Eukaryotic UDP-Galactopyranose Mutase (GLF Gene) in Microbial and Metazoal Pathogens†

Stephen M. Beverley,1,* Katherine L. Owens,1 Melissa Showalter,1 Cara L. Griffith,1 Tamara L. Doering,1 Victoria C. Jones,2 and Michael R. McNeil2

Department of Molecular Microbiology, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, Missouri 63110,1 and Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado 805232

Received 19 February 2005/Accepted 29 March 2005

Galactofuranose (Gal3) is a novel sugar absent in mammals but present in a variety of pathogenic microbes, often within glycocomplexes that play critical roles in cell surface formation and the infectious cycle. In prokaryotes, Gal3 is synthesized as the nucleotide sugar UDP-Gal3 by UDP-galactopyranose mutase (UGM) (gene GLF). Here we used a combinatorial bioinformatics screen to identify a family of candidate eukaryotic GLFs that had previously escaped detection. GLFs from three pathogens, two protozoan (Leishmania major and Trypanosoma cruzi) and one fungus (Cryptococcus neoformans), had UGM activity when expressed in Escherichia coli and assayed in vivo and/or in vitro. Eukaryotic GLFs are closely related to each other but distantly related to prokaryotic GLFs, showing limited conservation of core residues around the substrate-binding site and flavin adenine dinucleotide binding domain. Several eukaryotes not previously investigated for Gal3 synthesis also showed strong GLF homologs with conservation of key residues. These included other fungi, the alga Chlamydomonas and the algal phleovirus Feldmannia irregularis, parasitic nematodes (Brugia, Onchocerca, and Strongyloides), and Caenorhabditis elegans, and the urochordates Halocynthia and Ciona. The C. elegans open reading frame was shown to encode UGM activity. The GLF phylogenetic distribution suggests that Gal3 synthesis may occur more broadly in eukaryotes than previously supposed. Overall, GLF/GLF3 synthesis in eukaryotes appears to occur with a disjunct distribution and often in pathogenic species, similar to what is seen in prokaryotes. Thus, UGM inhibition may provide an attractive drug target in those eukaryotes where Gal3 plays critical roles in cellular viability and virulence.

In many microbes, the sugar galactofuranose (Gal3) is an important constituent of glycoconjugates comprising major portions of the cell surface (17). In prokaryotes, Gal3 constitutes a key part of the mycobacterial cell wall and occurs in lipopolysaccharide (LPS) O-antigen domains, extracellular capsules, and polysaccharides (17, 18). In fungi such as Aspergillus, Gal3 is a major component within the cell wall and structural glycoproteins (13, 17). In pathogenic protozoa, Gal3 residues are key components of abundant surface glycosyl-phosphoinositol-anched glycoconjugates, such as lipophosphoglycan (LPG) and glycoinositolphospholipids in Leishmania, and of mucins, glycosylphosphoinositol-anched proteins, and lipids in Trypanosoma cruzi (6, 18). In contrast, Gal3 residues have not been found in humans or other metazoaons (17), suggesting that inhibition of Gal3 synthesis could be an attractive target for chemotherapy in pathogens when its role(s) is critical for survival or virulence (17, 32).

The Gal3 synthetic pathway has been best studied for prokaryotes. Genetic and biochemical studies have shown that Gal3 arises through the action of UDP-galactopyranose mutase (UGM) (EC 5.4.99.9), which catalyzes the rearrangement of UDP-galactopyranose (Gal3p) to UDP-Gal3, the substrate of cellular UDP-Gal3 transferases which participate in pathways such as LPS or cell wall biosynthesis (15, 16). The gene encoding UGM, GLF, was first located and hypothesized as such by Reeves and colleagues as part of genetic and structural studies of Escherichia coli K12 O antigen (27). It was then definitively identified and studied in E. coli, Klebsiella, and Mycobacteria (10, 15, 16). UGM is a flavin-dependent enzyme, and the E. coli enzyme structure has been solved (21). A detailed picture of the enzymatic mechanism involving a novel form of flavin-dependent catalysis has been developed (24). High-throughput assays for inhibitor screens have been developed, and a number of compounds showing activity against UGM activity and/or bacteria have been identified (22, 25, 28).

Less is known about the Gal3 synthetic pathway in eukaryotes. Previous efforts had not yielded the eukaryotic enzyme responsible for synthesis of their UDP-Gal3 substrate, although as in prokaryotes this was thought to arise by conversion of UDP-Gal3 to UDP-Gal3 (29). In the parasitic protozoan Leishmania, several genes encoding putative UDP-Gal3 transferases have been identified, including LPG1, which has been implicated in the synthesis of the core of the abundant surface glycoconjugate LPG (7, 20). Notably, the Leishmania genome encodes at least 6 candidate UDP-Gal3 transferases (34), and there are more than 20 related genes present in Trypanosoma cruzi (34) (unpublished data). No candidate UDP-Gal3 transferases have been reported in fungi, although they must exist given the number of Gal3-containing glycoconjugates found in these organisms. In this report we have used a bioinformatics approach to identify the eukaryotic GLF gene family. We ex-

* Corresponding author. Mailing address: Department of Molecular Microbiology, Washington University Medical School, 660 S. Euclid Ave., St. Louis, MO 63110. Phone: (314) 747-2630. Fax: (314) 747-2634. E-mail: beverley@borcim.wustl.edu.
† Supplemental material for this article may be found at http://ec.asm.org.
pressed four diverse members of this family in *E. coli* and demonstrated their activity in vivo and/or in vitro.

**MATERIALS AND METHODS**

PCR and construction of pET3 derivatives. For *Leishmania major*, PCR oligonucleotides SMB2179 (5'-GATGCTCATATGGACATCGCGG-3') and SMB2180 (5'-GATGCTGTCCTCTACGCGCCTG-3') (underlined bases represent sites added for subsequent cloning) were based upon the open reading frame (ORF) Lm18.0200 annotated from the *Leishmania major* genome (www.genedb.org/genedbl/leish). The PCR template was genomic DNA of *L. major* Friedlin (HMOL/L8/1 Friedlin) clone V1, and amplification was performed with 35 cycles of 30 s at 95°C, 45 s at 50°C, and 150 s at 68°C. For *Trypanosoma cruzi*, PCR oligonucleotides SMB2182 (5'-GCACGCGCTAATGGAATTAGAT-3') and SMB2230 (5'-GCGATAGGATCCTACGAGGCCGTCGAC-3') were based upon the second of two ORFs (Tc00.1040735511277.600 and Tc00.1040735570993.160) annotated in release 3.0 of the *Trypanosoma cruzi* genome project (www.genedb.org/genedbl/terzi). The PCR template was DNA from the CL Brenner strain of *T. cruzi* used for genome sequencing (kindly provided by D. Barthomomeu, The Institute for Genomic Research, Rockville, MD), and amplification was performed with 35 cycles of 30 s at 95°C, 45 s at 50°C, and 150 s at 68°C. For *Coffea canephora*, PCR primers Gen-S (5'-GGAATTCCATAATCGGGTGGTGGT-3') and Gen-A (5'-GGAGATGAGGATCCTACGAGGCCGTCGAC-3') were based upon ORF77.620826 (The Institute for Genomic Research, www.tigr.org/tdb/e2k1/coffee/index.shtml). Total RNA was isolated from *C. canephora* serotype D strain JECS43 and converted to cDNA using the SuperScript First-Strand synthesis system as recommended by the manufacturer (Invitrogen, Carlsbad, CA). PCR amplification using HIFI Taq polymerase (Invitrogen, Carlsbad, CA) was performed with 30 cycles of 60 s at 94°C, 60 s at 55°C, and 120 s at 72°C. For *C. elegans*, gene predictions for the GLF ORF H04m03.4 were obtained from WormBase (http://www.wormbase.org). One internal region had not been determined experimentally, and we found by sequence analysis of several cDNAs (expressed sequence tags [ESTs] syk1442e06, syk1440e04, and syk1626a12 from Yuji Kohara, National Institute of Genetics, Mishima, Japan; GenBank accession no. D856302-6; and ORF clone 100136/E10, obtained from Open Biosystems, Inc., Huntsville, AL) that the current gene model had missed a short intron, resulting in a predicted 9-amino-acid insertion (SVYCVLFEV) not encoded by any of the three cDNAs sequenced. The *C. elegans* H04m03.4 ORF contained an internal NdeI site, requiring a series of PCR and cloning steps starting with the 100136/E10 cDNA template prior to insertion into the pET3a NdeI-BamHI sites.

All PCR-amplified products obtained with the expected sizes, which as necessary were digested with NdeI and BamHI ligated to pET-3A vector DNA previously digested with the same enzymes, and transformed into *E. coli* DH10B or DH5α, yielding pET3a-LmjGLF, pET3a-CnGLF, pET3a-TcGLF, or pET3a-TeGLF (tab labs strains B234, B300, B3530, and B4255, respectively). The authenticity of candidate recombinants was confirmed by DNA sequencing (GenBank accession numbers AY900625, AY900626, and BK005688, respectively). For *T. cruzi* GLF, our sequence corresponded well with the pET3a-LmjGLF construct.

**RESULTS**

Identification of candidate GLFs. Although the first prokaryotic GLF was identified in 1996, efforts to detect eukaryotic GLFs has not been successful. Recently the genomes of several eukaryotic pathogens known to synthesize Gal progressed towards completion, and we applied a combinational bioinformatics screen to these genomes in an effort to identify candidate GLFs. Briefly, these involved a combination of (i) BLAST searches with prokaryotic GLFs, (ii) searches for flavin binding domains, and (iii) appropriate phylogenetic distribution, specifically occurrence in fungi and protozoa known to synthesize Gal, but absence from taxa lacking Gal. Since the presence of Gal has not been systematically detected for many microbial and metazoan species, we applied the latter criterion conservatively. We focused these studies initially on the *Leishmania major* genome, using ORF predictions developed by the *Leishmania* genome project (release 4.0, August 2004; see www.genedb.org/genedbl/leish).

While many ORFs satisfied each of these criteria individually (often quite weakly), a single candidate *L. major* ORF was identified in the combined screen. The Lm18.0200 GLF sequence predicted a 491-amino-acid protein, which showed a weak relationship to the Pfam family of flavin-containing amine oxidoreductases (P010804; e = 6 × 10⁻⁴), including a potential FAD binding motif at amino acids 6 to 19 (Fig. 1, underlined by parallel lines). The three-dimensional structure of *E. coli* UGM has shown that the FAD binding domain actually encompasses a much larger portion of the protein (21), and limited sequence conservation homology was provided by Whitfield and Clarke, U. Guehl) at a titler of 1:10,000 for 2 h at room temperature in 5% milk in TBST (10 mM Tris, 150 mM NaCl, 0.05% Tween 20 [vol/vol]). Binding was visualized with peroxidase-conjugated goat anti-rabbit immunoglobulin G (Jackson ImmunoResearch, catalog no. 111-035-003) at a titer of 1:20,000 in TBST for 1 h at room temperature followed by detection by chemiluminescence (Western Lightning, Perkin-Elmer Life Sciences, catalog no. NEL100).

UGM enzyme assays. One-liter cultures of *E. coli* BL21(DE3)/pET3a-CnGLF expressing Cryptococcus GLF (BS303), *E. coli* BL21(DE3)/pET3a-LmjGLF expressing *Leishmania GLF* (BS235), or *E. coli* BL21(DE3) without plasmid were grown overnight at 37°C in LB broth. Samples containing pET3a-GLF constructs were grown in the presence of 100 μg/ml ampicillin. All cultures were induced by addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside, followed by incubation at 22°C with shaking at 90 rpm for 3 h. Cultures were harvested by centrifugation, and cell pellets were resuspended in 2 ml 100 mM potassium phosphate buffer (pH 7.0) containing 0.15 M NaCl, 15% glycerol, 10 mM flavin adenine dinucleotide (FAD), 0.1 mM phenylmethylsulfonyl fluoride, 1 μM pepstatin A, and 1 μM leupeptin. The cells were broken by three passages through a French press and centrifuged at 20,000 g for 40 min at 4°C. The supernatant fractions were frozen at −80°C until use.

UDP-Gal was synthesized enzymatically from UDP-Gal, and purified as described previously (14). Activity assays were performed in 96-well PCR plates. Each well contained 50 μM UDP-Gal, 100 mM phosphate buffer at pH 7, 1 mM MgCl₂, 15% glycerol, protein (80 to 10,000 ng) *E. coli* extract expressing Cryptococcus GLF, 35 to 4,400 ng *E. coli* extract expressing *Leishmania GLF*, or 27,400 ng control *E. coli* BL21(DE3), and 20 mM sodium dithionite (to reduce the FAD to FADH₂) in a 20-μl volume. FAD was added in various amounts depending on how much protein was added (final concentrations of 1.6 to 1,000 μM). Reactions were incubated at 30°C for 60 min and stopped by the addition of 40 μl 95% ethanol. Samples were analyzed by high-performance liquid chromatography (HPLC) as described previously (14), except that the solvent system was modified by elution of the sugar nucleotides for 10 min with 200 mM KH₂PO₄ wash. Yields of UDP-Gal were calculated from the relative fraction of UDP-Gal, and UDP-Gal, in the samples and the knowledge that 1 nmol of UDP-Gal was originally added to each well.

**ACKNOWLEDGMENTS**

This work was supported by Public Health Service grants AI25395, AI30804, and AI44153 awarded by the National Institute of Allergy and Infectious Diseases and the National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health.
also evident throughout this region (Fig. 1, underlined by dark lines). BLAST searches of a variety of databases, including unfinished microbial genomes, revealed a number of strong BLAST hits (P/H1102110/H1100240), most of which were annotated as hypothetical proteins although several were annotated as amine oxidases (Table 1; Fig. 2). Notably, in every case where sequence from organisms capable of Gal\textsubscript{f} synthesis was analyzed, strong BLAST hits to Lm18.0200 were present. These included the protozoans \textit{Leishmania infantum} and \textit{T. cruzi} and the fungi \textit{Aspergillus}, \textit{Cryptococcus}, and \textit{Neurospora} (Fig. 2; Table 1). Correspondingly, homologs of the candidate \textit{L. major} GLF were not detected in vertebrates, nor in \textit{Trypanosoma brucei}, \textit{Saccharomyces cerevisiae} or \textit{Schizosaccharomyces pombe}, species where Gal\textsubscript{f} may be absent. Prokaryotic GLF/UGMs yielded weak BLAST scores (P/H110220.2) across only the N-terminal portion of the \textit{Leishmania} UGM, which undoubtedly contributed to the eukaryotic genes being overlooked previously.

Comparison of the predicted proteins encoded by the \textit{L. major}, \textit{T. cruzi}, and \textit{Cryptococcus neoformans} GLFs to that of \textit{E. coli} UGM/GLF (Fig. 1; Table 1) shows that the overall similarity to prokaryote GLFs is relatively low. Most regions of sequence conservation fall within the FAD binding domain determined previously from the structure of \textit{E. coli} UGM (Fig. 1, underlined regions) (21). Significantly, genetic and structural analyses have identified a number of residues possibly involved in UDP-Gal\textsubscript{p} binding, and with few exceptions these residues were identical in prokaryotic and eukaryotic GLF proteins (Fig. 1, residues shaded in orange). These residues were also conserved within the strong database hits noted elsewhere in this section (data not shown). The candidate eukaryotic GLF proteins showed a high degree of identity to each other (Fig. 1 and 2; Table 1).

While the correlation of GLF candidates with species known to carry out Gal\textsubscript{f} synthesis was encouraging, strong BLAST hits were also obtained to sequences from a number of organisms not known to synthesize Gal\textsubscript{f}, including the nematodes \textit{C. elegans} and \textit{Caenorhabditis briggsae}, the urochordate \textit{Halocynthia}, and the algal \textit{Feldmannia irregularis} virus (Table 1). A number of ESTs encoding partial GLF ORFs were found in parasitic nematodes, including \textit{Brugia malayi}, \textit{Onchocerca volvolus}, \textit{Strongyloides stercoralis} and \textit{Strongyloides ratti}, \textit{Heterodura glycine}, and \textit{Meloidogyne hapla} and \textit{Meloidogyne arenaria} (see Table S1 in FIG. 1. Alignment of key eukaryotic and prokaryotic GLF proteins. The predicted proteins encoded by the \textit{L. major}, \textit{T. cruzi}, \textit{C. neoformans}, and \textit{C. elegans} GLF ORFs were aligned with that of \textit{E. coli} using methods implemented in the Clustal program (4); the unedited alignment is shown. The amino acid positions are shown on the right, and asterisks mark 10-amino-acid intervals in the alignment. Residues identical in 5/5 sequences are shaded in blue, and residues identical in 4/5 sequences are shaded gray. Residues proposed as part of UDP-Gal\textsubscript{p} binding site are shaded orange (all but one of these are 5/5 identities). The dark underlined regions correspond to the flavin-binding region of \textit{E. coli} UGM (21); a smaller region showing homology to the Pfam family flavin-containing amine oxidoreductases (PBO10804; residues 6 to 19 of the \textit{L. major} UGM) is underlined by two parallel lines.
the supplemental material), and we were able to assemble complete ORFs for the urochordate Ciona intestinalis and the alga Chlamydomonas reinhardtii (Table 1; see Tables S1 and S2 in the supplemental material). Several strong hits (P < 10^{-8}) were obtained to eubacterial sequences annotated as "amine oxidases," although enzymatic data supporting these assignments were lacking. Strong hits (P < 10^{-25} to 10^{-35}) were also obtained with sequences from three archa (Pyrococcus horikoshii, Pyrobaculum aerophilum, and Methanococcus maripaludis) (Table 1). A phylogenetic tree depicting relationships among the eukaryotic candidates is shown in Fig. 2, and an alignment of the C. elegans ORF H04M30.4 is included in Fig. 1. As observed for the fungal and protozoal homologs, candidate flavin-binding regions and UDP-Gal oxidation binding residues were conserved in these predicted proteins (Fig. 2; also data not shown). Genes or proteins showing significant relationship were not observed in BLAST searches of the genomes of Giardia lamblia, Trichomonas vaginalis, Entamoeba, apicomplexans including Toxoplasma and Plasmodium, Tetrahymena, or several plant genomes including Arabidopsis.

The imperfect concordance between the occurrence of eukaryotic GLFs and GalO synthesis could arise from incomplete knowledge of GalO synthesis and/or an incorrect assignment of this gene family (or individual members therein) as encoding active UGMs. Thus, we sought confirmation that representative candidate GLF ORFs encoded UGM activity.

**Rescue of GalO-dependent LPS synthesis in E. coli.** We developed an in vivo complementation assay for UDP-GalO synthesis in E. coli, adapted from work by Whitfield and colleagues (10). These authors showed that expression of the Klebsiella rfbKPO1 locus in an E. coli strain deleted for the endogenous LPS rfb locus led to the synthesis of LPS O1 antigen bearing the repeating unit \([-\beta-D-Gal-(1\rightarrow3)-\beta-D-Gal-(1\rightarrow3)]\). Notably, inactivation of the Klebsiella GLF gene rfbP\_KPO1 abrogated LPS biosynthesis, which could be restored by episomal expression of KpGLF. We surmised that this would provide a rapid and convenient test for the activity of the candidate eukaryotic GLF family ORFs encoded UGM activity.

<table>
<thead>
<tr>
<th>Species</th>
<th>Group</th>
<th>Contains Gal(^b)</th>
<th>Gene</th>
<th>Annotation(^c)</th>
<th>BlastP P value</th>
<th>% Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leishmania major</td>
<td>Protists</td>
<td>YES</td>
<td>Lmj18.200</td>
<td>UGM (query)</td>
<td>10^{-117}</td>
<td>100</td>
</tr>
<tr>
<td>Trypanosoma cruzi</td>
<td>Protists</td>
<td>YES</td>
<td>Tcp0.1047053511277.600</td>
<td>UGM</td>
<td>10^{-100}</td>
<td>58</td>
</tr>
<tr>
<td>Aspergillus nidulans</td>
<td>Fungi</td>
<td>YES</td>
<td>EAA63683</td>
<td>Hyp.</td>
<td>10^{-117}</td>
<td>45</td>
</tr>
<tr>
<td>Neurospora crassa</td>
<td>Fungi</td>
<td>YES</td>
<td>EAA23752</td>
<td>Hyp.</td>
<td>10^{-117}</td>
<td>44</td>
</tr>
<tr>
<td>Magneporthe grisea</td>
<td>Fungi</td>
<td>UNK</td>
<td>EAA55508</td>
<td>Hyp.</td>
<td>10^{-117}</td>
<td>44</td>
</tr>
<tr>
<td>Giberella zeae</td>
<td>Fungi</td>
<td>UNK</td>
<td>EAA25642</td>
<td>Hyp.</td>
<td>10^{-117}</td>
<td>44</td>
</tr>
<tr>
<td>Cryptococcus neoforans</td>
<td>Fungi</td>
<td>YES</td>
<td>EAL19520</td>
<td>UGM</td>
<td>10^{-117}</td>
<td>42</td>
</tr>
<tr>
<td>Chlamydomonas reinhardtii</td>
<td>Green algae</td>
<td>UNK</td>
<td>Table S2(^a)</td>
<td></td>
<td>10^{-117}</td>
<td>42</td>
</tr>
<tr>
<td>Ustilago madidis</td>
<td>Fungi</td>
<td>UNK</td>
<td>UM03994</td>
<td>Hyp.</td>
<td>10^{-104}</td>
<td>40</td>
</tr>
<tr>
<td>Geobacter sulfurreducens</td>
<td>Eubacteria</td>
<td>UNK</td>
<td>AAR34886</td>
<td>Flavin amine oxidase</td>
<td>10^{-103}</td>
<td>44</td>
</tr>
<tr>
<td>Desulfovibrio vulgaris</td>
<td>Eubacteria</td>
<td>UNK</td>
<td>AAS94778</td>
<td>Flavin amine oxidase</td>
<td>10^{-91}</td>
<td>40</td>
</tr>
<tr>
<td>Desulfovibrio desulfuricans</td>
<td>Eubacteria</td>
<td>UNK</td>
<td>ZP00346806</td>
<td>Protoporphyrinogen oxidase</td>
<td>10^{-35}</td>
<td>37</td>
</tr>
<tr>
<td>Feldmannia irregularis virus</td>
<td>Algal</td>
<td>UNK</td>
<td>AAR26880</td>
<td>Hyp.</td>
<td>10^{-84}</td>
<td>38</td>
</tr>
<tr>
<td>Caenorhabditis briggsae</td>
<td>Nematodes</td>
<td>UNK</td>
<td>CAE72630</td>
<td>Hyp.</td>
<td>10^{-56}</td>
<td>34</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>Nematodes</td>
<td>UNK</td>
<td>AAD12787</td>
<td>UGM</td>
<td>10^{-68}</td>
<td>34</td>
</tr>
<tr>
<td>Ciona intestinalis</td>
<td>Nematodes</td>
<td>UNK</td>
<td>Table S2</td>
<td></td>
<td>10^{-62}</td>
<td>31</td>
</tr>
<tr>
<td>Halocynthia rosetei</td>
<td>Urochordates</td>
<td>UNK</td>
<td>BAB20093</td>
<td>HorLTI5</td>
<td>10^{-46}</td>
<td>26</td>
</tr>
<tr>
<td>Pyroccocus horikoshii</td>
<td>Archea</td>
<td>UNK</td>
<td>B71115</td>
<td>Hyp.</td>
<td>10^{-27}</td>
<td>27</td>
</tr>
<tr>
<td>Pyrobaculum aerophilum</td>
<td>Archea</td>
<td>UNK</td>
<td>AAL62855</td>
<td>Hyp.</td>
<td>10^{-36}</td>
<td>27</td>
</tr>
<tr>
<td>Yersinia pseudotuberculosis</td>
<td>Eubacteria</td>
<td>PROB</td>
<td>CAB63294</td>
<td>O-antigen WboH</td>
<td>10^{-18}</td>
<td>22</td>
</tr>
<tr>
<td>Methanococcus maripaludis</td>
<td>Archea</td>
<td>UNK</td>
<td>CAF30324</td>
<td>Hyp.</td>
<td>10^{-14}</td>
<td>21</td>
</tr>
<tr>
<td>Klebsiella pneumonia</td>
<td>Eubacteria</td>
<td>YES</td>
<td>Q48485</td>
<td>UGM</td>
<td>0.2</td>
<td>&lt;10%</td>
</tr>
</tbody>
</table>

\(^a\) The genes are ordered based on their BlastP P value, based on searches with the L. major GLF query sequence.
\(^b\) YES, evidence showing occurrence of Gal in one or more glycoconjugates; PROB, Gal suspected to be present; UNK, not investigated or established.
\(^c\) Hyp., hypothetical protein. Annotations of L. major, T. cruzi, C. neoforans, and C. elegans GLFs reflect GenBank depositions arising from this work.

\(\text{GLF}\) by episomal expression of \(\text{KpGLF}\) abrogated LPS biosynthesis, which could be restored by episomal expression of \(\text{KpGLF}\). We surmised that this would provide a rapid and convenient test for the activity of the candidate eukaryotic GLF family ORFs encoded UGM activity.

**FIG. 1.** Phylogenetic tree of the eukaryotic GLF family ORFs. A minimum evolution tree for the predicted protein sequences of candidate eukaryotic GLFs was constructed using algorithms implicated in the MEGA3 software package (11). A "+" marks species where Gal has been detected; Gal has not been examined in the other taxa shown. Species where the GLF-encoded UGM activity was confirmed in this work are boxed. For information on the specific sequences analyzed, see Table 1 and Table S2 in the supplemental material. The scale at the bottom corresponds to the fraction of amino acid sequence difference.
potential heterologous GLFs. Accordingly, we obtained the L. major, T. cruzi, C. neoformans, and C. elegans GLF ORFs by PCR and introduced them into the bacterial expression vector pET3a. These were then introduced into E. coli strain CWG288 (deleted for the endogenous rfb locus) containing plasmid pWQ70, which bears the Klebsiella rfb locus with an inactivated GLF (rfbD) gene (10). LPS expression was monitored by Western blotting with rabbit anti-galactan I antibody, which recognizes a [Gal$_f$-Gal$_f$] repeating unit (5).

We confirmed that the intact Klebsiella rfb locus restored Gal$_f$-containing O1 LPS biosynthesis in E. coli strain CWG288 transformed with pWQ71, which bears an intact rfb locus, while plasmid pWQ70 bearing the inactivated GLF failed to do so (Fig. 3, lane 1 versus lane 2). Transformation of strain CWG288/pWQ70 with any of the four eukaryotic species’ pET3a-GLF constructs (L. major, T. cruzi, C. neoformans, or C. elegans) resulted in synthesis of LPS reactive with anti-Klebsiella galactan I antisera, at levels comparable to those previously described prokaryotic enzymes, the amino acid sequence of purified soluble and active eukaryotic UGMs will address the catalytic properties of these enzymes relative to those of prokaryotes in more detail.

**DISCUSSION**

By using a combinatorial bioinformatics approach incorporating criteria including sequence homology, protein motifs, and phylogenetic associations with Gal$_f$ synthesis, we identified a candidate Leishmania GLF gene and from this a large family of potential eukaryotic GLFs. We used a rapid in vivo assay of GLFs expressed in E. coli as a first screen for UGM activity of several candidate eukaryotic proteins (Fig. 3), which were confirmed subsequently in enzymatic assays (Fig. 4). While the eukaryotic UGM/GLF proteins show some relationship to the previously described prokaryotic enzymes, the amino acid sequence divergence was extensive, and the correlation of potential GLF homologs with organisms known to synthesize Gal$_f$ was imperfect. Perhaps for these reasons, eukaryotic GLFs had remained elusive despite efforts undertaken by many researchers.

Despite the extensive divergence, eukaryotic UGMs share key properties with those of prokaryotes. The eukaryotic proteins contain clear flavin binding motifs, and most importantly, a number of residues implicated by functional or structural criterion in substrate binding were conserved (Fig. 1). Preliminary analysis suggests that the eukaryotic UGMs can be readily modeled to the E. coli UGM structure determined previously, with a number of sequence insertions potentially mapping to external loops (data not shown). As in prokaryotes, eukaryotic UGMs lack obvious secretory signals and are likely to be suggested that the eukaryotic candidate GLFs mediated Gal$_f$ synthesis when expressed heterologously in E. coli.

While our data suggest that similar levels of LPS restoration were seen in the experimental versus control samples, some restraint is required in interpreting these data. As shown below, the eukaryotic UGMs are largely insoluble when expressed in E. coli. Thus, differences in the degree of eukaryotic GLF rescue of LPS expression may reflect either intrinsic differences in catalytic activity of the enzymes or “noise” arising from differences in expression and/or folding. Nonetheless, these data suggest that the eukaryotic UGMs are able to provide sufficient UDP-Gal$_f$ under these conditions for LPS synthesis comparable to that of authentic bacterial UGMs.

**UGM activity of L. major and C. neoformans GLFs.** To confirm directly that candidate GLFs encoded UGM activity, we first expressed high levels of protein in E. coli using the pET3a system. While abundant expression was evident in sodium dodecyl sulfate-PAGE analysis, the majority of protein appeared in insoluble inclusion bodies (data not shown). Nonetheless, soluble protein extracts were prepared, and UGM activity was assayed in the reverse direction, following the conversion of UDP-Gal$_f$ to UDP-Gal$_p$ by HPLC (14). Since the E. coli strain BL21(DE3) lacks GLF, there was no endogenous background UGM activity in crude preparations (15) (Fig. 4E and F). In contrast, with extracts from cells expressing L. major or C. neoformans GLFs, the UDP-Gal$_p$ substrate was converted to UDP-Gal$_f$ (Fig. 4C and D) in a protein-dependent manner (Fig. 4A and B). These data establish that these eukaryotic GLFs encode proteins with UGM activity. In the future, studies of purified soluble and active eukaryotic UGMs will address the catalytic properties of these enzymes relative to those of prokaryotes in more detail.
cytoplasmic. Since the bulk of eukaryotic glycoconjugate synthesis typically occurs within the secretory pathway, this suggests that one or more nucleotide sugar transporters must recognize and transport UDP-Galf to lumenal compartments.

We found a number of genes with strong homology to GLFs encoding eukaryotic UGMs in species not known to synthesize Galf (Fig. 1 and 2; Table 1; see Tables S1 and S2 in the supplemental material). These predicted proteins also showed conservation of key UGM substrate-binding residues (Fig. 1, data not shown). Since the phylogenetic tree presented in Fig. 2 showed that members of both major branches of the eukaryotic GLF family had UGM activity, by parsimony we believe that most or all of these candidate GLFs encode proteins with UGM activity as well. Given the paucity of information on Galf in eukaryotes outside of fungi and protozoa, these new eukaryotic GLFs raise the possibility that Galf synthesis may occur more widely in species than previously supposed. For fungi such as Ustilago, Magnaporthe, and Giberella, this was unsurprising, but the occurrence of eukaryotic GLFs in the nematodes C. elegans, the urochordates Ciona and Halocynthia, and the algae Chlamydomonas and algal virus Feldmannia (Fig. 1; Table 1; see Tables S1 and S2 in the supplemental material) was unanticipated. Notably, our data show that C. elegans GLF encodes a protein with UGM activity. Additionally, we found strong GLF homologs in a variety of unfinished genome and EST surveys, including Histoplasma capsulatum, which contains Galf within sphingolipids (2), and nematode parasites of humans (Brugia malayi and Onchocerca volvulus), animals (Strongyloides), and plants (Heterodura and Meloidogyne) (see Table S1 in the supplemental material) (3, 33).

Assuming that the presence of eukaryotic GLF within a species is predictive of Galf synthesis, our analyses suggest that the phylogenetic distribution of the GLF/Galf synthetic pathway shows some features reminiscent of those seen in prokaryotes. First, the GLF/Galf pathway occurs only sporadically, in widely divergent species, and second, it is often found in pathogenic/parasitic species from both the microbial and metazoan worlds. Why pathogens show an affinity for inclusion of Galf in their metabolomic repertoire is unknown; potential explanations might include the strong immunogenicity of Galf and/or its ability to adopt novel structural conformations (17, 31), both of which could contribute to microbial virulence and survival. Notably, the GLF relationships depicted in Fig. 2 or evident in Table 1 often do not closely follow those of the species involved. Potentially, the predilection of pathogens for Galf could be a contributing factor to the sporadic distribution of Galf synthesis among species, perhaps by lateral gene transfer mechanisms as seen in bacterial LPS O antigens (19). The presence of a GLF within the algal virus Feldmannia (Fig. 2) is

![Figure 4](https://example.com/figure4.png)
interesting in this light, since viruses often can be transmitted laterally among species.

Galα-containing glycoconjugates are often dominant features of the surface of many protozoans and fungi, making Galα synthesis a potential target for chemotherapy. However, the role of Galα conjugates in the survival and virulence of eukaryotic microbes appear to vary greatly among species. For example, in fungi, Galα comprises only a small portion of the Cryptococcus capsule (30) but a major fraction of the abundant galactofuranose-containing glycoconjugates. Preliminary tests of a panel of prospective prokaryotic UGM inhibitors (22) suggest that they fail to inhibit the eukaryotic UGMs studied here (unpublished work), in keeping with the extensive amino acid sequence divergence (Fig. 1). Thus, it will be necessary to identify eukaryote-specific UGM inhibitors in the future. Interestingly, it has recently been suggested that several of the inhibitors of the Mycobacterium tuberculosis UGM may also act against bacteria independently of UGM (28).

For Leishmania, genetic studies suggest that parasites lacking Galα glycoconjugates, such as LPG and glycoinositolphospholipids, retain virulence in mammalian infections, although their ability to be transmitted by the insect vector sand fly is greatly reduced (26, 34, 35). Whether a similar conclusion pertains to Trypanosoma cruzi, where Galα-containing molecules are also highly abundant in infective stages, awaits confirmation (18). Given that the Leishmania and T. cruzi genomes encode numerous potential UDP-Galα transferases (34), genetic inactivation of their GLFs (which occur in one or two copies, respectively; Table 1) may offer an easier definitive test (9). Interestingly, T. brucei, which is thought to be evolutionarily more closely related to T. cruzi than Leishmania, lacks GLF and Galα-containing conjugates. While this could arise from evolutionary loss of GLF in this lineage, the complex disjunct distribution and relationships of GLF genes in other eukaryotic taxa suggest that other explanations cannot yet be ruled out.

In summary, we have identified a large eukaryotic GLF family and provided evidence that members of the two major groups within this family possess UGM activity. This in turn has permitted predictions about the occurrence of Galα synthesis among species and its frequent appearance within both microbial and metazoan pathogens. Future work will focus on exploitation of UGM-targeted inhibitors in chemotherapy and determination of the structure and role of Galα in those lineages for which information is currently lacking.

ACKNOWLEDGMENTS

This work was supported by NIH grants AI31903 (S.M.B.), GM66303 (T.D.), AI33706, and AI057836 (M.R.M.).

We thank Natalia Akopyants for help with PCR amplification, Chad Rappleye, David Sanders, Sam Turco, Chris Whitfield, and Brad Clarke for helpful discussions and/or for providing E. coli strains and plasmids, Daniella Bartholomew for T. cruzi cDNA; Yuji Kohara for providing C. elegans EST clones, and A. Capul, D. E. Dobson, and K. Zhang for comments on the manuscript. We also express our appreciation to the many genome and EST projects contributing data incorporated into Table 1 and into Tables S1 and S2 in the supplemental material.

REFERENCES


