MOLECULAR GENETIC ANALYSIS OF THE *SACCHAROMYCES CEREVISIAE MAT*α LOCUS

By

SUSAN DOROTHY PORTER

B.Sc. (Hons.), The University of New Brunswick, 1980

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF BIOCHEMISTRY

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October, 1987

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Biochemistry

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date Dec. 30 '87

ABSTRACT

The *MAT*α locus of the yeast *Saccharomyces cerevisiae* encodes two regulatory proteins responsible for determining the α cell type. The *MAT*α*1* gene encodes α1, a positive regulator of α cell-specific genes, whereas the *MAT*α*2* gene encodes a negative regulator of *a* cell-specific genes (α2). *MAT*α*2* (in conjunction with the *MATa1* gene) also determines the *a*/α diploid cell type by repressing haploid-specific genes. α2 exerts its effect at the transcriptional level in the α cell by binding to a sequence located upstream of *a* cell-specific genes.

The present study undertook to examine, through *in vitro* genetic manipulation, the structure/function relationship of the *MAT*α regulatory proteins, particularly α2, in their role as gene regulators. The construction of mutant *MAT*α*2* genes containing termination codons at various points within the gene, and subsequent transformation of the mutant genes into *mat*α*2* yeast, indicated that the carboxy-terminal one-third of the gene product was necessary for full repressor activity in the haploid as well as in the diploid.

A segment within the carboxy-terminal one-third of α2 displays some homology to the higher eukaryote homeo domain as well as to a prokaryotic bihelical DNA-binding structural motif. This region of the gene was subjected to semi-random missense mutagenesis *in vitro* and the mutant genes were analyzed by transformation into strains containing chimaeric genes that encode β-galactosidase from α2 and a1/α2 repressible promoters.

In this manner it was demonstrated that most of those residues in α2 which correspond to conserved amino acids in the prokaryotic DNA-binding structure and in the homeo domain are essential for the two repressor activities of α2. Several mutations more severely affected the ability of α2 to repress *a*-specific genes than haploid-specific genes.

Analysis of the temperature dependence of the activities of some of the mutants was consistent with the existence of a helix-turn-helix structure at this region of the protein. Finally, further analysis

of some of these mutants *in vitro* confirmed that the observed defect correlated with a loss of DNA-binding activity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Dr. Michael Smith, for the opportunity to pursue this project, and for his encouragement and advice along the way. I thank the members of my supervisory committee, Drs. Caroline Astell and Anthony Griffiths, for their interest and instructive criticism. I am particularly grateful to the members of the Smith lab, past and present, for their helpful discussions and friendship, and especially thank Andrew Spence, Johnny Ngsee, Caroline Beard, Bryan McNeil, and David Goodin. I am indebted to David Goodin, as well, for the sharing of his time and knowledge of computers. I thank Dr. Ira Herskowitz for providing important strains, plasmids and advice. Last but definitely not least, thank-you Blake.

LIST OF ABBREVIATIONS

a1	the *MATa1* gene product
α1	the *MATα1* gene product
α2	the *MATα2* gene product
bp	base pair
BSA	bovine serum albumin
ddNTP	2',3'-dideoxynucleoside triphosphate
dNTP	2'-deoxynucleoside triphosphate
DTT	dithiothreitol
IgG	immunoglobulin G
kb	kilobase
MOPS	3-(N-morpholino)-2-hydroxypropanesulfonic acid
PAGE	polyacrylamide gel electrophoresis
ONPG	orthonitrophenylgalactoside
SDS	sodium dodecylsulfate
SSC	1 x SSC = 0.15 M NaCl, 0.015 M sodium citrate
SSPE	1 x SSPE = 0.15 M NaCl, 0.010 M sodium phophate, 0.001 M EDTA, pH 7.4
X-gal	5-bromo-4-chloro-3 indolyl-β-D-galactoside

INTRODUCTION

THE *MAT* LOCI

S. cerevisiae has three distinct cell types: two haploid cell types (*a* and α), which are capable of mating with each other to form the third, diploid, *a*/α cell type, which is non-mating but which is capable of meiosis and sporulation. The two partially nonhomologous alleles *a* and α, of a single regulatory locus, *MAT*, located on chromosome III, are responsible for the determination of each of these cell types (Mortimer and Hawthorne, 1969).

An additional aspect of the yeast life cycle is the ability (termed homothallism) of most naturally occurring yeast strains to interconvert between mating types. This is accomplished by the transfer of genetic information from one of two normally unexpressed *MAT* loci, *HML*α or *HMRa*, to the expressed *MAT* locus and requires the *HO* gene product (Hicks *et al.*, 1979; Hicks & Herskowitz, 1976; Nasmyth & Tatchell, 1980)

MacKay and Manney (1974a,b) and Strathern *et al.* (1981) first identified two complementation groups within the *MAT*α locus. By analysis of individual and combined mutant phenotypes, the *MAT*α*1* complementation group was deduced to positively regulate a series of unlinked genes responsible for the physiological aspects specific to α cells (α-specific genes), whereas the *MAT*α*2* complementation group was proposed to negatively regulate a normally constitutively expressed set of genes responsible for the *a* cell type (*a*-specific genes). The *MAT*α*2* gene and the *MATa1* gene (Kassir and Simchen, 1976), were shown to be necessary for determination of the diploid cell type. It was proposed that the combination of these two genes allowed the derepression (or activation) of a set of diploid-specific genes.

Hicks *et al.* (1979) and Nasmyth and Tatchell (1980) isolated recombinant plasmids containing mating type genes. This was accomplished by the transformation of either a *mata1* (Hicks *et al.*) or *mat*α*1* (Nasmyth and Tatchell) strain with a total yeast genomic recombinant library and the selection of transformants able to mate as α cells. Hicks *et al.* isolated *HML*α, and Nasmyth and Tatchell, *MAT*α, in

this manner. The DNA sequences of the *MATa* and *HMLα* loci were determined by Astell *et al.* (1981), and the location and direction of transcripts specific to the *MAT* loci were established by Nasmyth *et al.* (1981) and Klar *et al.* (1981) through the use of Northern blots and RNA-DNA heteroduplex analysis. Each *MAT* locus was shown to produce two divergently transcribed RNA species.

The previously defined genetic complementation groups were then localized to specific transcripts within each *MAT* locus by the use of *in vitro* linker mutagenesis and *in vivo* complementation analyses (Tatchell *et al.*, 1981). Linker mutagenesis (Heffron *et al.*, 1978) produced an array of disruptions ranging from deletions to insertions throughout the loci, and the transformation of the resultant clones into strains containing *MATα1*, *MATα2*, or *MATa1* defects enabled a genetic to physical correlation to be made. This is indicated schematically in Figure 1. Specifically, the rightwardly-transcribed *MATa* transcript was shown to correspond to the *MATa1* gene, the right *MATα* transcript to *MATα1* and the left *MATα* transcript to *MATα2*. Mutations in the leftwardly-transcribed *MATa* transcript (termed *MATa2*) exhibited no observed phenotypic disruptions. [*MATa2* in the genetically related yeast *Saccharomyces diastaticus*, however, has been shown to negatively regulate the expression of extracellular glucoamylase, in conjunction with *MATα2*, at the post-transcriptional level (Yamashita *et al.*, 1985).]

Regulation of unlinked genes by the mating type loci

Many of the propositions concerning the function of the mating type genes suggested from genetic analyses, illustrated schematically in Figure 2, have been confirmed by more recent molecular studies.

In the case of *MATα1*, the role as a positive regulator of α-specific genes was proposed because *MATα1* mutants are sterile and exhibit none of the characteristics of either mating type. That is, no α-factor (a pheromone produced by α cells and required for mating with *a* cells) is produced, and no barrier function (the ability to degrade α-factor externally - a characteristic of *a* cells) is observed (Strathern *et al.*, 1981). This model of *MATα1* function was further supported through the construction of *matα1/matα2* double mutants (described later). More recently, the molecular cloning of two α-

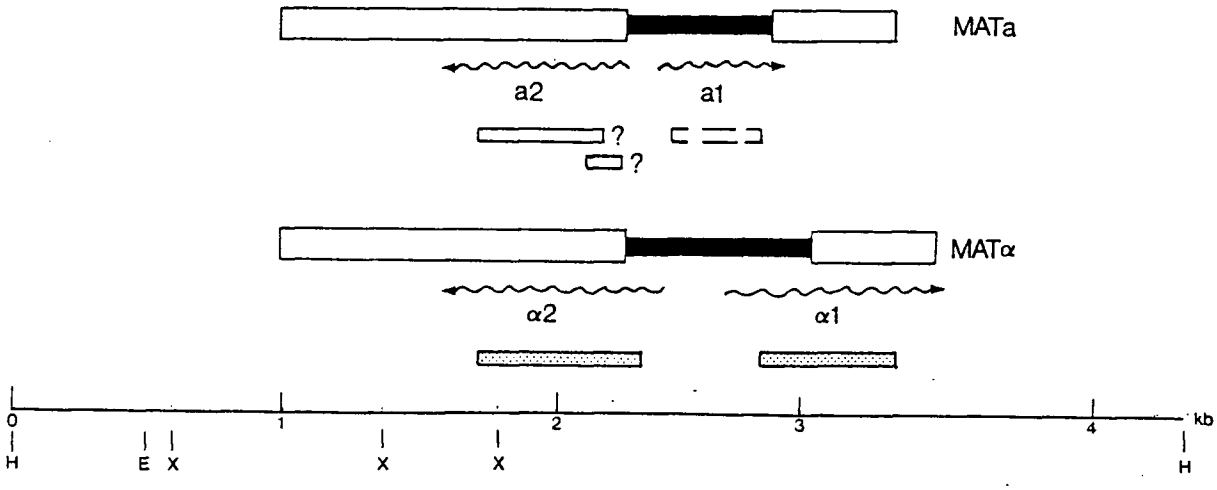


Figure 1. Organization of the Yeast *MAT* Loci.

The arrangement of transcripts and putative gene products of the *MATa* and *MATα* loci is indicated (Tatchell *et al.*, 1981; Astell *et al.*, 1981). Wavy lines indicate transcripts and bars represent gene products. The two gaps in the *MATa1* gene product indentify the two small introns in the *MATa1* gene (Miller *et al.*, 1985b). The *MATa2* transcript codes for two putative gene products ("?"), neither of which has been identified, genetically or biochemically. Open space within the loci represents DNA common to both *MAT* alleles, whereas the darkened area represents DNA unique to each one. Restriction site abbreviations are as follows: H, HindIII; E, EcoRI; X, XbaI.

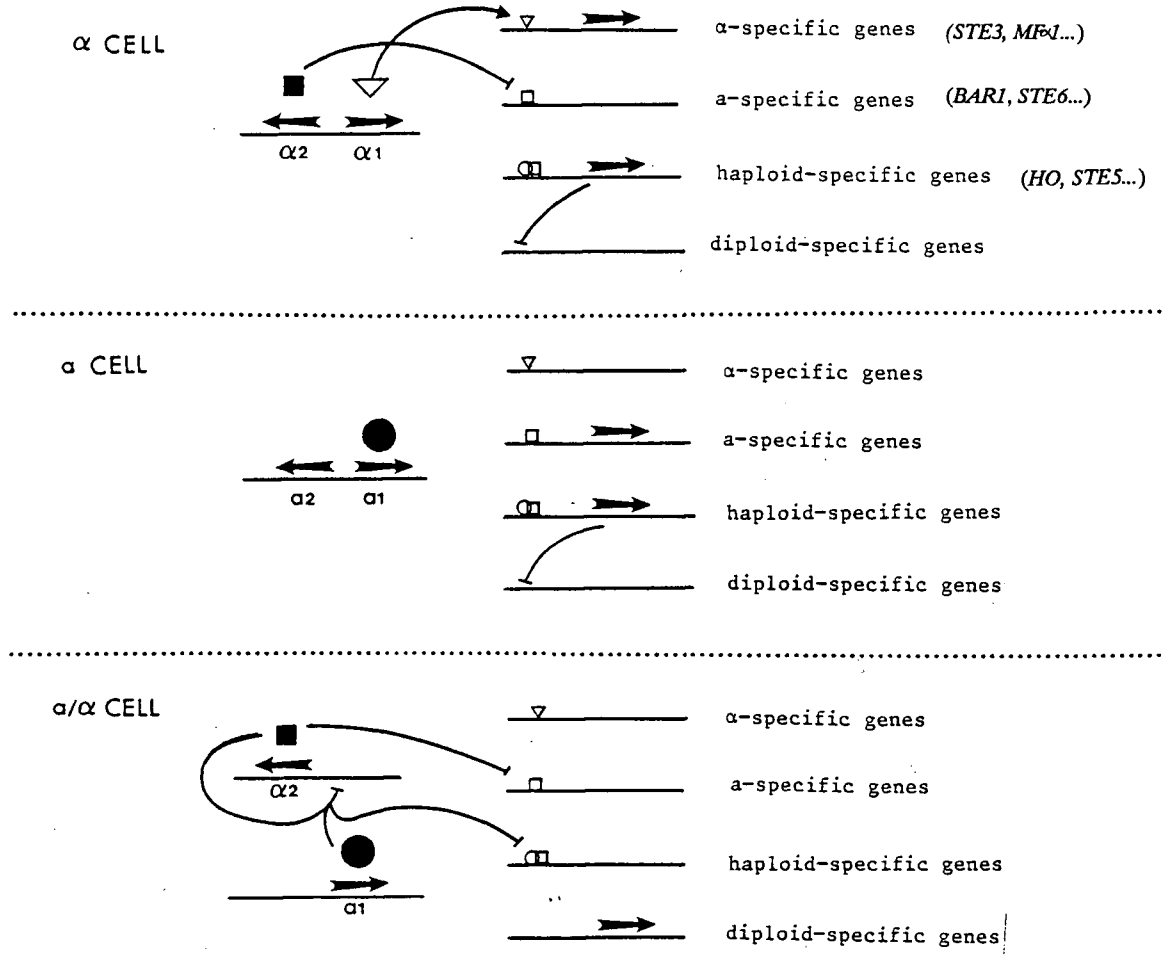


Figure 2. Regulatory Functions of the Mating Type Loci.

The products of the mating type locus in the cell, 1 and 2, regulate sets of unlinked -specific genes or *a*-specific genes, respectively. 1 is a positive regulator and 2 is a negative regulator. In the *a* cell type, the *a*-specific genes are constitutively expressed. A subset of the haploid-specific genes, which are needed and expressed in both mating types, are thought to negatively regulate certain diploid-specific genes. In the diploid cell type, 2 continues to turn off *a*-specific genes, as well as, in conjunction with a1, haploid-specific genes. Diploid-specific genes are then expressed.

specific genes has enabled a direct test of this model. The *STE3* gene [thought to code for the *a*-factor receptor (Hagen *et al.*, 1986)] and *MFα1* gene [an α-factor structural gene (Kurjan and Herskowitz, 1982)] have been shown to be controlled at the transcriptional level by the *matα1* gene product (α1) (Sprague *et al.*, 1983b). It is not known, however, if this control is direct or if intermediate factors are involved. There are several additional loci, identified by sterile mutants, that have been shown to be necessary for the expression of *a*-specific as well as α-specific genes [*STE7, STE11, STE12* and in some cases *STE4* and *STE5* (Fields and Herskowitz, 1985)]; these may encode factors acting in concert with α1.

Mutations in *MATα2* exhibit a distinct phenotype (MacKay and Manney, 1974a,b). Although *MATα2* mutants do not mate with *a* cells, they do mate poorly with α cells. In addition, they display the *a*-specific function of barrier (Bar^+). Thus, it appears that the defect in mating in *MATα2* mutants is caused by the simultaneous expression of *a*- and α-specific characteristics, and that the *MATα2* gene product (α2), in haploids, acts as a repressor of the constitutive *a*-specific characteristics. This was confirmed by the construction of *matα1/matα2* double mutants (Strathern *et al.*, 1981), which mated in all respects as *a* cells. This genotype is called Alf (for *a*-like faker) and appears to arise from the lack of expression of α-specific genes (due to the absence of α1) and the continued expression of the *a*-specific genes (due to the absence of α2). The role of α2 as a repressor has been studied in detail at the molecular level. The *STE6* gene (Wilson and Herskowitz, 1984), required in *a* cells for mating activity, is negatively regulated by α2 at the transcriptional level. In addition, evidence for a direct interaction between α2 and *a*-specific genes has been provided by the work of Johnson and Herskowitz (1985). This will be discussed in detail in a later section.

The requirement for both *MATα2* and *MATa1* in sporulation was demonstrated by the inability of either *MATα2* (Strathern *et al.*, 1981) or *MATa1* (called *mata**; Kassir and Simchen, 1976) mutants to support sporulation in a diploid. Mutations in *MATα1* had no effect. The existence of the recessive mutations *sca* (Gerlach, 1974), *csp1* (Hopper and Hall, 1975) and *rme* (Kassir and Simchen, 1976) that allowed sporulation in the absence of functional *MATα2* or *MATa1* alleles led to the proposal that

sporulation is under negative control and that the function of α2/a1 in promoting sporulation may be to repress these or other negative regulators. Verification of the role of a1 and α2 in transcriptional repression in diploids was provided by the demonstration of a dependence of negative transcriptional regulation of the *RME* gene (Mitchell and Herskowitz, 1986), the *HO* gene (Jensen *et al.*, 1983), and the *MATα1* gene (Nasmyth *et al*, 1981 and Klar *et al.*, 1981) on functional *MATa1* and *MATα2* alleles. Because this set of genes (*RME*, *HO*, etc.) is expressed in the haploid cell type only, they are referred to as haploid-specific genes.

THE α2 PROTEIN

Several studies have focussed on the structural and functional aspect of the activity of yeast α2 as a gene repressor. Reasons for this attention include: 1) α2 is a eukaryotic regulator of transcription. Although there is an increasing amount of research in this area (see, for example, Gluzman, 1985), little is known at present about *trans*-acting factors involved in eukaryotic gene expression, particularly at the molecular level. An understanding of the mechanism of negative regulation of a specific set of genes by a single, characterized, gene product would provide a useful model for the understanding of eukaryotic transcriptional regulation. 2) Yeast is an ideal organism for the purpose of studying eukaryotic gene regulation at the molecular level. The ease of genetic manipulation in yeast allows the rapid isolation of genes of interest, their manipulation *in vitro*, and their replacement back into the organism under identical conditions. A short doubling-time and simplicity of structure also contribute to the wide-spread use of yeast in genetic research. Of the genetic loci in yeast, those responsible for mating type determination have been most extensively studied. For this reason, the understanding of the function of the *MAT* loci in regulating gene expression at the genetic level is far advanced. Thus, the combination of an extensive genetic groundwork in an easily manipulatable organism has provided impetus for the further molecular analysis of *MATα2*. 3) α2 is a complex regulator in that it has two distinct functions, dependent on cell type. It appears to act in a combinatorial fashion with at least one

other gene regulator (a1), and thus may represent a simple prototype of the complex genetic mechanisms presumed to be necessary for development and differentiation in higher eukaryotes.

Methods so far employed in the analysis of α2 and its function as a repressor include mutagenesis (Tatchell *et al.*, 1981), subcellular localization through immuno-fluorescence (Hall *et al.*, 1984), *in vitro* DNA-binding analysis (Johnson and Herskowitz, 1985) and primary amino acid sequence comparison with other regulatory proteins (Shepherd *et al.*, 1984; Laughon and Scott, 1984).

Functional analysis of α2 by genetic manipulation and cytological studies

From DNA sequence data, the α2 gene product is proposed to be a 210 amino acid, basic protein (Astell *et al.*, 1981; see Figure 3). As discussed above, genetic evidence has led to the proposal that the function of the protein is the direct negative transcriptional regulation of a diverse set of unlinked genes. This was first tested biochemically by determining the cellular location of the gene product (Hall *et al.*, 1984). Various portions of the *MATα2* gene were fused to the *E. coli* β-galactosidase gene and the hybrid genes were transformed into yeast. The β-galactosidase activity was followed by cytological immunofluorescence assay and subcellular fractionation. In this way, it was demonstrated that α2 directed the localization of the fusion product to the nucleus, and that only the first 13 amino acid residues were necessary for this function. The sequence lys-ile-pro-ile-lys, located within this amino terminal region, was thought to be important, as similar sequences are found within the yeast nuclear proteins histones H2B (Wallis *et al.*, 1980), H2A (Choe *et al.*, 1982), and H4 (Smith and Andresson, 1983), and the presumed nuclear protein *MATα1* (Astell *et al.*, 1981). The yeast nuclear proteins GAL4 (Laughon and Gesteland, 1984), HO (Russell *et al.*, 1986) and histone H3 (Smith and Andresson, 1983), however, do not contain the sequence.

The next major step to an understanding of the mechanism of α2 action was the demonstration by Johnson and Herskowitz (1985) that α2 is a sequence-specific DNA-binding protein. This was accomplished, as well, through the use of a *MATα2*-β-galactosidase fusion. A construction (Hall *et al.* 1984) containing the entire *MATα2* coding sequence fused at the carboxy-terminus to the β-galactosidase coding sequence, was integrated into the yeast genome at the *MAT* locus. The fusion

```
                                        (M)  N   K   I   P   I   K   D   L   L (10)
  AAGAAAAAAAGGAAGATAAGCAAGAAAAA ATG AAT AAA ATA CCC ATT AAA GAC CTT TTA
  TTCTTTTTTTCCTTCTATTCGTTCTTTTT TAC TTA TTT TAT GGG TAA TTT CAG GAA AAT

 N   P   Q   I   T   D   E   F   K   S   S   I   L   D   I   N   K   K (28)
AAT CCA CAA ATC ACA GAT GAG TTT AAA TCC AGC ATA CTA GAC ATA AAT AAA AAG
TTA GGT GTT TAG TGT CTA CTC AAA TTT AGG TCG TAT GAT CTG TAT TTA TTT TTC

 L   F   S   I   C   C   N   L   P   K   L   P   E   S   V   T   T   E (46)
CTC TTT TCT ATT TGC TGT AAT TTA CCT AAG TTA CCA GAG AGT GTA ACA ACA GAA
GAG AAA AGA TAA ACG ACA TTA AAT GGA TTC AAT GGT CTC TCA CAT TGT TGT CTT

 E   E   V   E   L   R   D   I   L   V   F   L   S   R   A   N   K   N (64)
GAA GAA GTT GAA TTA AGG GAT ATA TTA GGA TTC TTA TCT AGG GCC AAC AAA AAC
CTT CTT CAA CTT AAT TCC CTA TAT AAT CCT AAG AAT AGA TCC CGG TTG TTT TTG

 R   K   I   S   D   E   E   K   K   L   L   Q   T   T   S   Q   L   T (82)
CGT AAG ATT AGT GAT GAG GAG AAG AAG TTG TTG CAA ACA ACA TCT CAA CTC ACT
GCA TTC TAA TCA CTA CTC CTC TTC TTC AAC AAC GTT TGT TGT AGA GTT GAG TGA

 T   T   I   T   V   L   L   K   E   M   R   S   I   E   N   D   R   S (100)
ACT ACC ATT ACT GTA TTA CTC AAA GAA ATG CGC AGC ATA GAA AAC GAT AGA AGT
TGA TGG TAA TGA CAT AAT GAG TTT CTT TAC GCG TCG TAT CTT TTG CTA TCT TCA

 N   Y   Q   L   T   Q   K   N   K   S   A   D   G   L   V   F   N   V (118)
AAT TAT CAA CTT ACA CAG AAA AAT AAA TCG GCG GAT GGG TTG GTA TTT AAT GTG
TTA ATA GTT GAA TGT GTC TTT TTA TTT AGC CGC CTA CCC AAC CAT AAA TTA CAC

 V   T   Q   D   M   I   N   K   S   T   K   P   Y   R   G   H   R   F (136)
GTA ACT CAA GAT ATG ATA AAC AAA AGT ACT AAA CCT TAC AGA GGA CAC CGG TTT
CAT TGA GTT CTA TAC TAT TTG TTT TCA TGA TTT GGA ATG TCT CCT GTG GCC AAA

 T   K   E   N   V   R   I   L   E   S   W   F   A   K   N   I   E   N (154)
ACA AAA GAA AAT GTC CGA ATA CTA GAA AGT TGG TTT GCA AAG AAC ATC GAG AAC
TGT TTT CTT TTA CAG GCT TAT GAT CTT TCA ACC AAA CGT TTC TTG TAG CTC TTG

 P   Y   L   D   T   K   G   L   E   N   L   M   K   N   T   S   L   S (172)
CCA TAT CTA GAT ACC AAG GGC CTA GAG AAT CTA ATG AAG AAT ACC AGT TTA TCT
GGT ATA GAT CTA TGG TTC CCG GAT CTC TTA GAT TAC TTC TTA TGG TCA AAT AGA

 R   I   Q   I   K   N   W   V   S   N   R   R   R   K   E   K   T   I (190)
CGC ATT CAA ATC AAA AAC TGG GTT TCG AAT AGA AGA AGA AAA GAA AAA ACA ATA
GCG TAA GTT TAG TTT TTG ACC CAA AGC TTA TCT TCT TCT TTT CTT TTT TGT TAT

 T   I   A   P   E   L   A   D   L   L   S   G   E   P   L   A   K   K (208)
ACA ATC GCT CCA GAA TTA GCG GAC CTC TTG AGC GGT GAG CCT CTG GCA AAG AAG
TGT TAG CGA GGT CTT AAT CGC CTG GAG TAC TCG CCA CTC GGA GAC CGT TTC TTC

 K   E   *
AAA GAA TGA GCCCGAAAAACAAATATGTATATATCTGTGTAGAATATATATATATATATTTCGCAAAA
TTT CTT ACT CGGGCTTTTTGTTTATACATATATAGACACATCTTATATATATATATATAAAGCGTTTT

ATACATAAACAATCAACCCTCTCCTCAGACACTACTAAGATGTTTG
TATGTATTTGTTAGTTGGGAGAGGAGTCTGTGATGATTCTACAAAC
```

Figure 3. DNA and Amino Acid Sequence of *MATα2*.

The DNA sequence of *MATα2*, starting at the most 5' transcript initiation site and ending at the most 3' transcript termination site, is displayed (Astell *et al.*, 1981).

protein was partially purified from a cell extract of the transformant. Four β-galactosidase-positive peaks were obtained upon separation on DNA-cellulose. These were postulated to consist of tetramers of β-galactosidase with none, one, three or four complete α2 moieties attached. The α2 portion of the isolates containing less than the complete α2 was thought, from the mobility on SDS polyacrylamide gels, to consist of only the carboxy-terminal 25-30 amino acids of the polypeptide. Proteolysis producing these truncated species occurred only upon isolation from yeast and was postulated to be a natural degradative pathway for α2.

The complete tetrameric fusion protein was mixed with a Hae III digest of a recombinant plasmid containing the complete *STE6* gene and immunoprecipitated with anti-β-galactosidase IgG. A single DNA fragment, corresponding to a region approximately 200bp 5' to the structural gene, co-precipitated with the complex. The segment of DNA within this fragment to which the fusion protein bound was further localized to an approximately 34 bp region by the method of DNase I protection. This sequence, synthesized chemically and placed within the yeast *CYC1* promoter, conferred *MATα2*-dependent regulation on the promoter *in vivo*. Thus, the *in vitro* binding of α2-β-galactosidase to this sequence has identified the target for *in vivo* transcriptional regulation by α2.

Information relating the function of α2 to the structure of the gene product has, thus far, been derived only from the study (Tatchell *et al.*, 1981) that correlated the transcripts of *MAT* with known genetic complementation groups. Random linker gene disruption was followed by transformation of the mutated genes on high-copy-number plasmids into yeast strains containing conventionally isolated *MATα2* mutants. Large deletions or frame shifts within the left transcript produced gene products which were unable to complement the resident *matα2* allele. There were two exceptions, however, that displayed interesting phenotypes. Mutation αx111 is an insertion of 10bp that causes a frameshift at codon 80 of the 210-codon gene and results in a truncated protein with the addition at this point of 9 C-terminal missense amino acids. This mutant, when present on a high copy number plasmid, is capable of complementing the sporulation deficiency of either a *MATa/matα2* or a *MATa/MATa* strain, yet is unable to complement (or very poorly complements) a *matα2* allele in promoting mating with *a* cells.

Mutation α1x75, on the other hand, has the reciprocal phenotype of partially complementing the mating, but not the sporulation, activity of a *matα2* allele. (It is able to allow a low level of sporulation in an *a/a* cell.) This mutation results in an insertion of 17 bp at the termination codon of the gene and would be expected to give rise to a gene product containing an additional 28 amino acid residues.

One explanation for the phenotypes of the two mutants is that the amino terminal third of the protein is sufficient for the sporulation, but not the mating, activity of α2, and that mutation α1x75 caused secondary structure alterations that disrupted the N-terminal (sporulation) but not the C-terminal (mating) functions of the protein. It should be noted that another mutant, α1x38, containing a frameshift mutation that would result in a truncated protein containing 14 more of the wild-type residues than α1x111, is defective in both mating and sporulation. Thus, the interpretation of the relationship of the phenotypes of these mutants in terms of the structure of α2 is not a simple matter. The conventionally isolated allele *matα2-4* (MacKay and Manney, 1974) displays a similar sporulation$^+$/mating$^-$ phenotype to mutant α1x111, however α1x75 is the first example of the reciprocal, sporulation$^-$/mating$^+$, phenotype.

Two additional comments may be made concerning these *MATα2* mutants. First, both α1x111 and α1x75 are non-functional in all respects when present in the cell in a single copy (i.e., integrated into the genome; K. Tatchell, personal communication). Thus, although the phenotypes of the mutants on high copy number plasmids may provide some information regarding the function of the α2 molecule, the activity observed may not be physiologically significant. Secondly, the fact that some mutations in α2 affect one activity and not the other (mating or sporulation) implies that the two functions of the protein are separate and therefore that the mechanisms of haploid-specific gene repression and *a*-specific gene repression are distinct. This possibility provides further reason for examining the molecular mechanisms in more detail both at the level of the target DNA and in the α2 protein. It also has implications for the model of α2 as a prototype for higher eukaryote regulators of gene expression.

Homology with the eukaryotic homeo domain and with the bihelical prokaryotic DNA-binding domain

i. The eukaryotic homeo domain

Research into the developmental processes of the fruitfly *Drosophila melanogaster* has resulted in an accumulation of a vast amount of genetic information regarding cellular (or segmental) determination in this organism. Within the past few years, this information has led to the beginnings of a more detailed analysis at the molecular level (see Gehring and Hiromi, 1986). The following is a brief description of the genetics of segmental development and the discovery of the highly conserved homeo box.

The organization of the *Drosophila* embryo into discrete segmental units is controlled by three sets of genes. Maternally active genes specify the dorso-ventral and antero-posterior axes of the embryo, the segmentation genes determine the number and polarity of the segments, and the homeotic genes are responsible for the specification of segment identity. E.B. Lewis (1978) proposed a strategy of segment determination whereby genes from the homeotic loci are activated in a stepwise fashion in each segment in an anterior to posterior direction. The specific set of homeotic 'selector' genes expressed in each segment is responsible for the expression of the appropriate 'realisator' genes that ultimately give each segment its identity (Garcia-Bellido, 1977). Although, in a general sense, this still is a valid model, a somewhat more complex model is now envisioned. This is due to the observed interactions between various homeotic genes (Struhl & White, 1985; Hafen *et al.*, 1984).

The role of the segmentation genes in establishing the developmental pattern of the embryo is perhaps less well understood. It seems that a very complex, spatially-restricted, hierarchichal expression of segmentation genes controls the boundaries, number and polarity of individual segments (see, for example, Weir and Kornberg, 1985; Ingham *et al.*, 1985). The initial pattern is apparently set up directly or indirectly by some maternally-controlled gradient of chemical information (Mlodzik and Gehring, 1987), although specifically localized cytoplasmic determinants are also involved (Frohnhofer and Nusslein-Volhard, 1986).

The existence of large deletion or rearrangement mutations in the two major homeotic gene clusters, the *bithorax* complex (BX-C) and the *Antennapedia* complex (ANT-C), has enabled the molecular cloning of several of the loci involved in segment determination (Bender *et al.*, 1983; Garber *et al.*, 1983; Scott *et al.*, 1983). During the process of the physical mapping of the *Antp* locus, located within ANT-C, DNA cross-hybridization was observed with another portion of the complex, later found to be part of the *fushi tarazu* (*ftz*) transcription unit, as well as with the *ultrabithorax* (*Ubx*) domain within BX-C (McGinnis *et al.*, 1984; Scott and Weiner, 1984). This region of DNA homology, termed the homeo box, was limited to approximately 180bp, encoding about 60 amino acids, and mapped to the 3' exon of each of the loci. DNA sequence determinations revealed that the extent of homology at the DNA level was 48% to 81%, with a corresponding protein sequence homology of 38% to 88% (Gehring and Hiromi, 1986). More than 10 additional homeo boxes within the *Drosophila* genome have now been isolated, and at least three sub-sets of homeo domain (the protein sequence encoded by the homeo box) have been shown to exist by the criterion of sequence homology. Although most of the homeo boxes occur in homeotic genes, several segmentation genes [*ftz* (Laughon and Scott, 1984), *eve* (MacDonald *et al.*, 1986), *prd* (Frigerio *etal.*, 1986) and *en* (Poole *et al.*, 1985)] and a maternal gene (Mlodzik *et al.*, 1985) also contain homeo boxes. Most dramatically, Southern analysis and subsequent gene isolation have identified homologous sequences in mouse, man, frog, and sea urchin genomes (for review, see Gehring & Hiromi, 1986). All of the homeo box-containg genes analyzed to date are expressed during development or in specific cell-types, and thus may play analogous roles in development and/or differentiation in all these animal species.

ii. The helix-turn-helix prokaryote DNA-binding structure

The structural determination of four prokaryotic DNA-binding regulatory proteins, phage λ repressor and Cro, *E. coli* catabolite activator protein (CAP) and the tryptophan repressor (Anderson *et al.*, 1981; Pabo & Lewis, 1982; McKay & Steitz, 1981; Schevitz *et al.*, 1985) has led to a model in which binding of the proteins to DNA is mediated by a common structural motif consisting of two α-helices separated by a turn. Data from the crystallographic structural determination of λ Cro at 2.8 Å

resolution (Anderson *et al.*, 1981) was used for the model of the protein-DNA complex illustrated in Figure 4. The model is supported by genetic and biochemical data (see Pabo and Sauer, 1984), and now as well as by the recent high resolution structural determination of the bacteriophage 434 repressor DNA-binding domain (which contains the helix-turn-helix motif) co-crystallized with its operator (Anderson *et al.*, 1987). As this represents the most informative data on the helix-turn-helix interaction with DNA, these results will be described here.

The 434 repressor is a 290 amino acid protein, and, like λ repressor has two structural domains and binds as a dimer to its 14 bp, bilaterally symmetrical operator. The operator exists in the protein-DNA structure as right-handed B-DNA that is overwound (with respect to the average) at the center and underwound at the ends. The helix axis is slightly bent towards the protein. This combination of alterations results in narrowed minor grooves in the center and widened minor grooves at the ends.

The DNA-binding domain of 434 repressor (the amino-terminal 69 residues) is largely α-helical. Helices α2 and α3 represent the consensus helix-turn-helix motif and fit into the DNA structure in a manner basically similar to that predicted for the other protein structures. α3 lies in the major groove and is connected by a short turn to α2, which is approximately perpendicular to the α3 helix. The N-termini of both helices point toward the DNA backbone. DNA backbone contacts are thought primarily to be made by either the peptide backbone groups or side chains of the residues of the turn between α3 and α4 and of the amino-terminus of α2. Contacts with base-pairs are made by the amino-terminal two residues and the sixth residue of α3 (residues 28, 29, and 33) through either hydrogen bonds or van der Waals contact.

Comparison of the amino acid sequence of the helix-turn-helix-containing DNA-binding proteins indicates the sequences comprising this structure contain a number of conserved residues. A number of other phage and bacterial regulatory proteins, whose three-dimensional structures are not known, share this homology. These are, thus, thought to have an analogous structure and to function in

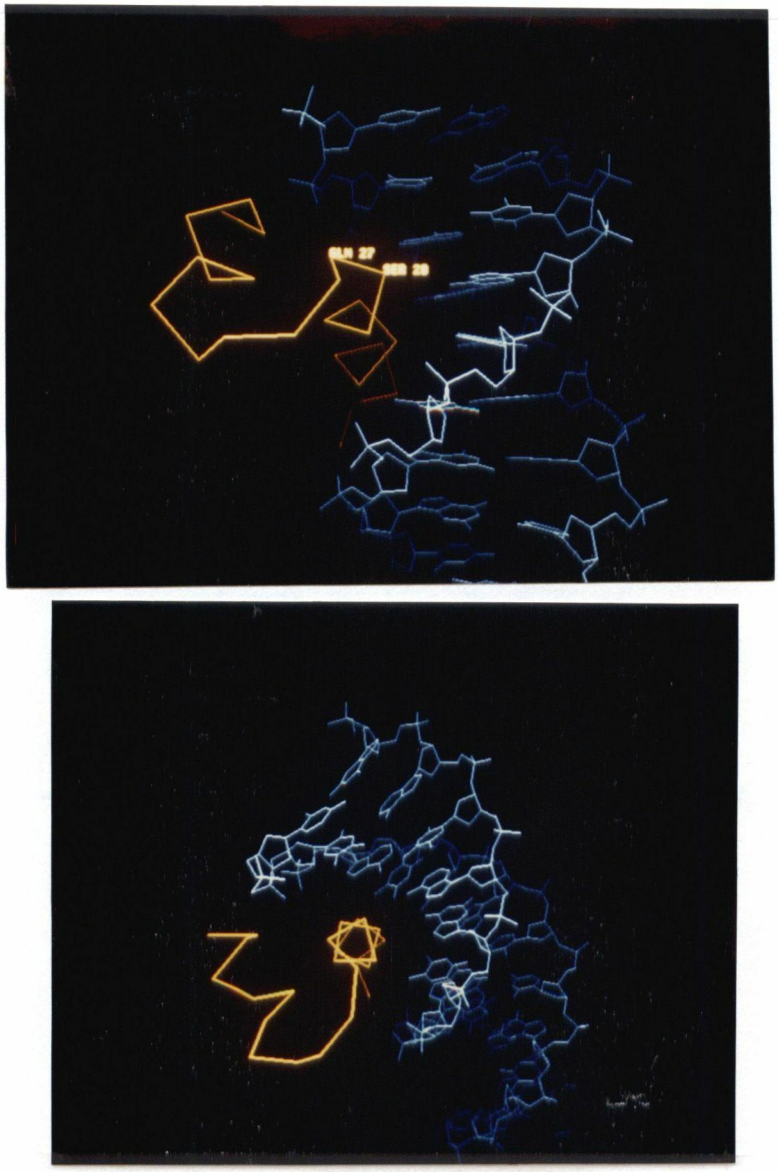


Figure 4. Model for the Interaction of λCro Protein with Operator DNA.

The λCro monomer is a 66 amino acid protein containing three strands of β-sheet and three α-helices (Anderson *et al.*, 1981). The carbon backbone of residues 15-38, comprising the helix-turn-helix DNA-binding region (yellow), in a model with its operator DNA (blue) are indicated from two perspectives. (Top) The view is perpendicular to the DNA axis. The carbon backbone of the helix-turn-helix proceeds from the upper left (N-terminus) to the lower right (C-terminus). The N-terminal two residues of the C-terminal helix are indicated (Gln-27, Ser-28) (Bottom). The view is down the axis of the C-terminal helix from the amino-terminal end.

a similar fashion (Sauer *et al.*, 1982; Ohlendorf *et al.*, 1983). A comparison of the putative helix-turn-helix DNA-binding domains of several regulatory proteins and their conserved characteristics is diagrammed in Figure 5. In particular, residues corresponding to the fifth position of the N-terminal helix (Ala), the first residue of the turn (Gly) and residue 4 (Ile or Val) of the C-terminal helix are almost invariant in this group of proteins. Residues 5 of the N-terminal helix and 4 of the C-terminal helix are in van der Waal's contact in the known structures, and could be important in maintaining the proper orientation of the two helices. The tight turn between the two helices is facilitated by the conserved Gly [although it is not absolutely required, as a double mutant of λ repressor containing a Gly to Glu change at this residue is functional (Hochschild *et al.*, 1983)]. In addition to these highly conserved residues, the chemical character of several other residues in the helix-turn-helix is maintained. Because the two helices are on the external surface of the protein, there is a characteristic pattern of hydrophobic (internal) and hydrophilic (external) residues corresponding to the 3.6 residues/turn of the α helix. That is, residues 4, 5 and 8 of the N-terminal helix, and residues 4, 7 and 8 of the C-terminal helix tend to be hydrophobic in character, whereas most of the remaining helical residues tend to be hydrophilic. There is a large amount of biochemical and genetic data for some of the proteins containing this homologous amino acid sequence (Lac repressor, λ *cII*, P22 repressor, and Tn10 Tet repressor) that supports the existence of the helix-turn-helix DNA-binding motif (see Pabo and Sauer, 1984).

iii. Homology of the *MATα2* gene product with the eukaryote homeo domain and with the prokaryote helix-turn-helix sequence.

A search of the Dayhoff protein sequence bank (of 2372 sequences) for homology to a consensus homeo domain sequence revealed that the greatest homologies occurred in the gene products of *MATα2* and *MATa1* (Shepherd *et al.*, 1984). As illustrated in Figure 6, although only 18 amino acids of the 60 amino acid homeo domain are conserved in α2, 11 of an 18 residue block near the carboxy-terminus of the homeo domain are conserved. This block, which is also the most conserved region of the different homeo domain sequences, contains many of the sequence elements pertaining to

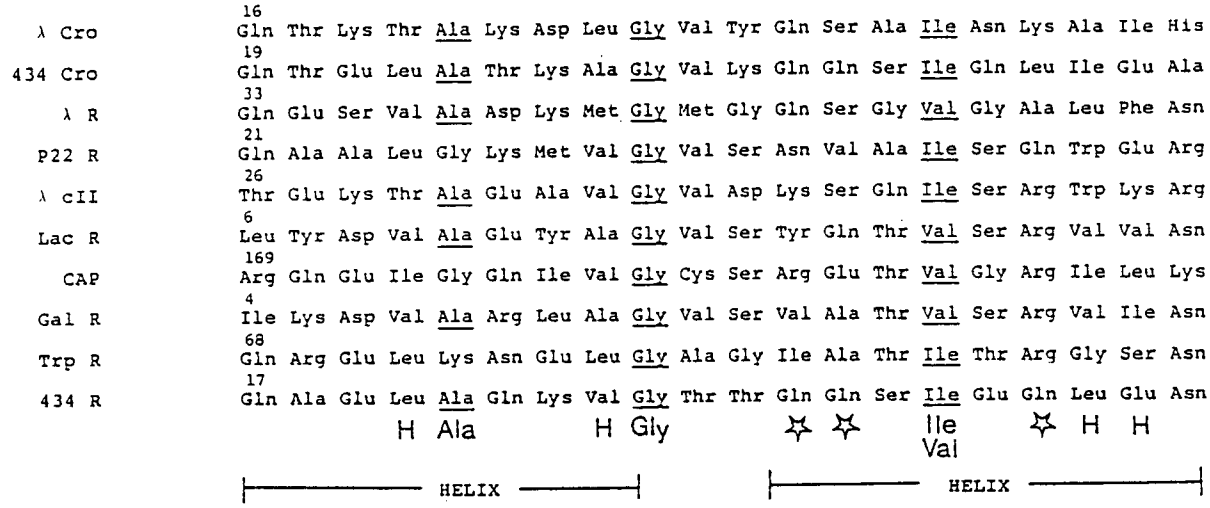


Figure 5. A Comparison of Prokaryotic Transcriptional Regulatory Protein Sequences Homologous to the Helix-Turn-Helix Motif.

Segments of the protein sequence of nine prokaryotic regulatory proteins homologous to the helix-turn-helix motif as described in λ Cro, repressor and *E. coli* CAP are indicated (Ohlendorf *et al.*, 1983). Helices correspond to the two α-helices of the conserved helix-turn-helix structure. Conservation of a hydrophobic residue is indicated by H, and * represents those residues expected to contact DNA.

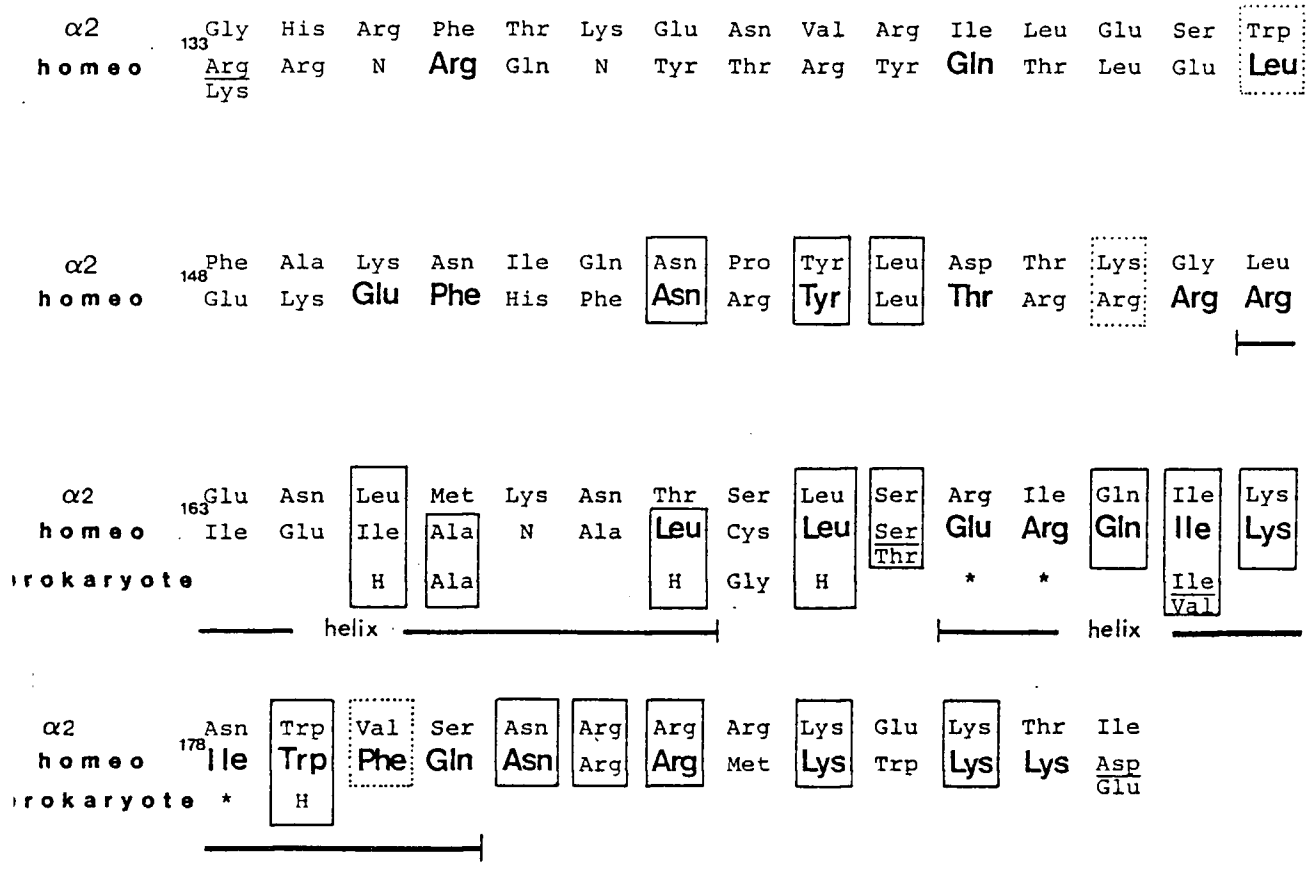


Figure 6. Comparison of Consensus Homeo Domain and Prokaryotic DNA-Binding Domain Sequences With α2.

The sequence of α2 is compared with the homeo domain consensus sequence as well as with conserved features of the prokaryotic helix-turn-helix structure. Residues highly conserved within the homeo domain are in bold type, and those conserved with α2 are boxed (dotted boxes represent conservative changes). The features of the helix-turn-helix structure illustrated are hydrophobic residues (H), conserved Ala, Gly or Ile/Val residues, and those residues postulated to contact DNA (*).

the conserved helix-turn-helix structure (Laughon & Scott, 1984). In particular, in the segment of α2 corresponding to the C-terminal helix, the Ile at position 4 and the hydrophobic nature of residues 7 and 8 are conserved (the remaining residues in this segment are hydrophilic). In the region corresponding to the N-terminal helix, the Ala at position 5 and the hydrophobic nature of 8, as well as the Gly in the turn, however, are not conserved. The consensus homeo domain has a better homology with the prokaryotic DNA-binding N-terminal helix consensus than does α2.

The homology between the homeotic domain and the helix-turn-helix sequence motif suggests a structural and functional analogy (Laughon and Scott, 1984). This is supported, in the case of α2, by the demonstration that it is a site-specific DNA-binding protein. In the case of the *Drosophila* homeo-domain-containing proteins, a finding that is consistent with a DNA-binding regulatory role of the proteins is that all of those tested by antibody-localization studies accumulate in the nucleus (White and Wilcox, 1984; Beachy *et al.*, 1985; Carroll and Scott, 1985; DiNardo *et al.*, 1985; Carroll *et al.*, 1986; Wirz *et al.*, 1986; Macdonald and Struhl, 1986; Mlodzik and Gehring, 1987) Furthermore, a portion of the *en* gene product containing the homeo domain exhibits sequence-specific DNA-binding activity (Desplan *et al.*, 1985).

REGULATORS OF TRANSCRIPTION

It is the purpose of this study to extend our understanding of the nature of the interactions between α2 and its target DNA; how, in molecular terms, the protein binds its "operator", what determines the specificity of binding and, eventually, how it is able to prevent transcription. A great many studies have focussed on these same aspects of prokaryotic transcriptional regulators, resulting in a very detailed understanding of several proteins (see Pabo and Sauer, 1984). There have been relatively few detailed analyses of eukaryotic regulators (particularly negative regulators), however, because of the need for detailed genetic analysis of the regulatory systems amid the greater inaccessibility of eukaryotic genes. For this reason, it is instructive to first examine the prokaryotic

models of transcriptional regulation, and subsequently to relate this to what is known of the regulation of eukaryote gene transcription.

Molecular aspects of the regulation of prokaryotic transcription

Analysis of the molecular aspects of prokaryotic DNA-binding and transcriptional regulation has been directed at protein:DNA and protein:protein interactions. Certain mechanistic themes have resulted by which proteins bind to genes and alter their rate of transcription. The following sections describe these themes, as exemplified in a small number of well-characterized cases. An overview of the prokaryotic promoter is first presented.

i. The prokaryotic promoter and its regulation by DNA-binding proteins

In the prokaryote promoter, two conserved DNA sequences, located at approximately 10 bp and 35 bp upstream from the transcription initiation site (Hawley and McClure, 1983), are the constitutive elements to which RNA polymerase binds (Pribnow, 1975; Schaller *et al.*, 1975). Negative regulatory elements generally overlap the polymerase binding site, and the repressors bound to these sites prevent transcription by physically blocking access to the promoter by the polymerase (Majors, 1975). As will be discussed, there appears also in some cases to be an element of long range protein interactions through DNA loops (see Ptashne, 1986), although the significance of these interactions in repressing transcription is not known.

The DNA element required for transcriptional activation in prokaryotes is generally situated close to the polymerase binding site, and, as suggested for the λ repressor, may enable the activator protein bound at that site to contact RNA polymerase and thereby increase the transcription rate (Meyer and Ptashne, 1980; Guarente *et al.*, 1982).

ii. Protein:DNA interactions of prokaryotic regulatory proteins

Definition of the interactions of prokaryotic DNA-binding regulatory proteins with their target DNA has resulted from a combination of genetic, physical and biochemical analyses (see Pabo and Sauer, 1984). As discussed previously, a model of DNA-binding based on a common helix-turn-helix motif has emerged.

The structure of the helix-turn-helix was first derived from X-ray crystallographic studies of the bacteriophage λ repressor (Pabo and Lewis, 1982) and Cro (Anderson *et al.*, 1981) proteins, and *E. coli* CAP (catabolite activator protein) (McKay and Steitz, 1981). The model for the interaction of the bihelical motif with DNA was based on steric and chemical complementarity with the DNA double helix as well as data that had localized the sites on the DNA that were involved in protein:DNA interactions. This was accomplished by establishing the DNA guanine residues protected from methylation by dimethyl sulfate, or phosphate groups protected from ethylation by ethylnitrosourea, in the protein:operator complex (Johnson *et al.*, 1978; Humayun *et al.*, 1977; Simpson, 1980).

Although there exist a large number of non-functional mutants throughout the helix-turn-helix in several proteins that are consistent with the function of the structure in binding DNA (see Pabo and Sauer, 1984), the most persuasive genetic evidence for the model has been derived from mutants with an altered specificity for DNA. Ebright *et al.* (1984) used one of two mutant (and non-functional) *E. coli* CAP operators that differed from the consensus by one nucleotide (the fifth on either side of the dyad axis), to select a complementing mutation in CAP. Three CAP mutants were obtained, each containing alterations in the second amino acid residue of the C-terminal helix of the helix-turn-helix motif. In a converse manner, Ebright (1986b), substituted the equivalent position of the Lac repressor by three different amino acid residues and examined the specificity of the altered proteins for target DNA. Although the mutant proteins bound all operator DNAs with much less affinity than that with which the wild-type protein bound its operator, the affinity of the mutant repressors for an operator altered in the fifth position from the dyad axis was equal to or greater than that for the wild-type operator. No other alterations in the operator sequence had this effect. These studies are consistent with the X-ray crystallographic data of the 434 repressor-operator complex (Anderson *et al.*, 1987) that suggests a contact by the second residue of the C-terminal helix with the fifth and sixth residues from the dyad axis of the operator (although, in the case of Lac repressor, it is possible only one of these contacts is made). Similar studies with λ repressor and Cro (Hochshild *et al.*, 1986) and 434 repressor

(Wharton and Ptashne, 1987) have confirmed the importance of the first and sixth residues of the C-terminal helix in contacting DNA specifically.

In addition to the DNA sequence-recognizing bihelical structure, other parts of the protein can make contact with DNA. Thus, in the case of λ repressor, the N-termini of the dimeric protein have "flexible arms" that wrap around the operator and provide additional specificity of binding (Pabo *et al.*, 1982; Eliason *et al.*, 1985). The C-terminus of λ Cro is also flexible and appears to contact the operator (Anderson *et al.*, 1981). Because there is no significant structural homology among the DNA-binding regulatory proteins apart from the helix-turn-helix, there is not a general form or location of DNA-contact other than that occurring at the helix-turn-helix.

Not all prokaryotic DNA-binding proteins follow the helix-turn-helix pattern. X-ray crystallography of *E. coli* EcoRI (McClarin *et al.*, 1986) and DNA polymerase I (Ollis *et al.*, 1985) has revealed a slightly different mode of binding, based as well on helices, but through a non-homologous structure. It should be noted, however, that these proteins are enzymes and have functions altogether different from transcriptional regulators.

434 repressor was shown to cause a small amount of DNA perturbation upon binding to its operator (Anderson *et al.*, 1987). Although no other structure of a DNA-binding regulatory protein with its operator is known, the models of protein-DNA interactions deduced from X-ray crystallography of λ repressor (Pabo and Lewis, 1982) and Cro (Anderson *et al.*, 1981), and the *E. coli* Trp repressor (Schevitz *et al.*, 1985) without operator do not necessitate alterations in the DNA structure apart from a slight bending towards the protein.

The angular placement of the helix-turn-helix N-terminal helices of the CAP dimer relative to the DNA major groove (McKay and Steitz, 1981), however, is different than that of the other structures analyzed. Non-denaturing gel electrophoresis of CAP-DNA complexes revealed that CAP causes a significant amount of bending of its operator (Liu-Johnson *et al.*, 1986). CAP was complexed with an array of identical-sized DNA fragments that contained the operator sequence at different points within the fragment. These complexes demonstrated different electrophoretic mobilities that could be

correlated to the location of CAP complex within the fragment. The altered electrophoretic mobility of the CAP-DNA complex was thought to result from a specific CAP-induced bend in the DNA, whose magnitude was calculated from steric considerations to be between 90^0 and 180^0. The energy of this structure is thought to be used for the opening of the helix to facilitate subsequent transcription initiation. EcoRI also induces bends (or "kinks") in DNA. However, in this case, the kinks cause an unwinding of the DNA helix that allows the proper recognition of the DNA by the protein (McClarin *et al.*, 1986).

A feature common to those prokaryotic regulators of transcription studied is the use of protein dimers to contact the DNA. The operators to which the proteins bind are symmetrical, and each half-site, in successive major grooves lying on the same face of the DNA double helix, is in identical contact with its monomer subunit. Deviations from this pattern, however, are apparent in the *E. coli* LexA repressor (Hurstel *et al.*, 1986), which appears to bind to major grooves separated by 1.5 turns of the helix, and the AraC protein (Hendrickson & Schleif, 1985), which contacts three adjacent major grooves of the operator DNA.

Postulated mechanisms by which DNA-binding activator proteins increase the rate of transcription of their target genes include conformational alteration of the DNA (as discussed for the case of CAP), or interaction with the RNA polymerase molecule. Mutations in the λ and P22 repressors (which have a transcriptional activator, as well as a repressor role) have been isolated that allow the proteins to bind DNA and repress transcription normally but that do not allow transcriptional activation (Guarente *et al.*, 1982; Poteete & Ptashne, 1982). This argues against a purely DNA-conformational change mechanism for transcription activation in these cases. The residues altered in each of these bacteriophage positive control mutants are on the surface of the molecule and are predicted to contact the polymerase. It is still unclear how this contact stimulates transcription or if this is a general phenomenon of transcription activation.

iii. Protein:protein interactions of prokaryotic regulatory proteins

The protein:protein interactions of prokaryote DNA-binding regulatory proteins are both complex (involving intramolecular and intermolecular associations) and dynamic (due to the involvement of the proteins in the kinetic process of transcription initiation), and for these reasons our understanding of the subject has lagged.

The interactions responsible for the dimerization of regulatory proteins have in some cases been defined. Dimer contacts are made in the C-terminal domains of λ and 434 repressors (Pabo *et al.*, 1979; Koudelka *et al.*, 1987) and *E. coli* LexA repressor (Hurstel *et al.*, 1986), and in the N-terminal domain of CAP (McKay *et al.*, 1982). The DNA-binding function of these proteins is contained in a separate domain, i.e. the N-terminal domain [λ repressor (Pabo *et al.*, 1979; Sauer *et al.*, 1979), *E. coli* LexA repressor (Little and Hill, 1985), 434 repressor (Anderson *et al.*, 1985)] or C-terminal domain [CAP (McKay *et al.*, 1982)]. The two domains of the λ and LexA repressor proteins are connected by a flexible hinge region which is a target for the RecA protease (Little and Mount, 1982).

In addition to the interactions between monomer units of DNA-binding regulatory proteins, and of the potential regulatory protein-polymerase interactions, there is evidence that interactions exist between DNA-bound proteins separated by as much as hundreds of base pairs (for review, see Ptashne, 1986). Efficient repression of the *E. coli gal*, *ara* and *deo* operons requires the presence of two operators, separated by 110, 200 and 600 base pairs, respectively (Dunn *et al.*, 1984; Irani *et al.*, 1983; Dandanell & Hammer, 1985). In each case, repressor molecules are bound to both sites in a cooperative manner (i.e. the binding of one molecule facilitates the binding of the second), and the repressors are thought to make contact by looping out the intervening DNA. Strong evidence for looping was the observation that repression of the *ara* operon was maintained when an integral number of DNA helix turns separated the sites (and they were therefore on the same face of the DNA), but not when they were separated by half-integral numbers of turns (and therefore on opposite sides) (Dunn *et al.*, 1984). The energy required to bend and twist the DNA to allow contact between protein molecules

on opposite sides of the helix was thought to be too great. λ repressors also bind their operator sites in a cooperative manner (Ackers *et al.*, 1982) and have been shown, by electron microscopy, to interact by a looping out of DNA when the sites are separated by an integral number of helix turns (Griffith *et al.*, 1986).

Molecular aspects of eukaryotic transcriptional regulation

Eukaryotic mRNA is synthesized by RNA polymerase II. The regulation of transcription initiation by this enzyme is apparently dependent on the interaction of a number of auxiliary protein factors which recognize and bind to specific DNA sequences located within the promoters of eukaryotic genes (for review, see Dyan and Tjian, 1985). Some of these factors have been isolated and characterized, and will be described after a brief review of the eukaryotic promoter. As the yeast promoter and its regulation differs in some respects from that of higher eukaryotes, the two systems will be examined separately.

i. The higher eukaryotic promoter

Most higher eukaryotic promoters contain an element (called the TATA box) that is located 25-30 bp upstream from the start-point of transcription (for review, see Breathnach and Chambon, 1981) and which is important for the correct location of transcript initiation (Grosschedl and Birnstiel, 1980). Other promoter elements located further upstream determine the frequency of initiation, and include, for example, sequences homologous to CCAAT [CAAT box (Efstratiadis *et al.*, 1980)] or GGGCGG [GC box (Benoist and Chambon, 1981)]. These promoter elements may have a very general (i.e. common to many varied promoters) or specialized (e.g. tissue-specific) role.

Many genes are further regulated by sequence elements called enhancers. These may be situated at varied distances from the promoter, either upstream or downstream from the transcription initiation site, and in either orientation. Some, such as the SV40 enhancer (Neuhaus *et al*, 1984; deVilliers, *et al.*, 1984), activate genes in a variety of cell types, whereas others are active only in specific cell types or environmental conditions (see Serfling *et al.*, 1985). The specificity of the enhancer is

apparently derived from the combination of sequence motifs present within the enhancer (Schirm *et al.*, 1987).

There are relatively few identified negative regulatory elements in higher eukaryotic promoters. Three 30-35 bp sites within the SV40 early promoter (overlapping the transcription initiation site) are responsible for the repression of the transcription unit by the T antigen (Tjian, 1978), and DNA sequences with a negative effect on transcription have been identified in the ovalbumin (Gaub *et al.*, 1985) and HMG coenzyme A reductase (Osborne *et al.*, 1985) promoters. These latter two elements are located far upstream, or within, the positive control regions of the promoters, respectively. The enhancers of the β-interferon gene (Goodbourn *et al.*, 1986), the insulin 1 gene (Laimins *et al.*, 1986; Nir *et al.*, 1986), the immunoglobulin heavy chain gene (Imler *et al.*, 1987), and SV40 (Borrelli *et al*, 1984), polyomavirus (Borrelli *et al*, 1984) and murine sarcoma virus (Gorman *et al.*, 1985) also contain sequences that mediate negative control.

ii. Higher eukaryotic transcription factors

Many of the promoter/enhancer elements discussed above have been shown to bind, and require for function, specific protein factors. One of the best characterized of these promoter-specific factors is Sp1. This factor, present as both 95 kd and 105 kd species (Briggs *et al.*, 1986), activates transcription from a variety of viral and cellular genes by binding to the GC box promoter element. Analysis of the protein-DNA complexes by DNase or dimethylsulfate protection has revealed that DNA contacts are made by the protein in the major groove and on one strand (Dynan and Tjian, 1983). The binding of Sp1 to DNA also causes an enhanced reactivity of part of the DNA sequence toward DNase I and dimethylsulfate, thus it is possible that Sp1 causes significant alterations in the DNA structure.

Sp1-responsive promoters generally contain multiple copies of the Sp1 element. Tandem sites in the AIDS virus LTR (long terminal repeat) and in the SV40 early promoter are arranged at intervals of 10-12 bp, i.e. approximately once per helical turn, and thus lie on the same face of the DNA helix (Kadonaga *et al.*, 1986). Sp1 does not, however, bind in a cooperative manner.

Although the mechanism of transcriptional activation by Sp1 or other transcription factors is unknown, there is evidence that long-range protein-protein interactions are important. In an experiment similar to the ones described previously for the *ara* and λ transcription units, Takahashi *et al.* (1986) have demonstrated that efficient transcription of the SV40 early promoter requires an integral number of helical turns between the TATA box, the Sp1 binding sites and the enhancer. It is likely that transcriptional activation results from contacts between proteins bound at each of these sequence motifs.

Several other sequence elements within eukaryotic promoters have been shown to bind specific protein species. The transcription factor that binds to the CAAT box (Jones *et al.*, 1987), as well as the enhancer-binding activator protein AP1 (Lee *et al.*, 1987) have been purified from human cells. The *Drosophila* transcription factor that mediates the heat shock response by binding to specific promoter elements has also been purified (Wiederrecht *et al.*, 1987). This heat shock transcription factor (HSTF) binds cooperatively to two sites within the *Drosophila hsp70* promoter (Topol *et al.*, 1985) and induces DNA bending (Shuey and Parker, 1986a). The conformation of the protein also appears to change upon specific cooperative binding (Shuey and Parker, 1986b).

The best characterized example of a higher eukaryotic transcriptional repressor is the SV40 T antigen (see Tjian, 1981), which, as mentioned, is responsible for the repression of the SV40 early transcription unit. It is a 96 kd phosphoprotein that binds as a tetramer to three tandem sites within the SV40 early promoter. The binding is cooperative and contacts are make by the protein to residues within the major groove of the DNA helix.

Although an increasing number of higher eukaryotic transcription factors are being characterized and purified, no factor has been studied to the extent that the molecular interactions responsible for specificity of binding and transcriptional activation have been elucidated. The combination of factor purification and cellular transformation of mutated genes should quicken the pace of understanding.

iii. The yeast promoter

Yeast promoters differ from those of higher eukaryotes in several respects. The yeast promoter generally contains a conserved "TATA" box, although it is located at a greater and more variable distance from the transcription initiation site (approximately 40-120 bp away). It is necessary for proper initiation of transcription, but the actual site of initiation is dependent on the "initiator" element, located near the site of initiation (Chen and Struhl, 1985; Hahn *et al.*, 1985; Nagawa and Fink, 1985; McNeil and Smith, 1986).

Upstream activator elements (UAS), located at variable distances up to 600 bp upstream of the transcription initiation site, are required for transcription and determine the regulatory properties of the promoter (Guarente, 1984). UASs share characteristics of higher eukaryotic enhancers, although they do not function when present downstream from the initiation site (Guarente and Hoar, 1984; Struhl, 1984). UASs are thought to function by binding transcriptional regulatory proteins, and in some cases, these factors have been isolated (to be discussed later).

Some promoters are regulated in a negative fashion, and sequence elements responsible for this control have been located (Brent, 1985). Although often located between the UAS and TATA elements [*MAT*α*1* (Siliciano and Tatchell, 1984), *HO* (Miller *et al.*, 1985), *CYC1* (Guarente and Mason, 1983), *STE6* (Johnson and Herskowitz, 1985)], they may occur far upstream of the transcription initiation site [*gal10* (Struhl, 1985), *ADR2* (Beier and Young, 1982)].

iv. Yeast transcriptional regulatory proteins

The gene products of the *GAL4*, *GCN4*, and *HAP1* genes are the best characterized *trans*-acting factors responsible for the positive regulation of transcription initiation in yeast.

GAL4 is required for the galactose-induced expression of the yeast *GAL1, GAL7,* and *GAL10* genes (Douglas and Hawthorne, 1966). It binds to a series of related 17 bp sequences located within the *GAL1* and *GAL10* UAS, which is situated approximately 250 bp upstream from the transcription start point (Johnson and Davis, 1984; Giniger *et al.*, 1985). The sequence to which GAL4 binds is bilaterally symmetrical (Giniger *et al.*, 1985), and thus GAL4 is thought to bind as a dimer (or

tetramer). Mutational analysis has allowed the location of domains of activity within the GAL4 protein. The amino terminal 74 amino acids contain the DNA-binding activity (Keegan *et al.*, 1986), and display homology to a cysteine-rich sequence comprising the DNA-binding domain of the eukaryotic RNA polymerase III transcription factor TFIIIA (Miller *et al.*, 1985b). The structure that is thought to bind DNA consists of several flexibly linked loops (fingers) that each enclose a zinc atom at its base through pairs of cysteines and histidines. Each "fingertip" is suggested to contact successive one half turns of the helix and provide specificty of binding. Homologies to the distinctive primary amino acid sequence of this domain have been found in several other eukaryotic DNA-binding proteins (see Berg, 1986). The transcriptional activation function of GAL4 is separable from its DNA-binding activity, and is located at either or both of two regions of the protein; that is, its carboxy terminal 113 amino acids, or a 48 residue stretch in the amino terminal half (Ma and Ptashne, 1987a). These regions are likely involved in contacting additional transcription factors. The carboxy terminal 28 amino acids interact with the negative regulatory protein GAL80 (Ma and Ptashne, 1987b; Johnston *et al.*, 1987), which abolishes transcriptional activation by GAL4 in the absence of galactose.

Although the *cis*-acting sequences required for negative regulation of several yeast genes have been identified [the mating type-responsive genes (Johnson and Herskowitz, 1985; Miller *et al.*, 1985)], the silent mating type loci (Feldman *et al.*, 1984), *ADR2* (Beier and Young), *GAL10* (Stuhl, 1985), *CAR1* (Sumrada and Cooper, 1987)], the *trans*-acting factors responsible for that function have not been well characterized (apart from α2), and DNA-binding of these factors has not been demonstrated. It is thus impossible at present to compare the properties of DNA-binding negative regulatory proteins within this organism.

Inferrence of the mechanism of transcriptional repression by *trans*-acting factors in yeast by analysis of the *cis*-acting sequences responsible has not been possible. Thus, although in many cases, the negative regulatory sequence is located between the UAS and the TATA (and could therefore function to prevent the polymerase from sliding from the the former to the latter element), it has been shown in several cases (Struhl, 1985; Beier & Young, 1982; Johnson & Herskowitz, 1985; Siliciano &

Tatchell, 1986) that the element also functions when located well upstream from the UAS. [In the case of repression of the silent mating type loci by the *cis*-acting "silencer" sequence, function is transmitted over a kilobase of DNA (Brand *et al.*, 1985).] Thus, whether the *trans*-acting factors cause a change in the DNA structure that is relayed to the promoter or transcription initiation site, or whether they are involved in complex auxiliary protein contacts that ultimately control the frequency of initiation, or whether it is a combination of these or a completely different mechanism, remains to be seen.

IS THERE UNIVERSALITY OF TRANSCRIPTION CONTROL MECHANISMS?

In spite of the many additional aspects of transcriptional control present in the eukaryote (i.e. through chromatin modification), it appears that there are certain fundamental similarities between prokaryotes and eukaryotes with respect to the mechanisms whereby *trans*-acting factors regulate transcription. Regulation of transcription initiation generally occurs in all organisms through the binding of the *trans*-acting factor to a DNA sequence located near (but at variable distances from) the transcription initiation point. Protein-protein interactions (sometimes cooperative) between transcription factors have been discussed for both eukaryotes and prokaryotes, and seem to play a central role in regulation in both types of organisms. Concrete evidence for common mechanisms of gene regulation has been demonstrated in at least four cases. Three constructions in yeast (Brent & Ptashne, 1984; Brent & Ptashne, 1985; Pinkham *et al.,* 1987) and one in mammalian cells (Hu & Davidson, 1987), employing both the *cis* and *trans* elements of prokaryotic regulators, have been shown to be appropriately regulated. The bacterial LexA repressor repressed the yeast *GAL1* gene when the *lexA* operator was present within the *GAL1* promoter (Brent and Ptashne, 1984). Similarly, the inducible *lac* operator/repressor system was functional in mouse cells (Hu and Davidson, 1987). Even more interesting, however, was the demonstration that bacterial-yeast hybrid proteins (LexA-Gal4, LexA-Hap2) bearing the DNA-binding specificity of the bacterial protein and the activation functionality of the yeast protein, were capable of activating transcription from synthetic promoters containing the bacterial operator (Brent and Ptashne, 1985; Pinkham *et al.*, 1987). These studies

suggest a similarity in mechanism of gene regulation and rule out the possibility, for example, that the *trans*-acting factors (in these cases) function primarily to alter the structure of the DNA.

So far, the only indication as to whether prokaryote and eukaryote regulatory proteins share structural similarities relating to the function of binding DNA is the existence of sequence homologies, as previously discussed, between eukaryotic proteins and the prokaryotic DNA-binding helix-turn-helix motif. Homologies between the DNA-binding domain of the RNA polymerase III transcription factor TFIIIA and several RNA polymerase II regulatory proteins (Berg, 1986) suggests that there is at least one other motif of DNA-binding in the eukaryote.

UNDERSTANDING THE MOLECULAR MECHANISM OF ACTION OF THE *MAT* LOCUS

Ultimately, a complete understanding of any given system of gene control will require the combined data of genetic manipulation (mutagenesis), *in vivo* and *in vitro* reconstruction analyses, and X-ray crystallography of the regulatory protein-DNA complex. The *MAT* locus in yeast is one of the genetically best defined eukaryotic regulatory systems. How the locus determines cell type at the genetic level is fairly well understood, however, the molecular interactions responsible for this control are certainly less clear. In the case of *MAT*α*1*, a first step in elucidating the molecular mechanism of action is the verification of the gene product amino acid sequence. This was accomplished in this study through the sequence analysis of a conventionally isolated nonsense mutant. This ensures that the predicted reading frame is that which encodes the gene product.

The fashion in which *MAT*α*2* exerts its control in the determination of the α cell type is better understood at the molecular level. However, although it is known that the gene product binds DNA, the structure of the protein-DNA complex, the mode of recognition of the protein, and the mechanism whereby specific binding prevents transcription from the corresponding gene are only very vaguely understood. This study addresses the first two of these questions through the use of mutagenesis and both *in vivo* and *in vitro* phenotypic analyses. The techniques involved (specific mutation of a gene *in vitro*, genetic manipulation of yeast, the *in vitro* synthesis of a protein, and the analysis of its interaction

with DNA) have only recently been developed and have proven to be extremely powerful tools for the elucidation of genetic interactions.

MATERIALS AND METHODS

REAGENTS

Enzymes

Restriction endonucleases were supplied by Bethesda Research Laboratories (BRL) or New England Biolabs, Inc. (NEBL). Bacterial alkaline phosphatase, T4 polynucleotide kinase, and T4 DNA ligase were supplied by BRL; DNA polymerase I (Klenow fragment), RQ1 DNase, SP6 RNA polymerase, and RNasin were supplied by Promega; pancreatic RNase was from Sigma; Zymolase 60,000 and Zymolase 5000 were from Seikagaku Kogyo Co. Ltd.; and Glusulase was from DuPont.

Nucleotides

Deoxyribonucleotide triphosphates, dideoxyribo-nucleotide triphosphates, m^7GpppG, and GpppG were purchased from Pharmacia; ribonucleotide triphosphates from Boehringer Mannheim; and ^{32}P-labelled nucleotides from Amersham.

Oligonucleotides

Oligonucleotides were either purchased from Pharmacia or were synthesized by T. Atkinson or Dr. R. Barnett on an Applied Biosystems 380A DNA synthesizer. Table I lists the seqeuence and source for each oligonucleotide. Oligonucleotides synthesized on the DNA synthesizer were purified by electrophoresis as described by Atkinson and Smith (1984).

Autoradiography materials

Kodak XRP-1 film was used for all autoradiography, a Dupont Cronex intensifying screen was used occasionally, and New England Nuclear (NEN) Enlightning was used for fluorography of ^{35}S-labelled proteins.

In vitro translation materials

Wheat germ extract, an amino acid mixture without methionine, and a potassium acetate

TABLE I

LIST OF OLIGONUCLEOTIDES

Name	Sequence	Source
TAA1	5'-d(ATCGTTTTATATGC)-3'	Pharmacia
TAA2	5'-d(TATCTAGTTATGGG)-3'	"
SP1	5'-d(TTTGTTTTTCGGGCTCAT)-3'	T. Atkinson
SP2	5'-d(CAGCTTAGAAGTGGGCAAGA AAAAAAGGAAGATAAGCAAG AAAAAATGA)-3'	"
SP3	5'-d(TTCTGGAGCTCTTGTTATTG)-3'	"
SP4	5'-d(TGATTTGAACCCGAGATAAACT)-3'	"
SP5	see Figure 10	"
SP6	see Figure 10	"
SP7	5'-d(GGATTTAAATTAATCTGTGAT)-3'	"
SP8	5'-d(TTTGATTTGACTTTGAGATAAACT)-3'	"
SP10	5'-d(TTTGATTTGAATCTCAGATAAACT)-3'	"
SP11	5'-d(TTTGATTTGACGCTCAGATAAACT)-3'	"
FP1	5'-d(TCACGACGTTGTAAAAC)-3'	R. Barnett
RP1	5'-d(TCACACAGGAAACAGCT)-3'	"

solution were purchased as a kit from Amersham. ^{35}S-methionine, at >2.96x10^{7} MBq/mmol specific activity was also from Amersham.

Media components

Amino acids, vitamins and ampicillin were supplied by Sigma. The remainder of media components were supplied by Difco.

Other

Low melting point agarose, acrylamide, and N'N-methylenebisacrylamide (all electrophoresis grade) were purchased from BRL. Formamide (from BDH, analytical grade) and phenol (Fisher reagent grade) were generally used without further purification.

BACTERIAL AND YEAST STRAINS

E. coli

Strain RR1, used for most bacterial transformations, has the following genotype: F$^-$, *hsdS20* (r_B, m_B), *ara-14, proA2, lacY1, galK2, rpsL20 (Sm^r), xyl-5, mtl-1, supE44, i*$^-$ (Bolivar et al. 1977).

JM101 (*Δ(lac pro), thi, supE,* F'*traD36, proAB, lacI^{q}Z M15*) was used for the isolation of single-stranded DNA and for the propogation of M13 vectors (Messing, 1983).

RZ1032 (HfrKL16 PO/45 [*lysA*(61-62K)], *dut1, ung1, thi1, relA1,* Zbd::Tn10, *supE44,* Kunkel et al. 1986) was used for the isolation of uracil-containing single-stranded DNA.

S. cerevisiae

Table II lists the genotypes and sources for all strains used in this study.

Strain SP2 (containing *MATΔ* and *ste6::lacZ* alleles) was constructed as follows: strain MH52-3C (*MATΔ*) was transformed with YEpMATa (to allow sporulation in the diploid) and was crossed with strain 1161 (*α ste6::lacZ*). The mating mixture was streaked on SC-leu medium (to select for the diploid or for the plasmid-transformed MH52-3C) and single colonies were transferred to sporulation plates. Approximately 10% of the colonies sporulated. Asci were disrupted and streaked on YPD, and several of the colonies tested for β-galactosidase activity. Those displaying a *LACZ*$^+$ phenotype were

TABLE II

S. cerevisiae STRAINS

Strain	Genotype	Source
2935-10C	*matα1-o hmla HMRa gal1 suc trp1 met8 leu2 lys1 can1-o rme1*	D. Hawthorne
23.75	*matα-x75*[a] *trp1 leu2 ura3 his4*	K. Tatchell
D311-3A	*a lys2 his1 trp2*	J. Boss
RC757	*α sst2-1 met1 his6 can1 cyh2*	V. MacKay
S91	*a leu2 his4 lys2 cry*^{r}	S. Roeder
S91-75	*αx75/a his4/his4 leu2/leu2 CRY*^{s}*/cry*^{r} *LYS2/lys2 trp1/TRP1 ura3/URA3*	This study
K80.148aa	*a/a leu2/LEU2 lys1/LYS1 ura3/URA3 trp1/trp1*	K. Tatchell
MH52-3C	*MATΔ130-141::CAN his3 his4 trp1 leu2 sac*^{R} *rme*	M. Hall
K1107	*a HMRa hmla ho::lacZ-46 ura3 ade2 can1 met*$^{-}$ *his3 leu2 trp1*	K. Nasmyth
1160	*a ste6::lacZ ura3 leu2 trp1 his3 his4*	K. Wilson
1161	isogenic to 1160, except α	K. Wilson
1402	*a ho::lacZ ura3 leu2 trp1*	R. Jensen
SP2	*matα::CAN ste6::lacZ ura3 leu2 trp1 his3 his4*	This study
SP3	*matα::CAN/MATa his3/+ his4/+ trp1/trp1 leu2/leu2 ura3/ura3 sac*^{R}*/+ rme/+*	This study

[a] This is a transplacement of mutant *matα-x75* (Tatchell *et al.*, 1981) into the *MAT* locus.

grown in YPD culture and screened for the loss of the YEpMATa-encoded *LEU2* marker. A single colony containing the *MAT* deletion and *ste6::lacZ* loci characteristics was chosen.

Strain SP3 was constructed by crossing strain MH52-3C, transformed with YCpMATα, with strain 1402, selecting for the diploid, and subsequently screening for loss of the plasmid.

PLASMIDS

Bacterial-yeast shuttle plasmids used in this study are indicated schematically in Figure 7. Other plasmids are indicated throughout the work. In most cases, standard enzymological procedures were employed for their construction. Plasmid pSP64-MATα2 (Figure 17) was constructed as follows: pEMBL8(+)-MATα was partially digested with AluI and the products separated by electrophoresis on low-melting-point agarose. The region of the gel containing approximately 1 kb fragments was excised and the DNA isolated. The mixture was ligated to HincII-digested pEMBL8(+), cleaved with SalI (to select against religated vector), transformed into *E. coli*, and those transformants displaying the appropriate restriction pattern were verified by sequence analysis. The HindIII-EcoRI insert was then sub-cloned into pSP64. Derivatives of plasmid pSP64-MATα2 containing mutations in the region of homeo domain homology were constructed by replacing the wild-type XbaI fragment with that containing the mutation. The conversion was verified by sequence analysis.

MEDIA AND GROWTH CONDITIONS

E. coli

Cells were grown at 37^0C in YT (0.8% Bacto-tryptone, 0.5% Bacto-yeast extract, 0.5% NaCl), LB (1% Bacto tryptone, 0.5% Bacto-yeast extract, 0.5% NaCl, 0.1% glucose) or in M9S (50 mM Na_2HPO_4, 25 mM KH_2PO_4, 8.5 mM NaCl, 20 mM NH_4Cl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 10 mM glucose, 0.001% thiamine). Ampicillin was added, when required, at 100 mg/l. Plates contained 15 g/l agar, soft agar overlays contained 7 g/l agar.

S. cerevisiae

Yeast were grown at 30^0C unless otherwise indicated in media described in Sherman *et*

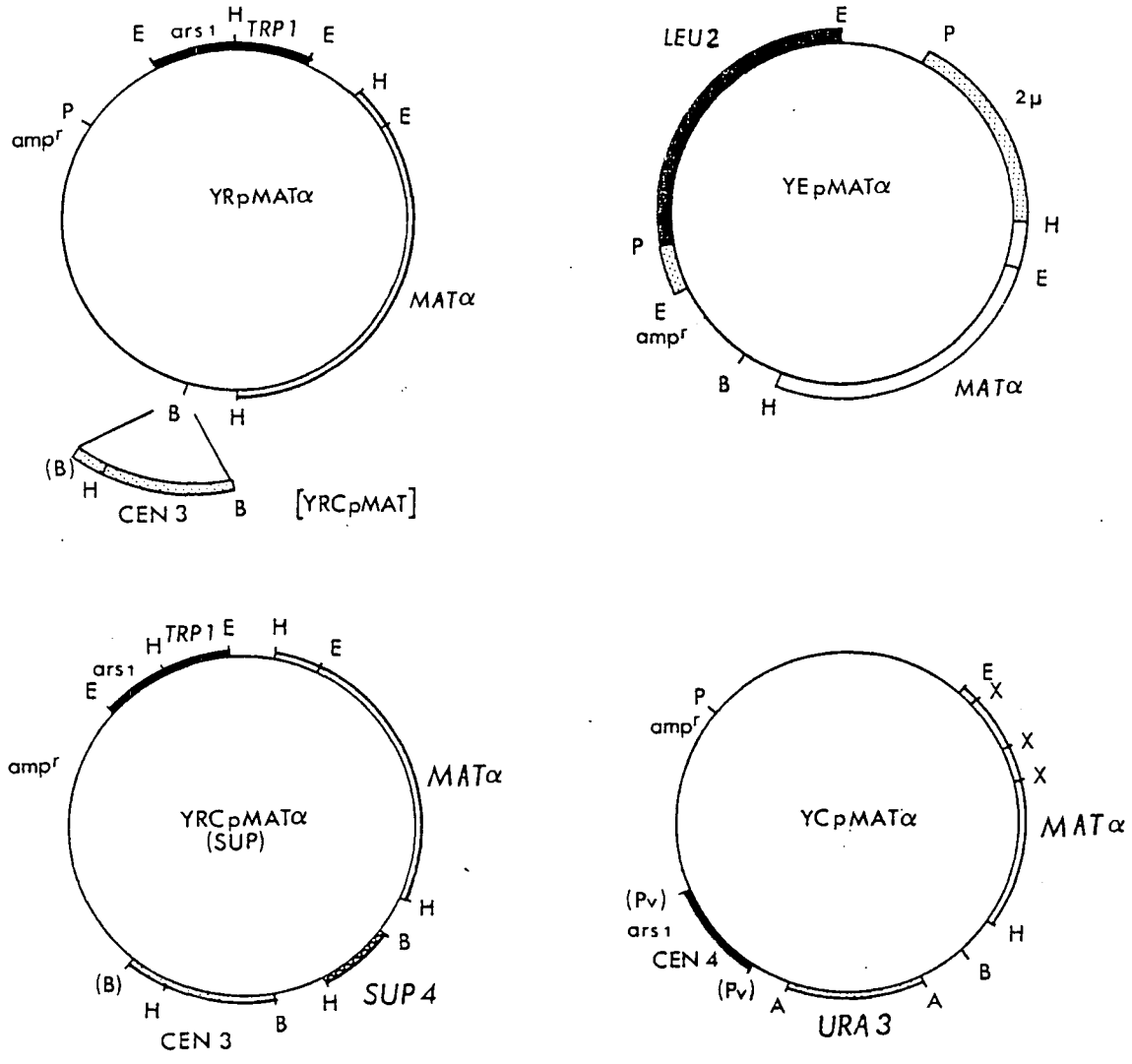


Figure 7. Structures of Yeast-Bacterial Shuttle Vectors Containing Wild-Type or Mutant *MATα* Loci. The single line represents pBR322 sequences. Rest-riction site abbreviations are as follows: A, AvaI; B, BamHI; E, EcoRI; H, HindIII; P, PstI; Pv, PvuII; X, XbaI. Restriction sites in parentheses are those destroyed during construction of the plasmid.

al.,1981. The following media were used:

YPD 2% Bacto-peptone, 1% Bacto-yeast extract, 2% glucose

SD 0.67% yeast nitrogen base without amino acids, 2% glucose

SC SD with amino acid supplementation as described in Sherman *et al.*, 1981, with an additional 120 mg/l leucine.

Solid media contained 2% agar.

Solid sporulation medium (Rothstein *et al.*, 1977) consisted of 2% potassium acetate, 0.25% yeast extract, 0.1% glucose, 1.5% agar, and appropriate amino acid supplementation. The pH was adjusted to 7.0.

TECHNIQUES FOR DNA MANIPULATION AND ANALYSIS

Restriction digests

Restriction digestions were performed at DNA concentrations of approximately 100-250 ng/µl in buffers described in Maniatis *et al.*, 1982 (containing 0 mM, 50 mM, or 100 mM NaCl, or 20 mM KCl). The salt concentration chosen for a particular enzymatic reaction buffer was that closest to that recommended by the manufacturer.

Gel electrophoresis

For analysis of DNA on polyacrylamide gel electrophoresis, the running buffer as well as the gel buffer was 50 mM Tris-borate pH 8.3, 1 mM EDTA. The ratio of acrylamide to bisacrylamide in the gel was 29:1. Agarose gels were made and run in the same buffer.

Isolation of DNA fragments

Fragments used for sequencing by the method of Maxam and Gilbert (1980) were isolated by polyacrylamide gel electrophoresis and electroeluted in dialysis bags in 25 mM Tris-borate pH 8.3, 0.05 mM EDTA.

Fragments used for cloning or other purposes were generally isolated from low-melting point agarose as described in Maniatis *et al.*, 1982.

Ligations

Ligations were performed in ligation buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM spermidine, 1 mM ATP, 0.5 mg/ml BSA) with approximately 200 ng cut vector and a three-fold molar excess of insert. The reaction was carried out either at 15^0C for 12-18 hours or at room temperature for 6 hours. If the vector was cut with one restriction enzyme only, it was dephosphorylated with bacterial alkaline phophatase as described by the manufacturer.

DNA sequence determination

DNA sequence analysis was performed either by the method of base-specific chemical cleavage (Maxam and Gilbert, 1980) or by the dideoxyribonucleotide chain-terminator method (Sanger, 1980).

Site-directed mutagenesis

Figure 8 indicates the three strategies used for the production of mutations at specified sites. Method A was carried out according to Zoller and Smith (1983) except that the extended and ligated template was separated from excess nucleotides by chromatography on Sephadex G-100 before purification on the sucrose gradient.

Method B is described in Zoller and Smith (1984) with the following exceptions: a pEMBL vector (Dente *et al.*, 1983) was used instead of an M13 mp vector, a 10-fold molar excess, only, of oligonucleotide over template was used, the extension and ligation were performed at room temperature for 7 hours, and the mixture was transformed into the *E. coli* strain RR1. Colony hybridization using the radioactively-labelled oligonucleotide was performed as follows: colonies grown on nitrocellulose filters were lysed and the DNA denatured by placing the filters on 10 ml 1M NaOH for 10 minutes, followed by 1 M Tris-HCl pH 8 for 1-2 minutes, and 1 M Tris-HCl pH 8, 1.5 M NaCl for 5 minutes. The filters were then inverted for several seconds on chloroform and baked *in vacuo* at 80^0C for 1 hour. Prehybridization of the filters was accomplished in 5 ml of 6X SSC, 10X Denhardt's, 0.2% SDS in a Petri dish at 37^0C for 1 hour. Hybridization was carried out in 4 ml of 6X SSC, 10X Denhardt's containing the probe, at room temperature for 1 hour. The filters were washed

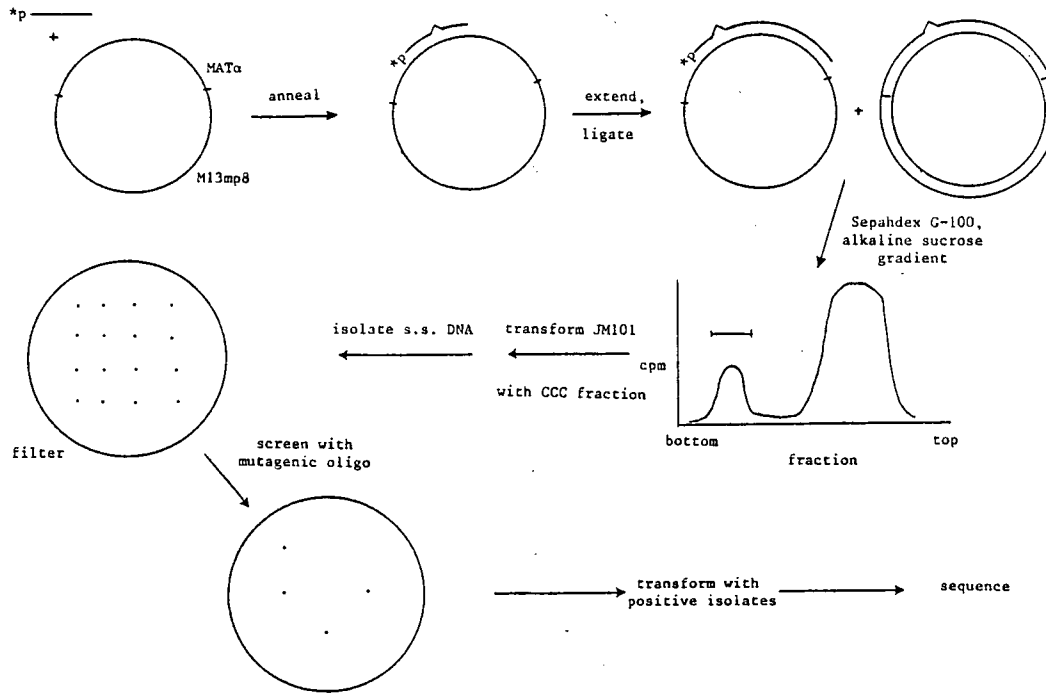
Figure 8. Schematic of Site-Specific Mutagenesis Methods.

Panel A: 5'-end labelled mutagenic oligonucleotide is annealed to single-stranded DNA template. After extension (with dNTPs) and ligation, completed double-stranded molecules are separated from incompletely extended molecules, after removal of excess nucleotides, by separation on an alkaline sucrose gradient. DNA corresponding to the CCC (covalently closed circular) fraction is used to transform strain JM101. Single-stranded DNA is isolated and hybridized on a filter to the radioactively-labelled mutagenic oligonucleotide. DNA hybridizing preferentially with the mutagenic oligonucleotide is used to transform *E. coli* to ensure clonality. The mutation is verified by DNA sequence analysis.

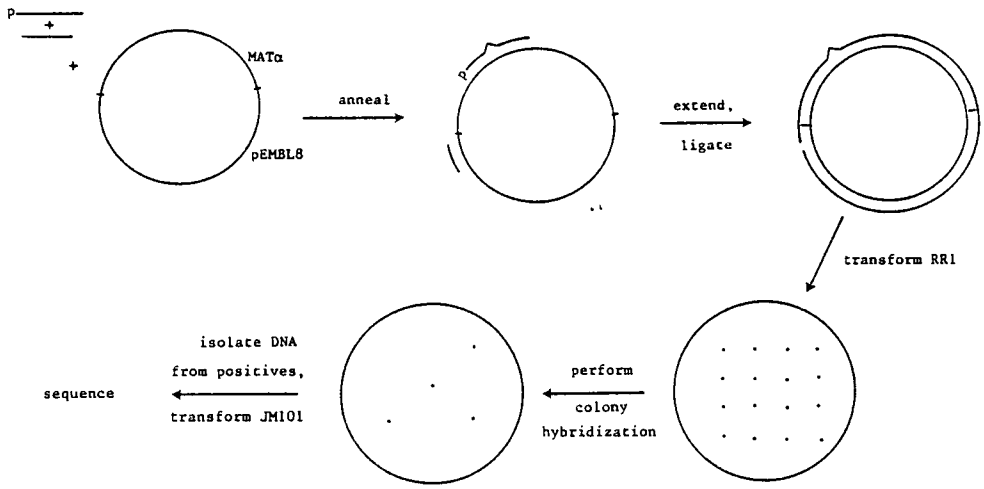
Panel B: A mixture of 5'-end labelled mutagenic oligonucleotide and unlabelled RP1 (which hybridizes to a portion of the vector sequence; Table I) was annealed to single-stranded pEMBL8(-)-derived DNA. The mixture was extended and ligated as before, and transformed into strain RR1. Colony hybridization with labelled mutagenic oligonucleotide was performed and the DNA from positive clones was isolated and used to transform strain JM101. The mutation was verified by sequence analysis.

Panel C: The mutagenic procedure as described in Panel B was performed on pEMBL8(-)-derived DNA that had been propogated through the *E. coli* strain RZ1032. A pool of DNA was isolated from strain RR1 after the initial transformation of mutagenized molecules, and single stranded DNA was subsequently isolated from JM101 for sequence determination.

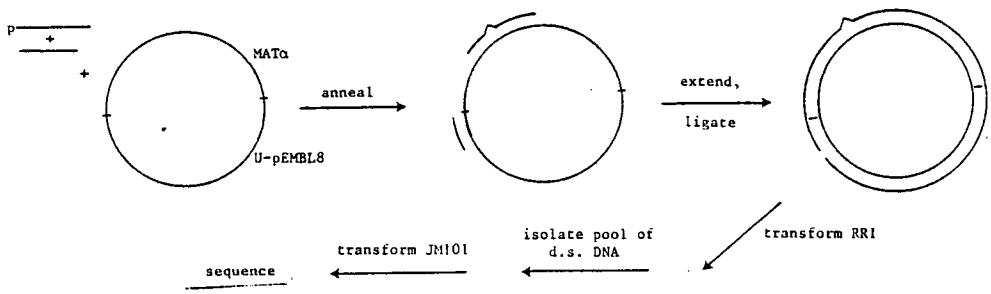
(A)


*p
MATα
M13mp8
anneal
*p
extend,
ligate
*p
Sepahdex G-100,
alkaline sucrose
gradient
filter
isolate s.s. DNA
transform JM101
with CCC fraction
cpm
bottom
top
fraction
screen with
mutagenic oligo
transform with
positive isolates
sequence

(B)


p
MATα
pEMBL8
anneal
p
extend,
ligate
transform RR1
isolate DNA
from positives,
transform JM101
sequence
perform
colony
hybridization

(C)


p
MATα
U-pEMBL8
anneal
extend,
ligate
transform RRI
isolate pool of
d.s. DNA
transform JM101
sequence

at successively higher temperatures in approximately 100 ml 6X SSC, and an autoradiograph was made for each temperature.

Panel (C) of Figure 8 displays the third strategy for creating site-specific changes. This is identical to the second method with the following exceptions: the single-stranded template used was propagated in strain RZ1032 in the presence of uridine as described in Kunkel (1986), the mutagenesis mixture was transformed initially into RR1, then a plasmid pool was obtained and was transformed into JM101 to isolate single-stranded DNA for sequence determination.

"Cassette" mutagenesis (Wells *et al.*, 1985) using a mixed synthetic DNA duplex insert (McNeil and Smith, 1985) was carried out as diagrammed in Figure 9. The region of *MAT*α2 with the greatest homology to the eukaryotic homeo box was replaced by a synthetic DNA duplex that contained, at specified positions, a mixture of mutant (4%) and wild-type (96%) bases. AvaI and SacI sites flanking this region, of plasmid pEMBL8(-)-MATα were constructed by method C of Figure 8. Double-stranded DNA was then cleaved with the two restriction enzymes and the vector-containing DNA was purified through low-melting-point agarose. This DNA was ligated with duplex DNA derived from oligonucleotides 5 and 6 (see Figure 10). (Duplex DNA was formed by mixing both oligonucleotides in equimolar amounts at 5 pmol/μl in 10 mM Tris-HCl pH7.5, 1 mM EDTA, placing them in boiling water for 1 min, and cooling them to 15^0C over a 2 hour period.) The ligation mixture was mixed with SacI (to cleave any remaining wild-type sequence) and transformed into *E. coli* strain RR1. DNA from pooled transformants (approximately 2600) was isolated and used to transform strain JM101. DNA from these transformants was isolated and individual isolates subjected to DNA sequence determination. The mutant *MAT* loci were transferred, as 4 kb HindIII-EcoRI fragments, into vector YCp50 (forming YCpMATα derivatives; see Figure 7).

TRANSFORMATIONS

Transformation of *E. coli* was carried out by the method of Mandel and Higa, 1970 as described in Maniatis *et al.*, 1982. Yeast transformation was carried out as follows: cells were grown in

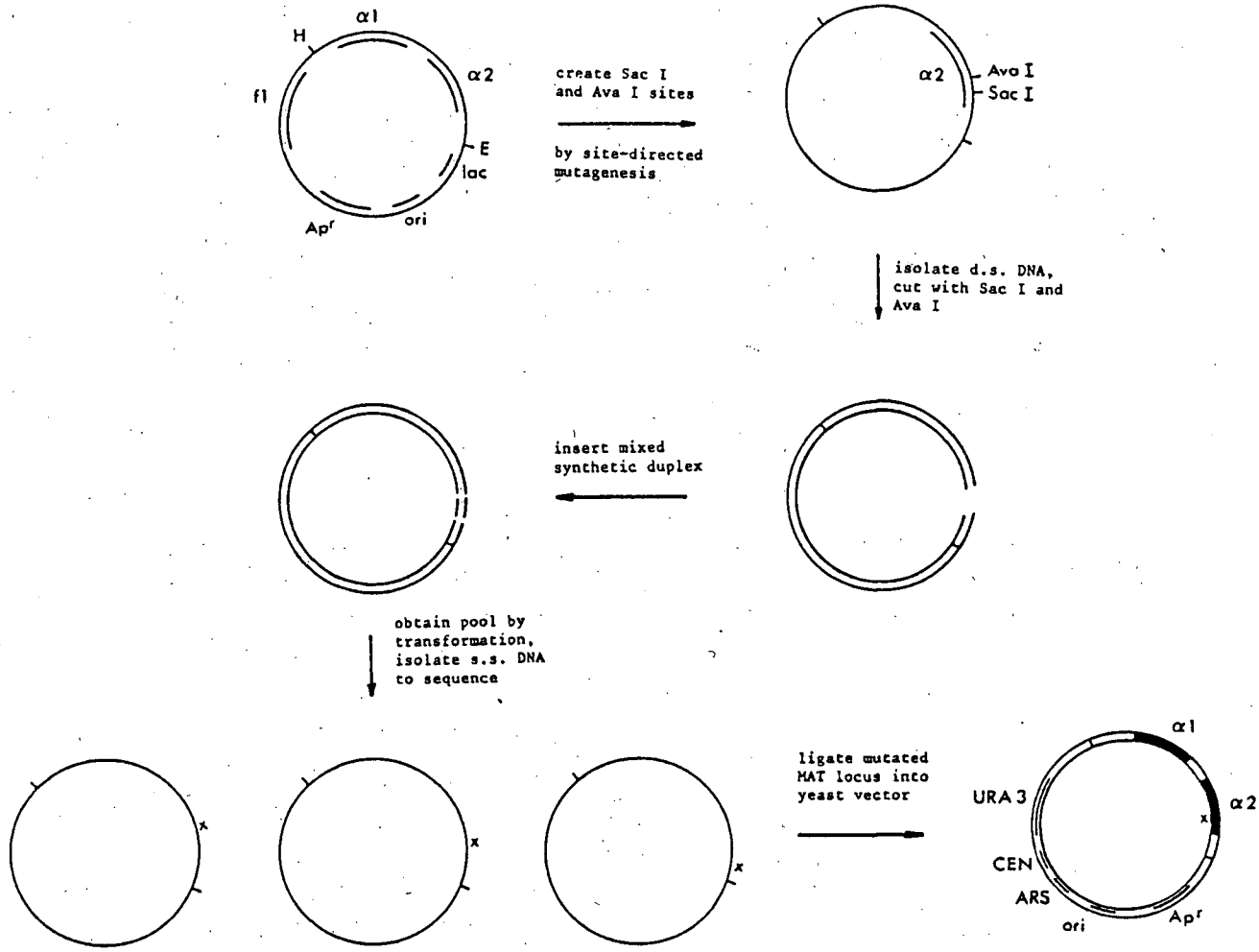


Figure 9. Outline of Strategy for Cassette Mutagenesis of Homeo Domain Homology Region.

SacI and AvaI sites that surround the homeo domain homology region of *matα2* were constructed by method C of Figure 7. Double stranded mutant DNA was cleaved with the two enzymes and a mixed synthetic oligonucleotide duplex was inserted. A pool of mutant plasmids was obtained, and individual clones were selected for DNA sequence analysis. Ap^r, ampicillin resistance gene; f1, the region of the f1 genome containing the *cis*-acting elements required for DNA replication and morphogenesis.

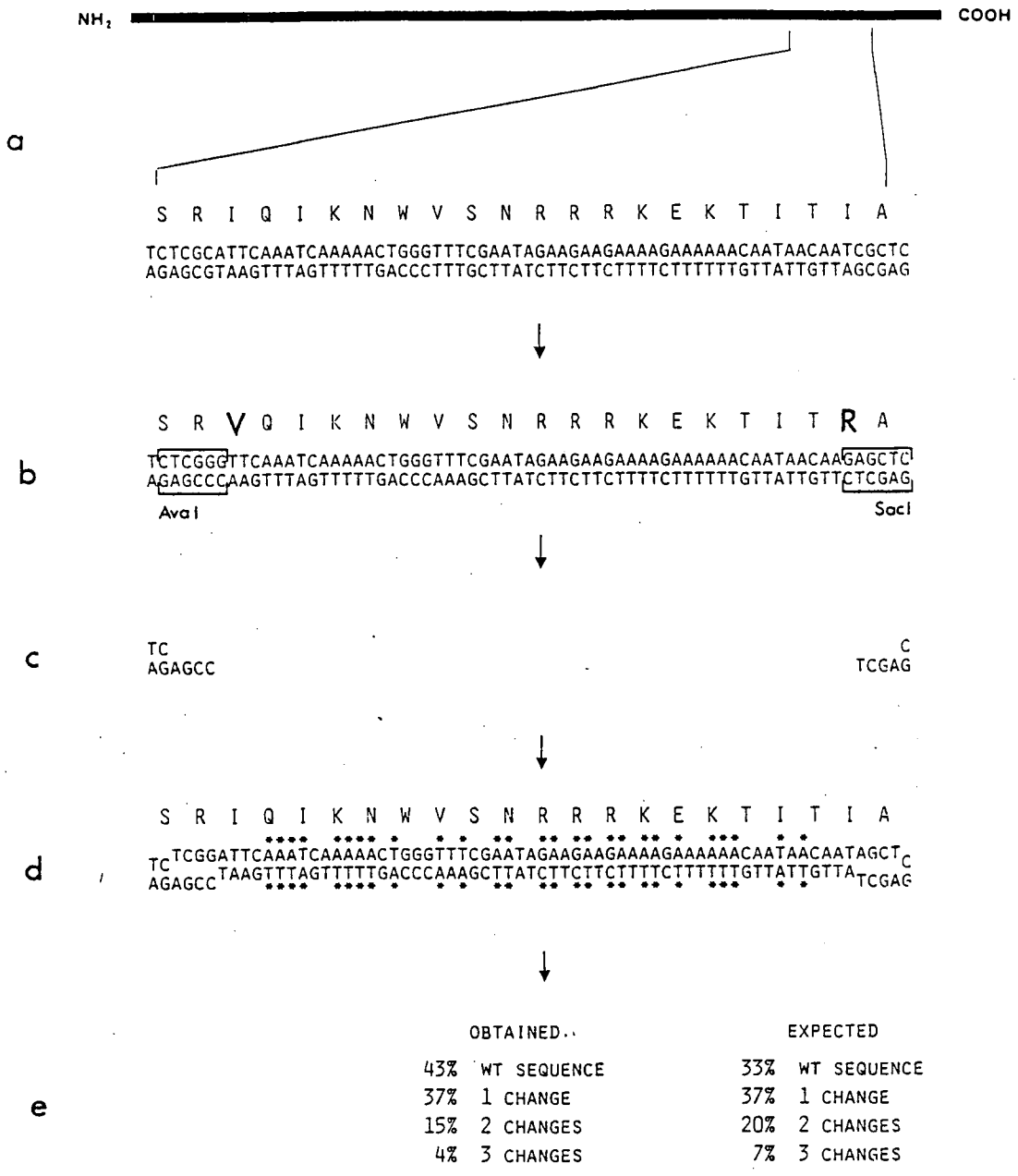


Figure 10. Details of Cassette Mutagenesis of the α2 Homeo Domain Homology Region.

a displays the wild-type sequence of the region subjected to mutagenesis. The two constructed restriction sites surrounding the cassette, and their corresponding amino acid residue changes, are indicated in **b**. **c** shows the removal of DNA sequence internal to the two sites, and **d** displays the nucleotide sequence of the insert. Asterisks denote those positions at which a mixture of 96% wild-type base and 4% of an equimolar mixture of the remaining three bases was used in the synthesis. **e** indicates the quantitative distribution of changes obtaining from sequencing approximately 75 clones in comparison to that expected.

YPD to mid-log phase, centrifuged, and washed in 1/10 volume 1 M sorbitol. Cells were resuspended in 1/10 volume 1 M sorbitol containing 20 mM DTT and 1% glusulase, and were gently agitated at 30^0C for 45-60 minutes until 70-80% of the cells were spheroplasts (as observed under phase contrast when a drop of suspension was mixed with an equal amount of 1% SDS). The cells were washed once in 1/10 volume 1 M sorbitol and once in 1/10 volume STC (1 M sorbitol, 10 mM Tris-HCl pH 8, 10 mM $CaCl_2$) and resuspended in 1/100 volume STC. DNA (approximately 1 ug) was incubated with 0.05 ml cells at 0^0C for 10-15 minutes, 0.25 ml of 20% PEG 3350, 10 mM Tris-HCl pH8, 10 mM $CaCl_2$ was then added and the mixture was incubated at 0^0C for 20 minutes. These cells were then added to approximately 15 ml of regeneration agar (1 M sorbitol, 0.67% yeast nitrogen base, 2% glucose, selective amino acids as described for SC, 2% YPD, 3% agar, pH 6) at 55^0C and plated on SC selective media. This procedure is based on that originally described by Beggs (1978).

ISOLATION OF PLASMID AND BACTERIOPHAGE DNAS

Isolation of double-stranded DNA from *E. coli*

Plasmid DNA was isolated from *E. coli* by the general method of Birnboim and Doly (1979) as modified by Maniatis *et al.* (1982) with several additional modifications. 1.5 ml of an overnight culture was centrifuged for several seconds in an Eppendorf centrifuge. The pellet was resuspended in 100 µl of an ice-cold solution of 25 mM Tris-HCl pH 8, 10 mM EDTA, and to this was added 200 µl of ice-cold 0.2 N NaOH, 1% SDS. The tube was inverted several times and stored on ice for 5 minutes. 150 µl of ice-cold 3 M KAc (pH 4.8) was added and the tube was vortexed and stored on ice for 5 minutes. The mixture was centrifuged for 5 minutes and the supernatant transferred to a new tube. The DNA was precipitated with 2 volumes ethanol at room temperature for 2 minutes, centrifuged for 5 minutes, and redissolved in 40 µl 10 mM Tris-HCl pH 7.5, 1 mM EDTA. 40 µl of 5 M LiCl was added, the solution was incubated for 5 minutes on ice, and centrifuged for 5 minutes. To the supernatant was added 2 volumes ethanol, the DNA was precipitated, and was redissolved in 50 µl 10 mM Tris-HCl pH 7.5, 1 mM EDTA. Large scale plasmid isolations (from a 40 ml or larger culture) were carried out in a

basically identical fashion. On occasion, the DNA was purified through a CsCl gradient as described by Maniatis *et al.*, 1982.

Isolation of replicative form (RF) M13 recombinants

M13 recombinants were grown and the replicative form isolated as described by Messing, 1983.

Isolation of single-stranded DNA

M13 recombinant single-stranded DNA was isolated as described by Messing, 1983.

Single-stranded pEMBL recombinant DNA from JM101 was isolated as described by Dente *et al.* (1983). Single-stranded DNA from RZ1032 in the presence of uridine was prepared as described in Kunkel (1986).

ISOLATION OF YEAST DNA

Plasmid recovery

The small-scale isolation of yeast DNA for the purpose of recovering plasmids is described in Sherman *et al.* (1981). Five mls of cells were generally used and the nucleic acid precipitate was not subjected to RNase treatment.

Isolation of genomic DNA

The large-scale isolation of DNA from strain 2935-10C for the purpose of cloning the *mat*$\propto$ locus was carried out as follows: 500 ml cells grown in YPD to saturation were centrifuged at 16,000 x g for 10 minutes, washed once in water and, in a 30 ml Corex tube, were resuspended in 2 ml/g cells of SCE mix (1 M sorbitol, 0.1 M sodium citrate, 0.06 M EDTA, 0.5 mg/ml Zymolase 5000, 0.1 M 2-mercaptoethanol, pH7) and incubated at 37^0C for 2 hours. 1.8 volumes of lysis buffer (3% sarcosyl, 0.5 M Tris, 0.2 M EDTA, pH9) was then added and the mixture placed at 65^0C for 15 minutes. After cooling on ice to room temperature, the solution was layered on to a 15%-50% sucrose gradient in 0.8 M NaCl, 0.02 M Tris-HCl pH 8, 0.01 M EDTA and centrifuged at 130,000 x g for 3.5 hours. The DNA was isolated from the gradient by removing the solution by Pasteur pipette from the top of the tube,

dialysed overnight against TE (10 mM Tris-HCl pH 8, 1 mM EDTA) and purified through a CsCl gradient (containing 10 g CsCl/8 ml solution of TE).

ISOLATION OF YEAST RNA

Cells were grown in selective medium until mid-log phase, then were incubated with cycloheximide at a concentration of 0.1 mg/ml for 5 minutes, and harvested by centrifugation at 1500 x g for 1 minute after being quickly cooled with crushed ice. The pellet was washed once in 4^0C water containing 0.1 mg/ml cycloheximide and frozen immediately at -70^0C. Silanized, acid-washed beads (3 g beads/g cell pellet), 3 ml/g cells of RNA extraction buffer (0.15 M NaCl, 0.1 M Tris-HCl pH7.5) and 50 µl/g of vanadyl ribonucleoside complexes (VRC, 0.2 M) were added to the cell pellet and the mixture vortexed for six 15-second intervals each followed by 45 seconds cooling on ice. Following centrifugation of the mixture at 10,000 x g for 10 minutes, the supernatant was incubated at 37^0C with 0.5% SDS and 0.5 mg/ml Proteinase K and the cell debris was extracted as before. The second supernatant was combined with the first, the SDS concentration adjusted to 0.5%, and the mixture incubated for 1 hour. The solution was extracted once with phenol/chloroform (1:1) and the nucleic acid precipitated with 0.1 volumes sodium acetate and 2.5 volumes ethanol. The pellet was then redissolved in 10 mM EDTA/2 M LiCl (per 100 ml original culture volume), incubated overnight at 4^0C, and centrifuged at 10,000 x g for 30-40 minutes. The pellet containing RNA was washed once with 4^0C 2 M LiCl/10 mM EDTA, was dissolved in TE and then precipitated with 0.1 volume sodium acetate and 2.5 volumes ethanol.

Small-scale isolation of RNA from 5 ml cultures was performed by a similar procedure. The isolation of poly-A^+ RNA was carried out if necessary by a batch procedure using fines derived from oligo-dT cellulose as described in Maniatis *et al.*, 1982.

ANALYSIS OF RNA

Agarose gel electrophoresis of RNA was carried out by the method of Lehrach *et al.*, 1977 as

described in Maniatis *et al.*, 1982. Gels contained 2.2 M formaldehyde in MOPS buffer (0.04 M MOPS pH7, 10 mM NaAc, 1 mM EDTA) and were run at 10 V/cm in MOPS buffer.

Transfer onto nitrocellulose for the purpose of Northern analysis was performed as described by Maniatis *et al.*, 1982. The conditions used for Northern blots are detailed in Zinn *et al.*, 1983. The radioactive RNA probe used for detection of *lacZ*-containing transcripts was prepared by transcribing plasmid pSP65-alacZ (1.9) (Figure 14) with SP6 RNA polymerase in the presence of radioactive nucleotides. This plasmid contains a portion of the *E. coli lacZ* gene in an orientation that allows the production of anti-sense RNA upon transcription. Specifically, 0.5 µg of HindIII-cleaved pSP65-alacZ (1.9) was transcribed by SP6 RNA polymerase as described by the manufacturer using 0.74 MBq ^{32}P-UTP/20 µl reaction. 0.5 units RNase-free DNase was added to reaction mixture for 10 minutes, and then the mixture was phenol-extracted. The RNA was precipitated three times with 33 µl 5 M NH_4Ac, 10 µl of 1 mg/ml tRNA and 3 volumes ethanol, and used directly.

CLONING OF *MATα1-oc* LOCUS

Preparation of DNA Library from yeast strain 2935-10C

DNA from yeast strain 2935-10C was prepared as previously described. 0.5 µg of the DNA was cleaved with HindIII and ligated in a volume of 50 µl with 0.1 µg of HindIII-cut, BAP-treated pBR322. This was transformed into *E. coli* RR1 and ampicillin-resistant colonies were selected. Ninety percent were tetracycline-sensitive and hence contained recombinant plasmids.

Identification of *matα* locus in strain 2935-10C genomic library

E. coli transformants containing the strain 2935-10C genomic library were grown on a grid in duplicate on LB-amp plates and on nitrocellulose filters placed on LB-amp plates. Colonies grown on the filters were lysed *in situ* as follows: the filters were placed, at room temperature for 10 minutes, on Whatman 3MM paper saturated with 1 M NaOH, followed by 1 minute on 3MM paper saturated with 1 M Tris-HCl pH 7.5, 5 minutes on 3MM paper saturated with 1 M Tris-HCl pH7.5, and 10 minutes on 3MM paper saturated with 1.5 M NaCl, 0.5 M Tris-HCl pH7.5. The nitrocellulose filters were

suctioned on a Buchner funnel for approximately 3 minutes, inverted onto a solution of 1.5 mg/ml Pronase (self-digested at 37^0C for 45 minutes) in 1XSSPE (0.18 M NaCl, 10 mM NaH_2PO_4 pH 7.4, 1 mM EDTA) for 15 minutes at room temperature, rinsed with ethanol, inverted onto $CHCl_3$, washed in 0.3 M NaCl and baked at 80^0C in a vacuum for 3 hours.

The probe used to detect *MAT* locus-specific DNA was prepared as follows: 0.5 µg of the 4.3 kb HindIII fragment of the *MATa* locus was labelled with 0.74 MBq of each of the four ^{32}P-dNTPs in the presence of DNA Polymerase I (Klenow fragment). The probe (approximately 10^6- 10^7 cpm) was denatured by boiling for 10 minutes together with 1.5 µg pBR322 and 500 µg *E. coli* DNA (donated by R.A.J. Warren). The filters were hybridized with the radioactive denatured *MATa* probe in 5X SSPE, 50% formamide, 0.3% SDS at 45^0C overnight. They were washed four times for 10 minutes each in 2X SSPE, 0.2% SDS at 45^0C and exposed to X-ray film.

Characterization of *mat*α clone from strain 2935-10C genomic library

Of 2040 colonies screened, 4 colonies hybridized with the *MATa* probe. Plasmid DNA from one of these was shown by restriction analysis to contain DNA corresponding to the *MAT*α locus and was further characterized by DNA sequence determination by the method of Maxam and Gilbert, 1980. Figure 11 indicates the strategy by which this was accomplished.

YEAST FUNCTIONAL ASSAYS

Mating

Qualitative mating tests were performed by mixing clumps of cells of both the known strain and that to be tested in a patch on either selective or complex medium. The patches were replica plated onto selective plates.

Quantitative mating tests were performed as follows: Cells were grown in a small culture (selective medium) to mid-log phase, and to 0.5 ml sterile water was added $2x10^6$ cells each of the strain to be tested and the tester strain. The mixture was collected on a nitrocellulose filter and placed on a YPD plate for 6 hours at 30^0C (unless otherwise specified). The cells were resuspended in 10 ml

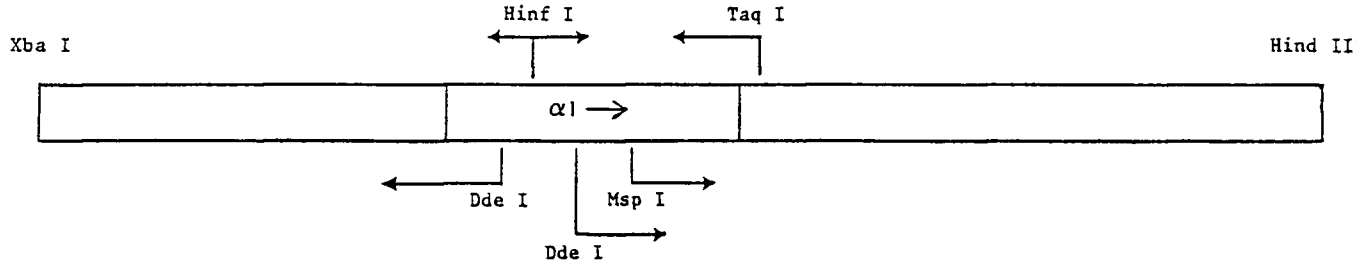


Figure 11. Strategy for the Sequence Determination of *matα1- oc* from 2935-10C.

The 2.3 kb XbaI-HindIII fragment from the *mat* locus-containing plasmid was digested with various enzymes. Fragments were 3'-end labelled with the Klenow fragment of DNA polymerase and one of four radioactive nucleotides. The choice of nucleotide and presence if necessary of unlabelled nucleotides in the reaction mixture was designed to allow labelling of one end of the fragment only. The sequence of the purified fragments were determined by the method of Maxam and Gilbert (1980).

sterile water, and 0.5 ml was removed, vortexed 1 minute, appropriately diluted (generally 10^{-3}) and plated in duplicate on plates selecting separately for the two haploid strains (and the diploid) or for the diploid alone. Efficiency of mating refers to the number of diploids with respect to that of the haploid of least concentration.

Sporulation

Sporulation was carried out by patching a freshly grown colony (grown on solid selective media) onto a sporulation plate. After 3 days at 30^0C (unless otherwise indicated; sporulation at room temperature was only complete after 8 days), a suspension of cells in 1M sorbitol (to prevent the lysis of un-sporulated cells) was examined under the microscope, and the percentage of asci with respect to the total number of cells was determined.

Budding pattern

A stationary phase culture was streaked on selective medium and individual cells were examined, for approximately 3-4 generations, under the microscope. If budding occurred consistently at the same site, the pattern was termed medial (Strathern *et al.*, 1979). If buds occurred at opposite ends of the cell, or at random sites on the cell surface, budding was termed random.

a-factor production

1.6 x 10^6 cells of strain RC757 [which contains the "super-sensitive" allele *sst2-1* (Chan and Otte, 1982)] was mixed with 2 ml of 0.75% molten agar at 45^0C and spread on selective plates. Cells to be tested were patched onto the lawn of RC757 cells and incubated overnight at 30^0C. A ring, or halo, around a patch (due to the absence of growth of RC757) indicated secretion of *a*-factor by the cells of the patch.

β-GALACTOSIDASE ASSAYS

Cells were grown to early to mid log phase and β-galactosidase activity measured after treatment with SDS and chloroform as described (Ruby *et al.*, 1983). For studies on the temperature-

dependence of gene expression, cells were grown at the described temperature, and harvested and analyzed at the usual temperatures.

Rapid, qualitative β-galactosidase tests were accomplished by floating a nitrocellulose filter containing patches of the cells of interest on chloroform for several seconds, then on a solution of approximately 0.2% X-gal in Z-buffer [60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM 2-mercaptoethanol (Ruby *et al.*, 1983)].

IN VITRO DNA-BINDING ANALYSES

In vitro DNA-binding was assessed by the method of Hope and Struhl (1985) with minor modifications. Radioactive α2 gene product was synthesized *in vitro* and its binding to DNA was monitored by the alteration of its electrophoretic mobility.

In vitro transcription of *MATα2* and mutants

Plasmids derived from pSP64- α2 were linearized with HpaI and the enzyme inactivated by heating. 0.5 μg of the DNA was then transcribed in a 25 μl solution of 40 mM Tris-HCl pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine-HCl_3, 0.5 mM each of rCTP, rUTP and rATP, 50 uM rGTP, 0.5 mM GpppG (or m^7GpppG), 10 mM DTT, 1.2 u/μl RNasin, and 0.15 u/μl SP6 RNA polymerase. Transcription was allowed to procede at 37^0C for 1-2 hours, and was then terminated by phenol extraction.

In vitro translation of *MATα2* and mutant mRNAs

Precipitated RNA derived from 0.2 μg DNA was redissolved in water (0.5 μl) and was translated for 50 minutes at 25^0C with 0.33 μl of a 1 mM solution of amino acids without methionine, 1 ul of 370 kBq/μl ^{35}S-methionine, 2.5 μl of wheat germ extract and 0.5 μl of 1M KAc.

Binding of wild-type and mutant α2 proteins to *BAR1* DNA

DNA-binding analysis of 0.25 μl of the translation mix was carried out in 15 μl of binding buffer (10 mM Tris-HCl pH 8, 5 mM $MgCl_2$, 50 mM KCl, 0.1 mM EDTA, 100 ug/ml BSA), with 8-40 nM pZV9 DNA (Figure 19) which had been digested with the appropriate restriction enzyme.

Incubation was at 4^0C overnight. 3 µl of loading buffer (binding buffer containing 20% glycerol and 1 mg/ml each of xylene cyanol and bromophenol blue) was added and the solution loaded onto a 0.75 mm thick, 5% polyacrylamide gel. Electrophoresis was carried out at 10-15 V/cm for 1.5 hours at 4^0C. The gel was fixed for 30 minutes in 10% acetic acid/40% methanol, for 10 minutes in 10% acetic acid/10% ethanol, and was then subjected to fluorography.

RESULTS AND DISCUSSION

The aim of the first part of this study was to demonstrate that the protein sequence predicted from the DNA sequence of the *MAT∝1* gene does correspond to the *MAT∝1* gene product.

SEQUENCE OF *mat∝1-oc*

The two complementation groups, *MAT∝1* and *MAT∝2*, within the *MAT∝* locus were first assigned to their respective coding sequence and transcripts by the analysis of *in vitro*-constructed linker mutants (Tatchell *et al.*, 1981; see Introduction). One group of mutants failed to complement a known *MAT∝1* mutant. This group of mutations lay within a major open reading frame which also corresponded to an *in vivo* transcript, thus identifying the putative *MAT∝1* gene product (Astell *et al.*, 1981). This was a 175 amino acid basic protein that is rich in hydrophobic residues. To verify that the proposed reading frame encodes ∝1 and thus correctly predicts its amino acid sequence, the DNA sequence of a *mat∝1* allele known to contain an ochre stop codon was determined.

Figure 12 displays the sequence of the *mat∝1* gene isolated from strain 2935-10C. This strain (provided by D. Hawthorne) contains a *mat∝1*, sterile mutation that is relieved in an ochre-suppressing background. The entire coding sequence was determined and two changes from the wild-type were noted; a G to T transversion converting the putative valine codon 22 to one for phenylalanine, and a C to A transversion converting serine codon 23 to an ochre stop codon. The remaining sequence was identical to that of the wild-type. As no other reading frames allow the generation of an ochre stop codon from these two changes, and no splicing is known to occur in this gene, the reading frame predicted must be the true reading frame.

PRODUCTION AND ANALYSIS OF SPECIFIC, TRUNCATED *MAT∝2* GENE PRODUCTS

Genotypic and phenotypic analyses of *in vitro* constructed *MAT∝2* mutants (Tatchell *et al.*,

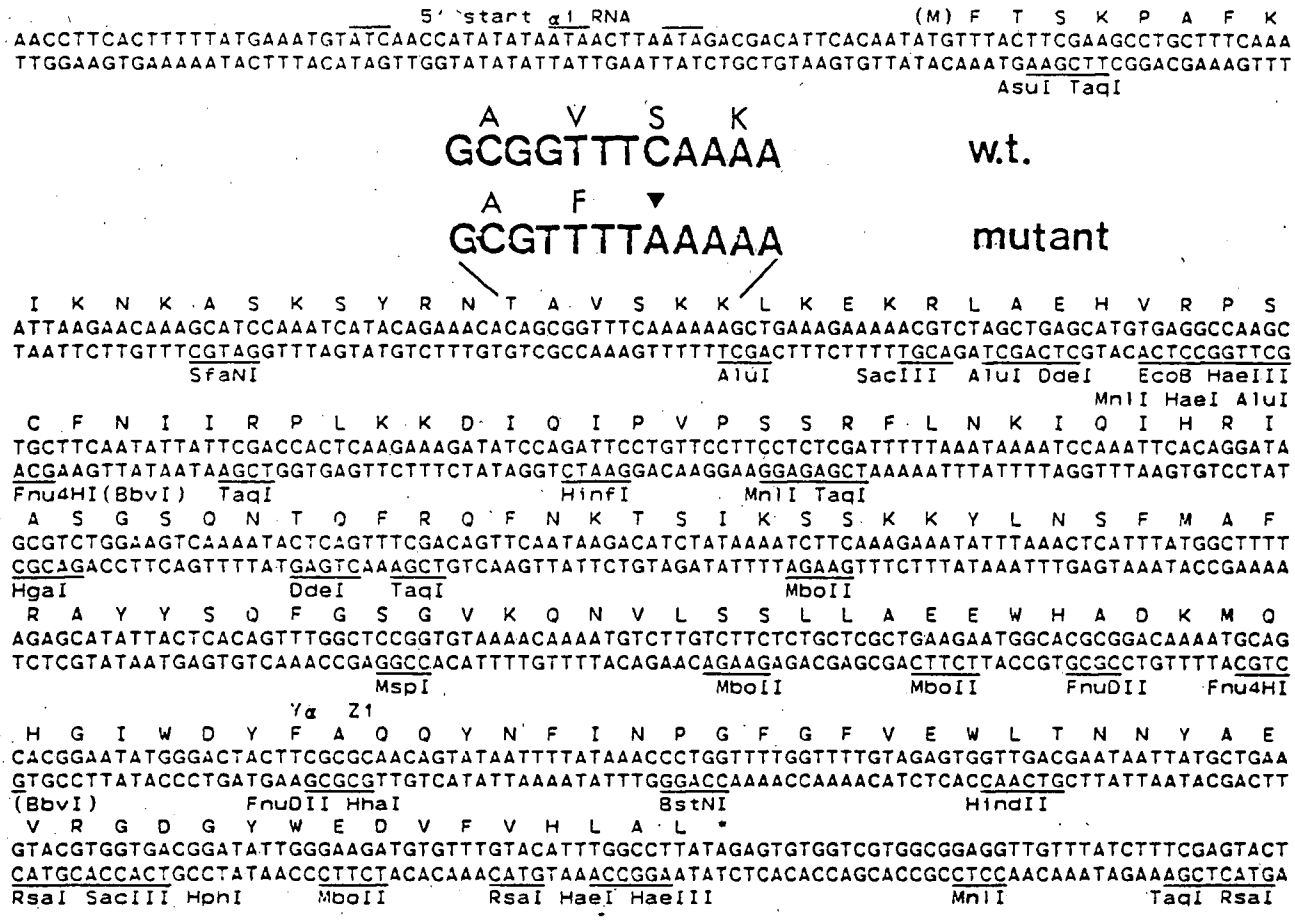


Figure 12. DNA and Predicted Amino Acid Sequences of the *matα1-oc* Allele of Yeast Strain 2935-10C.

The mutant allele was identical in sequence to that of the wild-type (w.t.) (Astell *et al.*, 1981) except for the region displayed in large type.

1981) have led to several conclusions. Most importantly, the activities of α2 in functioning to determine the α and a/α cell types were shown to be distinct (in that they are separable by mutation), and may be contained within separate domains of the gene product. This interpretation was complicated, however, by the fact that these mutations resulted in putative truncated *MATα2* products that contained several carboxy-terminal missense amino acid residues (which could conceivably alter the secondary structure or function of the protein), and by the possibility that re-initiation of protein synthesis had taken place in one or more of the mutants. (The translational stop point of several of the mutants preceded a methionine codon and it was suggested that the remaining portion of the gene may have been translated.)

To further analyze structural and functional aspects of α2 and to test the hypothesis of the division of α2 into separate functional domains in a more controlled manner, derivatives of *MATα2* that would produce distinctly truncated proteins with no additional C-terminal amino acid residues were constructed.

Construction of ochre stop codons in *MATα2*

For the purpose of constructing a set of ochre stop mutations in *MATα2*, oligonucleotides TAA1 and TAA2 were used, separately, by the method outlined in Figure 8, panel A, to produce a G to T change converting a glutamic acid codon to an ochre codon at amino acid residue 96 (termed *matα2-96-oc*), and a T to A change converting a tyrosine codon to an ochre one at position 156 (*matα2-156-oc*), respectively. These particular changes were made for the following reasons: 1) mutations producing approximately one-half and three-quarters of the gene product were desired, 2) both mutations produce the same termination codon, ensuring that an equivalent frequency of termination occurs, 3) the mutation at codon 156 eliminated the XbaI site at this location and simplified the mutant isolation procedure, 4) neither location immediately precedes a methionine codon (this reduces the rather unlikely chance that re-initiation of translation can occur).

Analysis of *matα2-oc* alleles in yeast

The 4.3 kb mutant *MATα* loci (*matα2-96-oc* and *matα2-156-oc*) were ligated into vectors YRp7-

CEN, YRp7, or YEp13 (see Figure 7) and transformed into yeast strains 23.75 (containing the *MAT*α2 mutant αx75) or S91-75 (αx75/*a*) (see Table II for genotypes). Some of the plasmids were re-isolated from the transformed yeast to permit confirmation of the mutant sequence. The activity of the *MAT*α2 mutants in haploid function, that is, their ability to determine the α cell type, was analyzed by examining the ability of 23.75 transformants to mate with *a* cells and by the production of *a*-factor by these transformants. *a*-factor secretion by the transformed cells was determined by comparing the extent of growth inhibition around a patch of transformed cells on a lawn of yeast strain RC757 (α *sst2-1*). *a*-factor arrests the cell division of α cells, and a mutation of the *SST2* gene in an α strain renders it "super-sensitive" to the pheromone (Chan and Otte, 1982). Classical haploid *MAT*α2 mutants are incapable of mating with *a* cells (MacKay and Manney, 1974a,b) and secrete *a*-factor (Sprague *et al.*, 1982), a product normally produced by *a* cells.

Analysis of diploid function (the ability to determine the *a*/α cell type) was carried out by assaying sporulation ability and the budding pattern of strain S91-75 transformants. Possession of sporulation competence and a polar (or random) budding pattern are unique to *a*/α cells; *mat*α2/*MATa* diploids behave in these respects like haploids in that they are unable to sporulate, and bud medially (Hicks *et al.*, 1977).

Table III displays the results of these analyses for the *MAT*α2 ochre mutants carried on YRp7. For the two haploid functions tested, both mutants behaved like true *mat*α2 alleles. That is, the mutant transformants were unable to mate with *a* cells and secreted *a*-factor, even though the mutant loci were on a high copy number plasmid. That the wild-type transformant also secreted some *a*-factor can be explained by the instability of the plasmid (Fitzgerald-Hayes *et al.*, 1982), since cells that have lost the plasmid would be expected to secrete *a*-factor.

The diploid phenotypes of the two ochre mutants were unanticipated. Although *mat*α2-*156-oc* behaved in all respects like a null mutant in the haploid and diploid cells, *mat*α2-*96-oc*, which should produce a shorter gene product than *mat*α2-*156-oc*, showed some α2 activity in diploid cells. That is, the *mat*α2/*MATa* diploid, when transformed with multiple copies of a plasmid containing the *mat*α2-*96-o*

TABLE III

PHENOTYPIC ANALYSIS OF *MATα2-OC* MUTANTS

	HAPLOID PHENOTYPE (in *matα2* strain)		DIPLOID PHENOTYPE (in *matα2/MATa* strain)	
Transformant	Mating[a]	*a*-factor[b]	Sporulation	Budding
YRpMATα	80-90%	+	20-30%	random (diploid)
YRpα2-96-oc	<0.4%	+ +	5-10%	random (diploid)
YRpα2-156-oc	<0.4%	+ +	0%	medial (haploid)
YRCpSUP4-oc	-[c]	n.d.[d]		
YRCpMATα/SUP4-oc	+	n.d.		
YRCpα2-96-oc/SUP4-oc	+	n.d.		
YRCpα2-156-oc/SUP4-oc	+	n.d.		

[a] Mating with the *a* strain D211-3A was assayed.
[b] + or + + refer to the extent of cell arrest of the lawn of cells (size of halo) around the patch tested.
[c] This and subsequent mating analyses were qualitative only.
- signifies no visible diploids formed, + signifies a distinct diploid patch of cells formed.
[d] Not determined.
Structures of the plasmids are given in Figure 7. Both YRp and YRCp plasmids are derived from YRp7, however the YRCp plasmids also contain *CEN3*.

allele, was rendered sporulation competent (although less than wild-type) and non-medially budding. This result is reminiscent of the phenotype of the truncated linker mutant α-x111 which supports sporulation in the diploid (Tatchell *et al.*, 1981), in contrast to mutant α-x75, producing a much longer gene product, which did not. It is possible, of course, that the nonsense codon of *mat-2-96-oc* is read-through with the resultant protein being active in the diploid tester strain. Ochre codon read-through in yeast can occur through the cytoplasmic factor [*psi* $^{+}$] (Cox, 1965) or through amplification of a normal $tRNA^{Gln}$ gene (Pure *et al.*, 1985). Genetic and biochemical evidence have also suggested that less α2 product is needed for diploid function than haploid function (discussed later); thus it is possible that the amount of putative read-through product in the diploid tester strain is sufficient for function, whereas not enough product is synthesized in the haploid tester for function (assuming that both tester strains have some ochre suppression).

Determination of the suppressability of *mat-2-96-oc* and *mat-2-156-oc*

To prove that the two contructed mutants were, in fact, functioning as ochre mutants, it was necessary to examine their suppressability by an ochre suppressor. Therefore, plasmids were constructed that contain, in addition to the mutated *MAT* loci, a tyrosine-inserting ochre suppressor tRNA locus, *SUP4-oc* (Goodman *et al.*, 1977). The structures of these plasmids are indicated in Figure 7. If the constructed *MAT-2* mutations are true ochre termination codons, their phenotypes should be suppressed in the *SUP4-oc*-containing plasmids.

The plasmids were transformed into strain 23.75 (*mat-2*) and the transformants were assayed for α mating ability. The transformants gave variable phenotypes, and for this reason, the plasmids were isolated from the yeast strain and analyzed by restriction digestion. A direct correlation was observed between rearrangement of the plasmid and a non-suppressing phenotype. Transformants that contained plasmids retaining their original structure showed significant levels of mating ability (see Table III, lower half). Thus, both *mat-2-96-oc* and *mat-2-156-oc* are true ochre nonsense alleles.

This has been the only case throughout this work that a rearrangement in yeast of an *in vitro*-constructed plasmid has been observed. Other than the *SUP4-oc* gene, the only major difference

between this and other plasmids is a direct repeat of the 0.34 kb HindIII-BamHI fragment of pBR322 DNA flanking the *SUP4-oc* gene. It is possible that, in yeast, recombination takes place frequently within this segment.

Construction and analysis of *matα2-16-oc*

To address the possibility that a low level of ochre codon read-through is occurring in some or all transformed yeast strains, and that this is accounting for the observed phenotypes of *matα2-96-oc* and *matα2-156-oc* in the diploid strain, a mutation converting glutamic acid codon 16 (GAG) to a TAA termination codon was constructed. Oligonucleotide SP7 was used in this case by the method illustrated in panel B of Figure 8. The rationale for this construction is that a termination codon near the beginning of the gene would yield an inactive product if no read-through is occurring. A glutamic acid codon was altered as this reproduces the change in mutant *matα2-96-oc*.

The diploid activity of *matα2-16-oc* in relation to the other ochre mutants was examined using an alternate, more quantitative strategy than assessment of sporulation efficiency or budding pattern. Specifically, a gene whose transcription is controlled by α2 and a1 was used as a monitor of α2 activity in the diploid. The *HO* gene, which is repressed by the combined action of α2 and a1, and which is required in the haploid cell for mating type switching (Jensen *et al.*, 1983) has been fused, in frame (at codon 38), with the *E. coli lacZ* gene and integrated into the yeast genome (Breeden and Nasmyth, 1975). The genotype of this strain, K1107, is indicated in Table II. As yeast do not normally synthesize β-galactosidase, the level of the enzyme in the cell should correlate directly with the level of transcription of the corresponding fusion gene and thus, inversely to the activity of the α2 gene product.

Mutants *matα2-16-oc*, *matα2-96-oc*, and *matα2-156-oc* were transferred to the multi-copy vector YEp13 and transformed into strain K1107 (*a ho::lacZ*). The results of β-galactosidase analyses on transformants are displayed in Table IV. *matα2-16-oc* has virtually no effect on the *ho::lacZ*-derived β-galactosidase activity of strain K1107, indicating that the mutant is non-functional with respect to the repression of *HO*. This suggests that there is no significant read through of an ochre codon in this strain. Both mutants *matα2-156-oc* and *matα2-96-oc* were partially active in repressing *ho::lacZ* (even

TABLE IV

REPRESSION OF *ho::lacZ* BY *matα2-oc* MUTANTS

Transformant	*ho::lacZ* β-galactosidase (strain K1107)
YEp13	68 (6.2)
MATα2	2.8 (1.4)
matα2-16-oc	57 (3.0)
matα2-96-oc	32 (7.8)
matα2-156-oc	23 (1.4)

β-galactosidase activity was measured as described in the Materials and Methods. Units are 1000 x A_{420} of β-galactosidase-catalyzed ONPG digestion / A_{600} of cell culture x time (min) x volume of cells (ml).
The results are a mean of the determined values of at least two individual transformants. Standard deviations are in parentheses.

when present on a single-copy plasmid; data not shown), which suggests that these genes do produce truncated gene products that can function in concert with a1. The observation that both *matα2-96-oc* and *matα2-156-oc* displayed partial diploid function activity in strain K1107 (*MATa*), whereas only *matα2-96-oc* supported sporulation in strain S91-75 (*matα2/MATa*) reflects the analogous observation that both mutants αx111 (producing the shorter gene product) and αx75 supported some sporulation in a *MATa/MATa* strain, whereas only αx111 allowed sporulation in a *matα2/MATa* strain (Tatchell *et al.*, 1981).

There are several additional explanations for the phenotypes of the ochre mutations in *MATα2*. For example, it is possible that a small amount of read-through is occurring in the strains used, but that the read-through product from *matα2-16-oc* is less active than that from *matα2-96-oc* or *matα2-156-oc* (due to the nature and location of the amino acid substitution). Alternatively, it is possible that the small amount of read-through full-length products from *matα2-96-oc* and *matα2-156-oc* are capable of forming functional heterodimers with their corresponding truncated gene products, whereas the full-length product from *matα2-16-oc* cannot interact with the truncated *matα2-16-oc* product.

Summary of analysis of nonsense mutations in *MATα2*

Truncated α2 products containing the first 15, 95 or 155 amino acid residues cannot repress *a*-specific genes. This is the case whether the mutant genes are present in the cell in single or multiple copies. In situations where α2 acts in concert with a1 (e.g. repression of *HO*), however, the *matα2-96-oc* and *matα2-156-oc* products are partially active. It is not clear whether this is due to read-through by the cellular translation apparatus, or if it is an intrinsic property of the truncated α2 proteins. In future similar experimentation designed to investigate the biological properties of truncated proteins, it may be advantageous to delete the unwanted portion of the gene or to construct a series of adjacent termination codons.

MUTAGENESIS OF THE REGION OF *MAT*α2 HOMOLOGOUS TO THE HOMEO DOMAIN

The Homeo domain homology

As discussed previously, there is significant homology between a sequence near the carboxy-terminus of α2 and a portion of the higher eukaryote homeo domain (Shepherd *et al.*, 1984). This segment also shares conserved characteristics of the helix-turn-helix DNA-binding motif found in many prokaryotic regulatory proteins (Laughon and Scott, 1984). It was thus hypothesized that this common region in the homeotic and α2 proteins could be responsible for a specific interaction with DNA (Laughon and Scott, 1984). The phenotypes of mutants *mat*α*2-96-oc* and *mat*α*2-156-oc* show that the carboxy-one-third of the α2 gene product is necessary for full activity. Hence, the region near the *MAT*α*2* 3'-terminus coding for the sequence homologous to the homeo domain and to the prokaryotic DNA-binding structure was chosen for analysis by targeted, semi-random, missense mutagenesis. It was hoped that the phenotypes of mutants would yield clues as to the significance of the homology with the prokaryotic regulatory proteins. This should provide insight into the function of at least a portion of the α2 gene product and, by inference, have implications for homeotic gene function.

Construction of semi-random missense mutations in *MAT*α*2*

Several methods of targeted, semi-random missense mutagenesis have been described (see Smith, 1985). These include incubation of the DNA with chemical reagents that either modify the base in such a way as to alter its base pairing specificity (Shortle and Botstein, 1983; Borrias *et al.*, 1976; Busby *et al.*, 1982; Hirose *et al.*, 1982) or that damage it so that it may no longer base-pair (Myers *et al.*, 1985). These modifications result in nucleotide misincorporation upon repair of the DNA. Alternatively, nucleotide misincorporation may be carried out *in vitro* using either nucleotide analogues that induce mispairing (Flavell *et al.*, 1974) or normal nucleotides under conditions that promote mispairing [e.g., the absence of a nucleotide (Zakour and Loeb, 1982), the use of nonproofreading polymerases (Zakour *et al*, 1984)].

Figure 9 displays the scheme by which a series of missense mutations was constructed in the region of *MAT*α*2* homologous to the homeo domain and to the prokaryotic bihelical motif. The

method involves the replacement of the wild-type sequence with a synthetic duplex oligonucleotide containing, at specific sites, a mixture of wild-type and mutant nucleotides. The strategy of "cassette" replacement of wild-type sequence has been used to generate a series of replacements of a specific codon (Wells *et al.*, 1985), and the synthesis of DNA containing a low level of mutant bases to provide a targeted, semi-random mutagen has also been described (McNeil and Smith, 1985). This method was chosen above other techniques of semi-random targeted mutagenesis because the target area and types of mutation can be precisely defined, and because a high efficiency of mutagenesis can be achieved (McNeil and Smith, 1985). Four criteria determined the conditions of mutagenesis. (1), the mutations were to be confined to the area most highly conserved with the homeo domain and DNA-binding structure (from codon Gln-175 to codon Thr-191). (2), termination codon formation was not desired. Hence, the identity of those base pairs at positions which would preclude the possibility of termination codon formation were left unaltered. (3), a minimal number of silent changes was desired and achieved by reducing the number of wobble site (or other positions of redundancy in amino acid coding) alterations. (4), one change per molecule was desired. Although this is achieved by setting the fraction of mutant bases in the synthetic mixture to 1/27 (for 27 positions), or 3.7%, the value of 4% was chosen so as to increase the liklihood of two changes as opposed to no changes (the probability of one change is only slightly smaller for the latter value). The distribution of probabilities for the different categories of mutant molecules is given by the binomial formula. That is, if x and y refer, respectively, to the fraction of wild-type and mutant bases in the synthetic mixture used at 27 positions, expansion of the term $(x + y)^{27}$ yields individual terms describing the probability of occurrence of each species (i.e. no changes, x^{27}; 1 change, $27x^{26}y$, 2 changes, $351x^{25}y^2$, etc.).

The expected and observed efficiency of the mutagenesis is listed in Figure 10d. Of approximately 75 clones analyzed by DNA sequence determination, 57%, were mutant, with 37% of the total clones containing a single base-pair substitution. Changes were noted at 22 out of a potential 27 positions, with no apparent bias towards the type or location of nucleotide substitution. Fourteen of the mutants were chosen for further study.

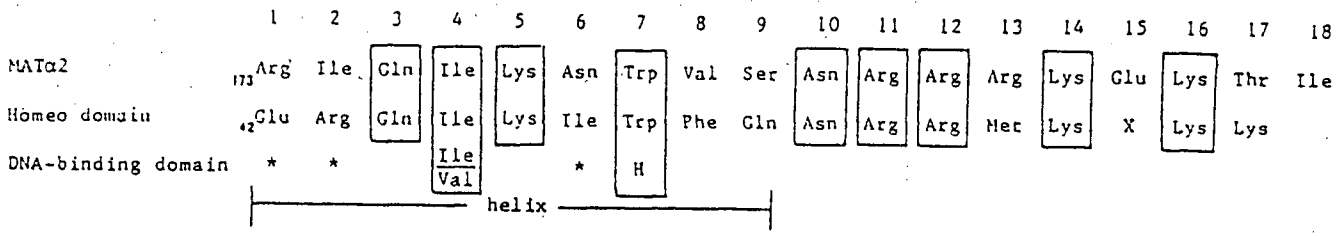
PHENOTYPE DETERMINATION OF *MATα2* MISSENSE MUTANTS

In vivo assay of mutants

For all *in vivo* assays, the approximately 4 kb HindIII-EcoRI fragment containing the mutated *MATα* locus was transferred from the pEMBL8(-) vector, in which the mutant was generated and characterized by DNA sequence determination, to YCp50 (see Figure 7) and transformed into yeast strains SP2 (*MATa ste6::lacZ*), SP3 (*MATa/MATa*) or K1107 (*MATa ho::lacZ*) (see Table II). SP2 contains a fusion of the yeast *STE6* gene [which is repressed by α2 and which is required for mating in the *a* cell type (Wilson and Herskowitz, 1984)] with the *E. coli lacZ* gene. The construct replaces the resident *STE6* in the strain 1160 genome (Wilson and Herskowitz, 1984). However, to eliminate both α2 and a1 activity from the strain, a *MAT* locus containing a deletion of the entire coding sequence (constructed by P. Siliciano and K. Tatchell *in vitro*) was incorporated into the strain through a cross with strain MH52-3C (see Materials and Methods).

Mating and sporulation abilities, and *ho::lacZ* and *ste6::lacZ*-derived β-galactosidase levels of the transformants were monitored. Several plasmids were re-isolated and subjected to DNA sequence analysis to ensure that the correct mutant was being studied. Figure 13 displays the β-galactosidase levels produced by the transformed strains as well as their sporulation and mating efficiencies. To confirm that the monitored β-galactosidase levels correspond to transcriptional activity of the respective genes, RNA levels from several transformants were determined by Northern analysis, using an RNA probe consisting of the anti-sense strand of a portion of the *lacZ* gene. The results of these experiments are indicated in Figure 14. The largest RNA from each yeast strain corresponds in intensity to the β-galactosidase activity of the respective transformants, indicating a correlation between the β-galactosidase levels and transcriptional activities of the *lacZ*-containing genes. Two other RNAs are detected in the autoradiogram. These appear consistently, but are not detected when poly-A^+ RNA is tested (not shown). It is not known how the poly-A^- RNAs are related to β-galactosidase mRNA.

a



b

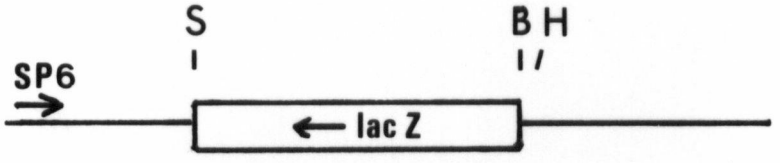
Clone	Position	Change	Haploid Functions: ste6::lacZ β-Gal	Haploid Functions: Mating	Diploid Functions: ho::lacZ β-Gal	Diploid Functions: Sporulation
101	7	Trp (TGG) → Arg (CGG)	7.0	-	13.7	-
37	10	Asn (AAT) → Asp (GAT)	6.7	-	4.6	-
5	4	Ile (ATC) → Asn (AAC)	6.5	-	10.7	-
66	3	Gln (CAA) → His (CAT)	6.3	-	1.6	-
52	10	Asn (AAT) → Thr (ACT)	5.9	-	0.6	+/-
96	16	Lys (AAA) → Ile (ATA)	4.2	-	0.2	+
56	14	Lys (AAA) → Asn (AAT)	3.5	-	0.2	+
58	6	Asn (AAC) → His (CAC)	0.8	+/-	0.2	+
41	18	Ile (ATA) → Arg (AGA)	0.5	+	0.1	+
105	5	Lys (AAA) → Asn (AAT)	0.4	+	0.2	+
82	11	Arg (AGA) → Ser (AGC)	0.3	+	0.1	+
59	8	Val (GTT) → Ala (GCT)	0.3	+	0.2	+
	15	Glu (GAA) → Ala (GCA)				
102	19	Thr (ACA) → Pro (CCA)	0.2	+	0.1	+
38	17	Thr (ACA) → Ser (TCA)	0.2	+	0.2	+
Wild type			0.2	+	0.2	+
Vector alone			6.4	-	11.8	-

Figure 13. *MATα2* Mutations in the Region of Homology with the Homeo Domain and their Phenotypes.

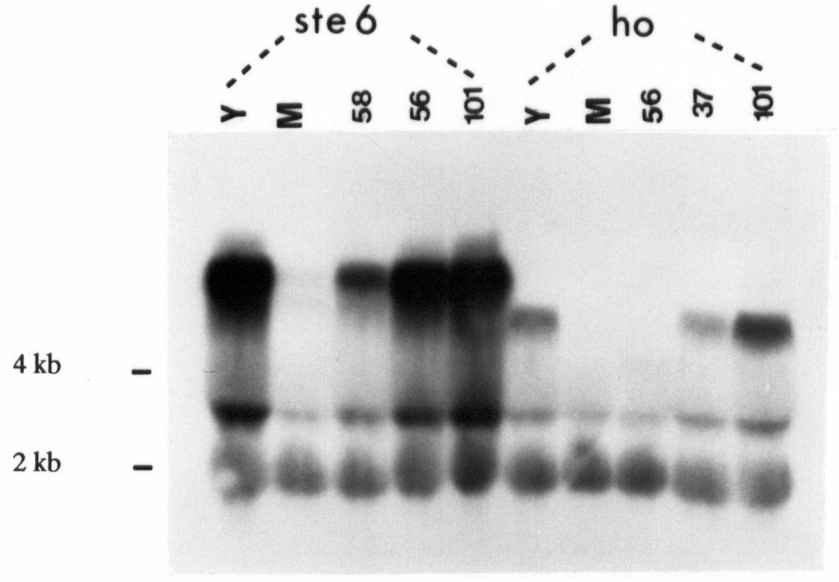
a Aligns a portion of *MATα2* sequence with the carboxy-terminus of the homeo domain and the carboxy-terminal helix of the prokaryote DNA-binding domain consensus (see Figure 6). Conserved residues are boxed, H refers to a hydrophobic residue, *, to a residue thought to contact DNA, and X, to a residue not conserved within the homeo domain. Changes at residues indicated by the numbering system in **a** and their corresponding phenotypes are shown in **b**. β-galactosidase units are described in the Materials and Methods section. The value indicated is a mean of at least 2 isolates, each in duplicate. For both mating and sporulation analyses, - refers to values <1% of controls, +/- refers to values 20-80% of controls, and + refers to values >90% of those of the wild-type.

Figure 14. Northern Analysis of *ste6::lacZ* and *ho::lacZ* RNA.

18 µg of total cellular RNA either from strain SP2 (*ste6::lacZ*) or K1107 (*ho::lacZ*) transformed with the indicated plasmids were subjected to gel electrophoresis and hybridization with SP6-generated radioactive RNA from plasmid pSP65-alacZ (1.9). This plasmid contains the 5'-terminal 1.9 kb (starting at codon 8) of *lacZ*, in such an orientation in pSP65 so as to allow anti-sense RNA to be synthesized from the SP6 promoter. The highest molecular weight RNAs in each strain correspond to *lacZ*-containing mRNA. The lower molecular weight RNAs are poly-A^- (data not shown) and do not appear to be regulated. Restriction site abbreviations are as follows: S, SstI; B, BamHI; H, HindIII. Plasmid abbreviations are: Y, YCp50; M, YCpMATα; 58, 56, 37, and 101 are YCpMATα containing the 58, 56, 37 or 101 *matα2* alleles respectively (see Figure 13).



pSP65 – alacZ (1.9)



Several observations can be made concerning the phenotypes of the mutants. Many of the mutations severely affect α2 function. Also, several mutations that cause a reduction or elimination of the haploid function of α2, i.e., repression of *STE6* and the ability to support mating, retain most or all of their diploid function in supporting sporulation and in repression of *HO*. The reciprocal phenotype, retention of haploid function with loss of diploid function, is not seen. Finally, mutations causing the most severe effects generally result from replacement of those residues which are strongly conserved within the homeo domain and within the DNA-binding structure consensus. Conversely, mutations having little or no effect on α2 function generally occur at variable positions within the homeo domain or are located outside the DNA-binding helix homology.

That there exist mutations affecting haploid function more severely than diploid function can be accounted for in several ways. Firstly, it is likely that less α2 product is needed to carry out diploid functions than to carry out haploid functions. The *MATα2* transcript in the diploid cell type is present at one-fifth the level of that in the α cell type (Nasmyth *et al.*, 1981; Hall *et al.*, 1984). Evidence supporting the possibility, however, that even the reduced amount present in the diploid is in greater excess (for function) than that in the haploid, is provided by Siliciano and Tatchell (1986), who described a diploid strain containing an *in vitro*-altered *MATα* locus which expressed an undetectable level of *matα2* transcript, yet which sporulated at normal efficiency. Another *MATα* mutant in a haploid produced 10% of the wild-type (haploid) level of *matα2* transcript but was incapable of repressing *a*-specific genes (Siliciano and Tatchell, 1984). Thus, the amount of α2 product required for haploid function appears to be greater than five times the amount required for diploid function. Consequently, a partially active α2 protein may have less of an effect on the phenotype of a diploid than on that of a haploid.

A second potential reason for a mating$^-$ / sporulation$^+$ phenotype involves the possibility that α2 in the diploid regulates gene expression through formation of a dimer with the *MATa1* gene product. The evidence for this lies in the codominant mode of regulation by α2 and a1, and by comparison of the sequences of the predicted DNA target sites for α2 and for a1/α2 (Miller *et al.*, 1985). Thus, although

both sites contain a related inverted repeat structure, the spacing between the repeats of the a1/α2 operator is much smaller (they are separated by 11 bp rather than by 24 bp, as for the α2 operator), and the repeats themselves may not be identically related to each other. It was therefore suggested that α2 binds to DNA as a homodimer in its haploid function and as a heterodimer with a1 in its diploid function. In this respect, it is reasonable to suggest that a mutation inactivating α2 in its haploid function might cause less of an effect in the diploid when the protein forms a complex with the wild-type a1 protein. It is also possible that other proteins or factors unique to diploids might act on α2 in the diploid in a similar fashion.

A third possible explanation for the phenotypes of the mutants showing different activities is that, in the genes analyzed so far, those controlled by a1/α2 contain multiple potential DNA target sites upstream of the coding sequence, whereas those controlled by α2 contain a single DNA target site (Miller *et al.*, 1985a). An exception to this is the single a1/α2 target site located upstream of the *MATα1* gene. [Lack of repression of *MATα1* in a diploid, however, does not affect any known diploid phenotype (Siliciano and Tatchell, 1984; Ammerer *et al.*, 1985)]. It is possible that a redundant number of target sites allows a wild-type level of repression by a partially inactivated repressor.

Finally, it is possible that certain residues may be less important for the diploid function of α2 than for its haploid function. That is, the mutant amino acids may be in a region important for α2/α2 dimerization but not a1/α2 dimerization, they may be more critical for haploid function DNA-binding than for diploid function DNA-binding, or they may be more critical to binding haploid regulation-specific transcription factors. In any case, the fact that there exist mutations that have different phenotypes in haploid and diploid cells suggests that the mechanisms of action of α2 differ, and/or that the regulation of the two sets of target genes is quantitatively different.

The most markedly defective mutants affecting haploid and diploid function involve amino acid residues which are strongly conserved within the homeo domain and with the region corresponding to the prokaryotic DNA-binding structure (positions 1 to 9 in Figure 13, corresponding to α2 residues 173-181). This supports the suggestion, based on the amino acid sequence homologies

and the regulatory nature of the gene products, that the homeo domain, α2, and the prokaryote DNA-binding helical structure share structural and functional similarities. In particular, mutation of the two residues in α2 which correspond to the residues in the prokaryote DNA-binding helical structure that are most highly conserved, Ile/Val 4 and the hydrophobic residue 7 (Figure 13), gives rise to the two most severely defective phenotypes. Ile/Val 4, in the C-terminal helix of proteins of known three dimensional structure, is in van der Waal's contact with the fifth residue of the N-terminal DNA-binding helix (see Figure 4). Substitution at this position with the more polar residue, asparagine (mutant 5), would be expected to disrupt this interaction and result in an altered or destabilized structure. The hydrophobic residue 7 is buried within the protein, and mutations that alter the corresponding residue in λ *cro* (Eisenbeis *et al.*, 1985), in Trp repressor (Kelley and Yanofsky, 1985) and in λ *cI* (Nelson *et al.*, 1982) all inactivate the protein, presumably by destabilizing the helix-turn-helix structure. Thus, the corresponding α2 mutation with a change to the charged residue arginine from tryptophan (mutant 101) would be expected to be particularly disruptive.

Replacement of residue 3 (Gln) by histidine (mutant 66) results in substantially defective α2 function. Although the corresponding residue is not conserved within the prokaryotic DNA-binding proteins, mutations that alter the corresponding position in Trp repressor (Kelley and Yanofsky, 1985), in Lac repressor (Weber *et al.*, 1972) and in the Tn10 Tet repressor (Isackson and Bertrand, 1985) greatly reduce activity. The precise function that this residue plays in the structure of the DNA-binding helical structure is unknown. It appears to be partially buried inside the protein, and may contact a point on the sugar-phosphate backbone of the target operator DNA (Schevitz *et al.*, 1985) and/or be important in maintaining the structure of the protein (Kelley and Yanofsky, 1985).

Three other mutations were obtained within the region of α2 which is homologous to the DNA-binding helix-turn-helix. Replacement of Asn-6 by His (mutant 58) has a slight effect on haploid functions and no effect on diploid functions; replacement of Lys-5 by Asn or Val-8 by Ala (with a corresponding mutation of Glu-15 to Ala) have no effect on haploid or diploid functions. Mutations at position 6 in λ *cI* (Hecht *et al.*, 1983) and in Trp repressor (Kelley and Yanofsky, 1985) significantly

reduce repressor activity, which is consistent with the suggestion that this residue contacts DNA in the major groove (Anderson *et al.*, 1987; Lewis *et al.*, 1983; Ohlendorf *et al.*, 1982). It is not known whether substitutions of this residue affect the specificity of DNA-binding; it is speculated that the DNA-protein contact at this point is more important for affinity of binding (Ebright, 1986a). Substitution of the corresponding residue in CAP (Arg-185), however, by Lys or Leu has little or no effect on *in vivo* or *in vitro* activity (Gent *et al.*, 1987). It is not clear whether this lack of effect is an indication of a unique structure at this position of CAP, or whether it is due to the nature of the particular substitutions. It is possible that the small reduction in activity of mutant 58 (Asn-6 → His) relates, also, to the nature of the amino acid replacement (thus, for the particular steric and electrostatic requirements of residue 6, histidine may be an adequate substitute for asparagine). Alternatively, this residue and its surrounding sequence in ∝2 may have a different structural and/or functional role. Finally, it is possible that ∝2 contacts the DNA at more sites (or with greater affinity at the same sites) than the prokaryotic structure, and that a decreased affinity of DNA contact at position 6 is less deleterious to ∝2 than to the prokaryotic structure (other than that of CAP). Lack of mutant phenotype for replacement of residue 5 may be explained in a similar fashion to the small effect of replacement of residue 6. It is interesting, however, to note that two mutations at the corresponding position in λ *cI* (Gly-48) actually improve repressor activity relative to wild-type activity (Hecht and Sauer, 1985; Nelson and Sauer, 1985). In this case, it was suggested that the substituted residues (Asn and Ser) made favorable contacts with the DNA that are not made by the wild-type protein. A replacement of Val-8 by Ala in 2 (mutant 59) is conservative, and the hydrophobic character of the residue is retained. It is not surprising, therefore, that it effects no phenotypic alteration. Mutant 59 contains a second change (Glu-15 → Ala) at a residue that is outside the prokaryotic helix homology and that is not conserved within the homeo domain.

Mutations at residue 10, adjacent to the C-terminal DNA-binding helix homology, (mutants 37 and 52) significantly diminish repressor activity. These have counterparts in λ *cI* (Hecht *et al.*, 1983) in which the corresponding residue (glycine) is changed to Cys or Asp. For both λ *cI* and ∝2, substitution of this residue by a negatively charged residue (aspartic acid in both cases; ∝2 mutant 37) results in a

more severe affect on function than does substitution by a less polar amino acid (cysteine, threonine; α2 mutant 52). The conformation of this position in the prokaryotic DNA-binding proteins of known structure is not conserved. In Trp repressor, λ Cro, and CAP, this residue still makes up part of the C-terminal DNA-binding helix (Schevitz *et al.*, 1985; Anderson *et al.*, 1981; McKay *et al.*, 1982), whereas in the 434 and λ repressors, this residue is part of an irregular turn before the next helix (Anderson *et al.*, 1987; Pabo and Lewis, 1982).

Mutants 96 and 56 (at positions 14 and 16) display interesting phenotypes in being mating$^-$ /sporulation$^+$. As mentioned, the corresponding region of prokaryotic DNA-binding proteins of known three-dimensional structure is not conserved. In both λ and 434 repressors, this region connects the C-terminal DNA-binding helix with the adjacent helix through an irregular structure (Anderson *et al.*, 1987; Pabo and Lewis, 1982); in λ Cro, this segment forms a very short turn followed by a β-sheet (Anderson *et al.*, 1981); in Trp repressor the C-terminal DNA-binding helix continues for 5 more residues beyond the helix-turn-helix homology and is then followed by a turn to the next helix (Schevitz *et al.*, 1985); and in CAP, the C-terminal helix continues for an additional 3 residues, followed by a short turn to a short β-sheet (McKay *et al.*, 1982). In the cases of 434 and λ repressors and λ Cro, this region has been implicated in contacting the DNA sugar-phosphate backbone (Anderson *et al.*, 1987; Pabo and Sauer, 1984), and, in the cases of λ repressor and λ Cro, in making potentially specific base-pair contacts (Pabo and Sauer, 1984). The side chains (as well as peptide NH groups) of 434 repressor Arg-41 and Arg-43 residues (corresponding to the lysines at positions 14 and 16 of Figure 13) are in position to interact with DNA phosphate groups (Anderson *et al.*, 1987). In addition, the aliphatic chain of Arg-41 in one subunit of 434 repressor lies against the aromatic ring of Phe-44 of the other subunit, and may be important in maintaining the proper orientation of the individual DNA-binding If this region in α2 is important for correct dimer orientation, then it is not surprising that substitution of residues 14 or 16 affects the haploid function of α2 (in which the protein may function as an α2:α2 dimer) more than diploid function (when the protein may function as an α2:a1 conjugate). If the region contributes affinity of binding to DNA (as in the cases of 434 and λ repressors and λ Cro), then one

would expect their conservation in DNA-binding structures, and their mutation to be deleterious to repressor activity. A lessened affinity of the α2 dimer for DNA could result in a more pronounced defect in haploid function than diploid function for the reasons previously discussed. Clearly, however, the true functional nature of these residues in α2 and in the homeo domain awaits physical investigation.

In conclusion, this first set of data, resulting from a set of partially random mutants, shows that some amino acid residues in α2 that correspond to residues in the homeo domain, are important for function. Further, a subset of these residues that is potentially homologous to the prokaryotic regulatory protein DNA-binding structure is especially crucial for function. This is supportive of the suggestion of structural and functional homology between the homeo domain, α2, and the prokaryotic helix-turn-helix DNA-binding structure.

The fact that α2 activity is not always affected when residues within the region corresponding to the homeo domain are mutated suggests either that the substitution fulfills the required function for that particular residue, that the homology of the α2 structure with the homeo domain is not functionally significant, that conservation of the residue within the homeo domain is unrelated to function, or that the homeo domain has an additional role in higher eukaryotes which is not a function of α2 in yeast.

As the mutational analysis has supported the possibility of a structural/functional homology between a portion of α2 and the prokaryotic DNA-binding bihelical structure, more extensive studies were undertaken to further define the relationship. The following questions were addressed: 1) Do additional mutations in this region, particularly in the first two residues of the second helix, affect DNA-binding specificity? 2) Are the physical characteristics of the α2 missense mutants similar to those of the equivalent mutants in the prokaryotic structures? 3) Are the mutants defective in DNA-binding, and if so, is the deficiency due only to an altered DNA-binding structure?

SITE-SPECIFIC MUTAGENESIS WITHIN THE HOMEO-DOMAIN HOMOLOGY

Construction of mutants

Mutations converting Arg-1 and Ile-2 of the C-terminal DNA-binding helix homology (residues 173 and 174 of the protein) to either Glu-Arg (mutant 11), Glu-Ile (mutant 10) or Gln-Ser (mutant 8), were constructed as outlined in Figure 8, panel C, and as described in the Materials and Methods section. This procedure differs from the previous methods of site-specific mutagenesis in that the template DNA contains a small fraction of uracil residues, derived through passage of the DNA through an *E. coli* strain lacking the enzymes for dUTPase (*dut*$^-$) and uracil N-glycosylase (*ung*$^-$) (see Kunkel, 1986). The elevated level of dUTP in the cell competes with TTP for incorporation into DNA, and the subsequent uracil-containing DNA is maintained due to the lack of uracil N-glycosylase activity. [The hydrolysis of the uracil residue by this enzyme produces an apyrimidinic (AP) site, which is the target for an AP endonuclease that produces strand breaks.] Transformation of wild-type *E. coli* with *in vitro*-synthesized double-stranded DNA selects for progeny of the strand complementary to the template uracil-containing DNA. The resultant high efficiency of mutant production (20-50%) allows screening by DNA sequence determination.

The rationale for the choice of substitutions at positions 1 and 2 of α2 described above is as follows. The residues at the corresponding positions in the homeo domain are Glu-Arg, and it was of interest to determine what effect the presence of one or both of these residues has on α2 function. The sequence of the first two amino acids of the C-terminal DNA-binding helix of λ *cI* is Gln-Ser. The three-dimensional structure of this protein and its interaction with DNA have been extensively characterized. Again, it was of interest to determine how these residues affect α2 function. In addition, the properties of all the mutant residues differ substantially from those of the wild-type, and the substitutions would be expected to have a significant effect on α2 activity if the positions are important.

The mutants were constructed in a pEMBL vector containing only the 0.5kb XbaI fragment of the *MAT*α locus. This fragment was then re-inserted into the *MAT* locus [without the 0.8kb XbaI fragment, which contains no coding or transcriptional information (see Figure 1)] within the yeast

vector, YCp50 (see Figure 7). SP2 and K1107 yeast strains were then transformed and assayed, as previously described, for β-galactosidase and mating activities.

Arg-1 and Ile-2 mutant phenotypes

Table V displays the β-galactosidase levels and mating efficiencies of yeast transformed by mutants containing altered residues at the positions corresponding to the first two residues of the C-terminal DNA-binding helix. Mutation of these residues results in a limited reduction of α2 activity in its haploid function, whereas no observable defect is apparent in diploid function.

The general trend of the mutations described in this work, which alter the haploid function activity of α2 more severely than its diploid function, is thus maintained for these mutants. Possible explanations for this difference have been discussed previously.

Mutations at the corresponding positions of CAP (Ebright *et al.*, 1984; Gent *et al.*, 1987), Lac repressor (Ebright, 1986b), Trp repressor (Kelley and Yanofsky, 1985),λ Cro repressor (Eisenbeis *et al.*, 1985), λ *cI* (Hecht *et al.*, 1983) and Tn10 Tet repressor (Isackson and Bertrand, 1985) have been isolated or constructed. The effects on activity of the resultant proteins are highly variable, ranging from a 23% decrease in CAP activity (Glu → Leu at position 2) to nearly complete loss of activity of Tn10 Tet repressor (Gln →Tyr at position 1, or Pro → Leu at position 2), λ Cro (Gln → Leu at position 1, or Ser → Ala at position 2), CAP (Arg →Lys or Leu at position 1), and λ repressor (Ser →Leu at position 2). Interestingly, Gln →Leu at position 1 of λ repressor, an identical change to that constructed in λ Cro, only partially inactivated activity. Double mutants have not been described. The mutants analyzed further for alterations in specificity (as opposed to affinity) of DNA-binding were shown to bind preferentially to mutant operators (Ebright *et al.*, 1984, Ebright, 1986b).

It is difficult to assess the significance of the partial inactivation of activity observed for the replacements of positions 1 and 2 of the C-terminal DNA-binding helix homology of α2. The limited reduction in activity observed for even the two double mutants in this region seems to be inconsistent with a crucial role of these two residues in binding DNA specifically. Some divergence in structure and function of the postulated homology between α2 and the helix-turn-helix is not unlikely, considering the

TABLE V

PHENOTYPIC ANALYSES OF MUTANTS 8, 10 AND 11

Mutant	Sequence[a]	Haploid Functions		Diploid Functions
		ste6::lacZ β-gal[b]	Mating[c]	*ho::lacZ* β-gal[b]
WT	Arg-Ile	1.3	0.8	0.08
8	Gln-Ser	4.5	0.2	0.08
10	Glu-Ile	1.8	0.5	0.1
11	Glu-Arg	2.1	0.8	0.1
Vector alone		16.1	<0.01	2.9

[a] The sequence at positions 1-2 of Figure 13 (corresponding to residues 173-174 of 2) is indicated.
[b] Refers to β-galactosidase activities derived from these genes. *ste6::lacZ* is from strain SP2 and *ho::lacZ* is from strain K1107. The values indicated are the means of at least two indiviual transformants.
[c] Mating efficiency with strain D311-3A is indicated.
[d] The cell density and colorimetric assays for this set of β-galactosidase activity determinations were performed using a different spectrophotometer from that used for the β-galactosidase determination of strain SP2. The values for individual strains cannot therefore be directly compared.

evolutionary distance between yeast and bacteria, and the differing complexity of transcriptional regulatory mechanisms. In particular, the larger size of the yeast genome relative to that of *E. coli* necessitates an increased degree of discrimination by a regulatory protein between operator and non-operator DNA. Although this may be accomplished by an increased repressor concentration in the nucleus or by cooperativity (for example, with transcription factors), it may also be accomplished by additional operator contacts by the protein. The fact that the $\alpha 2$ operator [approximately 30 bp (Johnson and Herskowitz, 1985)] is longer than the average prokaryotic operator (approximately 17-20 bp) may indeed suggest that more contacts are made with the protein. Even for prokaryotic regulatory proteins, the first two residues of the C-terminal DNA-binding helix are not the sole mediators of specific binding. In addition to a number of residues throughout λ repressor and Cro that are postulated to contact DNA, there is evidence that the overall conformation of the protein plays a part in specificity of binding. [A mutation in the fifth helix of λ *cI* has been shown to increase both the affinity and specificity of the protein:DNA interaction. This helix is involved in dimerization, and the mutation is postulated to alter the global structure of the domain. (Nelson & Sauer, 1985)]

A second interpretation of the observed phenotypes of the $\alpha 2$ mutations is that the amino acid residue substitutions are capable of the same specific interactions as those of the wild-type. Considering the severity of alterations constructed, and the fact that both sets of mutations had similar effects, this possibility is less likely.

Although the affinity of these mutants for DNA does not appear to be drastically reduced (as deduced by the *in vivo* activities), the specificity of binding is not known. That determination would have to be carried out either *in vitro*, using a sensitive DNA-binding assay with purified components, or by a systematic analysis *in vivo* of a variety of mutated operators, since the low level of reduction in repressor activity would preclude the possibility of using standard genetic procedures for isolating operators of greater binding potential. (This last problem might potentially be overcome, however, by further mutating $\alpha 2$ to reduce its affinity for the wild-type operator more fully.)

It is interesting that the two residues corresponding to the beginning of the C-terminal DNA-binding helix in the homeo domain are very highly conserved. If this region is responsible for binding DNA in a manner similar to that of the prokaryotic DNA-binding domain, this would imply either that all homeo-domain-containing proteins bind to the same sequence, or that other residues, or interactions with other proteins, determine specificity of binding.

THERMAL PROPERTIES OF *MATα2* MUTANTS

To further investigate whether the sequence homology between a portion of α2 and the helix-turn-helix is reflected in its secondary and tertiary structure, a physical characteristic of the mutant phenotypes (its dependence on temperature) was assessed and compared to that of DNA-binding mutant proteins of known structure. It was reasoned that the thermal properties of the mutants should be a function of their structure, or stability, and might indicate the probability of structural homology.

Thermal properties of phage λ repressor mutants

In an attempt to locate the region in λ *cI* involved in contacting DNA, Hecht and Sauer and co-workers (Hecht *et al.*, 1983; Hecht *et al.*, 1984; Hecht and Sauer, 1985) analyzed the repressor activities and thermal properties of a series of proteins containing amino acid substitutions throughout the amino-terminal DNA-binding domain of the molecule. Substitutions that substantially reduced repressor activity could be divided into two classes, relating to their location in the three-dimensional structure. Mutations of the first class altered internal, or buried residues, and these substitutions were located throughout the domain. Mutations of the second class affected solvent-exposed residues and were clustered in helices 2 and 3. (Substitutions by proline that reduced repressor activity, however, occurred both internally and externally, within and outside of helices 2 and 3.) It was reasoned that the cluster of external inactivating mutations defined the DNA-contact point of the protein, whereas those located internally inactivated DNA-binding indirectly by destabilizing or altering the global structure of the domain. This thesis was examined by analyzing the activity of the mutant repressors *in vivo* as a function of temperature and by the determination of the thermal stability of the proteins *in vitro* by

physical and biochemical means (that is, by differential scanning calorimetry, and by circular dichroism and thermolysin susceptibility changes at varying temperatures). Their results are displayed in Table VI. With only one exception (Ala-49 → Val), mutations occurring at amino acid residues on the surface of the protein exhibited a temperature-independent phenotype *in vivo* as well as (for the mutants analyzed) a wild-type pattern of thermal stability (given in the table as thermolysin sensitivity). Conversely, all those mutations occurring at internal or partially buried residues displayed temperature-dependent phenotypes and reduced thermal stabilities. Therefore, the *in vivo* temperature dependence of α2 activity for several of the constructed mutants was assessed to determine if the mutations within the putative helix-turn-helix region of α2 affect activity in a similar fashion.

Mutant and wild-type α2 activities as a function of temperature

Figure 15 displays the results of the determination of α2 activity in repressing *ste6::lacZ* and *ho::lacZ* expression as a function of temperature (at 25^0C and 30^0C). Replacement of residues 3, 4 and 7, which, together with residue 8, would be buried in a helix-turn-helix structure (the corresponding λ Cro helix-turn-helix model is displayed in Figure 16; the red side chains are internal) results in significantly decreased expression of *ste6::lacZ* and *ho::lacZ* when cells are grown at 30^0C rather than 25^0C. The temperature effects are particularly striking for diploid function (repression of *ho::lacZ*) where the activity of mutant 5 (Ile-4 → Asn) is reduced by about 75% and that of mutant 101 (Trp-7 → Arg) is reduced by close to 100% upon a temperature shift from 25^0C to 30^0C. [Substitution of the corresponding residue of λ *cI* (Leu-50 → Cys) also causes a temperature dependent phenotype; see Table VI]. Replacement of residue 3 (Gln → His) effects a biological *ts* phenotype in that the mutant, inactive in sporulation at 30^0C (in strain SP3; data not shown), sporulates at 64% of the wild-type level at 23^0C. Amino acid substitution of residues 1, 2, 5 or 6 results in activities that are virtually independent of temperature. These residues would be expected to lie on the surface of the helix-turn-helix (Figure 16, green side chains). Thus, the phenotypic characteristics of these α2 mutants are consistent with the existence of a structural homology with the prokaryotic DNA-binding helix-turn-helix.

TABLE VI

THERMAL PROPERTIES OF WILD-TYPE AND MUTANT λ REPRESSORS

	Proteolysis Temperature (^{0}C)	Activities *in vivo*		
		30^{0}C	37^{0}C	42^{0}C
Wild-type	47-51	R	R	R
Surface substitutions				
Lys-4 → Gln	47-51	S	S	S
Gln-33 → Ser	45-47	S	S	S
Gln-33 → Tyr	51-56	R	R	R
Gly-43 → Glu	47-51	S	S	S
Gln-44 → Leu	47-51	S	S	S
Gln-44 → Tyr	47-51	S	S	S
Ser-45 → Leu	47-51	S	S	S
Gly-48 → Asp	n.d.	S	S	S
Gly-48 → Asn	n.d.	R	R	R
Ala-49 → Val	37-42	R	S	S
Ala-49 → Asp	47-51	S	S	S
Asn-52 → Asp	47-51	S	S	S
Asn-55 → Lys	47-51	S	S	S
Buried substitutions				
Ala-15 → Glu	n.d.	R	S	S
Leu-18 → Phe	n.d.	R	S	S
Tyr-22 → Cys	n.d.	R	S	S
Tyr-22 → His	30-34	R	R	S
Leu-31 → Ser	n.d.	R	R	S
Leu-50 → Cys	n.d.	R	S	S[a]
Gly-53 → Cys	n.d.	R	R	S
Ala-66 → Thr	26-30	R	R	R[b]
Ser-77 → Asn	n.d.	R	S	S
Ile-84 → Ser	37-42	S	S	S

Data are taken from Hecht *et al.*, 1984; Hecht and Sauer, 1985. Proteolysis temeprature refers to the temperature at which the N-terminal domain becomes sensitive to proteolysis by thermolysin. Activities *in vivo* refer to resistance (R) or sensitivity (S) to infection by phage λ. Repressor proteins were synthesized in the cell from multiple copy plasmids, and resistance or sensitivity to superinfection was determined by the absence or presence of a plaque, respectively, following addition of wild-type λ. Residues of the DNA-binding helix-turn-helix are 44-52. The corresponding positions in λ Cro, with respect to their three-dimensional location in the protein, are indicated in Figure 16.

[a] This mutant is resistant to superinfection of wild-type in this assay, but displays this temperature-sensitive profile for immunity to a more viulent form of λ.

[b] This mutant is temperature-sensitive for a more virulent phage λ (Hecht *et al*, 1983).

n.d. Not determined.

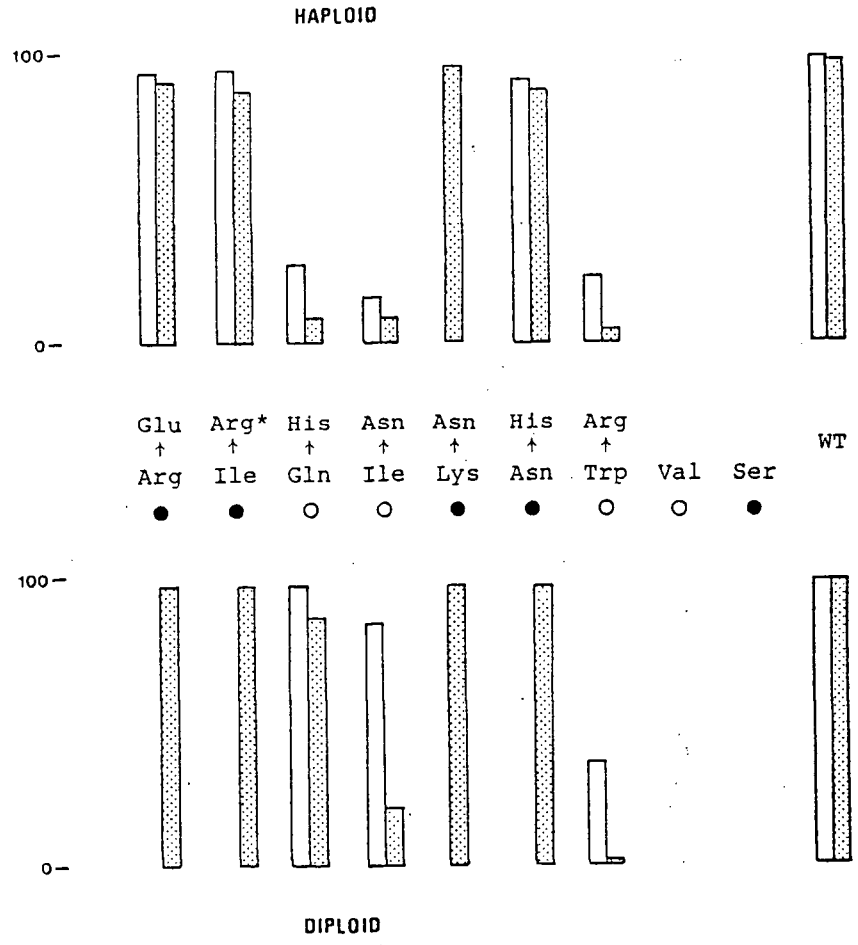


Figure 15. Temperature Dependence of Repressor Activity of Selected $\alpha 2$ Mutants.

Bars represent percentage repression of either *ste6::lacZ* (strain SP2, indicated by HAPLOID) or *ho::lacZ* (strain K1107, indicated by DIPLOID) as determined by assay of β-galactosidase in a strain transformed by the *mat*α-containing plasmid (YCp50-derived) relative to β-galactosidase in the strain transformed by YCp50. Shaded bars refer to activities obtained with a cell growth temperature of 30^0C, and open bars, to activities with cell growth at 25^0C. * at position 2 indicates a double mutant, that is Arg 1-Ile 2 → Glu 1-Arg 2 (mutant 11). Closed circles refer to those positions in the corresponding helix-turn-helix structure which are solvent-exposed, open circles refer to internal residues (Wharton and Ptashne, 1985; see Figure 16). WT indicates wild-type.

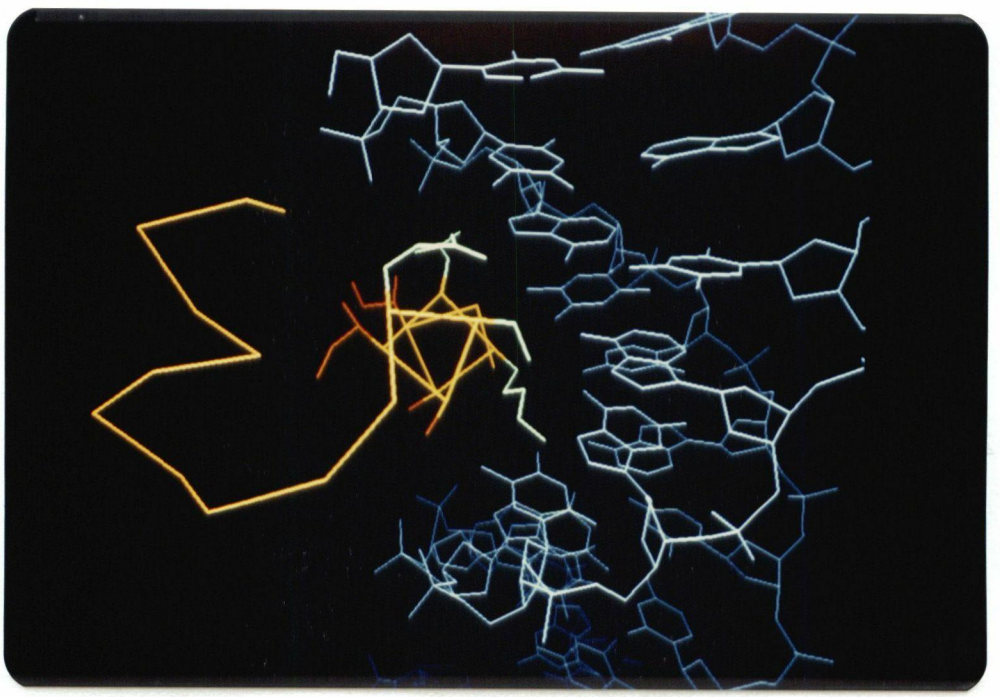


Figure 16. Model for the Interaction of λ Cro Protein with Operator DNA: Solvent-exposed and Buried Side Chains.

The carbon backbone of residues 15-38 of a λ Cro monomer, comprising the helix-turn-helix (yellow), in a model with its operator DNA (blue) (Anderson *et al.*, 1981). Red side chains, from residues 3, 4, 7, and 8 of the C-terminal helix are buried in the protein, whereas the green side chains, from residues 1, 2, 5, and 6 are solvent-exposed. The view is down the axis of the C-terminal helix from the N-terminal end.

Mutations in the region corresponding to the proposed N-terminal DNA-binding helix could be analyzed in a similar fashion. It is interesting to note that a mutant of the homeo domain-containing gene *fushi tarazu*, with an alteration at a position corresponding to the fifth residue of the N-terminal helix has a temperature-sensitive phenotype (Laughon and Scott, 1984). This residue of the helix-turn-helix structure is predicted to contact Ile/Val 4 of the C-terminal helix. The Ile 4 - Asn mutation of α2, as mentioned, displays a pronounced temperature-dependent diploid phenotype.

ANALYSIS OF THE BINDING OF α2 MUTANTS TO DNA

The concept of a functional homology of the portion of α2 under study to the prokaryotic DNA-binding helix-turn-helix was investigated, finally, by the *in vitro* analysis of the interaction of α2 with DNA. This served also to eliminate the possibility that reduction in α2 repressor activity by the mutations *in vivo* was due solely to altered interactions with other proteins in the transcription machinery.

The method used for the determination of DNA-binding activity is based on the procedure used by Hope and Struhl (1985) in their analysis of the interaction of yeast GCN4 protein with DNA. Fried and Crothers (1981) and Garner and Revzin (1981) first described the use of polyacrylamide gel electrophoresis for the separation of protein:DNA complexes from free DNA. The migration of DNA complexed with protein is considerably retarded with respect to free DNA, and this difference may be detected either by ethidium bromide staining or radioactive labelling of the DNA. Hope and Struhl (1985) modified this approach, however, by specifically labelling the protein being analyzed and by following its altered mobility upon complex formation with DNA. In this case, altered mobility is based on a charge difference between native protein (neutral at the pH used) and complexed protein (negatively charged). Also, because the mobility of the protein is monitored, DNA, rather than protein, is in excess.

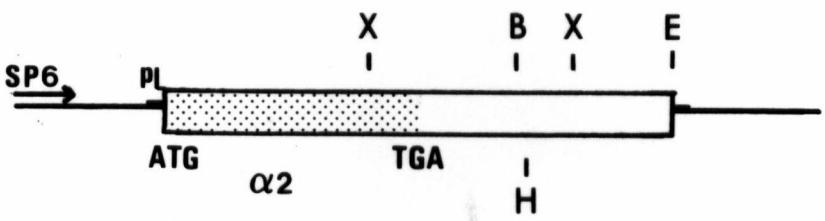
A simple method for the specific *in vitro* radioactive labelling of unpurified protein involves the use of vectors that are capable of directing the synthesis of pure, specific, biologically active transcripts.

These were first introduced by Melton *et al.* (1984) and exploit the fact that RNA polymerase from the bacteriophage SP6, under simple *in vitro* conditions, efficiently initiates transcription at a specific promoter and completes full-length transcripts of any composition. Transcription in the presence of the 5' terminal cap structure, m^7GpppN (or GpppG), yields RNA that is active in protein synthesis in a eukaryotic system *in vivo* or *in vitro*.

For the specific labelling of α2 protein, vector pSP64-α2 (see Figure 17) was used to direct the transcription of *MATα2*, in the presence of SP6 RNA polymerase. The transcripts were translated in the presence of ^{35}S-methionine in a wheat germ *in vitro* translation system, as described in the Materials and Methods section. Plasmids containing mutant *MATα2* genes were constructed by the replacement of the wild-type 0.5kb XbaI fragment with that from the mutant gene. To verify the structure of the construct and to determine whether specific, full length transcripts were synthesized (and were not terminated prematurely due to a fortuitous internal terminator sequence), pSP64-α2 was cleaved, separately, with different restriction enzymes, and the resultant linear molecules were used as transcription templates in the presence of α-^{32}P UTP. The results of this analysis are shown in Figure 17. Single transcripts are observed for each transcription reaction and correspond to the predicted length.

SP6-generated, unlabelled, 5'-capped transcripts were added to a commercial wheat germ translation system in the presence of ^{35}S-methionine. Figure 18 displays the translational products from wild-type and mutant *in vitro*-transcribed RNAs, as analyzed by SDS-polyacrylamide gel electrophoresis. All proteins correspond, approximately, in size to the expected molecular weight of α2 (27,640 d) and are produced in equivalent amounts.

The DNA used to assess DNA-binding ability of the α2 proteins (plasmid pZV9) is indicated in Figure 19. This plasmid contains the entire 2.7 kb *a*-specific *BAR1* gene in pUC13. *BAR1* expression is repressed *in vivo* by α2 (Kronstad *et al.*, 1987), and an α2-β-galactosidase fusion binds to the DNA *in vitro* (Johnson and Herskowitz, 1985). A 32bp sequence located 250bp upstream of the coding



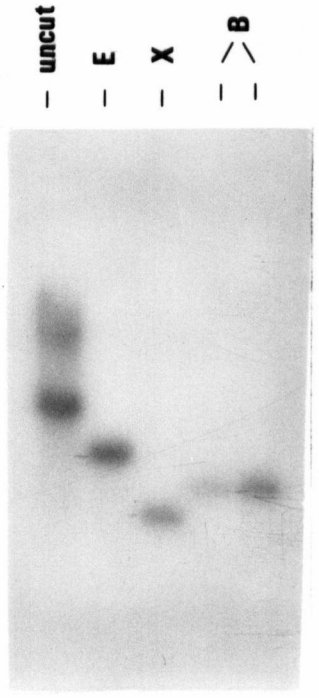


Figure 17. SP6 Transcription Analysis of Plasmid pSP64-α2.

Plasmid pSP64-α2 was cleaved with either EcoRI (E), XbaI (X), BstNI (B) or left uncut and subjected to transcription by SP6 RNA polymerase, in the presence of a small amount of α-^{32}P-UTP, as described in Materials and Methods. Transcripts were separated by formaldehyde agarose gel electrophoresis as described and exposed for autoradiography. Other abbreviations: PL, poly-linker cloning site; SP6, SP6 promoter; H, HpaI. The shaded region corresponds to the coding region of *MAT*α*2*.

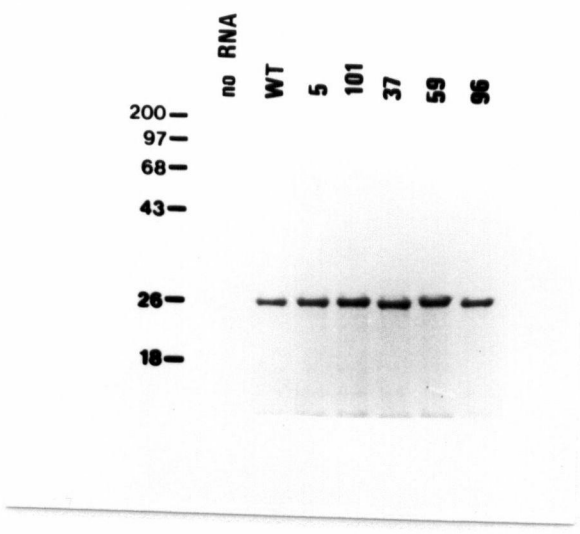


Figure 18. SDS-PAGE of *In Vitro*-Translated Wild-Type and Mutant *MATα2* Gene Products.

2 µl of translation mixes containing either wild-type or mutant *MATα2* genes were loaded and subjected to SDS-PAGE (12.5%) as described (Laemmli, 1970), and fluorographed for several hours. Mutants refer to those described in Figure 13, standard molecular weight markers are displayed at the left.

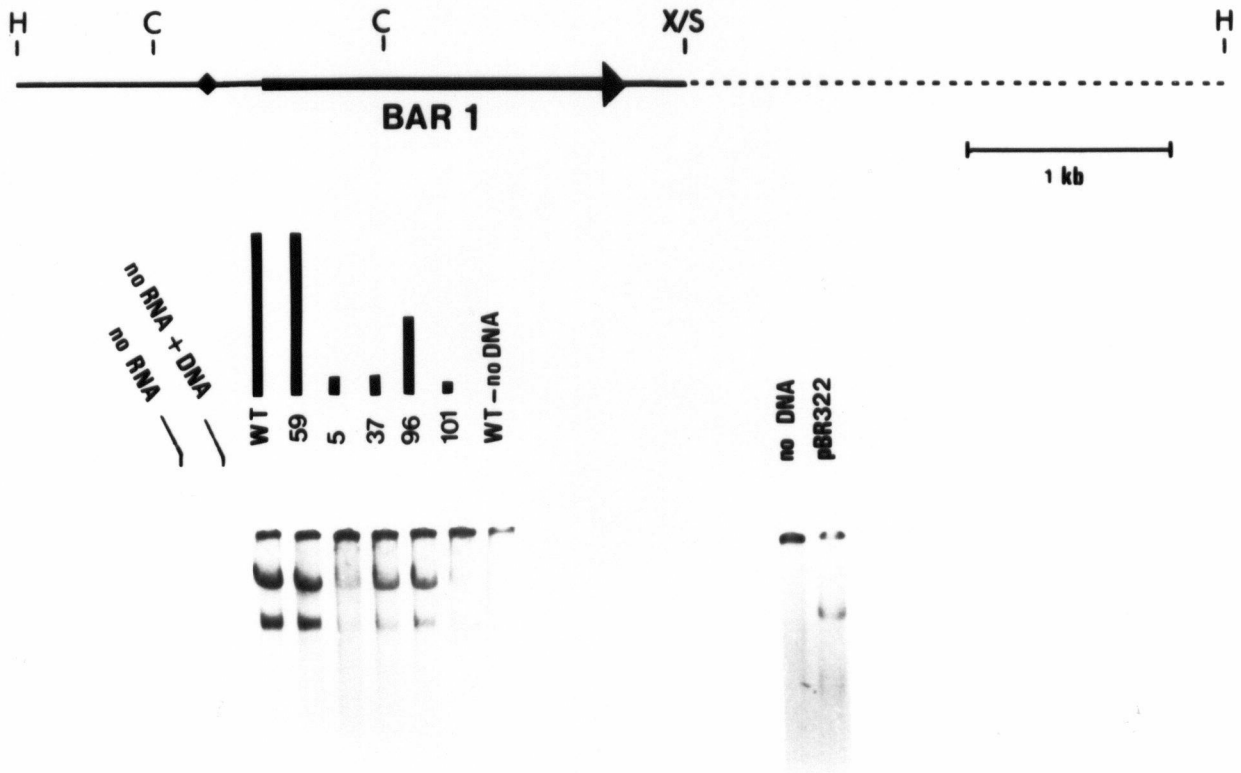


Figure 19. DNA-Binding Analysis of Wild Type and Mutant *MATα2* Gene Products.

Plasmid pZV9 (*BAR1* in pUC13) is indicated schematically at the top. The gel electrophoresis pattern of wild-type (WT) and mutant α2 (indicated by the mutant number) products bound to ClaI-cleaved pZV9, as described in Materials and Methods, is indicated at the left. The bars above each lane represent *in vivo* activity in repressing *ste6::lacZ* (see Figure 13). At the right is displayed a similar gel electrophoresis pattern of the wild type α2 product bound in this case to HinfI-cleaved pBR322 DNA.

sequence is highly homologous to the α2 operator and is contained within the fragment bound by the α2 fusion.

The conditions used in the DNA-binding analysis vary slightly from those described by Hope and Struhl (1985), primarily in the length of time, and temperature, of incubation of the protein with DNA. α2 is thought to bind DNA as a dimer. Since the protein is present as a dilute solution (and hence likely to a large extent as the monomer species), a longer incubation time of the protein with the DNA should increase the proportion of α2 bound to the DNA as a dimer. An increased amount of binding was in fact observed upon longer incubation of the α2/DNA mixture (not shown). The electrophoresis conditions were also modified in that the procedure was carried out at 4^0C rather than at room temperature. The rationale for this alteration was the liklihood of a greater stability of the complex at the lower temperature. Finally, as the presence of carrier DNA in the quantities used by Hope and Struhl had no effect on the pattern of DNA-binding, this component was omitted from the mixture.

Figure 19 (left) displays the gel electrophoresis pattern of wild-type and mutant α2 products bound to ClaI-cleaved pZV9. (ClaI digestion produces two DNA fragments from the 5.8 kb plasmid. The smaller, 1.1 kb fragment contains the α2 operator.) The bars above each lane represent the *in vivo* repressor activity of each mutant as determined previously by the analysis of *ste6::lacZ*-derived β-galactosidase activity of the yeast transformants. As the pH of the gel and running buffer (pH 8.3) is close to the estimated isoelectric point of α2 (8.2), labelled protein in the mixture containing no added DNA does not enter the gel. When a ClaI digest of pZV9 is included, however, two radioactive bands migrate into the gel. This indicates binding of α2 to both fragments of pZV9. The binding of α2 to the non-operator DNA is apparently sequence-specific, as α2 binds primarily to only one fragment of HinfI-cleaved pBR322 DNA (Figure 19, right). The largest HinfI fragment of pBR322 contains a sequence related to the consensus α2 operator inverted repeat structure (it is identical in 14 of 18 positions and displays complete identity in one-half of the inverted repeat), and could be acting as a "pseudo-operator". The ability of α2 to recognize more than one sequence in this assay can be contrasted with

the behaviour of the α2-β-galactosidase fusion, as assayed by Johnson and Herskowitz, which bound to only one fragment of a *BAR1* plasmid. Whether this difference is due to the structure of the fusion peptide or to the conditions of synthesis and/or DNA-binding is not known.

A comparison of the individual gel electrophoresis patterns (Figure 19) reveals a marked difference in the intensity of the migrating bands, that is, the affinity of the proteins for DNA, which corresponds roughly to the *in vivo* activity of the proteins. Thus, for all the mutants analyzed, the defect in *in vivo* repressor activity occurs at the DNA-binding level, and defects in auxiliary protein interaction, at least as the sole cause for inactivation, may be ruled out.

Binding activity of the mutants in the context of diploid function has not been analyzed. If α2 binds DNA in its diploid function as well as in its haploid function, it likely does so via a similar structure. Thus, one would predict that the disruption of this structure in α2 would result in proportionate decreases in binding activity for the two functions. If the *in vivo* pattern of activity for some of the mutants of this work (a more pronounced effect on α2 activity in its haploid function than in its diploid function), is due to the greater DNA-binding capacity of a mutant α2:wild-type a1 heterodimer than a mutant α2 homodimer, then the *in vitro* binding of α2 (in the presence of a1) to haploid-specific genes should correlate to its *in vivo* activity in the diploid. If, however, the difference in activity of the α2 mutants is due to interactions at the transcriptional (rather than DNA-binding) level, or to factors other than a1 that increase the binding potential of α2 (and which would not be included in the *in vitro* system), then the α2 mutants should bind *a*-specific and haploid-specific genes in a quantitatively similar fashion.

CONCLUSIONS AND FUTURE PROSPECTS

This work represents an initial attempt to localize and characterize the functional regions of the eukaryotic repressor MATα2. This was done in three consecutive steps. The first was the construction of mutations in *MATα2* that produce truncated gene products of various lengths. The purpose of this set of experiments was to test the hypothesis (Tatchell *et al.*, 1981) that the repressor

was divisible into domains of haploid and diploid activity. Although the results suggest that some diploid function of α2 is retained when the carboxy-terminal half of the gene product is missing, further studies would be required to verify this. The truncated gene products were defective in the haploid function of α2.

As a second step in the determination of functional domains of activity, based on the knowledge that 1) α2 is a DNA-binding protein (Johnson & Herksowitz, 1985), 2) the carboxy-terminal one-third of the protein is essential for full activity, and 3) this region contains a sequence homologous to a prokaryotic DNA-binding structure (Laughon and Scott, 1984), attention was focussed on this specific region of the protein. A procedure making use of synthetic oligonucleotides containing a low level of mutant bases was employed to saturate a specific sequence in the gene with missense mutations. This, coupled with the use of *E. coli lacZ* fusion genes expressed from promoters under the control of α2, allowed a verification of the importance of this region for repressor activity.

Single or double point mutations were then constructed at positions deduced from structural analyses, model-building, biochemical, and genetic studies of prokaryotic DNA-binding proteins, to be important in the specificity of the DNA:protein interactions. The phenotypes of the α2 mutants, a slight decline in *in vivo* activity, did not easily permit a determination of their DNA-binding specificity. The minor reduction in activity was thought to be inconsistent with a major role of these residues in specifically contacting DNA. It is not known whether this is because the structure of this region is not homologous to the helix-turn-helix, or because additional, specific contacts with DNA are made elsewhere in the protein.

In vitro DNA-binding analyses confirmed that the α2 missense mutants were deficient in binding DNA, and therefore that mutations in this region affected, either directly or indirectly, the DNA-binding structure of the protein. The results of studies *in vivo* to determine the temperature dependence of activity were consistent with the existence of a structure in this region of α2 homologous to that involved in the prokaryotic repressor function of binding DNA.

Several alternatives to the postulated structural/functional homology of the region to the helix-turn-helix have yet to be eliminated. Specifically, it could be important for dimerization, or for maintaining the structural integrity of the DNA-binding domain (but not involved in contacting DNA directly). Mutations affecting DNA-binding have been isolated throughout the 96 amino acid amino-terminal DNA-binding domain in λ *cI* (Nelson *et al.*, 1982); the α2 mutations constructed in this work could affect DNA-binding in a similar indirect fashion. The possibility that these mutations affect self-dimerization could be tested by transforming a *MATΔste6::lacZ* strain with the mutant genes on a high copy number plasmid. If the mutants are capable of dimerization, they should form heterodimers with the resident wild-type α2, thereby diminishing the α2 repressor activity in the strain. Mutations in Trp repressor (Kelley and Yanofsky, 1985), Lac repressor (Bennet and Yanofsky, 1978; Adler *et al.*, 1972) λ repressor (Nelson *et al.*, 1982) and Tn10 Tet repressor (Isackson and Bertrand, 1985) that confer such a dominant (or negatively complementing) phenotype have been described. The majority of these mutations lie within the helix-turn-helix structure.

One experiment that would help resolve the question of homology of this region of α2 with the helix-turn-helix is the selection of mutants with altered specificities for DNA. That is, one could insert an *in vitro*-constructed mutant α2 operator within a promoter of a gene that can be selected against (e.g. *CAN1* gene in the presence of canavanine, or the *URA3* gene in the presence of "suicide" substrates), and integrate it into a yeast genome. One could then transform the constructed yeast strain with a randomly mutated *MATα2* gene, and select those transformants that do not express the marker gene. To eliminate the possibility of selecting for α2 mutants with increased binding affinity for both mutant and wild-type operators, one should ensure that the mutants have lost the ability to recognize the wild-type operator. This could be accomplished by examining either the expression of wild-type genes normally under the control of α2 (e.g., screening for sterility) or the expression of a specific construct that places a selectable marker under the control of an α2-repressed promoter. Thus, one would select mutants that are either sterile (in the first case) or prototrophic for a particular nutrient (in the second), in addition to exhibiting resistance to the chosen substrate (canavanine or other). This

approach has been used successfully for the isolation of altered specificity mutants in the Mnt repressor of bacteriophage P22 (Youderian *et al.*, 1983) and in CAP (Ebright *et al.*, 1984).

If the mutations isolated by this method occurred in the region homologous to the helix-turn-helix, then its significance in specific DNA-binding would be independently verified, and further analysis of the region would be justified. If the mutations occurred elsewhere, particularly in a cluster, this would indicate a substantial deviation in structure and/or function of this region to that proposed, and would identify a location for the focus of further studies.

This study has concentrated on a region proposed, by sequence homology, to have a function in binding DNA. There are several other functions of the protein, for example, dimerization or gene repression (which may or may not be separable from its DNA-binding function), that could be studied in an analogous fashion. Because there is as yet no information to indicate the nature or location of these functions, the isolation of mutants through a selection scheme, rather than by directed *in vitro* manipulation, would most usefully address these problems. It would be particularly interesting and instructive to search for (and isolate, if they do exist) mutants that are able to bind DNA normally, but that are unable to repress gene transcription.

REFERENCES

Ackers, G.K., Shea, M.A. and Johnson, A.D. 1982. Proc. Natl. Acad. Sci. USA *79*:1129-1133.

Adler, K., Beyreuther, K., Fanning, E., Geisler, N., Gronenborn, B., Klemm, A., Muller-Hill, B., Pfah, M. and Schmitz, A. 1972. Nature *237*:322-327.

Ammerer, G., Sprague, G.F., Jr. and Bender, A. 1985. Proc. Natl. Acad. Sci. USA *82*:5855-5859.

Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. 1981. Nature *290*:754-758.

Anderson, J.E., Ptashne, M. and Harrison, S.C. 1987. Nature *326*:846-852.

Astell, C.R.A., Ahlstrom-Johnasson, L., Smith, M., Tatchell, K., Nasmyth, K.A. and Hall, B.D. 1981. Cell *27*:15-23.

Atkinson, T. and Smith, M. 1984. pp.35-82 in M. J. Gait, ed. Oligonucleotide Synthesis: a practical approach. IRL Press Ltd., Oxford.

Beachy, P.A., Helfand, S.L. and Hogness, D.S. 1985. Nature *313*:545-551.

Beggs, J. 1978. Nature *275*:104-108.

Beier, D.R. and Young, E.T. 1982. Nature *300*:724-728.

Bender, W., Akam, M., Karch, F., Beachy, P.A., Peifer, M., Spierer, P., Lewis, E.B. and Hogness, D.S. 1983. Science *221*:23-29.

Bennet, G.N. and Yanofsky, C. 1978. J. Mol. Biol. *121*:179-192.

Benoist, C. and Chambon, P. 1981. Nature *290*:304-310.

Berg, J.M. 1986. Science *232*:485-487.

Birnboim, N.C. and Doly, J. 1979. Nucl. Acids Res. *7*:1513-1523.

Bolivar, F. *et al.* 1977. Gene *2*:95-113.

Borrelli, E., Hen, R. and Chambon, P. 1984. Nature *312*:608-612.

Borrias, W.E., Wilschut, I.J.C., Vereijken, J.M., Weisbeek, P.J. and van Arkel, G.A. 1976. Virology *70*:195-197.

Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R. and Nasmyth, K. 1985. Cell *41*:41-48.

Breathnach, J. and Chambon, P.A. 1981. Ann. Rev. Biochem. *50*:349-383.

Breeden, L. and Nasmyth, K. 1985. Cold Spring Harbor Symp. Quant. Biol. *50*:643-650.

Brent, R. 1985. Cell *42*:3-4.

Brent, R. and Ptashne, M. 1984. Proc. Natl. Acad. Sci. USA *78*:4204-4208.

Brent, R. and Ptashne, M. 1985. Cell *43*:729-736.

Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tjian, R. 1986. Science *234*:47-52.

Brown, P.A. and Szostak, J.W. 1983. Meth. Enzymol. *101*:278-290.

Busby, S., Irani, M. and de Crombergghe, B. 1982. J. Mol. Biol. *154*:197-209.

Carroll, S.B., Laymon, R.A., McCutcheon, M.A., Riley, P.D. and Scott, M.P. 1986. Cell *47*:113-122.

Carroll, S.B. and Scott, M.P. 1985. Cell *43*:47-57.

Chan, R. and Otte, G. 1982. Mol. Cell. Biol. *2*:11-20.

Chen, W. and Struhl, K. 1985. EMBO J. *4*:3273-3280.

Choe, J., Kolodrubetz, D. and Grunstein, M. 1982. Proc. Natl. Acad. Sci. USA *79*:1484-1487.

Cox, B.S. 1965. Heredity *20*:505-521.

Dandanell, G. and Herman, K. 1985. EMBO J. *4*:3333-3338.

Dandanell, G., Valentin-Hansen, P., Larsen, J.E.L. and Hammer, K. 1987. Nature *325*:823-826.

Davison, B.L., Egly, J.-M., Mulvihill, E.R. and Chambon, P. 1983. Nature *301*:680-686.

Dente, L., Cesareni, G. and Cortese, R. 1983. Nucl. Acids Res. *11*:1645-1655.

Desplan, C. Theis, J. and O'Farrell, P.H. 1985. Nature *318*:630-635.

de Villiers, J. Olson, L., Tyndall, C., and Schaffner, W. 1982. Nucl. Acids Res. *10*:7965-7976.

DiNardo, S., Kuner, J.M., Theis, J. and O'Farrell, P.H. 1985. Cell *43*:59-69.

Douglas, H.C. and Hawthorne, D. 1966. Genetics *54*:911-916.

Dunn, T.M., Hahn, S., Ogden, S. and Schleif, R. 1984. Proc. Natl. Acad. Sci. USA *81*:5017-5020.

Dynan, W.S. and Tjian, R. 1983. Cell *35*:79-87.

Dynan, W.S. and Tjian, R. 1985. Nature *316*:774-778.

Ebright, R.H. 1986a. in Protein Structure, Function, and Design (ed. Oxender, D.) Liss, New York.

Ebright, R.H. 1986b. Proc. Natl. Acad. Sci. USA *83*:303-307.

Ebright, R.H., Cossart, P., Gicquel-Sanzey, B. and Beckwith, J. 1984. Nature *311*:232-235.

Efstradiatis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spiritz, R.A., DeRiel, J.K., Forget, B.G., Slighton, L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. 1980 Cell *21*:653-668.

Eisenbeis, S.J. *et al.* 1985. Proc. Natl. Acad. Sci. USA *82*: 1084-1088.

Eliason, J.L., Weiss, M.A. and Ptashne, M. 1985. Proc. Natl. Acad. Sci. USA *82*:2339-2343.

Feldman, J.B., Hicks, J.B., and Broach, J.R. 1984. J. Mol. Biol. *178*:815-834.

Fitzgerald-Hayes, M., Clarke, L. and Carbon, J. 1982. Cell *29*:235-244.

Flavell, R.A., Sabo, D.L., Bandle, E.F. and Weissmann, C. 1974. J. Mol. Biol. *89*:255-272.

Fried, M. and Crothers, D. 1981. Nucl. Acid Res. *9*:6505-6525.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. and Noll, M. 1986. Cell *47*:735-746.

Frohnhofer, H.G. and Nusslein-Volhard, C. 1986. Nature *324*:120-125.

Garber, R.L., Kuroiwa, A., and Gehring, W.J. 1983. EMBO J. *2*:2027-2034.

Garcia-Bellido, A. 1977. Am Zoologist *17*:613-629.

Garner, M. and Revzin, A. 1981. Nucl. Acid Res. *9*:3047-3060.

Gaub, M.P., Dierick, A., Astinotti, D., LePennec, J.-P. and Chambon, P. 1985. in Eukaryotic Transcription: The Role of *cis*- and *trans*-acting Elements in Initiation. (Gluzman, Y. ed.) pp. 123-131, Cold Spring Harbor Laboratory, N.Y.

Gehring, W.J. 1987. Cell *236*:1245-1252.

Gehring, W.J. and Hiromi, Y. 1986. Ann. Rev. Genet. *20*:147-173.

Gent, M.E., Gartner, S., Gronenborn, A.M., Sandulache, R. and Clore, G.M. 1987. Prot. Eng. *1*:201-203.

Gerlach, W.L. 1974. Heredity *32*:241-249.

Giniger, E., Varnum, S.M. and Ptashne, M. 1985. Cell *40*:767-774.

Gluzman, Y. (ed.) 1985. Eukaryotic Transcription: The Role of *cis*- and *trans*-acting Elements in Initiation. Cold Spring Harbor Laboratory, N.Y.

Godowski, P.J., Rusconi, S., Miesfeld, R. and Yamamoto, K.R. 1987. Nature *325*:365-368.

Goodbourn, S., Burstein, H. and Maniatis, T. 1986. Cell *45*: 601-610.

Goodman, H.M., Olson, M.V. & Hall, B.D. 1977. Proc. Natl. Acad. Sci. USA *74*:5453-5457.

Gorman, C.M., Rigby, P.W.J. and Lane, D.P. 1985. Cell *42*:519-526.

Graves, B.J., Johnson, PF., and McKnight, S.L. 1986. Cell *44*:565-576.

Griffith, J., Hochshild, A. and Ptashne, M. 1986. Nature *322*:750-752.

Guarente, L. 1984. Cell *36*:799-800.

Guarente, L. and Hoar, E.T. 1984. Proc. Natl. Acad. Sci. USA *81*:7860-7864.

Guarente, L. and Mason, T. 1983. Cell *32*:1279-1286.

Guarente, L., Nye, J.S., Hochschild, A. and Ptashne, M. 1982. Proc. Natl. Acad. Sci. USA *79*:2236-2239.

Hafen, E., Levine, M. and Gehring, W.J. 1984. Nature *307*: 287-289.

Hagen, D.C., McCaffrey, G. and Sprague, G.F. 1986. Proc. Natl. Acad. Sci. USA *83*:1418-1422.

Hahn, S., Hoar, E.T. and Guarente, L. 1982. Proc. Natl. Acad. Sci. USA *82*:8562-8561.

Hall, M.N., Hereford, L. and Herskowitz, I. 1984. Cell *36*: 1057-1065.

Hawley, D.K. and McClure, W.R. 1983. Nucl. Acids Res. *11*:2237-2255.

Hecht, M.H., Nelson, H.C.M. and Sauer, R.T. 1983. Proc. Natl. Acad. Sci. USA *80*:2676-2680.

Hecht, M.H. and Sauer, R.T. 1985. J. Mol. Biol. *186*:53-63.

Hecht, M.H., Sturtevant, J.M. and Sauer, R.T. 1984. Proc. Natl. Acad. Sci. USA *81*:5685-5689.

Heffron, F., So, M. and McCarthy, B.J. 1978. Proc. Natl. Acad. Sci. USA *75*:6012-6016.

Hendrickson, W. and Schleif, R. 1985. Proc. Natl. Acad. Sci. USA *82*:3129-3133.

Hicks, J. and Herskowitz, I. 1976. Genetics *83*:245-258.

Hicks, J., Strathern, J.N. and Herskowitz, I. 1977. Genetics *85*:395-405.

Hicks, J., Strathern, J. and Klar, A. 1979. Nature *282*:478-483.

Hill, D.E., Hope, I.A., Macke, J.P. and Struhl, K. 1986. Science *234*:451-457.

Hirose, S., Takeuchi, K. and Suzuki, Y. 1982. Proc. Natl. Acad. Sci. USA *79*:7258-7262.

Hochschild, A., Douhan, J. and Ptashne, M. 1986. Cell *47*: 807-816.

Hochschild, A., Irwin, N. and Ptashne, M. 1983. Cell *32*: 319-325.

Hochschild, A.H. and Ptashne, M. 1986. Cell *44*:681-687.

Hope, I.A. and Struhl, K. 1985. Cell *43*:177-188.

Hopper, A.K. and Hall, B.D. 1975. Genetics *80*:41-59.

Hu, M.C.-T. and Davidson, N. 1987. Cell *48*:555-566.

Humayun, Z., Kleid, D. and Ptashne, M. 1977. Nucl. Acids. Res. *4*:1595-1607.

Hurstel, S., Granger-Schnarr, M., Daune, M. and Schnarr, M. 1986. EMBO J. *5*:793-798.

Imler, J.-L., Lemaire, C., Wasylyk, C. and Wasylyk, B. 1987. Mol. Cell. Biol. *7*:2558-2567.

Ingham, P.W., Howard, K.R. and Ish-Horowicz, D. 1985. Nature *318*:439-445.

Irani, M.H., Orosz, L. and Adhya, S. 1983. Gene *32*:783-788.

Isackson, P.J. and Bertrand, K.P. 1985. Proc. Natl. Acad. Sci. USA *82*:6226-6230.

Jensen, R., Sprague, G.F. and Herskowitz, I. 1983. Proc. Natl. Acad. Sci. USA *80*:3035-3039.

Johnson, M. and Davis, R.W. 1984. Mol Cell. Biol. *4*:1440-1448.

Johnson, A.D. and Herskowitz, I. 1985. Cell *42*:237-247.

Johnson, A., Meyer, B.J. and Ptashne, M. 1978. Proc. Natl. Acad. Sci. USA *75*:1783-1787.

Johnston, S.A., Salmeron, J.M. and Dincher, S.S. 1987. Cell *50*:143-146.

Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. and Tjian, R. 1987. Cell *48*:79-89.

Jones, K.A., Yamamoto, K.R. and Tjian, R. 1985. Cell *42*: 559-572.

Kadonaga, J.T., Jones, K.A. and Tjian, R. 1986. Trends Bio. Sci. *11*:20-23.

Kassir, Y. and Simchen, G. 1976. Genetics *82*:187-206.

Keegan, L., Gill, G. and Ptashne, M. 1986. Science *231*:699-704.

Kelley, R.L. and Yanofsky, C. 1985. Proc. Natl. Acad. Sci. USA *82*:483-487.

Klar, A. J. S., Strathern, J. N., Broach, J. R. and Hicks, J. B. 1981. Nature *289*:239-244.

Koudelka, G.B., Harrison, S.C. and Ptashne, M. 1987. Nature *326*:886-888.

Kronstad, J.W., Holly, J.A. and MacKay, V.L. 1987. Cell *50*:369-377.

Kunkel, T.A., Roberts, J.D. and Zakour, R.A. 1986. Meth. Enzymol., in press.

Kurjan, J. and Herskowitz, I. 1982. Cell *30*:933-943.

Laimins, L., Holmgren-Konig, M. and Khoury, G. 1986. Proc. Natl. Acad. Sci. USA *83*:3151-3155.

Laughon, A. and Gesteland, R.F. 1982. Proc. Natl. Acad. Sci. USA *79*:6827-6831.

Laughon, A. and Scott, M.P. 1984. Nature *310*:25-31.

Lee, W., Mitchell, P. and Tjian, R. 1987. Cell *49*:741-752.

Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. 1977. Biochemistry *16*:4743-4751.

Lewis, E.B. 1978. Nature *276*:565-570.

Little, J.W. and Hill, S.A. 1985. Proc. Natl. Acad. Sci. USA *82*:2301-2305.

Little, J.W. and Mount, D.W. 1982. Cell *29*:11-22.

Liu-Johnson, H.-N., Gartenberg, M.R. and Crothers, D.M. 1986. Cell *47*:995-1005.

Ma, J. and Ptashne, M. 1987a. Cell *48*:847-853.

Ma, J. and Ptashne, M. 1987b. Cell *50*:137-142.

Macdonald, M. and Struhl, G. 1986. Nature *324*:537-545.

MacKay, V. L. and Manney, T. R. 1974a. Genetics *76*:255-271.

MacKay, V. L. and Manney, T. R. 1974b. Genetics *76*:272-288.

Majors, J. 1975. Proc. Natl. Acad. Sci. USA *72*:4394-4398.

Majumdar, A. and Adhya, S. 1984. Proc. Natl. Acad. Sci. USA *81*:6100-6104.

Mandel, M. and Higa, A. 1970. J. Mol. Biol. *53*:159-162.

Maniatis, T., Fritsch, E.F., and Sambrook, J. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Martin, K., Huo, L. and Schleif, R.F. 1986. Proc. Natl. Acad. Sci. USA *83*:3654-3658.

Matthews, B.W., Ohlendorf, D.H., Anderson, W.F. and Takeda, Y. 1982. Proc. Natl. Acad. Sci. USA *79*:1428-1432.

McClarin, J.A., Frederick, C.A., Wang, B.-C., Greene, P., Boyer, H.W., Grable, J. and Rosenberg, J.M. 1986. Science *234*:1526-1541.

McGinnis, W., Harber, R.L., Wirz, J., Kuroiwa, A. and Gehring, W.J. 1984. Cell *37*:403-408.

McKay, D.B. and Steitz, T.A. 1981. Nature *290*:744-749.

McKay, D.B., Weber, I.T. and Steitz, T.A. 1982. J. Biol. Chem. *257*:9518-9524.

McNeil, J.B. and Smith, M. 1985. Molec. Cell. Biol. *5*:3545-3551.

McNeil, J.B. and Smith, M. 1986. J. Mol. Biol. *187*:363-378.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. 1984. Nucl. Acid Res. *12*: 7035-7056.

Messing, J. 1983. Methods in Enzymology *101*:20-78.

Meyer, B.J. and Ptashne, M. 1980. J. Mol. Biol. *139*:195-205.

Miller, A.M. 1984. EMBO J. *3*:1061-1065.

Miller, A.M., MacKay,V.L. and Nasmyth, K.A. 1985a. Nature *314*:598-603.

Miller, J., McLachlan, A.D. and Klug, A. 1985b. EMBO J. *4*: 1609-1614.

Mlodzik, M., Fjose, A. and Gehring, W.J. 1985. EMBO J. *4*:2961-2969.

Mlodzik, M. and Gehring, W.J. 1987. Cell *48*:465-478.

Mortimer, R.K. and Hawthorne, D.C. 1969. pp. 385-460 in Rose, A.H. and Harrison, J.S., ed. The yeasts, Vol. 1. Academic Press, New York.

Myers, R.M., Lumelsky, N., Lerman, L.S. and Maniatis, T. 1985. Nature *313*:495-498.

Nagawa, F. and Fink, G.R. 1985. Proc. Natl. Acad. Sci. USA *82*:8557-8561.

Nasmyth, K. A. and Tatchell, K. 1980. Cell *19*:753-764.

Nasmyth, K.A., Tatchell, K., Hall, B.D., Astell, C. and Smith, M. 1981. Nature *289*:244-250.

Nelson, H.C.M., Hecht, M.H. and Sauer, R.T. 1983. Cold Spring Harbor Symp. Quant. Biol. *47*:441-449.

Nelson, H.C.M. and Sauer, R.T. 1985. Cell *42*:549-558.

Neuhaus, G., Neuhaus-Url, G., Gruss, P. and Schweiger, H.G. 1984. EMBO J. *3*:2169-2172.

Nir, U., Walker, M.D. and Rutter, W.J. 1986. Proc. Natl. Acad. Sci. USA *83*:3180-3184.

Ohlendorf, D.H., Anderson, W.F. and Matthews, B.W. 1983. J. Mol. Evol. (1983) 109-114.

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. and Steitz, T.A. 1985. Nature *313*:762-766.

Osborne, T.F., Goldstein, J.L. and Brown, M.S. 1985. Cell *42*:203-212.

Pabo, C.O., Krovatin, W., Jeffrey, A. and Sauer, R.T. 1982. Nature *298*:441-443.

Pabo, C.O. and Lewis, M. 1982. Nature *298*:443-447.

Pabo, C.O. and Sauer, R.T. 1984. Ann. Rev. Biochem. *53*:293-321.

Pabo, C.O., Sauer, R.T., Sturtevant, J.M. and Ptashne, M. 1979. Proc. Natl. Acad. Sci. USA *76*:1608-1612.

Payvar, F., Wrange, O., Carlstedt-Duke, J., Okret, S., Gustafsson, J.A. & Yamamoto, K.R. 1981. Proc. Natl. Acad. Sci. USA *78*:6628-6632.

P Pinkham, J.L., Olesen, J.T. and Guarente, L.P. 1987. Mol. Cell. Biol. *7*:578-585.

Poole, S.J., Kauvar, L.M., Drees, B. and Kornberg, T. 1985. Cell *40*:37-43.

Poteete, A.R. and Ptashne, M. 1982. J. Mol. Biol. *157*:21-48.

Pribnow, D. 1975. Proc. Natl. Acad. Sci. USA *72*:784-788.

Ptashne, M. 1986. Nature *322*:697-701.

Pure, G.A., Robinson, G.W., Naumovski, L. & Friedberg, E.C. 1985. J. Mol. Biol. *183*:31-42.

Rothstein, R.J., Esposito, R.E. and Esposito, M.S. 1977. Genetics *85*:35-54.

Ruby, S.W., Szostak, J.W. and Murray, A.W. 1983. Meth. Enzymol. *101*:253-269.

Russell, D.W., Jensen, R., Zoller, M.J., Burke, J., Errede, B., Smith, M. and Herskowitz, I. 1986. Mol. Cell. Biol. *6*:4281-4294.

Sauer, R.T., Pabo, C.O., Meyer, B.J., Ptashne, M. and Backman, K. 1979. Nature *279*:396-400.

Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M. and Pabo, C.O. 1982. Nature *298*:447-451.

Schaller, H., Gray, C. and Herrmann, K. 1975. Proc. Natl. Acad. Sci. USA *72*:737-741.

Schevitz, R., Otwinowski, Z., Joachimiak, A., Lawson, C.L. and Sigler, P.B. 1985. Nature *317*:782-786.

Schirm, S., Jiricny, J. and Schaffner, W. 1987. Genes and Dev. *1*:65-74.

Scott, M.P. and Weiner, A.J. 1984. Proc. Natl. Acad. Sci. USA *81*:4115-4119.

Scott, M.P., Weiner, A.J., Polisky, B.A., Hazelrigg, T.I., Pirrotta, V., Scalenghe, F., and Kaufman, T.C. 1983. Cell *35*:763-776.

Serfling, E., Jasin, M. and Schaffner, W. 1985. Trends Genet. *1*:224-230.

Shepherd, J.C.W., McGinnis, W., Carrasco, A.E., De Robertis, E.M. and Gehring, W.J. 1984. Nature *310*:70-71.

Sherman, F., Fink, G.R., and Hicks, J.B. 1981. Methods in Yeast Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Shortle, D. and Botstein, D. 1983. Meth. Enzymol. *100*:457-468.

Shuey, D.J. and Parker, C.S. 1986a. Nature *323*:459-461.

Shuey, D.J. and Parker, C.S. 1986b. J. Biol. Chem. *261*: 7934-7940.

Siliciano, P.G. and Tatchell, K. 1984. Cell *37*:969-978.

Siliciano, P.G. and Tatchell, K. 1986. Proc. Natl. Acad. Sci. USA *83*:2320-2324.

Simpson, R.B. 1980. Nucl. Acids Res. *8*:759-766.

Singh, A., Helms, C., Sherman, F. 1979. Proc. Natl. Acad. Sci. USA *76* 1952-1956.

Smith, M. 1985. Ann. Rev. Gen. *19*:423-462.

Smith, M.M. and Andresson, O.S. 1983. J. Mol. Biol. *169*:663-690.

Sprague, G.F., Blair, L.C. and Thorner, J. 1983a. Ann. Rev. Microb. *37*:623-660.

Sprague, G.F., Jensen, R. and Herskowitz, I. 1983b. Cell *32*: 409-415.

Strathern, J., Hicks, J. and Herskowitz, I. 1981. J. Mol. Biol. *147*:357-372.

Struhl, K. 1984. Proc. Natl. Acad. Sci. USA *82*:8419-8423.

Struhl, K. 1985. Nature *317*:822-824.

Struhl, G. and White, R.H. 1985. Cell *43*:507-519.

Sumrada, R.A. and Cooper, T.G. 1987. Proc. Natl. Acad. Sci. USA *84*:3997-4001.

Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, M. and Chambon, P. 1986. Nature *319*:121-126.

Tatchell, K., Nasmyth, K.A., Hall, B.D., Astell, C. and Smith, M. 1981. Cell *27*:25-35.

Tjian, R. 1978. Cell *13*:165-179.

Tjian, R. 1981. Cell *26*:1-2.

Topol, J., Ruden, D.M. and Parker, C.S. 1985. Cell *42*:527-537.

Wallis, J.W., Hereford, L. and Grunstein, M. 1980. Cell *22*:799-805.

Weber, K., Platt, T., Ganem, D. and Miller, J.H. Proc. Natl. Acad. Sci. USA *69*:3624-3628.

Weir, M.P. and Kornberg, T. 1985. Nature *318*:433-439.

Wells, J.A., Vasser, M. and Powers, D.B. 1985. Gene *34*:315-323.

Wharton, R.P. and Ptashne, M. 1985. Nature *316*:601-605.

Wharton, R.P. and Ptashne, M. 1987. Nature *326*:888-891.

White, R.A.H. and Wilcox, M. 1984. Cell *39*:163-172.

Wiederrecht, G., Shuey, D.J., Kibbe W.A. and Parker, C.S. 1987. Cell *48*: 507-515.

Wilson, K.L. and Herskowitz, I. 1984. Mol. Cell. Biol. *4*: 2420-2427.

Wirz, J., Fessler, L. and Gehring, W.J. 1986. EMBO J. *5*:3327-3334.

Wu, C. 1985. Nature *317*:84-87.

Yamashita, I., Takano, Y. and Fukui, S. 1985. J. Bact. *164*: 769-773.

Youderian, P., Vershon, A., Bouvier, S., Sauer, R.T. and Susskind, M.M. 1983. Cell *35*:777-783.

Zakour, R.A., James, E.A. and Loeb, L.A. 1984. Nucl. Acids Res. *12*:6615-6628.

Zakour, R.A. and Loeb, L.A. 1982. Nature *295*:708-710.

Zinn, K., DiMaio, D. and Maniatis, T. 1983. Cell *34*:865-879.

Zoller, M.J. and Smith, M. 1983. Meth. Enzymol. *100*:468-500.

Zoller, M.J. and Smith, M. 1984. DNA *3*:479-488.