Structure and Function of Glycosylated Tandem Repeats from *Candida albicans* Als Adhesins

Aaron T. Frank, Caleen B. Ramsook, Henry N. Otoo, Cho Tan, Gregory Soybelman, Jason M. Rauceo, Nand K. Gaur, Stephen A. Klotz, and Peter N. Lipke

Department of Biophysics, University of Michigan, Ann Arbor, Michigan 48109; Department of Biology, Brooklyn College of CUNY, Brooklyn, New York 11210; Department of Biology, John Jay College of Criminal Justice of CUNY, New York, New York 10019; and University of Arizona and Southern Arizona VA Health Care System, Tucson, Arizona 85724

Received 16 August 2009/Accepted 2 October 2009

Yeast adhesins are a diverse set of cell adhesion proteins that mediate adhesion to host cells, environmental substrates, other fungi, and coinfecting bacteria (6, 8, 20, 21, 23, 29). The adhesins share common features, including compact N-terminal domains similar to Ig or lectin domains, Thr-rich midpieces, often in tandem repeats, and long highly glycosylated Ser/Thr-rich C-terminal regions that extend the functional domains out from the cell surface. No structures for the Thr-rich midpieces are known, but they can mediate aggregation of fungal cells (33, 35, 47). The prevalence and conservation of such repeats argue that they are functionally important, despite limited data on their structure and function.

In *Candida albicans*, the Als adhesins are homologous proteins, products of 8 loci that encode numerous alleles of cell surface adhesins (16). In each mature Als protein, there are, from the N terminus, three tandem Ig-like domains, a β-sheet-rich conserved 127-residue amyloid-forming T region, a variable number of 36-residue tandem repeats (TRs), and a highly glycosylated stalk region that extends the N-terminal domains away from the cell surface (Fig. 1) (16, 33, 41). The C termini of these and other wall-associated adhesins are covalently cross-linked into the cell wall through transglycosylation of a modified glycosylphosphatidylinositol (GPI) anchor (18, 25). This modular design, including tandem repeats, is typical of fungal adhesins (8).

The Als protein Ig-like region, T region, and TR region all have protein-protein interaction activities (26, 33, 35). The Ig-like regions can interact with diverse mammalian proteins, presumably in a way analogous to antibody-antigen binding, as has been shown in the homologous protein α-agglutinin from *Saccharomyces cerevisiae* (8, 24, 26, 35). The T regions interact through formation of amyloid-like structures both *in vivo* and *in vitro* (33, 34a, 36). An insight into the function of the tandem repeats followed from observations that Als proteins initiate and maintain cell-to-cell aggregations, either spontaneously ("autoaggregation") or following adhesion to a bead-bound defined ligand (10, 11, 36). Aggregation is more extensive for Als proteins with more tandem repeats (26, 35). This result suggested that the tandem repeats are uniquely structured to facilitate or mediate the aggregative function. Circular dichroism spectroscopy of the TR region of Als5p shows a β-sheet-rich structure in the soluble protein (35).

In support of their direct involvement in aggregation, the repeat region of the *C. albicans* adhesin Als5p mediates cell-cell aggregation in the absence of the Ig-like and T domains (35). Moreover, the repeats can also potentiate binding of Als5p to fibronectin (35). Thus, the TR domains mediate cellular aggregation and increased binding to fibronectin. In addition, TR domains and their amino acid sequences are highly conserved across several *Candida* species (3). These properties need to be explained by their three-dimensional structure.
Materials and Methods

Nomenclature. In this paper, we refer to each individual repeat by the ALS locus that encodes the repeat followed by "R" and the number of the repeat as counted from the N terminus. The repeats chosen for modeling from Als5p derive from sequence NCBI accession number O13368.1; thus, Als5p-R3 is the third repeat in this Als5p sequence. Individual repeats from other Als proteins were chosen as those repeats with sequences most divergent from the Als5p sequences (Fig. 2). The NCBI accession numbers of the protein sequences are as follows: Als1p, P46590; Als3p, XP712666.1; Als6p, AAD42033.1; Als7p, AAF89068.1; Als9p, AAP34370.1. The prefix "L" in the chosen Als3-LR4 designates that the longer allele of Als3 was the source of the sequence.

Tertiary structure modeling. The conformational data-based protein-folding program Rosetta was used to generate structural models of eleven representative TR sequences of the Als family (Fig. 2 and Table 1). The Rosetta algorithm is based on a fragment assembly approach. In this approach, a library of 3- and 9-residue fragments is created based on segments of proteins of known structure. This library represents the local conformational space accessible to these short residue fragments. To create a model, the fragments are combined using a Monte Carlo approach, and assesses structures by combining short peptide structures extracted from the protein structural database PDB (38), then combines structures in a Monte Carlo approach, and assesses energetics of assembled structures. Rosetta has recently been shown to generate accurate models for protein-sized domains (40). We also predicted structures with LINUS, which generates randomized structures and rapidly estimates energetics to choose low-energy models (45). The models were supported by structural analyses with atomic force microscopy and circular dichroism spectroscopy. Functional assays showed that the TR domains can mediate binding activities predicted from the calculated structures.

Glycosylation of structural models. The Glycam Biomolecule Builder (http://glycam.cerc.uga.edu) (19) was used to add the disaccharide 2)-D-mannose-(1-4)-O-glycoside and conforms to the expected average stoichiometry of glycosylation of TRs synthesized by S. cerevisiae (12, 15, 35). The same sugars were similarly added to the exposed hydroxyl groups of the calculated structures. Because there are no homologous structures known, we modeled by two independent ab initio methods. Rosetta assembles structures by combining short peptide structures extracted from the protein structural database PDB (38), then combines structures in a Monte Carlo approach, and assesses energetics of assembled structures. Rosetta has recently been shown to generate accurate models for protein-sized domains (40). We also predicted structures with LINUS, which generates randomized structures and rapidly estimates energetics to choose low-energy models (45). The models were supported by structural analyses with atomic force microscopy and circular dichroism spectroscopy. Functional assays showed that the TR domains can mediate binding activities predicted from the calculated structures.

Table 1. RMSD values for Cn relative to the main structure for folds

| TR sequence | All Cn | β-Sheet | Turn 1 | Turn 2 |
|-------------|--------|---------|--------|--------|
| Als5-R1     | 2.6    | 1.2     | 1.7    | 2.5    |
| Als5-R2     | 2.0    | 1.6     | 2.5    | 1.7    |
| Als5-R3     | 1.5    | 1.2     | 1.9    | 2.1    |
| Als5-R4     | 1.7    | 1.5     | 1.6    | 2.0    |
| Als5-R5     | 2.3    | 1.8     | 4.5    | 2.5    |
| Als5-R6     | 1.5    | 1.2     | 2.3    | 1.7    |
| Als5-R7     | 2.5    | 2.5     | 2.6    | 2.5    |
| Als5-LR4    | 1.7    | 1.7     | 2.4    | 1.5    |
| Als5-R3     | 1.8    | 1.4     | 3.2    | 2.0    |
| Als5-R2     | 1.6    | 1.4     | 2.4    | 1.8    |

Als5p mean 1.9 1.5 2.7 2.2

Overall mean 1.9 1.6 2.7 2.1

*The RMSD values (in Å) were calculated with the RMSD Calculator in VMD (17).
top-scoring structure for the 72-residue, two-domain structure Als5-R1R2 (Als5 repeat sequences 1 and 2). N-glycosylation sites were derivatized with GlcNAc2Man9 yeast core glycan (15, 19). The folded and glycosylated domain models were subjected to brief Molecular Dynamics simulations to minimize energy.

The root mean square differences (RMSDs) of the Cα positions was estimated after alignment by a best-fit procedure. Als5R1 was used as the reference structure for the alignment. The RMSDs for the aligned structure were then calculated relative to the average structure. For our calculations, the β-strand regions were defined as residues 3 to 11, 17 to 23, and 28 to 34. Turn 1 consisted of residues 11 to 15, and turn 2 included residues 23 to 27.

Circular dichroism. A 36-mer peptide with the sequence of Als5-R3 (HNPTVTTTEFWSESYATTEITITNGPEGDSV1R3) was synthesized by the Arizona State University Peptide Facility and purified to ~90% purity. Far UV spectra were obtained using AVIV and Applied Photophysics Chirascan circular dichroism (CD) spectrophotometers. The peptide was suspended at 1 mg/ml in 10 mM phosphate buffer (pH 5.6), and 2,2,2-trifluoroethanol was added to 0, 10, 20, 30, and 40% (vol/vol) concentrations. For each condition, three spectra were obtained at 25°C and averaged, and the baseline was subtracted.

Secondary structural percentages were estimated from the raw data with CDStr, CONTINLL, and SELCON3 programs (43, 44, 46). All three programs gave similar secondary structure predictions; CDStr results are reported, because they gave the most consistent results, regardless of changes in wavelength range or basis set.

Soluble truncated versions of Als5p, Als5p1–443 (Als5p protein with residues 1 to 443), and Als5p1–664, were expressed and purified as previously described (33), and spectra were acquired within a week of purification. Denaturation and renaturation CD studies were carried out by raising (for denaturation) or lowering (for renaturation) the temperature in 10°C steps at 1°C min−1 and then equilibrating for 10 min after each step. After each step, 10 spectra were obtained, averaged, and then smoothed (43).

Binding assays. Enzyme-linked immunosorbant assays (ELISAs) for Als5p binding to substrates were carried out as previously described, using horseradish peroxidase-labeled anti-V5 at 1:500 dilution as the sole antibody (35). The antibody reacted equivalently with both forms of Als5p (35; data not shown). Peroxidase activity was assayed with Quanta Blu fluorescent substrate (Pierce Chemical). Soluble forms of Als5p were purified as previously described and were used within 2 weeks of isolation to minimize amyloid formation (33). All assays were performed in triplicate, and standard deviations are shown as error bars.

RESULTS

Predicted secondary structure of Als repeats. The individual tandem repeats (TRs) have highly conserved sequences and low rates of nonsynonymous substitutions, no insertions, and almost no deletions (Fig. 2 and data not shown) (33). The repeats vary in number from 2 to 36 copies in the alleles sequenced so far (16, 31, 49, 51). TR sequences include a high frequency of aliphatic β-branched amino acids, with 9 to 11 Thr residues and 6 Val or Ile residues in clusters within each repeat. All of these amino acid have high β-strand potential (blue in Fig. 2). These β-branched clusters are interspersed with clusters of residues with high turn or helical potential (red in Fig. 2). In the multiple alignments like Fig. 2, β-branched residues Ile, Val, and Thr can substitute for one another (blue regions), and the residues with turn-inducing or high helical regions also substitute only for each other (red and white regions).

The sequence conservation leads to uniform consensus predictions of β-sheet-rich structures (Fig. 2) (4, 7, 9). The sequence-based predictions were highly consistent in predicting structures composed of three β-strands in each of the Als5p repeats (7). The first and second strands sometimes varied in their predicted length by one or two residues, but the third strand predictions were invariant (Fig. 2, bottom). Neither JPRED nor other secondary structure prediction methods showed any helical regions (4, 7, 9). Therefore, we predict that the overall fold of the repeats is conserved.

Circular dichroism spectroscopy. The TR region has a β-sheet-rich secondary structure in soluble versions of Als5p (35). To establish whether the strongly predicted secondary structure of individual repeats was reflected in solution, we carried out far UV CD spectroscopy on a synthetic 36-mer peptide (HNPTVTTTEFWSESYATTEITITNGPEGDSV1R3) (Als5p-R3) dissolved at 1 mg/ml in buffer. The spectra showed a strong negative peak at 198 nm and a less negative shoulder at ~220 nm and varied little between pH 5.5 and 7.0, the approximate physiological pH range for this surface-exposed protein (data not shown). Curve fitting with CDStr gave secondary structure predictions of about 34% β-sheet content, slightly less β-turn and unstructured content, and very low α-helix (Fig. 3). These fractions are consistent with an all-β-domain structure (7, 14). Increasing trifluoroethanol (TFE), which generally stabilizes peptide structures but favors helix formation (50), shifted the negative peak to longer wavelength and made the shoulder more negative. The secondary structure analyses showed that TFE had little effect on secondary structure at concentrations up to 20% (vol/vol). At higher TFE concentrations, α-helix and aperiodic structures increased. Thus, the secondary structure of a single synthetic repeat reflected the secondary structure predictions in its all β structure under a variety of pH levels and at TFE concentrations up to 20%.

Three-dimensional models of Als5p repeats. Three independent analyses implied uniform structure for tandem repeats: sequence conservation, secondary structure predictions, and CD of the synthetic peptide and repeats in situ. To test whether this uniformity led to uniform tertiary structures, we used independent modelers to predict structure of many Als tandem repeat domain sequences. Rosetta modeling generated highly similar structures for all six repeats in Als5p (Fig. 4A). In each cluster of modeled structures, the lowest-energy structures had β-β-β antiparallel topology, with highly similar tertiary structure. The RMS distance differences for the Cα,s in the superimposed models were 1.9 Å for all residues and 1.6 Å for
residues in the β-strands (Table 1). RMSD values for the residues in the core β-sheet were considerably smaller (Fig. 4). These results thus showed uniform secondary and tertiary structure for the repeats.

Because Rosetta is based on libraries of known peptide conformations, we were concerned that it might predict known structures preferentially. We therefore repeated the modeling with LINUS, which assesses energetics for randomly generated structures. Interestingly, LINUS also predicted similar antiparallel three-strand β-sheet topology for all six repeats of Als5p. The LINUS results therefore provided an independent confirmation of the propensity of Als repeat sequences to form compact three-stranded antiparallel β-sheet domains and increased our confidence in the Rosetta models.

It was possible that the uniformity of the predictions was due purely to the unusual amino acid composition of the repeats. We tested whether this was so by modeling randomly permuted sequences of each repeat in Rosetta. The predicted structures were much different from those of the authentic sequence. There were all-helical structures as well as antiparallel β-sheet structures. Moreover, none of the models of permuted sequences showed the three-stranded β-β-β topology (for examples, see Fig. 4C). The energy values of the best structures from permuted sequences were substantially worse than those from the authentic sequences (see Fig. S1 in the supplemental material). Therefore, the consistency of the modeled structures resulted from both the amino acid composition and the conserved sequences of the repeats.

**Tandem repeats in other Als proteins.** The tandem repeats in other Als proteins have more diverse sequences than those in Als5p (Fig. 2). Therefore, we also modeled examples of the repeats, in each case choosing the repeats with sequences most divergent from the Als5p TR consensus. Rosetta models for these repeats also showed uniform antiparallel β-β-β strand topology (Fig. 4B). The β-sheets in these domains were superimposable on those of the Als5 TR domains (Fig. 4B). In four of the five models, RMSD variations were indistinguishable from those among the Als5p repeat domains (Table 1). In each case, random permutations of the repeat sequence gave different structures (data not shown), a result that reinforced the sequence-specific nature of the three-strand models.

The single exception to the uniform results was that the top-scoring models for Als9-R2 showed two sets of structures. For this sequence, too, all of the models gave three antiparallel β-strands as the other models did. However, the secondary structure of the domain differed in that the spatial position of the first strand varied in different trials. In about 70% of the trials, the topology, tertiary structure, and RMSD variation of the lowest-energy structure were similar to those of the lowest-energy models of the other domains (red in Fig. 4B). However, in about one-third of the models of Als9-R2, the first strand was modeled above the plane formed by second and third strands. Therefore, the structure was less strongly consistent.

**Glycosylation patterns in the tandem repeats.** Gel electrophoresis, lectin blotting, and stoichiometric analyses show that the TR domains are heavily O glycosylated when exogenously expressed in S. cerevisiae, with about 1.5 mannose units per Ser or Thr hydroxyl group (35). In the models, about two-thirds of the Ser and Thr hydroxyls were solvent accessible, as expected (see Table S1 in the supplemental material). Therefore, we would expect a mean glycosylation of about 2 mannose residues per exposed hydroxyl group. This mean ratio of 2:1 is consistent with known O-glycosylation structures in C. albicans and S. cerevisiae. In C. albicans, the O-linked oligosaccharides include α-mannose, α1,2-linked manno-biose, and mannotriose (12, 15).

Typically, residues Thr4, Thr6, Thr7, Thr8, Ser12, Ser14, Thr18, Ser30, and sometimes Thr16, Thr22 and Thr28 were accessible to solvent to a degree consistent with glycosylation (Fig. 5; see Table S1 in the supplemental material). Therefore, these residues were glycosylated with the disaccharide Manα1,2Manα1. The resulting glycosylated structures were tested for stability by Molecular Dynamic simulations on a time scale of a few nanoseconds. The structures were stable, as indicated by small RMSD fluctuation of ~0.03 Å². The glycosylated models showed that oligosaccharides surround exposed surface hydrophobic surface areas (Fig. 5). This pattern was conserved in all modeled repeats of Als5p (Fig. 5C) and other Als proteins (Fig. 5D).

Some TR domains, such as those from Als1p and Als3p,
have a potential N-glycosylation site at Asn2. These sites were modeled with a yeast core glycan of GlcNac2Mann, a minimal sized N-glycan in C. albicans (15). The result was a polar position for these glycans on the more hydrophilic end of the prolate ellipsoidal domains (Fig. 5D).

**Tandem domain models.** To gain structural insights about the structure of multiple repeats, Als5p repeat sequences 1 and 2 (Als5p-R1) were modeled together, as were repeats R3-R4, and a set of three sequences, R1 to R3, delaying nonlocal interactions in the folding. For these models, we generated 3,000 structures rather than our standard 1,000 in Rosetta and LINUS. Each of the repeats consistently folded as the monomers did, again showing the stability of the compact three-stranded domains (Fig. 6). Local folding of individual repeat domains is consistent with sequential domain folding after synthesis and endoplasmic reticulum (ER) translocation and is also consistent with force extension analyses by atomic force microscopy of Als5p (2, 27, 32).

**Functions of TR domains.** The structural models predicted that the hydrophobic effect should mediate the adherence of TR domains to surfaces and to each other. A single-antibody ELISA measured binding to polystyrene for equivalent molar amounts of Als5p1–443 (without TR domains) and Als5p1–664 (including six TR domains) (35). Inclusion of TR domains resulted in greater binding at lower protein concentrations, a result indicative of increased affinity of Als5p for polystyrene (Fig. 7A). Half-maximal binding was achieved at $5 \times 10^{-10}$ M for Als5p1–664, whereas the value was $1.5 \times 10^{-8}$ M for Als5p1–443, 30-fold-lower affinity (Fig. 7A). The geometric mean difference in affinity was 10-fold in three independent experiments. Thus, the TR domains showed greater affinity for polystyrene than the Ig-like and T domains did.

The single-antibody ELISA was also used to estimate the affinity of Als5p for a protein substrate. Excess Als5p with or without the TR domains bound to increasing amounts of immobilized fibronectin (Fig. 7B). Both Als5p1–443 and Als5p1–664 forms bound to fibronectin at similar concentrations, with half-maximal binding at $2.5 \times 10^{-8}$ M (approximately 1 µg/ml). Nevertheless, almost 10-fold-more Als5p1–664 bound than Als5p1–443. Thus, inclusion of TR domains did not affect the affinity of the Ig-T region for fibronectin but altered the maximal binding.

To determine whether the increased binding resulted from Als-Als aggregation, we titrated Als5p against limiting fibronectin. The results for Als5p1–443 were similar to those shown in Fig. 7B and showed saturation of binding, as expected for adhesin-ligand binding at limited discrete sites. On the other hand, titration with Als5p1–664 yielded binding proportional to the concentration of Als5p1–664 (Fig. 7C). This result showed that binding of Als5p1–664 was nonsaturable and increased linearly with added protein, a result demonstrating that each Als5p molecule that bound became a potential ligand for more Als5p. Thus, the linear increase in binding for this form of Als5p was characteristic of self-association of Als5p. Therefore, the TR domains mediated both increased affinity for polystyrene and Als5p-Als5p homotypic binding.

**Unfolding of TR domains.** A second strong prediction of the models is that the tandem repeats fold independently and thus should unfold independently. This independent unfolding has, in fact, been observed by atomic force microscopy (see Discussion) (2). In addition, the uniformity and reproducibility of the models implied that the unfolding of the TR domains would be a reversible equilibrium process in vivo. Furthermore, the TR
domains should refold to their native structure after denaturation. Therefore, we thermally denatured and renatured Als5p1–443, which lacks tandem repeats, and Als5p1–664, which includes them. For each protein, CD spectra were obtained during heating and cooling. Als5p1–664 denatured reversibly, and the unfolding was cooperative with an isosbestic point at 216 or 217 nm, indicative of an unfolding equilibrium reaction (Fig. 8A) (34). The spectra showed great stability of the Ig-like and T regions and equilibrium unfolding up to 80°C. The denaturation of each form was fully reversible upon recoiling to 20°C: the CD spectra were indistinguishable from those of the original untreated samples (Fig. 8B). The results were similar for Als5p1–443 (see Fig. S2 in the supplemental material). These data showed that the Als5p TR region has an intrinsic ability to refold to a stable native conformation, a result in accord with the consistency and stability of secondary structure predictors and the calculated structures.

**DISCUSSION**

We have calculated structures for the Thr-rich tandem repeats from Candida albicans Als proteins. We believe that these structures are the first for Thr-rich repeats in any fungal adhesin or cell wall glycoprotein. Structural calculation seems the most practical approach currently, because there are no structures in PDB on which to base homology models (the closest match has e = 0.14) and because O glycosylation makes nuclear magnetic resonance (NMR) structure determination impractical and crystallization is not yet possible for X-ray diffraction. The Als repeats showed highly regular structures, and these structures were well-supported by independent physical and functional data. In addition, the models explain many of the observed and predicted activities of the repeat regions in Als proteins.

Sequence conservation led to similar structures. The alignment shown in Fig. 2 shows high conservation of key residues, especially the aliphatic β-branched amino acids Ile, Thr, and Val that form the conserved β-strands. There are only 15 nonconservative substitutions in 396 amino acid positions in 11 sequences in Fig. 2, and this low frequency is reflected in the low nonsynonymous substitution rate $K_a$ (33). At the same time, the normal-to-high rate of synonymous substitution $K_s$ shows that this sequence is subject to mutation at normal frequency. Therefore, the region must be under strong purifying selection, and few amino acid substitutions are tolerated. The vast majority of the nonsynonymous mutations are in the second turn of each domain (residues 23 to 26 in Fig. 2), and they are all substitutions between residues common to turns (37). Thus, the Als proteins and their individual repeats have conserved structural propensities, implying that the structure itself is essential to domain function (3).

This conservation of sequence results in the remarkable convergence in secondary and tertiary structure models. Ten of eleven TR sequences generated the same fold as the lowest-energy structure in all trials, and the eleventh gave similar structures in the majority of tests. The results were also consistent between two independent prediction algorithms: Rosetta, based on a short sequence structural motif library, and LINUS, which finds stable structures from among randomly generated models. The programs also have different structure
This common structure was sequence specific: random sequences with the same composition as the repeats generated Rosetta models that were different from each other and different from the three-strand fold that the authentic sequences generated (e.g., Fig. 4C; also see Fig. S1 in the supplemental material).

Features of the conserved structures. The spatial patterns of surface hydrophobicity and glycosylation were similar in all modeled TR sequences (Fig. 5). The exposed hydrophobic side chains were highly conserved: one hydrophobic patch consisted of Val5, Ile/Val21, and Ile or Val residues in positions 31 to 33; another patch was more aromatic and included conserved residues Phe/Tyr11, Trp12, Phe/Tyr15, and Pro36. The positions of the glycosylated residues were also remarkably constant (see Table S1 in the supplemental material). Therefore, the conserved sequences resulted in similar structures with conserved surface hydrophobicity and glycosylations. Thus, the TR region of each Als protein consists of a string of compact tandem domains, each with hydrophobic binding surfaces with surrounding hydrophilic O-glycans (Fig. 5, 6, and 9A).

Supporting structural data. The models are supported by extensive physicochemical and functional data from independent approaches. Atomic force microscopy has demonstrated that each tandem repeat folds into a discrete domain, and in studies of Als5p with different numbers of repeats, the number of unfolding domains is equal to the number of repeats (2). Each TR domain unfolds to a length characteristic of extended peptide sequence, with a length increase corresponding to that expected for a 36-residue sequence with an original long axis dimension of 4.3 nm, the same dimension shown in the models (Fig. 5 and 6). Thus, each unfolding domain extends by a length equal to the difference between the folded three-stranded domains shown in Fig. 4 to 6 and the length of an extended peptide. We conclude that AFM confirms the number and linear dimensions of the domains.

Circular dichroism spectra also supported the models. CD spectroscopy of the intact proteins demonstrates that the TR domains have a β-sheet-rich structure, as predicted by secondary and tertiary modeling (35). That a synthetic sequence of a single repeat also forms a similar structure shows that the β-sheet pro-

FIG. 8. CD spectra of thermal unfolding (A) and refolding (B) of Als5p–664. The sample was heated (A) or cooled (B) at 1°C/min and equilibrated for 10 min after each 10°C step of temperature change. An isosbestic point is present at 217 or 218 nm in each experiment.

FIG. 9. Cartoons of multidomain structures of TRs. Each three-strand β-sheet domain is rendered as a prolate spheroid, with hydrophobic surfaces colored yellow and hydrophilic glycosylated regions colored green as in Fig. 5. (A) Drawing of three successive domains. (B) Model for interaction of three Als TR domains (colored as in panel A and rotated 90° around the x axis relative to that shown in panel A) with the surface of a larger protein. Hydrophobic areas (yellow), basic areas (blue), and acidic areas (red) are shown. (C) Model for binding to a hydrophobic surface such as polystyrene. (D) Model of interaction of TR domains from two Als proteins with antiparallel orientation. Similar models can be drawn with parallel orientations. Note that incomplete overlap of the two chains would result in unoccupied binding surfaces able to bind TR domains of other Als molecules or to heterologous ligands, making binding potentially noncompetitive.
penetres in each individual repeat, as the convergent structure calculations predict.

The experimentally demonstrated stability of the domain structure also supports the models. First, the in vitro secondary structure in the presence of TFE, which tends to induce α-helical structure, is remarkable for the synthetic peptide (Fig. 3) (50). Second, the stability to TFE is also seen in spectra of Als5p1–664 (data not shown). Finally, the stability and consistency of the calculated structures were also consistent with the observed equilibrium unfolding and renaturation of the TR region after thermal denaturation (Fig. 8) (44). Therefore, CD results supported the β-sheet potential of each individual repeat and also confirmed that the β-sheet-rich domain structures are thermodynamically favored.

Als TR domain activities. The structures of the Als tandem repeats showed the physical basis for their known activities. Specifically, the exposed hydrophobic surfaces of the TR domains should interact with other hydrophobic surfaces. Canonical protein-protein interaction domains consist of a core of hydrophobic residues surrounded by an annulus of interacting polar groups (5, 39, 40). These hydrophobic and polar regions each form complementary interactions with the same kinds of residues on the protein ligand. This type of structure is cartooned for the Als tandem repeat domains in Fig. 9.

Als tandem repeats and heterotypic binding. The TR domain surfaces have the generally “sticky” characteristics of protein-protein interaction surfaces. Exposed aliphatic and aromatic hydrophobic side chains form a central surface, and the polar interactions would be mediated through the mannosyl residues (Fig. 5, 6, and 9). These structures would have highly diverse interaction partners, because carbohydrates can donate or accept H bonds or mediate charge-dipole or dipole-dipole interactions with polar or charged amino acids or with other glycolys groups. Thus, each repeat domain presents a hydrophobic surface of amino acid side chains surrounded by flexible polar mannose units that can mediate carbohydrate-amino acid or carbohydrate-carbohydrate bonding. Such interactions are cartooned in Fig. 9B, which illustrates apposed hydrophobic surfaces with polar carbohydrates (green) interacting with polar acidic (red) and basic (blue) groups on the surface of a large protein ligand. This model illustrates heretofore unexplained observations about the broad specificity of Als binding to both heterologous ligands, perhaps including interactions of Als5p with hydrophobic regions of Hwp1p and ferritin (1, 8, 16, 28, 47, 49, 51). More repeats are associated with greater aggregation in the S. cerevisiae flocculins, as they are in the C. albicans Als proteins (16, 47, 48). These similarities imply that the approach taken here should also be applicable to tandem repeats in other fungal adhesins.

ACKNOWLEDGMENTS

We thank Janis Lee and Melissa Garcia for their consideration and comments on the manuscript.

This work was supported by NIGMS SCORE grants S06 GM070758 and SCI GM 083756 to Brooklyn College. Work in GLYCAM was supported by GM 55230 and the Center for Functional Glycomics through RR 05357. A.T.F. gratefully acknowledges support of an NSF Graduate Research Fellowship.

REFERENCES

1. Almeida, R. S., S. Brunke, A. Albrecht, S. Thewes, M. Laue, J. E. Edwards, S. G. Filler, and B. Hube. 2008. The hyphal-associated adhesin and invasive Als5p of Candida albicans mediates iron acquisition from host ferritin. PLoS Pathog. 4:e1000217.
2. Alsteens, D., V. Dupres, S. A. Klotz, N. K. Gaur, P. N. Lipke, and Y. F. Dufrene. 2009. Unfolding individual Als5p adhesion proteins on live cells. ACS Nano 3:1677–1682.
3. Butler, G., M. D. Rasmussen, M. F. Lin, M. A. Santos, S. Sakhikumar, C. A. Munro, E. Rheinbay, M. Grabherr, A. Forche, J. L. Reedy, I. Agrafioti, M. B. Arnaud, S. Bates, A. J. Brown, S. Brunke, M. C. Costanzo, D. A. Fitzpatrick, P. W. de Groot, D. Harris, L. L. Hoyer, B. Hube, F. M. Klis, C. Kodira, N. Lennard, M. E. Logue, R. Martin, A. M. Neiman, E. Nikolou, M. A. Quail, J. Quinn, M. C. Santos, F. F. Schmitzberger, G. Sherlock, P. Shah, K. A. Silverstein, M. S. Skryzep, D. Soll, R. Staggs, I. Stansfeld, M. P. Stumpf, P. E. Sudbery, T. Srikanthia, Q. Zeng, J. Berman, M. Berriman, J. Heitman, N. A. Gow, M. C. Lorenc, B. W. Birren, M. Kellis, and C. A. Cuomo. 2009. Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature 459:657–662.
4. Chou, P. Y., and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. Relat. Areas Mol. Biol. 47:45–148.
5. Clackson, T., and J. A. Wells. 1995. A hot spot of binding energy in a hormone-receptor interface. Science 267:383–386.
6. Coronado, J. E., S. M Nicar, S. L. Epstein, W. G. Giu, and P. N. Lipke. 2007. Conserved processes and lineage-specific proteins in fungal cell wall evolution. Eukaryot. Cell 6:2269–2277.
7. Cuff, J. A., M. E. Clamp, A. S. Siddiqui, M. Finlay, and G. J. Barton. 1998. JPred: a consensus secondary structure prediction server. Bioinformatics 14:892–893.
8. Dreganis, A. M., J. R. Rauceo, J. E. Coronado, and P. N. Lipke. 2007. A
biochemical guide to yeast adhesins: glycoproteins for social and antisocial occasions. Microbiol. Mol. Biol. Rev. 71:252–294.

9. Garnier, J., J. F. Gibrat, and B. Robson. 1996. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol.

10. Gaur, N. K., and S. A. Klotz. 2004. Accessibility of the peptide backbone of protein ligands is a key specificity determinant in Candida albicans SRS adherence. Microbiology 150:277–284.

11. Gaur, N. K., S. A. Klotz, and R. L. Henderson. 1999. Overexpression of the Candida albicans A1 gene in Saccharomyces cerevisiae results in aggregation following attachment of yeast cells to extracellular matrix proteins, adherence properties similar to those of Candida albicans. Infect. Immun.

12. Gemmill, T. R., and R. B. Trimble. 1999. Overview of N- and O-linked oligosaccharide structures found in various yeast species. Biochim. Acta 1426:227–237.

13. Goodard, T. C., Huang, and T. E. Ferrin. 2007. Visualizing density maps with UCSF Chimera. J. Struct. Biol. 157:281–287.

14. Grigorescu, A., M. H. Chen, H. Zhao, P. C. Kahn, and P. N. Lipke. 2000. A CD2-based model of yeast alpha-agglutinin elucidates solution properties and binding characteristics. J. Bacteriol. 182(5):133–139.

15. Hazen, K. C., D. R. Singleton, and J. Masuoka. 2007. Influence of outer region mannosylphosphorylation on N-glycan formation by Candida albicans: normal acid-stable N-glycan formation requires acid-labile mannosylphosphoryl addition. Glycobiology 17:1052–1060.

16. Hoyer, L. L., C. B. Green, S. H. Oh, and X. Zhao. 2008. Discovering the secrets of the Candida albicans agglutinin-like sequence (ALS) gene family—a sticky pursuit. Med. Mycol. 46:1–15.

17. Humphrey, W., A. Dalke, and K. Schulten. 1996. Visual molecular dynamics. J. Mol. Graph. 14:33–38, 27–28.

18. Kaptyn, J. C., L. L. Hoyer, J. E. Hecht, W. H. Muller, A. Andel, A. J. Verkleij, M. Makarov, H. Van Den Ende, and F. M. Klis. 2000. The cell wall architecture of Candida albicans wild-type cells and cell wall-defective mutants. Mol. Microbiol. 35:601–611.

19. Kirschner, K. N., A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley, and R. J. Woods. 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J. Comput. Chem. 29:622–655.

20. Klotz, S. A., N. K. Gaur, R. De Armond, D. C. Sheppard, N. Khardori, E. J. J. Richardson, and J. Kurjan. 2001. Estimation of protein secondary structure and error from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal. Biochem. 287:252–260.

21. Klotz, S. A., N. K. Gaur, D. F. Lake, V. Chan, J. Rauco, and P. N. Lipke. 2004. Degenerate peptide recognition by Candida albicans adhesins Als3p and Als1p. Infect. Immun. 72:2029–2034.

22. Kumamoto, C. A., and D. B. Vines. 2005. Alternative Candida albicans lifestyles: growth on surfaces. Annu. Rev. Microbiol. 59:113–133.

23. Lipke, P. N., and M. R. Yeaman. W. H. Welch, Q. T. Phan, Y. Fu, A. S. Ibrahim, S. G. Filler, M. Zhang, A. J. Waring, and J. E. Edwards, Jr. 2004. Functional and structural diversity in the Als protein family of Candida albicans. J. Biol. Chem. 279:30480–30489.

24. Lipke, P. N., and J. Kurjan. 2000. Roles for glycosylation of cell surface receptors involved in cellular immune recognition. J. Mol. Biol. 293:351–366.

25. Lipke, P. N., and J. Kurjan. 2004. Global cell surface conformational shift mediated by a Candida albicans adhesin. Infect. Immun. 72:4948–4955.

26. Pan, Z. P., and J. Kurjan. 2007. Yeast cell adhesions have functional amyloid-forming sequences. Eukaryot. Cell 7:776–782.

27. Pochez, F., C. M. Smales, S. J. Soll, and L. L. Hoyer. 2007. Analysis of ALS7 open reading frame in Candida albicans: Sixty alleles of the ALS7 open reading frame in Candida albicans ALS7: tandem repeats generate functional variability. Nat. Genet.

28. Povey, J., F. C. M. Smales, S. J. Soll, and M. J. Howard. 2007. Comparative analysis of Candida albicans ALS3 alleles discriminated by the number of copies of the tandem repeat sequence in the central domain. Microbiology 153:673–681.

29. Rauco, J. M., N. K. Gaur, K. G. Lee, J. E. Edwards, S. A. Klotz, and P. N. Lipke. 2007. Yeast cell adhesion molecules have functional amyloid-forming sequences. Eukaryot. Cell 7:776–782.

30. Rauco, J. M., R. De Armond, H. Otto, P. C. Kahn, S. A. Klotz, N. K. Gaur, and P. N. Lipke. 2006. Threonine-rich repeats increase fibronecin binding in the Candida albicans adhesin Als3p. Eukaryot. Cell 5:1664–1673.

31. Rauco, J. M., N. K. Gaur, K. G. Lee, J. E. Edwards, S. A. Klotz, and P. N. Lipke. 2006. Threonine-rich repeats increase fibronecin binding in the Candida albicans adhesin Als3p. Eukaryot. Cell 5:1664–1673.

32. Sheppard, D. C., M. R. Yeaman, W. H. Welch, Q. T. Phan, Y. Fu, A. S. Ibrahim, S. G. Filler, M. Zhang, A. J. Waring, and J. E. Edwards, Jr. 2004. Functional and structural diversity in the Als protein family of Candida albicans. J. Biol. Chem. 279:30480–30489.

33. Soll, D. R. 2008. Candida biofilms: is adhesion sexy? Curr. Biol. 18:R717–R720.

34. Sreevama, N., and R. W. Woody. 2000. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSTR methods with an expanded reference set. Anal. Biochem. 287:252–260.

35. Sreevama, N., and R. W. Woody. 2004. Computation and analysis of protein circular dichroism spectra. Methods Enzymol. 383:318–351.

36. Srivinasan, R., P. J. Fleming, and G. D. Rose. 2004. Ab initio protein folding using LINUS. Methods Enzymol. 383:48–66.

37. van Stokkum, I. H., J. H. Spoelder, M. Bloemendal, R. van Grondelle, and F. C. Groen. 1990. Estimation of protein secondary structure and error analysis from circular dichroism spectra. Anal. Biochem. 191:110–118.

38. Verstrepen, K. J., A. Jansen, F. Lewitter, and G. R. Fink. 2005. Intragenic tandem repeats generate functional variability. Nat. Genet. 37:986–990.

39. Verstrepen, K. J., R. T. Reynolds, and G. R. Fink. 2004. Origins of variation in the fungal cell surface. Nat. Rev. Microbiol. 2:533–540.

40. Zhang, N., A. L. Harrex, B. R. Holland, L. E. Fenton, R. D. Cannon, and J. Schmidt. 2003. Sixty alleles of the ALS7 open reading frame in Candida albicans: ALS7 is a hypermutable contingency locus. Genome Res. 13:2005–2017.

41. Zhao, H., M. H. Chen, Z. M. Shen, P. C. Kahn, and P. N. Lipke. 2001. Environmentally induced reversible conformational switching in the yeast cell adherence protein alpha-agglutinin. Protein Sci. 10:1113–1123.

42. Zhao, X., S. H. Oh, K. M. Yeater, X. Zhao, C. Pujol, D. R. Soll, and L. L. Hoyer. 2005. Functional specificity of Candida albicans Als3p proteins and clade specificity of ALS3 alleles discriminated by the number of copies of the tandem repeat sequence in the central domain. Microbiology 151:673–681.