DIVERGENT AND CONCERTED EVOLUTION OF THE TWO REGIONS ENCOMPASSING THE ISO-1-CYTOCHROME c AND ISO-2-CYTOCHROME c GENES OF YEAST

(Saccharomyces cerevisiae, *cytochrome* c *of yeast, evolution of yeast, cloning yeast genes, recombinant DNA, transpositions, composite genes*)

FRED SHERMAN, JOHN I. STILES*, LINDA R. FRIEDMAN, JOACHIM F. ERNST[†] and GARY L. McKNIGHT[‡]

Department of Radiation Biology and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642; *Department of Life Sciences, Indiana State University, Terre Haute, Indiana 47809; [†]Biogen S.A., Route de Troinex 3, 1227 Carouge, Geneva, Switzerland; [‡]Department of Genetics, University of Washington, Seattle, Washington 98195

SUMMARY

The yeast Saccharomyces cerevisiae *contains a cluster of three genes* CYC1-OSM1-RAD7, *denoted* COR, *located on the right arm of chromosome* X *and a cluster of three genes* ANP1-RAD23-CYC7, *denoted* ARC, *located on the left arm of chromosome* V. *Both* COR *and* ARC *are in approximately 6 kb long regions and there are striking similarities between* CYC1 *and* CYC7, *between* OSM1 *and* ANP1 *and between* RAD7 *and* RAD23. *The relationships between* COR *and* ARC *suggest that one of the regions arose during evolution by a transposition in which a circular intermediate containing the three genes integrated at a new chromosomal position. The order of the genes could have been altered if the circle closed and opened at different sites during the excision and integration, respectively. Subsequently, the primordial* COR *and* ARC *genes could have acquired slightly different properties or functions by divergent evolution. The hypothetical transposition that occurred during evolution superficially resembles transpositions observed in certain* COR2 *strains in which segments containing the entire* COR *region are transposed to various positions on different chromosomes.*

In addition to the evolutionary development by means of transpositions and divergencies, identical short stretches observed within the CYC1 *and* CYC7 *genes, as well as in other duplicated genes from yeast and other organisms, suggest the occurrence of concerted evolution. The possibility*

of concerted evolution has been directly demonstrated with the formation of the so-called composite genes, in which the central portions of the CYC1 locus is replaced with a corresponding segment from the CYC7 locus by gene conversion. Thus, by considering the structures of the COR and ARC regions, we suggest that these two regions developed by three evolutionary processes: a transposition, divergent evolution and concerted evolution. The transposition of COR regions and the formation of CYC1-CYC7 composite genes superficially resemble two of the evolutionary steps.

INTRODUCTION

The regions encompassing and adjacent to the structural genes of iso-1-cytochrome *c* and iso-2-cytochrome *c* in the yeast *Saccharomyces cerevisiae* have been systematically investigated by genetic analysis and by the analysis of DNA structures. Iso-1-cytochrome *c* is encoded by the *CYC1* gene located on the right arm of chromosome X (SHERMAN et al. 1966; LAWRENCE et al. 1975), whereas iso-2-cytochrome *c* is encoded by the *CYC7* gene located on the left arm of chromosome V (DOWNIE et al. 1977; SHERMAN et al. 1978). Because the *CYC1* and *CYC7* genes as well as their gene products are approximately 80% homologous (MONTGOMERY et al. 1980) it is generally accepted that they are evolutionarily related. In addition, the two genes *OSM1* and *RAD7* that are genetically linked to the *CYC1* locus show striking similarities, respectively, to the two genes *ANP1* and *RAD23* that are genetically linked to the *CYC7* locus; these similarities suggest that the two clusters of genes may have arisen from a common ancestral region (McKNIGHT et al. 1981). Consistent with this view and as described in this communication, the *CYC1, OSM1* and *RAD7* gene cluster, denoted COR (STILES et al. 1981a,b), and the *ANP1, RAD23* and *CYC7* gene cluster, denoted ARC (McKNIGHT et al. 1981), reside in regions that are both approximately 6 kb long (Fig. 1).
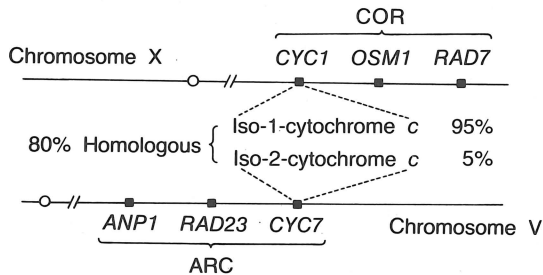

Fig. 1. The arrangement of the COR and ARC genes on, respectively, chromosomes X and V. The *CYC1* and *CYC7* genes products constitute, respectively, 95% and 5% of the total cytochrome *c* complement and are approximately 80% homologous.

Another feature revealed from the *CYC1* and *CYC7* sequences, as well as the sequences of numerous other duplicated genes from a wide range of organisms, is the nonrandom differences between pairs of nonallelic genes and small stretches of complete homology. These identical stretches within duplicated regions suggest the occurrence of recombination between the nonallelic genes after divergence, a process referred to as concerted evolution.

In addition to considering the structures of COR and ARC for deducing evolutionary pathways, one can take into account certain unusual mutations and rearrangements that appear to at least superficially resemble evolutionary developments. The entire COR region, when in a certain chromosomal constitution denoted COR2, can transpose to other sites on different chromosomes (STILES et al. 1981a,b). The occurrence of this type of transposition suggests that the COR and ARC regions may have duplicated from a common ancestral region through a transpositional rearrangement. Furthermore, recombination between the nonallelic *CYC1* and *CYC7* genes, leading to the formation of the so-called composite genes (ERNST et al. 1981; 1982), mimics the type of changes that are expected to occur during concerted evolution. Thus, by considering the COR and ARC structures and certain mutational and recombinational alterations, the following evolutionary pathways outlined in Figure 2 can be suggested: a duplication occurred by transposition of an ancient region; the two regions diverged into ancestral COR and ARC regions;
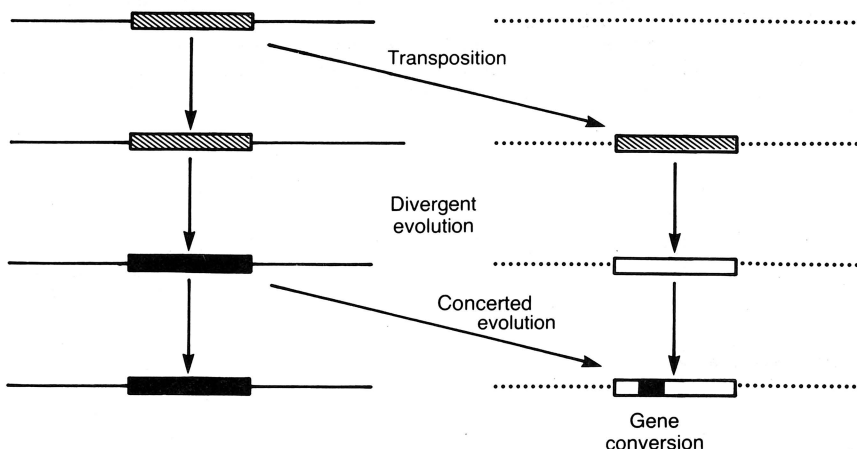

Fig. 2. The proposed evolutionary pathways for the COR and ARC regions.

and finally concerted evolution occurred by recombination between these ancestral COR and ARC regions. This paper summarizes the evidence bearing on these evolutionary pathways, and describes recent studies on the structures of COR and ARC, the COR2 transpositions and the formation of *CYC1* and *CYC7* composite genes.

STRUCTURE OF COR

Two different polymorphic regions, denoted COR1 and COR2 have been uncovered in normal wild-type strains of *S. cerevisiae*. Although both COR1 and COR2 strains have been and are being investigated, early studies of the *CYC1* locus were conducted with COR1 strains. Studies using DNA procedures were initiated after 44 nucleotides at one end of the translated portion of the *CYC1* gene were deduced from the altered iso-1-cytochrome *c* in frameshift revertants (SHERMAN & STEWART 1973; STEWART & SHERMAN 1974). A DNA restriction fragment containing the *CYC1* gene was identified with a deoxyribonucleotide probe that was 15 nucleotides long and that was complementary to a portion of this known 44 bp region (SZOSTAK et al. 1977; 1979). Fragments prepared by partial *Eco*RI digestion of yeast DNA were ligated into the bacteriophage λgtWES and the resulting bank was screened with the synthetic probe. A 9 kb fragment encompassing the *CYC1* locus was retrieved and identified by sequence analysis (SZOSTAK et al. 1979; STILES et al. 1981c). The *CYC1* gene was also cloned by MONTGOMERY et al. (1978) who used a similar procedure with a different synthetic probe.

Extensive regions adjacent to the *CYC1* locus and encompassing the entire COR1 region were cloned by integrating a plasmid into the COR1 region and then excising the plasmid along with new adjacent chromosomal regions (STILES et al. 1981a,b,c; STILES 1983). The integrating plasmid consisted of pBR322 with the *URA3* yeast gene and the *Hind*III-*Eco*RI fragment adjacent to, but not including the *CYC1* locus (Fig. 3). Genomic DNA from one of the yeast transformants was partially digested with *Eco*RI,
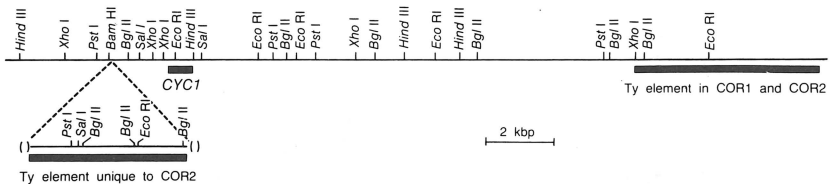

Fig. 3. The physical structures of the COR1 and COR2 regions are shown along with the positions of the *CYC1* locus, the Ty elements and the cleavage sites of various restriction endonucleases.

ligated and the resulting plasmids containing the *amp*[r] marker
were used to transform the *Escherichia coli* strain HB101. One of
plasmids, pAB40, was used to determine the restriction endo-
nuclease sites of the COR1 region shown in Figure 3 (STILES
1981a,b). In addition, segments of the COR1 regions were the
source of probes required for the analysis of deletion mutants
and transcription described below. A restriction map of the
same COR1 region, but including only *Eco*RI and *Hin*dIII sites,
was reported by SHALIT et al. (1981).

A series of 104 deletions encompassing at least portions
of the *CYC1* locus was isolated and genetically characterized by
SHERMAN et al. (1975). Four of these mutants contained addi-
tional phenotypes that could be ascribed to deletions of the
two genes *OSM1* and *RAD7*; two deletions encompassed *CYC1* and *OSM1*
and two encompassed *CYC1*, *OSM1* and *RAD7* (SINGH & SHERMAN 1978).
Mutations of *OSM1* caused increased sensitivity to osmotic agents
such as KCl, ethylene glycol, etc., whereas mutations of *RAD7*
caused increased sensitivity to UV. The deletion map suggested
that the *OSM1* and *RAD7* genes were contiguous to or nearby the *CYC1*
locus although the order of the genes could not be assigned
from these results alone. However, genetic analysis of *cyc1*
and *rad7* point mutations established that the *RAD7* locus was dis-
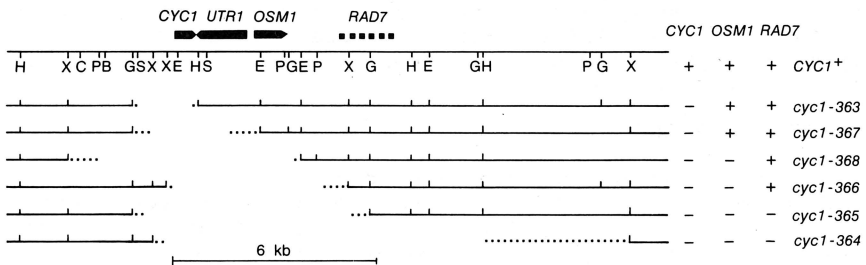tal to the *CYC1* locus (LAWRENCE et al. 1975).

Fig. 4. The physical structures of the *cyc1* deletions. The *CYC1*,
OSM1 and *RAD7* genotypes are indicated at the right (SINGH &
SHERMAN 1978). The *CYC1*, *UTR1* and *OSM1* transcripts (ZARET &
SHERMAN 1982) and the hypothetical *RAD7* transcripts are indicated
at the top. The cleavage sites for the following restriction
endonucleases are shown: *Hin*dIII, (H); *Xho*I, (X); *Sac*I, (C); *Pst*I,
(P); *Bam*HI, (B); *Bgl*II, (G); *Sal*I, (S); and *Eco*RI, (E). The
regions absent in the *cyc1* deletions are indicated by blank spaces
and the ambiguous regions are indicated by dotted lines. The
structures were determined by probing restriction fragments of
genomic DNA from the *cyc1* deletions (BARRY et al. in preparation).

Recently, the order and approximate physical positions of the *CYC1*, *OSM1* and *RAD7* gene cluster was determined by probing restriction fragments of genomic DNA from the deletion mutants (BARRY et al. unpublished results). The results, summarized in Figure 4, indicated that the order is *CYC1*, *OSM1* and *RAD7* and that the gene cluster is confined to an approximately 6 kb region.

The precise position of the *CYC1* locus, shown in Figure 4, was directly established by comparing the protein sequence to the DNA sequence (SMITH et al. 1979; STILES et al. 1981c) and by the location of the corresponding transcript (FAYE et al. 1981; BOSS et al. 1981). The results with the *cyc1-363* and *cyc1-367* deletions that are *cyc1⁻ OSM1⁺ RAD7⁺*, the *cyc1-368* and *cyc1-366*, deletions that are *cyc1⁻ osm1⁻ RAD7⁺* and the *cyc1-365* and *cyc1-364* deletions that are *cyc1⁻ osm1⁻ rad7⁻* establish that *OSM1* and *RAD7* are on the 3' end of the *CYC1* locus and that *OSM1* is between *CYC1* and *RAD7* as indicated in Figure 4.

Recently, ZARET & SHERMAN (1982) characterized two transcripts at the 3' end of the *CYC1* locus, a 1450 nucleotide long (nt) transcript immediately adjacent to the *CYC1* locus and a 1050 nt transcript immediately adjacent to the 1450 nt transcript. The approximate positions of the 1450 nt and 1050 nt transcripts were determined with various restriction fragments and the directions of transcription were determined with strand-specific probes. The results with the *cyc1* deletions suggest that the 1050 nt transcript corresponds to *OSM1* as shown in Figure 4. The locus corresponding to the 1450 nt transcript is denoted *UTR1* (unidentified transcript) in Figure 4. Mutants defective in *utr1*, such as the *cyc1-367* deletion (Figure 4) or *CYC1-512-Ins1* (ZARET & SHERMAN, in preparation) have no obvious phenotype. The transcript corresponding to *RAD7* is currently under investigation.

STRUCTURE OF ARC

The region encompassing and adjacent to the iso-2-cytochrome *c* structural gene, *CYC7*, has been characterized by analyzing mutants overproducing iso-2-cytochrome *c*. The three mutants *CYC7-H1*, *CYC7-H2* and *CYC7-H3* all overproduce 20 to 30 times the normal amount of iso-2-cytochrome *c*, all are dominant and all have gross alterations adjacent to the *CYC7* locus. The *CYC7-H1* mutation is a reciprocal translocation (SHERMAN & HELMS 1978), the *CYC7-H2* mutations is an insertion of a Ty element (ERREDE et al. 1980;1981) and the *CYC7-H3* is a deletion (McKNIGHT et al. 1981) which was instrumental for revealing the genetic information adjacent to the *CYC7* locus.

In addition to having the dominant phenotype of producing approximately 20 times the normal amount of iso-2-cytochrome *c*, the *CYC7-H3* mutant also has the following recessive traits: sensitivity to ultraviolet radiation; inability to grow on media made hypertonic with such agents as ethylene glycol or potassium chloride; sensitivity to amino nitrophenyl propanediol (ANP), a degradation product of chloramphenicol; flocculation. None of these recessive phenotypes found in the *CYC7-H3* mutant was observed in the *CYC7-H1* and *CYC7-H2* mutants. Furthermore, these

four phenotypes segregated concomitantly with the *CYC7-H3* allele, indicating that the multiple phenotypes can be attributed to a single-mutational alteration which, as described below, is a 5 kb deletion. Two point mutations *rad23-1* and *anpl-1* that failed to complement the recessive traits of the *CYC7-H3* deletion were isolated in a normal *CYC7⁺* strain. UV-sensitivity in the *CYC7-H3* deletion and the *rad23-1* point mutant is due to ʿdefects in the *RAD23* locus that is required for excision of UV damage. The phenotypes of osmotic sensitivity, ANP sensitivity and flocculation in the *CYC7-H3* deletion and the *anpl-1* point mutant are due to defects in the *ANP1* locus. Genetic analysis of the *rad23-1* and *anpl-1* point mutations and the *CYC7-H2* marker established that the three loci *ANP1*, *RAD23* and *CYC7* are within a genetic region of approximately 2 cM as shown at the top of Figure 5.

The physical alteration in the *CYC7-H3* mutant was shown to be a deletion by restriction mapping of a cloned 4.6 kb *Hind*III fragment containing the *CYC7-H3* gene and by using probes from the normal *CYC7⁺* region (Fig. 5) (McKNIGHT et al. 1981). Furthermore, the exact position of a deletion breakpoint was determined by DNA sequencing of a portion of the *CYC7-H3* gene as shown at the bottom of Figure 5 (KOSIBA et al. 1982). Thus, the *CYC7-H3* mutation is a deletion of an approximately 5 kb segment with a breakpoint 222 bp in front of the translated portion of the *CYC7* gene. The *CYC7-H3* mutant is osmotic and UV sensitive because the *ANP1* and *RAD23* genes are deleted and the mutant overproduces iso-2-cytochrome *c* because of fusion of an abnormal regulatory region to the *CYC7* locus.

SIMILARITIES BETWEEN COR AND ARC

The striking similarities between the genes in the COR region on chromosome X and the genes in the ARC region on chromosome V have been discussed by McKNIGHT et al. (1981). Both iso-1-cytochrome *c* and iso-2-cytochrome *c* which are encoded, respectively, by *CYC1* and *CYC7*, appear to carry out equivalent functions in mitochondrial oxidative phosphorylation (MATTOON & SHERMAN 1966; SHERMAN & STEWART 1971). One iso-cytochrome appears to be able to substitute for the other and no obvious growth differences are observed in strains when they contain the same amount of either iso-1-cytochrome *c* or iso-2-cytochrome *c*. Nevertheless, the presence of 95% of iso-1-cytochrome *c* and 5% of iso-2-cytochrome *c* in unrelated strains of *Saccharomyces* indicate that there is a selective advantage for maintaining both iso-cytochromes *c* and for maintaining each at their respective levels. Although the iso-cytochromes *c* perform equivalent or nearly equivalent functions, they differ in length (iso-1-cytochrome *c* and iso-2-cytochrome *c* are, respectively, 108 and 112 amino acids long) and they differ from each other by 21 amino acids. Furthermore, the *CYC1* and *CYC7* genes differ by 78 bp of the 327 bp coding region (MONTGOMERY et al. 1980). This approximately 80% homology supports the notion that *CYC1* and *CYC7* have a common evolutionary origin.

In contrast to the *CYC1* and *CYC7* genes, the *RAD7* and *RAD23* genes cannot substitute for each other in the normal excision
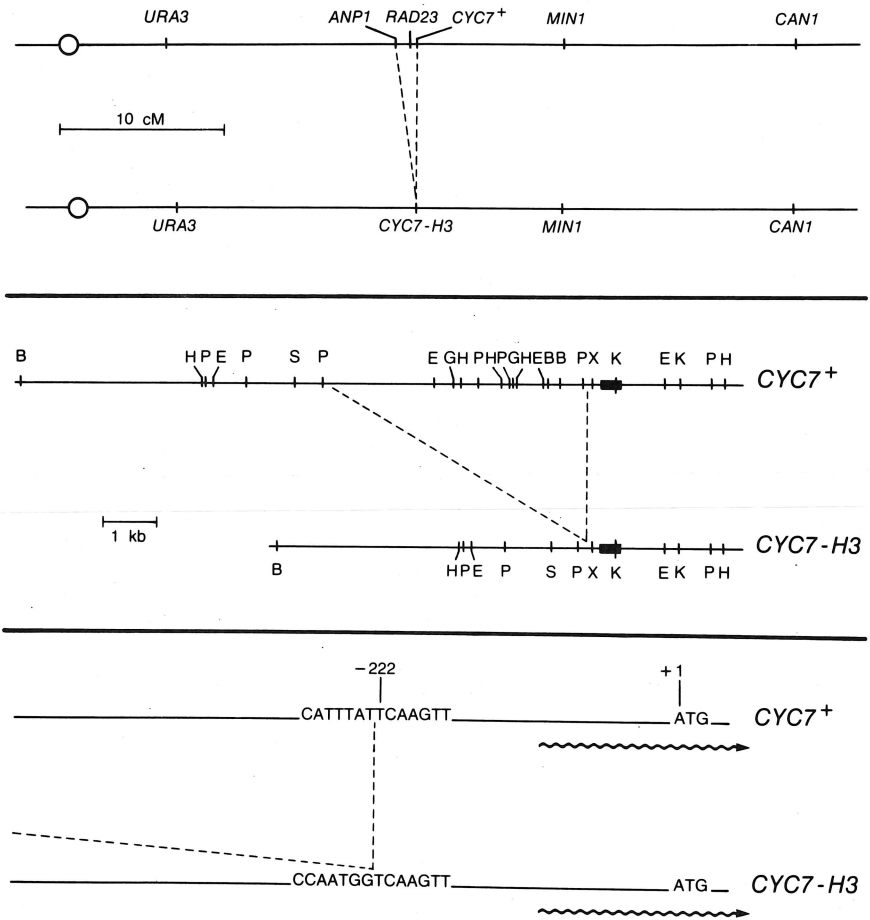
Fig. 5. Top. The genetic map of the left arm of chromosome V showing the deletion of the *ANP1* and *RAD23* genes and the formation of the *CYC7-H3* allele that causes overproduction of iso-2-cytochrome *c* (McKNIGHT et al. 1981). Middle. The physical structure of the normal *CYC7+* and mutant *CYC7-H3* regions. The translated portion of the *CYC7* locus is indicated by filled-in boxes. The sites of cleavages of the following restriction endonucleases are indicated: *Bam*HI, (B); *Hind*III, (H); *Pst*I, (P); *Eco*RI, (E); *Sal*I, (S); *Bgl*II, (G); *Xho*I, (X); and *Kpn*I, (K). The approximately 5 kb deletion is indicated by the dashed line (McKNIGHT et al. 1981). Bottom. The DNA sequence of the region encompassing the *CYC7-H3* breakpoint at the nucleotide position -222 and the DNA sequence of the corresponding normal region (KOSIBA et al. 1982). The *CYC7+* and *CYC7-H3* transcripts appear to be the same length (KOSIBA et al. 1982) and they may initiate around position -77 (MONTGOMERY et al. 1982).

of UV-induced damage (McKNIGHT et al. 1981). MILLER et al.
(1983) showed that *rad7* and *rad23* mutants are partially defective
in excising pyrimidine dimers and interstrand crosslinks in DNA
and that the *rad7 rad23* double mutant was more defective then
either of the single mutants. The study also suggested that
the *RAD7* and *RAD23* gene products are required for nicking damaged
DNA. Although it is still unclear if the *RAD7* and *RAD23* gene
products perform similar but not identical functions or if they
performed identical but incomplete functions, their common mode
of action is consistent with the view that they are evolution-
arily related.

The *OSM1* and *ANP1* gene products are both required for normal
osmotic sensitivity but the *ANP1* gene product is also required
for normal ANP sensitivity and for normal cell dispersion.
These different phenotypes of *osm1* and *anp1* mutants indicate that
the *OSM1* and *ANP1* gene products may be carrying out similar but
not identical functions.

In addition to the phenotypic similarities of mutants in
COR compared to mutants in ARC, both COR and ARC are approxi-
mately 6 kb long. However, the gene order of these three genes
in COR and ARC are not colinear with each other but can be re-
lated by a circular permutation as shown in Figure 6. The
rearranged order can be explained by the model proposed by
McKNIGHT et al. (1981), in which the primordial transposition
involved a circular intermediate. For example, if the COR
region was excised, formed a close circle, and subsequently
integrated by opening between *CYC1* and *OSM1*, the order of the
genes would change to the order found in ARC. A detailed com-
parison of the DNA sequences of COR and ARC and the orientations
of the *RAD7, RAD23* and *ANP1* and other genes within these regions
may clarify this hypothesis.

TRANSPOSITIONS OF COR2

The primordial transposition that is suggested to have
occurred during evolution superficially resembles transpositions
of COR regions that occur in certain laboratory strains. In
contrast to standard strains, certain laboratory strains were
found to give rise to unusually high frequencies of mutants
that either lacked or overproduced iso-1-cytochrome *c*. The var-
iation in the capacity to yield mutants has been attributed to
two polymorphic regions denoted COR1 and COR2 (STILES et al.
1981a,b). Although COR1 strains primarily yield mutants with
simple point mutations, COR2 strains yield mutants with extended
deletions, duplications and transpositions. The basis for this
difference resides in the difference in the physical structure
of COR1 and COR2 regions.

The COR2 region has been cloned by the site-specific inte-
gration of a plasmid and recovery of the plasmid along with ad-
jacent chromosomal material (STILES 1983). As shown in Figure
3, comparison of the restriction maps of the COR regions indi-
cates that COR2 strains contain an insertion of 5.5 kb at the
proximal end of the COR region. The extra sequence in COR2 has

been subcloned and used to probe *Hin*dIII restriction fragments
of genomic DNA. The analysis indicates that the extra segment
in the COR2 region is a reiterated sequence which is represented
in the haploid genome in approximately 35 copies. An analysis
of cloned COR2 fragments blotted to nitrocellulose and probed
with a plasmid containing a portion of the Ty1 element (CAMERON
et al. 1979) established that the extra sequence at the proximal
end of the COR2 region is a member of the Ty class of transpos-
able elements.

   Genetic studies established that COR2 but not COR1 strains
gave rise to high spontaneous frequencies of deletions, de-
noted *cyc1-Δ*, that encompassed the *CYC1*, *OSM1* and *RAD7* loci (STILES
et al. 1981a). Although different genetic backgrounds can
modify the efficiency of detection of *cyc1* mutations and the
rate of spontaneously arising *cyc1-Δ* deletions, some COR2 strains
yield *cyc1-Δ* deletions at 8×10^{-6} per cell per generation
whereas COR1 strains did not yield deletions or yielded dele-
tions at rates less than 10^{-18}. Over 98% of the *cyc1* mutants
derived from 49 different COR2 strains contained *cyc1-Δ* deletions
of the *CYC1*, *OSM1* and *RAD7* loci whereas more than 97% of the *cyc1*
mutants derived from 38 different COR1 strains contained point
mutations or at least retained the *OSM1* and *RAD7* loci. Further-
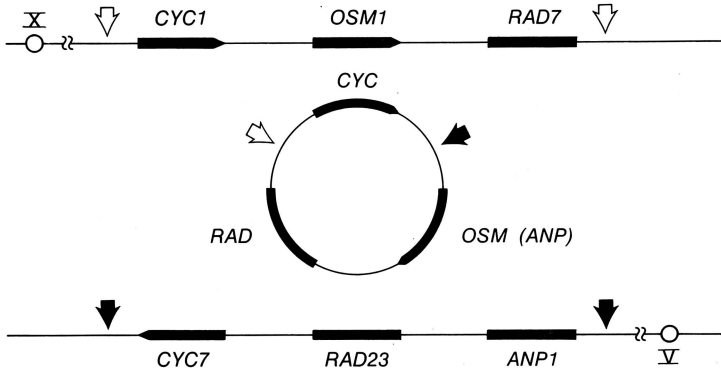more, there was no evidence from genetic tests that any of

Fig. 6. The hypothetical relationship between the COR and ARC
genes. The relative order of the *CYC1*, *OSM1* and *RAD7* genes on
chromosome X is shown at the top of the figure and the relative
order of the *CYC7*, *RAD23* and *ANP1* genes on chromosome V is shown at
the bottom of the figure. The direction of transcription of
the *CYC1*, *OSM1* and *CYC7* genes is indicated by pointed bars. The
gene orders are related by a circular permutation shown in the
center of the figure. The open and filled-in arrows indicate
the breakpoints that produce the order of the COR and ARC genes,
respectively (adapted from McKNIGHT et al. 1981).

the *cyc1-Δ* deletions extended into the *SUP4* locus which is closely
linked to the *RAD7* locus. COR2 strains appear to be identical
to *DEL1* strains (LIEBMAN et al. 1979), although COR2 and *DEL1*
studies were conducted with laboratory stocks not obviously
related.

The analysis of restriction fragments of genomic DNA that
hybridizes to COR1 probes indicated that various *cyc1-Δ* deletions
lack approximately 12 kb segments as shown in Figure 7. The
cyc1-Δ deletion end points are within or near the flanking Ty
elements.

COR2 strains also give rise to mutants that overproduce
iso-1-cytochrome *c* (STILES et al. 1981b). The overproduction is
due to increased copies of the COR region which, of course
results in increased copies of *CYC1* genes. Twenty-two
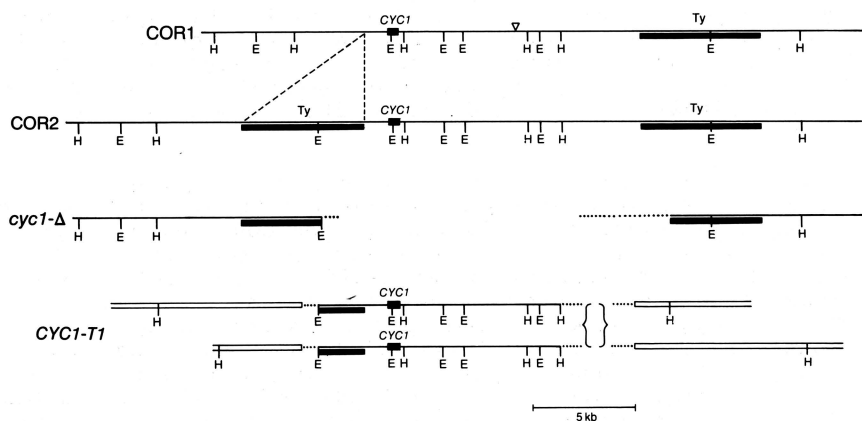mutants overproducing iso-1-cytochrome *c* were obtained after

Fig. 7. Physical structures of the COR1, COR2 *cyc1-Δ* and *CYC1-T1*
regions. Ty elements, the translated portion of the *CYC1* locus
and cleavage sites of restriction endocnucleases *Hin*dIII (H)
and *Eco*RI (E) are indicated. The triangle within the COR1 region
represents an approximately 300 base-pair segment found in COR1
but absent from COR2. The blank space denotes the material
absent from the represented *cyc1-Δ* deletion, *cyc1-1*; the dotted
lines denote material that is either deleted or homologous to
COR2 material. The *CYC1-T1* transposition contains two copies of
the COR2 region; the open lines denote flanking material from
either chromosome VII or from the 3' end of the COR2 region; the
dotted lines denote ambiguously assigned material; the brackets
indicate that connections of the end segments have not been
determined and that either of the two COR segments can be con-
tiguous to either of the two flanking segments (modified from
STILES et al. 1981a,b).

examining approximately 4.5 x 10^5 colonies from a COR2 diploid culture that was exposed to either x-ray or UV. Genetic analysis of 12 of these mutants suggested that overproduction could be ascribed to at least two classes of alterations denoted *CYC1-H* and *CYC1-T*. Alterations that are confined to the vicinity of the COR region and that do not involve rearrangements with other chromosomes are denoted *CYC1-H*. The *CYC1-H* mutations could be attributed to a number of different types of aberrations including tandem duplications of the COR segment. Transpositions of regions that encompass the *CYC1* gene and that are inserted at sites remote from the normal *CYC1* locus are denoted *CYC1-T*. So far we have uncovered eight *CYC1-H* mutations (*CYC1-H1*, *CYC1-H2*, etc.) and four *CYC1-T* mutations (*CYC1-T1*, *CYC1-T2*, etc.) from a COR2 strain. Mutants overproducing iso-1-cytochrome *c* have not been obtained in a limited number of experiments with COR1 strains.

Genetic analysis indicated that each of the four *CYC1-T* transpositions segregated as a single Mendelian unit independent of the normal *CYC1* locus. For example analysis of *cyc1 CYC1-T1* X *cyc1* crosses established that *CYC1-T1* is located approximately 15 cM from the centromere on the right arm of chromosome VII (STILES et al. 1981b). Further analysis indicated that *CYC1-T2*, *CYC1-T3* and *CYC1-T4* are not near the centromere of chromosome VII and are situated at sites distinct from *CYC1-T1* and the normal *CYC1* locus.

Complementation of the *cyc1*, *osm1* and *rad7* deficiencies by *CYC1-T1*, *CYC1-T2*, etc. established that the transpositions contain a region genetically equivalent to the COR region. For example, strains with the *cyc1-1* deletion are *cyc1⁻ osm1⁻ rad7⁻* whereas *cyc1-1 CYC1-T1* strains do not manifest these deficiencies. Measurements of the amounts of iso-1-cytochrome *c* in *cyc1 CYC1-T* and *CYC1⁺* segregants from the same pedigrees and measurements of the amounts in various diploid and triploid crosses indicated that the *CYC1-T1* transposition produced approximately two *CYC1⁺* equivalents of iso-1-cytochrome *c* (STILES et al. 1981b) and that the *CYC1-T2*, *CYC1-T3* and *CYC1-T4* produced approximately one equivalent. Thus it appears as if *CYC1-T* transpositions can include one or more copies of the COR region.

The physical structure of the *CYC1-T1* transposition has been investigated by hybridization of COR probes to restriction fragments of genomic DNA from *cyc1-Δ CYC1-T1* and *CYC1⁺ CYC1-T1* strains (STILES et al. 1981b). The *CYC1-T1* structure deduced from the restriction map, shown in Figure 7, is consistent with the measurements of iso-1-cytochrome *c* which indicated that *CYC1-T1* contains two copies of the *CYC1* gene. The analysis of restriction fragments showed that *CYC1-T1* is composed of two copies of the COR region including at least part of the proximal Ty element and that each of the two COR regions is equivalent to each other and to the COR region absent in *cyc1-Δ* deletions. Although the precise arrangement of the two COR regions in the *CYC1-T1* transposition is unknown, genetic analysis demonstrated that *CYC1-T1* segregated as a single unit. Current analysis of the cloned *CYC1-T1* transposition should reveal if the two COR regions in the *CYC1-T1* transposition are contiguous.

The *cyc1-Δ* deletions and the *CYC1-T1* transposition have end-points that lie within or near the two Ty elements that flank the COR region. We have suggested that these reiterated segments could promote *cyc1-Δ* deletions, *CYC1-H* duplications and *CYC1-T* transpositions (STILES et al. 1981a,b) by the mechanisms outlined in Figure 8. In a detailed analysis of *cyc1-Δ* deletions LIEBMAN et al. (1981) also emphasized the importance of the flanking Ty elements.

Although the *CYC1-T* transpositions observed in COR2 strains superficially resemble a transposition that could have occurred during evolutionary development of the COR and ARC regions, some differences should be noted. The *CYC1-T* transposition involves an approximate 12 kb COR region that retains the original order of the *CYC1*, *OSM1* and *RAD7* genes. Also the *CYC1-T* transpositions most likely have been inserted at sites having Ty homology. The COR and ARC regions, however, have a different order of similar genes (Fig. 6) and so far the similarities reside within 6 kb segments. The DNA sequences of the COR and ARC regions may reveal if Ty elements played a role in the primordial transposition.
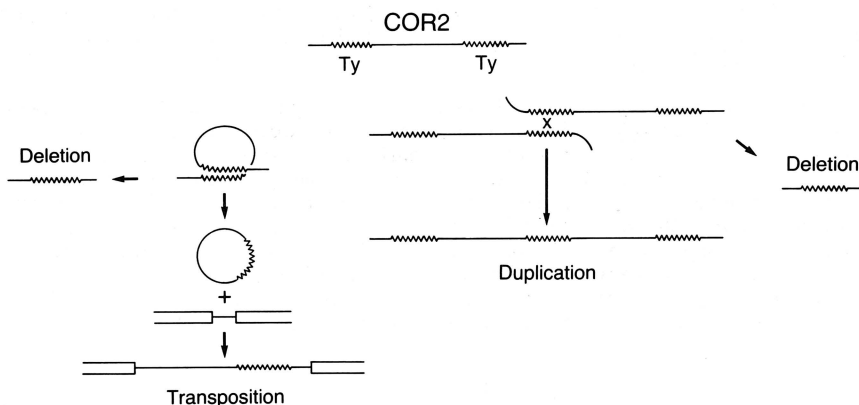

Fig. 8. A schematic representation of the COR2 region with the flanking Ty elements. Theoretically, deletions and duplication could be generated when Ty elements recombine by unequal exchanges of sister chromatids (right). Deletions also could be generated by recombination within a single chromatid (left). A transposition could arise by formation of a closed circle and integration of the circle at a new site (left). The order of the genes within the region would be altered if the circle opens at a different position during integration.

COMPOSITE GENES AND CONCERTED EVOLUTION

After a gene duplicates, it appears as if each duplicated locus may not necessarily evolve independently but may interchange sequences with each other by some recombinational process. This process has been denoted "concerted evolution" (ZIMMER et al. 1980) and has been suggested to occur in a wide range of eukaryotes (see JEFFREYS 1981; DOVER 1982; ERNST et al. 1982). Concerted evolution produces short stretches of extreme homology, similar to a recently described mutational process that leads to the formation of *CYC1-CYC7* composite genes (ERNST et al. 1981;1982).

Systematic reversion studies of numerous *cyc1* mutations revealed that approximately 90% of revertants arise by single base-pair changes and most of the remaining approximately 10% arise by two or sometimes three base-pair changes. In addition, at least some revertants of certain *cyc1* mutants contained composite genes in which the *cyc1* lesion was corrected by replacement of the central portion of the *CYC1* locus with a corresponding portion from the *CYC7* locus. Thus there are two distinct pathways leading to revertants that are schematically illustrated in Figure 9; intragenic reversion usually occurs at higher frequencies by single base-pair changes; composite genes occur at low frequencies by recombinational events between the *CYC1* and *CYC7* loci.

The examination of 549 revertants from over 20 different types of *cyc1* mutants indicated that the composite genes were formed by reversion of the five mutants *cyc1-136*, *cyc1-158*, *cyc1-11*, *cyc1-94* and *cyc1-156*. The results, summarized in Table 1, were obtained by examining iso-1-cytochrome *c*, or in some instances, the DNA sequence at the *CYC1* locus of the revertant strains (ERNST et al. 1981;1982). The revertants containing composite genes were obtained in haploid strains and arise spontaneously or were induced with either UV, methylmethane sulfonate, ethylmethane sulfonate, diethyl sulfate or 1-nitro-
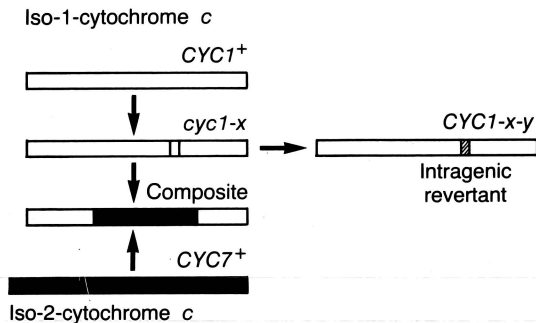
Iso-1-cytochrome *c*

CYC1⁺

Fig. 9. Schematic representation of the mutational pathways leading to either intragenic revertants or composite genes.

Table 1. The frequencies of revertants with composite genes

Allele	*cycl* mutation Amino acid position	Type	Reversion frequency*	Number of revertants Composite genes	Total analyzed
cycl-136	64	Frameshift	Low	7	7
cycl-158	62	Frameshift	Low	3	7
cycl-11	76	UAA	Low	3	11
cycl-94	79	UAA	Normal	1	14
cycl-156	26	UAA	Normal	1	19
cycl-9	2	UAA	high	0	84
cycl-31	3	Frameshift	Low	0	27
cycl-183	10	Frameshift	Normal	0	68
cycl-72	66	UAA	Normal	0	62
Others		Various		0	250

*Frequencies of revertants arising spontaneously or induced by various physical chemical agents. Normal refers to the reversion frequencies observed for most single base-pair substitutions; low and high refer to distinctly lower and higher frequencies, respectively. From ERNST et al. (1981;1982).

soimidazolidone-2. Composite genes occurred in most revertants from the *cycl-11*, *cycl-136* and *cycl-158* mutants that revert at low frequencies and in only a few revertants from the *cycl-94* and *cycl-156* mutants that revert at frequencies typical for single base-pair changes. The reason for the low reversion of the *cycl-11*, *cycl-136* and *cycl-158* mutations appears to be a rarity of single base-pair changes required to form functional iso-1-cytochromes *c*. The *cycl-11* mutation contains a two base-pair substitution, changing the proline 76 codon CCA to the nonsense codon UAA. Reversion of the *cycl-11* mutation can occur both by rare A·T → C·G transversions, producing serine replacements, and by the formation of composite genes. The frameshift mutations *cycl-136* and *cycl-158* contain deletions of, respectively, one base-pair at amino acid position 64 and two base-pairs at

amino acid positions 62 or 63.   These frameshift mutations are
within or near the codon for tryptophan 64 which has an essen-
tial and unique role for maintaining the normal structure of
iso-1-cytochrome c although a limited number of replacements at
position 64 leads to at least partially functional proteins
(SCHWEINGRUBER et al. 1979).   The low frequencies of intragenic
reversion of the cycl-136 and cycl-158 mutants appear to be due to
the limited pathways of base-pair changes required to generate
functional iso-1-cytochromes c and a rarity of the required
changes.   In contrast, the two UAA mutants cycl-94 and cycl-156
revert at frequencies typical of other mutants that revert by
single base-pair substitutions.

Seventeen composite structures were determined by protein
analysis of 14 composite iso-cytochromes c and by DNA sequencing
of five composite genes (ERNST et al. 1981;1982).   These 17
composite genes could be assigned to 8 different types shown in
Figure 10.   Because portions of the amino acid sequences and
DNA sequences of iso-1-cytochrome c and iso-2-cytochrome c are
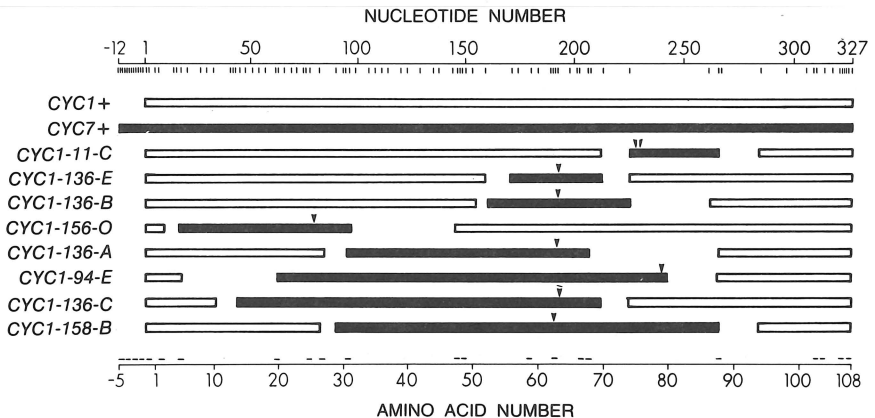completely identical, the termini of the replaced segments were

NUCLEOTIDE NUMBER

-12 1   50   100   150   200   250   300 327

CYC1+
CYC7+
CYC1-11-C
CYC1-136-E
CYC1-136-B
CYC1-156-O
CYC1-136-A
CYC1-94-E
CYC1-136-C
CYC1-158-B

-5 1   10   20   30   40   50   60   70   80   90   100 108

AMINO ACID NUMBER

Fig. 10.   Composite CYC1-CYC7 genes and composite iso-cytochrome c.
The nucleotide and amino acid numbering systems are for iso-1-
cytochrome c.   The nucleotides that are different in CYC1+ and
CYC7+ are indicated by vertical dashes at the top of the figure;
the amino acids that are different in iso-1-cytochrome c and iso-
2-cytochrome c are indicated by horizontal dashes at the bottom
of the figure.   The composite genes or proteins have iso-2-
cytochrome c material (filled-in bars) flanked by iso-1-cyto-
chrome c material (open bars).   The results are from DNA sequences
and represent CYC1-CYC7 composite genes except for the CYC1-156-O,
CYC1-136-A and CYC1-94-E results which are from protein sequences
and represent composite iso-cytochrome c.   The sites of the ori-
ginal cycl mutations are indicated by arrows (adapted from ERNST
et al. 1982).

equivocal and could be assigned only to regions spanned by dif-
ferent amino acid residues or nucleotides. The smallest *CYC7*
replacement, represented by *CYC1-136-E*, was 13% to 20% of the
total *CYC1* length whereas the longest, represented by *CYC1-158-B*,
was 54% to 61%. The endpoints of the replaced segments are not
random; 16 out of the 17 composite genes have one endpoint
within a highly homologous region encompassing nucleotides 208
through 305. Also 9 out of 17 of the composite genes corre-
sponded to the two types exemplified by *CYC1-136-B* and *CYC1-136-E*;
the two endpoints of these types are situated within positions
154 to 170 and positions 213 to 260, respectively, which have
high homology between *CYC1* and *CYC7*. The clustering of endpoints
of the replaced *CYC7* segments suggest that greater homology
enhances exchanges between the nonallelic genes. The unusual
composite gene, *CYC1-156-O*, does not have a *CYC7* endpoint within a
highly homologous region and also as a *CYC7* segment that extends
further toward the front of the gene.

Genetic analysis established that the composite genes are
produced by a recombinational process resembling mitotic gene
conversion and not by reciprocal exchanges. The replacement of
a portion of the *CYC1* locus with a homologous portion of the *CYC7*
locus conceivably could occur either by gene conversion of the
segment or by a double reciprocal exchange illustrated in Figure
11. No defective *cyc7* alleles were uncovered after testing the
17 characterized composite genes as well as 42 additional un-
characterized composite genes from the *cyc1-136* and *cyc1-158*
mutants. Genetic tests also revealed that the composite genes
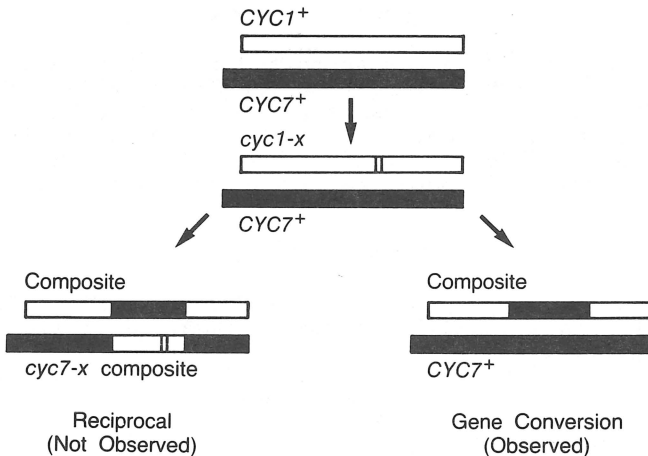were not associated with reciprocal translocations.

Fig. 11. Schematic representation of the formation of composite
genes. The expected structures of the *CYC7* alleles are shown
when the composite genes arise by either a reciprocal exchange
or by gene conversion.

Thus composite genes are formed by gene conversion of short segments without crossing over. Higher proportions of composite genes are detected among revertants when the *cyc1* mutations revert at lower frequencies. Composite genes are not observed among the revertants of most *cyc1* mutants because the reversion by base-pair changes mask their occurrence and because the converted segment may not necessarily encompass the *cyc1* lesion.

The structure of composite genes, containing short stretches of complete homology between nonallelic genes, resembles the structure of numerous duplicated genes in various eukaryotes. Such a region of perfect homology is illustrated by the segments encompassing nucleotides 226 through 260 in the *CYC1* and *CYC7* genes (Figure 10). The gene conversional process leading to composite genes may be identical to the process that occurs during evolution. Because of small stretches of complete or nearly complete homology are found in most duplicated genes, concerted evolution appears to be a common evolutionary process.

CONCLUDING REMARKS

The evolutionary pathways proposed in this and our earlier papers and outlined in Figures 2 and 6 should be clarified after transcriptional analysis and DNA sequencing of the entire COR and ARC regions. Nevertheless the hypotheses provide an incentive for investigating these regions in detail. In addition, further investigations may reveal the evolutionary relationship of material between genes and the differences in regulation of homologous genes, such as that observed for the *CYC1* and *CYC7* loci (MATNER & SHERMAN 1982).

LITERATURE CITED

BOSS, J. M., S. GILLAM, R. S. ZITOMER and M. SMITH 1981 Sequence of the yeast iso-1-cytochrome *c* mRNA. J.Biol.Chem. *256*: 12958-12961.
CAMERON, J. R., E. Y. LOH, and R. W. DAVIS 1979 Evidence for transposition of dispersed repetitive DNA families in yeast. Cell *16*: 739-751.
DOVER, G. 1982 A molecular drive through evolution. Bio. Science *32:* 526-533
DOWNIE, J. A., J. W. STEWART, N. BROCKMAN, A. M. SCHWEINGRUBER and F. SHERMAN 1977 Structural gene for yeast iso-2-cytochrome *c*. J. Mol. Biol. *113:* 369-384.

ERNST, J. F., J. W. STEWART and F. SHERMAN 1981 The *cyc1-11* mutation in yeast reverts by recombination with a nonallelic gene: Composite genes determining the iso-cytochromes *c*. Proc. Nat. Acad. Sci. USA *78:* 6334-6338.

ERNST, J. F., J. W. STEWART AND F. SHERMAN 1982 Formation of composite iso-cytochromes *c* by recombination between nonallelic genes of yeast. J. Mol. Biol. *161:* 373-394.

ERREDE, B., T. S. CARDILLO, F. SHERMAN, E. DUBOIS, J. DESCAMPS 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.

ERREDE, B., T. S. CARDILLO, G. WEVER and F. SHERMAN 1981 Studies on transposable elements in yeast. I. ROAM mutations causing increased expression of yeast genes: Their activation by signals directed toward conjugation functions and their formation by insertion of Ty1 repetitive elements. Cold Spring Harbor Symp. Quant. Biol *45:* 593-602.

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. Nat. Acad. Sci. USA *78:* 2258-2262.

JEFFREYS, A. J. 1981 Recent studies of gene evolution using recombinant DNA Pp 2-48. Genetic Engineering (Williamson, R., Ed.) vol. 2 Academic Press, London, New York.

KOSIBA, B. E., B. ERREDE, T. S. CARDILLO and F. SHERMAN 1982 Mutants of yeast overproducing iso-2-cytochrome *c*. Recent Advan. Yeast Mol. Biol. (Lawrence Berkeley Lab, Univ. of Calif.) *1:* 156-172.

LAWRENCE, C. W., F. SHERMAN, M. JACKSON and R. A. GILMORE 1975 Mapping and gene conversion studies with the structural gene for iso-1-cytochrome *c* in yeast. Genetics *81:* 615-629.

LIEBMAN, S. W., A. SINGH, and F. SHERMAN 1979 A mutator affecting the region of the iso-1-cytochrome *c* gene in yeast. Genetics *92:* 783-802.

LIEBMAN, S. W., P. SHALIT and S. PICOLOGLOU 1981 Ty elements are involved in the formation of deletions in *DEL1* strains of *Saccharomyces cerevisiae*. Cell *26:* 401-409.

MATNER, R. R. and F. SHERMAN 1982 Differential accumulation of two apo-iso-cytochromes *c* in processing mutants of yeast. J. Biol. Chem. *257:* 9811-9821.

MATTOON, J. R. and F. SHERMAN 1966 Reconstitution of phosphorylating electron transport in mitochondria from a cytochrome *c*-deficient yeast mutant. J. Biol. Chem. *241:* 4330-4338.

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.

MILLER, R. D., L. PRAKASH and S. PRAKASH 1983 Defective excision of pyrimidine dimers and interstrand DNA crosslinks in *rad7* and *rad23* mutants of *Saccharomyces cerevisiae*. Mol. Gen. Genetics (in press).

MONTGOMERY, D. L., B. D. HALL, S. GILLAM and M. SMITH 1978 Identification and isolation of the yeast cytochrome *c* gene. Cell *14:* 673-680.

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. Nat. Acad. Sci USA *77*: 541-545.

MONTGOMERY, D. L., J. M. BOSS, S. F. McANDREW, L. MARR, D. A. WALTHALL and R. S. ZITOMER 1982 The molecular characterization of the three transcriptional mutations in the yeast iso-2-cytochrome *c* gene. J. Biol. Chem. *257*: 7756-7761.

SCHWEINGRUBER, M. E., J. W. STEWART and F. SHERMAN 1979 Primary site and second site revertants of missense mutants of the evolutionarily invariant tryptophan 64 in iso-1-cytochrome *c* from yeast. J. Biol. Chem. *254*: 4132-4143.

SHALIT, P., K. LOUGHNEY, M. V. OLSON and B. D. HALL 1981 Physical analysis of the *CYC1-sup4* inverval in *Saccharomyces cerevisiae*. Mol. Cell Biol. *1*: 228-236.

SHERMAN, F. and C. HELMS 1978 A chromosomal translocation causing overproduction of iso-2-cytochrome *c* in yeast. Genetics *88*: 684-707.

SHERMAN, F. and J. W. STEWART 1971 Genetics and biosynthesis of cytochrome *c*. Ann. Rev. Genetics *5*: 257-296.

SHERMAN, F. and J. W. STEWART 1973 Mutations at the end of the iso-1-cytochrome *c* gene of yeast. Pp. 56-86. The Biochemistry of Gene Expression in Higher Organisms (Pollak, J. K., and J. W. Lee, Eds.) Australian and New Zealand Book Co., Sydney.

SHERMAN, F., J. W. STEWART, E. MARGOLIASH, J. PARKER and W. CAMPBELL 1966 The structural gene for yeast cytochrome *c*. Proc. Nat. Acad. Sci. USA *55*: 1498-1504.

SHERMAN, F., M. JACKSON, S. W. LIEBMAN, A. M. SCHWEINGRUBER and J. W. STEWART 1975 A deletion map of *cyc1* mutants and its correspondence of mutationally altered iso-1-cytochromes *c* of yeast. Genetics *81*: 51-73.

SHERMAN, F., J. W. STEWART, C. HELMS, and J. A. DOWNIE 1978 Chromosome mapping of the *CYC7* gene determining yeast iso-2-cytochrome *c*: Structural and regulatory regions. Proc. Nat. Acad Sci. USA *75*: 1437-1441.

SINGH, A. and F. SHERMAN 1978 Deletions of the iso-1-cytochrome *c* and adjacent genes of yeast: Discovery of the *osm1* gene controlling osmotic sensitivity. Genetics *89*: 653-665.

SMITH, M., D. W. LEUNG, S. GILLAM, C. R. ASTELL, D. L. MONTGOMERY, and B. D. HALL 1979 Sequence of the gene for iso-1-cytochrome *c* in *Saccharomyces cerevisiae*. Cell *16*: 753-761.

STEWART, J. W. and F. SHERMAN 1974 Yeast frameshift mutations identified by sequence changes in iso-1-cytochrome *c*. Pp. 102-207. Molecular and Environmental Aspects of Mutagenesis. (Prakash, L., F. Sherman, M.W. Miller, C.W. Lawrence and H.W. Taber, Eds.) C.C. Thomas, Springfield.

STILES, J. E., L. R. FRIEDMAN and F. SHERMAN 1981a Studies on transposable elements in yeast II. Deletions, duplications and transpositions of the COR segment that encompasses the structural gene of yeast iso-1-cytochrome *c*. Cold Spring Harbor Symp. Quant. Biol. *45*: 602-607.

STILES, J. I., L. R. FRIEDMAN, C. HELMS, S. CONSAUL and F.
    SHERMAN 1981b Transposition of the gene cluster, *CYC1-OSM1-
    RAD7*, in yeast. J. Mol. Biol. *148:* 331-346.
STILES, J. I., J. W. SZOSTAK, A. T. YOUNG, R. WU, S. CONSAUL
    and F. SHERMAN 1981c DNA sequence of a mutation in the
    leader region of the yeast iso-1-cytochrome *c* mRNA.
    Cell *25:* 277-284.
STILES, J. I. 1983 Use of integrative transformation of yeast
    in the cloning of mutant genes and large segments of con-
    tiguous chromosomal sequences. Meth. Enzymol. *100:* (in
    press).
SZOSTAK, J. W., J. I. STILES, C. P. BAHL and R. WU 1977 Specific
    binding of a synthetic oligodeoxyribonucleotide in yeast
    cytochrome *c* mRNA. Nature *265:* 61-63.
SZOSTAK, J. W., J. I. STILES, B.-K. TYE, P. CHIU, F. SHERMAN
    and R. WU 1979 Hybridization with synthetic oligonucleo-
    tides. Methods Enzymol. *68:* 419-428.
ZARET, K. S. and F. SHERMAN 1982 DNA sequence required for
    efficient transcription termination in yeast. Cell *28:* 563-
    573.
ZIMMER, E.A., S. L. MARTIN, S. M. BEVERLEY, Y. W. KAN and A. C.
    WILSON 1980 Rapid duplication and loss of genes coding for
    the α chains of hemoglobin. Proc. Nat. Acad. Sci. USA *77:*
    2158-2162.