Analysis of the *Candida albicans* Phosphoproteome

S. D. Willger,a Z. Liu,b R. A. Olarte,c M. E. Adarno,D J. E. Stajich,d L. C. Myers,e A. N. Kettenbach,b,d D. A. Hogana

Department of Microbiology and Immunology, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USAa; Department of Biochemistry, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA; Department of Plant Pathology and Microbiology, University of California, Riverside, California, USA; Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA

*Candida albicans* is an important human fungal pathogen in both immunocompetent and immunocompromised individuals. *C. albicans* regulation has been studied in many contexts, including morphological transitions, mating competence, biofilm formation, stress resistance, and cell wall synthesis. Analysis of kinase- and phosphatase-deficient mutants has made it clear that protein phosphorylation plays an important role in the regulation of these pathways. In this study, to further our understanding of phosphorylation in *C. albicans* regulation, we performed a deep analysis of the phosphoproteome in *C. albicans*. We identified 19,590 unique peptides that corresponded to 15,906 unique phosphosites on 2,896 proteins. The ratios of serine, threonine, and tyrosine phosphosites were 80.01%, 18.11%, and 1.81%, respectively. The majority of proteins (2,111) contained at least two detected phosphorylation sites. Consistent with findings in other fungi, cytoskeletal proteins were among the most highly phosphorylated proteins, and there were differences in Gene Ontology (GO) terms for proteins with serine and threonine versus tyrosine phosphorylation sites. This large-scale analysis identified phosphosites in protein components of Mediator, an important transcriptional coregulatory protein complex. A targeted analysis of the phosphosites in Mediator complex proteins confirmed the large-scale studies, and further in vitro assays identified a subset of these phosphorylations that were catalyzed by Cdk8 (Ssn3), a kinase within the Mediator complex. These data represent the deepest single analysis of a fungal phosphoproteome and lay the groundwork for future analyses of the *C. albicans* phosphoproteome and specific phosphoproteins.

*Candida* species are both human commensals and the most common human fungal pathogens. Candidiasis includes both superficial and invasive fungal infections (1–3). *Candida albicans* is the most common cause of invasive candidiasis worldwide (4), and this disease is associated with very high mortality rates, as high as 35% in some populations (5).

In the context of the host, *C. albicans* is exposed to numerous physical and chemical signals that can trigger a variety of responses, including morphological transitions between yeast and filamentous forms, switching between different cell types, the formation of multicellular structures, including biofilms, and altered filamentous forms, switching between different cell types, the formation of multicellular structures, including biofilms, and altered cell wall states (6–9). The ability of *C. albicans* to respond to different stimuli contributes to its virulence (3). Hyphal growth and the coordinated expression of hypha-specific genes are important virulence traits, since they participate in surface-associated growth and escape after phagocytosis (10, 11). Many kinases and phosphatases of biological importance have been described (12–15). Hyphal morphology in *C. albicans* is initiated and maintained by the Ras1-Cyr1 (adenylate cyclase)-protein kinase A (PKA) pathway, in which the PKA catalytic subunits Tpk1 and Tpk2 positively regulate hyphal growth (16–18). Mitogen-activated protein (MAP) kinase pathways also control many important processes, including filamentation, mating, biofilm formation, and stress resistance, through phosphorylation/dephosphorylation cycles (19).

Advances in mass spectrometry (MS)-based proteomics technologies and phosphopeptide enrichment methods have greatly enhanced our ability to identify and characterize phosphorylation sites (P sites) in the proteome of any organism with a sequenced genome (20–22). In many fungi, phosphoproteome characterization is at an early stage. The best-characterized fungal phosphoproteome is that in *Saccharomyces cerevisiae*. Using 12 publicly available data sets, Amoutzias et al. (21) identified 2,781 phosphorylated proteins with 9,783 unique P sites, with an average of four P sites per phosphoprotein. In the fungus *Fusarium graminearum*, 2,902 putative phosphopeptides on 1,496 different proteins were identified (23). Phosphoproteomics data sets for the fungi *Cryptococcus neoformans* (24), *Aspergillus nidulans* (25), *Alternaria brassicicola* (26), *Botrytis cinerea* (26), *Neurospora crassa* (27), and *Schizosaccharomyces pombe* (28) have also been published.

Here we report the first large-scale analysis of the phosphoproteome in *C. albicans* in cells grown under hypha-inducing conditions. We observed 19,590 unique P sites on 2,896 proteins. Consistent with phosphoproteomic analyses in *S. cerevisiae*, there was an enrichment of cytoskeleton proteins among those that were highly phosphorylated. P sites within the central regulatory Ras1-Cyr1-PKA signaling pathway in *C. albicans* were also found. Differences in the types of proteins, categorized by Gene Ontology (GO) term analysis, with serine and threonine P sites versus tyrosine P sites were identified. P sites within the global transcriptional coregulatory Mediator complex were found both in the large-scale data set and in a targeted follow-up study. Cdk8 is the sole kinase component of Mediator, and subsequent in vitro assays identified P sites within Mediator that were phosphorylated by Cdk8 in vitro at sites consistent with the previously described

Cdk8 target motif (phospho-S/T-P) (29–33). This data set gives a first in-depth analysis of the *C. albicans* phosphoproteome and is the largest fungal phosphopeptide data set at this time. These data should serve as an important resource for researchers studying pathways that are regulated by phosphorylation.

**MATERIALS AND METHODS**

**Preparation of *C. albicans* lysates.** A single colony of *C. albicans* strain SC5314 (34) was grown overnight in 50 ml of yeast extract–peptone–dextrose (YPD) medium. From that overnight culture, four flasks with 1 liter of YPD containing 5 mM N-acetylglucosamine (GlcNAc) were each inoculated to an optical density at 600 nm (OD600) of 0.05 and then incubated at 37°C at 150 rpm for 4 to 5 h. The cells were collected by centrifugation in 250-ml centrifugation cups for 5 min at 6,000 rpm. Cell pellets were washed once in 20 ml of phosphate-buffered saline and then pelleted in 50-ml Falcon tubes and stored at −80°C.

**Enrichment and analysis of *C. albicans* phosphopeptides and mass spectrometric analysis.** Lysophilized *C. albicans* powder was homogenized in ice-cold lysis buffer containing 8 M urea, 25 mM Tris-HCl (pH 8.1), 150 mM NaCl, phosphatase inhibitors (2.5 mM beta-glycerophosphate, 1 mM sodium fluoride, 1 mM sodium orthovanadate, 1 mM sodium molybdate, 1 mM sodium tartrate), and protease inhibitors (1 mini-complete EDTA-free tablet per 10 ml of lysis buffer; Roche Life Sciences) by means of sonication. The lystate was clarified by centrifugation, reduced, alkylated, and trypsin digested (35). After overnight digestion with trypsin at 37°C, the digest was acidified with trifluoroacetic acid (TFA), and peptides were desalted and lysophilized (22). Two rounds of phosphopeptide enrichment using titanium dioxide microspheres were performed, and phosphopeptides were separated by strong cation exchange chromatography (SCX) (22). After separation, the SCX fractions were dried, desalted, combined, and analyzed by liquid chromatography–tandem mass spectrometry (LC-MS/MS). LC-MS/MS analysis was performed on a Q-Exactive Plus mass spectrometer (Thermo Scientific) equipped with an Easy-NC-1000 (Thermo Scientific). Peptides were re-dissolved in 5% acetonitrile (ACN)–1% formic acid and loaded onto a trap column (ReproSil C18 AQ, 3 μm, 200-Å pore [Dr. Maisch, Ammerbuch, Germany]) and eluted across a flexrile analytical resolving column (30-cm length, 100-μm inner diameter, ReproSil C18 AQ, 3 μm, 200-Å pore) pulsed in-house with a 90-min gradient of 4 to 30% LC-MS buffer B (LC-MS buffer A includes 0.0625% formic acid and 3% ACN; LC-MS buffer B includes 0.0625% formic acid and 95% ACN) at 400 nl/min. Raw data were searched using COMET in high-resolution mode (36) against a target decoy (reversed) (37) version of the *C. albicans* proteome sequence database with a precursor mass tolerance of ±1 Da and requiring tryptic peptides with up to two miscleavages, carbamidomethylcysteine as fixed modification, and oxidized methionine and phosphorylated serine, threonine, and tyrosine as variable modifications. The *C. albicans* proteome sequence database was generated by combining the UniProt *C. albicans* database (downloaded February 2013; 18,226 total [forward and reverse] proteins) with proteins unique to the Candida Genome Database (CGD) (at least one amino acid difference between the sequence in the CGD and any sequence in the UniProt *C. albicans* database [downloaded February 2013]). The resulting peptide spectral matches were filtered to a <1% false-discovery rate (FDR) based on reverse-hit counting, mass measurement accuracy (MMA) within ±3.5 ppm, a delta-Xcorr (dCn) of more than 0.12, and appropriate Xcorr values for +2- and +3-charge state peptides (37). Probability of phosphorylation site localization was determined by PhosphoRS (38). Peptide motif analysis was performed based on reference 39.

**GO term analysis.** The Gene Ontology (GO)Slim annotation was used for GO term analysis. The gene association file, created 15 December 2014, was downloaded from the CGD website (http://www.candidagenome.org), and only annotations assigned to *C. albicans* (taxon 5476) were used. In total, 6,314 unique genes had at least one associated GO term and served as the background distribution of observed gene ontologies for the *C. albicans* genome in our study. GO enrichment analysis of the total phosphoprotein (P-all), phospho-S/T (P-S/T), and phospho-Y (P-Y) were evaluated independently using the gene ontologies Biological Process, Cellular Component, and Molecular Function. GO::TermFinder (40) as implemented via The Generic GO Term Finder (http://go.princeton.edu/cgi-bin/GOTermFinder) was used to test for ontology enrichment among phosphorylated peptides. To determine significance of enrichment of terms, a Bonferroni corrected P value cutoff of 0.05 was used. GO::TermFinder calculates a P value using the hypergeometric distribution comparing the number of terms found in a set (phosphorylated genes) to the overall background of the genome (40). No evidence codes were excluded from the analysis. The CGD Gene Ontology Slim Mapper available from the CGD website and an R script (GOstats.R, bioconductor) in which a GSEA/GHyperParams function is used for calculating a P value were also used to determine significant GO term enrichment, and similar term enrichments were found. Heat maps of the categories were generated by calculating nonparametric ranking of percent values for each GO term using the “heatmap.2” function in the “gplots” package (41) in R (R Foundation for Statistical Computing, Vienna, Austria).

**Strain construction and Mediator complex purification.** Wild-type *C. albicans* Mediator complex was purified from cTTR03 through 6×His-3×Flag-tagged Med8 (42). To allow for the affinity purification of Mediator complex from mutant strains, one or both copies of *MED8* were fused to a sequence encoding a C-terminal 6×His-3×Flag tag. In addition, Med8 was tagged in backgrounds lacking different Mediator components, including AZC52 ([med12Δ [42]], AZC63 ([med9Δ [42]], CTR01 ([med15Δ [42]], and SN913 ([ssn3Δ [43]]) using PCR products amplified by primer pair LM025/LM014 (44) from the plasmids pFA6-a6×3×His-FLAG-HIS1 and pFA6-a6×3×His-3×Flag-ARG4. These two template plasmids were generated by replacing the SAT1 marker on pFA6-a6×3×His-3×Flag-SAT1 (44) between the AscI and Pmel sites with HIS1 and ARG4 markers subcloned from pFA6-a6×3×His-3×Flag-TAP-HIS1 and pFA6-a6×3×Flag-TAP-ARG4 (45), respectively. Similarly, both copies of *SSN3* and *SSN8* were 6×His-3×Flag tagged in BWPL17 (46) by using primer pairs SSN3tag_sense/SSN3tag_anti and SSN8tag_sense/SSN8tag_anti (47) to generate yLM256 and yLM255, respectively.

Each species of Mediator complex used in this study was isolated from cultures grown at 30°C in YPD without addition of any filamentous-growth-inducing agents. The Mediator complex was purified from the high-salt elution fraction of the heparin column by Flag affinity purification through the 6×His-3×Flag tag placed on the C-terminus of the Med8 (cTTR03) or Sn3 (yLM256) subunit as previously described (44). Free Cdk8 module, containing only Sn3/Cdk8, Sn8/CycC, Sn5/Med12, and Sn2/Med13, was purified from the heparin flowthrough fraction of yLM255 lysate by Flag purification. The Flag purification method was identical to that described above except that the salt concentration of the heparin flowthrough was adjusted from 150 mM potassium acetate to a sequence encoding a C-terminal 6×His-3×Flag tag. In addition, Med8 was tagged in backgrounds lacking different Mediator components, including AZC52 ([med12Δ [42]], AZC63 ([med9Δ [42]], CTR01 ([med15Δ [42]], and SN913 ([ssn3Δ [43]]) using PCR products amplified by primer pair LM025/LM014 (44) from the plasmids pFA6-a6×3×His-FLAG-HIS1 and pFA6-a6×3×His-3×Flag-ARG4. These two template plasmids were generated by replacing the SAT1 marker on pFA6-a6×3×His-3×Flag-SAT1 (44) between the AscI and Pmel sites with HIS1 and ARG4 markers subcloned from pFA6-a6×3×His-3×Flag-TAP-HIS1 and pFA6-a6×3×Flag-TAP-ARG4 (45), respectively. Similarly, both copies of *SSN3* and *SSN8* were 6×His-3×Flag tagged in BWPL17 (46) by using primer pairs SSN3tag_sense/SSN3tag_anti and SSN8tag_sense/SSN8tag_anti (47) to generate yLM256 and yLM255, respectively.

**Analysis of Cdk8-dependent phosphorylation of Mediator.** For the analysis of *C. albicans* mediator Cdk8 kinase assay, trichloroacetic acid (TCA)-precipitated, reduced, and alkylated proteins were trypsin digested and dried. Reductive dimethyl labeling was performed essentially as described previously (48, 49). Dried peptides were resuspended in 100 μl of 100 mM triethyl ammonium bicarbonate (TEAB) by vortexing. For “light” labeling, 0.7 μl of 37% formaldehyde was added to the sample, followed by vortexing and addition of 4 μl of 0.6 M cyanoborohydride. For “heavy” labeling, 1 μl of 30% formaldehyde-13C2 was added to the

**Candida albicans** Phosphoproteome

May 2015 Volume 14 Number 5 Eukaryotic Cell ec.asm.org 475

Downloaded from http://ec.asm.org on October 24, 2017 by guest
resulting peptide spectral matches were filtered to a 1% FDR, based on reverse hit. Probability of phosphorylation site localization was determined by PhosphoRS. We identified 29,718 P sites (15,905 unique) on 19,590 unique peptides that can be clustered to 2,896 proteins. Probability of phosphorylation site localization was determined by PhosphoRS. We identified a total of 29,718 P sites (15,905 unique) on 19,590 unique peptides that can be clustered to 2,896 proteins.

FIG 1 Overview of the analytical workflow. C. albicans cells, grown as hyphae, were harvested, lyophilized, and ground to a fine powder. After trypsin digestion, peptides were enriched by two rounds of TiO$_2$ treatment, separated by a strong cation exchange (SCX) chromatography, and analyzed by LC-MS/MS. For peptide identification, spectra were searched against a custom C. albicans database using COMET. Probability of phosphorylation site localization was determined by PhosphoRS. We identified a total of 29,718 P sites (15,905 unique) on 19,590 unique peptides that can be clustered to 2,896 proteins.

FIG 2 Phosphorylation sites per peptide and distribution of phosphorylated amino acids. (A) Percentage of singly, doubly, triply, and quadruply phosphorylated peptides; (B) distribution of unique phosphosites by modified amino acids: serine (pS), threonine (pT), and tyrosine (pY) sites.

Eighty-six percent of the unique peptides (19,590 out of 22,783) identified contained a phosphorylation site (P site), indicating strong enrichment for phosphorylated peptides; this level of enrichment is similar to what has been reported in other studies (22). We found that 66.3% of identified phosphopeptides had a PhosphoRS score of 0.99 or higher, which corresponds to a false-localization rate (FLR) of approximately 1% (38, 52). The 19,590 phosphopeptides had an average length of 18.5 amino acids (aa), with the range being between 7 and 48 amino acids (see Table S2 in the supplemental material). We found that 11,348 (57.9%) peptides had one P site, 6,964 peptides (35.5%) had two P sites, 1,210 peptides (6.2%) had 3 P sites, and 68 peptides (0.3%) had 4 P sites (Fig. 2A).

The 19,690 phosphopeptides mapped to 15,905 unique P sites (see Table S2). We identified 12,736 (80.1%) serine, 2,881 (18.1%) threonine, and 288 (1.8%) tyrosine phosphorylation sites (Fig. 2B). This distribution is similar to what has been observed in other eukaryotic organisms, including C. neoformans (S, 81.7%; T, 17%; and Y, 1.3%) (24), Tetrahymena thermophila (S, 80.5%; T, 15.7%; and Y, 3.8%) (53), and humans (S, 85.3%; T, 12.9%; and Y, 1.8%) (54). Phosphorylation can also occur on aspartate, lysine, arginine, and histidine residues; however, those phosphorylations are comparatively difficult to detect under our experimental conditions due to their labile character under acidic conditions (55, 56). While our analysis approach was not designed to look for phosphohistidines, C. albicans does have three histidine kinases, Nik1, Snl1, and Chk1, which can autophosphorylate in response to specific cues in order to initiate a phosphorylation cascade that results in a change in transcriptional regulation (57, 58). We detected 6 P sites (3 serine and 3 threonine) on Nik1 and 25 P sites (22 serine and 3 threonine) on Snl1; Chk1 peptides were not detected (see Table S3 in the supplemental material).

Identification of phosphorylated proteins. To assign protein...
identities to the peptide spectra, we assembled a custom database by merging the Swiss-Prot/TrEMBL C. albicans database (59) and the Candida Genome Database (CGD) (60). Of the 19,590 detected phosphopeptides, 13,825 matched a single identifier (ID) in our database; 5,765 phosphopeptides matched with more than one ID. Peptides with two IDs generally reflected UniProt IDs for the same protein referring to the two alleles, since C. albicans is a diploid organism. For the peptides assigned three or more IDs, often there was a previous CGD ID that could be converted into one of the two UniProt IDs. We identified 54 phosphopeptides with 6 or more IDs. Interestingly, those peptides can be clustered into 3 groups. The first group consists of 42 phosphopeptides with the same amino acid sequence, varying only in phosphorylation pattern, within proteins belonging to a set of fungus-specific putative Rho GTPase-activating proteins (see Table S2). The second group consists of 11 phosphopeptides that matched with 7 open reading frames (ORFs) within a family of FGR6-related (fungal growth regulator) genes in the RB2 repeat sequence of the genome (61). The last group contains a phosphopeptide found within 13 of the 14 Tlo protein orthologs of S. cerevisiae Med2 (ScMed2), a component of Mediator complex (discussed further below); only Tlo protein Tlo4 does not contain this peptide sequence. To simplify the downstream analyses, CGD ORF numbers were replaced with UniProt IDs when available, and the first ID that was assigned to a cluster as the protein identifier was used. When nonoverlapping peptides with the same UniProt ID were clustered, we identified 2,896 unique UniProt IDs with the caveat that in some cases it is not clear if the peptide identified belongs to one or more than one distinct proteins.

Analysis of proteins with tyrosine phosphorylations. While C. albicans has numerous serine/threonine kinases (14), only a single putative tyrosine kinase has been described (62). In addition, some serine/threonine kinases are known to also phosphorylate tyrosine residues, like members of the MEK/Ste7 and Wee1/Mkl1 families (63, 64). We detected 253 proteins with unique UniProt identifiers (8.7% of all phosphopeptides) with tyrosine P sites. The majority of these proteins contained only one detected phosphotyrosine, only 23 proteins were found to have two phosphorylated tyrosine residues, and 6 proteins had more than three detected tyrosine P sites. The majority of proteins with phosphorytyrosines also had detected phosphorylations on serines and threonines. Only 16 proteins were found to have only phosphorylated tyrosines (see Table S3 in the supplemental material). For example, the only two P sites on ribosomal protein L15 were Y6 and Y81 (see Table S3). Ddr48 was found to have the most phosphoryrosines, with seven, and Gin4 was second, with four. Gin4, a serine/threonine protein kinase that plays a role in C. albicans hypha formation and septin organization (15), was the most phosphorylated protein, with 74 unique P sites (57 serine, 13 threonine, and 4 tyrosine residues) (see Table S3). In S. cerevisiae, Gin4 interacts physically with Swe1, a kinase with the ability to phosphorylate S, T, and Y residues. Swe1 inhibits the activity of Cdc28 (Cdk1) by phosphorylation of tyrosine residue 19 in S. cerevisiae and Y18 in C. albicans, and this phosphorylation controls a morphogenesis checkpoint that delays mitosis and prevents a switch from polarized growth to isometric growth (15, 65, 66). We detected phospho-Y18 in C. albicans Cdc28, along with Cdc28 P sites at S5 and T166, suggesting that this phosphorylation pattern also exists in C. albicans hyphae (see Table S2).

GO term analysis of phosphoproteins. To determine if any classes of proteins were enriched among the detected phosphoproteins, the percentage of proteins assigned to Gene Ontology (GO) term categories for all annotated proteins was compared to the percentages for proteins with either phospho-S/T or phospho-Y residues. The distribution of C. albicans genome-wide GO terms includes 6,314 reported ORFs with at least one GO term annotation, and this number was used to calculate the percentage of proteins within each GO category (Fig. 3). There were 2,577 proteins detected as phosphoproteins (Fig. 3, P-all). We also analyzed the GO term distribution for the 2,368 proteins found to have serine and/or threonine phosphorylations (Fig. 3, P-S/T) and the 255 proteins with tyrosine phosphorylation sites (Fig. 3, P-Y). Only GO terms that had significant enrichment of proteins were used for further analyses, and the significance cutoff for shared GO terms was a P value of <0.05 with a Bonferroni correction. The enrichment of proteins within a subset of GO Slim assignments within the general categories Biological Process, Molecular Function, and Cellular Component, are presented in Fig. 3; the percentage of proteins in the GO term category set is provided in Table S4 in the supplemental material. For each GO term where significance was met in all three phosphoprotein data sets, P-all equals the sum of P-S/T and P-Y minus any overlap between these two data sets. A gray box is in place for cases where a GO term did not make the P value cutoff of 0.05 (Fig. 3).

Comparison of the rank-based distribution of the GO assignments of the total phosphorylation data set to the whole predicted proteome revealed a similar distribution of terms, showing no detectable bias in what classes of proteins were phosphorylated (Fig. 3). As the vast majority of P sites were on serine or threonine residues, it was not surprising that P-S/T GO term distribution was similar to that for the genome and the complete phosphoprotein data set. In contrast, the small set of proteins with detected tyrosine phosphorylations differed in their GO Slim term distribution (Fig. 3). For example, the P-Y proteins had a larger fraction of proteins with Biological Process classification of “response to extracellular stimuli,” “signal transduction,” and “regulation of cellular process,” while there was no significant enrichment in the GO term “metabolic process,” which contained the largest percentage of proteins encoded in the genome-, P-all, and P-S/T data sets (Fig. 3A). In the Molecular Function ontology, tyrosine-phosphorylated proteins were present in only 3 of the 8 selected GO terms. Interestingly, there was a larger fraction of P-Y proteins in three Molecular Function categories (DNA binding, kinase, and signal transducer activity) than in the other data sets; only P-Y phosphoproteins were enriched for the “signal transducer activity” GO term category (Fig. 3B). Analyses of proteins with the GO terms in the Cellular Component ontology revealed that tyrosine phosphoproteins were predominantly in the cytoplasm and the nucleus and were generally absent from membranes or membrane-bound organelles. In contrast, S/T phosphoproteins were not specifically enriched in or excluded from either membrane-associated categories or the cytoplasm (Fig. 3C). Analysis of all phosphoproteins (P-all) revealed that the most represented Cellular Component categories were “nucleus” and “cytoplasm,” and this result is comparable to what was observed in Aspergillus nidulans (25).

Phosphorylation sites per protein. The average number of phosphorylation sites per protein is 5.5; however, the distribution of P sites across proteins is skewed. For example, 786 of 2,896 proteins had a single detected P site and 509 proteins contained
FIG 3 GO enrichment analysis of the C. albicans phosphoproteome. Heat maps indicate the rank abundance of proteins within selected significant GO Slim categories (P value cutoff = 0.05) within the Biological Process (A), Molecular Function (B), and Cellular Component (C) gene ontologies. For each set of proteins (Genome, P-all, P-S/T, and P-Y), the percentages of proteins associated with each GO term were used to rank the distribution of proteins across the GO terms analyzed. See Table S4 in the supplemental material for the data used in these analyses.
These proteins have been involved in septin ring formation, including Bud4, Bni4, and Spa2 enrichment of proteins that are localized to the bud neck and are other kinases.

Hsl1 is phosphorylated at S33, S782, and S1299 by Cdc28 and May 2015 Volume 14 Number 5 ec.asm.org

the findings of Umeyama et al. (68), which showed that among the highly phosphorylated proteins, we observed an enrichment of proteins that are localized to the bud neck and are involved in septin ring formation, including Bud4, Bni4, and Spa2 in addition to Gin4 and Hsl1 (69, 70). These proteins have been associated with hyphal growth and establishment or maintenance of cytoskeleton polarity (71). Future studies will reveal how changes in phosphorylation state of these proteins participate in the induction and maintenance of different growth states.

Phosphorylation sites within the Ras1-Cyr1-PKA pathway.

The Ras/cyclic AMP (cAMP)/PKA signaling pathway (72), a central regulatory pathway involved in the induction and maintenance of hyphal growth, may be regulated by phosphorylation in a variety of different ways (Fig. 5). Ras1, which is active in its GTP-bound state and inactive in its GDP-bound conformation, is controlled by a Ras-GTPase-activating protein (GAP), Ira2, and the guanine nucleotide exchange factor (GEF) Cdc25. We found P sites on both Ira2 and Cdc25. Ras1 interacts activates adenylate cyclase Cyr1 (73), and the resulting increase in CAMP activates the catalytic subunits of protein kinase A (Tpk1 and -2) by causing their release from the regulatory subunit Bcy1. The activated PKA subunits likely have many targets, such as transcription factors, including Efg1 (72, 74).

In S. cerevisiae, Bcy1 itself is regulated by phosphorylation, in part by PKA subunits. Bcy1 that has been phosphorylated at S145 by PKA catalytic subunits causes Bcy1 to be a more efficient inhibitor of Tpk1, Tpk2, and Tpk3 (75). An alignment of ScBcy1 and C. albicans Bcy1 (CaBcy1) shows that the two proteins are highly similar; ScBcy1 S145, found within the canonical RXS/T PKA motif, corresponds to CaBcy1 S180 (76) (Fig. 5B). In Fig. 5B, blue arrows indicate the six phosphorylation sites that we identified for Bcy1 in our data set, including S180.

Besides Bcy1, other proteins known to be phosphorylated by PKA in S. cerevisiae were also phosphorylated within a putative PKA motif in our data set. For example, the forkhead transcription factor Fhl1 and its regulator Ifh1 (77) and the autophagy regulators Atg1 and Atg13 (78) were found. Efg1, reported as a PKA target at T206 (18, 79, 80), was found to be phosphorylated at six sites (5 serine and 1 threonine) but not at the T206 position, suggesting that PKA-mediated phosphorylation of Efg1 is dynamic (79). The amino acid sequence of Efg1 around the important threonine 206 is IRPR\textsubscript{TTT} (the underline indicates three-

![FIG 4](image-url) Phosphorylation sites per protein. Of the 2,896 unique proteins detected, 685 had only 1 P site, 415 had 2 P sites, and 271 had 3 P sites. More highly phosphorylated proteins were less commonly observed. The protein with the most P sites (74 P sites) is the serine/threonine-protein kinase Gin4 (Table 1).

### TABLE 1 Phosphorylated proteins with >45 P sites

<table>
<thead>
<tr>
<th>No. of phosphorylation sites</th>
<th>Gene name</th>
<th>Length of protein (aa)</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>74</td>
<td>GIN4</td>
<td>1,349</td>
<td>Serine/threonine protein kinase Gin4</td>
</tr>
<tr>
<td>73</td>
<td>CaO19.14160</td>
<td>1,352</td>
<td>Putative uncharacterized protein</td>
</tr>
<tr>
<td>72</td>
<td>BUD4, INTI</td>
<td>1,711</td>
<td>Bud site selection protein Bud4</td>
</tr>
<tr>
<td>66</td>
<td>CTA3</td>
<td>1,217</td>
<td>Potential EH domains and endocytosis protein</td>
</tr>
<tr>
<td>63</td>
<td>SPA2</td>
<td>1,466</td>
<td>Putative uncharacterized protein Spa2</td>
</tr>
<tr>
<td>62</td>
<td>ZDS1</td>
<td>1,645</td>
<td>Potential regulator of cell polarity</td>
</tr>
<tr>
<td>62</td>
<td>HSL1</td>
<td>1,462</td>
<td>Serine/threonine protein kinase Hsl1</td>
</tr>
<tr>
<td>58</td>
<td>FAR1</td>
<td>2,624</td>
<td>Uncharacterized protein</td>
</tr>
<tr>
<td>57</td>
<td>BNI4</td>
<td>1,655</td>
<td>Possible bud neck-localizing protein phosphatase subunit Bni4</td>
</tr>
<tr>
<td>53</td>
<td>CaO19.9831</td>
<td>1,309</td>
<td>Uncharacterized protein</td>
</tr>
<tr>
<td>50</td>
<td>CaO19.11901</td>
<td>1,073</td>
<td>Potential GRAM domain protein</td>
</tr>
<tr>
<td>47</td>
<td>CaO19.11117</td>
<td>1,563</td>
<td>Uncharacterized protein</td>
</tr>
<tr>
<td>47</td>
<td>BOI2</td>
<td>1,176</td>
<td>Putative uncharacterized protein Boi2</td>
</tr>
<tr>
<td>47</td>
<td>RGA2</td>
<td>1,176</td>
<td>Putative uncharacterized protein Rga2</td>
</tr>
<tr>
<td>46</td>
<td>CaO19.8830</td>
<td>1,043</td>
<td>Uncharacterized protein</td>
</tr>
<tr>
<td>45</td>
<td>CaO19.7006</td>
<td>1,094</td>
<td>Putative uncharacterized protein</td>
</tr>
<tr>
<td>45</td>
<td>CaO19.7738</td>
<td>896</td>
<td>Uncharacterized protein</td>
</tr>
<tr>
<td>45</td>
<td>KIN1</td>
<td>1,212</td>
<td>Likely protein kinase</td>
</tr>
<tr>
<td>45</td>
<td>SEC16</td>
<td>1,947</td>
<td>Putative uncharacterized protein Sec16</td>
</tr>
</tbody>
</table>
Phosphorylation pattern of proteins in the Ras1-Cyr1-PKA pathway in *C. albicans*. (A) Schematic model and the phosphorylation states of the Ras1-Cyr1-PKA pathway in *C. albicans*. Ras1 activity is governed by Ira2 and Cdc25, which both contain P sites. Ras1 induces adenylate cyclase (Cyr1) to produce cAMP, which inhibits the formation of the protein kinase A complex consisting of the catalytic subunits Tpk1 and -2, and the regulatory subunit Bcy1, by directly binding to Bcy1. Phosphorylation states are depicted by color (open ovals, no phosphorylation; green ovals, 3 P sites; orange ovals, between 4 and 10 P sites; and red jagged ovals, >10 P sites). (B) An alignment of *Sc Bcy1* and *Ca Bcy1* shows high overall similarity. *S. cerevisiae* Ser145 is within an RRxS/T motif which is recognized by PKA (red box); this serine corresponds to Ser180 in *C. albicans*. The blue arrows indicate the Bcy1 P sites (Q9HEW1; see Table S3 in the supplemental material).
nine 206), which varies only by one amino acid from the PKA consensus sequence.

Cdk8-dependent and independent phosphorylation of Mediator. The multiprotein complex Mediator is a coactivator of RNA polymerase II transcription. In other species, the Mediator complex is known to be phosphorylated by multiple kinases, and the phosphorylation states of proteins within the complex can impact gene expression (81). The Mediator complex in C. albicans is composed of at least 23 subunits plus a Med2 ortholog, which in C. albicans is encoded by 14 different paralogs (referred to as Tlo subunits).

We observed P sites in 17 out of the 23 subunits (Supplemental Table S1), which are held in the tail module through interaction with Med3 (44). We observed P sites in 17 out of the 23 subunits (Supplemental Table S1), which are held in the tail module through interaction with Med3 (44). We observed P sites in 17 out of the 23 subunits (Supplemental Table S1), which are held in the tail module through interaction with Med3 (44). We observed P sites in 17 out of the 23 subunits (Supplemental Table S1), which are held in the tail module through interaction with Med3 (44).

Cdk8/Ssn3 is a kinase that is part of the CDK8 module of Mediator, and it has been shown to modulate Mediator and transcription factor activity in ways that impact hyphal growth, switching, metabolism, and stress resistance (42, 43, 47, 82). To directly determine if Cdk8 catalyzes the phosphorylation of Mediator subunits, we performed in vitro kinase assays using purified Mediator or mixtures of Mediator subcomplexes. First, we found that when the native complex, including the Cdk8 module, was incubated with [32P]ATP, we detected phosphorylations on proteins with mobilities that most likely corresponded to those of Cdk8, Med8, Med9, and Med15. In addition, we detected phosphorylation of protein(s) in a band whose molecular mass corresponded to either Med12 or Med13, as well as another band (Mp48) whose molecular mass corresponded to either Med4, Med7, or Med18 (Fig. 6A). Subsequent phosphoproteomics, discussed below, indicate that the bands that could not be unambiguously assigned were most likely Med13 and Med4, respectively.

**TABLE 2 Mediator complex P sites**

<table>
<thead>
<tr>
<th>UniProt ID</th>
<th>Gene ID</th>
<th>Large-scale phosphoproteomics</th>
<th>In vitro Cdk8 kinase assay (ATP dependent)</th>
<th>In vitro Cdk8 kinase assay (ATP independent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q5AHH0</td>
<td>Med1</td>
<td>S445, S446</td>
<td>T140</td>
<td>S445, S446</td>
</tr>
<tr>
<td>Q59Q94</td>
<td>Med3</td>
<td>S117, T125, T133</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q59U73</td>
<td>Med4</td>
<td>S9, S13, S17, S33, S322, T36, T37, Y321</td>
<td>S9, S33, T36, T37</td>
<td>S9, S13, S252</td>
</tr>
<tr>
<td>Q59KC0</td>
<td>Med5</td>
<td>S252</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5AZ21</td>
<td>Med6</td>
<td>S285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5AEN6</td>
<td>Med7</td>
<td>S102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q59TD3</td>
<td>Med8</td>
<td>T37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5A0D0</td>
<td>Med9</td>
<td>T109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5A3K2</td>
<td>Med10</td>
<td>S170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q59S43</td>
<td>Med11</td>
<td>S132, S142, T140, T144, T149</td>
<td>S120, S132, S142, T140</td>
<td>S63</td>
</tr>
<tr>
<td>Q5AG31</td>
<td>Med14</td>
<td>S385, T395</td>
<td>S14, S385, S678, S763, S1170</td>
<td></td>
</tr>
<tr>
<td>Q5PP6</td>
<td>Med16</td>
<td>S380</td>
<td>S898, S900, S182, S184, S1084, S1097, T903, T913, T1030, T1042</td>
<td></td>
</tr>
<tr>
<td>Q5AHZ7</td>
<td>Med17</td>
<td>S43, S73, T70</td>
<td>S898, S900, S182, S184, S1084, S1097, T903, T913, T1030, T1042</td>
<td></td>
</tr>
<tr>
<td>Q59X40</td>
<td>Med18</td>
<td>S207, T162</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4YFG8</td>
<td>Med19</td>
<td>S79, S83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5ACU4</td>
<td>Med21</td>
<td>S79, S83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5AH9</td>
<td>Med22</td>
<td>S79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5PR87</td>
<td>Med31</td>
<td>S79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5AH2</td>
<td>Sn3 (Cdk8)</td>
<td>S209</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5AH9</td>
<td>Tlo</td>
<td>S39</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a P sites detected in the large-scale phosphoproteomics data set and the ATP-dependent in vitro Cdk8 kinase assay are in bold, and P sites detected in the large-scale phosphoproteomics data set and the ATP-independent in vitro Cdk8 kinase assay are in italics.
with Mediator from a strain bearing either Cdk8-HF or Med8-HF (6×His-3×FLAG), we detected the expected mass shifts in the corresponding bands (Fig. 6B, lanes 3 and 4). All proteins other than Med8 that showed in vitro phosphorylation were also detected as phosphorylated proteins in our large phosphoproteomics data set (Fig. 7). The absence of Med8 from the large-scale data set is not surprising, since the tryptic peptide generated from the N terminus of Med8, which contains the most likely Cdk8 P sites, is above the molecular mass typically detected by this methodology.

To determine if Cdk8 can phosphorylate a Tlo protein family member and Med3, we performed an [32P]ATP kinase assay in which purified Mediator was incubated with a complex of recombinant Tlo12 bound to 6×His-tagged Med3 (44). The phosphorylation pattern showed that the recombinant proteins were detected in a dose-dependent manner at the expected sizes (Fig. 6C). Phosphorylated bands for the native Tlo proteins associated with Mediator were not observed, likely because any one of the 14 different Tlo paralogs, with varying molecular masses, were not sufficiently abundant.

To further define the P sites that were attributable to Cdk8 activity and compare these sites to those detected in the large-scale phosphoproteomics effort, we performed an in vitro analysis using purified Mediator. To quantitatively determine which Mediator P sites could be a consequence of Cdk8 activity, we analyzed purified Mediator, in which Cdk8 is the only protein with kinase activity, after incubation with and without ATP. Differential methyl labeling of the samples incubated with and without ATP prior to sample mixing allowed for direct quantitative analysis of newly phosphorylated sites versus those present on purified Mediator prior to ATP addition. The results confirmed and refined our previous assignments of phosphorylated subunits (Fig. 6) and identified the specific sites phosphorylated by Cdk8 in purified Mediator (Table 2). Consistent with the data shown in Fig. 6, Cdk8-dependent P sites were found in subunits Med3, Med8, Med9, Cdk8, and Med15 (Table 2). Of Med4, Med7, and Med18, which could not be distinguished in the [32P]ATP in vitro assays (Fig. 6A), Med4 was the only one that was found to be phosphorylated; 4 of the 5 Med4 P sites were heavily enriched among the ATP-dependent P sites. San2/Med13, but not Med12, had also many sites enriched in the ATP-dependent P sites. This analysis also revealed ATP-independent P sites already present in the purified Mediator, which could have resulted from the in vivo activity of Cdk8 or other kinases (Table 2).

Comparison of the ATP-dependent P sites identified in the in vitro Cdk8 kinase assay to those identified in a phosphoproteomics analysis of S. cerevisiae Mediator (81) revealed some general similarities but no single P site that was clearly highly conserved. For instance, both our data and the S. cerevisiae work showed high phosphorylation of the C terminus of Med13, which is thought to be associated with suppression of stress-induced transcription

**FIG 6** Western blot analysis of phosphorylation of Mediator subunits by Cdk8. (A) Autoradiographic images of in vitro kinase assay data are shown for reaction mixtures containing [32P]ATP and the Mediator complex purified from C. albicans via the Med8-HF tag ("HF" refers to a 6×His-3×Flag tag) (lane 1), Mediator purified from a Δmed12 strain (which lacks the Cdk8 module) without and with purified Cdk8 module (lanes 2 and 3, respectively), and Mediator purified from a cdk8Δ/Δ strain (Δcdk8) without and with purified Cdk8 module (lanes 4 and 5, respectively). Cdk8 module alone shows autophosphorylation within the module (lane 6). (B) Similar experiments were performed with med15Δ/Δ (Δmed15) and med9Δ/Δ (Δmed9) to confirm the identity of those bands by showing their absence in the appropriate strain. In vitro kinase assays with strains expressing either Cdk8-HF or Med8-HF also demonstrated mass shifts that confirmed protein identities. (C) Purified Mediator with Cdk8 as the sole kinase was used in an experiment to demonstrate that Cdk8 phosphorylates the Tlo protein family and Med3. Spiked-in recombinant Tlo12 and 6×His-tagged Med3 from C. albicans could be detected in a dose-dependent manner at the expected sizes.
Our work advances the *S. cerevisiae* study, since those authors were unable to identify P sites in the Cdk8 module due to the absence of the module in their purified Mediator sample. We found an autophosphorylation site in Ssn3/Cdk8 as well as several ATP (Cdk8)-dependent and independent P sites in Ssn2/Med13 that potentially play regulatory roles. The results also showed that the large-scale phosphoproteomics analysis of whole-cell extracts was reasonably adept at picking up both Cdk8-dependent and independent P sites in Mediator. In the large-scale phosphoproteomics data set, we identified 68 P sites on 18 of the 24 Mediator subunits plus the Tlo protein family. The *in vitro* Cdk8 kinase analyses found 20 ATP-dependent P sites, with 12 that were also present in the large-scale experiment (Fig. 7; Table 2). It is interesting that a majority of the Mediator P sites detected in the large-scale proteomics were not detected in the purified sample; however, the discrepancy could be explained by the different growth conditions that were used to perform the two experiments, and it is well known that phosphorylation levels of the Mediator change in response to external stimuli (81). This finding also suggests that the large-scale methodology may be highly valuable in identifying highly labile P sites that do not survive purification of the complex.

Conclusions. To our knowledge, this report is the first comprehensive phosphoproteomics study of hyphae in the human fungal pathogen *C. albicans*. In this data set, we found phosphorylation targets that have been previously identified experimentally both in either *C. albicans* and in *S. cerevisiae*, such as those in the highly conserved Ras1-Cyr1-PKA pathway and in the Mediator complex. It will be interesting to see if the level of conservation of phosphorylation sites across other species in different pathways as more studies characterizing the phosphoproteome are published. These phosphoproteomics data can be used as a resource for future targeted and large-scale phosphoprotein analyses in this fungus or as a reference for other fungi.

**ACKNOWLEDGMENTS**

Research reported in this publication was supported by National Institutes of Health (NIH) grant R01GM108492 (D.A.H. and J.E.S.), the Hitchcock Foundation (L.C.M.), American Cancer Society research grant IRG-82-003-30 (A.N.K.), and the Cystic Fibrosis Research Development Program (STANTC00780) (S.D.W.). The Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) was acquired with support from the NIH (S10-OD016212).

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

We thank Allia Lindsay for her analysis of Cdk8 phosphorylation sites in preliminary studies.

**REFERENCES**

7. Soll DR. 2014. The role of phenotypic switching in the basic biology and.

FIG 7 Summary of phosphorylation analyses of Mediator complex. The overall structure of the Mediator complex is shown. Subunits phosphorylated *in vitro* by the Cdk8 kinase are indicated by a bold outline. Subunits with P sites identified in the large-scale phosphoproteomics analysis are shown in color. Subunits labeled in bold letters have been identified in a phosphoproteomics analysis of *in vitro* Cdk8 activity using purified Mediator. P sites in purified Mediator that are enriched after incubation with ATP are indicated in parentheses.

(81).


Lindsay AK, Morales DK, Liu Z, Grahl N, Zhang A, Willger SD, Myers LC,
The Valot
Uhl
Ciesára
Inglis
Marx
Fattaey


AR, Hermjakob H. nated proteomics data submission and dissemination. Nat Biotechnol 31:


