Identification of Candida albicans ALS2 and ALS4 and Localization of Als Proteins to the Fungal Cell Surface

L. L. HOYER,* T. L. PAYNE, AND J. E. HECHT
Department of Veterinary Pathobiology, University of Illinois, Urbana, Illinois
Received 26 June 1998/Accepted 18 August 1998

Additional genes in the growing ALS family of Candida albicans were isolated by PCR screening of a genomic fosmid library with primers designed from the consensus tandem-repeat sequence of ALS1. This procedure yielded fosmids encoding ALS2 and ALS4. ALS2 and ALS4 conformed to the three-domain structure of ALS genes, which consists of a central domain of tandemly repeated copies of a 108-bp motif, an upstream domain of highly conserved sequences, and a domain of divergent sequences of the tandem repeats. Alignment of five predicted ALS protein sequences indicated conservation of N- and C-terminal hydrophobic regions which have the hallmarks of secretory signal sequences and glycosylphosphatidylinositol addition sites, respectively. Heterologous expression of an N-terminal fragment of Als1p in Saccharomyces cerevisiae demonstrated function of the putative signal sequence with cleavage following Ala17. This signal sequence cleavage site was conserved in the four other Als proteins analyzed, suggesting identical processing of each protein. Primary-structure features of the five Als proteins suggested a cell-surface localization, which was confirmed by indirect immunofluorescence with an anti-Als antiserum. Staining was observed on mother yeasts and germ tubes, although the intensity of staining on the mother yeast decreased with elongation of the germ tube. Similar to other ALS genes, ALS2 and ALS4 were differentially regulated. ALS4 expression was correlated with the growth phase of the culture; ALS2 expression was not observed under many different in vitro growth conditions. The data presented here demonstrate that ALS genes encode cell-surface proteins and support the conclusion that the size and number of Als proteins on the C. albicans cell surface vary with strain and growth conditions.

The opportunistic pathogen Candida albicans is a fungus that exists in a diverse range of associations with its human or animal host. C. albicans can survive in the host without overt disease symptoms and, under the appropriate circumstances, can cause disease that varies in site and severity. C. albicans infection can be localized and superficial or systemic and disseminated to a wide range of organs (reviewed in reference 7). Adherence of initial point of contact between fungus and host (reviewed in references 7, 9, 17, 23, and 29). Adherence of C. albicans to host surfaces is also involved in the process of colonization, which may occur without accompanying pathogenesis. Investigations to understand C. albicans adhesion have involved characterization of the cell surface, since this is the initial point of contact between fungus and host (reviewed in references 7, 9, 17, 23, and 29). Adherence of C. albicans to numerous cell types, cellular components, and nonliving substances has been examined to further define adhesive relationships. Numerous obstacles exist to hamper the study of C. albicans adhesion, including widely noted growth-medium-dependent effects and differences between C. albicans strains tested in the same assay (reviewed in references 15, 30, and 45). Despite these obstacles, several C. albicans molecules involved in adhesive interactions have been identified (reviewed in references 7, 9, 17, and 23).

Characterization of the C. albicans ALS family has yielded data that address the major themes discussed above. The first gene in the ALS family, ALS1, was isolated in a differential screen to identify hypha-specific genes (26). Although subsequent studies demonstrated that ALS1 is not strictly hypha-specific, its sequence has significant identity with the sequence of AGA1 from Saccharomyces cerevisiae, which encodes α-agglutinin, a cell-surface adhesion glycoprotein that facilitates contact between haploid cells during mating (19, 37). Because C. albicans has not been observed to undergo meiosis or mating, it may be less likely that the function of Als1p is directly analogous to that of Aga1p (26). However, conservation between sequences required for the adhesive function of α-agglutinin and those at the N terminus of Als1p raised the intriguing possibility that Als1p is an adhesion glycoprotein (26). Data supporting this conclusion have been published recently (16, 18).

In addition to its potential to encode an adhesion glycoprotein, other features of ALS1 prompted further study. ALS1 encodes a central domain of tandemly repeated copies of a highly conserved 108-bp sequence that, when translated, predicts a highly conserved 36-amino-acid motif (26). The tandem-repeat sequence hybridizes to several genomic fragments from C. albicans, suggesting that ALS1 belongs to a gene family (26). The existence of a gene family defined by the tandem-repeat-hybridizing fragments was demonstrated by the characterization of ALS3 (25). The size of the ALS family is difficult to estimate because of the presence of additional tandem-repeat-hybridizing fragments and of other genomic sequences that hybridize to a probe derived from the 5′ end of ALS1 (25, 26).
Experiments to characterize the ALS genes and to understand their regulation were initially undertaken to lay the groundwork for studying Als protein function in C. albicans. Studying Als protein function in C. albicans is challenging because, assuming redundancy of function among proteins in the family, creation of a truly null mutant requires characterization of the entire family and disruption of many genes. We reasoned that the number of gene disruption steps could be reduced by knowing which genes were expressed under a particular growth condition. Combining a particular growth condition with specific disruptions could effectively create a null mutant. Studies of ALS gene regulation demonstrated that ALS1 and ALS3 are differentially expressed (25, 26). ALS1 expression in vitro is regulated by components of growth media, and ALS3 is hypha-specific (25, 26). In ALS1-specific expression of ALS1 was shown to vary among strains of C. albicans (25).

In this study, we present data to further characterize the ALS genes and their encoded proteins. Two new genes, ALS2 and ALS4, are described. Similar to previously characterized ALS genes, ALS2 and ALS4 are shown to be differentially regulated. Comparison of ALS gene sequences yielded a generalized ALS gene structure that fits another C. albicans gene, ALA1 (18). Here, we recognize the place of ALA1 in the ALS family. Analysis of sequence features of the five predicted Als proteins suggests that they are localized on the C. albicans cell surface, a property demonstrated by indirect immunofluorescence with an anti-Als antiserum. Taken together, the data presented here demonstrate that the cell-surface-localized Als proteins could account for a significant portion of the strain- and growth-medium-dependent differences in adhesion commonly noted in the C. albicans literature.

MATERIALS AND METHODS

Media and strains. All standard growth media and strains were described previously (25, 26). C. albicans B311 used in this study was purchased from the American Type Culture Collection; other isolates of B311 used in previous studies were not used here.

Library screening and DNA sequencing of ALS genes. Construction of the fosmid library and PCR screening with primers specific for the consensus tandem-repeat sequence were described previously. Fosmids that were positive in the PCR screen were grouped on the basis of PCR products (25); a representative fosmid was chosen from each group for subcloning and DNA sequencing of each Als allele. Fosmids chosen were 19F-1 (ALS2-1), 20F-3 (ALS2-2), 20E-6 (ALS4-1), and 29F-9 (ALS4-2). These fosmids were purified and digested with a variety of restriction enzymes. Southern blots of these digests were probed with an 870-bp KpnI fragment from ALS1 that encodes only tandem-repeat sequence (26) to identify fragments likely to encode related ALS genes. These fragments were subcloned into pUC vectors (57) and transformed into E. coli DH5α (Invitrogen) and transformed into E. coli TOP10F (Invitrogen). The resulting plasmids were subcloned into pCR2.1 (Invitrogen) and transformed into E. coli TOP10F cells (Invitrogen). The DNA sequence of the cloned PCR fragment conformed to the consensus tandem-repeat sequence of ALS1 (18).

Nucleic acid blots. Southern blotting was performed as described previously (26) with the digoxigenin nonradioactive nucleic acid labeling and detection system (Boehringer Mannheim). Blots probed with the ALS1 or ALS5 tandem-repeat sequence were hybridized at 65°C overnight. Blots were washed in 2× SSC (1× SSC is 0.1 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate (SDS) at room temperature for 30 min and then with 0.5× SSC–0.1% SDS at 65°C for 1 h.

Cultures of strains SC5314 and 3153A were used to demonstrate the growth-specific expression of ALS4 on Northern blots. Cells from a culture grown in YPD (yeast extract, peptone, dextrose) were cultured in water and inoculated into 500 ml of fresh YPD at 5 × 106 cells/ml. An aliquot of this culture was extracted to measure total RNA for the 0 h time point. Cells for RNA preparation were washed twice with diethyrylacetonate buffer, sonicated in an ethanol dry ice bath, and stored at −80°C until RNA was extracted. The freshly inopulated YPD culture was cultured at 30°C, with shaking at 200 rpm. Samples were removed every hour for 8 h and processed for RNA extraction as described above. Three or four independent cell counts were performed at each point to construct a growth curve.

Total RNA extraction, formaldehyde gel electrophoresis, and Northern blotting were performed as described previously (25). Fifty micrograms of total RNA was loaded into each gel lane. ALS4-specific message was detected with a 2× end-labelled ALS4-specific oligonucleotide. Hybridization with oligonucleotide probes followed the method of Sundstrom et al. (52). Equal loading of total RNA on Northern blots was evaluated with a fragment from the C. albicans TEF1 gene (25) as described above (26).

Production of anti-Als antiserum. Hyperimmune anti-Als serum was raised in a New Zealand White rabbit was a gift from George Liv (SmithKline Beecham Pharmaceuticals). The anti-Als serum was raised against four 10-mer peptides derived from the N-terminal domain of Als1p (26). These peptides were chosen because they were likely to be in a hydrophilic, surface-exposed region of the mature, folded protein as predicted by secondary-structure algorithms (14). Peptides from the N-terminal region of Als1p were selected because this portion of the protein is freely exposed to the outside of the cell. Peptides from other regions of the protein (26). The peptides selected were GWSLGDTSAN (amino acids 53 to 62), FYSGREENFTF (amino acids 98 to 107), TGGSTDLEDS (amino acids 139 to 148), and NTYFNDGBK (amino acids 156 to 165). These peptides were linked to keyhole limpet hemocyanin (KLH) and emulsified in equal quantities to emulsification in Freund’s complete adjuvant (Sigma). The emulsion was injected into the rabbit at multiple subcutaneous sites. A blood sample was collected from the marginal ear vein 14 days after. The anti-immun serum was administered 4 weeks after the initial immunization. The booster immunization was performed with the mixture of four KLH-linked peptides emulsified in incomplete Freund’s adjuvant (Sigma). A blood sample was collected 14 days after the booster immunization, and the anti-Als titer was assayed on a Western blot of a heterologously produced soluble N-terminal fragment of Als1p (see below). Four total-booster injections were performed, with the anti-Als titer increasing following each round. Increasing titer was judged by increasing dilutions of serum required to obtain an equivalent Western blot signal. Serum collected from the rabbit was stored in small aliquots at −80°C. Precipu- mum serum collected from the same rabbit in which the anti-Als serum was raised and a commercially purchased anti-KLH serum (ICN) were both utilized as negative controls.

Indirect immunofluorescence of C. albicans cells. Cells of strain SC5314 were grown in YPD until they reached late stationary phase; at this stage of growth, cultures typically have a density of approximately 5 × 106 cells/ml. This culture was washed twice in phosphate-buffered saline (PBS) (per liter: 10 g of NaCl, 0.25 g of KCl, and 1.43 g of NaH2PO4; pH 7.2 to 7.3) and counted. A fresh culture of RPMI 1640 (catalog no. 11875-085; Gibco BRL) was inoculated at a density of 5 × 106 cells/ml. This culture was incubated at 37°C and 120 rpm for 1 h 45 min. One hundred microtiter of this culture was spread into an area of a clean glass slide that had been delineated by etching with a diamond pen. Slides were washed thoroughly in PBS before we proceeded. Slides were blocked with 200 μl of 1.5% normal goat serum (Jackson Research Laboratories) diluted in PBS and incubated for 10 min at room temperature in a humid chamber. Excess normal serum was drained from slides, and 200 μl of a 1:100 dilution of detected only with oligonucleotide probes. The resulting ALS4-specific probe, 5′-TTGACGTTTACAAGTAAAGCCTGTTACATTT-3′ (located at nucleotide 897 in the ALS2-coding region), and the ALS4-specific probe 5′-CCGCGCTTCTCTTGGATGACTCATTTACTCATACTT-3′ (located at nucleotide 900 of the ALS4-coding region) were used in Southern blotting. The reverse primer of each probe was synthesized for use in Northern blotting as described below.

ALS1 probe. A probe encoding only tandem-repeat sequences from ALA1/ALS3 (18) was synthesized by PCR with the forward primer 5′-GTTGACGTTTACAAGTAAAGCCTGTTACATTT-3′ and the reverse primer 5′-CCGCGCTTCTCTTGGATGACTCATTTACTCATACTT-3′ (located at nucleotide 900 of the ALS4-coding region) were used in Southern blotting. The reverse primer of each probe was synthesized for use in Northern blotting as described below.

ALS3 probe. A probe encoding only tandem-repeat sequences from ALA1/ALS3 (18) was synthesized by PCR with the forward primer 5′-GTTGACGTTTACAAGTAAAGCCTGTTACATTT-3′ and the reverse primer 5′-CCGCGCTTCTCTTGGATGACTCATTTACTCATACTT-3′ (located at nucleotide 900 of the ALS4-coding region) were used in Southern blotting. The reverse primer of each probe was synthesized for use in Northern blotting as described below.

ALS4 probe. A probe encoding only tandem-repeat sequences from ALA1/ALS3 (18) was synthesized by PCR with the forward primer 5′-GTTGACGTTTACAAGTAAAGCCTGTTACATTT-3′ and the reverse primer 5′-CCGCGCTTCTCTTGGATGACTCATTTACTCATACTT-3′ (located at nucleotide 900 of the ALS4-coding region) were used in Southern blotting. The reverse primer of each probe was synthesized for use in Northern blotting as described below.
the anti-Als serum purified on a protein G column according to the manufacturer’s instructions (MABTrap G II column; Amersham Pharmacia Biotech) and a 1:500 dilution of the preimmune serum from the rabbit in which anti-Als serum was raised. Immunoglobulin G (IgG) concentrations of these preparations were roughly equivalent; a lower dilution of the protein-G-purified serum was used to account for the dilution that occurs during the purification procedure. Primary antiserum was incubated on slides for 1 h at 4°C in a humid chamber. Slides were washed thoroughly in ice-cold PBS. The secondary antibody, fluorescein-isothiocyanate (FITC)-conjugated goat anti-rabbit IgG (heavy plus light chains) (catalog no. 11-095-003; Jackson Laboratory Research Laboratories) was diluted 1:2,500 from a 0.75 mg/ml stock and incubated on slides for 1 h at 4°C in a humid chamber.

Slides were washed thoroughly in ice-cold PBS and stored in darkness at 4°C until they were viewed later the same day.

Immunofluorescence images were obtained with an Olympus BX60 microscope with an oil immersion lens at an approximately 400× magnification. Images were captured with a Photometrics Ltd. system consisting of a charge-coupled device camera (model CH250), an electronic unit (model CE 200A), equipped with a 50-Hz 16-bit A/D converter, and a controller board (model NU 200). Images were acquired and evaluated with Adobe Photoshop software and a Macintosh Quadra 840 AV computer (Apple Computer, Inc.).

RESULTS

ALS2 and ALS4 complete the set of genes that hybridize with the ALS1 tandem-repeat probe. An 870-bp KpnI fragment derived from within the tandem-repeat domain of ALS1 hybridizes with several genomic fragments from C. albicans and C. stellatoidea; the number of hybridizing fragments depends upon the strain examined (26). The 10 copies of the 108-bp tandem-repeat element from ALS1 in strain B792 were aligned to derive a consensus sequence (26). PCR primers based on the most-conserved regions of the tandem-repeat sequence were used to screen a fosmid library from C. albicans (1161). By this technique, fosmids were noted to yield different PCR product patterns and were grouped accordingly (25). Characterization of a representative fosmid from one of the groups yielded ALS3 (25); fosmids from the other groups encoded alleles of two other ALS genes, designated ALS2 and ALS4. Restriction enzyme analysis and hybridization with ALS-gene-specific probes indicated that ALS2 and ALS4 accounted for the remaining fragments that hybridize with the ALS1 tandem-repeat probe in genomic DNA from C. albicans 1161 (Fig. 1).

A similar analysis with 12 C. albicans and two C. stellatoidea strains indicated that, in each strain, all four ALS genes are present and account for all of the restriction fragments that hybridize with the ALS1 tandem-repeat probe (data not shown). Varying numbers of tandem-repeat-hybridizing fragments originally noted in each strain (26) were due to differences between allelic fragments encoding the same gene and to comigration of larger restriction fragments (data not shown).

The DNA sequence was derived for both alleles of ALS2 and ALS4. ALS2 and ALS4 conform to the basic three-domain structure of ALS genes, which includes a central domain of varying numbers of copies of a tandemly repeated 108-bp sequence, a 5′ domain that is approximately 1.3 kb in length and conserved between ALS genes, and a 3′ domain that is variable in length and sequence (25). Characterization of fosmids encoding ALS2 and ALS4, as well as physical map-
Table 1. Comparison of Als protein features from predicted amino acid sequences

<table>
<thead>
<tr>
<th>Als protein feature</th>
<th>Predicted amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Als1p</td>
</tr>
<tr>
<td>No. of amino acids</td>
<td></td>
</tr>
<tr>
<td>N-terminal domain</td>
<td>433</td>
</tr>
<tr>
<td>C-terminal domain</td>
<td>470</td>
</tr>
<tr>
<td>No. of consensus N glycosylation sitesX</td>
<td></td>
</tr>
<tr>
<td>N-terminal domain</td>
<td>0</td>
</tr>
<tr>
<td>C-terminal domain</td>
<td>6</td>
</tr>
<tr>
<td>% Serine and threonine</td>
<td></td>
</tr>
<tr>
<td>N-terminal domain</td>
<td>28</td>
</tr>
<tr>
<td>C-terminal domain</td>
<td>45</td>
</tr>
<tr>
<td>No. of repeat copiesX</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Size of nonglycosylated protein (kDa)Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>133</td>
</tr>
</tbody>
</table>

a The Als1p sequence is from C. albicans B728 (26); the Als5p sequence (18) is from a clinical isolate. All other sequences are from C. albicans 1161. In strain 1161, ALS3 alleles are similar in size, but ALS1 alleles are detectably different lengths (Fig. 1). Allelic differences have been noted previously for other ALS genes, although this has been mainly in the tandem-repeat region, where alleles have been demonstrated to encode different numbers of head-to-tail copies of the 108-bp motif (26) (Table 1). In the current study, alleles of ALS2 and ALS4 were sequenced to study conservation of nucleotides in the non-tandem-repeat domains. In these two domains, alleles of ALS2 and alleles of ALS4 were more than 97% identical in nucleotide sequence (see below). Although the complete tandem-repeat region in both ALS2 and ALS4 was not sequenced, sufficient sequencing was completed in each allele to determine that the tandem-repeat region conforms to the consensus tandem-repeat sequences derived for ALS1 and ALS3 and to confirm that only tandem-repeat sequences were present (data not shown). Sequencing of large regions of tandemly repeated DNA has been omitted in other studies such as the S. cerevisiae genome project (28).

Designation of C. albicans ALS5 and its place in the ALS family. Characterization of ALS1, ALS2, ALS3, and ALS4 yielded a generalized three-domain structure for ALS genes. Another recently characterized C. albicans gene, called ALA1 (18), also fits this basic motif and belongs to the ALS family. Because the name ALA1 has been previously used to denote an alanyl tRNA synthetase in S. cerevisiae (44), and because C. albicans nomenclature follows the S. cerevisiae precedent (48), we propose that the gene described by Gaur and Klotz (18) be called ALS5.

ALS5 was not identified in the original screening of genomic DNA by hybridization with the ALS1 tandem-repeat probe. Therefore, although both genes encode a domain of tandem repeats, the exact tandem-repeat sequences were sufficiently dissimilar to escape detection by cross-hybridization under the experimental conditions employed. Comparison of the consensus tandem-repeat sequences of ALS1 (26) and ALS5 (18) indicated matches between only 39% of the nucleotides (data not shown). Because hybridization between two sequences depends on matches between individual tandem-repeat copies rather than an idealized consensus, Southern blotting was done to see if any genomic fragments hybridized with both probes. An ALS5 tandem-repeat probe was amplified by PCR with primers that flank the tandem-repeat region in this gene (18). The DNA sequence of this probe fragment indicated that it closely matches the consensus sequence of the tandem repeats of ALS5 in the clinical isolate used by Gaur and Klotz (data not shown; 18). Southern blots of BglII-digested genomic DNA from C. albicans and C. stellatoidea strains indicated that the ALS5 tandem repeats hybridized to a different set of genomic fragments than did the ALS5 tandem repeats in most strains (Fig. 2). Based on this result, the small number of fragments of similar size that were detected with both probes most likely encode different genes. Additional ALS genes have been isolated which possess the general ALS three-domain structure but which have more sequence similarities to ALS5 than to ALS1. Analysis of these gene sequences suggested that variability in the tandem-repeat sequences was a potential criterion for the division of the ALS family into subfamilies (24).

Sequences in the N-terminal domain of ALS proteins are highly conserved and encode a secretory signal peptide. Sequences N-terminal of the tandem-repeat domain were highly conserved in each predicted ALS protein, with amino acid identity between different proteins ranging from 68 to 86% (Fig. 3). Identity of the nucleotide sequences encoding the N-terminal domain was 73 to 90%. At the start of each coding region was a hydrophobic sequence with hallmarks of a secretory signal peptide (54). To test whether the hydrophobic N terminus functioned as a signal sequence, a construct expressing the N-terminal 433 amino acids of Als1p under control of the CUP1 promoter was transformed into S. cerevisiae (see Materials and Methods). Supernatants from cultures of the CuSO4-induced construct and a control strain harboring the blank expression plasmid were run on SDS-polyacrylamide gels (31). Coomassie blue and silver staining of these gels indicated that a major protein species was present at approximately 65 kDa in supernatant from cells with the Als1p construct and absent from cells transformed with blank vector (data not shown). N-
terminal amino acid sequencing of the 65-kDa protein yielded the sequence Lys-Thr-Ile-Thr and indicated true secretion of the N-terminal Als1p fragment with resultant loss of the first 17 amino acids (Fig. 3). Cleavage following Ala17 was correctly predicted by the Signalase program (21), which is based on the predictive algorithms of von Heijne (54) and identifies sites for signal peptide cleavage. Analysis of the other Als protein sequences with the Signalase program predicted signal peptide cleavage at the same site, which was conserved in each of the Als amino acid sequences characterized to date (Fig. 3). Because signal sequences have been shown to be processed similarly in S. cerevisiae and C. albicans (39), it is likely that the same processing site is utilized by C. albicans.

The N-terminal domain of the Als protein molecule, particularly within the first 330 amino acids, was predicted to be relatively free of glycosylation. In this region, all predicted Als protein sequences, with the exception of Als2p, lacked consensus sequences for N glycosylation (Fig. 3). After the first 330 amino acids, each predicted Als protein had a threonine-rich region, raising the possibility that O glycosylation may be added; the increased frequency of serine and proline residues also supported this possibility (Fig. 3).

Sequences C-terminal of the tandem-repeat domain are divergent but have a serine-threonine-rich composition. Of the three domains present in Als proteins, the domain C-terminal of the tandem repeats was the least conserved across the family. This domain varied in sequence and length in the predicted Als proteins but exhibited similar amino acid sequence compositions (Table 1). The serine-threonine richness of this domain was consistent with the possibility of abundant O glycosylation (27). This feature, along with the presence of consensus sites for N glycosylation, predicted that the C-terminal domain was heavily glycosylated (Table 1). Both of these features were also observed in the tandem-repeat domain, which was similarly predicted to be heavily glycosylated.

Although sequences of the C-terminal domain were divergent, those residues within 50 amino acids of the stop codon were highly conserved (Fig. 4). Within these conserved residues was a hydrophobic region with hallmarks of the consensus site for glycosylphosphatidylinositol (GPI) addition. The features of a GPI addition site have been well characterized (reviewed in references 8 and 12); following these rules, cleavage at the conserved Gly or Ser was predicted (Fig. 4). Yeast proteins to which GPI is added can be localized to either the cell membrane or, after truncation of the GPI, cross-linked in the cell wall (reviewed in reference 8). Analysis of amino acid sequences predicted from the S. cerevisiae genome sequencing project identified a dibasic motif immediately preceding the GPI attachment site that is present in proteins localized to the cell membrane (8). The lack of this dibasic motif in the Als protein sequences (Fig. 4) suggested that if C. albicans followed the same rules as S. cerevisiae, Als proteins were likely to be localized in the cell wall.

Although the C-terminal domain was the most highly divergent of the three ALS domains, the C-terminal domains of Als2p and Als4p and the nucleotide sequences which encode them were more than 95% identical. Multiple alignment of the nucleotide sequences of the ALS2 and ALS4 alleles indicated that only 3.9% of the sequence positions were mismatched. ALS2-1 and ALS4-1 each had a small gap; in each case, this gap was apparently due to duplication of a trinucleotide sequence in the other allele. Allelic sequences were also examined; ALS2 alleles had 2.8% of sequences mismatched, whereas ALS4 sequences varied in only 0.5% of nucleotides. The mismatch between allelic sequences in the 3′ domain was slightly less, with 0.3% variation between the ALS2 alleles and 0.4% between alleles of ALS4. The nucleotide sequence identity between the 3′ domains of ALS2 and ALS4 extended beyond the coding region; a region approximately 500 bp 3′ of each coding region was sequenced and found to be over 95% identical for each gene (data not shown).

Cell-surface localization of Als proteins by indirect immunofluorescence. The predicted amino acid sequences of the Als proteins suggested that they were localized on the cell surface. Features such as an N-terminal signal peptide, a C-terminal GPI addition site, repeated sequences, and C-terminal regions rich in serine and threonine have all been noted in other yeast cell-surface proteins (reviewed in reference 8). Cell-surface localization of Als proteins was demonstrated by indirect immunofluorescence with a rabbit polyclonal antiserum raised against four KLH-linked 10-mer peptides from Als1p. Subsequent characterization of additional ALS genes indicated that the 10-mer peptide sequences were highly conserved in the predicted amino acid sequences for other proteins of the Als
family (Fig. 3). Recognition of Als1p and Als3p by the anti-Als antiserum was demonstrated by Western blotting of heterologously produced protein fragments (data not shown); other Als proteins remain to be tested.

For immunofluorescence studies, yeast-form cells were grown in YPD medium and transferred to RPMI 1640 to induce germ tube formation. Both the mother yeast and germ tube stained with the anti-Als antiserum (Fig. 5). The specificity of this staining was demonstrated in competition experiments in which staining of both the mother yeast and the germ tube was blocked by the addition of soluble N-terminal Als1p fragment (data not shown).

*C. albicans* cells treated either with preimmune serum from the rabbit in which the polyclonal serum was raised (Fig. 5F) or with commercially purchased anti-KLH antiserum in place of the anti-Als serum did not stain (data not shown).

Although YPD-grown yeast forms stained with anti-Als serum was demonstrated in competition experiments in which staining of both the mother yeast and the germ tube was blocked by the addition of soluble N-terminal Als1p fragment (data not shown). *C. albicans* cells treated either with preimmune serum from the rabbit in which the polyclonal serum was raised (Fig. 5F) or with commercially purchased anti-KLH antiserum in place of the anti-Als serum did not stain (data not shown).

FIG. 3. Alignment of N-terminal amino acid sequences predicted from genes in the ALS family. Amino acid sequences were predicted by the translation of ALS gene sequences of the tandem-repeat domain. Sequences included are Als1p (26), Als2p and Als4p (this study), Als3p (25), and Als5p/Ala1p (18). Sequences were aligned with default parameters of the PILEUP program of Genetics Computer Group software (14). A consensus sequence (Cons), indicating amino acids conserved in all sequences, is provided below the alignment. The vertical line (designated SSC) between residues 17 and 18 denotes the site of signal sequence cleavage demonstrated biochemically for Als1p and predicted by computer algorithm to be conserved for the remaining proteins. Boxed regions labeled 1, 2, 3, and 4 correspond to the four 10-mer peptides from Als1p used to raise the rabbit polyclonal anti-Als antiserum used in indirect immunofluorescence studies. The boxed Als2p and Als4p sequences (labeled Probe) correspond to the region in the nucleotide sequence from which *ALS2* - and *ALS4* -specific oligonucleotides were derived. Boxed sequences between alleles of Als2p or Als4p indicate nonconserved amino acid sequences predicted from allelic nucleotide sequences. Consensus N-glycosylation sites (2) are underlined in the Als2p sequences at positions 253 and 315. All CUG codons have been changed from Leu to Ser (46, 55).

FIG. 4. Alignment of the C-terminal amino acid sequences of Als proteins. Approximately the last 50 amino acids of each predicted Als protein sequence were aligned, to demonstrate sequence conservation in this region. A consensus sequence indicating amino acids conserved in every protein is indicated below the multiple alignment. The putative GPI addition sites are indicated by arrows. The larger arrow over the Gly residue suggests that this is the more likely GPI attachment site.
Variability in ALS gene size and expression pattern. It is well documented that the sizes of ALS genes in any C. albicans strain are highly variable (25, 26). In certain strains, alleles of a given ALS gene produce different-sized proteins due to variation in the numbers of tandem-repeat copies present in each allele (Table 1). Also, certain ALS proteins are likely to be larger than others, with Als1p in one strain, for example, being twice the size of Als1p in another strain (26).

Variability also exists within the ALS family with respect to patterns of gene expression. Previous work demonstrated the differential expression of ALS1 (26) and a hypha-specific expression of ALS3 (25). Northern analysis established that ALS4 expression was correlated with the growth phase of a C. albicans culture. This effect was first noted in a pilot experiment in which C. albicans cells from an overnight YPD culture were subcultured into fresh media and incubated at 30°C with shaking at 200 rpm. RNA was analyzed at 3-h time points by Northern blotting with an ALS4-specific probe. ALS4-specific message was present except when cells were in early- to mid-log phase (data not shown). Cultures were followed for 33 h, at which time ALS4 message was abundant (data not shown). To more carefully analyze the time period in which ALS4-specific message was absent, the experiment was repeated with 1-h time intervals (Fig. 6). A Northern blot hybridized to detect ALS4-specific messages was deliberately overexposed to identify lanes in which ALS4-specific message was absent and to pinpoint the time when the synthesis of ALS4-specific message began (Fig. 6). Because the half-life of the ALS4-specific RNA was not known, it was unclear whether signals present at the 0-, 1- and 2-h time points were due to new synthesis or to dilution and decay of message present in the stationary-phase cells used to inoculate the culture. Synthesis of ALS4-specific message began during the fifth hour, when cells reached mid-log phase, and increased as the culture reached stationary phase. In previous experiments, the increase in ALS4-specific signal continued as the culture reached stationary phase (data not shown). These experiments were also done with C. albicans 3153A, and the same pattern of expression was noted (data not shown).

In contrast to the definable pattern of ALS4 expression, ALS2-specific message was not detected in cultures grown under a wide variety of in vitro conditions, including all growth stages in YPD and YND (neopeptone substituted for peptone); RPMI 1640-induced germ tubes and hyphae; Lee (33) and Soll medium (supplemented Lee medium [3]) at pHs 4.5, 5.5, 6.5, and 7.5; Emmons-modified Sabouraud medium (32) with various carbon sources, including dextrose, galactose, maltose, and sucrose; and hyphal cells induced by adding 10% serum to YPD, 10 mM imidazole buffer (49), and PBS. Cells in these media were grown at various temperatures, including 25, 30, and 37°C. C. albicans strains in these studies included B311,
B792, SC5314, 1177, 3153A, and WO-1. The lack of detection of an ALS2-specific message under so many in vitro conditions suggested a number of possibilities, including the possibility that ALS2 was a pseudogene and the possibility that ALS2 required in vivo signals for expression.

**DISCUSSION**

PCR screening of a *C. albicans* fosmid library with primers based on the consensus tandem-repeat sequence of ALS1 yielded fosmids encoding ALS3 (25), ALS2, and ALS4. These genes account for the ALS1-tandem-repeat-hybridizing fragments detected in high-stringency Southern blots of *C. albicans* genomic DNA (26). An additional gene described in the literature, ALA1 (18), also belongs in the ALS family and is designated ALS5. Although ALS5 encodes tandem repeats similar to those in ALS1, there is sufficient nucleotide sequence divergence between the two consensus repeat sequences that they detect different genomic fragments on high-stringency Southern blots.

Sequences N-terminal of the tandem repeats are highly conserved in the five aligned Als proteins; however, sequences C-terminal of the tandem repeats are divergent. N-terminal hydrophobic sequences function as a signal peptide which is cleaved following Ala17, a site conserved in each Als protein. C-terminal conserved hydrophobic sequences within the last 50 amino acids of each predicted Als protein have characteristics of the site for GPI addition. These observations are consistent with cell-surface localization of Als proteins, a feature demonstrated by indirect immunofluorescence with an anti-Als serum. The anti-Als serum stains both mother yeasts and germ tubes, with the intensity of staining of the mother yeast diminishing as the germ tube elongates.

Analysis of expression patterns of the two newly characterized ALS genes indicates that the expression of ALS4 is correlated with the growth phase of the culture. ALS2 message was not detected in vitro despite the fact that a wide variety of growth conditions were tested. These data support previous evidence that genes in the ALS family are differentially regulated.

**Als protein profile on the *C. albicans* cell surface.** The profile of Als proteins on the *C. albicans* cell surface is highly variable and depends upon several factors, including growth conditions and strain (Table 1) (25, 26). Indirect immunofluorescence experiments demonstrated that the Als protein profile on the *C. albicans* surface is dynamic. Both mother yeasts and germ tubes stain with an anti-Als serum, and staining of the mother yeast diminishes with germ tube elongation. Possible explanations for the diminishing fluorescence of the mother yeast include migration of the Als proteins from the mother yeast to the growing germ tube, masking of the Als protein antigens on the mother yeast by synthesis of new cell wall material, and shedding of the Als protein antigens into the culture medium. Migration of cell wall material from mother yeast to germ tube is unlikely in light of data published by Staebell and Soll (51), which demonstrated growth of a germ tube predominantly by apical expansion. Masking of cell-surface proteins by carbohydrate or by glycosylation of other proteins has been discussed in the context of *C. albicans* cell-surface hydrophobicity (20, 41). Because the N-terminal epitope recognized by the anti-Als
antiserum is predicted to be glycosylation free, it is unlikely that the epitope is masked by the direct addition of carbohydrate. However, modification of preexisting glycosylated cell-surface proteins or the production of new ones might explain the diminished fluorescence of the mother yeast. Another possible reason for the diminishing fluorescence of the mother yeast is the shedding of antigens into the culture medium, a phenomenon that is well documented (reviewed in reference 40). High-molecular-weight mannoproteins within the range of sizes predicted for Als proteins are among those shed (1, 40). Previous work by Brawner and Cutler described a C. albicans cell-surface antigen with the same pattern of expression we observed with our anti-Als serum (4–6). Immunogold electron microscopy with a monoclonal IgM indicated that the antigen studied by Brawner and Cutler was associated with the outer flocculent layer of the cell surface (4–6); they discussed shedding of the antigen into the growth medium as an explanation for its diminished intensity on the mother yeast during germ tube elongation.

Nature of the putative binding domain. Als proteins were named because of the similarities between predicted Als sequences and the sequence of α-agglutinin of S. cerevisiae. Information about the well-characterized α-agglutinin has provided numerous clues about the localization and function of Als proteins (36). The most-significant sequence identity between Als proteins and α-agglutinin exists between the first 300 amino acids of each protein (26). Within this region, α-agglutinin has an immunoglobulin fold structure that is characteristic of many different cell adhesion molecules (10, 13, 35, 56). Because the Als proteins have significant sequence similarity to Aggl1p across domain-sized blocks, they are likely to have a threedimensional structure similar to that of α-agglutinin (34). This observation implicates the first 300 amino acids of the Als protein N terminus as a putative binding domain. Examination of Cys residues in the two sequences indicates that the six Cys residues in the first 300 amino acids of α-agglutinin are conserved in the first 300 amino acids of all Als proteins; however, in this region, each Als protein contains two additional Cys residues that are not found in α-agglutinin. Conservation of the eight Cys residues in the first 300 amino acids of each mature Als protein strongly suggests structural similarity in this portion of each molecule. Additional studies are planned to predict the structure of the putative binding domain of Als proteins and to evaluate its relatedness to the immunoglobulin fold. The tandem-repeat region and C-terminal domain are predicted to exhibit the putative binding domain or to another portion of the molecule. One portion of the molecule with the potential to mediate an adhesive effect is the region predicted to be highly glycosylated. The effect of protein must be separated from that of the likely carbohydrate in order to define the nature of the adhesive interaction.

The availability of amino acid sequences for several Als proteins allows initial speculation about the conservation of function among proteins in the family. Presumed adhesive function due to extensive glycosylation suggests that Als proteins require only a sequence that serves as the scaffold for the addition of carbohydrate; this is provided by the tandem-repeat and C-terminal domains which are rich in consensus N glycosylation sites and in serine and threonine, which are potential sites for the addition of O-linked carbohydrate (2, 27). The amino acid sequence of the C-terminal domain is highly variable between Als proteins, but each protein has features conducive to abundant carbohydrate addition. Adhesive function due to an N-terminal binding domain requires conservation of sequence determining similar structural features. This is observed for each mature Als protein, which has eight conserved Cys residues, a similar N-terminal domain length, and a high degree of sequence identity. Of the five Als protein sequences, Als4p is least like the others in regard to its N-terminal domain sequence, suggesting that it is the least likely to exhibit conserved function. More-meaningful information about conservation of function will be gained from structural predictions with even-less-similar sequences, such as that of Als7p, which is only about 50% identical to Als1p in the N-terminal domain (24). Differential glycosylation at the N terminus, which is possible for Als2p (Table 1), could also lead to alterations in the three-dimensional structure of the region and to corresponding variations in function. Whether adhesive function is due to protein, carbohydrate, or both, proteins encoded by the differentially regulated, multigene ALS family have the potential to explain much of the strain- and growthmedium-dependent differences in adhesion commonly observed for C. albicans.

ACKNOWLEDGMENTS

We thank George Livi and Megan McLaughlin of SmithKline Beecham Pharmaceuticals for their gift of the anti-Als antiserum, Steven Klotz for making the sequence of ALA1 available to us prior to publication, and Roberto Docampo and Hong-Gang Lu for the use of and assistance with the fluorescence microscope and digital imaging equipment. We also thank the Iowa State University DNA Sequencing and Synthesis Facility for sequencing the ALS2 and ALS4 genes. L.L.H. is grateful to Allan Shatzman, George Livi, Stewart Scherer, and Alan Myers for their support of studies of the ALS gene family; without their support, this work would not have been possible. This work was supported by Public Health Service Grant AI39441 (to L.L.H.) by Cooperative State Research, Education and Extension, U.S. Department of Agriculture, under project no. ILLU-70-0305 (to L.L.H.); and by the University of Illinois Campus Research Board (to L.L.H.).

REFERENCES