Supporting Online Material for

Crystal Structure of Glycoprotein B from Herpes Simplex Virus 1

Ekaterina E. Heldwein,* Huan Lou, Florent C. Bender, Gary H. Cohen, Roselyn J. Eisenberg, Stephen C. Harrison

*To whom correspondence should be addressed. E-mail: heldwein@crystal.harvard.edu

DOI: 10.1126/science.1126548

This PDF file includes:

- Materials and Methods
- Figs. S1 to S3
- Tables S1 and S2
- References
Supporting Online Material

Materials and Methods

Expression and purification of recombinant gB730
DNA encoding residues Ala\(^{31}\) to Ala\(^{730}\) of HSV-1 strain KOS gB was inserted into the pVT-Bac vector in frame with the mellitin signal sequence coding region (SI). We refer to the resulting construct as "gB730". SF9 cells were grown to a density of 4x10\(^6\) cells per mL in a Celligen bioreactor and infected with the recombinant virus at an MOI of 4. Supernatant was separated from cells when the cell viability dropped to 70 percent (typically, 72 hours post-infection). gB730 was purified from the culture supernatant by immunoaffinity chromatography using the gB-specific monoclonal antibody, DL16 (SI), a non-neutralizing antibody that recognizes a discontinuous, trimer-dependent epitope. After extensive washing with 10 mM Tris·HCl, pH 7.2 and 500 mM NaCl, bound gB730 was eluted from the column with 3M KSCN. The recovered protein was then dialyzed against PBS and concentrated prior to storage at -80°C. For crystallization and biochemical experiments, gB samples were thawed and further purified by gel-filtration chromatography in 10mM Tris·HCl, pH 7.6, 100 mM NaCl, an 1mM EDTA (Superdex S200, Amersham Pharmacia) and concentrated using Millipore Ultra-4 (MW cutoff 50 kDa).

Crystallization and data collection
To obtain well-ordered crystals, we incubated gB730 with trypsin in 100:1 w/w ratio for up to 4 hours at room temperature. The reaction was stopped by passing the digest mix over a benzamidine sepharose column. The flow-through and the wash were concentrated and passed over a gel-filtration column, Superdex S200, equilibrated in 10mM Tris·HCl, pH 7.6 and 100 mM NaCl. The eluted fractions containing the trypsin-trimmed core of gB, residues Asp\(^{103}\) to Ala\(^{730}\), were concentrated to 4.5-6 mg/ml using Millipore Ultra-4 (MW cutoff 50 kDa). Trypsin cleaves gB between residues Arg\(^{102}\)/Asp\(^{103}\), Arg\(^{335}\)/Ala\(^{336}\), and Arg\(^{474}\)/Lys\(^{475}\), as determined by N-terminal sequencing. Cleavage effectively removed residues Ala\(^{31}\) to Arg\(^{102}\), yielding a stable homogenous product. The tryptic core of gB forms a trimer. Crystals of trypsin-trimmed gB730 were grown by vapor diffusion at room temperature in hanging drops with 2 \(\mu\)l protein sample and 2 \(\mu\)l reservoir solution (8% PEG 6000, 0.1 M Na·Hepes, pH 7.0, and 15% 2-methyl-2,4-pentanediol). Crystals in space group P1 with a=83 Å, b=99 Å, c=100 Å, α=67°, β=71°, γ=78° appeared after several days and grew to their final size over 3 weeks. For data collection, crystals were quickly transferred to 8% PEG 6000, 0.1 M Na·Hepes, pH 7.0, 0.1 M NaCl, and 18% 2-methyl-2,4-pentanediol and plunged into liquid N\(_2\). A native data set was collected at beamline 24ID (APS) and processed to 2.1 Å resolution using DENZO and SCALEPACK (S2) (Table S1). The self-rotation function as implemented in GLRF (S3) unequivocally showed a non-crystallographic (molecular) three-fold symmetry axis with three gB molecules (a gB trimer) in the asymmetric unit.

To obtain experimental phases, we expressed and purified a selenomethionine-substituted version, SeMet gB730, using methionine-deficient insect cell media supplemented with L-selenomethionine. High SeMet incorporation, 80-90%, was confirmed by mass-spectrometry. After limited proteolysis using trypsin, SeMet gB730 was crystallized
under conditions similar to those for the native trypsin-cleaved gB730, (8% PEG 6000, 0.1 M Na·Hepes, pH 7.5, and 13% 2-methyl-2,4-pentanediol). A single-wavelength anomalous dispersion (SAD) dataset was collected at a peak wavelength at beamline 24ID (APS) and processed to 2.6 Å resolution using DENZO and SCALEPACK (S2) (Table S1).

**Structure determination**

Eight out of 36 predicted Se sites were found using direct methods as implemented in Shake-n-Bake (S4) and refined using SHARP (S5). The residual maps revealed the locations of 12 additional sites. The 3-fold noncrystallographic symmetry (NCS) operators were determined using the refined positions of 20 Se sites. After phase improvement by density modification using DM (S6), which included 3-fold NCS averaging, solvent-flattening, and histogram-matching, the experimental electron-density map allowed tracing and sequence assignment for 75% of the polypeptide chain. At this point, model phases were used to locate 9 additional Se sites, which were used to confirm the correctness of the sequence assignment. Data from 41.2 to 2.6 Å resolution were used to refine a thermal parameter for each domain, followed by gradient minimization, torsion-angle simulated annealing, further gradient minimization, and group thermal parameter refinement, all as implemented in CNS (S7). Subsequent simulated-annealing refinement to 2.6 Å in CNS allowed the building of the rest of the polypeptide chain. Model rebuilding was done with O (S8). Prior to refinement, 5% of the data, in thin resolution shells, were set aside for cross-validation using DATAMAN (S9). Test set flags were transferred to the native dataset; additionally, 5% of the native data between 2.6 and 2.1 Å, also in thin resolution shells, were selected. The model was then rigid-body refined against the native data to 2.1 Å resolution using CNS. Initially, the model was refined with strict NCS constraints and later with tight NCS restraints. After several rounds of simulated annealing refinement, group B-factor refinement, rebuilding, and addition of solvent molecules, the NCS restraints were released for the final round of gradient minimization and individual B-factor refinement. The current model has an R of 24.8% and an R\textsubscript{free} of 27.5% (Table S1). It contains three copies of the trypsin-cleaved gB, residues Asp\textsuperscript{103} to Ala\textsuperscript{730}, and 227 solvent molecules and ions. The final model is missing residues Asp\textsuperscript{103} to Asp\textsuperscript{110}, Asp\textsuperscript{329} to Thr\textsuperscript{337}, Leu\textsuperscript{460} to Ser\textsuperscript{491}, and Asp\textsuperscript{726} to Ala\textsuperscript{730} of protomer A; residues Asp\textsuperscript{103} to Asn\textsuperscript{108}, Thr\textsuperscript{331} to Ala\textsuperscript{336}, Glu\textsuperscript{462} to Ser\textsuperscript{491}, and Ala\textsuperscript{725} to Ala\textsuperscript{730} of protomer B; and residues Asp\textsuperscript{103} to Asn\textsuperscript{108}, Arg\textsuperscript{328} to Ala\textsuperscript{336}, Leu\textsuperscript{460} to Ser\textsuperscript{491}, and His\textsuperscript{724} to Ala\textsuperscript{730} of protomer C. According to PROCHECK, 89.1% of the residues lie in the most favored, 10.6% in the additionally allowed, and 0.3% in the generously allowed region of the Ramachandran plot (S10).

**Crystallization and structure determination of the uncleaved gB ectodomain**

Crystals of uncleaved gB730 were grown by vapor diffusion at room temperature in hanging drops with 2 µl protein sample and 2 µl reservoir solution (10% PEG 4000 and 0.1 M Na·citrate, pH 5.5). Crystals in space group P321 with a=b=118.2 Å, c=316.4 Å, \( \alpha=\beta=90^\circ, \gamma=120^\circ \) appeared after several days and grew to their final size over 2 weeks. To obtain experimental phases, crystals were soaked in 5 mM solution of K\textsubscript{2}Pt(SCN)\textsubscript{6} for 4 days. For data collection, crystals were transferred stepwise to 10% PEG 4000, 0.1 M Na·citrate, pH 5.5, 0.1 M NaCl, and 15% mesoerythritol and plunged into liquid N\textsubscript{2}. A
two-wavelength multiple anomalous dispersion (MAD) data set was collected at the peak and the high-energy remote wavelengths at beamline 8.2.1 (ALS) and processed to 3.53 Å and 4.10 Å resolution, respectively, using DENZO and SCALEPACK (S2) (Table S1). Inspection of the native Patterson maps, calculated in CCP4 (S11), revealed a translational non-crystallographic symmetry with two gB protomers in the asymmetric unit arranged top-to-bottom along the crystal c axis. Molecular replacement using AMORE (S12) and the refined model of trypsin-cut gB ectodomain located both protomers. Ten Pt sites in the peak data set were found using anomalous differences and phases of the molecular replacement solution. The sites were refined against the 2-wavelength MAD data set using SHARP (S5). After phase improvement by solvent flipping using Solomon (S13), the experimental electron-density map showed that the structure of the uncleaved ectodomain is essentially identical to that of the trypsin-cut ectodomain. In addition, it allowed tracing of residues Leu$^{460}$ to Pro$^{484}$, unresolved in the maps of the trypsin-cut gB ectodomain. The molecular replacement solution was used as a starting model. Data from 62.6 to 3.53 Å resolution were used to refine a thermal parameter for each domain, followed by rigid-body refinement, gradient minimization, torsion-angle simulated annealing, further gradient minimization, and group thermal parameter refinement, all as implemented in CNS (S7). After a single round of refinement, current model has an R of 30% and an R$_{free}$ of 37%.

**Analytical Ultracentrifugation**

The mass of the gB ectodomain, residues Ala$^{31}$ to Ala$^{730}$, in solution was determined using sedimentation equilibrium analytical centrifugation (Optima XL-A analytical ultracentrifuge, Beckman-Coulter, with an AN-60 Ti rotor and a six-channel cell). Absorbance at 280 nm was used to monitor concentration gradients. Samples were centrifuged at 3 different speeds, 7000, 10000, and 13000 rpm, and 3 different concentrations, OD$_{280} = 0.3$, 0.6, and 0.9. gB samples subjected to the sedimentation analysis contained protein in 0.1 M Na·citrate, pH 5.5, 0.1 M NaCl, and 1mM EDTA. Protein partial specific volume was calculated from the amino acid composition. Datasets at 3 different rotor speeds and 3 different sample concentrations were analyzed using global parameters for the molecular weight and local parameters for the baseline offset. The curves of absorbance versus radius fit a distribution for molecular mass of ~240 kDa, corresponding to a trimer.

**Characterization of monoclonal antibodies to gB**

Anti-HSV mAbs SS106 and SS144 to gB were generated as described (S14). Specificity for HSV type 1 or 2 was measured by Western blot on extracts of cells infected by HSV-1 (strain KOS) or HSV-2 (strain 333). Neutralization was measured as 50% plaque reduction on Vero cells (S14). To identify specific epitopes, we used a panel of peptides derived from gB and screened mAbs using ELISA as described (S15). Peptides, numbered 1 to 82, were 20-amino-acid long and spanned the entire extra-cellular region of gB, residues Ala$^{31}$ to Asn$^{773}$, and overlapped one another by 11 amino acids. Each peptide had biotin coupled to its N-terminus so that it could be immobilized on ELISA plates coated with streptavidin. Binding of mAb to gB was measured using ELISA (S14) and optical biosensor (BIAcore-X) (S15). For biosensor experiments, gB with an engineered His$_6$ tag following position Ala$^{725}$ was first captured with an anti-His$_6$ mAb.
immobilized on a BIAcore chip (Qiagen Inc.) and then reacted with different antibodies to gB. This assay measured mAb binding to the surface of the native molecule. MAb blocking studies were also performed using optical biosensor as described (S16).
Table S1. Crystallographic data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>Native</th>
<th>SeMet</th>
<th>K₂Pt(SCN)₆ peak</th>
<th>K₂Pt(SCN)₆ remote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.97944</td>
<td>0.97932</td>
<td>1.07206</td>
<td>0.87000</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>41.20-2.10 (2.18-2.10)</td>
<td>48.6-2.60 (2.69-2.60)</td>
<td>62.60-3.53 (3.68-3.53)</td>
<td>72.55-4.10 (4.25-4.10)</td>
</tr>
<tr>
<td>Total observations</td>
<td>2,809,776</td>
<td>706,214</td>
<td>1,027,153</td>
<td>743,254</td>
</tr>
<tr>
<td>Unique observations</td>
<td>145,627</td>
<td>68,278</td>
<td>32,336</td>
<td>21,279</td>
</tr>
<tr>
<td>R_merge* (%)</td>
<td>6.6 (33.1)</td>
<td>10.0 (37.3)</td>
<td>18.1 (56.2)</td>
<td>27.4 (78.2)</td>
</tr>
<tr>
<td>I/σI</td>
<td>18.1 (3.7)</td>
<td>11.6 (2.2)</td>
<td>10.2 (3.4)</td>
<td>6.3 (2.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>87.8 (58.8)</td>
<td>77.0 (28.0)</td>
<td>99.8 (99.7)</td>
<td>99.8 (99.8)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.7 (3.7)</td>
<td>3.8 (2.6)</td>
<td>7.7 (7.3)</td>
<td>5.6 (5.7)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>41.2-2.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflections (free set)</td>
<td>138,285 (7,446)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_crystal/R_free † (%)</td>
<td>24.8/27.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino acid residues</td>
<td>1728</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>13965</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solvent atoms</td>
<td>227</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-factors (Å²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>57.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solvent</td>
<td>40.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R.m.s.d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.0068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.264</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot (%) of non-glycine and non-proline residues‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favored regions</td>
<td>89.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additionally allowed regions</td>
<td>10.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generously allowed regions</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*R_merge = ΣΣI_{hkl} - <I_{hkl}> /ΣI_{hkl}, where I_{hkl} is observed intensity and <I_{hkl}> is the final average value of intensity.

†R_{cryst} and R_{free} = ΣF_{obs} - F_{calc} /ΣF_{obs} for the reflections in the working and test sets, respectively.

‡As determined by PROCHECK (S10).
Table S2. Epitope mapping of monoclonal antibodies to the ectodomain of HSV-1 gB.

<table>
<thead>
<tr>
<th>mAb</th>
<th>Neutr.</th>
<th>Epitope peptide sequence</th>
<th>gB binding ELISA</th>
<th>Competition</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS106</td>
<td>+</td>
<td>75, 76 697-725</td>
<td>+</td>
<td>SS144</td>
<td>1 and 2‡</td>
</tr>
<tr>
<td>SS144</td>
<td>+</td>
<td>75, 76 697-725</td>
<td>+</td>
<td>SS106</td>
<td>1 and 2‡</td>
</tr>
<tr>
<td>H1838</td>
<td>+*</td>
<td>41 391-410§</td>
<td>ND†</td>
<td>ND†</td>
<td>1 and 2‡</td>
</tr>
<tr>
<td>H1781</td>
<td>+*</td>
<td>48 454-473§</td>
<td>ND†</td>
<td>no</td>
<td>1 and 2‡</td>
</tr>
</tbody>
</table>

*Previously reported (SI7).
†Not determined.
‡1 refers to HSV-1 (strain KOS); 2, to HSV-2 (strain 333).
§The mapping of H1838 agrees with that reported previously (SI7). However, the mapping of H1781 does not. H1781 was previously mapped to residues 470-487 (SI7). Our results are reproducible, and we are confident in our mapping. Possibly, differences in the techniques used for epitope mapping can account for this discrepancy. Nonetheless, the two mapped regions are adjacent and exposed.
Fig. S1. Secondary structure of the gB ectodomain showing the organization of the polypeptide chain. The coloring scheme is the same as in Fig. 1. β-strands are shown as arrows and α-helices as cylinders. β-strands are numbered numerically, and α-helices alphabetically. Locations of cysteines are indicated with blue dots and residue numbers; disulfide bridges are shown as blue lines.
Fig. S2. Alignment of secondary structure with the sequences of gB proteins from the most prominent representatives of three major herpesvirus subfamilies. α-herpesviruses: HSV1, strain KOS; HSV2, strain HG52; PRV, strain Indiana-Funkhauser; and VZV. β-herpesviruses: HCMV, strain AD169. γ-herpesviruses: KSHV; and EBV, strain LCL8664. Residue numbering for HSV1 is shown above its sequence. The coloring of the secondary structure elements matches that in Fig. 1. Signal sequences, absent from mature glycoproteins, are shown in red letters. Conserved cysteines are labeled sequentially underneath the alignment. Identical residues are highlighted in yellow, homologous residues, in grey, cysteines, in blue. Furin cleavage sites in PRV, VZV,
HCMV, EBV, and KSHV are highlighted in purple. This figure was generated using ALSCRIPT (S18).
Domains of HSV-1 gB (A) and VSV G (B) in a side-by-side representation. The domains of gB are shown in the same orientation as in Fig. 1. Domains of VSV G were aligned to homologous domains of gB using DALI (S19) or LSQMAN (S20). Residues at the boundaries of domains are labeled. Residues 661-723 of gB belong to a different protomer than the rest of the shown residues.
References


