MINIREVIEW

Comprehensive Analysis of Glycosylphosphatidylinositol-Anchored Proteins in Candida albicans

Mathias L. Richard* and Armeél Plaine

Laboratoire de Microbiologie et Génétique Moléculaire, INA P-G UMR-INRA1238 UMR-CNRS2585, 78850 Thiverval-Grignon, France

There are two types of membrane proteins: the integral membrane proteins and the lipid-anchored proteins. Integral membrane proteins contain one or several transmembrane domains that allow for the formation of hydrophobic α-helices, which ultimately embed the protein in a lipid bilayer. We count four types of lipid-anchored proteins divided into two groups: one group generally links the proteins to the internal side of the plasma membrane and the other generally links them to the outside. In the first group are found the amide-linked myristoyl anchors (calcineurin B) (6), the (thio)ester-linked fatty acyl anchors, and the thioether-linked prenyl anchors (G protein alpha subunits) (13). The second group is composed only of the proteins that contain a C-terminal signal sequence that allows for linkage to a glycosylphosphatidylinositol (GPI) anchor. The anchor of the GPI-anchored proteins (GpiPs) can be removed by the action of specific phospholipases (24), converting the protein into a water-soluble form. GpiPs are thought to associate with lipid rafts, specialized regions of the membrane containing elevated levels of cholesterol and sphingolipids, common to human cells, for example (48). Proteins destined to be GPI anchored share conserved features: an N-terminal signal sequence for localization to the endoplasmic reticulum (ER), a C-terminal hydrophobic domain (9 to 24 residues) for transient attachment to the ER membrane, and the so-called ω site, where the protein is cleaved to be ligated to the GPI anchor (57). The ω site is localized 9 to 10 amino acids before the C-terminal hydrophobic domain, and its amino acid environment determines whether the protein has a high probability to be GPI anchored. Amine acids at positions ω-1 to ω-11 form a linker region usually with no charge and no secondary structure (α-helix or β-sheet). The ω site region is composed of small amino acids to fit the transamidase protease catalytic pocket, and finally, the spacer region (ω+3 to ω+9) is comparable to the linker, having no charge and being flexible (16). Shortly after protein synthesis in the ER, the preformed GPI anchor replaces the C-terminal transmembrane region. The core GPI anchor consists of a lipid group (serving as the membrane anchor), a myoinositol group, an N-acetylgalcosamine group, three mannose groups, and a phosphoethanolamine group, which ultimately connects the GPI anchor to the protein via an amide linkage (58) (Fig. 1). The addition of mannose groups and the positioning of side chains like phosphoethanolamine on the GPI anchor contributes to the variety of GPI anchors identified so far (Fig. 1). We will not discuss GPI anchor biosynthesis in this review since we have decided to focus the review on the proteins only, but it is important to note that the GPI anchors are essential for viability in yeast since any deletion of proteins key in GPI biogenesis is lethal (15, 37, 58).

GPI anchoring is encountered in every eukaryotic cell, including unicellular yeast cells, cells of several parasites, and the highly specialized mammalian cells (7, 17, 38). Nevertheless, this wide occurrence does not dictate specific functions, as many types of proteins are GPI anchored (glucosidase, alpha amylase, aspartyl protease, superoxide dismutase, and phospholipase, etc.) and many GpiPs have unknown functions (2, 11, 18, 39). From the human perspective, GpiPs have been under investigation for two specific reasons: (i) a rare defect in the biosynthesis of the GPI anchor itself triggers a disease called paroxysmal nocturnal hemoglobinuria affecting the protection of erythrocytes against complement factors (28), and (ii) PrP* is a GPI-anchored protein derived from the normal PrP(c) that causes various fatal neurodegenerative diseases, such as the well-known Creutzfeldt-Jacob disease (36). In Saccharomyces cerevisiae, the characterized GpiPs, most are involved in cell wall compound biosynthesis, flocculation, protease activity, sporulation, and mating (9).

CANDIDA ALBICANS POSSESSES 115 PUTATIVE GPI-ANCHORED PROTEINS

The availability of more and more eukaryotic genome sequences has given us the resources to do large-scale comparative-genomics studies of GpiPs. The only way to anticipate the addition of a GPI anchor is to look for the three characteristics described above. However, none of these characteristics are strictly defined, thus confounding our interpretations. In the last 4 years, several studies have increased our understanding of the numbers and functions of the GpiPs. In terms of the human genome, no large-scale analysis of GpiPs has been published yet, but in the B. Eisenhaber laboratory, a preliminary survey of potentially GPI-modified proteins in Homo sapiens sapiens was done in 2004 (http://mendel.imp.ac.at/gpi/Hs/hs.html). Lower eukaryotes, such as Arabidopsis thaliana (5)
and Caenorhabditis elegans (14), and microorganisms, such as Aspergillus nidulans, C. albicans, Candida glabrata, Neurospora crassa, S. cerevisiae, and Schizosaccharomyces pombe (9, 12, 14, 16, 55, 61), however, have been subjected to a more thorough analysis. It is notable that S. cerevisiae and Schizosaccharomyces pombe seem to possess a smaller number of GpiPs than C. albicans.

Four different lists of C. albicans GpiPs are published in the literature (12, 16, 21, 55). In this review, we will not include the results of Paula Sundstrom’s analysis since her work was done using assembly 6, an incomplete genome sequence. In an attempt to simplify our analysis, we have made a comprehensive list of putative GpiPs by combining the three data sets identified from assembly 19 (12, 16, 21). The three different teams presenting lists of C. albicans GpiPs identified different numbers of putative GpiPs (102, 104, and 169), but a careful analysis of each data set allowed us to refine our list (Table 1).

Other than the differences in the algorithms used to define GpiPs, the main reasons for the differences among the data sets were the presence in the data sets of open reading frames (ORFs) that have since been deleted and the presence in the data sets of two alleles coding for the same GpiP. The recent publication online of release 20, a complete and curated haploid genome sequence for C. albicans (http://candida.bri.nrc.ca/candida/index.cfm), has allowed for the creation of a robust base for further research on this group of proteins in C. albicans or for comparison with other organisms.

Our final list consists of 115 putative GpiPs corresponding to the C. albicans genome (Table 1), 70 of which are common to the three published data sets, 6 of which are common to the data sets presented by de Groot et al. and Eisenhaber et al., 24 of which are specific to the study by de Groot et al. (12), 9 of which are specific to the study by Eisenhaber et al. (16), and 6 of which are specific to the study by Garcerà et al. (21). The number of putative GpiPs identified in C. albicans (115) is almost twice that identified in S. cerevisiae (58 GpiPs); however, we should keep in mind that lists for both C. albicans and S. cerevisiae are the results of computerized predictions and may still retain some annotation artifacts.

MORE THAN 65% OF THE GpiPs IN C. ALBICANS HAVE UNKNOWN FUNCTIONS

Looking at each putative GpiP provides us with very little functional information, as a vast majority of the proteins are of unknown function. Nevertheless, newly available genomic sequences and transcriptomic tools now allow us to make informed inferences about genes with unknown functions. Thus, among the 115 GpiPs identified from the C. albicans genome, four classes can be described (Fig. 2): (i) the GpiPs with completely unknown functions (corresponding to 76 genes, or 66% of the GpiP genes), (ii) the GpiPs with putative functions related to cell wall biogenesis or remodeling (corresponding to 15 genes, or ~13%), (iii) the GpiPs with putative functions related to cell-cell adhesion and interactions (corresponding to 13 genes, or ~11%), and (iv) the GpiPs with diverse enzymatic properties, such as superoxide dismutase activity and aspartyl protease activity, etc. (corresponding to 11 genes, or 9.5% altogether). The largest class, representing 66% of genes and having completely unknown functions, may be relevant to the future discovery of novel genes involved specifically in C. albicans pathogenicity. Firstly, because of the putative localization of GpiPs at the cell surface, they are in a very good position to be involved in any interaction with the surrounding environment, the host cells. Secondly, C. albicans is highly adapted to its environment compared to other opportunistic fungi (in terms of pH, oxidation, phagosomes, etc.), which suggests that during evolution it developed several mechanisms to colonize its host; these original functions might be carried out by these surface proteins of unknown function.
Careful analysis of the 76 proteins of unknown function reveals that 41 of them have been reported to be regulated by various transcription factors and conditions (Table 1). Rim101 is the most common regulator of these proteins; 22% of them are regulated by Rim101 (9 out of 41). Changes in iron levels, the morphogenetic switch, Nrg1, and Tup1, as well as cell wall regeneration, also regulate them; for each of the aforementioned regulators, we count eight genes (19.5%) reported to be regulated and six after caspofungin treatment. Even if this number is not statistically robust considering the sizes of the groups, these data might be worth further attention. For instance, in the literature, two papers report experimental data on the monitoring of the genes regulated by Rim101 (4, 34). The results vary between 132 (M.L.R., unpublished results) and 186 (4, 34) genes out of more than 5,000; thus, around 3% of the genome is regulated by Rim101, which is a much smaller amount than the 22% indicated in our data set. This finding would suggest that, as well as being involved in pH response, in adaptation to high salt concentrations, and in morphogenesis, Rim101 might also play a role in regulating the expression of some GpiPs. Alternatively, it is possible that GpiPs may simply have an important role in the three aforementioned functions. For example, the two glucosyltransferases, Phr1 and Phr2, are both regulated by Rim101 and are essential for growth at alkaline and acidic pHs, respectively (10, 18).

FAMILIES WITHIN THE GROUP OF GPI-ANCHORED PROTEINS

In order to identify families within the GpiP gene group, we used BLASTP to compare each gene against the entire group. From this analysis, we identified eight families with only two members (Table 2). Interestingly, two of these gene couples are adjacent on the chromosome (families 1 and 8), which might be the result of a tandem duplication event during evolution. We identified six families with three members, two families with four members, and two families with at least eight members, the ALS genes and the HYR/IFF genes. This analysis indicates that within the GpiP group, more than 50% of the proteins (67 out of 115 proteins) have a paralogue. There are at least two hypotheses for the utility of a gene family in a genome: (i) the function carried out by the family is essential, and thus the redundancy of the genes protects the cell against any lethal loss in functionality, and (ii) the function carried out requires fine-tuning that proteins within the family can provide only if they are expressed under different conditions or if their affinities for diverse substrates are slightly different, for instance. A good example is the DFG5 family, where the DFG5 or PGA51/DCW1 gene is dispensable to the cell but the deletion of both genes is lethal in C. albicans (26), as is the deletion of the respective orthologues of these genes in S. cerevisiae (31), suggesting that the function carried out by these proteins is probably essential to the cell.

In addition to families within the GpiP group, we looked for GpiP genes that could be members of gene families within the whole C. albicans genome. Our analysis indicated two types of scenarios (Table 2): (i) GpiP genes that belong to an “outside” family, that is, those including members other than GpiP genes (seven families), and (ii) GpiP genes that are already members of “inside” families but that also have paralogues outside of the GpiP gene group (eight families). For example, several gene families already characterized are heterogeneous: the product of CHT3 is not GPI anchored, while those of CHT1 and CHT2 are; the products of SAP9 and SAP10 are GPI anchored, but those of the eight other SAP genes (SAP1 through SAP8) are not; and the products of PLB1 and PLB2 are not GPI anchored, while those of PLB3 through PLB5 are.

With these heterogeneous families, it is difficult to assess whether the GPI anchor prediction was wrong in these cases or whether the cells use two proteins with similar functions but distinct localizations. Additionally, several GpiPs were found to have no GpiP homologues but instead one to four non-GpiP homologues (Table 2).

Our limited understanding of the functions and the regulation of these proteins hinders our abilities to interpret how C. albicans modulates its genome to end up with such diversity in localization. Nevertheless, it seems that the percentage of gene families within the GpiP group is relatively high, with many heterogeneous families composed of genes for proteins with or without GPI anchors.

One of the most studied GpiP gene families to date is the ALS (agglutinin-like sequence) family. The entire ALS family was first described in detail by Lois Hoyer in 2001 (26) as a family of genes homologous to the alpha-agglutinin genes of S. cerevisiae. Since then, numerous research papers on this family have been published: in the last 10 years at least 55 papers have discussed the expression or the roles of various ALS genes in C. albicans biology. An entire review would be necessary to discuss the data available for the ALS gene family of C. albicans; thus, we will try here to emphasize just two particularly interesting aspects of this family: the high potential for variation among the ALS gene sequences and the role of the ALS genes in virulence (see “GPI-Anchored Proteins, Cell Walls, and Virulence” below).

Looking closely at the sequences of all of the ALS proteins reveals that there is a high level of similarity among ALS family members both in nucleotide and corresponding peptide sequences, a similarity that also includes the upstream and downstream regions of the ORF. For example, sequences of ALS2 and ALS4 appear to be extremely close, as well as sequences of ALS1, ALS3, and ALS5.

An interesting point raised by several recent papers is the variation in ALS sequences between alleles of the same gene or in ALS sequences of the same gene among strains. In 2003, Zhang et al. studied ALS7 in 66 different strains of C. albicans and found 60 different forms of ALS7 generating 49 different genotypes (62). The domains of high variation in the corresponding proteins are the middle stretch of tandem repeats and the C-terminal portion containing VASES repeats. An analysis of the genotype data in parallel with the infection caused by the strains revealed that isolates representing a more pathogenic general-purpose genotype cluster tend to have more tandem repeats than other strains. In another study we learned that, in addition to differences in the sequences of the highly varied domain of tandem repeats, the two alleles of the same ALS gene (here ALS9) vary in the 5′ part of the ORF sequence (66).

There is also evidence suggesting that variation in the intragenic repeat numbers provides the cell a way to escape host defenses. Verstrepen and coworkers studied the variation gen-
| Protein | Gene number$^a$ | Identified homologue(s)$^c$ | % Identity/% similarity$^d$ | Localization$^e$ | Regulator(s)$^f$ | Virulence of deletion mutant | Predicted function |
|---------|----------------|-----------------------------|-----------------------------|------------------|------------------|-----------------------------|------------------|
| Als1    | orf6 1115 5741| ScSag1                      | 26/43                       | Biofilm formation, Rfg1, Ssk1, growth | Defective       | Involved in cell-cell adhesion |
| Als5    | orf19 4915 5736| ScSag1                      | 27/45                       | ND (no mutant described) | By homology, involved in cell-cell adhesion |
| Als6    | orf19 8574 7414| ScSag1                      | 26/44                       | ND (no mutant described) | By homology, involved in cell-cell adhesion |
| Cht2    | orf6 2344 3895| ScCts1                      | 43/55                       | Caspofungin, yeast-form cells, Cyr1, Efg1, pH | ND (not tested) | By homology, chitinase |
| Crh11   | orf19 1231 2706| ScCrh1                      | 56/71                       | Caspofungin       | ND (no mutant described) | By homology, cell wall glycosidase of the cell wall |
| Csa1    | orf6 8439 7114| None                        | CW (V)                      | Morphogenesis switch, Rim101, Efg1, Cph1 | ND (not tested) | Surface antigen on elongating hyphae and buds |
| Dfg5    | orf6 7480 2075| ScDew1                      | 50/70                       | Caspofungin, morphogenesis switch | ND (not tested) | Unknown |
| Ecm331  | orf6 3969 4255| ScEcm33                     | 33/52                       | Caspofungin, Plc1, Rim101, Hog1, Cyr1, Sn6 | ND (no mutant described) | Unknown |
| Exg2    | orf6 6664 2952| ScExg1                      | 42/57                       | Cell wall regeneration | ND (no mutant described) | Putative exo-1,3-
β-glucosidase by homology |
| Hwp1    | orf6 4883 1321| None                        | CW (I)                      | Morphogenesis, opaque and cell specific | ND (not tested) | Hyphal cell wall protein; covalently cross-linked to epithelial cells |
| Hyr1    | orf6 857 4975| None                        | CW (N, I)                   | Morphogenesis switch, Rfg1, Efg1, Nrg1, Tup1, Cyr1 | ND (not tested) | Unknown |
| Hyr3/Iff2| orf6 4725 575 | None                        | CW (I)                      | ND (no mutant described) | Unknown |
| Iff3    | orf6 2179 4361| None                        | CW (Y, I)                   | ND (no mutant described) | Unknown |
| Iff4    | orf6 8724 7472| None                        | CW (I)                      | ND (not tested) | Unknown |
| Iff5    | orf6 297 2879| None                        | CW (I)                      | ND (no mutant described) | Unknown |
| Iff6    | orf6 2490 4072| None                        | White-opaque switching      | ND (no mutant described) | Unknown |
| Hyr4/Iff7| orf6 8279 3279| None                        | CW (N)                      | Rim101 | ND (no mutant described) | Unknown |
| Mid1    | orf6 496 3212| ScMid1                      | 36/55                       | Mb (K, K)        | ND (no mutant described) | By homology, involved in Ca$^{2+}$ influx during mating |
| Pga1    | orf6 8973 7625| ScKrc1                      | 33/57                       | Cell wall regeneration | ND (no mutant described) | By homology, cell wall glycoprotein involved in β-glucan assembly |
| Pga2/Sod4| orf6 7493 2062| D. hansenii protein 41/61   | White-opaque switching, caspofungin | ND (no mutant described) | Superoxide dismutase domain |
| Pga3/Sod5| orf6 7495 2060| D. hansenii protein 41/58   | Neutrophil contact, morphogenesis, caspofungin | Defective | Superoxide dismutase domain |
| Gene | Start | End | Protein | Length (K, K) | Phenotype | Notes |
|------|-------|-----|---------|--------------|-----------|-------|
| Pga4/Gas1 | 7448 | 4035 | *D. hansenii* protein | 55/73 Mb | Mb (K, K) | ND (no mutant described) |
| Pga6/Flo9 | 4590 | 4765 | ScFlo9 | 28/45 | Iron, Als2 | ND (no mutant described) |
| Pga10/Rbr51 | 6914 | 5674 | *D. hansenii* protein | 46/58 | Ketoconazole, morphogenesis, Rim101 | ND (not tested) |
| Pga11 | 8957 | 7609 | None | | | Iron uptake using hemoglobin |
| Pga13 | 7834 | 6420 | None | | CW (I) | ND (no mutant described) |
| Pga14 | 8217 | 968 | ScYnl190w | 46/62 | Mb (K, R) | Unknown |
| Pga15 | 4387 | 2878 | None | | Mb (K) | Unknown |
| Pga17 | 7284 | 893 | None | | | Unknown |
| Pga18 | 4292 | 301 | None | | CW (V) | Nrg1, Tup1 |
| Pga20/Rbr1 | 450 | 535 | None | | Mb (K) | Rim101, Nrg1, Rim101, white-opaque switching |
| Pga21/Rbr2 | 6744 | 532 | None | | | Unknown |
| Pga22 | 5468 | 3738 | None | | | Unknown |
| Pga23 | 83 | 3740 | None | | | Rim101, Cyr1, Ssn6 |
| Pga24/YWP1 | 3288 | 3618 | *D. hansenii* protein | 34/48 | Mb (K, K) | Ssk1, Ssn6, Efg1, Efhl, phosphate |
| Pga26 | 4960 | 2475 | None | | CW (V) | Iron, cell wall regeneration |
| Pga28 | 311 | 5144 | None | | | Unknown |
| Pga29/Rhd3 | 8294 | 5305 | None | | CW (V) | Iron, morphogenesis switch |
| Pga30 | 8292 | 5303 | None | | CW (Y, V) | Unknown |
| Pga33 | 7301 | 876 | None | | | Unknown |
| Pga34 | 4494 | 2833 | None | | Mb (K, K) | ND (no mutant described) |
| Pga36/Ihd1 | 6198 | 5760 | None | | pH, morphogenesis switch, Nrg1, Rtg1, Tup1, Tsa1p, Tsa1Bp | ND (no mutant described) |
| Pga37 | 4053 | 3923 | None | | | Unknown |
| Pga39 | 5495 | 6302 | None | | | Unknown |

*Continued on following page*
| Protein       | Gene number<sup>a</sup> | Identified homologue(s)<sup>c</sup> | % Identity<sup>b</sup> / % similarity<sup>d</sup> | Localization<sup>e</sup> | Regulator(s)<sup>f</sup> | Virulence of deletion mutant | Predicted function                                      |
|---------------|--------------------------|-------------------------------------|-----------------------------------------------|--------------------------|--------------------------|-----------------------------|----------------------------------------------------------|
| Pga41         | 6116 2906                | None                                |                                               | Mb (K)                    | Overexpression of CDR1 and CDR2 or MDR1 | ND (no mutant described)                                      | Unknown                                                  |
| Pga42         | 6115 2907                | None                                |                                               | Mb (K)                    | ND (no mutant described)                                      | Unknown                                                  |
| Pga44         | 3113 1714                | None                                |                                               | CW (Y)                    | ND (no mutant described)                                      | Unknown                                                  |
| Pga45         | 1247 2451                | None                                |                                               | CW (V, I)                 | Caspofungin, Hog1, Ssn6                                       | ND (no mutant described)                                  | Unknown                                                  |
| Pga46         | 2753 3638                | None                                |                                               | Mb (K)                    | ND (no mutant described)                                      | Unknown                                                  |
| Pga47/Eap1    | 5354 1401                | None                                |                                               | Efg1                      | ND (no mutant described)                                      | Involved in binding human epithelial cells               |
| Pga48         | 5514 6321                | None                                |                                               | CW (V)                    | Iron                                                     | ND (no mutant described)                                  | Weak homologue of ScSed1 and ScSpi1                      |
| Pga49         | 2824 4404                | None                                |                                               | Mb (R, R, R)              | ND (no mutant described)                                      | Unknown                                                  |
| Pga50         | 3431 1824                | None                                |                                               | CW (I) and Mb (K)         | ND (no mutant described)                                      | Unknown                                                  |
| Pga52         | 799 1911 ScYj171c        | 52/68                               |                                               | Mb (K, K)                 | Fluconazole                                               | ND (no mutant described)                                  | Unknown                                                  |
| Pga53         | 305 4651                 | None                                |                                               | CW (I) and Mb (K)         | Cyr1                                      | ND (no mutant described)                                  | Unknown                                                  |
| Pga56         | N/D 1105.2               | None                                |                                               | CW (N)                    | Morphogenesis switch, cell wall regeneration, Ssn6         | ND (no mutant described)                                  | Unknown                                                  |
| Pga57         | 461 4689                 | None                                |                                               | Mb (R)                    | ND (no mutant described)                                      | Unknown                                                  |
| Pga58         | 1698 4334                | None                                |                                               | Ssn6                      | ND (no mutant described)                                      | Unknown                                                  |
| Pga59         | 6585 2767                | None                                |                                               | ND (no mutant described)                                      | Unknown                                                  |
| Pga60         | 4430 5588                | None                                |                                               | CW (N)                    | Morphogenesis switch                                       | ND (no mutant described)                                  | Unknown                                                  |
| Pga61         | 6196 5762                | None                                |                                               | CW (V)                    | ND (no mutant described)                                      | Unknown                                                  |
| Pga62         | 6587 2765                | None                                |                                               | Fluconazole, iron, cell wall regeneration, Cyr1, Ras1 | ND (no mutant described)                                  | Unknown                                                  |
| Phr1          | 7524 3829 ScGas1         | 55/69                               |                                               | Rim101                    | Defective by homology, \(\beta\)-1,3-galactosyltransferase, required for cell wall assembly |
| Phr2          | 1067 6081 ScGas2         | 59/69                               |                                               | Mb (K)                    | Rim101, fluconazole                                       | Defective by homology, \(\beta\)-1,3-galactosyltransferase, required for cell wall assembly |
| Pb3           | 6206 6594 ScPb1          | 53/67                               |                                               | Mb (K)                    | ND (no mutant described)                                      | By homology, phospholipase B                             |
| Pb5           | 4037 5102 ScPb3          | 46/63                               |                                               | Mb (K, K, K)              | ND (no mutant described)                                      | By homology, phospholipase B                             |
| Gene   | Accession | Description | Function | Requirement | Status          | Notes                                                                 |
|--------|-----------|-------------|----------|-------------|----------------|----------------------------------------------------------------------|
| Rbt5   | 4505      | None        | CW (V)   | RFg1, Rim101, Sfu1, Hog1, Tup1, serum, iron, ketoconazole | Normal          | Unknown                                                              |
| Sap9   | 7314      | ScYap3      | 33/50    | Antifungal drugs, stationary phase, white phase | Defective      | By homology, aspartyl proteinase                                        |
| Ssr1   | 7956      | ScCcW14/ScSsr1 | 57/67   | Cell wall regeneration, yeast form                | ND (not tested) | Required for cell wall integrity                                      |
| Cs4/Utr2 | 1639    | ScUtr2      | 42/58    | ND (not tested)                                   | Defective      | By homology, 1,3-1,4-β-glucanase                                      |
| Pga7   | 4504      | D. hansenii protein | 51/63   | Rim101, ketoconazole, planktonic growth           | ND (no mutant described) | Unknown                                                               |
| Pga8/Hwp2 | 2933  | None        | CW (Y, I) | Ef1, Tup1                                          | ND (no mutant described) | By homology, extracellular α-1,4-glucan glucosidase                   |
| Pga38  | 1332      | None        | 26/36    | Cell wall regeneration                            | ND (no mutant described) | Unknown                                                               |
| Pga55  | 5166      | None        |          | Morphogenesis, Nrg1, Tup1                         | ND (no mutant described) | Unknown                                                               |
| Rbt1   | 4889      | None        | CW(I)    | Tup1, serum, morphogenesis, farnesol, alpha factor, Rim101 | Defective      | Unknown                                                               |
| Spr1   | 2395      | ScSpr1      | 41.57    | ND (not tested)                                   | ND (no mutant described) | By homology, exo-1,3-β-glucanase precursor                           |

Proteins specific to de Groot et al. study

| ALS9   | 2924      | None        |          | ND (no mutant described)                          | ALS9           | By homology, cell-cell adhesion                                       |
| ALS2   | ND        | ScSag1      | 26/45    | Morphogenesis switch, ketoconazole, iron, cell wall regeneration | Protein essential? | By homology, cell-cell adhesion                                      |
| ALS3   | 1614      | ScSag1      | 27/44    | Morphogenetic switch, pH, Nrg1, Eff1, Tup1 Vaginal contact | Defective      | By homology, cell-cell adhesion                                      |
| ALS4   | 3075      | ScFlo1      | 22/39    | Biofilm formation                                 | Normal         | By homology, cell-cell adhesion                                      |
| ALS7   | 8560      | ScSag1      | 23/40    | ND (not tested)                                   | Unknown        | By homology, cell-cell adhesion                                      |
| Cht1   | 8769      | ScCts1      | 51.63    | Mb (K, R)                                         | ND (not tested) | By homology, chitinase                                               |
| Ecm33  | ND        | ScEcm33     | 29/43    | Mb (K, K)                                        | ND (no mutant described) | Unknown                                                               |
| Iff8   | 4720      | None        | CW (I)   | ND (no mutant described)                          | Unknown        | By homology, protein involved in 1,3-1,4-β-glucan biosynthesis       |
| Kre1   | 4542      | ScKre1      | ND (sequence too short) | CW (N) and Mb (K) | ND (no mutant described) | By homology, protein involved in β-1,3-Glucanosyl transferase         |
| Pga5   | 3873      | C. glabrata protein | 50/65   | Caspofungin                                      | ND (no mutant described) | Unknown                                                               |
| Pga9/Sod6 | 4753 | D. hansenii protein | 47/64 | Planktonic growth                                | ND (no mutant described) | Superoxide dismutase domain                                         |
| Pga12  | 8945      | D. hansenii protein | 44/59   |                                                | ND (no mutant described) | Unknown                                                               |

Continued on following page
| Protein       | Gene number<sup>a</sup> | Identified homologue(s)<sup>b</sup> | % Identity/% similarity<sup>c</sup> | Localization<sup>d</sup> | Regulator(s)<sup>f</sup> | Virulence of deletion mutant | Predicted function          |
|--------------|--------------------------|-----------------------------------|----------------------------------|---------------------------|-----------------------------|-----------------------------|-----------------------------|
| Pga16        | orf6 901 orf19 848       | None                              | ND (no mutant described)         | Fluconazole               | Fluconazole                 | ND (no mutant described)    | Unknown                     |
| Pga19        | orf6 5972 orf19 2033     | None                              | ND (not tested)                  | Fluconazole               | Fluconazole                 | ND (no mutant described)    | Unknown                     |
| Pga25        | orf6 5529 orf19 6336     | None                              | ND (no mutant described)         | Fluconazole               | Fluconazole                 | ND (no mutant described)    | Unknown                     |
| Pga27        | orf6 5983 orf19 2044     | None                              | Mb (K, K, K)                     | Fluconazole               | Fluconazole                 | ND (no mutant described)    | Unknown                     |
| Pga31        | orf6 3932 orf19 5302     | D. hansenii protein               | CW (V) and Mb (K, K)             | Fluconazole               | Fluconazole                 | ND (no mutant described)    | Unknown                     |
| Pga32        | orf6 8847 orf19 6784     | None                              | Mg (K, K, K)                     | Fluconazole               | Fluconazole                 | ND (no mutant described)    | Unknown                     |
| Pga35/Fgr41  | orf6 3757 orf19 4910     | D. hansenii protein               | CW (V) and Mb (K, K)             | Fluconazole               | Fluconazole                 | ND (no mutant described)    | Unknown                     |
| Pga40/Fgr23  | orf6 555 orf19 1616      | None                              | CW (N)                           | Alpha factor              | Alpha factor                | ND (no mutant described)    | Unknown                     |
| Pga43        | orf6 6112 orf19 2910     | Mb (K)                            | ND (no mutant described)         | White-opaque switch, cell wall regeneration | White-opaque switch, cell wall regeneration | ND (no mutant described)    | Unknown                     |
| Pga51/Dew1   | orf6 1293 orf19 1989     | ScDew1                            | CW (L)                           | Cyr1                      | Cyr1                        | ND (no mutant described)    | Unknown                     |
| Pga54        | orf6 4795 orf19 2685     | None                              | Hog1, Cyr1, Efg1, Ssn6           | Hog1, Cyr1, Efg1, Ssn6    | Hog1, Cyr1, Efg1, Ssn6      | ND (no mutant described)    | Unknown                     |
| Pb4.5        | orf6 1422 orf19 1442     | ScPb1                             | CW (V, V) and Mb (K, R)           | Hog1, Ssn6                | Hog1, Ssn6                  | ND (no mutant described)    | Unknown                     |

Proteins specific to Eisenhaber et al. study

| Protein | Gene number<sup>a</sup> | Identified homologue(s)<sup>b</sup> | % Identity/% similarity<sup>c</sup> | Localization<sup>d</sup> | Regulator(s)<sup>f</sup> | Virulence of deletion mutant | Predicted function          |
|---------|--------------------------|-----------------------------------|----------------------------------|---------------------------|-----------------------------|-----------------------------|-----------------------------|
| Crh1    | orf6 3505 orf19 3966     | D. hansenii protein               | 38/53                            | CW (I)                    | Nrg1, Tup1, Riml01, cell wall regeneration | ND (no mutant described)    | Unknown                     |
| Sap10   | orf6 7534 orf19 3839     | ScYap3                            | 25/38                            | Mb (K)                    | Iron                        | ND (no mutant described)    | Unknown                     |
| Iff9    | orf6 353 orf19 465       | None                              | CW (V, I)                        | Iron                      | Defective                   | ND (no mutant described)    | unknown by homology, aspartyl proteinase |
| Unknown | orf6 129 orf19 206       | None                              | CW (N, V)                        | Iron                      | Defective                   | ND (no mutant described)    | Unknown                     |
| Unknown | orf6 66 orf19 4652       | None                              | Mb (K)                           | Iron                      | Defective                   | ND (no mutant described)    | Unknown                     |
| Unknown | orf6 4704 orf19 4653     | None                              | Mb (K)                           | Iron                      | Defective                   | ND (no mutant described)    | Unknown                     |
| Unknown | orf6 4705 orf19 4654     | None                              | Mb (K)                           | Iron                      | Defective                   | ND (no mutant described)    | Unknown                     |
| Unknown | orf6 7864 orf19 6390     | None                              | Mb (K)                           | Iron                      | Defective                   | ND (no mutant described)    | Unknown                     |
| Unknown | orf6 8030 orf19 7130     | None                              | CW (Y) and Mb (R, R)             | Iron                      | Defective                   | ND (no mutant described)    | Unknown                     |
Proteins specific to *D. hansenii*

| Description | Function | Case | Notes |
|-------------|----------|------|-------|
| Unknown     | Unknown  | ND (not tested) | |
| Unknown     | Unknown  | ND (no mutant described) | |
| Unknown     | Unknown  | ND (no mutant described) | |
| Unknown     | Unknown  | ND (no mutant described) | |
| Unknown     | Unknown  | ND (not tested) | |
| Unknown     | Unknown  | None | |
| Unknown     | Unknown  | None | |
| Unknown     | Unknown  | None | |
| Unknown     | Unknown  | None | |

This final list has been obtained by compiling data from three different in silico studies of the *C. albicans* genome, those from de Groot et al., Eisenhaber et al., and Garcia et al. (12, 16, 21). Boldface indicates a protein with a function related to cell-to-cell adhesion; italics indicate a protein with a function related to cell wall biogenesis or remodeling. ND, not determined.

The numbers in the orf6 column are the different numbers given to the same sequences through the steps of the *C. albicans* genome sequencing project.

The localization prediction: CW, cell wall; Mb, membrane. The residues found in the sequence that permitted the prediction are displayed in parentheses.

Data were extracted from transcriptional analyses, mainly using the *Candida* genome database website (http://www.candidagenome.org/).

The largest family present in this GpiP gene class, with 12 genes, is the group of genes orthologous to *HYR1* (for hyphally regulated), also known as the *IFF* genes (for individual protein family F). As we noted above, 10 genes are predicted to have GpiP anchors and two genes, *IFF10* and *IFF11*, do not have any signal sequences for Gpi anchor linkages. Very little is known about this family; only one member has been studied to date: *HYR1* (3). This gene is strongly induced under conditions that trigger the morphogenetic switch. Mutation, however, did not show a strong phenotype under the conditions tested. Analysis of the coding sequences in this family revealed gene products with a large N-terminal domain of 340 amino acids shared by each member of the family with high homology. The proteins diverge strongly after this domain, both in size and sequence. Phylogenetic analysis of this family confirmed that the two non-GPI-anchored proteins Iff10 and Iff11 diverge obviously from the original group but the others remain united. Interestingly, comparison with *Candida dubliniensis* sequences shows that *C. albicans* has a subgroup of *IFF2*, *3*, *4*, and *5* while *C. dubliniensis* has only one orthologue (Fig. 3). The tree representation suggests either that (i) *C. dubliniensis* has lost *IFF2*, *3*, and *4* orthologues but kept the *IFF5* orthologue or that (ii) in *C. albicans*, amplification of a gene from an *IFF5* ancestor occurred. In either case, such differences might be one of the reasons explaining the difference in virulence levels of these two very close species (54). One can envision, for instance, that these proteins change the adherence properties of the cells and thus modify the capacity to colonize a specific environment.

All these proteins possess serine/threonine-rich regions accounting for more than 30% of the sequence, as well as 8 to 48 N glycosylation sites; these characteristic suggest that these proteins are likely to be transferred to the cell wall. No function so far has been suggested for any of the Iff proteins, but
the presence of the genes in 12 copies in the genome implies that they perform a useful function in terms of C. albicans biology, an idea reinforced by the fact that S. cerevisiae has no clear orthologue.

**GPI-ANCHORED PROTEINS PRESENT AT THE CELL SURFACE**

GpiPs in eukaryotes are largely thought to be localized only to the plasma membrane, although an alternate hypothesis (lacking direct experimental evidence thus far) is that a few of them may stay or be redirected to other cell compartments. A recent study of membrane lipid rafts in C. albicans shows that (lacking direct experimental evidence thus far) is that a few of the 115 proteins (46 proteins, or 40%) such an analysis of C. albicans genome sequences.

**TABLE 2. Families of genes corresponding to GpiPs of C. albicans**

| Family | Member genes |
|--------|--------------|
| 1      | orf19.4652, orf19.4653 |
| 2      | CHT1, CHT2, CHT3 |
| 3      | DPG5, PGA51/DCW1 |
| 4      | ECM331, ECM33.3 |
| 5      | EXG2, SPR1, EXG1 |
| 6      | PGA37, PGA57 |
| 7      | PGA59, PGA62 |
| 8      | SAP9, SAP10, SAP1, SAP2, SAP3, SAP4, SAP5, SAP6, SAP7, SAP8 |
| 9      | CRH1, UTR2, CRH11 |
| 10     | PGA2/SOD4, PGA3/SOD5, PGA4/SOD6 |
| 11     | PGA15, PGA41, PGA42 |
| 12     | HWP1, RBT1, PGA8/HWP2 |
| 13     | PGA30, PGA31, PGA32, orf19.1745 |
| 14     | PLB3, PLB4, PLB5, PLB1, PLB2 |
| 15     | CSA1, PGA7, PGA10/RBT8, RBT5, orf19.3117 |
| 16     | PGA4/GAS1, PGA5/GAS2, PHR1, PHR2, PHR3 |
| 17     | ALS1, ALS2, ALS3, ALS4, ALS5, ALS6, ALS7, ALS9 |
| 18     | HYR1, IFF1, IFF2, IFF3, IFF4, IFF5, IFF6, IFF7, IFF8, IFF9, IFF10, IFF11 |
| 19     | PGA52, TOS1 |
| 20     | PGA24, orf19.3621 |
| 21     | orf19.4014, CDC1 |
| 22     | KRE1, KRE2 |
| 23     | orf19.220, orf19.2783 |
| 24     | YCK3, YCK2, HRR25 |
| 25     | PGA12, orf19.7596, orf19.6809, orf19.2607 |

*Genes corresponding to predicted GpiPs are in bold; the other genes listed correspond to proteins that are not GPI anchored. Families have been classified according to a BLASTP analysis of C. albicans genome sequences.

Among the 29 proteins identified only 2 were GpiPs, Ecm33 and a Gas1 homologue (29), which seems surprisingly low compared to what was expected considering the results obtained with higher eukaryotes (48). A study of Cryptococcus neoformans, however, highlights the notion of enrichment with proteins in lipid rafts. Siafakas and coworkers also found only two GpiPs, but they were in much greater concentrations in lipid rafts than in the membrane (50). No quantitative data are yet available for C. albicans lipid rafts, but the situation might be comparable to that of C. neoformans: high concentration but low diversity.

On the other hand, it has been shown with several fungi, S. cerevisiae and C. albicans in particular, that some of the GpiPs are released from the GPI anchor and are covalently bound to cell wall sugars, thus being only transiently localized at the plasma membrane (9). In S. cerevisiae, this cleavage event occurs between the N-acetylgalactosamine group and the first mannose group of the anchor (Fig. 1); the protein with the GPI anchor remnant section is then transported to the cell wall network and linked to the β-1,6-glucans by an unidentified protein or protein complex. Figure 4 summarizes the different stages of synthesis and transport of a fungal GPI-anchored protein.

Based on studies with S. cerevisiae, there are few known sequence features common among the cell wall GpiPs. Two kinds of signal sequences for cellular localization in S. cerevisiae have been proposed: (i) the specific amino acid residues V, I, or L at the site ω-4 or ω-5 of the GPI attachment site (the ω site) and Y or N at the site ω-2 for cell wall localization and (ii) dibasic residues, R and K, for instance, in the region upstream of the ω site for plasma membrane localization (25). No similar studies have been done with C. albicans, given that we can extrapolate from our knowledge of S. cerevisiae. We applied these signal sequence rules to the 115 GpiPs presented earlier in the review and to the 58 putative GpiPs listed by Caro et al. (9) for S. cerevisiae (Table 3). These rules of localization give similar results for S. cerevisiae and C. albicans, with equal percentages of cell wall and membrane-bound proteins, although in C. albicans the percentage of unclassified proteins is much higher, comprising almost half of the 115 proteins (46 proteins, or 40%). Such an analysis of Candida, however, is not necessarily consistent with exper-
mental data. Dfg5, for example, contains several dibasic amino acids upstream of the predicted ω site. However, Spreghini et al. showed that Dfg5 is localized to the cell wall after cell fractionation and Western blotting (52). Again, if we try to predict the localization patterns of the 12 GpiPs identified in a proteomic analysis (see below), 2 are incorrectly predicted to be localized to the membrane, 4 are correctly predicted, and 6 are associated with no clear prediction (11). Thus, one cannot yet predict with confidence where a GpiP will be localized and experimental testing is still the only way to clearly know.

Another approach to establishing the localization of cell wall GpiPs is to directly extract the cell wall and subsequently identify the GpiPs present. This strategy was used by investigators in the F. Klis laboratory; they prepared C. albicans cell walls cleared of weakly bound or unbound cell wall proteins and digested the glucan network to free the covalently bound cell wall proteins (11). This protein mix was then subjected to liquid chromatography-tandem mass spectrometry analysis for protein identification. Interestingly, they identified only 14 proteins, 12 of which are GpiPs, supporting the idea that the “surfoproteome” is quite reduced. It is indeed highly surprising to identify 12 GpiPs on the cell wall when 115 are encoded by the genome, of which 30 to 50% are supposedly cell wall bound. There are at least three simple hypotheses to explain this finding. (i) The technique in itself needs a large amount of proteins in order to detect them and so the majority of the cell wall-bound proteins expressed at low levels are not distinguished from the background noise. (ii) The surfoproteome is highly dependent on environmental conditions, and so for each condition there is a specific pool of cell wall GpiPs. (iii) The last obvious hypothesis is that cell wall localization is a relatively rare event and that most GpiPs are indeed membrane bound. The first hypothesis is partially dismissed by de Groot and coworkers since they state that their technique allows them to identify cell wall proteins when there are about 500 molecules present per wall (~0.05% of the covalently bound cell wall molecules), which is a rather small amount of protein (11).

Concerning the second hypothesis, a study on the stability of cell wall protein pools was done in the laboratory of R. Sentandreu and the results support the idea that the turnover of cell wall proteins is very slow (47). Indeed, during the pulse-chase experiments conducted by Ruiz-Herrera et al., the investigators did not observe any protein recycling after 4.5 h of chase, suggesting that the proteins linked to the cell wall components are very stable. Considering the results reported in these two publications, it would be interesting to repeat the same type of experiments using a chase under completely different conditions. One may find a cell wall protein turnover that becomes induced by adaptations to external changes, such as pH or temperature.

**GPI-ANCHORED PROTEINS, CELL WALLS, AND VIRULENCE**

Cell surface proteins and virulence are often linked in the literature about both eukaryotes and prokaryote pathogens. To date, several GpiPs have been found to be involved in viru-
lence, but it is not clear whether the effects are direct or not, since subsequent molecular analysis is often missing.

A global survey of the 115 GpiP genes tells us that for 87 genes no mutant exists, 13 mutants have been constructed but not tested in virulence assays, and of the 15 mutants tested, 3 exhibit normal virulence and 12 have reduced virulence compared to that of the corresponding reference strain (Table 1). The sample is too small to allow statistical comparison but still suggests that GpiPs probably support one or several aspects of C. albicans virulence traits. The small number of genes tested makes the classification difficult, but the virulence defect of the mutants can be explained by very different reasons, usually indirect consequences of the gene deletion. We propose the following subdivision.

Cell wall biosynthesis and modeling. Phr1, Phr2, Utr2, Ecm33, and Cht1 are clearly linked to some aspect of the cell wall structure, which is critical in the cell cycle and often has important consequences for the capacity of C. albicans to switch from yeast to hyphae or vice versa, a well-known virulence trait. In addition, the modification of the cell wall structure or composition may change the surface epitope recognized by the host and thus modify the level of virulence.

Specific enzymatic functions. The two aspartyl proteinases Sap9 and Sap10 seem to affect virulence through indirect effects on more-essential GpiPs, as was observed for a gpi7 null mutant (45, 46) (Gpi7 is responsible for the addition of a side chain to the GPI core structure [Fig. 1]). Indeed, in contrast to Sap1 through Sap6, which digest proteins surrounding the cell, Sap9 and Sap10 are involved in protein processing and activation, thus regulating protein function. While a mutant lacking gpi7 secretes GpiPs into the medium, mutants lacking SAP9 and SAP10 do not process essential proteins and thus block the functions of these proteins, but the consequences are the same: a defect in cytokinesis and a disorganized cell wall network (2).

A recent study of the phospholipase B family from C. albicans (five members) showed that strains lacking the GpiP Plb5 have a reduced level of in vivo organ colonization (56). It seems, however, that Plb5’s role in pathogenesis is weaker than that of Plb1, a secreted non-GPI-anchored phospholipase B, since the phenotype is visible only in organ colonization for Plb5 (33, 40). The hypothesis for both proteins is that such secreted or cell-associated enzymes may be especially important while Candida is in close contact with host cells for the breakdown of the host cell membranes.

Rbt1 is one interesting target for further research, since the function of this protein is poorly understood, but its relation to infection is still unclear. Braun and coworkers suggest that it may interact with and impair the immune system of the host and thus the mutant lacking RBT1 would be cleared more easily during infection (8).

Oxidant adaptation. Superoxide dismutase appears to be involved in virulence by increasing C. albicans resistance to the immune system. Mutants lacking SOD5 are more sensitive to neutrophils (19) but not to macrophage phagocytosis (35). The attenuated in vivo virulence of a sod5/−/− mutant suggests that oxidative burst might be an important way to kill C. albicans and also confirms the crucial role of neutrophils in preventing candidiasis.

Adherence. Other than the Als proteins, Hwp1 is one of the most studied proteins and probably the protein that fits best into the current virulence models. Moreover, Hwp1 and Als proteins both appear to act on the same aspect of C. albicans.

![Diagram](https://example.com/diagram.png)
TABLE 3. Prediction of GpiP localization in *S. cerevisiae* and *C. albicans* based on the amino acid environment of the ω-minus region

| Yeast          | % of GpiPs with: |
|---------------|------------------|
|               | Cell wall        | Membrane localization | Both localization patterns | No particular residues for prediction of localization |
| *C. albicans* | 49               | 25                  | 6                       | 40                |
| *S. cerevisiae* | 45               | 45                  | 2                       | 7                 |

* The calculation of the percentages has been done using the 115 sequences (Table 1) from *C. albicans* and the 58 sequences from *S. cerevisiae* extracted from the study by Caro et al. (9). We followed the rules of prediction published by Hamada et al. (25).

biology; interactions with cells. The ability of *C. albicans* to adhere to various constituents plays an important role in host colonization and in the initiation and maintenance of an infection.

The function of Hwp1 in interacting with host cells has been clearly demonstrated, since Hwp1 was shown to be the substrate of human transglutaminases that enable *C. albicans* to be cross-linked to host epithelial cells (53). Defects in adherence in this case and maybe in other types of interactions with the host impair the virulence of the strains.

Among the eight ALS genes present in the *Candida* genome, only ALS1 (20), ALS3 (64), ALS4 (65), and ALS7 (42) have been disrupted; as for ALS2, the gene seems to be essential to *C. albicans* viability (65). In the work of Zhao et al. with ALS1, ALS3, and ALS9 in 2003, only heterozygotes were produced but the investigators did not report the phenotypes (66). Of the four ALS gene deletion mutants of *C. albicans*, the als4 null mutant has wild-type virulence, the als7 null mutant has not been tested, and the als1 and als3 null mutants are described as having reduced virulence but in particular models (20, 64). The strain lacking ALS3 is defective in adherence to both endothelial and epithelial cells, reducing the destruction of the tissue in an in vitro model (64). As for ALS1, the deletion strain shows slightly reduced virulence in a mouse model and reduced adherence on mouse tongues (1, 30). The relatively weak effect of deleting ALS1 and ALS3 may be explained by the redundancy of functions in the ALS family.

In each case, the virulence is tightly linked to a phenotype in adherence tests; only als1, als3, and als4 mutants have been tested for adherence and were found to be defective, thus confirming the hypothesis that at least part of the function of the ALS family is in cell-cell adherence (20, 64, 65). Strategies other than deletion were used to confirm the involvement of ALS proteins in the adhesion phenomenon. Als5 has been the subject of several studies in which the protein was expressed in *S. cerevisiae*, resulting in binding to a variety of substrates, extracellular matrix-coated beads, human buccal cells, epithelial cells, and various peptides, etc. (22, 23, 32). A recent study deciphered the role of Als protein domains by again using Als5 and *S. cerevisiae*. Raueo and coworkers concluded that the N-terminal region contributes to cell adhesion activity through increased binding to fibronectin and that the central domain constituted of threonine-rich tandem repeats participates in cell-cell interactions (44). Finally, recent work confirms the role of Als3 in adhesion and in particular its crucial role during biofilm formation; in these studies, the investigators interestingly developed in vitro and, more importantly, in vivo models of biofilm formation (41, 43, 63).

**CONCLUSION**

Gpi-anchored proteins in *C. albicans* represent an extraordinary group of proteins localized to the cell surface. A variety of phenomena occur at the plasma membrane or at the outer surface of a cell, including molecule and ion transport, ligand-receptor interactions, protein activation by enzymatic processing, glucan biosynthesis and destruction, and many types of host-pathogen interactions involving proteins and glucans. A thorough analysis of this special class of proteins would not have been possible without all of the recently published data and the availability of the sequence of the *C. albicans* genome.

There are 70 Gpi-anchored proteins in *C. albicans* if we count only the ones in common among all the studies discussed or 115 if we combine the results these studies together. In both cases, the count is higher than that for *S. cerevisiae*, and it does not seem that this increase is only the consequence of the amplification of several gene families since between *S. cerevisiae* and *C. albicans* GpiP genes, families appear or increase in size but also disappear or reduce in size (for instance, FLO and TIR families in *S. cerevisiae* do not exist in *C. albicans*). This variation suggests that several unique attributes of *C. albicans* are likely to be encoded by these genes. This idea is reinforced by the high percentage of genes with unknown functions in this group, more than 65% of the group of 115 (and almost 65% of the group of 70); for *S. cerevisiae*, for instance, of the 58 predicted GpiPs (9), we find only 36% of these proteins to be without a known function by using *Saccharomyces* genome database tools (http://www.yeastgenome.org/). This comparison between *C. albicans* and *S. cerevisiae* has to be carefully used since it seems that the percentage of GpiPs of unknown function varies among fungi, with 60% for *Aspergillus nidulans*, 58% for *Neurospora crassa*, and 39% for *Schizosaccharomyces pombe* (16). It is interesting that among the proteins of unknown function, some are *C. albicans* specific and some are fungus specific; in either case, these proteins are of particular interest. The data set differs greatly in size and composition from *S. cerevisiae* GpiP lists, so the probability of finding targets original and specific to fungi increases. It is thus likely that by studying *C. albicans* GpiPs we will discover genes with functions that interest other scientific communities specializing in fungal human or plant pathogens like *Aspergillus, Cryptococcus, Magnaporthe*, or *Ustilago*, genes that we would not have unraveled by studying exclusively the model yeast *S. cerevisiae*.

In this review, we have shown the high heterogeneity of GpiPs in predictability of function, gene family arrangement, and cellular localization. The only general conclusion that we can draw so far is that GpiPs seem to be involved directly or indirectly in virulence through their functions in cell adhesion/ cell recognition or cell wall biosynthesis and remodeling. Thus, there is still a lot to learn from the Gpi-anchored proteins that may provide insight into the diagnosis and treatment of *C. albicans* infections. The recent publication on an Als1-based vaccine (27, 51) may give us a little glimpse into what the future will hold for us.
ACKNOWLEDGMENTS

We thank Cécile Névcéguise for assistance with the manipulation of genome sequences of C. albicans and phylogenetic tree construction. We are grateful to Dr. Nobile for her critical comments on the manuscript. We also thank the reviewers for their comments that greatly helped the improvement of our review.

Support for this research was provided by the INRA, and Armel Plaine has an MNERT fellowship funded by the French government.

REFERENCES

1. Alberti-Segui, C., A. J. Morales, H. Xing, M. M. Kessler, D. A. Williams, K. G. Weinstock, G. Cottearel, K. Fechtel, and B. Rogers. 2004. Identification of potential cell-surface proteins in Candida albicans and investigation of the role of a putative cell-surface glycosidase in adhesion and virulence. Yeast 21:285–302.

2. Albrecht, A., A. Felk, I. Pichova, J. R. Naglik, M. Schaller, P. de Groot, D. Maccallum, F. C. Odds, W. Schafer, F. Klis, M. Monod, and B. Hube. 2006. Glycosylphosphatidylinositol-anchored proteases of Candida albicans target proteins necessary for both cellular processes and pathogen-host interaction. J. Biol. Chem. 281:685–694.

3. Bailey, D. A., P. J. Feldmann, M. Bovey, N. A. Gow, and A. J. Brown. 1996. The Candida albicans HVR1 gene, which is activated in response to hyphal development, belongs to a gene family encoding yeast cell wall proteins. J. Bacteriol. 178:6533–6540.

4. Bensen, E. S., S. J. Martin, M. Li, J. Berman, and D. A. Davis. 2004. Transcriptional profiling in Candida albicans reveals new adaptive responses to extracellular pH and functions for Rim101p. Mol. Microbiol. 54:1335–1351.

5. Borner, G. H., D. J. Sherrier, T. J. Stevens, I. T. Arkin, and P. Dupree. 2002. Prediction of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A genomic analysis. Plant Physiol. 129:486–499.

6. Boutin, J. A. 1997. Myristoylation. Cell. Signal. 9:15–35.

7. Bowman, S. M., A. Piwowar, M. Al Dabbous, J. Vierula, and S. J. Free. 2006. Amino acid residues in the omega-minus region participate in cellular localization of yeast glycosylphosphatidylinositol-anchored proteins. J. Bacteriol. 188:3886–3889.

8. De Bernardis, F., F. A. Muhlschlegel, A. Cassone, and W. A. Fonzi. 1997. In silico identification of glycosyl-phosphatidylinositol-anchored (GPI) anchor pathway. Yeast 13:1477–1489.

9. De Bernardis, F., A. M. de Groot, G. van den Hoon, F. Klis, and W. A. Fonzi. 1998. The pH of the host niche controls gene expression in and virulence of Candida albicans. Infect. Immun. 66:3317–3325.

10. de Groot, P. W., A. de Boer, J. Cunningham, H. L. Dekker, L. de Jong, K. J. Hellingwerf, C. de Koster, and F. Klis. 2004. Proteomic analysis of Candida albicans cell wall reveals covalently bound carbohydrate-active enzymes and adhesins. Eukaryot. Cell 3:385–395.

11. de Groot, P. W., K. J. Hellingwerf, and F. P. Klis. 2003. Genome-wide identification of fungal GPI proteins. Yeast 20:787–796.

12. Dohlmann, H. G. 2002. G proteins and pheromone signaling. Annu. Rev. Physiol. 64:129–152.

13. Eisenbarth, B., P. Bork, and F. Eisenhaber. 2001. Post-translational GPI lipid anchor modification of proteins in kingdoms of life: analysis of protein sequence data from complete genomes. Protein Eng. 14:17–25.

14. Eisenbarth, B., S. Maurer-Stroh, M. Novatchkova, G. Schneider, and F. Eisenhaber. 2003. Enzymes and auxiliary factors for GPI lipid anchor modification of proteins in kingdoms of life: analysis of protein sequence data from complete genomes. Yeast 19:147–163.

15. Eisenbarth, B., G. Schneider, M. Wildpaner, and F. Eisenhaber. 2004. A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for Aspergillus nigerans, Candida albicans, Neurospora crassa, Saccharomyces cerevisiae and Schizosaccharomyces pombe. Mol. Biol. J. 37:243–253.

16. Ferguson, M. A. 1999. Structure, biosynthesis and functions of glyco-sphingolipidphosphatidylinositol anchors, and the contributions of trypanosome research. J. Cell Sci. 112:2799–2809.

17. Fonzi, W. A. 1999. PHR1 and PHR2 of Candida albicans encodes putative glycosidases required for proper cross-linking of beta-1,3- and beta-1,6-glycans. J. Bacteriol. 181:7070–7079.

18. Fu, A. S., D. C. Sheppard, Y. C. Chen, S. W. French, J. E. Cutler, S. G. Filler, and J. E. Edwards. Jr. 2002. Candida albicans Als1p: an adhesin that is a downstream effector of the EFG1 filamentation pathway. Mol. Microbiol. 44:61–72.
46. Richard, M., S. Ibata-Ombetta, F. Dromer, F. Bordon-Pallier, T. Jouault, and C. Gaillardin. 2002. Complete glycosylphosphatidylinositol anchors are required in Candida albicans for full morphogenesis, virulence and resistance to macrophages. Mol. Microbiol. 44:841–853.

47. Ruiz-Herrera, J., A. I. Martinez, and R. Sentandreu. 2002. Determination of the stability of protein pools from the cell wall of fungi. Res. Microbiol. 153:373–378.

48. Sharam, F. J., and M. T. Lehto. 2002. Glycosylphosphatidylinositol-anchored proteins: structure, function, and cleavage by phosphatidylinositol-specific phospholipase C. Biochem. Cell Biol. 80:535–549.

49. 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.

50. Siafakas, A. R., L. C. Wright, T. C. Sorrell, and J. T. Djordjevic. 2006. Lipid rafts in Cryptococcus neoformans concentrate the virulence determinants phospholipase B1 and Cu/Zn superoxide dismutase. Eukaryot. Cell 5:488–498.

51. Spellberg, B. J., A. S. Ibrahim, V. Avenissian, S. G. Filler, C. L. Myers, Y. Fu, and J. E. Edwards, Jr. 2005. The anti-Candida albicans vaccine composed of the recombinant N terminus of Als1p reduces fungal burden and improves survival in both immunocompetent and immunocompromised mice. Infect. Immun. 73:6191–6193.

52. Spreghini, E., D. A. Davis, R. Subaran, M. Kim, and A. P. Mitchell. 2003. Roles of Candida albicans Dfg5p and Dcw1p cell surface proteins in growth and hypha formation. Eukaryot. Cell 2:746–755.

53. Staab, J. F., S. D. Bradway, P. L. Fidel, and P. Sundstrom. 1999. Adhesive and mammalian transglutaminase substrate properties of Candida albicans Hwp1. Science 283:1535–1538.

54. Sullivan, D., and D. Coleman. 1997. Candida dubliniensis: an emerging opportunistic pathogen. Curr. Top. Med. Mycol. 8:15–25.

55. Sundstrom, P. 2002. Adhesion in Candida spp. Cell. Microbiol. 4:461–469.

56. Theiss, S., G. Ishdorj, A. Brenot, M. Kretschmar, C. Y. Lan, T. Richterlein, J. Hacker, S. Nigam, N. Agabian, and G. A. Kohler. 2006. Inactivation of the phospholipase B gene PLB5 in wild-type Candida albicans reduces cell-associated phospholipase A(2) activity and attenuates virulence. Int. J. Med. Microbiol. 296:405–420.

57. Thomas, J. R., R. A. Dwek, and T. W. Rademacher. 1990. Structure, biosynthesis, and function of glycosylphosphatidylinositol. Biochemistry 29:5413–5422.

58. Tiede, A., I. Bastisch, J. Schubert, P. Orlean, and R. E. Schmidt. 1999. Biosynthesis of glycosylphosphatidylinositol in mammals and unicellular microbes. Biol. Chem. 380:503–523.

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

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

61. Weig, M., L. Jansch, U. Gross, C. G. De Koster, F. M. Klis, and P. W. De Groot. 2004. Systematic identification in silico of covalently bound cell wall proteins and analysis of protein-polysaccharide linkages of the human pathogen Candida glabrata. Microbiology 150:3129–3144.

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

63. Zhao, X., K. J. Daniels, S. H. Oh, C. B. Green, K. M. Yeater, D. R. Soll, and L. L. Hoyer. 2006. Candida albicans Als3p is required for wild-type biofilm formation on silicone elastomer surfaces. Microbiology 152:2287–2299.

64. Zhao, X., S. H. Oh, G. Cheng, C. B. Green, J. A. Noessens, K. Yeater, R. P. Leng, A. J. Brown, and L. L. Hoyer. 2004. ALS3 and ALS8 represent a single locus that encodes a Candida albicans adhesin: functional comparisons between Als3p and Als1p. Microbiology 150:2415–2428.

65. Zhao, X., S. H. Oh, K. M. Yeater, and L. L. Hoyer. 2005. Analysis of the Candida albicans Als2p and Als4p adhesins suggests the potential for compensatory function within the Als family. Microbiology 151:1619–1630.

66. Zhao, X., C. Pujol, D. R. Soll, and L. L. Hoyer. 2003. Allelic variation in the contiguous loci encoding Candida albicans AL5, ALS1 and ALS9. Microbiology 149:2947–2960.