INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9218244

Sequence analysis and identification of the DNA-binding domain of the *Saccharomyces* MAL-activator encoded by *MAL63*

Kim, Jeong Hwan, Ph.D.

City University of New York, 1992

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

SEQUENCE ANALYSIS AND IDENTIFICATION OF THE DNA-BINDING DOMAIN OF THE *SACCHAROMYCES* MAL-ACTIVATOR ENCODED BY *MAL63*

by

Jeong H. Kim

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirementS for the degree of Doctor of Philosophy, The City University of New York.

1992

This manuscript has been read and accepted for the Graduate Faculty in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

November 14, 1991
date

Chairman of Examining Committee
Dr. Corinne Michels, Queens College

10 Jan 1992
date

Executive Officer
Dr. Peter C. Chabora

Dr. David Calhoun, City College

Dr. Wilma Saffran, Queens College

Dr. Susan Rotenberg, Queens College

Dr. Diana Bartelt, St. John's University

Supervisory Committee

The City University of New York

Abstract

SEQUENCE ANALYSIS AND IDENTIFICATION OF THE DNA-BINDING DOMAIN OF THE *SACCHAROMYCES* MAL-ACTIVATOR ENCODED BY *MAL63*

by

Jeong H. Kim

Advisor: Professor Corinne A. Michels

Inducible maltose fermentation by *Saccharomyces* species requires three products of any one of the five identified *MAL* loci. Two of these gene products are maltose permease (GENE 1) and maltase (GENE 2), the proteins needed for transport of the sugar into the cell and its cleavage to glucose. The third gene product GENE 3 has been suggested to be an activator protein controlling the expression of the structural genes encoding the maltose fermentative enzymes perhaps by binding to DNA sequences upstream of these genes. We report the sequence of GENE 3 of the *MAL6* locus, *MAL63*. A single open reading frame is seen capable of encoding a protein of 470 amino acid residues. A cysteine-basic amino acid-rich sequence at the N-terminal end of the protein is characteristic of one class of DNA-binding proteins supporting the hypothesis that the *MAL63* gene product is a DNA-binding transcriptional activator.

To test if the cysteine-basic amino acid-rich region is a part

of DNA-binding domain, we mutagenized cysteine residues 18, 27 and 34 to leucine, serine and glycine, respectively, using site-directed mutagenesis. The resulting *mal63* mutant alleles were incapable of activating maltose fermentation in yeast. Also, using a gel mobility shift assay, protein extracts from *E. coli* strains expressing the mal63 mutant proteins all failed to bind to an oligonucleotide from the UAS_{MAL} under conditions where the wild type MAL63 protein was able to bind. In addition, we altered the invariant proline found in this cysteine-rich region to leucine and found that the encoded product was partially functional *in vivo* but unable to bind DNA *in vitro* suggesting that the DNA binding domain was structurally altered. These results provide genetic evidence that the cysteine-basic amino acid-rich region is required for DNA-binding by the MAL63 protein and support the proposal that the DNA-binding domain by the MAL63 protein forms a zinc cluster similar to the structure proposed for the DNA-binding domain of the GAL4 protein (Pan and Coleman, 1990; Gardner et al., 1991).

Acknowledgements

I wish to thank all the people that I know of or already forgotten. I appreciate their tiny influences in my life.

I thank my mentor Dr. Corinne Michels for thoughtful guidance and patience. Also, I thank my second examination and dissertation defense committee members, Drs. Bartelt, Brandriss, Calhoun, Rotenberg, Saffran and Tropp, for helpful suggestions on my thesis project. I also thank Dr. Needleman for long cooperative works. Many people whom I have not yet met, helped me by providing the up-to-dated knowledges, techniques, plasmids and cells. These include Drs. Johnston, Langdon, Sikorski and Studier. I thank the City University faculty members: Drs. Szalay, Boylan, Alsop, Minutoli and Chabora are especially appreciated. The staff members like Victor, Bob, Terry and Eileen were so great to me. Esther was always helpful for my work and righteous life.

My seniors, Drs. Dubin and Charron, were guiding lights in pursuing my career goal. Both of you performed excellent Postdoctoral work and are still working hard as faculty members. I will remember Drs. Levine and Cotty as the most humorous men.

We had an unique population in our lab. Very well mixed races, characters and talents. My peculiar character would probably not be acceptable in any other lab. Thanks to all of you, Andy, Lori, Qi, Sheryl, Chantal, Evan and Caryl for open mindedness and forgivings.

My neighbors: Helen, Kashmira, Jane, Dido and Ron! You were the best friends when I felt like nobody.

I thank my late father Young Geun Kim for making it possible for me to begin study in the USA. You did more than you possible could to have your children educated better than yourself. You had no time to rest until your final day. Here's the first Ph.D degree from your children that you would have wanted to see. I thank my mother for taking great care of my children. When any grandchild was not able to sleep you also did not sleep. No one could put more heart, mother.

I shouldn't forget my brother-in-law Heung and sister Jung for financial as well as mental support for studying in the UAS.

I thank my wife for giving me three children. You put your Ph.D. program off for mine to support family. My son Dauhyun and daughters Suhyun and Gahyun, all were born at the busiest time in my life and did not get enough time with me. But all of you were always cheerful. I just thank you for growing up fine.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

To function as a regulatable transcription activator a protein must : 1) localize to the nucleus following translation; 2) bind to controlling elements on the DNA; 3) interact with the transcription machinery to activate transcription of genes located next to the controlling element; and 4) bind small metabolites or other protein factors so as to regulate the above functions. Studies of several yeast and other eukaryotic transcription activators indicate that, to a first approximation, distinct regions (or domains) of the activator proteins are separately responsible for these functions (Keegan and Ptashne, 1986). My thesis project is a study of a *Saccharomyces* transcription activator, the MAL-activator encoded by the *MAL63* gene. Previous studies from our laboratory, in collaboration with Richard Needleman and coworkers, had clearly demonstrated that the product of the *MAL63* gene (or its *MAL1* homologue, *MAL13*) is required for expression of the genes encoding the maltose fermentative enzymes (Charron et al., 1986; Chang et al., 1988). In Chapter 1 of this thesis, I describe the sequence analysis of *MAL63*. *MAL63* is shown to encode a protein containing a cysteine-basic amino acid-rich region at its N-terminal end exhibiting homology to one class of DNA-binding domain. In Chapter 2, I describe experiments demonstrating that this region of MAL63 functions as the DNA-binding domain and specifically binds to a sequence from the upstream activation site of the *MAL6* structural genes.

Eukaryotic DNA-binding domains.

The DNA-binding domains of different eukaryotic activator or repressor proteins have been identified and shown to fall into a few distinct classes referred to as the homeobox domains (or helix-turn-helix), the leucine zipper domain, and cysteine-rich domains.

Helix-turn-helix motif. The helix-turn-helix motif was first identified in DNA-binding proteins encoded by bacteriophage lambda (the CRO and CI proteins), and by *E. coli*, catabolite activator protein (CAP). All of these proteins bind DNA as dimers and their binding sites on DNA are known to have dyad symmetry. Upon resolution of their respective three-dimensional structure (Anderson et al., 1981; Mckay and Steitz, 1981; Pabo and Lewis, 1982), these proteins were found to have a distinctive arrangement of two alpha helical regions separated by a beta turn (Steitz et al., 1982). Thus this structure is termed the helix-turn-helix motif. In yeast, the DNA-binding domain of the proteins MATa1 and MATα2 are known to form this structure (Johnson and Herskowitz, 1985). Laughon and Scott (1984), and Shepherd et al. (1984) noted that the sequence similarities between yeast and bacterial helix-turn-helix proteins could be further extended to products of the *Drosophila* homeotic genes, *antennapedia*, *fushi tarazu*, and *ultrabithorax*. DNA sequence similarities predict that the amino acid sequence similarities in these three proteins exceed 90% over a 60 amino acid segment, which is termed the homeobox. A pattern of hydrophobic amino acids consistent with the helix-turn-helix motif occurs within a 30-amino acid segment disposed toward the carboxyl half of canonical homeobox sequence.

Leucine-zipper motif. The existence of this type of DNA-binding motif was first hinted at by a study that noted primary amino acid sequence similarities between the products of several oncogenes (FOS, MYC and JUN) and a yeast regulatory protein (GCN4) known to be capable of sequence-specific recognition of DNA (Vogt et al., 1987). This motif is characterized by, in the DNA-binding domain, a heptad repeat of leucine residues termed the leucine zipper and high density of oppositely charged amino acids (acidic and basic residues) juxtaposed in a manner suitable for intrahelical ion pairing. This high frequency of ion pairing predicted an unusually stable helical domain.

Zinc-finger motif. Miller et al. (1985) observed an unusual sequence in TFIIIA. There were 9 tandem units, each about 30 amino acid residues in length and having the following structure: Cys-X_{2-5}-Cys-X_{12}-His-X_{3-4}-His. TFIIIA binds to the Internal Control Region of the 5S rRNA genes. To provide an explanation of how RNA polymerase passes through this DNA-protein complex, Miller et al. (1985) hypothesized a finger-like structure for each of these 9 repeats with a zinc ion stabilizing the loop-like structure. Miller et al. (1985) also examined the requirement of zinc ions to stabilize the DNA-TFIIIA complex. Copper and cadmium could be substituted for zinc. However, when they examined purified DNA-TFIIIA particles zinc was the only significant ion found in each particle. Thus each seems to contain one zinc ion. The ADR1 receptor contains two C_2H_2 structures (Blumberg et al., 1987) and thus it is most likely to belong to this zinc-finger class of DNA-

binding proteins.

The zinc-twist motif. The glucocorticoid receptor shows a slight variation in structure from the zinc-finger described above. Instead of a C_2H_2 structure it shows C_2C_2 and contains two such structures. Since this protein chelates two zinc ions per molecule this was thought to form 2 zinc-fingers (Freedman et al., 1988). However, NMR studies show that this protein forms a structure distinct from the zinc-finger motif which is termed "zinc twist" (Vallee et al., 1991). This type of structure is found in the steroid hormone, thyroid hormone, retinoic acid and vitamin D_3 receptors (Vallee et al., 1991). The DNA-binding domain of the human glucocorticoid receptor (hGR) has been localized to amino acid residues 421 to 486 (Hollenberg and Evans, 1988). This region shows two adjacent cysteine rich-regions. Deletion of either the first (del 420-451) or the second region (del 450-487) leads to complete loss of DNA-binding activity *in vitro* and thus transcription *in vivo*. The first region (closer to N-terminal end) appears to be responsible for specific binding to the Glucocorticoid Responsive Element (GRE) (Freeman et al., 1988). Substitution (missense) mutations in the hGR gene were made introducing glycine residues into the first region (Hollenberg and Evans, 1988). Whenever a cysteine residue was altered, there was complete loss of DNA-binding activity *in vitro* and transcription activation *in vivo* (Hollenberg and Evans, 1988). In hGR, residues 416 to 487 containing both cysteine-rich regions were replaced with the first 74 amino acids of the yeast GAL4 protein (Hollenberg and

Evans, 1988). This chimeric protein activated the transcription of a reporter gene containing an insertion of the UAS_{GAL} 17mer sequence in a dexamethasone dependent manner. This sequence did not trans-activate a reporter gene which had the GRE but lacked the UAS_{GAL} 17mer. This indicates that, in the intact hGR, there is only one DNA-binding domain and it is contained within the 416-487 region of hGR.

The third motif. *GAL4* encodes a protein that activates transcription of the *GAL* structural genes (except for *GAL5* and *GAL4*) by binding to DNA sites located upstream of each gene. The GAL4 protein is 881 amino acids long (Laughon and Gesteland, 1984). The 74 N-terminal residues of the GAL4 protein are sufficient for nuclear localization (Silver et al., 1984). This 74 amino acid long peptide also contains the DNA-binding domain and has been shown by foot-printing analysis, to bind to the UAS_{GAL} (Keegan et al., 1986) but is unable to activate transcription, implying that the transcription activation domain is lacking (Keegan et al., 1986). The N-terminal 74 residues of the GAL4 protein also protect the 17 base pair consensus sequence of UAS_{GAL} from DNAse digestion (Keegan et al., 1986). When a peptide containing the N-terminal residues from 1 to 147 is used to foot-print the UAS_{GAL} there are 4 DNAse protected sites (Keegan et al., 1986). The binding affinity of the 4 UAS_{GAL} sites to the protein are not the same. Transcription activation is increased if more than one UAS_{GAL} is present although more than two sites do not further increase levels of gene expression significantly (Lorch and Kornberg, 1985).

Johnston and Dover (1987; 1988) randomly mutagenized yeast selecting for *GAL4* mutants using a method designed to allow only *gal4* noninducible mutants to survive. The resulting genes were sequenced and a rule emerged (Johnston and Dover, 1987; 1988). Firstly, most of the mutations (27/31) in the 5'-terminal portion of *GAL4* were missense mutations and all of these reduced the DNA-binding activity. The sequence changes caused by these mutations were located within a 48 residue region containing six cysteine residues and having the sequence $C\text{-}X_2\text{-}C\text{-}X_6\text{-}C\text{-}X_6\text{-}C\text{-}X_2\text{-}C\text{-}X_6\text{-}C$ (Johnston and Dover, 1987). Secondly, mutations in the remainder of the gene were mostly of the frameshift or nonsense type (35/38). It implies that the DNA-binding domain is sensitive even to single amino acid residue changes but the transcription activation domain is not and requires more severe changes such as frameshift or nonsense mutations. Noninducible, non-DNA-binding mutations altering proline-26 of the GAL4 DNA-binding domain to leucine or serine could be phenotypically cured with a high concentration of zinc in the growth medium. This result suggests that the mutationally altered structure of this site might have a decreased affinity for zinc ions (Johnston, 1987; Johnston and Dover, 1988).

The GAL4 protein was thought to bind DNA using a motif termed the zinc cluster (Pan and Coleman, 1990). In this structure two zinc ions are proposed to interact with all six cysteines with the residues between cysteines 2 and 3, 3 and 5, and 5 and 6 looping out to form a kind of cloverleaf. However, this structure was determined based on NMR analysis using cadmium ions instead of

zinc, which is the naturally occurring ion (Pan and Coleman, 1990). Each mole of the GAL4 protein found to bind 2 moles of cadmium ions while 1 to 1.5 moles of zinc ion were found to bind. Similar results have been obtained in several other laboratories and have led researchers to question the validity of using cadmium in place of zinc and thus the validity of the binuclear ion binding structure (Lena Basile, personal communication).

Maltose fermentation in *Saccharomyces*.

Saccharomyces strains capable of fermenting maltose carry one of five unlinked, telomere-associated *MAL* loci: *MAL1*, *MAL2*, *MAL3*, *MAL4* or *MAL6* (Barnett, 1976). Each locus is a cluster of 3 genes. GENE 1 and GENE 2 encode maltose permease and maltase, respectively (Goldenthal et al., 1983; ; Needleman et al., 1984; Cohen et al., 1985; Dubin et al., 1985; Chang et al., 1988). GENE 3 (referred to as *MAL13* at the *MAL1* locus and *MAL63* at the *MAL6* locus) encodes a regulator of maltose fermentation whose function is required for induction. This conclusion is based on the finding that several noninducible *mal6* mutations mapped within a restricted region of the *mal63* gene as well as on the finding that a deletion/disruption of the coding region of *MAL63* was also noninducible (Chang et al., 1988). Maltose fermenting revertants of the *mal63* mutations frequently were found to express the GENE 1 and GENE 2 products constitutively (ten Berge et al., 1973b; ten Berge et al., 1974; Dubin et al., 1986; Dubin et al., 1988). In addition, the GENE 3 homologue at the *MAL4* locus (*MAL43*) has been shown to encode a naturally occurring constitutive, glucose-repression-insensitive

allele (Charron and Michels, 1987). All of these results taken together provide strong evidence that the GENE 3 product is a *trans* acting regulator required for maltose induction. The MAL63 protein is proposed to activate transcription of GENE 1 and GENE 2 by binding to a specific DNA sequence located upstream to both GENE 1 and GENE 2, which is referred to as the UAS_{MAL} (see Figure 1). Deletion analysis of the *MAL61-MAL62* intergenic region revealed a 68 basepair region required for the maltose inducible expression of both *MAL61* and *MAL62* (Levine, Tanouye and Michels, submitted). This sequence, placed upstream of a heterologous gene, showed the maltose inducible expression of this gene and induction required a functional MAL-activator. Additionally, MAL63 protein binds within this 68 base pair region (Ni and Needleman, 1990).

My thesis research involved an analysis of *MAL63* with the goal of identifying the DNA-binding domain. I sequenced *MAL63* and found that residues 8 to 34 were rich in cysteine and basic amino acids and that this region of MAL63 was highly homologous to the DNA-binding domain of GAL4 (Kim and Michels, 1988). This work is described in Chapter 1 of this thesis. Three of the cysteine residues from this region and the proline at position 23 were altered by *in vitro* mutagenesis of the cloned *MAL63* gene. Strains carrying these mutant alleles were noninducible (or weakly inducible in the case of the proline alteration) and none of the mutant proteins were able to bind *in vitro* to an oligonucleotide containing the UAS_{MAL}. Thus, *MAL63*, and probably each of the GENE 3 homologues, encodes a DNA-binding protein which binds to the

UAS_{MAL} and functions in the transcription activation of the *MAL* structural genes.

Chapter 1

Sequence analysis of the *MAL63* gene of *Saccharomyces*

Introduction

The ability of a *Saccharomyces* yeast strains to ferment maltose is dependent upon the presence of at least one member of a polygenic family of loci referred to as the *MAL* loci (Barnett, 1976). Complete structural and functional analyses of two members of this family (*MAL1* and *MAL6*) have been reported (Needleman et al., 1984; Charron et al., 1986). Both are highly sequence homologous over an approximately 9.0 kbp region encoding three genes. GENE 1 (referred to as *MAL11* at the *MAL1* locus and *MAL61* at the *MAL6* locus) and GENE 2 (referred to as *MAL12* at the *MAL1* locus and *MAL62* at the *MAL6* locus) encode transcripts which are maltose inducible and several lines of evidence indicate that GENE 1 and GENE 2 encode maltose permease and maltase, respectively (Goldenthal et al., 1983; Needleman et al., 1984; Cohen et al., 1985; Dubin et al., 1985; Chang et al., 1988). GENE 3 (referred to as *MAL13* at the *MAL1* locus and *MAL63* at the *MAL6* locus) encodes a regulator of maltose fermentation whose function is required for induction. This conclusion is based on two findings: (1) that several noninducible *mal6* mutations mapped within a restricted region of the *mal63* gene, and (2) that a deletion/disruption of the coding region of *MAL63* was also noninducible (Chang et al., 1988). Maltose fermenting revertants of the *mal63* mutations frequently were found to express the GENE 1 and GENE 2 products constitutively (ten Berge et al., 1973b; ten Berge et al., 1974; Dubin et al., 1986; Dubin et al., 1988). In addition, the GENE 3 homologue at the *MAL4* locus (*MAL43*) has been shown to encode a

naturally occurring constitutive, glucose-repression-insensitive allele (Charron and Michels, 1987). All of these results taken together provide strong evidence that the GENE 3 product is a *trans* acting regulator required for maltose induction. It has been suggested that the *MAL63* gene encodes a protein that is a MAL activator whose function is comparable to that of the GAL4 protein in the induction of the galactose fermentative enzymes. This report provides evidence in support of this hypothesis by demonstrating that the single large open reading frame found in the region containing the *MAL63* gene encodes a protein with several features characteristic of yeast transcription activators and thus is a candidate for the MAL-activator.

Materials and Methods

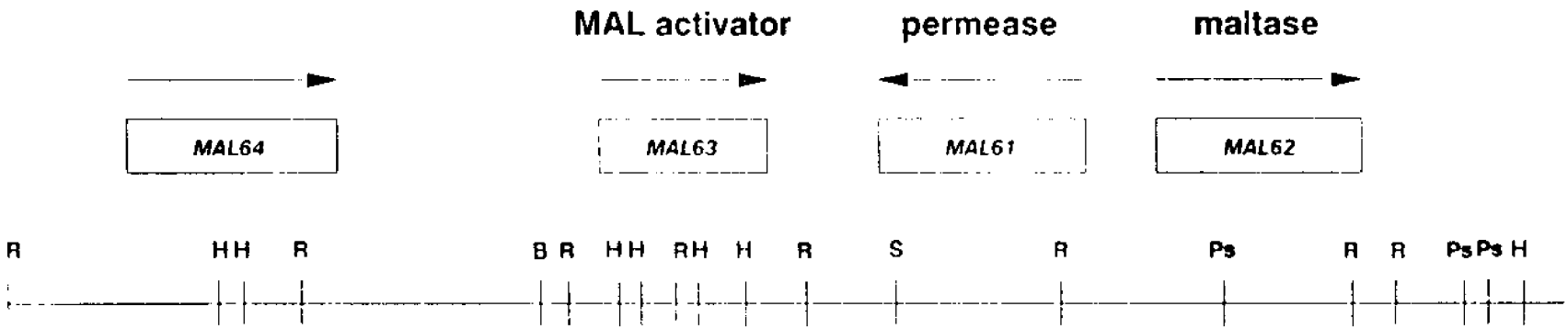
A *SalI* fragment containing the *MAL63* gene was subcloned into the M13 phage vector mp19 in both orientations and nested sets of deletions were constructed by unidirectional degradation of the insert fragment using exonuclease III. Linearization of the phage DNA was carried out by cleaving both *BamHI* and *SstI* sites in the polylinker (Henikoff, 1984). Sequencing was done by the method of Sanger et al. (1977) using the universal primer (Messing, 1983). Data analysis was carried out on BIONET, Intelligenetics, Mountainview, California.

Results and Discussion

Figure 1 shows a restriction map of the DNA region containing *MAL63*. The sequence shown in Figure 2 extends 1816 bases from the *EcoRI* site upstream of the *MAL63* gene to a downstream *SspI* site.

Figure 1: Restriction map of the *MAL6* locus.

The location of the *MAL63* gene in the *MAL6* locus is shown. The direction and size of the arrows show the direction of transcription and the size of open reading frame. The scale shown at the bottom of the figure (200 base pair) is for the expanded *MAL63* gene only. The *MAL64* gene is a nonfunctional homolog of the *MAL63* gene (Dubin et al., 1988). Recognition sites for restriction endonucleases are abbreviated as follows: B, *Bam*HI; C, *Cla*I; H, *Hind*III; N, *Nco*I; Ps, *Pst*I; R, *Eco*RI; S, *Sal*I; Sp, *Spe*I; Ss, *Ssp*I; X, *Xba*I.


MAL activator
permease
maltase
MAL64
MAL63
MAL61
MAL62
R
H H
R
B R
H H
R H
H
R
S
R
Ps
R
R
Ps Ps H


B
R
HSs HX
R
XH
NCSp
H
Ss
R
Sp
S

0 200

A single open reading frame 1,410 base pairs in length is seen encoding a protein of 470 amino acid residues. The orientation of this ORF is in agreement with the reported direction of transcription of the *MAL63* gene (Needleman et al., 1984). In addition, both the *MAL63* and *MAL13* transcripts are estimated to be 1.6 kb in length which is consistent with the postulated size of the MAL63 protein deduced from this nucleotide sequence (Needleman et al., 1984; Charron et al., 1986). While we have yet to isolate the protein, we refer to the product of this ORF as the MAL63 protein.

Several AT-rich sequences are seen in the region 5' to the open reading frame. One, at position -222 to -217, corresponds to a reported functional yeast "TATA" element (Chen and Struhl, 1988). Whether it functions as such in the *MAL63* gene remains to be determined. The region 3' to the open reading frame is also AT-rich and does contain the sequence 5'-TAG-space-TATGT-space-TTT-3', starting at base 1,445, postulated as a yeast transcription termination signal but the length of the spacer regions are longer than the consensus sequence of Zaret and Sherman (1982). These transcription signals are underlined in Figure 2.

The most striking feature of the sequence of the MAL63 protein is the presence of a cysteine-lysine-arginine rich region near the N-terminal end of the protein indicated in Figure 2 beginning at residue 8 of the deduced protein. Sequences similar to this are seen in many eukaryotic DNA-binding and transcriptional regulatory proteins including TFIIIA, ADR1, glucocorticoid receptor, estrogen

Figure 2: Nucleotide and deduced amino acid sequence of the *MAL63* gene.

The numbering of the DNA sequence is from the start of translation. The region of cysteine-zinc finger is *shaded*. Regions rich in acidic residues are *double underlined*. Potential "TATA" element and transcription termination signal are *single underlined*. A potential nuclear localization signal is *underlined with a filled bar*.

```
                -258
                GAATTCTT TTAACTCAAT AGTAATATGC ATTGTTCTTA TCTAAAAAAT TGCAGGTACC TGCAGACGAA TCCGGGTCAT
-180
GATCTGCGCT GCGCCGTCAT CCCACCCCGT GCTGCCTGCC ACTTAAAGCT ACCCCGGGTT TAATAATTCG TTCTTTAAGT TCTACAACTT
-90
AAATACAGGC AGCTAAAAAA CTGGGTTCGA GAGTTTTCCA CTTTACAGAC AAAAATAAAA ATACTGCCAG AAAATTTATC ATATAATAAT

1                                                30                                                60
ATG GGT ATT GCG AAA CAG TCT TGC GAC TGC TGT CGC GTT CGT CGA GTA AAG TGT GAC AGG AAT AAA CCA TGT AAT
 M   G   I   A   K   Q   S   C   D   C   C   R   V   R   R   V   K   C   D   R   N   K   P   C   N

                90                                              120                                             150
CGC TGC ATT CAG CGC AAT TTG AAC TGC ACT TAT CTT CAA CCG TTG AAA AAG AGA GGT CCA AAA TCC ATT AGA GCA
 R   C   I   Q   R   N   L   N   C   T   Y   L   Q   P   L   K   K   R   G   P   K   S   I   R   A

                180                                             210
GGA AGC TTA AAA AAA ATA GCG GAA GTG CAG ATG GTG AGT ATG AAT AAT AAT ATT ATG GCC GCT CCG GTG GTA TGT
 G   S   L   K   K   I   A   E   V   Q   M   V   S   M   N   N   N   I   M   A   A   P   V   V   C

                240                                             270                                             300
AAG AAG GTT CCG AAA AAC CTG ATT GAT CAA TGT TTG AGG TTG TAT CAT GAT AAC TTA TAT GTA ATT TGG CCA ATG
 K   K   V   P   K   N   L   I   D   Q   C   L   R   L   Y   H   D   N   L   Y   V   I   W   P   M

                330                                             360
CTT TCC TAC GAT GAT CTT CAC AAG CTT CTA GAA GAG AAA TAC GAT GAC CGT TGC GCC TAC TGG TTT CTG GTA TCC
 L   S   Y   D   D   L   H   K   L   L   E   E   K   Y   D   D   R   C   A   Y   W   F   L   V   S

                390                                             420                                             450
CTT TCG GCA GCT ACT CTT AGC GAC TTG CAA ATT GAA ATA GAG TAT GAG GAA GGA GTC ACT TTT ACT GGA GAG CAG
 L   S   A   A   T   L   S   D   L   Q   I   E   I   E   Y   E   E   G   V   T   F   T   G   E   Q

                480                                             510
TTG TGC ACT CTT TGC ATG TTA TCT CGG CAA TTC TTT GAC GAC CTT AGT AAC AGC GAC ATA TTT CGA ATC ATG ACA
 L   C   T   L   C   M   L   S   R   Q   F   F   D   D   L   S   N   S   D   I   F   R   I   M   T

                540                                             570                                             600
TAC TAT TGT TTG CAC CGT TGT TAC GCG CAG TTT GCG GAT ACG AGA ACT TCA TAT AGA CTT TCT TGT GAA GCC GTG
 Y   Y   C   L   H   K   C   Y   A   Q   F   A   D   T   R   T   S   Y   R   L   S   C   E   A   V

                630                                             660
GGC CTC ATC AAG ATA GCT GGA TTC CAT CGG GAA GAA ACC TAT GAA TTC CTT CCC TTC GGT GAA CAA CAA CTC AGA
 G   L   I   K   I   A   G   F   H   R   E   E   T   Y   E   F   L   P   F   G   E   Q   Q   L   R

                690                                             720                                             750
AGG AAA GTT TAC TAT TTA CTT CTT ATG ACA GAG AGA TTT TAC GCT GTA TAT ATT AAG TGT GTC ACG AGC CTA GAT
 R   K   V   Y   Y   L   L   L   M   T   E   R   F   Y   A   V   Y   I   K   C   V   T   S   L   D

                780                                             810
GCA ACA ATA GCG CCA CCA CTA CCA GAG GTT GTA ACA GAC CCT CGT CTT TCT CTA GAA AGC TTC CTT GAG GTG ATT
 A   T   I   A   P   P   L   P   E   V   V   T   D   P   R   L   S   L   E   S   F   L   E   V   I

                840                                             870                                             900
AGA GTT TTC ACT ATA CCA GGA AAG TGT TTT TAT GAT GCT TTG GCT ACT AAC TGT GTC GAT GAT TCC TGC ACC GAA
 R   V   F   T   I   P   G   K   C   F   Y   D   A   L   A   T   N   C   V   D   D   S   C   T   E

                930                                             960
GAC TCT CTA AAA AGG ATA CGG ACC GAA CTT CAT ACC ACA TCA CTT GAT ATA GAG CCA TGG TCT TAC GGA TAC ATC
 D   S   L   K   R   I   R   N   E   L   H   T   T   S   L   D   I   E   P   W   S   Y   G   Y   I

                990                                            1020                                            1050
GAT TTT CTG TTT TCG AGG CAT TGG GTC AGG ACA CTA GCG TGG AAA CTA GTA CTT CAT ATG AAA GGT ATG CGG ATG
 D   F   L   F   S   R   H   W   V   R   T   L   A   W   K   L   V   L   H   M   K   G   M   R   M

               1080                                            1110
AAT TTT CTT TCG AAT ACT AAT AAC ACA CAT ATA CCA GTC GAA ATT GCT AGA GAC ATG TTG GGA GAC ACG TTT TTA
 N   F   L   S   N   T   N   N   T   H   I   P   V   E   I   A   R   D   M   L   G   D   T   F   L

               1140                                            1170                                            1200
ACT CCG AAA AAC CTG TAT GAT GTA CAT GGT CCT GGA ATA CCG ATG AAG GCA TTA GAA ATA GCC AAT GCA TTG GTA
 T   P   K   N   L   Y   D   V   H   G   P   G   I   P   M   K   A   L   E   I   A   N   A   L   V

               1230                                            1260
GAT GTC GTA AAT AAG TAT GAT CAC AAT ATG AAG TTG GAA GCT TGG AAT GTT TTG TAT GAT GTA TCC AAG TTT GTT
 D   V   V   N   K   Y   D   H   N   M   K   L   E   A   W   N   V   L   Y   D   V   S   K   F   V

               1290                                            1320                                            1350
TTT TCT CTG AAA CAT TGC AAT AAT AAA ATG TTC GAC AGA TTT TCA ACG AAA TGT CAA GGT GCC CTA ATT ACT CTG
 F   S   L   K   H   C   N   N   K   M   F   D   R   F   S   T   K   C   Q   G   A   L   I   T   L

               1380                                            1410
CCC ATT TCT AAA CCT TTG CAA TTA AAT GAT AAC TCC AAA GAT GAA GAC GAC ATA ATT CCT TAA TTTATTG
 P   I   S   K   P   L   Q   L   N   D   N   S   K   D   E   D   D   I   I   P   *

      1430                                            1450                                                    1500
TTCACGCCGT TCACTTATAC GAGATAGATA TACTGATAGA GTGTGAGCGA TATTCTTAAG TCTTGCTTTT CGAGGGTGTA

                                                      1550
AGAAGCTATG TTCAGGCGAGATTATTCTAC TCCTGCCTTA CTTGTTTGTA ATATT
```

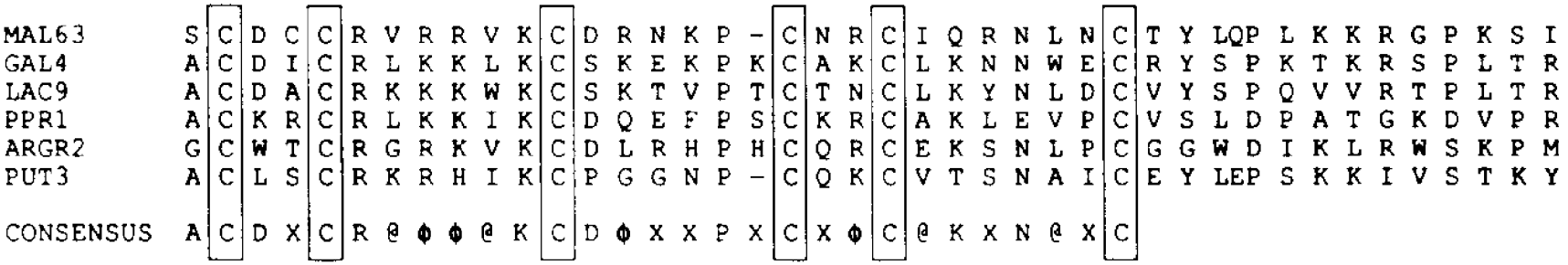
receptor (Harrison, 1991). In each of these, it is suggested that zinc ions interact with the cysteine residues or cysteine and histidine residues to form the DNA-binding structure. Direct evidence that the cysteine-rich region is part of the DNA-binding domain comes from studies by Johnston and Dover (1987) on the GAL4 protein. Of 30 noninducible *gal4* missense mutations, 27 were found to abolish or severely reduce the DNA-binding activity of the protein and all of these were alterations in the cysteine-rich region near the N-terminal end of the protein. The remaining three mutations mapped to the middle of the GAL4 protein and did not affect the ability of the GAL4 protein to bind DNA.

The results of Johnston and Dover (1987) are consistent with those of Keegan et al. (1986) and clearly delineate the essential region of the DNA binding domain of GAL4 protein as lying within the 42 amino acid residues shown in Figure 3. Figure 3 compares this sequence from GAL4 to that from MAL63 protein and several other yeast transcription activation proteins. The striking sequence similarity between the cysteine-rich region of the activator shown in Figure 3 offers strong support for the proposal that each of these proteins is a DNA-binding protein (Evans and Hollenberg, 1988).

In several recent reports, regions rich in acidic amino acid residues have been implicated in the transcriptional activation function of proteins such as GAL4 and GCN4 (Hope and Struhl, 1986; Gill and Ptashne, 1987; Ma and Ptashne, 1987; Giniger and Ptashne, 1987). Similar acidic regions are found in the MAL63 protein and

Figure 3: Homologies among the cysteine-rich region of several yeast transcription activation proteins.

The sequences of these proteins are reported in references Kammerer et al., 1984 (PPR1); Laughon and Gesteland, 1984 (GAL4); Messenguy et al., 1986 (ARGR2); Salmeron and Johnston, 1986 (LAC9) and Wray et al., 1987 (LAC9); Marczak and Brandriss, 1991 (PUT3). -: no amino acid residue at the position; @: hydrophobic; ϕ: Lysine or Arginine. The boxed cysteine residues in MAL63 are the 8th, 11th, 18th, 24th, 27th and 34th amino acid residues, starting from the left.


MAL63 S C D C C R V R R V K C D R N K P - C N R C I Q R N L N C T Y LQP L K K R G P K S I
GAL4 A C D I C R L K K L K C S K E K P K C A K C L K N N W E C R Y S P K T K R S P L T R
LAC9 A C D A C R K K K W K C S K T V P T C T N C L K Y N L D C V Y S P Q V V R T P L T R
PPR1 A C K R C R L K K I K C D Q E F P S C K R C A K L E V P C V S L D P A T G K D V P R
ARGR2 G C W T C R G R K V K C D L R H P H C Q R C E K S N L P C G G W D I K L R W S K P M
PUT3 A C L S C R K R H I K C P G G N P - C Q K C V T S N A I C E Y LEP S K K I V S T K Y
CONSENSUS A C D X C R @ ϕ ϕ @ K C D ϕ X X P X C X ϕ C @ K X N @ X C

these are indicated in Figure 2 by the double underline. The extreme C-terminus of MAL63 is very acidic in character and similar to the C-terminal region of GAL4 which has been shown to function as the transcription activation domain (Ma and Ptashne, 1987). Interestingly, the sequence contains a serine residue which is within a potential casein kinase II phosphorylation site (Cohen, 1988). Phosphorylation of this residue could possibly play a role in the functional activation of MAL63.

Nuclear targeting sequences have been identified in several proteins (Rihs et al., 1991) These sequences appear to consist of two clusters of basic residues separated by approximately 10 other residues. The MAL63 sequence from residues 41 to 55 conforms well to the reported consensus sequence and lies just C-terminal to the proposed DNA-binding domain. It is interesting to note that many of the known nuclear localization signals have closely associated casein kinase II phosphorylation sites and that phosphorylation of these sites enhances the nuclear targeting ability of the protein (Rihs et al., 1991).

If MAL63 protein is responsible for the transcription activation of the structural genes encoding the maltose fermentative enzymes, it is expected to localize to the nucleus, bind to the upstream activation sequence of the *MAL* structural genes, interact with the transcription machinery, and possibly to bind maltose and/or interact with other regulators of maltose fermentation (although none has as yet been identified). The results reported here strongly support the hypothesis that MAL63

protein is a DNA-binding protein. It has yet to be demonstrated directly that the cysteine-rich region contained in this protein functions in DNA-binding.

Chapter 2

Identification of the DNA-binding Domain of the *Saccharomyces* MAL-activator Encoded by *MAL63*

Introduction

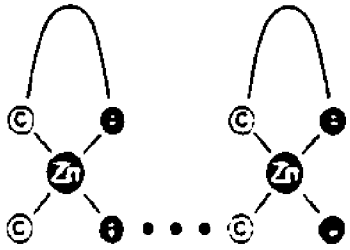
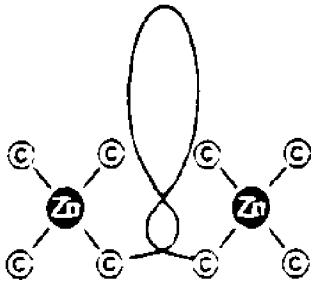
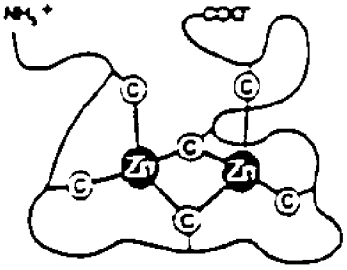
Saccharomyces strains capable of fermenting maltose carry one of five unlinked, telomere-associated *MAL* loci: *MAL1*, *MAL2*, *MAL3*, *MAL4* or *MAL6* (Barnett, 1976). Each locus is a cluster of 3 genes. GENE 1 and GENE 2 encode maltose permease and maltase, respectively (Goldenthal et al., 1983; Needleman et al., 1984; Cohen et al., 1985; Dubin et al., 1985; Chang et al., 1988). GENE 3 encodes the *trans* activator required for maltose-induced transcription of GENE 1 and GENE 2 (Chang et al., 1988). The GENEs at the different *MAL* loci are named using a two digit numbering system with the first digit indicating the locus number and the second the GENE function. Thus *MAL63* encodes the activator at *MAL6*.

MAL63 mutants, including a deletion/disruption, are noninducible for the *MAL* structural genes and are unable to ferment maltose (Chang et al., 1988). MAL63 protein activates transcription of GENE 1 and GENE 2 by binding to a specific DNA sequence located upstream to both GENE 1 and GENE 2, which is referred to as the UAS_{MAL}. Deletion analysis of the *MAL61-MAL62* intergenic region revealed a 68 basepair region containing the UAS_{MAL} (see Figure 1) and regulating expression of both *MAL61* and *MAL62* (Levine, Tanouye and Michels, submitted). This maltose-inducible expression was dependent upon a functional MAL-activator. The evidence that MAL63 protein directly binds to this sequence comes from footprint analysis and a filter-binding assay using *E. coli* synthesized MAL63 protein (Ni and Needleman, 1990). We wished to define the DNA-binding domain of MAL63.

We reported the *MAL63* gene sequence (Kim and Michels, 1988). It shows that the gene encodes a protein of 470 amino acid residues with certain features characteristic of the class of DNA-binding proteins containing six cysteine residues and in a basic amino acid-rich cluster, similar to one found in the GAL4 protein of *Saccharomyces* (Laughon and Gesteland, 1984). Three cysteine-rich DNA-binding motifs have been described in the literature all containing zinc ions (showed in Vallee et al., 1991). These are shown in Figure 4 (taken from Vallee et al., 1991). In zinc-finger motif the first, second, fourth and fifth cysteine residues of the cysteine-basic amino acid-rich region are involved in chelating a single zinc ion while in the binuclear cluster structure all six cysteine residues are involved in the chelation of two zinc ions (Pan and Coleman, 1989). In the study described here, we present evidences that the cysteine-rich, N-terminal region of MAL63 is required for binding to the UAS_{MAL}. In addition, we used mutation analysis to investigate the importance for DNA-binding of the various cysteine residues and the single proline in this region. If zinc-finger motif is formed, the third and sixth cysteine residues (see Figure 3) in the cysteine-basic amino acid-rich region might not be critical for DNA-binding while the first, second, fourth and fifth cysteine residues would be expected to be very important. If this region of MAL63 forms a zinc cluster, all six cysteines would expected to be critical for DNA-binding. Thus we decided to alter the third, fifth and sixth cysteines to examine the effect of these changes on DNA-binding. All of the cysteines

Figure 4: Proposed structure of the cysteine-rich DNA-binding motifs.

This figure is taken from Vallee et al. (1991).

DOMAIN	STRUCTURE	PROTEIN
ZINC FINGER		TFIIIA
ZINC TWIST		Glu Rec
ZINC CLUSTER		GAL4

that we tested were found to be critical for DNA-binding.

Materials and Methods

Strains and growth conditions. *Saccharomyces cerevisiae* strain 340-2A (*MATα MAL11-2 MAL12 ura3-52 ade*) was used to determine the *in vivo* ability of the altered MAL63 protein to function as a transcription activator of the structural genes, *MAL11-2* and *MAL12*, encoding maltose permease and maltase, respectively. The strain lacks sequences encoding a MAL-activator (Charron et al., 1986). Maltose fermentation is defined as the production of acid and gas in 1-5 days after inoculation and determined in 5 ml of YEP medium [1% (wt/vol) yeast extract/1%(wt/vol) peptone] plus 2% (wt/vol) maltose in Durham tubes. Maltase levels were determined using cell extracts as described by Dubin et al. (1985). The cells were grown in synthetic media (SM) supplemented with adenine plus 2% lactate, 3% glycerol and 2% maltose for induced conditions or 2% lactate and 3% glycerol for uninduced conditions (Hopper et al., 1978). Activities are reported as μmoles of p-nitrophenol released from PNPG (p-nitrophenyl-α-D-glucopyranoside) per minute per mg of protein. Protein determinations were carried out using the Bio-Rad Protein Determination kit.

E. coli strain MV1190 (*Δlac-pro*AB), *thi*, *sup*E, *Δ*(*sr*l-*rec*A) 306::Tn10(*tet*^{r}) [F': *tra* D36, *pro*AB, *lac*I^{q}ZΔM15] and CJ236 (dut, ung, thi, relA; pCJ105(Cm^r) were obtained from Bio-Rad as part of the *in vitro* Mutagenesis Kit. MV1190 was used to propagate plasmids for large scale preparation and for the generation of single stranded DNA's for sequencing as well as in site directed

mutagenesis to select against uracil containing DNA. Strain CJ236 was used to prepare the uracil containing single stranded DNA used as a template in site-directed mutagenesis. The BL21(DE3)pLysS strain was used for protein expression. This strain was a gift from William Studier (Studier and Moffatt, 1986).

Plasmids. The mp19 RF DNA was purchased from Pharmacia. This vector was used for site-directed mutagenesis and sequencing. The pRS316, a yeast-*E. coli* shuttle vector was a gift from Robert S. Sikorski (Johns Hopkins University) (Sikorski and Hieter, 1989). This vector bears the yeast *URA3*, *CEN6*, and *ARSH4* sequences and thus is present as a single copy in yeast. The pET-8c plasmid, a protein expression vector, was a gift from William Studier (Rosenberg et al., 1987).

Oligonucleotide-Directed *in vitro* Mutagenesis. A *Sal*I fragment containing the *MAL63* gene flanked by approximately 500 base pairs on each sides of the open reading frame was inserted into mp19 at the *Sal*I site of the multiple cloning site with the 5' end of the gene oriented toward the *Hind*III site of the vector. This fragment is derived from plasmid p40Leu (Charron et al., 1986). It contains the 3 kb *Bgl*II-*Sal*I fragment from *MAL6* containing *MAL63* and fully complements *MAL*-activator gene mutations (Charron et al., 1986). This construction was used for all site-directed mutagenesis.

Oligonucleotide-directed mutagenesis was carried out using a kit purchased from Bio-Rad. The method provides for enhanced replication of mutated sequences by utilizing the inability of

uracil-containing DNA to replicate in wild-type *E. coli*. The sequence to be mutagenized was cloned into an M13 vector and uracil-containing single stranded DNA is prepared in strain CJ236 (see above). Oligonucleotides complementary to the cloned sequence but containing sequence alterations, are annealed to this single stranded DNA and act as primers for T4 polymerase which synthesizes the full complementary strand: this amplified strand is then transformed into strain MV1190 (see above) which preferentially inhibits the replication of the uracil-containing wild-type strand.

Oligonucleotides were synthesized with the model 381A DNA synthesizer from Applied Biosystems. The following mutations were created using the indicated oligonucleotides. The *Nco*I site at codons 319 to 320 nucleotide of *MAL63* was removed without affecting amino acid sequence using the oligonucleotide: 5'-GAT GTA TCC GTA AGA CCA GGG CTC TAT ATC AAG TGA TGT-3'. An *Nco*I site was created at codon 1 using the oligonucleotide 5'-AGA CTG TTT CGC AAT ACC CAT GGT ATT ATA TGA TAA ATT TTC TG-3'. This altered *MAL63* allele is referred to as *MAL63/N* and the M13 plasmid containing this gene mp19-*MAL63/N*. The *Sal*I fragment containing the *MAL63/N* gene was cloned into pRS316 and tested for its ability to activate maltose fermentation in yeast. This altered gene showed virtually no difference in maltose fermentation from the wild type *MAL63* gene. Using the MAL63/N allele the following changes were made using the indicated oligonucleotides: Cys18Leu (5'-ACA TGG TTT ATT CCT GTC AAG CTT TAC TCG ACG AAC GC), Cys27Ser (5'-GTT CAA ATT GCG CAA GTT TAA CGC CTG AAT CGA TCG ATT ACA TGG TTT ATT), Cys34Gly (5'-TTT CAA

CGG TTG AAG ATA AGT ACT GTT CAA ATT GCG CTG AAT), and Pro23Leu (5'-CTG AAT GCA GCG ATT ACA AAG CTT ATT CCT GTC ACA CTT TAC). All of these altered *MAL63* genes were sequenced in their entirety to confirm that no changes other than those intended had been made. These alleles are referred to as *MAL63/N*Leu18, *MAL63/N*Ser27, *MAL63/N*Gly34 and *MAL63/N*Leu23, respectively.

Expression of *MAL63* in *E. coli*. The T7 expression system (Studier and Moffatt 1986; Rosenberg et al., 1987) was used to synthesize high levels of MAL63 protein in *E. coli*. The *Nco*I-*Bam*HI fragment from mp19MAL63/N, containing the full *MAL63/N* open reading frame from codon 1 to the *Bam*HI site in the multiple cloning site, was subcloned into plasmid pET-8c so as to fuse the T7 gene10 upstream sequences at the *Nco*I site. Initial experiments using this system showed that expression of the full length 470 amino acid long MAL63 protein was poor even with the use of rifampicin which reportedly enhances expression. Expression was undetectable in Coomassie stained gels and barely detected in ^{35}S-methionine pulse-labeled cells. Thus we decided to express a truncated MAL63 protein. Two truncation constructions were made. In the first, an *Spe*I fragment was deleted from the *MAL63/N* sequences in mp19MAL63/N so as to fuse codon 341 of *MAL63/N* to an *Spe*I site located after the termination codon, thereby creating a 129 residue truncated MAL63 protein containing the N-terminal 342 residues of MAL63 plus one random residue (see Figure 1). The predicted molecular weight of this deletion product is 39.7 Kd. The resulting construction is referred to as mp19-*MAL63/N*Δ*Spe*I. In the second construction, the

*Xba*I fragment from plasmid mp19-*MAL63/N* was deleted, resulting in the fusion of the *Xba*I site at codon 109 of *MAL63/N* to an *Xba*I site in the mp19 multiple cloning site. This construction encodes residues 1-111 of MAL63 protein and 21 residues from the vector sequences. The predicted size of the deletion product is 15.3 Kd. The resulting construct is referred to as mp19-*MAL63/NΔXba*I. The *Nco*I/*Bam*HI fragment containing *MAL63/NΔSpe*I or *MAL63/NΔXba*I extending from the *Nco*I site at codon 1 to the *Bam*HI site of the multiple cloning site was inserted in the *Nco*I/*Bam*HI site of pET-8c vector for expression.

Strain BL21(DE3) pLysS was used as the *E. coli* host. It contains the T7 RNA polymerase gene under the control of the *lac* promoter, which can be induced by IPTG. This strain also carries plasmid pLysS containing the gene encoding T7 lysozyme, *lys*, and the chloramphenicol resistance gene, *chr*^{r}. The T7 lysozyme encoded by the *lys* gene prevents even basal level expression of the T7 RNA polymerase target gene. Synthesis of lysozyme in induced cells also renders them sensitive to lysis following repeated freezing and thawing. The pET-8c*MAL63/NΔSpe*I and pET-8c*MAL63/NΔXba*I plasmid constructions were transformed into this strain for expression. Transformants were maintained on solid ZY medium (1% wt/vol Bactotryptone, 0.5% wt/vol Bacto yeast extract and 0.5% wt/vol NaCl) plus 20 μg/ml ampicillin and 25 μg/ml chloramphenicol. Only freshly grown colonies were used for protein expression.

A single transformant colony was grown overnight in 5 ml of ZY medium containing chloramphenicol and ampicillin, and a 0.5 ml

aliquot of the overnight was inoculated into 20 ml ZY medium plus chloramphenicol and ampicillin. At an OD_{600} of 0.5 to 0.7, IPTG was added to a final concentration of 0.4 mM. After 2 hours of induction the cells were processed for protein extraction. Cells were harvested by centrifugation for 5 minutes at 5900 g and pellet was saved. The pelleted cells were washed with 1 ml of K(50) buffer [25 mM Hepes (pH 7.5), 0.1 mM EDTA, 0.5 mM ß-mercaptoethanol, 10% glycerol]. To this washed pellet 1 ml of K(50) was added and the sample was kept frozen at -70°C until used. This frozen pellet was vortexed while being thawed to lyse the cells. These cell extracts were centrifuged at 9500 g and the supernatant used in the gel mobility shift DNA-binding assays.

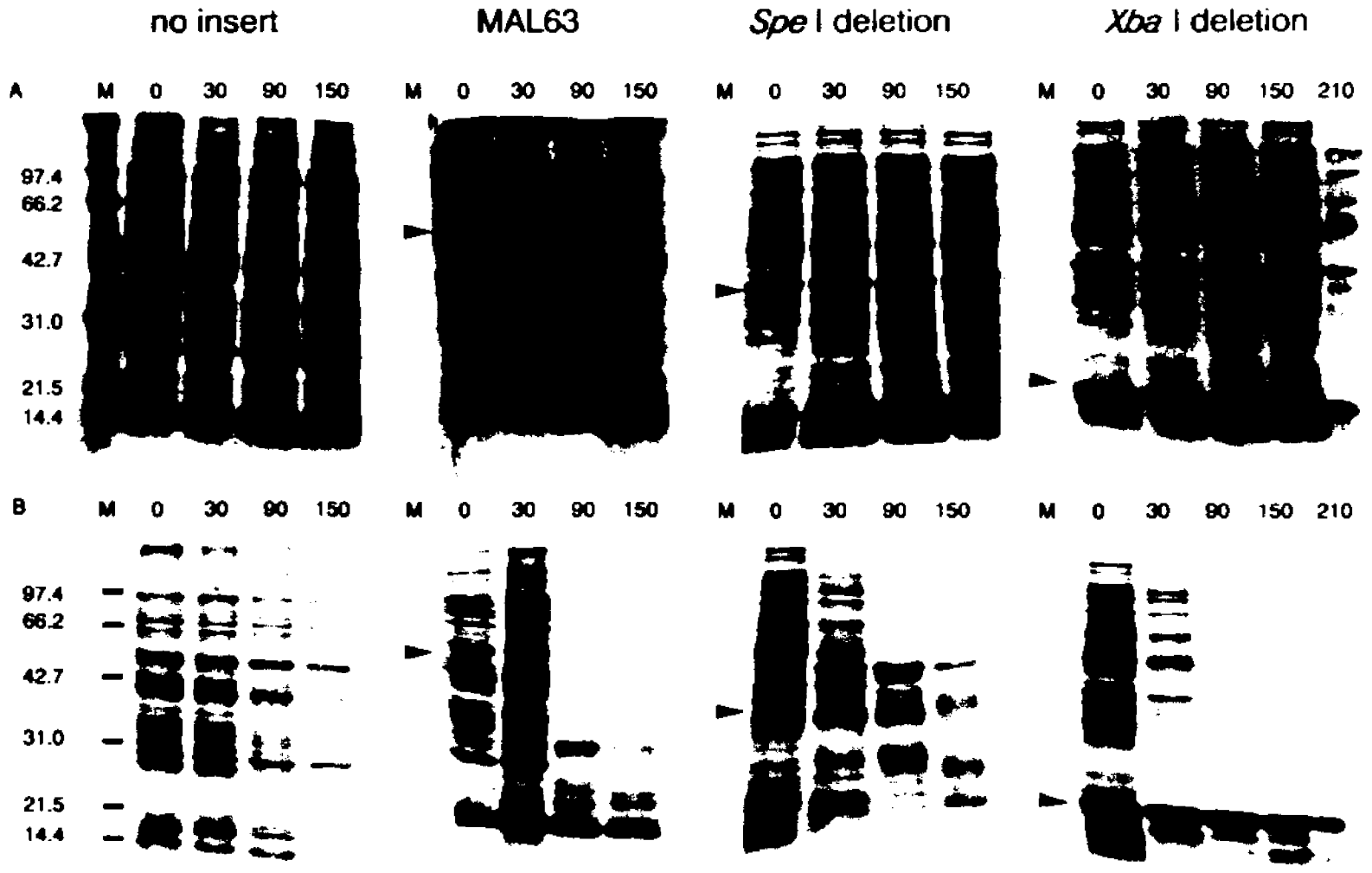
The K(50) buffer used in these experiments is essentially the same as the A(50) buffer used for GAL4 protein-UAS_{GAL} filterbinding tests with yeast extract (Bram and Kornberg, 1985), with *E. coli* extract (Johnston and Dover, 1987) and gel mobility shift assay with yeast extract (Salmeron et al., 1989) with the following changes. $MgCl_2$ was deleted to suppress DNAse activity in the cell extract, ß-mercaptoethanol (1mM) was used instead of dithiothreitol to prevent any metal chelation activity, and protease inhibitors were not included.

To estimate protein expression, the induced cells were pulse-labelled with ^{35}S-methionine as follows (Studier and Moffatt, 1986). For this experiment M9 medium (0.1%, NH_4Cl, 0.13% KH_2PO_4, 0.16% Na_2HPO_4, 0.14% Glucose, 1 mM $MgSO_4$) instead of ZY medium was used to enhance ^{35}S-methionine incorporation into the protein being

synthesized. 1.0 ml of *E. coli* culture was taken immediately before and 1.5 hours after induction with IPTG and labelled with 0.5 μCi of ^{35}S-methionine for 5 minutes at 37°C. The labeled cells were harvested by centrifugation, lysed with 200 μl of 2X loading buffer (0.125 M Tris-Cl pH6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol) (Laemmli, 1970), and heated in boiling water for 1.5 minutes. The lysed cells were centrifuged for 5 minutes and the supernatant proteins were analyzed by PAGE-SDS gel analysis according to the procedure of Laemmli (1970). The gel was stained with Coomassie Blue, dried, and autoradiographed. Figure 5 shows the results of this expression system using the full-length gene from plasmid pET-8c*MAL63/N* and the two truncated constructions, pET-8c*MAL63/NΔSpe*I and pET-8c*MAL63/NΔXba*I. The Coomassie blue stained gel (upper panel) is compared to an autoradiograph of the same gel (lower panel). Nc maltose-inducible protein of the expected size (55 kd) is seen from the full length *MAL63* gene. Expression of an approximately 36 kd IPTG-induced protein is barely detectable by autoradiogrph in the *Spe*I deletion construction.

Figure 5: MAL63 Protein Expressions in *E. coli*.

The *MAL63* gene was fused to the T7 gene10 promoter as described in **Materials and Methods.** Expression from this full length gene (plasmid pET8c-*MAL63/N*) and two shorter length truncations (plasmids pET8c-*MAL63/NΔSpe*I and pET8c-*MAL63/NΔXba*I) was followed following induction with IPTG. Proteins were labeled by a 5 minute pulse with ^{35}S-methionine. The upper panel (A) shows Coomassie stained gels. The lower panel (B) shows the autoradiograph of the same gel. The control follows expression in cells containing the pET-8c plasmid lacking an insert. The arrows alongside the autoradiographs indicate the anticipated size of the MAL63 product.


no insert
MAL63
Spe I deletion
Xba I deletion
A
B
M 0 30 90 150
M 0 30 90 150 210
97.4
66.2
42.7
31.0
21.5
14.4

In contrast, the *Xba*I deletion construction clearly shows an abundant approximately 17 kd IPTG-induced protein which was the sole protein expressed at one and half hours of induction.

The gel mobility shift DNA-binding assay. DNA-binding activity was determined using a 40 basepair oligonucleotide synthesized as follows. Two partially complementary oligonucleotides, 5'-GTT TAC AGG ATT TAT CCG GAA ATT TTC GCG G-3' and 5'-TA GGC CTT TAA AAG CGC CTG GGG TGT G-3', were annealed and the ends were filled in with ^{32}P-dCTP and nonradioactive nucleotides using the Sequenase kit (purchased from USB). This labelled DNA was precipitated with 80% ethanol (vol/vol) and resuspended in TE at concentration of 1.0 μg/μl.

In each 1.5 ml eppendorf tube, 5 μl of protein extract prepared as described above (2.5 μg of protein), 0.15 μl of labelled DNA (50,000 cpm or 0.15 μg), 0.15 μl of poly dIdC (0.2 μg), and 19.7 μl of K(50) buffer were mixed and left at room temperature for 5 to 10 minutes. This reaction mix was loaded onto a 4% polyacrylamide gel (80:1 acrylamide and bisacrylamide) and run for 2.5 hours at 22 mA. The polyacrylamide gel [50 mM Tris base (pH 8.5), 380 mM glycine, 2.1 mM EDTA, 4% acrylamide, 0.05% bisacrylamide, 2.5% glycerol] and running buffer [50 mM Tris base (pH 8.5), 380 mM glycine, 2.1 mM EDTA] were prepared as described by Chodosh (1988). After electrophoresis the gel was dried and subjected to auto-radiography.

Results

The isolation of maltose nonfermenting mutations in *MAL63* has

been reported by ten Berge et al. (1973a) and Chang et al. (1988). Sequence analysis of several of these mutant alleles (to be reported elsewhere) revealed that none of the four missense mutations sequenced had an alteration in the cysteine-basic amino acid-rich region which had been proposed to be the DNA-binding domain of the protein (Kim and Michels, 1988). This result was surprising in light of the results of Johnston and Dover (1987, 1988) who found that 27 of 31 galactose nonfermenting *gal4* mutant alleles contained alterations in the DNA-binding domain of this transcription activator. The GAL4 DNA-binding domain is highly homologous to the cysteine-basic amino acid-rich region near the N-terminal end of MAL63. The experiments described here are designed to demonstrate whether or not this region of MAL63 is the DNA-binding domain of the protein.

Functional activity of mutant *mal63* in yeast. When the cysteine-basic amino acid-rich region of MAL63 is aligned with the DNA binding domain of GAL4, six cysteine residues of MAL63, (Cys8, Cys11, Cys18, Cys24, Cys27 and Cys34) were found to lie at sites comparable to those in the DNA-binding domain of GAL4 (see Figure 3). Analysis of the structure of the GAL4 DNA-binding domain using NMR suggests that these six cysteines are involved in the chelation of two zinc ions to form a binucleated structure quite distinct from the zinc-finger structure proposed for TFIIIA (Miller et al., 1985; Pan and Coleman, 1990). Given this structure all six cysteine residues should be expected to be required for the proper formation of the DNA-binding domain. If the homologous region of

MAL63 is the DNA-binding domain then alterations of these cysteine residues should abolish DNA binding. Thus, we chose to mutate cysteine-encoding codons 18, 27 and 34 to test the importance of the encoded residues in DNA-binding. Additionally, the DNA-binding domain of GAL4 and several other similar yeast DNA-binding proteins contains a proline residue just N-terminal to the fourth cysteine residue (see Chapter 1, Figure 3). In GAL4, mutation of Pro26 to leucine or serine led to the loss of DNA-binding and this was suppressed by increased concentration of zinc ions in the binding reaction buffer demonstrating that zinc is the key metal ion in the DNA-binding domain. Based on this finding, we also chose to alter Pro23 of MAL63 to leucine. These mutations were created in the cloned *MAL63* gene by site-directed mutagenesis as described in Materials and Methods. In preparation for the DNA-binding studies to be carried out on the altered mal63 proteins (described below), changes in the restriction map of *MAL63* were made. The *Nco*I site within the *MAL63* coding sequence was deleted and an *Nco*I site was created at the translation initiation site. The sequence changes created no amino acid substitutions. This allele is referred to as *MAL63/N*, and a fragment carrying the gene was subcloned into the yeast CEN vector pRS316 to form plasmid pMAL63/N.

The ability of the *MAL63/N* allele to activate *MAL* structural gene transcription was tested by transforming plasmid pMAL63/N into strain 340-2A which contains the *MAL11-2* and *MAL12* structural genes encoding the maltose fermentation enzymes maltose permease and maltase, respectively, but lacks *MAL*-activator gene sequences and

all other *MAL* sequences (Charron and Michels, 1988). Strain 340-2A is a nonfermentor. Table 1 shows that the *MAL63/N* allele complements the structural genes in 340-2A. In this experiment, it was seen that wild type levels of maltase are synthesized, and are comparable to levels produced by the same strain carrying the wild-type *MAL63* gene in the same vector. Thus, the changes in *MAL63/N* have no apparent effect on the ability of *MAL63/N* to activate transcription.

Four point mutations were created by site-directed mutagenesis in the cysteine-rich region of *MAL63/N* encoding the proposed DNA-binding domain as described in Materials and Methods. The alterations in the MAL63 protein are: Cys18Leu; Cys27Ser; Cys34Gly and Pro23Leu. Each mutant allele was subcloned into pRS316 and these plasmids were transformed into strain 340-2A. The activation of maltase expression was tested for each transformant. The results are shown in Table 1. None of the alleles containing an altered cysteine complemented strain 340-2A. None were able to activate maltase inducible expression of the *MAL12* gene of 340-2A. However, transformants containing the Pro23Leu allele showed partially inducible expression of maltase. When two copies of the *mal63*-Pro23Leu allele were inserted into the pRS316 vector, yeast transformed with this plasmid showed increased maltase expression over the single copy plasmid but still less than that of the wild type allele.

These results are in good agreement with the results of growth tests on YP + 2% maltose. Transformants containing the wild type

Table 1: The ability of MAL63 mutant alleles to activate structural gene expression.

CEN plasmids carrying the *MAL63*, *MAL63/N* and the four *mal63* mutant alleles were transformed into strain 340-2A (*MAL11-2*, *MAL12*). Maltase levels of the transformants were determined following growth in induced (3% glycerol, 2% lactate, 2% maltose) and uninduced (3% glycerol, 2% lactate) conditions. Maltase activity is shown as μM of p-nitrophenol released/min/mg of protein.

	Maltose Fermentation	Maltase Activity Induced	Maltase Activity Uninduced
pRS316 Vector only	-	25	17
pRS316-MAL63	++	563	31
pRS316-MAL63/N	++	819	35
pRS316-MAL63/NCys18Leu	-	20	13
pRS316-MAL63/NCys27Ser	-	25	15
pRS316-MAL63/NCys34Gly	-	25	14
pRS316-MAL63/NPro23Leu	+/-	291	22
pRS316-MAL63/NPro23Leu (2 copies)	+	386	26

MAL63 or *MAL63/N* alleles showed gas production in 2 days. No gasproduction was exhibited by the 340-2A transformants containing any of the Cys mutant alleles following 14 days or more. Those transformed with the single copy Pro23Leu allele showed gas production in 5-7 days while those containing the two copies of Pro23Leu allele fermented after only 3 days. Comparable results were obtained by measuring growth on solid YP + 2% maltose media and observing colony size and acid production.

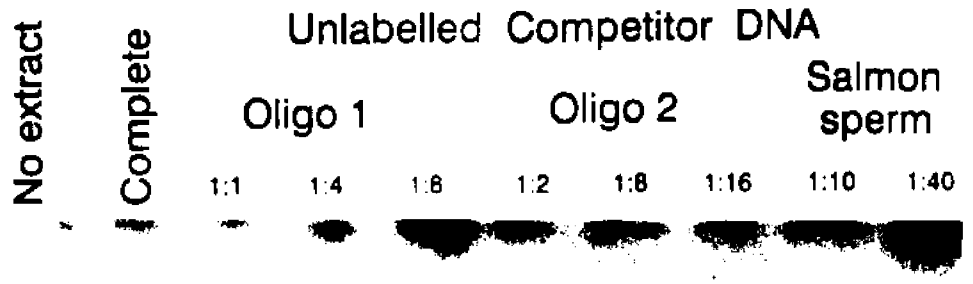
These results clearly indicate that the cysteines at residues 18, 27 and 34 are required for the MAL-activator to be functional whereas the proline at residue 23 is involved in MAL-activation function but is not required.

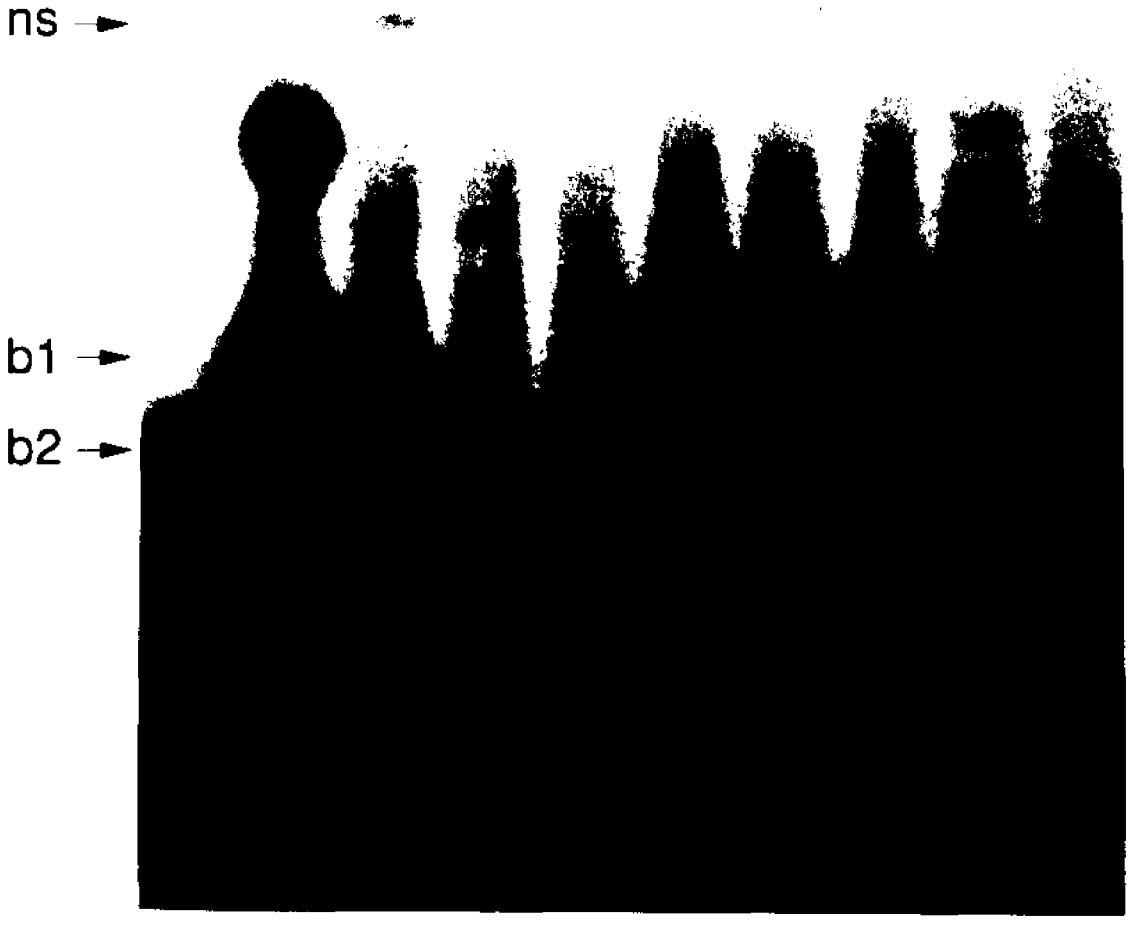
Gel mobility shift DNA-binding assay in *E. coli* extracts. To test the DNA-binding capacity of the wild type MAL63 and altered proteins we carried out a gel mobility shift assay using *E. coli* synthesized proteins. The N-terminal fragment of the MAL-activators containing residues 1-111 (plus approximately 40 other residues encoded by vector sequence) was expressed using the T7 expression system as described in the Materials and Methods (Studier and Moffat, 1986). The binding assay utilized a 40 basepair oligonucleotide, referred to as Oligo 1, derived from the UAS_{MAL} and identified by Levine, Tanouye and Michels (unpublished results) and by Ni and Needleman (1990) as the MAL63 binding site.

Figure 6 illustrates the specificity of the binding assay. Three binding complexes labelled b1, b2 and ns are seen with wild type MAL63 extract. (The intense signal seen just below the ns

Figure 6: Binding specificity of MAL63 to UAS_{MAL}.

The truncated MAL63 protein was expressed in *E. coli* from pET-8c*MAL63/NdXba*I. Oligo 1 is a 40 basepair sequence derived from the UAS_{MAL} of *MAL6* and shown to bind MAL63 proteins (Ni and Needleman, 1990; Levine, Tanouye and Michels, submitted). Oligo 2 is a 30 basepair sequence from a site adjacent to Oligo 1. The specific binding complexes are labeled b1 and b2, and the nonspecific complex is labelled ns. The ratio of labeled to unlabeled DNA in the binding mix is indicated.


No extract
Complete
Unlabelled Competitor DNA
Oligo 1
Oligo 2
Salmon sperm
1:1
1:4
1:8
1:2
1:8
1:16
1:10
1:40


ns
b1
b2

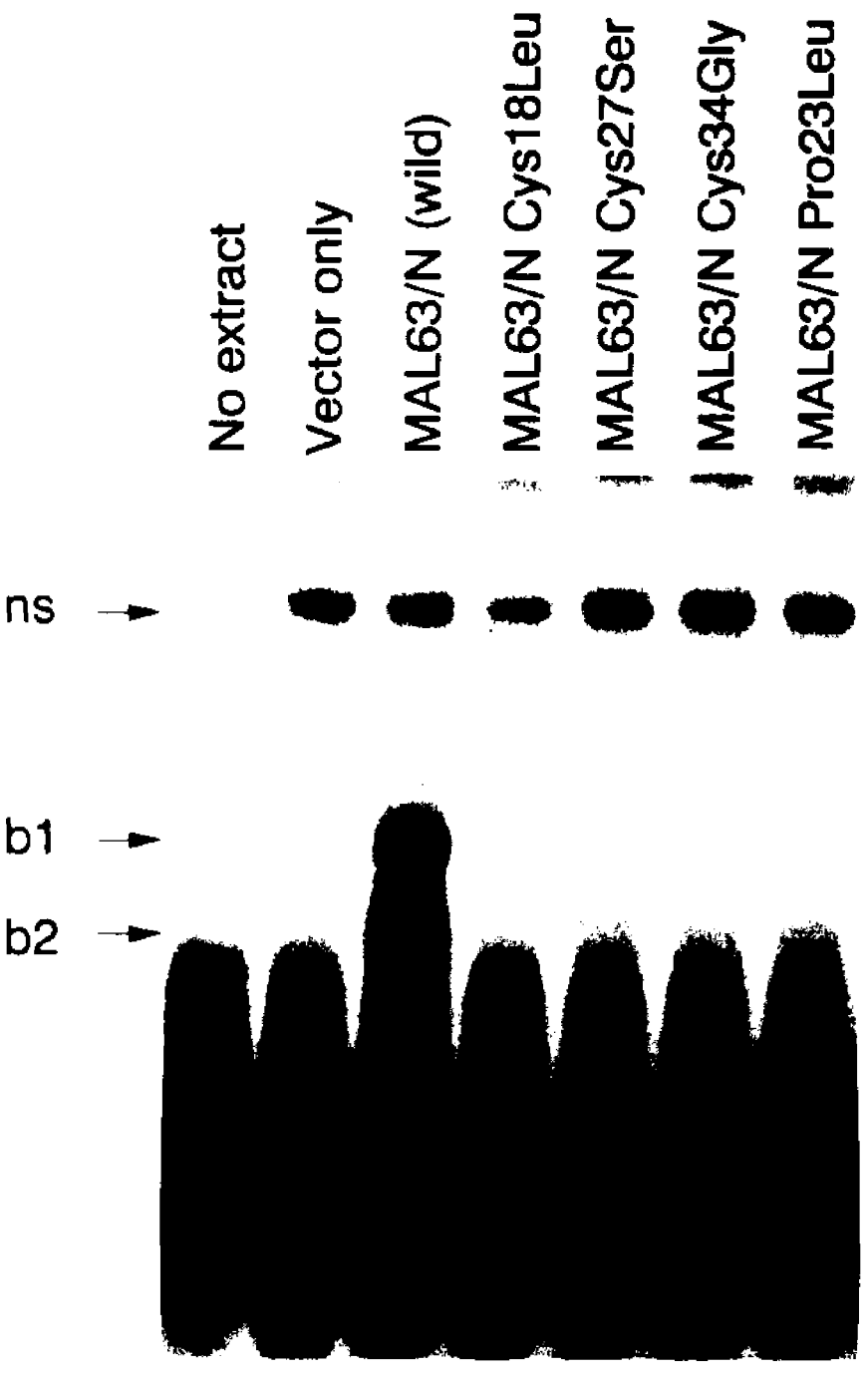
band in the lane containing the Complete mix appears to represent an artifact since it has not been seen in any other experiments.) Bands b1 and b2 appear to represent specific binding to MAL63 based on competition studies shown in Figure 6. The addition of an equal quantity of unlabeled Oligo 1 significantly reduced the intensity of bands b1 and b2 and the addition of a four-fold excess of unlabelled Oligo 1 DNA to the binding mix almost completely eliminated bands b1 and b2. The sequence of Oligo 2 DNA is derived from the region of the *MAL62* promoter immediately adjacent to the Oligo 1 sequence. A sixteen-fold excess of cold Oligo 2 DNA did not effectively compete with Oligo 1 in the formation of complexes b1 or b2. More than forty-fold excess of unlabelled salmon sperm DNA was necessary to demonstrate competition with Oligo 1 to form complexes b1 and b2. The complex labelled ns was easily competed by all of the cold DNAs indicating that it represents nonspecific binding. The intensity of this band acts as a convenient internal control indicating that equal amounts of extract were used in all samples. A third band, an apparently specific binding species, is seen above b1 in samples containing Oligo 2 or salmon sperm DNA. This band was only occasionally seen, particulary when higher DNA or extract concentrations were used.

The mutant *MAL63* alleles were introduced into the pET-8c*MAL63/NΔXba*I construction and expressed in *E. coli*. Extracts were prepared from ^{35}S-methionine pulse labeled cells and these were analyzed by PAGE to demonstrate the level of MAL63 protein present in the extract. The amount of MAL63 protein sythesized was

estimated by Coomassie blue staining and by autoradiography. Equal amounts of extract containing approximately equal amounts of the MAL63 wild-type or mutant proteins were used in the DNA-binding assay. The results of gel mobility shift assays on each of the mutant proteins is shown in Figure 7. None of the mutant mal63 proteins was able to bind UAS_{MAL} DNA. Despite the ability of the Pro23Leu mutation to ferment maltose and activate maltase induction, the *E. coli* expressed Pro23Leu mal63 protein did not show any DNA-binding activity. Thisis probably because the *in vitro* binding conditions are not the best environment for demonstrating binding and any weakening of the DNA-binding specificity *in vivo* will have dramatic effects on the binding *in vitro*. Thus, we conclude that all four of the altered residues are involved in the formation of the DNA-binding domain of MAL63 but that Cys18, Cys27 and Cys34 are critically important.

Figure 7: Gel mobility shift of UAS_{MAL} DNA by MAL63 or mal63 mutant proteins.

Extracts were prepared from *E. coli* transformants carrying the pET-8c*MAL63/NΔXba*I construction. Each of the mutant alleles were individually introduced into this construction. The conditions of the binding reaction are described in Methods and Materials. Approximately equal amounts of protein extract were used in all reaction mixes. The level of mal63 protein expression was determined by pulse labeling with ^{35}S-methionine.


No extract
Vector only
MAL63/N (wild)
MAL63/N Cys18Leu
MAL63/N Cys27Ser
MAL63/N Cys34Gly
MAL63/N Pro23Leu
ns
b1
b2

DISCUSSION

We have identified the cysteine and basic amino acid-rich region at the N-terminal end of the MAL63 protein as the DNA-binding domain of the protein and have demonstrated the importance of the cysteine residues in the binding activity. Four missense mutations were generated in this region of *MAL63* using site-directed mutagenesis: Cys18Leu, Pro23Leu, Cys27Ser and Cys34Gly. The cysteines represent the third, fifth and the sixth cysteines in this C_6 class of DNA-binding domain. All cysteine mutations were nonfermentors and were unable to induce maltase gene expression. On the other hand, the proline mutation was able to ferment maltose and induce maltase gene expression at lower rate. In *in vitro* DNA-binding studies using a gel-mobility shift assay, none of the *E. coli* synthesized mutant proteins were able to bind to the UAS_{MAL}.

Three distinct cysteine-rich DNA-binding motifs have been proposed: zinc finger, zinc twist and zinc cluster (Vallee et al., 1991). The DNA-binding domain of TFIIIA consists of nine repeats having a Cys2 His2 (C_2H_2) structure, each of which forms the zinc finger structural motif to bind to DNA (Miller et al., 1985; for review see Evans and Hollenberg, 1988). This class of protein contains a minimum of two fingers and chelates one zinc atom per finger (Evans and Hollenberg, 1988). The steroid and thyroid receptors contain two apparently unrelated cysteine-rich regions encoded by separate exons (Huckaby et al., 1987) with four (C_4) and five (C_5) conserved cysteines (Evans and Hollenberg, 1988). Glucocorticoid receptors are members of this class of DNA-binding

proteins and studies on these proteins have shown that the structure of the DNA-binding domain forms a zinc twist motif (Vallee et al., 1991). In this motif eight cysteine residues and two zinc atoms form two tetrahedral zinc binding sites. The GAL4 protein had been thought to form a zinc finger involving the two pairs of cysteine residues (Cys2-Cys2) with all four involved in the chelation of a single zinc atom (Johnston and Dover, 1987). More recent studies of the structure of the GAL4 DNA-binding domain using NMR have proposed an alternate structure in which all six cysteine residues of this region interact with two zinc atoms to form a binuclear zinc thiolate cluster (Pan and Coleman, 1990). This structure was proposed based on NMR analysis of the GAL4 protein in which the native bound zinc atom had been replaced by cadmium (Pan and Coleman, 1990). Although 2 moles of cadmium were shown to bind per mole of the GAL4 protein, researchers in several laboratories have not been able to demonstrate that comparable levels of zinc ions are bound (Pan and Coleman, 1989; Lena Basile, personal communication). These results cast some doubt on the validity of the zinc cluster structure (Harrison, 1991). In addition to GAL4, MAL63 and several other yeast transcription activators also contain six, highly conserved, cysteine residues in their proposed DNA-binding domains and thus are also members of the C_6 class of proteins (Evans and Hollenberg, 1988). This class also includes LAC9 (Salmeron and Johnston, 1986), LEU3 (Friden and Schimmel, 1987), PUT3 (Brandriss, 1990), PPRII (Kammerer et al., 1984) and ARGRII (Messenguy et al., 1986). Presumably, the DNA-

binding domains of each of these proteins forms a structure similar to that of the GAL4 DNA-binding domain.

The zinc cluster structure predicts that all six cysteine residues participate in the structure in the region of the DNA-binding domain. If the C_6 class of proteins were to form a zinc finger type of structure the third and sixth cysteine residues might not be important to the formation of a functional structure. Our results indicate that these cysteine residues as well as cysteine number five are critically important to the *in vivo* functional activity of MAL63 and to the *in vitro* DNA-binding capacity of the protein and strongly suggests that all of the cysteines are involved in forming the DNA-binding structure. These results are consistent with the zinc cluster structure but do not exclude other possibilities.

Despite large numbers of missense mutations in the cysteine-basic amino acid-rich region of GAL4, Johnston and Dover (1987; 1988) did not isolate any mutations in Cys21, which corresponds to Cys18 of MAL63, the third cysteine. This result could be interpreted to indicate that either the codon was unreachable by the mutagen or that this cysteine residue is not essential for the functional activity of GAL4. In our experiment, we were able to specifically alter Cys18 to Leu and to test the effect of this change on MAL63 function. We would expect that alteration of Cys21 of GAL4 would have a similar deleterious effect on the activity of GAL4. The NMR study shows this cysteine residue is involved in zinc ion chelation (Pan and Coleman, 1990).

Pro23 of MAL63 does not appear to be critically important to the integrity of the DNA-binding structure. Our *in vivo* results found that transcription activation by MAL63 was only partially affected by alteration of this residue. Alteration of the comparable residue in GAL4, Pro23 had a more severe effect on GAL4 activity but this was suppressible by increased zinc in the growth medium, again, indicating that this residue is not critical to formation of the DNA-binding structure (Johnston, 1987). As can be seen in Figure 3 of Chapter 1, one difference between GAL4 and MAL63 is that in GAL4 a lysine residue is present between the proline residue and cysteine residue four but no intervening residue is present in MAL63. This sequence difference could be responsible for the different functional response of the two proteins to the activation of this residue.

In order to determine the stoichiometry of zinc binding, we provided Lena Basile and Joseph Coleman of Yale University with our MAL63 expression plasmids. Using partially purified MAL63 N-terminal fragment, they have shown by NMR that 1 mole of protein binds 1 to 1.5 moles of zinc ions and that these preparations function *in vitro* in DNA-binding assays (Lena Basile, personal communication). This stoichiometry of zinc binding is similar to that observed for GAL4 suggesting, again, that the structure of DNA-binding domain of both MAL63 and GAL4 is very similar. NMR analysis of MAL63 has not been possible. Purified protein samples have been difficult to obtain because of insolubility problems. Work on both GAL4 and MAL63 proteins will be continued in an effort

to determine the structure of the C_6 class of DNA-binding domain.

BIBLIOGRAPHY

1. Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Mathews, B.W. 1981. Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. Nature 290:754-758.

2. Barnett, J.A. 1976. The utilization of sugars by yeasts. Adv. Carbohydr. Chem. Biochem. 32: 125-234.

3. Bram, R.J. and R.D. Kornberg. 1985. Specific protein binding to far upstream activating sequences in polymerase II promoters. Proc. Natl. Acad. Sci. USA. 82: 43-47.

4. Bram, R.J., Lue, N.F. and Kornberg, R.D. 1986. A *GAL* family of upstream activating sequences in yeast: roles in both induction and repression of transcription. EMBO J. 5: 603-608.

5. Chang, Y.S., Dubin, R.A., Perkins, E., Forrest, D., Michels, C.A. and Needleman, R.B. 1988. *MAL63* codes for a positive regulator of maltose fermentation in *Saccharomyces cerevisiae*. Curr. Genet. 14: 201-209.

6. Charron, M.J., Dubin, R.A. and Michels, C.A. 1986. Structural and functional analysis of the *MAL1* locus of *Saccharomyces cerevisiae*. Mol. Cell. Biol. 6: 3891-3899.

7. Charron, M.J. and Michels C.A. 1987. The constitutive, glucose repression-insensitive mutation of the yeast *MAL4* locus is an alteration of the *MAL43 gene*. Genetics 116: 23-31.

8. Charron, M.J. and Michels, C.A. 1988. The naturally occurring alleles of *MAL1* in *Saccharomyces* species evolved by various mutagenic processes including chromosomal rearrangement. Genetics 120: 83-93.

9. Charron, M.J., Read, E., Haut, S.R., and Michels, C.A. 1989. Molecular evolution of the telomere-associates *MAL1* loci of *Saccharomyces*. Genetics 122: 307-316.

10. Chen, W. and Struhl, K. 1988. Saturation mutagegesis of a yeast *his3* "TATA element": genetic evidence for a specific TATA-binding protein. Proc. Natl. Acad. Sci. USA. 85: 2691-2695.

11. Chodosh, L. A. 1988. Mobility shift DNA-binding assay using gel electrophoresis. in the Current Protocols in Molecular Biology. vol. 2. ed. Ausubel et al. pp. 12.2.1-12.2.10.

12. Cohen, J., Goldenthal, M., Chow, T., Buchferer. B. and Marmur, J. 1985. Organization of the *MAL* loci of

Saccharomyces: physical identification and functional characterization of three genes at the *MAL6* locus. Mol. Gen. Genet. 200: 1-8.

13. Cohen, P. 1988. Protein phosphorylation and hormone action. Proc. Royal soc. London B 234: 115-144.

14. Dubin, R.A., Needleman, R.B., Gossett, D. and Michels, C.A. 1985. Identification of the structural encoding maltase within the *MAL6* locus of *Saccharomyces carlsbergensis*. J. Bacteriol. 164: 605-610.

15. Dubin, R.A., Perkins, E.L., Needleman, R.B. and Michels, C.A. 1986. Identification of a second *trans*-acting gene controlling maltose fermentation in Saccharomyces carlsbergensis. Mol. Cell. Biol. 6: 2757-2765.

16. Dubin, R.A., Charron, M.J., Haut, S.R., Needleman, R.B. and Michels, C.A. 1988. Constitutive expression of the maltose fermentative enzymes in *Saccharomyces carlsbergensis* is dependent upon the mutational activation of a nonessential homolog of *MAL63*. Mol. Cell. Biol. 8: 1027-1035.

17. Evans, R.M. and Hollenberg, S.M. 1988. Zinc fingers: Gilt by association. Cell 52: 1-3.

18. Freedman, L.P., Luisi, B.F., Korszun, Z.R., Basavappa, R., Sigler, P.B. and Yamamoto, K.R. 1988. The function and structure of the metal coordination sites within the glucocorticoid receptor DNA-binding domain. Nature 334: 543-546.

19. Friden, P. and Schimmel, P. 1987. LEU3 of *Saccharomyces cerevisiae* encodes a fator for control of RNA levels of a group of leucine-specific genes. Mol. Cell. Biol. 7: 2708-2717.

20. Gardner, K.H., Pan, T., Narula, S., Rivera, E. and Coleman, J.E. 1991. Structure of the binuclear metal-binding site in the GAL4 transcription factor. Biochemistry 30: 11292-11302.

21. Gill, G. and Ptashne, M. 1987. Mutations of GAL4 protein altered in an activation function. Cell 51: 121-126.

22. Giniger, E. and Ptashne, M. 1987. Transcription in yeast activated by a putative amphipathic alpha helix linked to a DNA binding unit. Nature 330: 670-672.

23. Goldenthal, M.J., Cohen, J.D. and Marmur, J. 1983. Isolation and characterization of a maltose transport mutant in the yeast *Saccharomyces cerevisiae*. Curr. Genet. 7: 195-

199.

24. Harrison, S.C. 1991. A structural taxonomy of DNA-binding domains. Nature 353: 715-719.

25. Henikoff, S. 1984. Unidirectional digestion with exonucleaseIII creates targeted breakpoints for DNA sequencing. Gene 28: 351-359.

26. Hollenberg, S.M. and Evans, R. 1988. Multiple and cooperative trans-activation domains of the human glucocorticoid receptor. Cell 55: 899-906.

27. Hope, I.A and Struhl, K. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell 46: 885-894.

28. Hopper, J.E., Broach, J.R. and Rowe, L.B. 1978. Regulation of the galactose pathway in *Saccharomyces cerevisiae*: induction of uridyl transferase mRNA and dependency on *GAL4* gene function. Proc. Natl. Acad. Sci. USA. 75: 2878-2882.

29. Huckaby, C.S., Conneely, O.M., Beattie, W.G., Dobson, A.D.W., Tsai, M-J. and O'Malley, B.W. 1987. Structure of the chromosomal chicken progesterone receptor gene. Proc. Natl. Acad. Sci. USA. 84: 8380-8384.

30. Johnson, A.D. and Herskowitz, I. 1985. A repressor (MAT alpha 2 product) and its operator control expression of a set of cell type specific genes in yeast. Cell 42: 237-247

31. Johnston, M. 1987. Genetic evidence that zinc is an essential co-factor in the DNA binding domain of GAL4 protein. Nature 328: 353-355.

32. Johnston, M. and Dover, J. 1987. Mutations that inactivate a yeast transcriptional regulatory protein cluster in an evolutionarily conserved DNA binding domain. Proc. Natl. Acad. Sci. USA. 84: 2401-2405.

33. Johnston, M. and Dover, J. 1988. Mutational Analysis of the *GAL4*-Encoded Transcriptional Activator Protein of *Saccharomyces cerevisiae*. Genetics 120: 63-74.

34. Kammerer, B., Guyonvarch, A. and Hubert, J.C. 1984. Yeast regulatory gene PPR1: I. Nucleotide sequence, restriction map and codon usage. J. Mol. Biol. 180: 239-250.

35. Keegan, L., Gill, G. and Ptashne, M. 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science 231: 699-703.

36. Kim, J. and Michels, C.A. 1988. The *MAL63* gene of *Saccharomyces* encodes a cysteine-zinc finger protein. Curr. Genet. 14: 319-323.

37. Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

38. Laughon, A. and Scott, M.P. 1984. Sequence of a Drosophila segmentation gene protein structure homology with DNA-binding proteins. Nature 310: 25-31.

39. Laughon, A. and Gesteland, R.F. 1984. Primary structure of the *Saccharomyces cerevisiae GAL4* gene. Mol. Cell. Biol. 4: 260-267.

40. Lorch, Y. and Kornberg, R.D. 1985. A region flanking the *GAL7* gene and a binding site for GAL4 protein as upstream activating sequences in yeast. J. Mol. Biol. 186: 821-824.

41. Ma, J. and Ptashne, M. 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. Cell 48: 847-853.

42. Marczak, J. and Brandriss, M. 1991. Analysis of constitutive and noninducible mutations of the PUT3 transcriptional activator. Mol. Cell. Biol. 11: 2609-2169.

43. McKay, D.B. and Steitz, T.A. 1981. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. Nature 290: 744-749.

44. Messenguy, F., Dubois, E. and Descamps, F. 1986. Nucleotide sequence of the ARGII regulatory gene and amino acid sequence homologies between ARGII, PPRI and GAL4 regulatory proteins. Eur. J. Biochem. 157: 77-81.

45. Messing, J. 1983. New M13 vectors for cloning. Methods. Enzymol. 101: 20-78.

46. Miller, J., McLachlan, A.D. and Klug, A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. 4: 1609-1614.

47. Needleman, R., Kaback, D.B., Dubin, R.A., Perkins, E.L., Rosenberg, N.G., Sutherland, K.A., Forrest, D.B., and Michels, C.A. 1984. *MAL6* of *Saccharomyces*: A complex genetic locus containing three genes required for maltose fermentation. Proc. Natl. Acad. Sci. U.S.A. 81: 2811-2815.

48. Ni, B. and Needleman, R.B., 1990. Identification of the upstream activating sequence of *MAL* and the binding sites for

the MAL63 activator of *Saccharomyces cerevisiae*. Mol. Cell. Biol. 10: 3797-3800.

49. Pabo, C. O. and Lewis, M. 1982. The operator-binding domain of lambda repressor: Structure and DNA recognition. Nature 298: 443-447.

50. Pan, T. and Coleman, J.E. 1989. Structure and function of the Zn(II) binding site within the DNA-binding domain of the GAL4 transcription factor. Proc. Natl. Acad. Sci. USA. 86: 3145-3149.

51. Pan, T. and Coleman, J.E. 1990. GAL4 transcription factor is not a "zinc finger" but forms a $Zn(II)_2Cys_6$ binuclear cluster. Proc. Natl. Acad. Sci. USA. 87: 2077-2081.

52. Rihs, H-P., Jans, D.A., Fan, H. and Peters, R. 1991. The rate of nuclear cytoplasmic protein transport is determined by the casein kinase II site flanking the nuclear localization sequence of the SV40 T-antigen. EMBO J. 10: 633-639.

53. Rosenberg, A.H., Lade, B.N, Chui, D., Lin, S., Dunn, J.J. and Studier, F.W. 1987. Vectors for selective Expression of cloned DNAs by T7 RNA polymerase. Gene 56: 125-135.

54. Salmeron, J.M. and Johnston, S. 1986. Analysis of the *Kluyveromyes lactis* positive regulatory gene *LAC9* reveals functional homology to, but sequence divergence from, the *Saccharomyces cerevisiae GAL4* gene. Nucleic Acids Res. 14: 7767-7781.

55. Salmeron, J.M., Langdon, S.D. and Johnston, S.A. 1989. Interaction between transcriptional activator protein LAC9 and negative regulatory protein GAL80. Mol. Cell. Biol. 9: 2950-2956.

56. Sanger, F., Nicklen, S. and Coulson, A.R. 1977. Sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74: 5463-5467.

57. Shepherd, J.C.W, Mcginnis, W., Carrasco, A.E., De Robertis, E.M. and Gehring, W.J. 1984. Fly and frog homeo domains show homologies with yeast mating type regulatory proteins. Nature 310: 70-71.

58. Sikorski, R.S. and Hieter, P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics 122: 19-27.

59. Silver, P.A., Keegan, L.P. and Ptashne, M. 1984. Amino

terminus of the yeast *GAL4* gene product is sufficient for nuclear localization. Proc. Natl. Acad. Sci. USA. 81: 5951-5955.

60. Steitz, T.A., Ohlendorf, D.H., McKay, D.B., Anderson, W.F., Mathews, B.W. 1982. Structure similarity in the DNA-binding domain of catabolite gene activator and cro repressor proteins. Proc. Natl. Acad. Sci. USA. 79: 3097-3100.

61. Studier, F.W. and Moffat, B.A. 1986. Use of bacteriophage T7 polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189: 113-130.

62. ten Berge, A.M.A., Zoutewelle, G. and de Poll, K.W. 1973a. Regulation of maltose fermentation in *Saccharomyces carlsbergensis*. I. The function of the gene MAL_6, as recognized by mal_6-Mutants. Mol. Gen. Genet. 123: 233-246.

63. ten Berge, A.M.A., Zoutewelle, G., van de Poll, K.W. and Bloemers, H.P.J. 1973b. Regulation of maltose fermentation in *Saccharomyces carlsbergensis*: II. Properties of a constitutive MAL_6-mutant. Mol. Gen. Genet. 125: 139-146.

64. ten Berge, A.M.A., Zoutewelle, G. and Needleman, R.B. 1974. Regulation of maltose fermentation in *Saccharomyces carlsbergensis*: III. Constitutive mutations at the MAL_6 locus and suppressors changing a constitutive phenotype into a maltose negative phenotype. Mol. Gen. Genet. 131: 113-121.

65. Vallee, B.L., Coleman, J.E. and Auld, D.S. 1991. Zinc fingers, zinc clusters, and zinc twists in DNA-binding protein domains. Proc. Natl. Acad. Sci. USA. 88: 999-1003.

66. Vogt, P.K., Bos, T.J. and Doolittle, R.F. 1987. Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxy-terminal region of a protein coded for by the oncogene *jun*. Proc. Natl. Acad. Sci. UAS. 84: 3316-3319.

67. Wray, L.V. Jr., Witte, M.M., Dickson, R.C. and Riley, M.I. 1987. Characterization of a positive regulatory gene, *LAC9*, that controls induction of the lactose-galactose regulon of *Kluyveromyces lactis*: structural and functional relationships to GAL4 of *Saccharomyces cerevisiae*. Mol. Cell. Biol. 7: 1111-1121.

68. Zaret, K.S. and Sherman, F. 1982. DNA sequence required for efficient transcription termination in yeast. Cell 28: 563-573.