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 Ta_6Br_{12} dataset using SOLVE (1). The Native, Ta_6Br_{12}, and K_2IrCl_6 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 glycoprotein H (gH)/gL molecules per asymmetric unit was identified, consistent with the expectations from the unit cell volume. Further averaging was carried out using the program DM, resulting in an improved electron density map, clearer connectivity between α-helices, additional model building, and improved molecular 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/flipping was done using SOLOMON (2) 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, $KReO_4$, $YbCl_3$, $(NH_4)_2WS_4$, $AuCN_2$, and SeMet data, were combined with the Native, Ta_6Br_{12}, and K_2IrCl_6 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/flipping using the initial gH/gL model with 643 residues. A total of 40 Ta_6Br_{12} clusters, 34 K_2IrCl_6, 38 $KReO_4$, 7 $YbCl_3$, 14 $(NH_4)_2WS_4$, 24 $AuCN_2$, and 27 SeMet sites were used to calculate the phase, and the map from solvent flattening/flipping 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 0.3937 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.

1. Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55:849–861.
2. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
3. de La Fortelle E, Bricogne G (1997) Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods in Enzymology: Macromolecular Crystallography*, eds Sweet RM, Carter CW (Academic, New York), Vol 276, pp 472–494.
4. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
5. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62:1002–1011.
6. Brünger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
7. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12–21.
8. Wrigley NG, et al. (1983) Electron microscopy of influenza haemagglutinin-monoclonal antibody complexes. *Virology* 131:308–314.
9. Tang G, et al. (2007) EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* 157:38–46.

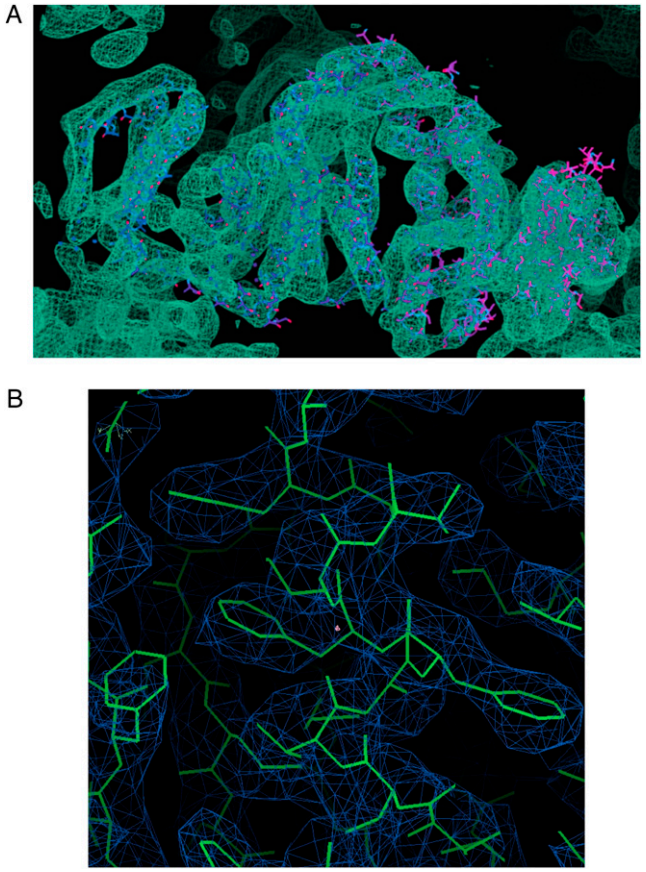

Fig. S1. Electron density maps for the EBV gH/gL structure. Initial phases were determined from anomalous signals obtained from crystals soaked in Ta_6Br_{12} and further improved with other heavy-atom derivatives such as K_2IrCl_6. (*A*) α-helices were fitted manually in the initial low-resolution (~6 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.

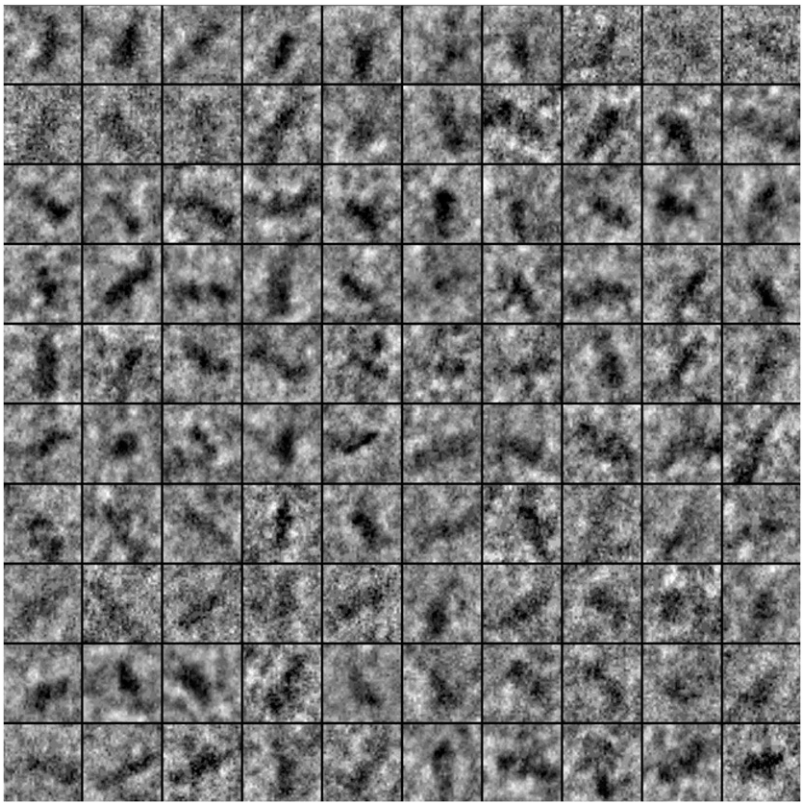

Fig. S2. EM of EBV gH/gL. EBV gH/gL molecules were selected from negatively stained electron microscope images using the program BOXER in the EMAN image processing package (9).

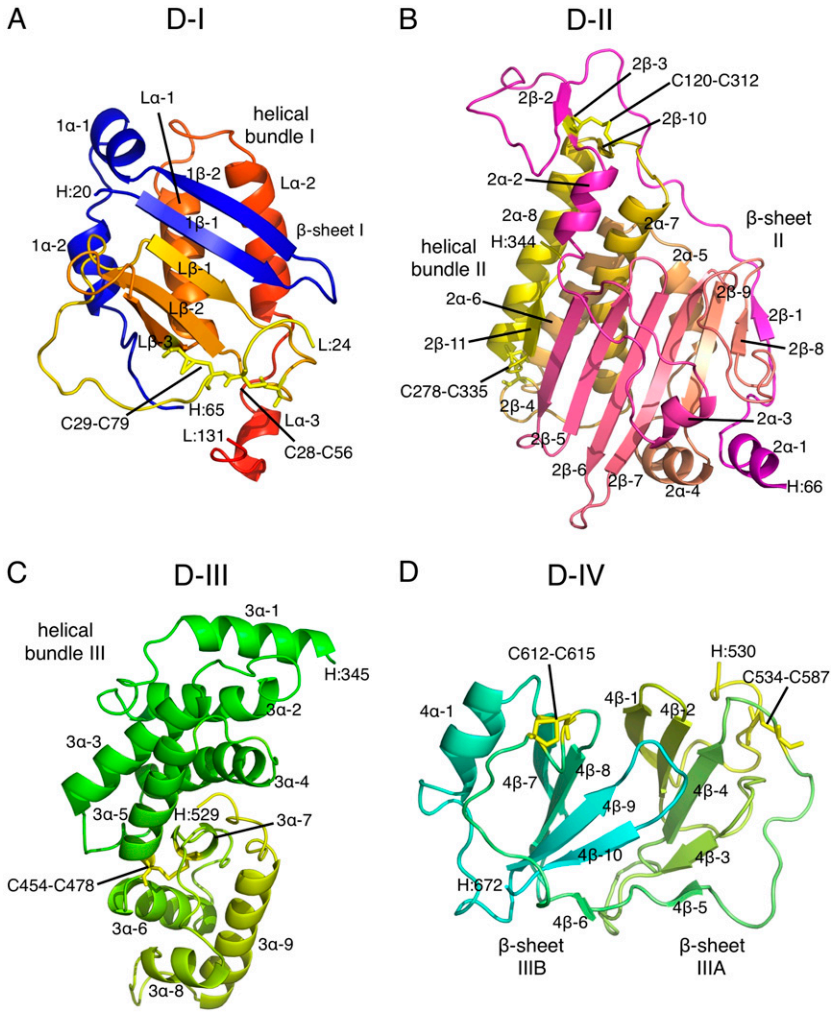

Fig. S3. EBV gH and gL domain structures. Each domain is shown as a ribbon diagram with secondary structure elements and residue numbers labeled. A numerical prefix on the secondary structure element indicates the associated domain. gH and gL residue numbers are indicated with a preceding H or L, respectively. 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. 1*B*. 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. 1*B*.

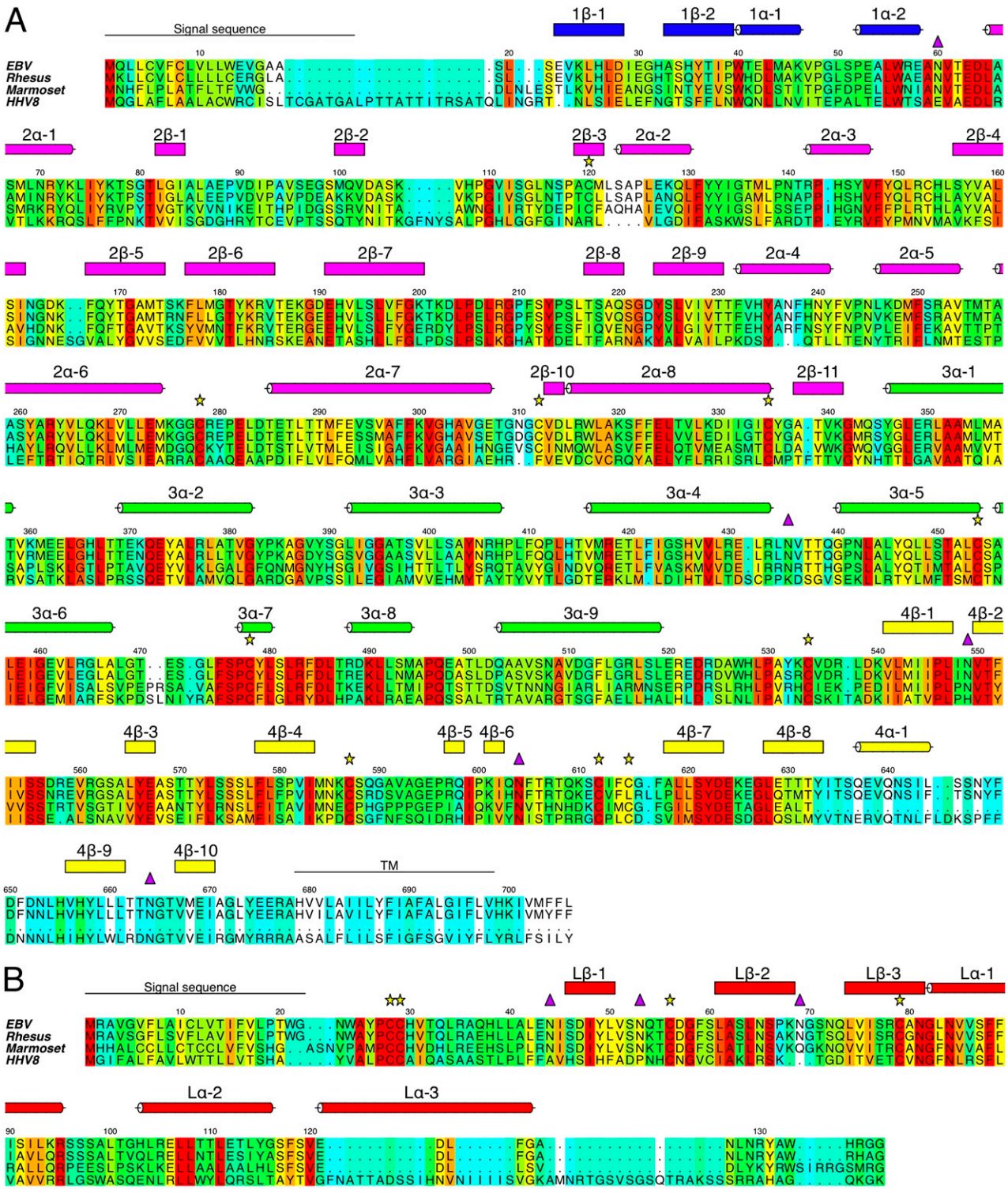

Fig. S4. Structure-based multiple sequence alignment of gH and gL proteins from related γ-herpesviruses. The amino acid sequences of gH (*A*) and gL (*B*) from Rhesus strain LCL8664 (AY037858), Marmoset strain CJ0149 (AF319782), and Kaposi's Sarcoma herpesvirus (HHV8) strain GK18 (AF148805) were aligned 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.

Table S1. Data collection and refinement statistics

	Native	Ta_6Br_{12}	K_2IrCl_6	$KReO_4$	$YbCl_3$	$(NH_4)_2WS_4$	$AuCN_2$	SeMet
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	50–5.4	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	4.5	3.9	13	3.6	5.7	3.8	6.2
Phasing from MIRAS† data								
No. of heavy-atom sites		40 Ta_6Br_{12}clusters	34	38	7	14	24	27
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	0.659/0.503	0.165/0.128
Phasing $power_{anomalous}$ (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 gH/gL molecule.
†Multiple isomorphous replacement with anomalous scattering.