Three-Dimensional Structure of a Protozoal Double-Stranded RNA Virus That Infects the Enteric Pathogen *Giardia lamblia*

Mandy E. W. Janssen, a Yuko Takagi, c Kristin N. Parent, a Giovanni Cardone, a Max L. Nibert, c Timothy S. Baker a,b

Department of Chemistry and Biochemistry a and Division of Biological Sciences, b University of California–San Diego, La Jolla, California, USA; Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts, USA c

**ABSTRACT**

*Giardia lamblia* virus (GLV) is a small, nonenveloped, nonsegmented double-stranded RNA (dsRNA) virus infecting *Giardia lamblia*, the most common protozoan pathogen of the human intestine and a major agent of waterborne diarrheal disease worldwide. GLV (genus *Giardiavirus*) is a member of family *Totiviridae*, along with several other groups of protozoal or fungal viruses, including *Leishmania* RNA viruses and *Trichomonas vaginalis* viruses. Interestingly, GLV is more closely related than other *Totiviridae* members to a group of recently discovered metazoan viruses that includes penaeid shrimp infectious myonecrosis virus (IMNV). Moreover, GLV is the only known protozoal dsRNA virus that can transmit efficiently by extracellular means, also like IMNV. In this study, we used transmission electron cryomicroscopy and icosahedral image reconstruction to examine the GLV virion at an estimated resolution of 6.0 Å. Its outermost diameter is 485 Å, making it the largest totivirus capsid analyzed to date. Structural comparisons of GLV and other totiviruses highlighted a related “T=2” capsid organization and a conserved helix-rich fold in the capsid subunits. In agreement with its unique capacity as a protozoal dsRNA virus to survive and transmit through extracellular environments, GLV was found to be more thermoresistant than *Trichomonas vaginalis* virus 1, but no specific protein machinery to mediate cell entry, such as the fiber complexes in IMNV, could be localized. These and other structural and biochemical findings provide a basis for future work to dissect the cell entry mechanism of GLV into a “primitive” (early-branching) eukaryotic host and an important enteric pathogen of humans.

**IMPORTANCE**

Numerous pathogenic bacteria, including *Corynebacterium diphtheriae*, *Salmonella enterica*, and *Vibrio cholerae*, are infected with lysogenic bacteriophages that contribute significantly to bacterial virulence. In line with this phenomenon, several pathogenic protozoa, including *Giardia lamblia*, *Leishmania* species, and *Trichomonas vaginalis* are persistently infected with dsRNA viruses, and growing evidence indicates that at least some of these protozoal viruses can likewise enhance the pathogenicity of their hosts. Understanding of these protozoal viruses, however, lags far behind that of many bacteriophages. Here, we investigated the dsRNA virus that infects the widespread enteric parasite *Giardia lamblia*. Using electron cryomicroscopy and icosahedral image reconstruction, we determined the virion structure of *Giardia lamblia* virus, obtaining new information relating to its assembly, stability, functions in cell entry and transcription, and similarities and differences with other dsRNA viruses. The results of our study set the stage for further mechanistic work on the roles of these viruses in protozoal virulence.

*Giardia lamblia* (also called *G. intestinalis* or *duodenalis*) is a flagellated, parasitic protozoan and an important cause of diarrheal disease worldwide, especially in developing countries. It is a member of the supergroup Excavata, one of the earliest diverging eukaryotic taxa, to which the parasitic protozoa *Leishmania*, *Naegleria*, *Trichomonas*, and *Trypanosoma* also belong, although in distinct groups or subgroups relative to *Giardia* (1). Infection by *G. lamblia* occurs after ingestion of cysts from contaminated water, food, or soil or sometimes by direct contact. After excystation and multiplication, *G. lamblia* trophozoites adhere to the surface of epithelial cells in the small intestine. Specific pathogenic mechanisms remain poorly characterized, but the resulting diarrheal disease is characterized by malabsorption, malnutrition, chloride hypersecretion, and increased intestinal transit rates (2, 3). Histopathological findings include lymphocyte infiltration and diffuse blunting of the microvilli, enterocyte apoptosis, and disruption of epithelial tight junctions.

Epidemiologically, *G. lamblia* is the most common protozoan pathogen of the human small intestine, estimated to cause ≥280 million cases yearly (2, 4), and is also common in companion animals and livestock. Most infections are self-limiting, but in immunocompromised persons, including undernourished children and people with AIDS, *G. lamblia* disease may be more chronic, severe, and atypical and may be associated in children with growth retardation and cognitive impairment (5, 6). Nota-
Interestingly, GLV is the only one of the known protozoal dsRNA viruses that can transmit efficiently by extracellular means (8, 21), meaning that GLV virions uniquely include the protein machinery for mediating efficient cell entry. The GLV-Giardia virus-host system thus provides an unusual opportunity to determine the steps involved in cell entry of a “primitive” (early-branched) eukaryote, with a simplified endo/lysosomal vacuolar system (22, 23). Attachment of GLV to the G. lamblia host cell has been shown to involve a cell surface receptor, which is missing from some G. lamblia strains (21, 24), but the identity of this receptor remains unknown. Otherwise, the entry mechanism of GLV is poorly characterized, except for one report suggesting that it is susceptible to inhibition by ammonium chloride and chloroquine and thus may require endocytosis and vacuolar acidification (25).

Resemblance of GLV to other Totiviridae members has not been previously confirmed at the structural level, except by negative-stain transmission electron microscopy (TEM) showing non-enveloped particles with single-layered capsids, appearing to measure 340 to 370 Å in diameter (8, 26). Moreover, in addition to its virions containing the protein machinery for efficient cell entry, which the other Totiviridae members from protozoal or fungal hosts do not, GLV appears to be more closely related to a group of nonsegmented dsRNA viruses isolated in recent years from metazoan hosts, both arthropods (crustaceans and insects) and vertebrates (bony fish) (27–31) and tentatively also assigned to family Totiviridae (Fig. 1). Notably, these metazoan viruses share with GLV the ability to transmit efficiently by extracellular means. The virions of one of these viruses, penaeid shrimp infectious myonecrosis virus (IMNV), have been analyzed by transmission electron cryomicroscopy (cryo-TEM) and icosahedral three-dimensional (3D) image reconstruction, which revealed an unusual protein-fiber complex that projects above the so-called “T=2” capsid at the icosahedral 5-fold axes and is likely to play one or more key roles in cell entry (32, 33). Such “T=2” capsid organization (34) reflects a T=1 structure in which a CP dimer occupies each of the icosahedral asymmetric units (120 CP subunits per capsid) and is common to many dsRNA viruses, including other members of family Totiviridae (35–39).

Considering the variety of interesting biological aspects of GLV and also the importance of its protozoal host G. lamblia as a...
human disease agent, we undertook this study to determine the virion structure of GLV via cryo-TEM and icosahedral 3D image reconstruction. In parallel, we also investigated several of the in vitro biochemical properties of GLV virions. The results of these studies reveal both similarities and differences between GLV and other Totiviridae members, including TVV1, prototype fungal totivirus *Saccharomyces cerevisiae* virus L-A (ScV-L-A), and tentatively assigned member IMNV, and set the stage for ongoing studies of cell entry by GLV.

**MATERIALS AND METHODS**

Protozoan cultures and virion purifications. A frozen aliquot of *G. lamblia* strain WBI (GLV-infected WBI) cells was kindly provided by C. C. Wang and coworkers (University of California—San Francisco). The batch culture was maintained as described previously (26) and expanded in cell numbers and volume over several days before harvest for virion purification. At harvest time, containers were chilled in an ice bath and swirled to remove attached cells from the glass surface. GLV virions were customarily purified from a batch of 2 × 10^9 WBI cells, using the same protocol of sonication and centrifugations as previously described for TVV1 virions (18, 39). GLV virions obtained from the last centrifugation step (banding in a CsCl density gradient) were dialyzed overnight against HNM buffer (50 mM HEPES [pH 7.2], 500 mM NaCl, 20 mM MgCl_2) and then kept at 4°C for up to 3 weeks or at −80°C for longer periods before use. Concentrations of the dialyzed preparations routinely measured near 200 μg/ml (−10^{13} virions/ml) according to a modified Bradford protein assay (Bio-Rad). *T. vaginalis* isolate UH9 was cultured, and TVV1-UH9 virions were purified as previously described (18, 39). The original *T. vaginalis* UH9 stock, which was kindly provided by Raina N. Fichorova and coworkers (Brigham and Women’s Hospital and Harvard Medical School), had been serially cloned twice in soft agar before use in this study. GLV CP and CP/RdRp sequences. Analyses were otherwise performed as described previously (39).

**TEM.** Prior to performing cryo-TEM, the integrity and homogeneity of samples were evaluated by using negative-stain TEM for this, 3.5-μl aliquots of sample were absorbed to a carbon-coated copper grid that had been glow-discharged for ~60 s in an Emitech K350 evaporation unit. The absorbed sample was then stained with 1% (unbuffered) uranyl acetate and rinsed with deionized distilled water. Micrographs were recorded on a 4K² Gatan Ultrascan charge-coupled device camera in an FEI G2 Tecnai (Polara) microscope. Initial experiments performed to judge proper storage conditions for purified GLV virions and to provide an initial gauge of viral astigmatism and specimen drift were digitized at 6.35-μm intervals (1.09-Å pixel size at the specimen) on a Nikon Super Coolscan 8000 mi-

crodensitometer. RobEM (http://cryoem.ucsd.edu/programs.shtm) was used to extract 14,125 particle images, each 601 × 601 pixels in size, and to pre-process the images as described previously (41). The defocus and astigmatism levels of each micrograph were estimated using the program CTFIND3 (42). The random-model computation procedure (43) was used to generate an initial 3D density map at an ~29-Å resolution from the images of 150 particles. This map served as a starting model to initiate full orientation and origin determinations of the entire set of images using AUTO3DEM v4.03.1 (44). After several cycles of orientation and origin refinement with the POVR and P3DR functions of AUTO3DEM, the resolution converged to ~6.0 Å (based on the 0.5 threshold of the Fourier shell correlation criterion [45]). The effects of the microscope contrast-transfer function were compensated in part by inverting image phases in spatial frequency bands where contrast reversals occurred (46). An inverse temperature factor of 1/(80 Å²) was applied to the final reconstruction (47).

The Segger v1.7 tool in the Chimera visualization package (48) was used to segment out individual CP-A and CP-B subunits in the final map and to examine interactions between them as described previously (38). Briefly, segmentation is obtained with this tool by means of an iterative procedure, starting from an initial arbitrary assignment of density regions to the two subunits and then refining those assignments in a semiautomatic manner by imposing additional restrictions. The additional restrictions in this case were based on evidence from the estimated capsid volume that the capsid indeed comprises 120 CP subunits, i.e., two CP subunits per icosahedral asymmetric unit, and on the following assumptions: (i) each subunit forms a compact domain, (ii) global morphology and particular features are repeated in CP-A and CP-B, (iii) symmetry constraints of the icosahedral arrangement of subunits within the capsid are maintained, and (iv) no obvious capsid density is left unassigned. Under these constraints, the final result does not depend on the initial arbitrary assignment to the two subunits. GLV subunit A was manually docked into GLV subunit B for comparison and the Chimera “fit-in-map” tool was then used to refine the fit. The normalized correlation value of the final fitting was 0.91 as provided by Chimera (48). Location and length of density features that could be assigned to putative α-helical segments in the segmented A subunits of the different totiviruses were predicted automatically using the VolTrac algorithm (49) implemented in the Sculptor graphics software (50). The procedure was performed as described by Dunn et al. (38).

**Temperature stability of virions.** GLV or TVV1 virions (2 × 10^9 to 5 × 10^10 in 10 μl) were incubated in HNM buffer for 5 min at the indicated temperatures, using the gradient setting of a thermal cycler (MultiGene; Labnet International). RNase III (New England BioLabs) was then added to a final concentration of 0.02 U/μl, followed by incubation for 5 min at 37°C to digest dsRNA that was released from, or made accessible inside, destabilized virions. RNase III digestion was quenched by addition of EDTA to a final concentration of 100 mM. Protected dsRNA was then released from intact virions by heating the mixture for 5 min at 65–75°C. The genomic dsRNA bands of GLV and TVV1 were visualized on a 0.7% agarose gel, stained with ethidium bromide, and quantitated by using ImageJ software (http://imagej.nih.gov/ij/). The plot and fit for Fig. 6A were generated using DataGraph 3.1.1.

**In vitro transcription.** GLV or TVV1 virions (10^10 in 20 μl) were incubated in transcription buffer (150 mM NaCl; 2 mM MgCl_2; 200 μM concentrations each of ATP, CTP, and UTP; 20 μM [α-32P]GTP; and 50 mM Tris-HCl or Tris/acetate buffer at the indicated pH value) for 30 min at 37°C. The transcription reaction was terminated by addition of formamide and EDTA to final respective concentrations of 30% and 17 mM. Transcripts were separated on an 8% denaturing polyacrylamide gel containing 7 M urea in 0.5× Tris-borate-EDTA buffer. Radiolabeled RNA bands were then visualized by phosphorimaging on a Typhoon 9400 variable mode imager and quantitated using ImageQuant TL software (GE Healthcare). The plot and fit for Fig. 6B were generated using DataGraph 3.1.1.
Phylogenetic tree and polarity plot. A maximum-likelihood phylogenetic tree was reconstructed from sequences encompassing the predicted CP+RdRp products of each analyzed virus (see Fig. 1). Multiple-sequence alignments were conducted using MAFFT 7.1, as implemented with default settings at http://mafft.cbrc.jp/alignment/software. Model testing was then performed using ProtTest 3.3 (51), which showed the Blosum62 substitution matrix and I+G+F distributions (proportion of invariable sites, 0.01; gamma shape parameter, 2.975; and empirically determined amino acid frequencies) to score most highly according to both Akaike and Bayesian information criteria. Phylogenetic analyses were next performed with these predetermined parameters using PhyML 3.0 (52) as implemented at http://www.hiv.lanl.gov/. The starting tree was optimized by both branch length and tree topology, and tree improvement was performed according to the best of nearest neighbor interchange and subtree pruning and regrafting. Branch support values (%) were estimated by the approximate likelihood ratio test with SH-like supports. The final tree was refined for display using FigTree 1.4.0. The abbreviations (along with the corresponding GenBank accession numbers) for the viruses included in the tree are as follows: GLV (NC_003555); TVV1 (NC_003824); TVV2 (NC_003873); TVV3 (NC_004034); TVV4 (HQ607522); LRV1 (NC_002063); LRV2 (NC_002064); EbRV1, Eimeria brunetti RNA virus 1 (NC_002701); SrRV2, Sphaeropsis sapinea RNA virus 2 (NC_001964); GaRV1-L1, Gremmeniella abietina RNA virus L1 (NC_003876); MoV2, Magnaporthe oryzae virus 2 (NC_010246); CmRV, Coniothyrium minitans RNA virus (NC_007523); EfV1, Epichloë festucae virus 1 (AM264127); HmTV1-17, Helicobasidium monsporiotus virus 1-17 (NC_005074); MoV1, Magnaporthe oryzae virus 1 (NC_006367); SrRV1, Sphaeropsis sapinea RNA virus 1 (NC_001963); AsSV1, Aspergillus fumigatus slow virus (HE588147); BbV1, Beauveria bassiana victorivirus 1 (HE572591); RnV1, Rosellinia necatrix victorivirus 1 (AB698489); CeRV1, Chalara elegans RNA virus 1 (NC_005883); HtV1, Helminthosporium victoriae virus 1905 (NC_003607); TeV1, Tolyphocladium cylindrosporum virus 1 (NC_014823); XdV-L1A, Xanthophyllomyces dendrorhous virus L1A (NC_020903); XdV-L1B, Xanthophyllomyces dendrorhous virus Lib (JN997473); TaV1, Tuber aestivum virus 1 (HQ158596); ScV-L-A (L1) (NC_003745); ScV-L-BC (La) (NC_001641); ScV-L, Scheffersomyces segobiiensis virus 1 (KC610514); UmV-H1, Ustilago maydis virus H1 (NC_003823); ORV, Omono River virus (AB555544); DmTV, Drosophila melanogaster totivirus (NC_013499); AsV, Armigeres subalbatus virus (NC_014609); IMNV (NC_007915); and PMCV, piscine myocarditis virus (NC_015639).

A polarity plot was generated for the full-length GLV ORF 1 product (CP precursor) according to the method of Grantham (53) as implemented at http://web.expasy.org/protscale/ with a running window size of 15 (see Fig. 3B).

Accession number. The 3D density map of the cryo-reconstruction of the GLV virion has been deposited in the Electron Microscopy Data Bank at the European Bioinformatics Institute with accession code EMD-5948.

RESULTS

Initial characterizations of virions used for cryo-TEM studies. G. lambia strain WBI (strain WB infected with GLV strain Portland I) was analyzed, because it has been extensively studied by Wang and coworkers (8–14, 21, 24–26). WBI was serially passaged in liquid batch culture, with a doubling time of ~24 h. After harvest, cells were sonicated to release virions, centrifuged to deplete debris, and sedimented into a CsCl density gradient. The gradient was fractionated, and fractions containing virions were identified by SDS-PAGE to track GLV CP. Virions consistently concentrated in only one or two consecutive fractions, corresponding to a visible band in the gradient, well separated from any debris. The virus fractions were dialyzed against buffer and stored at either 4°C or ~80°C before use.

SDS-PAGE of the purified GLV virions showed a major Coomassie-stained band migrating near 100 kDa, a finding consistent with the sequence-predicted mass of 95 to 98 kDa for GLV CP (95 kDa if missing its N-terminal 32 amino acids [aa] [10]) (Fig. 2A). A minor Coomassie blue-stained band migrating near 200 kDa was also routinely visible, consistent with the lower copy number and sequence-predicted mass of 211 kDa for GLV CP/RdRp (only 1 or 2 copies per virion versus 119 or 118, respectively, for CP). An ethidium-stained band migrating near the top of the SDS-PAGE gel, and consistent with the GLV dsRNA genome, was additionally seen (Fig. 2A).

To evaluate suitability of the sample for structural analysis, we...
FIG 3 Mass spectrometry to assess protein contents of GLV virions. (A) Sequence of GLV CP precursor is shown. Arrow indicates position of the N-proximal processing cleavage (overlining, N-terminal sequencing residues) are translated from panel A. (B) A polarity plot was generated as described in Materials and Methods. Highlighted features (arrow, underlines) are translated from panel A.

negatively stained the purified GLV virions and inspected them by TEM. The particles were seen to be well distributed, with minimal breakage and consistent widths near 450 Å (Fig. 2B). Most displayed a regular, circular outline. Essentially all particles appeared to be penetrated by stain, but most had a heterogeneous distribution of interior staining consistent with the presence of centrally packaged GLV genome. A few "empty" particles, with more homogeneous interior staining consistent with the absence of genome, were also observed. Similar findings were obtained for purified virions that had been stored at 4°C or at −80°C for several weeks, with or without addition of glycerol. Storage at −20°C was less reliable, however, as particles in clumped, empty, and broken states were observed in increased numbers.

Purified GLV virions were submitted for LC-MS/MS to assess their protein contents. Parallel analyses were performed: one for peptides generated by trypsin treatment and the other by chymotrypsin treatment. The identified peptides, respectively, covered 56 and 46% of the deduced CP sequence, or 65% coverage by the combined results (Fig. 3A). A notable finding was that no peptides were identified from the N-terminal 32 aa of the ORF1-encoded CP sequence, which have been indicated by Yu et al. (10) to be removed by proteolytic cleavage in infected cells and to be missing from virions. Peptides were also recovered from the RdRp region of CP/RdRp in these analyses, although at lower frequency, 25%

coverage of the predicted RdRp sequence by the combination of peptides from either treatment, consistent with the lower copy number of CP/RdRp. Peptides spanning the CP/RdRp fusion junction were not recovered.

Cryo-TEM and 3D reconstruction: surface views and equatorial section. Purified GLV virions vitrified and visualized by cryo-TEM appeared consistently sized and shaped, with mostly unbroken capsids (Fig. 2C). The genomic dsRNA enclosed within the capsid layer appeared as distributed density, without the "fingerprint" motif that is observed in virions of other totiviruses, including tentatively assigned member IMNV (32, 35, 36, 39). A few particles, however, exhibited lower central densities, and linear strands suggestive of free dsRNA were seen at places in the solvent background, consistent with loss of genome from a few particles in the preparation (Fig. 2C). The capsids in this mixture of mostly "full" and few "empty" particles displayed a very regular, circular outline, without any obvious protrusions. An icosahedral 3D image reconstruction was computed from 14,125 particle images recorded on 101 micrographs. The resolution of the final reconstruction was estimated at 6.0 Å based on a Fourier-shell correlation threshold of 0.5 (45).

As expected for totiviruses (32, 35–39), a surface view of the 3D reconstruction revealed that GLV appears to possess a so-called "T=2" capsid (34) comprising 120 CP subunits in a T=1 lattice (Fig. 4A and C). The estimated volume of the capsid shell is also consistent with this conclusion. Large portions of the CP outer surfaces form a raised but mostly flat "plateau" centered on each of the icosahedral 5-fold axes, with the portion of each subunit contributing to this plateau having the appearance of an elongated, triangular "pizza slice." The wider ("crust") end of each slice lies distal to the 5-fold axis. From this perspective, it is notable that each CP subunit occupies either of two nonidentical positions within the capsid (Fig. 4A). In one position (designated "A") as in other "T=2" capsids, the CP subunits approach one another around each 5-fold axis via the narrow ("toppings") end of the figurative slice, while in the other position (designated "B"), the CP subunits lie wedged between neighboring A subunits, in a similar orientation but more distant from the 5-fold axis. Each plateau thus appears to be constituted by portions of 10 CP subunits: five A and five B. By comparison, each of the icosahedral 3-fold axes, which are low points on the capsid surface, appears to be immediately surrounded by three B subunits, each from a different decameric plateau (Fig. 4A). Each of the icosahedral 2-fold axes, on the other hand, appears to be immediately surrounded by four CP subunits: one A and one B subunit from one decameric plateau and one A and one B subunit from the adjacent decameric plateau (Fig. 4A). Moreover, each of these A subunits appears to make contact in a quasi-2-fold-symmetrical, "tail-to-tail" (or "crust-to-crust") fashion with the B subunit from the adjacent decamer, thereby forming two elongated A:B ridges that pass alongside the icosahedral 2-fold axis, which is another low point on the capsid surface.

The equatorial section of the 3D reconstruction shows both punctate and linear features, which at this level of resolution suggest a prevalence of α-helical secondary structure in the CP subunits (Fig. 4B). It also shows that the single-layered capsid of GLV has a maximum thickness of ~60 Å (median, ~40 Å). The outermost capsid diameter is ~485 Å, making this the largest totivirus capsid studied to date (35–39), including the capsid (sans spikes) of IMNV (32). The central section also reveals an interior capsid...
surface that is smoother than the exterior, except for short inward protrusions near the icosahedral 3-fold axes.

Unlike what has been observed to date for other totiviruses (32, 35, 36, 38, 39), the packaged genome of GLV does not appear in the reconstruction as a series of concentric shells of density in the particle interior. This finding is consistent with the lack of central fingerprint densities in the raw cryo-micrographs of GLV (Fig. 2C). Instead, densities attributable to genome are distributed almost homogeneously throughout the interior of the capsid, except for a hint of greater density in some localized regions nearest the shell (Fig. 4B). The most notable contacts between capsid and RNA densities appear to occur near the icosahedral 3-fold axes, as evident in the equatorial section (Fig. 4B, filled white arrowhead). An additional, smaller contact appears to be located between the icosahedral 5-fold and 2-fold axes (Fig. 4B, unfilled white arrowhead).

Subunit arrangements observed in radial sections. Radial sections of the 3D reconstruction were found to complement surface views for gaining a better understanding of GLV capsid organization. Clearly distinguishable, triangular “pizza-slice” elements fully constitute the sections at higher radii (e.g., 236 to 214 Å in Fig. 4C). Visualized especially well at radius 214 Å, but also at 223 and 217 Å, is how the capsid is organized with four of these elements, representing two A and two B subunits, arrayed around each icosahedral 2-fold axis (Fig. 4C). As also inferred from the surface view, these four subunits around each icosahedral 2-fold axis appear to be arranged as two antiparallel A:B pairs, within each of which the A and B subunits make quasi-2-fold-symmetric tail-to-tail contacts across a broad, flat interface. These A:B contacts appear to be the major ones that connect adjacent decamers at these higher radii, although B:B contacts around the icosahedral 3-fold axes appear also to contribute. Progressing to lower radii in the sections, it is nearly impossible to distinguish individual CP subunits at radii 209 and 199 Å (Fig. 4C). Instead, at these radii almost continuous protein densities are seen, representing the core of the capsid shell. At the lowest radii in the capsid (e.g., radii 191 and 182 Å in Fig. 4C), however, the A and B subunits are again distinguishable. The close correspondence in features between A and B subunits, across the full capsid thickness, is notable in the radial sec-
This correspondence pertains not only to their similar outlines but also their similar internal features appearing to represent secondary-structure elements. The A and B subunits thus appear to lie at nearly the same radial heights throughout the capsid, although close inspection suggests that the A subunits may lie just a few angstroms (i.e., 1 to 3 Å) higher. In addition, a small clockwise rotation along the longitudinal axis of the A subunits (relative to B subunits; see below) may explain subtle differences observed between the two classes of subunits, especially at lower radii (e.g., see radius 191 Å).

The positions of channels through the GLV capsid are of interest, in particular with regard to RNA transcription and RNA release from particles. Channels 10 to 15 Å wide would be large enough to permit translocation of nucleotide triphosphates and perhaps newly synthesized RNA transcripts, but not the dsRNA genome in the absence of movements of the capsid elements. In the equatorial section, open channels that span the full capsid thickness appear to be present at the icosahedral 5-fold axes (Fig. 4C). Using the radial sections for size estimates, however, we found that this channel narrows in width to only 10 to 11 Å at radius 209 Å and remains nearly as narrow at radius 199 Å (Fig. 4C). Open channels are not evident at the icosahedral 3- and 2-fold axes, although the capsid in these locations does appear to be quite thin.

Capsid segmentation. To ascribe specific features in the GLV 3D density map to the A and B subunits (Fig. 5), we used the ab initio approach previously used to segment the maps of HvV190S and TVV1 (38, 39). The segmentation procedure was most difficult to perform near the icosahedral 2-fold axes, where the A and B subunits from adjacent decamers seem to interact extensively in a quasi-2-fold-symmetrical arrangement (see radii 214 and 209 Å in Fig. 4C). Nonetheless, iterative refinement of density assignments, obeying the constraints described previously (39) and carried out independently by two persons, provided consistent iden-
tification of the molecular envelopes and boundaries of CP-A and CP-B (Fig. 5). Secondary-structure elements similar in shape and placement were visible as repeating units in both segmented subunits, and no evidence for major domain swapping between them was detected (Fig. 5).

Despite having different local environments, CP-A and CP-B as defined by segmentation have very similar envelopes. Superposition of CP-A and CP-B densities yielded a correlation value of 0.91 (see Materials and Methods), which strongly suggests they share a similar tertiary structure. Indeed, the most prominent tubular density features in each subunit (ascribed to α-helical elements) are nearly identically located and oriented. Each subunit has a length of maximally ~100 Å, and as viewed radially from inside or outside the capsid, the “crust” region is about twice as wide as the “toppings” region.

CP-A and CP-B as defined by segmentation do differ at their boundaries, however. One major difference occurs at the narrow (toppings) end of each subunit. Here, CP-A contains densities that line the icosahedral 5-fold channel (Fig. 5F, arrow) and contact an adjacent, 5-fold-related A subunit (e.g., A1 contacting A2 in Fig. 5B) but that are not apparent at the comparable locations in CP-B. Instead, features of similar volume and shape in CP-B are oriented toward a neighboring A subunit of the asymmetric A:B pair (e.g., B2 contacting A1 in Fig. 5B and dotted ellipse in Fig. 5F), such that they appear to be extended outward and flipped by 90° from their position in CP-A (Fig. 5G, red arrowhead). Two other notable differences occur near the wider (crust) end of each subunit. One occurs at the outer edge of the quasi-2-fold-symmetrical A:B interaction (e.g., for A1, B2 in Fig. 5B and black arrowhead in Fig. 5G), on the side distal to the icosahedral 2-fold axis, where a helix-like density from CP-A extends outward toward two neighboring B subunits (e.g., A1 contacting B2 and B3 in Fig. 5B). In contrast, this same density in CP-B does not protrude away from the subunit but instead extends in a radial direction toward the outside of the capsid. The other notable difference in this domain is a slightly shifted loop of density in CP-B near the icosahedral 3-fold axis, which contacts the other 3-fold-related CP-B subunits (e.g., B2 contacting B3 and B4 in Fig. 5B). This same region in CP-A, though more compact compared to that in CP-B, is the one that contacts another CP-A subunit directly across the icosahedral 2-fold axis (e.g., A1 contacting A4 in Fig. 5B).

Different responses of GLV and TVV1 virions to temperature and pH. The capacity of GLV to transmit between cells extracellularly (8, 21), whereas other known protozoal or fungal totiviruses cannot, led us to hypothesize that GLV virions may be more stable than many of these other viruses in order to withstand extracellular environments. This idea is consistent with the preceding observation that in our preparations of GLV virions (Fig. 2B and C) only a few particles appeared to have lost their dsRNA genomes after purification, in contrast to the greater number of such particles seen in preparations of TVV1 and ScV-L-A (35, 36, 39). To examine this stability difference more quantitatively, we adapted a previously described assay using RNase III to digest accessible viral genomic dsRNA (36) and used TVV1 virions for comparison with GLV virions. This assay allows for measurement of the amount of genome that becomes exposed from its protected location inside capsules upon preheating virions to different temperatures. The results showed that GLV virions were substantially more thermostable than TVV1 virions (Fig. 6A), a finding consistent with the greater stability of GLV.

Optimal growth of T. vaginalis trophozoites occurs at more acidic conditions (pH < 6.5) than does that of G. lamblia trophozoites (pH > 7) (54–56). Moreover, Tritrichomonas foetus, a relative of T. vaginalis, has been shown to possess a more acidic cytoplasm (pH 6.3) than do most eukaryotic cells, including G. lamblia (pH > 7) (55, 57). Hence, we hypothesized that the transcriptase activities of GLV and TVV1 may exhibit different pH optima, reflecting these different cytoplasmic pH values. By adapting in vitro transcriptase assays previously described for these viruses (58, 59), we found that GLV virions mediate transcription across a fairly broad pH range, peaking near pH 8.0, whereas TVV1 virions mediate transcription across a narrower range, peaking near pH 6.5 (Fig. 6B). These results are consistent with the different intracellular conditions in which these two viruses are commonly found in their two distinct hosts.

**DISCUSSION**

**Lack of projecting fibers in GLV.** GLV appears to be more closely related than other taxonomically approved totiviruses to a group of recently discovered, nonsegmented dsRNA viruses from meta
zoan hosts, including IMNV (Fig. 1). It is thus noteworthy that GLV has no projecting surface fibers, as are found at the 5-fold axes in IMNV (32) (Fig. 7). This absence of fibers is consistent with the lack of coding capacity for a fiber protein in the GLV genome, which IMNV and the related insect “totiviruses” contain upstream of their respective CPs in ORF1 (27–30). Coding for the proteolytically separated, 32-aa N-terminal peptide of the GLV CP precursor (see Fig. 3) is similarly positioned in the genome (i.e., upstream of CP in ORF1), but this peptide would be much too small to form a long fiber. It follows that in GLV, required functions for cell entry and tropism (e.g., receptor binding and membrane penetration) are likely to be mediated by as-yet-unlocalized regions of the capsid, rather than by any surface fibers as in IMNV.

In the tentatively assigned vertebrate (piscine) totivirus PMCV, a putative receptor-binding protein, containing a chemokine-superfamily motif, is encoded by a unique downstream ORF, designated ORF3 (31). This additional, recently discovered GLV-related virus (see Fig. 1) also transmits efficiently by extracellular means via virion-associated cell entry machinery. Since this property is not shared with other approved totiviruses of protozoal or fungal hosts (see Fig. 1), we suggest that GLV, IMNV and related insect viruses, and PMCV might be usefully classified in either a distinct taxonomic family or, more conservatively, in a distinct subfamily within the family Totiviridae.

**Sequence features with potential relevance to RNA interactions and cell entry.** Examination of the GLV CP precursor sequence (GenBank NC_003555) reveals a few, previously unnoted features of interest. Immediately following the 32-aa N-terminal region that is separated by proteolysis from the CP region (10) is a strongly polar, 71-aa region dominated by charged residues (see Fig. 1). This region seems reminiscent to us of the disordered, N-terminally positioned, small R (RNA-interacting) domain of many plus-strand RNA virus CPs (60), although the R domain in those viruses is rich in basic residues, whereas this part of the sequence in GLV is acidic (calculated pI = 4.6 for this region). We hypothesize that this strongly polar region projects into the central cavity enclosed by the GLV capsid, where its negatively charged residues may promote sliding movements of the negatively charged genomic RNA during transcription. The apparent CP contacts with RNA densities reflect where these R-domain-like regions of GLV CP are proposed to project into the interior (see Fig. 4B). These regions might also be involved in interacting with and positioning the RdRp domain of the CP/RdRp subunit(s), as has been suggested, for example, for a similar N-terminal domain in rotaviruses (61, 62).

Given the probable role of GLV CP in membrane penetration during cell entry by GLV virions into G. lamblia cells, it seemed informative to us to identify regions of CP sequence with hydrophobic character. Figure 3 highlights 14 regions of predominantly uncharged residues (regions ≈16 aa in length, including at most one charged residue), 12 of which are spread across the central portions of the CP sequence, suggestive of the folded core of this protein. In addition, there is one such region isolated near the C terminus and another near the N terminus. In fact, the N-terminal 32 aa cleaved from the CP precursor (MVWGTGYGNPSPLNPYGFAHLHRRGGGLNPLILA) include only two charged residues (underlined), as well as a high proportion of small residues (44% Gly, Ala, Pro, and Ser), reminiscent of the membrane-penetration peptides of birnaviruses, orthoreoviruses, and others (63–66). Since only 65% of the GLV CP sequence was covered by the tryptic or chymotryptic peptides identified by LC-MS/MS analysis in this study (see Fig. 3), the possibility remains that this N-terminal peptide, or smaller peptides derived from it, might still be present in the GLV virion.

**Conserved capsid elements in GLV and other Totiviridae members.** Despite very little primary-sequence homology among the CPs of GLV, ScV-L-A, HvV190S, TVV1, and IMNV (pairwise-identity scores all <18% using EMBoss Stretcher [http://www.ebi.ac.uk/Tools/psa/]), their capsid organizations are highly similar (Fig. 7), other than the projecting fibers in IMNV. Although the GLV capsid is somewhat larger than the others (17.5% larger than ScV-L-A; ~5–7% larger than HvV190S, TVV1, and IMNV), comparable decameric plateaus are seen on the outer surfaces of

---

**FIG 7** Comparison of GLV to four other taxonomically approved or tentatively assigned totiviruses. All particles are viewed along an icosahedral 2-fold axis, and the surfaces are color coded by radius according to the spectrum shown at right, which lists radii in angstroms. IMNV is tentatively assigned to family Totiviridae.
each, although HvV190S is notably the smoothest. The larger capsid of GLV correlates with it having a larger CP than the other approved Totiviridae members shown here, although the IMNV CP is the largest (IMNV, 99 kDa > GLV, 98 kDa > HvV190S, 81 kDa > ScV-L-A, 76 kDa > TVV1, 75 kDa). The genome size of IMNV is also larger than that of GLV and the other totiviruses (IMNV, 7,561 bp > GLV, 6,277 bp > HvV190S, 5,179 bp > TVV1, 4,647 bp > ScV-L-A, 4,579 bp). Using the inner capsid diameter to estimate the central cavity volume from the density maps of these viruses, however, it turns out that GLV appears to have more space available for its genome than does IMNV (GLV 20,035 to 21,497 nm$^3$ > IMNV 16,362 to 19,858 nm$^3$; since the viral capsids are not perfectly round, the range reflects use of either the smallest or the largest inner diameter of the capsid for calculation). Assuming a single genome molecule per virion, this would result in a lower genome packing density for GLV. By being less densely packed, the GLV genome might be relatively too disorganized to give rise to concentric spherical shells of dsRNA densities in the icosahedral reconstruction as seen in IMNV and other totiviruses (32, 36, 38, 39).

Besides overall capsid organization, the individual GLV A and B subunits show a striking resemblance to those of ScV-L-A, HvV190S, and TVV1 (Fig. 8A). In addition, the orientation of GLV CP-A and CP-B with respect to each other is almost identical to the orientation of these subunits in these three viruses. The resolution of the GLV cryo-reconstruction is too low to permit a reliable trace of the entire CP polypeptide backbone, but it is clear that even at a 6.0-Å resolution the GLV CP is rich in twisted, tubular elements likely to represent α-helices, as are also the CPs of ScV-L-A, HvV190S, and TVV1 (Fig. 8B). Of particular note are four putative, interacting long α-helices (colored blue in Fig. 8B) that are similarly located and oriented in the CP subunits of all four viruses (best seen in the 90° and 180° side views in Fig. 8B). Thus, despite some differences in length, orientation, and angular position in the different viruses, these putative α-helices appear to constitute a conserved core of the CP subunits that might be common to all totiviruses (38), including GLV.

**Capsid channels and relative stability of GLV and TVV1.** Following entry into the *G. lamblia* cytoplasm, the particle-enclosed GLV genome is presumably transcribed by the RdRp domain of...
the CP/RdRp subunit(s) still inside the confines of the “\(T=2\)" icosahedral capsid, and the plus-strand transcripts are then released from the particle (e.g., see evidence for this \textit{in vitro} activity in Fig. 6B), allowing translation of the GLV proteins and packaging into nascent, progeny GLV virions. This strategy for transcription inside the icosahedral capsid is thought to be common to all dsRNA viruses except possibly birnaviruses (67), which lack the “\(T=2\)" capsid organization (68). The only solvent-accessible channels that span the GLV capsid appear to occur at each of the 12 5-fold axes, and though somewhat narrow (ca. 10 to 11 Å in width), they are probably wide enough to permit both entry of nucleoside triphosphates and exit of newly synthesized transcripts during transcription. It is also conceivable, and perhaps likely, that conformational changes in the capsid during cell entry and/or in association with transcription may enlarge these channels or perhaps open other ones.

The 5-fold channels in GLV are considerably smaller than those in TVV1, which narrow to only ~16 Å in the lower portions of the capsid shell (39). Although these 5-fold channels, absent conformational changes, are not wide enough in either virus for genomic dsRNA to pass through, we have found that purified preparations of TVV1 virions contain a substantial fraction of “empty" particles, i.e., particles devoid or depleted of genome, with accompanying evidence for dsRNA release following purification (39). Indeed, we have suggested that the relatively wide 5-fold channels in TVV1 may serve as the conduit for this release of genome upon limited conformational widening of these channels. This may in fact have significant biological relevance as the source of dsRNA for TLR3/TRIF-dependent signaling of proinflammatory responses by mammalian cells that encounter TVV1-positive \textit{Trichomonas vaginalis} parasites in the human genitourinary tract, thereby contributing to human disease (18, 19).

Regarding a similar possible effect of GLV virions in contributing to human enteric disease by \textit{G. lamblia}, we note that our purified preparations of GLV virions contain a much lower fraction of empty particles or evidence for dsRNA release (see Fig. 2B and C), a finding consistent with our experimental evidence that GLV virions are substantially more thermostable than TVV1 virions (Fig. 6A). This greater stability of GLV virions is furthermore consistent with its capacity for transmission between cells by extracellular means, entailing a greater need for it to withstand a wider range of environments but potentially reducing its propensity for proinflammatory signaling as observed for TVV1 (18, 19).

**Possible pathways of GLV/totivirus capsid assembly.** Assembly pathways for the capsids of icosahedral viruses often include symmetrical dimers or trimers as initial intermediates (69). Assembly of a “\(T=2\)" capsid like that of GLV, however, could not simply involve CP trimers, because only half of the subunits (60 CP-B) are arranged as 20 trimers in the capsid; the other half (60 CP-A) are arranged as 12 pentamers. Although not impossible, the formation of two such notably different intermediates by the chemically identical CP subunits seems unlikely.

“\(T=2\)" assembly pathways potentially involve compact decamers as intermediates (69), 12 of which would coalesce to form the complete capsid. Such a pathway involving compact decamers would rule out the participation of either symmetrical trimers or symmetrical dimers as intermediates and would instead rely on the formation of asymmetrical CP dimers (one A and one B subunit arranged in a near-parallel, side-by-side manner). Although this asymmetrical dimer followed by decamer assembly is certainly plausible, inspections of the GLV and other totivirus capsid structures suggest that their “\(T=2\)" capsid assembly pathway may instead involve 2-fold-symmetrical, extended CP dimers, two of which could associate to form compact, 2-fold-symmetrical CP tetramers. A related scenario is that these extended CP dimers might be transient and reversible, such that the tetramer represents the first stable intermediate. In any case, 30 such tetramers could then coalesce to form the complete capsid, possibly by first assembling into trimers of tetramers (69), 10 of which would then combine. Yet another, conceivable “\(T=2\)" assembly pathway involving a stable tetramer intermediate would have extended CP dimers continuing to accrete to an initial, nucleating tetramer.

Interestingly, studies of \textit{Pseudomonas} bacteriophages \(\phi 6\) and \(\phi 8\) (family \textit{Cystoviridae}) (70, 71), three fungal partitiviruses (family \textit{Partitiviridae}) (72, 73), and rabbit picobirnavirus (family \textit{Pico- birnaviridae}) (65) have all identified 2-fold-symmetrical CP tetramers as the most probable assembly intermediates. Possibly, some “\(T=2\)" capsids assemble via a decamer pathway (e.g., family \textit{Reoviridae}), whereas others assemble via a tetramer pathway (e.g., families \textit{Cystoviridae}, \textit{Partitiviridae}, and \textit{Pico- birnaviridae}) (69). In the case of family \textit{Totiviridae}, the greater CP mass and more elongated CP structure of its members (Fig. 8) might suggest these viruses are more like members of family \textit{Reoviridae} and assemble via the decamer pathway, but we consider the tetramer pathway to remain a viable alternative. In fact, the substantial interactions seen between CP-A and CP-B within each quasi-2-fold-symmetrical dimer in the capsid (segmentation was most difficult to perform at this subunit interface) provide direct evidence in favor of the dimer/tetramer assembly pathway.

Beyond these considerations strictly about CP subunit assembly, other key questions remain about how the CP/RdRp fusion subunit(s) and the genomic RNA assemble into mature virions. For example, regardless of the precise manner by which the GLV/totivirus CP subunits form assembly intermediates, do the one or two CP/RdRp molecules per virion always occupy either an A or a B position, or can the CP/RdRp molecule(s) assemble and function in either of these positions? Also, is the viral RNA recruited by a particular intermediate, presumably one containing a CP/RdRp subunit? These and other challenging questions need to be addressed in future studies.

**ACKNOWLEDGMENTS**

We are grateful to C. C. Wang and coworkers (University of California, San Francisco) for providing a stock of \textit{G. lamblia} WBI cells and to Raina N. Fichorova and coworkers (Brigham and Women’s Hospital and Harvard Medical School) for providing an original stock of \textit{T. vaginalis} UH9 cells. We also thank Norman H. Olson for technical assistance with microscopy, Maria Ericsson for negative-stain microscopy of GLV virions that had been stored at different conditions, and John M. Asara and coworkers (Beth Israel Deaconess Medical Center and Harvard Medical School) for LC-MS/MS analyses of GLV virions.

This study was supported in part by NIH grants R37 GM033050 (to T.S.B.) and R56 AI091889 (subcontract to M.L.N.). Support from the University of California San Diego and the Agouron Foundation (all to T.S.B.) was used to establish the cryo-TEM facilities used in this study.

**REFERENCES**


2. Upcroft P, Upcroft JA. 2001. Drug targets and mechanisms of resistance...


