Microbiology (2004), **150**, 3129–3144 DOI 10.1099/mic.0.27256-0

Systematic identification *in silico* of covalently bound cell wall proteins and analysis of protein–polysaccharide linkages of the human pathogen *Candida glabrata*

Michael Weig,[1] Lothar Jänsch,[2] Uwe Groß,[1] Chris G. De Koster,[4] Frans M. Klis[3] and Piet W. J. De Groot[3,4]

Correspondence
Piet W. J. de Groot
pgroot@science.uva.nl

[1]Institute of Medical Microbiology and National Reference Center for Systemic Mycoses, University of Göttingen, Germany

[2]GBF, Braunschweig, Germany

[3,4]SILS-Microbiology[3] and SILS-Biomacromolecular Mass Spectrometry[4], University of Amsterdam, The Netherlands

Candida glabrata is an important cause of systemic candidiasis in humans. This paper reports a systematic analysis of the putative glycosylphosphatidylinositol-modified (GPI) proteins of *C. glabrata*, a large part of which are covalently bound to the cell wall glucan network and the remainder of which are retained in the plasma membrane, and of cell wall proteins (CWPs) which are covalently bound in a mild-alkali-sensitive manner. *In silico* genomic analysis revealed 106 putative GPI proteins. Fifty-one of these GPI proteins could be categorized as adhesive proteins, potentially implicated in fungus–host interactions or biofilm formation during the development of fungal infections. Eleven proteins belonged to well-known GPI protein families of glycoside hydrolases, probably involved in cell wall expansion and remodelling during growth. Other identified GPI proteins included phospholipases, aspartic proteases, homologues of *Sc*Ecm33p and *Sc*Kre1p, and structural CWPs. Interestingly, the GPI algorithm predicted three orthologues of an abundant CWP in *S. cerevisiae*, Cwp1p, which is absent in *Candida albicans*. To evaluate the *in silico* predictions, isolated cell walls were extracted using HF-pyridine, which specifically cleaves phosphodiester bonds, to release GPI-CWPs. Immunological analysis of the extract using one-dimensional SDS-PAGE and anti-*Sc*Cwp1p antiserum indicated the presence of a Cwp1p homologue in *C. glabrata* cell walls. Further analysis by two-dimensional gel electrophoresis and electrospray ionization tandem mass spectrometry (ESI-MS/MS) confirmed the presence of two of the predicted Cwp1p proteins, Cwp1.1p and Cwp1.2p. Crh1p, a putative 1,3-β-glucan remodelling enzyme, was also identified. *In silico* genomic analysis further revealed five putative Pir proteins (Pir1–5p) and five members of the Bgl2 glycoside hydrolase family 17, belonging to a class of putative CWPs that can be extracted with NaOH. Immunological analysis of mild-alkali-extracted CWPs showed the presence of a *Sc*Pir2p homologue. Together, these experimental data and *in silico* predictions represent the first systematic analysis of the *C. glabrata* cell wall proteome.

Received 19 April 2004
Revised 6 August 2004
Accepted 12 August 2004

INTRODUCTION

The opportunistic yeast *Candida glabrata* is now the second most common *Candida* species isolated from the bloodstream of patients in the USA and accounts for approximately 18 % of all yeast blood culture isolates (Pfaller *et al.*, 2003b). *C. glabrata* infections are difficult to treat, as the organism is naturally less susceptible to fluconazole and amphotericin B compared to *Candida albicans* (Rex *et al.*, 2000; Pfaller *et al.*, 2002, 2003a). In the pleomorphic yeast *C. albicans*, virulence is related to its ability to switch between yeast and hyphal morphology (Lo *et al.*, 1997), and both yeast and hyphal specific cell wall proteins (CWPs) have been identified (Klis *et al.*, 2001). Although *C. glabrata* is capable of forming germ-like, nucleus-free tubes (Lachke *et al.*, 2002) and 'pseudohypha-like projections' at the

Abbreviations: 1D, one-dimensional; 2D, two-dimensional; CAI, codon adaptation index; ConA, concanavalin A; CWP, cell wall protein; ER, endoplasmic reticulum; ESI-MS/MS, electrospray ionization tandem mass spectrometry; GPI, glycosylphosphatidylinositol (-modified).

periphery of colonies grown on nitrogen-starvation medium (Csank & Haynes, 2000; Calcagno *et al.*, 2003), true hyphae have not been observed during infection. This suggests that, in contrast to *C. albicans*, pathogenicity of *C. glabrata* is solely mediated by the yeast growth form (Fidel *et al.*, 1999).

The first contact between pathogenic yeast cells and human tissue occurs at the cell surface and is attributed to fungal cell wall components. For the pathogen *C. albicans* and for the baker's yeast *Saccharomyces cerevisiae*, models that describe their cell wall structure and organization have been developed. The major components of their walls are the polysaccharides 1,3-β-glucan, 1,6-β-glucan and chitin, and two types of covalently incorporated CWPs, glycosylphosphatidylinositol-modified (GPI) proteins, and proteins with internal repeats (Pir proteins). These proteins undergo post-translational modifications during secretion and they are often heavily mannosylated. The inner part of the wall is a skeletal layer formed by covalently connected 1,3-β-glucan and chitin molecules to which 1,6-β-glucan molecules are attached that in turn tether GPI-CWPs. These GPI-CWPs have been associated with various functions, such as maintaining cell shape, limiting permeability, hydrophobicity, and cell wall biosynthesis and remodelling (Scherrer *et al.*, 1974; Zlotnik *et al.*, 1984; Van der Vaart *et al.*, 1995; Mouyna *et al.*, 2000; Rodriguez-Pena *et al.*, 2000). Pir proteins are directly linked to 1,3-β-glucan via a currently unknown linkage, which is sensitive to NaOH treatment (Mrsa & Tanner, 1999; De Groot *et al.*, 2004). In *C. albicans* several CWPs have been identified that mediate adhesion to host tissue and virulence (i.e. adhesins of the Als-family and Hwp1p) (Hoyer *et al.*, 1999; Hoyer, 2001; Sundstrom, 2002; Cheng *et al.*, 2003). Also, other functions related to *Candida* virulence, such as adhesion to plastics, biofilm formation and antigenicity, can be attributed to CWPs (Gomez *et al.*, 1996; Chaffin *et al.*, 1998; Nisini *et al.*, 2001; Li & Palecek, 2003; Garcia-Sanchez *et al.*, 2004). Recently, a genome-wide *in silico* survey identified 66 and 104 putative GPI proteins in *S. cerevisiae* and *C. albicans* (De Groot *et al.*, 2003), respectively. The final destiny of these fungal GPI proteins will be either the cell wall or the plasma membrane and is influenced by the protein sequence in the C-terminal region after GPI modification (Vossen *et al.*, 1997; Hamada *et al.*, 1998, 1999; Terashima *et al.*, 2003; Frieman & Cormack, 2003). However, the precise mechanism of cell wall incorporation of GPI proteins is still unknown. Besides GPI-CWPs, in both organisms a small family of Pir-CWPs was also identified. In *S. cerevisiae*, this family consists of four members which differ mainly in the number of internal tandem repeats. In *C. albicans*, two proteins have Pir protein characteristics, but only Pir1 was shown to be covalently bound to the cell wall (De Groot *et al.*, 2004; Martínez *et al.*, 2004). Recently, we have demonstrated that *C. albicans* Scw1p/MP65, a presumed 1,3-β-glucanase/1,3-β-glucanosyl transferase, is also (partly) covalently bound to the cell wall in a Pir-like manner (De Groot *et al.*, 2004).

For *C. glabrata*, much less is known with respect to the glycan and protein composition of the cell wall. Only a small number of GPI proteins have been identified in this organism so far: the Epa1p adhesin (Cormack *et al.*, 1999; Frieman *et al.*, 2002), Epa2–5p (De Las Penas *et al.*, 2003), Epa6–7p (NCBI database), and a family of three Gas/Phr orthologues (Weig *et al.*, 2001). Using the second release of the annotated *C. glabrata* proteome, which was made publicly available in June 2004 by the Génolevures consortium (http://cbi.labri.fr/Genolevures/elt/CAGL), we have now performed a systematic, genome-wide prediction of GPI proteins of *C. glabrata*. GPI proteins can be identified by screening the genome for C-terminal consensus sequences for GPI modification. Essentially, we have applied the same algorithm that was used to predict GPI proteins of *S. cerevisiae*, *C. albicans*, *Schizosaccharomyces pombe* and *Neurospora crassa* (De Groot *et al.*, 2003), but made a minor modification to make it more comprehensive. Additionally, putative Pir proteins were identified by screening for proteins with a characteristic pattern of four cysteine residues in the C-terminal region, whereas Bgl2 family members were identified by performing BLAST searches. In order to validate our *in silico* predictions, we also aimed to identify covalently linked CWPs from isolated cell walls using tandem mass spectrometry, and we studied the covalent linkages between CWPs and the cell wall polysaccharides.

METHODS

***In silico* analysis.** The second release of the *C. glabrata* proteome, which was made publicly available in June 2004, was retrieved from the Génolevures website (http://cbi.labri.fr/Genolevures/elt/CAGL). In this release, comprising 5272 translated ORF sequences, two systematic numbers are provided for each ORF. One is a systematic individual protein file (IPF) number, used during the annotation process; the other indicates the position of the ORF on artificially assigned chromosomes (numbered A–M), and can be used to link the ORFs to NCBI protein files. Pattern searches for proteins having a GPI protein consensus sequence or the Pir-specific four-cysteine domain at their C-terminus (Fig. 1) were performed using the program FUZZPRO from the EMBOSS software package at http://www.hgmp.mrc.ac.uk/Software/EMBOSS/. The selection criteria to identify GPI proteins are described in De Groot *et al.* (2003). In short, ORFs were screened for (1) the presence of a C-terminal, fungal-specific, consensus sequence for GPI modification, (2) the presence of an N-terminal signal peptide for translocation across the membrane of the endoplasmic reticulum, and (3) the absence of internal transmembrane domains (Fig. 1a). The C-terminal GPI-modification algorithm is described as follows: [GNSDAC]-[GASVIETKDLF]-[GASV]-X(4,19)-[FILMVAGPSTCYWN](10)>, in Prosite format, where > indicates the C-terminal end of the protein. To optimize this GPI algorithm, we allowed the presence of glutamine in the hydrophobic tail. The presence of an N-terminal ER-import signal was analysed using the program SignalP V3.0 at http://www.cbs.dtu.dk/services/SignalP-3.0/. The standardized threshold value for signal peptides in the two algorithms [SignalP-NN (Smean) and SignalP-HMM (Sprob)] used in SignalP V3.0 was 0·5. The presence of internal membrane domains and protein localization was analysed using TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) and PSORTII (http://psort.nibb.ac.jp). BLASTP searches were performed at NCBI (http://www.ncbi.nlm.nih.gov/blast). The results of

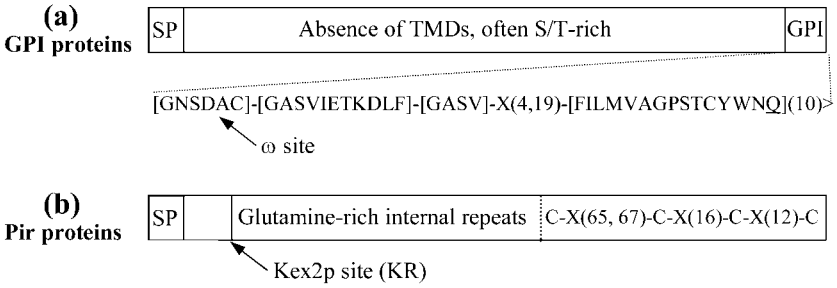

Fig. 1. Schematic structure of GPI proteins (a) and Pir proteins (b). C-terminal motifs are in Prosite format. The GPI algorithm was adapted from De Groot *et al.* (2003). A minor modification in that algorithm, concerning the possible presence of glutamine (Q) in the hydrophobic tail, is underlined. GPI modification (ω) and Kex2p cleavage sites are indicated by arrows. SP, signal peptide for translocation across the membrane of the endoplasmic reticulum; TMD, transmembrane domain; S/T, serine/threonine; GPI, C-terminal GPI-modification signal peptide.

our *in silico* GPI protein analysis are accessible at http://www.pasteur.fr/recherche/unites/Galar_Fungail/. The big-PI GPI protein prediction webserver is available at http://mendel.imp.univie.ac.at/gpi/fungi/gpi_fungi.html (Eisenhaber *et al.*, 2004). Global protein sequence identity levels were calculated with the Needleman–Wunsch method of DNA Strider 1.3 using default settings. Theoretical pI values, molecular masses and the serine and threonine content of proteins were also calculated using DNA Strider 1.3. Multiple protein alignments and an unrooted phylogenetic tree were created using the webserver CLUSTALW at http://www.ebi.ac.uk/clustalw/index.html. Codon adaptation index (CAI) values were determined with the codon usage measurement tool CAI Calculator 2 (UMBC, University of Maryland; http://www.evolvingcode.net/codon/CalculateCAIs.php) using *S. cerevisiae* as a template. The maximum value of the CAI is 1·0 and the minimum value is 0. Highly expressed proteins have CAI values greater than 0·500.

Yeast strains and growth conditions. *C. glabrata* wild-type stain ATCC 90876 was used in this study. The isolate was routinely cultured on YPD plates at 30 °C. For cell wall analysis yeast cells were grown at 30 °C in synthetic complete (SC) medium, containing 2 % (w/v) glucose, 1·1 % (w/v) Casamino acids (Difco), 0·8 % (w/v) yeast nitrogen base with ammonium sulphate without amino acids (Becton Dickinson), 110 μg ml^{-1} L-leucine (Sigma), 55 μg ml^{-1} L-tyrosine (Sigma), 55 μg ml^{-1} L-tryptophan (Sigma) and 55 μg ml^{-1} adenine sulphate (Fluka).

Isolation of cell walls. Cell walls of *C. glabrata* were isolated using a method that was previously described for *C. albicans* (Kapteyn *et al.*, 1995; De Groot *et al.*, 2004). Briefly, yeast cells were grown to exponential phase, washed and homogenized using glass beads in a FastPrep Instrument FP120 (Bio 101). The cell walls were isolated and washed extensively with 1 M NaCl. Subsequently, the cell walls were boiled twice for 5 min in the presence of 2 % SDS, 100 mM sodium EDTA, 40 mM β-mercaptoethanol and 50 mM Tris/HCl, pH 7·8, to remove any contaminants derived from the cytosol and/or plasma membrane from the walls (Mrsa *et al.*, 1997; Klis *et al.*, 1998). Purified walls were freeze-dried, weighed and stored at −20 °C until use.

Chemical and enzymic release of cell wall proteins. GPI proteins were extracted from isolated cell walls using hydrogen fluoride (HF)-pyridine as described previously (De Groot *et al.*, 2004). Briefly, freeze-dried walls were incubated with HF-pyridine at 0 °C for 3 h. The reaction was quenched by diluting the reaction mixture with an equal amount of ice-cold H_2O, followed by dialysis against H_2O. After centrifugation, the supernatant containing the GPI-CWPs was freeze-dried. Alternatively, GPI-CWPs were released from cell walls using purified recombinant *Trichoderma harzianum* endo-1,6-β-glucanase (Bom *et al.*, 1998) (0·8 U per g wet weight of cell walls) as described elsewhere (Kapteyn *et al.*, 1996, 1999). In order to extract alkali-sensitive CWPs, isolated cell walls were incubated with 30 mM NaOH for 17 h at 4 °C. The reaction was stopped by adding neutralizing amounts of acetic acid (Kapteyn *et al.*, 1999).

Analysis of cell wall proteins. Solubilized CWP fractions were separated by SDS-PAGE using a linear gradient of 2·2 % to 20 %. Proteins were stained either with silver (Bio-Rad) or colloidal Coomassie according to Neuhoff *et al.* (1988). For Western blot analysis, the proteins were transferred onto an Immobilon PVDF membrane (Millipore). To verify that the isolated cell wall fractions were not contaminated with cytosolic proteins after SDS/β-mercaptoethanol extraction, the PVDF membranes were probed with polyclonal antibodies raised against the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of *S. cerevisiae* (Gozalbo *et al.*, 1998) (results not shown). To visualize mannosylated proteins, PVDF membranes were probed with peroxidase-labelled concanavalin A (ConA; 1 μg ml^{-1}, Sigma) in phosphate-buffered saline (PBS), containing 3 % (w/v) BSA, 2·5 mM $CaCl_2$ and 2·5 mM $MnCl_2$, as described elsewhere (Klis *et al.*, 1998). We also used polyclonal antisera raised against the *S. cerevisiae* CWPs Pir2p (Russo *et al.*, 1992) and Cwp1p (Shimoi *et al.*, 1995), following the protocol described by Kapteyn *et al.* (1996, 1999). To increase the specificity of the antisera against Cwp1p and Pir2p, the membranes were pre-treated with 50 mM periodic acid and 100 mM sodium acetate (pH 4·5) for 30 min prior to the blocking step. To visualize binding of polyclonal antibodies to cell wall components, goat anti-rabbit IgG peroxidase secondary antibodies (Pierce) at a dilution of 1 : 10 000 in PBS/5 % (w/v) milk powder were used. Blots were developed using ECL Western blotting detection reagents (Amersham).

Two-dimensional (2D) gel electrophoresis. For first-dimension gel electrophoresis, precast immobilized pH-gradient (IPG) strips (pH 3–10) 18 cm long were used for isoelectric focusing (IEF). IEF was carried out in an IPGphor device (Amersham). Approximately 200 μg protein was loaded for colloidal Coomassie stain and 50 μg was loaded for silver staining. Protein concentrations were determined using the Bradford protein assay (Merck). Rehydration was done for 10 h (30 V) using 8 M urea, 2 % w/v CHAPS, 1 % IPG buffer (pH 3–10, Amersham) and 70 mM DTT in a volume of 360 μl at 20 °C. The following gradient was applied: 300 V for 3 h, 600 V for 3 h and 1000 V for 3 h. The final focusing step at 8000 V was extended to 100 000 Volt-hours (Vh). For second-dimension gel electrophoresis, gradient gels (8 % to 16 %) were applied. After treatment of the IPG strips in SDS equilibration buffer, electrophoresis was performed in a Hoefer DALT SE 600-system (Amersham). To determine the relative molecular masses of the proteins, a pre-stained broad-range Marker (Bio-Rad) was added in the second

dimension. Gels for mass spectrometry purposes were stained with Coomassie colloidal blue R250 as described by Neuhoff *et al.* (1988). The reproducibility of the protein profiles on 2D gels was evaluated by comparing HF-pyridine extracts from three independent cultures and cell wall preparations (data not shown).

Protein identification using electrospray ionization tandem mass spectrometry (ESI-MS/MS). Protein spots were excised from the gel and incubated once in 100 µl 25 mM NH_4HCO_3 and twice in a solution containing 50 % acetonitrile and 25 mM NH_4HCO_3 for 15 min. Subsequently, the excised gel pieces were dried at room temperature under reduced pressure. In-gel digestion was performed on the dried gel pieces by treatment with 20 to 40 µl trypsin (2 µg ml^{-1}) in 50 mM NH_4HCO_3 (sequencing grade; Promega) overnight at 37 °C. The supernatant was removed and stored, and the gel pieces were incubated in 30 to 60 µl 5 % formic acid for 30 min. The same volume of acetonitrile was added, and incubation was continued for 15 min. The supernatant was removed and pooled with the first supernatant, and the volume was reduced with a Speed Vac to approximately 10 µl. The peptides generated were desalted by using C18-ZipTip (Millipore). The purified peptides were eluted with 3 µl 65 % methanol/0·5 % formic acid.

Sequence data of internal peptides were acquired with a quadrupole time-of-flight mass spectrometer (Q-TOF-MS) by performing ESI MS/MS. For electrospray analysis and subsequent peptide sequencing, Au/Pd-coated nanospray glass capillaries (Protana) were filled with 3 µl ZipTip purified sample. The tip of the capillary was placed orthogonally in front of the entrance hole of a quadrupole time-of-flight mass spectrometry instrument (Q-TOF I; Micromass) equipped with a nanospray ion source. A capillary voltage between 750 V and 1000 V and a cone voltage of 35 V were applied. Doubly and triply charged peptides were chosen for collision-induced MS/MS fragmentation experiments and the corresponding parent ions were selectively transmitted from the quadrupole mass analyser into the collision cell. Argon was used as collision gas and the kinetic energy was set between 20 eV and 35 eV. The resulting daughter ions were separated by an orthogonal time-of-flight mass analyser. *De novo* peptide sequencing was carried out with the program PeptideSequencing within the BioLynx software (version 3.4; Micromass). Protein identifications were obtained by comparing the sequences of trypsin fragments with the proteins predicted from the *C. glabrata* genome sequence. For *in silico* digestion of proteins we used PeptideMass (http://ca.expasy.org/tools/peptide-mass.html).

RESULTS

In silico identification of GPI proteins

To make an inventory of the proteins that are putatively covalently linked to the cell wall of *C. glabrata*, we first searched the proteome of *C. glabrata* for putative GPI proteins, comprising the major class of CWPs and plasma-membrane proteins. For this, we utilized a selection method that was developed to identify GPI proteins from *S. cerevisiae*, *C. albicans*, *Sch. pombe*, and *N. crassa* (see Methods). Applying the GPI-modification algorithm described by De Groot *et al.* (2003) as the first selection step (Fig. 1a), we identified 163 candidates in the *C. glabrata* proteome. Of these preselected ORFs, 81 met the additional selection criteria (see Methods) and were therefore designated putative GPI proteins (Table 1). For two ORFs, the transmembrane (TM) domain predictions were not fully unambiguous because PSORTII indicated a possible internal TM domain whereas TMHMM did not. These ORFs, CAGL0I10362g/CAGL-IPF6446 and CAGL0J02552g/CAGL-IPF6528, have a serine/threonine content of 28 % and 27 %, respectively, have low pI values, and show similarity to the Flo adhesin GPI protein family of *S. cerevisiae* (Halme *et al.*, 2004) and to other *C. glabrata* adhesive proteins. Therefore, we have included them in our predicted GPI protein list. Additionally, 15 other ORFs, most of which lack a start codon, do not give a positive N-terminal signal peptide prediction but otherwise show all features that are often encountered in GPI proteins and are homologous to other GPI proteins (adhesive proteins and aspartic proteases). These ORFs probably represent C-terminal parts of larger GPI proteins and are therefore also included in Table 1.

In *C. glabrata* strain BG2, Epa1p, which is a member of a large GPI protein family, mediates 95 % of the adherence to human epithelial cells (Cormack *et al.*, 1999). Five members (Epa1–5p) of this family have been published so far (De Las Penas *et al.*, 2003), and recently, two new members (Epa6p and Epa7p) of this family have been submitted to the NCBI database. A BLAST search with Epa6p identified CAGL0C00110g/CAGL-IPF9053 as its closest homologue in the proteome of strain CBS138. This protein was not identified with our algorithm because it contains a glutamine residue in the hydrophobic tail (the C-terminal 10 amino acids), which is not commonly found in fungal GPI proteins (De Groot *et al.*, 2003). This prompted us to test a modified GPI algorithm that would allow Q within the hydrophobic tail. This analysis resulted in 43 additional preselected ORFs, eight of which appear to be (part of) genuine GPI proteins. Thus, glutamines, although not frequently, may occur in the hydrophobic tail of GPI signal peptides. Allowing Q in the hydrophobic tail therefore has made our GPI algorithm less selective but more comprehensive. With this approach, we have in total identified 106 putative GPI proteins of *C. glabrata*. Eighty-six of these predicted GPI proteins are confirmed by applying the recently published big-PI Fungal Predictor (Eisenhaber *et al.*, 2004). Of the 20 others, at least 17 have homology with known fungal GPI proteins, as predicted both by our method and by Eisenhaber's method, confirming that our method is comprehensive.

Functional classification of GPI proteins of *C. glabrata*

The 106 identified GPI proteins can be classified into different functional categories (Table 1).

Adhesive proteins. No fewer than 51 of the predicted *C. glabrata* GPI proteins have features that suggest adhesive functions at the cell surface and thus are destined for cell wall incorporation. For 27 of these proteins, significant similarity with 'adhesive' proteins was directly indicated by homology searches using NCBI-BLAST. Among them are 13 proteins that can be grouped in the *Cg*Epa-family. Other proteins in this group show homology with flocculins of *S. cerevisiae* and to a lesser extent with the Epa

Table 1. *In silico* identification and functional classification of putative GPI proteins of *C. glabrata*

Génolevures systematic ORF numbers	Protein name	S/T* (%)	pI	Size (aa)	big-PI prediction	Dibasic motif	NCBI BLAST result
Adhesive proteins							
CAGL0A01386g/CAGL-IPF9643		18	4·0	1102	+	+	Epa1–7p
CAGL0C00110g/CAGL-IPF9053†	Epa6p	31	5·3	715	+	−	Epa1–7p
CAGL0C05643g/CAGL-IPF4055		24	4·6	379	−	−	Epa1–7p
CAGL0D06732g/CAGL-IPF8202		22	5·0	428	−	−	Epa1–7p
CAGL0E06644g/CAGL-IPF5973	Epa1p	34	8·3	1034	+	−	Epa1–7p
CAGL0E06666g/CAGL-IPF5969	Epa2p	30	5·3	1420	+	−	Epa1–7p
CAGL0E06688g/CAGL-IPF9553	Epa3p	27	5·6	975	−	−	Epa1–7p
CAGL0I00220g/CAGL-IPF1990		32	4·4	864	+	−	Epa1–7p
CAGL0J11968g/CAGL-IPF8937		28	8·7	958	−	−	Epa1–7p
CAGL0L13332g/CAGL-IPF9478		27	5·1	1035	+	−	Epa1–7p
CAGL0M00132g/CAGL-IPF5389		29	9·1	922	−	−	Epa1–7p
CAGL0G10219g/CAGL-IPF1499		27	4·7	992	+	−	*S. cerevisiae* flocculins
CAGL0H10626g/CAGL-IPF8809		28	4·3	1326	+	−	*S. cerevisiae* flocculins
CAGL0I10098g/CAGL-IPF9938		28	4·4	1618	+	−	*S. cerevisiae* flocculins
CAGL0I10197g/CAGL-IPF9318		26	4·4	661	+	−	*S. cerevisiae* flocculins
CAGL0I10274g/CAGL-IPF6456†		23	4·5	940	+	−	*S. cerevisiae* flocculins
CAGL0I10340g/CAGL-IPF6449†		28	4·1	1012	+	−	*S. cerevisiae* flocculins
CAGL0I10362g/CAGL-IPF6446‡		28	4·1	1203	+	−	*S. cerevisiae* flocculins
CAGL0M14069g/CAGL-IPF6033		30	4·8	866	+	−	*S. cerevisiae* flocculins
CAGL0G04125g/CAGL-IPF7100		40	5·1	763	+	−	*S. cerevisiae* agglutinin Sag1p
CAGL0K00110g/CAGL-IPF9379		26	4·9	832	+	−	*C. albicans* Hyr1p
CAGL0I07293g/CAGL-IPF1493		33	4·6	1681	+	−	*C. albicans* Hyr1p
CAGL0M00110g/CAGL-IPF5376		22	9·5	186	−	−	*C. albicans* Hyr1p
CAGL0J01727g/CAGL-IPF7834		31	8·3	979	+	−	*S. cerevisiae* Awa1p and Yol155c
CAGL0J01771g/CAGL-IPF13525		33	4·4	1553	+	−	*S. cerevisiae* Awa1p and Yol155c
CAGL0J02530g/CAGL-IPF6521		28	3·9	754	−	−	*S. cerevisiae* Awa1p and Yol155c
CAGL0L09911g/CAGL-IPF8781		28	4·3	1351	+	−	*S. cerevisiae* Awa1p and Yol155c
CAGL0C00253g/CAGL-IPF6626		28	4·3	1608	+	−	*C. glabrata* adhesins, internal repeats
CAGL0C01133g/CAGL-IPF9456		30	4·5	1036	+	−	*C. glabrata* adhesins, internal repeats
CAGL0C00968g/CAGL-IPF9514		31	4·6	1034	+	−	*C. glabrata* adhesins, internal repeats
CAGL0E01661g/CAGL-IPF5474		26	4·4	1415	+	−	*C. glabrata* adhesins, internal repeats
CAGL0J02508g/CAGL-IPF6516		30	3·9	870	+	−	*C. glabrata* adhesins, internal repeats
CAGL0J02552g/CAGL-IPF6528†		27	3·9	895	−	−	*C. glabrata* adhesins, internal repeats
CAGL0K13024g/CAGL-IPF4601		27	5·1	1075	+	−	*C. glabrata* adhesins, internal repeats
Adhesive proteins without N-terminal signal peptide							
CAGL0C05665g/CAGL-IPF4054†§	Epa7p	40	7·7	334	−	−	Epa1–7p
CAGL0K00143g/CAGL-IPF5199§	Epa5p	30	5·1	830	+	−	Epa1–7p
CAGL0E00110g/CAGL-IPF15041§		29	5·9	339	+	−	*S. cerevisiae* agglutinin Sag1p
CAGL0L10098g/CAGL-IPF4133§		32	4·4	725	+	−	*C. albicans* Hyr1p
CAGL0D00110g/CAGL-IPF16506§		31	6·5	309	+	−	*S. cerevisiae* Awa1p and Yol155c
CAGL0E00121g/CAGL-IPF10387†§		30	4·1	609	−	−	*C. glabrata* adhesins, internal repeats
CAGL0H00132g/CAGL-IPF11660§		32	9·2	202	+	−	*C. glabrata* adhesins, internal repeats
CAGL0I00110g/CAGL-IPF11079§		31	4·3	957	+	−	*C. glabrata* adhesins, internal repeats
CAGL0J05148g/CAGL-IPF7006§		32	5·4	435	+	−	*C. glabrata* adhesins, internal repeats
CAGL0J12056g/CAGL-IPF10262§		35	4·5	290	+	−	*C. glabrata* adhesins, internal repeats
CAGL0A00110g/CAGL-IPF16553§		31	5·4	385	+	−	*C. glabrata* adhesive proteins
CAGL0F00110g/CAGL-IPF9600§		30	4·9	360	+	−	*C. glabrata* adhesive proteins
CAGL0G00110g/CAGL-IPF16464§		32	5·6	399	+	−	*C. glabrata* adhesive proteins
CAGL0I10120g/CAGL-IPF7689§		32	5·0	395	+	−	*C. glabrata* adhesive proteins
CAGL0J00110g/CAGL-IPF16371		32	5·0	428	+	−	*C. glabrata* adhesive proteins
CAGL0L00110g/CAGL-IPF11724§		33	6·4	409	+	−	*C. glabrata* adhesive proteins
CAGL0H00110g/CAGL-IPF11661†		25	7·1	366	−	−	*C. glabrata* adhesive proteins

Table 1. cont.

Génolevures systematic ORF numbers	Protein name	S/T* (%)	pI	Size (aa)	big-PI prediction	Dibasic motif	NCBI BLAST result
Glycoside hydrolases‖							
CAGL0E01595g/CAGL-IPF5482†		12	5·9	468	+	+	Gas/Phr family, GH72
CAGL0F03883g/CAGL-IPF1728		18	5·1	523	+	+	Gas/Phr family, GH72
CAGL0G00286g/CAGL-IPF8924	Gas1p	21	4·7	559	+	+	Gas/Phr family, GH72
CAGL0G01056g/CAGL-IPF3881	Gas3p	14	5·2	542	+	−	Gas/Phr family, GH72
CAGL0M13849g/CAGL-IPF6049	Gas2p	22	4·8	565	+	+	Gas/Phr family, GH72
CAGL0C02211g/CAGL-IPF1070	Utr2p	26	5·1	481	+	−	Crh family, GH16
CAGL0G09449g/CAGL-IPF1559	Crh1p	32	4·8	452	+	−	Crh family, GH16
CAGL0L01727g/CAGL-IPF7426		13	4·8	446	+	−	Dfg5p/Dcw1p, GH76
CAGL0M05049g/CAGL-IPF7200		13	4·6	453	+	−	Dfg5p/Dcw1p, GH76
CAGL0A01452g/CAGL-IPF3033		13	6·3	814	−	−	*Sc*Cwh41p, glucosidase I, GH63
CAGL0D02530g/CAGL-IPF4864		31	6·3	676	+	−	*Sc*Egt2p, endoglucanase
CAGL0M08756g/CAGL-IPF8154		16	5·2	569	+	−	*Sc*Exg2p, exo-1,3-β-glucanase, GH5
Aspartic proteases							
CAGL0A02431g/CAGL-IPF2954		21	4·9	587	−	+	*S. cerevisiae* aspartic proteases
CAGL0E01727g/CAGL-IPF8396		17	6·3	539	+	+	*S. cerevisiae* aspartic proteases
CAGL0E01771g/CAGL-IPF8398		19	5·9	519	−	+	*S. cerevisiae* aspartic proteases
CAGL0E01793g/CAGL-IPF8399		18	4·8	516	+	+	*S. cerevisiae* aspartic proteases
CAGL0E01815g/CAGL-IPF8400		16	6·7	519	+	+	*S. cerevisiae* aspartic proteases
CAGL0E01837g/CAGL-IPF8402		17	5·3	521	+	+	*S. cerevisiae* aspartic proteases
CAGL0E01859g/CAGL-IPF8403		18	7·7	505	+	+	*S. cerevisiae* aspartic proteases
CAGL0M04191g/CAGL-IPF8287		23	5·3	601	+	+	*S. cerevisiae* aspartic proteases
Partial aspartic protease							
CAGL0E01419g/CAGL-IPF5500§		19	4·6	591	+	+	*S. cerevisiae* aspartic proteases
Phospholipases							
CAGL0E02321g/CAGL-IPF7955		15	5·1	704	−	−	Phospholipase B
CAGL0J11770g/CAGL-IPF2933†	Plb1	17	4·7	659	+	+	Phospholipase B
CAGL0J11748g/CAGL-IPF2931		20	4·7	695	+	+	Phospholipase B
Cell wall biogenesis							
CAGL0B00616g/CAGL-IPF3695		13	9·0	416	−	−	Ecm33 family
CAGL0E04620g/CAGL-IPF2526		26	4·9	429	+	−	Ecm33 family
CAGL0H01661g/CAGL-IPF5829		11	4·9	458	−	−	Ecm33 family
CAGL0M01826g/CAGL-IPF6362		26	5·2	421	+	+	Ecm33 family
CAGL0M04169g/CAGL-IPF8285		34	7·8	373	−	−	*S. cerevisiae* Kre1p
Structural wall proteins							
CAGL0C00209g/CAGL-IPF9350		43	5·7	437	+	−	*S. cerevisiae* Srp1p/Tip1p family
CAGL0C03872g/CAGL-IPF930		33	4·5	255	+	−	*S. cerevisiae* Srp1p/Tip1p family
CAGL0F01463g/CAGL-IPF9579		26	7·7	221	+	−	*S. cerevisiae* Srp1p/Tip1p family
CAGL0F01485g/CAGL-IPF9637		28	7·7	337	+	−	*S. cerevisiae* Srp1p/Tip1p family
CAGL0G10175g/CAGL-IPF1506		38	6·4	577	+	−	*S. cerevisiae* Srp1p/Tip1p family
CAGL0H09592g/CAGL-IPF65		28	7·7	236	+	−	*S. cerevisiae* Srp1p/Tip1p family
CAGL0H09614g/CAGL-IPF62		28	7·7	236	+	−	*S. cerevisiae* Srp1p/Tip1p family
CAGL0L07502g/CAGL-IPF8864		26	5·0	228	+	−	*S. cerevisiae* Srp1p/Tip1p family
CAGL0F07579g/CAGL-IPF543	Cwp1.2p	23	7·7	212	+	−	*S. cerevisiae* Cwp1p
CAGL0F07601g/CAGL-IPF545	Cwp1.1p	23	7·7	218	+	−	*S. cerevisiae* Cwp1p
CAGL0J01463g/CAGL-IPF7854	Cwp1.3p	23	6·6	251	+	−	*S. cerevisiae* Cwp1p
CAGL0I06644g/CAGL-IPF1440		31	6·6	224	+	−	*S. cerevisiae* Spi1p and Sed1p
CAGL0H06413g/CAGL-IPF310		39	9·2	212	+	−	*S. cerevisiae* Ssr1p
CAGL0M11726g/CAGL-IPF6265		31	7·7	146	+	−	*S. cerevisiae* Ccw12p
Others							
CAGL0A04081g/CAGL-IPF1171		35	9·5	374	+	+	*S. cerevisiae* Ylr194cp
CAGL0L10670g/CAGL-IPF4179		18	5·1	263	+	+	*S. cerevisiae* Rot1p
CAGL0M02717g/CAGL-IPF6590		17	4·7	262	−	−	*S. cerevisiae* Spo19p, sporulation

Table 1. cont.

Génolevures systematic ORF numbers	Protein name	S/T* (%)	pI	Size (aa)	big-PI prediction	Dibasic motif	NCBI BLAST result
CAGL0J11462g/CAGL-IPF2905		23	11·0	208	+	−	*S. cerevisiae* Ynl190wp
CAGL0F01683g/CAGL-IPF11187		14	3·9	124	−	−	*S. cerevisiae* Bud28p, bud-site selection
Unknown							
CAGL0A03608g/CAGL-IPF1222		30	4·3	214	+	−	No hits
CAGL0B00594g/CAGL-IPF3694		19	11·5	65	+	−	No hits
CAGL0C03575g/CAGL-IPF954		28	7·1	668	+	−	No hits
CAGL0H02563g/CAGL-IPF13439		12	10·9	58	+	−	No hits
CAGL0K10164g/CAGL-IPF3204		32	9·1	217	+	−	No hits
CAGL0L06424g/CAGL-IPF7627		28	10·7	222	+	−	No hits
CAGL0M03773g/CAGL-IPF6786		32	10·8	193	+	−	No hits

*Serine/threonine.

†Glutamine in C-terminal hydrophobic domain.

‡TM prediction was not unambiguous.

§No start codon.

||Glycoside hydrolase (GH) classification according to Coutinho & Henrissat (1999).

family, or to sexual agglutinins of *S. cerevisiae* or Hyr1p of *C. albicans*. All these proteins are predicted to mediate interaction of the fungal cell with its surroundings, be they other fungal cells, host cells, or inert materials like the plastics of a catheter. Five other proteins have significant similarity with the GPI proteins Yol155c of *S. cerevisiae* and Awa1p of sake yeast. Awa1p and Yol155c are not yet fully characterized; however, the former was shown to render yeast cells more hydrophobic (Shimoi *et al.*, 2002). Both Yol155c and Awa1p have adhesin-like features: they are large proteins, have a high serine/threonine content and contain internal tandem repeats. For 19 translated ORFs, BLAST analysis indicated that they have homology with other members in the 'adhesive proteins' category. Only seven of these proteins appear to be full-length. All these seven proteins are large ($\geqslant$870 aa), have a serine/threonine content of at least 26 %, pI values of 5·1 or below, and they all contain internal repeats, indicating that they also belong to the 'adhesive proteins' category. Interestingly, among the 'adhesive' proteins are 29 proteins, 14 of which seem to comprise partial ORFs, that represent subtelomeric sequences. All these subtelomeric proteins are homologous to 'adhesive proteins' of different subcategories, including the Epa family. Consistent with this, Epa1p–Epa5p have all been localized to subtelomeric loci, and at least six other subtelomeric Epa-related proteins have been identified (De Las Penas *et al.*, 2003).

Glycoside hydrolases. Twelve of the predicted GPI proteins have homology with glycoside hydrolases (Coutinho & Henrissat, 1999). Five proteins belong to the Gas/Phr family of 1,3-β-glucanosyl transferases (Mouyna *et al.*, 2000). In *S. cerevisiae* and *C. albicans*, also five members of this family have been identifed. Two *C. glabrata* GPI proteins belong to the Crh family, which are classified in the glycoside hydrolase (GH) family 16, mainly consisting of 1,3-β-glucanases and 1,3-1,4-β-glucanosyl transferases. By performing a BLAST search, we identified a third member of the Crh family, CAGL0J08910g/CAGL-IPF6994, which like its closest homologues in *S. cerevisiae* (Crr1p) and *C. albicans* (Crh12p) does not conform to the GPI algorithm. With the nine Crh proteins from these organisms we constructed an unrooted phylogenetic tree to reveal possible functional subgroups within this family (Fig. 2a). Interestingly, Crh1p-like proteins from each organism grouped together, as was also the case for Utr2p/Crh2p-like proteins. The Crr1p-like proteins of *C. glabrata* and *S. cerevisiae* are also closely related to each other, but not to *Ca*Crr1p. All nine proteins have a putative catalytic domain similar to that of bacterial endo-1,3-1,4-β-glucanases (Fig. 2b) (Tabernero *et al.*, 1994).

Two other proteins are homologous to the Dfg5/Dcw1 family belonging to glycoside hydrolase family 76, which comprises α-1,6-mannanases, whereas CAGL0D02530g/CAGL-IPF4864 is homologous to glucan exo-1,3-β-glucosidases of glycoside hydrolase family 5 and CAGL0M08756g/CAGL-IPF8154 to the endoglucanase Egt2p of *S. cerevisiae*. All these proteins are orthologues of GPI proteins that are also present in other *in silico*-searched ascomycetous proteomes (De Groot *et al.*, 2003; Eisenhaber *et al.*, 2004), suggesting that they are ubiquitous in Ascomycetes. Finally, CAGL0A01452g/CAGL-IPF3033 is an orthologue of *Sc*Cwh41p, a processing glucosidase I enzyme (glycoside hydrolase 63) involved in modifying nascent *N*-glycosylation chains in the ER. Since *Sc*Cwh41p is not GPI-modified, and the big-PI algorithm did not detect GPI modification sites in CAGL0A01452g/CAGL-IPF3033, we suspect that this protein may be a false positive in our list. On the other hand, deletion of *CWH41* in *S. cerevisiae* causes a 50 % reduction of 1,6-β-glucan in the cell wall (Jiang *et al.*, 1996).

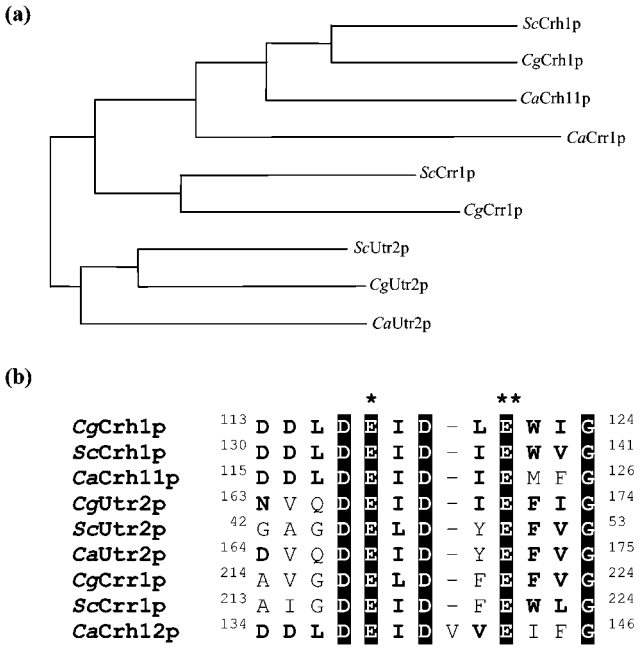

Fig. 2. The Crh families of *C. glabrata*, *C. albicans* and *S. cerevisiae*. (a) Phylogram of the Crh isoenzyme families based on amino acid sequences. (b) Alignment of the amino acid sequences of the putative catalytic domains of CAGL0G09449g/CAGL-IPF1559 (*Cg*Crh1p), CAGL0C02211g/CAGL-IPF1070 (*Cg*Utr2p), CAGL0J08910g/CAGL-IPF6994 (*Cg*Crr1p), IPF14360.2 (*Ca*Crh11p), IPF8080.2 (*Ca*Utr2p), IPF5772.2 (*Ca*CRh12p), Ygr189c (*Sc*Crh1p) Yel040w (*Sc*Utr2p) and Ylr213c (*Sc*Crr1p). The numbers show the positions of the given amino acids in the protein sequence. The position of the proposed catalytic nucleophile is indicated by a single asterisk; the proposed catalytic proton donor is indicated by two asterisks (Hoj *et al*., 1992; Viladot *et al*., 1998). Conserved amino acid residues are given in boldface and amino acids that are identical in all sequences are boxed.

Aspartic proteases. Nine proteins have significant similarity with aspartic proteases of *S. cerevisiae*. Of the six aspartic proteases that have been identified in *S. cerevisiae*, only three are predicted GPI proteins according to our selection criteria. In the human pathogen *C. albicans*, ten aspartic proteases (Saps) have been identified; however, only Sap9p has GPI protein characteristics that conform to our GPI algorithm (De Groot *et al*., 2003).

Phospholipases. The GPI algorithm identified three phopholipase B proteins, which belong to a class of well-established plasma-membrane-localized fungal GPI proteins.

Cell wall biogenesis. Five proteins are proposed to play an important, as yet uncharacterized, role in cell wall biogenesis. Four have homology to the Ecm33 family, also an ubiquitous fungal GPI protein family. Deletion of *ECM33* in *S. cerevisiae* results in swollen cells that secrete increased levels of 1,6-β-glucosylated CWPs into the growth medium and are hypersensitive to drugs that interfere with cell wall biosynthesis (De Groot *et al*., 2001). Recently, we have shown that Ecm33.3p of *C. albicans* is covalently attached to the cell wall glycan network (De Groot *et al*., 2004). The last protein in this category, CAGL0M04169g/CAGL-IPF8285, has similarity with Kre1p of *S. cerevisiae* and *C. albicans*. Neither CAGL0M04169g/CAGL-IPF8285 nor *Sc*Kre1p and *Ca*Kre1p is predicted by big-PI to be GPI-modified. However, immunofluorescence and cell fractionation studies using epitope-tagged fusion proteins confirmed that *Sc*Kre1p is localized at the cell surface (Roemer & Bussey, 1995). Furthermore, the C-terminal region of *Sc*Kre1p, comprising the putative GPI-modification signal peptide, was shown to be vital for its proper localization and functioning (Breinig *et al*., 2002). Absence of *KRE1* in *S. cerevisiae* resulted in a 40 % reduced level of cell wall 1,6-β-glucan and a shorter chain length of this cell wall polymer (Boone *et al*., 1990).

Structural wall proteins. Fourteen of the predicted GPI proteins have homology to CWPs of *S. cerevisiae* that are suggested to have a structural function. Deletion of their corresponding genes in *S. cerevisiae* did not result in strong alterations of the cell wall composition. However, some cell-wall-related phenotypes, such as increased sensitivities to the cell-wall-perturbing agents Congo red and Calcofluor white, have been observed (Van der Vaart *et al*., 1995; Garcia *et al*., 2004). Eight of the *C. glabrata* proteins in this group are homologous to the Srp1p/Tip1p family and three others to the abundant GPI-CWP Cwp1p of *S. cerevisiae* (Fig. 3). Homologues of the latter protein have not been identified in *C. albicans*.

Others. Five proteins have homology with *S. cerevisiae* proteins that cannot be directly associated with the cell wall. For the *C. glabrata* proteins with similarity to *Sc*Spo19p and *Sc*Bud28p, it is not certain whether they are true GPI proteins, because their closest homologues in *S. cerevisiae* are not predicted to be GPI-modified.

Proteins of unknown function. The remaining seven proteins have unknown functions.

In silico identification of putative mild-alkali-sensitive proteins

Established members of the class of covalently bound CWPs that can be extracted with mild alkali are Pir proteins (Kapteyn *et al*., 2001). Pir proteins share common features, namely, an N-terminal signal peptide for secretion, a Kex2p protease cleavage site, a central domain with a variable number of characteristic tandem repeats, and a C-terminal region composed of four cysteine residues with conserved spacing (Fig. 1b). Using the conserved cysteine pattern as the primary selection criterion, we identified five putative Pir proteins in *C. glabrata*. These five proteins also have sequence homology with known Pir proteins; therefore, we tentatively named these proteins *Cg*Pir1–5p

Fig. 3. Alignment of the *Cg*Cwp1p family of *C. glabrata* and *Sc*Cwp1p of *S. cerevisiae*. The degree of conservation in each position is indicated by '*' for identical residues, by ':' for conserved substitutions, and by '.' for semi-conserved substitutions in all sequences in the alignment. Peptide sequences of *Cg*Cwp1.1p and *Cg*Cwp1.2p derived from the Q-TOF MS/MS analysis are depicted in bold. Predicted N-terminal ER-import signals are underlined. The signal peptidase cleavage site of *Cg*Cwp1.1p was experimentally validated by ESI-MS/MS. Putative GPI attachment sites according to our GPI algorithm in the C-terminal region are boxed. Pir protein repeat-like sequences (in the C-terminal half of the proteins) are also underlined.

(Table 2). *Cg*Pir1–4p have all of the above-described Pir protein features, but *Cg*Pir5p has no Kex2p cleavage site, no internal repeats, and a very low pI value compared to *Cg*Pir1–4p (Table 2). Since studies with mutated versions of *Sc*Pir4p/Cis3p indicated that the presence of the glutamine-rich repetitive sequence is required for its covalent attachment to wall carbohydrates (Castillo *et al.*, 2003), it is uncertain whether *Cg*Pir5p can be covalently bound to the cell wall.

Recently, by mass spectrometric analysis of cell wall fractions of *C. albicans*, we discovered that besides Pir1p, also (part of) Scw1p/MP65 is covalently bound to the cell wall and can be extracted with mild alkali (De Groot *et al.*, 2004). Scw1p is orthologous to the Bgl2 family in *S. cerevisiae*, generally known as proteins that can be released from cell walls with reducing agents (Cappellaro *et al.*, 1998). Using the four *S. cerevisiae* orthologues and *C. albicans* Scw1p, BLAST analysis identified five members of this protein family in *C. glabrata* (Table 3). Interestingly, only two of these proteins, CAGL0G00308g/CAGL-IPF3933 and CAGL0M13805g/CAGL-IPF6053, appear to

Table 2. The Pir protein family of *C. glabrata*

Protein (ORF no.)	Sequence identity with *Sc*Pir1/2/3/4p (%)	Sequence identity with *Ca*Pir1p (%) (both alleles)	Size (aa)	No. of internal repeats	Kex2p site (KR)* position	S/T† (%)	pI
*Cg*Pir1p (CAGL0I06204g/CAGL-IPF1408)	76/62/62/40	43/43	349	7	63	18·9	9·5
*Cg*Pir2p (CAGL0I06182g/CAGL-IPF1404)	74/64/66/41	46/46	340	8	63	18·0	8·9
*Cg*Pir3p (CAGL0M08492g/CAGL-IPF8003)	75/62/67/41	48/45	335	8	64	17·1	8·5
*Cg*Pir4p (CAGL0I06160g/CAGL-IPF1401)	41/33/40/68	30/25	233	2	62	22·7	8·8
*Cg*Pir5p (CAGL0M08514g/CAGL-IPF8000)	31/26/31/50	24/22	210	0	n/a	26·6	4·9

*Lysine followed by arginine residue.
†Serine/threonine.

Table 3. Homology (% sequence identity) between *C. glabrata* Bgl2 family members and orthologues of *S. cerevisiae* and *C. albicans*

The size (number of amino acids) of the proteins is indicated in parentheses. Close relatives (sequence identity > 50 %) are indicated in bold.

C. glabrata protein	*Sc*Scw4p/ Ygr279c (386)	*Sc*Scw10p/ Ymr305c (389)	*Sc*Bgl2p/Ygr282c (313)	*Sc*Scw11p/ Ygl028c (542)	*Ca*Scw1p/ MP65 (378)
CAGL0G00308g/CAGL-IPF3933 (374 aa)	**64**	**57**	14	13	**53**
CAGL0M13805g/CAGL-IPF6053 (371 aa)	**63**	**60**	14	12	**57**
CAGL0A01474g/CAGL-IPF3032 (592 aa)	11	14	9	**51**	12
CAGL0E02915g/CAGL-IPF9200 (498 aa)	14	21	12	**52**	13
CAGL0G00220g/CAGL-IPF8919 (308 aa)	17	18	**83**	11	15

be closely related to *C. albicans* Scw1p, as well as to *S. cerevisiae* Scw4p and Scw10p (Table 3). CAGL0A01474g/CAGL-IPF3032 and CAGL0E02915g/CAGL-IPF9200, on the other hand, seem to be close relatives of *Sc*Scw11p, whereas CAGL0G00220g/CAGL-IPF8919 is much more closely related to *Sc*Bgl2p. Therefore, subgroups may exist within the Bgl2 protein family, belonging to glycoside hydrolase family 17, with respect to protein localization and function.

SDS-PAGE and Western blot analysis

As indicated by studies on the adhesin Epa1p (Frieman *et al.*, 2002; Frieman & Cormack, 2003) and by our *in silico* analysis, *C. glabrata* contains covalently bound GPI-CWPs. In order to study CWPs on a proteomic level, cell walls of *C. glabrata* strain ATCC 90876 were isolated, treated with SDS, and incubated with HF-pyridine to extract GPI proteins (De Groot *et al.*, 2004). The released material was separated by one-dimensional (1D) gradient SDS-PAGE and stained with silver, resulting in a major band of approximately 37 kDa and a high-molecular-mass smear (Fig. 4, lane 2). When the HF-pyridine-released material was blotted and probed with ConA–peroxidase, a 37 kDa protein, three additional bands, of approximately 80 kDa, 150 kDa and 250 kDa, and a high-molecular-mass smear were detected (Fig. 4, lane 4), suggesting that several GPI-CWPs were released. Immunoblot analysis further showed that the major 37 kDa band reacted with antiserum raised against *S. cerevisiae* Cwp1p, indicating that this abundant CWP of *C. glabrata* is homologous to *Sc*Cwp1p (Fig. 4, lane 3).

As *in silico* analysis indicated that *C. glabrata* also contains five Pir-related proteins, we analysed its cell walls for the presence of Pir CWPs. NaOH-released material, separated by gradient SDS-PAGE and stained with silver, appeared as a smear of high molecular mass (Fig. 4, lane 5). Western blot analysis using antiserum raised against *S. cerevisiae* Pir2p detected a distinct protein band of approximately 75 kDa (Fig. 4, lane 6). This indicates that at least one Pir protein is incorporated in the cell wall of *C. glabrata* via an alkali-sensitive linkage. Probing the same sample with ConA–peroxidase also visualized the 75 kDa protein, as well as additional bands and smears of approximately 30 kDa, 37 kDa, 150 kDa and > 250 kDa, showing that multiple mannoproteins can be released from the cell wall of *C. glabrata* by mild alkali treatment (Fig. 4, lane 7). Together with the *in silico* identification of five Pir proteins and members of the *Sc*Bgl2p/*Ca*Scw1p protein family, these results indicate that multiple Pir proteins or Bgl2 family members or both are present in cell walls of *C. glabrata*.

Mass spectrometric identification of HF-pyridine-released proteins

To identify GPI-CWPs of *C. glabrata* by *de novo* sequencing, HF-pyridine-released material was separated by 2D gel electrophoresis. Three major groups of protein spots were

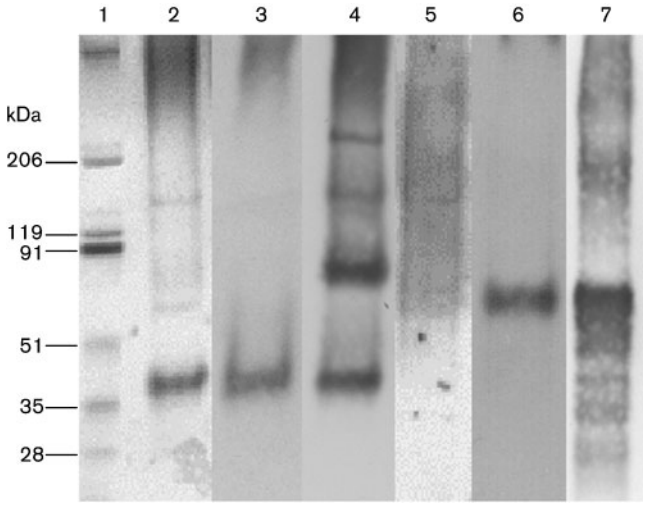

Fig. 4. *C. glabrata* contains HF-pyridine- and NaOH-extractable covalently bound CWPs. SDS-PAGE of protein fractions extracted with HF-pyridine (lanes 2–4) or mild-alkali (lanes 5–7) from isolated cell walls of *C. glabrata*. Visualization by silver staining (lanes 2 and 5), or by Western blot analysis using ConA–peroxidase (lanes 4 and 7), antiserum raised against *Sc*Cwp1 (lane 3) and antiserum raised against *Sc*Pir2p (lane 6). Lane 1, Size markers.

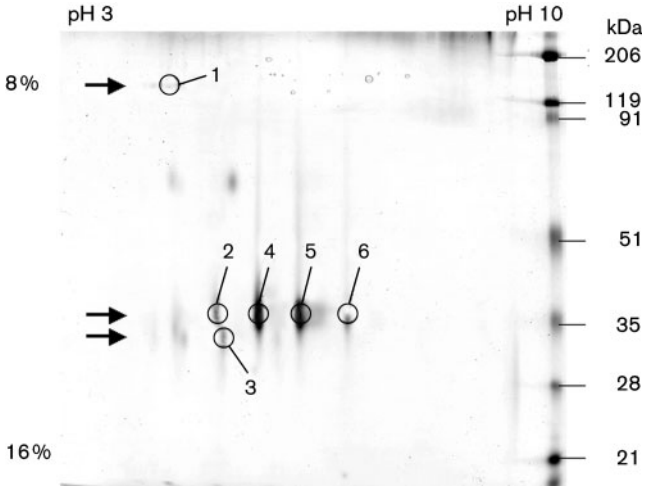

Fig. 5. HF-pyridine-extracted CWPs separated by 2D gradient gel electrophoresis. Proteins were extracted from isolated cell walls of *C. glabrata*, separated by 2D gradient gel electrophoresis (8 % to 16 %) and stained with colloidal Coomassie. Isoelectric focusing was performed using precast IPG-strips (pH 3–10). The arrows point to the three groups of protein spots at 150, 37 and 34 kDa, respectively. Spots that were excised for ESI-MS/MS purposes are indicated by numbers. Size markers are shown on the right.

observed after staining the gels with colloidal Coomassie. Each group consisted of protein spots of similar molecular mass but different pI (Fig. 5, arrows). The first group had a molecular mass of approximately 150 kDa and acidic pIs of 4–5. These spots probably correspond to the 150 kDa band that was detected when 1D gels with HF-pyridine extracts were blotted and probed with ConA–peroxidase (Fig. 4, lane 4). For the second group we estimated a molecular mass of approximately 37 kDa and pIs of 6–8, and the third group of protein spots showed a molecular mass of 33–34 kDa. The latter two groups seem to correspond to the abundant 37 kDa band that was observed on the 1D PAGE gel. We therefore speculated that these two groups of spots represented modifications of a single CWP. Six reproducible protein spots (1 to 6) were selected and excised for mass spectrometry purposes (Fig. 5).

Determined peptide masses and sequences by tandem mass spectrometry of trypsin fragments of the excised spots are given in Table 4. All these masses and sequences perfectly matched with *in silico* tryptic digests of proteins from the *C. glabrata* proteome, leading to unambiguous identifications of all six protein spots. Spot 1 corresponded to CAGL0G09449g/CAGL-IPF1559, BLAST analysis of which indicated that its closest homologue in *S. cerevisiae* is Crh1p. Hence we named this protein *Cg*Crh1p. Spots 2, 4, 5 and 6 were identified as CAGL0F07601g/CAGL-IPF545, and spot 3 as CAGL0F07579g/CAGL-IPF543 (Table 4). These two *C. glabrata* proteins are 95 % identical to each other and both have 50 % identity with Cwp1p of *S. cerevisiae*. These proteins are therefore termed *Cg*Cwp1.1p (CAGL0F07601g/CAGL-IPF545) and *Cg*Cwp1.2p (CAGL0F07579g/CAGL-IPF543). An alignment of *Sc*Cwp1p with the *Cg*Cwp proteins, in which the sequenced peptides of Cwp1.1p and Cwp1.2p are indicated in bold, is given in Fig. 3. These MS/MS data confirmed the results obtained from SDS-PAGE and Western blot analyses, namely, the presence of (an) abundant homologue(s) of *Sc*Cwp1p in the cell wall of *C. glabrata*. Furthermore, the results of the 2D gels and ESI-MS/MS analyses unequivocally show that Cwp1.1p and Cwp1.2p are both expressed and covalently bound to the cell wall, and also revealed the signal peptidase cleavage site of Cwp1.1p (Table 4).

*Cg*Cwp1.1p is a 218 aa protein with a predicted molecular mass of 21·3 kDa (unprocessed precursor). *Cg*Cwp1.2p contains 212 aa and its calculated mass is 20·7 kDa. The higher molecular mass of 33–37 kDa observed on the gel as compared to the theoretical molecular masses of the proteins are probably due to post-translational modifications such as *O*-glycosylation. *Cg*Cwp1.1p and *Cg*Cwp1.2p possess serine/threonine-rich regions that may act as multiple acceptor sites for *O*-glycosylation but they do not contain potential acceptor sites for *N*-glycosylation (N-X-S/T). In *S. cerevisiae*, *O*-linked carbohydrate side-chains may contain mannosylphosphate groups, which confer a negative charge upon the proteins (Nakayama *et al.*, 1998). Additions of phosphorylated *O*-linked carbohydrate side-chains on Cwp1.1p and Cwp1.2p molecules may therefore explain why protein spots comprising these proteins are found at different pIs. In the C-terminal region immediately upstream of the predicted GPI-attachment sites, adjacent basic residues are not present in these proteins, and also not in Crh1p. This supports the idea that adjacent basic residues may negatively influence cell wall incorporation of GPI proteins (Vossen *et al.*, 1997; Frieman & Cormack, 2003), and is in agreement with the notion that Cwp1.1p and Cwp1.2p are among the major protein constituents of the cell wall of *C. glabrata*. High expression levels of Cwp1.1p and Cwp1.2p are also in agreement with codon adaptation index (CAI) values that were determined with the codon usage measurement tool CAI Calculator 2 (Sharp & Li, 1987). When *S. cerevisiae* was chosen as the template, CAI values that were obtained for *Cg*Cwp1.1p and *Cg*Cwp1.2p were 0·919 and 0·917, respectively. As a result of the *in silico* GPI protein search, a third homologue of *Sc*Cwp1p was discovered in the proteome of *C. glabrata*, named *Cg*Cwp1.3p (Table 1, Fig. 3). This protein is composed of 251 aa and has 40 % and 42 % identity with *Cg*Cwp1.1 and *Cg*Cwp1.2, respectively. This protein also does not contain a dibasic motif, suggesting that it is predominantly cell wall localized. However, we did not detect it with ESI-MS/MS. Consistent with its lower CAI value of 0·366, this protein is possibly expressed at lower levels compared to *Cg*Cwp1.1p and *Cg*Cwp1.2p. Interestingly, in the C-terminal region of all three GPI-modified *Cg*Cwp1 proteins a short sequence similar to the internal

Table 4. Identification of HF-pyridine-extracted CWPs from *C. glabrata* cell walls by ESI-MS/MS

Protein spots 1 to 6 were excised from 2D gels, digested with trypsin and analysed using ESI-MS/MS.

Trypsin fragments of spots	Measured mass $[M+H]^+$ (Da)	Calculated mass $[M+H]^+$ (Da)	Residues	Determined sequence of peptide	Identified protein	Functional class
1a	1618·58	1618·76	36–50	ALATSFSENFTSESK	*Cg*crh1p/CAGL0G09449g/ CAGL-IPF1559	Glycoside hydrolase
1b	2898·98	2899·36	234–259	SLVVTDYSTGKEYTYG-DQSGSWQSIK		
2a	1268·38	1268·61	52–65	TGGSFAGTVTDAGK	*Cg*cwp1.1p/CAGL0F07601g/ CAGL-IPF545	Structural cell wall protein
2b	2115·68	2115·98	84–104	TGSESEGTSGFALSGSH-LTYK		
2c	1732·78	1732·92	120–136	FSTAQGTGAIDIVISPR		
3a	1254·38	1254·60	32–45	SGGSFAGTVTDAGK	*Cg*cwp1.2p/CAGL0F07579g/ CAGL-IPF543	Structural cell wall protein
3b	1560·58	1560·73	85–99	GNSGFFAIPSGSEYK		
3c	1732·78	1732·92	100–116	FSTAQGTGAIDIVISPR		
4a	1268·38	1268·61	52–65	TGGSFAGTVTDAGK	*Cg*cwp1.1p/CAGL0F07601g/ CAGL-IPF545	Structural cell wall protein
4b	3163·48	3163·64	20–51	DSQAFGLLAIHSGSPVQ-NTPVDSQNGALVLK*		
5a	1268·38	1268·61	52–65	TGGSFAGTVTDAGK	*Cg*cwp1.1p/CAGL0F07601g/ CAGL-IPF545	Structural cell wall protein
5b	1732·93	1732·92	120–136	FSTAQGTGAIDIVISPR		
6a	1268·60	1268·61	52–65	TGGSFAGTVTDAGK	*Cg*cwp1.1p/CAGL0F07601g/ CAGL-IPF545	Structural cell wall protein
6b	1560·58	1560·73	105–119	GNSGFFAIPSGSEYK		

*N-terminus of mature protein.

repeat sequences of Pir proteins is present. These internal repeats have been shown to be essential for covalent coupling of *Sc*Pir4 to the cell wall glycan network (Castillo *et al.*, 2003). Such a sequence is also present in *Sc*Cwp1, which has been shown to become linked to the cell wall both through GPI modification and in a mild-alkali-sensitive manner (Kapteyn *et al.*, 2001).

DISCUSSION

In this paper, we present a first systematic analysis of the cell wall proteome of the human pathogen *C. glabrata*. We have also studied the covalent linkages between CWPs and the glycan network, and performed confirmative mass spectrometric identifications of CWPs. In *S. cerevisiae* and *C. albicans*, the major class of covalently bound CWPs are GPI-CWPs. Extraction of CWPs with HF-pyridine, which specifically cleaves phosphodiester bridges under the conditions used, indicated that *C. glabrata* also contains mannosylated GPI-CWPs that are linked to 1,6-β-glucan through phosphodiester bridges. This architectural model was strengthened by data of 1,6-β-glucanase digestions, which also liberated mannosylated CWPs (results not shown). Finally, mass spectrometric analysis of HF-pyridine-extracted CWPs resulted in the identification of three predicted GPI proteins (Crh1p, Cwp1.1p and Cwp1.2p). A similar strategy has been applied by Pitarch and coworkers to characterize the interactions of CWPs with other wall components in *C. albicans* (Pitarch *et al.*, 2002). Our preliminary mass spectrometric analysis of 2D-separated proteins was not expected to yield a complete image of the cell wall proteome. However, using a combination of sequential protein fractionation, 2D PAGE and mass spectrometry we confirmed the presence of GPI-CWPs in *C. glabrata* and validated our *in silico* method to predict GPI proteins in this organism. In addition, particularly for Cwp1.1p, our analyses demonstrated that CWPs undergo post-translational modifications.

As shown by SDS-PAGE and Western blotting, the identified Cwp1.1p and Cwp1.2p correspond with the most abundant protein band on 1D gels of the HF-pyridine extracts. The homologous Cwp1p, the first identified GPI-CWP of *S. cerevisiae*, is also one of the most abundant CWPs (Van der Vaart *et al.*, 1995), indicating a structural similarity between cell walls of these evolutionarily closely

related yeasts. In contrast, a homologue of *Sc*Cwp1 is not present in *C. albicans*.

To make a comprehensive inventory of the putative GPI proteins in *C. glabrata*, we performed an *in silico* analysis by applying a slightly modified version of a fungal-specific algorithm which was previously developed to identify GPI proteins in *S. cerevisiae* and *C. albicans* (De Groot *et al.*, 2003). Of the 106 predicted GPI proteins, 51 could be classified as putative adhesive proteins, possibly involved in interactions of *C. glabrata* with other yeast cells or adhesion to host cells or inert materials, and thus are likely to be incorporated in the cell wall. In agreement with this, only one of these proteins has adjacent basic residues immediately upstream of putative GPI-modification sites (Table 1). These proteins therefore may play an important role in establishing fungal infections. Consistent with this idea, it has been shown that the GPI-modified protein *Cg*Epa1p mediates adhesion to cultured human epithelial cells (Cormack *et al.*, 1999). Four other members of this family have been described so far, but more Epa-like proteins have been discovered (De Las Penas *et al.*, 2003; NCBI database). Of the 13 Epa family members that we have identified with our GPI algorithm, CAGL0D06732g/CAGL-IPF8202 and CAGL0C05643g/CAGL-IPF4055 comprise 428 and 379 amino acids, respectively, and have significant similarity with only the N-terminal part of Epa proteins. This might suggest that these translated ORFs constitute C-terminally truncated versions of proteins that may not be present in full length in the current release of the *C. glabrata* proteome. On the other hand, in *C. albicans*, members of the Rbt5p/Csa1p GPI protein family, which also contain features that are indicative of proteins with adhesive properties, vary in length and are homologous only in their N-terminal regions (De Groot *et al.*, 2003). Also, CAGL0D06732g/CAGL-IPF8202 and CAGL0C05643g/CAGL-IPF4055 were recognized by the GPI algorithm, suggesting that they are smaller full-length representatives of the Epa family. In contrast, CAGL0A00121g/CAGL-IPF9521, the closest homologue of Epa4p in the proteome of strain CBS138, was not selected by our algorithm and seems to comprise only the N-terminal part of Epa4p, as indicated by BLAST analysis. Possibly, large allelic variations may occur in the Epa adhesin family of *C. glabrata*, as is also the case for the Als family of *C. albicans* (Hoyer, 2001; Zhao *et al.*, 2003).

The nonpathogenic *S. cerevisiae* contains only 11 known adhesive proteins, including agglutinins, Awa1p-like proteins, which mediate the foam-forming capacity of sake yeast (Shimoi *et al.*, 2002), and flocculins. The human pathogens *C. glabrata* and *C. albicans* seem to contain a much larger number of putative adhesive proteins. In *C. albicans*, 24–26 of the predicted GPI proteins have adhesin-like properties, including the Als, Hwp1p/Rbt1p, Csa1p/Rbt5p and Hyr1p families. Conceivably, the high number of adhesive proteins may be related to the ability of the fungus to cause infections. In the polymorphic yeast *C. albicans*, differential regulation has been shown for Als3p, Hwp1p and Hyr1p, which are all specifically expressed or upregulated during hyphal development (Bailey *et al.*, 1996; Hoyer *et al.*, 1998; Staab *et al.*, 1999), a growth form that does not seem to occur in *C. glabrata*. Hyphal-specific Hwp1p has been shown to mediate virulence in a mouse model of systemic infection (Staab *et al.*, 1999; Sundstrom, 2002). Nevertheless, the fact that the nonfilamentous *C. glabrata* is a frequent cause of fungal infections indicates that, rather than the ability to undergo morphological development, a large repertoire of isofunctional adhesins may be required to develop infections. Possibly, differentially regulated adhesins enable human pathogenic fungi to make quick adaptations of their adhesion capacity to the altering conditions that are present in different human tissues, and to escape the hostile environment that is created by the host immune response.

At least 12 of the identified putative GPI proteins specify glycoside hydrolases, and most of these belong to GPI protein families that seem ubiquitous among ascomycetous fungi (De Groot *et al.*, 2003; Eisenhaber *et al.*, 2004). In *C. albicans*, we have shown recently that GPI-modified enzymes belonging to glycoside hydrolase families are present in HF-pyridine extracts of cell walls and are covalently bound to 1,6-β-glucan (De Groot *et al.*, 2004). We hypothesize that such a mechanism allows fungi to retain a desired enzyme activity in the wall, which otherwise would be lost to the surrounding growth environment. This would enable the organism to perform an effective and targeted remodelling and expansion of its cell wall structure during growth. The Crh1p protein that we have identified in HF-pyridine extracts is more closely related to *Sc*Crh1p and *Ca*Crh11p than to other members of this protein family. Interestingly, *Ca*Crh11p is also covalently bound to the cell wall lattice (De Groot *et al.*, 2004), and, using green fluorescent protein as a fusion marker, *Sc*Crh1p was shown to be localized at the cell surface in a cell-cycle-dependent manner. *Sc*Crh1p expression seems to correlate with chitin deposition in the cell wall (Rodriguez-Pena *et al.*, 2000), suggesting a role, possibly indirect by remodelling the 1,3-β-glucan network, for this enzyme in chitin incorporation in the wall. Notably, in contrast to the filamentous *C. albicans* and *N. crassa*, GPI proteins specifying chitinase activity were not identified in the nonfilamentous yeasts *C. glabrata*, *S. cerevisiae* and *Sch. pombe*. GPI modification of *C. albicans* chitinase 2 was confirmed by its identification in HF-pyridine extracts of cell walls (De Groot *et al.*, 2004). We speculate that, in terms of the glycan composition, in particular chitin, differences in cell wall structure between fungi may be related to the ability to undergo filamentation.

Nine aspartic proteases of *C. glabrata* are predicted to be GPI-modified. These proteins are probably predominantly localized in the plasma membrane since they all have adjacent basic residues in their C-terminal regions. Interestingly, in *C. albicans*, secreted aspartic proteases (Saps) have

been demonstrated to be virulence attributes, crucial for mucosal and systemic infections and involved in evasion of host immune responses (Hube & Naglik, 2001). However, in contrast to *C. glabrata*, only *Ca*Sap9p is predicted to be GPI-modified, suggesting differences in their defence strategy when approaching hostile environments.

In *C. albicans* and *S. cerevisiae*, a minor class of covalently incorporated CWPs can be released with mild alkali. The best-characterized members of this class are Pir proteins, five of which were identified in *C. glabrata*. In *C. albicans* cell walls, only a single Pir protein was found (De Groot *et al.*, 2004), whereas *S. cerevisiae* walls contain at least three (Mrsa *et al.*, 1997). In *C. albicans*, a 1,3-β-glucanase with homology to the Bgl2 family of *S. cerevisiae*, *Ca*Scw1p, is also incorporated in a mild-alkali-sensitive manner. Of the five Bgl homologues that we identified in *C. glabrata*, two are closely related to *Ca*Scw1p, making them likely candidates to be incorporated in a similar way. SDS-PAGE and Western blot analysis using ConA confirmed the presence of mannosylated mild-alkali-sensitive proteins in *C. glabrata*. Among those are Pir proteins, as was shown by reactivity of the extracts with anti-*Sc*Pir2p antiserum. Thus, although its CWP composition may differ significantly, in terms of protein–glycan linkages, *C. glabrata* conforms to the cell wall architectural models that have been described for *S. cerevisiae* (Klis *et al.*, 2002) and *C. albicans* (Klis *et al.*, 2001). These models can therefore successfully be used to explore the cell wall structure of the human pathogen *C. glabrata*.

ACKNOWLEDGEMENTS

Candida glabrata genome sequencing was performed by the Génolevures consortium. We thank B. Dujon and C. Hennequin for communication of unpublished results, A. Goretzki for excellent technical assistance, and D. Gozalbo, H. Shimoi and P. Russo for providing antibodies. P. W. J. de G. was supported by the European Commission (QLRT-1999-30795). M. W. was supported by the Forschungsförderprogramm of the University of Göttingen, by the MSD-grant Mycology 2001 and by the DFG (WE 3537/1-1).

REFERENCES

Bailey, D. A., Feldmann, P. J., Bovey, M., Gow, N. A. & Brown, A. J. (1996). The *Candida albicans HYR1* gene, which is activated in response to hyphal development, belongs to a gene family encoding yeast cell wall proteins. *J Bacteriol* **178**, 5353–5360.

Bom, I. J., Dielbandhoesing, S. K., Harvey, K. N., Oomes, S. J., Klis, F. M. & Brul, S. (1998). A new tool for studying the molecular architecture of the fungal cell wall: one-step purification of recombinant trichoderma beta-(1-6)-glucanase expressed in *Pichia pastoris*. *Biochim Biophys Acta* **1425**, 419–424.

Boone, C., Sommer, S. S., Hensel, A. & Bussey, H. (1990). Yeast *KRE* genes provide evidence for a pathway of cell wall beta-glucan assembly. *J Cell Biol* **110**, 1833–1843.

Breinig, F., Tipper, D. J. & Schmitt, M. J. (2002). Kre1p, the plasma membrane receptor for the yeast K1 viral toxin. *Cell* **108**, 395–405.

Calcagno, A. M., Bignell, E., Warn, P., Jones, M. D., Denning, D. W., Muhlschlegel, F. A., Rogers, T. R. & Haynes, K. (2003). *Candida glabrata STE12* is required for wild-type levels of virulence and nitrogen starvation induced filamentation. *Mol Microbiol* **50**, 1309–1318.

Cappellaro, C., Mrsa, V. & Tanner, W. (1998). New potential cell wall glucanases of *Saccharomyces cerevisiae* and their involvement in mating. *J Bacteriol* **180**, 5030–5037.

Castillo, L., Martínez, A. I., Garcera, A., Elorza, M. V., Valentin, E. & Sentandreu, R. (2003). Functional analysis of the cysteine residues and the repetitive sequence of *Saccharomyces cerevisiae* Pir4/Cis3: the repetitive sequence is needed for binding to the cell wall beta-1,3-glucan. *Yeast* **20**, 973–983.

Chaffin, W. L., Lopez-Ribot, J. L., Casanova, M., Gozalbo, D. & Martinez, J. P. (1998). Cell wall and secreted proteins of *Candida albicans*: identification, function, and expression. *Microbiol Mol Biol Rev* **62**, 130–180.

Cheng, S., Clancy, C. J., Checkley, M. A., Handfield, M., Hillman, J. D., Progulske-Fox, A., Lewin, A. S., Fidel, P. L. & Nguyen, M. H. (2003). Identification of *Candida albicans* genes induced during thrush offers insight into pathogenesis. *Mol Microbiol* **48**, 1275–1288.

Cormack, B. P., Ghori, N. & Falkow, S. (1999). An adhesin of the yeast pathogen *Candida glabrata* mediating adherence to human epithelial cells. *Science* **285**, 578–582.

Coutinho, P. M. & Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering*, pp. 3–12. Edited by H. J. Gilbert, G. Davies, B. Henrissat & B. Svensson. Cambridge: Royal Society of Chemistry.

Csank, C. & Haynes, K. (2000). *Candida glabrata* displays pseudohyphal growth. *FEMS Microbiol Lett* **189**, 115–120.

De Groot, P. W., Ruiz, C., Vázquez de Aldana, C. R. & 14 other authors (2001). A genomic approach for the identification and classification of genes involved in cell wall formation and its regulation in *Saccharomyces cerevisiae*. *Comp Funct Genomics* **2**, 124–142.

De Groot, P. W., Hellingwerf, K. J. & Klis, F. M. (2003). Genome-wide identification of fungal GPI proteins. *Yeast* **20**, 781–796.

De Groot, P. W., De Boer, A. D., Cunningham, J., Dekker, H. L., De Jong, L., Hellingwerf, K. J., De Koster, C. & Klis, F. M. (2004). Proteomic analysis of *Candida albicans* cell walls reveals covalently bound carbohydrate-active enzymes and adhesins. *Eukaryot Cell* **3**, 955–965.

De Las Penas, A., Pan, S. J., Castano, I., Alder, J., Cregg, R. & Cormack, B. P. (2003). Virulence-related surface glycoproteins in the yeast pathogen *Candida glabrata* are encoded in subtelomeric clusters and subject to *RAP1*- and *SIR*-dependent transcriptional silencing *Genes Dev* **17**, 2245–2258.

Eisenhaber, B., Schneider, G., Wildpaner, M. & Eisenhaber, F. (2004). A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *J Mol Biol* **337**, 243–253.

Fidel, P. L., Jr, Vazquez, J. A. & Sobel, J. D. (1999). *Candida glabrata*: review of epidemiology, pathogenesis, and clinical disease with comparison to *C. albicans*. *Clin Microbiol Rev* **12**, 80–96.

Frieman, M. B. & Cormack, B. P. (2003). The omega-site sequence of glycosylphosphatidylinositol-anchored proteins in *Saccharomyces cerevisiae* can determine distribution between the membrane and the cell wall. *Mol Microbiol* **50**, 883–896.

Frieman, M. B., McCaffery, J. M. & Cormack, B. P. (2002). Modular domain structure in the *Candida glabrata* adhesin Epa1p,

a beta1,6 glucan-cross-linked cell wall protein. *Mol Microbiol* **46**, 479–492.

Garcia, R., Bermejo, C., Grau, C., Perez, R., Rodriguez-Pena, J. M., Francois, J., Nombela, C. & Arroyo, J. (2004). The global transcriptional response to transient cell wall damage in *Saccharomyces cerevisiae* and its regulation by the cell integrity signaling pathway. *J Biol Chem* **279**, 15183–15195.

Garcia-Sanchez, S., Aubert, S., Iraqui, I., Janbon, G., Ghigo, J. M. & D'Enfert, C. (2004). Biofilms of *Candida albicans*: a developmental state associated with specific and stable gene expression patterns. *Eukaryot Cell* **3**, 536–545.

Gomez, M. J., Torosantucci, A., Arancia, S., Maras, B., Parisi, L. & Cassone, A. (1996). Purification and biochemical characterization of a 65-kilodalton mannoprotein (MP65), a main target of anti-Candida cell-mediated immune responses in humans. *Infect Immun* **64**, 2577–2584.

Gozalbo, D., Gil-Navarro, I., Azorin, I., Renau-Piqueras, J., Martinez, J. P. & Gil, M. L. (1998). The cell wall-associated glyceraldehyde-3-phosphate dehydrogenase of *Candida albicans* is also a fibronectin and laminin binding protein. *Infect Immun* **66**, 2052–2059.

Halme, A., Bumgarner, S., Styles, C. & Fink, G. R. (2004). Genetic and epigenetic regulation of the *FLO* gene family generates cell-surface variation in yeast. *Cell* **116**, 405–415.

Hamada, K., Terashima, H., Arisawa, M. & Kitada, K. (1998). Amino acid sequence requirement for efficient incorporation of glycosylphosphatidylinositol-associated proteins into the cell wall of *Saccharomyces cerevisiae*. *J Biol Chem* **273**, 26946–26953.

Hamada, K., Terashima, H., Arisawa, M., Yabuki, N. & Kitada, K. (1999). Amino acid residues in the omega-minus region participate in cellular localization of yeast glycosylphosphatidylinositol-attached proteins. *J Bacteriol* **181**, 3886–3889.

Hoj, P. B., Condron, R., Traeger, J. C., McAuliffe, J. C. & Stone, B. A. (1992). Identification of glutamic acid 105 at the active site of *Bacillus amyloliquefaciens* 1,3-1,4-beta-D-glucan 4-glucanohydrolase using epoxide-based inhibitors. *J Biol Chem* **267**, 25059–25066.

Hoyer, L. L. (2001). The *ALS* gene family of *Candida albicans*. *Trends Microbiol* **9**, 176–180.

Hoyer, L. L., Payne, T. L., Bell, M., Myers, A. M. & Scherer, S. (1998). *Candida albicans ALS3* and insights into the nature of the *ALS* gene family. *Curr Genet* **33**, 451–459.

Hoyer, L. L., Clevenger, J., Hecht, J. E., Ehrhart, E. J. & Poulet, F. M. (1999). Detection of Als proteins on the cell wall of *Candida albicans* in murine tissues. *Infect Immun* **67**, 4251–4255.

Hube, B. & Naglik, J. (2001). *Candida albicans* proteinases: resolving the mystery of a gene family. *Microbiology* **147**, 1997–2005.

Jiang, B., Sheraton, J., Ram, A. F., Dijkgraaf, G. J., Klis, F. M. & Bussey, H. (1996). *CWH41* encodes a novel endoplasmic reticulum membrane *N*-glycoprotein involved in beta 1,6-glucan assembly. *J Bacteriol* **178**, 1162–1171.

Kapteyn, J. C., Montijn, R. C., Dijkgraaf, G. J., Van den, E. H. & Klis, F. M. (1995). Covalent association of beta-1,3-glucan with beta-1,6-glucosylated mannoproteins in cell walls of *Candida albicans*. *J Bacteriol* **177**, 3788–3792.

Kapteyn, J. C., Montijn, R. C., Vink, E., de la, Cruz. J., Llobell, A., Douwes, J. E., Shimoi, H., Lipke, P. N. & Klis, F. M. (1996). Retention of *Saccharomyces cerevisiae* cell wall proteins through a phosphodiester-linked beta-1,3-/beta-1,6-glucan heteropolymer. *Glycobiology* **6**, 337–345.

Kapteyn, J. C., Van Egmond, P., Sievi, E., Van den Ende, H., Makarow, M. & Klis, F. M. (1999). The contribution of the *O*-glycosylated protein Pir2p/Hsp150 to the construction of the yeast cell wall in wild-type cells and beta 1,6-glucan-deficient mutants. *Mol Microbiol* **31**, 1835–1844.

Kapteyn, J. C., ter Riet, B., Vink, E., Blad, S., De Nobel, H., Van den Ende, H. & Klis, F. M. (2001). Low external pH induces *HOG1*-dependent changes in the organization of the *Saccharomyces cerevisiae* cell wall. *Mol Microbiol* **39**, 469–479.

Klis, F. M., Ram, A. F. J., Montijn, R. C., Kapteyn, J. C., Caro, L. H. & Vossen, J. H. (1998). Posttranslational modifications of secretory proteins. *Methods Microbiol* **26**, 223–238.

Klis, F. M., de Groot, P. & Hellingwerf, K. (2001). Molecular organization of the cell wall of *Candida albicans*. *Med Mycol* **39 Suppl 1**, 1–8.

Klis, F. M., Mol, P., Hellingwerf, K. & Brul, S. (2002). Dynamics of cell wall structure in *Saccharomyces cerevisiae*. *FEMS Microbiol Rev* **26**, 239–256.

Lachke, S. A., Joly, S., Daniels, K. & Soll, D. R. (2002). Phenotypic switching and filamentation in *Candida glabrata*. *Microbiology* **148**, 2661–2674.

Li, F. & Palecek, S. P. (2003). *EAP1*, a *Candida albicans* gene involved in binding human epithelial cells. *Eukaryot Cell* **2**, 1266–1273.

Lo, H. J., Kohler, J. R., DiDomenico, B., Loebenberg, D., Cacciapuoti, A. & Fink, G. R. (1997). Nonfilamentous *Candida albicans* mutants are avirulent. *Cell* **90**, 939–949.

Martínez, A. I., Castillo, L., Garcerá, A., Elorza, M. V., Valentín, E. & Sentandreu, R. (2004). Role of Pir1 in the construction of the *Candida albicans* cell wall. *Microbiology* **150**, 3151–3161.

Mouyna, I., Monod, M., Fontaine, T., Henrissat, B. & Lechenne, B., Latgé, J. P. (2000). Identification of the catalytic residues of the first family of beta(1-3)glucanosyltransferases identified in fungi. *Biochem J* **347**, 741–747.

Mrsa, V. & Tanner, W. (1999). Role of NaOH-extractable cell wall proteins Ccw5p, Ccw6p, Ccw7p and Ccw8p (members of the Pir protein family) in stability of the *Saccharomyces cerevisiae* cell wall. *Yeast* **15**, 813–820.

Mrsa, V., Seidl, T., Gentzsch, M. & Tanner, W. (1997). Specific labelling of cell wall proteins by biotinylation. Identification of four covalently linked *O*-mannosylated proteins of *Saccharomyces cerevisiae*. *Yeast* **13**, 1145–1154.

Nakayama, K., Feng, Y., Tanaka, A. & Jigami, Y. (1998). The involvement of *mnn4* and *mnn6* mutations in mannosylphosphorylation of *O*-linked oligosaccharide in yeast *Saccharomyces cerevisiae*. *Biochim Biophys Acta* **1425**, 255–262.

Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**, 255–262.

Nisini, R., Romagnoli, G., Gomez, M. J., La Valle, R., Torosantucci, A., Mariotti, S., Teloni, R. & Cassone, A. (2001). Antigenic properties and processing requirements of 65-kilodalton mannoprotein, a major antigen target of anti-Candida human T-cell response, as disclosed by specific human T-cell clones. *Infect Immun* **69**, 3728–3736.

Pfaller, M. A., Messer, S. A., Hollis, R. J., Jones, R. N. & Diekema, D. J. (2002). In vitro activities of ravuconazole and voriconazole compared with those of four approved systemic antifungal agents against 6,970 clinical isolates of *Candida* spp. *Antimicrob Agents Chemother* **46**, 1723–1727.

Pfaller, M. A., Diekema, D. J., Messer, S. A., Boyken, L. & Hollis, R. J. (2003a). Activities of fluconazole and voriconazole against 1,586 recent clinical isolates of *Candida* species determined by Broth microdilution, disk diffusion, and Etest methods: report from the ARTEMIS Global Antifungal Susceptibility Program, 2001. *J Clin Microbiol* **41**, 1440–1446.

Pfaller, M. A., Messer, S. A., Boyken, L., Tendolkar, S., Hollis, R. J. & Diekema, D. J. (2003b). Variation in susceptibility of bloodstream isolates of *Candida glabrata* to fluconazole according to patient age and geographic location. *J Clin Microbiol* **41**, 2176–2179.

Pitarch, A., Sanchez, M., Nombela, C. & Gil, C. (2002). Sequential fractionation and two-dimensional gel analysis unravels the complexity of the dimorphic fungus *Candida albicans* cell wall proteome. *Mol Cell Proteomics* **12**, 967–982.

Rex, J. H., Walsh, T. J., Sobel, J. D., Filler, S. G., Pappas, P. G., Dismukes, W. E. & Edwards, J. E. (2000). Practice guidelines for the treatment of candidiasis. Infectious Diseases Society of America. *Clin Infect Dis* **30**, 662–678.

Rodriguez-Pena, J. M., Cid, V. J., Arroyo, J. & Nombela, C. (2000). A novel family of cell wall-related proteins regulated differently during the yeast life cycle. *Mol Cell Biol* **20**, 3245–3255.

Roemer, T. & Bussey, H. (1995). Yeast Kre1p is a cell surface *O*-glycoprotein. *Mol Gen Genet* **249**, 209–216.

Russo, P., Kalkkinen, N., Sareneva, H., Paakkola, J. & Makarow, M. (1992). A heat shock gene from *Saccharomyces cerevisiae* encoding a secretory glycoprotein. *Proc Natl Acad Sci U S A* **89**, 3671–3675.

Scherrer, R., Louden, L. & Gerhardt, P. (1974). Porosity of the yeast cell wall and membrane. *J Bacteriol* **118**, 534–540.

Sharp, P. M. & Li, W. H. (1987). The Codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**, 1281–1295.

Shimoi, H., Iimura, Y. & Obata, T. (1995). Molecular cloning of *CWP1*: a gene encoding a *Saccharomyces cerevisiae* cell wall protein solubilized with *Rarobacter faecitabidus* protease I. *J Biochem* **118**, 302–311.

Shimoi, H., Sakamoto, K., Okuda, M., Atthi, R., Iwashita, K. & Ito, K. (2002). The *Awa1* gene is required for the foam-forming phenotype and cell surface hydrophobicity of sake yeast. *Appl Environ Microbiol* **68**, 2018–2025.

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

Sundstrom, P. (2002). Adhesion in *Candida* spp. *Cell Microbiol* **4**, 461–469.

Tabernero, C., Coll, P. M., Fernandez-Abalos, J. M., Perez, P. & Santamaria, R. I. (1994). Cloning and DNA sequencing of *bgaA*, a gene encoding an endo-beta-1,3-1,4-glucanase, from an alkalophilic *Bacillus* strain (N137). *Appl Environ Microbiol* **60**, 1213–1220.

Terashima, H., Hamada, K. & Kitada, K. (2003). The localization change of Ybr078w/Ecm33, a yeast GPI-associated protein, from the plasma membrane to the cell wall, affecting the cellular function. *FEMS Microbiol Lett* **218**, 175–180.

Van der Vaart, J. M., Caro, L. H., Chapman, J. W., Klis, F. M. & Verrips, C. T. (1995). Identification of three mannoproteins in the cell wall of *Saccharomyces cerevisiae*. *J Bacteriol* **177**, 3104–3110.

Viladot, J. L., de Ramon, E., Durany, O. & Planas, A. (1998). Probing the mechanism of *Bacillus* 1,3-1,4-beta-D-glucan 4-glucanohydrolases by chemical rescue of inactive mutants at catalytically essential residues. *Biochemistry* **37**, 11332–11342.

Vossen, J. H., Muller, W. H., Lipke, P. N. & Klis, F. M. (1997). Restrictive glycosylphosphatidylinositol anchor synthesis in *cwh6*/*gpi3* yeast cells causes aberrant biogenesis of cell wall proteins. *J Bacteriol* **179**, 2202–2209.

Weig, M., Haynes, K., Rogers, T. R., Kurzai, O., Frosch, M. & Muhlschlegel, F. A. (2001). A *GAS*-like gene family in the pathogenic fungus *Candida glabrata*. *Microbiology* **147**, 2007–2019.

Zhao, X., Pujol, C., Soll, D. R. & Hoyer, L. L. (2003). Allelic variation in the contiguous loci encoding *Candida albicans ALS5*, *ALS1* and *ALS9*. *Microbiology* **149**, 2947–2960.

Zlotnik, H., Fernandez, M. P., Bowers, B. & Cabib, E. (1984). *Saccharomyces cerevisiae* mannoproteins form an external cell wall layer that determines wall porosity. *J Bacteriol* **159**, 1018–1026.