Volume 12 Number 16 1984 Nucleic Acids Research

Analysis of full-length cDNA clones carrying *GAL1* of *Saccharomyces cerevisiae:* a model system for cDNA expression

Atsushi Miyajima, Naoki Nakayama, Ikuko Miyajima, Naoko Arai, Hiroto Okayama[1] and Ken-ichi Arai*

DNAX Research Institute of Molecular and Cellular Biology, Palo Alto, CA 94304, and [1]Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, Bethesda, MD 20205, USA

Received 18 June 1984; Accepted 27 July 1984

ABSTRACT

A cDNA cloning vector that allows expression in Saccharomyces cerevisiae has been developed using the plasmid primer approach described by Okayama and Berg [Mol. Cell. Biol. 2:161-170(1982)]. The vector contains ARS1 and TRP1 for plasmid maintenance in yeast and the ADC1 or GAL1 promoter and the TRP5 terminator for expression of the cloned cDNA. Using this system, several recombinants with nearly full-length GAL1 cDNA inserts in a cDNA library made with galactose-induced yeast mRNA were identified. By measurement of galactokinase mRNA and its protein, the expression of GAL1 cDNA was shown to be under the control of the promoter placed upstream of the cDNA insert. Nucleotide sequence analysis revealed that the 3'-ends of the GAL1 cDNA inserts were not unique, indicating that polyA tails were added to GAL1 transcripts at multiple sites in the GAL1 gene. Genetic complementation of appropriate yeast mutants permitted the isolation of clones containing the coding sequences for GAL1, HIS3, and LEU2 from the same cDNA library.

INTRODUCTION

The wealth of genetic and biochemical information about Saccharomyces cerevisiae makes it an attractive organism for the application of recombinant DNA technology to the analysis of gene structure and function. Transformation of yeast is efficient (1-3) with vectors such as the 2 μm plasmid (1) or with plasmids carrying an autonomously replicating sequence (ARS) (2). The transforming genes are maintained either in an autonomously replicating state or as a chromosomal integrated segment (4); when the vector contains a yeast centromere in addition to ARS, the plasmid copy number is stabilized at approximately one (5,6).

Many yeast genes (7), as well as several eukaryotic (8) and prokaryotic (9) genes, have been isolated by genetic complementation of mutant yeast cells; others have been cloned using DNA

© IRL Press Limited, Oxford, England.

probes for nucleic acid hybridization (10). Although cloning of cDNAs in *S. cerevisiae* has usually not been done, it has some obvious advantages; i.e., (i) since cDNAs do not have their own promoters or introns, expression of the cloned cDNAs can be controlled by fusing them to a desired promoter, and (ii) since cDNAs are a copy of mRNA, the genes actively transcribed under a certain condition can be selectively cloned. These advantages may provide a new approach for isolating a cDNA from various sources by its functional expression in yeast.

Okayama and Berg (11) devised a way to achieve the expression of cDNA in mammalian cells. We have adapted this approach and developed a multiple-element shuttle vector for cDNA expression in yeast. Our plasmid vector contains the β-lactamase gene and the ColEl origin of replication to permit selection and propagation of the resulting recombinants in *E. coli*; the yeast *ARS1* and *TRP1* gene achieve the same purpose in yeast. A linker segment, placed 5' to the cDNA segment, supplies the yeast *ADC1* or *GAL1* promoter and the 3'-end of the yeast *TRP5* gene, located 3' to the cDNA segment, provides the signals for termination and polyadenylation of the transcripts (Fig. 1). Basic questions important in developing this new approach in yeast include: (i) Is transcription of the cloned cDNA controlled by the *ADC1* or *GAL1* promoter? (ii) Where is the transcription of cDNA terminated? (iii) Is expression efficient? (iv) Does expressed cDNA complement yeast mutations?

To test this multiple-element expression vector, a cDNA library was prepared with galactose-induced yeast mRNA. *GAL1* cDNA clones were isolated and analysed. Most of the *GAL1* cDNA clones were complete or nearly full-length. The 3' ends of the cDNA inserts were heterogeneous indicating that polyA tails are attached at various sites of the *GAL1* transcripts. The expression of the cDNA insert was regulated by the transcriptional and processing signals that straddle the cloned cDNA. The level of expression of the individual cDNA clones correlated well with their ability to complement a *gal1* mutation. Furthermore, yeast *GAL1*, *HIS3*, and *LEU2* cDNAs were isolated by genetic complementation from the same library. Recently McKnight and McConaughy adopted a similar approach (12) and constructed a cDNA library with yeast mRNA using the *ADC1* promoter and *CYC1* terminator for cDNA expression. They

isolated ADC1, HIS3, URA3, and ASP5 genes by direct complementation of yeast mutants.

MATERIALS AND METHODS

S. cerevisiae, E. coli, and Plasmids.

S. cerevisiae YNN214 (a, Gal^+, ura3-52, ade2-102, lys82-801) and YNN23 (α, gal1Δ107, trp1-289, lys1) were from M. Johnston (Stanford University, Stanford, CA) and SHY3 (a, steCV9, trp1-289, leu2-3, leu2-112, his3-Δ1, ade1-101, can1-100) from the yeast Genetic Stock Center (University of California, Berkeley, CA). E. coli MC1061 (araD139, Δ(ara,leu)7697, ΔlacX74, galU, galK, hsr, strA) was used as a host strain for plasmid preparations. Plasmids YRp7 and pNN78 were from R. Davis (Stanford University, Stanford, CA), pYe(trp1, trp5) and pYe(CEN3)30 from J. Carbon (University of California, Santa Barbara, CA) and pAAH5 from B. Hall (Washington University, Seattle, WA).

Media, enzymes and chemicals.

The synthetic medium contained 0.67% yeast nitrogen base without amino acids, 2% glucose, galactose or ethanol, 0.5% casamino acids, and 50 μg/ml each of adenine and uracil, as indicated. YPGal medium contained 2% Bacto Peptone, 1% yeast extract, and 2% galactose. EBGal medium was YPGal medium with 20 μg/ml ethidium bromide. Restriction endonucleases and DNA polymerase I were from Bethesda Research Laboratories, terminal deoxynucleotidyltransferase and RNase H were from PL Biochemicals and E. coli DNA ligase was a gift from R. Lehman (Stanford University, Stanford, CA). Reverse transcriptase obtained from Life Sciences was purified on Sephacryl S200. Oligo(dT)- and oligo(dA)-cellulose were products of Collaborative Research.

Preparation of $polyA^+$ RNA.

YNN214, grown exponentially in YPGal medium, was harvested, washed with water, resuspended in 4 M guanidine thiocyanate (13), disrupted by shaking with glass beads (Glasperlen, 0.45-0.50 mm) and centrifuged at 12,000 rpm for 10 min to produce a clear supernatant. RNA was sedimented through a 5.7 M CsCl cushion, separated at 28,000 rpm for 24 hr and the $polyA^+$ fraction was obtained by chromatography on oligo(dT)-cellulose.

Construction of the cDNA libraries.

The plasmid primer was prepared from pTRP56 by cleavage with *KpnI* endonuclease follwed by addition of about 70 deoxythymidylic acid residues at each end as described previously (14). The poly(dT)-tailed DNA was digested with *SalI* and the 4.9 kb plasmid fragment was isolated by agarose gel electrophoresis and purified on oligo(dA)-cellulose (14). The promoter-linker fragment was prepared from pA1 and pG1 by cleavage with *PstI* endonuclease and attachments of about 10 and 17 deoxyguanylic acid residues, respectively, at the ends. After digestion with *BamHI* endonuclease, a promoter linker fragment was purified by agarose gel electrophoresis and chromatography on DEAE-cellulose. PolyA$^+$ RNA was annealed to the primer DNA fragment and cDNA was synthesized as described previously (11). After cDNA synthesis, about 10 deoxycytidylic acid residues were added to the end of the cDNA followed by *BamHI* cleavage.

E.coli DNA ligase was used to ligate cDNA and the linker fragment and RNA strand was replaced with DNA by Klenow fragment of DNA polymerase I and RNase H as described previously (11). *E. coli* MC1061 was used as a recipient for transformation.

Galactokinase assay.

Galactokinase activity was measured according to Nogi *et al* (15). Exponentially growing cells were harvested, treated with 40% dimethylsulfoxide (16), and incubated with [^{14}C]galactose. The resultant [^{14}C]galactose-1-phosphate was adsorbed to DE-81 filter paper and measured in a liquid scintilation spectrometer.

Genetic complementation.

S. cerevisiae SHY3 was transformed with the plasmid-cDNA library to generate a pool of Trp$^+$ transformants. About 4 x 10^4 independent transformants were mixed, suspended in water and various dilutions were plated onto synthetic medium lacking either histidine or leucine. YNN23 was the host to detect the complementation of *gal1* and SHY3 was the recipient for complementation of *his3* and *leu2*.

Transformation and other methods.

The transformation protocol with lithium acetate (3) was used to introduce the cDNA library and individual clones into yeast. PolyA$^+$ RNAs of *GAL1* cDNA clones were analysed by blot hybridization

(17). DNA segments were sequenced according to the procedure of Maxam and Gilbert (18) and Sanger et al (19).

RESULTS

The design of a cDNA expression vector for use in yeast.

We have adopted the procedure described by Okayama and Berg (14) for high-efficiency cloning of full-length cDNA to devise a system for expressing cloned cDNA in yeast. To promote expression of a cDNA segment in yeast, the cDNA cloning vector had the following elements: (i) a replication origin that functions in E. coli; (ii) a selective marker for transformation of E. coli; (iii) a yeast replication origin for maintenance of the cDNA as an autonomously replicating plasmid; (iv) a selective marker for yeast transformation; (v) yeast regulatory sequences that can direct the expression of the cDNA segment in yeast; and (vi) several unique restriction sites for molecular cloning.

The starting plasmid was the yeast-E. coli shuttle vector YRp7 that carries TRP1 and ARS1 in a 1.45 kb yeast DNA fragment inserted at a unique EcoR1 site. TRP1 provides an auxotrophic selection marker for yeast transformation, while the ARS1 sequence ensures that the plasmid replicates as an episome. A yeast promoter and terminator for transcription, as well as a polyadenylation sequence, are placed appropriately to provide efficient expression of the cloned cDNA insert. Since the method of constructing the cDNA fixes the orientation of the cDNA segment in the plasmid, the promoter and terminator are placed at the 5' side and 3' side, respectively, of the site occupied by the cloned cDNA segment. YRp7 was converted to the plasmid pTRP56 by inserting a 390 bp fragment carrying the yeast TRP5 terminator (20,21) into the TthIII-1 site, and converting the PvuII restriction site, which is located upstream of that terminator, to a KpnI restriction site. The plasmid that provides the promoter-linker segment was constructed by inserting the yeast alcohol dehydrogenaseI (ADCI) (22) or galactokinase (GAL1) (23) promoter fragment into plasmid pUC8 (24); these promoters contain the regulatory sequence and transcription initiation site, but lacks the initiation codon (Fig. 1).

The pTRP57, the low copy number derivative of pTRP56, was

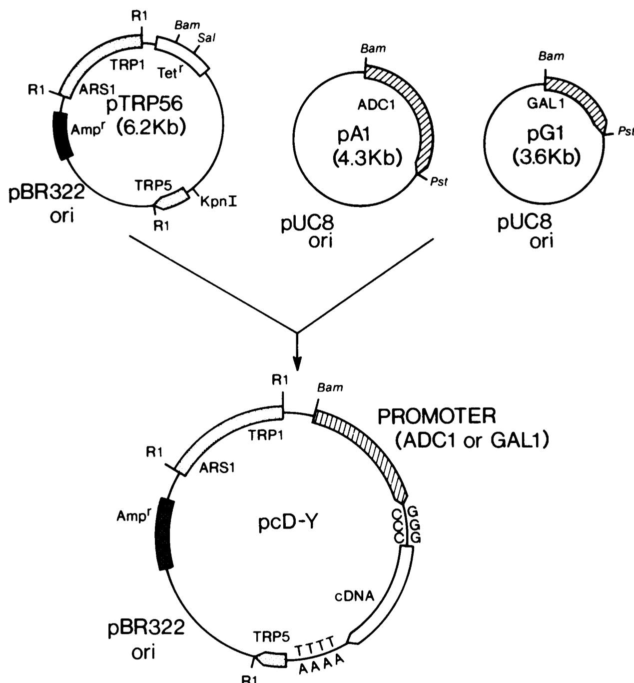

Fig. 1. Structure and component parts of the yeast cDNA-expression vector system. The primer plasmid pTRP56 was constructed by insertion of a 390 bp fragment carrying the yeast TRP5 terminator and polyadenylation signal (21) into the TthIII-1 site of YRp7. A KpnI linker was inserted into the unique PvuII restriction site upstream of the fragment to provide for the poly(dT) tailing. pA1 was constructed by subcloning the promoter region (between nucleotides -1500 to -12 from the ATG codon) of yeast alcohol dehydrogenase I (ADC1) between the BamHI-HincII restriction sites of pUC8. pG1 was constructed by subcloning about 800 nucleotides of the GAL1 promoter region starting from nucleotide 4 upstream of GAL1 ATG initiation codon. The fragment was cloned into the HincII restriction site of pUC8. The unique PstI restriction site of pA1 and pG1 were used for the oligo(dG) tailing. pcD-Y depicts the generalized structure of cDNA recombinants constructed with the plasmid primer and oligo(dG)-tailed linker fragment.

constructed by introducing a fragment, about 600 nucleotides, carrying centromere III (5) into the pTRP56 (see Fig. 4).

Construction of a cDNA library from galactose-induced S. cerevisiae

To test the effectiveness of these plasmid vectors in the Okayama-Berg cDNA cloning protocol, we first constructed a cDNA library with rabbit globin mRNA with pA1 as a linker plasmid.

Approximately 60% of the cDNA clones in the library contained α- or β-globin cDNA inserts of nearly full-length (data not shown). Accordingly, a yeast cDNA library was constructed with polyA$^+$ RNA prepared from galactose-induced cells ("Materials and Methods"). Approximately 1.2 x 10^5 ampicillin resistant E. coli transformants were obtained. Colony hybridization with a cDNA probe prepared from yeast polyA$^+$ RNA showed that at least 40% of the clones contained cDNA inserts; this is a minimum estimate since the intensity of the hybridization signal is dependent on the abundance of each mRNA species and was variable from colony to colony, some colonies being hardly distinguishable from the background (data not shown).

Isolation and characterization of GAL1 cDNA clones.

The yeast enzymes for the Leloir pathway are inducible by galactose (25). Each of the galactose-inducible mRNAs constitutes 0.25% to 1% of the total yeast polyA$^+$ RNA (23). Since the polyA$^+$ RNA used to construct the cDNA library was prepared from Gal$^+$ yeast grown in a medium containing galactose, the library should include galactose induction specific cDNA clones. To test this supposition, the cDNA library was screened by hybridization with a 2 kb EcoRl fragment, Sc4812, which carries most of the GAL1 and a small portion of the GAL10 gene (23). Fifteen clones out of approximately 3 x 10^4 E. coli transformants were positive by colony hybridization, nine of which were characterized further. The restriction enzyme analysis of the plasmids (Fig. 2) and the nucleotide sequence analysis of the junction at the 5' side of the putative cDNA and at the 3' side of the ADC1 promoter (Fig. 3A) showed that most of the plasmids had nearly full-length GAL1 cDNAs. The transcription initiation site of GAL1 has been localized at -71 or -69 nucleotides before the ATG codon (26, and M. Johnston, personal communication). Three clones (pGAL106, pGAL109, pGAL112) had cDNA inserts starting at -69 or -68. Since -69 is a guanine residue and an oligodeoxyguanylic acid linker was used in the present construction, we could not distinguish which nucleotide was the start of the 5' end of these cDNA inserts. Two cDNA clones (pGAL103, pGAL105) started at -58, another two (pGAL110, pGAL111) at -30 and one (pGAL107) at -49. The number of oligodeoxyguanylic acid residues varied from 8 to 24 (Table 1). One plasmid, pGAL108, carries a deletion within the ADCI promoter region; two AccI sites

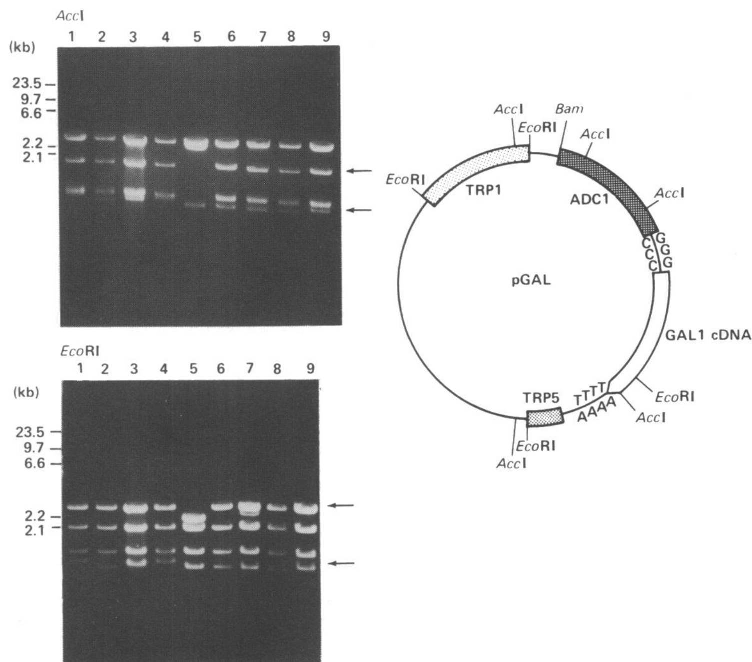

Fig. 2. Restriction enzyme analysis of plasmids carrying GAL1 cDNA segments. Left: GAL1 cDNA plasmids were digested with AccI (upper panel) or EcoRl (lower panel) restriction enzymes and products were separated by agarose gel electrophoresis. Lane 1, pGAL103; lane 2, pGAL105; lane 3, pGAL106; lane 4, pGAL107; lane 5, pGAL108; lane 6, pGAL109; lane 7, pGAL110; lane 8, pGAL111; lane 9, pGAL112. The upper arrow indicates the fragment carrying the 5' side of the GAL1 cDNA and the lower arrow shows the fragment carrying the 3' side of the GAL1 cDNA. Right: A physical map of the putative pGAL plasmids.

within the ADC1 promoter fragment (Fig. 2) were lacking in this plasmid, but restriction analysis showed that it has a nearly full-length cDNA insert similar to the other cDNA clones (Fig. 2, lane 5).

It should be noted that the sizes of the EcoRl and the Accl fragments carrying the 3' end of the cDNA inserts were heterogeneous (Fig. 2). Nucleotide sequence analysis has revealed that the junction between the 3' end of the cDNA inserts and the polyA block is not unique (Fig. 3B). Out of 8 cDNA clones analyzed, four (pGAL105, PGAL109, pGAL110, pGAL111) terminated at 51 nucleotides, two (pGAL106, pGAL112) at 95 nucleotides, pGAL103

TABLE I

Expression of GAL1 cDNA Clones Using the ADC1 Promoter

	cDNA Start Point*	Length of Oligo(dG) Block	Galactokinase Activity*** (units/OD600) Medium: Glucose	Galactose	Ethanol	Growth on EBGal Plate
YNN214 (Gal+)	--	--	0.5	214	0.1	+
YNN23 (Δgal1)						
pTRP56	--	--	0.3	2	1	-
pGAL103	-58	13	55	83	25	+
pGAL105	-58	18	8	7	1	-
pGAL106	-68 (or -69)	8	60	84	25	+
pGAL107	-49	8	87	83	39	+
pGAL108	ND**** (promoter deletion)	ND	1	3	2	-
pGAL109	-68 (or -69)	24	35	42	15	+
pGAL109CEN**	-68 (or -69)	24	7	ND	ND	-
pGAL110	-30	12	43	68	12	+
pGAL111	-30	12	60	66	27	+
pGAL112	-68 (or -69)	8	84	97	11	+
pGAL112CEN**	-68 (or -69)	8	29	ND	ND	-

*Nucleotide residues upstream of the GAL1 ATG initiator codon

**CEN3 was introduced into pGAL109 or pGAL112

***1 unit is that amount of galactokinase catalyzing the conversion of 1 nmol of galactose to galactose-1-phosphate in 10 min at 30°C. The enzyme activity is expressed as units per optical density at 600 nm of the yeast culture.

****Not determined

at 155 nucleotides, and pGAL107 at 160 nucleotides downstream of the translation termination condon, respectively (Fig. 3B). The length of the polyA block was from 49 to 61 (see the legend to Fig. 3B). The consensus sequence (TAG•••TAGT•••TTT or TAG•••TATGT•••TTT) for yeast transcription termination (27) was found between 64 and 107 nucleotides after the termination codon (Fig. 3B). Mammalian polyadenylation signals AATAAA (28) were found at 85 and 134 nucleotides downstream of the termination codon.

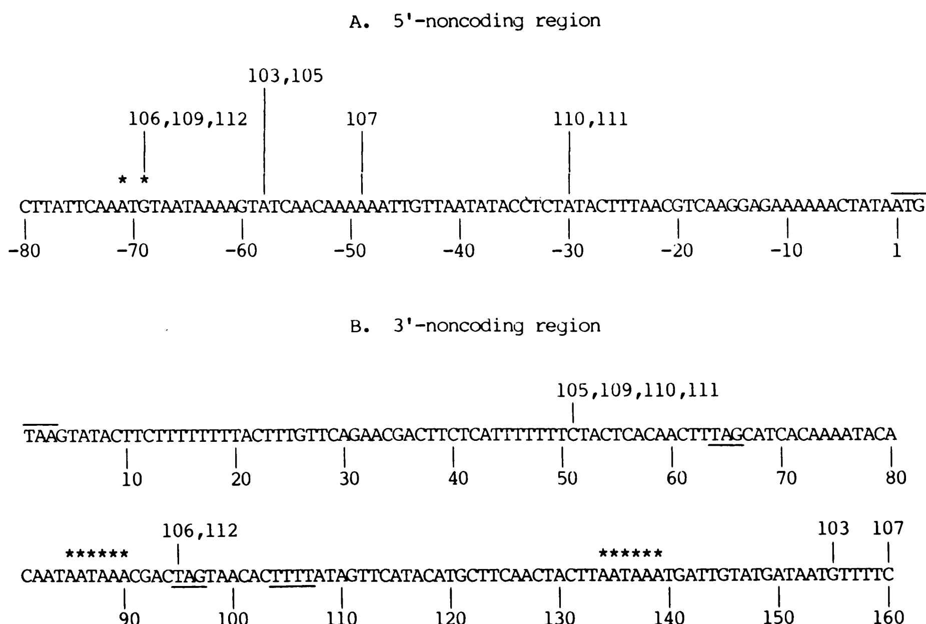

Fig. 3.A. Nucleotide sequence of the 5'-noncoding region of GAL1 and analysis of the junction between the ADC1 promoter and the 5' end of the GAL1 cDNA clones. An AccI restriction site at position -190 from the ATG codon of ADCI was labeled either by T4 polynucleotide kinase or the Klenow fragment of DNA polymerase I and the nucleotide sequence was determined according to Maxam and Gilbert (18). The 5' ends of the cloned cDNA are marked by the vertical lines. The nucleotide sequence of GAL1 and the transcription initiation sites (marked by asterisk) were determined by M. Johnston (personal communication). Initiation codon (ATG) is marked by the overline. B. Nucleotide sequence of the 3' noncoding region and polyA addition sites. The AccI restriction site located near the 3' end of GAL1 (Fig. 2) was labeled with Klenow fragment and the nucleotide sequence was determined according to Maxam and Gilbert (18). The sequence shown in the figure is from the translation termination codon to the polyA addition site of pGAL107. The 3' ends of the cDNA inserts are marked by the vertical lines. The length of the polyA tail is; pGAL103 (60), pGAL105 (57), pGAL106 (58), pGAL107 (49), pGAL109 (61), pGAL110 (58) pGAL111 (58), and pGAL112 (61). Termination codon (TAA) is marked by the overline. The consensus sequence of yeast transcription termination is shown by the underlines and mammalian polyA signals (28) are marked by an asterisk.

Expression of GAL1 cDNA is under control of the ADC1 promoter.

Since the linker fragment contained the yeast ADCI promoter, transcription of the cloned cDNA segment should be mediated and controlled by the ADCI promoter. We examined GAL1 cDNA expression

in three ways: complementation of a gal1 mutant, measurement of galactokinase activity, and characterization of GAL1 mRNA in yeast cells carrying the GAL1 cDNA plasmid. First, YNN23(gal1Δ107, trp1-289, lys1) was transformed with the GAL1 cDNA plasmids, and Trp^+ transformants were selected and streaked on EBGal agar. Because ethidium bromide causes elimination of functional mitochondria (29), only yeast cells carrying a functional GAL1 gene and, therefore, capable of fermenting galactose, can survive on this media. As shown in Table 1, YNN23 carrying pGAL103, pGAL106, pGAL107, pGAL109, pGAL110, pGAL111, and pGAL112 can grow under these conditions, while YNN23 carrying pGAL105 and pGAL108 cannot; YNN23 carrying pGAL109 grows more slowly than the others.

To quantitate the GAL1 cDNA expression, the activity of galactokinase was determined under various growth conditions. The growth rate of yeast carrying the putative GAL1 cDNA clones on EBGal agar correlates well with the amount of galactokinase produced (Table I). This may be related to the length of the 5'-noncoding region of the GAL1 cDNA or to the length of the oligo(dG) stretch, or both. The results suggest that the efficiency of cDNA expression may be affected by the length of the dG:dC segment. For example, pGAL106, pGAL109, and pGAL112 have the same 5'-flanking regions, but pGAL109 has a longer oligo(dG) sequence (Table I). YNN23 carrying pGAL109 grows more slowly in EBGal medium than when carrying pGAL106 or pGAL112 (data not shown). The level of galactokinase in cells containing pGAL109 is also lower than in cells with pGAL106 and pGAL112. Similar differences were observed for cells with pGAL103 and pGAL105 (Table I); however, pGAL105 might carry an additional defect, since pGAL109, which has a longer dG:dC block, can grow slowly.

Since these plasmids have an ARS1 origin and are unstable (5, 6), the differences in galactokinase activity might be due to variations in the plasmid copy number. To stablilize the plasmid and maintain a copy number of approximately one, we introduced the CEN3 sequence (5) into pGAL109 and pGAL112 (Fig. 4). YNN23 carrying pGAL112CEN produces more galactokinase than pGAL109CEN, but neither produces enough to grow on EBGal medium (Table I). This finding is consistent with the suggestion that the differences in expressibility of the GAL1 cDNA may be due to an inhibition by

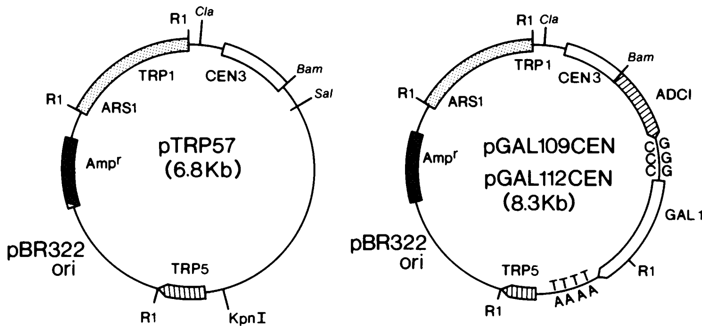

Fig. 4. Structure of the low copy number plasmids. Left. pTRP57 was constructed by introducing the BamHI-ClaI 900 bp fragment carrying 600 nucleotide long centromere III fragment (5) into the BamHI-ClaI site of pTRP56. The centromere fragment was derived from pYe(CEN3)30 (5). Right. The same fragment was inserted into the BamHI-ClaI site of pGAL109 and pGAL112 to obtain pGAL109CEN and pGAL112CEN.

the dG:dC block.

The wild type chromosomal GAL1 gene of YNN214 is induced by galactose and repressed by glucose or ethanol (23) (Table I). In contrast, galactokinase production by the GAL1 cDNA plasmids is only partially repressed if the cells are grown in ethanol and is not appreciably induced in galactose media (Table I). This regulation is characteristic of the ADC1 gene (30). The result indicates that expression of the GAL1 cDNA is regulated by the ADC1 promoter, a suggestion borne out by the analysis of the structure of the GAL1 cDNA transcript.

Analysis of GAL1 cDNA transcripts.

Northern analysis of the GAL1 cDNA transcripts also shows that the transcription of the cloned GAL1 cDNA is controlled by the ADC1 promoter, because the GAL1 mRNA was detected in glucose media (Fig. 5). Furthermore pGAL108, which has a deletion within the ADC1 promoter, does not express the GAL1 cDNA (Fig. 2, lane 5 and Table I). These results indicate that the ADC1 promoter is responsible for the transcription and regulation of the GAL1 cDNA by alcohol (Table I).

Two GAL1 mRNAs, slightly longer than the wild type (YNN214) GAL1 mRNA, were produced in glucose medium in cells harboring the GAL1 cDNA plasmids (Fig. 5A). The lengths of these mRNAs is

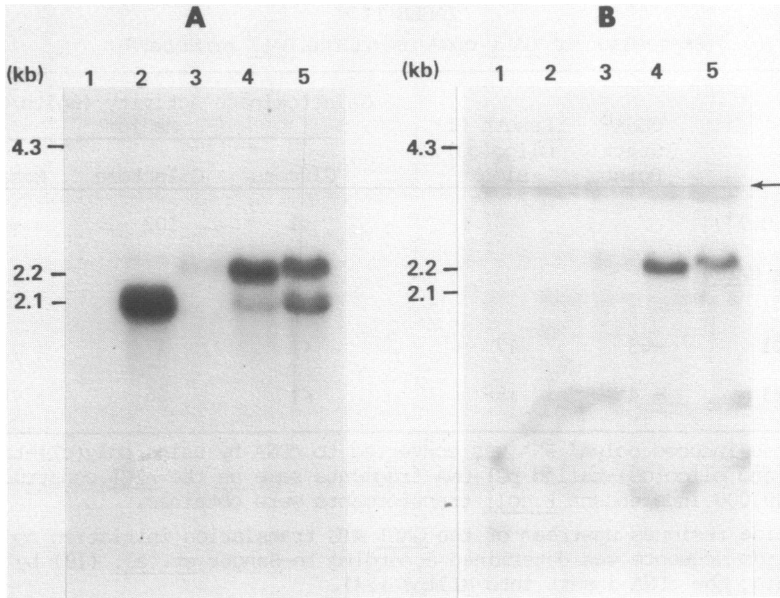

Fig. 5. Northern analysis of GAL1 mRNA. A. PolyA+ RNAs prepared from various yeast cultures were separated by agarose gel electrophoresis ("Materials and Methods") and transferred to nitrocellulose filters. Each filter was hybridized with a nick-translated DNA probe prepared from the 2.0 kb EcoR1 restriction fragment (Sc4812) carrying most of the GAL1 sequence but lacking the 3' end (23, see also Fig. 2). To remove the GAL1 probe from the nitrocellulose filter, the sheet was extensively washed at 68°C with a hybridization solution containing 50% formamide, 5 x SSPE (1 x SSPE: 0.18 M NaCl, 10 mM sodium phosphate pH 7.4, and 1 mM EDTA), 100 μg/ml salmon sperm DNA and 0.5% sodium dodecyl sulfate. B. The sheet was rehybridized with a nick-translated 0.6 kb KpnI-EcoRI fragment carrying the TRP5 terminator prepared from pTRP56 (Fig. 1). The arrow indicates the position of the TRP5 transcript. Lane 1, YNN214(Gal+) grown in glucose medium; lane 2, YNN214(Gal+) grown in galactose medium; lane 3, YNN23(Δgal1) carrying pTRP57 grown in glucose medium. Lane 4, YNN23 carrying pGAL109CEN grown in glucose medium; lane 5, YNN23 carrying pGAL112CEN grown in glucose medium.

consistent with their being transcribed from the ADC1 promoter; an analysis of the RNA obtained from cells carrying the GAL1 cDNA plasmids using blot hybridization with the TRP5 terminator as a hybridization probe revealed that the longer GAL1 mRNA hybridized with this probe (Fig. 5B). This indicates that the GAL1 mRNA produced by these recombinants probably terminates within the cDNA

TABLE II

Expression of GAL1 cDNA using the GAL1 promoter[a]

	cDNA[b] Start Point	Length of Oligo(dG) Block	Galactokinase Activity (units/OD_{600}) Medium		
			Glucose	Galactose	Ethanol
YNN214 (Gal^+)			<1	102	<1
YNN23 (Δ GAL1)					
pTRP56			<1	<1	<1
pGAL121	-63	17	<1	17	1
pGAL123	- 4	15[c]	<1	26	<1

[a]Galactose-induced polyA$^+$ RNA was converted to cDNA by using poly(dT)-tailed pTRP56 and oligo(dG)-tailed pG1 DNA fragments same as the ADC1 construction. About 40,000 independent E.coli transformants were obtained.

[b]Nucleotide residues upstream of the GAL1 ATG translation initiation codon. Nucleotide sequence was determined according to Sanger et. al. (19) by subcloning the cDNA insert into M13mp8 (24).

[c]One dC residue is present within the stretch of oligo (dG).

segment itself and also in the TRP5 segment just beyond.

Construction of cDNA library with the inducible GAL1 promoter.

Since the ADC1 promoter promotes the expression of the cDNA inserts rather constitutively, a promoter whose expression is more strictly regulated, may be desirable for cloning of the genes whose constitutive expression is harmful to the cells. To assess the inducible GAL1 promoter, we constructed another cDNA library from galactose-induced polyA$^+$ RNA by using pG1 as the promoter linker plasmid (Fig 1 and Table 2). Several GAL1 cDNA clones were identified by colony hybridization and two of them were characterized further (Table 2). These two clones had cDNA inserts starting at -63 (pGAL121) and -4 nucleotides (pGAL123) before the initiation codon, respectively. The length of the oligo dG tail was 17 and 15, respectively. The GAL1 cDNA plasmids were introduced into YNN23(Δgal1) and the galactokinase activity was measured in different media (Table 2). The galactokinase activity of YNN23 carrying GAL1 cDNA plasmids was detected only in galactose media similar to wild type cells YNN214 (Gal^+) and the inducibility of the GAL1 expression was high, however, the level of the expression was not as high as the chromosomal GAL1 gene.

Detection of GAL1, LEU2, and HIS3 cDNA clones by direct genetic complementation.

Since GAL1 cDNA is expressed in the cloning vector, we attempted to isolate cDNA clones corresponding to other genes by genetic complementation. The same cDNA library was introduced into yeast SHY3 (trp1, leu2, his3, ade1, ura3) cells to generate about 4 x 10^4 Trp^+ transformants. Screening these transformants for their requirement for histidine or leucine revealed several Leu^+ and His^+ colonies. Gal^+ transformants of YNN23 (trp1, lys1, gal1) were also detected when the Trp^+ transformants were tested for their ability to grow on galactose. The frequency of Gal^+, Leu^+, and His^+ transformants was one in 2 x 10^3, 2 x 10^4, and 2 x 10^4, respectively. The GAL1, LEU2, or HIS3 phenotype cosegregated with TRP1, suggesting that the GAL1, LEU2, or HIS3 cDNA segment is on the same plasmid as TRP1. Recovery of the plasmids in E. coli yielded plasmids that could transform gal1, leu2, or his3 yeast to Gal^+, Leu^+, or His^+ phenotypes, respectively, at high frequency.

DISCUSSION

In this paper, we describe a high-efficiency cDNA cloning vector which permits expression of full-length cDNAs in yeast. The design of the vector is the same, in principle, as that utilized for cDNA expression in mammalian cells (11). Since the orientation of the cloned cDNA segment is fixed by the method, a yeast promoter and polyadenylation signal was arranged to permit expression of the cDNA segment. In the present work, the strong ADC1 promoter was placed upstream and the TRP5 terminator region downstream from the cDNA to ensure efficient initiation and termination of transcription, as well as polyadenylation of the transcripts.

The system was used to construct a cDNA library with mRNA from galactose-induced yeast cells. GAL1 cDNA clones were identified by colony hybridization, and the nucleotide sequence at the junction of the promoter fragment and the inserted cDNA segment was determined. Most of the plasmids had complete or nearly full-length cDNA inserts separated from the ADC1 promoter by a stretch of 8 to 24 dG:dC base pairs. About 0.05% of the E. coli transformants contained clones with full-length GAL1 cDNA. Since

GAL1 mRNA represents between 0.25% to 1% of the yeast polyA$^+$ RNA in galactose-grown cells (23) and about 40% of the cDNA library contains inserts, the efficiency of cloning full-length cDNA is reasonably high. The HIS3 and LEU2 cDNA clones that complemented the his3 and leu2 mutations comprised each about 0.005% of the population. These values are consistent with the abundance of HIS3 mRNA and LEU2 protein which represent about 0.01% and 0.05% of the total yeast polyA$^+$ RNA (31) and protein (32), respectively.

Our data indicate that the GAL1 cDNA sequences cloned behind the ADC1 promoter are expressed from that regulatory sequence. The galactokinase activity was several-fold higher in glucose or galactose media than in ethanol medium, which is typical of the regulation of the ADC1 gene (Table I) . The level of GAL1 cDNA expression reflected in the galactokinase activity was lower than expected from the level of GAL1 cDNA transcripts (Fig. 4), suggesting that the mRNA produced from GAL1 cDNA might be inefficiently translated. Quite possibly, the abnormaly long stretch of G's preceeding the coding sequence (Table I and II) affects transcription and/or translation of GAL1 mRNA since the 5'-noncoding region of yeast mRNA is generally AT rich (28). This same difficulty was experienced by McKnight (personal communication), whose vector placed an oligo(dC) sequence instead of oligo(dG) on the coding strand (12).

Two modifications in the structure of the cDNA cloning vehicle have expanded the use of this approach. The inducible GAL1 (23) has been substituted for the ADC1 promoter in the linker segment. With the GAL1 promoter, the galactokinase activity was detected only in galactose media (Table II). This change permits regulation of the cloned cDNA segments and minimize the possibility that continuous or high level expression of the cDNA is harmful to the host. A centromere sequence reduced the expression of the cDNA insert (Table I). We have also produced a derivative, pTRP57 (Fig. 4), that contains a yeast centromere sequence (CEN3)(5). This plasmid may be useful where a low copy number of the transforming plasmid is required.

The junction between the 3' end of the cDNA insert and the polyA tail was determined by nucleotide sequencing. At least four different junctions were located within the GAL1 gene indicating

that the GAL1 transcripts were polyadenylated at multiple sites. Short GAL cDNA transcripts which were not extended beyond the KpnI site of the cloning vector pTRP56 (Fig. 1 and 5) might be polyadenylated using the signal(s) within the cloned cDNA insert. Even pGAL109, which lacks the consensus sequence for transcription termination located between the translation termination codon and the downstream TRP5 terminator region (Fig 1 and 3), directed the synthesis of the short GAL1 cDNA transcript (Fig. 5).

The ratio of the production of the shorter and longer GAL1 cDNA transcripts (Fig. 5) varied with pGAL plasmids employed. pGAL112, the plasmid which has two polyadenylation sites within the cDNA insert, synthesized a shorter transcript at a much higher level than pGAL109 which has only one polyadenylation site. Thus the number of polyadenylation sites within the cDNA clone may affect the size distribution of GAL1 cDNA transcripts. The cDNA cloning system described above may provide a useful approach to the study of transcription termination and polyadenylation.

AKNOWLEDGEMENT

We thank M. Johnston and R. Davis for communicating results prior to publication, F. Lee, and T. Yokota for helpful discussions, and P. Berg, LeRoy Bertsch, and C. Yanofsky for critical reading of the manuscript. We thank J. Carbon and B. Hall for providing plasmids. N.N. is a recipient of a DNAX Research Institute Graduate Fellowship.

*To whom inquiries should be directed

REFERENCES

1. Beggs, J.D. (1978) Nature, 275; 104-109.
2. Struhl, K., Stinchcomb, D.T., Scherer, S., and Davis, R.W. (1979) Proc. Natl. Acad. Sci. U.S.A., 76; 1035-1039.
3. Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983) J. Bacteriol., 153; 163-168.
4. Schere, S. and Davis, R.W. (1979) Proc. Natl. Acad. Sci. U.S.A., 76; 4951-4955.
5. Fitzgerald-Hayes, M., Clark, L., and Carbon, J. (1982) Cell, 29; 235-244.
6. Stinchcomb, D.T., Mann, C., and Davis, R.W. (1982) J. Mol. Biol., 158; 157-179.
7. Nasmyth, K.A. and Reed, S.I. (1980) Proc. Natl. Acad. Sci. U.S.A., 77; 2119-2123.
8. Henikoff, S., Tatchell, K., Hall, B.D., and K. Nasmyth.

(1981) Nature, 289; 33-37.

9. Storms, R.K., Holouachwck, E.W., and Friesen, J.D. (1981) Mol. Cell. Biol., 1; 836-842.
10. Gallwitz, D. and Seidel, R. (1980) Nucleic Acids Res., 8; 1043-1059.
11. Okayama, H. and Berg, P. (1983) Mol. Cell. Biol., 3; 280-289.
12. McKnight, G.L. and McConaughy, B.C. (1983) Proc. Natl. Acad. Sci. U.S.A., 80; 4412-4416.
13. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979) Biochemistry, 18; 5294-5299.
14. Okayama, H. and Berg, P. (1982) Mol. Cell. Biol., 2; 161-170.
15. Nogi, Y., Matsumoto, K., Toh-e, A., and Oshima, Y. (1977) Mol. Gen. Genet., 152; 137-144.
16. Adams, B.G. (1972) J. Bacteriol., 111; 308-315.
17. Thomas, S.P. (1980) Proc. Natl. Acad. Sci. U.S.A., 77; 5201-5205.
18. Maxam, A. and Gilbert, W. (1980) Methods in Enzymol., 65; 499-560.
19. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A., 74; 5463-5467.
20. Zalkin, H. and Yanofsky, C. (1982) J. Biol. Chem., 257; 1491-1500.
21. Miyajima, A., Miyajima, I., Arai, K., and Arai, N. (1984) Mol. Cell. Biol. 4; 407-414.
22. Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem., 257; 3018-3025.
23. St. John, T.P. and Davis, R.W. (1981) J. Mol. Biol., 152; 285-315.
24. Vieira, J. and Messing, J. (1982) Gene, 19; 259-268.
25. Douglas, H.C. and Hawthorne, D.C. (1964) Genetics, 49; 837-844.
26. Citron, B.A. and Donelson, J.E. (1984) J. Bacteriol., 158; 269-278.
27. Zaret, K.S. and Sherman, F. (1982) Cell, 28; 563-573.
28. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature, 263; 211-214.
29. Sherman, F., Fink, G.R., Lawrence, C.W. (1981) Method in Yeast Genetics Cold Spring Harbor Laboratory, New York.
30. Denis, C.L., Ferguson, J., and Young, E.T. (1983) J. Biol. Chem, 258; 1165-1171.
31. Struhl, K. and Davis, R.W. (1981) J. Mol. Biol., 152; 535-552.
32. Hsu, Y.P. and Kohlhaw, G.B. (1982) J. Biol. Chem., 257; 39-41.