Volume 13 Number 23 1985 Nucleic Acids Research

# Nucleotide sequence and transcriptional mapping of the yeast *pet56-his3-ded1* gene region

Kevin Struhl

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

Received 28 August 1985; Revised 24 October 1985; Accepted 29 October 1985

## ABSTRACT

Genes of the baker's yeast *Saccharomyces cerevisiae* are densely clustered on 16 linear chromosomes. Here, I characterize a 1.8 kb region of chromosome XV containing the entire structural gene for the histidine biosynthetic enzyme imidazoleglycerolphosphate (IGP) dehydratase (*his3*) as well as the promoter sequences and 5'-proximal mRNA coding regions for the adjacent genes. The *his3* gene encodes several mRNA species averaging 820 bases in length, all of which contain an open reading frame of 219 codons. The location of this open reading frame coincides with the *his3* gene as defined by functional criteria, suggesting that the primary translation product of yeast IGP dehydratase has a molecular weight of 23,850. Phenotypic analysis of mutations constructed *in vitro* indicate that one of the adjacent genes (*pet56*) is required for mitochondrial function, whereas the other gene (*ded1*) is essential for cell viability. The *pet56* and *his3* genes are transcribed divergently from initiation sites that are separated by only 192 bp. Transcription of the *ded1* gene is initiated only 130 bp beyond the 3'-end of the *his3* mRNA coding region. These results suggest that these unrelated genes are located extremely close together and that the spacer regions between them consist largely of promoter and terminator sequences.

## INTRODUCTION

The genome of the yeast *Saccharomyces cerevisiae* contains approximately 10,000 kb of DNA, about half of which is transcribed under normal growth conditions (1,2). However, genes encoding related functions are not clustered, but rather are distributed randomly on the 16 chromosomes that constitute the yeast genome. For example, the genes encoding the histidine biosynthetic enzymes are located on 6 different chromosomes (3), whereas in *E.coli*, they are organized as a single operon (4).

The *his3* gene, which encodes imidazoleglycerolphosphate (IGP) dehydratase, is located on chromosome XV (5). The gene, originally obtained as a 10.1 kb *Eco*RI-generated DNA fragment, was isolated by virtue of its ability to be functionally expressed in *E.coli* and hence to permit strains lacking the analogous bacterial enzyme to grow in the absence of histidine (6). The original *his3* DNA fragment encodes 6 discrete transcripts that presumably represent the products of six different genes (7). The *his3* promoter/regulatory region has been investigated in great detail by constructing mutant derivatives *in vitro* and analyzing the expression in *E.coli* (8) and yeast cells (9-12). However, these analyses have essentially ignored the adjacent transcripts as well as the *his3* structural gene itself.

Transcriptional mapping (7) indicates that a 1.7 kb *Bam*H1 DNA fragment, Sc2676, contains the entire *his3* structural gene as well as the 5'-proximal mRNA coding regions for a 1.2 kb species located upstream of *his3* and a 2.3 kb species located downstream (Fig. 1). The *his3* and 1.2 kb RNAs are transcribed divergently, whereas the 2.3 kb RNA is transcribed in the same direction as *his3*. Under normal conditions, the *his3* and 1.2 kb transcripts are transcribed at the average level for yeast genes (1-2 mRNA molecules/cell at the steady state)(7), whereas the level of the 2.3 kb transcript is about five-fold higher. *His3* transcription levels are regulated as a function of amino acid starvation, whereas neither of the adjacent genes are subject to this control mechanism. Thus, this region of chromosome XV should be typical of the yeast genome in general in that the genes encode unrelated functions and are closely packed.

This paper describes the genetic organization of this chromosomal region in more detail by determining the nucleotide sequence, the positions of mRNA initiation and termination within the region, and the functions of the adjacent genes. Of particular interest are the boundaries between different genes. In one case, promoter regions for divergently transcribed genes that are not co-regulated are located in close proximity, whereas in the other case, the promoter region of one gene is located very near the termination region of the neighboring gene.

## MATERIALS AND METHODS

### Microbiological techniques

The yeast strains used in this work were KY114 (a *ura3*-52 *ade2*-101 *trp1*-Δ1 *lys2*-801 *HIS3*$^+$), KY463 (same as KY114 except *cyh2*$^r$), KY115 (α version of KY114), and KY462 (α *ura3*-52 *cyh2*$^r$ *ade2*-101 *trp1*-Δ1 *LYS2*$^+$ *his3*-Δ200). Methods for cell growth, mating sporulation, and ascus dissection have been described previously (13). Hybrid DNA molecules were propagated in *E.coli* cells using strains EQ82 (for plasmid DNAs)(9) or JM101 (for M13 derivatives)(14).

### DNA mechanics

Methods for the preparation of plasmid (15) and single stranded phage DNAs (16) have been described previously. To construct hybrid DNA molecules, appropriate DNA segments were purified by electrophoresis in low temperature gelling/melting agarose (SeaPlaque from Marine Colloids, Rockland, Me.). Gel slices containing these segments were melted, combined, and then incubated with T4 DNA ligase at 15°C; the ligation reaction occurred in the resolidified gel. After the reaction, the gel slice was remelted and the ligated products were introduced into *E.coli*. Details for this procedure can be found in (15).

For DNA sequencing, DNA segments produced by restriction endonuclease cleavage of pUC8-Sc2676, which contains the 1.7 kb *Bam*H1 DNA fragment including the *his3* gene (8) cloned into the pUC8 vector (17), were subcloned into the single stranded vectors mp18 and mp19 (14). In addition, 2.1 kb *Eco*RI-*Bam*H1 segment, Sc2732, that is adjacent to Sc2676 in the genome (8) was also subcloned. The subcloned segments are shown diagrammatically in Fig. 1.

The *pet56* mutation, YIp5-Sc3317, was constructed by a triple ligation involving *Eco*RI and *Xho*I cleaved YIp5-Sc3302 (18), the 173 bp *Pst*I-*Xho*I fragment of pUC8-Sc2676 (nucleotides +707 to +880), and the 2.5 kb *cyh2* segment generated by complete *Pst*I and partial *Eco*RI digestion of pBR322-*CYH2* (15,19). The *ded1* deletion, YIp5-Sc3390, was generated by cleavage of YIp5-Sc2812 with *Bgl*II and subsequent circularization.

Replacement of *his3* chromosomal sequences by mutant DNAs

The DNA molecules described above were introduced into KY463 (a *ura3*-52 *cyh2*$^r$ *ade2*-101 *trp1*-Δ1 *lys2*-801 *HIS3*$^+$ ) cells by a modification of the method described by Ito *et. al.* (20). An exponential culture (80 ml.) of yeast cells grown overnight in YPD broth to the exponential phase ($A_{600}$ = 1.0) was washed twice in TE (10mM Tris pH 7.5, 1mM EDTA) and resuspended in 1 ml of TE containing 0.1M LiCl for 20 minutes at 30°C. For each transformation reaction, 0.1-1μg of the transforming DNA and 15μg of sonicated, salmon sperm DNA were added to 0.1ml of cells. The transforming DNA was cleaved with *Xba*I (for YIp5-Sc3317) or with *Hpa*I (for YIp5-Sc3390) in order to increase the integration frequency and to direct the integration to the *his3* locus (21). The transforming DNA was purified by phenol extraction following electrophoresis in low gelling agarose; this removed transformation-inhibiting materials present in the rapid plasmid DNA preparations. After incubation for 10 minutes at 30°C, 1 ml of polyethylene glycol 4000 (Fisher) was added; after an additional 20 minutes, the cells were heated to 42°C for 5 minutes, subjected to centrifugation, washed twice with water, and spread on 2 or 3 plates containing medium lacking uracil. The transformation events were analyzed by hybridizing $^{32}$P-labelled Sc2676 DNA to electrophoretically separated, *Bam*H1 cleaved, genomic DNA (15,22). Hybridization to the wild type *his3* locus reveals a 1.7 kb fragment, whereas integration of Sc3317 produces an additional 3.3 kb fragment and integration of Sc3390 produces an additional 2.7 kb fragment.

To replace the wild type *his3* region by Sc3317 DNA, the Ura$^+$ transformants were mated to KY462 (α *ura3*-52 *cyh2*$^r$ *ade2*-101 *trp1*-Δ1 *LYS2*$^+$ *his3*-Δ200), and Ura$^-$ His$^-$ segregants of the resulting diploids were identified after growth in non-selective medium. These strains were sensitive to 40μM cycloheximide because drug sensitivity (encoded by the wild type *cyh2* gene present in the Sc3317 DNA) is dominant to resistance (encoded by the genomic *cyh2*$^r$ mutation)(17). The gene replacement event was confirmed by hybridization as described above. To replace the *his3* region by Sc3390 DNA, Ura$^+$ transformants were mated to KY115 (α *ura3*-52 *ade2*-101 *trp1*-Δ1 *lys2*-801 *HIS3*$^+$), Ura$^-$ segregants were selected, and the desired gene event was identified by genomic hybridization.

RNA analysis

Total RNA was isolated from exponential cultures of KY114 as described previously (7,13). 50μg of RNA was mixed with approximately 50,000 cpm of each hybridization probe (roughly 0.5 ng) in 30μl of 1M NaCl, 50mM HEPES pH 7.5. From previous estimates (7), this amount of RNA corresponds to about 10-50 pg RNA that is complementary to each hybridization

probe. The hybridization mixture was incubated for 16-24 hours at 75°C, and it was occasionally spun in a microcentrifuge to prevent excessive evaporation to the top of the tube. After hybridization;, 270 µl of buffer (60 mM sodium acetate pH 4.5, 250 mM NaCl, 1 mM zinc acetate) containing 100 units of S1 nuclease (Sigma) was added. After incubation for 30-60 minutes at 37°C, the reaction was terminated by the addition of 3 µl 0.5M EDTA, 1 µl *E.coli* tRNA (10 mg/ml), and 0.7 ml ethanol. The nucleic acid was precipitated, washed once with 95% ethanol, and resuspended in 10 µl of 0.1M NaOH, 1 mM EDTA. Half of this sample was subjected to electrophoresis in a thin 6% acrylamide gel containing 7M urea (16).

*His3* probe B and the *pet56* probe were made by labelling the 5'-ends of the appropriate DNA fragments with T4 polynucleotide kinase and separating the DNA strands on a 4% native polyacrylamide gel (23). The *his3* probe was derived from the 322 bp *Msp*I-*Dde*I fragment (nucleotides +242 to -80), and the *pet56* probe was derived from the 358 bp *Bam*H1- *Dde*I fragment (nucleotides -442 to -84). *His3* probe A and the *ded1* probe were synthesized by a primer extension method similar to that employed for DNA sequencing (16). Synthetic oligonucleotide primers (100 ng) were labelled at their 5'-ends with T4 polynucleoltide kinase using a molar excess of $^{32}$P γ-ATP (7000 Ci/mMole) and then hybridized to 20µg of appropriate M13 hybrid DNAs. The *his3* 22-mer corresponding to nucleotides +174 to +153 was annealed to mp8-Sc3292, and the *ded1* 23-mer corresponding to nucleotides +1232 to +1210 was annealed to mp9-Sc2677 (see Fig.1). The hybridized primers were extended with *E.coli* DNA polymerase I (large fragment) in the presence of all four deoxy nucleotide triphosphates (0.2 mM each), cleaved with the appropriate restriction endonuclease (*Eco*RI for *his3* and *Xho*I for *ded1*, and the resulting products were separated in 1.7% low gelling agarose, 30 mM NaOH, 1 mM EDTA. The gel slices containing the desired probes were melted at 73°C, extracted with phenol, and precipitated with ethanol after the addition of 100 µg of *E.coli* tRNA.

## RESULTS AND DISCUSSION

### Nucleotide sequence

The restriction endonuclease cleavage map of the 1.7 kb *Bam*H1 DNA fragment containing the entire *his3* gene (Sc2676) has been determined previously (8) and is shown in Fig. 1. A number of DNA fragments produced by restriction endonuclease cleavage were subcloned in the single-stranded vectors mp18 and mp19 (14), and their nucleotide sequences determined as described by Sanger *et. al.* (16). At each position, the nucleotides of both strands have been determined; in many cases, particular regions of a given strand have been analyzed on molecules containing different subcloned DNA fragments. The DNA sequence of this region is shown in Fig. 2, and the positions of the restriction endonuclease cleavage sites are listed in Table 1. The nucleotide coordinates are defined with respect to the the site of *his3* transcriptional initiation (defined as +1)(9).

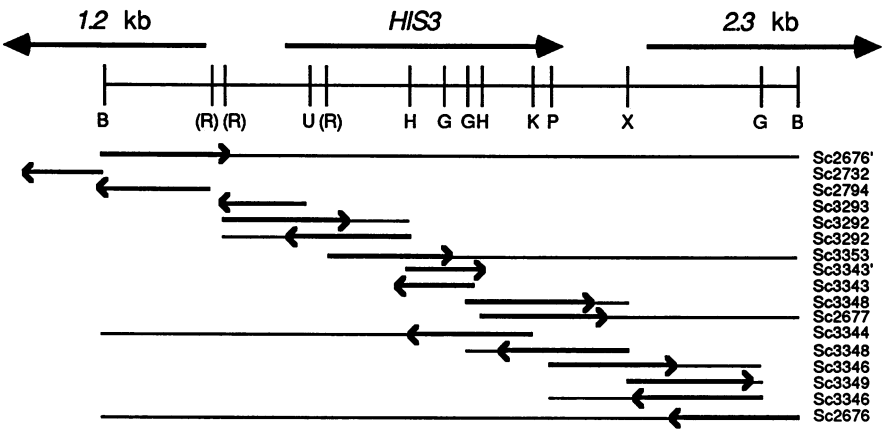



Figure 1: Structure of Sc2676, a 1.7 kb *Bam*HI generated DNA fragment containing the *his3* gene. Restriction sites for the following enzymes are shown as vertical lines: B-*Bam*HI, (R)-*Eco*RI sites introduced artificially with *Eco*RI linkers and do not occur in the natural DNA fragment, U-*Sau*3A (only the site at +82 is indicated), H-*Hin*dIII, G-*Bgl*II, K-*Kpn*I, P-*Pst*I, X-*Xho*I. The orientations and approximate locations of the *his3*, 1.2 kb, and 2.3 kb transcripts (7) are indicated by arrows above the restriction map. The extent of DNA segments from Sc2676 subcloned into mp18 and mp19 are shown below the map as thin, horizontal lines. The thick arrows indicate the direction and length of DNA sequences determined by the di-deoxy chain termination method (16). Segments with leftward arrows were cloned into mp18; segments with rightward arrows were cloned into mp19. The entire figure is drawn to scale.

The *his3* structural gene

Previously, the DNA sequence of the *his3* promoter region and the beginning of the structural gene had been determined (9). In particular, this nucleotide sequence clearly implicates one particular AUG codon as the site of translational initiation. This codon is the 5'-proximal AUG in the *his3* mRNA and there are no other AUG codons between nucleotides +1 and +100, a region known to be critical for *his3* function (8).

The complete nucleotide sequence shown in Fig. 2 indicates that the open reading frame originally proposed to encode the *his3* gene extends for 219 codons and ends with 2 termination codons of different specificity (TAG,TGA). The C-terminal boundary of this open reading frame (nucleotide +680) is in accord with that expected for the *his3* structural gene. A DNA fragment extending to the *Pst*I site (nucleotide +707) supports *his3* expression in *E.coli* (8) and yeast cells (9), whereas a segment extending only to the *Kpn*I site (nucleotide +646) does not. These results indicate that the 219 amino acid protein (with a calculated molecular weight of 23,850) represents the primary translation product of yeast imidazoleglycerolphosphate dehydratase. The amino acid composition of this protein is typical for yeast proteins in general except that histidine is approximately 5-10 fold overrepresented. Codon usage for the *his3* gene is not biased in the manner described for highly expressed yeast genes. The codon bias index (24) for *his3* is 0.01,

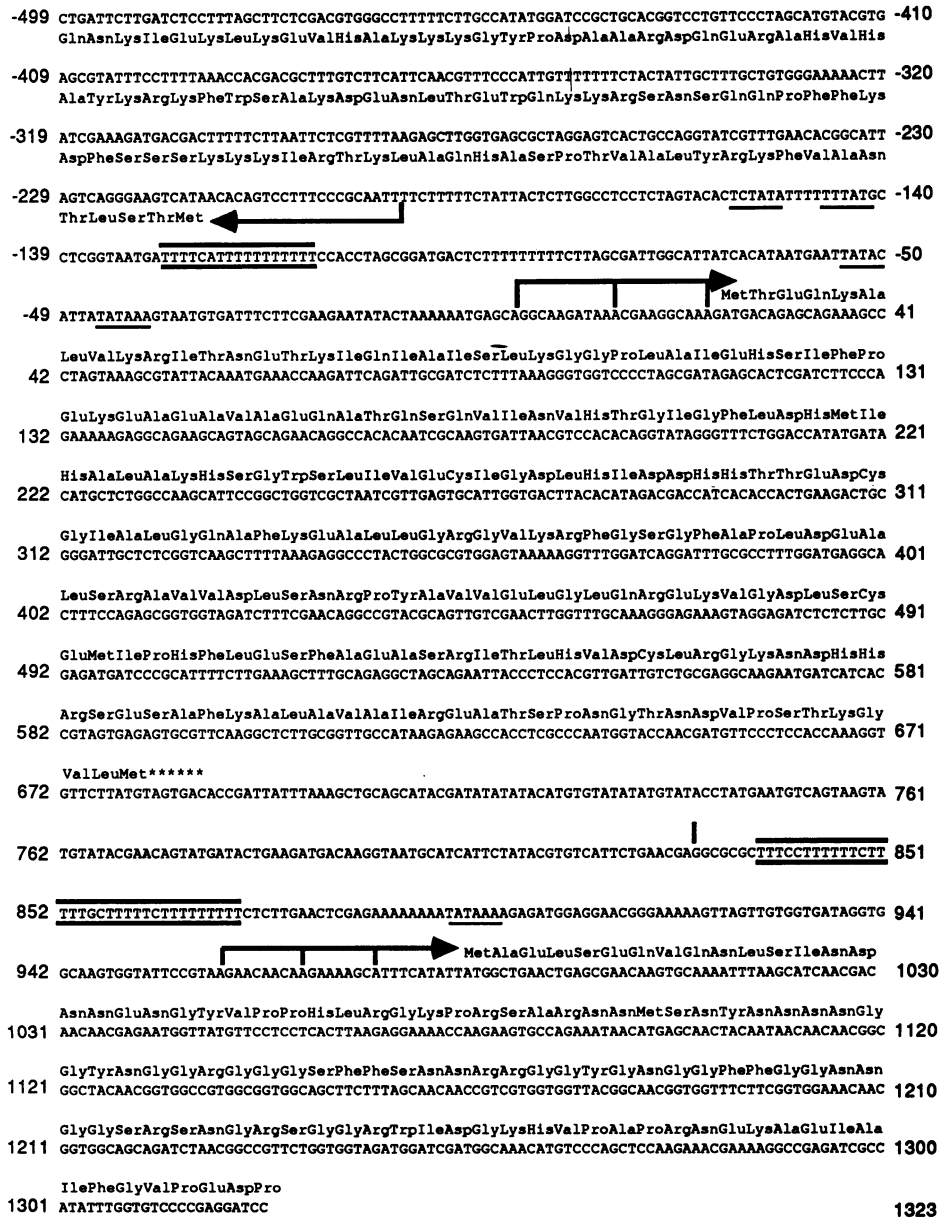



Figure 2: Nucleotide sequence of the *pet56-his3-ded1* gene region. Only the *his3* coding strand is shown. Nucleotide +1 is defined by the upstream-most *his3* mRNA start site. The deduced *his3* and *ded1* protein sequences are shown above the DNA sequence, whereas the deduced *pet56* protein sequence (which is encoded on the opposite strand) is shown below. mRNA initiation sites are diagrammed as vertical lines connected to horizontal arrows which indicate the

direction of transcription. The major *his3* mRNA termination site is indicated by a vertical line at position +830. Tracts of $(dA:dT)_n$ located between the genes are indicated as thick horizontal lines above and below the DNA sequence, and TATA sequences are shown as thin horizontal lines below the sequence. The sequence between -160 and +120 was determined previously (9), and two errors are corrected here. More accurate mRNA mapping indicates that the +1 position is 3 bp further downstream than reported previously. In addition, the published sequence indicates that the *his3* 5'-untranslated mRNA leader contains 2.4 copies of a 15 bp sequence. However, one of these copies (i.e. 15 bp) was inserted accidentally during the preparation of the original figure and does not exist.

which indicates that codon selection is essentially random and that *his3* expression is likely to be relatively poor.

Construction and analysis of mutations in the genes that are adjacent to *his3*

To determine the functions of the adjacent genes, mutated derivatives of YIp5-Sc2812 DNA that delete *his3* structural sequences as well as the sequences for one of the two adjacent genes were generated as described in the materials and methods and illustrated in Fig. 3A,B. YIp5 is an "integration vector" that contains the $ura3^+$ gene as a selectable marker (22), and Sc2812 is a 6.1 kb DNA fragment with *his3* centrally located (17; Fig.3). Because of the possibility that loss of either function encoded by the adjacent genes might be incompatible with life, the mutations were generated in diploid yeast strains that contained the wild type genes on the homologous chromosome.

Table 1. Restriction endonuclease cleavage sites.

| Enzyme | Sites | Enzyme | Sites | Enzyme | Sites |
|---|---|---|---|---|---|
| *Acc*I | 738,764 | *Aha*III | -395,91,335,698 | *Alu*I | -279,330,517,702, |
| *Asu*II | -25,426 | *Ava*I | 880,1313 | | 1149,1271 |
| *Ava*II | -434,99,209 | *Bal*I | 231 | *Bam*HI | -447,1318 |
| *Ban*I | 642 | *Bbv*I | -442,701,704, | *Bcl*I | 572 |
| *Bgl*I | 348 | | 1145,1214 | *Bgl*II | 419,479,1220 |
| *Bsp*1286 | 118 | *Bss*HII | 832 | *Bst*NI | -253 |
| *Bst*XI | 413 | *Cla*I | 1251 | *Dde*I | -84,995 |
| *Fnu*4HI | -440,703,706, | *Fnu*DII | 834 | *Fok*I | -103,391 |
| | 1120,1147,1216 | *Hae*II | -267 | *Hae*III | -465,-172,162,231, |
| *Hga*I | -388 | *Hgi*AI | 118 | | 342,434,1134,1230,1289 |
| *Hha*I | -268,353,386, | *Hin*dIII | 328,515 | *Hin* fI | -497,-263,-98,71, |
| | 834,836 | *Hph*I | -276,269,577,930 | *Kpn*I | 646 |
| *Mbo*II | -379,-29,-24,123, | *Msp*I | 241 | *Nde*I | -451,214 |
| | 302,783,1193 | *Nla*III | -416,225,727, | *Nla*IV | -445,101,644,1320 |
| *Nsi*I | 803 | | 1096,1265 | *Pst*I | 707 |
| *Rsa*I | -415,-161,438,644 | *Sau*3A | -491, -447,82,121, | *Sau*96I | -467,-434,99, |
| *Scr*FI | -253 | | 372,419,479,496,572, | | 209,341 |
| *Sfa*NI | 800,1019 | | 1220,1248,1293,1318 | *Taq*I | -474,-318,-25,120, |
| *Taq*II | 321 | *Xho*I | 880 | | 426,448,881,1251 |
| *Xho*II | -447,419,479,1220,1318 | | | *Xma*III | 1228 |

Coordinates are defined by the position in the top strand (see Fig. 2) just before the cleavage point. Restriction sites are not found for *Aat*I, *Aha*II, *Apa*I, *Ban*II, *Bst*EII, *Eco*RI, *Eco*RV, *Hin*cII, *Hpa*I, *Mlu*II, *Mst*I, *Mst*II, *Nae*I, *Nar*I, *Nci*I, *Nco*I, *Not*I, *Nru*I, *Pvu*I, *Pvu*II, *Sac*I, *Sac*II, *Sal*I, *Sca*I, *Sfi*I, *Sma*I, *Sna*BI, *Sph*I, *Stu*I, *Tth*111I, *Xba*I, *Xmn*I.

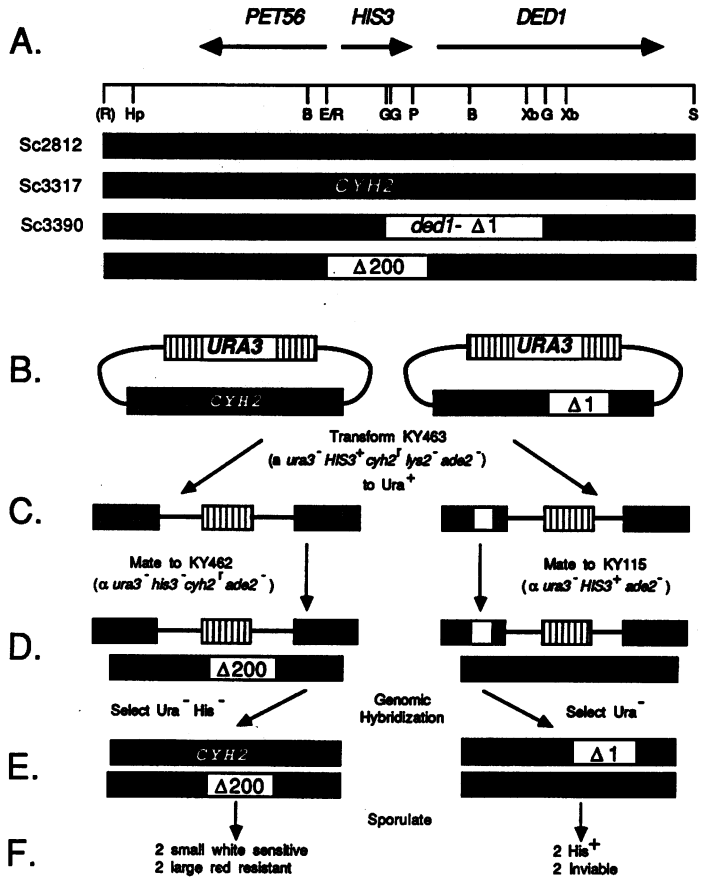



Figure 3: Structure and phenotypic analysis of *pet56* and *ded1* mutations. A) Restriction map of Sc2812, a 6.1 kb DNA segment containing the entire *pet56-his3-ded1* gene region. Additional abbreviations for restriction endonuclease sites not shown in Fig. 1 are as follows: (R)-mutated *Eco*RI site, Hp-*Hpa*I, E/R-*Eco*RI site produced by insertion of an *Eco*RI linker at the -172 *Hae*III site (18), Xb-*Xba*I, S-*Sal*I. The locations and orientations of the mRNA transcripts are shown above the restriction map, and the structure of the relevant mutant DNAs are shown below. Normal chromosomal sequences are indicated by gray bars, deleted sequences are indicated by open bars, and *cyh2* sequences are indicated by black bars. B) Structures of the transforming DNA molecules. C) "Duplication structures" after integration at the *his3* locus. D) Chromosomal structures of diploid strains obtained after mating. E) Chromosomal structures after "gene replacement". F) Phenotypes of spores obtained by tetrad dissection.

The phenotypes *in vivo* conferred by these mutations were determined as shown in Fig. 3. First, a single copy of each mutant DNA was integrated at the *his3* chromosomal location of yeast strain KY463 by selecting for Ura$^+$ transformants. This produces a "duplication" structure containing the wild type and the mutated versions of the 6.1 kb *his3* region which are separated by the YIp5 vector sequences (Fig. 3C). Second, the resulting transformants were mated to

appropriate α strains to create diploids containing a single copy of the duplication structure (Fig. 3D). Third, after growth of the diploid strains in non-selective medium, Ura⁻ segregants were identified by replica plating (Fig. 3E). Segregants due to gene replacement of the wild type *his3* locus by the mutated derivative constructed *in vitro* were identified by genomic hybridization methods (data not shown). Fourth, the phenotypes conferred by these mutations were determined following tetrad dissection of the appropriate diploids.

The 1.2 kb RNA species is essential for mitochondrial function

The adjacent gene located upstream of *his3* was mutated by replacing the region between the *Hae*III site at -172 and the *Pst*I site (+707) by a 2.5 kb DNA fragment (Sc3316) containing the wild type *cyh2* gene. The resulting molecule, YIp5-Sc3317, lacks the entire *his3* structural gene and, as will become apparent later, it also removes most of the promoter region for the gene encoding the 1.2 kb transcript. The wild type chromosomal locus was replaced by Sc3317 DNA as described above.

After sporulation of normal diploid strains, greater than 90% of the asci give rise to 4 colonies that with occasional exceptions are roughly equal in size. However, each of 23 tetrads dissected from a diploid strain containing Sc3317 DNA produced 2 normal sized colonies and 2 small colonies. These colony sizes were maintained upon further restreaking. Addition of cycloheximide (to 40μM) severely inhibited the growth of all the small colonies but did not affect the growth of the normal sized colonies. This demonstrates that the small colony phenotype is due to Sc3317 DNA, because this is the only source of the wild type (drug sensitive) *cyh2* gene.

The small colony phenotype is similar to that of petite (*pet*) mutants which are defective in mitochondrial function. Two additional observations indicate this to be the case. First, although every colony was genotypically *ade2*⁻, all the small colonies were white, whereas all the large colonies were red. Normally, *ade2* mutants accumulate an intermediate in adenine biosynthesis which is oxidized and excreted as a red pigment (25). *Pet* mutants, however, remain white because they are unable to oxidize this intermediate (26). Second, all the small colonies were unable to grow in medium containing 2% glycerol as the sole source of carbon, conditions that require mitochondrial function (26).

Thus replacement of the normal chromosomal region by Sc3317 DNA results in strains that are defective in mitochondrial function. This effect can not be due to deletion of the *his3* gene, because the *his3*-Δ200 allele removes the entire *his3* structural gene yet does not cause any of the phenotypes described above. Instead, the petite phenotype must be due to effects on the 1.2 kb transcript; indeed, the small colonies do not produce detectable quantities of this RNA species (unpublished results). Therefore, these results demonstrate that the 1.2 kb RNA encodes a gene necessary for mitochondrial function, which is now called *pet56*.

The 2.3 kb RNA species is essential for cell viability

The adjacent gene located downstream of *his3* was mutated by deleting 1.8 kb of DNA between the *Bgl*II sites of YIp5-Sc2812 (Fig. 3). The resulting molecule, YIp5-Sc3390, deletes

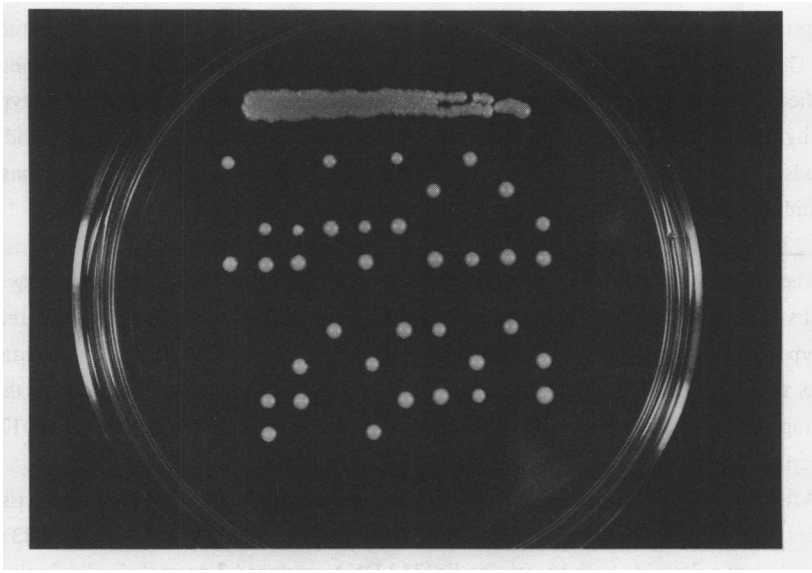


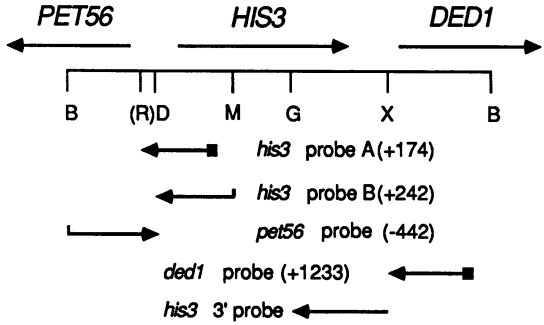
Figure 4: Tetrad analysis of *ded1*-Δ1/+ diploid strain. The four spores of individual tetrads are arranged vertically.

about half of the *his3* structural gene as well as the promoter region and about 1000 bp of the 2.3 kb mRNA coding sequences. The wild type chromosomal locus was replaced by Sc3390 DNA as described above, and the resulting strain was analyzed by tetrad dissection.

In 17 out of 19 tetrads, only two of the four spores were viable, and in the remaining 2 tetrads, only one spore was viable (Fig. 4). This segretation pattern indicates that the diploid strain contains a single recessive mutation that is lethal in haploid cells. Moreover, this mutation is due to Sc3390 DNA because none of the viable spores were $His^-$, the expected phenotype of Sc3390. Thus replacement of the normal chromosomal region by Sc3390 DNA results in strains that are inviable. This demonstrates that the 2.3 kb RNA encodes a gene necessary for cell viability, which is now called *ded1* (defines essential domain).

Location of mRNA and protein coding sequences

In order to map the 5'-termini of the *pet56*, *his3*, and *ded1* RNA species, single stranded DNA hybridization probes labelled with $^{32}P$ at their 5'-ends were prepared for each gene (see materials and methods; Fig. 5). In order to map the 3'-termini of *his3* RNAs, a single stranded probe extending from +884 to +480 and uniformly labelled with $^{32}P$ was employed. Each hybridization probe was incubated with total RNA from KY114 (wild type for the *his3* gene region), and the reaction products were treated with S1 nuclease and analyzed by gel electrophoresis (Fig. 5). Each distinct 5' or 3' end was mapped by the position of a major radioactive band. Fainter bands, observed within ±2 nucleotides of a major band were not



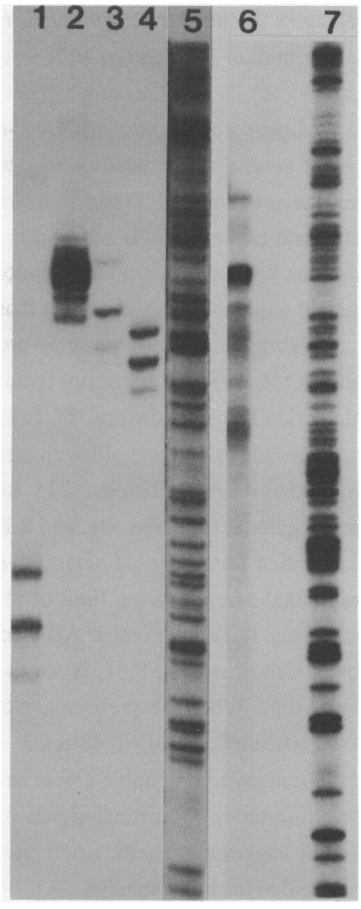




Figure 5: mRNA mapping. The hybridization probes are diagrammed below the restriction map (additional abbreviations are D-*Dde*I, M-*Msp*I). *His3* probe A and the *ded1* probe were generated by extention of a 5'-end labelled oligonucleotide primer (black square) in the direction indicated. *His3* probe Band the *pet56* probe were prepared by separating the DNA strands of 5'-end labelled restriction fragments. The vertical line indicates the position of the label, and the direction of the arrow indicates the 5'-3' polarity of the strand. The *his3* 3'-probe was uniformly labelled by extention of the M13 DNA sequencing primer; the arrow indicates the 5'-3' polarity. The autoradiogram represents analysis of total RNA from wild type strain KY114 with the following hybridization probes: *his3* probe A (lane 1), *ded1* (lane 2), *pet56* (lane 3), *his3* probe B (lane 4), and *his3* 3'-probe (lane 6). Lanes 5 and 7 are size standards produced by the chain termination method (16). The products shown in lanes 1-5 were separated on a different gel from those shown in lanes 6-7.

counted; presumably, these represent incomplete S1 digestion and/or "nibbling" of RNA:DNA duplexes. The locations of these mRNA initiation and termination sites with respect to the DNA sequence are indicated in Fig. 2.

Analysis of *his3* transcripts with probe A reveals two major bands corresponding to 174 and 163 bases in length and a minor band corresponding to 153 bases (Fig. 5, lane 1). The bands obtained with probe B correspond respectively to lengths of 242, 231, and 221 bases (lane 4). The 5' end points of the two major *his3* RNA species are defined as positions +1 and +12, and that of the minor RNA species defined as +22. Since the *his3* translational initiation codon maps between positions +24 and +26 (Fig. 2), these results indicate that the *his3* RNA species contain very short non-translated leader sequences. Presumably, the +1 transcript (leader of 23 bases) is translated into functional *his3* enzyme because the open reading frame begins with the 5'-proximal AUG codon (27). The +12 transcript (leader of 12 bases) is translatable because

galactose induction of *gal-his3* fusions, which is associated almost exclusively with transcriptional initiation at +12, results in the expected increase in IGP dehydratase enzyme activity (18). It is unknown whether the +22 transcript (leader of 2 bases) is translatable.

Mapping of the *his3* mRNA 3'-ends results in a major band that is 280 bases in length as well as several minor bands whose lengths range from 230-280 bases (lane 6). This indicates that the predominant 3'-end is located at position +830, whereas the less abundant species terminate between +780 and +830. These positions are roughly 100-150 bases downstream from the *his3* translational termination codons. The *his3* termination region does not appear to contain sequences homologous to those proposed to be involved in 3'-end formation (24,28).

Hybridization with the *pet56* probe reveals one predominant band 250 bases in length that defines the *pet56* +1 transcript (lane 3). The 5' end point corresponds to position -192 of the nucleotide sequence in Fig. 2. Faint bands are also observed at coordinates -178 and -210 (corresponds to +14 and -18 with respect to the *pet56* +1 site). The 5'-proximal AUG encoded by *pet56* RNA (positions -215 to -217) is located 23-25 bases downstream from the transcriptional initiation site, and it is extremely likely to be the translational initiation codon. Conceptual translation of *pet56* RNA initiating at this AUG codon produces an open reading frame that extends to the limit of the DNA sequence information (95 amino acids), whereas translation initiating at other AUG codons terminates within this region. The codon bias index for this short region, 0.01, is similar to that for *his3* and it is suggestive of relatively poor expression. When this presumptive coding region is compared to proteins in several data bases, no significant homology is detected. However, it is worth noting that the N-terminal 34 residues contain many hydrophobic amino acids, several basic amino acids, and no acidic amino acids. This resembles N-terminal signal sequences that specify transport to the mitochondria (29) thereby suggesting that the *pet56* protein, which is necessary for mitochondrial function, may be localized to the mitochondria.

Hybridization with the *ded1* probe reveals one very intense band (270 bases in length) as well as two other bands (260 and 250 bases in length). These end points map at nucleotides +970, +980, and +990, and they correspond to *ded1* transcripts initiated at +1, +10, and +18 respectively. The increased intensity is probably due to the fact that the overall *ded1* transcription levels are roughly five-fold higher than the *his3* or *pet56* RNA levels (7). The 5'-proximal AUG (nucleotides +986 to +988) is located very near the terminus of *ded1* RNA. It is extremely likely to be the translational initiation codon because it defines an open reading frame that extends for at least 113 amino acids. The codon bias index for *ded1* is 0.3, which suggests that its expression level is higher than *his3* or *pet56* although not as high as genes such as alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, and enolase (24); this correlates well with the respective mRNA levels. The N-terminal portion of the presumptive *ded1* protein is unusual in that it is very rich in asparagine (19 out of the first 81 amino acids) and glycine (21 out of 86). A computer search reveals statistically significant homology to keratin proteins from a number of organisms. Perhaps the *ded1* gene encodes a protein that is part of the yeast cytoskeleton.

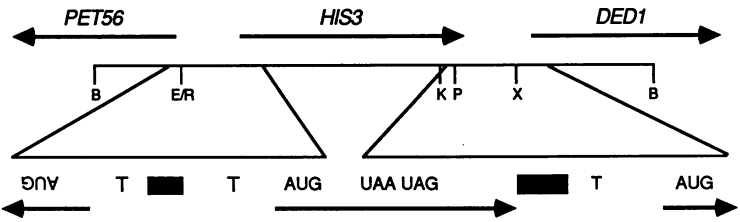



Figure 6: Organization of the *pet56-his3-ded1* gene region. The top part of the figure shows the positions of the mRNA transcripts drawn to scale with respect to the restriction map. Expanded views of this region including the AUG translational initiation codons, the UAA UAG translational termination codons for the *his3* protein, the $(dA:dT)_n$ spacer sequences (black boxes) and TATA sequences (T) are shown below (not to scale).

Organization of the *pet56-his3-ded1* region

A schematic view of the *pet56-his3-ded1* region is shown in Fig. 6. In most ways, these genes are typical for yeast. All of them encode RNAs whose 5' untranslated leaders are short (20-30 bp) and are rich in A residues. As expected (27), the 5'-proximal AUG codons appear to serve as translational initiation sites . The DNA sequences around the mRNA initiation sites are not particularly conserved, but like many yeast genes are rich in purine (especially A) residues. All the genes contain TATA-like sequences located 40-90 bp upstream of their respective RNA start sites. The *his3* TATA sequences are known to be critical for transcription (10,11), and unpublished results from this laboratory are suggestive of similar roles for the *pet56* and *ded1* TATA sequences. The regions between the genes contain relatively large tracts of poly dA:dT. For example, the *his3-pet56* intergenic region contains 9 and 11 bp tracts, and the *his3-ded1* region contains 8 and 9 bp tracts. Assuming that the yeast genome contains approximately $10^7$ bp with an average AT composition of 60%, an 8 bp stretch of poly dA:dT should occur at a frequency of $10^{-4}$ or once per 10 average genes. However, a computer analysis of all yeast DNA sequences in GEN.BANK indicates that 80% of them contain such poly dA:dT regions, often more than once per sequence. Moreover, with a single exception, these regions are always found in non-coding sequences. Thus, such homopolymer tracts are diagnostic of intergenic regions in yeast.

The *his3-pet56* intergenic region contains 6 copies of a short sequence, TGACTC. Two perfect copies are found at -99 to -94 and -263 and -258 (opposite strand), and four copies with 5 out of 6 matches are found at -142 to -137, -181 to -176, -216 to -221 (opposite strand), and -225 to -230 (opposite strand). This sequence is repeated in the 5'-flanking regions of genes whose expression is induced coordinately in response to conditions of amino acid starvation (12,30,31). Moreover, deletion analyses of the *his3* (12) and *his4* (32) promoter regions indicate that this sequence is critical for proper regulation *in vivo*. However although these sequences lie upstream of both the *his3* and *pet56* TATA elements, only *his3* expression is regulated (7).

In general, functional yeast promoters require DNA sequences located more than 100 bp upstream from the mRNA initiation sites (reviewed in 33). The most striking feature of the *pet56-his3-ded1* region is that even though the genes are unrelated, they are extremely close together. The *pet56* and *his3* genes are transcribed divergently from initiation sites that are separated by only 192 bp. Thus, the *his3* and *pet56* promoter regions almost certainly overlap even though the expression of these genes is not regulated in a common manner (7). In fact, it appears that a poly dA:dT stretch between these two genes is important for the constitutive transcription of both (unpublished results). Similarly, it is very possible that the *ded1* promoter region overlaps the *his3* termination region because transcription of the *ded1* gene is initiated only 133 bp beyond the 3'-end of the *his3* mRNA coding region. In particular, *his3* transcription terminates just before an extended poly dA:dT region (28 out of 34 dT residues in the coding strand). This region, which is only about 20 bp upstream of the *ded1* TATA sequence, appears to be essential for wild type levels of *ded1* transcription (unpublished results). These results suggest the intergenic regions between unrelated genes are not composed simply of inert spacer sequences, but rather consist largely of promoter and terminator sequences. In addition, the observations that unrelated yeast genes are closely packed and that poly dA:dT tracts are often found in 5' and 3' non-coding regions suggests that the genomic organization of the *pet56-his3-ded1* region will be typical of yeast chromosomal regions in general.

ACKNOWLEDGEMENTS

I thank David Hill for performing the 3' mapping experiment and Marjorie Oettinger for conducting the computer search and for suggesting the name *ded1*. This work was supported by grants from the National Institute of Health (GM30186) and from the Chicago Community Trust (Searle Scholars Program).

REFERENCES

1. Hereford, L.M. and Rosbash, M. (1977) *Cell* **10,** 453-462.
2. Kaback, D.B., Angerer, L.M. and Davidson, N. (1979) *Nucl. Acid Res.* **6,** 2499-2517.
3. Mortimer, R.K. and Schild, D. (1980) *Microbiol. Rev.* **44,** 519-571.
4. Garrick-Silversmith, L. and Hartmann, P.E. (1970) *J. Bacteriol.* **66,** 231-244.
5. Fink, G.R. (1964) *Science* **146,** 525-527.
6. Struhl, K., Cameron, J.R. and Davis, R.W. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73,** 1471-1475.
7. Struhl, K. and Davis, R.W. (1981) *J. Mol. Biol.* **152,** 535-552.
8. Struhl, K. and Davis, R.W. (1980) *J. Mol. Biol.* **136,** 309-332.
9. Struhl, K. and Davis, R.W. (1981) *J. Mol. Biol.* **152,** 553-568.
10. Struhl, K. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78,** 4461-4465.
11. Struhl, K. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79,** 7385-7389.
12. Struhl, K. (1982) *Nature* **300,** 284-287.
13. Oettinger, M.A. and Struhl, K. (1985) *Mol. Cell. Biol.* **5,** 1901-1909.
14. Messing, J. (1983) *Methods in Enzymology* **101,** 20-79.
15. Struhl, K. (1983) *Gene* **26,** 231-241.
16. Sanger, F., Coulson, A.R., Barell, B.G., Smith, A.J. and Roe, B.A. (1980) *J. Mol. Biol.* **143,** 161-178.
17. Viera, J. and Messing, J. (1982) *Gene* **19,** 259-268.

18. Struhl, K. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81,** 7865-7869.
19. Fried, H. M and Warner, J.R. (1982) *Nucl. Acids Res.* **10,** 3133-3148.
20. Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) *J. Bacteriol.* **153,** 163-168.
21. Orr-Weaver, T., Rothstein, R.J. and Szostak, J.W. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78,** 6354-6358.
22. Struhl, K., Stinchcomb, D.T., Scherer, S. and Davis, R.W. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76,** 1035-1039.
23. Hayward, G.S. (1972). *Virology* **49,** 342-344.
24. Bennetzen, J. L. and Hall, B.D. (1982) *J. Biol. Chem.* **257,** 3026-3031.
25. Roman, H. (1956) *Cold Spring Harbor Symposium on Quantitative Biology* **21,** 175-185.
26. Sherman, F. and Slonimski, P.P. (1964) *Biochim. Biophys. Acta* **90,** 1-15.
27. Sherman, F., Stewart, J.W., and Schweingruber, A.M. (1980) *Cell* **20,** 215-222.
28. Zaret, K.S. and Sherman, F. (1982) *Cell* **28,** 563-573.
29. Horwich, A.L., Kalousek, F., and Rosenberg, L.E. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82,** 4930-4934.
30. Zalkin, H. and Yanofsky, C. (1982) *J. Biol. Chem.* **257,** 1491-1500.
31. Hinnebusch, A.G. and Fink, G.R. (1983) *J. Biol. Chem.* **258,** 5238-5247.
32. Donohue, T.F., Daves, R.S., Lucchini, G., and Fink, G.R. (1983) *Cell* **32,** 89-98.
33. Struhl, K. (1985) In From gene to protein: Steps dictating the maximal level of gene expression, W. Reznikoff and L. Gold, eds.(California: Benjamin/Cummings Publishing Co.), in press.