Crystal structure of the Epstein-Barr virus (EBV) glycoprotein H/glycoprotein L (gH/gL) complex

Hisae Matsuraa,b, Austin N. Kirschnerc, Richard Longnecker, and Theodore S. Jardetzky

The Epstein-Barr virus (EBV) is a γ-herpesvirus that infects B cells and epithelial cells and that has been linked to malignancies in both cell types in vivo. EBV, like other herpesviruses, has three glycoproteins, glycoprotein B (gB), gH, and gL, that form the core membrane fusion machinery mediating viral penetration into the cell. The gH and gL proteins associate to form a heterodimeric complex, which is necessary for efficient membrane fusion and also implicated in direct binding to epithelial cell receptors required for viral entry. To gain insight into the mechanistic role of gH/gL, we determined the crystal structure of the EBV gH/gL complex. The structure is comprised of four domains organized along the longest axis of the molecule. Comparisons with homologous HSV-2 gH/gL and partial pseudorabies virus gH structures support the domain boundaries determined for the EBV gH/gL structure and illustrate significant differences in interdomain packing angles. The gL subunit and N-terminal residues of gH form a globular domain at one end of the structure, implicated in interactions with gB and activation of membrane fusion. The C-terminal domain of gH, proximal to the viral membrane, is also implicated in membrane fusion. The gH/gL structure locates an integrin binding motif, implicated in epithelial cell entry, on a prominent loop in the central region of the structure. Multiple regions of gH/gL, including its two extreme ends, are functionally important, consistent with the multiple roles of gH/gL in EBV entry.

The Epstein-Barr virus (EBV) is a double-stranded DNA virus belonging to the Herpesviridae family, which is divided into three subfamilies (α, β, and γ) (1–3). EBV and Kaposis Sarcoma herpesvirus (KSHV or HHV-8) are the two human host-specific viruses forming the γ-herpesviridae subfamily (2). EBV infects both B lymphocytes and epithelial cells and causes infectious mononucleosis. Nearly 95% of the population is infected by EBV by adulthood and carries EBV DNA throughout life. EBV is maintained in a latent state in infected B lymphocytes, with periodic reactivation of lytic replication. Numerous malignancies are associated with EBV, such as Burkitt's lymphoma, Hodgkin's lymphoma, and nasopharyngeal carcinoma (1, 3). In healthy individuals, the pathological effects of the virus are controlled by the immune system. However, in immunosuppressed individuals, such as transplant recipients and AIDS patients, EBV can cause tumor outgrowth and trigger fatal lymphoproliferative disease.

EBV uses different pathways for the infection of epithelial cells and B lymphocytes (4, 5). For both cell types, the minimal viral glycoprotein components that mediate membrane fusion have been identified (4, 5). As with other herpesviruses, EBV uses the core viral entry glycoproteins, glycoprotein B (gB) and the gH/gL complex. For the infection of B lymphocytes, EBV requires an additional protein, gp42, which binds to host HLA class II molecules, triggering the membrane fusion step (4, 5). gp42 has multiple functional sites for interaction with gH/gL, HLA class II, and potentially, another unknown binding ligand that could be engaged through a large surface-exposed hydrophobic pocket (6–12). The gp42 protein binds to the gH/gL complex with nanomolar affinity through its N-terminal region, and this interaction can be recapitulated with a synthetic peptide of ~35 aa residues (10, 13, 14). EBV glycoprotein-mediated membrane fusion with epithelial cells does not require gp42 but only gB and gH/gL, and fusion can be completely blocked by saturating amounts of either gp42 or short gp42-derived peptides (13–16), consistent with the hypothesis that gp42 levels in the virion regulate the cellular tropism of the virus in vivo (16). Recent observations indicate that EBV gH/gL engages integrins αvβ6 and/or αvβ8 on epithelial cells to trigger membrane fusion and entry (17).

Among the core glycoproteins for EBV-induced membrane fusion, the crystal structures of gB (18), the gp42:HLA complex (6), and gp42 alone (10) have been determined. The EBV gB protein belongs to the recently identified class III viral fusion glycoproteins (19), which includes the herpesvirus gB proteins (18, 20), the Vesicular Stomatitis Virus G (VSV G) protein (21, 22), and the baculovirus gp64 protein (23). Both VSV G and baculovirus gp64 act alone as the fusogenic proteins for virus entry and both are activated by low pH changes during endocytosis of the virus (19, 21–23). The VSV G fusion protein undergoes reversible conformational changes that are pH-dependent in contrast to the irreversible transitions characterized for both class I and class II viral fusion proteins (24, 25). The VSV G protein structure has been solved in two conformational states, interpreted as representing the pre- and postfusion forms. The EBV and HSV-1 gB protein structures are most similar to the postfusion conformations of VSV G (19). A prefusion model of the EBV gB structure has been proposed (18), but there is no experimental evidence to date for this conformational transition in the herpesvirus glycoproteins. In contrast to the VSV and baculovirus fusion proteins, which act independently, herpesvirus gB requires gH/gL for the most efficient membrane fusion (15, 26, 27).

To better understand the role of gH/gL in virus entry and membrane fusion, we determined the crystal structure of the EBV gH/gL heterodimer. EBV gH and gL form a multidomain structure similar to the recently determined HSV-2 gH/gL (28) and partial pseudorabies virus (PRV) gH structures. The structure consists of four domains, forming a flat, elongated shape, with individual domains sequentially folded along the length of the protein. Each of the four domains adopts different tertiary folds and secondary structures, and the N-terminal domain seems more disordered in the postfusion form. The EBV and HSV-1 gB protein structures are most similar to the postfusion conformations of VSV G (19).

Edited* by Robert A. Lamb, Northwestern University, Evanston, IL, and approved November 12, 2010 (received for review August 9, 2010)


The authors declare no conflict of interest.

This Direct Submission article had a prearranged editor.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 3PHF).

1To whom correspondence should be addressed. E-mail: tjardetz@stanford.edu.

This article contains supporting information online at www.pnas.org/cgi/suppl/doi:10.1073/pnas.1011806108/DCSupplemental.
for the first and second domains of the EBV and HSV-2 structures. Analysis of the gH/gL complex structure, based on previous mutational and functional experiments, points to specific regions in multiple domains that are critical for this protein complex to act as part of the core herpesvirus fusion machinery.

Results and Discussion

**Determination of the EBV gH/gL Protein Structure.** The ectodomains of EBV gH and gL were coexpressed in insect cells. The two proteins form a stable complex and were purified by affinity chromatography with an antibody (E1D1) that recognizes an epitope dependent on both gH and gL (13, 29). The complex elutes in gel filtration chromatography with the expected molecular weight of the heterodimer (13).

EBV gH/gL crystallized in the P21 space group and a native dataset to 3.58-Å resolution was collected (Table S1). The structure was determined by heavy-atom methods, and the asymmetric unit was expected to contain 10–20 copies of gH/gL. Initial phases were obtained using a TaqBr12 cluster and further improved with a second (KBrCl) derivative (Table S1). A low-resolution selenium derivative dataset was used to validate the chain tracing and model building but not in initial phasing. Sixteen copies of gH/gL were identified in the asymmetric unit, and phases were extended to 3.58 Å using nonsymmetry averaging. Side-chain electron density was significantly improved using B-factor sharpened electron density maps (Fig. S1). The final Rfree is 31.3%, and the refinement and model statistics are collected in Table S1. The structure includes the majority of the gH and gL residues included in the expression construct.

**Overall Description of the Structure of EBV gH/gL.** The EBV gH/gL structure reveals an elongated rod-like shape ~100 Å in length and 30–60 Å in width, with the middle of the molecule being the widest (Fig. 1). The overall shape agrees with the rod-like shapes observed in negatively stained EM images (Fig. S2), gH/gL is divided into four major domains laid out along the length of the molecule, whose folds involve only immediately adjacent polypeptide segments. There are no domain insertions or long interdomain excursions in the structure.

The gL protein forms intimate contacts with the N-terminal residues of gH, which together define the first domain (D-I) of the heterodimer structure (Fig. 1B). The central two domains of gH (D-II and D-III) create a core region of predominantly helical structure, surrounded on either side by β-sheet. The exclusive involvement of gL with the N-terminal residues of gH in D-I is consistent with previous mapping experiments (30) and the HSV-2 gH/gL structure (28).

There are 11 cysteine residues in gH and 4 in gL. Five disulfide bonds were observed within the gH/gL structure between residues 120 and 312, 278 and 335, 454 and 478, 534 and 587, and 612 and 615. C153 does not form a disulfide bond with any other cysteine residues and bound Hg in heavy atom-soaked crystals. gL contains two disulfide bonds between residues 28 and 56 and 29 and 79. Extra density consistent with N-linked glycosylation was observed in negatively stained EM images (Fig. S2), gH/gL is divided into four major domains laid out along the length of the molecule, whose folds involve only immediately adjacent polypeptide segments. There are no domain insertions or long interdomain excursions in the structure.

**N-Terminal Residues of gH and gL Form D-I.** The gL subunit is important for gH folding (28, 30, 31) and is critical in the specificity of EBV gB activation (32). In EBV gH/gL (residues 24–131) plays a key structural role in D-I (Fig. 1), constructing an α/β-domain that is intimately intertwined with the N-terminal residues of gH (20–65). D-I is a flat, semiglobular domain formed by a layer of five-stranded mixed parallel/antiparallel β-sheets supported by a layer of three α-helices (Fig. S3A). The secondary structure and hydrophobic core of D-I seems dependent on both gH and gL residues and is similar at its core to the HSV-2 gH/gL structure (28). The five-stranded β-sheet contains three strands from gL and two strands from gH, whereas the underlying three-helix layer is formed by helixes from gL and one from gH. A long protrusion from the surface of D-I is formed by strands Lβ-2 and Lβ-3 (Fig. 1B and Fig. S3A). Two neighboring cysteine residues in gL (residues 28 and 29) form disulfide bonds with residues 56 and 79, located in the loop between strands Lβ-1 and Lβ-2 and in Lβ-3, respectively. The N terminus of gL forms a long loop leading into the central Lβ-1 strand beginning at residue 47. In HSV-2 gL, the N-terminal residues form additional structure on the face of the D-I β-sheet (28).

The three-helix layer separates D-I from D-II, forming a wall that includes a number of charged and polar residues (Fig. 1B). The polypeptide link between D-I and the rest of gH is mediated by a single α-helix (residues 66–74) at the base of the interdomain interface (Fig. 1B), which might allow repositioning of the
the two domains, with the helix acting as an internal hinge or lever. Superposition of the D-I domain from the EBV and HSV-2 structures reveals a common core tertiary structure, with the HSV-2 domain exhibiting significant variations at the N terminus of gL and in extended loop structures throughout the domain.

gH Residues 66–672 Fold into Three Sequential Globular Domains. The remainder of gH folds into three additional and potentially, semiautonomous domains with extensive interdomain interfaces, similar to the structures of the HSV-2 gH/gL (28) and PRV gH fragment.

D-II starts with the C-terminal end of the D-I/D-II linker α-helix (Fig. S3B), followed by a long extended chain that wraps around the outside surface of D-II. After the linker from D-I, residues 79–97 climb ∼60 Å diagonally across D-II to the opposite corner, forming flaps that cover the core helical bundle of the domain. Electron density of this segment is more disordered, and it is likely less constrained and flexible. D-II can be split into two major secondary structure sections (Fig. 1 and Fig. S3B), with a prominent eight-stranded antiparallel β-sheet (β-sheet II), followed by an antiparallel five-helix bundle (helical bundle II). β-sheet II forms a picket fence separating D-I from helical bundle II (Fig. 1B). The β-sheet II face facing D-I is covered by two α-helical polypeptide loops (residues 133–154 and 201–217) that include a short helix (2α-3) at the tip of the first loop (residues 143–148), forming the interface with D-I (Fig. 1B).

The three main α-helices of helical bundle II lie parallel to the long axes of β-sheet II, forming the central and widest region of the entire gH/gL molecule. Residues 245–336 form the majority of this helical region that is mainly helix–loop–helix structure. The core three helices consist of residues 258–274, 285–307, and 315–335 (2α-6–2α-8), and two shorter α-helices from residues 123–130 and 246–254 (2α-2 and 2α-5) add to the outside of this three-helix core, with five helices total. A final helix forms the transition between D-II and D-III.

D-III (Fig. S3C), starting from S345 and extending to L529, is a mainly α-helical structure, with its helices lying nearly perpendicular to the length of helical bundle II. The domain begins with five helices arranged in four layers of a helix–loop–helix structure that spirals down the side of D-II, involving residues 345–454 (Fig. 1B and Fig. S3C). The final four helices form a distinct subdomain bundle, making a total of nine helices in D-III. There is one disulfide bond near the center of D-III (C454–C478).

D-IV (Fig. S3D), beginning at P530 and ending at G672, forms a β-sandwich domain, with two antiparallel β-sheets (β-sheet IIIA and IIIB). Both β-sheets consist of four core antiparallel β-strands (4β-1–4β-4 and 4β-7–4β-10), which are linked by a single cross-over loop (residues 584–617). The β-sandwich is aligned with one edge abutting D-III and the axis of the β-strands aligned with the helix and sheet axis of D-II (Fig. 1B). Extensive loops connect the β-strands within each sheet, and the two disulfide bonds within this domain are located on the exterior chains decorating the core β-sandwich.

Comparison of the EBV, HSV-2, and PRV Structures. The recent structure determinations of HSV-2 gH/gL (28) and a three-domain PRV gH fragment allow a structural comparison across the α- and γ-herpesviruses (33). The core secondary structures of the domains, as defined here for EBV gH/gL, are very similar in the three structures. However, the interdomain packing arrangements differ significantly, particularly in comparisons of the EBV and HSV-2 structures (Fig. 2). The PRV structure lacks D-I but exhibits slightly different packing at both the D-II/D-III and D-III/D-IV domain interfaces. However, much larger packing angle differences are observed for the HSV-2 D-I/D-II and D-II/D-III interfaces, of ∼90° and ∼45°, respectively. Thus, whereas the EBV structure seems linear and elongated, the HSV-2 structure adopts a boat-like (28) configuration (Fig. 2).

The difference in D-I/D-II packing creates a prominent groove in the EBV gH/gL structure that is not present in HSV-2 gH/gL.

Mutations in D-I Influence Membrane Fusion Activity and gB Specificity. Mutations that affect EBV gH/gL activity in membrane fusion map to D-I and the D-I/D-II interface (Fig. 34) (32, 33). Mutants in leucine residues in gH were generated in the N-terminal region spanning residues 54–74, which are found in D-I or the helical junction to D-II (34). Mutations of L65A and L69A showed reduction in membrane fusion with both B cells and epithelial cells, whereas mutants L55A and L74A enhanced viral fusion. Residues 65 and 69 are both located in the N-terminal portion of the D-I/D-II linker helix and are buried in a hydrophobic interface with gL (Fig. 3A and B). L55 is located at the interface of D-I/D-II and may destabilize the interdomain interface, whereas L74 is exposed on the surface of the D-I/D-II linker helix at the beginning of D-II (Fig. 3A and C). The deleterious mutations of L65A and L69A may disrupt interactions with gL and the proper folding of D-I. The monoclonal antibody E1D1 recognizes an epitope generated by both gH and gL, and its binding to the L65A mutant is diminished, consistent with an al-
Fig. 3. D-I residues implicated in membrane fusion activity of gH/gL. (A) Residues implicated in gH/gL function are shown as semitransparent CPK spheres colored by chain and domain as in Fig. 1. The backbone trace is represented as coils, and the residues are labeled as described in Fig. 1. (B) Mutations of EBV gH L56A and C54A, referred to as H:56 and H:69, respectively, reduce gHgL membrane fusion activity with both B cells and epithelial cells. H:65 and H:69 are shown in yellow sticks, with surrounding residues shown as lines colored by the gH/gL chain as in A and labeled. Both residues form hydrophobic contacts with gL at the edge of the D-I hydrophobic core. (C) gH mutants L55A and L74A enhance viral fusion activity. H:55 is located at the D-I/D-II interface, whereas H:75 is exposed on the surface of the D-I/D-II linker helix at the beginning of D-II. Residues are colored as in B (D) EBV gL Q54 and K94 (yellow sticks and boxes) control the specificity of gB activation in membrane fusion. Mutation to Rhesus lym- phocryptovirus (LCV) gL (82% identity with EBV gL) residues K54 and Q94 reduces membrane fusion with EBV gB. L:Q54 is exposed at the surface of D-I in the loop between the first and second a2-strands (Lp1-1 and Lp1-2). Adjacent residues (L:W24, H:129, H:E30, and H:H32) are also surface-exposed and form a prominent ridge adjacent to Q54. K94 is located at the D-I/D-II interface in an area rich in polar and charged residues (as described in the text). L:Q54 and L:K94 are shown as yellow sticks, with identified neighboring residues shown as sticks colored by their associated chain and domain.

D-I/D-II interactions and orientation, possibly leading to a en- hanced fusogenic activity. Because K94 is located in a groove at the top of the D-I/D-II interface, it is also possible that this could form part of a binding surface that could engage gB (Fig. 3A).

Residues in D-II and D-IV Are Implicated in Membrane Fusion. Other regions of gH are also implicated in membrane fusion activity (Fig. 4A). The C-terminal end of gH has been identified as a critical functional region (35, 36). The anti-gH antibody CL59 blocks membrane fusion with epithelial cells but not B cells, and its binding epitope has been mapped to gH residues 501–628 (35, 36), forming the C-terminal end of D-III and the majority of D-IV. Insertion and point mutant studies have also implicated specific C-terminal residues in regulating differential gH fusion activity with B cells and epithelial cells (35, 36).

The mutation of G594 to alanine abolishes fusion with both B cells and epithelial cells, whereas the mutation of E595A has only moderate effects on epithelial cell fusion but dramatically increases fusion activity with B cells (35, 36). G594 and E595 are located on the cross-over loop in D-IV and exposed at the extreme C-terminal end of the gH structure (Fig. 4B). Additional mutations of V592A, R597A, and R607A have moderate effects on gH activity (35, 36), decreasing fusion with B cells ~40%. These three residues are also located in the cross-over loop in D- IV (Fig. 4B), near the gH C terminus and the expected position of the transmembrane domain and viral membrane.

Recently, the interaction of EBV gHgL with αβδα and αβδβ integrins has been implicated in triggering viral fusion with epithelial cells (17). In the structure, the identified KGD motif is located in a prominent D-II loop between strands 2β-6 and 2β-7 (Fig. 4A and C). The motif is near the D-I/D-II linker α-helix (2α-1) and the D-I/D-II interface, where mutations in both gH and gL affect the fusogenic activity of gHgL. Nearby mutations in gH residues 65 and 69 disrupt E1D1 antibody binding. E1D1 inhibits membrane fusion with epithelial cells (but not B cells) (36) and blocks gH binding to integrin (37), consistent with the possibility that E1D1 and integrins may bind overlapping surfaces in this functional region between D-I and D-II.

Final Remarks. The structure of the EBV gHgL complex adds to the previous structures of the EBV gB and gp42 proteins (6, 10, 18), which together define the essential membrane fusion machinery of this γ-herpesvirus. The EBV gHgL protein has a four- domain architecture arranged linearly along the length of the polypeptide chain, with the N-terminal D-I being formed by both gH and gL subunits, similar to the structures of HSV-2 gHgL (28) and a PRV gH fragment (33). Functional regions in EBV gHgL have been mapped to at least four areas, including D-I, the interface between D-I and D-II, D-IV, and a prominence displayed KGD sequence in D-II implicated in integrin binding and epithelial cell entry (17). The integrin binding loop (residues 188–190) lies next to the functionally important L74 (34). L74 forms part of a ridge defining a hydrophobic groove adjacent to the integrin binding site and surrounded by a number of charged residues (Fig. 4C), potentially defining an extended surface region important to gHgL function in EBV entry. gL residues that regulate the specificity of gB activation (32) have been mapped to the surface of D-I (Q54) and to the D-I/D-II interface (K94). Q54 is surface-exposed and surrounded by other prominentaly displayed D-I residues, which together could form a gB binding site. The binding region of gp42 for gHgL has been mapped to a 33-aa peptide (14). We have not yet identified the corresponding binding site on gHgL, but the prominent groove be-
between D-I and D-II and the surface between the D-I/D-II linker helix and the integrin binding loop stand out as potential peptide binding sites. These are also shown in stick representation, labeled, and colored with carbon atoms in green.

![Diagram](image_url)

**Fig. 4.** Residues in gH D-II to D-IV implicated in EBV entry. (A) Residues in EBV gH D-II and D-IV that are implicated in EBV entry are shown on the structure as CPK spheres, colored by domain as in Fig. 1. The D-IV residues are located at the extreme C-terminal end of the molecule in a cross-over loop between the two halves of the domain. The D-II residues, defining an integrin binding motif implicated in epithelial cell entry, are located in D-II, near the D-I/D-II linker helix. (B) gH residues G594 and E595 are located in the D-IV cross-over loop connecting the two β-sheets of the domain fold. The view shown is oriented ∼90° about the vertical axis from the view in A (bold arrow). The residues are shown as dark gray sticks with O and N atoms colored red and blue, respectively. gH residues 592, 597, and 607, which have moderate effects on gH/gL fusion activity, are also exposed on the D-IV cross-over loop. These residues are shown as orange sticks, with O and N atoms colored red and blue. (C) The gH/gL integrin binding site is adjacent to an extended hydrophobic groove including L74. The view shown is oriented ∼90° about the horizontal axis from the view in A. The gH/gL chain is shown as a coil beneath a transparent protein surface. The KGD motif is shown as sticks with carbon atoms colored yellow and mapped onto the gH/gL surface. The nearby residue L74, located in the D-I/D-II linker helix, is surrounded by gH residues that form a primarily hydrophobic groove lined with charged residues. These are also shown in stick representation, labeled, and colored with carbon atoms in green.

**Materials and Methods**

Production and Purification of gH/gL. The expression and purification of soluble EBV gH/gL has been previously described (13). In the expression constructs, gH residues 18–679 and gL residues 24–137 were fused to the baculovirus gp64 signal sequence, resulting in the addition of three N-terminal residues (AMT) into gH and three (AMD) into gL. For protein production, Sf+ insect cells were grown to a density of 1.5 million cells/mL and infected with recombinant gH/gL baculovirus stock. Cell supernatants were harvested 72 h postinfection, and gH/gL was isolated using an anti-gH/gL mouse monoclonal antibody (E1D1) column. The E1D1 hybridoma was kindly provided by L. Hutt-Fletcher (Louisiana State University Health Sciences Center, Shreveport, LA). Typically, 4 L of infected cell supernatant were loaded onto the E1D1 affinity column and washed with PBS. The protein was eluted with 0.1 M sodium citrate, pH 2.5, and immediately neutralized by the addition of 1 M sodium citrate Tris(vinyl) phosphate (pH ~8) and 5 M NaCl, yielding a final pH of 5.5. The eluted protein solutions were combined, concentrated, and exchanged into 50 mM sodium citrate, pH 5.5, and 50 mM NaCl buffer. The protein was concentrated, filtered, and further purified using a 5200 gel filtration column in 50 mM sodium citrate, pH 5.5, and 50 mM NaCl buffer.

**gH/gL Crystallization.** The gH/gL protein was concentrated to ~10 mg/mL using Vivaspin 2 and Vivaspin 500 (Sartorius) concentrators, and the final concentration was confirmed using a Nanodrop spectrophotometer with an extinction coefficient of 87,515 (M⁻¹ cm⁻¹) at 280 nm. gH/gL crystals were grown in sitting drop format (Cryschem Plate HM-158; Hampton Research), initially using a well solution consisting of 24% PEG1000 and 100 mM sodium citrate, pH 5.5, and drops containing 24% PEG1000, 100 mM sodium citrate, pH 5.5, 200–400 mM NDSB256-4T, and protein at room temperature. The crystallization was further optimized by seeding. For seeded crystal growth, the well solution contained 18–19% PEG1000, 100 mM sodium citrate, pH 5.5, and 200 mM NDSB251, and the drop consisted of a mixture of PEG1000, citrate buffer, NDSB256-4T, and protein. After >1 h of equilibration, 0.1 μL gH/gL crystallization solution was added to the gH/gL crystal seed stock, generated using Seed Beads (Hampton Research) and incubated at 23°C. In the optimized conditions, gH/gL crystals typically appeared after 2 d and grew up to a size of 50–100 μm × 200–300 μm, and occasionally larger, after 7 d. The gH/gL crystals were harvested in 2–3% higher PEG1000 con-
concentrations and transferred in gradual steps of increasing PEG1000 concentration to a final concentration of 30% before freezing in liquid nitrogen.

**Heavy-Atom Derivatization and Data Collection.** Initial phases were obtained from datasets with significant anomalous signals from crystals soaked in 2 mM Ta6Br12 for 4 h and crystals soaked in 1 mM K2IrCl6 for 4 d. The heavy-atom-saturated crystals were transferred into freezing buffer conditions, maintaining constant concentrations of heavy atom before freezing. Selenium-methionine-substituted protein was expressed in HEK cells by replacing the medium with methionine-free Ex-Cell 405 and by adding selenomethionine at the final concentration of 60.6 mM at the time of infection.

The native dataset was collected at the Advanced Light Source beamline 8.3.1 and processed to 3.5 Å using Mosfi and SCALA (41). The Ta6Br12 single-wavelength anomalous dispersion (SAD) data were collected at 24ID-C beamline operated by the Northeastern Collaborative Access Team at the Advanced Photon Source at the Ta peak wavelength of 1.2543 Å and processed with HKL2000 (42) to 4.6 Å. The K2IrCl6 dataset was collected at beamline BL9-2 at the Stanford Synchrotron Radiation Laboratory at 1.10519 Å and processed with HKL2000 to 5.5 Å. Additional heavy-atom datasets are collected in Table S1.

**Structure Determination and Refinement.** The Native, Ta6Br12, and K2IrCl6 datasets were combined and scaled together (41), and SHARP (43) was used for heavy-atom refinement and phasing. Initial heavy atom-based phases were generated at low resolution (6 Å) and used to identify noncrystallographic symmetry (NCS) relationships and helical features of the gHgL structure. Electron density maps were improved by a combination of density modification, NCS averaging, and heavy-atom refinement. Final model of 40 Ta6Br12 clusters, 34 K2IrCl6, 38 KReO4, 7 YbCl3, 14 (NH4)2WS4, 24 AuCN2, 27 SeMet sites, and 16 NCS-related gHgL molecules were identified in the crystallographic asymmetric unit. The gHgL model was built and refined using a combination of automated and manual approaches. Further details of the structure determination are included in **SI Materials and Methods**, and the final model statistics are presented in Table S1. The final model has an R-free of 31.3%, overall good geometry, Ramachandran statistics, and no unfavorable side-chain rotamers.

**ACKNOWLEDGMENTS.** We thank Lindsey Hurt-Fletcher (Louisiana State University) and Kishan Aravind (Centers, Shreveport, LA) for the E1D1 monoclonal antibody and members of the T.S.J. and R.L. laboratories. Portions of this research were conducted at the Advanced Photon Source, the Stanford Synchrotron Radiation Laboratory, and the Advanced Light Source (**SI Materials and Methods**). This work was supported by Public Health Service Grants AI076183 and CA117794 from the National Cancer Institute (to R.L.I.).
Supporting Information

Matsuura et al. 10.1073/pnas.1011806108

SI Materials and Methods

Structure Determination and Refinement. Five initial heavy-atom sites were found from the TaBr12 dataset using SOLVE (1). The Native, TaBr12, and KtIrCl6 datasets were combined and scaled together (2), and SHARP (3) was used for heavy-atom refinement and phasing. An expanded set of 9 Ta clusters and 37 Ir sites were found and refined in SHARP. The handedness of the heavy-atom sites were determined by comparing the overall correlations on [E]2 in the initial cycle of SIGMAA (2) at different solvent content percentages. From these initial heavy-atom sites, four possible noncrystallographic symmetry (NCS) relationships were identified. The initial map from SHARP was averaged in DM (2) using the four NCS operators, and the improved electron density map revealed low-resolution secondary structure features.

COOT (4) was used to manually fit α-helices into the observed electron density, and independent repeating helical clusters were identified. One of the modeled clusters, consisting of seven α-helices, together with the initial heavy-atom phase information from SHARP was used to search for additional NCS-related molecules using the Phased translation function implemented in the BRUTEPTF server (http://zonker.bioc.aecom.yu.edu/cgi-bin/inhouse/bruteptf/bruteptf.cgi). A total of 16 NCS-related heavy-atom sites were found and redefined using SOLOMON (2) in the SHARP environment, and the handedness of the cluster was confirmed. One of the modeled clusters, consisting of seven α-helices, combined with the heavy-atom phase information and additional model building, and improved model boundaries between NCS and symmetry-related molecules. The Hendrickson–Lattman coefficients from DM averaging were included in SHARP to improve the heavy-atom refinement, and a total of 15 Ta clusters and 44 Ir sites were found, refined, and used for calculating improved experimental phases. Solvent flattening/fitting was done using COOT (4) in the SHARP density modification interface, with the preliminary model used to define the initial solvent envelope. The resolution was extended to 3.58 Å in the solvent flattening step, greatly increasing the map quality and revealing side-chain density. Further electron density map improvements were obtained by averaging the four domains of gH/gL independently in the 16 NCS copies. The resulting electron density map showed clear β-sheet structure between the helical bundles and clear connectivity of the majority of the main chain. Other datasets with lower phasing power, KReO4, YbCl3, (NH4)2WS4, AuCN2, and SeMet data, were combined with the Native, TaBr12, and KtIrCl6 data by CAD and scaled by SCALEIT (Table S1). The combined datasets and phases improved by averaging were entered into SHARP for further heavy-atom search and refinement followed by solvent flattening/fitting using the initial gH/gL model with 643 residues. A total of 40 TaBr12 clusters, 34 KtIrCl6, 38 KReO4, 7 YbCl3, 14 (NH4)2WS4, 24 AuCN2, and 27 SeMet sites were used to calculate the phase, and the map from solvent flattening/fitting showed density with better connectivity.

The partial model of the gH/gL was improved by automated building using Buccaneer (5) and extended manually using COOT (4). A poly-alanine chain trace was used in an initial round of refinement using Crystallography and NMR System (CNS) (6). After rigid body, B-domain, minimization, and B-group refinement in CNS, the R-free was 0.3957 for strict NCS and 0.3824 for restrained NCS definitions. A 2fo-fc map was calculated from the restrained NCS CNS refinement, and the electron density map showed loops that were not visible before refinement. B-factor sharpening, with an empirically determined optimal value of B = −120, was used to improve side-chain density, allowing for the modeling of individual residues. The B-sharpened map showed density for most of the gH/gL side chains and clear disulfide bond connections. A low-resolution Se-Met dataset was also used to further ensure reliability of the model building. The final model quality was checked with Molprobity (7) and Coot (4), and the statistics are presented in Table S1.

EM. Carbon-coated grids were first glow-discharged, and soon after, the protein sample was mounted and followed with 2% Uranyl Acetate for staining. A JOEL 1230 transmission electron microscope at Northwestern University Biological Imaging Facility was used at 100 kV at room temperature as described (8). The negative stained image was scanned and digitized with a Microtek Scanner using ScanWizard Pro.

Acknowledgments. Portions of this research were conducted at the Northeastern Collaborative Access Team beam lines of the Advanced Photon Source supported by Award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source is supported by the US Department of Energy, Office of Basic Energy Sciences, under contract W-31-109-ENG-38. Research was also carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The Stanford Synchrotron Radiation Laboratory (SSRL) Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and National Institute of General Medical Sciences. Additional research was conducted at beamline 8.3.1 at the Advanced Light Source, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under Contract DE-AC02-05CH11231.

Fig. S1. Electron density maps for the EBV gH/gL structure. Initial phases were determined from anomalous signals obtained from crystals soaked in Ta$_6$Br$_{12}$ and further improved with other heavy-atom derivatives such as K$_2$IrCl$_6$. (A) α-helices were fitted manually in the initial low-resolution (≈6 Å) electron density map. Multiple helical clusters were found, allowing for further phase improvement by averaging; 16 copies of the gH/gL complex were found in the asymmetric unit. (B) Phases were extended to 3.58 Å resolution, and a full model was built using automated and manual approaches, including information from B factor–sharpened electron density maps.
EM of EBV gH/gL. EBV gH/gL molecules were selected from negatively stained electron microscope images using the program BOXER in the EMAN image processing package (9).

Fig. S2. EM of EBV gH/gL. EBV gH/gL molecules were selected from negatively stained electron microscope images using the program BOXER in the EMAN image processing package (9).
Fig. S3. EBV gH and gL domain structures. Each domain is shown as a ribbon diagram with secondary structure elements and residue numbers labeled. A numerical prefix on the secondary structure element indicates the associated domain. gH and gL residue numbers are indicated with a preceding H or L, respectively. Disulfide bonds are indicated in yellow sticks, and residue numbers are indicated. (A) Ribbon diagram of domain I (D-I). gH is shown in blue; gL is shown colored from N to C as a gradient from yellow to red. The view is an end-on view, ∼90° from the view shown in Fig. 1B. Helical bundle I is indicated. (B) Ribbon diagram of D-II colored as a gradient from N (magenta) to C (yellow). The view is to D-II from D-I. Helical bundle II is located behind β-sheet II. (C) Ribbon diagram of D-III colored as a gradient from N (green) to C (yellow) along the peptide chain. D-II lies to the right and D-IV lies to the left in this orientation. (D) D-IV is colored in a gradient from N (yellow) to C (cyan). The view is an end-on view rotated ∼90° to the view shown in Fig. 1B.
Fig. S4. Structure-based multiple sequence alignment of gH and gL proteins from related γ-herpesviruses. The amino acid sequences of gH (A) and gL (B) from Rhesus strain LCL8664 (AY037858), Marmoset strain CJ0149 (AF319782), and Kaposi’s Sarcoma herpesvirus (HHV8) strain GK18 (AF148805) were aligned with EBV gH and gL (YP_401700.1). Secondary structure elements of EBV gH and gL are labeled corresponding to Fig. 2 and colored by domain as in Fig. 1. N-glycosylation sites are labeled with purple triangles, and cysteine residues that are forming disulfide bonds are labeled with yellow stars according to the EBV gH and gL structure. Signal sequence and transmembrane region (TM) are indicated with lines. The alignment was generated using ESPRESSO and edited using ALINE. Residues are colored from cyan to red according to bulk similarity properties, with red indicating the most similar residues.

Matsuura et al. www.pnas.org/cgi/content/short/1011806108

Fig. 54. Structure-based multiple sequence alignment of gH and gL proteins from related γ-herpesviruses. The amino acid sequences of gH (A) and gL (B) from Rhesus strain LCL8664 (AY037858), Marmoset strain CJ0149 (AF319782), and Kaposi’s Sarcoma herpesvirus (HHV8) strain GK18 (AF148805) were aligned with EBV gH and gL (YP_401700.1). Secondary structure elements of EBV gH and gL are labeled corresponding to Fig. 2 and colored by domain as in Fig. 1. N-glycosylation sites are labeled with purple triangles, and cysteine residues that are forming disulfide bonds are labeled with yellow stars according to the EBV gH and gL structure. Signal sequence and transmembrane region (TM) are indicated with lines. The alignment was generated using ESPRESSO and edited using ALINE. Residues are colored from cyan to red according to bulk similarity properties, with red indicating the most similar residues.
### Table S1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Source</th>
<th>Wavelength (Å)</th>
<th>Space group</th>
<th>Resolution (Å; last shell)*</th>
<th>Mosaicity (°)</th>
<th>Measured reflections</th>
<th>Unique reflections</th>
<th>R_{sym} (%)</th>
<th>I/σ</th>
<th>Completeness (%)</th>
<th>Redundancy</th>
<th>Phasing from MIRAS† data</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALS 8.3.1</td>
<td>1.00753</td>
<td>P21</td>
<td>28.87–3.58</td>
<td>0.61</td>
<td>883,556</td>
<td>222,472</td>
<td>11.4</td>
<td>6.3</td>
<td>99.6</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>APS 24ID-C</td>
<td>1.2543</td>
<td>P21</td>
<td>50–4.6</td>
<td>0.63</td>
<td>1,485,087</td>
<td>116,119</td>
<td>14.8</td>
<td>16.98</td>
<td>99.5</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>SSRL 9–2</td>
<td>1.10519</td>
<td>P21</td>
<td>50–5.5</td>
<td>0.88</td>
<td>243,820</td>
<td>63,276</td>
<td>11.1</td>
<td>15.588</td>
<td>95.4</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>SSRL 9–2</td>
<td>1.17609</td>
<td>P21</td>
<td>50–5.4</td>
<td>1.4</td>
<td>883,745</td>
<td>71,118</td>
<td>15.1</td>
<td>17.471</td>
<td>95.5</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>SSRL 11–1</td>
<td>1.16962</td>
<td>P21</td>
<td>50–7.75</td>
<td>1.6</td>
<td>171,548</td>
<td>47,880</td>
<td>8.9</td>
<td>17.97</td>
<td>98.1</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>SSRL 9–2</td>
<td>1.21455</td>
<td>P21</td>
<td>50–7.25</td>
<td>0.30</td>
<td>168,746</td>
<td>29,887</td>
<td>17.8</td>
<td>9.264</td>
<td>99.6</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>SSRL 11–1</td>
<td>1.03635</td>
<td>P21</td>
<td>50–7.0</td>
<td>0.96</td>
<td>517,211</td>
<td>137,437</td>
<td>15.5</td>
<td>19.250</td>
<td>97.7</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>APS DND</td>
<td>0.97829</td>
<td>P21</td>
<td>50–5.4</td>
<td>0.98</td>
<td>410,142</td>
<td>34,187</td>
<td>14.8</td>
<td>18.2</td>
<td>99.4</td>
<td>6.2</td>
<td></td>
</tr>
</tbody>
</table>

**Data collection**

- **Source**: ALS 8.3.1, APS 24ID-C, SSRL 9–2, SSRL 9–2, SSRL 11–1, SSRL 9–2, SSRL 11–1, APS DND
- **Wavelength (Å)**: 1.00753, 1.2543, 1.10519, 1.17609, 1.16962, 1.21455, 1.03635, 0.97829
- **Space group**: P21, P21, P21, P21, P21, P21, P21, P21
- **Unit cell dimension**
  - **a (Å)**: 151.5, 151.1, 151.6, 150.9, 150.5, 150.3, 151.3, 151.3
  - **b (Å)**: 244.9, 244.5, 242.6, 242.6, 244, 247.2, 244.6, 246.3
  - **c (Å)**: 288.0, 288.2, 286.0, 287.4, 286.6, 289.0, 287.3, 288.7
- **Resolution (Å; last shell)**: 28.87–3.58, 50–4.6, 50–5.5, 50–5.4, 50–7.75, 50–7.25, 50–7.0
- **Mosaicity (°)**: 0.61, 0.63, 0.88, 1.4, 1.6, 0.30, 0.96, 0.98
- **Measured reflections**: 883,556, 1,485,087, 243,820, 883,745, 171,548, 168,746, 517,211, 410,142
- **Unique reflections**: 222,472, 116,119, 63,276, 71,118, 47,880, 29,887, 137,437, 34,187
- **R_{sym} (%)**: 11.4, 14.8, 11.1, 15, 8.9, 17.8, 15.5, 14.8
- **I/σ**: 6.3, 16.98, 15.588, 17.471, 17.97, 9.264, 19.250, 18.2
- **Completeness (%)**: 99.6, 99.5, 95.4, 95.5, 98.1, 99.6, 97.7, 99.4
- **Redundancy**: 4.0

**Phasing from MIRAS† data**

- **No. of heavy-atom sites**: 40 Ta_{6}Br_{12} clusters
- **Phasing power**
  - **isomorphous** (acentric/centric): 1.216/1.236, 0.398/0.322, 0.154/0.128, 0.187/0.143, 0.219/0.155, 0.65/0.503, 0.165/0.128
  - **anomalous** (acentric): 0.502, 0.550, 0.215, 0.107, 0.188, 0.197, 0.162

**Refinement**

- **Resolution (Å)**: 29–3.58
- **Reflections (free set)**: 245,098 (2,461)
- **R_{work}/R_{free} (%)**: 28.4/31.3
- **Amino acid residues**: 12,112
- **Protein atoms**: 94,656
- **Root mean square deviations**
  - **Bond length (Å)**: 0.003
  - **Bond angle (°)**: 0.88
  - **Mean B value (Å^2)**: 121.5
- **Ramachandran plot**
  - **Favored (%)**: 85.7
  - **Allowed (%)**: 11.8
  - **Outlier (%)**: 2.5

*Calculated with Molprobity for one representative gHgL molecule.
†Multiple isomorphous replacement with anomalous scattering.