Growth-dependent secretome of *Candida utilis*

Christoph Buerth,¹ Clemens J. Heilmann,² Frans M. Klis,² Chris G. de Koster,² Joachim F. Ernst¹ and Denis Tielker¹

¹Molecular Mycology, Heinrich-Heine-University, Duesseldorf, Universitaetsstraße 1, 40225 Duesseldorf, Germany
²Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

Recently, the food yeast *Candida utilis* has emerged as an excellent host for production of heterologous proteins. Since secretion of the recombinant product is advantageous for its purification, we characterized the secreted proteome of *C. utilis*. Cells were cultivated to the exponential or stationary growth phase, and the proteins in the medium were identified by MS. In parallel, a draft genome sequence of *C. utilis* strain DSM 2361 was determined by massively parallel sequencing. Comparisons of protein and coding sequences established that *C. utilis* is not a member of the CUG clade of *Candida* species. In total, we identified 37 proteins in the culture solution, 17 of which were exclusively present in the stationary phase, whereas three proteins were specific to the exponential growth phase. Identified proteins represented mostly carbohydrate-active enzymes associated with cell wall organization, while no proteolytic enzymes and only a few cytoplasmic proteins were detected. Remarkably, cultivation in xylose-based medium generated a protein pattern that diverged significantly from glucose-grown cells, containing the invertase Invl as the major extracellular protein, particularly in its highly glycosylated S-form (slow-migrating). Furthermore, cultivation without ammonium sulfate induced the secretion of the asparaginase Asp3. Comparisons of the secretome of *C. utilis* with those of *Kluyveromyces lactis* and *Pichia pastoris*, as well as with those of the human fungal pathogens *Candida albicans* and *Candida glabrata*, revealed a conserved set of 10 and six secretory proteins, respectively.

INTRODUCTION

For more than six decades, *Candida utilis* has represented an industrially important yeast, being classified as GRAS (generally recognized as safe) by regulatory authorities. Another advantage of this yeast is that it assimilates and easily adapts to a number of different carbon and nitrogen sources, including pentose sugars, organic acids, alcohols, urea, ammonium salts, pyrimidine, and various amino acids. Furthermore, the growth of *C. utilis* is scarcely affected by extremes in pH, and being Crabtree-negative it does not produce ethanol in aerated cultures, which limits growth in other yeast species (Boze et al., 1992).

Initially, *C. utilis* was used as a food yeast to produce single-cell protein of high quality from cheap, biomass-derived waste substrates, including spent sulfite liquor and sugar molasses (Inskeep et al., 1951). Subsequently, *C. utilis* served as a source of different endogenous products such as glutathione (Liang et al., 2008), biotin (Hong et al., 2006), glucomannan (Ruszova et al., 2008), L-phenylacetylcarbinol (Khan & Daugulis, 2010), ribonucleic acids, NAD (reviewed by Chakravorty et al., 1962) and several enzymes, including invertase (Belcarz et al., 2002). The development of genetic tools, including expression vectors, resistance cassettes and the cre-loxP recombination system, has allowed the efficient production of heterologous products in *C. utilis*, e.g. the sweetener monellin from *Dioscoreophyllum cumminnsii* (Kondo et al., 1997), three carotenoids from *Erwinia uredovora* (Miura et al., 1998), α-amylase from *Sulfolobus solfataricus* (Miura et al., 1999) and xylanase from *Streptomyces olivaceoviridis* (Wei et al., 2010). Moreover, Ikushima et al. (2009) have constructed a pyruvate decarboxylase-deficient *C. utilis* strain that expresses L-lactate dehydrogenase from *Bos taurus*, and which produces the highest amounts of the polylactic acid precursor L-lactate ever reported in yeast. These reports suggest that *C. utilis* is an efficient host for the high-level production of recombinant proteins, and may become an alternative to more established yeast expression hosts, including *Pichia pastoris, Kluyveromyces lactis* and *Saccharomyces cerevisiae*. In all fungal species, secretion of recombinant proteins to the culture medium is a preferred...
route of production, because protein purification is greatly facilitated and proteins are glycosylated. This raises questions about the number, functions and amounts of proteins that are naturally secreted into the medium. Such studies have been initiated for the yeasts P. pastoris (Mattanovich et al., 2009) and K. lactis (Swaim et al., 2008; Madinger et al., 2009). Importantly, in the culture medium of K. lactis, extracellular proteases have been detected whose presence can diminish the yields and complicate the isolation of pure and functional products. The secretome of C. utilis, however, has not yet been characterized. Furthermore, despite its importance, its precise ploidy status and genomic or proteomic sequence are largely unknown. Some members of the genus Candida belonging to the so-called CUG clade (reviewed by Massey et al., 2003) translate the CUG codon, which according to the standard genetic code encodes leucine, to serine. The taxonomic position of C. utilis, especially its relation to the CUG clade, remains uncertain.

In this study we present for the first time the C. utilis strain DSM 2361 secretory proteome (secretome), which includes proteins that possess either a signal peptide or a glycosylphosphatidylinositol (GPI) anchor for wall and membrane anchorage, or both. We demonstrate that the C. utilis secretome is strongly dependent on the growth phase, and also on the types of carbon and nitrogen source used. Furthermore, we have established the C. utilis genome sequence by massively parallel sequencing, allowing the identification of potential secretome proteins and the conclusion that C. utilis is not a member of the CUG clade of Candida species. Furthermore, we report on a conserved core set of secretory proteins among different yeasts of biotechnological and medical relevance.

METHODS

Strain. C. utilis strain DSM 2361 (ATCC 9950, CBS 5609) was purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ) and used for the preparation of genomic DNA for sequencing and for all experimental work.

Genomic DNA isolation. For the preparation of genomic DNA, 50 ml of a C. utilis cell culture grown for 16 h in YPD medium [1 % yeast extract, 2 % peptone, 2 % glucose (w/v)] on a horizontal shaker at 100 r.p.m. at 30 °C was harvested by centrifugation for 5 min at 1800 g, washed once with H2O, and suspended in 4 ml SCE buffer, pH 8 (0.9 M sorbitol, 0.1 M EDTA, 14 mM β-mercaptoethanol), containing 800 µg Zymolyase (MP Biomedicals). After incubation for 2 h at 30 °C, the cells were centrifuged, carefully resuspended in 5.5 ml 50 mM EDTA, pH 7.4, containing 1 % SDS, and incubated for 30 min at 72 °C. After the addition of 1 ml 5 M potassium acetate buffer, pH 6, cells were stored on ice for 60 min and then centrifuged for 15 min at 16 200 g. The supernatant was saved, 2.5 volumes of ethanol were added, and the DNA was precipitated overnight at −20 °C. The nucleic acids were isolated by centrifugation for 15 min at 9600 g, and the pellet was air-dried and dissolved in 4 ml 150 mM sodium acetate buffer, pH 7.5, containing 800 µg RNase A (Qiagen). After incubation for 3.5 h at 37 °C, protein contaminants were removed by phenol/ chloroform extraction, and the genomic DNA was ethanol-precipitated and dissolved in 200 µl 10 mM Tris/HC1, pH 8.

Sequencing and gene prediction. Genomic DNA was sequenced by Goettingen Genomics Laboratory (Goettingen, Germany) on a Roche GS FLX pyrosequencer, and the reads were assembled with the Newbler software. Genes were predicted with the eukaryotic gene finder AUGUSTUS after it had been pre-trained on a Candida albicans dataset (Stanke et al., 2006).

Preparation of secretome samples. C. utilis was pre-cultured in YPD liquid medium on a horizontal shaker at 100 r.p.m. After overnight incubation at 30 °C, 100 ml SD medium [0.67 % yeast nitrogen base without amino acids containing 5 g ammonium sulfate l−1 (w/v) (BD Diagnostics), 50 ml amino acid mix l−1 (0.28 g adenine l−1, 0.48 g arginine l−1, 0.72 g tyrosine l−1, 1.44 g isoleucine l−1, 0.72 g lysine l−1, 1.2 g phenylalanine l−1, 0.72 g valine l−1, 0.72 g threonine l−1, 0.48 g methionine l−1, 0.40 g histidine l−1)] and 2 % glucose (w/v) was inoculated from the pre-culture into a 500 ml Fernbach culture flask to an initial OD595 of 0.1 at 30 °C. For growth on xylose, the medium was supplemented with 2 % (w/v) xylose (SX medium) instead of glucose. For growth in the absence of ammonium sulfate, yeast nitrogen base without amino acids and without ammonium sulfate (BD Diagnostics) was used. Samples for the mass spectrometric analysis of the exponential phase secretome after growth on SD medium were collected after 9 h of incubation, and after 48 h for the stationary phase, by centrifugation for 10 min at 3833 g. Cell pellets were dried at 60 °C and weighed. The supernatant was sterile-filtered (0.2 µm pore-size) to remove residual cells. To concentrate the medium, 10 kDa cut-off spin filters (Vivaspin 20 columns, Sartorius Stedim) were used. All steps were carried out at 4 °C. Protein concentrations were determined by Bradford assay (Smith et al., 1985) with BSA as standard. The concentrated supernatants were subsequently used for MS analyses.

In order to investigate the effects of the different sugars and of ammonium sulfate on the secretome composition, 6 µg of secretory proteins was separated by 10 % SDS-PAGE, and the most prominent protein bands were excised and analysed by MS (ZBA/ZMMK, Cologne, Germany).

PNGase F treatment of medium proteins. To de-N-glycosylate secreted proteins, 4.5 µg of the medium proteins was treated with a PNGase F Deglycosylation kit (New England Biolabs) according to the manufacturer’s protocol. Proteins were subsequently separated by SDS-PAGE, followed by immunoblotting using a polyclonal antisera specific for Suc2 of S. cerevisiae.

Mass spectrometric analyses. The concentrated culture supernatants were analysed as described previously (Sorgo et al., 2010). Briefly, the protein solutions were reduced (10 mM DTT in 100 mM NH4HCO3) and alkylated (65 mM iodoacetamide in 100 mM NH4HCO3) to disrupt and prevent disulfide bridge formation, and were then concentrated by 10 kDa cut-off spin filters (Amicon). Subsequently the proteins were digested for 18 h with Trypsin Gold (Promega), and the peptide concentration was measured at 205 nm on a NanoDrop ND-1000 spectrophotometer (Isogen Life Science) (Desjardins et al., 2009). Peptides (250 ng) were loaded onto an Ultimate 2000 nano-HPLC system (LC Packings) and eluted from a PepMap100 C18 reversed-phase column (75 µm internal diameter, 25 cm length; Dionex) with a 45 min gradient and an increasing acetonitrile (ACN) concentration (0–50 % ACN at a flow rate of 300 nl min−1). The eluting peptides were directly ionized by electrospray in a Q-ToF mass spectrometer (Micromass). The resulting spectra were processed with MassLynx ProteinLynx software and submitted to a licensed in-house version of MASCOT (Matrix Science). Two miscleavages were allowed and the mass tolerance for peptides and MS/MS was set to 0.3 Da. Results were searched against the predicted ORF translation of the C. utilis genome. Probabilistic MASCOT scoring was used, and a p value of less than 0.05 was considered significant for peptide identification. Four biological
replicates of each condition were analysed three times (MS/MS switching time 2 s; 1.5 s; 1.25 s) and compared in a semiquantitative manner described previously by calculating for each condition the mean of the total number of peptide identifications (MOPI) per biological replicate (Sorgo et al., 2010).

The SignalP 3.0 prediction server was used to find signal sequences in the proteins detected by MS. Each protein possessing a signal peptide was screened for the presence of a putative GPI anchor sequence using the big-PI Fungal Predictor (Eisenhaber et al., 2004). Additionally, the TMHMM Server 2.0 was used to identify transmembrane domains (TMDs). S. cerevisiae homologues were found by BLASTP searches (E-value of 1.0e-20) (Altschul et al., 1990).

RESULTS

Genomic sequencing and prediction of *C. utilis* proteins

The draft genome sequence of *C. utilis* strain DSM 2361 (ATCC 9950, CBS 5609) was determined by large-scale parallel pyrosequencing. In total, 12 556 397 bases were sequenced with a 17-fold coverage. A total of 6417 putative ORFs were predicted using the gene finder software AUGUSTUS pre-trained on a *C. albicans* dataset, since this well-characterized species belongs to the same genus. Functional annotations were assigned according to the best hit assignments obtained by BLAST searches in the Swiss-Prot database (E-value of 1.0e-20). Manual curation of the automatic annotation is currently in progress (C. Buerth and others, unpublished results).

Growth phase-dependent composition of the *C. utilis* secretome

To characterize the secretome of *C. utilis*, the cells were cultivated in SD medium to the exponential (OD_{595} 6) and stationary (OD_{595} 18) growth phases. The culture medium was concentrated, and its protein concentration was determined and normalized to the biomass. These measurements revealed that stationary phase cells released 54% more proteins per milligram of biomass than exponential phase cells (Table 1), which may have been caused by cell lysis during the stationary phase. However, since the increase in medium protein concentration occurred during early stationary phase (data not shown), it might also have been caused by increased secretion of proteins.

Table 2 shows the liquid chromatography-electrospray ionization (LC-ESI)-MS/MS identification of proteins present in the culture supernatants. Among the 37 extracellular proteins detected, at least 24 contained a putative signal peptide. In four of the identified proteins the presence of a leader peptide remained unclear, since the corresponding genes in the predicted ORF library were truncated at their 5’ ends. For example, the ORF encoding the Tos1 protein lacked 5’ sequences, but in view of its considerable degree of identity to *S. cerevisiae* Tos1p (43%), a cell wall-bound protein of unknown function, it seems likely that the protein possesses a signal peptide. Likewise, three presumed GPI-anchored proteins, namely a hypothetical protein encoded by ORF g5050, a protein containing a Flo11 domain and the endochitinase Cts1, could not be verified because their sequences were truncated at their 3’ ends, making big-PI predictions of GPI anchor sequences impossible. However, six of the secretory proteins were predicted to contain a GPI anchor signal, including the phospholipase Plb3, the chitin transglycosidase Crh1 and Ecm33, a cell wall protein of unknown function.

For six of the extracellular proteins, BLAST searches did not identify any homologues with annotated functions or conserved domains in other fungal species (E-value of 1.0e-20). The remaining secretory proteins were grouped according to their functional annotations (Fig. 1). Most of the proteins identified were related to cell wall assembly and maintenance, such as (trans)glucanases. Three proteins, the ferro-O₂-oxidoreductase Fet3 (type I TMD) and two proteins possessing either a Tip1 (type I TMD) or a Flo5 (type II TMD) domain, contained a single TMD. Ten proteins (Emg1, Met2, Hht1, Osh6, Rpl8b, Tdh3, Trx1, Zps1, g3555, g4439) had a predicted cellular localization in the cytosol or the nucleus, and probably represent intracellular contaminants. However, a glyceraldehyde-3-phosphate dehydrogenase has been detected extracellularly or in the cell wall of other fungi before (Barbosa et al., 2006; Mattanovich et al., 2009; Stead et al., 2010; Swaim et al., 2008). Osh6 has been found to localize to patch-like or punctate structures in the vicinity of the plasma membrane in *S. cerevisiae*, from where it may be released into the medium (Wang et al., 2005).

In total, 20 proteins were detected in the exponential phase, with three proteins being exclusively present in this growth phase. Interestingly, we identified a peptide with 31%
identity to the signalling mucin CaMsb2, which in *C. albicans* is involved in hyphal formation (Román et al., 2009). In both *C. albicans* and *S. cerevisiae* (E. Schneider & J. F. Ernst, unpublished results; Vadaie et al., 2008), Msb2 contains a single transmembrane domain separating a highly O-glycosylated extracellular N-terminal domain,

<table>
<thead>
<tr>
<th>ORF ID</th>
<th><em>S. cerevisiae</em> homologue (% identity)</th>
<th>Predicted function</th>
<th>SP</th>
<th>GPI</th>
<th>MOPI*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>exp</td>
<td>stat</td>
<td></td>
</tr>
<tr>
<td><strong>Exponential phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g4347</td>
<td>1Ecm33† (44 %)</td>
<td>Cell wall protein of unknown function</td>
<td>Yes</td>
<td>Yes</td>
<td>0.5</td>
</tr>
<tr>
<td>g3158</td>
<td>2Msb2 (27 %)</td>
<td>Mucin family member</td>
<td>NP</td>
<td>No</td>
<td>0.5</td>
</tr>
<tr>
<td>g2592/g281</td>
<td>3/No homologues</td>
<td>Contains domain homologous to Flo11 superfamily</td>
<td>NP</td>
<td>NP</td>
<td>0.25</td>
</tr>
<tr>
<td><strong>Exponential and stationary phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g5908</td>
<td>6Suc2 (Inv1) (48 %)</td>
<td>Invertase, sucrose-hydrolysing enzyme</td>
<td>Yes</td>
<td>No</td>
<td>12.5</td>
</tr>
<tr>
<td>g1121</td>
<td>3Scw4† (67 %)</td>
<td>Cell wall protein with similarity to glucanases</td>
<td>Yes</td>
<td>No</td>
<td>14.5</td>
</tr>
<tr>
<td>g3105</td>
<td>6Bgl2 (65 %)</td>
<td>Endo-β,1-3-glucanase, cell wall maintenance</td>
<td>Yes</td>
<td>No</td>
<td>7.5</td>
</tr>
<tr>
<td>g4105</td>
<td>7Exg2† (56 %)</td>
<td>Exo-β,1-3-glucanase, cell wall β-glucan assembly</td>
<td>Yes</td>
<td>No</td>
<td>5.5</td>
</tr>
<tr>
<td>g165</td>
<td>6Scw10 (60 %)</td>
<td>Cell wall protein with similarity to glucanases</td>
<td>Yes</td>
<td>No</td>
<td>3.5</td>
</tr>
<tr>
<td>g4226</td>
<td>9Tos1 (43 %)</td>
<td>Cell wall protein of unknown function</td>
<td>NP</td>
<td>No</td>
<td>4.5</td>
</tr>
<tr>
<td>g3474</td>
<td>10Uth1† (70 %)</td>
<td>SUN family member, cell wall biogenesis</td>
<td>Yes</td>
<td>No</td>
<td>3.75</td>
</tr>
<tr>
<td>g6254</td>
<td>13Hsp150 (59 %)</td>
<td>Heat-shock protein, cell wall stability</td>
<td>Yes</td>
<td>Yes</td>
<td>2.75</td>
</tr>
<tr>
<td>g943</td>
<td>13Pry1† (64 %)</td>
<td>Protein of unknown function</td>
<td>Yes</td>
<td>No</td>
<td>1.75</td>
</tr>
<tr>
<td>g1681</td>
<td>13Gas1† (55 %)</td>
<td>β-1,3-Glucanosyltransferase, cell wall assembly</td>
<td>Yes</td>
<td>Yes</td>
<td>2.5</td>
</tr>
<tr>
<td>g5579</td>
<td>13Scw11† (64 %)</td>
<td>Cell wall protein with similarity to glucanases</td>
<td>Yes</td>
<td>No</td>
<td>2.25</td>
</tr>
<tr>
<td>g3771</td>
<td>10Fet3 (56 %)</td>
<td>Ferro-O2-oxidoreductase, high-affinity iron uptake</td>
<td>Yes</td>
<td>No</td>
<td>1.5</td>
</tr>
<tr>
<td>g464</td>
<td>11Khn1 (59 %)</td>
<td>Similar to Krc9 (cell wall β-1,6-glucan synthesis)</td>
<td>Yes</td>
<td>No</td>
<td>1.5</td>
</tr>
<tr>
<td>g3498</td>
<td>2Msb2 (27 %)</td>
<td>Mucin family member</td>
<td>NP</td>
<td>No</td>
<td>0.5</td>
</tr>
<tr>
<td>g1137</td>
<td>6Osh6 (66 %)</td>
<td>Oxysterol-binding protein, sterol metabolism</td>
<td>No</td>
<td>No</td>
<td>0.75</td>
</tr>
<tr>
<td>g6286</td>
<td>13No homologue</td>
<td>Hypothetical protein with Tip1 domain</td>
<td>Yes</td>
<td>Yes</td>
<td>0.25</td>
</tr>
<tr>
<td>g174</td>
<td>14Met2 (59 %)</td>
<td>Homoserine O-acetyltransferase, Met biosynthesis</td>
<td>No</td>
<td>No</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>Stationary phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g457</td>
<td>14No homologue</td>
<td>Hypothetical protein</td>
<td>Yes</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>g3555</td>
<td>13No homologue</td>
<td>Hypothetical protein</td>
<td>Yes</td>
<td>No</td>
<td>3.25</td>
</tr>
<tr>
<td>g6140</td>
<td>13Zps1 (32 %)</td>
<td>Zinc- and pH-regulated surface protein</td>
<td>No</td>
<td>No</td>
<td>1.5</td>
</tr>
<tr>
<td>g5411</td>
<td>13Trx1 (68 %)</td>
<td>Thioredoxin isoenzyme, oxidative/reductive stress</td>
<td>No</td>
<td>No</td>
<td>1.5</td>
</tr>
<tr>
<td>g5050</td>
<td>10No homologue</td>
<td>Hypothetical protein</td>
<td>Yes</td>
<td>NP</td>
<td>–</td>
</tr>
<tr>
<td>g4439</td>
<td>8No homologue</td>
<td>Hypothetical protein</td>
<td>No</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>g5643</td>
<td>5No homologue</td>
<td>Contains domain homologous to Flo5</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>g6668</td>
<td>5No homologue</td>
<td>Contains domain homologous to Flo11 superfamily</td>
<td>NP</td>
<td>No</td>
<td>0.75</td>
</tr>
<tr>
<td>g1357</td>
<td>5Crh1† (57 %)</td>
<td>Chitin transglycosylase, induced by cell wall stress</td>
<td>Yes</td>
<td>Yes</td>
<td>0.75</td>
</tr>
<tr>
<td>g1791/g2236</td>
<td>1/13No homologues</td>
<td>Hypothetical protein with Tip1 domain</td>
<td>Yes</td>
<td>No</td>
<td>0.75</td>
</tr>
<tr>
<td>g2175</td>
<td>3Hht1 (100 %)</td>
<td>Core histone protein H3</td>
<td>No</td>
<td>No</td>
<td>0.5</td>
</tr>
<tr>
<td>g6131</td>
<td>10Tdh3† (81 %)</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>No</td>
<td>No</td>
<td>0.5</td>
</tr>
<tr>
<td>g2628</td>
<td>9Pbh3 (57 %)</td>
<td>Lyso phospholipase 3, phospholipid metabolism</td>
<td>Yes</td>
<td>Yes</td>
<td>0.5</td>
</tr>
<tr>
<td>g4412</td>
<td>6Rpl8b (78 %)</td>
<td>Ribosomal protein L4 of the large ribosomal subunit</td>
<td>No</td>
<td>No</td>
<td>0.25</td>
</tr>
<tr>
<td>g3620</td>
<td>6Cts1† (61 %)</td>
<td>Endochitinase, required for cell separation</td>
<td>Yes</td>
<td>NP</td>
<td>0.25</td>
</tr>
<tr>
<td>g19</td>
<td>1Emg1 (83 %)</td>
<td>x/β-Knot fold methyltransferase</td>
<td>No</td>
<td>No</td>
<td>0.25</td>
</tr>
<tr>
<td>g2360</td>
<td>10No homologue</td>
<td>Hypothetical protein</td>
<td>No</td>
<td>No</td>
<td>0.25</td>
</tr>
</tbody>
</table>


*MOPI* refers to *M. oryzae* orthologues indicating the fungal species with the closest homologue: 1 *M. oryzae*, 2 *M. grisea*, 3 *M. grisea*, 4 *M. oryzae*, 5 *M. oryzae*, 6 *M. oryzae*, 7 *M. oryzae*, 8 *M. oryzae*, 9 *M. oryzae*.  

*Abbreviations: exp, exponential growth phase; stat, stationary growth phase; SP, predicted signal peptide; GPI, predicted GPI anchor; NP, signal sequence and/or GPI anchor not predictable due to truncations of the putative ORF.*
which is released by proteolytic cleavage, and a short cytosolic C-terminal tail. Sequence analysis revealed a similar organization of Msb2 in *C. utilis*, and in agreement with this topology, the peptide detected by MS corresponds to the N-terminal portion of the protein (data not shown).

The second protein, with 44% identity to ScEcm33, which in *S. cerevisiae* has a possible role in apical bud growth (Terashima *et al.*, 2003), has already been detected in the extracellular milieu of *Candida glabrata* and *C. albicans* (Stead *et al.*, 2010; Sorgo *et al.*, 2010). The third secreted protein contained a domain that occurs in the Flo11 flocculin, which is required in *S. cerevisiae* for diploid pseudohyphal formation and haploid invasive growth (Lo & Dranginis, 1996).

In total, 34 different proteins were detected within the stationary phase medium. Seventeen proteins were found exclusively in this growth phase, including proteins related to the stress response, such as the thioredoxin Trx1 and the chitin transglycosylase Crh1. In addition, two hydrolases were identified, the phospholipase PlpB and the endochitinase Cts1. Cts1 is involved in cell separation during cytokinesis. Eight out of 17 stationary phase-specific proteins have a predicted localization within the cytosol, suggesting that they are at least partially released by cell lysis or translocated via unconventional secretion pathways.

The semiquantitative analysis revealed that secretory proteins, which were present in both growth phases, were detected with similar abundances in the exponential and stationary phase media (Table 2).

**Effects of different carbon and nitrogen sources on secretome composition**

To study the effects of different carbon and nitrogen sources on the secretome composition, cells were cultivated to the stationary phase in either a glucose- or a xylose-based substrate, with or without ammonium sulfate. Extracellular proteins were separated by SDS-PAGE before mass spectrometric identification of selected bands (Fig. 2), and only two proteins per band with the highest exponentially modified protein abundance indices (emPAIs) (Ishihama *et al.*, 2005) were considered.

*C. utilis* is known to efficiently grow on xylose-based substrates, one of the main advantages of *C. utilis* over other yeast expression systems (Chakravorty *et al.*, 1962). Extracellular proteins from stationary phase cultures grown in SX medium (containing xylose as carbon source) were analysed (Fig. 2a); in parallel, a culture grown in SD medium (containing glucose) was studied similarly. After 48 h, the two cultures reached a final OD<sub>595</sub> of 16 or 18 in SX or SD medium, respectively. This result confirms that *C. utilis*, in contrast to other yeast species, uses xylose as a carbon source very efficiently.

Proteins secreted in SX medium showed a different pattern from those secreted in SD medium. Firstly, in xylose medium, a diffuse major protein band was detected in the range 110–140 kDa, which was less prominent in the supernatants of glucose-grown cells. This difference was more pronounced in media without ammonium sulfate (Fig. 2b). Secondly, in xylose medium, a protein band with a molecular mass of 55 kDa was detected, which was absent in the glucose medium. The most prominent protein bands were excised and subjected to mass spectrometric protein identification. Most proteins were identified in multiple bands, probably due to different isoforms and extensive post-translational modifications, particularly glycosylation. MS analysis identified two different proteins in the 110–140 kDa band, i.e. a protein encoded by the putative ORF g4439 and the invertase Inv1, which has been shown to functionally complement a *SUC2*-deficient *S. cerevisiae* strain (Chávez *et al.*, 1998). Inv1 also represented the major protein in the xylose-specific 55 kDa band. Consistent with our observations, *C. utilis* has been shown to secrete two variants of invertase, a 60 kDa unglycosylated F-form (fast-migrating) and a heavily glycosylated S-form (slow-migrating), consisting of two identical 150 kDa monomers (Belcarz *et al.*, 2002). Invertase was also identified in a third band with a molecular mass of about 72 kDa, which...
presumably contains an incompletely glycosylated Inv1 species. Moreover, in SD medium without ammonium sulfate, Inv1 was detected in a fourth band with a molecular mass of about 48 kDa (Fig. 2b).

Inv1 contains a total of 13 potential N-glycosylation sites, one of them located within the predicted signal peptide (Chávez et al., 1998). In order to verify N-glycosylation of Inv1 experimentally, we treated the supernatant obtained after growth in either a glucose- or a xylose-based substrate with the endoglycosidase PNGase F. Proteins were then separated by SDS-PAGE, which was followed by immuno-blotting using a ScSuc2 polyclonal antiserum (Fig. 2c). In SX medium, a strong signal ranging from 110 to 140 kDa was detected before PNGase F treatment, corresponding to the S-form of Inv1. This signal was much less pronounced in SD medium. After de-N-glycosylation this signal disappeared, whereas a double band was detected with a mass of about 65 kDa, which corresponds to the apparent molecular mass of 60 kDa of the deglycosylated Inv1 found by Chávez et al. (1997). In addition, about six further signals appeared in the lower molecular mass range. We assume that these signals either arose from truncated forms of Inv1 or represent proteolytic fragments, although proteolytic enzymes were not detected in the supernatants analysed.

One protein, which was exclusively detected in the xylose supernatant(s), showed identity to the putative endomannosidase Dfg5, which in S. cerevisiae is involved in cell growth and cell wall biogenesis, possibly mediating the transfer of GPI-anchored cell wall proteins from the plasma membrane to the cell wall (Kitagaki et al. 2002). In C. albicans, Dfg5 is an N-glycosylated, GPI-anchored protein which is involved in cell growth and hyphal formation (Spreghini et al., 2003).

Notably, in both SD and SX media, C. utilis cultures reached the same final cell densities after 48 h of growth, irrespective of whether cells were grown in the presence or absence of ammonium sulfate (data not shown). Moreover, the protein patterns obtained by SDS-PAGE did not significantly differ in the two media (Fig. 2a, b). This finding demonstrates how efficiently C. utilis utilizes amino acids as a sole source of nitrogen. All proteins identified in the media without ammonium sulfate had already been detected by LC-ESI-MS/MS (Table 2) or by SDS-PAGE analysis of media with ammonium sulfate (Fig. 2a), except for a protein with 73 % identity to the cell wall asparaginase Asp3, which was detected in numerous protein bands. In S. cerevisiae, Asp3 is a glycosylated mannan cell wall protein that converts L-asparagine to ammonia and aspartate, and is induced upon nitrogen starvation (Kim et al., 1988; Huang et al., 2010).

CUG decoding in C. utilis

Among species of the genus Candida, some, including the human fungal pathogen C. albicans, translate the leucine

Fig. 2. SDS-PAGE analysis of the extracellular proteome of stationary phase cells of C. utilis grown in the presence of either 2 % glucose (SD medium) or 2 % xylose (SX medium), with (a) or without (b) ammonium sulfate. Each lane was loaded with 6 μg of medium proteins and the gel was stained with Coomassie blue. The indicated bands were subjected to MS analysis for peptide identification. Molecular weights (MW) are indicated in kDa. (c) Immunoblot of medium proteins using an anti-ScInv1 serum before and after treatment with the endoglycosidase PNGase F.
CUG codon as serine by a novel serine tRNA (Massey et al., 2003). Consequently, in C. albicans, serine CUG decoding prevents the functional expression of reporter proteins from CUG-containing transcripts, including GFP, *Escherichia coli* β-galactosidase and *S. cerevisiae* orotidine-5′-phosphate decarboxylase (Ura3p), unless adapted to the codon usage of *C. albicans* (Cormack et al., 1997).

In *C. utilis*, genetic code alterations would interfere with the expression of heterologous genes derived from organisms that use the standard genetic code. To explore whether *C. utilis* belongs to the CUG clade of the genus *Candida* we created a phylogenetic tree based on 18S rRNA sequences of several yeast species, including *C. utilis* and several members of the CUG clade, and this indicated that *C. utilis* uses the standard genetic code (Fig. 3a). To verify this hypothesis we used the peptide identification data obtained by the mass spectrometric analyses to search for tryptic peptides containing leucine or serine residues. As illustrated in Fig. 3(b), one of the detected Inv1 secretory peptides contains a leucine, which is encoded by a CUG codon. Likewise, we detected a secretory Met2 peptide containing a CUG codon, which was also found to be translated to leucine (data not shown), indicating that in contrast to *C. albicans*, *C. utilis* translates CUG as leucine according to the standard genetic code.

**Secretome comparison among yeast species**

In the last few years the secretomes of several yeast species of biotechnological or medical relevance have been characterized (Swaim et al., 2008; Madinger et al., 2009; Mattanovich et al., 2009; Stead et al., 2010; Sorgo et al., 2010). To define the core secretome of these yeasts we compared the secretomes of *C. utilis*, *K. lactis*, *P. pastoris*, *C. albicans* and *C. glabrata*.

Yeast species important in biotechnology differ considerably in the number of secreted proteins (Fig. 4a), and this may be only partially caused by different experimental setups. Interestingly, we found a subset of 10 proteins that are present in all three secretory proteomes. The largest functional group consists of the glucanases Scw4, Scw11, Bgl2 and Exg2 or its isofrom Exg1 (*K. lactis* and *P. pastoris*). Additionally, the core secretome contains the transglucosylase Gas1, the chitinase Cts1, the SUN family member Uth1, the glyceraldehyde-3-phosphate dehydrogenase Tdh3 or its isofrom Tdh2 (*K. lactis*), and two cell wall proteins of unknown functions, Tos1 and Pry1 or its isofrom Pry2 (*P. pastoris*). Notably, except for the highly abundant proteins Tdh3 and Pry1, eight out of the 10 core secretome proteins have a function related to cell wall assembly or function.

We also compared the secretome of *C. utilis* with the published secretomes of its pathogenic *Candida* relatives.

---

**Fig. 3.** (a) Phylogenetic tree based on 18S rRNA sequences, showing the relationships of the CUG clade of the genus *Candida*. The phylogenetic tree was constructed according to Dereeper et al. (2008). (b) Section of the *C. utilis* INV1 gene encoding a tryptic peptide that was detected by MS (blue box). The CUG codon, which was found to be translated to leucine, is highlighted in red.
C. utilis and C. glabrata (Fig. 4b). Interestingly, the secretome of C. utilis turned out to be very similar to the secretome of C. albicans grown in sucrose as carbon source (Sorgo et al., 2010). In contrast to C. utilis, C. albicans also secretes proteins related to pathogenicity, such as the secreted aspartyl proteases of the Sap family (Parra-Ortega et al., 2010), and the Op4 protein, which is involved in phenotypic switching (Soll et al., 1994). Correspondingly, C. glabrata, a second pathogenic Candida species, has also been shown to secrete mainly cell wall proteins, cell wall-modifying enzymes and aspartyl proteinases after growth in a glucose-containing medium (Stead et al., 2010).

**DISCUSSION**

In the last few years it has been demonstrated that C. utilis represents a promising expression host, generating high levels of recombinant proteins (Kondo et al., 1997; Miura et al., 1998, 1999; Ikushima et al., 2009; Wei et al., 2010). However, up to now, heterologous protein production has only been carried out intracellularly. Since extracellular protein production simplifies downstream processing, because of fewer by-products and simplified purification protocols, we analysed the secretome of C. utilis to provide basic knowledge of the secretory potential of this yeast. In the C. utilis secretome we identified a total of 37 proteins by direct LC-ESI-MS/MS analysis, which, in comparison with the secretomes of K. lactis with 81 and P. pastoris with 20 extracellular proteins, represents a moderate number of potentially contaminating proteins. However, it should be considered that the K. lactis secretome (Swaim et al., 2008; Madinger et al., 2009) was characterized by the sensitive 2D-LC methodology, while the P. pastoris secretome (Mattanovich et al., 2009) was analysed by MS identification subsequent to gel electrophoresis, which may underestimate the detection of highly glycosylated proteins.

The number of extracellular proteins in C. utilis depended greatly on the growth phase. During exponential growth, 20 proteins were detected, whereas in the stationary phase, 34 different proteins were released from the cells. This increase appears to be largely due to cell lysis, since half of the proteins found only in stationary phase medium are not predicted to enter the secretory pathway. In contrast, nearly 90 % of the proteins detected in both growth phases are predicted secretory proteins. In agreement with the occurrence of cell lysis in the stationary phase, the chitin transglycosylase Crh1 was detected, which itself is predicted to be secretory but is induced in response to cell wall damage (Bermejo et al., 2008). Another stationary phasespecific, non-secretory protein that we detected was the thioredoxin Trx1, which is also induced in stationary phase cultures of S. cerevisiae (Jakubowski et al., 2000) and upon oxidative stress in C. albicans (Enjalbert et al., 2003). Furthermore, the highly abundant glycolytic protein Tdh3 was detected in stationary phase medium, which agrees with similar findings in other yeast species (Swaim et al., 2008; Mattanovich et al., 2009; Stead et al., 2010). The same holds true for cytosolic proteins involved in ribosome assembly, such as Rpl8B and Emg1. However, considering both the semiquantitative MOPI values and the amounts of protein in SDS-PAGE, it can be assumed that the nonsecretory proteins are not abundant in stationary phase medium. The most prominent secretory proteins were found with similar abundances in both the exponential and the stationary growth phases.

Classification of the identified secretory proteins according to their functional annotations revealed that most of the proteins had as yet unknown functions. The next most abundant group were (trans)glucanases, which is consistent with the composition of the known secretomes of other yeast species (Swaim et al., 2008; Madinger et al., 2009; Mattanovich et al., 2009; Sorgo et al., 2010; Stead et al., 2010). One secretory protein was detected that was expressed in response to nutrient deprivation. In S. cerevisiae, the ferroxidase Fet3 is part of a cell-surface high-affinity iron-uptake system, which imports iron during growth in low-iron media (Askwith et al., 1994). In agreement with this, Fet3 was more abundantly produced in the stationary phase medium, in which iron becomes limiting.

As is typical for extracellular proteomes of yeasts, we identified several proteins involved in assembly or
maintenance of the cell wall. We found non‐covalently bound cell wall proteins, including Scw4, Scw10, Scw11 and Bgl2, which presumably are released by shaking-induced shear-stress. A second group contained six GPI-anchored proteins covalently attached to the cell wall, including the phospholipase Plb3 and the chitin transglycosidase Crh1. The presence of an extracellular phospholipase B in C. utilis has already been described (Fujino et al., 2006). In a previous study of the C. albicans secretome it was proposed that GPI-anchored proteins may be released into the medium before they are covalently attached to cell wall carbohydrates, during cell wall remodelling (Sorgo et al., 2010) or by partial proteolysis (Swaim et al., 2008). Miller et al. (2010) obtained evidence for the release of some misfolded or excess GPI-anchored proteins by membrane-bound yapsins in S. cerevisiae. Notably, the secretome of C. utilis, unlike those of many other fungal species, did not contain any proteolytic enzymes. The presence of proteolytic enzymes in the secretome could potentially lower the yield of recombinant proteins produced by protease-secreting fungi.

In a preliminary approach, we used our draft genomic sequence to determine the number of proteins which would be predicted to enter the secretory pathway (K. Wolstencroft, personal communication). Among the 6417 predicted ORFs, 403 were found to encode proteins which possessed a signal peptide, 91 additionally contained at least one transmembrane domain and 33 contained a GPI anchor. These findings are in agreement with previous studies of the secretomes of K. lactis (Swaim et al., 2008), P. pastoris (Mattanovich et al., 2009) and C. albicans (Lee et al., 2003), showing that computational analyses of the hypothetical secretomes tend to overestimate the number of secreted proteins. Most likely, this discrepancy is caused by the strict regulation of the production of many secreted proteins.

For K. lactis it has recently been shown that the carbon source has a significant impact on the composition of the extracellular proteome (Madinger et al., 2009). Likewise, our SDS-PAGE analysis indicated that C. utilis secretes a different set of proteins when grown in xylose-based substrates. Most noticeably, growth on xylose led to an induction of the invertase Inv1, particularly of the hyperglycosylated S-form of Inv1, confirming earlier studies (Dworschack & Wickerham, 1961). Invertase induction by xylose and N-glycosylation of Inv1 could be verified by treatment with PNGase F. This mode of invertase regulation is unlike that of S. cerevisiae (Hackel & Khan, 1978; Roca et al., 2004) or Aspergillus nidulans (Vainstein & Peberdy, 1991), in which xylose inhibits invertase production. Chávez et al. (1997) only detected a periplasmatic invertase with an approximate molecular mass of 150 kDa in C. utilis grown in glycerol and glucose. Belcarz et al. (2002) identified two different forms of invertase in C. utilis grown on 1% sucrose, one within cells and one in the culture medium, consisting of a glycosylated S-form and a non-glycosylated F-form. Here, we found a third form with a molecular mass of about 72 kDa, which may represent a partially glycosylated form, e.g. only containing N-glycan chains lacking outer chains connected to the 13 potential N-glycosylation sites of Inv1. In addition, Inv1 was identified in a fourth band with a molecular mass of about 45 kDa. Secretion of the apparently non-glycosylated F-form of invertase in C. utilis is surprising and contrasts with results obtained for S. cerevisiae and other fungal species. The invertase promoter and its secretion signal sequence may therefore become convenient tools to mediate regulated secretion of desired recombinant proteins in C. utilis. However, regulation of the invertase promoter is not fully understood in C. utilis. Chávez et al. (1998) identified putative binding sites for the Mig1 repressor in the 5′ sequence of INV1. Moreover, they found that invertase activity was decreased in the presence of high glucose concentrations, which suggests that INV1 is regulated by carbon catabolite repression, as occurs in S. cerevisiae.

C. utilis cultures reached roughly the same cell densities after growth on glucose and on xylose. However, growth rates during exponential growth were lower in SX medium (0.34 h⁻¹) compared with SD (0.45 h⁻¹). Moreover, cultures grown in xylose had an extended lag phase, since overnight cultures were prepared in YPD and cells had to adapt to the new carbon source. Glucose-grown cells entered stationary phase approximately 15 h before cells grown on xylose. In order to exclude that the retarded growth on xylose led to the aberrant secretome composition described, we also compared secretomes of cultures grown for 48 h on xylose with those grown only for 33 h on glucose. The protein profiles obtained were the same as those described for supernatants which were analysed after the same period of time (data not shown).

SDS-PAGE analysis of culture supernatants of cells grown with or without ammonium sulfate revealed that the L-asparaginase Asp3 is regulated by nitrogen starvation, since it was detected in five of six different protein bands in SD medium only in the absence of ammonium sulfate. The presence of an extracellular L-asparaginase in C. utilis after growth without ammonium sulfate was demonstrated by Kil et al. (1995) and is confirmed here. In S. cerevisiae, Asp3 is known to convert asparagine into aspartate and ammonia, and to be regulated by nitrogen catabolite repression (Kim et al., 1988; Huang et al., 2010).

Those members of the genus Candida that belong to the CUG clade translate the CUG codon to serine instead of leucine. Before this code alteration in C. albicans was discovered, serine CUG decoding delayed the use of functional reporter proteins for studies of gene expression or protein localization. Moreover, in C. utilis, serine CUG decoding would interfere with the production of heterologous target proteins, making synthetic adaptation of the gene sequence necessary. However, both comparative rRNA sequence analysis and our mass spectrometric data indicate for the first time that C. utilis is not a member of the CUG clade and uses the standard genetic code.
Recently, the secretomes of several yeast species have been analysed (Swaim et al., 2008; Madinger et al., 2009; Mattanovich et al., 2009; Sorgo et al., 2010; Stead et al., 2010). Comparisons of the secretome of C. utilis with those of K. lactis and P. pastoris grown in a glucose-containing medium identified 10 proteins that occur in all three species. This core secretome is mainly composed of proteins involved in cell wall-related functions, including glucanases, transglucosylases and chitinases. P. pastoris has been shown to secrete the lowest number of proteins, while K. lactis secretes four times as many proteins as P. pastoris and twice as many proteins as C. utilis. The secretome of K. lactis was determined after fed-batch fermentation at 30 °C and pH 6, whereas P. pastoris was cultivated at 25 °C and pH 5. Both species were cultivated on minimal medium with glucose as the limiting carbon source. Although these results may be due in part to different experimental set-ups, we tentatively conclude that C. utilis, because of its relatively moderate number of native secretory proteins, has significant potential as a secretory platform. Comparisons of the C. utilis secretome with those of its pathogenic Candida relatives C. glabrata (Stead et al., 2010) and C. albicans (Sorgo et al., 2010) also revealed extensive similarities, although Sorgo et al. (2010) used sucrose instead of glucose as carbon source. Thus, the basal extracellular proteomes are highly conserved among different yeast species.

ACKNOWLEDGEMENTS

We thank Dr Katy Wolsentcroft (School of Computer Science, University of Manchester) for her work on the predicted secretome, and Dr Andrea Thuermer, Goettingen Genomics Laboratory (Goettingen, Germany), for kindly supporting the processing of the genome sequence. This work was supported by the Cluster of Industrial Biotechnology NRW CLIB2021 and by the EU program FP7-214004-2 FINSysB.

REFERENCES


Edited by: D. Mattanovich

http://mic.sgmjournals.org

2503

IP: 54.70.40.11
On: Sat, 08 Dec 2018 02:42:52