Point Mutations Implicate Repeated Sequences as Essential Elements of the *CYC7* Negative Upstream Site in *Saccharomyces cerevisiae*

CYNTHIA F. WRIGHT† AND RICHARD S. ZITOMER*

Department of Biological Sciences, State University of New York at Albany, Albany, New York 12222

The transcription of the *CYC7* gene of *Saccharomyces cerevisiae*, encoding the iso-2-cytochrome *c* protein, is controlled by two upstream regulatory elements, a positive element and a negative element. The nature of the DNA sequences in the negative element were investigated in a two-part approach. The first involved the construction of a *CYC7-galK* fusion gene which placed the coding sequence of the *Escherichia coli* galactokinase gene under the regulation of the *CYC7* upstream sequences. This fusion allowed the quantitation by galactokinase enzyme assays of the effects on gene expression of a variety of previously isolated deletion mutations within the negative site. The results suggested that the negative site contained three related sequences. This hypothesis was tested in the second part of these studies, the selection of point mutations within the region of the negative site which led to increased *CYC7* expression. Point mutations were introduced by a technique which induced mutations within a localized region at high efficiency. All but one of the mutations involved more than a single base-pair change. The mutations followed the pattern that multiple base-pair changes occurred in one repeat or single base-pair changes occurred in two repeats, with the exception of one mutant, which had a single base-pair change in one repeat. This pattern of mutations and the base pairs that were altered strongly supported the hypothesis that the repeats are integral elements of the negative site.

Transcriptional regulation of genes in *Saccharomyces cerevisiae* is mediated through sequences located one hundred to several hundred base pairs 5' to the coding sequence (6, 10–12, 15–18, 29, 31, 32, 34; J. L. Weiss, C. V. Lowry, and R. S. Zitomer, submitted for publication). In all cases reported, there is a sequence which stimulates transcription and is often the site of positive regulation or activation; such sites have been designated upstream activation sites. In addition, in a relatively rare number of cases, there is also a site for repression or negative regulation (32, 34). The *CYC7* gene, encoding the iso-2-cytochrome *c* protein, contains both the positive site and the rarer negative site and, therefore, offers the opportunity to study the nature of both types of sites. Fortunately, the *CYC7* gene also offers an excellent genetic system that can be exploited in such studies.

S. cerevisiae has two unlinked nuclear genes encoding the cytochrome *c* protein; *CYC1* and *CYC7* encode the iso-1 and iso-2 proteins, respectively (7, 28). Although the proteins are functionally indistinguishable, the genes are expressed at different levels and differ in their regulation. Transcription of the *CYC1* gene is induced by oxygen through two upstream activation site sequences (10), and the resulting protein constitutes 95% of the cytochrome *c* in aerobically grown cells. Transcription of the *CYC7* gene, on the other hand, occurs at a low level which is only marginally sensitive to oxygen (Weiss et al., submitted). The low level of expression is apparently due to the antagonistic effects of the positive and negative sites located 5' to the coding sequence, and the low sensitivity to oxygen is hypothesized to be due to the poor oxygen regulatory consensus sequence which the pos-

itive site represents (34; Weiss et al., submitted). A cell lacking a functional *CYC1* gene contains only 5% the wild-type level of cytochrome *c*, all in the form of iso-2 and, although it is capable of growing on some nonfermentable energy sources, it cannot grow on lactic acid. This growth limitation can be exploited for the selection of *CYC7*-overexpressing mutants, and a number of *cis*-acting, *CYC7*-linked mutations have been obtained (4, 5, 8, 19, 21, 33). However, molecular analysis revealed that these were not point mutations but larger DNA rearrangements (8, 19, 21, 33). With the discovery of the negative site through the use of deletions created in vitro, the possibility was raised that point mutations might be extremely difficult to obtain because of the presence of repeated sequences within this region. If the information were redundant, multiple mutations would be required. In this study, localized in vitro mutagenesis of this site was carried out to test this hypothesis and reveal the base pairs important for the functioning of this site.

MATERIALS AND METHODS

Strains and transformations. For the analysis of mutant *CYC7* phenotypes, *S. cerevisiae* GM3C-2 (*MATα trp1-1 leu2-3 leu2-112 his4-519 gal cyc1-1 cyp1-3 [cyc7] osm^r*) (9) was used. For the analysis of galactokinase phenotypes, *S. cerevisiae* aGH1 (*MATa trp1-289 leu2-3 leu2-112 gal1-Δ152*) was used. The *gal1-Δ152* mutation is a deletion constructed by Mark Johnston (Washington University). Yeast transformations were carried out as described previously (14).

Selections for and analyses of *CYC7* mutant phenotypes were carried out on lactic acid and glycerol plates (27). For the preparation of extracts for galactokinase assays, cells were grown to the mid-log phase (2×10^7 cells per ml) in YPR (2% raffinose) or YPD (2% glucose) medium (15) with vigorous shaking or, for anaerobic growth, in YPD with ultrapure nitrogen (99.999% N_2) bubbling through.

---

* Corresponding author.

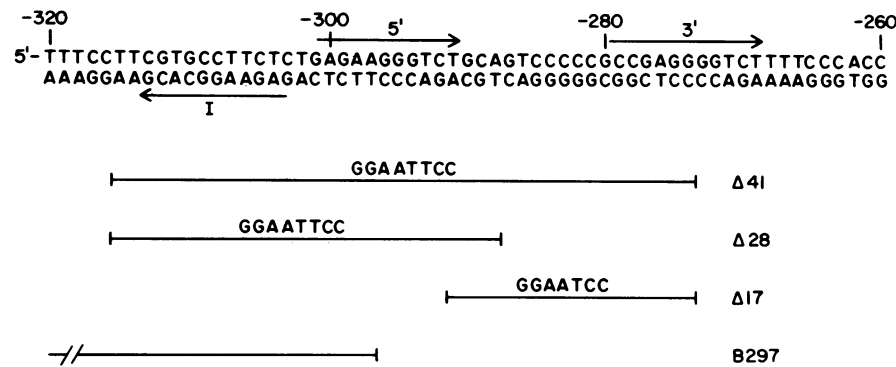† Present address: Laboratory of Biochemistry, National Cancer Institute, Bethesda, MD 20205.

FIG. 1. Deletions of the negative site. The *CYC7* upstream region from −320 to −260 is shown. The deletions in various plasmids are represented by bars. The base pairs above the bars indicate the presence of an *Eco*RI linker in place of the deleted sequence. Deletion B297 contains a *Bam*HI linker and is deleted to −716. The repeated sequences designated 3′, 5′, and I are indicated by arrows, which also indicate the direction of the repeats.

The bacterial strain HB101 (2) was used for bacterial transformations (13), except for the initial transformations with sodium bisulfite-mutagenized plasmids. In these cases, strain BD1528 (23) was used.

Nomenclature. The base pairs in the plasmids are numbered with respect to the *CYC7*-coding sequence. The A in the ATG initiation codon represents base pair 1; bases 3′ are numbered consecutively in positive integers, and bases 5′ are numbered consecutively in negative integers. The sequence for the coding region and a portion of the upstream sequence were determined by Montgomery et al. (22) and subsequently extended (34).

Plasmids. The plasmid YCp*CYC7*(2)r has been described previously (34). It contains pBR322 sequences (1), the yeast *TRP1* selectable marker and *ARS1* autonomous replicating sequence (30), the centromere from chromosome III (*CEN3*) (3), and a 2-kilobase (kb) insert from the *CYC7* locus containing the coding sequence and 1 kb of 3′- and 716 base pairs (bp) of 5′-flanking sequences (22). The deletion plasmids Δ41, Δ28, Δ17, B297, X238, and X347 have been described previously (34). They were derived from YCp*CYC7*(2)r, and their deletions are shown in Fig. 1 and Table 1.

The YCpWZ plasmid series was constructed to contain a fusion of *CYC7* and the *Escherichia coli galK* gene encoding galactokinase. The initial fusion was constructed by using the YCp*CYC7*(2)r deletion plasmid B297. Base pairs −14 to 263 of B297 were deleted and replaced with a 1.2-kb *Eco*RI

fragment containing the entire *E. coli galK*-coding sequence (from plasmid YRp72 [24, 35]) plus 27 bp of 5′- and 51 bp of 3′-flanking sequences (including the *Eco*RI linkers added to B297). This new plasmid, shown in Fig. 2, was designated YCpWZ-B297. To obtain *CYC7-galK* fusions for the other plasmids containing mutations in the *CYC7* 5′-flanking sequences, YCpWZ-B297 was digested with *Xho*I and *Bgl*II. *Xho*I cleaved the plasmid once at −142, and the 3-kb *Xho*I-*Bgl*II fragment containing the fusion was gel purified and used to replace the *Xho*I-*Bgl*II fragment in the target plasmid. Since HB101 is *galK⁻* and the *CYC7-galK* fusion was expressed in *E. coli*, bacterial transformants carrying the desired plasmids were selected and identified as purple-staining colonies on MacConkey agar (Difco Laboratories) supplemented with 0.4% galactose and 50 μg of ampicillin per ml. Restriction analysis was used to confirm that the appropriate construction was obtained. The fusion plasmid containing the wild-type upstream sequences of *CYC7* was designated YCpWZ2. The mutant fusion plasmids were designated YCpWZ-X, where X represents the name of the original mutant plasmid.

Δ41-trp⁻ was constructed by digesting Δ41 with *Xba*I, which cleaves once within the *TRP1*-coding sequence, filling in the single-stranded ends with DNA polymerase I, and recircularizing the plasmid with T4 DNA ligase. The resulting plasmid contained a frameshift mutation in the *TRP1* gene and, therefore, was incapable of transforming *trp1* yeasts to prototrophy.

TABLE 1. Galactokinase levels in deletion mutants

Plasmid	Base pairs deleted	Repeats deleted	Galactokinase activity (pmol/min per μg of protein) under indicated conditions			Mutant/wild-type ratio[a]
			Aerobic		Anaerobic (glucose)	
			Raffinose	Glucose		
YCpWZ2			1.8	0.94	0.94	
YCpWZ-Δ41	−275 to −315	3′, 5′, I	11.6	4.9	2.4	6.4
YCpWZ-Δ28	−291 to −315	5′, I	5.8	2.3	NT[b]	3.2
YCpWZ-Δ17	−275 to −292	3′	3.3	1.4	NT	1.8
YCpWZ-B297	−716 to −297	I, (5′)[c]	3.5	1.1	NT	1.9
YCpWZ-X238	−145 to −238		0.11	<0.02	<0.02	0.06

[a] Calculated by dividing the galactokinase activity in the mutant raffinose extract by that in the wild-type raffinose extract.
[b] NT, Not tested.
[c] Only part of the 5′ repeat was deleted, and the *Bam*HI linker restored a number of the critical base pairs.

Mutagenesis. Sodium bisulfite in vitro mutagenesis of the single-stranded regions in heteroduplexed plasmids was carried out as described previously (23). Two different types of heteroduplexes were mutagenized. In the first case, YCpCYC7(2)r cleaved at the unique *Bam*HI site was hybridized with Δ41-trp⁻ cleaved at the unique *Bgl*II site. In heteroduplex molecules, the 42 nucleotides present in the wild-type plasmid but not in the Δ41-trp⁻ deletion plasmid formed a single-stranded loop and, therefore, were exposed to sodium bisulfite attack. The mutagenized heteroduplexes were transformed into the *ung⁻ E. coli* strain BD1528 and, after amplification, the plasmids were transformed into *cyc1 cyc7* yeasts. Trp⁺ transformants were selected, excluding any Δ41-trp⁻ plasmids that were still present. The yeast transformants were pooled and replated on complete medium lacking tryptophan and containing lactic acid as the sole energy source. This medium selected for transformants containing up mutations in the *CYC7* gene. The plasmids were recovered from these mutants (30), and restriction analysis was performed to ensure that no rearrangements had occurred. Those plasmids containing a wild-type restriction pattern were retransformed into *cyc1 cyc7* cells to ensure that the Lac⁺ phenotype was a function of the plasmid. The sequences of the upstream regions of plasmids that passed these tests were determined.

In the second mutagenesis, heteroduplexes were formed between YCpCYC7(2)r cleaved at the unique *Bgl*II site and X347 cleaved at the unique *Xho*I site. In this case, the heteroduplexes that formed contained a single-stranded gap

ACTACATTAC**GGAATTCCCGAATCCGGAGTGTAAGAA ATG AGT CTG AAA**
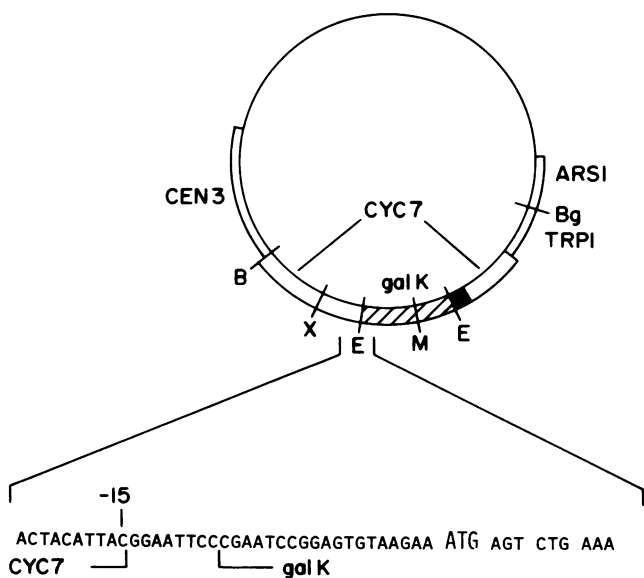CYC7 ⌐   ⌐⌐⌐ gal K

FIG. 2. *CYC7-galK* fusion plasmid. A 1.2-kb *Eco*RI fragment containing the *galK*-coding sequence plus short segments of both the 5'- and 3'-flanking regions was substituted for base pairs −14 to 263 of the *CYC7* gene in plasmid YCpCYC7(2)r. In this figure, the line represents pBR322 sequences, the thin boxes represent the *TRP1-ARS1* and *CEN3* sequences, the open thick box represents the *CYC7* noncoding sequences, the solid box represents the remaining 3' *CYC7*-coding sequence, and the striped box represents the *galK* sequence. The symbols for the restriction enzyme sites are as follows: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; M, *Mlu*I; and X, *Xho*I. The nucleotide sequence below the plasmid diagram represents the 5' fusion junction, with the *galK* ATG initiation codon indicated in larger letters.

rather than a loop, since the *Xho*I site in X347 lies at the site of the deletion. After mutagenesis, this gap was filled in with T4 DNA polymerase. The advantage of this procedure was that no X347 deletion mutants were recovered after transformation of *E. coli* cells. The remainder of the procedure was identical to that described above.

Sequence analyses and galactokinase assays. The procedure of Sanger et al. (25) was used for DNA sequence analyses with *Xho*I-*Bam*HI fragments cloned into the *Sal*I and *Bam*HI sites of the M13 cloning vector mp8 (20). Galactokinase assays were performed as described previously (24, 36). For every mutant plasmid, assays were performed on extracts from at least two independent yeast transformants. The variation in the galactokinase levels determined in extracts prepared from independent transformants containing the same plasmid was less than 20%.

Materials. Enzymes were purchased from Boehringer-Mannheim Biochemicals, New England BioLabs, Inc., Bethesda Research Laboratories, Inc., and International Biotechnologies, Inc. T4 DNA ligase was purchased from New England Nuclear Corp. Enzyme reactions were carried out in accordance with the instructions of the manufacturers.

RESULTS

The transcription of the *CYC7* gene, encoding the iso-2-cytochrome *c* protein, is influenced by the presence of both a positive site located around 240 bp 5' from the coding sequences (−240) and a negative site located around −300. These sites have been defined by deletion studies (34) in which the deletion of the positive site caused decreased expression and the deletion of the negative site resulted in increased expression. In these studies (34), it was postulated that the negative site contained redundant information both because of the pattern of phenotypes obtained with various small deletions within the region and because of the presence of repeated sequences. The purpose of the present study was to test this hypothesis and, to do so, we required both a more quantitative analysis for gene expression than could be achieved either through Northern analysis or cytochrome *c* assays and point mutations which would precisely define the base pairs in the negative site.

To achieve a more quantitative assay for *CYC7* expression, we fused the *CYC7* regulatory sequences to the coding sequence of the *E. coli* galactokinase gene, *galK* (Fig. 2). We had previously shown that fusions between the yeast *CYC1* gene and *galK* complemented mutations in the yeast galactokinase gene, *GAL1*, and that the enzyme produced could be easily and quantitatively assayed (24, 35). The fusion that was constructed contained the 5'-flanking sequences of *CYC7* up to −15, including the transcriptional regulatory signal plus part of the leader sequence of the mRNA (21), followed by 27 bp of the *galK* leader sequence, the *galK* ATG initiation codon, which is efficiently recognized in *S. cerevisiae* (36), and the *galK*-coding sequence. When the deletions within the *CYC7* upstream sequences shown in Fig. 1 were placed upstream from this fusion, the levels of galactokinase produced in *gal1* yeast transformants followed the same pattern as that for the expression of the intact *CYC7* gene (Table 1). Cells containing the wild-type fusion conatined relatively low levels of galactokinase (as compared with a *CYC1* fusion; data not shown). which was subject to glucose repression but not oxygen regulation, in agreement with previous studies of the expression of the *CYC7* gene (21, 34; Weiss et al., submitted). Cells containing fusion plasmids having deletions in the region of the negative site all had increased galactokinase levels, with YCpWZ-Δ41

transformants showing the greatest increase (sixfold over wild-type levels). We had previously found that the Δ41 deletion had the highest cytochrome c levels and best growth on lactic acid (34). Also, as expected, YCpWZ-X238 transformants contained extremely low galactokinase levels, consistent with the previous finding that the positive site was deleted in X238 (34). Thus, there was an extremely good correlation between the expression of the intact CYC7 gene and the levels of galactokinase in extracts of cells transformed with the various fusion plasmids, indicating the usefulness of this fusion as a tool for the quantitation of mutant phenotypes.

It is also interesting to note that in the absence of a functional negative site, the expression of the fusion gene was inducible by oxygen, as can be seen for YCpWZ-Δ41 in Table 1, although the effect was negligible (twofold) as compared with the orders-of-magnitude induction observed for the CYC1 gene (15, 16). Finally, and most significantly for the purpose of this study, these results indicated a correlation between galactokinase activity and the number of repeats deleted from the negative site (Fig. 1 and Table 1).

Point mutations in the negative site. The most direct approach to the question of which base pairs are in the negative site would be to identify single base-pair mutations which would cause an up phenotype. However, such mutations have proven exceedingly difficult to obtain by both in vivo mutagenesis and random in vitro mutagenesis of a plasmid-borne gene (33). These failures suggested that the target was either very small or that multiple base-pair changes were required. Therefore, we chose an in vitro mutagenesis procedure devised by Peden and Nathans (23) which produces mutations in only a very limited region at high efficiency and which gives a high proportion of multiple hits. This procedure involved the formation of heteroduplex molecules between a wild-type plasmid and a plasmid containing a deletion in the negative site. In such heteroduplexes, the negative site of the wild-type plasmid was exposed as a single-stranded region and therefore was much more sensitive to mutagenesis by sodium bisulfite.

Initially, this scheme was carried out with heteroduplexes between YCpCYC7(2)r and the Δ41 deletion plasmid, in which the TRP1 gene had been destroyed to avoid the recovery of the Δ41 plasmid when the Trp$^+$ Lac$^+$ phenotype was selected. (This scheme was not entirely successful, as Δ41-trp$^+$ plasmids were recovered at a low frequency [1%], presumably because of a conversion-type event in the heteroduplex.) Mutagenesis was carried out for 1 and 2 h, and the DNA was then transformed into E. coli cells. Over 10^6 independent transformants were obtained. After amplification of the mutant plasmid pool in bacteria, the plasmids were transformed into cyc1 cyc7 yeasts. Lac$^+$ transformants were found at frequencies of 3 and 6% in the 1-and 2-h pools, respectively, each pool containing over 10^4 independent transformants. The plasmids were rescued from these transformants, restriction analysis was performed to verify that no large DNA rearrangements had occurred, and then the plasmids were retransformed into yeast cells to ensure that the mutant phenotype was a property of the plasmids. The sequences of the upstream regions from −142 to −350 of those plasmids which fulfilled the above criteria were determined, and the sequences for part of those regions are shown in Fig. 3. The most striking feature of the pattern of mutations was that all involved more than a single base-pair change. Also, as expected from the specificity of sodium bisulfite (23), all were GC-to-AT transitions and, because only one strand in each heteroduplex molecule was exposed

to sodium bisulfite attack, the changes in any given mutant were all either C to T or G to A for the strand shown.

The detection of only multiple hits reflected either that the distribution of hits within the mutagenized pool was greatly weighted towards multiple hits or that more than a single base-pair change was required to answer the selection, single base-pair mutations being either impossible or extremely rare. In an attempt to distinguish between these two possibilities, we determined by sequence analysis the number of mutations within the upstream regions of five plasmids selected at random from the 2-h pool, r1, r2, r5, r6, and r11. These plasmids were also tested for their ability to transform yeasts to the Lac$^+$ phenotype. Two, r6 and r11, proved to have the mutant phenotype, and their sequences are included among the mutant sequences in Fig. 3. The sequences for the other three are shown as nonmutant sequences in Fig. 3. Also shown as nonmutant sequences are the base-pair changes in an additional six plasmids which were originally thought to have mutant phenotypes but proved negative in a second transformation. Although it is obvious from these data that multiple hits are common (in fact, no plasmids were found with no base-pair changes), more than half of the plasmids with the wild-type phenotype contained only one hit. Therefore, we concluded that single hits were not rare within the plasmid population.

When the Δ41-trp$^-$ deletion plasmid was used for heteroduplex formation, the range of mutations that could be obtained was restricted to a small region. To determine whether mutations could be obtained outside this region, which would cause increased CYC7 expression, we used the same basic mutagenesis procedure but with a larger deletion plasmid, X347. In this plasmid, base pairs −145 to −347 are deleted and, therefore, all these base pairs in the wild-type plasmid were exposed to mutagenesis. The mutations obtained from a 45-min mutagenesis are shown in Fig. 3. All the mutant plasmids found in this experiment contained mutations within the same region (although some did contain base-pair changes outside this region, too). In this case, a single point mutation, 2, which caused increased CYC7 expression, was found. (The upstream region of this plasmid was subcloned into M13mp8, and the sequence was determined in two independent experiments to ensure that the sequence was correct.) The significance of this mutation and the general pattern of mutations are discussed below.

Galactokinase levels in the point mutants. To quantitate the extent of increased expression from the various mutant plasmids, the CYC7 sequences from the XhoI site at −142 through the coding and 3′-flanking sequences were replaced by an XhoI-BglII fragment containing the CYC7-galK fusion. These fusion plasmids were transformed into gal1 yeast cells, and the levels of galactokinase in each were determined (Table 2). The galactokinase levels ranged from 1.9 to 3.7 times those in the wild type. In addition, the fact that these fusion plasmids manifested the up phenotype demonstrated that this phenotype resulted from mutations in the region sequenced and not from some downstream mutation which had gone undetected. Fusions were also constructed for some of the nonmutant genes whose sequences had been determined and, as expected, the galactokinase levels in cells transformed with these plasmids were similar to those in the wild type (Table 2).

DISCUSSION

Within the region defined as the negative site of CYC7, there are three sequences that contain striking homology.
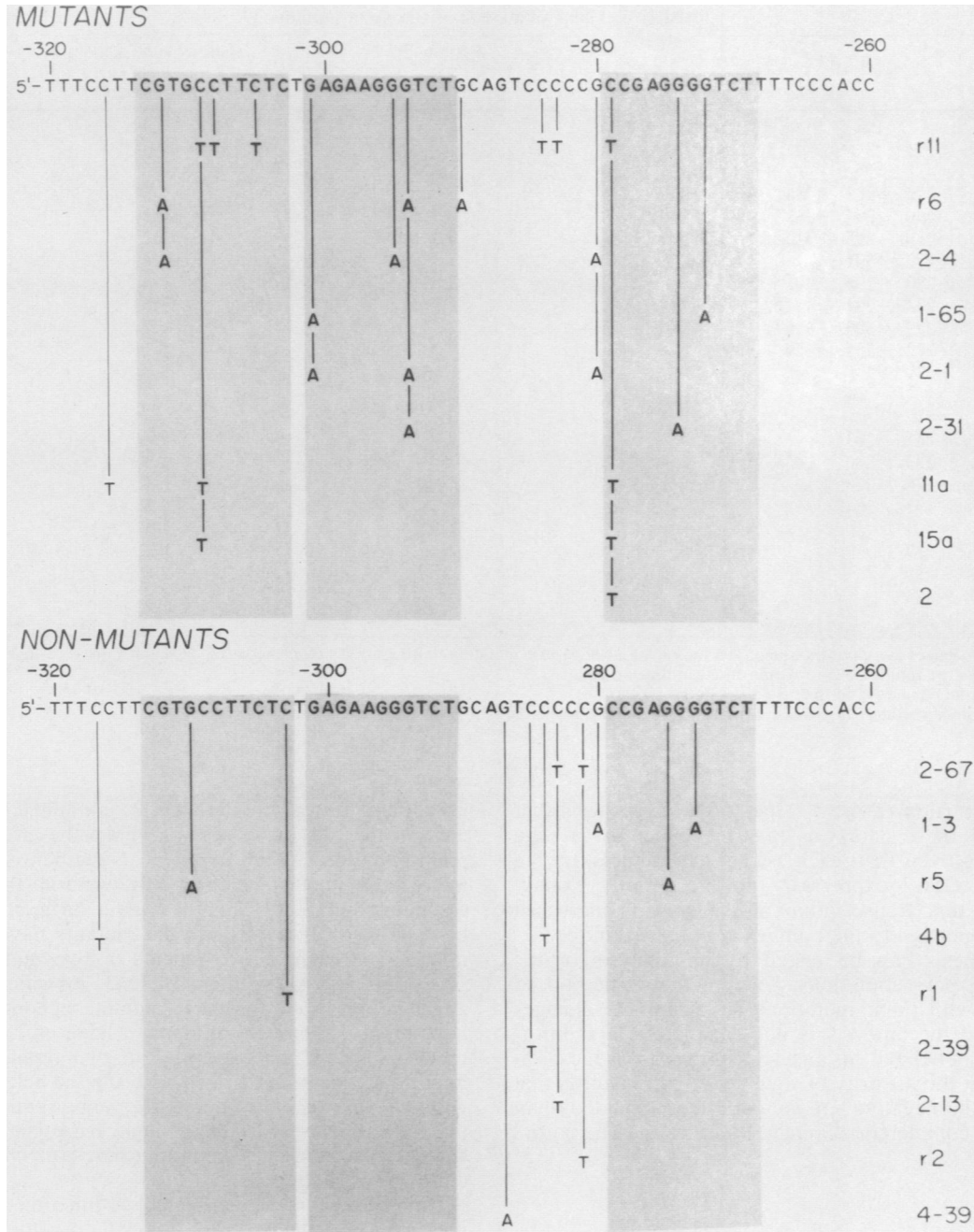
FIG. 3. Point mutations in the negative site. The coding strand from −320 to −260 is represented at the top and in the middle of the figure. The base-pair changes for the various mutants (plasmids giving increased expression of *CYC7*) are indicated below the affected base in the upper half of the figure, and those for the nonmutants are indicated in the lower half. Mutant 15a and nonmutant 4b have additional changes (see Table 2). The regions of the three repeats are shaded. Plasmids 1-65, 1-3, 2-13, 2-39, 2-67, and 4-39 were derived from the 1-h mutagenesis, and plasmids 2-1, 2-4, 2-31, r1, r2, r5, r6, and r11 were derived from the 2-h mutagenesis; both mutagenesis experiments were carried out with wild-type–Δ41-trp⁻ heteroduplexes. Plasmids 2, 4b, 11a, and 15a were derived from a 45-min mutagenesis experiment with wild-type–X347 heteroduplexes.

These repeated sequences are shown in Fig. 4. Two are oriented in the same direction, whereas the third is inverted; therefore, these repeats have been designated the 3′ (−279 to −269), 5′ (−301 to −291), and I (−313 to −303) repeats. An analysis of the effects of the deletion mutations suggested that these repeats are integral components of the negative site. As can be seen in Fig. 1, the Δ41 deletion removed all of the 5′ and I repeats and most of the 3′ repeat. This

deletion caused the greatest increase in galactokinase levels, sixfold, when placed upstream from the fusion gene. The Δ28 deletion removed the 5′ and I repeats but left the 3′ repeat intact and caused about a threefold increase in galactokinase levels. The Δ17 deletion removed only one repeat and caused a rather small increase in galactokinase levels, about twofold. The B297 deletion also caused only a small increase in galactokinase levels and removed the I

TABLE 2. Galactokinase levels in point mutants

Plasmid	Mutation[a]	Repeat affected[b]			Galactokinase activity (pmol/min per μg of protein)	Mutant/wild-type ratio[c]
		3'	5'	I		
YCpWZ2					1.8	
Mutant						
r11	−279, −283, −284, −305, −308, −309*	1		3, 6, 7	4.6	2.5
r6	−290, −294, −312		8	10	6.7	3.7
2-4	−280, −295, −312		7	10	5.2	2.9
1-65	−272, −301	8	1		4.4	2.4
2-1	−280, −294, −301		1, 8		4.0	2.0
2-31	−274, −294	7	8		3.5	1.9
11a	−279, −309, −316*	1		7	4.7	2.6
15a	−172, −255, −279, −309*	1		7	5.1	2.8
2	−279*	1			3.8	2.1
Nonmutant						
2-67	−281, −283*				1.8	1.0
1-3	−273, −280	7			1.3	0.7
r5	−275, −310	5		8	NT[d]	NT
4b	−192, −284, −317*				1.8	1.0
r1	−303*			1	NT	NT
4-39	−287				NT	NT
2-39	−285*				1.6	0.9
2-13	−283*				2.3	1.2
r2	−281*				NT	NT

[a] The base-pair changes are all of the same type for each mutant. Those mutations with asterisks after the base-pair numbers have C-to-T transitions in the coding strand; those without have G-to-A transitions in that strand.

[b] See Fig. 4 for the numbering of the base pairs in the repeats.

[c] Calculated by dividing the galactokinase activity in the mutant extract by that in the wild-type extract.

[d] NT, Not tested.

repeat and a portion of the 5' repeat. The hypothesis that emerged from these data was that the effect of deleting repeats is additive, that each repeat contributes to the repression of *CYC7* expression (not necessarily exactly equally), and that transcription of the gene is increased roughly in proportion to the number of repeats removed.

This hypothesis can be tested by an analysis of the location of the point mutations. Unfortunately, the bulk of the plasmids with point mutations have multiple changes and, therefore, the analysis is not completely straightforward. We do not believe it is an accident that this is the case; we believe that the majority of single base-pair changes that occur within the negative site cannot affect a single repeat sufficiently to completely eliminate its function. The partial

```
3'      C C G A G G G G T C T

5'      G A G A A G G G T C T

I       G A G A A G G C A C G

CONSENSUS C/G − G A − G G C/G − C −

        1 2 3 4 5 6 7 8 9 10 11
```
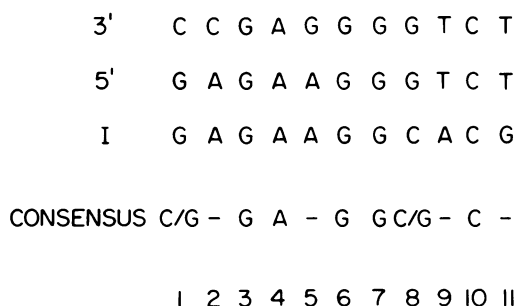
FIG. 4. Consensus sequence of the negative-site repeats. The sequences of the 3' and 5' repeats represent the coding strand, and the sequence of the I repeat represents the noncoding strand. The consensus sequence, as well as the numbering of the bases in the repeats, is indicated at the bottom of the figure.

functioning of the mutated repeat, combined with the presence of the two intact repeats, make the effect of a single point mutation difficult to detect. Nonetheless, a number of interesting conclusions can be drawn from the data. First, excluding mutant 2, which has just a single change and will be discussed separately, all the mutants have at least one mutation in each of two repeats or two mutations in one repeat, as can be seen in Table 2, in which the repeats affected are listed for each mutant. Second, among the nonmutants, although a number have mutations in one repeat, only r5 has mutations in two repeats; this is an important exception that will be discussed below. Therefore, with the two exceptions noted, the general rule appears to be that two repeats must have single base-pair changes or a single repeat must have multiple base-pair changes to generate an up phenotype.

If these repeats are equivalent in function, as, for example, targets for the binding of a regulatory complex, another rule should hold: only mutations in base pairs that are common to the three repeats should disrupt the function of the repeats. Base pairs 3, 4, 6, 7, and 10 (as numbered in Fig. 4) are common to all three repeats, so these base pairs should be most sensitive to mutation. Clearly, the rule in this form does not hold; only mutant 2-4 has two of these base pairs altered in two repeats (Table 2). However, this rule can be extended to include base pairs 1 and 8 as equivalent if we assume that these base pairs are recognized by a protein interacting with the minor groove. Seeman et al. (26) postulated that proteins could not distinguish GC from CG base pairs in the minor groove but that either base pair could be easily distinguished from an AT or TA base pair. Thus, when mutations in base pairs 1 and 8 as well as 3, 4, 6, 7, and 10 are included as being potentially disruptive to repeat function, all the mutants, except mutant 2, satisfy both rules. Signifi-

cantly, nonmutant r5, which was noted above to have mutations in two repeats, does not satisfy the second rule and should not have a mutant phenotype; one of its mutations is in base pair 5 of the 3' repeat,which is not a conserved site. Thus, with the exception of mutant 2, the pattern of base-pair changes in all the mutants and nonmutants fit the hypothesis that the three repeats are in the negative site.

Mutant 2 represents a special and perhaps the most interesting case; it has a single base-pair change at base pair 1 of the 3' repeat. This site obviously is very sensitive, and the same mutation is present in three other mutants, r11, 11a, and 15a. Each of these three mutants also has a change in one of the other repeats, and it is difficult to assess whether these additional changes have any effect; galactokinase expression in each of these mutant plasmids was only marginally greater than that in mutant 2. Interestingly, *CYC7* expression is not as sensitive to changes at the equivalent base pair of the I repeat, since nonmutant r1 has base pair 1 of this repeat altered. This suggests that the repeats are not entirely equivalent in function either in terms of their effect on transcription or in the way a regulatory element binds to them. Perhaps the nonequivalence is related to their location with respect to the positive site; the 3' repeat, which is closer, may have a stronger antagonistic effect. Such a differential effect would not be detected in the deletions, since removal of the 3' repeat by deletion simply places the 5' repeat closer to the positive site. The ability to isolate mutant 2 indicates that it is not impossible to obtain single base-pair changes, but the overall data seem to indicate that the target for such changes is small, perhaps solely this base pair.

In addition to base pair 1 of the 3' repeat, one other base pair can unequivocally be assigned an essential role in the function of the negative site, base pair 8 of the 5' repeat. Nonmutant 1-3 has a mutation in basepair 7 of the 3' repeat, but this change alone does not cause an up phenotype. Mutant 2-31, on the other hand, has this same base pair plus base pair 8 of the 5' repeat altered. Thus, either the two mutations in combination or the latter mutation alone is responsible for the up phenotype of this mutant; in either case, the mutation in base pair 8 of the 5' repeat plays a critical role. None of the other conserved base pairs in the repeats can be assigned essential roles unambiguously, although the overall pattern of mutations strongly suggests that they play some role.

The location of the three repeats with respect to each other is interesting and should be noted. The 5' and I repeats in combination represent a sequence with a twofold axis of symmetry around base pair −302. This arrangement immediately suggests a symmetric binding site with which a dimeric protein might interact. However, such an interpretation is not supported by the limited data. No such symmetry involving the 3' repeat exists, and the deletion and point mutation data, especially the existence of mutant 2, indicate that this repeat plays as important a role in the negative site as the other two. If a dimer bound cooperatively to the symmetric sequence involving the 5' and I repeats, only half of the dimer would bind at the 3' repeat. As a result, the protein would not bind cooperatively, resulting in a greatly reduced role for this repeat in decreasing *CYC7* expression. Also, the loss of the I repeat in deletion B297 would be expected to have a much more than additive effect in increasing gene expression if it represented half the target for the cooperative binding of a dimeric protein. Therefore, although an attractive symmetric sequence is present, we cannot at present assign it an important role as a binding site for a regulatory protein.

There are other regulatory sequences in *S. cerevisiae* which involve repeated information. The activation site for general amino acid control of the *HIS4* gene contains a number of repeated sequences, at least one of which appears to be required for activation (6, 17). The *CYC1* oxygen regulatory sites are also repeated (10). Although the reason for this redundancy is not clear, it may not be uncommon, and the negative site of the *CYC7* gene may be another example.

It should be noted that the increase in *CYC7* expression caused by most of the deletion mutations and the point mutations was not large (two- to threefold), and the greatest effect, seen for the Δ41 mutation, was only sixfold. We believe that the point mutations could be isolated only because of a combination of the extreme sensitivity of a *cyc1 CYC7* cell to increased levels of cytochrome *c* when grown on lactic acid and the high, localized mutation rate, which limited the numbers of cells that had to be screened and raised the rate of multiple hits above that for large deletions, transposable element insertions, and *trans*-acting mutations, all of which can cause increased *CYC7* expression. This mutagenesis scheme has the additional value that it introduces random changes in a specified region, which can be expanded or contracted as desired, and therefore limits the bias of researchers in selecting what base pairs to alter, a step that is necessary in site-directed mutagenesis. However, it is clear that to extract additional information about the negative site, specific base-pair changes are required.

LITERATURE CITED

1. **Bolivar, F., R. L. Rodriguez, P. J. Green, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 2:95–113.
2. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. 4:459–472.
3. **Clarke, L., and J. Carbon.** 1980. Isolation of a yeast centromere and the construction of functional small circular chromosomes. Nature (London) 287:504–509.
4. **Clavalier, L., G. Pere-Aubert, and P. P. Slonimski.** 1969. Mise en evidence de plusieurs loci independants impliques dans la synthese de l'iso-2-cytochrome c chez la levure. Mol. Gen. Genet. 104:195–218.
5. **Clavalier, L., G. Pere-Aubert, M. Somlo, and P. P. Slonimski.** 1976. Reseau d'interactions entre des genes non lies: regulation synergique ou antagoniste de la synthese de l'iso-1-cytochrome c, de l'iso-2-cytochrome c et du cytochrome b₂. Biochimie 58:155–172.
6. **Donahue, T. F., R. S. Daves, G. Lucchini, and G. R. Fink.** 1983. A short nucleotide sequence required for regulation of *HIS4* by the general control system in yeast. Cell 32:89–98.
7. **Downie, J. A., J. W. Stewart, N. Brockman, A. Schweingruber, and F. Sherman.** 1977. Structural gene for iso-2-cytochrome c. J. Mol. Biol. 113:369–384.
8. **Errede, B., T. S. Cardillo, F. Sherman, E. Dubois, J. Deschamps,**

and J. M. Wiame. 1980. Mating signals control expression of mutations resulting from insertion of a transposable repetitive element adjacent to diverse yeast genes. Cell **22**:427–436.

9. **Faye, G., D. W. Leung, K. Tatchell, B. D. Hall, and M. Smith.** 1981. Deletion mapping of sequences essential for *in vivo* transcription of the iso-1-cytochrome c gene. Proc. Natl. Acad. Sci. USA **78**:2258–2262.

10. **Guarente, L., B. Lalonde, P. Gifford, and E. Alani.** 1984. Distinctly regulated tandemly upstream activation sites mediate catabolite repression of the *CYC1* gene of *S. cerevisiae.* Cell **36**:503–511.

11. **Guarente, L., and T. Mason.** 1983. Heme regulates the transcription of the *CYC1* gene of *S. cerevisiae* via an upstream activation site. Cell **32**:1279–1286.

12. **Guarente, L., R. R. Yocum, and P. Gifford.** 1982. A *GAL10-CYC1* hybrid yeast promoter identifies the *GAL4* regulatory region as an upstream site. Proc. Natl. Acad. Sci. USA **79**:7410–7414.

13. **Hanahan, H.** 1983. Studies on the transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.

14. **Ito, H., Y. Fukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153**:163–168.

15. **Lowry, C. V., J. L. Weiss, D. A. Walthall, and R. S. Zitomer.** 1983. Modulator sequences mediate the oxygen regulation of *CYC1* and a neighboring gene in yeast. Proc. Natl. Acad. Sci. USA **80**:151–155.

16. **Lowry, C. V., and R. S. Zitomer.** 1984. Oxygen regulation of aerobic and anaerobic genes mediated by a common regulatory factor in yeast. Proc. Natl. Acad. Sci. USA **81**:6129–6133.

17. **Lucchini, G., A. G. Hinnebusch, C. Chen, and G. R. Fink.** 1984. Positive regulatory interactions of the *HIS4* gene of *Saccharomyces cerevisiae.* Mol. Cell. Biol. **4**:1326–1333.

18. **Matinez-Arias, A., H. Yost, and M. Casadaban.** 1984. The role of an upstream regulatory element in leucine repression of the *Saccharomyces cerevisiae LEU2* gene. Nature (London) **307**:740–742.

19. **McKnight, G. L., T. S. Cardillo, and F. Sherman.** 1981. An extensive deletion causing overproduction of yeast iso-2-cytochrome c. Cell **25**:409–419.

20. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.

21. **Montgomery, D. L., J. M. Boss, S. J. McAndrew, L. Marr, D. A. Walthall, and R. S. Zitomer.** 1982. The molecular characterization of three transcriptional mutations in the yeast iso-2-cytochrome c gene. J. Biol. Chem. **257**:7756–7761.

22. **Montgomery, D. L., D. W. Leung, M. Smith, P. Shalit, G. Faye,**

and B. D. Hall. 1980. Isolation and sequence of the gene for iso-2-cytochrome c in *Saccharomyces cerevisiae.* Proc. Natl. Acad. Sci. USA **77**:541–545.

23. **Peden, K. W. C., and D. Nathans.** 1982. Local mutagenesis with deletion loops of DNA heteroduplexes. Proc. Natl. Acad. Sci. USA **79**:7214–7217.

24. **Rymond, B. R., R. S. Zitomer, D. Schumperli, and M. Rosenberg.** 1983. The expression in yeast of the *Escherichia coli galK* gene on *CYC1::galK* fusion plasmids. Gene **25**:249–262.

25. **Sanger, F., S. Nicklen, and S. A. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

26. **Seeman, N. C., J. M. Rosenberg, and A. Rich.** 1976. Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. USA **73**:804–808.

27. **Sherman, F., J. W. Stewart, M. Jackson, R. A. Gilmore, and J. H. Parker.** 1974. Mutants in yeast defective in iso-1-cytochrome c. Genetics **77**:255–284.

28. **Sherman, F., J. W. Stewart, E. Margoliash, J. Parker, and W. Campbell.** 1966. The structural gene for yeast cytochrome c. Proc. Natl. Acad. Sci. USA **55**:1498–1504.

29. **Struhl, K.** 1982. Regulatory sites for *HIS3* gene expression in yeast. Nature (London) **300**:284–287.

30. **Struhl, K., D. T. Stinchcomb, S. Scherer, and R. W. Davis.** 1979. High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. Proc. Natl. Acad. Sci. USA **76**:1035–1039.

31. **West, R. W., Jr., R. R. Yocum, and M. Ptashne.** 1984. *Saccharomyces cerevisiae GAL1-GAL10* divergent promoter region: location and function of the upstream activating sequence UAS$_G$. Mol. Cell. Biol. **4**:2467–2478.

32. **Wilson, K. L., and I. Herskowitz.** 1984. Negative regulation of *STE6* gene expression by the α2 product of *Saccharomyces cerevisiae.* Mol. Cell. Biol. **4**:2420–2427.

33. **Wright, C. F., D. A. Walthall, J. M. Boss, and R. S. Zitomer.** 1983. DNA insertions which affect the expression of the yeast iso-2-cytochrome c gene. Curr. Genet. **7**:117–122.

34. **Wright, C. F., and R. S. Zitomer.** 1984. A positive regulatory site and a negative regulatory site control the expression of the *Saccharomyces cerevisiae CYC7* gene. Mol. Cell. Biol. **4**:2023–2030.

35. **Zitomer, R. S., B. R. Rymond, D. Schumperli, and M. Rosenberg.** 1983. Expression vectors for fusions of the *E. coli galK* gene to yeast *CYC1* gene regulatory sequences. UCLA Symp. Mol. Cell. Biol. New Ser. **8**:523–541.

36. **Zitomer, R. S., D. A. Walthall, B. R. Rymond, and C. P. Hollenberg.** 1984. *Saccharomyces cerevisiae* ribosomes recognize non-AUG initiation codons. Mol. Cell. Biol. **4**:1191–1197.