conserved nucleotides. The TTGGCA sequence between −94 and −99 represents a five out of six fit with the canonical −35 element. Second, *E. coli* RNA polymerase initiates transcription at −64, a position that lies the expected distance from the −10 region. Third, a G to A transition mutation at position −96, which results in a perfect fit to the consensus −35 element, causes increased expression of *HIS3*.

Phenotypic analysis of the deletion mutants generally supports the view that these sequences constitute the promoter that is necessary for expression in *E. coli*. Deletions that remove sequences upstream or downstream from the

Table 3
Phenotypes of M13 substitiuion mutations

DNA fragment	*his3* allele	End points	M13 inserts	Orientation	*E. coli*	*S. cerevisiae*
Wild type					+	+
Sc3141	Δ50	−106 −78	1	A	−	+
Sc3142	Δ51	−106 −78	1	B	++	+
Sc3150	Δ52	−106 −78	2	A	−	+
Sc3151	Δ53	−106 −78	3	B	++	+
Sc3164	Δ54	−106 −78	5	A	+	+
Sc3139	Δ55	−106 −58	1	A	−	+
Sc3140	Δ56	−106 −58	1	B	++	−
Sc3147	Δ57	−106 −58	2	A	−	+
Sc3162	Δ58	−106 −58	2	B	−	−
Sc3148	Δ59	−106 −58	3	A	−	+
Sc3163	Δ60	−106 −58	3	B	−	−
Sc3149	Δ61	−106 −58	4	A	−	+
Sc3145	Δ62	−132 −78	1	A	±	+
Sc3146	Δ63	−132 −78	1	B	++	+
Sc3153	Δ64	−132 −78	2	A	±	+
Sc3154	Δ65	−132 −78	3	A	±	+
Sc3143	Δ66	−132 −58	1	A	±	+
Sc3144	Δ67	−132 −58	1	B	±	−
Sc3152	Δ68	−132 −58	2	A	−	+

For each derivative, the original end points and the number and orientation of the M13 segment (A and B as described in Fig. 4) are indicated (Struhl, 1982*a*). The phenotypes in yeast and in *E. coli* are listed as described in Table 2. These derivatives have not been examined for their ability to induce *HIS3* expression in yeast cells during amino acid starvation.

proposed elements do not generally affect levels of *HIS3* expression. The only exception, Δ36, is a deletion that should shorten the RNA leader before the AUG initiation codon; hence the increased *HIS3* expression observed in these derivatives may be due to more efficient translation. In contrast, deletions that remove one or both of the elements usually result in reduced expression. Although occasional mutations delete part of the promoter region without causing detectable effects on *HIS3* expression, the most likely explanation of these "aberrant" phenotypes is that the novel joint created by the deletion generates a sequence that functions as a promoter element. This probably explains the variable phenotypes of derivatives with a downstream end point at −78, in which different DNA sequences are fused to the −10 region. The basis for increased expression seen in derivatives with a downstream end point at −98 is obscure. Thus, although the genetic data presented here are insufficient precisely to localize the promoter, the overall patterns of *HIS3* expression strongly support the location inferred from the DNA sequence and determined by transcription *in vitro*.

The phenotypes of the M13 substitution mutants, all of which disrupt the promoter region, also fit a pattern. In general, insertion of the M13 segment in orientation A does not restore normal levels of *HIS3* expression, whereas insertion in orientation B often results in increased expression. This strongly suggests that orientation B of the M13 segment contains an element that, when fused to appropriate sequences, can produce a functional *E. coli* promoter. Although the basis for these effects is unclear, the most likely explanation is that the B orientation contains several potential −35 regions, including the sequence TTTACA, a five out of six match.

(b) *Comparison of* HIS3 *expression in* E. coli *and in yeast*

In yeast cells, *HIS3* transcription is initiated at two distinct sites that map at positions −23 and −12 with respect to the AUG initiation codon (Struhl, 1985). However, although *HIS3* transcription is initiated from different positions in a prokaryotic or eukaryotic host, the results presented above and summarized in Figure 5 indicate that the −10 element of the *E. coli* promoter is located within the region that encodes the TATA element necessary for transcription in yeast (Struhl, 1982*a*, 1984). This suggests the possibility that prokaryotic −10 regions have mechanistic similarities to eukaryotic TATA elements, even though the basic initiation processes are clearly different. As the derivatives tested in this paper have been examined for their ability to support *HIS3* expression in yeast (Struhl, 1981*b*, 1982*a*,*b*), and as all of them contain the entire *HIS3* structural gene, it is possible directly to compare prokaryotic and eukaryotic promoter function (Tables 2 and 3).

The clear result is that the phenotypic patterns observed in *E. coli* are very different from those observed in yeast. For many individual mutations, the level of *HIS3* expression in one organism does

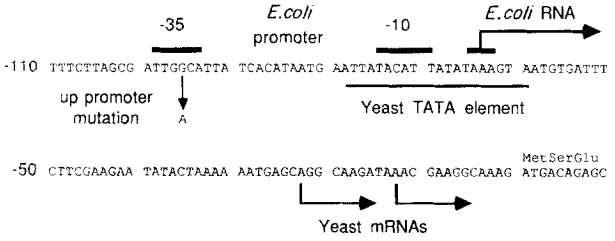


Figure 5. *HIS3* promoter elements for expression in *E. coli* and in yeast. The nucleotide sequence of the *HIS3* coding strand from nucleotides −110 to +10 is shown. The −10 and −35 elements for expression in *E. coli* as well as the site of transcription initiation *in vivo* are shown as lines above the nucleotide sequence. The location of the yeast TATA region and the eukaryotic mRNA initiation sites *in vivo* are shown below the sequence (Struhl, 1982*a*, 1984, 1985). The G to A transition at position −96 that causes increased expression in *E. coli* is indicated.

not correlate with the expression level in the other. In some of these cases, the relative expression is higher in *E. coli*, whereas in others the relative expression is higher in yeast. Perhaps the most obvious differences are observed in derivatives containing the M13 segment. In *E. coli*, *HIS3* expression in mutants with the segment in orientation A is poor in comparison with those containing orientation B, whereas the situation is reversed in yeast. In considering small deletions of the −10/TATA region, Δ45 functions very well in *E. coli* but poorly in yeast, whereas Δ21, Δ23 and Δ24 function normally in yeast but poorly in *E. coli*. In the M13 substitutions, orientation A can substitute functionally for the yeast *HIS3* TATA element, whereas orientation B cannot. In contrast, orientation A can not substitute for the *E. coli* −10 element, although orientation B appears to substitute for the −35 element. Thus, many of the phenotypic differences can be ascribed to functional differences between the −10 and TATA elements.

(c) *The use of deletion and substitution mutations to study promoter function*

Detailed analysis of eukaryotic genes became possible with the advent of recombinant DNA technology. One consequence of this is that eukaryotic promoters have been defined almost exclusively with deletion and substitution mutations because these are obtained more easily by DNA manipulations. In contrast, prokaryotic promoter mutations were isolated primarily by classical genetic techniques and hence are due typically to single-base-pair changes, the most common lesion *in vivo*. This means that comparisons of prokaryotic and eukaryotic promoters are complicated by the different methodologies used to elucidate their functional components.

A unique aspect of the work described here is that the identical mutants are used to examine a prokaryotic and a eukaryotic promoter. As discussed previously (Struhl, 1981*b*, 1982*a*,*b*, 1984; see Tables 2 and 3), it is apparent that the phenotypic patterns observed in yeast cells are extremely consistent, thereby facilitating the localization of the important elements. In contrast, the phenotypic patterns in *E. coli* are much less well defined. In the absence of other information, the genetic data in this paper would lead to a crude localization of the promoter region, but they would be insufficient to define the promoter elements.

As the clear difference in the interpretability of the results can not be explained by the methodology, it must reflect a basic difference in the mechanism of transcription. The deletion and substitution mutations described here cause relatively large rearrangements of the local DNA sequence, and hence are crude genetic tools. Thus, the observation that the phenotypes in yeast can be arranged into a consistent pattern strongly suggests that the yeast promoter is modular, in that it consists of separable elements whose precise spacing relationshp is unimportant. Consequently, it suggests that the separate elements are recognized by different proteins. In contrast, the relative variability of the phenotypes in *E. coli* suggests that the promoter is a compact unit whose components are relatively inseparable. Indeed, it is well-established that an *E. coli* promoter is defined by the precise interaction of RNA polymerase with properly spaced −10 and −35 elements. Thus, the direct comparison of the yeast and the *E. coli HIS3* promoters provides clear genetic evidence of the mechanistic differences between the prokaryotic and eukaryotic transcription machinery.

(d) *Evolutionary significance of the conservation beween prokaryotic and eukaryotic elements*

Although the same region of the wild-type *HIS3* gene serves both as a prokaryotic −10 region and as a eukaryotic TATA element (Fig. 5), the clear difference in phenotypic patterns in *E. coli* and in yeast strongly suggests that this is a coincidence with little functional basis. However, the homology between the prokaryotic −10 region (consensus TATAAT) and the eukaryotic TATA element (consensus TATAAA) indicates that such a coincidence will occur frequently. This may explain why native yeast genes are often expressed in *E. coli*.

If these promoter elements are mechanistically unrelated, why are their DNA sequences so conserved? One hypothesis is that promoter elements contain dA-dT-rich DNA sequences because these favor localized denaturation, a common and important step of transcriptional initiation. However, this is unlikely to be the entire explanation because (1) the homology between the prokaryotic −10 region and the eukaryotic TATA element goes beyond simple dA-dT richness, and (2) many equally dA-dT rich sequences do not function as promoter elements. As an alternative suggestion,

perhaps the common ancestor to prokaryotic and eukaryotic organisms used this DNA sequence for transcriptional initiation. Presumably, this hypothetical organism contained many genes, thus making it difficult to change the "common transcription sequence" without wreaking havoc. Similar arguments have been used to explain why the genetic code is essentially universal, even though prokaryotic and eukaryotic mechanisms of translation are quite different. By analogy, prokaryotic and eukaryotic organisms may have evolved different transcriptional mechanisms to utilize the original common sequence. Thus, although the prokaryotic −10 region and the eukaryotic TATA element probably do not have identical nucleotide requirements, their sequences remain strongly conserved.

I thank Judith Jaehning for communication of unpublished results, Nancy Kleckner for strain NK5154, and Stephen Apfelroth for determining the DNA sequence of the up-promoter mutations. This work was supported by a grant from the National Institutes of Health (GM30186).

References

Breathnach, R. & Chambon, P. (1981). *Annu. Rev. Biochem.* **50**, 349–383.

Brennan, M. B. & Struhl, K. (1980). *J. Mol. Biol.* **136**, 333–338.

Degnan, G. E. & Cox, E. C. (1974). *J. Bacteriol.* **117**, 477–487.

Grosveld, G. C., Shewmaker, C. K., Jat, P. & Flavell, R. A. (1981). *Cell*, **25**, 215–226.

Hawley, D. K. & McClure, W. R. (1983). *Nucl. Acids Res.* **11**, 2237–2255.

Mathis, D. J. & Chambon, P. (1981). *Nature (London)*, **290**, 310–315.

McKnight, S. L. & Kingsbury, R. (1982). *Science*, **217**, 316–324.

Messing, J. & Vieira, J. (1982). *Gene*, **19**, 269–276.

Oettinger, M. A. & Struhl, K. (1985). *Mol. Cell. Biol.* **5**, 1901–1909.

Rosenberg, M. & Court, D. (1979). *Annu. Rev. Genet.* **13**, 319–353.

Sanger, F., Coulson, A. R., Barell, B. G., Smith, A. J. & Roe, B. A. (1980). *J. Mol. Biol.* **143**, 161–178.

Struhl, K. (1981*a*). *J. Mol. Biol.* **152**, 517–533.

Struhl, K. (1981*b*). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 4461–4465.

Struhl, K. (1982*a*). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 7385–7839.

Struhl, K. (1982*b*). *Nature (London)*, **300**, 284–287.

Struhl, K. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 7865–7869.

Struhl, K. (1985). *Nucl. Acids Res.* **13**, 8587–8601.

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**, 553–568.

Struhl, K., Cameron, J. R. & Davis, R. W. (1976). *Proc. Nat. Acad. Sci., U.S.A.* **73**, 1471–1475.

Struhl, K., Stinchcomb, D. T. & Davies, R. W. (1980). *J. Mol. Biol.* **136**, 291–307.

Vieira, J. & Messing, J. (1982). *Gene*, **19**, 259–268.

Edited by I. Herskowitz