Functional and Structural Diversity in the Als Protein Family of *Candida albicans*

Donald C. Sheppard‡**, Michael R. Yeaman‡§, William H. Welch¶, Quynh T. Phan‡, Yue Fu‡§, Ashraf S. Ibrahim‡§, Scott G. Filler‡§, Mason Zhang‖, Alan J. Waring‡§, John E. Edwards Jr‡§.

From ‡St. Johns Cardiovascular Research Center, Division of Infectious Diseases, Department of Medicine, Harbor-UCLA Research and Education Institute, Torrance, California 90502, the §David Geffen School of Medicine at UCLA, Los Angeles, California 90024, the ¶Department of Biochemistry, University of Nevada, Reno, Nevada 89507, and the ‖Department of Biological Sciences, California State University, Long Beach, CA 90840.

**Corresponding author. Donald C. Sheppard. Phone: (310) 222-3813. FAX: (310) 782-2016.
E-mail: dsheppard@rei.edu

Running title: *Candida albicans ALS* gene family

Key Words: *Candida*, adherence, invasion, Als protein, endothelial cell
Summary

The human fungal pathogen *Candida albicans* colonizes and invades a wide range of host tissues. Adherence to host constituents plays an important role in this process. Two members of the *C. albicans* Als protein family (Als1p and Als5p) have been found to mediate adherence; however, the functions of other members of this family are unknown. In this study, members of the *ALS* gene family were cloned and expressed in *Saccharomyces cerevisiae* to characterize their individual functions. Distinct Als proteins conferred distinct adherence profiles to diverse host substrates. Using chimeric Als5p-Als6p constructs, the regions mediating substrate-specific adherence were localized to the N-terminal domains in Als proteins. Interestingly, a subset of Als proteins also mediated endothelial cell invasion, a previously unknown function of this family. Consistent with these results, homology modeling revealed that Als members contain anti-parallel β-sheet motifs interposed by extended regions, homologous to adhesins or invasins of the immunoglobulin superfamily. This finding was confirmed using circular dichroism and Fourier-transform infrared spectrometric analysis of the N-terminal domain of Als1p. Specific regions of amino acid hypervariability were found among the N-terminal domains of Als proteins, and energy-based models predicted similarities and differences in the N-terminal domains that likely govern the diverse function of Als family members. Collectively, these results indicate that the structural and functional diversity within the Als family provides *C. albicans* with an array of cell wall proteins capable of recognizing and interacting with a wide range of host constituents during infection.
Introduction

_Candida albicans_ is the most common fungal pathogen in humans (1,2). While normally a harmless commensal, this organism can cause a variety of conditions ranging from superficial mucocutaneous infection, to deep organ involvement in disseminated candidiasis (1). Prior to causing disease, the fungus colonizes the gastrointestinal tract, and in some cases skin and mucous membranes. Adherence to host mucosal surfaces is a key prerequisite for this initial step (3). After colonization, _C. albicans_ enters the bloodstream via infected intravascular devices or by transmigration through gastrointestinal mucosa compromised by chemotherapy or stress ulcerations (1). Organisms then disseminate via the bloodstream, bind to and penetrate the vascular endothelium to egress from the vascular tree and invade deep organs such as liver, spleen and kidney (3). Thus, _C. albicans_ must be capable of adhering to a variety of biological substrates at different stages of infection.

There is abundant experimental evidence to support the role of adherence in candidal virulence. Initial observations noted a correlation between the degree of endothelial cell adherence and virulence of different yeast species (4). More recently, several adhesins of _C. albicans_ have been isolated and characterized (5-11). Mutants deficient in the genes encoding these adhesins not only exhibit decreased adherence to host substrates _in vitro_, but also a corresponding reduction in virulence in several experimental models of _C. albicans_ infection (6,8,12-14). The extent of these reductions has been quite variable, likely due to the actions of other adhesins with redundant or overlapping function.
Previously, we isolated and characterized the *C. albicans* *ALS1* gene by heterologous complementation of non-adherent *Saccharomyces cerevisiae* (9). *ALS1* encodes a cell surface protein that mediates adherence to endothelial and epithelial cells. Disruption of both copies of this gene in *C. albicans* is associated with a 35% reduction in adherence to endothelial cells, and overexpression of *ALS1* increases adherence by 125% (8).

*ALS1* is a member of a large *C. albicans* gene family consisting of at least 8 members originally described by Hoyer et al. (15,16). These genes encode cell surface proteins that are characterized by three domains. The N-terminal region contains a putative signal peptide, and is relatively conserved among Als proteins. This region is predicted to be poorly glycosylated (16,17). The central portion of these proteins consists of a variable number of tandem repeats, (~36 amino acids in length) and is followed by a serine-threonine rich C-terminal region that contains a glycosylphosphatidylinositol anchor sequence (16,17). While the proteins encoded by this gene family are known to be expressed during infection (18,19), the function of the different Als proteins has not been investigated in detail.

We therefore used heterologous expression of Als proteins in non-adherent *Saccharomyces cerevisiae* to evaluate the function of Als proteins in isolation, and to avoid the high background adherence mediated by the multiple other adhesins expressed by *C. albicans*. This heterologous expression system has been used extensively for the study of *C. albicans* genes, including the isolation and characterization of the adhesins *ALS1*, *ALS5* and *EAP1* (7,9,10). Using this model system we demonstrated that Als proteins have diverse adhesive and invasive functions. Consistent with these results, homology modeling indicated that Als proteins are closely related in structure to adhesin and invasin members of the immunoglobulin superfamily of proteins.

---

1 The abbreviations used are: ALS, Agglutin-Like-Sequence; CD, circular dichroism; FTIR, Fourier-transform infrared spectrometry; HBSS, Hanks balanced salt solution; PBS, phosphate buffered saline; TE, Tris-EDTA; CR, conserved region; HVR, hypervariable region.
Structural analyses using circular dichroism (CD) and Fourier-transform infrared (FTIR) spectrometry confirmed that the N-terminal domain of Als1p is composed of anti-parallel \( \beta \) sheet, turn, alpha-helical, and unstructured domains consistent with the structures of other members of the immunoglobulin superfamily. Finally, comparative energy-based models suggest differences in key physicochemical properties of the N-terminal domains among different Als proteins that may govern their distinct adherence and invasive biological functions.

**Experimental Procedures**

**Fungal Strains and Culture** - *S. cerevisiae* strain S150-2B (*leu2 his3 trp1 ura3*) was used for heterologous expression and has been described previously (20). *C. albicans* strain SC5314 was used for genomic cloning. All strains were grown in minimal defined media [1X yeast nitrogen base broth (Difco), 2% glucose, and 0.5% ammonium sulphate, supplemented with 100 µg/ml of L-leucine, L-tryptophan, L-histidine and adenine sulphate] solidified with 1.5% Bacto-agar (Difco) as needed. Growth of ura- strains was supported by the addition of 80µg/ml of uridine (Sigma-Aldrich). Plasmids pGK103, containing *ALS5*, pYF5 containing *ALS1* and pALSn containing *ALS9*, have been described previously (9,10,21). Plasmid pADH1, a generous gift from A. Brown (Aberdeen, UK) contains the *C. albicans* alcohol dehydrogenase gene (*ADH1*) promoter and terminator which are functional in *S. cerevisiae* (22). This plasmid was used for constitutive expression of *ALS* genes in *S. cerevisiae*.

**Human Oral Epithelial and Vascular Endothelial Cells** - The FaDu oral epithelial cell line, isolated from a pharyngeal carcinoma, was purchased from the American Type Culture Collection, and maintained as per their recommended protocol. Endothelial cells were isolated
from umbilical cord veins and maintained by our previously described modification of the method of Jaffe et al. (8,23). All cell cultures were maintained at 37°C in a humidified environment containing 5% CO₂.

**Cloning of ALS Genes** - The genomic sequences of members of the ALS family were identified by BLAST searching of the Stanford database (http://www-sequence.stanford.edu/group/candida/search.html). PCR primers were generated to specifically amplify each of the open reading frames which incorporated a 5’ BglII and a 3’ XhoI restriction enzyme site (Table I). Each gene was cloned by PCR using the Expand® High Fidelity PCR system (Roche Diagnostics). ALS3, ALS6, and ALS7 were amplified from C. albicans SC5314 genomic DNA, whereas ALS1, ALS5 and ALS9 were amplified from plasmids that had been previously retrieved from C. albicans genomic libraries (9,10,21). PCR products were ligated into pGEM-T-Easy (Promega) for sequencing. Sequence verified ALS open reading frames were then released from pGEM-T-Easy by BglII-XhoI co-digestion and ligated into pADH1, such that the ALS gene of interest was under the control of the ADH1 promoter. S. cerevisiae strain S150-2B was transformed with each of the ALS overexpression constructs, as well as the empty pADH1 construct using the lithium acetate method. Expression of each ALS gene in S. cerevisiae was verified by Northern blot analysis before phenotypic analyses were performed.

**Flow cytometry** - To confirm the cell surface expression of each of the Als constructs, we performed indirect immunofluorescence using two different polyclonal anti-Als antisera. Antiserum A consisted of anti-Als1p antibodies, generated by immunization of rabbits with a 417 amino acid N-terminal fragment of Als1p. Antiserum B was rabbit anti-C. albicans mannan.
factor 5 that recognizes *C. albicans* cell wall components, but does not cross-react with *S. cerevisiae* (Iatron Laboratories). For each strain, $10^7$ blastospores were isolated from overnight culture, blocked with 100µl of goat serum, and then stained with either polyclonal antisera A or B at a 1:25 dilution, followed by fluorescein isothiocyanate-labeled goat anti-rabbit IgG at 1:100.

A FACSCaliber (Becton Dickinson) instrument equipped with an argon laser emitting at 488 nm was used for flow cytometric analyses. Fluorescence emission was detected with a 515/40 band-pass filter. Fluorescence data for 10,000 events were collected, and the distribution of cells with fluorescence above baseline (i.e., *S. cerevisiae* transformed with the empty plasmid) was analyzed for each strain using CELLQUEST software (Becton Dickinson).

**Chimeric Als Protein Construction and Expression** - To test the hypothesis that N-terminal sequences were responsible for mediating substrate specific adherence, we constructed chimeric Als5/Als6 proteins by exchanging the N-termini of each protein. Chimeric *ALS5/6* genes were constructed as follows. A *BglII*-HpaI fragment of *ALS5* encompassing the 5’ 2117 bp of the gene was isolated. pGEM-T-*ALS6* was then digested with *BglII* and *HpaI* to release the corresponding 5’ 2126 bp of *ALS6*, and the fragment consisting of pGEM-T-Easy plus the 3’ sequences of *ALS6* was isolated and ligated to the 5’ *ALS5* fragment to generate plasmid pGEM-T-5N6C. An identical approach using the corresponding 5’ fragment of *ALS6* and 3’ fragment of *ALS5* was used to generate plasmid p-GEM-T-6N5C. After sequence confirmation, each chimeric *ALS* gene was released by *BglII*-XhoI digestion and subcloned into p*ADH1* as above. *S. cerevisiae* S150-2B was then transformed with these constructs, and expression was verified by Northern Blot analysis before characterization of their adherence properties.
**Fungal Adherence Assays -**

I. **Six-Well Plate Assay** - To determine the adherence properties of transformed *S. cerevisiae* strains, we used a modification of our previously described adherence assay (8). Briefly, adherence plates were coated by adding 1 ml of a 0.01 mg/ml solution of gelatin (Sigma Aldrich), laminin (Sigma Aldrich) or fibronectin (Becton Dickinson) to each well of a six-well tissue culture plate (Costar) and incubating overnight at 37°C. For endothelial cells, second-passage cells were grown to confluency in 6-well tissue culture plates coated with a 0.2% gelatin matrix; and for epithelial cells, FaDU cells were grown to confluence (3 days) in 6-well tissue culture plates coated with a 0.1% fibronectin matrix. Before adherence testing, wells were washed twice with 1 ml of warm HBSS. *S. cerevisiae* strains to be tested were grown overnight in minimal defined media at 30°C then harvested by centrifugation, washed with Hanks balanced salt solution (HBSS, Irvine Scientific) and enumerated using a hemacytometer. Three hundred organisms were added to each well of a six-well tissue culture plate coated with the substrate of interest and incubated for 30 minutes at 37°C in CO₂. Non-adherent organisms were removed by washing twice in a standardized manner with 10 ml of HBSS. The wells were overlaid with YPD agar [1% yeast extract (Difco), 2% bacto-peptone (Difco), 2% D-glucose, 1.5% agar], and the inoculum confirmed by quantitative culture. Plates were incubated for 48 hours at 30°C, and the colonies counted. Adherence was expressed as a percentage of the initial inoculum. Differences in adherence were compared using a single factor Analysis of Variance test, with p<0.01 considered significant.

II. **Magnetic Bead Assay** - Als5p was originally identified by virtue of the proteins ability to induce agglutination of fibronectin coated beads when expressed on the surface of *S. cerevisiae* (10). We therefore tested *S. cerevisiae* strains transformed with *ALS5, ALS6, 5N6C* and *6N5C*. 
for fibronectin bead adherence using this methodology (10,11). Briefly, tosylated magnetic beads (Dynal Biotech) were coated with fibronectin following the manufacturer’s instructions. Next, 10µl of coated beads (approximately 10^6 beads) were mixed with 1X10^8 transformed S. cerevisiae in 1 ml of 1 X Tris-EDTA (TE) buffer pH 7.0, and incubated with gentle mixing for 45 minutes. The tubes were placed in a magnet to separate beads and adherent S. cerevisiae from non-adherent organisms. The supernatant containing non-adherent organisms was removed by aspiration and the remaining beads were washed three times by resuspending in 1 ml of TE buffer, followed by magnetic separation and aspiration of the supernatant. Finally the washed beads and adherent organisms were resuspended in 100µl of TE buffer and examined microscopically for co-agglutination.

**Invasion Assay** - The ability of Als proteins to mediate endothelial cell invasion was determined using a modification of our previously described differential fluorescence assay (24). Briefly, endothelial cells were grown to confluency on 12 mm diameter glass coverslips coated with fibronectin and placed in a 24-well tissue culture plate (Corning Inc.). Cells were then infected with 10^5 blastospores of each S. cerevisiae strain in RPMI 1640 medium (Irvine Scientific). As a positive control, cells were infected with a similar number of C. albicans blastospores. After incubation for 90 minute, the cells were rinsed twice with 0.5 ml of HBSS in a standardized manner and fixed with 3% paraformaldehyde. Organisms remaining adherent to the surface of the endothelial cells were stained for 1 hr with the rabbit anti-*Candida albicans* antiserum (Biodesign) which had been conjugated with Alexa 568 (Molecular probes) which fluoresces red. This antiserum cross-reacts with S. cerevisiae, at a 2-fold higher dilution. The endothelial cells were then permeablized in 0.2% Triton X-100 in PBS for 10 min, after which the cell-associated
organisms (the internalized plus adherent organisms) were again stained with the anti-\textit{Candida albicans} antiserum conjugated with Alexa 488 which fluoresces green. The coverslips were then observed under epifluorescence. The number of organisms that had been internalized by the endothelial cells was determined by subtracting the number of adherent organisms (fluorescing red) from the number of cell-associated organisms (fluorescing green). At least 100 organisms were counted on each coverslip, and all experiments were performed in triplicate on at least three separate occasions.

\textbf{Molecular Modeling} - Homology and energy-based modeling was conducted to compare overall physicochemical features of Als proteins. First, a knowledge-based method (SWISS-MODEL; (25,26)) was used to analyze and compare combinatorial extension structural alignments of structures in the Swiss and Brookhaven protein databases for proteins with homologous conformation (27). This approach included the BLASTP2 algorithm (28) to search for primary sequence similarities in the ExNRL-3D database. In parallel, the dynamic sequence alignment algorithm SIM (29) was used to select candidate templates with greatest sequence identity. Subsequently, ProModII was used to conduct a primary and refined match analyses. Resulting proteins were used as templates for homology modeling of Als protein backbone trajectories.

Robust models of the N-terminal domains of Als proteins (eg., amino acids 1 – 480; preceding initial tandem repeats) were generated through complementary approaches. The N-terminal domains of Als proteins were converted to putative solution conformations by sequence homology (Composer; (30)) and threading methods (Matchmaker (31), Gene-Fold (32-35) using SYBYL 6.9.1 software (Tripos Associates) operating on Silicon Graphics workstations (SGI, Inc.). Resulting conformers and amino acid side chains of target Als domains were refined by
molecular dynamics, and strain energies minimized using the AMBER95 force field method (36) and the Powell minimizer (37). These approaches optimize side chain interactions where positions of the peptide backbone atoms are fixed. Preferred conformations were determined from extended molecular dynamics in aqueous solvent. Next, the torsion angles of all peptide bonds were adjusted to 180 ± 15°, with minimal constraints. In some cases, molecular dynamics were executed, either with no constraints, or with α-helical regions constrained by applying a 0.4 kJ penalty to the canonical Ramachandran φ and ψ angles. Final global energy minimizations were performed for each model after removal of all constraints and aggregates. Resulting Als N-terminal domain models were prioritized based on three criteria: (i) most favorable strain energy (molecular mechanics); (ii) empirical positional energy functions; and (iii) preservation of the spatial arrangement of potential disulfide bridging (31,38-41). Als models were assessed for validity in relationship to homology templates using standard measures (e-values; (42,43)). Finally, the physicochemical properties of the Als models were visualized by MOLCAD (44); as implemented in SYBYL and HINT platforms (45), such that the physical properties were projected onto the water-accessible surface of the Als N-terminal domains.

**Determination of the Structure of the N-terminus of Als1p** – To test the hypotheses generated by our homology modeling, we determined the structural features of the N-terminal domain of Als1p using the complementary approaches of CD and FTIR spectrometry. This protein, encompassing amino acids 17-432 of Als1p was produced in *S. cerevisiae* and has been described previously (8).

**I. Circular Dichroism Spectrometry** - Circular dichroic spectra were recorded with an AVIV 62DS spectropolarimeter (Aviv Biomedical Inc.) fitted with a thermoelectric temperature
controller. Aqueous solutions of Als1p (10 µM in phosphate buffered saline) were scanned using 0.1 mm light path demountable quartz cells at a rate of 10 nm/min from 260 nm to 185 nm and a sample interval of 0.2 nm. Spectra from buffer lacking peptide were subtracted from sample solutions to minimize light scattering artifacts, and final spectra were an average of 8 scans recorded at 25°C. The instrument was routinely calibrated with (+)-10-camphorsulfonic acid (1 mg/ml in a 1 mm pathlength cell; (46) and ellipticity expressed as the mean residue ellipticity \([\theta]_{MRE}\) (deg-cm² dmol⁻¹). The protein concentration was determined by absorbance at 280nm based on aromatic amino acid composition of the expressed Als1p domain (47). The CD spectra were deconvoluted into helix, β sheet, turn and disordered structures using Selcon (48), through the internet-based Dichroweb (49) interface.

II. Fourier-Transform Infrared (FTIR) Spectrometry – Infrared spectra of Als1p self-films were recorded at 25°C on a Bruker Vector 22 FTIR spectrometer (Bruker Optics) fitted with a deuterated-triglycine sulfate detector at a gain of 4, averaged over 256 scans, and at a resolution of 2 cm⁻¹. Fifty micrograms of the protein in 50 µl of phosphate buffered saline were spread onto the surface of a 50 x 20 x 2 mm germanium attenuated-total-reflectance sample crystal (Pike Technologies) and allowed to dry. The dry protein self-film was then hydrated with D₂O for 1h prior to recording the infrared spectra. Amide I bands of the infrared spectra were analyzed for secondary conformations by area calculations of component peaks with curve fitting software (GRAMS/32, Version 5, Galactic). The frequency limits for the various conformations were as follows: α-helix (1662-1645 cm⁻¹), β-sheet (1637-1613 cm⁻¹ and 1710-1682 cm⁻¹), β-turn – loops (1682-1662 cm⁻¹) and disordered structures (1645-1637 cm⁻¹) (50-52).
Results

Cloning of ALS Family Members and Expression in S. cerevisiae - ALS1, 3, 5, 6, 7, and 9 were successfully amplified and expressed in S. cerevisiae. ALS mRNA expression was detected by Northern blot analysis for each construct (data not shown). Despite the use of three sets of primers, we were unable to amplify ALS2 and ALS4 from genomic DNA of C. albicans SC5314. Given the difficulty of sequencing and assembling across the tandem repeats of ALS genes, it is possible that this outcome reflects errors in the sequence assembly currently available on the published genome database.

Flow cytometry confirmed that each of the Als proteins was expressed on the surface of their respective S. cerevisiae hosts. Two distinct anti-sera demonstrated that all of the Alsp expressing strains exhibited at least a four-fold increase in fluorescence when compared with S. cerevisiae transformed with the empty plasmid (Table II). Consistent with the predicted structural diversity among members of the Als family, the antisera displayed differences in recognition of individual Als expression strains.

Als Proteins Display Different Profiles of Substrate Specific Adherence - S. cerevisiae clones that expressed the various Als proteins were examined for their ability to adhere to a variety of host substrates. There were striking differences in the adherence profiles of the S. cerevisiae transformants to the different substrates (Figure 1). While Als1p-, Als3p- and Als5p-expressing strains bound to all substrates tested, Als6p-expressing S. cerevisiae adhered only to gelatin, and Als9p-expressing S. cerevisiae adhered above background levels only to laminin. Further, there were quantitative differences in adherence to the various substrates. For example, when compared with Als3p, Als1p conferred greater adherence to gelatin, but less adherence to
epithelial cells (p<0.01, single factor ANOVA). Only *S. cerevisiae* expressing Als7p adhered to none of the substrates tested. While small differences in levels of Als protein expression cannot be ruled out by the immunofluorescence studies shown in Table II, such differences are unlikely to be responsible for the substrate-specific binding patterns found in this study. Such a global increase or decrease in the amount of Als protein expressed on the cell surface would be expected to produce a commensurate increase or decrease in adherence across all substrates, and not result in the substrate specific differences that were observed.

*N-terminal Sequences of Als Proteins Confer Substrate Specificity* - Als5p expression in *S. cerevisiae* confers adherence to multiple substrates, including gelatin and endothelial cells, whereas Als6p expression results in adherence to gelatin alone. Despite this marked difference in function, Als5p and Als6p are more than 80% identical at the amino acid level. The tandem repeat and C-terminal portions of these proteins are virtually identical, and the majority of the sequence differences are concentrated in the N-termini of these two proteins. These data suggest that N-terminal sequence variability confers substrate specificity.

This hypothesis was supported by the results of studies determining the adherence phenotypes of chimeric *ALS5/ALS6* constructs. *S. cerevisiae* expressing a chimeric fusion of the N-terminus of Als5p to the C-terminus of Als6p adhered to both gelatin and endothelial cells in a manner similar to Als5p (*Figure 2*). Likewise, strains expressing the chimeric fusion of the Als6 N-terminus to the C-terminus of Als5p adhered only to gelatin, as did *S. cerevisiae* expressing Als6p (*Figure 2*). Further, strains expressing Als5p and chimeric Als5N6C protein agglutinated fibronectin coated beads, while those expressing Als6p and chimeric Als6N5C protein had little to no affinity for these beads. Collectively, these data suggest that the adherence profiles of
these transformed *S. cerevisiae* strains were governed by the N-terminal portion of the Als protein.

**A Subset of Als Proteins Mediate Endothelial Cell Invasion by *S. cerevisiae*.** - *C. albicans* invades endothelial cells by inducing its own endocytosis (53,54). This endocytosis occurs after the organism adheres to endothelial cells, however the *C. albicans* ligands required for this process are unknown. Further, it is unclear if distinct candidal ligands are required for both adherence and endocytosis. In addition to being non-adherent, *S. cerevisiae* does not undergo significant endocytosis by endothelial cells. Therefore to test if Als proteins could serve as invasins as well as adhesins, we determined the ability of *S. cerevisiae* strains expressing Als proteins to invade endothelial cells. As expected, *S. cerevisiae* expressing Als1p, Als3p and Als5p displayed a significant increase in the percentage of cell-associated organisms, reflecting their ability to adhere to endothelial cells. In addition, organisms expressing Als3p, and to a lesser extent Als1p and Als5p, demonstrated significant endothelial cell invasion (**Figure 3**).

**Als Proteins are Homologous to Adhesins and Invasins of the Immunoglobulin Superfamily** – As an initial step in the molecular modeling of Als proteins, we used a knowledge-based search algorithm to identify molecules that share significant structural similarity with Als family members. These models indicate that the N-terminal domains of all Als proteins contain multiple anti-parallel $\beta$-sheet domains, consistent with members of the immunoglobulin superfamily (**Table III**). These proteins typically consist of complex 7-stranded anti-parallel $\beta$-sheet domains, from which project loop / coil structures. The $\beta$-sheet domains are separated from one another by interposing regions. This structure is often referred to as a beads-on-a-
string motif. Of great interest, virtually all of the Als proteins modeled to known adhesin or invasin homologs (Table III). Different patterns of similarity were observed among the Als proteins analyzed. For example, all Als proteins examined, except Als7p, shared significant homology with collagen-binding protein of S. aureus. However, the specific primary, secondary and tertiary homologs varied for most family members (Table III). Indeed, only Als2p and Als9p shared an identical primary, secondary and tertiary homolog.

**Als Proteins Contain N-terminal Hypervariable Regions that Map to Predicted Loop / Coil Structures** - Despite the striking differences in substrate specific adherence mediated by individual Als proteins, large regions of sequence in the N-terminal domains are conserved across this family. However, seven regions of significant divergence among Als proteins designated HyperVariable Regions [HVRs] 1-7, were found. These regions (comprised of 8 or more amino acids in length) contained no consensus identity across Als proteins, and less than 50% consensus conservation. In contrast, the intervening Conserved Regions [CRs] 1-7, displayed more than 30% consensus identity and more than 50% consensus conservation across Als proteins. An identity plot and schematic alignment of these amino acid sequences comprising the N-terminal domains (residues 1-420) of Als proteins with known function is presented in **Figure 4A and B.** Importantly, homology modeling revealed that the HVRs of different Als proteins, while distinguishable in sequence, are predicted to conform to similar loop / coil structures that project from the β-sheet components of the CRs. Thus, these HVRs are likely available to interact with host constituents.
Circular dichroism of the N-terminal domain of Als1p (Figure 5, Panel A) showed a dichroic minimum at 217 nm, and strong positive dichroic maximum near 200 nm. These features are characteristic of a protein having a dominant anti-parallel β sheet component. Deconvolution of the CD spectra indicated the protein assumed conformations of 50.1% β sheet, while other structure class contributions include disordered (26.9 %), turn structures (19.3 %), and alpha helix (3.7%). FTIR measurements of a self-film of the hydrated Als1p strongly corroborated that the sample has a dominant β sheet conformation (Figure 5, Panel B). These spectra revealed strong low frequency amide I bands with peaks centered at 1634 and 1628 cm\(^{-1}\), and a weak high frequency band centered at 1685 cm\(^{-1}\). This frequency splitting of the protein amide I infrared spectra into high and low frequency components has been shown to be typical of the effect of transition dipole coupling between intermolecular anti-parallel β sheets (55). Curve fitting of the spectra indicated that the protein construct is approximately 57.2% antiparallel β sheet. Other secondary structural conformations from curve fitting of the IR spectra include disordered structures (20.5%), turn components (13.3%), and alpha helix (9.0%).

Taken together, the FTIR and CD data strongly support our hypothesis that the N-terminus of Als1p contains predominant domains of anti-parallel β sheet structure containing minor alpha-helical and turn components, interposed by less structured regions.

Three-dimensional Models Suggest Physicochemical Distinctions among Als N-terminal Domains - Molecular models indicated differences in predicted physicochemical attributes of the N-terminal domains of Als proteins that likely influence their interactions with host cells and substrates several. Als proteins appear to be separable into three distinct groups based on surface
distributions of hydrophobicity, charge, and hydrogen-bonding potential (Figure 6). Als1p, Als3p, and Als5p each share similar patterns of these properties, and thus are considered the Als-group-A. In contrast, the predicted physicochemical properties of Als6p and Als7p N-terminal domains (Als-group-B) have striking differences from those of the Als-group-A (Figure 6). While the cationic potential in Als-group-A members is typically segregated from their neutral or anionic facets, positive charge is broadly distributed across the entire surface of the Als-group-B members Als6p and Als7p. Finally, the N-termini of Als2p, Als4p and Als9p appear to constitute a third group of Als proteins (the Als-group-C) that differ structurally from either the Als-group-A or -B proteins. The Als-group-C proteins, would appear to be more similar to the Als-group-A than Als-group-B proteins in terms of hydrophobic or electrostatic distribution.

Discussion

Several proteins with adhesive function have been identified in C. albicans. Hwp1p has been shown to mediate adherence to buccal epithelial cells by acting as a substrate for mammalian transglutaminase (5). EAP1 was recently identified by heterologous expression in S. cerevisiae, and mediates adherence to polystyrene and renal epithelial cells in vitro (7). Of the 8 members of the Als protein family, only Als1p and Als5p have been studied from a functional perspective. Heterologous expression of Als1p has been shown to mediate binding to human vascular endothelial cells and epithelial cells, a finding which has been confirmed in C. albicans through gene disruption studies (8,9). Heterologous expression of ALS5 in S. cerevisiae confers adherence to collagen, fibronectin, bovine serum albumin and laminin (10,11,56). No large scale comparison of the substrate specificities of C. albicans adhesins has been performed. In this study, we compared the adhesive properties of a structurally diverse group of Als protein family
members. Our data demonstrate that the Als proteins comprise a diverse family of surface proteins with an overlapping spectrum of specificities for adherence to a variety of human substrates (Figure 1). Further, results from the present domain exchange experiments indicate that the N-terminal domains of Als proteins confer the specificity of their substrate adherence profiles.

In addition to mediating adherence, our data suggest that Als proteins may also function as invasins. Interestingly, while both Als1p and Als3p expressing *S. cerevisiae* demonstrated similar endothelial cell adherence, Als3p-expressing *S. cerevisiae* underwent internalization at a much higher rate. These results suggest the intriguing hypothesis that endocytosis is not simply an extension of adherence, but rather a distinct process that can be influenced by the ligand-receptor interaction. It is possible that differences in N-terminal sequences in Als proteins mediate these distinct functions, as is the case with adherence. However, we were unable to test this hypothesis using chimeric Als5/6 proteins since adherence is likely a prerequisite for invasion, and non-adherent chimeric proteins would by definition fail to mediate invasion. Further studies to delineate the specific Als domains involved in the invasion process are underway.

The physicochemical properties of protein domains as distributed in three-dimensional space are crucial structural features governing receptor-ligand interactions (57-59). The Als proteins share conformational features characteristic of other adhesins and invasins of the immunoglobulin superfamily. However, individual Als proteins differed in their primary homolog, a finding consistent with the experimental data that members of the Als family exhibit different substrate-binding profiles. Collectively, these patterns of Als homologies suggest that, while Als protein members share a global similarity in structure and predicted fold, there are
important structural differences among distinct Als proteins that are likely responsible for their differences in function.

The results of our structural determinations corroborate our homology modeling, which suggests that the N-terminal regions of Als1p, is comprised predominantly of anti-parallel β sheet domains containing loop/coil structures, with lesser amounts of relatively unstructured regions. These features are hallmark motifs of members of the immunoglobulin superfamily. These results show significant predictive correlation with circular dichroism studies of Als5p (60) indicating that the N-terminal domain of Als5p is characterized by a relative predominance of anti-parallel β-sheet and loop / coil regions. Thus it is highly likely that all members of the Als protein family exhibit this overall structure. Importantly, our structural data are also consistent with our homology models that suggest that many of the HVRs correspond to the flexible loop / coil structures projecting from β-sheet domains in the N-termini of distinct Als proteins. We hypothesize that these structures are integral to substrate-specific binding by Als proteins (Figure 7). Consistent with our data, analogous regions of mannose binding lectin, α-agglutinin, and other members of the immunoglobulin superfamily appear to confer substrate binding specificity (61,62). Furthermore, mutations of these variable loop regions significantly alter substrate binding in these homologous proteins (63,64).

Our three-dimensional modeling strongly suggests that N-terminal domains of individual Als proteins possess distinctive molecular signatures that relate to their adhesive profiles. These signatures incorporate parameters such as surface area, hydrophobicity, and electrostatic charge, yielding configurations that distinguish structural relationships among Als proteins. For example, Als proteins that bind to multiple substrates, such as the Als-group-A members (Als1p, Als3p, and Als5p), have similar predicted N-terminal profiles in terms of steric bulk, hydrophobic
distribution, and electrostatic potential. Yet, even within this group, specific physicochemical
distinctions exist that may govern functional differences within the group (Figure 6). In contrast,
Als proteins with reduced adhesive capacity have surface features predicted to be distinct from
the Als-group-A proteins in multiple physicochemical properties, including hydrophobicity and
electrostatic potential. It is highly likely that the aggregate effects of differences in these
structural features confer the specific functional properties of distinct Als proteins.

Extensive genetic variability has been demonstrated within the ALS gene family. Sequence
variation in specific ALS genes of different isolates of C. albicans have been observed (19,60),
and not all members of the ALS family are present in all isolates. Even significant sequence
divergence between two different alleles in a single isolate have been found (16,19). This degree
of genetic variability would suggest these proteins may undergo rearrangement or mutation at a
relatively high frequency. Such a mechanism would provide the organism with the ability to
generate the high degree of structural and functional diversity demonstrated in this study.
Indirect support for this hypothesis is provided by a recent study of allelic variation of ALS7,
which suggested that this gene is both hypermutable, and that these mutations are subject to
selective pressure (19).

Collectively, these studies suggest an analogy between antibodies and Als proteins at both
the structural and functional level. For example, our homology modeling underscores the
similarities in structural configurations of these families, with hypervariability targeted to
localized domains within an otherwise stable framework (eg., HVRs of Als proteins and F_{ab}
regions in immunoglobulins). Further, as with antibodies, the genetic variability of the ALS gene
family may provide the opportunity for Candida to display a diverse array of proteins with a
spectrum of specificity in adherence and invasion. The availability of such a group of related
proteins is likely to improve the ability of the organism to colonize and invade different anatomical and physiological niches during infection.

**Acknowledgements:** We would like to thank the nurses at Harbor-UCLA Medical Center for collecting umbilical cords, Valentina Avanesian for technical assistance in preparing the Als1 protein for analysis, and Toyota USA for donating the Olympus phase-contrast microscope. This research was supported by Public Health Service grants PO1 AI37194, RO1 AI19990 to J.E.E. and MO1 RR00425. D.C.S. is supported by a Clinician Scientist training grant from the Canadian Institutes of Health Research. M.R.Y. and W.H.W. are supported by RO1AI-48031 from the National Institutes of Health. A.S.I is supported by Public Health Service grant RO3 AI054531 and a Burroughs Wellcome New Investigator Award in Molecular Pathogenic Mycology. S.G.F is supported by National Institute of Health grants 5RO1 DE13974 and 1RO1 AI054928. J.E.E. is the recipient of Bristol Myers Squibb Unrestricted Research Award. We thank Professor James Bowie (UCLA) for use of the AVIV 62DS spectropolarimeter. Selcon CD analysis used the DICHROWEB website located at: http://www.cryst.bbk.ac.uk/cdweb/html/home.html, and is supported by grants to the BBSRC Centre for Protein and Membrane Structure and Dynamics (CPMSD), Daresbury Laboratory, Birkbeck College, University of London. The Bruker Vector 22 FTIR spectrometer was funded by a grant from the Research & Education Institute Scientific Equipment Program.
References

1. Edwards, J. E., Jr. (2000) in Mandell, Douglas, and Bennett's principals and practice of infectious diseases (Mandell, G. L., Bennett, J. E., and Dolin, R., eds), 5th Ed., pp. 2656-2674, Churchill Livingston, Philadelphia

2. Edmond, M. B., Wallace, S. E., McClish, D. K., Pfaller, M. A., Jones, R. N., and Wenzel, R. P. (1999) Clin Infect Dis 29, 239-244

3. Klotz, S. A. (1992) Clin Infect Dis 14, 340-347

4. Rotrosen, D., Calderone, R. A., and Edwards, J. E., Jr. (1986) Rev Infect Dis 8, 73-85

5. Staab, J. F., Bradway, S. D., Fidel, P. L., and Sundstrom, P. (1999) Science 283, 1535-1538

6. Tsuchimori, N., Sharkey, L. L., Fonzi, W. A., French, S. W., Edwards, J. E., Jr., and Filler, S. G. (2000) Infect Immun 68, 1997-2002

7. Li, F., and Palecek, S. P. (2003) Eukaryot Cell 2, 1266-1273

8. Fu, Y., Ibrahim, A. S., Sheppard, D. C., Chen, Y. C., French, S. W., Cutler, J. E., Filler, S. G., and Edwards, J. E., Jr. (2002) Mol Microbiol 44, 61-72

9. Fu, Y., Rieg, G., Fonzi, W. A., Belanger, P. H., Edwards, J. E., Jr., and Filler, S. G. (1998) Infect Immun 66, 1783-1786

10. Gaur, N. K., and Klotz, S. A. (1997) Infect Immun 65, 5289-5294.

11. Gaur, N. K., Klotz, S. A., and Henderson, R. L. (1999) Infect Immun 67, 6040-6047.

12. Kamai, Y., Kubota, M., Hosokawa, T., Fukuoka, T., and Filler, S. G. (2002) Infect Immun 70, 5256-5258.
13. Gale, C. A., Bendel, C. M., McClellan, M., Hauser, M., Becker, J. M., Berman, J., and
Hostetter, M. K. (1998) Science 279, 1355-1358
14. Sundstrom, P., Cutler, J. E., and Staab, J. F. (2002) Infect Immun 70, 3281-3283.
15. Hoyer, L. L. (2001) Trends Microbiol 9, 176-180.
16. Zhao, X., Pujol, C., Soll, D. R., and Hoyer, L. L. (2003) Microbiology 149, 2947-2960
17. Hoyer, L. L., Fundyga, R., Hecht, J. E., Kapteyn, J. C., Klis, F. M., and Arnold, J. (2001)
Genetics 157, 1555-1567.
18. Hoyer, L. L., Clevenger, J., Hecht, J. E., Ehrhart, E. J., and Poulet, F. M. (1999) Infect
Immun 67, 4251-4255
19. Zhang, N., Harrex, A. L., Holland, B. R., Fenton, L. E., Cannon, R. D., and Schmid, J.
(2003) Genome Res 13, 2005-2017
20. Fu, Y., Filler, S. G., Spellberg, B. J., Fonzi, W., Ibrahim, A. S., Kanbe, T., Ghannoum, M.
A., and Edwards, J. E., Jr. (1998) Infect Immun 66, 2078-2084
21. Lucindo, N. V., Sheppard, D. C., Filler, S. G., Edwards, J. E., and Zhang, M. (2002) in
Proceedings of the 102nd Annual Meeting of the American Society of Microbiology, pp.
204, Salt Lake City
22. Bailey, D. A., Feldmann, P. J., Bovey, M., Gow, N. A., and Brown, A. J. (1996) J
Bacteriol 178, 5353-5360
23. Jaffe, E. A., Nachman, R. L., Becker, C. G., and Minick, C. R. (1973) J Clin Invest 52,
2745-2756
24. Phan, Q. T., Belanger, P. H., and Filler, S. G. (2000) Infect Immun 68, 3485-3490
25. Guex, N., and Peitsch, M. C. (1997) Electrophoresis 18, 2714-2723
26. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) *Nucleic Acids Res* **31**, 3381-3385

27. Shindyalov, I. N., and Bourne, P. E. (1998) *Protein Eng* **11**, 739-747

28. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J Mol Biol* **215**, 403-410

29. Huang, X., and Miller, M. (1991) *Adv Appl Math* **12**, 337-367

30. Topham, C. M., Thomas, P., Overington, J. P., Johnson, M. S., Eisenmenger, F., and Blundell, T. L. (1990) *Biochem Soc Symp* **57**, 1-9

31. Godzik, A., Kolinski, A., and Skolnick, J. (1992) *J Mol Biol* **227**, 227-238

32. Jaroszewski, L., Rychlewski, L., Zhang, B., and Godzik, A. (1998) *Protein Sci* **7**, 1431-1440

33. Godzik, A. (1995) *Protein Eng* **8**, 409-416

34. Godzik, A., and Skolnick, J. (1992) *Proc Natl Acad Sci U S A* **89**, 12098-12102

35. Godzik, A., Kolinski, A., and Skolnick, J. (1993) *J Comput Aided Mol Des* **7**, 397-438

36. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. (2003) *J Comput Chem* **24**, 1999-2012

37. Powell, M. J. D. (1977) *Mathem Program* **12**, 241-254

38. Bowie, J. U., Luthy, R., and Eisenberg, D. (1991) *Science* **253**, 164-170

39. Eisenberg, D., Luthy, R., and Bowie, J. U. (1997) *Methods Enzymol* **277**, 396-404

40. Fischer, D., Rice, D., Bowie, J. U., and Eisenberg, D. (1996) *Faseb J* **10**, 126-136

41. Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) *Nature* **356**, 83-85
42. Welch, W., Sutko, J. L., Mitchell, K. E., Airey, J., and Ruest, L. (1996) Biochemistry 35, 7165-7173
43. Welch, W., Ahmad, S., Airey, J. A., Gerzon, K., Humerickhouse, R. A., Besch, H. R., Jr., Ruest, L., Deslongchamps, P., and Sutko, J. L. (1994) Biochemistry 33, 6074-6085
44. Heiden, W., Goetze, T., and Brickmann, J. (1993) J Comput Chem 14, 246-250
45. Kellogg, G. E., Semus, S. F., and Abraham, D. J. (1991) J Comput Aided Mol Des 5, 545-552
46. Johnson, W. C., Jr. (1990) Proteins 7, 205-214
47. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) Protein Sci 4, 2411-2423
48. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (1999) Protein Sci 8, 370-380
49. Lobley, A., Whitmore, L., and Wallace, B. A. (2002) Bioinformatics 18, 211-212
50. Byler, D. M., and Susi, H. (1986) Biopolymers 25, 469-487
51. Surewicz, W. K., and Mantsch, H. H. (1988) Biochim Biophys Acta 952, 115-130
52. Goormaghtigh, E., Raussens, V., and Ruysschaert, J. M. (1999) Biochim Biophys Acta 1422, 105-185
53. Filler, S. G., Swerdloff, J. N., Hobbs, C., and Luckett, P. M. (1995) Infect Immun 63, 976-983
54. Belanger, P. H., Johnston, D., Fratti, R. A., Zhang, M., and Filler, S. G. (2002) Cell Microbiol, (in press)
55. Halverson, K. J., Sucholeiki, I., Ashburn, T. T., and Landsbury, J., P.T. (1991) J. Am. Chem. Soc. 113, 6701-6703
56. Gaur, N. K., Smith, R. L., and Klotz, S. A. (2002) Cell Commun Adhes 9, 45-57
57. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) *J Mol Biol* **179**, 125-142

58. Waring, A. J., Harwig, S. S. L., and Lehrer, R. I. (1996) *Protein and Peptide letters* **3**, 177-184

59. Hancock, R. E. (1997) *Lancet* **349**, 418-422

60. Hoyer, L. L., and Hecht, J. E. (2001) *Yeast* **18**, 49-60.

61. Zhao, H., Wakamiya, N., Suzuki, Y., Hamonko, M. T., and Stahl, G. L. (2002) *Hybridomics* **21**, 25-36

62. Wojciechowicz, D., Lu, C. F., Kurjan, J., and Lipke, P. N. (1993) *Mol Cell Biol* **13**, 2554-2563

63. Renz, M. E., Chiu, H. H., Jones, S., Fox, J., Kim, K. J., Presta, L. G., and Fong, S. (1994) *J Cell Biol* **125**, 1395-1406

64. Viney, J. L., Jones, S., Chiu, H. H., Lagrimas, B., Renz, M. E., Presta, L. G., Jackson, D., Hillan, K. J., Lew, S., and Fong, S. (1996) *J Immunol* **157**, 2488-2497
| ALS gene | Sense [5’ to 3’] | Antisense [5’ to 3’] |
|---------|------------------|----------------------|
| ALS1    | AGATCTCAGATGCTTCAACAATTACATTG | CTCGAGTCACCTAAATGAACACAGGACATA |
| ALS3    | GAAGATCTATGCTACAAATATACTACATTGCTTC | CCGCTCGAGTTAAATAACAAAGGATAATAATGTGATC |
| ALS5    | AGATCTCAACTACAACTGCTAAC   | CTCGACCATATTATTGGTGACAAATC |
| ALS6    | AGATCTCATTCCACCACAATGAAAGCA  | CTCGAGTTGCTCAAATCCCGGTTGA |
| ALS7    | AGATCTTCAACAGTCTAATACCTATGA | CTCGAGACTTGGATTGAATTATACCATATA |
| ALS9    | AGATCTCGAAATGCTACCACAATCCTCTA | CTCGAGTTACCCCTGACGTAGCT |

**Table I.** PCR primers used to amplify the coding regions of ALS genes for heterologous expression in *S. cerevisiae.*
| Als construct | % Cells Above Background (fold increase) |
|---------------|-----------------------------------------|
|               | Antisera A | Antisera B |
| Empty plasmid | -- (1)     | -- (1)     |
| Als1p         | 47.8 (17)  | 50.1 (19)  |
| Als3p         | 24.5 (9)   | 54.0 (20)  |
| Als5p         | 23.5 (8)   | 28.2 (11)  |
| Als6p         | 12.7 (4)   | 16.2 (6)   |
| Als7p         | 22.1 (8)   | 15.7 (6)   |
| Als9p         | 11.4 (4)   | 33.9 (13)  |

**Table II.** Detection of Als proteins on the surface of *S. cerevisiae* by flow cytometric analysis. Blastospores of each strain were stained using indirect immunofluorescence with either polyclonal anti-Als1p antiserum (A) or polyclonal anti-*C. albicans* cell wall antiserum (B), then analyzed using flow cytometry. Results are expressed as percentage of positive cells above background (*S. cerevisiae* transformed with empty plasmid), with fold increase in parentheses.
| Protein | Homolog 1 | Homolog 2 | Homolog 3 |
|---------|-----------|-----------|-----------|
| Als1p   | Invasin / Integrin Binding Protein  
*Yersinia pseudotuberculosis*  
[PDB – 1cwv] <sup>a</sup> | Collagen-Binding Protein  
*Staphylococcus aureus*  
[PDB – 1d2p] <sup>a</sup> | Clumping Factor  
*Staphylococcus aureus*  
[PDB – 1n67A] <sup>b</sup> |
| Als2p   | Collagen-Binding Protein  
*Staphylococcus aureus*  
[PDB – 1d2p] <sup>a</sup> | Invasin / Integrin Binding Protein  
*Yersinia pseudotuberculosis*  
[PDB – 1cwv] <sup>b</sup> | Surface Layer Protein  
*Methanosarcina mazei*  
[PDB – 1L0QA] <sup>d</sup> |
| Als3p   | Collagen-Binding Protein  
*Staphylococcus aureus*  
[PDB – 1d2p] <sup>a</sup> | Invasin / Integrin Binding Protein  
*Yersinia pseudotuberculosis*  
[PDB – 1cwv] <sup>b</sup> | Clumping Factor  
*Staphylococcus aureus*  
[PDB – 1n67A] <sup>c</sup> |
| Als4p   | Collagen-Binding Protein  
*Staphylococcus aureus*  
[PDB – 1d2p] <sup>a</sup> | Invasin / Integrin Binding Protein  
*Yersinia pseudotuberculosis*  
[PDB – 1cwv] <sup>b</sup> | NS |
| Als5p   | Invasin / Integrin Binding Protein  
*Yersinia pseudotuberculosis*  
[PDB – 1cwv] <sup>b</sup> | Surface Layer Protein  
*Methanosarcina mazei*  
[PDB – 1L0QA] <sup>b</sup> | Collagen-Binding Protein  
*Staphylococcus aureus*  
[PDB – 1d2p] <sup>c</sup> |
| Als6p   | Collagen-Binding Protein  
*Staphylococcus aureus*  
[PDB – 1d2p] <sup>b</sup> | Invasin / Integrin Binding Protein  
*Yersinia pseudotuberculosis*  
[PDB – 1cwv] <sup>c</sup> | Neuraminidase  
*Influenza Virus Type B*  
[PDB – 1nscA] <sup>c</sup> |
| Als7p   | Surface Layer Protein  
*Methanosarcina mazei*  
[PDB – 1L0QA] <sup>b</sup> | NS | NS |
| Als9p   | Collagen-Binding Protein  
*Staphylococcus aureus*  
[PDB – 1d2p] <sup>a</sup> | Invasin / Integrin Binding Protein  
*Yersinia pseudotuberculosis*  
[PDB – 1cwv] <sup>b</sup> | Surface Layer Protein  
*Methanosarcina mazei*  
[PDB – 1L0QA] <sup>c</sup> |

**Table III.** Comparison of homologs among Als proteins. Homologs of each Als protein were identified by the knowledge-based algorithm described, and ranked in descending order of structural correlation from 1 to 3. **Key** – statistically significant homology (correlation coefficient, $r^2$) of: <sup>a</sup> $≥ 95\%$ homology; <sup>b</sup> $≥ 90\%$ homology; <sup>c</sup> $≥ 80\%$ homology. NS, no significant model identified for homology modeling ($r^2 ≤ 70\%$). PDB, Protein DataBank code per the National Center for Biotechnology Information format.
Figure 1.

Sheppard et al.
| ALS Construct | Adherence Phenotype |
|---------------|---------------------|
| **Empty Plasmid** | ![Empty Plasmid](image1) |
| ALS5 N | ![ALS5 N](image2) |
| ALS6 N | ![ALS6 N](image3) |
| 5N6C N | ![5N6C N](image4) |
| 6N5C N | ![6N5C N](image5) |

**Figure 2.**

Sheppard *et al.*
Figure 3.

Sheppard et al.
Figure 4.

Sheppard et al.
Figure 5.

Sheppard et al.
Figure 6.
Sheppard et al.
Figure 7. Sheppard et al.
Figure 1. Als proteins confer substrate specific adherence properties when heterologously expressed in *S. cerevisiae*. Each panel demonstrates the % adherence of one Alsp expression strain (filled bars) to a variety of substrates to which *C. albicans* is known to adhere. Adherence of *S. cerevisiae* transformed with the empty vector (empty bars) is included in each panel as a negative control. Gel = gelatin, FN = fibronectin, LN = laminin, FaDU = FaDU epithelial cells, EC = endothelial cells. * p<0.01 when compared with empty plasmid control by single factor ANOVA. Results are the mean ± S.D. of at least three experiments performed in triplicate.
Figure 2. Domain swapping demonstrates that substrate specific adherence is determined by the composition of the N-terminal domain of Als proteins. A representation of the ALS gene or construct being tested is depicted as a bar composed of sequences from ALS5 (black) or ALS6 (white). Adherence properties of each mutant are displayed as a photomicrograph illustrating the adherence of transformed S. cerevisiae to fibronectin-coated beads, and a graph demonstrating the adherence to gelatin (black bars) and endothelial cells (grey bars) as measured in the 6-well plate assay. Results are mean ± S.D. of at least three experiments, each performed in triplicate.
Figure 3. A subset of Als proteins mediate endothelial cell invasion when expressed in *S. cerevisiae*. (A) Endothelial cell adherence of *S. cerevisiae* strains expressing Als proteins or transformed with the empty plasmid (control). Data represents total number of endothelial cell associated organisms and is expressed as cells per high power field. (B) Degree of endothelial cell invasion of Alsp expressing *S. cerevisiae* strains presented as number of intracellular organisms per high power field. * p<0.01 when compared with empty plasmid control by single factor ANOVA. Results are the mean ± S.D. of at least three experiments performed in triplicate.
Figure 4. Alignment of the N-terminal amino acid sequence of Als proteins of known function demonstrates an alternating pattern of conserved regions (CR) and hypervariable regions (HVR).

(A) Percent consensus identity among the N-terminal regions of Als proteins of known function. Note that the signal peptide region (amino acids 1-20) is not shown. Open boxes indicate the regions designated as HVRs 1-7. (B) Schematic alignment of Als proteins showing the composition of the individual HVRs. The sequences are arranged to compare proteins with an affinity to multiple substrates, with those that bind few or no identified substrates. The number of amino acids in each conserved region is indicated in parentheses.
Figure 5. CD and FTIR spectra of the Als1 protein N-terminal domain. (A) Circular dichroism spectrum of 10 uM Als1p in phosphate buffered saline. (B) FTIR spectrum of Als1p self-film hydrated with D$_2$O vapor.
**Figure 6.** Comparison of predicted physicochemical properties of N-terminal domains among the Als protein family. Hydrophobic, electrostatic, or hydrogen-bonding features are projected onto water-accessible surfaces of each domain. **Keys:** hydrophobics – most hydrophobic, brown; most hydrophilic, blue; electrostatics – (spectral continuum) most positive charge (+ 10 kcal/mol), red; most negative charge (-10 kcal/mol), blue; hydrogen-bonding potential – donor, red; acceptor, blue. Als proteins are distinguishable into 3 groups based on the composite of these properties. For example, note the similar hydrophobic, electrostatic, and hydrogen-bonding profiles among **Als-group-A** proteins, Als1p, Als3p, and Als5p. In contrast, **Als-group-B** members, Als6p and Als7p, display striking differences in hydrophobic and electrostatic features from those of **Als-group-A**. In addition to biochemical profiles, note the differences in predicted structure among these domains.
Figure 7. Conceptual model of structural-functional relationships in Als family proteins. Als proteins are composed of three general components: an N-terminal domain, tandem repeats, and a serine/threonine-rich C-terminal domain containing a GPI anchor that interfaces with the *C. albicans* cell wall. As illustrated, Als proteins contain multiple conserved anti-parallel β-sheet regions (conserved regions, CR<sub>1-n</sub>) that are interposed by extended spans, characteristic of the immunoglobulin superfamily. Projecting from the β-sheet domains are loop / coil structures containing the hypervariable regions (HVR). The three-dimensional physicochemical properties of specific Als protein HVRs likely govern interactions with host substrates that confer adhesive and invasive functions to *Candida*. For illustrative purposes, only three N-terminal β-sheet / coil domains and their respective CR/HVR components are shown. Note that this projection is viewed at right angles to the structural images shown in Figure 6.