Volume 14 Number 19 1986 Nucleic Acids Research

Analysis of the *Kluyveromyces lactis* positive regulatory gene *LAC9* reveals functional homology to, but sequence divergence from, the *Saccharomyces cerevisiae GAL4* gene

John M.Salmeron, Jr and Stephen A.Johnston

Department of Botany, Duke University, Durham, NC 27706, USA

Received 16 June 1986; Revised and Accepted 3 September 1986

ABSTRACT

The galactose metabolism positive regulatory gene from Kluyveromyces lactis, LAC9, has been isolated through its ability to activate expression of galactose metabolism enzyme genes in Saccharomyces cerevisiae. The LAC9 gene also activates expression of the S. cerevisiae α-galactosidase (MEL1) and K. lactis β-galactosidase (LAC4) genes in S. cerevisiae. Although LAC9-activated gene expression in K. lactis is not glucose repressed, activation of MEL1 gene expression by LAC9 in S. cerevisiae is. The LAC9 gene is expressed at an extremely low level as a ~2.9-kb mRNA, and encodes a protein of 865 amino acids. Although the LAC9 gene is functionally analogous to the S. cerevisiae GAL4 gene, the bulk of its protein sequence shows little homology to that of GAL4. Two of the three regions of homology that do exist, however, are restricted to areas of GAL4 protein already implicated in nuclear localization, DNA binding, and transcriptional activation.

INTRODUCTION

Much information about the regulation of gene expression in prokaryotes has been gained from comparative analysis of regulatory proteins sharing similar functions. Sequence comparisons and deletion analyses of prokaryotic regulatory proteins have yielded information about the structural requirements for regulatory protein functional domains involved in DNA binding (1), sugar binding (2), and interaction with auxiliary transcription factors (3,4). While it can be assumed that similar work will be important in elucidating regulatory mechanisms in eukaryotes, to date the sequences and activities of only a few eukaryotic positive regulatory genes have been determined (5-10).

The best candidates for such comparative studies are systems in which the regulatory proteins are similar, yet somewhat divergent in both sequence and function. Such systems are yeast galactose metabolism regulons (circuits in which multiple, independently transcribed genes are coordinately regulated). The Saccharomyces cerevisiae GAL4 gene, which encodes the galactose regulon positive regulatory element (11-14), has been

functionally characterized (15), sequenced (5), and subjected to preliminary functional domain mapping (16-20). Another organism in which the galactose regulon is well defined genetically is the milk yeast Kluyveromyces lactis. In this yeast, galactose is utilized through the action of the three enzymes encoded by the GAL1, GAL7 and GAL10 genes (21), the expression of which requires activation by the product of the LAC9 positive regulatory gene (22). The LAC9 gene also activates expression of the LAC4 gene (22-24), which encodes β-galactosidase and enables K. lactis to metabolize lactose (25). Unlike the analogous regulon in S. cerevisiae (26), expression of galactose regulon structural genes in K. lactis is not glucose repressed (27). We report here the isolation of the LAC9 regulatory gene from K. lactis through its ability to activate galactose metabolism in a heterologous system, that of S. cerevisiae. We present a functional characterization of the LAC9 gene in S. cérevisiae, and its complete nucleotide sequence.

MATERIALS AND METHODS

Strains

E. coli strain DH1 (28) was used as a host for transformation and propagation of plasmids. For characterization of plasmid activity in S. cerevisiae, strain 21 (gal4 ura3 leu2, ref. 15) or YJ1 (gal4$^{\Delta}$ ura3 leu2 his3) was used. Strain YJ1 is the product of a cross between strains 21 and YM582 (gal4$^{\Delta}$ his3 ura3, a gift of Dr. Mark Johnston). Wild type K. lactis strain Y-1140 was obtained from the United States Department of Agriculture.

Enzymes and biochemicals

Restriction endonucleases and T4 DNA ligase were obtained from Bethesda Research Laboratories or Pharmacia Fine Chemicals. Large fragment of E. coli DNA polymerase I was a gift of Deepak Bastia. Calf alkaline phosphatase was purchased from Boehringer Mannheim. [α-^{32}P]dATP was obtained from ICN or New England Nuclear, and [α-^{35}S]dATP from New England Nuclear or Amersham.

Construction of K. lactis genomic library

Digestion of DNA, ligation, propagation of bacterial strains, and preparation of plasmid DNA were performed as described by Maniatis et al. (29). For construction of the genomic library, K. lactis strain Y-1140 genomic DNA was partially digested with BglII and ligated into the BamHI site of yeast plasmid YEp24 (30). 8,000 independent recombinant plasmids

were generated, with an average insert size of 10.5 kilobases, representing statistically all K. lactis genes of size 3-kb and below with 99% probability (31).

Southern and Northern blot analyses

Plasmid DNA was extracted from yeast cells as described (32). Genomic DNA was isolated from S. cerevisiae or K. lactis cells by the detergent/alkaline lysis method (33). RNAs were purified as described (29). For Southern analysis, restriction endonuclease digestion products were electrophoresed through 0.8% agarose gels and transferred to nitrocellulose filters (34). For Northern analysis, RNA samples were electrophoresed through a 1% agarose, 3.7% formaldehyde gel in MOPS buffer (29) and transferred to nitrocellulose filters as described by Thomas (35). Hybridization to LAC9 DNA (nick-translated or probe-primer labelled to specific activity >1.5 x 10^8 cpm/μg) was carried out in 4xSSC at 65°C, in the presence of 10% dextran sulfate, and the filters were washed in 0.1xSSC, 0.1% SDS at 65°C. Autoradiography was performed with intensifying screens at -70°C for 24 hr, using XAR-5 film (Kodak). Molecular weights were estimated using fragments of λ phage DNA cleaved with restriction enzyme HindIII as standards.

Enzyme assays

Transformed yeast strains were grown in selective media to densities of 1.5 to 4.0x10^7 cells/ml. Between 0.1 and 5.0 ml of cells were harvested and assayed for α-galactosidase (36) or β-galactosidase (37).

DNA sequencing

To sequence the LAC9 gene, the 4.3-kb BglII-HindIII insert from pJ-LAC9 was cloned into M13 bacteriophage (38) and subjected to sequence analysis using the dideoxynucleotide chain termination method (39) as modified for use with [α-^{35}S]dATP (40). Primers included 13- to 17-mers synthesized by an Applied Biosystems Model 380A DNA synthesizer, and the M13 universal primer. DNA and protein sequence analyses were assisted by use of the Beckman Microgenie program.

RESULTS

Cloning of the LAC9 gene

A possible approach to cloning the LAC9 gene would be complementation of a lac9 mutation of K. lactis. However, at the outset of this study lac9 mutants of K. lactis were not available to us. Our alternative approach was to isolate the LAC9 gene by heterologous complementation in S.

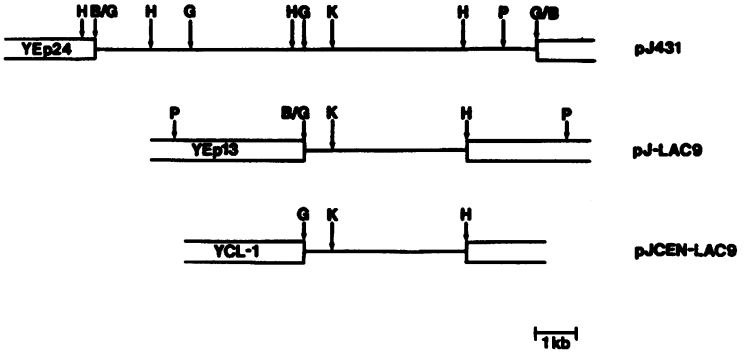

Fig. 1. Structures of yeast shuttle vectors containing K. lactis LAC9 sequence. Open boxes represent parental vector sequences; thin line, K. lactis genomic sequences. Restriction endonuclease site abbreviations: B, BamHI; G, BglII; H, HindIII; K, KpnI; P, PstI.

cerevisiae. The gal4 mutant S. cerevisiae strain 21 was transformed (41) with a library of K. lactis genomic DNA in plasmid YEp24, which contains the URA3 gene as a selectable marker. Approximately 40,000 Ura^+ transformants were obtained, and subsequently selected for ability to grow on galactose medium. Several Gal^+ transformants were identified, and plasmid DNA was extracted from each colony. A single plasmid was then isolated (pJ431, Fig. 1) which conferred the Gal^+ phenotype when used again to transform strain 21. After growth of this transformed strain under non-selective conditions, strictly concomitant loss of the Ura^+ and Gal^+ phenotypes was observed (data not shown), indicating that the complementing activity was borne on the plasmid. The gal4-complementing activity, presumed to be the K. lactis LAC9 gene, was localized to a 4.3-kb BglII-HindIII fragment internal to the genomic insert of the plasmid, and was subcloned into the multicopy vector YEp13 (pJ-LAC9, Fig. 1).

The identity of the LAC9 gene was confirmed by complementation of a lac9 mutant strain of K. lactis (generously provided by Dr. Cor Hollenberg). The 4.3-kb BglII-HindIII fragment containing the LAC9 gene was cloned into the BamHI site of pKARS2 (42), and the resulting plasmid was used to transform K. lactis strain RWJ 15d (lac9 trp1) to Trp prototrophy (42). All Trp^+ transformants exhibited growth on galactose medium and synthesized constitutively β-galactosidase (data not shown), the phenotype expected if the plasmid contained the LAC9 gene.

To confirm the genomic origin of the LAC9 gene, a 5.2-kb BglII-PstI fragment from pJ431 (Fig. 1) was hybridized to BglII-PstI digests of total

Table 1. Disaccharide metabolic activities activated by the presence of GAL4 and LAC9 genes. Growth media carbon source abbreviations: Glu = 2% glucose; Gly = 3% glycerol, 2% lactic acid; GlyGal = 3% glycerol, 2% lactic acid, 2% galactose; Gal = 2% galactose.

A. α-galactosidase (MEL1) activities. The wild-type level of activity of α-galactosidase observed in an S. cerevisiae strain under inducing conditions has been assigned a value of 100.0. Strain 21 is a gal4 ura3 MEL1 S. cerevisiae strain, and 21R is a GAL4 revertant of strain 21. YEpGAL4 (5) is a multicopy plasmid bearing a wild-type GAL4 gene.

			α-GALACTOSIDASE ACTIVITIES			
			growth media			
	strain	regulatory gene plasmid (copy level)	Glu	Gly	GlyGal	Gal
1.	21	none	<0.1	<0.1	<0.1	<0.1
2.	21R	none	<0.1	1.3	-*	100.0
3.	21	YEpGAL4 (multi)	0.3	79.2	-	133.7
4.	21	pJCEN-LAC9 (single)	<0.1	2.2	5.5	N.G.†
5.	21	pJ-LAC9 (multi)	0.4	22.6	31.2	32.3

* not determined

† no growth

B. β-galactosidase (LAC4) activities in glucose medium. The uninduced level of activity of β-galactosidase in the wild-type K. lactis strain Y-1140 has been assigned a value of 100.0. Strain YJ1 is a $gal4^{\Delta}$ ura3 leu2 his3 MEL1 S. cerevisiae strain.

	strain	regulatory gene plasmid	β-GALACTOSIDASE ACTIVITIES
1.	Y-1140 (K. lactis)	none	100.0
2.	YJ1 (S. cerevisiae)	pJ-LAC9	1.7
3.	YJ1[LAC4]	none	5.3
4.	YJ1[LAC4]	pJ-LAC9	192.1

genomic DNA from K. lactis and S. cerevisiae. Although there was no evidence for hybridization to the S. cerevisiae DNA under high-stringency hybridization conditions, a single band of the expected size of 5.2-kb hybridized strongly in the K. lactis DNA (data not shown).

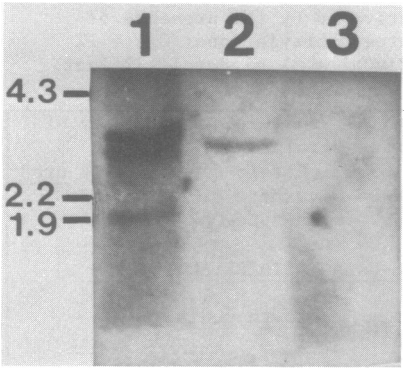

Fig. 2. Hybridization of LAC9 DNA to S. cerevisiae and K. lactis RNAs. Lane contents: lane 1, 10 µg total RNA from S. cerevisiae strain 21 transformed with pJ-LAC9; lane 2, 10 µg K. lactis poly A^+ RNA; lane 3, 10 µg S. cerevisiae poly A^+ RNA. Molecular weight markers are given in kilobase pairs on the left.

The few eukaryotic positive regulatory genes cloned and tested display dosage effects upon the level of expression of the structural genes they regulate (15,16,43-47); i.e., an increase in the copy number of the regulatory gene in the cell results in an increased level of expression of the regulated structural genes. To test whether our cloned LAC9 gene exhibited this characteristic, the 4.3-kb BglII-HindIII fragment containing the LAC9 gene was subcloned into the CEN vector YCL-1 (16; pJCEN-LAC9, Fig. 1), which is maintained at one to two copies per cell (48). The level of galactose regulon structural gene expression in S. cerevisiae declined when the copy number of the LAC9 gene was lowered, as evidenced by the inability of the CEN vector-bearing strain to utilize galactose, and by a drop in the level of MEL1 gene expression activated by LAC9 (Table 1A, see below).

Expression of the LAC9 gene

In order to investigate the expression of the LAC9 gene both in our transformed S. cerevisiae strain and in K. lactis, RNAs from wild-type K. lactis strain Y-1140 and from S. cerevisiae strain 21 transformed with pJ-LAC9 were probed with LAC9 DNA (Fig. 2). In K. lactis, the gene is expressed as a single mRNA approximately 2.9-kb in length (lane 2). The LAC9 mRNA level in K. lactis is roughly 0.0005% of the poly-A^+ RNA (data not shown). In the transformed S. cerevisiae strain (lane 1), two sizes of mRNA are detected, one equal in size to the K. lactis mRNA, and one ~100 nucleotides shorter, as well as a low amount of a much shorter RNA.

Expression of disaccharide metabolism structural genes in S. cerevisiae

In S. cerevisiae, the GAL4 protein positively regulates not only the galactose metabolism structural genes, but also the MEL1 (α-galactosidase) gene, which is responsible for melibiose metabolism (36). To determine whether our cloned K. lactis gene could provide this GAL4 function,

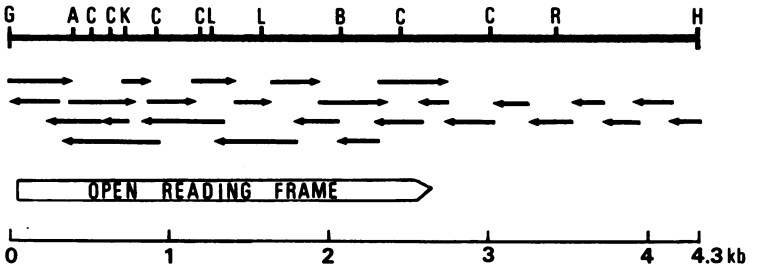

Fig. 3. Restriction map of the 4.3-kb BglII-HindIII fragment containing the LAC9 gene, and strategy for determination of its nucleotide sequence. The arrows indicate the length and direction of the sequenced fragments. Restriction endonuclease site abbreviations: A, AccI; B, BalI; C, ClaI; G, BglII; H, HindIII; K, KpnI; L, BclI; R, EcoRI.

α-galactosidase activity was measured under inducing (Gal or GlyGal) and non-inducing (Gly) conditions in transformed and wild-type S. cerevisiae strains (Table 1A). We find that in a multicopy state, the LAC9 gene can activate expression of MEL1, to an extent equal to approximately one-fourth the level of multi-copy, GAL4-dependent expression (lines 3 and 5, columns Gly and Gal). A single copy of the LAC9 gene is also able to activate a basal level of MEL1 expression under non-inducing conditions (line 4, column Gly). LAC9-activated expression of the MEL1 gene is less sensitive to galactose induction than is GAL4-activated expression (columns Gly, GlyGal and Gal).

In K. lactis, the LAC9 gene product activates expression of the K. lactis β-galactosidase (LAC4) gene, as well as the GAL structural genes (22-25). To determine whether LAC9 can execute this function in a heterologous system, we measured β-galactosidase activity in an S. cerevisiae strain containing an integrated copy of the LAC4 gene (YJ1[LAC4] strain), in the presence of the LAC9 gene. In this strain we detect a β-galactosidase activity level nearly twice that of wild-type K. lactis (Table 1B, lines 1 and 4).

LAC9-activated expression of MEL1 in S. cerevisiae is glucose repressed

In S. cerevisiae, expression of the structural genes of the galactose metabolism regulon (including MEL1) is glucose repressed (26), while glucose does not repress expression of genes regulated by LAC9 in K. lactis (27). To test whether glucose repression circuits are operative in S. cerevisiae when the galactose regulon is under control of the LAC9 gene, we assayed for α-galactosidase in S. cerevisiae cells transformed with the

```
                                                                50
GATCTGAACTTTACTCCGACTGATTGTTTTTACTATACGAA ATG GGT AGT AGG GCC TCC AAT TCG CCT TCT TTT TCA AGT
                                            Met Gly Ser Arg Ala Ser Asn Ser Pro Ser Phe Ser Ser

                                100                                                             150
AAG GCG GAA ACG TTA CTG CCA TCG GAG TAT AAA AAG AAT GCG GTT AAG AAG GAA ACA ATA CGC AAT GGC AAG
Lys Ala Glu Thr Leu Leu Pro Ser Glu Tyr Lys Lys Asn Ala Val Lys Lys Glu Thr Ile Arg Asn Gly Lys

                                                        200
AAA AGG AAA TTG CCT GAT ACA GAA TCC TCA GAT CCT GAG TTT GCA AGT CGG CGT TTG ATA GCT AAT GAA ACT
Lys Arg Lys Leu Pro Asp Thr Glu Ser Ser Asp Pro Glu Phe Ala Ser Arg Arg Leu Ile Ala Asn Glu Thr

                                250
GGC ACT GAT GCG GTG AGT AAT GGT AAC AAA AAT GAT AGC AAT GCC AAC AAC AAC AAC AAC AAC AAC AAC AAG
Gly Thr Asp Ala Val Ser Asn Gly Asn Lys Asn Asp Ser Asn Ala Asn Asn Asn Asn Asn Asn Asn Asn Lys

  300                                                           350
AAA TCA AGT GAA GTA ATG CAC CAG GCG TGC GAT GCT TGC AGG AAG AAG AAG TGG AAA TGT TCC AAG AGA GTA
Lys Ser Ser Glu Val Met His Gln Ala Cys Asp Ala Cys Arg Lys Lys Lys Trp Lys Cys Ser Lys Thr Val

                                        400
CCG ACT TGC ACG AAC TGT CTG AAA TAC AAT TTA GAC TGT GTC TAC TCT CCG CAA GTT GTT AGG ACT CCG TTG
Pro Thr Cys Thr Asn Cys Leu Lys Tyr Asn Leu Asp Cys Val Tyr Ser Pro Gln Val Val Arg Thr Pro Leu

          450                                                                   500
ACA AGA GCA CAT TTA ACA GAG ATG GAA AAT AGG GTT GCA GAG TTG GAA CAG TTT TTG AAA GAA CTT TTC CCA
Thr Arg Ala His Leu Thr Glu Met Glu Asn Arg Val Ala Glu Leu Glu Gln Phe Leu Lys Glu Leu Phe Pro

                                                550
GTT TGG GAT ATC GAT AGG TTA CTT CAG CAA AAA GAT ACA TAC AGG ATT AGG GAA TTG CTT ACT ATG GGT TCT
Val Trp Asp Ile Asp Arg Leu Leu Gln Gln Lys Asp Thr Tyr Arg Ile Arg Glu Leu Leu Thr Met Gly Ser

                    600                                                             650
ACA AAT ACT GTT CCG GGA CTT GCA TCG AAT AAT ATC GAT TCA TCG TTA GAA CAG CCC GTT GCC TTT GGT ACT
Thr Asn Thr Val Pro Gly Leu Ala Ser Asn Asn Ile Asp Ser Ser Leu Glu Gln Pro Val Ala Phe Gly Thr

                                        700
GCG CAG CCG GCA CAA TCT TTG TCA ACT GAT CCA GCA GTA CAA TCT CAA GCC TAT CCA ATG CAA CCG GTA CCG
Ala Gln Pro Ala Gln Ser Leu Ser Thr Asp Pro Ala Val Gln Ser Gln Ala Tyr Pro Met Gln Pro Val Pro

                        750                                                             800
ATG ACA GAG CTT CAA TCT ATC ACC AAT CTT CGA CAC ACG CCA TCA CTT CTG GAT GAA CAG CAA ATG AAC ACG
Met Thr Glu Leu Gln Ser Ile Thr Asn Leu Arg His Thr Pro Ser Leu Leu Asp Glu Gln Gln Met Asn Thr

                                                850
ATT TTC ACG GCA ACG GTC CGG AAC ATG TAC TCT TCA GGT AAC AAT AAT AAC AAC TTG GGT AAC ATC TCT GGT
Ile Ser Thr Ala Thr Leu Arg Asn Met Tyr Ser Ser Gly Asn Asn Asn Asn Asn Leu Gly Asn Ile Ser Gly

                                900
CTA TCA CCT GTT ACA GAG GCA TTC TTC CGT TGG CAG GAA GGT GAA ACG TCA ATC GAT AAT AGT TAT TTT GGA
Leu Ser Pro Val Thr Glu Ala Phe Phe Arg Trp Gln Glu Gly Glu Thr Ser Ile Asp Asn Ser Tyr Phe Gly

    950                                                                 1000
AAA GGT TCA ATT TTG TTT TGG TTT AAC CAA TTA CTA TCA TCA GAA AAG ATC GCT GGC GTT ACA TCA AAA GTA
Lys Gly Ser Ile Leu Phe Trp Leu Asn Gln Leu Leu Ser Ser Glu Lys Ile Ala Gly Val Thr Ser Lys Val

                                        1050
GGC AAT GAC ATT AAC ACT AAT AAT AAT AAT ATA AAC CAT CAG AAG CTA CCT CTA ATA CTA AAC AAT AAT ATT
Gly Asn Asp Ile Asn Thr Asn Asn Asn Asn Ile Asn His Gln Lys Leu Pro Leu Ile Leu Asn Asn Asn Ile

          1100                                                                  1150
ACT CAT AAT GTG TCG GAC ATA ACC ACA ACA AGT ACA TCT TCA AAC AAA AGG GCA ATG TCT CCT CTT TCT GCC
Thr His Asn Val Ser Asp Ile Thr Thr Thr Ser Thr Ser Ser Asn Lys Arg Ala Met Ser Pro Leu Ser Ala

                                                1200
AAT GAC TCT GTA TAT CTC GCT AAA AGA GAG ACA ATA TTC GCG TAT ATC GAT GCG TAC TTC AAG CAC TAT CAT
Asn Asp Ser Val Tyr Leu Ala Lys Arg Glu Thr Ile Ser Ala Tyr Ile Asp Ala Tyr Phe Lys His Tyr His

                    1250                                                            1300
GCG CTA TAT CCG TTG GTC AGT AAG GAA ATG TTT TTC GCT CAC TAT AAT GAT CAA ATT AAA CCA GAG AAC GTT
Ala Leu Tyr Pro Leu Val Ser Lys Glu Met Phe Phe Ala His Tyr Asn Asp Gln Ile Lys Pro Glu Asn Val

                                        1350
GAG ATA TGG CAC ATC TTA CTA AAC GCG GTA TTA GCT TTG GGT TCA TGG TGC TCT AAT TCA TGT TCA AGT CAC
Glu Ile Trp His Ile Leu Leu Asn Ala Val Leu Ala Leu Gly Ser Trp Cys Ser Asn Ser Cys Ser Ser His

                        1400                                                            14
CAT ACT CTC TAT TAC CAA AAC GCA TTA TCA TAT TTG TCC ACC GCT GTA TTG GAA ACA GGG TCC ACA GAT TTA
His Thr Leu Tyr Tyr Gln Asn Ala Leu Ser Tyr Leu Ser Thr Ala Val Leu Glu Thr Gly Ser Thr Asp Leu

50                                                              1500
ACC ATA GCA CTC ATA CTT TTA ACG CAT TAT GTT CAA AAG ATG CAT AAG CCA AAC ACT GCA TGG AGT CTC ATA
Thr Ile Ala Leu Ile Leu Leu Thr His Tyr Val Gln Lys Met His Lys Pro Asn Thr Ala Trp Ser Leu Ile

                                        1550
GGA CTT TGT AGC CAT ATG GCT ACA TCG TTG GGA TTA CAC CGG GAT CTA CCA AAC TCA ACG ATA CAT GAT CAG
Gly Leu Cys Ser His Met Ala Thr Ser Leu Gly Leu His Arg Asp Leu Pro Asn Ser Thr Ile His Asp Gln
```

```
     1600                                                    1650
CAA CTC CGT AGA GTA TTG TGG TGG ACT ATT TAT TGC ACG GGA TGC GAT CTC TCA TTA GAG ACT GGA AGG CCC
Gln Leu Arg Arg Val Leu Trp Trp Thr Ile Tyr Cys Thr Gly Cys Asp Leu Ser Leu Glu Thr Gly Arg Pro

                                        1700
TCA TTA TTG CCC AAT CTT CAG GCT ATT GAT ATA CCA TTA CCA GCT TCA TCT GCC ACT ATC AAA GAA CCA AGC
Ser Leu Leu Pro Asn Leu Gln Ala Ile Asp Ile Pro Leu Pro Ala Ser Ser Ala Thr Ile Lys Glu Pro Ser

              1750                                                    1800
ATA TAT TCC TCC ATC ATA CAA GAA TCC CAA TGG TCT CAA ATA TTG CAA CAG AAA TTG TCA AAT AAC TCA TAT
Ile Tyr Ser Ser Ile Ile Gln Glu Ser Gln Trp Ser Gln Ile Leu Gln Gln Lys Leu Ser Asn Asn Ser Tyr

                                        1850
CAG CAA AGT GCA GGT GAA TGT CTC TCA TGG TTC GAT AGT GTT CAA GCA TTT TTA GAC CAC TGG CCT ACT CCT
Gln Gln Ser Ala Gly Glu Cys Leu Ser Trp Phe Asp Ser Val Gln Ala Phe Leu Asp His Trp Pro Thr Pro

                    1900                                                    1950
AGT ACC GAA GCT GAA CTC AAA GCC TTA AAT GAA ACT CAA CTA GAT TGG CTA CCA TTA GTG AAG TTC CGG CCA
Ser Thr Glu Ala Glu Leu Lys Ala Leu Asn Glu The Gln Leu Asp Trp Leu Pro Leu Val Lys Phe Arg Pro

                                        2000
TAC TGG ATG TTC CAT TGT TCC CTA ATA TCA CTT TTC TCA GTT TTT TTT GAA GAA GAT GCC CCA ACC GAC AAC
Tyr Trp Met Phe His Cys Ser Leu Ile Ser Leu Phe Ser Val Phe Phe Glu Glu Asp Ala Pro Thr Asp Asn

                          2050
AAC GTC ATA CGG TGC AAG GAG TTA TGC CTT CAA CTT TCA AGC AGA AAT ATA TTT AGC GTG GCC ACT TTT GTA
Asn Val Ile Arg Cys Lys Glu Leu Cys Leu Gln Leu Ser Ser Arg Asn Ile Phe Ser Val Ala Thr Phe Val

                                              2150
CGG AGC TAT GCA TTC AAC TCA CTT TCC TGT TGG TAC GCG ACA CAT TAT CTT GTT AGA AGC GCA TTA GTG CCT
Arg Ser Tyr Ala Phe Asn Ser Leu Ser Cys Trp Tyr Ala Thr His Tyr Leu Val Arg Ser Ala Leu Val Pro

                                    2200
CTA CAT TTC GCA TCT CGG ATA TCT CCA CAG CAC GCC TTG TGG GAG ACA GTT AAA GCG CAA TTA TTA TCA GCC
Leu His Phe Ala Ser Arg Ile Ser Pro Gln His Ala Leu Trp Glu Thr Val Lys Ala Gln Leu Leu Ser Ala

          2250                                                    2300
CAT GAA GCG ATG GGT ATA TTG TCA CAA GAA TCT TCC TTG GCC GCT AAA TTT GAT GGG ATA TTA ACC AAG AAT
His Glu Ala Met Gly Ile Leu Ser Gln Glu Ser Ser Leu Ala Ala Lys Phe Asp Gly Ile Leu Thr Lys Asn

                                        2350
TAT TCT GAA ATA CTA CAA AGA GAA GGC ATC AAC AAA AGC CAA CTG ATG CCA CCA CCA ACT CCA TTG CTA CAA
Tyr Ser Glu Ile Leu Gln Arg Glu Gly Ile Asn Lys Ser Gln Leu Met Pro Pro Pro Thr Pro Leu Leu Gln

                    2400                                                    2450
TCA ACC AGT TTC TCG GAC CTA CTT TCA CTG TGG TCA GCA AAC GCA GAA GAC GCT CCG AGA GTC AGT AAT TCC
Ser Thr Ser Phe Ser Asp Leu Leu Ser Leu Trp Ser Ala Asn Ala Glu Asp Ala Pro Arg Val Ser Asn Ser

                                        2500
CAG ATG CCT CAA TCG ATC ACT ATC ACG GAC TCT TTG CTA CAG TCA TCA ACA ACT CAA ATG AGA CCT CCA ACC
Gln Met Pro Gln Ser Ile Thr Ile Thr Asp Ser Leu Leu Gln Ser Ser Thr Thr Gln Met Arg Pro Pro Thr

                          2550                                                    2600
ACA TCT GGA TGG CCT GAT ACC AAC AAC TTC CTG AAT CCA TCG ACC CAA CAG CTA TTC AAC ACC ACA ACA ATG
Thr Ser Gly Trp Pro Asp Thr Asn Asn Phe Leu Asn Pro Ser Thr Gln Gln Leu Phe Asn Thr Thr Thr Met

                                        2650
GAC GAT GTG TAC AAC TAT ATA TTT GAT AAC GAC GAG TAA GAAATCTCTCTTTTCCGTAGTCAATTGGGACAGCATCAATTCA
Asp Asp Val Tyr Asn Tyr Ile Phe Asp Asn Asp Glu End

              2700                                                    2750
TGTATTTACTTTTTGTTCACTAGCTATCAAATAGCTATCCAACGAGACCACTGGTACGAACAGTGTCCATCATGCACA
```

Fig. 4. Nucleotide sequence of the LAC9 gene and predicted amino acid sequence of the encoded protein.

LAC9 gene, grown in glucose medium (Table 1A). We find that MEL1 gene expression is repressed in these cells (lines 4 and 5, columns Glu and Gly).

Nucleotide sequence analysis

The complete nucleotide sequence of the LAC9 gene, determined by the strategy depicted in Fig. 3, is shown in Fig. 4. The LAC9 open reading frame is 2595 nucleotides long and encodes a predicted 865 amino acid

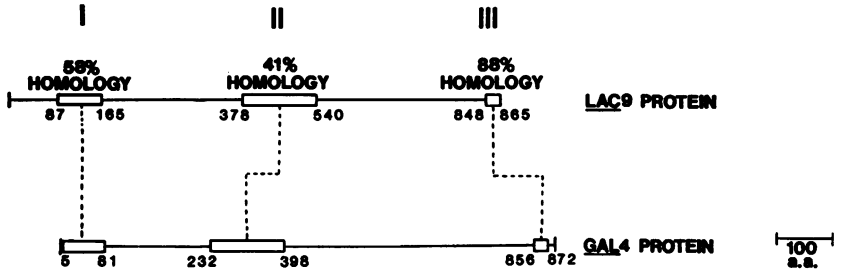

Fig. 5. Regions of homology (open boxes) between LAC9 and GAL4 proteins. Thin lines represent regions possessing no more than 17% homology between the two proteins.

protein with a derived molecular weight of 97,057. Sequence comparison to the GAL4 protein (5) reveals three regions of homology (Fig. 5), which are discussed below.

DISCUSSION

We report here the isolation of the LAC9 positive regulatory gene from the milk yeast K. lactis. The gene was cloned through its ability to complement an S. cerevisiae gal4 mutation, allowing for growth on galactose medium. The gene does not function as efficiently as GAL4 in S. cerevisiae, since multiple copies of the gene are required for complementation. In wild-type S. cerevisiae a single copy of the GAL4 gene is sufficient for growth of cells on galactose medium. We find that in S. cerevisiae, the LAC9 gene also activates expression of two disaccharide metabolism structural genes, the K. lactis β-galactosidase gene LAC4, which is normally regulated by LAC9 in K. lactis, and the S. cerevisiae α-galactosidase gene MEL1. The activation of MEL1 expression was not necessarily expected, since evidence exists that different sets of GAL4 functional domains activate the MEL1 and GAL structural genes (16,18), and since K. lactis itself lacks a MEL1 gene and does not utilize melibiose (49).

In K. lactis, the LAC9 gene is expressed at an extremely low level, a phenomenon noted for other fungal regulatory genes (5,6,50-53). In an S. cerevisiae strain transformed with pJ-LAC9, the gene is transcribed into two sizes of mRNA, as has been found for the GAL4 gene in S. cerevisiae (5). We also find that the increased LAC9 gene copy number in these cells is reflected in an increased level of LAC9 mRNA, relative to the single-copy situation existing in K. lactis.

Activation of MEL1 gene expression by LAC9 in S. cerevisiae is glucose repressed, although LAC9 function in wild-type K. lactis is not affected by glucose (27). There are three possible explanations for this result. First, the presence of glucose may cause a decrease in the level of LAC9 gene expression. However, we detect no significant difference in LAC9 mRNA level between transformed S. cerevisiae cells grown in glucose medium as opposed to glycerol medium (data not shown). Second, a glucose repression mediator that antagonizes both GAL4 and LAC9 proteins may be present in S. cerevisiae but absent in K. lactis. However, the GAL4 gene can activate expression of the LAC4 gene in K. lactis, and this expression becomes subject to glucose repression (M.I. Riley, S.A.J., R. Dickson and J.E. Hopper, unpublished data). Finally, our results may reflect the presence of glucose repression circuits unique to S. cerevisiae that act directly on the regulated structural genes, via mechanisms not involving the regulatory proteins (since the effects persist when GAL4 is replaced by LAC9). Matsumoto et al. (54) have previously isolated S. cerevisiae mutations in genes other than GAL4 that affect glucose repression, and have proposed a tripartite mechanism for glucose repression of the GAL/MEL regulon, including participation by GAL4 and auxillary factors. Mutations in other unlinked loci that relieve glucose repression of several genes, including those of the galactose regulon, have been described previously (55,56). We may be detecting the effects of one of these pleiotropic systems on the expression of the MEL1 gene.

LAC9-activated expression of the MEL1 gene is less responsive to galactose induction than is GAL4-activated expression (Table 1A, columns Gly, GlyGal and Gal). In S. cerevisiae, induction by galactose is believed to result from the inactivation of the GAL80 protein, which blocks GAL4 function under non-inducing conditions (57). Thus, our results may indicate that the interaction between the GAL80 and LAC9 proteins is either weak or non-existent, although alternative explanations certainly exist.

Analysis of the codon usage for the LAC9 protein reveals an almost complete departure from the codon bias observed in moderately to highly expressed genes in S. cerevisiae (58). Such a departure has been found for other yeast genes expressed at low levels (5,6). Although the LAC9 protein has a fairly typical amino acid composition, the distribution of these amino acids is not uniform. The N-terminal quarter of the protein contains two segments (residues 1→55 and 87→178), which are rich in the basic amino acids lysine and arginine (23 and 16 percent, respectively). These are

```
LAC9    87  Ser Ser Glu Val Met His Gln Ala Cys Asp Ala Cys Arg Lys Lys Lys Trp
GAL4     5  Ser Ser     Ile     Glu Gln Ala Cys Asp Ile Cys Arg Leu Lys Lys Leu
PPR1    26  Ser Lys     Ser     Arg Thr Ala Cys Lys Arg Cys Arg Leu Lys Lys Ile

LAC9   104  Lys Cys Ser Lys Thr Val Pro Thr Cys Thr Asn Cys Leu Lys Tyr Asn Leu
GAL4    20  Lys Cys Ser Lys Glu Lys Pro Lys Cys Ala Lys Cys Leu Lys Asn Asn Trp
PPR1    43  Lys Cys Asp Gln Glu Phe Pro Ser Cys Lys Arg Cys Ala Lys Leu Glu Val

                 ←REGION IA|REGION IB→
LAC9   121  Asp Cys Val Tyr Ser Pro Gln Val Val Arg Thr Pro Leu Thr Arg Ala His
GAL4    37  Glu Cys Arg Tyr Ser Pro Lys Thr Lys Arg Ser Pro Leu Thr Arg Ala His
PPR1    60  Pro Cys Val Ser Leu Asp Pro Ala Thr Gly Lys Asp Val Pro Arg Cys Tyr

LAC9   138  Leu Thr Glu Met Glu Asn Arg Val Ala Glu Leu Glu Gln     Phe Leu Lys
GAL4    54  Leu Thr Glu Val Glu Ser Arg Leu Glu Arg Leu Glu Gln Leu Phe Leu Leu
PPR1    77  Val Phe Phe Leu Glu Asp Arg Leu Ala Val Met Met Arg Val     Leu Lys

LAC9   155  Glu Leu Phe Pro Val Trp Asp Ile Asp Arg Leu Leu
GAL4    71     Ile Phe Pro Arg Glu Asp Leu Asp Met Ile Leu
PPR1    93  Glu Tyr Gly Val Asp Pro Thr Lys Ile Arg Gly Asn
```

Fig. 6. Amino acid sequence of homology region I in LAC9, GAL4 and PPR1 proteins. Boxed-in areas indicate exact matches between two or more amino acid sequences.

separated by a 29 amino acid stretch that contains 13 asparagines. Another asparagine-rich segment (22%) occurs between residues 267 and 352.

Comparison of predicted sequences of LAC9 and GAL4 proteins reveals that the regulatory elements have diverged extensively. Three regions of homology exist, together comprising only ~30% of the total protein sequence. The locations of these regions, however, are enlightening when analyzed in the context of what is known about the functional domain map of the GAL4 protein.

Region I (see Fig. 5), near the N-terminus of the LAC9 protein, is homologous to a 77 amino acid segment near the N-terminus of GAL4 protein. This includes the first 76 amino acids of GAL4 protein, which have been shown to be required by GAL4 for nuclear localization (20) and for binding to specific DNA sequences upstream of S. cerevisiae galactose metabolism structural genes (19). These are two functions that must be executed by both LAC9 and GAL4 proteins. Comparison of sequences in region I to those of other yeast regulatory proteins reveals that region I can be divided into two subregions, IA and IB, based on the presence or absence, respectively, of homology to the other protein sequences (Fig. 6). Region

IA contains the highly basic, cysteine-rich motif found in the S. cerevisiae PPR1, ADR1 and Xenopus laevis TFIIIA positive regulatory proteins (8,10), whereas region IB is homologous only between the GAL4 and LAC9 proteins. This suggests that functions required by all these positive regulators, such as those involved in nuclear localization, are executed by sequences within region IA, whereas regulator-specific functions are executed by sequences within region IB. We therefore propose that DNA binding in a certain class of eukaryotic regulatory proteins results from the activities of two functional domains. The first, defined by region IA, is likely to form a small, compact, basic structure that can displace histones and behave as a "chromatin plow" to allow access of the regulatory protein to the DNA. The second, defined by region IB, functions in recognition of regulatory sequences upstream of specific structural genes (e.g. GAL-promoters), allowing for transcriptional activation.

GAL4 deletion studies define a functional domain in the C-terminal 129 amino acids of GAL4 protein required for activation of GAL gene expression, but not for binding to regulatory sequences upstream of these genes (18,19). This functional domain presumably interacts with another protein required for GAL gene transcription (17). Interestingly, homology region III is a short (18 amino acid), but almost completely conserved, region located within this C-terminal area. We suspect that these amino acids may comprise this functional domain, since we detect little homology between the LAC9 and GAL4 proteins directly upstream of this region.

Region II lies near the middle of both LAC9 and GAL4 proteins. The region might play a role in yet another function held in common by the GAL4 and LAC9 proteins, such as monomer oligomerization (16,59), or interaction with a negative regulatory factor (57,60). Additional work is required before any ideas about this region may be formulated.

The lack of homology observed in the remaining areas of the LAC9 and GAL4 protein sequences strengthens the arguments in favor of functional roles for the three regions of homology mentioned above. Further analysis of the LAC9 gene should lead not only to insights into the nature of galactose metabolism regulation in K. lactis, but also to a greater understanding of mechanisms of gene regulation in S. cerevisiae and other eukaryotes.

ACKNOWLEDGEMENTS

We thank Dr. R. Rogers Yocum (Biotechnica, Inc.) for providing us with the K. lactis LAC4 integrating plasmid, Twila Jackson for technical

assistance, and Debbie Gooch for preparing the manuscript. J.M.S. was supported by NIH Genetics Training Grant 5T32-GM07754. This work was supported by an NSF grant DCB-8502626 to S.A.J.

REFERENCES

1. Pabo, C. O. and Sauer, R. T. (1984) Ann. Rev. Bioch., 53, 293-321.
2. Müller-Hill, B. (1983) Nature, 302, 163-164.
3. Buikema, W. J., Szeto, W. W., Lemley, P. V., Orme-Johnson, W. H. and Ausubel, F. M. (1985) Nucl. Acids Res., 13, 4539-4555.
4. Drummond, M., Whitty, P. and Wootton, J. (1986) EMBO J., 5, 441-447.
5. Laughon, A. and Gesteland, R. F. (1982) Proc. Natl. Acad. Sci. U.S.A., 79, 6827-6831.
6. Legrain, M., DeWilde, M. and Hilger, F. (1986) Nucl. Acids Res., 14, 3059-3073.
7. Hinnebusch, A. G. (1984) Proc. Natl. Acad. Sci. U.S.A., 81, 6442-6446.
8. Kammerer, B., Guyonvarch, A. and Hubert, J. C. (1984) J. Mol. Biol., 180, 239-250.
9. Huiet, L. Ph.D. thesis, Univ. of Georgia.
10. Hartshorne, T. A., Blumberg, H. and Young, E. T. (1986) Nature, 320, 283-287.
11. Klar, A. J. S. and Halvorson, H. O. (1974) Mol. Gen. Genet., 125, 203-212.
12. Hopper, J. E. and Rowe, L. B. (1978) J. Biol. Chem., 253, 7566-7569.
13. Hopper, J. E., Broach, J. R. and Rowe, L. B. (1978) Proc. Natl. Acad. Sci. U.S.A., 75, 2878-2882.
14. Matsumoto, K., Toh-e, A. and Oshima, Y. (1978) J. Bact., 134, 446-457.
15. Johnston, S. A. and Hopper, J. E. (1982) Proc. Natl. Acad. Sci. U.S.A., 79, 6971-6975.
16. Johnston, S. A., Zavortink, M. J., Debouck, C. and Hopper, J. E. (1986) Proc. Natl. Acad. Sci. U.S.A., 83, in press.
17. Brent, R. and Ptashne, M. (1985) Cell, 40, 729-735.
18. Dincher, S., Salmeron, J. M. Jr. and Johnston, S. A. Manuscript in preparation.
19. Keegan, L., Gill, G. and Ptashne, M. (1986) Science, 231, 699-704.
20. Silver, P. A., Keegan, L. P. and Ptashne, M. (1984) Proc. Natl. Acad. Sci. U.S.A., 81, 5951-5955.
21. Riley, M. and Dickson, R. (1984) J. Bact., 150, 705-712.
22. Sheetz, R. M. and Dickson, R. C. (1980) Genetics, 95, 877-890.
23. Das, S., Breunig, K. D. and Hollenberg, C. P. (1985) EMBO J., 4, 793-798.
24. Lacy, L. R. and Dickson, R. C. (1981) Mol. Cell. Biol., 1, 628-634.
25. Sheetz, R. M. and Dickson, R. C. (1981) Genetics, 98, 729-745.
26. Adams, B. G. (1972) J. Bact., 111, 308-315.
27. Dickson, R. C. and Markin, J. S. (1980) J. Bact., 142, 777-785.
28. Hanahan, D. (1983) J. Mol. Biol., 166, 577-580.
29. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
30. Botstein, D., Falco, S. C., Stewart, S. E., Brennan, M., Scherer, S., Stinchcomb, D. T., Struhl, K. and Davis, R. W. (1979) Gene, 8, 17-24.
31. Clarke, L. and Carbon, J. (1976) Cell, 9, 91-99.
32. Hirt, B. (1967) J. Mol. Biol., 26, 365-369.
33. Cryer, D. R., Ecclestall, R. and Marmur, J. (1975) Meth. Cell Biol., 12, 39-44.
34. Southern, E. M. (1975) J. Mol. Biol., 98, 503-517.

35. Thomas, P. (1984) Proc. Natl. Acad. Sci. U.S.A., 77, 5201-5205.
36. Kew, O. M. and Douglas, H. C. (1976) J. Bact., 125, 33-41.
37. Miller, J. H. (1974) Experiments in Molecular Genetics, pp. 352-355. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
38. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene, 33, 103-119.
39. Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. U.S.A., 74, 5463-5467.
40. Biggin, M., Gibson, T. and Hong, G. (1983) Proc. Natl. Sci. U.S.A., 80, 3963-3965.
41. Hinnen, A., Hicks, J. B. and Fink, G. R. (1978) Proc. Natl. Acad. Sci. U.S.A., 75, 1929-1933.
42. Das, S. and Hollenberg, C. P. (1982) Curr. Genet., 6, 123-128.
43. Hashimoto, H., Kikuchi, Y., Nogi, Y. and Fukasama, T. (1983) Mol. Gen. Genet., 191, 31-38.
44. Hinnenbusch, A. G. and Fink, G. R. (1983) Proc. Natl. Acad. Sci. U.S.A., 80, 5374-5378.
45. Liljelund, P., Losson, R., Kammerer, B. and Lacroute, F. (1984) J. Mol. Biol., 180, 251-265.
46. Metzenberg, R. L. and Chia, W. (1979) Genetics, 93, 625-643.
47. Denis, C. Personal communication.
48. Bloom, K. S., Fitzgerald-Hayes, M., and Carbon, J. (1983) Cold Spring Harbor Symp. Quant. Biol., 47, 1175-1185.
49. Barnett, J. A., Payne, R. W. and Yarrow, D. (1983) Yeasts: Characteristics and Identification, pp. 344-345 and 467-469, Cambridge University Press, London.
50. Losson, R. and Lacroute, F. (1981) Mol. Gen. Genet., 184, 394-399.
51. Denis, C. L. and Young, E. T. (1983) Mol. Cell. Biol., 3, 360-370.
52. Pinkham, J. L. and Guarente, L. (1985) Mol. Cell. Biol., 5, 3410-3416.
53. Patel, V. B. and Giles, N. H. (1985) Mol. Cell. Biol., 5, 3593-3599.
54. Matsumoto, K., Yoshimatsu, T. and Oshima, Y. (1983) J. Bact., 153, 1405-1414.
55. Michels, C. A. and Romanowski, A. (1980) J. Bact., 143, 674-679.
56. Bailey, R. B. and Woodward, A. (1984) Mol. Gen. Genet., 193, 507-512.
57. Douglas, H. C. and Hawthorne, D. C. (1966) Genetics, 54, 911-916.
58. Bennetzen, J. L. and Hall, B. D. (1982) J. Biol. Chem., 257, 3026-3031.
59. Matsumoto, K., Adachi, Y., Toh-e, A. and Oshima, Y. (1980) J. Bact., 141, 508-527.
60. Dickson, R. C., Sheetz, R. M. and Lacy, L. R. (1981) Mol. Cell Biol., 1, 1048-1056.