Transcriptional Activation Domains of the *Candida albicans* Gcn4p and Gal4p Homologs†‡

Mikhail Martchenko,¹ Anastasia Levitin,² and Malcolm Whiteway¹,²,*

Department of Biology, McGill University, Stewart Biology Building, 1205 Avenue Docteur Penfield, Montreal, Quebec, Canada H3A 1B1,¹ and Genetics Group, Biotechnology Research Institute, National Research Council, Montreal, Quebec, Canada H4P 2R2²

Received 14 June 2006/Accepted 28 November 2006

Many putative transcription factors in the pathogenic fungus *Candida albicans* contain sequence similarity to well-defined transcriptional regulators in the budding yeast *Saccharomyces cerevisiae*, but this sequence similarity is often limited to the DNA binding domains of the molecules. The Gcn4p and Gal4p proteins of *Saccharomyces cerevisiae* are highly studied and well-understood eukaryotic transcription factors of the basic leucine zipper (Gcn4p) and Cz zinc cluster (Gal4p) families; *C. albicans* has *C. albicans* Gcn4p (CaGcn4p) and CaGal4p with DNA binding domains highly similar to their *S. cerevisiae* counterparts. Deletion analysis of the CaGcn4p protein shows that the N’ terminus is needed for transcriptional activation; an 81-amino-acid region is critical for this function, and this domain can be coupled to a lexA DNA binding module to provide transcription-activating function in a heterologous reporter system. Deletion analysis of the *C. albicans* Gal4p identifies a C-terminal 73-amino-acid-long transcription-activating domain that also can be transferred to a heterologous reporter construct to direct transcriptional activation. These two transcriptional activation regions show no sequence similarity to the respective domains in their *S. cerevisiae* homologs, and the two *C. albicans* transcription-activating domains themselves show little similarity.

Transcriptional regulators control the expression of genes to coordinate the availability of cellular function with the physiological needs of the cell. Gene-specific transcriptional activation is often regulated by the binding of positively acting proteins to upstream activating sequences (UAS) in the DNA where they recruit and control the activities of chromatin-modifying and remodeling complexes and the transcription apparatus (34). A typical transcriptional activator then interacts with the RNA polymerase II complex through binding to an adaptor complex termed Mediator; this Mediator complex consists of about 20 proteins and is conserved from yeasts to humans (5). Eukaryotic transcriptional activator proteins are generally bipartite in nature, with separate domains for DNA binding and transcriptional activation (40, 53). The transcriptional activation domains are classified according to their amino acid composition: rich in acidic residues (e.g., *Saccharomyces cerevisiae* Gal4p and Gcn4p) or basic residues (tobacco BBC1) or rich in glutamine (*S. cerevisiae* Mcm1p), threonine/serine (human OCT2), or isoleucine (NTF1) (2, 7, 10, 15, 26, 30, 37, 38). The DNA binding modules also fall into many classes, such as zinc finger, leucine zipper, and helix-loop-helix motifs (14, 16, 57). Although the activation domain is critical for function and can provide a level of regulation, the functional targets of such transcriptional activators are determined by the DNA binding address of the protein.

In both *S. cerevisiae* and *Candida albicans*, GCN4 encodes a transcriptional activator of amino acid biosynthetic genes that responds to amino acid starvation (25, 54). *S. cerevisiae* Gcn4p (ScGcn4p) is tightly regulated at both the transcriptional and translational levels. The 5’ leader region of *GCN4*, which codes for a transcriptional activator of amino acid biosynthetic genes in response to amino acid starvation, contains four small upstream open reading frames (uORF1 to -4). These uORFs act as negative regulators of translation: the ribosome initiates translation at uORF1 and becomes reactivated for translation at subsequent uORFs. Under environmental stresses, such as amino acid starvation, the translation of *S. cerevisiae GCN4* (ScGCN4) is induced: ScGcn2 kinase phosphorylates eukaryotic initiation factor 2, and the scanning ribosome is not reactivated until it bypasses the uORFs and initiates translation at the *GCN4* open reading frame (24). The unusually long 5’ leader sequence on the *GCN4* mRNA, which carries four upstream open reading frames is conserved in *C. albicans* (54). It was recently shown that the protein kinase Gcn2, which is involved in transcriptional and translational regulation of Gcn4p in *S. cerevisiae*, is not involved in the regulation of *C. albicans* Gcn4p (51).

Gcn4p binds as a homodimer with its basic leucine zipper found in its carboxy terminus to a TGACTC sequence located upstream of many genes induced during amino acid starvation (1, 27). In *S. cerevisiae* Gcn4p, there are two transcription activation domains: one resides in an acidic segment in the center of the protein between residues 107 and 144 (13, 26, 28), and the second Gcn4p activation domain is located in the N-terminal 100 amino acids. The two activation domains are functionally redundant and can work independently to produce high-level activation (13, 28).

In *S. cerevisiae*, Gal4p is a second well-studied transcription factor and functions as the transcriptional activator of galac-
TABLE 1. C. albicans strains used in this study

| Strain                  | Genotype                                                                 |
|-------------------------|--------------------------------------------------------------------------|
| BWPI                   | ura3Δ::immm434/ura3Δ::immm434 his1:::JasG/his1:::JasG arg4:::JasG/arg4:::JasG |
| CAI8                    | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG                       |
| CRC103                  | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pCRlacZ]             |
| CRC106                  | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM10                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM11                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM12                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM13                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM14                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM15                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM16                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM17                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM18                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM19                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM20                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM21                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM22                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM23                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM24                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM25                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM26                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM27                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM28                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM29                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM30                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM31                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM32                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM33                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM34                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM35                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM36                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM37                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM38                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM39                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM40                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM41                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM42                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM43                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM44                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |
| CMM45                   | ura3Δ::immm434/ura3Δ::immm434 ade2:::hisGade2:::hisG [pOPlacZ]             |

The primary role of a transcription activator is to recruit the RNA polymerase II machinery to the promoter to which it is bound. To achieve this, the transcriptional activation domains of both Gcn4p and Gal4p interact with Gal11p, which is a component of the mediator complex that binds the RNA polymerase II machinery (3, 22, 29, 44, 47).

Three general models have been proposed to characterize the structure of the transcriptional activation domain. In the first hypothesis, it was proposed that activation domains are unstructured “acidic blobs” that interact with their targets via ionic interactions. This model is supported by the observation that the removal of the residues of activation domains decreases the activity gradually, rather than abruptly (50). A second model proposes that acidic activation domains form amphipathic α helices, in which acidic residues are aligned on one face of the helix. This model is supported by the observation that an artificial 15-residue peptide, designed to fold into amphipathic α helix, shows transcription-activating abilities when fused to the Gal4 DNA binding domain (38). The authors of the third model argue that the most likely secondary structure is the antiparallel β sheet (55).
We have recently established a detailed annotation of the \textit{C. albicans} genes (6). Intriguingly, although many of the \textit{C. albicans} transcription factors have sequence similarity to transcription factors in \textit{S. cerevisiae}, the similarities occur primarily in the DNA binding motifs of those proteins. In addition, it has been previously shown that although \textit{C. albicans} Rfg1p, Rap1p, Gat1p, Msn2p, and Msn4p have \textit{S. cerevisiae} DNA binding domain homologs, these transcription factors control the regulation of different processes in the two organisms (4, 32, 35, 41). Here we have investigated the transcription activation domains of \textit{C. albicans} homologs of the Gcn4p and Gal4p transcription factors; these domains are serine-threonine rich and lack sequence similarity to the \textit{S. cerevisiae} homologs.

### Materials and Methods

\textit{C. albicans} and bacterial strains. The \textit{C. albicans} strains are listed in Table 1. Strain CaHy8 (18) was used to generate strains CRC103 and CRC106 (49). Strain CRC106 carries the \textit{Staphylococcus aureus} lexA operon; CRC103 does not and serves as a negative control for lexA binding. Strains CRC106 and BW17 (56) were used to define the transcription activation domains of \textit{GCN4} and \textit{GAL4}. The \textit{Escherichia coli} strain MC1061 was used for all plasmid constructions.

**Plasmid construction.** Plasmids and oligonucleotides are shown in Tables 2 and 3, respectively. To create plexA-HIS1, we PCR amplified the \textit{C. albicans} \textit{HIS1} (CaHIS1) open reading frame and its termination sequence from pFA-HIS1 with oligonucleotides OMM46 and OMM47, which contain the lexA binding site, and the CaADH2 TATA box. The PCR product, which contains the lexA binding site, TATA box, \textit{HIS1} open reading frame and termination sequence, was cloned into pFA-ARG4 using the Sall and Sunl restriction sites. Two more lexA binding sites were added by annealing oligonucleotides OMM50 and

| Table 2. Plasmids used in this study |
|-------------------------------------|
| **Plasmid** | **Genes** | **Reference** |
| pFA-HIS1 | pFAdA backbone, CaHIS1 | 20 |
| pFA-ARG4 | pFAdA backbone, CaARG4 | 20 |
| pOPlacZ | CaADE2, lacZ, lexA operon, CaADH1 basal promoter | 49 |
| pOPlacZ-HIS1 | CaADE2, lacZ, lexA operon, CaADH1 basal promoter, CaHIS1 | This study |
| pCRlacZ | CaADE2, lacZ, CaADH1 basal promoter | 49 |
| plexA-HIS1 | pFA-ARG4 backbone, promoterless CaHIS1, lexA operon, ADH2 TATA box | This study |
| Clp-lexA | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4 | 49 |
| Clp-lexA-GCN4 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4 | This study |
| Clp-lexA-GAL4 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGAL4 | This study |
| Clp-lexA-GCN4Δ161-323 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4Δ161-323 | This study |
| Clp-lexA-GCN4Δ247-323 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4Δ247-323 | This study |
| Clp-lexA-GCN4Δ247-323 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4Δ247-323 | This study |
| Clp-lexA-GCN4Δ23-323 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4Δ23-323 | This study |
| Clp-lexA-GCN4Δ82-323 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4Δ82-323 | This study |
| Clp-lexA-GCN4Δ90-323 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4Δ90-323 | This study |
| Clp-lexA-GCN4Δ56-323 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4Δ56-323 | This study |
| Clp-lexA-GCN4Δ42-323 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4Δ42-323 | This study |
| Clp-lexA-GCN4Δ1-81 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4Δ1-81 | This study |
| Clp-lexA-GCN4Δ81Δ247-323 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGCN4Δ1-81Δ247-323 | This study |
| Clp-lexA-GAL4Δ188-261 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGAL4Δ188-261 | This study |
| Clp-lexA-GAL4Δ10-187 | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGAL4Δ10-187 | This study |
| Clp-lexA-GAL4-GCN4AD | CaURA3, CaRP5, promoter of CaACT1, lexA, CaGAL4-GCN4AD | This study |

| Table 3. Oligonucleotides used in this study |
|-----------------------------------------------|
| **Oligonucleotide** (restriction site) | **Sequence** | **Used to create plasmid** |
| OMM46 (Sall) | cagttcgcAAGAATGCCTATTTGACTTTTAAAGG | plexA-HIS1 R |
| OMM47 (Sall) | ctctggacGCAAAATTTTCTGATACACACATACATAATGATTGATTA | plexA-HIS1 F |
| OMM50 (Sall) | gtaaATCAAGAACAATATTCAGAAATAATCCCTGAAACgggggg | plexA-HIS1 R |
| OMM50 (Sall) | gtaaAAGGCGATTTTCTGATTGAGATTTGGCAATTTTCTGATACACACATACATAATGATTGATTA | plexA-HIS1 F |
| OMM56 (MluI) | GTAATTGTTTTGTTACCCGAATG | Clp-lexA-GCN4Δ5 F |
| OMM57 (PstI) | CAACGTAGCGTCGACGTGC | Clp-lexA-GCN4Δ1-81 R |
| OMM58 (PstI) | aatctgcatatATCAGTCTTCGAGAAAGAGCA | Clp-lexA-GCN4Δ161-323 R |
| OMM59 (PstI) | aatctgcatatatATCAGTCTTCGAGAAAGAGCA | Clp-lexA-GCN4Δ123-323 R |
| OMM60 (PstI) | aatctgcatatatATCAGTCTTCGAGAAAGAGCA | Clp-lexA-GCN4Δ82-323 R |
| OMM61 (PstI) | aatctgcatatatATCAGTCTTCGAGAAAGAGCA | Clp-lexA-GCN4Δ82-323 R |
| OMM62 (PstI) | aatctgcatatatATCAGTCTTCGAGAAAGAGCA | Clp-lexA-GCN4Δ82-323 R |
| OMM63 (PstI) | aatctgcatatatATCAGTCTTCGAGAAAGAGCA | Clp-lexA-GCN4Δ82-323 R |
| OMM71 (PstI) | ccacgctaatATCAGTCTTCGAGAAAGAGCA | Clp-lexA-GCN4Δ82-323 R |
| OMM66 (MluI) | ccacgctaatATCAGTCTTCGAGAAAGAGCA | Clp-lexA-GAL4 F |
| OMM67 (PstI) | aatctgcatatatATCAGTCTTCGAGAAAGAGCA | Clp-lexA-GAL4 R |
| OMM239 (BamHII) | ttgaacgatATCTGCCTGCATCTCCCTATTTTTTTATGAAGA | Clp-lexA-GAL4-FC4AD |
| OMM240 (ZraI) | ttgaacgatATCTGCCTGCATCTCCCTATTTTTTTATGAAGA | Clp-lexA-GAL4-FC4AD |

\[ a \] Lowercase letters indicate the sequences added for the restriction sites.

\[ b \] F, forward; R, reverse.
OMM125 and OMM57 were used to PCR amplify HIS1 out from pFA-open reading frame, its promoter, and termination sequence were cut into AatII- and PstI-cut CIp- the genome of C. gcn4 the locus. Strains CMM86, CMM87, CMM88, and CMM89 were constructed in CIp-lexA-GAL4-GCN4AD. CaHIS1 open reading frame, its promoter, and termination sequence were cut out from pFA-HIS1 using Notl and ligated into Notl-cut pOPlacZ to create pOPlacZ-HIS1. The CMM85 strain (gen4 with pOPPlac-Z-HIS1) was created by first converting gen4 (42) into an URA3 auxotroph (gen4-ura3) using 5'-fluoro-orotic acid-containing media, followed by an integration of XcmI-cut pOP and pGO to integrate them at the HPS1 locus. All of the constructs created in this study were integrated into the genome of C. albicans; all of the constructs, except for pOPPlacZ and plexA-HIS1, were digested with Stul to integrate them at the RPS1 locus, pOPPlacZ was digested with BamHI and plexA-HIS1 was digested with AgeI to integrate them at ADE2 and ARG4 loci, respectively. All of the DNA constructs were transformed into S. cerevisiae with lithium acetate (8).

β-Galactosidase assays. The expression level of the lacZ gene was assayed in two ways: by β-galactosidase overlay assay using independently isolated transformants grown on solid yeast extract-peptone-dextrose (YPD) medium, yeast extract-peptone-galactose (YPGal) medium, or synthetic complete medium with amino acids (SC-aa), or by β-galactosidase assays performed on mid-exponential shaking flask YPD, YPGal, or SC-aa-grown liquid cultures (48). The β-galactosidase activity was expressed in Miller units; the values are shown as means and standard deviations from three independent transformants.

RESULTS

Definition of the C. albicans Gcn4p (CaGcn4p) activation domain. We have recently performed a detailed genome annotation of the C. albicans genome (6), which showed that frequently the transcription factors of this organism share homology to transcription factors of other organisms only within the DNA binding domain. We have defined 198 S. cerevisiae genes whose products contain a DNA binding domain and are classified as transcription factors by combining the list of transcriptional regulators of Harbison et al. (23) with the list from http://www.yeastact.com/tflist.php. Of these, 32 were experimentally shown to be transcriptional repressors in S. cerevisiae. Ninety-nine of the remaining 166 S. cerevisiae transcriptional activators were found to have C. albicans homologs, half of which share homology only within a DNA binding domain (Fig. 1). A detailed assessment of global and transcriptional activation domain similarities is provided (see Table S1 in the supplemental material). Since there is no primary sequence that defines the activation domain as a module, the nature of the activation domain is based on the experimentally defined part of the transcription factor. A majority of S. cerevisiae transcription factors, such as Gal4p, Gcn4p, Upe2p, Leu3p, and Arg81p (11, 13, 26, 28, 38, 46, 58), were experimentally shown to have acidic activation domains. When we compared the transcription factors of S. cerevisiae with the transcription factors of C. albicans, we observed that in some cases the sequence of the experimentally defined activation domains of S. cerevisiae is well conserved in C. albicans transcription factor homologs, such as Upc2, Leu3, and Arg81 (11, 46, 58). In other cases, the sequence of the experimentally defined activation domain of S. cerevisiae is not detectable in C. albicans transcription factor homologs, such as Gal4p and Gcn4p. The presence of homology in the DNA binding domain of the C. albicans transcription factors like Gcn4p and Gal4p tells us that these might be transcriptional regulators, but the absence of homology in the activation domain makes it difficult for us to predict whether these could work as activators or repressors. We therefore directly investigated the functions of the regions outside the DNA binding motif.

We have investigated the roles of the nonhomologous regions of candidate transcription factors to establish if they play a role in transcriptional activation. Within the DNA binding domain, CaGcn4p shares strong sequence similarity (88%) with S. cerevisiae Gcn4p. To establish whether this protein possesses a functional activation domain, we made use of the S. aureus lexA one-hybrid system (49). This system contains the C. albicans actin promoter (pACT1) placed upstream of the S. aureus lexA open reading frame to create plasmid CIP-lexA, and the S. aureus lexA operator upstream of both an ADH1 basal promoter and a lacZ open reading frame, creating pOPlacZ (49); in addition, we placed the S. aureus lexA operator upstream of the HIS1 open reading frame to create plexA-HIS1. We integrated pOPlacZ into strain CAI8 to yield reporter strain CRC106 (49) and integrated plexA-HIS1 into strain BW17 to yield reporter strain CMM25. Fusions were constructed in CIP-lexA and introduced into these two reporter strains. In the absence of any transcriptional activator fused to
lexA, the reporter CRC106 derivative yielded basal levels of β-galactosidase, and the reporter CMM25 derivative produced no growth in the absence of histidine; when a transactivator is fused to lexA, the system yielded higher levels of β-galactosidase and permitted growth in the absence of histidine. A full-length CaGCN4 cloned downstream of the lexA open reading frame, creating Clp-lexA-GCN4, was fully capable of transactivation in the C. albicans assays. This construct generated fivefold-higher β-galactosidase activity when transformed into CRC106 to generate CMM14 and permitted growth in 1 day in the absence of histidine when transformed into CMM25 to create CMM30, compared to the appropriate controls CMM10 and CMM26, which contain Clp-lexA (Fig. 2). These results suggest that C. albicans GCN4 contains a transcription activation domain (49).

The C-terminal systematic deletions of lexA-GCN4 identified the N-terminal 81-amino-acid region serving as an activation domain; Clp-lexA-GCN4Δ247-323, Clp-lexA-GCN4Δ161-323, Clp-lexA-GCN4Δ23-323, and Clp-lexA-GCN4Δ18-23 were as active as full-length Clp-lexA-GCN4 in both the lacZ background and HIS1 background. Further C-terminal deletions gradually reduced both lacZ and HIS1 activities, suggesting that the Gcn4p activation domain is at least 81 amino acids long (see Clp-lexA-GCN4Δ249-323, Clp-lexA-GCN4Δ56-323, and Clp-lexA-GCN4Δ2-323 in Fig. 3). The deletion of the proposed N-terminal activation domain in the context of the full-length GCN4 (Clp-lexA-GCN4Δ1-81) showed the same β-galactosidase activity in strain CMM21 as that of full-length GCN4 in CMM14, while Clp-lexA-GCN4Δ1-81 in CMM37 showed a slightly reduced HIS1 activity compared to that of CMM30. This observation could be explained either by a second transcription activation domain (as in S. cerevisiae Gcn4p) located between the N-terminal activation domain and C-terminal DNA binding domain or by the Clp-lexA-GCN4Δ1-81 interaction with the endogenous wild-type Gcn4p through a dimerization domain located at the C’ terminus of the protein (27), generating an activating heterodimer. We directly tested the capacity of the region between the DNA binding domain and the N-terminal activating domain to allow transcriptional activation by creating Clp-lexA-GCN4Δ1-81Δ247-323, in which both the activation and DNA binding domains were deleted; this construct generated background β-galactosidase activity and no HIS1 activity (strains CMM64 and CMM65), suggesting that either the activation domain at the N’ terminus is the only CaGcn4p activation domain, that Clp-lexA-GCN4Δ1-81Δ247-323 yields an unstable protein, or that the Gcn4p DNA binding domain directs transcriptional activation. To distinguish between these hypotheses, we tested lexA-GCN4Δ1-81, which lacks the N-terminal activation domain, for its ability to activate the expression of lacZ in the absence of endogenous Gcn4p (gcn4 strain). Although lexA-GCN4Δ1-81 resulted in high expression of lacZ in the wild-type strain (CMM21), the lacZ expression was dropped down to background levels in the absence of endogenous GCN4 (CMM88). At the same time, lexA-GCN4Δ1-81 showed high lacZ expression in both wild-type and gcn4 backgrounds (CMM14 and CMM87), while lexA and lexA-GCN4Δ1-81Δ247-323 showed low lacZ levels of expression in both GCN4 and gcn4 backgrounds (CMM10, CMM86, CMM64, and CMM89). These results suggest that the activation domain at the N’ terminus of C. albicans Gcn4p is the only CaGcn4p activation domain. At the amino acid level, this activation domain is nucleophilic (has a composition of 20% serine-threonine) and shares no similarity with the activation domain of ScGcn4.

Definition of the CaGal4p-homologous activation domain. We examined the activation domain of a candidate C. albicans version of the S. cerevisiae Gal4p protein. In S. cerevisiae, Gal4p is a highly studied transcription factor, and the structures of its DNA binding and transcriptional activation modules as well as the target promoters have been extensively investigated (36, 45, 52). Within the DNA binding domain, the putative Gal4p protein, encoded by C. albicans ORF19.5338, shares strong sequence similarity (86%) with S. cerevisiae Gal4p; the DNA binding domain has the six cysteine residues, the linker region, and the dimerization region all well conserved. A BLAST search of the ScGal4p sequence in the C. albicans genome yields Orf19.5338 as its closest homolog; at the same time, searching the Orf19.5338 sequence in the S. cerevisiae genome yields ScGal4p as its closest homolog. Since ScGal4p and Orf19.5338 form a “reciprocal best hit” relationship, we named Orf19.5338 CaGal4p. Although this C. albicans Gal4p homolog binds 5′-CGGN11CCG-3′, the upstream activating sequence (UASc) to which Gal4p binds in S. cerevisiae (data not shown), the promoters of C. albicans GAL genes lack UASc. Rather, UASc are found upstream of C. albicans sub-
FIG. 3. Identification of CaGcn4p transcriptional activation domain. The results of structural (right) and functional (left) analyses of lexA-GCN4 derivatives are shown. (Right) The structures of various proteins derived from lexA-GCN4 are indicated by gray and black bars, with the N- and C-terminal residues defined as in the wild-type proteins. DBD, DNA binding domain. (Left) The results of liquid β-galactosidase assays expressed in Miller units are shown in the β-gal1 column. The expression of the HIS1 reporter gene is shown in the HIS12 column as follows: strong (S) (it took 1 day for a strain to grow in the absence of histidine), weak (W) (it took 2 days for a strain to grow in the absence of histidine), or no growth (N).
telomeric and glycolysis genes (data not shown). In addition, this *C. albicans* gene encodes a much smaller 261-amino-acid-long protein compared to the *S. cerevisiae* Gal4p of 881 amino acids. The regions outside the DNA binding domain of those two proteins share no similarity, and the negatively charged region that serves as the interaction domain for ScGal80p is missing in CaGal4p. Interestingly enough, although *C. albicans* can grow on galactose (19), its genome also lacks a Gal80p homolog.

To establish whether CaGal4p contains a transcriptional activation domain, full-length GAL4 was cloned downstream of the *lexA* open reading frame to create plasmid CIp-*lexA-GAL4*. We transformed this construct into strain CRC106 to create CMM11; the transactivating ability of CIp-*lexA-GAL4* was five times more than the activity of the control vector CIp-*lexA*, which was transformed into CRC106 to create strain CMM10 (Fig. 2). Similarly, strain CMM25, which contains the *lexA* operator in front of *HIS1*, was transformed with CIp-*lexA* and CIp-*lexA-GAL4* to create CMM26 and CMM27, respectively. As was found for the *lacZ* system, *lexA-GAL4* showed high transcription-activating ability compared to the activity of the vector alone, since only CMM27 grew in the absence of histidine. These results suggest that *C. albicans* Gal4p can act as a transcriptional activator.

We examined which part of CaGal4p was essential for the transactivating capacity. Deletion of the C-terminal 71 amino acids, creating CIp-*lexA-GAL4Δ188-261*, abolished the transactivating ability of *lexA-GAL4* when introduced into both strain CRC106 to create CMM13 and into strain CMM25 to create CMM29. In addition, fusion of *lexA* to the nucleotides encoding the C-terminal 71 amino acids of CaGal4p showed transcription-activating abilities similar to that of the full-length GAL4 (see CMM60 and CMM61) (Fig. 4). Similarly to the Gcn4p activation domain, the CaGal4p activation domain showed a nucleophilic nature; it has a 30% serine-threonine composition but shares no other similarity with the activation domain of ScGcn4p.

It is currently believed that the gene specificity of the transcription factor comes from its DNA binding domain: this domain binds to a nucleotide motif on the promoters and recruits the RNA polymerase II machinery (34). To see whether the transcriptional activation domain plays a role in the transcriptional selectivity of CaGal4p, we replaced the 71-amino-acid-long CaGal4p activation domain (Gal4AD) with the 81-amino-acid-long Gcn4p activation domain (Gcn4AD). We observed that in vivo the *lexA-GAL4-GCN4AD* construct in strains CMM12 and CMM28 showed the same transcription-activating ability as the *lexA-GAL4-GAL4AD* in strains CMM11 and CMM27, respectively (Fig. 4).

**DISCUSSION**

Eukaryotic transcription factors are typically bipartite in nature, with a region (the DNA binding or DB domain) specifically designed to interact with a defined DNA sequence and a region (the transcriptional activation or TA domain) required to interface the factor with the transcriptional machinery. There are several classes of each of these modules, and they are connected together in a variety of ways. Within the transcriptional activation modules, there are domains rich in acidic
or basic residues or rich in glutamine, threonine/serine, or isoleucine residues (2, 7, 10, 15, 26, 30, 37, 38). In this study we defined the transcription activation domains in a pair of C. albicans transcription factors that share sequence similarity with their S. cerevisiae homologs only within their DNA binding domains. The Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA binding domain is one of the largest classes of fungal transcription factors that share sequence similarity either to each other or to the activation domains of Gcn4p and Gal4p. However, the C. albicans homologs only within their DNA binding domains. The Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA binding domain is one of the largest classes of fungal transcription factors that share sequence similarity either to each other or to the activation domains of Gcn4p and Gal4p. Gcn4p and Gal4p proteins each contain two transcriptional activation domains that are positioned at the N’ and C’ termini, respectively (13, 26, 28, 37, 38). We analyzed the transcription activation domains of the C. albicans Gen4p (CaGNC4) and Gal4p (CaGAL4) homologs and found that just as in S. cerevisiae, they are positioned at the N’ and C’ termini of the respective proteins. However, the C. albicans Gen4p and Gal4p activation domains do not share sequence similarity either to each other or to the activation domains of their S. cerevisiae homologs, and C. albicans Gal4p and Gen4p have nucleophilic activation domains. Nucleophilic transcriptional activation regions have been previously seen almost exclusively in higher eukaryotic transcription factors (9, 10, 21). A screen for C. albicans transcriptional activation domains using a genomic library fused downstream of lexA yielded an active fragment containing a normally noncoding region that expressed 33% serines and threonines in the fusion construct (data not shown), which also suggests that nucleophilicity can be an important feature of C. albicans activation domains. The serine and threonine amino acids could potentially be converted into an acidic form by phosphorylation.

The C. albicans Gal4p and Gen4p proteins each contain two transcriptional activation domains (13, 27, 28, 37, 38). In contrast, the C. albicans Gen4p and Gal4p homologs appear to each contain only one transcriptional activation domain (Fig. 3 and 4). Each of the two ScGcn4p activation domains seems to be composed of two or more small subdomains that have additive effects on transcription and that can cooperate in different combinations to promote high-level expression of the Gcn4p-dependent genes (13, 28). These results are consistent with our observation that the C- to N-terminal deletions within the CuGcn4p activation domain lead to a gradual, rather than to an abrupt, reduction of the transcription-activating abilities of the fusion protein (Fig. 3). To determine when the changes in the activation domains of Gcn4p and Gal4p occurred during the evolution of the yeast species, we used available genomic data of the ascomycota (Schizosaccharomyces pombe, Neurospora crassa, Aspergillus niger, S. cerevisiae, S. paradoxus, S. mikatae, S. kudriavzevii, S. bayanus, Candida glabrata, S. castelli, Kluyveromyces lactis, Ashbya gossypii, Debaryomyces Hansenii, Candida tropicalis, Candida dubliniensis, and C. albicans) (Fig. 5A and B). Archiascomycetes were observed to lack the activation domains of either ScGcn4p or CaGcn4p. Euascomycetes possessed the ScGcn4p activation domain II (ADI). We noted that the common ancestor of D. hansenii, C. tropicalis, C. dubliniensis, and C. albicans lost activation domain I (ADI) of ScGcn4p and acquired the activation domain of CaGcn4p. We also observed that the ancestor of C. tropicalis, C. dubliniensis, and C. albicans lacked ScGcn4p ADII (Fig. 5A). In addition, the ancestor of C. tropicalis, C. dubliniensis, and C. albicans lacked ScGal4p ADI, while the ancestor of D. hansenii, C. tropicalis, C. dubliniensis, and C. albicans lacked ScGal4p ADII (Fig. 5B). In both cases, D. hansenii represents an intermediate with both S. cerevisiae and C. albicans activation domains. These observations show that the changes in the activation domains of Gal4p and Gcn4p of C. albicans occurred relatively recently on the evolutionary scale.

The lack of homology in the activation domains of transcriptional activators between S. cerevisiae and C. albicans might suggest a concomitant reduced structural similarity in the activation domain-interacting complexes between the two species. A pairwise sequence comparison of the transcriptional machinery between C. albicans and S. cerevisiae shows a high level of conservation in the RNA polymerase II complex. The exceptions for this are transcription factor IIA and the Mediator complex: S. cerevisiae and C. albicans show low levels of homology with respect to the proteins of those two complexes, and these are the complexes that interact with transcriptional regulators.

The characterization of the bipartite structure of eukaryotic transcription factors like S. cerevisiae Gal4p was a fundamental conceptual advance (33) and has led to important technical developments like the yeast two-hybrid system (17). In general, C. albicans transcription factors follow the pattern of distinct DNA binding and transcriptional activation domains, and many show strong sequence similarity, extending to both domains, to specific S. cerevisiae transcription factors. However, a large number of C. albicans proteins have strong sequence similarity that is limited only to the DNA binding module of an S. cerevisiae transcription regulator. We have shown that al-
though the well-studied Gal4p and Gcn4p proteins of *S. cerevisiae* share similarity only to the DNA binding regions of the Gcn4p and Gal4p proteins of *C. albicans*, the *Candida* proteins still contain transcriptional activation capacity. Further work will be necessary to establish the molecular logic of linking common DNA binding modules to distinct activation domains in these two fungi, in particular in cases such as Gcn4p where similar cellular processes are regulated by the two proteins.

**ACKNOWLEDGMENTS**

We thank A. J. Brown and Clair Russel for CRC103 and CRC106 strains and for the Clp-lexA and Clp-lexA-GCN4 DNA constructs and A. Mitchell for the *C. albicans* mutant libraries.

This work was supported by Canadian Institutes of Health Research grant MOP-42516 (to M.W.). M.M. gratefully acknowledges a FRSQ-FCAR-Sante Scholarship and National Research Council Graduate Student Scholarship Supplement. This is National Research Council publication 47514.

**REFERENCES**

1. Arndt, K. and G. R. Fink. 1986. GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5′ TGAACA 3′ sequences. Proc. Natl. Acad. Sci. USA 83:4855–4857.
2. Attardi, L. D. and R. Tjian. 1993. Drosophila tissue-specific transcription factor NTF-1 contains a novel isoleucine-rich activation motif. Genes Dev. 7:1341–1353.
3. Barberis, A., J. Pearlberg, N. Simkovich, S. Farrell, P. Reinagel, C. Bamdad, K. Biswas, K., K. J. Rieger, and J. Morschhauser. 2005. One-step transformation of *Candida* strains during infection. Eukaryot. Cell 4:156–165.
4. Gola, S., R. Martin, A. Walther, A. Dunkler, and J. Wendland. 2003. New modules for PCR-based gene targeting in *Candida* albicans: rapid and efficient gene targeting using 100 bp of flanking homology region. Yeast 20:1339–1347.
5. Hagman, J. M. J. Guthe, H. Lin, and R. Grosschedl. 1995. EBF contains a novel zinc coordination motif and multiple dimerization and transcriptional activation domains. EMBO J. 14:2907–2916.
6. Han, S. J., Y. C. Lee, B. S. Gim, G. H. Ryu, S. J. Park, W. S. Lane, and Y. J. Kim. 1999. Activator-specific requirement of yeast mediator proteins for RNA polymerase II transcriptional activation. Mol. Cell. Biol. 19:979–988.
7. Harbison, C. T., D. B. Gordon, T. L. Lee, N. J. Rinaldi, K. D. Mac Issac, T. W. Danford, N. M. Hannett, J. B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, and R. A. Young. 2004. Transcriptional regulatory codes of a eukaryotic genome. Nature 431:109–114.
8. Hinnebusch, A. G. 1984. Evidence for translational regulation of the activator of general amino acid control in yeast. Proc. Natl. Acad. Sci. USA 81:6442–6446.
9. Hinnebusch, A. G. and G. R. Fink. 1983. Positive regulation in the general amino acid control of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 80:5374–5378.
10. Hope, I. A., S. Mahadevan, and K. Struhl. 1988. Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. Nature 333:635–640.
11. Hope, I. A. and K. Struhl. 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. EMBO J. 6:2781–2784.
12. Jackson, B. M., C. M. Drysdale, K. Natarajan, and A. G. Hinnebusch. 1996. Identification of seven hydrophobic clusters in GCN4 making redundant contributions to transcriptional activation. Mol. Cell. Biol. 16:5557–5571.
13. Jeong, C. J., S. H. Yang, Yie, X. Zhang, S. A. Johnston, and T. Kodadek. 2001. Evidence that Gal1p is a target of the Gal4 transcription factor in the mediator. Biochemistry 40:9421–9427.
14. Johnson, P. F., E. Sterneck, and S. C. Williams. 1993. Activation domains of transcriptional regulatory proteins. J. Nutr. Biochem. 4:386–398.
15. Johnston, M. 1987. A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. Microbiol. Rev. 51:458–476.
16. Kadosh, D., and A. D. Johnson. 2001. Rgl1, a protein related to the *Saccharomyces cerevisiae* hypoxic regulator Rox1, controls filamentous growth and virulence in *Candida albicans*. Mol. Cell. Biol. 21:2496–2505.
17. Keegan, L., G. Gill, and M. Plashe. 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science 231:699–704.
18. Lee, T. I., and R. A. Young. 2000. Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. 34:177–137.
19. Limjindaporn, T., R. A. Khalaf, and W. A. Fonzi. 2003. Nitrogen metabolism and virulence of *Candida albicans* require the GATA-type transcriptional activator encoded by GAT1. Mol. Microbiol. 50:1004–1015.
20. Lohr, D., P. Venkov, and J. Zlatanova. 1995. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. 9:777–787.
21. Ma, J., and M. Plashe. 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. Cell 48:847–853.
22. Ma, J., and M. Plashe. 1987. A new class of yeast transcriptional activators. Cell 51:113–119.
23. Marmurstein, R., M. Carey, M. Ptashne, and S. C. Harrison. 1992. DNA recognition by GAL4: structure of a protein-DNA complex. Nature 356:408–414.
24. Mitchell, P. J., and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245:371–378.
25. Nicholls, S., M. Straubon, B. Enjeljhart, A. Nantel, S. Macaskill, M. Whiteway, and A. J. Brown. 2003. *Man1* and *Man2* type transcription factors play no obvious roles in the stress responses of the fungal pathogen *Candida albicans*. Eukaryot. Cell 3:1111–1123.
26. Nobile, C. J., and A. F. Mitchell. 2005. Regulation of cell-surface genes and biofilm formation by the *C. albicans* transcription factor Bcrp1. Curr. Biol. 15:1150–1155.
27. Nogi, Y., and T. Fukasawa. 1989. Functional domains of a negative regulatory protein, GAL80, of *Saccharomyces cerevisiae*. Mol. Cell. Biol. 9:3009–3017.
28. Park, J. M., S. H. Kim, S. J. Han, M. S. Hwang, Y. C. Lee, and Y. J. Kim. 2000. In vivo requirement of activator-specific binding targets of mediator. Mol. Cell. Biol. 20:8709–8719.
45. Peng, G., and J. E. Hopper. 2002. Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein. Proc. Natl. Acad. Sci. USA 99:8548–8553.
46. Qui, H. F., E. Dubois, and F. Messenguy. 1991. Dissection of the bifunctional ARGRII protein involved in the regulation of arginine anabolic and catabolic pathways. Mol. Cell. Biol. 11:2169–2179.
47. Reeves, W. M., and S. Hahn. 2005. Targets of the Gal4 transcription activator in functional transcription complexes. Mol. Cell. Biol. 25:9092–9102.
48. Rupp, S. 2002. LacZ assays in yeast. Methods Enzymol. 350:112–131.
49. Russell, C. L., and A. J. Brown. 2005. Expression of one-hybrid fusions with Staphylococcus aureus lexA in Candida albicans confirms that Nrg1 is a transcriptional repressor and that Gcn4 is a transcriptional activator. Fungal Genet. Biol. 42:676–683.
50. Sigler, P. B. 1988. Transcriptional activation. Acid blobs and negative noodles. Nature 333:210–212.
51. Tournu, H., G. Tripathi, G. Bertram, S. Macaskill, A. Mavor, L. Walker, F. C. Odds, N. A. Gow, and A. J. Brown. 2002. Global role of the protein kinase Gcn2 in the human pathogen Candida albicans. Eukaryot. Cell 4:1687–1696.
52. Traven, A., B. Jelicic, and M. Sopka. 2006. Yeast Gal4: a transcriptional paradigm revisited. EMBO Rep. 7:496–499.
53. Triezenberg, S. J. 1995. Structure and function of transcriptional activation domains. Curr. Opin. Genet. Dev. 5:190–196.
54. Tripathi, G., C. Wiltshire, S. Macaskill, H. Tournu, S. Budge, and A. J. Brown. 2002. Gcn4 co-ordinates morphogenetic and metabolic responses to amino acid starvation in Candida albicans. EMBO J. 21:5448–5456.
55. Van Hoy, M., K. K. Leuther, T. Kodadek, and S. A. Johnston. 1993. The acidic activation domains of the GCN4 and GAL4 proteins are not alpha helical but form beta sheets. Cell 72:587–594.
56. Wilson, R. B., D. Davis, and A. P. Mitchell. 1999. Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions. J. Bacteriol. 181:1868–1874.
57. Wolberger, C., A. K. Vershon, B. Liu, A. D. Johnson, and C. O. Pabo. 1991. Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell 67:517–528.
58. Zhou, K. M., and G. B. Kohlhaw. 1990. Transcriptional activator LEU3 of yeast. Mapping of the transcriptional activation function and significance of activation domain tryptophans. J. Biol. Chem. 265:17409–17412.